# APPLICATION OF BIG DATA, DEEP LEARNING, MACHINE LEARNING, AND OTHER ADVANCED ANALYTICAL TECHNIQUES IN ENVIRONMENTAL ECONOMICS AND POLICY

EDITED BY: Tsun Se Cheong, Xunpeng (Roc) Shi, Yanfei Li and Yongping Sun

PUBLISHED IN: Frontiers in Environmental Science and Frontiers in Energy Research

frontiers Research Topics

# frontiers

**Frontiers eBook Copyright Statement**

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88976-596-6
DOI 10.3389/978-2-88976-596-6

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

# APPLICATION OF BIG DATA, DEEP LEARNING, MACHINE LEARNING, AND OTHER ADVANCED ANALYTICAL TECHNIQUES IN ENVIRONMENTAL ECONOMICS AND POLICY

Topic Editors:
**Tsun Se Cheong,** Hang Seng University of Hong Kong, China
**Xunpeng (Roc) Shi,** University of Technology Sydney, Australia
**Yanfei Li,** Hunan University of Technology and Business, China
**Yongping Sun,** Huazhong University of Science and Technology, China

**Citation:** Cheong, T. S., Shi, X., Li, Y., Sun, Y., eds. (2022). Application of Big Data, Deep Learning, Machine Learning, and Other Advanced Analytical Techniques in Environmental Economics and Policy. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-596-6

# Table of Contents


06 ***Editorial: Application of Big Data, Deep Learning, Machine Learning, and Other Advanced Analytical Techniques in Environmental Economics and Policy***
Tsun Se Cheong, Xunpeng (Roc) Shi, Yanfei Li and Yongping Sun

11 ***Analysis of Information Feedback on Residential Energy Conservation and the Implications: The Case of China***
Zhuangai Li and Xia Cao

24 ***Regional Differences of Manufacturing Green Development Efficiency Considering Undesirable Outputs in the Yangtze River Economic Belt Based on Super-SBM and WSR System Methodology***
Gen Li, Ying Zhou, Fan Liu and Tao Wang

39 ***Analyzing Information Disclosure in the Chinese Electricity Market***
Luosong Jin, Weidong Liu, Xiangyang Wang, Jing Yu and Panting Zhao

47 ***Dynamic Risk Assessment of the Overseas Oil and Gas Investment Environment in the Big Data Era***
Xuqiang Duan, Xu Zhao, Jianye Liu, Shuquan Zhang and Dongkun Luo

61 ***Tripartite Coordinated Regulation of New Energy Vehicles Based on Dynamic Bayesian Game***
Changyu Liu, Yadong Song, Le Chang and Guanglong Dong

71 ***Symmetric and Asymmetric Impact of Poverty, Income Inequality, and Population on Carbon Emission in Pakistan: New Evidence From ARDL and NARDL Co-Integration***
Salim Khan and Wang Yahong

84 ***A Method to Deal With Inter-regional and Inter-provincial Transaction Settlement Deviation Quantity Based on Kernel Density–Entropy Weight***
Dongfa Wang, Weibin Ding, Feng Gao, Yang Xu and Qin Miao

95 ***Does Anti-corruption Policy Influence Energy Efficiency in China?***
Jie Liu, Jianming Guo, Xiangfeng Liu, Xiao Bai and Farhad Taghizadeh-Hesary

110 ***How to Promote Compliance Management in the Electricity Market? An Analysis Based on the Evolutionary Game Model***
Luosong Jin, Cheng Chen, Yun Li, Xiangyang Wang and Yuanyuan Cheng

126 ***Evaluating the Impact of Urban Blue Space Accessibility on Housing Price: A Spatial Quantile Regression Approach Applied in Changsha, China***
Huang Tuofu, He Qingyun, Yang Dongxiao and Ouyang Xiao

141 ***Does the New Urbanization Influence Air Quality in China?***
Yan Wu, Zhixiang Cui and Cong Hu

155 ***The Impacts of Two Revisions of the China's Environmental Protection Fee on Firm Performance: Evidence From Stock Markets***
Huan Zheng and Yu He

173 ***How Does Target-Based Performance Evaluation Affect the Accuracy of Energy-Saving Data: Evidence From China***
Ping Zhang, Tiantian Yu, Shengfeng Lu and Ruting Huang

183 ***Analysis of the Upgrading Effect of the Industrial Structure of Environmental Regulation: Evidence From 113 Cities in China***
Wei Shao, Yufei Yin, Xiao Bai and Farhad Taghizadeh-Hesary

195 ***The Main Research Characteristics of the Development of the Concept of the Circular Economy Concept: A Global Analysis and the Future Agenda***
Víctor Meseguer-Sánchez, Francisco Jesús Gálvez-Sánchez, Valentin Molina-Moreno and Gonzalo Wandosell-Fernández-de-Bobadilla

211 ***The Spatial Network Contagion of Environmental Risks Among Countries Along the Belt and Road Initiative***
Ling Lu, Kai Fang, Chuan Ming Liu and Cheng Sun

225 ***The Major Driving Factors of Carbon Emissions in China and Their Relative Importance: An Application of the LASSO Model***
Wai Yan Shum, Ning Ma, Xiaomei Lin and Tingting Han

232 ***Time-Varying Impact of Economic Growth on Carbon Emission in BRICS Countries: New Evidence From Wavelet Analysis***
Lijin Xiang, Xiao Chen, Shuling Su and Zhichao Yin

251 ***Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption***
Yang Xu, Weijun Gao, Fanyue Qian and Yanxue Li

265 ***Driving Factors of $CO_2$ Emissions: Further Study Based on Machine Learning***
Shanshan Li, Yam Wing Siu and Guoqin Zhao

281 ***Effects of Age Dependency and Urbanization on Energy Demand in BRICS: Evidence From the Machine Learning Estimator***
Zhou Lu, Mantu Kumar Mahalik, Hemachandra Padhan, Monika Gupta and Giray Gozgor

292 ***Tracking the Domestic Carbon Emission Intensity of China's Construction Industry: A Global Value Chain Perspective***
Xiaoping Li, Yuan Yu, Xunpeng Shi and Xin Hu

307 ***Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the $CO_2$ Emissions Analysis Using Gaussian Process Regression***
Ning Ma, Wai Yan Shum, Tingting Han and Fujun Lai

315 ***The Role of Economic Policy Uncertainty in Renewable Energy-Growth Nexus: Evidence From the Rossi-Wang Causality Test***
Zhou Lu, Linchuang Zhu, Chi Keung Marco Lau, Aliyu Buhari Isah and Xiaoxian Zhu

321 ***Stability Analysis of the World Energy Trade Structure by Multiscale Embedding***
Leyang Xue, Feier Chen, Guiyuan Fu, Qiliang Xia and Luhui Du

332 ***Assessments of Air Pollution Control Effectiveness Based on a Sharp Regression Discontinuity Design —Evidence From China's Environmental Big Data***
Ren Wang, Jiaqi Huang, Lizhi Zhang, Yu Xia, Xu Xu and Tongli Nong

344 ***The Spatio-Temporal Evolution of China's Hydrogen Fuel Cell Vehicle Innovation Network: Evidence From Patent Citation at Provincial Level***
Pei Yu, Feng Jiang, Zhengfang Cai and Yongping Sun

361 ***Carbon Neutrality Policies and Technologies: A Scientometric Analysis of Social Science Disciplines***
Yuan Zhang, Chung-Lien Pan and Han-Teng Liao

385 ***A Counterfactual Baseline for Assessing Future Environmental Impact: A Case Study of the Belt and Road Initiative***
Jing Victor Li, Tsun Se Cheong and Xunpeng Shi

399 ***Exposing the Effects of Environmental Regulations on China's Green Total Factor Productivity: Results From Econometrics Analysis and Machine Learning Methods***
Junyi Feng, Jianjun Yan and Xia Tao

413 ***Assessing the Environmental Efficiency of Grain Production and Their Spatial Effects: Case Study of Major Grain Production Areas in China***
Hanxiao Xu, Bei Ma and Qiang Gao

428 ***Tracking Cloud Forests With Cloud Technology and Random Forests***
Pasky Pascual and Cam Pascual

445 ***Electricity-Water Consumption and Metropolitan Economic Growth: An Empirical Dual Sectors Dynamic Equilibrium Model***
Yiming He and Shaoui Gao

457 ***Financial Inclusion and Carbon Reduction: Evidence From Chinese Counties***
Zhenkai Yang, Lu Yu, Yinwei Liu, Zhichao Yin and Zumian Xiao

467 ***Does Green Credit Policy Move the Industrial Firms Toward a Greener Future? Evidence From a Quasi-Natural Experiment in China***
Zumian Xiao, Lu Yu, Yinwei Liu, Xiaoning Bu and Zhichao Yin

478 ***Dynamic Supervision and Control of VOCs Emission From China's Furniture Manufacturing Based on Big Data and IoT***
Zi-Gui Chen, Wei-Ping Wu, Jing Li and Yu-Heng Zeng

frontiers | Frontiers in Environmental Science

EDITORIAL
published: 24 June 2022
doi: 10.3389/fenvs.2022.953659

# Editorial: Application of Big Data, Deep Learning, Machine Learning, and Other Advanced Analytical Techniques in Environmental Economics and Policy

*Tsun Se Cheong*[1]*, *Xunpeng (Roc) Shi*[2], *Yanfei Li*[3] and *Yongping Sun*[4]

[1]Hang Seng University of Hong Kong, Hong Kong SAR, China, [2]University of Technology Sydney, Sydney, Australia, [3]Hunan University of Technology and Business, Changsha Shi, China, [4]Huazhong University of Science and Technology, Wuhan, China

**Keywords: big data, machine learning, deep learning, advanced analytical techniques, environmental economics, policy, environmental protection, energy**

## Editorial on the Research Topic

## Application of Big Data, Deep Learning, Machine Learning, and Other Advanced Analytical Techniques in Environmental Economics and Policy

OPEN ACCESS

***Edited by:***
*Cosimo Magazzino,*
*Roma Tre University, Italy*

***Reviewed by:***
*Mucahit Aydin,*
*Sakarya University, Turkey*

***Correspondence:**
*Tsun Se Cheong*
*jamescheong@hsu.edu.hk*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *26 May 2022*
***Accepted:*** *09 June 2022*
***Published:*** *24 June 2022*

***Citation:***
*Cheong TS, Shi X, Li Y and Sun Y (2022) Editorial: Application of Big Data, Deep Learning, Machine Learning, and Other Advanced Analytical Techniques in Environmental Economics and Policy. Front. Environ. Sci. 10:953659. doi: 10.3389/fenvs.2022.953659*

Environmental science has attracted the attention of more and more researchers around the globe, yet most of the analyses are based on traditional analytical techniques. It is noteworthy that although big data, deep learning, and other machine learning techniques have been applied in many different disciplines, including engineering, computer science, and medical science, these state-of-the-art analytical techniques have not been applied widely in the field of environmental science, nor the areas of environmental economics and management. There are only a few articles based on machine learning, for example, Magazzino et al. (2021a), Magazzino et al. (2021b), Magazzino et al. (2021c), and Magazzino et al. (2021d). Given the powerful capability of these techniques and the increasing availability of big data, their application can supplement existing research by providing a new perspective on environmental economics and management, and providing accurate forecasts and pragmatic policy suggestions.

To fill this gap, the International Society for Energy Transition Studies (ISETS) collaborated with Frontiers in an attempt to promote the application of big data, deep learning, machine learning, and other advanced analytical techniques in analyzing environmental economics and policy by inviting members of the ISETS and other non-member researchers to contribute to a dedicated research topic. The objectives are to facilitate the engagement and the advancement of research centred around energy systems. There are four participating journals for this research topic, namely, *Frontiers in Energy Research*, *Frontiers in Environmental Science*, *Frontiers in Ecology and Evolution*, and *Frontiers in Earth Science*.

The goal of this research topic is to re-examine important environmental economics and management issues by employing cutting-edge research methods based on big data, deep learning, and other machine learning techniques, as well as other advanced analytical methods. Given that many important issues in environmental economics and management are exceptionally complex, and the underlying relationships with the determinants are nonlinear, applying these Frontier research methods may prove particularly valuable because of their capability in modelling various complex and nonlinear relationships.

Thirty-six articles based on significant environmental economics and management issues were published under this research topic. All the analyses are based on state-of-the-art analytical techniques. Most of the authors addressed critical issues from an empirical and quantitative

point of view by revisiting the issues with the application of big data, deep learning, other machine learning techniques, as well as other Frontier techniques. Some authors compared the findings derived from existing research studies based on traditional analytical methods with the proposed Frontier research methods. Moreover, many authors delved into burning issues or heated debates to provide insights into environmental economics and management policy formulation. The articles published under this research topic can be broadly divided into two major areas: environmental protection and energy. However, the focuses of the articles are varied and include many important issues in the two areas.

For the area of environmental protection, some researchers conducted research on the relationship between environmental protection and growth. Xiang et al. examined the impact of economic growth on carbon emission in BRICS countries by using the multivariate wavelet analysis. Khan and Wang examined the short and long-run effects of poverty, income inequality, population, and GDP per capita on carbon emission in Pakistan by applying the Autoregressive Distributive Lag (ARDL) and Non-linear Autoregressive Distributive Lag (NARDL) co-integration approach. Li et al. investigated the relation between carbon emissions and economic growth, industry structure, urbanization, research and development (R&D) investment, use of foreign capital, and growth rate of energy consumption in China based on machine learning.

Other researchers focus on emission analysis. For example, Shum et al. investigated the relative importance of carbon emissions drivers in China by employing the Least Absolute Shrinkage and Selection Operator (LASSO) model in ranking the relative importance of the independent variables. In addition, Ma et al. evaluated the feasibility of using machine learning in carbon emission analysis by employing the Gaussian Process Regression (GPR) algorithm.

Moreover, many authors conducted analyses on policy formulation. Zhang et al. examined carbon neutrality policies and technologies by adopting the scientometric analysis. Shao et al. studied the impact of environmental regulation on industrial structure upgrading by using the Pollution Information Transparency Index (PITI) to measure environmental regulation (ER) and examined the effect of ER on industrial structure upgrading. Feng et al. examined the impact of environmental regulations on China's green total factor productivity by using econometrics analysis and machine learning. Xiao et al. investigated the effect of the green credit policy implemented by the Chinese government on firm-level industrial pollutant emissions by employing a quasi-natural experiment, propensity score matching and the difference-in-difference approach (PSM-DID). Yang et al. studied the relationship between financial inclusion and carbon reduction in Chinese counties. Wu et al. examined the impacts of the new urbanization pilot policy on air quality and related air pollutants. Wang et al. implemented a sharp regression discontinuity (RD) design and assessed air quality control effectiveness in China based on the high-volume big data acquired from 173 cities. Zheng and He evaluated the impacts of two revisions of China's environmental protection fee on firm performance based on evidence from the stock markets.

Some researchers focus on the emissions of specific industries. For example, Li et al. examined the change in China's construction industry's domestic carbon emission intensity and analyzed the reason behind the change. Chen et al. studied dynamic supervision and control of volatile organic compounds (VOCs) emission from China's furniture manufacturing industry based on big data and the internet of things (IoT).

Given that environmental issue is a significant concern to many countries, some researchers focus on the Belt and Road Initiative (BRI) countries and aim to evaluate the impacts of BRI on various environmental issues. Li et al. offered an evolutionary and counterfactual baseline to assess the environmental impact of BRI based on the distribution dynamics approach and the mobility probability plots (MPPs). In addition, Lu et al. analyzed the environmental risk contagion relations among the BRI countries and the characteristics of their network structure by using social network analysis (SNA).

Other environmental issues are also examined. For example, Xu et al. examined the environmental efficiency of grain production and its spatial effects in China's major grain production areas by the global super-efficiency SBM model and the Spatial Durbin model.

For the field of energy, many researchers focus on electricity. Li and Cao compared the effectiveness of information feedback between emailing electricity bills to households and installing smart meters in promoting electricity conservation by employing empirical survey data from the Chinese General Social Survey and the propensity score matching method. Jin et al. analyzed the effects of sensitive information disclosure and compared the market-clearing results under different scenarios in the Chinese electricity market. He and Gao developed a dual-sector dynamic equilibrium model, and they introduced electricity consumption and water consumption in a growth model using a time series data set from 1950 to 2014 in Guangzhou, China. Jin et al. studied the way to effectively promote compliance management in the electricity market by using an evolutionary game model under two different scenarios, i.e., the scenario without governmental supervision and the scenario with governmental supervision, and explicitly described the strategic behaviours and dynamic evolution process of power enterprises and regulators in the power market. As the mismatch between energy distribution and power load in China can be alleviated by inter-regional and inter-provincial power transactions, Wang et al. studied a method to deal with inter-regional and inter-provincial transaction settlement deviation quantity based on the kernel density–entropy weight approach.

Other authors examined the efficiency, and Li et al. explored the evolution of manufacturing green development efficiency in the Yangtze River Economic Belt by considering the resource inputs and undesirable outputs in the production process using the WSR methodology, the super-SBM model, and the Tobit model. Liu et al. analyzed the impact of government corruption on energy efficiency (EE) in China from the perspective of energy regulations through statistical methods.

**TABLE 1 |** Summary of the methodologies employed in the articles.

| Article | Authors | Methodologies |
|---|---|---|
| Analysis of Information Feedback on Residential Energy Conservation and the Implications: The Case of China | Zhuangai Li and Xia Cao | Propensity score matching (PSM) method, Logit model |
| Regional Differences of Manufacturing Green Development Efficiency Considering Undesirable Outputs in the Yangtze River Economic Belt Based on Super-SBM and WSR System Methodology | Gen Li, Ying Zhou, Fan Liu and Tao Wang | SBM Model and Green Development Efficiency Measurement of Manufacturing Industry, Tobit Model |
| Does Anti-corruption Policy Influence Energy Efficiency in China? | Jie Liu, Jianming Guo, Xiangfeng Liu, Xiao Bai and Farhad Taghizadeh-Hesary | Fixed effects model, panel threshold regression model |
| Dynamic Risk Assessment of the Overseas Oil and Gas Investment Environment in the Big Data Era | Xuqiang Duan, Xu Zhao, Jianye Liu, Shuquan Zhang and Dongkun Luo | Cloud parameter Bayesian network |
| Does the New Urbanization Influence Air Quality in China? | Yan Wu, Zhixiang Cui and Cong Hu | Difference-in-difference (DID) method |
| Tripartite Coordinated Regulation of New Energy Vehicles Based on Dynamic Bayesian Game | Changyu Liu, Yadong Song, Le Chang and Guanglong Dong | Bayesian dynamic game model |
| Analyzing Information Disclosure in the Chinese Electricity Market | Luosong Jin, Weidong Liu, Xiangyang Wang, Jing Yu and Panting Zhao | Game model |
| How to Promote Compliance Management in the Electricity Market? An Analysis Based on the Evolutionary Game Model | Luosong Jin, Cheng Chen, Yun Li, Xiangyang Wang and Yuanyuan Cheng | Evolutionary Game Model |
| Symmetric and Asymmetric Impact of Poverty, Income Inequality, and Population on Carbon Emission in Pakistan: New Evidence From ARDL and NARDL Co-Integration | Salim Khan and Wang Yahong | Autoregressive Distributive Lag (ARDL), Non-linear Autoregressive Distributive Lag (NARDL) co-integration approach |
| A Method to Deal With Inter-regional and Inter-provincial Transaction Settlement Deviation Quantity Based on Kernel Density–Entropy Weight | Dongfa Wang, Weibin Ding, Feng Gao, Yang Xu and Qin Miao | Kernel function weighting method, improved entropy weight method, deviation electricity responsibility determination model, deviated electric quantity calculation model |
| The Impacts of Two Revisions of the China's Environmental Protection Fee on Firm Performance: Evidence From Stock Markets | Huan Zheng and Yu He | Difference-in-differences (DID) method |
| Analysis of the Upgrading Effect of the Industrial Structure of Environmental Regulation: Evidence From 113 Cities in China | Wei Shao, Yufei Yin, Xiao Bai and Farhad Taghizadeh-Hesary | Panel regression model |
| Evaluating the Impact of Urban Blue Space Accessibility on Housing Price: A Spatial Quantile Regression Approach Applied in Changsha, China | Huang Tuofu, He Qingyun, Yang Dongxiao and Ouyang Xiao | Two-stage instrumental method (2SLS) hedonic model, spatial quantile regression (SQR) model |
| How Does Target-Based Performance Evaluation Affect the Accuracy of Energy-Saving Data: Evidence From China | Ping Zhang, Tiantian Yu, Shengfeng Lu and Ruting Huang | Difference-in-differences (DID) method |
| The Main Research Characteristics of the Development of the Concept of the Circular Economy Concept: A Global Analysis and the Future Agenda | Víctor Meseguer-Sánchez, Francisco Jesús Gálvez-Sánchez, Valentin Molina-Moreno and Gonzalo Wandosell-Fernández-de-Bobadilla | Bibliometric analysis |
| Time-Varying Impact of Economic Growth on Carbon Emission in BRICS Countries: New Evidence From Wavelet Analysis | Lijin Xiang, Xiao Chen, Shuling Su and Zhichao Yin | Decoupling analysis, multivariate wavelet analysis |
| The Spatial Network Contagion of Environmental Risks Among Countries Along the Belt and Road Initiative | Ling Lu, Kai Fang, Chuan Ming Liu and Cheng Sun | Social network analysis (SNA) model |
| Driving Factors of CO2 Emissions: Further Study Based on Machine Learning | Shanshan Li, Yam Wing Siu and Guoqin Zhao | Linear regression (LR), least absolute shrinkage and selection operator (LASSO), and Elastic Net (EN), classification and regression Tree, support vector regression, and k-nearest neighbors regression, ensemble methods, artificial neural networks |
| A Counterfactual Baseline for Assessing Future Environmental Impact: A Case Study of the Belt and Road Initiative | Jing Victor Li, Tsun Se Cheong and Xunpeng Shi | Distribution dynamics approach, mobility probability plot (MPP) |
| Assessments of Air Pollution Control Effectiveness Based on a Sharp Regression Discontinuity Design —Evidence From China's Environmental Big Data | Ren Wang, Jiaqi Huang, Lizhi Zhang, Yu Xia, Xu Xu and Tongli Nong | Sharp regression discontinuity (RD) design |
| The Major Driving Factors of Carbon Emissions in China and Their Relative Importance: An Application of the LASSO Model | Wai Yan Shum, Ning Ma, Xiaomei Lin and Tingting Han | Least absolute shrinkage and selection operator (LASSO) |
| Tracking the Domestic Carbon Emission Intensity of China's Construction Industry: A Global Value Chain Perspective | Xiaoping Li, Yuan Yu, Xunpeng Shi and Xin Hu | Multi-regional input output (MRIO) model, structural decomposition analysis (SDA) method |

(Continued on following page)

**TABLE 1 |** (*Continued*) Summary of the methodologies employed in the articles.

| Article | Authors | Methodologies |
|---|---|---|
| Stability Analysis of the World Energy Trade Structure by Multiscale Embedding | Leyang Xue, Feier Chen, Guiyuan Fu, Qiliang Xia and Luhui Du | Dynamical similarity analysis |
| Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption | Yang Xu, Weijun Gao, Fanyue Qian and Yanxue Li | Attentional-based LSTM network (A-LSTM) |
| The Spatio-Temporal Evolution of China's Hydrogen Fuel Cell Vehicle Innovation Network: Evidence From Patent Citation at Provincial Level | Pei Yu, Feng Jiang, Zhengfang Cai and Yongping Sun | Social network analysis |
| Effects of Age Dependency and Urbanization on Energy Demand in BRICS: Evidence From the Machine Learning Estimator | Zhou Lu, Mantu Kumar Mahalik, Hemachandra Padhan, Monika Gupta and Giray Gozgor | Stochastic Impacts by Regression on Population, Affluence and Technology (STIRPAT) model |
| The Role of Economic Policy Uncertainty in Renewable Energy-Growth Nexus: Evidence From the Rossi-Wang Causality Test | Zhou Lu, Linchuang Zhu, Chi Keung Marco Lau, Aliyu Buhari Isah and Xiaoxian Zhu | Vector Autoregression (VAR)-based Granger-causality test of Rossi-Wang |
| Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the CO2 Emissions Analysis Using Gaussian Process Regression | Ning Ma, Wai Yan Shum, Tingting Han and Fujun Lai | Gaussian Process Regression (GPR) |
| Carbon Neutrality Policies and Technologies: A Scientometric Analysis of Social Science Disciplines | Yuan Zhang, Chung-Lien *Pan* and Han-Teng Liao | Scientometric analysis |
| Assessing the Environmental Efficiency of Grain Production and Their Spatial Effects: Case Study of Major Grain Production Areas in China | Hanxiao Xu, Bei Ma and Qiang Gao | Global super-efficiency SBM mode, spatial Durbin model |
| Exposing the Effects of Environmental Regulations on China's Green Total Factor Productivity: Results From Econometrics Analysis and Machine Learning Methods | Junyi Feng, Jianjun Yan and Xia Tao | Least absolute shrinkage and selection operator (LASSO), orthogonal least square regression, threshold model, mediation model |
| Financial Inclusion and Carbon Reduction: Evidence From Chinese Counties | Zhenkai Yang, Lu Yu, Yinwei Liu, Zhichao Yin and Zumian Xiao | Panel regression |
| Electricity-Water Consumption and Metropolitan Economic Growth: An Empirical Dual Sectors Dynamic Equilibrium Model | Yiming He and Shaoui Gao | Dynamic equilibrium model |
| Tracking Cloud Forests With Cloud Technology and Random Forests | Pasky Pascual and Cam Pascual | Random forest model |
| Dynamic Supervision and Control of VOCs Emission From China's Furniture Manufacturing Based on Big Data and IoT | Zi-Gui Chen, Wei-Ping Wu, Jing Li and Yu-Heng Zeng | Analysis based on application scenarios |
| Does Green Credit Policy Move the Industrial Firms Toward a Greener Future? Evidence From a Quasi-Natural Experiment in China | Zumian Xiao, Lu Yu, Yinwei Liu, Xiaoning Bu and Zhichao Yin | Propensity score matching and difference-in-difference approach (PSM-DID) |

Energy markets has also been studied, for example, Xue et al. analyzed the dynamic trading network structure of the international crude oil and gas market by employing the dynamical similarity analysis at different time scales by inducing a multiscale embedding for dimensionality reduction. Duan et al. studied how the uncertainties and risks of the overseas oil and gas investment environment changed over time and revealed the specific occurrence probabilities of risk on different levels.

Many authors conduct research on the energy transition issue. Using the Bayesian dynamic game model, Liu et al. analyzed the tripartite coordinated regulation for the manufacturers, consumers, and governments in the new energy vehicle (NEV) market. Yu et al. employed social network analysis and patent citation information of hydrogen fuel cell vehicle-related invention patents to construct China's hydrogen fuel cell vehicle innovation network.

As would be expected, renewable energy is also a subject in this research topic. Lu et al. studied the causal relationship between renewable energy consumption and economic growth in four countries: Brazil, Germany, Japan, and the United States, by using the recent vector autoregression (VAR)-based Granger-causality test of Rossi-Wang. In addition, Lu et al. investigated the effects of age dependency ratio and urbanization on renewable and non-renewable energy consumption in Brazil, India, China, and South Africa.

Some other energy issues are also explored in this research topic. Zhang et al. examined how target-based performance evaluation affects the accuracy of energy-saving data and found that the accuracy of the indicator improves after the central government has included energy intensity in the performance appraisal system for local officials. Xu et al. evaluated the potential of using an attentional-based LSTM network (A-LSTM) to predict heating, ventilation, and air-conditioning (HVAC) energy consumption in practical applications.

Besides the articles above, many researchers conduct research on other important topics. For example, Meseguer-Sánchez et al. analyzed the degree of advancement of the circular economy in the scientific field through a bibliometric analysis. Huang et al. studied the proximity effects of different types and sizes of urban

blue spaces on property value in Changsha metropolis, China. They examined the spatial quantile effect across housing prices by using the two-stage instrumental method (2SLS) hedonic model and spatial quantile regression (SQR). Pascual and Pascual explored the cloud forests by running simulations from a predictive model which is based on artificial intelligence, satellite images, and cloud technology.

**Table 1** shows a summary of the methodologies employed by different authors and it can be observed that many researchers employed big data, deep learning, and other machine learning techniques as well as other advanced analytical methods in analyzing environmental protection, energy, environmental economics and management. The application of these state-of-the-art analytical techniques offers a new perspective on many significant issues and contributes significantly to the literature.

The Topic Editors, namely Tsun Se Cheong, Xunpeng Shi, Yanfei Li, and Yongping Sun, would like to express their gratitude to the authors for their contribution to this research topic and look forward to working with the research community in research collaboration soon again in the coming future.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## REFERENCES

Magazzino, C., Mele, M., and Morelli, G. (2021a). The Relationship between Renewable Energy and Economic Growth in a Time of Covid-19: A Machine Learning Experiment on the Brazilian Economy. *Sustainability* 13 (3), 1285. doi:10.3390/su13031285

Magazzino, C., Mele, M., and Schneider, N. (2021b). A D2C Algorithm on the Natural Gas Consumption and Economic Growth: Challenges Faced by Germany and Japan. *Energy* 219, 119586. doi:10.1016/j.energy.2020.119586

Magazzino, C., Mele, M., Morelli, G., and Schneider, N. (2021c). The Nexus between Information Technology and Environmental Pollution: Application of a New Machine Learning Algorithm to OECD Countries. *Util. Policy* 72, 121056. doi:10.1016/j.jup.2021.101256

Magazzino, C., Mele, M., Schneider, N., and Shahbaz, M. (2021d). Can Biomass Energy Curtail Environmental Pollution? A Quantum Model Approach to Germany. *J. Environ. Manag.* 287, 112293. doi:10.1016/j.jenvman.2021.112293

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Cheong, Shi, Li and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 12 February 2021
doi: 10.3389/fenvs.2021.626890

# Analysis of Information Feedback on Residential Energy Conservation and the Implications: The Case of China

*Zhuangai Li and Xia Cao**

*School of Law, Shanxi University of Finance and Economics, Taiyuan, China*

Household electricity consumption has increased significantly over the last two decades, leading to a series of policy interventions targeting to reduce electricity consumption in the residential sector. Information feedback regarding household electricity consumption is considered a cost-effective approach to promoting energy conservation. Various studies have so far covered the effects of information feedback on electricity consumption, yet the variations in both the types of feedback provided and the research methodologies have made it difficult to draw plausible conclusions. In China's case, less due regard has been given to the effectiveness of various types of information feedback. This study has compared the effectiveness of information feedback between emailing electricity bills to households and installing smart meters in terms of promoting electricity conservation by employing empirical survey data from the Chinese General Social Survey with the help of the propensity score matching method. The results show that information provision via information bills curtails electricity consumption by around 20%, whereas information feedback via smart meters installation has no positive effects on household electricity conservation due to lack of access and knowledge to understand the displays of advanced meters. In light of the above results, the study suggests that policy-makers stress the importance of information feedback-based initiatives and improve the information feedback capacity of smart meters through training and education.

**Keywords: residential sector, information feedback, energy conservation, propensity score matching, electricity bills, smart meters**

OPEN ACCESS

***Edited by:***
*Xunpeng Shi,*
*University of Technology Sydney, Australia*

***Reviewed by:***
*Keying Wang,*
*Hubei University Of Economics, China*
*Yu Hao,*
*Beijing Institute of Technology, China*

****Correspondence:***
*Xia Cao*
*caoxia_502@163.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *07 November 2020*
***Accepted:*** *11 January 2021*
***Published:*** *12 February 2021*

***Citation:***
*Li Z and Cao X (2021) Analysis of Information Feedback on Residential Energy Conservation and the Implications: The Case of China. Front. Environ. Sci. 9:626890. doi: 10.3389/fenvs.2021.626890*

## INTRODUCTION

Residential energy consumption (REC), the second largest energy consumer in China, is responsible for about 12.8% of total energy demand with an average annual increase of 7.24% from 2000 to 2017 (Fan et al., 2020). With both direct and indirect $CO_2$ emissions taken into consideration, the residential sector, the major contributor to global warming, has contributed more emissions than the industrial sector (Qi and Li, 2020). Thus, it is of great significance for China to conduct energy conservation and carbon abatement in the residential sector (Zhang H et al., 2020). More recently, China's structure of energy consumption has largely shifted toward electricity and away from direct burning of coal (Xie et al., 2020). The electricity consumption in the dwellings has seen even faster growth, with an average growth rate of 11.49% from 1990 to 2017 (NBS, 2019), so there is enormous untapped potential for energy saving in the residential electricity consumption (Zhang M et al., 2020). In this context, residential electricity saving in China is becoming a priority. Understanding the policy measures to reduce residential electricity consumption is an essential part of mitigating

global warming, making it a prime target for intervention (Davis and Metcalf, 2016). One such intervention is the information feedback, which provides households with the amount of their electricity consumption and monetary costs over a specified time period to promote energy conservation (Ito, 2014). Information feedback addresses problems of incomplete information in markets, which is at the crux of the energy efficiency gap, thus helping the residents overcome cognitive or behavioral biases and make better decisions of electricity consumption (Ratner et al., 2008). This method has been implemented and proven to be successful for promoting behavior change from individual to group levels in various fields, such as public health and education (Aydin et al., 2018).

In the field of electricity, bounded rationality constrains effective planning for electricity consumption, because households have limited information on electricity consumption due to separation of consumption and payment (Gilbert and Zivin, 2014), along with complicated electricity tariff structures (Bushnell and Mansur, 2005), making it difficult for households to understand actual electricity use. Studies have found that only 56% of the households understand their monthly energy consumption (Brounen et al., 2013) and only about 27% of the residents have knowledge about the average price of electricity (Blasch et al., 2017). Therefore, provision of more information feedback will increase awareness and transparency of electricity use. Among feedback-based intervention, electricity bills are considered a low-cost strategy for motivating household electricity conservation and has received considerable research attention (Henryson et al., 2000). A second resource for electricity feedback is from smart metering technology, which helps to remove informational constraints on residential consumption due to advances in energy infrastructure (Carroll et al., 2014). All that feedback information could motivate people to increase awareness of and knowledge about energy consumption levels, and to develop new energy-efficient habits and adapt to real-time pricing more effectively.

Policymakers are increasingly looking to information feedback for guidance. EU Directive (2006/32/EC) on Energy End-Use Efficiency and Energy Services (Energy Services Directive), dating from April 2006, requires the Member States to provide information feedback, including metering, frequent informative billing, historic and normative comparisons, contact information and advice for improving energy end-use efficiency (EUR, 2006). Denmark sets forth a legal obligation to offer informative electricity bills presenting historic and normative comparisons along with environmental impact. Sweden is the first country in Europe to have started a large-scale deployment of smart meters and enacted statutes to encourage provision of monthly billings by smart meters for all customers from July 2009 (Fischer, 2008). Likewise, the United States government allocated over $4 billion to install smart meters in 2009, encouraging utilities to supply residents with access to their energy information (Chopra, 2011). Aware of the significance of information feedback for residential electricity conservation, China has implemented a multitude of policies and programs to increase information feedback capacity, such as deployment of in-house display and smart meters. However, these initiatives are often designed without accurate resident's requirement in mind (Du et al., 2017).

In this paper, we have employed the data of majority of Chinese provinces from the large-scale fieldwork through the Chinese General Social Survey (CGSS) launched jointly by Renmin University and the Hong Kong University of Science and Technology in 2018 to analyze how the information feedback can influence electricity consumption in the residential sector. The CGSS is the first national representative survey project run by academic institution in China mainland for researchers and scientists to study the behavior of Chinese households (Li et al., 2019). In this survey, the information feedback is made available by sending emails to households about their electricity bills or made the consumption shown on the dashboard by using the smart meters. The findings have indicated that providing feedback by sending emails contributes to a substantial reduction in residential electricity consumption, but the deployment of smart meters has failed to achieve the similar result as electricity bills. These results have indicated that information feedback via electricity bills can be used as a strategy to address behavioral barriers and ultimately conserve electricity. Advanced meters alone are not sufficient to change household energy consumption practices in China. The real-time display of electricity expenditure can lead to energy-conservative behavior in households through the learning channel (Lynham et al., 2016). As such, further training about those infrastructures are needed to make consumers gain adequate information in an instantaneous and easily accessible way. Understanding the underlying mechanism of electricity information feedback thus helps to identify challenges and opportunities and offer concrete suggestions for China's future energy policy.

As the effectiveness of information feedback is receiving increasingly more attention, the research has shifted from the laboratory to field experiments (Levitt and List, 2009). However, we still have insufficient knowledge of the effectiveness of information feedback on residential energy consumption (Bernedo et al., 2014), especially in the context of China. In this paper, we evaluate the effects of information feedback via both electricity bills and smart meters using a large-scale fieldwork and a large sample in China. The study uses the propensity score matching to focus on the effectiveness of feedback intervention by collecting the data of households from the Chinese General Social Survey (CGSS).

## LITERATURE REVIEW AND HYPOTHESES

Much literature has discussed over the effectiveness of information feedback on electricity consumption, but with a variety of conclusions by applying the different feedback mechanisms. A set of meta-analysis results indicate that information feedback is effective, and can encourage favorable changes in energy consumption behavior (Delmas et al., 2013; Karlin et al., 2015; Nemati and Penn, 2020). According to some information feedback studies, feedback initiatives could lower residential electricity consumption ranging from 1.1% to 20% (Fischer, 2008; Vine et al., 2013).

Accurate and frequent billing, as interactive feedback service, will give the households a much better sense of the electricity expenditure at different times of a year. Pon (2017) found that billing reports contributed to reducing residential electricity use. Carroll et al. (2014) demonstrated that households could curtail the electricity consumption by 8.7% and 5.4% when receiving monthly billing and bi-monthly billing respectively. The Electric Power Research Institute (2009) concluded that monthly feedback and daily/weekly feedback could promote energy saving by 9% and 8% respectively (Neenan et al., 2009). Gleerup et al. (2010) showed that providing information via messages and emails led to averagely 3% of reduction in total annual electricity consumption. Other studies, however, have found smaller effects or no effects. Andor et al. (2020) found that the electricity feedback based on home energy report could cut down electricity consumption only by 0.7%. Restrepo and Morales-Pinzón (2020) and Khanna et al. (2016) demonstrated that electricity consumption seems unaffected by the delivery of feedback information due to the fact that consumers do not know which activity generates the waste of electricity.

Smart meters could help residents to monitor and better understand their electricity consumption. Gans et al. (2013) and Schleich et al. (2017) indicated that the installation of a smart metering could help reduce electricity consumption by 11–17% and 5% per household respectively. Aydin et al. (2018) showed that feedback initiatives through the use of in-home displays could save electricity by around 20% on average. In China, Zhang et al. (2016) tested the effectiveness of real-time in-home displays on residential electricity consumption, and showed that the in-home displays reduced monthly electricity consumption by around 9.1% in Shanghai City (one of the big and highly developed cities in China). However, Matsukawa (2018) found that feedback from in-home displays only reduced electricity on the part of energy-using households rather than energy-saving households. Hargreaves et al. (2013) found that smart meters did not necessarily bring electricity conservation in the United Kingdom, which is close to the estimates documented in the study by Nilsson et al. (2014) and Khanna et al. (2016) who also found that provision of information feedback via smart meters had little significant impact on electricity consumption on the side of residents. The researchers have ascribed this result to residents' lack of ability to comprehend and utilize the information from advanced intelligent devices.

The effectiveness of information feedback via electricity bills and smart meters in cutting down energy consumption varies widely across studies. Given the variety in reported results, it is essential to identify what types of information feedback can have a positive impact on electricity conservation and to understand the mechanisms behind households' behavior in China. In particular, so far, few studies have identified the causal effects between information-based initiatives and residential electricity consumption at a national level covering both urban and rural residents in China, which is critical to policy intervention. This study tries to fill in this gap.

In light of the previous literature, this paper studies the effects of information feedback via emails and smart meters on households' electricity consumption. Therefore, following hypotheses are put forward based on the arguments presented above.

Hypothesis 1: Feedback information could reduce electricity consumption through sending emails of electricity bills.

Information feedback promotes public awareness and knowledge, which makes residents change their consumption behavior and promoting energy efficiency (Oltra et al., 2013). These changes in household behavior contribute to a reduction in electricity expenditure. Numerous studies have concluded that the provision of electricity bills affects residential electricity consumption (Darby, 2006; Gleerup et al., 2010; Carroll et al., 2014; Pon, 2017). Thus we hypothesize that households that receive electricity bills tend to change their electricity-use behavior, resulting in lower electricity consumption.

Hypothesis 2: The installation of smart meters is conducive to the reduction of electricity consumption.

Smart meters could help households to better monitor their electricity consumption. Studies have examined whether and how smart meters influence electricity usage, but with different conclusions (Gans et al., 2013; Hargreaves et al., 2013; Nilsson et al., 2014; Schleich et al., 2017; Aydin et al., 2018; Matsukawa 2018). We assume that the Chinese households with smart meters are likely to consume less electricity.

# RESEARCH DESIGN

## Sample

We adopt the propensity score matching (PSM) method with the data from the Chinese General Social Survey (CGSS) conducted by the National Survey Research Center of China, which is a continuous large-scale nationwide survey covering 28 provinces in both urban and rural China in 2015. This CGSS survey consists of Core module, Work and Economy module, East Asian Social Survey (EASS) module, International Social Survey Program (ISSP) model, Energy module and Law module with different sample size in each module to systematically investigate the changing relationship between social structure and quality of life. By the end of 2018, about 2,470 journal papers and 645 master's dissertations were published based on CGSS data. In this study, we employed the Core module (sample size 10,968) and the Energy module (sample size 3,557) to discuss the relationship between information feedback via electricity bills or smart meters and electricity consumption in the residential sector. The final sample size is 2,584 after removing the observations with missing values.

To verify that our statistics in the sample are able to provide a sound representation of China's case, the distribution of the demographic factors in our sample survey is compared against the 2015 statistics from the National Bureau of Statistics (NBS, 2015). The education levels of the respondents are presented in **Figure 1A** and the overall trend is similar to that of NBS. **Figure 1B** displays the age distributions of the modeling samples and NBS. Since the head of the household typically responds to the sample, there are some differences about age level 1 and 5. The remaining categories are similar to those of NBS.

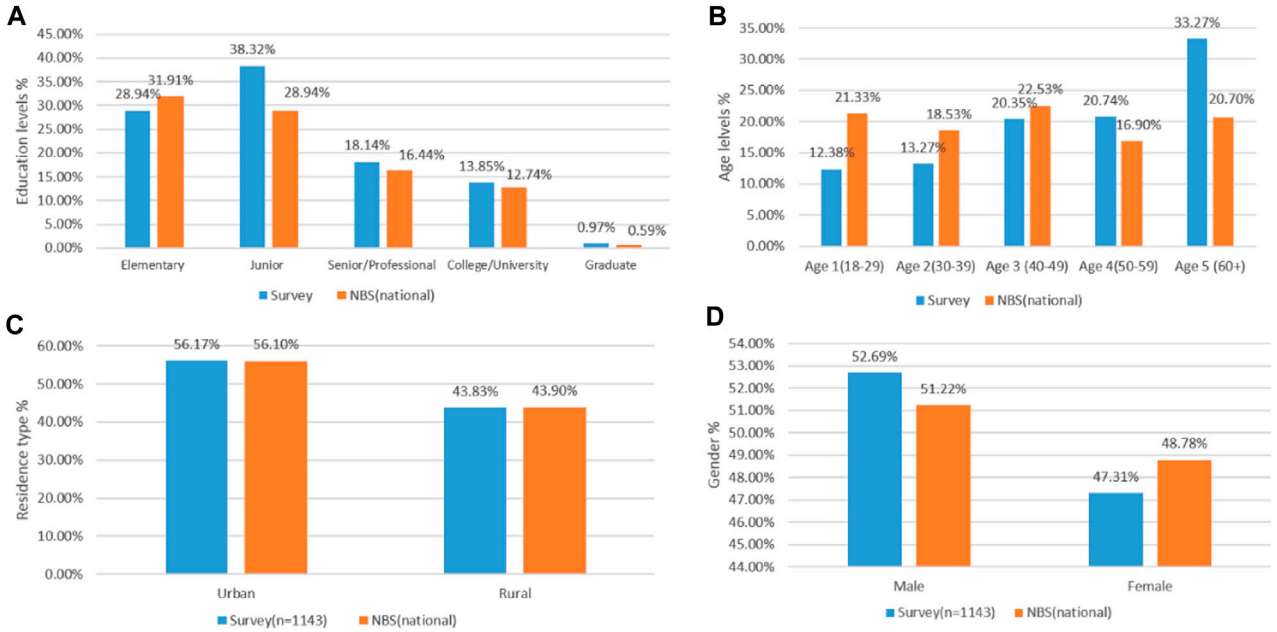



**FIGURE 1 |** Major socioeconomic attributes: survey samples and national distribution. **(A)** education level; **(B)** age level; **(C)** residence type; **(D)** gender.

**Figures 1C,D** show that the distributions of registered residence and gender in the survey sample closely match those of NBS. Besides, the average household size of this survey (2.9 persons per household) is similar to that in the statistical yearbook (3.1 persons per household). Overall, the survey sample has covered major demographic factors.

## Variables

### Electricity Consumption

Electricity consumption is the dependent variable in our study. Sample data of the households' average monthly electricity usage are selected from the dataset of the Chinese General Social Survey (CGSS), in which the respondents are requested to provide the average monthly consumption data in kilowatt hour (kWh) spent on electricity.

### Information Feedback

The independent variable of this study is information feedback by sending the emails about electricity bills and making the information available on the dashboard via smart meters. Information feedback is a dichotomy variable (as receiving information feedback = 1, without receiving information feedback = 0). In the analysis of propensity score matching, the group with information feedback is the treatment group, while the group without information feedback is the control group.

### Control Variables

Further, in order to control the interference of other factors of households, a group of control variables are employed in this study. Previous studies have suggested that the physical attributes of residential dwelling impact household electricity consumption. For example, house size is confirmed as an important indicator of the residential electricity consumption (Santamouris et al., 2007). As the house size increases, the electricity consumption of household is expected to increase. Other dwelling characteristics identified as significant parameters affecting the amount of electricity are the number of electrical appliances (Huebner et al., 2016), house heating (Romero-Jordan et al., 2014), house ownership (Hamilton et al., 2013). As the households own more electrical appliances and house heating, more electricity is consumed. In this study, the total number of residential electrical appliances is calculated, including cooking equipment, refrigerator/freezer, washer/dryer, TV, computer, water heater, television, computers and laptop, air conditioners and washing machine. Previous studies have proved that electricity consumption is strongly correlated with household characteristics such as household income (Huang, 2015), number of individuals in a household (Ndiaye and Gabriel, 2011), length of residency (Xu et al., 2020) and residence type. According to China's Residential Energy Consumption Report, urban residents in China consume approximately 1.25 times as much electricity as rural residents (Yang et al., 2019). Household size, or the number of individuals in a household, is positively associated with residential electricity consumption, as a direct effect (Karatasou and Santamouris, 2019). The residents with larger household size tend to consume more electricity. Besides, there is a positive relationship between the amount of electricity used in a building and income (Bao and Li, 2020). Furthermore, electricity price is also confirmed as an important parameter affecting electricity consumption (Blázquez et al., 2013). Therefore, the control variables include household characteristics with the quantity of occupants, total yearly income in 2014, length of residency and residence type, dwelling characteristics with floor

**TABLE 1 |** Variable definition.

| Variable name | Variable index | Definition |
|---|---|---|
| Electricity consumption | Consumption | in kWh in [0, +∞] |
| Information feedback via electricity bills | Electricity bills | 1 = monthly electricity bills; 0 = otherwise |
| Information feedback via smart meters | Smart meters | 1 = adoption of smart meters; 0 = otherwise |
| Yearly income in 2014 | Income | 5 levels: Well above average, above average, average, below average, well below average |
| Length of residency | Length | Continuous: From 1 to ≥7 |
| Residence type | Residence | 2 levels: Urban residents, rural residents |
| Number of electrical appliances | Number | Continuous: From 1 to ≥51 |
| Floor area | Area | Continuous: From 5 to ≥1,050 |
| Qty. occupants | Qty. | Continuous: From 1 to ≥9 |
| House heating | Heating | 1 = with house heating; 0 = without house heating |
| House ownership | Ownership | 1 = privately owned house;<br>0 = rented dwellings |
| Electricity price (CNY) | Price | Continuous: From 0.36 to ≥0.62 |

**TABLE 2 |** Effect factors of electricity consumption.

| Variable | Variable index | Coefficient | Std. Error |
|---|---|---|---|
| Yearly income in 2014 | Income | 0.127*** | 0.021 |
| Length of residency | ln length | 0.862*** | 0.048 |
| Residence type | Residence | 0.328*** | 0.034 |
| Number of electrical appliances | ln number | 0.704*** | 0.033 |
| Floor area | ln area | 0.249*** | 0.050 |
| Qty. occupants | ln Qty | 0.286*** | 0.031 |
| House heating | Heating | 0.077** | 0.031 |
| House ownership | Ownership | −0.754* | 0.049 |
| Electricity price (CNY) | ln price | 0.973*** | 0.128 |
| LR Chi Square = 134.12 | | | |
| Prob > Chi2 = 0.0000 | | | |
| Pseudo R2 = 0.975 | | | |

*Note. N = 2584. *, **, and *** indicate statistical difference at 10%, 5%, and 1%, respectively.*

area, number of electrical appliances, house heating and house ownership and electricity price which have been identified as the factors influencing the electricity consumption. Variables are shown in **Table 1**. In order to verify the relationship between electricity consumption and control variables, the Ordinary (OLS) methods are employed to estimate the variables' effect on electricity consumption by using the logarithmic forms of electricity consumption and the continuous control variables. **Table 2** illustrates the regression results of electricity consumption, which helps to prove that those control variables indeed influence the electricity consumption. In order to ensure the trustworthy regression results, the White test is used to check for the presence of heteroscedasticity on these variables (White, 1980). The test statistics is 0.29, which fails to reject the null hypothesis of homoscedasticity.

## Model

The issue studied in this paper is the influence of the implementation of information feedback of electricity bills and smart meters on the consumption of electricity. We analyze the effects of information feedback on the residential energy consumption using PSM which applies quasi-experimental option for interventional effects analysis (Rosenbaum and Rubin, 1983). Therefore, taking electricity consumption as the independent variable; taking the information feedback via electricity bills and smart meters as dependent variables; controlling the factors household characteristics with the household size, total yearly income in 2014, length of residency and residence type, dwelling characteristics with including floor area, number of electrical appliances, house heating and house ownership, and electricity price, a flexible Logit model to estimate propensity score is performed in Eq. 1.

$$\boldsymbol{P}(\boldsymbol{x_i}) = \boldsymbol{P}(\boldsymbol{D} = 1|\boldsymbol{x_i}) = \frac{\boldsymbol{exp}(\boldsymbol{\alpha x_i})}{1 + \boldsymbol{exp}(\boldsymbol{\alpha x_i})} \tag{1}$$

$\mathrm{P}(x_i)$ is the probability that the households receive the information feedback, that is, the propensity score of the observed value. $x_i$ is the control variables. The average treatment effect for the treated (ATT) is assessed after estimating the propensity scores. To analyze the effect of information feedback on electricity consumption, average treatment effect $E(Y_1 - Y_0)$ or average treatment effect on treated $E(Y_1 - Y_0)|D = 1$ can be calculated. Here, a binary treatment indicator $D$ equals 1 if households have received the feedback information and 0 otherwise. $Y_0$ shows the electricity consumption of the household without receiving information feedback and $Y_1$ shows electricity consumption of household when receiving information feedback and Average effect of Treatment on the Treated (ATT) can be written as follows:

$$\boldsymbol{ATT} = \boldsymbol{E}(\boldsymbol{Y}_1|\boldsymbol{D} = 1) - \boldsymbol{E}(\boldsymbol{Y}_0|\boldsymbol{D} = 0), \boldsymbol{P}(\boldsymbol{x}) \tag{2}$$

Naturally, the counterfactual mean for those receiving the information feedback via electricity bills and smart meters i.e., $\mathrm{E}[Y_0 \,|\, \mathrm{D} = 1]$ is not observable (the fundamental problem of causal inferences). The treatment effect of receiving information feedback is the difference between the outcomes of electricity consumption for the same household in these two states with and without information feedback. Since unobservable residential characteristics that influence information feedback might also have an effect on electricity consumption and lead to selection bias, propensity score matching offers a way to select a subsample of adopter and non-adopters with characteristics that are observationally similar to obtain the treatment effect

**TABLE 3 |** The descriptive statistics.

| Variable name | Obs | Mean | Sd.dev | Min | Max |
|---|---|---|---|---|---|
| Electricity consumption | 2,585 | 137.1157 | 144.6854 | 3 | 3000 |
| Information feedback via electricity bills | 2,585 | 0.5014 | 0.3594 | 0 | 1 |
| Information feedback via smart meters | 2,585 | 0.3594 | 0.4799 | 0 | 1 |
| Yearly income in 2014 | 2,585 | 2.6789 | 0.7151 | 1 | 5 |
| Length of residency | 2,585 | 6.8143 | 0.8260 | 1 | 7 |
| Residence type | 2,585 | 0.5617 | 0.4963 | 0 | 1 |
| Number of electrical appliances | 2,585 | 7.2272 | 3.3587 | 1 | 51 |
| Floor area | 2,585 | 121.2719 | 88.97107 | 5 | 1050 |
| Qty. occupants | 2,585 | 2.9219 | 1.3994 | 1 | 9 |
| House heating | 2,585 | 0.4674 | 0.4990 | 0 | 1 |
| House ownership | 2,585 | 0.8878 | 0.3157 | 0 | 1 |
| Electricity price (CNY) | 2,585 | 0.5287 | 0.0495 | 0.36 | 0.62 |

estimates. It is assumed that any subject of causal analysis has results by matching each adopter with one or more non-adopters with similar observable attributes. Thus, matching adopting household and non-adopting households can effectively reduce the selection bias and heterogeneity and make sure that the results of the study are more reasonable and reliable.

# EMPIRICAL ANALYSIS

## Descriptive Statistics

**Table 3** gives information about the descriptive statistics. There are 2,584 observations after removing the observations with missing values in the final sample. Through mean, standard deviation, minimum, maximum, and observations, some information could be delivered. The maximum value of household electricity consumption is 3,000 while the minimum value is 3, which means that there is a great difference in household electricity consumption. The mean values of household receiving information feedback via electricity bills and smart meters are 0.5014 and 0.3594 respectively, indicating 50.14% of households receive electricity bills and only 35.94% of households install the smart meters. Then, we compare electricity consumption of households receiving information feedback and those without by PSM analysis, which reduces confounding factors. Specifically, three different matching algorithms: single nearest neighbor methods and four nearest neighbor methods with caliper (0.01) along with the kernel density matching are used to balance the probability of receiving the information feedback. Nearest neighbor matching matches a household from the control unit to a household in the treated unit, based on the closest propensity score. Four nearest neighbor method matches four control cases for each treated household, which could decrease variance. In kernel density matching, household in the treated unit is matched to weighted averages of the households who have similar controls, with greater weight being given to the households with closer scores.

## Determinants of Information Feedback

**Table 4** reports the empirical estimate of the logistic regression of PSM by using the logarithmic forms of electricity consumption and the continuous control variables, which shows the factors influencing the adoption of information feedback. Residence type has a positive effect on the information feedback, which is statistically significant at the 1% level. Urban residents have higher potential to receive information feedback, including both electricity bills and smart meters. In China, households in most third-tiered cities and rural areas fail to install smart meters (Mi et al., 2020), so residents in those areas have limited access to their electricity information from electricity supplies. Thus the government should formulate regulations in infrastructure development to informing the residents of their electricity consumption. Electricity price is also found to be a significant determinant of information feedback adoption via both electricity bills and smart meters. Households paying relatively expensive electricity prices are more likely to receive the electricity bills and install smart meters. This is due to the fact that the households with higher electricity expenditure are willing to monitor their electricity information. Those residents who stay in the dwellings shorter have a higher probability of receiving electricity bills. The households with more electrical appliances have a significantly higher possibility to use smart meters. People in privately owned houses are more interested in the installation of smart meters than those living in rented dwellings. House heating has a negative effect on the adoption of smart meters. This may be ascribable to electric heating's significant share of a household's total electricity consumption. Because of this high share, households are unlikely able to deploy smart meters.

## Checking Overlap Assumption

The effects of adoption of information feedback are estimated using electricity consumption. To meet this end, the research first produces the distribution of density propensity score of electricity bills and smart meters between the treatment and control groups shown in **Figure 2A** and **Figure 3A**, respectively.

**TABLE 4 |** Determinants of information feedback.

| Variable | Electricity bills | | Smart meters | |
|---|---|---|---|---|
| | Coefficient | Std. Error | Coefficient | Std. Error |
| Yearly income in 2014 | 0.049 | 0.062 | 0.087 | 0.059 |
| Length of residency | –0.185*** | 0.049 | –0.055 | 0.049 |
| Residence type | 0.877*** | 0.099 | 0.306*** | 0.093 |
| Number of electrical appliances | 0.012 | 0.014 | 0.060*** | 0.014 |
| Floor area | 0.110 | 0.075 | –0.086 | 0.072 |
| Qty. occupants | –0.020 | 0.032 | –0.006 | 0.030 |
| House heating | –0.110 | 0.091 | –0.426*** | 0.086 |
| House ownership | 0.120 | 0.138 | 0.325*** | 0.137 |
| Electricity price (CNY) | 2.089** | 0.954 | 4.371*** | 0.908 |
| LR Chi Square = 129.87 | | | LR Chi Square = 130.73 | |
| Prob > Chi2 = 0.0000 | | | Prob > Chi2 = 0.0000 | |
| Pseudo R2 = 0.0385 | | | Pseudo R2 = 0.0365 | |

*Note. N = 2584. *, **, and *** indicate statistical difference at 10%, 5%, and 1%, respectively.*

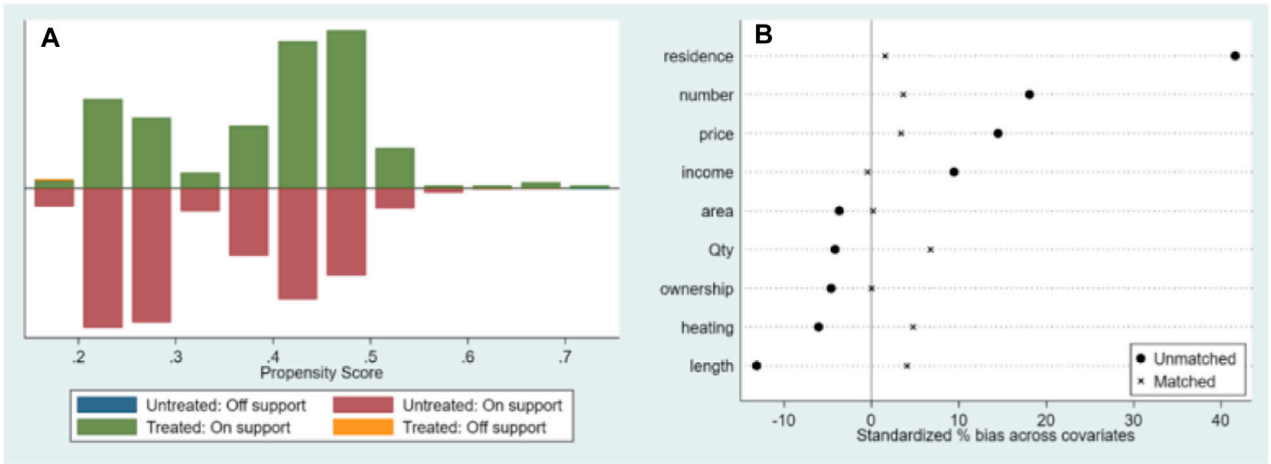



**FIGURE 2 |** Matching quality and balancing tests of electricity bills. **(A)** Distribution of propensity score; **(B)** Standard bias across covariates after matching.

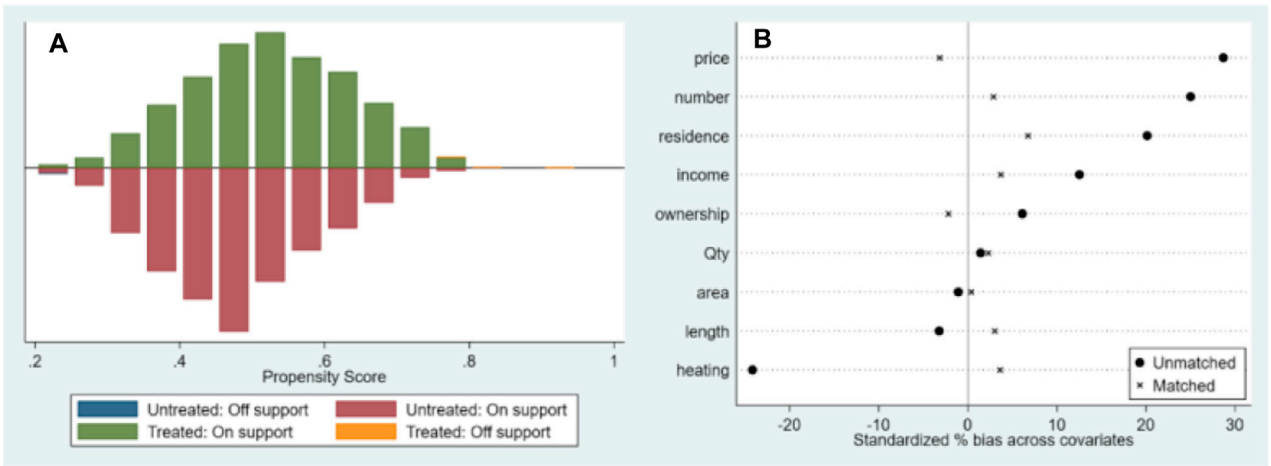



**FIGURE 3 |** Matching quality and balancing tests of smart meters. **(A)** Distribution of propensity score; **(B)** Standard bias across covariates after matching.

This is to testify the quality of the matching process after predicting the propensity scores. From **Figure 2A** and **Figure 3A**, a significant portion of observations of electricity bills and smart meters are among the common ranges. **Figure 2B** and **Figure 3B** present biases of the control variables of electricity bills and smart meters prior to and post the propensity score matching, which demonstrates the balance of the control variables before and after matching. The biases of all control variables have been dramatically reduced for the post matching as shown by the black cross. The distribution of propensity scores of the treated group and the control group of electricity bills and smart meters from the matched sample shows that they greatly overlap after matching, effectively reducing the data deviation. **Figures 4**, **5** illustrate distribution of tendency score of electricity billings and smart meters before matching and after matching respectively, in which the treated group and the control group have relatively good common support after matching. This means that, after matching, treatment and control are very similar to each other with respect to all variables.

**Tables 5, 6** illustrate the balancing test on the control variables of electricity billing and smart meters. From the Tables, the households with information feedback and those without present similar characteristics because the propensity score test presents a significant reduction in bias after matching, which is all below strict criterion of 10. Besides, there is no statistically significant difference between their mean control variables after the matching for both electricity bills and smart meters. Hence, as for both electricity bills and smart meters, apart from quantity of the occupants, the treated and the control groups are comparable conditioned on observable attributes regarding the household size, total yearly income in 2014, length of residency, residence type, floor area, number of electrical appliances, house heating,

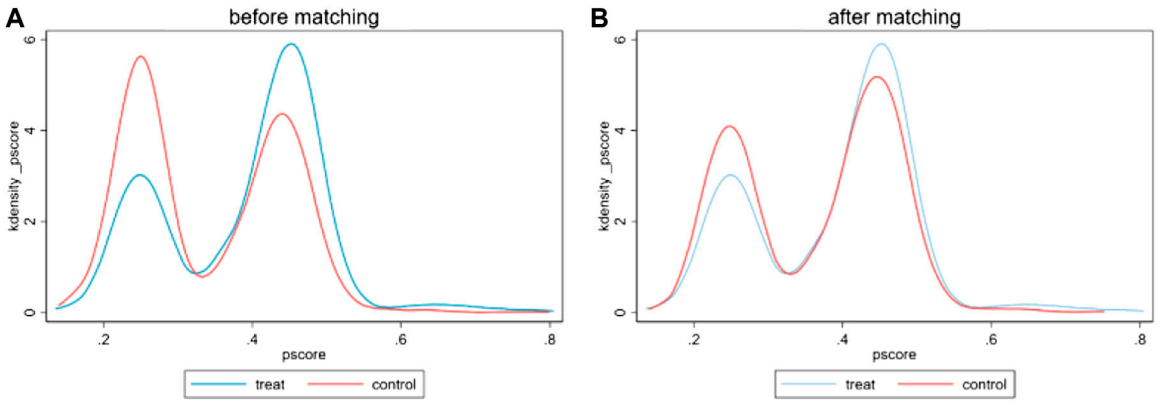



**FIGURE 4 |** Kernel density estimate of electricity billing. **(A)** Before matching; **(B)** After matching.

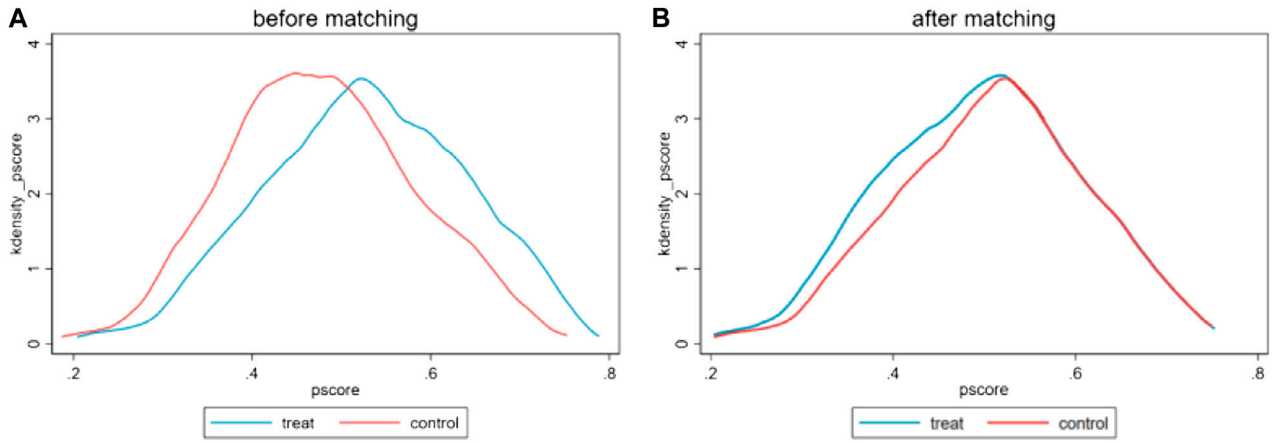



**FIGURE 5 |** Kernel density estimate of smart meters. **(A)** Before matching; **(B)** After matching.

house ownership and electricity price. Further, **Table 7** shows the statistical tests to evaluate the matching of electricity bills and smart meters. The pseudo $R^2$ of the estimated logit model for both electricity bills and smart meters is high before matching and low after matching, which demonstrates the control variables could well explain the probability of receiving information feedback. Similarly, the *p*-values of the likelihood ratio test for both electricity bills and smart meters are all insignificant after matching, indicating that no systematic differences remain in the distribution of control variables between the household receiving information feedback via electricity bills and smart meters and those without after matching. The joint significant effect of the control variables on households receiving both electricity bills and smart meters, as represented by the significant $\chi^2$, could not be rejected prior to matching but is rejected after matching in all the three matching methods. Finally, the mean and median bias are all below 20% as required after matching (Rosenbaum and Rubin, 1983), and are all even below stricter criterion of 10, indicating a very good match.

## ESTIMATION RESULTS OF PSM

Several matching algorithms are offered to further estimate the effects of information feedback on electricity consumption for robustness check of the regression results, as shown in **Table 8**. This is the estimated results of the average treatment effects on treated (ATT) by the single and four nearest neighbor methods with caliper (0.01) along with the kernel estimator and bootstrapped standard errors based on 500 replications of the data. All those matching estimators have provided similar results.

The study results of ATT show that electricity bills have significant influence on the reduction of electricity consumption for all matching algorithms. Households that receive electricity bills have their electricity consumption decreased between 15% and 23%. The significant effect of electricity bills on electricity consumption is in accordance with Vine et al. (2013) and Fischer (2008), who argued that information feedback could cut down consumption by up to 20%. The electricity bills provide the residents with receipts of printed bills, including electricity consumption quantity and expenditure.

**TABLE 5 |** Test of equality of means of each variable before and after matching for electricity bills.

| Variable | Unmatched | Mean | | %Bias | % Reduct \|bias\| | *t*-test | |
|---|---|---|---|---|---|---|---|
| | Matched | Treated | Control | | | t | $p > \|t\|$ |
| Residence | U | 0.6899 | 0.4894 | 41.6 | | 10.05 | 0.000 |
| | M | 0.6915 | 0.6839 | 1.6 | 96.2 | 0.35 | 0.726 |
| Income | U | 2.7223 | 2.655 | 9.4 | | 2.30 | 0.022 |
| | M | 2.7238 | 2.7271 | –0.5 | 95.2 | –0.10 | 0.921 |
| Length | U | 6.7406 | 6.8556 | –13.1 | | –3.40 | 0.001 |
| | M | 6.74 | 6.7044 | 4.1 | 69.0 | 0.72 | 0.471 |
| Number | U | 7.6114 | 7.0115 | 18.1 | | 4.37 | 0.000 |
| | M | 7.6214 | 7.5005 | 3.6 | 79.9 | 0.79 | 0.431 |
| Area | U | 2.0958 | 2.119 | –3.7 | | –0.89 | 0.372 |
| | M | 2.0982 | 2.0971 | 0.2 | 95.4 | 0.04 | 0.970 |
| Qty. | U | 2.8827 | 2.9408 | –4.2 | | –1.02 | 0.310 |
| | M | 2.8803 | 2.7864 | 6.7 | –61.5 | 1.51 | 0.132 |
| Heating | U | 0.4478 | 0.4780 | –6.0 | | –1.47 | 0.141 |
| | M | 0.4488 | 0.4250 | 4.8 | 21.3 | 1.03 | 0.303 |
| Ownership | U | 0.8784 | 0.8931 | –4.6 | | –1.13 | 0.257 |
| | M | 0.8781 | 0.8781 | 0.0 | 100.0 | 0.00 | 1.000 |
| Price | U | 0.5330 | 0.5259 | 14.5 | | 3.57 | 0.000 |
| | M | 0.5333 | 0.5317 | 3.4 | 76.6 | 0.76 | 0.446 |

**TABLE 6 |** Test of equality of means of each variable before and after matching for smart meters.

| Variable | Unmatched | Mean | | %Bias | % Reduct \|bias\| | *t*-test | |
|---|---|---|---|---|---|---|---|
| | Matched | Treated | Control | | | t | $p > \|t\|$ |
| Residence | U | 0.6111 | 0.5117 | 20.1 | | 5.12 | 0.000 |
| | M | 0.61068 | 0.5774 | 6.7 | 66.5 | 1.72 | 0.085 |
| Income | U | 2.7238 | 2.6343 | 12.5 | | 3.18 | 0.001 |
| | M | 2.7229 | 2.6966 | 3.7 | 70.6 | 0.95 | 0.342 |
| Length | U | 6.8009 | 6.8276 | –3.2 | | –0.82 | 0.411 |
| | M | 6.805 | 6.7802 | 3.0 | 7.3 | 0.74 | 0.459 |
| Number | U | 7.6427 | 6.809 | 25.0 | | 6.36 | 0.000 |
| | M | 7.5906 | 7.4954 | 2.9 | 88.6 | 0.71 | 0.477 |
| Area | U | 2.1073 | 2.1141 | –1.1 | | –0.28 | 0.783 |
| | M | 2.1068 | 2.1045 | 0.4 | 66.2 | 0.10 | 0.923 |
| Qty. | U | 2.9298 | 2.9099 | 1.4 | | 0.36 | 0.718 |
| | M | 2.9296 | 2.8978 | 2.3 | –59.9 | 0.58 | 0.562 |
| Heating | U | 0.4074 | 0.5272 | –24.2 | | –6.14 | 0.000 |
| | M | 0.4087 | 0.3909 | 3.6 | 85.1 | 0.92 | 0.356 |
| Ownership | U | 0.8974 | 0.8781 | 6.1 | | 1.55 | 0.121 |
| | M | 0.8971 | 0.9040 | –2.2 | 63.9 | –0.59 | 0.554 |
| Price | U | 0.5353 | 0.5216 | 28.7 | | 7.29 | 0.000 |
| | M | 0.5351 | 0.5366 | –3.2 | 88.9 | –0.84 | 0.400 |

This kind of information feedback may inform households of their monthly consumption, so that they can compare it with historical records. However, the installation of smart meters increases electricity consumption. This is also in conformity with Nilsson et al. (2014) and Hargreaves et al. (2013) whose studies proved that the installation of smart meters did not necessarily reduce electricity consumption. A possible explanation is that the residents have limited access to electricity consumption information in spite of the installation of smart meters. In this survey, only 2% of the households' smart meters are located inside the house and 16% of those households who have deployed smart meters do not know their electricity bills. Most smart meter screens in China only show the accumulated electricity consumption in kWh, so some households may not fully understand the meaning of the display. In addition, households have limited access to historical electricity consumption records via smart meters. By comparison, Zhang et al. (2016) showed that the in-home displays reduced monthly electricity consumption by around 9.1% in Shanghai City due to the fact that the study focused on real-time in-home displays, which could make households check bills online and compare history electricity usage and seek

**TABLE 7 |** Statistical tests to evaluate the matching.

| Matching method | | Pseudo $R^2$ | Likelihood ratio $Chi^2$ | $p > Chi^2$ | Mean bias | Median bias |
|---|---|---|---|---|---|---|
| Before matching | Electricity bills | 0.038 | 129.18 | 0.000 | 22.8 | 19.4 |
| | Smart meters | 0.036 | 130.60 | 0.000 | 23.6 | 22.5 |
| Neighbor matching | Electricity bills | 0.001 | 5.63 | 0.776 | 2.8 | 3.4 |
| | Smart meters | 0.002 | 6.15 | 0.725 | 3.1 | 3.0 |
| Radius matching | Electricity bills | 0.001 | 1.95 | 0.996 | 1.5 | 1.1 |
| | Smart meters | 0.000 | 1.32 | 0.998 | 1.0 | 0.4 |
| Kernel matching | Electricity bills | 0.001 | 2.66 | 0.976 | 2.1 | 2.6 |
| | Smart meters | 0.000 | 1.27 | 0.999 | 1.0 | 0.7 |

**TABLE 8 |** Average treatment effects of different matching algorithms.

| Matching algorithms | Influencing factors | ATT | Std. Err. | Treated | Controls |
|---|---|---|---|---|---|
| Nearest neighbor matching (1:1) | Electricity bills | −22.113*** (−3.19) | 6.938 | 141.529 | 163.642 |
| | Smart meters | 29.021*** (3.95) | 7.351 | 177.810 | 148.788 |
| Nearest neighbor matching (1:4) | Electricity bills | −19.060** (−2.50) | 7.615 | 141.529 | 160.589 |
| | Smart meters | 19.117 *** (2.72) | 7.0264 | 177.613 | 158.496 |
| Kernel-based matching (bandwidth 0.06) | Electricity bills | −15.614** (−2.19) | 6.925 | 141.529 | 157.143 |
| | Smart meters | 19.993*** (3.02) | 6.628 | 177.809 | 157.817 |

*Note. Numbers of t-values are in parentheses. *, **, and *** indicate statistical difference at 10%, 5%, and 1%, respectively.*

advices about potential energy saving. In this case, the residents are presented the electricity consumption data visually. Another reason why installation of smart meters increases electricity consumption smart is that the installation of smart meters is usually associated with high-income urban residents. **Table 4** shows that the urban residents have higher potential to install smart meters. Those residents may have little motivation to conserve electricity because electricity expenditure only takes up a small proportion of their household income. Besides, a boomerang effect of smart meters will occur among the households whose electricity expenditure is relatively modest before the experiment (Matsukawa, 2018). Providing the households with smart meters would increase their electricity consumption by informing them of excessive electricity saving. Fischer (2008) also illustrated that the households with extremely low usage may increase their usage overall when provided with more information. Therefore, further studies are needed to investigate how smart meters could be wisely used to conserve electricity consumption.

## CONCLUSIONS AND RECOMMENDATIONS

The current paper has looked into the effects of information feedback on electricity consumption in the residential sector with the micro-level data from the Chinese General Social Survey (CGSS) by using a propensity score matching (PSM) method. The study shows that the feedback in the form of electricity bills proves effective in encouraging the residents to make efficient use of electricity, which is consistent with our hypothesis. This is because in the absence of feedback, the households lack information on their electricity consumption and are unaware of the potential for further electricity conservation (Tiefenbeck et al., 2018).

Providing feedback via electricity bills can inform the residents of actual electricity consumption and costs frequently enough to motivate them to control their own electricity consumption, thus resulting in adjustments and changes of households' behavior and improving energy efficiency in the residential sector (Gans et al., 2013). Besides, historical information feedback serves as a self-comparison that can drive the residents by establishing personal norms (Promann and Brunswicker, 2017), which could lead to 10% reduction in electricity consumption (Wood and Newborough, 2007). Therefore, the households should be provided with the amount of electricity compared to a previous period or some pre-set standards. In addition, evidence has shown that peers' or neighbors' behavior significantly influences the residents' behavior change and proves the effectiveness of social norms in response to information feedback (Allcott, 2011). Information feedback via social comparison can cut down energy consumption of private households, ranging from 1.2% to 30% (Andor and Fels, 2018). Those households, who receive both individual and comparative information feedback, are successful in curtailing energy consumption in both the short run and long run, especially when targeting high consumers of energy due to a higher energy-saving potential (McCalley and Midden, 2002).

Concerning frequency of electricity bills, Abrahamse et al. (2005) concludes that energy saving potential of feedback relies heavily on its frequency through the systematic analysis. Frequent information feedback has proven to be a successful policy intervention for promoting energy conservation. Therefore, the persistence of feedback should be taken into consideration when planning this type of intervention because the effects of feedback

often disappear when it is removed (Ehrhardt-Martinezet al., 2010). Hence, from a policy perspective, the information provided to the households via electricity bills should include knowledge and tips regarding energy conservation, a comparison of usage history and energy consumption of neighbors, pro-environmental behaviors and their potential impact on the environment and climate change. Besides, the information feedback should be offered more frequently to remain the effects of feedback.

As for smart meters, our studies show that the installation of smart meters fails to exert positive effects on electricity conservation, which contradicts the hypothesis. The differences of the studies' duration, sample size, country, methodologies applied among researches may make studies come to different conclusions. This finding is consistent with the results drawn by Hargreaves et al. (2013) and Khanna et al. (2016). The previous studies prove that the feedback via smart meters is closely associated with positive environmental attitudes and understanding of information that smart meters provide (Darby, 2006). However, a majority of the residents in China lack information about smart meters and thus cannot convert the information into action to conserve electricity. Therefore, the government and public utility should provide trustworthy advice and technical training to increase consumers' understanding of smart meters to make the conservation possible.

Although the smart meters boast a relatively high penetration in China, making up more than 50% of the market, most resident users of those smart meters have limited access to the information as meters are installed outside the house. Smart meters for the households only perform functions of measuring usage data and fail to deliver adequate information to the residents (Du et al., 2017). Thus, smart meters limit the capability to inform the residents as it relies on the residents' self-initiation with low residents' engagement in energy conservation. Therefore, from a policy perspective, smart meters should bring all this information onto residents' computers and mobile phones inside home in a form that is accessible, clear and easy to understand, facilitating the residents to curb their electricity consumption. Besides, as learning plays a prominent role in enhancing energy conservation in the residential sector (Lynham et al., 2016), the policy-makers should provide training and educational outreach programs about the usage of smart meters. With the help of additional training and education, smart meters help the households to access to the information about usage in real time and keep informed of the use of specific appliances where they can conserve electricity most effectively. Provision of appliance-specific feedback by using smart meters keeps the households better informed of the potential ways through which they could reduce electricity consumption, in particular, from high energy-use electrical devices. Overall, instantaneous direct feedback via smart meters and frequent, accurate electricity billing are needed as a basis for sustained electricity demand conservation, which provides the guidance for future policy formation.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

All authors have made a substantial, direct and intellectual contribution to this article and approved it for publication.

## REFERENCES


Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *J. Environ. Psychol.* 25 (3), 273–291. doi:10.1016/j.jenvp.2005.08.002

Allcott, H. (2011). Social norms and energy conservation. *J. Publ. Econ.* 95 (9–10), 1082–1095. doi:10.1016/j.jpubeco.2011.03.003

Andor, M. A., and Fels, K. M. (2018). Behavioral economics and energy conservation–a systematic review of non-price interventions and their causal effects. *Ecol. Econ.* 148, 178–210. doi:10.1016/j.ecolecon.2018.01.018

Andor, M. A., Gerster, A., Peters, J., and Schmidt, C. M. (2020). Social norms and energy conservation beyond the United States. *J. Environ. Econ. Manag.* 103, 102351. doi:10.1016/j.jeem.2020.102351

Aydin, E., Brounen, D., and Kok, N. (2018). Information provision and energy consumption: evidence from a field experiment. *Energy Econ.* 71, 403–410. doi:10.1016/j.eneco.2018.03.008

Bao, H. X., and Li, S. H. (2020). Housing wealth and residential energy consumption. *Energy Pol.* 143, 111581. doi:10.1016/j.enpol.2020.111581

Bernedo, M., Ferraro, P. J., and Price, M. (2014). The persistent impacts of norm-based messaging and their implications for water conservation. *J. Consum. Pol.* 37 (3), 437–452. doi:10.1007/s10603-014-9266-0

Blasch, J. E., Filippini, M., Boogen, N., and Kumar, N. (2017). Explaining electricity demand and the role of energy and investment literacy on end-use efficiency of Swiss households. *Energy Econ.* 68, 89–102. doi:10.1016/j.eneco.2017.12.004

Blázquez, L., Boogen, N., and Filippini, M. (2013). Residential electricity demand in Spain: new empirical evidence using aggregate data. *Energy Econ.* 36, 648–657. doi:10.1016/j.eneco.2012.11.010

Brounen, D., Kok, N., and Quigley, J. M. (2013). Energy literacy, awareness, and conservation behavior of residential households. *Energy Econ.* 38, 42–50. doi:10.1016/j.eneco.2013.02.008

Bushnell, J. B., and Mansur, E. T. (2005). Consumption under noisy price signals: a study of electricity retail rate deregulation in San Diego. *J. Ind. Econ.* 53 (4), 493–513. doi:10.1111/j.1467-6451.2005.00267.x

Carroll, J., Lyons, S., and Denny, E. (2014). Reducing household electricity demand through smart metering: the role of improved information about energy saving. *Energy Econ.* 45, 234–243. doi:10.1016/j.eneco.2014.07.007

Chopra, A. (2011). Modeling a green energy challenge after a blue button. White House Office of Science and Technology Policy. Available at: http://www.whitehouse.gov/blog/2011/09/15/modeling-green-energychallenge-after-blue-button (Accessed September 27, 2020).

Darby, S. (2006). The effectiveness of feedback on energy consumption: a Review for DEFRA of the literature on metering, billing and direct displays (Oxford, UK: Environmental Change Institute). Available at: http://wwweci.ox.ac.uk/research/energy/downloads/smart meteringreport.pdf (Accessed September 30, 2020).

Davis, L. W., and Metcalf, G. E. (2016). Does better information lead to better choices? Evidence from energy-efficiency labels. *J. Assoc. Environ Reso.* 3 (3), 589–625. doi:10.1086/686252

Delmas, M. A., Fischlein, M., and Asensio, O. I. (2013). Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975 to 2012. *Energy Pol.* 61, 729–739. doi:10.1016/j.enpol.2013.05.109

Du, L., Guo, J., and Wei, C. (2017). Impact of information feedback on residential electricity demand in China. *Resour. Conserv. Recycl.* 125, 324–334. doi:10.1016/j.resconrec.2017.07.004

Ehrhardt-Martinez, K., Donnelly, K. A., and Laitner, S. (2010). Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities (Washington, DC: American Council for an Energy-Efficient Economy. Available at: http://aceee.org/research-report/e105 (Accessed October 11, 2020).

EUR (2006). Directive 2006/32/EC of the European Parliament and of the Council of 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC. Available at: http://eurlex.europa.eu/LexUriServ/site/en/oj/2006/l_114/l_11420060427en00640085.pdf(Accessed September 20, 2020).

Fan, J. L., Chen, K. Y., and Zhang, X. (2020). Inequality of household energy and water consumption in China: an input-output analysis. *J. Environ. Manag.* 269, 110716. doi:10.1016/j.jenvman.2020.110716

Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy?. *Energy Effic.* 1 (1), 79–104. doi:10.1007/s12053-008-9009-7

Gans, W., Alberini, A., and Longo, A. (2013). Smart meter devices and the effect of feedback on residential electricity consumption: evidence from a natural experiment in Northern Ireland. *Energy Econ.* 36 (3), 729–743. doi:10.1016/j.eneco.2012.11.022

Gilbert, B., and Zivin, J. G. (2014). Dynamic salience with intermittent billing: evidence from smart electricity meters. *J. Econ. Behav. Organ.* 107, 176–190. doi:10.1016/j.jebo.2014.03.011

Gleerup, M., Larsen, A., Leth-Petersen, S., and Togeby, M. (2010). The effect of feedback by text message (SMS) and email on household electricity consumption: experimental evidence. *Energy J.* 31 (3), 113–132. doi:10.5547/ISSN0195-6574-EJ-Vol31-No3-6

Hamilton, I. G., Steadman, P. J., Bruhns, H., Summerfield, A. J., and Lowe, R. (2013). Energy efficiency in the British housing stock: energy demand and the Homes Energy Efficiency Database. *Energy Pol.* 60, 462–480. doi:10.1016/j.enpol.2013.04.004

Hargreaves, T., Nye, M., and Burgess, J. (2013). Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term. *Energy Pol.* 52, 126–134. doi:10.1016/j.enpol.2012.03.027

Henryson, J., Håkansson, T., and Pyrko, J. (2000). Energy efficiency in buildings through information–Swedish perspective. *Energy Pol.* 28 (3), 169–180. doi:10.1016/S0301-4215(00)00004-5

Huang, W. H. (2015). The determinants of household electricity consumption in Taiwan: evidence from quantile regression. *Energy* 87, 120–133. doi:10.1016/j.energy.2015.04.101

Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z., and Oreszczyn, T. (2016). Understanding electricity consumption: a comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Appl. Energy* 177, 692–702. doi:10.1016/j.apenergy.2016.04.075

Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *Am. Econ. Rev.* 104 (2), 537–563. doi:10.1257/aer.104.2.537

Karatasou, S., and Santamouris, M. (2019). Socio-economic status and residential energy consumption: a latent variable approach. *Energy Build.* 198, 100–105. doi:10.1016/j.enbuild.2019.06.013

Karlin, B., Zinger, J. F., and Ford, R. (2015). The effects of feedback on energy conservation: a meta-analysis. *Psychol. Bull.* 141 (6), 1205–1227. doi:10.1037/a0039650

Khanna, N. Z., Guo, J., and Zheng, X. (2016). Effects of demand side management on Chinese household electricity consumption: empirical findings from Chinese household survey. *Energy Pol.* 95, 113–125. doi:10.1016/j.enpol.2016.04.049

Levitt, S. D., and List, J. A. (2009). Field experiments in economics: the past, the present, and the future. *Eur. Econ. Rev.* 53 (3), 1–18. doi:10.1016/j.euroecorev.2008.12.001

Li, J., Zhang, J., Zhang, D., and Ji, Q. (2019). Does gender inequality affect household green consumption behaviour in China?. *Energy Pol.* 135, 111071. doi:10.1016/j.enpol.2019.111071

Lynham, J., Nitta, K., Saijo, T., and Tarui, N. (2016). Why does real-time information reduce energy consumption?. *Energy Econ.* 54, 173–181. doi:10.1016/j.eneco.2015.11.007

Matsukawa, I. (2018). Information acquisition and residential electricity consumption: evidence from a field experiment. *Resour. Energy Econ.* 53, 1–19. doi:10.1016/j.reseneeco.2018.02.001

McCalley, L. T., and Midden, C. J. (2002). Energy conservation through product-integrated feedback: the roles of goal-setting and social orientation. *J. Econ. Psychol.* 23 (5), 589–603. doi:10.1016/S0167-4870(02)00119-8

Mi, L., Qiao, L., Du, S., Xu, T., Gan, X., Wang, W., et al. (2020). Evaluating the effect of eight customized information strategies on urban households' electricity saving: a field experiment in China. *Sust. Cities Soc.* 62, 102344. doi:10.1016/j.scs.2020.102344

National Bureau of Statistics of China (2015). China statistical yearbook 2019. Available at http://www.stats.gov.cn/tjsj/ndsj/2015/indexch.htm (Accessed October 10, 2020).

National Bureau of Statistics of China (2019). China statistical yearbook 2019. Available at http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm (Accessed September 10, 2020).

Ndiaye, D., and Gabriel, K. (2011). Principal component analysis of the electricity consumption in residential dwellings. *Energy Build.* 43 (2–3), 446–453. doi:10.1016/j.enbuild.2010.10.008

Neenan, B., Robinson, J., and Boisvert, R. N. (2009). *Residential electricity use feedback: a research synthesis and economic framework*. Palo Alto, CA: Electric Power Research Institute.

Nemati, M., and Penn, J. (2020). The impact of information-based interventions on conservation behavior: a meta-analysis. *Resour. Energy Econ.* 62, 101201. doi:10.1016/j.reseneeco.2020.101201

Nilsson, A., Bergstad, C. J., Thuvander, L., Andersson, D., Andersson, K., and Meiling, P. (2014). Effects of continuous feedback on households' electricity consumption: potentials and barriers. *Appl. Energy* 122, 17–23. doi:10.1016/j.apenergy.2014.01.060

Oltra, C., Boso, A., Espluga, J., and Prades, A. (2013). A qualitative study of users' engagement with real-time feedback from in-house energy consumption displays. *Energy Pol.* 61, 788–792. doi:10.1016/j.enpol.2013.06.127

Pon, S. (2017). The effect of information on TOU electricity use: an Irish residential study. *Energy J.* 38 (6), 55–79. doi:10.5547/01956574.38.6.spon

Promann, M., and Brunswicker, S. (2017). "Affordances of eco-feedback design in home energy context," in Proceedings of 33rd American Conference Information Systems, 2017 August, Boston, MA Vol. 23, 1–10. Available at: https://aisel.aisnet.org/amcis2017/GreenIS/Presentations/7/ (Accessed October 15, 2020)

Qi, W., and Li, G. (2020). Residential carbon emission embedded in China's inter-provincial population migration. *Energy Pol.* 136, 111065. doi:10.1016/j.enpol.2019.111065

Ratner, R. K., Soman, D., Zauberman, G., Ariely, D., Carmon, Z., Keller, P. A., et al. (2008). How behavioral decision research can enhance consumer welfare: from freedom of choice to paternalistic intervention. *Market. Lett.* 19 (3–4), 383. doi:10.1007/s11002-008-9044-3

Restrepo, J. D. C., and Morales-Pinzón, T. (2020). Effects of feedback information on the household consumption of water and electricity: a case study in Colombia. *J. Environ. Manag.* 262, 110315. doi:10.1016/j.jenvman.2020.110315

Romero-Jordan, D., Penasco, C., and Rio, P. D. (2014). Analysing the determinants of household electricity demand in Spain. an econometric study. *Int. J. Electr. Power Energy Syst.* 63, 950–961. doi:10.1016/j.ijepes.2014.06.005

Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1), 41–55. doi:10.1093/biomet/70.1.41

Santamouris, M., Kapsis, K., Korres, D., Livada, I., Pavlou, C., and Assimakopoulos, M. N. (2007). On the relation between the energy and social characteristics of the residential sector. *Energy Build.* 39 (8), 893–905. doi:10.1016/j.enbuild.2006.11.001

Schleich, J., Faure, C., and Klobasa, M. (2017). Persistence of the effects of providing feedback alongside smart metering devices on household electricity demand. *Energy Pol.* 107, 225–233. doi:10.1016/j.enpol.2017.05.002

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., et al. (2018). Overcoming salience bias: how real-time feedback fosters resource conservation. *Manag. Sci.* 64 (3), 1458–1476. doi:10.1287/mnsc.2016.2646

Vine, D., Buys, L., and Morris, P. (2013). The effectiveness of energy feedback for conservation and peak demand: a literature review. *Open J. Energy Effic.* 2 (1), 7–15. doi:10.4236/ojee.2013.21002

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4), 817–838. doi:10.2307/1912934

Wood, G., and Newborough, M. (2007). Energy-use information transfer for intelligent homes: enabling energy conservation with central and local displays. *Energy Build.* 39, 495–503. doi:10.1016/j.enbuild.2006.06.009

Xie, L., Yan, H., Zhang, S., and Wei, C. (2020). Does urbanization increase residential energy use? Evidence from the Chinese residential energy consumption survey 2012. *China Econ. Rev.* 59, 101374. doi:10.1016/j.chieco.2019.101374

Xu, X., Xiao, B., and Li, C. Z. (2020). Critical factors of electricity consumption in residential buildings: an analysis from the point of occupant characteristics view. *J. Clean. Prod.* 256, 120423. doi:10.1016/j.jclepro.2020.120423

Yang, Y., Liu, J., Lin, Y., and Li, Q. (2019). The impact of urbanization on China's residential energy consumption. *Struct. Change Econ. Dynam.* 49, 170–182. doi:10.1016/j.strueco.2018.09.002

Zhang, H., Shi, X., Cheong, T. S., and Wang, K. (2020). Convergence of carbon emissions at the household level in China: a distribution dynamics approach. *Energy Econ.* 92, 104956. doi:10.1016/j.eneco.2020.104956

Zhang, M., Cheng, C. H., and Ma, H. Y. (2020). Projection of residential and commercial electricity consumption under SSPs in Jiangsu province, China. *Adv. Clim. Change Res.* 11 (2), 131–140. doi:10.1016/j.accre.2020.06.005

Zhang, X., Shen, J., Yang, T., Tang, L., Wang, L., Liu, Y., et al. (2016). Smart meter and in-home display for energy savings in residential buildings: a pilot investigation in Shanghai, China. *Intell. Build. Int.* 11 (1), 4–26. doi:10.1080/17508975.2016.1213694

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Li and Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fenvs.2020.631911

# Regional Differences of Manufacturing Green Development Efficiency Considering Undesirable Outputs in the Yangtze River Economic Belt Based on Super-SBM and WSR System Methodology

*Gen Li*[1], *Ying Zhou*[1], *Fan Liu*[2]* *and Tao Wang*[1]

[1]School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, China, [2]School of Business Administration, Zhongnan University of Economics and Law, Wuhan, China

To explore the evolution mechanism of manufacturing green development efficiency is of great significance to realize green transformation of manufacturing industry in the Yangtze River Economic Belt. This paper fully considers the resource inputs and undesirable outputs in the production process and applies WSR methodology to construct the index system of influencing factors. Based on the panel data of 11 provinces and cities in the Yangtze River Economic Belt from 1998 to 2017, the super-SBM model is used to calculate the manufacturing green development efficiency. Then, the regional differences of manufacturing green development efficiency in the Yangtze River Economic Belt are deeply analyzed. Finally, Tobit model is applied to analyze the influencing factors of the manufacturing green development efficiency. And it turns out, during the statistics period, manufacturing green development efficiency in the Yangtze River Economic Belt is "U" shaped distribution, the mean value of each province over the years is 0.812, which is at the medium development level; the manufacturing green development efficiency in the Yangtze River Economic Belt is on the rise, and the low scale efficiency is the main reason that restricts the manufacturing green development efficiency in the Yangtze River Economic Belt. All the influencing factors have different effects on the manufacturing green development efficiency in different regions. Therefore, this paper puts forward corresponding policy suggestions from the three dimensions of Wuli, Shili and Renli.

**Keywords: manufacturing industry, green development efficiency, WSR methodology, super-SBM model, tobit model**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Yongbing Chen,*
*Xiamen University, China*
*Qingping Jin,*
*Wuhan University of Science and Technology, China*

***Correspondence:**
*Fan Liu*
*fanliusfsu@gmail.com*

***Specialty section:***
*This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science*

***Received:*** *24 November 2020*
***Accepted:*** *24 December 2020*
***Published:*** *18 February 2021*

***Citation:***
*Li G, Zhou Y, Liu F and Wang T (2021) Regional Differences of Manufacturing Green Development Efficiency Considering Undesirable Outputs in the Yangtze River Economic Belt Based on Super-SBM and WSR System Methodology. Front. Environ. Sci. 8:631911. doi: 10.3389/fenvs.2020.631911*

## INTRODUCTION

Manufacturing industry is the main source of modern material wealth, but also an important industry that causes damage to the ecological environment. The manufacturing enterprises in the Yangtze River Economic Belt are dense, and there are many environmental risks. The unreasonable industrial structure and layout result in prominent ecological and environmental problems of accumulation, superposition and potential, which have become the main bottleneck restricting the

green development of the Yangtze River Economic Belt. To this end, China issued the guidance on strengthening the green development of industry along the Yangtze River Economic Belt in 2017. In april 2018, general secretary Xi Jinping pointed out at a symposium on promoting the development of the Yangtze River Economic Belt in Hubei province that manufacturing industry was the main body and important force in the ecological and environmental protection construction along the Yangtze River Economic Belt, sticking to the path of ecological priority and green development. In June 2018, the Chinese audit office announced that the total amount of wastewater discharge in the Yangtze River Economic Belt accounted for more than 40% of China's total, and the emission intensity of chemical oxygen demand, ammonia nitrogen, nitrogen oxides, sulfur dioxide and volatile organic compounds per unit region was about 1.5–2.0 times the national average level. Incremental expansion of manufacturing development in the Yangtze River Economic Belt, such as high water production mode is still dominant; the ecological environment pressure is still large. Under the background of calling for ecological priority and green development, the Yangtze River Economic Belt should focus on improving the green development level of manufacturing industry and reducing the waste of resources and environmental pollution as much as possible. As the regions of the Yangtze River Economic Belt have great differences in economic development level and other aspects, and the development model of manufacturing industry is also different, so it is urgent to consider the green balanced development of manufacturing industry among regions. Therefore, it is of great significance to explore the evolution mechanism of manufacturing green development efficiency in the Yangtze River Economic Belt and find out the causes of unbalanced manufacturing green development efficiency in different regions for realizing green transformation of manufacturing industry in the Yangtze River Economic Belt.

At present, scholars have carried out a lot of researches on the connotation, mechanism and realization path of green development. In 1989, Pierce first put forward the concept of "green economy", and then the concept of green development gradually took shape. Green development emphasizes saving resources and protecting the ecological environment while developing economy. Jiang and Qu, 2020 conducted an in-depth discussion on core journals, core authors, core countries, discipline distribution, key word co-occurrence and literature co-citation by using Cite Space software, taking 6,591 papers from core journals of Science Net (1999–2019) as research samples. It found that the influence of Chinese research institutions and scholars was increasing, mainly in the fields of engineering and environmental science. The sustainable innovation ability of the manufacturing industry was generally at a high level. There are relatively few researches on the green development efficiency measurement of manufacturing industry in the Yangtze River Economic Belt. The existing researches on the green development efficiency measurement mainly focus on the following aspects: 1) Object of the study. Most existing scholars choose regions or industries as research objects and analyze the evolution characteristics of the green development efficiency of regions or industries. Wang et al. (2014) calculated the green technology efficiency of provincial units from 2001 to 2010 based on the SFA model, and then analyzed the factors from the perspectives of technology, system and industry. Jing and Zhang, 2014 constructed the global Luenberger index based on SBM model to measure the green technology progress of Chinese industry. Li and Zhang, 2016 used the entropy weight TOPSIS model to evaluate the industrial green development level of 108 prefecture-level cities in the Yangtze River Economic Belt from 2004 to 2013, and applied the Tin index to analyze the difference characteristics and composition of industrial green development level among and within the three major city clusters in the Yangtze River Economic Belt. Duman and Kasman, 2018 adopted the parametric hyperbolic distance function to study the environmental technology efficiency of EU member states and candidate countries from 1990 to 2011. Singh C. et al. (2020) identified and examined the various key performance parameters (KPPs) of Green-Lean practices in manufacturing industries to develop green manufacturing by utilizing resources and reducing waste. Future research would focus on ranking these KPPs of green lean manufacturing using appropriate Multi-Criteria Decision Analysis (MCDA) technique. 2) Research indicator. Since the index system method contains many indicators, it is difficult to obtain statistical data. Many scholars adopt the input-output method, which takes capital, labor and energy as input variables, GDP or industrial added value as desirable output, and industrial waste as undesirable output, to calculate the efficiency of green development. Wu and Wen, 2013 calculated China's industrial green productivity and emission reduction cost from 1995 to 2009 respectively by SML index method and DEA method based on the direction distance function. Then, based on the panel data model estimation method, it discussed the influencing factors of performance and cost in industrial emission reduction. Chen (2010) re-estimated the total factor productivity of Chinese industry since the reform based on the directional distance function, and found that the actual total factor productivity with correct consideration of environmental constraints was much lower than the traditional estimate without correct consideration of environmental factors, which was also supported by the comparison with the results of the main literature. Yi et al. (2018) used DEA-Malmquist index method and exploratory spatial data analysis method to calculate and study the spatial and temporal differences of green total factor productivity (GTFP) in 11 provinces (cities) of the Yangtze River Economic Belt from 2004 to 2015. (3)Research method. Based on the regional or industrial input and output data, most scholars apply SFA, DEA and comprehensive index evaluation to calculate the regional or industrial green development efficiency. Considering the undesirable output in the production process, Li et al. (2019)constructed an SBM model to estimate the energy ecological efficiency of 30 provinces and cities in China from 2000 to 2016. Then the regional differences in the energy ecological efficiency of the manufacturing industry were analyzed in depth. Finally, the Tobit model was used to empirically analyze the factors affecting the energy ecological efficiency of the manufacturing industry in China. Li and Lin, 2017 evaluated

and compared the ecological total factor energy efficiency of heavy and light industries and evaluated their technology gaps. Wang et al. (2017) built the Epsilon-Based Measure (EBM)-Tobit two-stage efficiency analysis model by taking regional industry as the research object and integrating industrial water and water pollution emissions into the analysis framework. Sun et al. (2017) measured the green efficiency of water resource based on data enveloped technology, comparing and analyzing the results with the traditional measures of water resource economic efficiency and environmental efficiency, and studying the spatial pattern by using ESDA method. Nukman et al. (2017) established a green manufacturing index (GMI) technology that would determine the effectiveness of green manufacturing processes from an economic and environmental perspective. Zhang J. et al. (2018) used the DEA-Malmquist method to measure the green total factor productivity of various types in China's food industry from 2006 to 2014. The results showed that from 2006 to 2014, the environmental pollution index of China's food industry showed an upward - downward - upward trend. From 2006 to 2014, green total factor productivity in China's food industry showed an upward trend. The mean technical efficiency of China's food industry was low, but the technical efficiency was on the rise. The above-mentioned scholars mainly focus on the macro research of the whole country or province, and the micro research results of a certain region or subdivided industry are relatively less. They mostly use the traditional input-output indicators and adopt the SBM model of undesirable output to measure the green development efficiency. The scientificity of the results needs to be improved.

There are relatively few studies on the factors affecting the manufacturing green development efficiency in the Yangtze River Economic Belt, and more scholars pay attention to the effects of different factors on the green development efficiency of the industry, which are mainly reflected in the following five aspects: 1) Technical factor. Zhou and Wu, 2013; Wu and Du, 2018 found that technological progress was an important factor to improve the efficiency of industrial green transformation. Gandhi et al. (2018) determined the drivers of integrated lean and green manufacturing industry by using the ideal solution similarity sorting method (TOPSIS) and simple additive weighting method (SAW). The results showed that top management commitment, technology upgrading, current legislation, green brand image and future legislation were the five driving forces of lean and green manufacturing integration for SMEs in India's manufacturing sector. Singh M. et al. (2020) obtained answers from 36 senior managers of SMEs in different geographical locations in India through the Green Manufacturing Questionnaire. According to the empirical data, the adoption of new technology had become an important factor affecting the business performance of enterprises. Lu and Zhao, 2016 held different views. 2) Economic factor. Huang and Shi, 2015; Ban and Yuan, 2016 believed that the relationship between green development efficiency and per capita income was u-shaped, while Zhang Y. H. et al. (2018) thought that the level of economic development had a positive effect on green development efficiency. 3) Structural factor. Zhang (2016) believed that industrial structure restricted the improvement of green development efficiency, while Hu and Li, 2012 held the opposite view. 4) Factor of opening to the outside world. Yuan and Xie, 2015 had a negative impact on the overall efficiency of industrial green development, while the impact direction of FDI depended on the intensity of environmental control. Han and Lan, 2012; Lin and Liu, 2015; Chen et al. (2017) believed that the degree of opening to the outside world had a significantly positive effect on the efficiency of green industrial growth. 5) Regulatory factors. Zhang and Wang, 2013; Borghesi et al. (2015), Zhao et al. (2018) thought that the support of environmental protection had a significantly positive effect on the efficiency of green development, while Wei and Zheng, 2017 supported that the greater the intensity of environmental protection or market segmentation, the greater the efficiency of green development would be inhibited. Based on the methodology of systematic literature review through the content analysis of literary resources. Vrchota et al. (2020) found that the most often common sustainability outcomes were energy saving, emission reduction, resource optimization, cost reduction, higher economic performance, human resources development, social welfare and workplace safety. Wang et al. (2020) used a spatial-temporal geographical weighted regression (GTWR) model to analyze the regional influencing factors of the high-carbon manufacturing industry. The industrial structure and economic scale were the main reasons for the regional carbon lock-in of the high-carbon manufacturing industry, and the strength of the lock-in had continued to increase. On the premise of estimating the green development efficiency, most scholars have explored the effect of one or fewer factors on the green development efficiency. There are significant differences in the conclusions, and the system of influencing factors needs to be improved.

On the whole, the existing literature has formed a good research foundation, but there are relatively few studies on the manufacturing green development efficiency in the Yangtze River Economic Belt. The research of relevant scholars is divided into the following aspects: 1) Research object. It mainly focuses on macro studies such as national or provincial studies, but there are relatively few micro studies on a certain region, so it is necessary to strengthen the research on green development efficiency of subregion. 2) Research indicator. More attention is paid to capital, human and energy input, while less attention is paid to water resource input. 3) Research method. The traditional DEA model and the undesirable output SBM model are mostly adopted, which fails to further distinguish the decision making unit, resulting in the deviation of the obtained efficiency value, so the research method needs to be improved. 4) Influencing factor. Most scholars explore the effect of one or fewer factors on the efficiency of green development, so it is necessary to construct a systematic and scientific system of influencing factors. To this end, this paper makes improvements from the following three aspects: 1) Fully considering the resource input and undesirable output in the production process, it selects the use of water resource in the manufacturing industry as the input variable and the emissions of waste gas, waste water and solid waste in the

manufacturing industry as the output variables; 2) Based on the panel data of 11 provinces and cities in the Yangtze River Economic Belt from 1998 to 2017, super-SBM is used to measure the manufacturing green development efficiency in the Yangtze River Economic Belt, so as to fully solve the problem that multiple decision-making units may be effective at the same time; 3) It uses the WSR methodology to construct a more scientific influencing factor system of manufacturing green development efficiency and applies Tobit model to analyze the causes of regional differences in the manufacturing green development efficiency. In order to improve the overall efficiency of manufacturing green development and narrow the region differences, the corresponding suggestions are put forward.

# STUDY DESIGN

## SBM Model and Green Development Efficiency Measurement of Manufacturing Industry

Data envelopment method (DEA) is mainly used in domestic and foreign researches on green development efficiency measurement. This method can evaluate the relative efficiency between multiple input and output decision making units, which has the advantages of not setting weights and specific forms of production frontier functions in advance. Traditional DEA models mainly include CCR model and BCC model (Han and Lan, 2012). With the proposal of these two models, DEA method has been continuously improved through the joint efforts of domestic and foreign scholars. After Tone improved DEA method twice, super efficiency SBM model has become an important method to calculate the green development efficiency. This method can not only measure the relaxation variable, but also evaluate the unit with an efficiency value of 1. Therefore, this method is adopted in this paper to measure the manufacturing green development efficiency. The model construction process is as follows:

Suppose the number of decision making units is $n$. In each decision making unit, there are three vectors, namely input, desirable output and undesirable output. The three matrices are $x \in R^q, y^w \in R^{m1}, y^b \in R^{m2}$, we can define the matrix $X$, $Y^w$, $Y^b$, As follows:

$$X = [x_1, \ldots x_n] \in R^{q\times n} > 0$$
$$Y^w = \left[y_1^w, \ldots y_n^w\right] \in R^{m_1\times n} > 0$$
$$Y^b = \left[y_1^b, \ldots y_n^b\right] \in R^{m_2\times n} > 0$$

You can then construct a set of production possibilities that includes the undesirable output $P$,

$$P = \left\{\left(x, y^w, y^b\right)\middle| x \geq X\lambda, y^w \leq Y^w\lambda, y^b \geq Y^b\lambda, \lambda \geq 0\right\}$$

According to the processing method of SBM model, the time-sharing planning form of SBM model (variable return to scale) considering undesirable output is shown in formula:

$$\begin{aligned} \rho = \min \quad & \frac{1 - 1/q\sum_{i=1}^{q} s_i^- / x_{i0}}{1 + 1/m_1 + m_2\left(\sum_{r=1}^{u_1} s_r^w / y_{r0}^w + \sum_{i=1}^{u_2} s_l^b / y_{l0}^b\right)} \\ \text{s.t.} \quad & x_0 = X\lambda + s^- \\ & y_0^w = Y^w\lambda - s^w \\ & y_0^b = Y^b\lambda + s^b \\ & \sum_{i=1}^{n} \lambda_i = 1, s^- \geq 0, s^w \geq 0 s^b \geq 0. \end{aligned} \tag{1}$$

In **Eq. 1**, $x$, $y^w$ and $y^b$ represent input variables, desirable output variables and undesirable output variables of the decision making unit. $s^-, s^w, s^b$ are the slack variables of inputs, desirable and undesirable outputs, $\lambda$ is expressed as the weight. The subscript "0" in the formula is the evaluated unit. The objective function value $\rho$ is the manufacturing green development efficiency. $\rho$ on $S^-$, $S^w$, $S^b$ strictly monotone decreasing, and its values range from 0 to 1. When $\rho = 1$. $S^-$, $S^w$, $S^b$ are 0, the decision-making unit is efficient; When $\rho < 1$, it represents the efficiency loss of the decision-making unit, so it is necessary to make corresponding improvements in input and output. SBM model can set three types: input-oriented, output-oriented and non-oriented. Input-oriented refers to finding the minimum input when the output is guaranteed to be certain. Output-orientation refers to finding the maximum output when the input is guaranteed to certain. Non-guidance refers to the calculation from the perspectives of input and output at the same time, so it is also called input-output bidirectional. This paper will choose the optimal guidance according to the empirical results.

Considering that the SBM model with undesirable output may be effective for multiple decisions at the same time, which makes it difficult to distinguish and arrange these decision making units, this paper uses the super-SBM model with undesirable output to solve problem. A finite set of possibilities excluding DMU ($x_0$,$y_0$) is:

$$P\backslash(x_0, y_0) = \left\{\left(\overline{x}, \overline{y}^w, \overline{y}^b\right)\middle| \overline{x} \geq \sum_{i=1}^{n} \lambda_i x_i,\right.$$
$$\left.\overline{y}^w \leq \sum_{i=1}^{n} \lambda_i y_i^w, \overline{y}^b \geq \sum_{i=1}^{n} \lambda_i y_i^b, \overline{y}^w \geq 0, \lambda \geq 0\right\}$$

The fractional programming form of super-SBM model (variable return to scale) considering undesirable output is shown in formula:

$$\begin{aligned} \rho = \min \quad & \frac{1/q\sum_{i=1}^{q} \overline{x}_i / x_{i0}}{1/m_1 + m_2\left(\sum_{r=1}^{u_1} \overline{y}_r^w / y_{r0}^w + \sum_{i=1}^{u_2} \overline{y}_l^b / y_{l0}^b\right)} \\ \text{s.t.} \quad & \overline{x} \geq \sum_{i=1,\neq 0}^{n} \lambda_i x_i, \\ & \overline{y}^w \geq \sum_{i=1,\neq 0}^{n} \lambda_i y_i^w, \\ & \overline{y}^b \geq \sum_{i=1,\neq 0}^{n} \lambda_i y_i^b, \\ & \overline{x} \geq x_0, \overline{y}^w \leq y_0^w, \overline{y}^b \geq y_0^b \\ & \sum_{i=1,\neq 0}^{n} \lambda_l = 1, \overline{y}^w \geq 0, \lambda \geq 0. \end{aligned} \tag{2}$$

In **Eq. 2**, $\rho\bullet$ is the target efficiency value, and the meanings of other variables are the same as **Eq. 1**.

## TOBIT Model

When the dependent variable is the fragment value or the cut value, the Tobit model should be adopted. This model belongs to the restricted dependent variable model, and the maximum likelihood method is adopted to estimate this model, which can better avoid the problem of inconsistent and biased estimation of parameters. Its mathematical expression is shown in formula:

$$\begin{aligned} \rho_i^* &= \beta_0 + \sum_{j=1}^{l} \beta_j x_{ij} + \varepsilon_i \\ &\begin{cases} \rho_i = \rho_i^* & 0 \le \rho_i^* \le 1 \\ \rho_i = 0 & \rho_i^* < 0 \\ \rho_i = 1 & \rho_i^* > 1 \end{cases} \end{aligned} \tag{3}$$

In **Eq. 3**, $\rho_i^*$ is the latent variable, $\rho_i$ is the observed actual dependent variable, $x_{ij}$ is the independent variable, $\beta_0$ is the constant term, $\beta_j$ is the correlation coefficient vector, $\varepsilon_i$ is independent and $\varepsilon_i \sim N(0, \sigma^2)$.

## Variable Selection and Data Description

### Input-Output Variables

***Input Variables***

① Capital input variable. In the selection of input indicators, most of the relevant literatures are based on the idea of Cobb-Douglas production function and take labor and invested capital as inputs. In the past research, capital investment index is mostly represented by capital stock (Khairunnisa et al., 2015), this paper selects fixed assets investment (unit: one hundred million yuan) of the manufacturing industry to replace the capital stock (Zhang and Zhang, 2003; Guo and Sun, 2013). In order to eliminate the influence of the price factor, the paper sets the year of 1998 for the base period and uses price index of fixed assets investment to convert fixed assets investment into comparable price capital stock. ② Labor input variable. In this paper, labor input is expressed by the number of workers (unit: ten thousand) at the end of the year. ③ Energy input variable. In order to reflect the manufacturing green development efficiency in each province, energy input is also added to the input variable. The energy consumption of manufacturing industry in each province over the years (unit: 10,000 tons of standard coal) is selected to represent the energy input variable. Due to the difficulty in obtaining data, industrial data is used as a substitute. ④ Water resource input variable. In order to highlight the green development, this paper selects the water resource consumption of manufacturing industry (unit: 100 million cubic meters) as the variable of water resource input. As the data is difficult to obtain, it is replaced by industrial data.

***Output Variables***

① Desirable output variable. In relevant studies on the green development efficiency, GDP is generally chosen as output or the index related to the research topic is constructed based on GDP. The desirable output variable refers to the desired output that is beneficial to the overall goal. From the perspective of development, the output value created each year is the most intuitive desirable output of manufacturing industry. Manufacturing added value (unit: 100 million yuan) is adopted as the desirable output in this paper. Due to the difficulty in obtaining data, it is replaced by industrial data. In order to exclude the influence of price factors, the annual price added value is converted into comparable price added value by using the industrial GDP index (1998 = 100).

② Undesirable output variables. An undesirable output is an accompaniment to a beneficial output, which is produced with the beneficial output. When the toxic gases and substances contained in the "three wastes of manufacturing industry" are discharged into the environment, they will not only cause environmental pollution, but also cause air pollution, which will have a serious effect on human health. The discharged substances may have physical and chemical changes, and the harmful substances can enter the human body through different ways, thus harming human health. In order to consider the problem of urban air and environmental pollution caused by the development of manufacturing industry, the three wastes of manufacturing industry are selected as the undesirable output variable. Due to the difficulty in obtaining data, three industrial wastes (the units of discharge of industrial waste gas, waste water and solid waste are 100 million standard cubic meters, 100 million tons and 10 thousand tons respectively) are used to replace the undesirable output variables.

## Influencing Factor Variables

This paper adopts the WSR method (Wuli-Shili-Renli theory system method), and puts forward the factors affecting the manufacturing green development efficiency from three different perspectives: Wuli, Shili and Renli theory. In 1994, Gu Jifa and Zhu Zhichang jointly proposed the WSR system methodology at the university of Hull, United Kingdom (Meng and Zou, 2018). The "principles" in "Wuli", "Shili" and "Renli" are not only the deepening of the understanding of "Wuli", "Shili" and Renli, but also the key to improve the universality of the methodology (Gu et al., 2007). WSR system methodology is a research method combining quantitative and qualitative. Its core idea is to make use of Wuli, Shili and Renli intelligently and rationally to solve problems. When dealing with complex problems, it is necessary to consider the objectivity of the research object (Wuli level), and how to deal with existing problems (Shili level). Then, it is necessary to consider that people, as objects or subjects, always exist in the practice of dealing with problems (Renli level). The essence of WSR system methodology is to coordinate the relationship of "Wuli", "Shili" and "Renli" in the system practice. Earlier before year 2000, WSR methodology was listed as an integrated system methodology by foreign scholars, which had its own uniqueness and traditional Chinese philosophical thinking (Gu, 2011). At present, this method has gradually matured and has been applied in various fields (Gu and Gao, 1998). Specifically, Wuli factors include economic and technological factors; Shili factors include efficiency and structural factors; Renli factors include human and regulatory factors. Economic development level,

urbanization level, openness to the outside world, technology marketization level, foreign investment level are economic and technological factors; energy productivity, labor productivity, r&d input rate, energy consumption structure, industrial structure and property rights structure are efficiency and structural factors; labor force quality, environmental protection strength and energy price are human and regulatory factors. All the selected variables are shown in **Figure 1**.

### Economic Development Level

The level of economic development refers to the size, speed or level of economic development. Generally speaking, regions with a higher level of economic development have a developed economy, a high level of technology, and a relatively high level of development in various industries, making a great contribution to the green development of manufacturing industry. In this paper, per capita GDP is used to represent the level of economic development of provinces and cities, and it is assumed that the level of economic development has a promoting effect on the manufacturing green development efficiency. In this variable, GDP has been converted into comparable GDP by using the GDP deflator (1998 = 100).

### Urbanization Level

Urbanization is an important symbol of the level of modernization. With the acceleration of technological progress and industrialization, non-agricultural industries continue to gather in urban regions, creating many employment opportunities, resulting in the continuous migration of rural population to cities, the continuous expansion of the number and scale of cities, and the continuous spread of urban civilization to the countryside. In general, the population transformation, industrial adjustment, industrial development and scientific and technological progress brought by urbanization will have positive effects on the manufacturing green development efficiency. This paper adopts the proportion of urban population in the total population in each region to represent the urbanization level.

### Openness to the Outside World

The degree of openness mainly measures the degree to which enterprises participate in the international market. The higher the degree of openness to the outside world, the greater the competitive pressure enterprises face, the more advanced management technology and mode they can learn, and thus they produce the significantly positive effect on the manufacturing green development efficiency. In this paper, the proportion of export delivery value in the main business income of manufacturing industry in each region is adopted to represent the degree of openness to the outside world.

### Technology Marketization Level

The level of technology marketization reflects the importance that a region attaches to scientific and technological research and development. The higher the level of technology marketization, the higher the level of manufacturing technology in the region, and the higher the level of pollution control. In this paper, the ratio of technology market turnover to manufacturing added value is used to represent the level of technology marketization, and it is assumed that this variable has a positive effect on the manufacturing green development efficiency.

### Foreign Investment Level

Foreign investment has become an important force to promote China's economic and social development and greatly promote the improvement of manufacturing technology. In order to further promote high-quality development of foreign investment, the second session of the 13th National People's Congress (NPC) passed the law of the People's Republic of China on foreign investment on March 15, 2019, the first comprehensive and systematic legislation on foreign investment in China's history. In this paper, the proportion of foreign investment in GDP by region is used to represent the level of foreign investment, and it is assumed that this variable has a promoting effect on the manufacturing green development efficiency.

### Energy Productivity

Energy productivity can be expressed by the ratio of manufacturing value added to energy use in different regions, which is mainly used to compare the comprehensive utilization efficiency of energy in different regions, reflecting the economic benefits of energy utilization. In recent years, China has introduced a series of energy conservation and emission reduction policies to promote energy technology innovation, which is conducive to improving China's energy productivity. In this variable, manufacturing added value has been converted into comparable added value by using the industrial GDP index (1998 = 100). This paper assumes that this variable has a positive effect on the manufacturing green development efficiency.

### Labor Productivity

Labor productivity is usually determined by the development level of social productivity, and is also affected by various economic and technological factors. In this paper, the ratio of the value added of manufacturing industry to the number of employed people in manufacturing industry in each region is adopted to represent labor productivity, and it is assumed that this variable has a significantly positive effect on the manufacturing green development efficiency. In this variable, manufacturing added value also has been converted into comparable added value by using the industrial GDP index (1998 = 100).

### R&d Input Rate

R&d investment is usually positively correlated with technological progress and has a positive effect on resource efficiency. Advanced technology can promote the development of high-end manufacturing industry. In this paper, the ratio of the internal r&d expenditure of manufacturing industry to the added value of manufacturing industry in each region is adopted as the proxy variable of r&d input rate. Since it is difficult to obtain data on r&d expenditure in manufacturing industry, industrial data are used instead, and it is assumed that this variable has a positive effect on the manufacturing green development efficiency.

### Energy Consumption Structure

China is the world's largest producer and consumer of coal. In 2018, China consumed 1.91 billion tons of standard coal, accounting for 50.5% of global consumption. Excessive consumption of coal will cause serious damage to the ecological environment, so it is urgent to optimize the energy consumption structure. In this paper, the ratio of coal consumption to total energy consumption in manufacturing industry in each region is used to represent the energy consumption structure, and it is assumed that this variable has a significantly negative effect on the manufacturing green development efficiency.

### Industrial Structure

The proportion of different industries is the industrial structure, which directly affects the manufacturing green development efficiency. After the reform and opening up, China's industrial structure is developing toward a higher and reasonable direction, but the proportion of the secondary industry in GDP is still high. In this paper, the ratio of the total output value of heavy industry to the total industrial output value in each region is used to represent the industrial structure, and it is assumed that this variable has a significantly negative effect on the manufacturing green development efficiency.

### Property Rights Structure

The rational operation and coordinated development of modern enterprises mainly depend on the rationalization and scientific property rights structure of enterprises. Some scholars have shown that the operation efficiency of state-owned enterprises is relatively low, which may hinder the green development of the industry. In this paper, the proportion of manufacturing national capital in total paid-in manufacturing capital in each region is used as the proxy variable of property rights structure, and it is assumed that this variable has a negative effect on the manufacturing green development efficiency.

### Labor Force Quality

With the continuous improvement of science and technology, the productivity of enterprises depends more and more on the intelligence level of laborers. The more talents with high knowledge level in the manufacturing industry, the faster the transformation and upgrading speed of enterprises will be. This paper chooses the average education years to measure the labor quality and the calculation method refers to Peng Guohua (Luo et al., 2017). The calculation formula is: the average number of years of education in the workforce = the proportion of working population with illiteracy, semiliterate * 1.5 + the proportion of the working population with primary education * 7.5 + the proportion of the working population with secondary education * 10.5 + the proportion of the working population with a high school education * 13.5 + the proportion of the working population with tertiary education and above * 17 and assumes that the variable has a significantly positive effect on the manufacturing green development efficiency.

### Environmental Protection Strength

Environmental protection can promote enterprises to strengthen the treatment of three wastes and reduce the waste of production capacity, and thus promote the green development of manufacturing industry. In this paper, the ratio of total investment in environmental pollution control to the added value of manufacturing industry in each region is adopted to represent the environmental protection strength, and it is assumed that this variable has a positive effect on the manufacturing green development efficiency.

### Energy Price

Energy price mainly includes the price of coal, oil, natural gas, new energy and other products. Energy price policy can promote lean production of enterprises, encourage using various energy resources reasonably and efficiently, and thus improve the green development level of manufacturing industry. In this paper, the purchasing price index of raw materials, fuels and power is used to represent energy price, and it is assumed that this variable has a significantly positive effect on the manufacturing green development efficiency. Description of influencing factor variables can be found in **Table 1**.

In order to reduce multicollinearity and eliminate the influence of dimension to some extent, the data of the above influencing factors (excluding labor force quality, technology marketization level, foreign investment level, environmental protection strength and r&d input rate) are logarithmized in this paper.

Based on the above analysis, the specific regression equation is shown in formula:

$$\begin{aligned}\rho_{it} = {} & \beta_0 + \beta_1 pcgdp_{it} + \beta_2 urbanization_{it} + \beta_3 openness_{it} \\ & + \beta_4 techno\log y_{it} + \beta_5 fdl_{it} + \beta_6 enprod_{it} + \beta_7 laborp_{it} \\ & + \beta_8 rdir_{it} + \beta_9 enstr_{it} + \beta_{10} gzb_{it} + \beta_{11} prostr_{it} \\ & + \beta_{12} laborq_{it} + \beta_{13} enprotect_{it} + \beta_{14} enprice_{it} + \varepsilon.\end{aligned} \tag{4}$$

In **Eq. 4**, *it* represents the value corresponding to the *i*th province in period *t*, $\rho$ is the manufacturing green development efficiency, $\beta_i$ is the regression coefficient, and $\varepsilon$ is the random interference term.

## Data Description

Considering the availability and effectiveness of data, this article selects 1998–2017 panel data of manufacturing industry of 11 provinces and cities in the Yangtze River Economic Belt as a sample. Data are mainly from various provinces and cities statistical yearbook, China statistical yearbook, China industrial economic statistical yearbook, China environment statistical yearbook, China population statistics yearbook and China energy statistical yearbook.

**TABLE 1 |** Description of influencing factor variables.

| Influencing factor variables | Symbol | Definition | Desirable effect |
|---|---|---|---|
| Economic development level | *pcgdp* | GDP/annual average resident population (yuan/person), 1998 as the base period (conversion basis by GDP deflator) | + |
| Urbanization level | *urbanization* | Urban population/total population (%) | + |
| Openness to the outside world | *openness* | The ratio of export delivery value/main business income of manufacturing industry (%) | + |
| Technology marketization level | *technology* | Technology market turnover/added value of manufacturing industry (%) | + |
| Foreign investment level | *fdl* | Total foreign investment/gross regional product (%) | + |
| Energy productivity | *enprod* | Manufacturing value added/energy use (million yuan/100 tons of standard coal) | + |
| Labor productivity | *laborp* | Manufacturing value added/number of manufacturing employment (million yuan/100 people) | + |
| r&d input rate | *rdir* | Internal expenditure of manufacturing r&d funds/added value of manufacturing industry (%) | + |
| Energy consumption structure | *enstr* | Coal consumption of manufacturing industry/total energy consumption of manufacturing industry (%) | - |
| Industrial structure | *gzb* | Gross output value of heavy industry/gross industrial output value (%) | - |
| Property rights structure | *prostr* | National capital of manufacturing industry/total paid-in capital of manufacturing industry (%) | - |
| Labor force quality | *laborq* | The average number of years of education in the workforce | + |
| Environmental protection strength | *enprotect* | Total investment in environmental pollution control/added value of manufacturing industry (%) | + |
| Energy price | *enprice* | Purchasing price index of raw materials, fuel and power | + |

# EMPIRICAL RESULTS AND ANALYSIS

## Analysis on Green Development Efficiency and Its Regional Differences of Manufacturing Industry

This paper uses MAXDEA 7.6 software to calculate the manufacturing green development efficiency in 11 provinces and cities of the Yangtze River Economic Belt from 1998 to 2017. The distributions of green development efficiency in 1998, 2004, 2011 and 2017 are made, and the distributions in 1998 and 2017 reflect the overall change of green development efficiency in 11 provinces and cities of the Yangtze River Economic Belt. In 2004 and 2011, the intermediate years are equally separated, and they can reflect the steady growth of efficiency values in some provinces and cities and the change of efficiency values in other provinces and cities. The following is the distribution diagrams of green development efficiency made by ARCGIS software in four years. And the whole result can be gained in **Supplementary Appendix Table A1**.

**Figures 2**, **3** and **Supplementary Appendix Table A1** show that the mean green development efficiency of the regions along the Yangtze River Economic Belt is not high, but the overall development trend is rising. It indicates that the manufacturing industry of the Yangtze River Economic Belt is still in a relatively low development stage and the task of transformation and upgrading is still arduous. Among them, the mean values of Shanghai, Jiangsu, Zhejiang and Yunnan are all greater than 1, indicating that the four provinces and cities have taken effective environmental protection measures in the past 20 years. The mean efficiencies of green development in Jiangxi, Anhui, Hunan, Guizhou and other regions are relatively low, which are close to 0.5, accounting for only 1/2 of that in Shanghai and Jiangsu. On the whole, the efficiency of green development in the downstream region is higher than that in the midstream and upstream regions. By analyzing **Figures 2**, **3** and **Supplementary Appendix Table A1**, the following conclusions can be drawn: 1) The efficiency of green development in most regions is lower than one except for some regions. But the overall trend of green development efficiency is rising. In recent years, the government's environmental protection policy of the Yangtze River Economic Belt region is effective and still needs to be carried out; 2) The differences among different regions are obvious. Green development efficiency in the downstream region is higher, green development efficiencies in upstream and downstream regions are low. The efficiencies of Shanghai, Jiangsu and Zhejiang provinces remain above 1 year by year, but efficiencies of Jiangxi, Anhui, Hunan and Guizhou are too low, mainly due to the lower level of manufacturing technology; 3) The high efficiency of green development in the downstream region is mainly due to its high level of economic development, huge manufacturing capital and higher level of scientific and technological research. Most studies show that technological

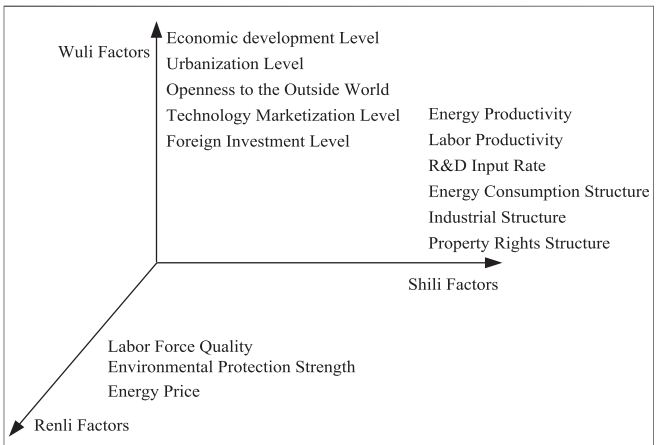



**FIGURE 1 |** 3D diagram of influencing factors of WSR methodology.

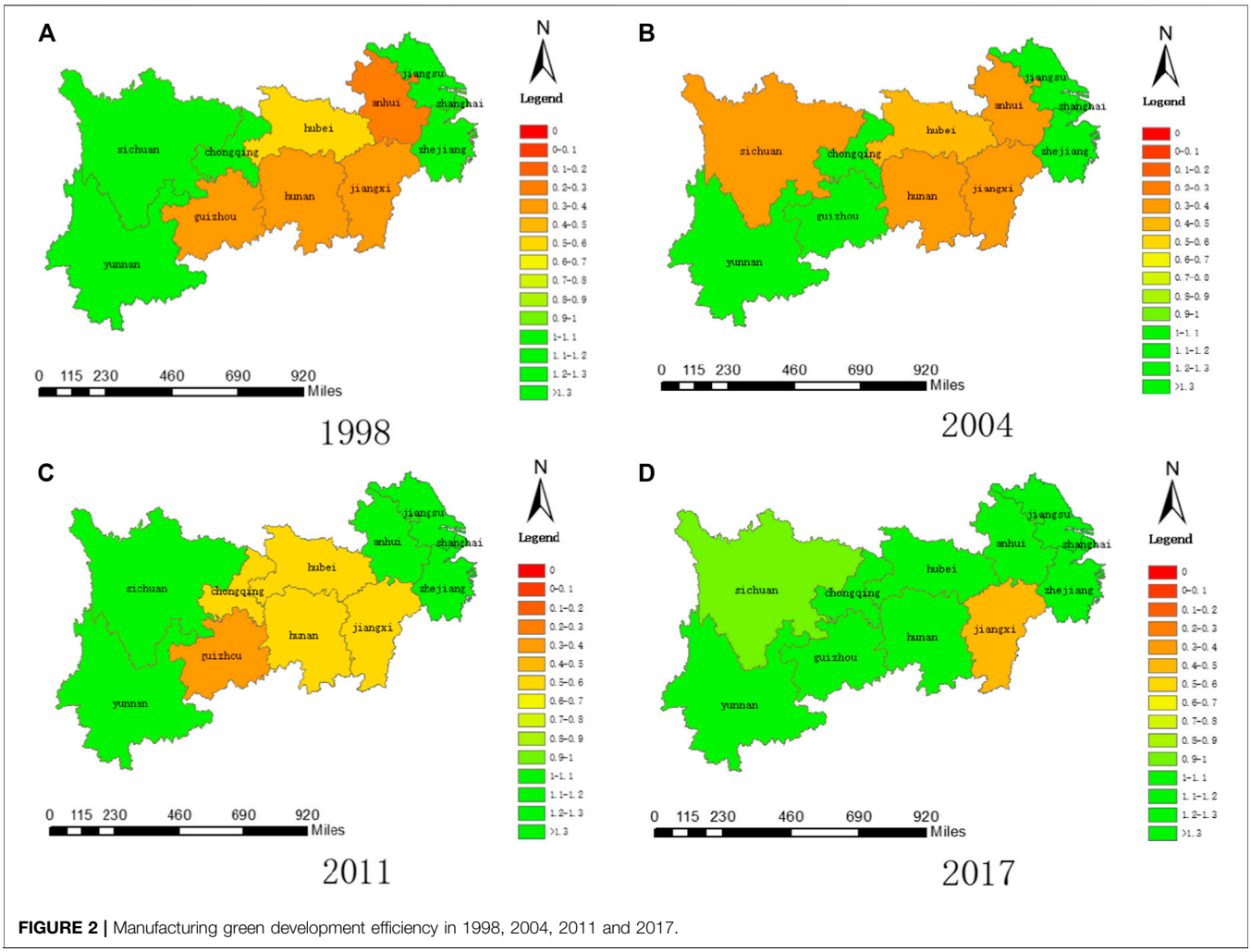



**FIGURE 2 |** Manufacturing green development efficiency in 1998, 2004, 2011 and 2017.

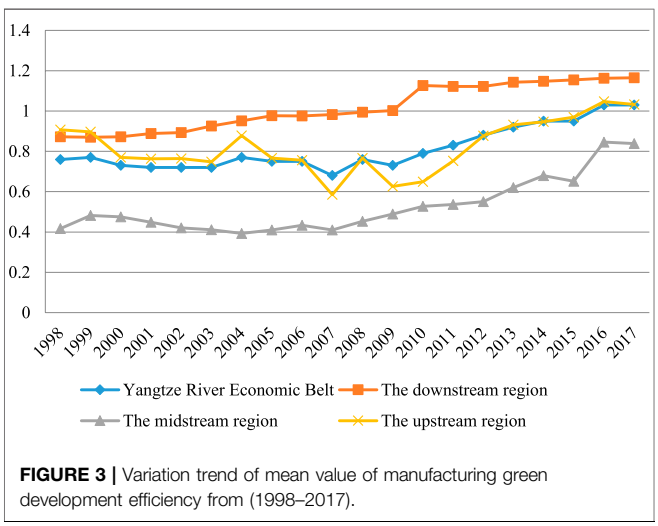



**FIGURE 3 |** Variation trend of mean value of manufacturing green development efficiency from (1998–2017).

level has a huge effect on the green development of manufacturing industry. Therefore, the manufacturing green development efficiency in developed regions is relatively higher.

From the perspective of inter-provincial differences, the mean manufacturing green development efficiency in Jiangsu over the years is 1.156, followed by 1.152 in Shanghai. Regions with higher efficiency, such as Shanghai, Jiangsu and Zhejiang have maintained a relatively stable growth trend. Although the mean value of green development efficiency of Yunnan's manufacturing industry is 1.08, from the perspective of time, it decreases slightly year by year, mainly because the scale of Yunnan's manufacturing industry is relatively small. In January 2015, during his visit to Yunnan, President Xi Jinping put forward the strategic positioning of "building Yunnan into a radiating center for South Asia and southeast Asia". In the same year, Yunnan province proposed to develop eight major industries, including biomedicine, advanced equipment, food and consumer goods processing and other manufacturing industries. As can be seen from **Figures 2**, **3**, the efficiency value of green development of manufacturing industry in the midstream and upstream regions is relatively low, and the rising rate is also relatively low, which forms an obvious contrast with that in the downstream region. The plan for ecological and environmental protection of the Yangtze River Economic Belt also points out that the regional development of the Yangtze River Economic

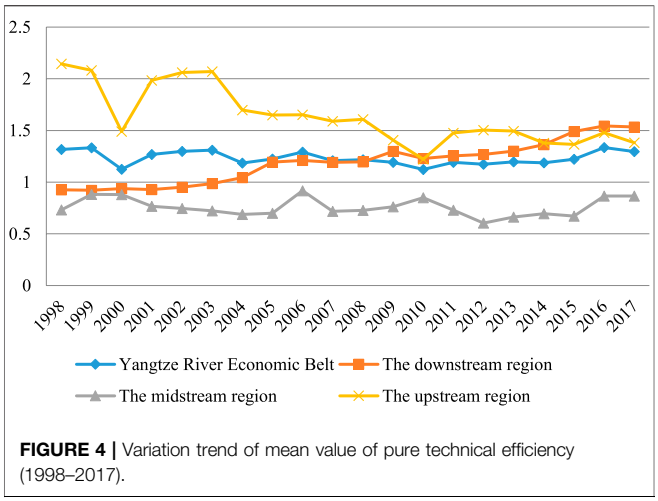



**FIGURE 4 |** Variation trend of mean value of pure technical efficiency (1998–2017).

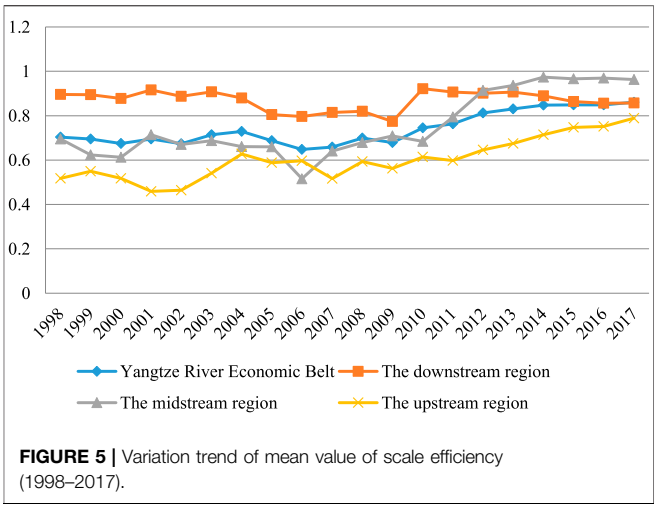



**FIGURE 5 |** Variation trend of mean value of scale efficiency (1998–2017).

Belt is unbalanced, most regions still practice the traditional extensive development mode, and the industrial isomorphism in the upstream, midstream and downstream regions is increasingly prominent. Some polluting enterprises are gradually shifting to the midstream and upstream regions. Although this has increased employment and output value in the midstream and upstream regions, environmental pollution in some provinces such as Jiangxi and Guizhou has increased, thus slowing down the growth rate of manufacturing green development efficiency in the midstream and upstream regions.

In this paper, green development efficiency is decomposed into pure technical efficiency and scale efficiency. The results are shown in **Figures 4**, **5**. During the statistical period, the comprehensive efficiency value of the downstream region is significantly higher than that of the upstream region, mainly because the scale efficiency value of the downstream region is larger than that of the upstream region, that is, the manufacturing scale advantage of the downstream region is larger than that of the upstream region. However, it cannot be ignored that the pure technical efficiency value of the downstream region is slightly lower than that of the upstream region, indicating that the upstream region puts more emphasis on technology and management level rather than scale when developing manufacturing industry. In undertaking the transfer of industries from the downstream region, the midstream region should focus on improving the quality of manufacturing industry and make full use of the benefits of economies of scale. While expanding the enterprise scale, the upstream region should continue improving the technology and management level.

## TOBIT Regression Analysis

With the help of EViews9.0 software and panel Tobit model, this paper conducts an empirical analysis on the factors influencing the manufacturing green development efficiency in the upstream, midstream and downstream regions of the Yangtze River Economic Belt. The specific effect of each variable is shown in **Table 2**. **Table 2** shows that the coefficient of each factor variable is different from 0, and most variables reject the assumption that the value is 0 at the corresponding significance level, and the data fitting is good.

### Wuli Factors Analysis

Per capita GDP has the largest effect on the manufacturing green development efficiency in the Yangtze River Economic Belt, with an influence coefficient of 0.474, which is consistent with the expectation. But influence coefficients of per capita GDP in upstream, midstream and downstream regions are −0.529, −0.025 and −0.086 respectively, the causes of this phenomenon lies in that the rapid economic development usually destroys the resources and environment beyond the capacity of the environment, thus reducing the efficiency of green development of the manufacturing industry. Therefore, although the overall economic development plays a positive role in the manufacturing green development efficiency, different regions also need to consider that the economic development should match the environmental carrying capacity.

The influence coefficients of urbanization level on the green development efficiencies of manufacturing industry in the Yangtze River Economic Belt and the upstream, midstream and downstream regions are negative, which are −0.073, −0.293, −0.179 and −0.019 respectively. It is inconsistent with the expectation. On the one hand, the development of urbanization will indeed bring the positive effect of reducing transaction costs and increasing economies of scale, but on the other hand, the technical knowledge level of farmers still needs to be improved, which will affect the quality of human capital. At the same time, the development of urbanization will inevitably lead to the expansion of city scale, thus consuming a large amount of resources, which may be the main reason why the influence coefficient of urbanization level is negative.

Openness to the outside world has a positive effect on the green development efficiencies of manufacturing industry in whole basin, upstream and downstream regions, which is in line with the expectation. Although openness to the outside world has a positive effect on the upstream and downstream regions, the effect is not significant, indicating that manufacturing exports of the Yangtze River Economic Belt are still relatively

**TABLE 2 |** Regression results of influencing factors of Yangtze River Economic Belt and its sub-regions.

| Variables | Influence coefficient | | | |
|---|---|---|---|---|
| | Yangtze river economic belt | The upstream | The midstream | The downstream |
| *pcgdp* | 0.474*** (0.000) | −0.529*** (0.008) | −0.025 (0.790) | −0.086 (0.312) |
| *urbanization* | −0.073 (0.465) | −0.293** (0.025) | −0.179 (0.529) | −0.019 (0.813) |
| *openness* | 0.186*** (0.000) | 0.104 (0.133) | −0.045 (0.395) | 0.062** (0.038) |
| *technology* | 0.059*** (0.000) | 0.043** (0.074) | 0.026*** (0.005) | 0.014 (0.276) |
| *fdl* | −0.001 (0.9472) | −0.003 (0.8146) | −0.003 (0.7681) | 0.023*** (0.001) |
| *enprod* | −0.038 (0.604) | 0.768*** (0.000) | 0.266*** (0.001) | 0.740** (0.000) |
| *laborp* | 0.114*** (0.000) | 0.329*** (0.000) | 0.018 (0.660) | −0.008 (0.6360) |
| *rdir* | −0.040** (0.027) | −0.071 (0.168) | 0.092*** (0.000) | −0.016** (0.044) |
| *enstr* | −0.101** (0.014) | 0.030 (0.768) | 0.001 (0.982) | −0.013 (0.635) |
| *gzb* | −0.153 (0.132) | 0.597*** (0.006) | −0.150** (0.071) | −0.003 (0.9498) |
| *prostr* | 0.048 (0.216) | 0.207*** (0.002) | −0.026 (0.469) | −0.031 (0.185) |
| *laborq* | −0.223*** (0.000) | 0.170** (0.035) | −0.004 (0.919) | 0.019 (0.495) |
| *enprotect* | −0.016 (0.121) | −0.071*** (0.000) | −0.004 (0.818) | 0.024** (0.034) |
| *enprice* | −0.455** (0.059) | −0.079 (0.852) | 0.007 (0.971) | 0.015 (0.901) |
| Cons | 1.261 (0.373) | −0.762 (0.765) | 1.092 (0.273) | −1.303 (0.103) |

*Note:*** and ** mean that the influence coefficient is significant at the level of 1% and 5% respectively, and p value is in brackets.*

small, which also confirms that the Yangtze River Economic Belt should still increase the export volume of manufacturing industry and reduce the proportion of energy-intensive products. For the midstream region, the influence coefficient of openness to the outside world is -0.045, which is not in line with the expectation. This may be due to the high energy consumption products in the exports of the midstream region, leading to the low green development efficiency of the manufacturing industry.

Technology marketization level has a positive effect on the green development efficiencies of the manufacturing industry in the Yangtze River Economic Belt and the upstream, midstream and downstream regions, which are 0.059, 0.043, 0.026 and 0.014 respectively. It is consistent with the expectation. Technology marketization level will directly affect the level of technology. According to existing studies, technology level can promote the manufacturing green development efficiency, and the use of technology innovation can solve the problem of resource constraint in the green transformation of manufacturing industry (Peng, 2005).

Foreign investment level has a negative effect on the green development efficiencies of the manufacturing industry in the Yangtze River Economic Belt and the midstream and upstream regions, with the influence coefficients of −0.001, −0.003 and −0.003, which are relatively small, while the influence coefficient on the manufacturing green development efficiency in the downstream regions is 0.023. The downstream region is located in the coastal region, which is more closely connected with the international market and more convenient for foreign investment. Therefore, the foreign investment level in the downstream region has a significant effect on the manufacturing green development efficiency.

## Shili Factors Analysis

Energy productivity has a positive effect on the green development efficiencies of manufacturing industry in the upstream, midstream and downstream regions, with influence coefficients of 0.768, 0.266 and 0.740. This shows that the energy productivity has a significantly positive effect on the green development efficiencies of manufacturing industry in all regions. The energy productivity has no significantly negative effect on the green development efficiency of the overall manufacturing industry in the Yangtze River Economic Belt, which may be related to the rebound effect of energy consumption. In general, the manufacturing industry should improve energy productivity, thus promoting the green development of the manufacturing industry (Zhang and Song, 2019).

Labor productivity has a positive effect on the green development efficiencies of manufacturing industry in the Yangtze River Economic Belt, the midstream and upstream regions, with influence coefficients of 0.114, 0.329 and 0.018 respectively. Labor productivity directly affects the production efficiency of enterprises. The improvement of labor productivity will promote the improvement of industrial capacity utilization rate and energy utilization rate, which plays a good role in promoting the manufacturing green development efficiency. Labor productivity has no significantly negative effect on the manufacturing green development efficiency in the downstream region of Yangtze River Economic Belt. Labor productivity is mainly affected by the quality and technical level of laborers. Although labor productivity is relatively high in the downstream region, the high labor cost and large investment in technology research and development affect the manufacturing green development efficiency.

R&d input rate has a negative effect on the green development efficiencies of the manufacturing industry in the Yangtze River Economic Belt, the upstream and downstream regions, with influence coefficients of −0.040, −0.071 and −0.016. The research and development investment rate has less negative effect on the downstream region, but more negative effect on the upstream region and the whole Yangtze River Economic Belt. The reason is that the cost of r&d is huge and the technological

innovation dividend brought by r&d investment is not enough to cover the cost of r&d (Wang et al., 2013). The influence coefficient of r&d input rate on the manufacturing green development efficiency in the midstream regions is 0.092, which is related to the relatively low r&d input rate in the midstream region, highlighting the benefits brought by r&d input to some extent.

Energy consumption structure has a negative effect on the green development efficiencies of the manufacturing industry in the Yangtze River Economic Belt and the downstream region, which is in line with the expectation. The influence coefficients are −0.101 and −0.013. The energy consumption structure has a positive effect on the midstream and upstream regions, with influence coefficients of 0.03 and 0.001. For the midstream and upstream regions, the cost of coal acquisition is low and the economic benefits of coal are relatively obvious. However, for the green development of manufacturing industry, midstream and upstream regions should reduce the use of coal and increase the use of clean energy.

Industrial structure has a negative effect on the green development efficiencies of the manufacturing industry in the Yangtze River Economic Belt, midstream and downstream regions, with influence coefficients of −0.153, −0.150 and −0.003, which are in line with the expectation. The industrial structure has a positive effect on the manufacturing green development efficiency in the upstream region, with an influence coefficient of 0.597. The reason is that the economic benefits of the development of heavy industry in the upstream region outweigh the negative environmental effects. The government should formulate strict emission regulations for heavy industry to promote green and sustainable development.

Property rights structure has a negative effect on the green development efficiencies of the manufacturing industry in the midstream and downstream regions, with influence coefficients of −0.026 and −0.031, but they are not significant. The property rights structure has a positive effect on the green development of manufacturing industry in the Yangtze River Economic Belt and the upstream region, with influence coefficients of 0.048 and 0.207. The effect in the upstream region is more significant. It can be seen that the injection of state capital can promote the expansion of upstream enterprises, thus generating economies of scale. Therefore, the higher the proportion of state capital in total paid-in capital, the more conducive to improving the manufacturing green development efficiency.

### Renli Factors Analysis

Labor force quality has a positive effect on the green development efficiencies of manufacturing industry in upstream and downstream regions, with influence coefficients of 0.170 and 0.019, which are consistent with the expectation. Generally speaking, the improvement of labor quality can improve employees' awareness of environmental protection, thus promoting the green development of manufacturing industry. The influence of labor quality on the green development of manufacturing industry in the Yangtze River Economic Belt and the midstream region is negative, with influence coefficients of −0.223 and −0.004, which are inconsistent with the expectation. On the whole, the role of labor quality in promoting the manufacturing green development efficiency in Yangtze River Economic Belt is still not significant, so it is necessary to improve the quality and quantity of talent introduction in manufacturing industry.

Environmental protection strength has a promoting effect on the manufacturing green development efficiency in the downstream region, with an influence coefficient of 0.024, but it is not significant. Environmental protection has a negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the upstream and midstream regions. This is mainly because the environmental damage of the Yangtze River Economic Belt is serious and the initial treatment cost is too large, which has exceeded the environmental benefits it brings. This also shows that the government should continue strengthening environmental protection efforts and improving the environmental protection mechanism.

Energy price has a significantly negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt, with an influence coefficient of −0.455, which is not in line with the expectation, while the effect on the other three regions is not significant. This shows that energy price has different effects on the manufacturing green development efficiency in different regions. The increase in energy price will not significantly reduce the use of energy consumption in the short term, but will significantly increase the production costs of manufacturing enterprises, thus reducing the economic benefits of manufacturing enterprises. Therefore, when formulating energy price policies, the government should pay more attention to the long-term effect of energy price and give full play to the promoting role of energy price in the green development of manufacturing industry.

## CONCLUSIONS AND POLICY RECOMMENDATIONS

In this paper, based on super-SBM model, the fixed assets, labor, energy and water resource of manufacturing industry are set as input variables, the added value of manufacturing industry is desirable output variable, and the emissions of "three wastes of industry" are undesirable output variables. Manufacturing green development efficiencies of 11 provinces and cities in the Yangtze River Economic Belt from 1998 to 2017 are calculated. On the basis of Wuli-Shili-Renli system methodology, the influencing factors of economic development level, urbanization level, and openness to the outside world and so on are selected. It applies a limited dependent variable panel Tobit model to analyze the factors affecting the manufacturing green development efficiency. The results show that there are significant differences in the manufacturing green development efficiency in different provinces and cities. Specific conclusions are as follows:

(1) As a whole, during the statistical period, the mean manufacturing green development efficiency in the provinces and cities of the Yangtze River Economic Belt over the years is 0.812, which is at the medium level of

development. The whole Yangtze River Economic Belt has a "U" shaped distribution, that is, the midstream low and the two sides high. From the comprehensive efficiency, the downstream region is higher than the upstream region, and the upstream region is higher than the midstream region. Therefore, from the overall perspective, optimizing the enterprise management model, vigorously developing energy-saving and emission reduction technologies, and formulating appropriate environmental protection policies by the government according to the actual situation are effective measures to improve the manufacturing green development efficiency. The manufacturing industry in the midstream and upstream regions should learn advanced technology and management experience from the downstream region, introduce a large number of high-tech talents, improve the infrastructure and supporting facilities of the manufacturing industry, implement the brand building strategy and enhance the public service capacity. From the results of comprehensive efficiency decomposition, the manufacturing green development efficiency in the Yangtze economic belt presents an overall rising trend. Low scale efficiency is the main reason that restricts the manufacturing green development efficiency. The scale efficiency of the downstream region is significantly higher than that of the upstream region, while the pure technical efficiency of the downstream region is slightly lower than that of the upstream region. Downstream region should continue learning and innovating excellent management models to ensure the quality of manufacturing green development; while expanding the scale of enterprises, the midstream and upstream regions should focus on the quality of manufacturing development and achieve sustainable development.

(2) From the regression results of influencing factors, the effect of economic development level on the manufacturing green development efficiency in the Yangtze River Economic Belt is the most significantly and positively correlated. The urbanization level is negatively correlated with the manufacturing green development efficiency in the Yangtze River Economic Belt and various regions. The openness to the outside world has a positive effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the upstream and downstream regions, but a negative effect on the midstream regions. The technology marketization level has a significant and positive influence on the Yangtze River Economic Belt and all regions. The foreign investment level has a significantly positive effect on the downstream region, but a significantly negative effect on other regions. The energy productivity has a significantly positive effect on the upstream, midstream and downstream regions. The labor productivity has a significantly positive effect on the Yangtze River Economic Belt and the upstream region. The r&d input rate has a negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the upstream and downstream regions. The energy consumption structure has a negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the downstream region. The industrial structure has a negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the midstream and downstream regions. The property rights structure has a negative effect on the manufacturing green development efficiency in the midstream and downstream regions, but a positive effect on the Yangtze River Economic Belt and the upstream region. The labor force quality has a positive effect on the manufacturing green development efficiency in the upstream and downstream regions, but a negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt and the midstream region. The environmental protection strength has a significantly positive effect on the manufacturing green development efficiency in the downstream region, while it has a significantly negative effect on the upstream region. The energy price has a significantly negative effect on the manufacturing green development efficiency in the Yangtze River Economic Belt.

It can be seen that various influencing factors have different effects on the manufacturing green development efficiency in different regions. In order to improve the manufacturing green development efficiency and narrow regional differences, Wuli factors should promote high-quality economic development, improve the quality of urbanization, reduce the proportion of energy-intensive products in exports, improve the level of technology marketization and attract high-quality foreign investment; Shili factors should improve the level of energy technology, heighten labor productivity, increase investment in research and development, optimize the energy consumption structure, adjust the structure of industry and property rights; Renli factors should cultivate high-quality personnel, strengthen emission management and rationalize the energy price system. Each region should also formulate corresponding policies based on the specific effects of various influencing factors.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: China statistical yearbook, China industrial economic statistical yearbook, China environment statistical yearbook, China's population statistics yearbook and China energy statistical yearbook from 1998 to 2017.

## AUTHOR CONTRIBUTIONS

GL: Conceptualization, Methodology, Software; YZ: Data curation, Visualization, Investigation; FL: Software, Validation, Writing-Original draft preparation; TW: Reviewing and Editing.

## FUNDING

This paper is funded by the Key Projects of National Social Science Fund (No.16AJL008), National Natural Science Foundation of China (No.71503106, No. 71803197), Jiangsu Social Science Fund (No.19EYB016), Jiangsu College Philosophy Social Science Outstanding Innovation Team Construction Project, CSSC Modern Logistics Research Center Project (No.2045081603), Ministry of Education Humanities and Social Sciences Foundation (No. 18YJC630094), Fundamental Research Funds for the Central Universities (No. 31511910801), Soft Science Project 563 of Technological Innovation in Hubei Province (No. 2019ADC029).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2020.631911/full#supplementary-material.

## REFERENCES

Ban, L., and Yuan, X. L. (2016). Differences in the efficiency of green economy in eight regions of China and spatial influence mechanism. *J. Xi'an Jiaot. Univ.* 36 (3), 22–30. doi:10.15896/j.xjtuskxb.201603004

Borghesi, S., Cainelli, G., and Mazzanti, M. (2015). Linking emission trading to environmental innovation: evidence from the Italian manufacturing industry. *Res. Pol.* 44 (3), 669–683. doi:10.1016/j.respol.2014.10.014

Chen, N., Xu, L., and Chen, Z. (2017). Environmental efficiency analysis of the yangtze river economic zone using super efficiency data envelopment analysis (sedea) and tobit models. *Energy* 134 (9), 659–671. doi:10.1016/j.energy.2017.06.076

Chen, S. Y. (2010). China's green industrial revolution: an explanation from the perspective of environmental total factor productivity (1980–2008). *Econ. Res. J.* 45 (11), 21–34+58.

Duman, Y. S., and Kasman, A. (2018). Environmental technical efficiency in EU member and candidate countries: a parametric hyperbolic distance function approach. *Energy* 147, 297–307. doi:10.1016/j.energy.2018.01.037

Gandhi, N. S., Thanki, S. J., and Thakkar, J. J. (2018). Ranking of drivers for integrated lean-green manufacturing for Indian manufacturing SMEs. *J. Clean. Prod.* 171, 675–689. doi:10.1016/j.jclepro.2017.10.041

Gu, J. F., and Gao, F. (1998). On the system methodology of physics-matter-human theory from the perspective of management science. *Syst. Eng.-Theory and Pract.* 18 (8), 2–6.

Gu, J. F. (2011). Practice of the methodology of physical and human systems. *Chin. J. Manag.* 8 (3), 317–322+355. doi:10.3969/j.issn.1672-884X.2011.03.001

Gu, J. f., Tang, X. J., and Zhu, Z. X. (2007). A review of physics-matter-theory system methodology. *J. Trans. Syst. Eng. Infor. Tech.* 7 (6), 51–60. doi:10.3969/j.issn.1009-6744.2007.06.007

Guo, W., and Sun, T. (2013). Research on ecological total factor energy efficiency of Chinese industry. *Chin. J. Manag.* 10 (11), 1690–1695. doi:10.3969/j.issn.1672-884x.2013.11.018

Han, J., and Lan, Q. X. (2012). Measurement of industrial greening degree in China and research on influencing factors. *China Population, Resour. Environ.* 22 (5), 101–107.

Hu, D. S., and Li, Y. (2012). Regional differences in environmental efficiency assessment and its influencing factors. *Finance Econ.* 56 (4), 116–124. doi:10.3969/j.issn.1000-8306.2012.04.014

Huang, Y. C., and Shi, Q. P. (2015). Research on regional environmental efficiency and total environmental factors in China -- based on the analysis of SBM model including R&D input. *China Population, Resour. Environ.* 25 (12), 25–34. doi:10.3969/j.issn.1002-2104.2015.12.004

Jiang, J., and Qu, L. (2020). Evolution and emerging trends of sustainability in manufacturing based on literature visualization analysis. *IEEE Acc.* 8, 121074. doi:10.1109/ACCESS.2020.3006582

Jing, W. M., and Zhang, L. (2014). Environmental regulation, opening up and the progress of green technology in Chinese industry. *Econ. Res. J.* 49 (9), 34–47.

Khairunnisa, A. R., Yusof, M. Z. M., Salleh, M. N. M., and Leman, A. M. (2015). The development of energy efficiency estimation system (eees) for sustainable development: a proposed study. *Energy Proc.* 79, 513–519. doi:10.1016/j.egypro.2015.11.527

Li, G., Liu, J. G., and Li, T. Q. (2019). Study on regional differences of energy ecological efficiency in manufacturing industry with undesired output considered - a two-stage analysis based on SBM and Tobit models. *Chinese Journal of Management Science* 27 (11), 76–87. doi:10.16381/j.cnki.issn1003-207x.2019.11.008

Li, J., and Lin, B. (2017). Ecological total-factor energy efficiency of China's heavy and light industries: which performs better?. *Renew. Sustain. Energy Rev.* 72 (2), 83–94. doi:10.1016/j.rser.2017.01.044

Li, L., and Zhang, J. (2016). Differences and decomposition of industrial green development in the Yangtze River Economic Belt - a comparative study of 108 cities from 2004 to 2013. *Soft Sci.* 30 (11), 48–53. doi:10.13956/j.ss.1001-8409.2016.11.11

Lin, B. Q., and Liu, H. C. (2015). Whether foreign trade is conducive to improving energy and environmental efficiency - a case study of China's industrial sector. *Econ. Res. J.* 50 (9), 127–141.

Lu, W., and Zhao, Y. F. (2016). Study on regional environmental efficiency evaluation and influencing factors in China. *J. Beijing Univ. Aeronaut. Astronaut.* 29 (3), 30–35. doi:10.13766/j.bhsk.1008-2204.2015.0127

Luo, J. Q., Li, W. P., Zhao, Y. P., and Cheng, F. X. (2017). Research on service derivative status of manufacturing enterprises based on WSR and its evaluation. *Manag. Rev.* 29 (6), 129–140.

Meng, F. S., and Zou, Y. (2018). Dynamic evaluation of energy eco-efficiency based on PP-SFA—take the 30 autonomous regions of China as an example. *Syst. Eng.* 36 (5), 47–56.

Nukman, Y., Farooqi, A., Al-Sultan, O., Alnasser, A. R. A., and Bhuiyan, M. S. H. (2017). A strategic development of green manufacturing index (GMI) topology concerning the environmental impacts. *Procedia engineering* 184, 370–380. doi:10.1016/j.proeng.2017.04.107

Peng, G. H. (2005). Analysis of income gap, total factor productivity and its convergence in China. *Econ. Res. J.* 15 (9), 19–29.

Singh, C.Singh, D., and Khamba, J. S. (Forthcoming 2020). Understanding the key performance parameters of green lean performance in manufacturing industries. *Mater. Today: Proceedings* doi:10.1016/j.matpr.2020.06.328

Singh, M.Singh, K., and Singh Sethi, A. (Forthcoming 2020). A pilot study on relation between various input parameters of green manufacturing and business performance parameters in Indian SME's. *Mater. Today: Proceedings* 33, 1493. doi:10.1016/j.matpr.2020.02.752

Sun, C. Z., Jiang, K., and Zhao, L. S. (2017). Measurement and spatial pattern of green efficiency of water resource in China. *J. Nat. Resour.* 32 (12), 1999–2011. doi:10.11849/zrzyxb.20161076

Vrchota, J., Pech, M., Rolínek, L., and Bednář, J. (2020). Sustainability outcomes of green processes in relation to industry 4.0 in manufacturing: systematic review. *Sustainability* 12 (15), 5968. doi:10.3390/su12155968

Wang, K. L., Liu, Y., Shi, L. J., Liu, L., Meng, X. R., and Yang, B. C. (2017). Spatio-temporal differentiation and influencing factors of green water resource efficiency in the Yangtze River Economic Belt industry-two-stage analysis based on ebm-tobit model. *Resour. Sci.* 39 (8), 1522–1534. doi:10.18402/resci.2017.08.09

Wang, K. L., Yang, L., Yang, B. C., and Cheng, Y. H. (2013). Energy economic efficiency, energy environmental performance and regional economic growth. *Journal of Management Science* 26 (3), 86–99. doi:10.3969/j.issn.1672-0334.2013.03.009

Wang, X., Zhang, L., Qin, Y., and Zhang, J. (2020). Analysis of China's manufacturing industry carbon lock-in and its influencing factors. *Sustainability* 12 (4), 1502. doi:10.3390/su12041502

Wang, Z. P., Tao, C. Q., and Shen, P. Y. (2014). Research on regional green technology efficiency based on ecological footprint and its influencing factors. *China Population, Resources and Environment* 24 (1), 35–40. doi:10.3969/j.issn.1002-2104.2014.01.006

Wei, C., and Zheng, X. Y. (2017). A new perspective on energy efficiency - based on market segmentation test. *Soc. Sci. China* 38 (10), 90–111+206.

Wu, C. Q., and Du, Y. (2018). Research on the influence of biased technological progress on total factor energy efficiency of the Yangtze River Economic Belt. *China Soft Sci.* 33 (3), 110–119. doi:10.3969/j.issn.1002-9753.2018.03.011

Wu, Y. Z., and Wen, Y. C. (2013). Industrial green productivity, emission reduction performance and emission reduction cost in China. *Sci. Res. Manag.* 34 (2), 105–111+151.

Yi, M., Li, G., Peng, J. C., and Chen, W. L. (2018). Spatiotemporal differentiation of green total factor productivity in the Yangtze River Economic Belt. *Manag. World* 34 (11), 178–179. doi:10.3969/j.issn.1002-5502.2018.11.018

Yuan, Y. J., and Xie, R. H. (2015). FDI, environmental regulation and China's industrial green total factor productivity growth: an empirical study based on the Luenberger index. *J. Inter. Trade* 41 (8), 84–93.

Zhang, F., and Song, X. N. (2019). Decoupling and equilibrium relationship between environmental regulation, resource endowment and green growth of manufacturing. *Science of Sci. Manag. S.& T.* 40 (4), 32–47.

Zhang, J.Qu, X., and Sangaiah, A. K. (2018). A study of green development mode and total factor productivity of the food industry based on the industrial Internet of things. *IEEE Commun. Mag.* 56 (5), 72–78. doi:10.1109/MCOM.2018.1700789

Zhang, J. X., and Wang, X. W. (2013). Study on the green growth index of China's regional industry and its influencing factors. *Soft Science* 27 (10), 92–96. doi:10.3969/j.issn.1001-8409.2013.10.020

Zhang, J., and Zhang, Y. (2003). Re-estimation of capital stock K in China. *Econ. Res. J.* (7), 35–43.

Zhang, W. (2016). Ways to realize green and low-carbon modern industrial system and influencing factors. *Sci. Res. Manag.* 37 (S1), 426–432.

Zhang, Y. H.Chen, J. L., and Cheng, Y. (2018). Research on the influence mechanism of environmental regulation on the efficiency of regional green economy in China - empirical analysis based on super-efficiency model and spatial panel econometric model. *Resour. Environ. Yangtze Basin* 27 (11), 2407–2418. doi:10.11870/cjlyzyyhj201811002

Zhao, X., Liu, C., and Yang, M. (2018). The effects of environmental regulation on China's total factor productivity: an empirical study of carbon-intensive industries. *J. Clean. Prod.* 179, 325–334. doi:10.1016/j.jclepro.2018.01.100

Zhou, W. Q., and Wu, G. (2013). Differences in sources of green TFP growth and its impact on industrial carbon productivity - an empirical analysis based on China's inter-provincial panel data. *Modern Finance and Economics-Journal of Tianjin University of Finance and Economics* 33 (12), 38–47. doi:10.19559/j.cnki.12-1387.2013.12.004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Li, Zhou, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 15 March 2021
doi: 10.3389/fenrg.2021.655006

# Analyzing Information Disclosure in the Chinese Electricity Market

*Luosong Jin[1], Weidong Liu[1], Xiangyang Wang[1], Jing Yu[2] and Panting Zhao[3]**

[1] Zhejiang Electricity Trading Center Co., Ltd., Hangzhou, China, [2] Zhejiang Huayun Mdt Info Tech Ltd., Hangzhou, China, [3] China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, China

Electricity generation relies heavily on fossil fuels in China, and this has posed great challenges for sustainable development. In 2015, China launched a new series of power reforms, aimed primarily at sustainable development and building a competitive power market system, where an information disclosure system plays an important role. This paper analyzes the effects of sensitive information disclosure, constructs different information disclosure scenarios, and compares the market clearing results under different scenarios. The results show that information transparency is conducive to the promotion of market efficiency. However, some problems, especially collusion, arise and inevitably bring negative impacts to the power market. Therefore, more attention should be paid to the content and quality of data transparency to avoid manipulations in the market. The data of power systems are complex and various, and thus big data applications may be conducive to effective information disclosure and better market regulation. Moreover, disclosure delay will help the electricity market become more transparent by reducing the risk of collusion. Besides, the information's scope and contents, timing should also be taken into consideration. These findings may provide some references for the sustainable development of the electricity industry and have certain policy implications for policymakers.

**Keywords: sustainable development, electricity market, information, disclosure rules, collusion risk**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Ailun Wang,*
*Southwestern University of Finance and Economics, China*
*Houyin Long,*
*Fuzhou University, China*

****Correspondence:***
*Panting Zhao*
*rainbowzpt@163.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *18 January 2021*
***Accepted:*** *24 February 2021*
***Published:*** *15 March 2021*

***Citation:***
*Jin L, Liu W, Wang X, Yu J and Zhao P (2021) Analyzing Information Disclosure in the Chinese Electricity Market. Front. Energy Res. 9:655006. doi: 10.3389/fenrg.2021.655006*

## INTRODUCTION

With the increasing development of the economy, the demand for energy sources increases continually (Song et al., 2020). Sustainable and green development has become a major concern for all industries worldwide (Rane and Thakker, 2019). Environmental pollution brought about by massive fossil energy burning has posed great challenges for sustainable development (Song et al., 2019). In China, coal is the most important element of electricity production and consumption in response to energy endowment characterized by "rich in coal, poor in oil, and little in gas." The energy structure of electric power causes many environmental issues; however, the cost of clean power is much higher than that of thermal power, leading to difficulties in power supply structure adjustment (Xie and Zhao, 2018; Khan, 2020). Increasingly strict environmental supervision and energy policy adjustment have put forward requirements for sustainable electricity supply, while traditional power networks have been unable to meet these requirements. In the era of big data, with the attention of the national power sector, the power industry is also developing toward informatization. With the continuous improvement of the power grid, the degree of integration of cloud technology and other emerging technologies with the power industry is continuously strengthened, and data in the power industry has also witnessed a substantial growth. Big data applications bring opportunities and challenges for the sustainable development of the power industry.

In 2015, China launched a new round of power reforms, aimed primarily at sustainable development of the electricity industry and building an "open, orderly, competitive, and complete" power market system (Song and Cui, 2016; Liu et al., 2019). To facilitate the implementation of this reform, a series of supplementary actions have been taken, for example the construction of an electric power trading center and eight pilots for electric power spot transaction. The spot market is an important part of the electricity market system, which plays a fundamental supporting role in coordinating market transactions and ensuring system security. With the advent of electricity spot trading, the transaction mode has been changed and transaction frequency has increased substantially. Consequently, the electricity information system has become a necessity for the electricity market.

In 2020, the National Energy Administration released a draft regulation entitled "Measures for Information Disclosure in the Electricity Spot Market," which stipulates the principles, methods, and contents of information disclosure. The electricity market is complex, where the information disclosure mechanism plays an important role in promoting market efficiency (Jin et al., 2020). Power generators, retailers, grid enterprises, and consumers are important participants in the electricity market. Thus, these players are also the main subjects for information disclosure. The foundation of ensuring effectiveness and enhancing competition in the electricity market depends on the fulfilling of obligations and responsibilities of market participants. The information disclosure mechanism is also of immense significance for market fairness and transparency as it reduces information asymmetry.

Information will be updated when it flows into the market. The participants will act on this new information and make decisions with more accurate analysis and prediction. Thus, high quality electricity information is the basis that forms effective pricing mechanisms. It is believed that the advanced information disclosure regime will be conducive to increasing transparency, improving market competition, reducing transaction costs, and ensuring electricity security (Healy and Palepu, 2001; Hooper et al., 2009; Delmas et al., 2010; Niefer, 2014).

Fairness, justice, and openness are the basic principles of the electricity market, whereas severe information asymmetry lies in the power sector which hinders the establishment of an effective electricity market. The effective operation and market development need to maintain a certain degree of transparency for the discovery of market inequities and inefficiencies. At present, the information system has not been fully explored, though research on this topic has gradually increased since the new electricity reform. It is worth noting that ways in which the information disclosure system could be perfected is still being explored in China.

Information in the electric power market can be divided into public information, open information, and private information. However, the electricity markets in China are still in their infancy, with an information disclosure system where basic trading information and practical operating conditions are the main focus. With the development and perfection of the market, more information could do with transparency, such as the bidding strategies used by power generators. However, sometimes problems will arise when sensitive information becomes open, for example, tacit collusion. That is, frequent interaction of bidders will increase their chances of transmitting information, and this interaction also gives them more opportunities to learn how to cooperate, which is particularly likely to lead to tacit collusion between power generators.

As mentioned above, it is widely recognized that the information disclosure mechanism plays an important role in the electricity market. However, the study on information disclosure is not enough, especially for the extent and quality of optimal market data transparency in the context of new electricity reform in China. Based on the game theory, this paper mainly focuses on the effects of disclosing sensitive information, taking power generators' bidding data as an example, at the early stage of market liberalization. How to use big data to further perfect the information disclosure mechanism has also been discussed in the era of big data. These findings may help to provide some references for further development of information disclosure systems and electricity reform.

This paper is organized as follows. Section "Literature Review" reviews the related literature. Section "Model Description" presents model constructs. In section "Model Analysis," we analyze and compare the optimal bidding strategies of power generators under different scenarios. Section "Further Discussion" discusses how to further perfect the information disclosure mechanism with the help of big data application. Section "Conclusions and Policy Implications" concludes.

## LITERATURE REVIEW

The electricity industry has the characteristic of a natural monopoly, which reaps a good deal of benefits and also results in operational inefficiency (Peng and Tao, 2018). The electricity market reform in China can be traced to the 1980s, while progress was very slow until 2002 (Wang and Chen, 2012). In the 2002 reform, the state grid company was split into five generation enterprises and two grid enterprises to break the monopoly, aimed at improving the generation efficiency and sustainable development of the electricity industry (Meng et al., 2016; Zeng et al., 2016; Deng et al., 2018). However, there still exists some problems in the electricity sector, for example, the regulated price and weak competition between two grid companies (Yu, 2012). For the further development of the electricity sector, China launched a new round of electricity reforms in 2015, mainly focusing on the retail market (Lin and Purra, 2019; Lin et al., 2019). More participants are encouraged to take part in the electricity market to introduce the competition mechanism into the sell side (Peng and Tao, 2018). The essence of this new reform is to form a competitive mechanism and thus to reduce electricity costs for other industries and improve operational efficiency (Yao et al., 2019).

With further introduction of a competition mechanism, the electricity trading system in China has been gradually perfected and market efficiency has improved considerably (Meng et al., 2016; Li et al., 2017). Regional electric power transaction markets have been established in some provinces, such as Zhejiang and

Guangdong. In today's marketization trend for power industries, information flow becomes particularly important. To be more specific, the information disclosure mechanism is essential to the efficiency and fairness of the whole power market (Yang et al., 2018). Markard and Holt (2003) find that critical information disclosure will help to improve competition in the electricity market by increasing consumer's ability to choose electricity offerings. Darudi et al. (2015) quantitatively measure the impacts of the transparency of bidding data of generating enterprises on unilateral exercises of market power and short-term market price. Their results on an actual market (Alberta) indicate that inappropriate bidding disclosure may allow power generators to increase price significantly, which in turn increases end user consumers' expenses. Brown et al. (2018) examine the role of information transparency in Alberta's wholesale electricity market and find that certain firms would respond to rival offer changes with a lag by using data on the bidding behavior of firms. Yang et al. (2018) review the information disclosure mechanism of several typical electricity markets and stress that timely and precise information should be disclosed to the market participants through a centralized and authorized information disclosure mechanism. Jin et al. (2020) investigate and compare the power retailers' information disclosure strategies under new electricity reforms in China and find that information non-transparency will inevitably result in market inefficiency.

Some literature has also focused on information disclosure and environmental performance and finds that mandatory information disclosure may be a vital component for improving the utilization of clean energy. For example, as stressed by Boardman and Palmer (2007), electricity labels which provide information about the source and the electricity environmental effects are unique opportunities for consumers to transfer suppliers to more sustainable sources of electricity. Delmas et al. (2010) find that mandatory information disclosure programs in the electricity market will achieve the targeted goals, which lead to a reduction in the average proportion of fossil fuels and increasing utilization of clean fuels. Bae and Yu (2018) examine the relationship between mandatory disclosure and states' adoption of renewable energy policy, and they stress that mandatory disclosure on fuel fix and greenhouse gas emissions have significantly positive effects.

In the previous research, it is widely recognized that information disclosure mechanisms play an important role in operational efficiency in the electricity market. However, information disclosure in some literature is sometimes referred to as environmental disclosure and few studies have discussed the extent and quality of optimal market data transparency of China's information disclosure mechanism under their new electricity reforms. The information disclosure system is still in its infancy, and thus this paper mainly focuses on the effects of disclosing sensitive information, taking power generators' bidding strategies for example, which are considered as private information within the existing disclosure rules. Moreover, we also discuss the feedback of "private information becoming open," and the way to perfect the information disclosure system with the help of big data applications, in order to promote the sustainable development of information disclosure and the whole industry.

# MODEL DESCRIPTION

The electricity market is as complex as the product is intangible. With the development of power markets and electricity reformation, the information disclosure system will be perfected. We adopt the game model that is often applied to the analysis of market participants' behavior. We assume that there are two power generators in this market: firm 1 and firm 2. The production cost of firm $i$ ($i \in \{1, 2\}$) is assumed as $C_i = a_i + bq_i + 0.5c_iq_i^2$ (a similar assumption could be seen in Liu et al., 2019), where $q_i$ is the product quantity of firm $i$, $a_i$ is the fixed costs, $b$ and $c_i$ are parameters related to the variable cost.

According to the trading rules, the power generator quotes the quantity and related prices to the trading center. However, the consumers only declare their needs. We assume the market demand is $p = D - \lambda(q_1 + q_2)$, where $D$ is the basic electricity demand and $\lambda$ is the sensitivity to the price. The pricing strategy of firm $i$ is assumed as $q_i = \beta_i(p_i - b)$, where $\beta_i$ is the parameter for optimal pricing strategy. The market will be cleared in terms of power generators' bidding strategies and electricity demand. After that, the market clearing price is decided and requirements for firm $i$ will be settled finally, which is $q_i = \beta_i(P - b)$. When the electricity market clears, there is $Q = q_1 + q_2$.

The revenue of firm $i$ is $\pi_i(q_i) = Pq_i - C_i(q_i)$, where $P$ is clearing pricing and $C_i$ is production cost. What should be noted is that power producers will decide their bidding strategies in response to their rivals' behavior. Under different information disclosure mechanisms, the information obtained by power producers is different, so the optimal bidding strategies also differ, leading to various market clearing results. The notations of variables are as shown in **Table 1**.

# MODEL ANALYSIS

To compare bidding strategies of power generators under different cases, we first formulate basic models with and without full private information disclosure.

## Information Disclosed (ID)

According to the trading rules, the trading center announces the demand and requests each power generator to make a quotation. It is assumed that private information is fully disclosed in this scenario. The power generator formulates a quotation strategy

**TABLE 1 |** The notations of variables.

| Notations | Definition |
|---|---|
| $P$ | Market clearing price |
| $q$ | Production quantity |
| $a$ | Fixed costs |
| $b$ | Parameters related to the variable cost |
| $c$ | Wholesale price of the distributor |
| $\beta_i$ | Parameter for optimal pricing strategy |
| $D$ | Basis of the electricity demand |
| $\lambda$ | Sensitivity to the price. |

in response to electricity demand and market information. After the quotation, the market clearing price will be decided with the balance of electricity supply and demand. Finally, the power generators charge the market clearing prices and are settled with respect to its quotation strategy.

When formulating bidding strategies, power generators will determine optimal policy based on different market conditions, comprehensively considering market information and the behavior of competitors. That is, as the market is becoming transparent, the power generators will choose to act on this new information. Under this case, the pricing strategies of firm 1 and firm 2 can be shown as $q_1 = \beta_1(p - b)$ and $q_2 = \beta_2(p - b)$. And thus, we can obtain the market clearing price, which is a function of power generators' pricing parameters.

$$P = \frac{D + \lambda\beta_1 b + \lambda\beta_2 b}{\lambda\beta_1 + \lambda\beta_2 + 1} \tag{1}$$

According to Eq. (1), the demand for firm 1 can be shown as follows.

$$q_1 = \beta_1 \frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1} \tag{2}$$

And then, we can have the profit of firm 1.

$$\max_{\beta_1} \pi_1 = (\beta_1 - 0.5c_1\beta_1^2)(\frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1})^2 - a_1 \tag{3}$$

The first order derivatives of $\pi_m$ with respective to $\beta_1$ is expressed as follows.

$$\frac{\partial \pi_1}{\partial \beta_1} = \frac{(D - \lambda b)^2}{(\lambda\beta_1 + \lambda\beta_2 + 1)^3}(-\lambda\beta_1 + \lambda\beta_2 + 1 - \lambda c_1\beta_1\beta_2 - c_1\beta_1)$$

With some algebra, we can obtain the optimal pricing parameter for firm 1.

$$\beta_1 = \frac{1 + \lambda\beta_2}{\lambda + \lambda c_1\beta_2 + c_1} \tag{4}$$

Similarly, the optimal pricing parameter for firm 2 is shown as Eq. 5.

$$\beta_2 = \frac{1 + \lambda\beta_1}{\lambda + \lambda c_2\beta_1 + c_2} \tag{5}$$

Based on the analysis above, we find that the optimal pricing strategies of power generation enterprises will meet the condition of Eqs. 4, 5. By solving these two equations, the market clearing prices and quantity can be settled. We will carry out simulation analysis to prove this in the following part. Also, we have $\frac{\partial P}{\partial \beta_i} > 0$, $\frac{\partial q_i}{\partial \beta_i} < 0$; this indicates that the higher optimal pricing parameter, the higher the market clearing price and the lower the production quantity, leading to low market efficiency.

## No Information Disclosed (NID)

According to the pricing strategies of the power generators, the relationship between the market clearing electricity price and the quotation strategy can be obtained. However, the assumption of full information disclosure cannot be satisfied in reality. When the information is missing or the decision-making cost is too high, the power generation company will ignore the influence of the competitors. Therefore, we assume private information is not disclosed in this part and discuss the consequences. In this case, the optimal pricing parameter only concerns cost parameters and price sensitivity. Hence, the pricing strategies for firm 1 and 2 are shown as follows.

$$\beta_1 = \frac{1}{\lambda + c_1} \tag{6}$$

$$\beta_2 = \frac{1}{\lambda + c_2} \tag{7}$$

Accordingly, we can obtain the market clearing price and production quantity.

$$P = \frac{D + \lambda\beta_1 b + \lambda\beta_2 b}{\lambda\beta_1 + \lambda\beta_2 + 1} \tag{8}$$

$$q_1 = \beta_1 \frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1} \tag{9}$$

$$q_2 = \beta_2 \frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1} \tag{10}$$

Where $\beta_1$ and $\beta_2$ satisfy Eqs. 6, 7. Through comparison, we can find that the market efficiency and social welfare have been improved by information disclosure, with lower market clearing price and higher production quantity.

The electricity market is similar to the ordinary trading market to some extent, but it has certain particularities at the same time. The entire market constitutes a system network of electrical information trading. The importance of information is also apparent. The key information mainly includes electricity supply and demand, grid congestion, electricity price, load, etc. The player in the market will act on the information disclosed by other participants. Effective disclosure of the private information plays a key role in improving market efficiency, reducing transaction costs, and increasing electricity quantity.

## Information Disclosed With Collusion (IDC)

Collusion means that there is an explicit or implicit non-competitive agreement between rivals to increase bid prices for higher market prices (Esmaeili Aliabadi et al., 2016). In the repeated quotation market, power generators have the motivation to form a tacit collusion in order to avoid "price wars." The supplier can gain more with a high enough market clearing price, though they end up selling less (Palacio, 2020). Disclosure of quotation information will exacerbate the risk of collusion. Under the case of tacit collusion, the total profits of firm 1 and firm 2 can be shown as follows.

$$\max_{\beta_1,\beta_2} \pi = (\beta_1 - 0.5c_1\beta_1^2)\left(\frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1}\right)^2 - a_1 + (\beta_2 - 0.5c_2\beta_2^2)(\frac{D - \lambda b}{\lambda\beta_1 + \lambda\beta_2 + 1})^2 - a_2 \tag{11}$$

We can obtain the first order derivatives of π with respective to $\beta_1$ and $\beta_2$ which are shown as follows.

$$\frac{\partial \pi}{\partial \beta_1} = \frac{(D-\lambda b)^2}{(\lambda\beta_1+\lambda\beta_2+1)^3} (-\lambda\beta_1+\lambda\beta_2+1-c_1\lambda\beta_1\beta_2-c_1\beta_1+\lambda c_2\beta_2^2)$$

$$\frac{\partial \pi}{\partial \beta_2} = \frac{(D-\lambda b)^2}{(\lambda\beta_1+\lambda\beta_2+1)^3} (-\lambda\beta_2+\lambda\beta_1+1-c_2\lambda\beta_1\beta_2-c_2\beta_2+\lambda c_1\beta_1^2)$$

Using algebra, the optimal bidding strategy parameters of power generators can be calculated. The optimal bidding strategies would satisfy the following formula at equilibrium.

$$\beta_1(\lambda+\lambda c_1\beta_2+c_1) = 1-\lambda\beta_2+\lambda c_2\beta_2^2 \quad (12)$$

$$\beta_2(\lambda+\lambda c_2\beta_1+c_2) = 1-\lambda\beta_1+\lambda c_1\beta_1^2 \quad (13)$$

## Information Disclosed With Time Delay

When it comes to the case that private information is disclosed with time delay, the power market efficiency will be improved inevitably. Disclosure delay combines the advantages of information disclosed and no information disclosed while improving the market transparency and avoiding the possibility of tacit collusion at the same time. It can be inferred from the ID and NID cases that when private information in the power market is disclosed with time delay, power generators cannot fully predict their rivals' behavior but can make their decisions based on the historical information which partially reflects the market conditions. Therefore, we can draw the conclusion that market efficiency will be greatly improved with disclosure delay. We will explore how to prove this using model analysis in our future research.

## Simulation and Comparison

We assume that the production cost for power generators is as shown in **Table 2**. The demand function is $p = 500 - 0.05Q$. Based on market demand and production costs, **Table 3** presents market clearing results under different scenarios. Case 1 shows the scenario with complete information disclosure, while case 2 is the scenario when the information is missing and the power generation company neglects the strategy of rival enterprises. In addition, case 3 presents the scenario when there are collusions among bidders in the electricity market.

By comparing, it is found that the clearing price is highest under case 3, while the quantity is the lowest. Information disclosure will make things better, resulting in a lower price and higher quantity compared with case 2. In the case of collusion, the profit of the manufacturer is greatly improved compared to the case without collusion, thus there is a good reason to believe that the power generation enterprises have strong collusion motivations. However, the collusion between power generators will result in the loss of social welfare and market inefficiency, i.e., a higher price and lower quantity.

**TABLE 2 |** Cost parameters of power generators.

| Firm | C | b | a |
|---|---|---|---|
| 1 | 0.21 | 150 | 1,000 |
| 2 | 0.24 | 150 | 900 |

**TABLE 3 |** Market clearing results under different scenarios.

| | | Case 1 | Case2 | Case3 |
|---|---|---|---|---|
| **Price** | | **215.14** | **217.06** | **243.22** |
| Quantity | Firm 1 | 258.03 | 257.91 | 265.90 |
| | Firm 2 | 231.22 | 231.23 | 221.94 |
| | Sum | 489.24 | 489.15 | 487.84 |
| Revenue | Firm 1 | 881.77 | 931.05 | 1636.3 |
| | Firm 2 | 769.66 | 813.97 | 1382.8 |
| | Sum | 1651.4 | 1745.0 | 3019.1 |

As mentioned above, the essence of electricity industry reform is to reduce the energy costs for manufacturing and other industries by introducing competition into this sector. Moreover, the government has taken many measures to lower electricity prices in recent years and will continue to explore ways to achieve this goal (Yang and Faruqui, 2019). We find that the disclosure of some sensitive information will improve the market operational efficiency. However, some problems arise at the same time. Collusion among players will inevitably have negative impacts on fairness and efficiency.

Currently, the construction of the spot market is still at an early stage, thus the required information which needs to be disclosed is just basic information. With the progress of marketization, there must be a better way to solve this dilemma. Disclosure delay may help turn the electricity market into a more transparent one by reducing the collusion risk. In addition to the principles, methods, and contents of information disclosed, timing also should be taken into consideration.

# FURTHER DISCUSSION

As China's economy steps into a new normal phase, its growth has slowed down in recent years. The electricity industry, a national fundamental industry, is of vital importance to the sustainable development of the national economy. The essence of China's electricity reform is to improve operational efficiency by introducing a competition mechanism, thus reducing energy costs for other industries and finally boosting the economic activities. To facilitate the implementation of the new reform in 2015, a series of supplementary actions have been taken, where the construction of an information disclosure system is of vital importance.

Market information is necessary for players in the electricity trading market to make decisions. A large amount of data will be generated during power generation and consumption. An adequate disclosure principle is essential to the perfection of market mechanisms. Market operation is efficient only when the decisions of participants are made with complete information. Missing information or information asymmetry will lead to

resource misallocation and market inefficiency, which do great harm to both the developing market and future market. Under the current system, the power grid enterprises are the main links between power producers and end-consumers, which makes the transaction price unable to reflect the market demand and transmit effective signals to power generation enterprises and investors correctly. The formation mechanism of the electricity price and the corresponding process are hidden from consumers, which leads to information asymmetry and is closely related to resource allocation. Failure of information transmission would increase the risk in market transactions, posing a big obstacle to the construction of an "open, competitive, and orderly" electricity market.

If the information disclosed is incomplete, for example, information on the bidding strategies of rivals, then the decision of power generators will not conform to the actual situation, leading to higher market clearing prices and lower production quantity, i.e., low social welfare. Information cost is a major component of transaction cost which includes the negotiation cost paid by the market participants due to the lack of relevant information. Meanwhile, effective information is also conducive for regulatory agencies to reduce corresponding supervision costs.

Information unfairness is an obstacle to ensure the efficiency of the electricity market. For example, a power generator can obtain more information on supply, demand, price, and so on to make its own decisions better. To sum up, the development and perfection of the information disclosure mechanism has a long-term impact on market efficiency and fairness. Therefore, it is of great significance for the power market to perfect the current information disclosure mechanism and improve the quality of circulating information.

The power trading center platform and data service website are the main platforms for power information disclosure. However, the data of the power system is various and complex. Different sources would make it more difficult for market participants to obtain useful information. Thus, a unified platform of information disclosure is needed. Moreover, big data mining may be conducive to effective information disclosure and better regulation.

Information can be divided into public information, open information, and private information with the current information disclosure rules. Private information is required to be disclosed to specific players and regulators. Based on the analysis above, we find that the disclosure of some private information or sensitive information will improve the market operational efficiency. On the other hand, the disclosure of private information will also create some problems, for example, collusion of generators' quotations. Therefore, how to deal with this private information is a problem that needs to be solved. Information disclosure delay may be a better way to solve this dilemma, which can refine information in the market and reduce collusion risk at the same time. It should be stressed here that electricity spot system and information disclosure system are still in their infancy. Consequently, the improvement of the electricity market and information disclosure timing are worthy of our focus.

## CONCLUSION AND POLICY IMPLICATIONS

In recent years, China's power industry has been moving toward informatization and stepping into the "big data era." Power industry data is experiencing blowout growth. The importance of data in the power industry has become increasingly prominent. Based on the game theory, this paper analyzes the effects of sensitive information disclosure and compares the market clear results under different scenarios. The results show that information transparency is conducive to the promotion of market efficiency, i.e., lower market clearing prices and higher production quantity. However, some problems, for example, collusion, arise and will inevitably have negative impacts. More attention should be paid to the content and quality of data transparency to avoid manipulations in the market. What's more, disclosure delay may help make the electricity market more transparent by reducing the collusion risk. In addition to the scope and contents of information disclosed, timing also should be taken into consideration. These findings may provide some references for the future development of information disclosure systems and electricity reforms.

In summary, information transparency has a long-term positive impact on market efficiency and fairness by reducing information asymmetry. Therefore, improving the current information disclosure mechanism is of great significance to the power market. The power trading center platform and data service website are the main platforms for power information disclosure. Various sources may lead to difficulty in accessing information. Thus, a unified platform of information disclosure is urgently needed. And the main information disclosed by electric power generation companies mainly includes basic information, such as company name, equity status, unit number, capacity, power generation business license, energy consumption level, etc. The data of power systems are complex and various; big data technology can effectively collect and process massive amounts of power data and display them in a centralized manner, so as to explore the value of the data and provide guidance for related work. Therefore, big data applications will be useful for effective information disclosure and better regulation.

The theory of information disclosure is based on mature markets, but China's power market is still in its infancy. In the initial stage, this market faces various unexpected risks, thus, the smooth operation is currently the primary goal. The basic electricity information may be enough as some participants would unreasonably use the sensitive or private information to form a conspiracy, damaging market efficiency. To reduce operational risks caused by system weakness, the information disclosed should meet the present stage's need. With the development and perfection of the market, information disclosure by all market players must be complete. For some sensitive information, disclosure delay may be a better way to solve this dilemma, which can refine market information and reduce collusion risk at the same time.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

LJ and WL: conceptualization, methodology, and writing—original draft. XW: software, data curation, writing–review, and editing. JY and PZ: conceptualization, methodology, writing–original draft, writing—review, and editing, supervision. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China [Grant Nos. 71874149 and 71934001] and National Social Science Fund of China [Grant No. 20ZDA084].

## REFERENCES

Bae, H., and Yu, S. (2018). Information and coercive regulation: the impact of fuel mix information disclosure on states' adoption of renewable energy policy. *Energy Policy* 117, 151–159. doi: 10.1016/j.enpol.2018.03.010

Boardman, B., and Palmer, J. (2007). Electricity disclosure: the troubled birth of a new policy. *Energy Policy* 35, 4947–4958. doi: 10.1016/j.enpol.2007.04.021

Brown, D. P., Eckert, A., and Lin, J. (2018). Information and transparency in wholesale electricity markets: evidence from Alberta. *J. Regul. Econ.* 54, 292–330. doi: 10.1007/s11149-018-9372-z

Darudi, A., Moghadam, A. Z., and Bayaz, H. J. (2015). "Effects of bidding data disclosure on unilateral exercise of market power," in *Proceedings of the 2nd International Congress on Technology, Communication and Knowledge (ICTCK 2015)*, Mashhad. doi: 10.1109/ICTCK.2015.7582641

Delmas, M., Montes-Sancho, M. J., and Shimshack, J. P. (2010). Information disclosure policies: evidence from the electricity industry. *Econ. Inq.* 48, 483–498. doi: 10.1111/j.1465-7295.2009.00227.x

Deng, N., Liu, L., and Deng, Y. (2018). Estimating the effects of restructuring on the technical and service-quality efficiency of electricity companies in China. *Util. Policy* 50, 91–100. doi: 10.1016/j.jup.2017.11.002

Aliabadi, D. E., Kaya, M., and Şahin, G. (2016). Determining collusion opportunities in deregulated electricity markets. *Electr. Power Syst. Res.* 141, 432–441. doi: 10.1016/j.epsr.2016.08.014

Healy, P. M., and Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature. *J. Account. Econ.* 31, 405–440. doi: 10.1016/S0165-4101(01)00018-0

Hooper, L., Twomey, P., and Newbery, D. (2009). *Transparency and Confidentiality in Competitive Electricity Markets. Energy and Infrastructure Division of the Bureau of Europe and Eurasia.* Working Paper. Washington, DC: United States Agency for International Development.

Jin, L., Chen, C., Wang, X., Yu, J., and Long, H. (2020). Research on information disclosure strategies of electricity retailers under new electricity reform in China. *Sci. Total Environ.* 710:136382. doi: 10.1016/j.scitotenv.2019.136382

Khan, I. (2020). Sustainability challenges for the south Asia growth quadrangle: a regional electricity generation sustainability assessment. *J. Clean. Prod.* 243:118639. doi: 10.1016/j.jclepro.2019.118639

Li, H., Kopsakangas-savolainen, M., Xiao, X., and Lau, S. (2017). Have regulatory reforms improved the efficiency levels of the Japanese electricity distribution sector? A cost metafrontier-based analysis. *Energy Policy* 108, 606–616. doi: 10.1016/j.enpol.2017.06.032

Lin, J., Kahrl, F., Yuan, J., Liu, X., and Zhang, W. (2019). Challenges and strategies for electricity market transition in China. *Energy Policy* 133:110899. doi: 10.1016/j.enpol.2019.110899

Lin, K., and Purra, M. (2019). Transforming China's electricity sector: politics of institutional change and regulation. *Energy Policy* 124, 401–410. doi: 10.1016/j.enpol.2018.07.041

Liu, H., Zhang, Z., Chen, Z., and Dou, D. (2019). The impact of China's electricity price deregulation on coal and power industries: two-stage game modeling. *Energy Policy* 134:110957. doi: 10.1016/j.enpol.2019.110957

Markard, J., and Holt, E. (2003). Disclosure of electricity products-lessons from consumer research as guidance for energy policy. *Energy Policy* 31, 1459–1474. doi: 10.1016/S0301-4215(02)00201-X

Meng, M., Mander, S., Zhao, X., and Niu, D. (2016). Have market-oriented reforms improved the electricity generation efficiency of China's thermal power industry? An empirical analysis. *Energy* 114, 734–741. doi: 10.1016/j.energy.2016.08.054

Niefer, M. J. (2014). Information and competition in electric power markets: is transparency the holy grail. *Energy* 35:375.

Palacio, S. M. (2020). Predicting collusive patterns in a liberalized electricity market with mandatory auctions of forward contracts. *Energy Policy* 139:111311. doi: 10.1016/j.enpol.2020.111311

Peng, X., and Tao, X. (2018). Cooperative game of electricity retailers in China's spot electricity market. *Energy* 145, 152–170. doi: 10.1016/j.energy.2017.12.122

Rane, S. B., and Thakker, S. V. (2019). Green procurement process model based on blockchain-IoT integrated architecture for a sustainable business. *Manage. Environ. Qual.* 31, 741–763. doi: 10.1108/MEQ-06-2019-0136

Song, M., and Cui, L. (2016). Economic evaluation of Chinese electricity price marketization based on dynamic computational general equilibrium model. *Comput. Ind. Eng.* 101, 614–628. doi: 10.1016/j.cie.2016.05.035

Song, M., Fisher, R. J., and Kwoh, Y. (2019). Technological challenges of green innovation and sustainable resource management with large scale data. *Technol. Forecast. Soc. Change* 144, 361–368. doi: 10.1016/j.techfore.2018.07.055

Song, M., Zhu, S., Wang, J., and Zhao, J. (2020). Share green growth: regional evaluation of green output performance in China. *Int. J. Prod. Econ.* 219, 152–163. doi: 10.1016/j.ijpe.2019.05.012

Wang, Q., and Chen, X. (2012). China's electricity market-oriented reform: from an absolute to a relative monopoly. *Energy Policy* 51, 143–148. doi: 10.1016/j.enpol.2012.08.039

Xie, B., and Zhao, W. (2018). Willingness to pay for green electricity in Tianjin. China: based on the contingent valuation method. *Energy Policy* 114, 98–107. doi: 10.1016/j.enpol.2017.11.067

Yang, Y., Bao, M., Ding, Y., Song, Y., Lin, Z., and Shao, C. (2018). Review of information disclosure in different electricity markets. *Energies* 11:3424. doi: 10.3390/en11123424

Yang, Y., and Faruqui, A. (2019). Reducing electricity prices and establishing electricity markets in China: dos and don'ts. *Electr. J.* 32:106633. doi: 10.1016/j.tej.2019.106633

Yao, X., Huang, R., and Du, K. (2019). The impacts of market power on power grid efficiency: evidence from China. *China Econ. Rev.* 55, 99–110.

Yu, Y. (2012). How to fit demand side management (DSM) into current Chinese electricity system reform. *Energy Econ.* 34, 549–557. doi: 10.1016/j.eneco.2011.08.005

Zeng, M., Yang, Y., Wang, L., and Sun, J. (2016). The power industry reform in China 2015: policies, evaluations and solutions. *Renew. Sustain. Energy Rev.* 57, 94–110. doi: 10.1016/j.rser.2015.12.203

**Conflict of Interest:** LJ, WL, and XW were employed by company Zhejiang Electricity Trading Center Co., Ltd. JY was employed by company Zhejiang Huayun Mdt Info Tech Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Jin, Liu, Wang, Yu and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

METHODS
published: 30 March 2021
doi: 10.3389/fenrg.2021.638437

# Dynamic Risk Assessment of the Overseas Oil and Gas Investment Environment in the Big Data Era

*Xuqiang Duan*[1], *Xu Zhao*[1], *Jianye Liu*[2], *Shuquan Zhang*[1] *and Dongkun Luo*[1]*

[1] China University of Petroleum, Beijing, China, [2] State Grid Energy Research Institute, Beijing, China

Our research aims to analyze how the uncertainties and risks of the overseas oil & gas investment environment change over time and reveal the specific occurrence probabilities of risk on different levels. In the process of long-drawn overseas oil & gas investment that can last for 30 years or longer, it is difficult for investment decision-makers to grasp the occurrence probabilities and trends of some specific risks accurately and in a timely manner. The overseas risk assessment system has made great progress; however, it has remained elusive due to the challenge of too many complex and interweaved factors. With the advent of big data and artificial intelligence, more precise and specific risk evaluations can be conducted. Our research selects 25 indicators from six dimensions and applies a Cloud parameter Bayesian network algorithm to dynamically assess the oil and gas overseas investment risk of 10 countries. The results reveal how risk dynamics have changed over the past two decades. Our research may serve as a reference in future overseas oil & gas investment risk decision-making, and is also significant to outbound investing, engineering, and service projects. The proper use of risk assessment results can be conducive to potential investors who may invest in potential countries in the future.

**Keywords: overseas oil and gas investment, dynamic, risk assessment, CPBN, big data**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Muhammad Mohsin,*
*Jiangsu University, China*
*Ehsan Rasoulinezhad,*
*University of Tehran, Iran*

***Correspondence:**
*Dongkun Luo*
*chouyunfat@163.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *06 December 2020*
***Accepted:*** *22 February 2021*
***Published:*** *30 March 2021*

***Citation:***
*Duan X, Zhao X, Liu J, Zhang S and Luo D (2021) Dynamic Risk Assessment of the Overseas Oil and Gas Investment Environment in the Big Data Era. Front. Energy Res. 9:638437. doi: 10.3389/fenrg.2021.638437*

## INTRODUCTION

As strategic resources are of paramount importance in the world today, oil & gas can directly affect economic development, political stability, and even the national security of a country. Although alternative energy has attracted increasing attention recently, both research institutions like EIA and giant oil companies like BP believe that alternative energy is unlikely to replace petroleum in the foreseeable future. For China, a developing country experiencing rapid economic growth, the growth of renewable and other non-fossil energies is insufficient in replacing the demand for coal. In the long-term, the use of oil and natural gas can reduce the use of coal (BP, 2020).

As of 2019, China's dependence on imported oil and natural gas has rapidly climbed to 72.6 and 42.1%, respectively. The security of the oil and gas supply is the most important issue for Chinese policy makers. Many scholars believe that overseas oil and gas investment is an important measure to improve the security of China's oil and gas supply (Duan et al., 2018). Through regression analysis of China's global investment tracking data and oil and gas import data, empirical tests have shown that overseas investments in oil and gas are conducive to improving China's oil and gas supply security (Zhao et al., 2020).

For overseas oil and gas investment projects, the most significant elements include a long life cycle, great uncertainty, high capital requirement, and irreversibility of the investments (Chapman and Ward, 2004). Due to the decade-long life cycles of such projects, they are exposed to a

multitude of risk factors including the political environment of the resource countries (Bøe et al., 2019), economic conditions (Ma et al., 2020), cultural factors, religious factors (Weldegiorgis et al., 2017), as well as geological conditions (Asrilhant et al., 2007) and oil price risks (Chorn and Shokhor, 2006). Because of the intra relevance and complexity of risk management, the identification of risks is usually decided by managers and experts.

The risk analysis method was initially introduced to the petroleum industry by Kaufman in 1963 to facilitate the exploration and development of decision-making processes (Walls and Dyer, 1992). After Kaufman, studies of risk assessments for overseas oil and gas investments mainly focused on risk identification, risk impact analysis, and risk model construction (Huang et al., 2020). There are two main methods to analyze or measure the impact of risks: one is the statistical measurement method, based on historical data, and the other is the subjective probabilistic measurement method (Kumar and Gregory, 2013). A lot of research has been dedicated to building and improving project risk assessment models for the oil and gas industry. So far, a number of achievements have been put into practice (Xie et al., 2010). For example, after a series of studies of certain international oil and gas companies' management processes, Midttun et al. (2007) established a risk assessment model which introduced the real option method. However, the method has been used sparingly due to the difficulty of thoroughly and accurately quantifying uncertainty (Hsieh et al., 2019).

In the exploratory work of constructing risk assessment models, the bond between qualitative and quantitative methods is getting tighter and closer. Early studies on the risk assessment of emergencies were mainly focused on risk expectation, risk probability matrix, and the Borda value sequence method (Sachse et al., 2012). With the development of numerical methods and algorithms based on probability theory, analytical hierarchy processes (Kul et al., 2020), neural networks (Wang et al., 2013), cloud technologies (Ali et al., 2017), Bayesian networks (Huang et al., 2020), and knowledge elements (Duan et al., 2018) have been gradually applied in the risk assessment area. Driven by these methods, risk assessment systems have made great progress. However, each method has its limitations. Specifically, the analytics process is quite subjective and is greatly influenced by expert judgments. Neural network algorithms require a large scale of objective data at the starting point. Cloud theory's reasoning capability is low, therefore it is not suitable for complex and dynamic evaluation systems (Olusola et al., 2019). Further, the Bayesian network cannot function well when the node size is large.

To overcome the shortcoming of single approaches, hybrid risk assessment methods have been developed. Two examples follow here; in the assessment of CFPP investment risks of "One Belt, One Road" countries, a hybrid evaluation model (Yuan et al., 2019) has been built based on network process, entropy value, and the TODIM method. It was applied to rank the overall risk level of CFPP investment risks for the 23 "One Belt, One Road" countries. In assessing the renewable energy investment risks in countries along the "One Belt, One Road," an ANP-cloud framework, which considers randomness, was constructed. The cloud model has been used to fully describe the randomness and fuzziness of information. The risk levels of 54 countries were tested under this framework. These hybrid methods have been applied in investment risk assessments in transportation, thermal power plants, non-ferrous metals, and renewable energy, but rarely in overseas oil and gas investments.

Cloud modeling is a powerful tool for uncertainty analysis in the field of big data and artificial intelligence. In the process of qualitative and quantitative conversion, it shows better expression ability for uncertain information. The Bayesian Network (BN) has outstanding abilities in expressing probability, handling uncertainty, and fusing multi-source information. These two models play important roles in the field of uncertain dynamic reasoning. Combining the two and applying the hybrid model into the dynamic assessment of environmental risks of overseas oil and gas investments has multiple advantages. Cloud modeling and BN are models capable of handling uncertainty. This fits well with the overseas oil and gas investment environment, which is often full of unexpected events and changes. Cloud modeling can effectively deal with randomness and fuzziness at the same time, resulting in more sophisticated and comprehensive consideration of risks. The BN has a flexible structure and can be applied to environmental assessment research in different countries.

There are three main contributions of this article. First, the index system of environmental risk assessment for overseas oil and gas investment was improved. Environmental constraints and state-related factors were added, enhancing the model's fit with the current status of overseas oil and gas investments. Second, a Cloud parameter Bayesian network has been constructed through the research. The new recommended model is capable of predicting the probability of high risks and evaluating specific risks within a time frame(s). Thus, the accuracy and dynamics of the risk assessment method are improved. Finally, this study contributes great practical importance. A comparative analysis was conducted of the overall evaluation of the investment environments of several oil and gas resource countries.

The following part of this article is organized into four sections. Section Methodology takes a closer look at the methods, the design of the index system, data sources, as well as the cloud parameter Bayesian network model we have built. Section Results presents the results drawn from running the model and conducting a simulation analysis of emergency events. Section Conclusion and Policy Implication provides the conclusions and policy recommendations.

## METHODOLOGY

In the field of investment risk assessment, the Analytic Hierarchy Process (AHP), Fuzzy Comprehensive Evaluation (FCE), Regression Approach, Time Series Method, Gray System Method, Neural Network Method, and other methods are extensively used. However, these methods have their limitations. AHP and FCE face difficulties in reflecting the objective state

of things due to their subjectivity, since investors are likely to have different subjective probability assessments of the same project-specific uncertainty (Minli and Wenpo, 2012; Yuan et al., 2019). The Regression Approach strictly requires sample data and lacks the ability to analyze uncertainties. The Time Series Method and Gray System Method are mainly aimed at the prediction of the trend of the data series itself, without considering the internal influencing factors of the predicted object and the relationship between various factors, so they are not suitable for the prediction of the risk of international investment (Ke et al., 2012). As a machine learning method developed in recent years, the neural network method is of great value in solving complex non-linear problems (Wang et al., 2013). It automatically extracts reasonable rules by learning a large number of samples and has high learning and prediction ability. However, too many uncertainties in international investment will lead to training failure or over-fitting of the neural network model (Chandrinos et al., 2018). Overfitting refers to the uncertainty of international investment risk being analyzed as deterministic regularity by the neural network model, which reduces the prediction accuracy of the neural network model. Risk analysis often involves fuzzy and inconsistent decision information, which may be caused by a variety of reasons (Gómez-Fuster and Jiménez, 2020). However, determinism-based analysis methods often regard such uncertainties simply as the noise or error of the system, thus neglecting the important decision information that may be contained therein (Sobczyk et al., 2017).

As an intelligent data mining and knowledge discovery method, the Bayesian network's intelligent reasoning ability plays an important role in dealing with such uncertain and inconsistent information, and has widely been used in disease prediction and diagnosis, fault diagnosis, risk assessment, and prediction (Yan et al., 2008). The construction of the Bayesian network model mostly depends on expert knowledge, which is bound to be unconvincing and extraordinarily complex. If a Bayesian network with binary nodes have n parent nodes, there are 2nd table entries of the node in the CPT table, that is, the number of table entries grow exponentially with the number of parent nodes. Over-complex Bayesian networks will lead to problems of computational complexity and overfitting. Most scholars use the Noisy-OR gate model to solve this problem, but this method is limited to the case of binary node state, which has great limitations (Feng et al., 2020).

Wang was the first to combine the membership cloud with the Bayesian network to generate the conditional probability table of the Bayesian network with less expert workload (Wei, 2016). This model can be constructed with multi-valued nodes. Experimental results show that the statistical properties of the Bayesian network parameters generated by this method are basically consistent with the proportion of nodes given by experts, and it can effectively replace the work of experts.

## The Construction of the Risk Assessment Model

When dealing with complex problems of uncertainties, the Bayesian Network (BN) uses the description of associations between system variables to replace the joint probability distribution, thus the amount of information in the system description is reduced. At the same time, because of the conditional independence between variables, the computational complexity of learning and reasoning has effectively been reduced.

Therefore, the Bayesian network has been widely used in pattern recognition, data mining, computational intelligence, and other fields. For a Bayesian network of a given structure, its Conditional Probability Table (CPT) can be constructed based on an expert knowledge method or data parameter learning method. Since different experts often provide different opinions based on their own professional judgments, the artificial subjective factors are key reasons in reducing the accuracy of networking reasoning in the expert knowledge method. The data parameter learning method helps the CPT fit the actual situation well. However, when a continuous node is the parent of the discrete node in the Bayesian network, problems such as difficulty in determining conditional probability and excessive construction workload may occur, resulting in modeling inefficiency.

Cloud modeling is a powerful tool for uncertainty analysis in the field of big data and artificial intelligence. The model functions well in conversing between qualitative information and quantitative data. Based on the reasoning ability of the Bayesian network and the knowledge expression ability of cloud theory, it is effective to combine the Bayesian network with cloud modeling theory (Jin et al., 2011). This paper adopts the Bayesian network parameter construction method, based on the cloud membership degree, to evaluate the risk of overseas oil and gas investments. The research aims to fully utilize the advantages of the two theories to provide scientific reference for decision-making and planning of overseas oil and gas investments.

### Bayesian Network (BN)

The Bayesian network was initiated by Pro. According to the conditional independence between variables, the chain rule of joint probability solution is simplified and presented in the form of a directed acyclic graph (DAG). The Bayesian network can be represented by a binary group <G, P>. G represents a directed acyclic graph (DAG), which is composed of a random variable node V and an edge E showing the dependence between variables. P represents the conditional probability table of the dependence between the random variable node and its parent node. Given a set of random variables V, the Bayesian network is a joint conditional probability distribution on V, and its joint probability can be expressed as:

$$p\,(v1, v2, \ldots vn) = \prod_{i=1}^{n} p\left(vi \middle| Pa(vi)\right)$$

Where Pa ($v_i$) represents all parent nodes of $v_i$.

### Membership Cloud Models

Elements in the universe of discourse x = {X} can map X to another ordered universe of discourse X' according to some rule F. In X', there is one and only one x' corresponding to x, then X'

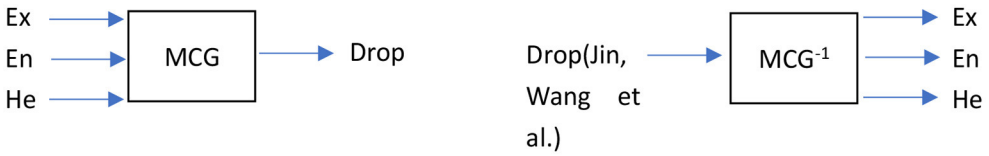



**FIGURE 1 |** Forward and backward membership cloud generator (MCG &MCG$^{-1}$).

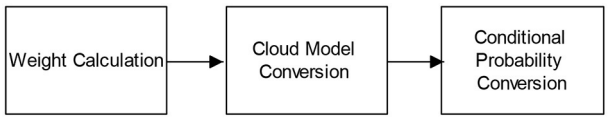



**FIGURE 2 |** The CPT generation process.

is the basic variable, and the distribution of membership degree in X 'is called membership cloud, and x is called a cloud drop.

$$F : X \rightarrow [0, 1], \forall x \in X, x \rightarrow F(X)$$

The cloud model can be represented by three eigenvalues: Expectation (Ex), Entropy (En), and Hyper Entropy (He).

Expectation (Ex) refers to the expectation of the spatial distribution of cloud drops. Ex is the most typical sample of concept quantification. Entropy (En) refers to the blur length of qualitative concepts in the cloud model. En reflects the range of qualitative concepts. Hyper entropy (He) refers to the uncertainty measurement of En. He is the En of En, which reflects the dispersion degree of cloud drops. The higher En is, the more discrete the distribution of cloud drops is.

The forward membership cloud generator (MCG) maps qualitative concepts to quantitative denotations, that is, the concrete realization of cloud drops generated by three eigenvalues of the cloud. The backward membership cloud generator (MCG$^{-1}$) maps quantitative denotations to the qualitative concept, which is shown in **Figure 1**.

## Cloud Parameter Bayesian Network

Cloud theory is superior to the Bayesian network in knowledge representation, while the Bayesian network is superior to cloud theory in inferential capability (Jian et al., 2009). The cloud parameter Bayesian network proposed by Wang Wei combines the knowledge expression ability of the cloud and the reasoning ability of the Bayesian network. This model integrates the membership cloud model into the node parameters of the Bayesian network, transforming the cloud model into a CPT table of single a Bayesian network node. This integration greatly reduces the workload of experts in designing CPT tables and improves the design efficiency of the Bayesian network. The CPT generation process is as shown in **Figure 2**.

1) Weight calculation

Weights include parent node weight, state weight, state impact factor, and state combination weight. The final result of weight calculation is the state combination weight, which is the basis of constructing the cloud cluster and generating CPT.

The parent node weight (WA) represents the influence of a single parent on a child node. State weight (Asrilhant et al., 2007) represents the degree of influence of the state changes of the parent node on the child nodes, and each state is divided into a corresponding set of state weights. Both WA and WS are constructed by expert knowledge and are inputs to the entire parameter generation model.

State impact factor (WAS) is the product of the weight of the parent node and its weight of each state. WAS represents how each state of a single parent node influences the child nodes. We set the weight of the i-th parent node as $WA_i$ and it has S states; $WS_i^s$ represents the state weight of state I, the influence factor of each state of the parent node is: $\{WAi^* WS_1^S, WAi^* WS_{2,\ldots,}^S WAi^* WS_S^S\}$. State combination weight (WCS) represents the influence of state combination of all parent nodes on child nodes, and the sum of WAS of each node state under this state combination. For example, node B has n parent nodes; the state impact factor of these parent nodes of state y is represented as:

$$WCS_y = \sum_{i=1}^{n} was_i.$$

2) Cloud model conversion

The cloud model conversion discussed in this paper refers to the definition of a cloud cluster in the universe of discourse of state combination weights, and the design of a group of cloud generators according to the definition of a cloud cluster. A universal normal cloud model is adopted in this paper, since the first and last states of the nodes are monotonous, the states at both ends use the semi-cloud model. To be more specific, the first state uses the semi-falling cloud model, the last state uses the half-rising cloud model, and the intermediate state uses the standard cloud model. The Fibonacci method is generally used to construct the eigenvalues of the cloud. The rationale is that the closer to the center of universe of discourse, the lower the En and He of the cloud will be.

3) Conditional probability conversion

Conditional probability conversion is substituting the state combination weight of the parent node into the cloud generator to obtain a cloud drop, namely, the determination degree of the node under this state combination weight. Then, the determination degree of each state is normalized to obtain a table item of the CPT table. Repeating the above

steps to calculate all the table entries, the entire CPT table is established.

The main algorithm steps are summarized as follows:

Let the state combination weight of the i-th parent node X of the desired node be $WCS_i$, WCS= $\{WCS_i | i \in N, 0 < i \leq s\}$, s is the number of parent node state combinations. The eigenvalues of the cloud generator corresponding to the j-th state of X are $He_j$, $En_j$, and $Ex_j$, respectively.

$$H_e = \{H_{ei} | i \in N, 0 < i \leq m\}$$
$$E_n = \{E_{ni} | i \in N, 0 < i \leq m\}$$
$$E_x = \{E_{xi} | i \in N, 0 < i \leq m\}$$

Where M represents the state number of node X.

The CPT table is generated as follows:

a. Iterate over each value of WCS
b. Iterate over the state space of node X, let the eigenvalues of the J-th state cloud generator be $He_j$, $En_j$ and $Ex_j$
c. Generate a normal random number $En_j^{'}$ with the expected value $En_j$ and the standard deviation Hej
d. Calculate the certainty degree of $WCS_i$, $\mu = e^{-\frac{(WCS_i - Ex_j)^2}{2(En^{'})^2}}$
e. Output a drop with degree of certainty, drop ($WCS_i$,μ)
f. Iterate over all states of X, normalize all cloud drops of the X state space, and save the normalized result to the CPT as one of the entries.
g. Iterate over WCS
h. Output CPT.

The cloud parameter Bayesian network is constructed and the inference operation is conducted according to the collected evidence. The only difference between the CPBN and BN is the CPT table construction process. Therefore, CPBN can use the BN's reasoning algorithm (such as elimination variable algorithm, clique tree reasoning algorithm, MCMC algorithm, etc.) to carry out probabilistic reasoning, to obtain the posterior probability of query variables, and to complete the final risk assessment.

Clique tree is an accurate BN inference algorithm. It can use the messaging mechanism to achieve convergence. The clique tree inference algorithm first transforms the Bayesian network into a quadratic structure and obtains the accurate result of the Bayesian network inference through the quadratic structure. SS=(CT, PP), clique tree (CT)=(C,S), where C is the cluster of the Bayesian network, S is the edge set in CT, probability potential (Chen et al., 2019) can be calculated from the probability distributions of the variables in each clique.

We use Netica, a Bayesian network analysis software which is widely used worldwide, to incorporate the Bayesian network into the node tree, to perform rapid probabilistic inference and to obtain the risk assessment result.

## Oil and Gas Investment Environment Assessment Index System

There are many rating institutions that provide professional ratings and analyses of macro investment risks worldwide (**Table 1**). Among them, six institutions study the overall comprehensive risks of countries, eight institutions focus on the risks of national debts, and the others focus their attention on countries' economic freedom, transparency index, global connectivity index, corruption index, and confidence index. Each of these institutions has their different rating objectives and classification methods. In addition, their index composition and emphasis are quite different (**Table 2**).

**TABLE 1 |** Agencies studies macro investment risks.

| Comprehensive risk | National debts | Others |
|---|---|---|
| BERI (Business Environment Risk Index) | S & P(standard and poors) | World Bank (Single Index) |
| Nord Sud Export | Fitch | Fraser Institute (Economic Freedom) |
| Political risk service | Institutional Investor | Heritage Foundation (Economic Freedom) |
| Economist Intelligence Unit (EIU) | DBRS (Dominion Bond Rating Service) | Transparency International (Corruption Index) |
| Political and economic risk consultancy | Frost & Sullivan | Human development index of UNDP |
| World Economic Forum (GCI)[i] | Credit Risk International | Price water house Coopers(PWC) (Transparency Index) |
| | OECD Credit rating | At Kearney (Global Connectivity and Confidence Index) |
| | Coface & Ducroire | |

The risks faced by overseas oil and gas investments are aligned with the overall comprehensive risks of resource countries. However, BERI is the only institution, among the six institutions which study the overall risk of countries, that specifically considers the characteristics of the oil and gas industry when it provides the Mineral Extraction Risk Assessment (MERA) for the mining industry. MERA integrates the following risks into its analysis: contract risk, personnel and equipment damage risk, nationalization risk, and expropriation risk of oil and gas assets. However, the most important risk for oil and gas investments, the resource risk, has not been considered.

Overseas oil and gas investment projects face many risk factors due to their long time horizons. This research divided the risk factors affecting overseas oil and gas investments into six dimensions based on the principles of the importance of indicators, the accessibility of the basic data, and the feasibility of the evaluation process. The six dimensions are the resource risk, the market risk, the political risk, the economic risk, the investment environment risk, and the bilateral relations between countries. Twenty-five layers were selected under these six dimensions. As shown in **Table 3**, the research established a risk evaluation index system for overseas oil and gas investments.

(1) Resource Risk

Resource risk refers to the uncertainty of the reserves of oil and gas investment projects. The reserves, which can also be referred to as the resource condition, are the basis for investment decisions. The uncertainty is highly related to the investment stage. In early stages of exploration, due to an insufficient

**TABLE 2 |** Institutions that study the overall risk of countries and their index.

| Organization | Publication | Factors | Countries |
|---|---|---|---|
| Business Environment Risk Intelligence (BERI) | Business Risk Service (BRS) | Business climate:1/3<br>Political stability:1/3<br>Currency and repayment risk:1/3 | 50 |
| | Mineral Extraction Risk Assessment (MERA) | Contract risk:25%<br>Physical risk: 25%<br>Operation risk: 25%<br>Financial risk:25% | 70–145 |
| Nord Sud Export (NSE) | Exporters index | Sovereign financial risk:30%<br>Financial market risk:40%<br>Political risk:10%<br>Business environment risk:20% | 100 |
| | Direct investors index | Sovereign financial risk:10%<br>Financial market risk:30%<br>Political risk:30%<br>Business environment risk:30% | 100 |
| Political Risk Service (PRS) | International country risk guide | Political:50%<br>Economic:25%<br>Financial:25% | 140 |
| Economist Intelligence Unit (EIU) | Country Risk Service (CRS) | Sovereign risk:22%<br>Economic structure risk: 27%<br>Banking sector risk: 28%<br>Currency risk:23% | 125 |
| | Operating risk ratings | Security<br>Political stability<br>Government effectiveness<br>Legal and regulatory<br>Macroeconomic<br>Foreign trade and payments<br>Financial<br>Tax policy<br>Infrastructure<br>Labor market | 180 |
| Global Competitiveness Index (GCI) | Basis elements | Infrastructure<br>Legal and administrative framework<br>Macroeconomic stability<br>Health and basic education | 144 |
| | Efficiency enhancement | Higher education and training<br>Goods market efficiency<br>Labor market efficiency<br>Completeness of financial markets<br>Technology<br>Market size | |
| | Innovation and maturity | Completeness of the business environment<br>Innovation | |

understanding of underground resources, large errors and high risks are inevitable in the prediction of proved reserves.

With an accumulated understanding through exploration, the uncertainty of reserves, as well as risk, decreases. Resource risks can be expressed as the impact of the uncertainty of the estimated proved resource quantity on economic benefits. Countries with abundant resources and great potential have high investment value (Wang et al., 2020).

(2) Market Risk

Market risk refers to the risk arising from market changes during the course of project operation. These changes affect the decision making through the entire life cycle of the oil and gas investment project. Market risk is an objective risk. Market risk encompasses many elements and factors. Generally speaking, the fluctuations of oil prices and exchange rates in resource countries are the most significant factors determining a project's profitability (Wu et al., 2019, 2020). Therefore, oil price risk and exchange rate risk were selected as market risk indicators.

(3) Political Risk

Due to the long life cycle of overseas oil and gas investment projects, the political situation, policies, and laws of resource countries are quite likely to change. The instability of national policies and the legal environment may affect the normal operation of a project, generating great risks.

**TABLE 3 |** Oil and Gas Overseas Investment Environment Assessment Indicators (OGOIEAI).

| Classifications of risks | Index | Abbreviation and units | Data resource |
|---|---|---|---|
| Resource risk | Total proved reserves of oil &gas | PR, Billion barrels | BP statistical review of world energy 2000–2020 |
| | Reserves-to-production ratio | R/P | |
| | Consumption-to-production ratio | C/P | |
| Market risk | Oil price (Brent) | P, $/barrel | Word Bank |
| | Exchange rate | ER | |
| Political risk | Government stability | GS | International Country Risk Guide 2019 |
| | Socioeconomic conditions | SC | |
| | Corruption | C | Corruption Perceptions Index |
| | Religious tensions | RT | International Country Risk Guide 2019 |
| | Ethnic tensions | ET | |
| | Democratic accountability | DA | |
| | Bureaucracy quality | BQ | |
| Economic Risk | GDP growth rate | GDPR | Word Bank |
| | Inflation rate | IR | |
| | Price fluctuation | PF | |
| Investment environment | Cost of business start-up procedures, male (% of GNI per capita) | CBS | Word Bank |
| | Net inflows of foreign direct investment (% of GDP) | NIFI | |
| | Environmental sustainability policy and institutional ratings | ESPIR | |
| | Forest area (% of land area) | FA | |
| | Renewable electricity output (% of total) | REO | |
| | $Co_2$ emissions (kg /PPP $GDP) | CO | |
| Bilateral relations between countries | Bilateral investment treaty signed or not | BIT, signed scores 1, not scores 0 | China's Ministry of Commerce |
| | Bilateral political relations | BPR | Delphi Method |
| | Dependence on foreign trade | DRT | CEIC |
| | Dependence on investment | DI | CEIC |

Government efficiency reflects a local governments' capability and willingness to provide services and facilitates to foreign investors (Duan et al., 2018). Democratic accountability and regulatory quality reflect the extent of a government's corruption. If government officials cannot handle things fairly and impartially, foreign companies will find it difficult to conduct business with the local community, and thus have higher probability to suffer a loss.

(4) Economic Risk

A sound economic foundation of resource countries is essential to ensure the safety and return of overseas oil and gas investments. The GDP growth rate can reflect the economic environment for a country's long-term investment. The inflation rate has high impacts on the operating cost of enterprises, and the exchange rate can also cause uncertainty with respect to project financing costs.

(5) Investment Environment Risk

To evaluate the investment environment, this paper not only considers various obstacles in the process of contract execution, but also the compulsory requirements on registered assets by the resource countries (Olusola et al., 2019). In addition, the trend of tighter and higher environmental constraints in resource countries should be integrated into the consideration. Although oil and gas investments are warmly welcomed to promote local economic development, the waste discharge in the process of resource development has a long-lasting negative impact on the local environment. More and more environmental concerns have been raised and transformed into operating costs. In other words, the concerns shake the investment profitability and decision-making process. This paper selects the following indicators for investment environment risk: Cost of business start-up procedures (% of GNI per capita), Net inflows of foreign direct investment (% of GDP), Environmental sustainability policy and institutional ratings, Forest area (% of land area),

Renewable electricity output (% of total), and $Co_2$ emissions (kg /PPP $GDP).

(6) Bilateral Relations between Countries

The success of Chinese companies' overseas investments can be largely influenced by the strength of diplomatic relationships and trade tightness. Therefore, bilateral relations should be incorporated into the index system. This paper selects the following indicators for bilateral relation risk: bilateral trade volume, exports to China /GDP, and imports from China/economic and trade cooperation with China.

## Data Resource

The geographically uneven distribution of oil and gas resources and the mismatch between supply and demand stirs up overseas oil and gas investment (Tan and Barton, 2017). China has invested and operated over 100 oil & gas projects worldwide. According to investment scale and potential, we chose 10 representative countries for our research objects. They are Brazil and Ecuador in South America, Chad and South Sudan in Africa, Iraq and the United Arab Emirates in the Middle East, Kazakhstan and Turkmenistan in Central Asia, Indonesia in Southeast Asia, and Russia. Data sources of relevant indicators are shown in **Table 3**, to facilitate the construction of CPBN, we abbreviate all the indicators due to the maximum character of node name in Netica is 30.

# RESULTS

## Overall Assessment of Oil and Gas Investment Environment Risk

The CPBN is established based on the logical level between the factors of oil & gas overseas investment environment assessment indicators, as shown in **Figure 3**.

We input the probability distribution of the two-layer root nodes and the CPT of the sub-nodes into Netica. The number of CPT table entries is exponential with parent nodes and the workload of the CPT table explodes when parent nodes increase. With 200 BN results of cloud parameters, we comprehensively evaluate the investment environment of 10 typical oil & gas resource countries from 2000 to 2019 and determine the degree of subordination of overseas investment risk as low (L), moderate (M), and high (H). Due to limited space, we present the cloud parameter BN inference visualization result of Brazil's oil & gas investment assessment in 2019, as shown in **Figure 4**.

A silent transforming investment environment in Brazil can be visually observed in the form of a graph, as shown in **Figure 5**.

The Netica results show that the risk of investment in Brazil's oil & gas industry over the years exhibits a downward trend from being a moderate risk to being a low risk. The investment risk (low risk, medium risk, high risk) has changed from 32, 42, and 26 to 42, 35, and 23 which is the result of the relatively good investment environment in Brazil.

Since 2015, there have been some fluctuations in the investment environment in Brazil due to the country's favorable resource prospects and investment incentive policies, the probability of high risks is gradually decreasing, as it turns out. China won the Ribella field in a consortium bid with Total and Shell in 2013. The Ribella field, located in the Santos basin in 2,000 meters of water south-east of Rio de Janeiro, is estimated to hold 12 billion barrels of oil reserves, making it the world's largest offshore field. In 2019, China won Buzios and Aram field, and in the foreseeable future, Chinese oil companies will invest billions in funds for exploration and development in Brazil. Thus, Chinese companies need to be more cautious because of recent evolving uncertainties in in Brazil.

## Assessments From Different Risk Dimensions

With Netica, we obtain the specific values of the risk assessment for various dimensions and illustrate the following representative sub dimensions.

The result of resource risk assessment is shown in **Figure 6**. The oil resource risk in Brazil has always been low. This risk trend chart exhibited a downward trend from 2001 to 2014, which increased in 2015 and has been relatively stable with a slight decline since then, and the probability of low risk remains above 70%. This situation is mainly affected by the growth of oil storage and exploitation in recent years.

From a macro point of view, the oil and gas market in Brazil is a magnet for foreign investment. Brazil's oil & gas exploration and development started from onshore resources, developed into a shallow continental shelf, and completed the historical leap by developing deep water pre-salt oil & gas. Since Brazil opened its oil and gas market to FDI in 1997, the market has been booming thanks to the influx of international companies. However, compared to the 60,000 wells in neighboring Argentina and four million wells in the US, only 30,000 wells have been drilled in Brazil. The oil exploration market is still in its infancy.

The result of political risk assessment is shown in **Figure 7**. The political situation in Brazil showed a clear process of gradual deterioration from 2007 to 2012 and then a gradual recovery since 2012. Brazil has built a representative democratic political system. Since 2015 Brazil has been suffering a series of political upheavals. Brazil has maintained smooth relations with China in the past couple of years. There is good development momentum in bilateral relations and trade.

The result of economic risk assessment is shown in **Figure 8**, which shows a dramatic picture. In recent years, Brazil has been suffering an economic downturn, currency depreciation, and fiscal and inflation pressures. Consecutive quarters of negative economic growth in 2015 and 2016 compelled the government to implement a series of adjustment measures with a focus on restoring fiscal balance. The economy remained stable in 2017 and 2018, with a real GDP growth of 1.1%. However, due to the impact of the international economic environment, the economy declined in 2019. Brazil will actively attract foreign capital and expand foreign trade, but the probability fluctuations of the risk level are more volatile in the near future.

As shown in **Figure 9**, although there are still many problems in Brazil's environmental constraints, Brazil's investment environment risk has been low and stable. Brazil has done a good job of developing clean energy, largely relying on domestic

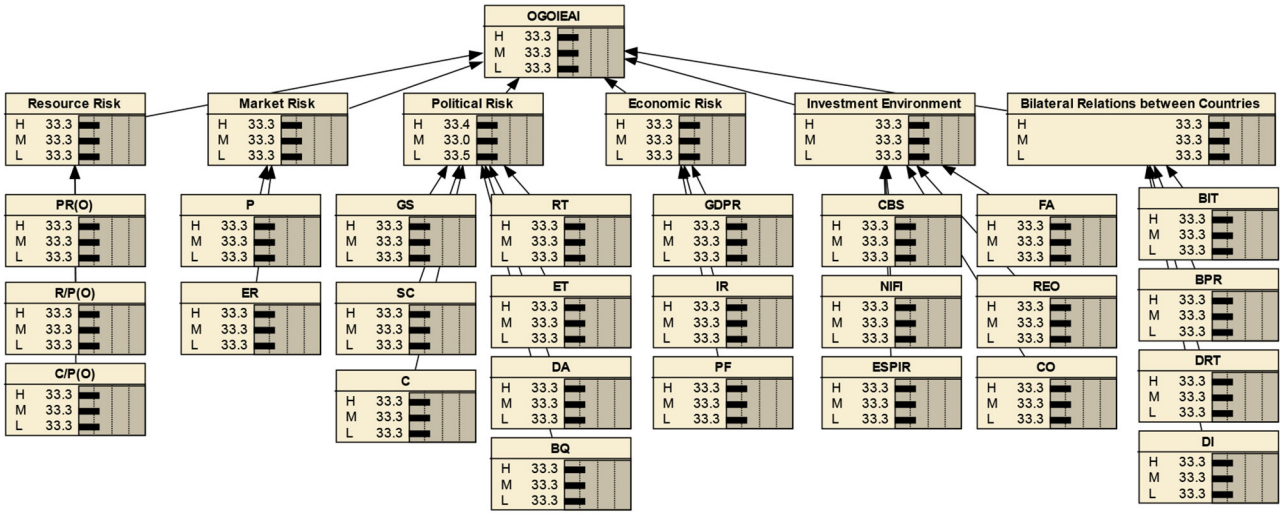



**FIGURE 3 |** CPBN of oil and gas overseas investment environment risk.

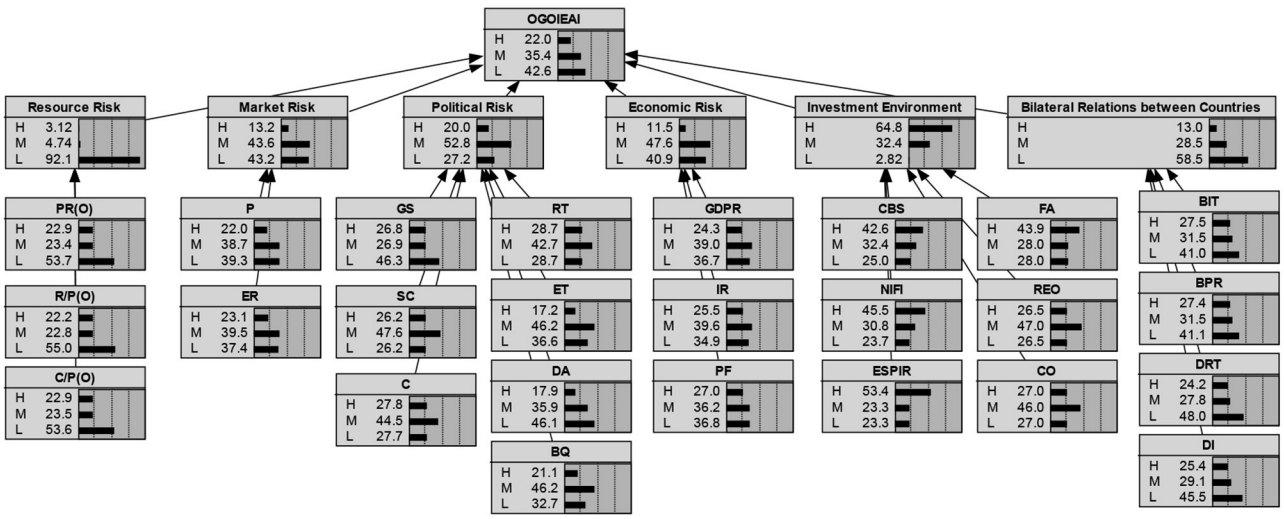



**FIGURE 4 |** Oil and gas investment risk assessment of Brazil in 2019.

capital. The carbon emissions have been falling every year since 2014. Brazil has the highest proportion of renewable energy in the context of global economic power, the proportion of renewable energy is 75%. Less than 4.5% of electricity is generated by coal in Brazil, while the same input generates 70% in China. Strict environmental protection policies impose binding forces on oil and gas investment enterprises.

The risk analysis of four dimensions analyzes the risk changes of oil & gas investment in Brazil from various perspectives, the comprehensive action of various factors formed the final investment environment risk trend.

## Emergency Simulation

The above simulation results are based on existing data in the data resource. We then revise the model parameters to consider the impact of possible sudden political or economic risks on oil investment risk. We introduce the new node information into the cloud parameter BN and use the joint tree reasoning algorithm to carry out reasoning calculation, so as to obtain the new membership degree of the emergency risk of overseas oil & gas investment. In this paper, the variable control method is used to analyze the impact of emergencies on the investment risk in Brazil in combination with the investment situation of Brazil in 2019.

### Political Emergencies Simulation

There is no doubt that the more stable the government, the less likely the country is to experience political upheaval and the more attractive it will be to oil & gas investors. We consider the main factors affecting the stability of the government, simulate the

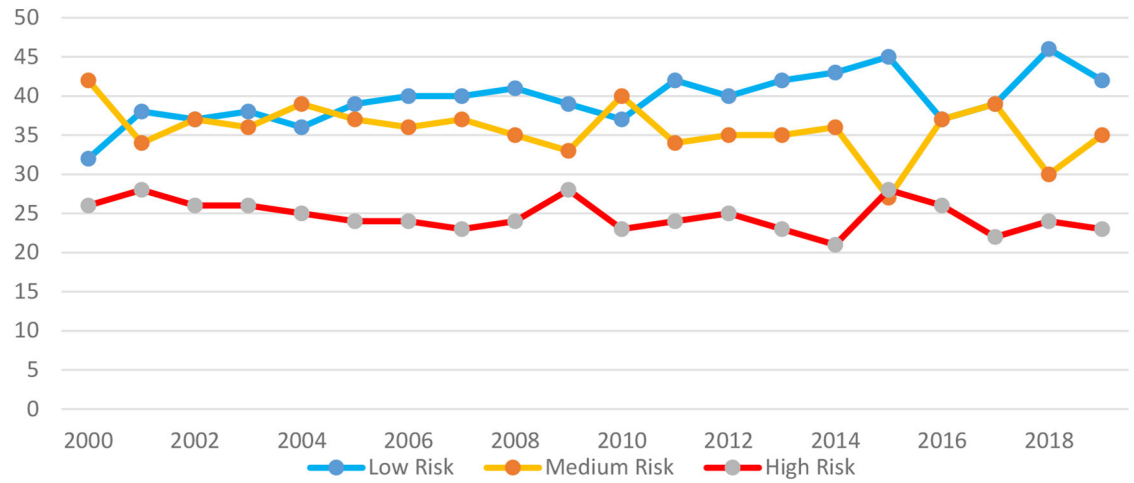



**FIGURE 5 |** Oil &gas investment environment risk in Brazil from 2000 to 2019.

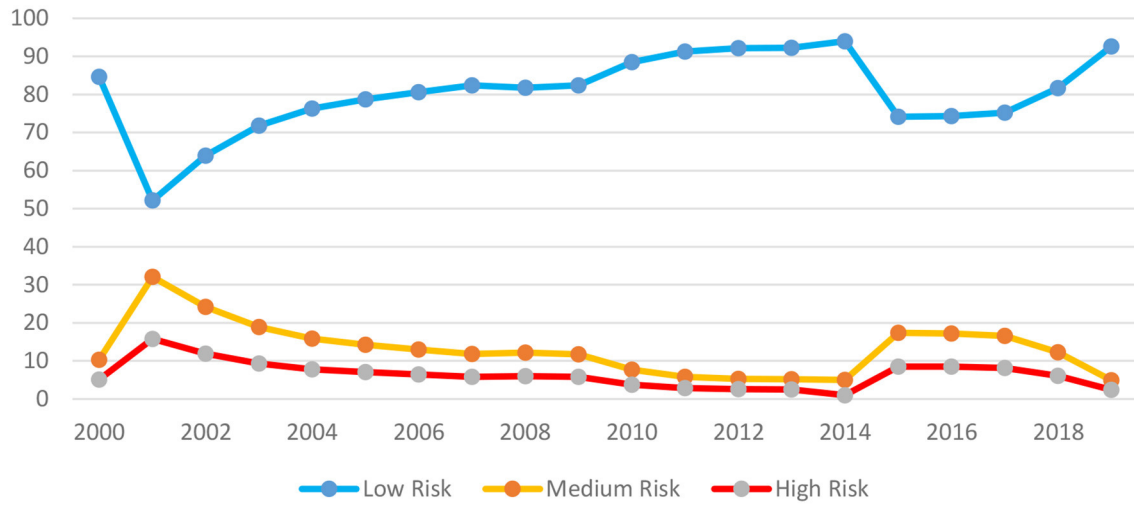



**FIGURE 6 |** Resource risk in Brazil from 2000 to 2019.

sudden political events, and analyze the dynamic trend of the risk level probability of the sudden political events.

Through the analysis of the original data of political risk dimension indicators from 2000 to 2018, we can identify several indicators that were significantly affected by political events at several time points when political risk changes greatly. According to various forecasting reports issued by the Brazilian central bank and other institutions, combined with the political risk assessment results of Brazil, the change range of indicators under the scenario simulation is roughly estimated. Therefore, based on the existing information, the risk assessment model node of Brazil's oil & gas investment is modified as follows:

(a) Government stability is at a high risk level, and the degree of membership (low, moderate, high) may be set to 0, 30, or 70.
(b) Socioeconomic Conditions is at a moderate risk level, and the degree of membership (low, moderate, high) may be set to 0, 70, 30.
(c) Oil price fluctuation is at a low risk level, and the degree of membership (low, moderate, high) may be set to 80, 20, 0.
(d) Inflation rate is at a moderate risk level, and the degree of membership (low, moderate, high) may be set to 0, 90, 10.
(e) Exchange rate is at a moderate risk level. It is advisable to set the degree of membership (low, moderate, high) to 40, 60, 0.
(f) GDP growth rate is at a moderate risk level, and the degree of membership (low, moderate, high) may be set to 20, 80, 0.
(g) Net inflows of foreign direct investment is at a high risk level, and the degree of membership (low, moderate, high) may be set to 0, 20, 80.

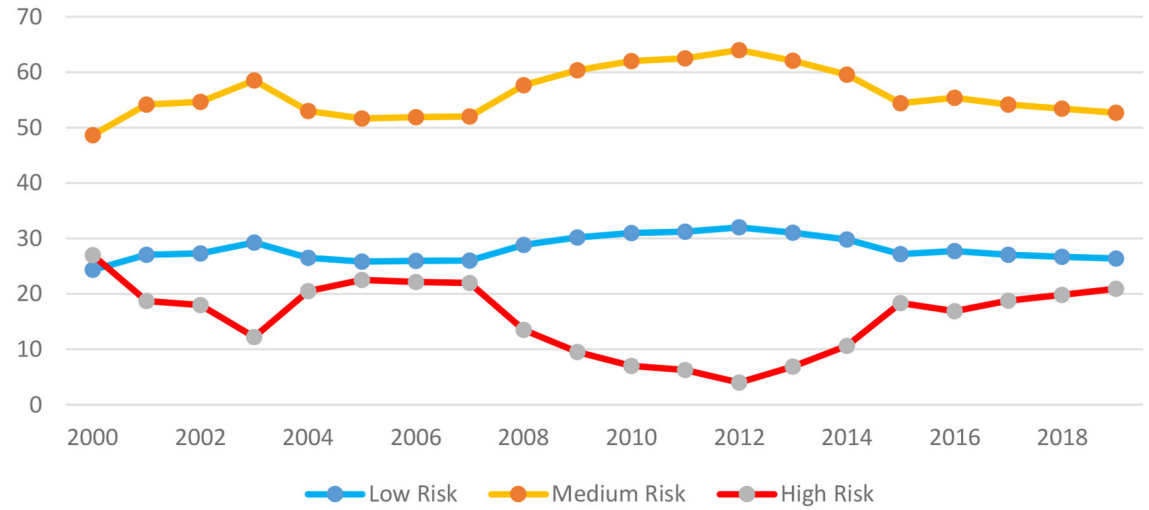



**FIGURE 7 |** Political risk in Brazil from 2000 to 2019.

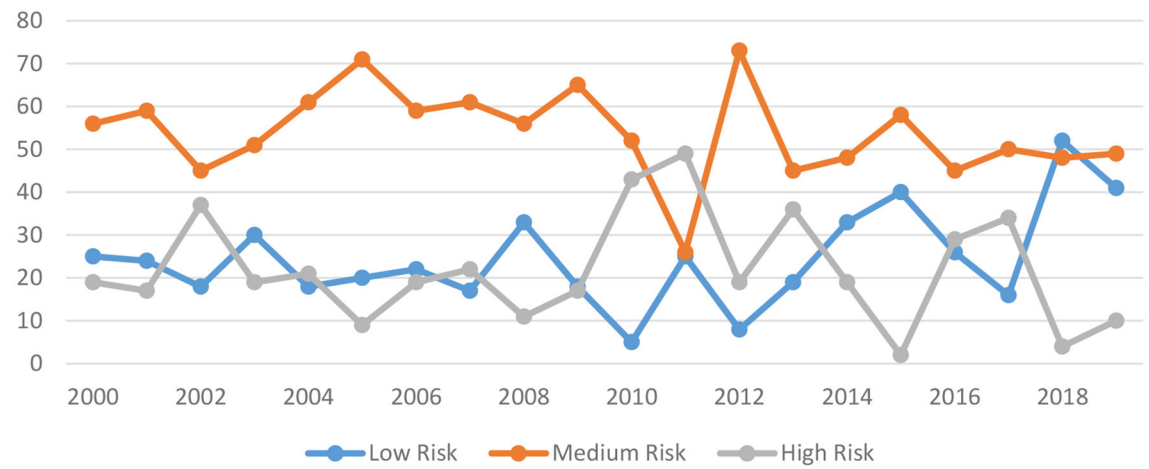



**FIGURE 8 |** Economical risk in Brazil from 2000 to 2019.

As illustrated in the above figure, under the simulated political event scenario, the variation of oil & gas investment risk in Brazil (low, medium, high) is 30.0, 37.9, 32.1. In other words, sudden political events make Brazil a medium-risk investment target country. Analyzing this sudden political event and understanding the risk composition affected by it can help overseas investors to formulate countermeasures quickly, adjusting investment decisions in time and preventing losses effectively.

## Economic Emergencies Simulation

As the largest country in South America, Brazil has suffered several economic and financial crises. The economic growth of Brazil has been declining since 2011. In particular in 2015 and 2016, Brazil's GDP declined 3.8%. In February 2016 Moody's cut Brazil's sovereign debt rating to junk status resulting in a very bad outlook. The International Monetary Fund's latest World Economic Outlook has further downgraded its growth forecasts for Brazil over the next 2 years and sees the country mired in a long and deep recession.

We assume that there was a major shock to the Brazilian economy in 2019. The selection criteria of indicators in the scenario simulation refers to 4.3.1. According to the existing information, the node of the risk assessment model for investment in Brazil is modified as follows:

(a) Oil price fluctuation is at a low risk level, and we still set the degree of membership (low, moderate, high) to 80, 20, 0.
(b) Exchange rate is at a moderate risk level, and the degree of membership (low, moderate, high) may be set to 60, 60, 0.
(c) Government stability is at a moderate risk level, and the degree of membership (low, moderate, high) may be set to 0, 60, 40.

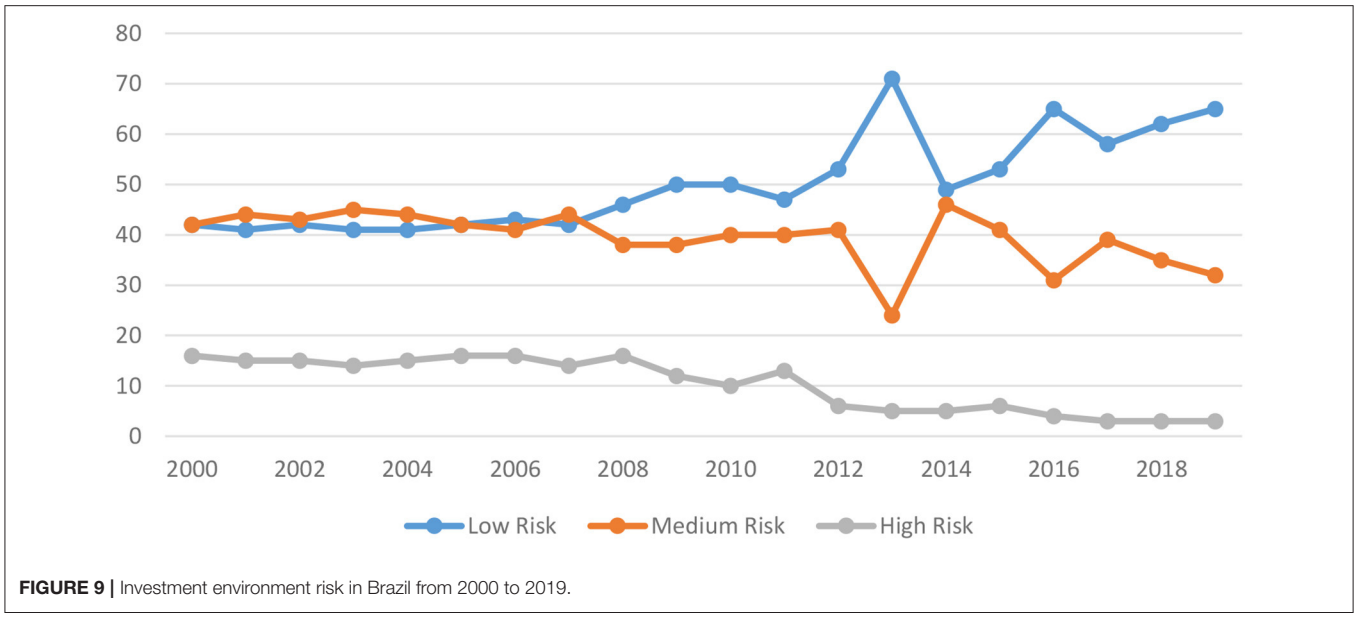



**FIGURE 9 |** Investment environment risk in Brazil from 2000 to 2019.

(d) Socioeconomic Conditions is at a moderate risk level, and it may be appropriate to set the degree of membership (low, moderate, high) to 20, 80, 0.
(e) GDP growth rate is at a high risk level, and the degree of membership (low, moderate, high) may be set to 0, 20, 80.
(f) Cost of business start-up procedures is at a moderate risk level. It is advisable to set the (low, moderate, high) membership degree to 20, 80, 0.
(g) Net inflows of foreign direct investment is at a high risk level, and the degree of membership (low, moderate, high) may be set to 0, 20, 80.

The newly acquired simulation results indicate that oil & gas investment risk (low, medium, high) in a sudden economic event scenario is (35.5, 38.0, 26.5). It can therefore be inferred that the oil & gas investment risk in Brazil is at a moderate risk level.

In general, there is no possibility of sudden change in resource risk, and the political, economic, and investment environment will have an emergency situation, so we simulated the model, considering the sudden political or economic risk for oil as the influence of investment risks.

## CONCLUSION AND POLICY IMPLICATION

Combining industry consensus and classification criteria of professional risk rating agencies, we selected 25 indicators from six dimensions to measure the oil and gas overseas investment environment; environmental constraints and state-related factors were added, improving the system's fit with the current status of overseas investments. We applied the Cloud parameter Bayesian network algorithm to dynamically evaluate oil and gas overseas investment risk of 10 resource countries. This new recommended model is capable of quantizing the probability of high risks and evaluating specific risks within time frame(s). Compared with those static numbers for risk rating agencies worldwide, our research results show the annual variation of risk dynamics over the past two decades. The evaluation period of overseas oil and gas investment is short and difficult; by using CPBN, a rapid and efficient quantitative evaluation of environment risk can be vividly conducted. In summary, our simulation results show that the risk level of investing in oil & gas projects in Brazil, Kazakhstan, Turkmenistan, Indonesia, and Russia is low. In terms of investing in Ecuador and United Arab Emirates, the risk levels are moderate. The other three countries which include Chad, South Sudan, and Iraq have been suffering political or economic crises, and their investment risks have therefore been experiencing increasing trends. Experts involved in oil and gas investment decisions verified our results. Their risk management experience of investing oil & gas projects in Brazil is similar to our model. This method can also be applied to investment, engineering, and service projects. The proper use of risk assessment results can be conducive to those potential investors who will invest in potential countries in the future.

From the results of evaluation, investors should prioritize the expansion of projects under the premise that the economic value meets the requirements in low-risk countries such as Russia, Indonesia etc. For those countries with medium risk levels, it is important for executives and policy makers to evaluate the investment risk according to different dimensions to avoid foreseeable increasing risks affecting feasibility and profitability of investments. Investors should identify risks in a timely manner from high-risk countries like South Sudan and Iraq and avoid worthless investments. Our simulation results are based on existing data in the data resource, which is still macroscopic and not updated regularly. This leads to our model only being able to evaluate the investment

risk at the national level. But resource risks and contract-related risks vary widely from region to region within the same country. We therefore suggest that governments and national oil companies establish a targeted data and information service pool to monitor the changing investment environment risks of potential investment projects in real time. Corporations and academic institutions should remain in touch via information sharing to construct the cornerstone of a prospective overseas investment insurance system. The more data accumulates, the more precise and specific the risk evaluation is in the age of big data and artificial intelligence (Desai et al., 2020).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

DL: proposer and supervisor of this research. XZ: chief of method application exploration. JL and SZ: providing related data. All authors contributed to the article and approved the submitted version.

## REFERENCES


Ali, A., Warren, D., and Mathiassen, L. (2017). Cloud-based business services innovation: a risk management model. *Int. J. Inform. Manag.* 37, 639–649. doi: 10.1016/j.ijinfomgt.2017.05.008

Asrilhant, B., Dyson, R. G., and Meadows, M. (2007). On the strategic project management process in the UK upstream oil and gas sector. *Omega* 35, 89–103. doi: 10.1016/j.omega.2005.04.006

Bøe, K. S., Jordal, T., Mikula, S., and Molnár, P. (2019). Do political risks harm development of oil fields? *J. Econ. Behav. Organ.* 157, 338–358. doi: 10.1016/j.jebo.2018.01.005

BP (2020). Statistical Review of World Energy 2020.

Chandrinos, S. K., Sakkas, G., and Lagaros, N. D. (2018). AIRMS: a risk management tool using machine learning. *Expert Syst. Appl.* 105, 34–48. doi: 10.1016/j.eswa.2018.03.044

Chapman, C., and Ward, S. (2004). Why risk efficiency is a key aspect of best practice projects. *Int. J. Project Manage.* 22, 619–632. doi: 10.1016/j.ijproman.2004.05.001

Chen, C., Limao, Z., and Kong Tiong, R. L. (2019). A novel learning cloud Bayesian network for risk measurement. *Appl. Soft Comp.* 87. doi: 10.1016/j.asoc.2019.105947

Chorn, L. G., and Shokhor, S. (2006). Real options for risk management in petroleum development investments. *Energy Econ.* 28, 489–505. doi: 10.1016/j.eneco.2006.03.002

Desai, J. N., Pandian, S., and Vij, R. K. (2020). Big data analytics in upstream oil and gas industries for sustainable exploration and development: a review. *Environ. Technol. Innov.* 21:101186. doi: 10.1016/j.eti.2020.101186

Duan, F., Ji, Q., Liu, B.-Y., and Fan, Y. (2018). Energy investment risk assessment for nations along China's Belt & Road Initiative. *J. Clean. Prod.* 170, 535–547. doi: 10.1016/j.jclepro.2017.09.152

Feng, X., Jiang, J.-, and c., Wang, W.-,f. (2020). Gas pipeline failure evaluation method based on a Noisy-OR gate bayesian network. *J. Loss Prev. Process Ind.* 66. doi: 10.1016/j.jlp.2020.104175

Gómez-Fuster, J. M., and Jiménez, P. (2020). Probabilistic risk modelling for port investments: a practical approach. *Case Stud. Transport Policy* 8, 822–831. doi: 10.1016/j.cstp.2020.06.001

Hsieh, H.-C., Sofia, B., and Vu, T. H. C. (2019). The effects of economic policy uncertainty on outward foreign direct investment. *Int. Rev. Econ.* 64, 377–392. doi: 10.1016/j.iref.2019.08.004

Huang, J., Liu, J., Zhang, H., and Guo, Y. (2020). Sustainable risk analysis of China's overseas investment in iron ore. *Res. Policy* 68:101771. doi: 10.1016/j.resourpol.2020.101771

Jian, W., Xiao, W.-,j., Sheng, W., and Wang, S.-,w. (2009). *An improved effectiveness evaluation method based on grey cloud model.* Microcomp. Inf.

Jin, B., Wang, Y., Liu, Z., and Xue, J. (2011). A trust model based on cloud model and bayesian networks. *Proc. Envi. Sci.* 11, 452–459. doi: 10.1016/j.proenv.2011.12.072

Ke, L., Xiaoliu, S., Zhongfu, T., and Wenyan, G. (2012). Grey clustering analysis method for overseas energy project investment risk decision. *Syst. Eng. Proc.* 3, 55–62. doi: 10.1016/j.sepro.2011.11.008

Kul, C., Zhang, L., and Solangi, Y. A. (2020). Assessing the renewable energy investment risk factors for sustainable development in Turkey. *J. Clean. Prod.* 276. doi: 10.1016/j.jclepro.2020.124164

Kumar, M., and Gregory, M. J. (2013). An exploration of risk management in global industrial investment. *Risk Manage.* 15, 272–300. doi: 10.1057/rm.2013.8

Ma, H., Sun, Z., and Fang, C. (2020). Risk assessment of transnational oil investment in Central Asia using a fuzzy comprehensive evaluation method. *Regional Sustainability* 1(1): 11–19. doi: 10.1016/j.regsus.2020.06.002

Midttun, A., Dirdal, T., Gautesen, K., Omland, T., and Wenstøp, S. (2007). Integrating corporate social responsibility and other strategic foci in a distributed production system: a transaction cost perspective on the North Sea offshore petroleum industry. *Corp. Govern. Int. J. Business Soc.* 7, 194–208. doi: 10.1108/14720700710739822

Minli, Z., and Wenpo, Y. (2012). Fuzzy comprehensive evaluation method applied in the real estate investment risks research. *Phys. Proc.* 24, 1815–1821. doi: 10.1016/j.phpro.2012.02.267

Olusola, A., Nurse, J. R. C., Martin, A., and New, A. (2019). Cyber risk assessment in cloud provider environments: current models and future needs. *Comp. Security* 87. doi: 10.1016/j.cose.2019.101600

Sachse, H., Jungermann, H., and Belting, J. M. (2012). Investment risk – the perspective of individual investors. *J. Econ. Psychol.* 33, 437–447. doi: 10.1016/j.joep.2011.12.006

Sobczyk, E. J., Kicki, J., Wiktoria, S., and Szuwarzyński, M. (2017). Support of mining investment choice decisions with the use of multi-criteria method. *Res. Policy* 51, 94–99. doi: 10.1016/j.resourpol.2016.11.012

Tan, S. H., and Barton, P. I. J. (2017). Optimal shale oil and gas investments in the United States. *Energy* 141, 398–422. doi: 10.1016/j.energy.2017.09.092

Walls, M. R., and Dyer, J. S. (1992). Risk propensity and firm performance: a study of the petroleum exploration industry. *Manage. Sci.* 42, 939–1092. doi: 10.1287/mnsc.42.7.1004

Wang, D., Tong, X., and Wang, Y. (2020). An early risk warning system for outward foreign direct investment in mineral resource-based enterprises using multi-classifiers fusion. *Res. Policy* 66. doi: 10.1016/j.resourpol.2020.101593

Wang, Z., Li, H., and Jia, Y. (2013). A neural network model for expressway investment risk evaluation and its application. *J. Transp. Syst. Eng. Inform. Technol.* 13, 94–99. doi: 10.1016/S1570-6672(13)60118-X

Wei, W. J. (2016). *An threat assessment method based on cloud parameters bayesian network*. Comp. Technol. Dev.

Weldegiorgis, F. S., Mesfin, B., and Sturman, K. (2017). Looking for oil, gas and mineral development in Ethiopia: Prospects and risks for the political settlement. *Extr. Ind. Soc.* 4, 151–162. doi: 10.1016/j.exis.2016.12.003

Wu, D., Huang, J.- b., Zhong, M.-r. (2020). Prediction of metal futures price volatility and empirical analysis based on symbolic time series of high-frequency. *Trans. Nonferrous Met. Soc. China* 30, 1707–1716. doi: 10.1016/S1003-6326(20)65332-8

Wu, Y., Wang, J., and Ji, S. (2019). Renewable energy investment risk assessment for nations along China's Belt & Road Initiative: an ANP-cloud model method. *Energy* 190. doi: 10.1016/j.energy.2019.116381

Xie, G., Yue, W., Wang, S., and Lai, K. K. (2010). Dynamic risk management in petroleum project investment based on a variable precision rough set model. *Technol. Forecast. Soc. Change* 77, 891–901. doi: 10.1016/j.techfore.2010.01.013

Yan, D. W., Liang-Xian, G. U., and Pan, L. (2008). A study of new method for weapon system effectiveness evaluation based on bayesian network. *Comp. Technol. Dev.* 4, 209–213.

Yuan, J., Li, X., Xu, C., Zhao, C., and Liu, Y. (2019). Investment risk assessment of coal-fired power plants in countries along the Belt and Road initiative based on ANP-Entropy-TODIM method. *Energy* 176, 623–640. doi: 10.1016/j.energy.2019.04.038

Zhao, Y., Shi, X., and Song, F. (2020). Has Chinese outward foreign direct investment in energy enhanced China's energy security? *Energy Policy* 146. doi: 10.1016/j.enpol.2020.111803

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Duan, Zhao, Liu, Zhang and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 09 April 2021
doi: 10.3389/fenvs.2021.649008

# Tripartite Coordinated Regulation of New Energy Vehicles Based on Dynamic Bayesian Game

*Changyu Liu[1]*†, Yadong Song[1]†, Le Chang[1,2]* and Guanglong Dong[3]**

[1] Business School, Shandong Normal University, Jinan, China, [2] School of Economics and Management, Shandong Agriculture and Engineering University, Jinan, China, [3] School of Management Engineering, Shandong Jianzhu University, Jinan, China

OPEN ACCESS

**Edited by:**
*Xunpeng Shi, University of Technology Sydney, Australia*

**Reviewed by:**
*Li Xu, Nanjing University of Aeronautics and Astronautics, China*
*Javier Bustos-Salvagno, Universidad del Desarrollo, Chile*

***Correspondence:**
*Changyu Liu*
*sdliuchangyu@126.com*
*Le Chang*
*andy.changle@yahoo.com.sg*
*Guanglong Dong*
*dongguanglong18@sdjzu.edu.cn*

†*These authors have contributed equally to this work and share first authorship*

**Specialty section:**
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

**Received:** *03 January 2021*
**Accepted:** *15 March 2021*
**Published:** *09 April 2021*

**Citation:**
*Liu C, Song Y, Chang L and Dong G (2021) Tripartite Coordinated Regulation of New Energy Vehicles Based on Dynamic Bayesian Game. Front. Environ. Sci. 9:649008. doi: 10.3389/fenvs.2021.649008*

Manufacturers are disseminating false or ambiguous information regarding new energy vehicles (NEVs), which has led to skepticism from consumers about the quality of NEVs. In this research, we simultaneously considered the relationship among manufacturers, consumers, and governments from the perspective of stakeholders, and then we analyzed the tripartite coordinated regulation. In view of the serious information asymmetry of NEVs, we innovatively developed the Bayesian dynamic game model. By solving refined Bayesian equilibrium strategies, this study explores the effects of key influencing factors on strategic choices. On the basis of the conclusion, relevant countermeasures and suggestions are put forward to engender effective regulation by governments.

**Keywords: new energy vehicles, coordinated regulation, tripartite game, Bayesian game model, environmental policy**

## INTRODUCTION

Environmental deterioration issues such as high emission levels and global warming have become a global focus during the past decades. Energy-saving products (ESPs) with the features of carbon emission reduction and low-energy consumption are key measures to solve pollution emissions and resource depletion (Yu et al., 2019; Ji et al., 2020). There are many ESPs, such as energy-saving automobiles, air conditioners, water dispensers, and so on. In this paper, we choose typical new energy vehicles (NEVs) as representative products. With the support of governmental policies, the NEV market scale continues to expand. However, the misuse of eco-labels and fake products has become more widespread. This has caused a series of defrauding incidents (Chen and Chang, 2013; Guyader et al., 2017). For example, in 2015, regulators from the United States confirmed that the German automaker Volkswagen had installed fake equipment that allowed the vehicles to pass emissions tests. However, in reality, the vehicles emitted up to 40 times the acceptable levels of harmful gasses and nitrogen oxides. In 2016, the Chinese government conducted a special inspection of NEVs and revealed that the production of 76,000 NEVs was associated with fraud, including five major manufacturers, such as Suzhou Jimsey Bus Manufacturing Co., Ltd.

Many researchers have indicated that consumers are willing to pay higher prices for NEVs, which encourages manufacturers to produce them (Liu et al., 2012; Chander and Muthukrishnan, 2015). However, dishonest manufacturers have caused poor purchasing decisions for consumers as well as a loss of credibility for NEVs. Despite the tighter regulations in place to stop speculations, including legislation and financial instruments, the results reveal that regulations in NEVs do not

effectively reduce the incidence of fake and inferior behaviors (Zhao et al., 2020). Thus, solving how to reduce the increasingly severe deception problem and ensure the quality of NEVs is an urgent concern. Considering these practical problems, the relevant regulation policies need to be developed urgently.

Consumers are important stakeholders, who not only affect the environmental performance of manufacturers but also influence the efficiency of regulation (Hammami et al., 2018; Yang et al., 2020). However, the current research mainly pays close attention to discussing how manufacturers respond to government policies. In particular, the role of consumers was relatively limited and concentrated on in previous studies. Studies on multi-agent relationships and collaborative regulation mechanisms have remained scarce. Therefore, excluding manufacturers and governments and adding consumers as stakeholders in the game can help further explore the regulation of NEVs. Therefore, tripartite coordinated regulation of NEVs represents more realistic situations, and the interaction relationship among the three participants can also be analyzed.

Stakeholders involve multiple members who may not want to share private information, especially manufacturers whose product quality is faked. With regard to the quality of NEVs, serious information asymmetry is also an important factor that cannot be ignored (Zhang et al., 2019; Wang Y. et al., 2020). The assumption in the previous game that participants in NEVs provide complete information is too idealistic (Shen et al., 2019). To the best of our knowledge, no NEV studies have taken asymmetry information and tripartite game into consideration simultaneously.

In this paper, we take manufacturers, consumers, and governments into account concurrently. We study the tripartite coordinated regulation of NEVs under the condition of asymmetric information. Several questions are of interest:

1. What are the game relationships among governments, manufacturers, and consumers?
2. With the asymmetric information, how can the tripartite game model among governments, manufacturers, and consumers in NEVs be formulated?
3. What are the key influencing factors affecting the different participants in the equilibrium strategy?

The motivations of our research for solving the above problems are the following: (1) This paper proposes a Bayesian game model, which differs from the previous perfect information game, to analyze the decisions of government regulation, consumers' consumption, and manufacturer production in asymmetric information. (2) Discuss the tripartite coordinated regulation on the quality of NEVs to investigate the decisions of different participants and obtain the optimal strategy. (3) Analyze the interplay and key factors affecting the choices of game players. Targeted regulatory measures will also be proposed.

The remainder of this paper is organized as follows. In section 2, the relevant literature is reviewed. In section 3, we provide our problem assumptions and basic model, and derive the utility functions of the game players. In section 4, we present and analyze the equilibrium possibility. Finally, in section 5, we summarize our concluding remarks and discuss directions for future research.

## LITERATURE REVIEW

The literature is related to various research aspects on environmental regulation. Our literature review primarily relates to three research streams: the effects of environmental regulation on manufacturers, the behavior strategy selection of stakeholders, and the differential game models in ESPs. Key studies from the different research streams are reviewed briefly in the following subsections.

### Effects of Environmental Regulation on Manufacturers

As more and more regulations on environmental protection are issued in practice, extensive literature on the effects of different types of regulation on manufacturers, such as carbon taxes, subsidies, and cap and trade, has emerged (Zhou and Huang, 2016; Liao and Shi, 2018).

Many scholars have discussed that government intervention and policies are critical to the environmental performance of manufacturers. Hafezi and Zolfagharinia (2018) proposed that governments should impose regulations with caution as firms may opt for a strategy that provides a larger profit at the expense of total environmental performance. Hafezalkotob (2015) developed a competition model of two green and regular supply chains under environmental protection and revenue seeking policies of government. Peng et al. (2019) investigated the main causes of environmental regulation failure and found that governments should strengthen regulations from the aspects of improving laws, establishing a monitoring system, and innovating incentive and constraint mechanisms. Murali et al. (2019) developed a framework for studying the impact of voluntary eco-labels and mandatory environmental regulation on green product development among competing firms.

Most studies from the perspective of governments have been performed on the policy of regulation and analyzing whether the tools and mechanisms effectively influence the decisions of manufacturers. Even though Gouda et al. (2016) addressed a manufacturer's product quality choice problem and proposed a composite regulatory solution, the study only discussed traditional and environmental qualities. However, the especially for manufacturers to fraud have rarely been analyzed. Ensuring that the quality of NEVs meets the requirements is the premise of developing them. Hence, when considering the incidence of fake and inferior behaviors, it is necessary to examine how to achieve effective regulation policy in its game with both manufacturers and governments.

### Behavior Strategy Selection of Stakeholders

The behavior strategy selection of stakeholders is becoming increasingly prominent in research (Geng et al., 2021). Thus far, the literature is rich on green supply chains and renewable energy development in terms of different stakeholders (Xu X. F. et al., 2019). Considering the manufacturers and retailers,

**TABLE 1 |** Papers most related to our research.

| Authors | Energy-saving products | Government regulation | Asymmetric information | Tripartite game |
|---|---|---|---|---|
| Rocha and Salomao (2019) | | ✓ | | |
| Zhang et al. (2020) | ✓ | ✓ | | |
| Zhao et al. (2020) | ✓ | ✓ | | |
| Sun and Zhang (2019) | | ✓ | ✓ | |
| Yang et al. (2018) | | | ✓ | ✓ |
| Feng et al. (2019) | ✓ | ✓ | | |
| Deng et al. (2013) | | | ✓ | ✓ |
| This paper | ✓ | ✓ | ✓ | ✓ |

*The symbol ✓ indicates that paper has the corresponding topic.*

Zhang et al. (2020) analyzed the optimal decisions of a green supply chain and revealed the impacted difference between manufacturer and retailer behavior on equilibrium solutions in the context of a government subsidy. Deng et al. (2013) proposed a principal-agent model to study a supply chain consisting of a manufacturer and a retailer with asymmetric information. Zhou et al. (2018) discussed a cooperative advertising and ordering issue in a two-echelon supply chain in which a risk-averse manufacturer sells a product through a risk-averse retailer. Choi et al. (2018) and Wang et al. (2019) explored how the stochastic risk preference of retailers affects the values of quick response to the supply chain and its members. Feng et al. (2019) formalized the game's theoretic model of manufacturers associated with governments in a low-carbon technology market and derived the mixed strategy Nash equilibrium. Ye et al. (2019) introduced the environmental governance cost prediction and showed that different risk coefficients have a greater impact on the investment in environmental governance. Liao et al. (2020) analyzed the effects of behavioral intention on the choice to purchase energy-saving appliances in China and revealed that behavioral intention has a significantly positive effect on the choice of consumers.

The literature mentioned above mainly indicates that behaviors between governments and manufacturers have been widely discussed. The role of consumers and their environmental awareness are also investigated as critical factors for manufacturers to promote ESPs (Li and Li, 2016; Liao and Shi, 2018). Very little research has looked at the impact of consumers' behavior on NEVs when governments provide regulations for manufacturers. In particular, the incorporating role of synergistic regulatory effects should be discussed in depth. This study takes the perspective of associated production, consumption, and regulation in considering the whole process of NEVs. This could have a great impact on the efficiency of environmental regulation.

## Differential Game Models in ESPs

Game theory provides a powerful tool with which to unpack the interactive strategies of governments and manufacturers, focusing particularly on the conflict of interests between them under environmental regulation. Assuming the participant is completely rational, Zu et al. (2018) used a Stackelberg differential game with three progressive environmental regulation situations to consider a two-echelon supply chain consisting of one manufacturer and one supplier that tries to increase sustainable profits by making efforts to reduce $CO_2$ emission. Zhao et al. (2020) constructed a perfect game model including the subsidy policies of governments, manufacturers, and customers. During their study, the crucial influencing factors of the governments' and manufacturers' strategies were discussed. Zhou and Huang (2016) modeled a three-stage game and presented an optimal design of the contracts under government regulation for ESPs.

Recently, on the basis of the bounded rationality hypothesis, evolutionary game theory has been widely used on new energy products (Yu et al., 2020). Jiang et al. (2019) implemented a multi-agent environmental regulation strategy under Chinese fiscal decentralization by using an evolutionary game theoretical approach. Xing et al. (2017) constructed an evolutionary game of environmental regulation among local governments in China and analyzed the dynamic evolution rules of the environmental regulation strategy. Sun and Zhang (2019) divided enterprises into two types, dominant and inferior, and analyzed the evolutionarily stable strategy of heterogeneous enterprises in preventing greenwashing. Rocha and Salomao (2019) presented an evolutionary game to study the interaction between polluting firms and auditors. For a tripartite game model, Duan et al. (2016) developed two systems of dynamics-based tripartite evolutionary game models: a government environmental regulation-static punishment model and a dynamic punishment model. Liu and Xia (2020) analyzed the strategies of governments, manufacturers, and consumers according to the evolutionary game model in new energy products.

Most game models assume that information is symmetric and are mainly for two-player games, such as non-cooperative, cooperative and evolutionary game. As far as we know, few works have studied the strategy choices of tripartite coordinated regulation for NEVs with incomplete information. The manufacturer has private information about product quality, and the governments and consumers can only observe the quality level of the products posteriori. Therefore, tripartite dynamic game relationships should be discussed to explore the predicament and countermeasures of governance (Wang D. L. et al., 2020). A Bayesian game is an efficient way to solve the incomplete information game problem and is a more reasonable means to describe the strategies chosen by participants in NEVs (Yang et al., 2018). Therefore, this study attempts to explore the impacts of the interactions of governments, manufacturers, and consumers via the Bayesian game.

To highlight the innovation of this study and clearly show its difference from previous literature, we summarize the literature most related to our paper in **Table 1**.

# PROBLEM ASSUMPTIONS AND MODEL DEVELOPMENT

The game model of product quality regulation involves three participants: governments, manufacturers, and consumers. The relationships among NEVs stakeholders is shown in **Figure 1**.

Manufacturers are responsible for the production, quality, and sale of NEVs, and governments design regulation standards to motivate manufacturers to improve their quality and protect consumers.

## Problem Assumptions

Considering the reality of NEVs, such as the tripartite dynamic game model presented by Yang et al. (2019) and Xu et al. (2020), incorporating the characteristics of NEVs and regulation policies, research assumptions are proposed to conduct a better analysis of the game problem in decision-making behaviors. Additionally, the parameters involved in the models are accurately defined.

**Assumption 1.** We introduce "Nature," which is considered under different product qualities, and with the products divided into two types $i(i = H, L)$. The strategy set of nature is high-quality products and low-quality products. This is consistent with the literature, such as Cai et al. (2019). So, we only consider two types of products in the market. High-quality products refer to products that adopt energy-saving technology and meet the emission standards set by governments. For NEVs, governments propose standards that mainly include energy consumption over 100 km and emissions. Of course, manufacturers can also implement higher standards to improve product quality. The probability that manufacturers design high-quality products is $P(i = H) = \theta, \theta \in [0, 1]$. From the perspective of quality standards formulated by the government, only unqualified product standards are considered low-quality products. Thus, the probability is denoted by $P(i = L) = 1 - \theta$. The quality of NEVs is private information.

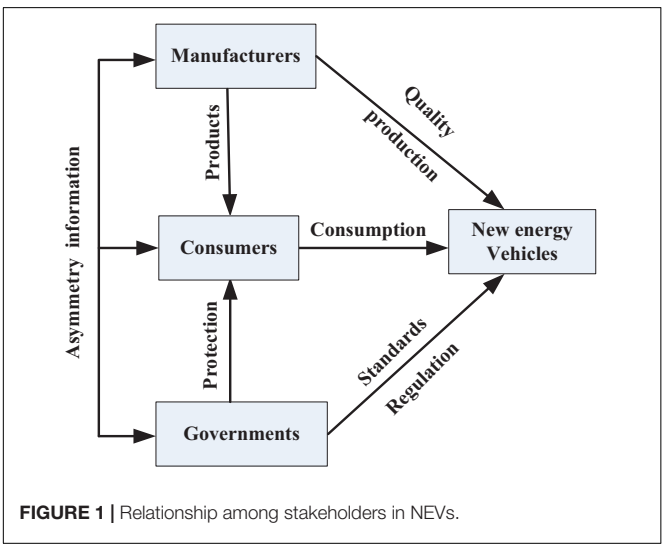



**FIGURE 1 |** Relationship among stakeholders in NEVs.

**TABLE 2 |** Parameters and implications.

| | |
|---|---|
| **For governments** | |
| $p_g$, $1 - p_g$ | Probability of government inspection or non-inspection |
| $R_g$ | Benefit of government when manufacturers sell high-quality products |
| $C_g$ | Cost of inspection |
| $F_L$ | Fines for low-quality NEVs |
| $L_{g1}$, $L_{g2}$ | Loss of non-inspection when manufacturers sell low-quality NEVs |
| **For manufacturers** | |
| $\theta, 1-\theta$ | Probability of high-or low-quality NEVs |
| $R_i$ | Benefit of high- or low-quality NEVs |
| $C_i$ | Production cost of high- or low-quality NEVs |
| $P(m_j\|H)$ $P(m_j\|L)$ | Conditional probabilities of manufacturers |
| $P(H\|m_j)$ $P(L\|m_j)$ | Posteriori probabilities of manufacturers |
| **For consumers** | |
| $P_c$, $1 - P_c$ | Probability of consumers' consumption or non-consumption |
| $R_c$ | Benefit of buying high-quality NEVs |
| $C_{ci}$ | Selection cost of high- or low-quality NEVs |

**Assumption 2.** The strategy set of manufacturers is $m_j = (m_1, m_2) =$ (sales, non-sales), where $m_1$ is the sales, and $m_2$ is the non-sales, and the corresponding probabilities are $p_m$ and $1 - p_m$. We set the benefit of the sales of different products as $R_i$, and the production cost as $C_i$. If the governments choose an inspection strategy, the low-quality products produced by manufacturers will be punished, and the penalty is designated as $F_L$. For example, on 7 February 2017, the Ministry of Industry and Information Technology issued a new ticket for NEV fraudulent compensation manufacturers and announced administrative penalty decisions for seven fraudulent repair manufacturers, including Jinhua Youth Automobile and Chongqing Lifan.

**Assumption 3.** The behavior set of the governments for environmental regulation is $b_j = (b_1, b_2) =$ (inspection, non-inspection). The probabilities of governments inspection and non-inspection are $P_g$ and $1 - P_g$, respectively. We define the inspection cost as $C_g$, as a constant. Here, we assume that $F_L > C_g$. If manufacturers fail to take green actions, then the governments should bear the loss of non-inspection. There are two scenarios. The first is that manufacturers produce low-quality products and sell them, but governments do not inspect low-quality products. At this time, even though consumers are not buying, there are inferior products on the market. Therefore, governments will be punished by the higher authorities for the lack of inspection. The losses that the government should bear is denoted by $L_{g2}$. In the second scenario, consumers buy inferior products. In addition to being punished by superiors, governments can also suffer loss of reputations as consumers distrust them because of poor regulation. $L_{g1}$ is the losses of governments when consumers purchase NEVs. Here, we know that $L_{g1} > L_{g2}$.

**Assumption 4.** The strategy set of consumers is $x_j = (x_1, x_2) =$ (consumption, non-consumption). The benefit of buying high-quality NEVs is represented by $R_c$. The cost of consumers' investment in the selection and confirmation of product quality is given by $C_{ci}$ e.g., time and effort. To purchase high-quality NEVs, consumers compare the performance of different brands and products.

**Assumption 5.** The conditional probabilities of the products are $P(m_j|H)$ and $P(m_j|L)$. The consumers will update their beliefs by following the Bayesian rule, where the posteriori probabilities

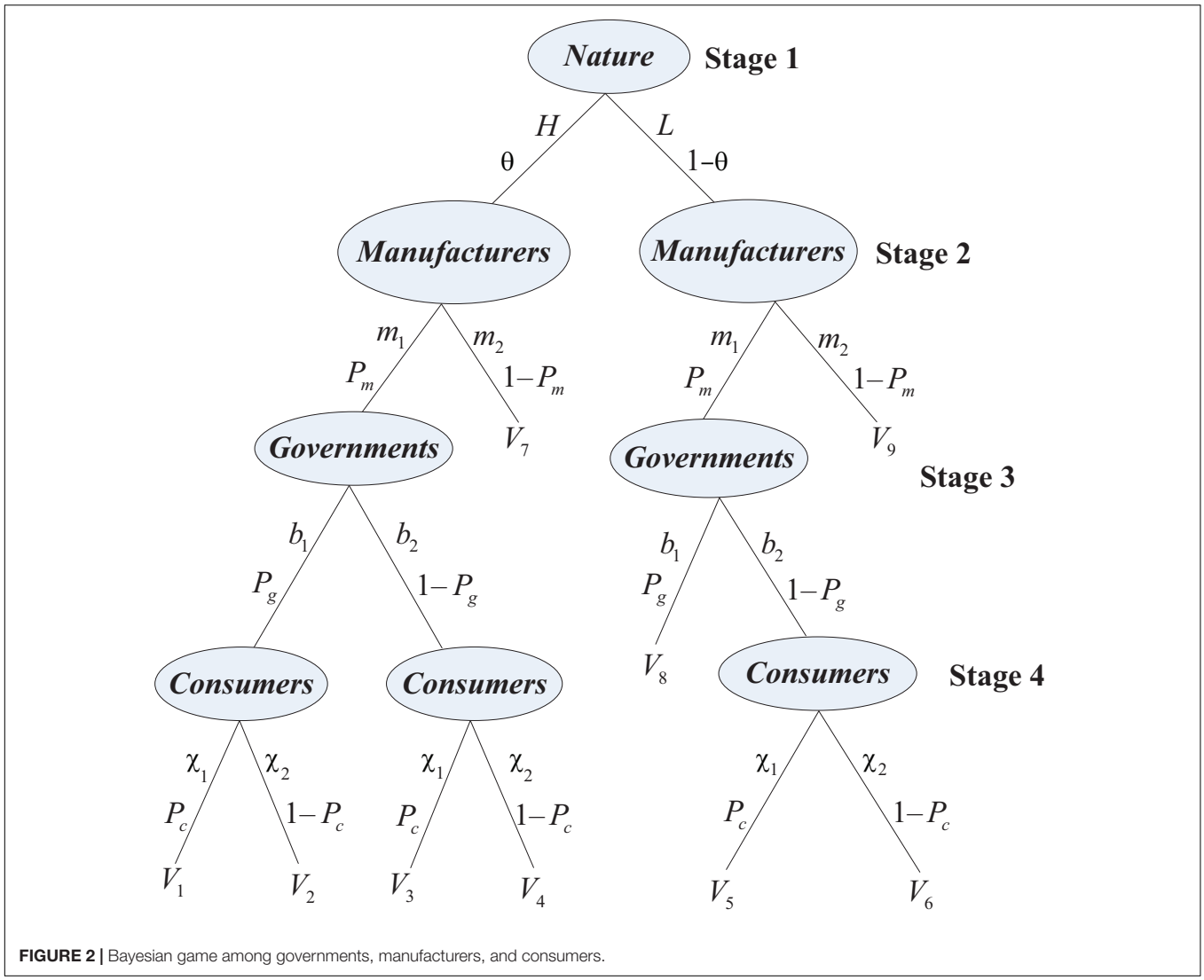



**FIGURE 2 |** Bayesian game among governments, manufacturers, and consumers.

are denoted by $P(H|m_j)$ and $P(L|m_j)$. The above parameters and implications are summarized in **Table 2**.

## Model Construction

In this section, the game process among governments, manufacturers and consumers is analyzed. **Figure 2** illustrates the game tree of the three players on the basis of their strategic combinations. The Bayesian game process includes four stages, which can be clearly expressed.

Referring to the game built by Xu L. et al. (2019), we analyze the regulation of the circulation market of product quality, which is a prior-regulation. We set the game order beginning in stage 1. In this stage, the game participant "Nature" randomly picks a possible type of product (for example, high-product NEVs have the probability θ). In stage 2, manufacturers completely know their types of products, while the other two game players do not. Manufacturers then proceed to choose the action from the strategy set. In stage 3, governments receive the signal of the manufacturers' action, speculate on the types of products, and then determine further action that would gain them the most profit. In stage 4, consumers finally choose one strategy.

Overall, the payoffs of governments, manufacturers, and consumers can be obtained according to the aforementioned assumptions. **Table 3** gives the payoffs in different nodes.

**TABLE 3 |** Payoffs of governments, manufacturers, and consumers.

| Node | Manufacturers | Governments | Consumers |
|---|---|---|---|
| $V_1$ | $R_H - C_H$ | $R_g - C_g$ | $R_c - C_{cH}$ |
| $V_2$ | $-C_H$ | $R_g - C_g$ | 0 |
| $V_3$ | $R_H - C_H$ | $R_g$ | $R_c - C_{cH}$ |
| $V_4$ | $-C_H$ | $R_g$ | 0 |
| $V_5$ | $R_L - C_L$ | $-L_{g1}$ | $-C_{cL}$ |
| $V_6$ | $-C_L$ | $-L_{g2}$ | 0 |
| $V_7$ | $-C_H$ | 0 | 0 |
| $V_8$ | $-C_H - F_L$ | $F_L - C_g$ | 0 |
| $V_9$ | $-C_L$ | 0 | 0 |

## Equilibrium Solution

The equilibrium solution of the dynamic game can be solved by using reverse induction (Chen et al., 2011). We analyze this game by considering the condition of separating equilibrium. If "Nature" chooses the high-quality NEVs, then the manufacturers will implement the pure strategy, which means $P(m_1|H) = 1$. If low-quality NEVs are chosen, then the manufacturers will choose to sell with the probability of δ, that is, $P(m_1|L) = \delta$. According to the Bayesian rule, we can obtain the posteriori probabilities $P(H|m_1)$ and $P(L|m_1)$ as follows:

$$P(H|m_1) = \frac{P(m_1|H) * P(H)}{P(m_1|H) * P(H) + P(m_1|L) * P(L)} = \frac{\theta}{\theta + (1-\theta)\delta}$$

$$P(L|m_1) = \frac{P(m_1|L) * P(L)}{P(m_1|H) * P(H) + P(m_1|L) * P(L)} = \frac{(1-\theta)\delta}{\theta + (1-\theta)\delta}$$

If manufacturers choose to sell low-quality products, then the governments will choose to conduct or not conduct an inspection, and the expected utility functions of the manufacturers will be denoted by $E_m(b_1, m_1|L)$ and $E_m(b_2, m_1|L)$, respectively. These expected utility functions are:

$$E_m(b_1, m_1|L) = -C_L - F_L,$$

$$E_m(b_2, m_1|L) = p_c(R_L - C_L) + (1-p_c)(-C_L).$$

Then, the whole expected utility function of manufacturers that sell low-quality NEVs can be obtained:

$$E_m(m_1|L) = p_g * E_m(b_1, m_1|L) + (1-p_g) * E_m(b_2, m_1|L)$$

When manufacturers select the non-sales strategy, we derive the expected utility functions of manufacturers as $E_m(m_2|L) = -C_L$.

The expected utility $E_m(m_1|L)$ equals the expected profit $E_m(m_2|L)$, and $E_m(m_1|L) = E_m(m_2|L)$ yields the Bayesian equilibrium probability of consumption $p_c^*$, that is, $p_c^* = \frac{p_g F_L}{(1-p_g)R_L}$. Therefore, the equilibrium probability of the non-consumption of consumers is $1 - p_c^* = 1 - \frac{p_g F_L}{(1-p_g)R_L}$.

When $p_c \in (p_c^*, 1]$, and $E_m(m_1|L) > E_m(m_2|L)$, $p_m = 1$ can become the optimal strategy, and the optimal response strategy for manufacturers is to sell low-quality NEVs. When $p_c \in [0, p_c^*)$, the optimal strategy of manufacturers is $m_1$, that is, $p_m = 0$. If $p_c = p_c^*$, then it makes no difference which strategy the manufacturers select. Hence, the different scenarios can be expressed as follows:

$$P_m = \begin{cases} 1 & if \quad P_c > P_c^* \\ [0,1] & if \quad P_c = P_c^* \\ 0 & if \quad P_c < P_c^* \end{cases}.$$

The equilibrium probability of consumers is affected not only by fines for low-quality NEVs, but also by the benefit of low-quality NEVs. Additionally, it is still influenced by the probability of government regulation.

Similarly, for governments, the expected utility when they choose inspection and non-inspection are denoted by $E_g(b_1|m_1)$ and $E_g(b_2|m_1)$, respectively. By using **Table 3**, we have the following equations:

$E_g(b_1|m_1) = p(H|m_1)(R_g - C_g) + p(L|m_1)(F_L - C_g) = \frac{\theta}{\theta+(1-\theta)\delta}(R_g - C_g) + \frac{(1-\theta)\delta}{\theta+(1-\theta)\delta}(F_L - C_g)$

$E_g(b_2|m_1) = p(H|m_1)R_g + p(L|m_1)p_c(-L_{g1} - L_{g2}) = \frac{\theta}{\theta+(1-\theta)\delta}R_g + \frac{(1-\theta)\delta}{\theta+(1-\theta)\delta}(-L_{g1} - L_{g2})$ With $E_g(b_1|m_1) = E_g(b_2|m_1)$, we obtain the equilibrium probability of the manufacturers selling the low-quality NEVs, δ*:

$$\delta^* = \frac{\theta C_g}{(1-\theta)[F_L - C_g + p_c(L_{g1} + L_{g2})]}.$$

According to the optimal equilibrium probability of manufacturers $\delta^* = \frac{\theta C_g}{(1-\theta)[F_L - C_g + p_c(L_{g1}+L_{g2})]}$. When $\delta > \delta^*$ and $E_g(b_1|m_1) > E_g(b_2|m_1)$, the optimal strategy is $p_g = 1$; at this point, the optimal strategy for the governments is to perform an inspection. When $\delta > \delta^*$, $p_g = 0$ is the optimal strategy, that is, the governments will choose non-inspection; if the optimal strategy is $\delta = \delta^*$, then it does not matter which strategy the governments choose. The above analysis can be expressed by:

$$P_g = \begin{cases} 1 & if \quad \delta < \delta^* \\ [0,1] & if \quad \delta = \delta^* \\ 0 & if \quad \delta > \delta^* \end{cases}.$$

The equilibrium probability shows that the optimal equilibrium probability of manufacturers is due to several factors, mainly influenced by the governments' behavioral strategies. That includes the regulation cost, fines for low-quality NEVs, and governmental losses of non-inspection when manufacturers sell low-quality NEVs.

(3) For consumers, the expected utility of consumers who choose or do not choose the consumption of NEVs will be denoted by $E_c(x_1|m_1)$ and $E_c(x_2|m_1)$ respectively. These expected utility functions are:

$$E_c(x_1|m_1) = p(H|m_1)(R_c - C_{cH}) + p(L|m_1)(1-p_g)(-C_{cL}) = \frac{\theta}{\theta+(1-\theta)\delta}(R_c - C_{cH}) + \frac{(1-\theta)\delta}{\theta+(1-\theta)\delta}(1-P_g)(-C_{cL}),$$

$$E_c(x_2|m_1) = 0.$$

When $E_c(x_1|m_1) = E_c(x_2|m_1)$, we can obtain the equilibrium probability of the governments' inspection as $P_g^* = 1 - \frac{\theta(Rc - C_{cH})}{(1-\theta)\delta C_{cL}}$. Therefore, the equilibrium probability of governments' non-inspection is $1 - P_g^* = \frac{\theta(Rc - C_{cH})}{(1-\theta)\delta\ C_{cL}}$.

When $P_g > P_g^*$ and $E_c(x_1|m_1) > E_c(x_2|m_1)$, $p_c = 1$ is the optimal consumer strategy. When this condition is met, the consumers' optimal strategy is the consumption of NEVs. Alternatively, when $P_g < P_g^*$, $p_c = 0$ becomes the optimal strategy, and the consumers will choose non-consumption. When

$P_g < P_g^*$, it does not matter which strategy the consumers choose. The above analysis can be expressed as follows:

$$P_c = \begin{cases} 1 & if \quad P_g < P_g^* \\ [0, 1] & if \quad P_g = P_g^* \\ 0 & if \quad P_g > P_g^* \end{cases}.$$

The equilibrium probability of the government's inspection is affected by consumer behavior strategy. The benefit of buying high-quality NEVs and the selection cost of high-or low-quality NEVs are important influencing factors.

# ANALYSIS OF THE EQUILIBRIUM RESULTS OF THE DYNAMIC BAYESIAN GAME

Through solving and analyzing the equilibrium solution of the tripartite game model, the most important influencing factors of strategy selection analysis are carried out for each party. The internal correlation between the game participants and strategy selection is discussed in this section.

## Equilibrium Probability of Manufacturers

The equilibrium probability of manufacturers selling the low-quality NEVs is associated with $C_g$, $F_L$, $L_{g1}$ and $L_{g2}$.

**Proposition 1:** The probability of manufacturers selling the low-quality NEVs, $\delta^*$, increases as, $C_g$, increases.

By calculating the partial derivative of $\delta^*$, we calculate that

$$\frac{\partial \delta^*}{\partial C_g} = \frac{\theta(1-\theta)[F_L + p_c(L_{g1} + L_{g2})]}{\{(1-\theta)[F_L - C_g + p_c(L_{g1} + L_{g2})]\}^2}$$

Given that $\theta \in [0, 1]$, it is concluded that $\frac{\partial \delta^*}{\partial C_g} > 0$.

From Proposition 1, an increased cost of government inspection lead to an increase im manufacturers with inferior eco-friendly operations. Because the increased cost of government inspection will reduce the probability of inspection, so it will increase the rate of low-quality products. The manufacturers do not care about the risk of deception in NEVs and degrade the quality of their products. In other words, the higher the regulation cost of NEVs, the more likely it is that manufacturers will commit fraud. Accordingly, there will be no reduction in carbon emissions.

**Proposition 2:** The probability of manufacturers, $\delta^*$, decreases as the losses $L_{g1}$, $L_{g2}$ and governmental fines, $F_L$ increase.

Given the solutions for $L_{g1}$, $L_{g2}$ and $F_L$ for the first derivatives of $\delta^*$, we have

$$\frac{\partial \delta^*}{\partial L_{g1}} = \frac{-(1-\theta)p_c}{\{(1-\theta)[F_L - C_g + p_c(L_{g1} + L_{g2})]\}^2} < 0,$$

$$\frac{\partial \delta^*}{\partial L_{g2}} = \frac{-(1-\theta)p_c}{\{(1-\theta)[F_L - C_g + p_c(L_{g1} + L_{g2})]\}^2} < 0,$$

$$\frac{\partial \delta^*}{\partial F_L} = \frac{-(1-\theta)}{\{(1-\theta)[F_L - C_g + p_c(L_{g1} + L_{g2})]\}^2} < 0.$$

Proposition 2 illustrates that whether consumers purchase NEVs, the losses of governments that affect the behavior of manufacturers are the same; when these losses increase, the problem of illegal emissions gets better. Proposition 2 also reveals that enhancing fines will decrease illegal emission behaviors. That is, governments contribute to ensuring product quality when they strengthen the punishment mechanism.

## Equilibrium Probability of Governments

The equilibrium probabilities of governments' inspection, $P_g^*$, are determined by $R_c$, $C_{cH}$ and $C_{cL}$.

**Proposition 3:** The probability of governments' inspection, $P_g^*$, decreases as the consumers' benefit of high-quality NEVs, $R_c$, and the selection cost of low-quality NEVs, $C_{cL}$, increases.

The first derivatives of the $P_g^*$ on $R_c$ and $C_{cL}$ are as follows:

$$\frac{\partial P_g^*}{\partial R_c} = \frac{-\theta}{[\theta(1-\theta)\delta C_{cL}]^2} < 0,$$

$$\frac{\partial P_g^*}{\partial C_{cL}} = \frac{-(1-\theta)\delta}{[(1-\theta)\delta C_{cL}]^2} < 0.$$

Proposition 3 shows that as the benefit of high-quality NEVs and the cost of low-quality NEVs increases, consumers will benefit from purchasing the conforming products and will buy the NEVs. Thus, the probability of governments' inspection will be reduced, and the governments will be unlikely to regulate the manufacturers. Meanwhile, when the cost of identifying low-quality products is low, it is easy for consumers to find substandard products and buy high-quality products. The probability of governments' inspection can also be decreased.

**Proposition 4:** The probability of governments' inspection, $P_g^*$, increases as the selection cost of high-quality NEVs, $C_{cH}$, increases.

The solution for $C_{cH}$ for the first derivatives of $P_g^*$ is as follows:

$$\frac{\partial P_g^*}{\partial C_{cH}} = \frac{\theta}{(1-\theta)\delta C_{cL}} > 0.$$

Proposition 4 gives the relationships between the cost of consumption and the probability of governmental inspection. As the values of $C_{cH}$ grow, $P_g^*$ will increase. More specifically, when the selection cost of buying high-quality NEVs is high, and the information asymmetry of product quality is more serious, consumers will be more likely to buy low-quality NEVs. At this point, the government will increase the probability of regulation.

## Equilibrium Probability of Consumers

The equilibrium probability of consumers is related to $F_L$ and $R_L$.

**Proposition 5:** The probability of consumers' consumption $P_c^*$ is an increasing function of $F_L$.

The first derivative of the equilibrium probability of consumers in terms of $F_L$ is obtained by:

$$\frac{\partial P_c^*}{\partial F_L} = \frac{p_g}{(1-p_g)R_L} > 0.$$

From Proposition 5, it is clear that the more the governments penalize manufacturers of substandard products, the less they

**TABLE 4 |** Changes in equilibrium probability as model parameters increase.

| Equilibrium probability | | Parameters |
|---|---|---|
| Manufacturers | $\delta^*$ | $C_g$ $F_L$ $L_{g1}$ $L_{g2}$<br>+ − − − |
| Governments | $Pg^*$ | $R_c$ $C_{CH}$ $C_{CL}$<br>− + − |
| Consumers | $Pc^*$ | $R_L$ $F_L$<br>− + |

+, *Increase;* −, *Decrease.* * *is the identification of equilibrium probability.*

will violate the rules, and thus the more likely that consumers will choose to buy NEVs. This case is good for governments because the lower the carbon emission reduction, the higher the environmental quality.

**Proposition 6:** The probability of consumers' consumption, $P_c^*$, is a negative function of $R_L$.

The first derivative of $P_c^*$ in terms of $R_L$ is as follows:

$$\frac{\partial P_c^*}{\partial R_L} = \frac{-(1-p_g)}{[(1-p_g)R_L]^2} > 0.$$

Proposition 6 reveals that the higher the low-quality products that manufacturers produce, the more likely they are to forego compliance emissions. To consumers, the probability of consumers' consumption of high-quality products will be reduced. In other words, when manufacturers make high profits by producing low-quality products, they will produce more of these products. Manufacturers have no incentive to produce high-quality products, and consumers are inclined to buy low-quality products.

In summary, the changes in equilibrium probability as the model parameters increase are presented in **Table 4**.

## CONCLUSION

In the incomplete information dynamic game, the Bayesian game model is used to describe and explain the game mechanism chosen by participants in NEVs. We provide optimal decisions for manufacturers, governments, and consumers, and analyze their key influencing factors. The results of this research reveal the following: (1) The strategy of manufacturers depends on the inspection cost, the punishment incurred, and the losses of the governments for low-quality NEVs. (2) The strategy of governments is influenced by the production cost of high-and low-quality NEVs and the benefit of the manufacturers from selling high-quality products. (3) The strategy of consumers is related to the fines and benefits for low-quality products by manufacturers. On the basis of the above conclusions, we provide the following recommendations for governments to improve the quality of NEVs.

First, governments should enhance their constraint penalty management mechanisms and punishment policies. According to Proposition 2 and Proposition 5, the fines of NEVs are an important factor affecting the game equilibrium of both manufacturers and consumers. They are the key to comprehensive environmental governance and are an effective means of encouraging manufacturers to abide by environmental policies. Additionally, in consideration of Proposition 1, reducing the cost of government regulation will improve product quality. Therefore, to realize the integration of online and offline regulations, governments can set up a regulatory platform for manufacturers. By collecting data from social media or consumers to update information from the platform, once the manufacturers' violation is discovered, the government can investigate the violation and deal with it accordingly. This is an effective way to reduce the cost of regulation especially when governments have budgetary constraints.

Second, it is also necessary for the government to enhance the ability of consumers to identify product quality. From Proposition 3 to Proposition 4, we know that reducing the cost for consumers to distinguish product quality cannot only reduce the government supervision input, but also reduce the information asymmetry of product quality. On the one hand, through timely disclosure of regulatory results of NEVs, it can provide consumers with information and channels to identify product quality. On the other hand, strengthening the product quality certification provides information support for consumers to buy products and introduces clear compensation measures for low-quality products to better protect the legitimate rights of consumers.

Third, the manufacturers' production costs of NEVs should be reduced. It is well known that producing NEVs with carbon emission reduction is often extremely costly for manufacturers (Gouda et al., 2016). In the market competition, the NEVs industry needs to rely on support from the government (Olson, 2018). To reduce the high levels of risk and uncertainty, governments should make preferential policies to support manufacturers. For example, the appropriate subsidies can provide incentives and encourage manufacturers to develop low-carbon emission technologies. On the contrary, manufacturers that produce high-carbon emission products need to pay taxes, such as pollution taxes.

Lastly, there should be an increase in consumers' returns on NEVs. Consumers must realize some benefits, including reduced environmental degradation, for them to be willing to take on the burden of additional costs. Thus, many governments have subsidized consumers who purchase NEVs. For example, in China, consumers can receive certain government subsidies when they purchase NEVs. Meanwhile, through improved consumer environmental awareness, consumers are now more familiar with carbon-emission products and are willing to pay extra for NEVs.

The study expounds on the dynamic interaction among government regulation, manufacturer production, and consumer consumption. Some recommendations on the formulation and improvement of environmental policies for governments are provided. Several future research directions for this study are likewise presented. The behaviors of governments, manufacturers, and consumers can be analyzed through empirical research, which may provide more objective and practical results. Extending the tripartite game to

the case with heterogeneous game players could also be a challenging and interesting direction for future research.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design, read, and approved the final manuscript.

## FUNDING

This research was funded by the Shandong Provincial Natural Science Foundation (Grant Nos. ZR2017QG003 and ZR2016GM10). National Natural Science Foundation of China (Grant No. 41801173), Shandong Provincial Social Science Foundation (Grant No. 17DGLJ09), and the China Postdoctoral Science Foundation (Grant No. 2017M622265).

## ACKNOWLEDGMENTS

We gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## REFERENCES

Cai, D., Guo, C. X., and Tan, Y. (2019). Design of incentive contract for technological innovation of new energy vehicles with asymmetric information. *Int. J. Environ. Res. Public Health* 16:4544. doi: 10.3390/ijerph16224544

Chander, P., and Muthukrishnan, S. (2015). Green consumerism and pollution control. *J. Eco. Behav. Organization* 114, 27–35. doi: 10.1016/j.jebo.2015.02.013

Chen, C. J. P., Li, Z. Q., Su, X. J., and Sun, Z. (2011). Rent-seeking incentives, corporate political connections, and the control structure of private firms: chinese evidence. *J. Corp. Finance* 17, 229–243. doi: 10.1016/j.jcorpfin.2010.09.009

Chen, Y. S., and Chang, C. H. (2013). Greenwash and green trust: the mediation effects of green consumer confusion and green perceived risk. *J. Business Ethics* 114, 489–500. doi: 10.1007/s10551-012-1360-0

Choi, T. M., Zhang, J., and Cheng, T. C. E. (2018). Quick response in supply chains with stochastically risk sensitive retailers. *Decision Sci.* 49, 932–957. doi: 10.1111/deci.12303

Deng, X. X., Xie, J. X., and Xiong, H. C. (2013). Manufacturer-retailer contracting with asymmetric information on retailer's degree of loss aversion. *Int. J. Produc. Eco.* 142, 372–380. doi: 10.1016/j.ijpe.2012.12.013

Duan, W., Li, C., Zhang, P., and Chang, Q. (2016). Game modeling and policy research on the system dynamics-based tripartite evolution for government environmental regulation. *Cluster Comput.* 19, 2061–2074. doi: 10.1007/s10586-016-0642-1

Feng, W., Ji, G. J., and Pardalos, P. M. (2019). Effects of government regulations on Manufacturer's behaviors under carbon emission reduction. *Environ. Sci. Pollu. Res.* 26, 17918–17926. doi: 10.1007/s11356-017-0891-4

Geng, J. B., Du, Y. J., Ji, Q., and Zhang, D. Y. (2021). Modeling return and volatility spillover networks of global new energy companies. *Renewable Sustain. Energy Rev.* 135:16. doi: 10.1016/j.rser.2020.110214

Gouda, S. K., Jonnalagedda, S., and Saranga, H. (2016). Design for the environment: Impact of regulatory policies on product development. *Eur. J. Operat. Res.* 248, 558–570. doi: 10.1016/j.ejor.2015.07.043

Guyader, H., Ottosson, M., and Witell, L. (2017). You can't buy what you can't see: Retailer practices to increase the green premium. *J. Retail. Consumer Serv.* 34, 319–325. doi: 10.1016/j.jretconser.2016.07.008

Hafezalkotob, A. (2015). Competition of two green and regular supply chains under environmental protection and revenue seeking policies of government. *Comput. Indus. Eng.* 82, 103–114. doi: 10.1016/j.cie.2015.01.016

Hafezi, M., and Zolfagharinia, H. (2018). Green product development and environmental performance: Investigating the role of government regulations. *Int. J. Produc. Eco.* 204, 395–410. doi: 10.1016/j.ijpe.2018.08.012

Hammami, R., Nouira, I., and Frein, Y. (2018). Effects of customers' environmental awareness and environmental regulations on the emission intensity and price of a product. *Decision Sci.* 49, 1116–1155. doi: 10.1111/deci.12302

Ji, Q., Zhang, D. Y., and Kutan, A. M. (2020). Energy financialization, risk and challenges. *Int. Rev. Financial Anal.* 68:101485. doi: 10.1016/j.irfa.2020.101485

Jiang, K., You, D., Merrill, R., and Li, Z. (2019). Implementation of a multi-agent environmental regulation strategy under chinese fiscal decentralization: an evolutionary game theoretical approach. *J. Cleaner Produc.* 214, 902–915. doi: 10.1016/j.jclepro.2018.12.252

Li, X., and Li, Y. J. (2016). Chain-to-chain competition on product sustainability. *J. Cleaner Produc.* 112, 2058–2065. doi: 10.1016/j.jclepro.2014.09.027

Liao, X. C., Shen, S. V., and Shi, X. P. (2020). The effects of behavioral intention on the choice to purchase energy-saving appliances in china: the role of environmental attitude, concern, and perceived psychological benefits in shaping intention. *Energy Efficiency* 13, 33–49. doi: 10.1007/s12053-019-09828-5

Liao, X. C., and Shi, X. P. (2018). Public appeal, environmental regulation and green investment: evidence from china. *Energy Policy* 119, 554–562. doi: 10.1016/j.enpol.2018.05.020

Liu, C. Y., and Xia, T. S. (2020). Strategy analysis of governments and new energy product manufacturers and consumers based on evolutionary game model. *Soft Comput.* 24, 6445–6455. doi: 10.1007/s00500-019-04571-6

Liu, Z. G., Anderson, T. D., and Cruz, J. M. (2012). Consumer environmental awareness and competition in two-stage supply chains. *Eur. J. Operat. Res.* 218, 602–613. doi: 10.1016/j.ejor.2011.11.027

Murali, K., Lim, M. K., and Petruzzi, N. C. (2019). The effects of ecolabels and environmental regulation on green product development. *Manuf. Serv. Oper. Manag.* 21, 519–535. doi: 10.1287/msom.2017.0703

Olson, E. L. (2018). Lead market learning in the development and diffusion of electric vehicles. *J. Cleaner Produc.* 172, 3279–3288. doi: 10.1016/j.jclepro.2017.10.318

Peng, B., Wang, Y., Elahi, E., and Wei, G. (2019). Behavioral game and simulation analysis of extended producer responsibility system's implementation under environmental regulations. *Environ. Sci. Pollu. Res.* 26, 17644–17654. doi: 10.1007/s11356-019-05215-w

Rocha, A. B. D., and Salomao, G. M. (2019). Environmental policy regulation and corporate compliance in evolutionary game models with well-mixed and structured populations. *Eur. J. Operat. Res.* 279, 486–501. doi: 10.1016/j.ejor.2019.05.040

Shen, B., Choi, T. M., and Minner, S. (2019). A review on supply chain contracting with information considerations: information updating and information asymmetry. *Int. J. Produc. Res.* 57, 4898–4936. doi: 10.1080/00207543.2018.1467062

Sun, Z. Y., and Zhang, W. W. (2019). Do government regulations prevent greenwashing? an evolutionary game analysis of heterogeneous enterprises. *J. Cleaner Produc.* 231, 1489–1502. doi: 10.1016/j.jclepro.2019.05.335

Wang, D. L., Liu, Y. F., Wang, Y. D., Shi, X. P., and Song, X. F. (2020). Allocation of coal de-capacity quota among provinces in china: a bi-level multi-objective combinatorial optimization approach. *Energy Eco.* 87:104709. doi: 10.1016/j.eneco.2020.104709

Wang, L. Y., Ye, M. H., Ma, S. S., and Sha, Y. P. (2019). Pricing and coordination strategy in a green supply chain with a risk-averse retailer. *Math. Prob. Eng.* 2019, 1–12. doi: 10.1155/2019/7482080

Wang, Y., Zhang, D. Y., Ji, Q., and Shi, X. P. (2020). Regional renewable energy development in china: a multidimensional assessment. *Renewable Sust. Energy Rev.* 124:109797. doi: 10.1016/j.rser.2020.109797

Xing, L., Yao, W., and Li, C. (2017). Evolution of environmental regulation strategy among local goverments and its impact on reginal ecological efficifency in china. *Environ. Eng. Manag. J.* 16, 2425–2433.

Xu, L., Zhang, Q., and Shi, X. P. (2019). Stakeholders strategies in poverty alleviation and clean energy access: a case study of china's PV poverty alleviation program. *Energy Policy* 135:13. doi: 10.1016/j.enpol.2019.111011

Xu, L., Zhang, Q., Wang, K. Y., and Shi, X. P. (2020). Subsidies, loans, and companies' performance: evidence from china's photovoltaic industry. *Appl. Energy* 260:114280. doi: 10.1016/j.apenergy.2019.114280

Xu, X. F., Wei, Z. F., Ji, Q., Wang, C. L., and Gao, G. W. (2019). Global renewable energy development: Influencing factors, trend predictions and countermeasures. *Res. Policy* 63:15. doi: 10.1016/j.resourpol.2019.101470

Yang, D. X., Qiu, L. S., Yan, J. J., Chen, Z. Y., and Jiang, M. X. (2019). The government regulation and market behavior of the new energy automotive industry. *J. Cleaner Produc.* 210, 1281–1288. doi: 10.1016/j.jclepro.2018.11.124

Yang, J., Tian, Z. H., Ma, K., Wang, C. S., and Liu, Z. X. (2018). A Bayesian game approach for noncooperative pricing among multiple utility companies in smart grid. *Ieee Access* 6, 68576–68585. doi: 10.1109/access.2018.2879882

Yang, M. Y., Chi, Y. Y., Mamaril, K., Berry, A., Shi, X. P., and Zhu, L. M. (2020). Communication-based approach for promoting energy consumer switching: some evidence from ofgem's database trials in the united kingdom. *Energies* 13:5179. doi: 10.3390/en13195179

Ye, F. F., Yang, L. H., and Wang, Y. M. (2019). A new environmental governance cost prediction method based on indicator synthesis and different risk coefficients. *J. Cleaner Produc.* 212, 548–566. doi: 10.1016/j.jclepro.2018.12.029

Yu, P., Zhang, J., Yang, D. F., Lin, X., and Xu, T. Y. (2019). The evolution of china's new energy vehicle industry from the perspective of a technology-market-policy framework. *Sustainability* 11:1711.

Yu, X. R., Dong, G. L., and Liu, C. Y. (2020). The tripartite regulation game of carbon financial products based on the prospect theory. *Front. Environ. Sci.* 8:610732. doi: 10.3389/fenvs.2020.610732

Zhang, L. H., Zhou, H., Liu, Y. Y., and Lu, R. (2019). Optimal environmental quality and price with consumer environmental awareness and retailer's fairness concerns in supply chain. *J. Cleaner Produc.* 213, 1063–1079. doi: 10.1016/j.jclepro.2018.12.187

Zhang, Z. Y., Fu, D. X., and Zhou, Q. (2020). Optimal decisions of a green supply chain under the joint action of fairness preference and subsidy to the manufacturer. *Discrete Dynamics Nat. Soc.* 2020, 1–18. doi: 10.1155/2020/9610503

Zhao, N., Xia, T. S., Yu, T., and Liu, C. Y. (2020). Subsidy-related deception behavior in energy-saving products based on game theory. *Front. Energy Res.* 7:154. doi: 10.3389/fenrg.2019.00154

Zhou, W. H., and Huang, W. X. (2016). Contract designs for energy-saving product development in a monopoly. *Eur. J. Oper. Res.* 250, 902–913. doi: 10.1016/j.ejor.2015.10.024

Zhou, Y. W., Li, J. C., and Zhong, Y. G. (2018). Cooperative advertising and ordering policies in a two-echelon supply chain with risk-averse agents. *Omega-Int. J. Manag. Sci.* 75, 97–117. doi: 10.1016/j.omega.2017.02.005

Zu, Y. F., Chen, L. H., and Fan, Y. (2018). Research on low-carbon strategies in supply chain with environmental regulations based on differential game. *J. Cleaner Produc.* 177, 527–546. doi: 10.1016/j.jclepro.2017.12.220

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Liu, Song, Chang and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 14 April 2021
doi: 10.3389/fenvs.2021.666362

# Symmetric and Asymmetric Impact of Poverty, Income Inequality, and Population on Carbon Emission in Pakistan: New Evidence From ARDL and NARDL Co-Integration

*Salim Khan*[1,2†] *and Wang Yahong*[1*†]

[1] School of Tourism and Management, Zhengzhou University, Zhengzhou, China, [2] School of Business, Zhengzhou University, Zhengzhou, China

OPEN ACCESS

**Edited by:**
*Xunpeng Roc Shi, University of Technology Sydney, Australia*

**Reviewed by:**
*Festus Victor Bekun, Gelişim Üniversitesi, Turkey*
*Muhammad Kamran Khan, Northeast Normal University, China*
*Abbas Ali Chandio, Sichuan Academy of Agricultural Sciences, China*

***Correspondence:**
*Wang Yahong*
*wyahong2009@zzu.edu.cn*

*†These authors have contributed equally to this work*

**Specialty section:**
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

**Received:** *10 February 2021*
**Accepted:** *16 March 2021*
**Published:** *14 April 2021*

**Citation:**
*Khan S and Yahong W (2021) Symmetric and Asymmetric Impact of Poverty, Income Inequality, and Population on Carbon Emission in Pakistan: New Evidence From ARDL and NARDL Co-Integration. Front. Environ. Sci. 9:666362. doi: 10.3389/fenvs.2021.666362*

Several researchers have studied the relationship between poverty and environmental degradation, as these concerns are remained at top priority in achieving Sustainable Development Goals (SDGs). However, the symmetric and asymmetric impact of poverty and income inequality along with population and economic growth on carbon emissions ($CO_2e$) has not been studied in the case of Pakistan. For this purpose, the short and long-run impact of poverty, income inequality, population, and GDP per capita on $CO_2e$ investigated by applying the Autoregressive Distributive Lag (ARDL) along with Non-linear Autoregressive Distributive Lag (NARDL) co-integration approach in the context of Pakistan for period 1971–2015. The symmetric results of the current study show poverty and population density along with GDP per capita increase carbon emissions in both the short and long-run, while income inequality has no impact on carbon emissions in the short-run. While in the long-run the symmetric results show that income inequality weakens environmental degradation in terms of carbon emissions. The analysis of NARDL also supports the results obtained from ARDL and suggests a positive effect of poverty, population, and economic growth on carbon emission in Pakistan. The empirical findings of the current study provide policy implications in light of the United Nation's SDGs for the development of Pakistan.

**Keywords: environmental degradation, poverty, income inequality, autoregressive distributed lag, non-linear ARDL, carbon emission**

## INTRODUCTION

Increasing environmental degradation in terms of carbon dioxide emissions, ($CO_2e$) is a major threat to the quality of the environment and for the sustainable development of many countries. The ongoing environmental degradation has placed numerous countries under pressure to decrease $CO_2e$ and to focus on low-carbon emitter economies. As a result, many small and large series of agreements, such as the "Paris Agreement," "Kyoto Protocol," and the "United Nations Framework Convention on Climate Change" (UNFCCC) have been made among different

developing and developed countries to control and decrease $CO_2e$. The reduction of $CO_2e$ has become important in recent years because many industrialized and developed countries have applied their policies to mitigate $Co_2e$ (Zhang et al., 2014; Adusah-Poku, 2016; Yang et al., 2017). However, the ongoing concerns for sustainable development have increased over the years as the quality of the environment has declined because of economic activities. Environmental changes and alleviation of poverty are the main concerns and agendas in the sustainable development goals (SDGs) of the UN. All the countries around the globe, including developing countries, have decided to release mankind from extreme poverty and to provide them with clean and safe atmosphere (Maji, 2019). To achieve higher economic growth, mostly less developed countries boost their economic activities to reduce extreme poverty by encouraging production units and the level of industrialization. However, the main aim of economic activities to accelerate economic growth, which raises greater energy demand that further leads to a rise in $CO_2e$, adversely affect sustainable development and the well-being of humans (Sani et al., 2017; Danish, 2019; Qamruzzaman and Jianguo, 2020). Sustainable development and economic growth are supportive to the alleviation of extreme poverty, particularly, if they are not associated with environmental degradation and income inequality. Numerous research studies (Baloch et al., 2020) have explored the nexus between $CO_2e$ and income inequality and advocated that a rise in the level of income gap depreciates the quality of the environment at a certain point. Ultimately, from the perspective of underdeveloped nations like Pakistan, the nexus between $CO_2e$ and income inequality might decrease poverty, but at the same time, it increases environmental degradation (Grunewald et al., 2017). However, the association of extreme poverty and income inequality along with population, economic growth and environmental pollution has been overlooked in many countries, particularly, in developing countries like South Asia, including Pakistan. Many studies have focused mainly on other environmental issues in the developed countries, while the empirical studies on developing economies like Pakistan, to the extent of past literature reviews, are very few.

South Asia is economically less developed and one of the poorest regions of the world. Many developing countries of South Asia face extreme poverty, income inequality, population growth, and environmental degradation. It is also observed that less developed countries (LDCs) are more vulnerable to achieve the targets of SDGs as compared to other developed countries (DCs). We have concentrated on the economy of Pakistan because Pakistan is continuously struggling to improve its economic growth and development, reduce all types of inequalities (particularly income), eradicate extreme poverty, and to control population growth so as to raise the quality of life of its impoverished people. Every productive activity aims to diminish extreme poverty which is responsible for high production and high consumption of energy and thus $CO_2e$ in Pakistan, the energy supply in Pakistan, is based on fossil fuels (Baloch et al., 2018). With the higher energy consumption, Pakistan is facing serious issues regarding environmental degradation since the emission of deadly and toxic wastes contaminate the quality of water and air. Furthermore, Pakistan has an agricultural economy and the increasing usage of insecticides (pesticides) badly affects the quality of water (groundwater). Besides, the higher degree of deforestation and rapid urbanization also increase the environmental degradation. For the last many years, Pakistan and other developing countries have continued to post a substantial increase in environmental degradation in terms of $CO_2e$ and have postulated very severe threats to the quality of the environment. Among other developing countries, Pakistan only accounts for 0.70% of the world's $CO_2e$ in 2013. As reported by the International Energy Agency (IEA) in 2013, the total production of $CO_2e$ is more than 321.2 million tons in Pakistan, while this is augmented more than 2-fold from about 151 million tons during the nineteen-nineties (Heubaum and Biermann, 2015). However, the contribution of Pakistan to Greenhouse emission (GHS) is not very high, but the emission of carbon has been increasing severely and thus has become the main problem to environmental degradation.

Moreover, despite the current climate and environmental changes in Pakistan, the country has still faced extreme poverty, a dramatic increase in income inequality, and high population growth. The negative environmental changes might place a negative and significant impact on the economy, as many people in the developing countries are mostly in a vulnerable position and are suffering from the absence of human capital (education and health) to alleviate the negative impacts of environmental degradation (Awodumi and Adewuyi, 2020; Yahong and Khan, 2021). The Pakistani economy is gradually growing, but it is still on the list of the poorest country in the world. From the prevailing economic situations, it is observed that the extreme poverty level may not be alleviated in Pakistan by 2030, but it can be reduced through equal distribution of income and continuous economic growth. Most notably, it is also recorded that Pakistan could not achieve the target of poverty reduction and millennium development goals (MDGs), because of the extending income gap (Abdullah et al., 2015; Arshed et al., 2020). Although, struggles are being made against poverty alleviation, many regions in the country are static and suffering from extreme poverty due to lack of basic needs, such as education, health, and clean drinking water, as well as low income of <$1.9 a day (Munir and Khan, 2014; Akbar et al., 2018). According to the World Bank (Khan M. K. et al., 2020), a decreasing tendency is seen in the poverty headcount ratio (HCR), subsequently after 1995, but millions of people are still living below the extreme poverty line. For instance, according to Global Consumption and Income Project (Lahoti et al., 2015), inequality in terms of the Gini coefficient has risen from 22% in 1960 to 43% in 2015.

In the meantime, an increasing trend in income inequality badly affects different economic and social classes of the society. The underprivileged and poor societies are very badly affected by an uneven income distribution, since they are predominantly a vulnerable part of the society. Further, the extended income gap leads to high $CO_2e$ by creating hurdles in the path of environmental policies. Likewise, it might decrease ecological security and ultimately cause greater $CO_2e$. Meanwhile, large number of studies stated that the stability of environmental quality could be a substantial factor to reduce the intensity of

poverty (Schleicher et al., 2018; Butt et al., 2020). Environment and natural resources are helpful to reduce extreme poverty levels by providing better health, access to healthy foods, and provide opportunities for income generation. It is also reported that the environmental quality can be improved by reducing extreme poverty, because most poor people living in rural areas are highly dependent on natural resources, consequently manipulating environmental resources in an unsustainable way that leads to the degradation of the quality of the environment (Qin, 2020; Baloch et al., 2021). In 1987, the World Commission published a report on Development and Environment and stated that extreme poverty degrades the quality of the environment. Moreover, the degradation of environmental quality can be controlled by reducing extreme poverty, as extreme poverty level, coupled with a growing population leads to depreciation in the quality of the environment and discourages the absorbing capacity (Rai, 2019).

In 2016, WB reported that more than half of the poor people belong to African and Asian countries. Most of them are not well-educated and (or) even uneducated and they enormously rely on natural (environmental) resources for their daily survival. Consequently, around 299 million hectares of land have been degraded due to defenestration. Many studies have been conducted on poverty–environment nexus, for instance, the work of Brundtland (1987) identified the association between poverty and environmental degradation and hypothesized that poverty is the main indicator of environmental pollution. Baloch et al. (2018) explored that relative poverty (income inequality) degrades the quality of the environment. Similarly, Khan M. I. et al. (2020) explored that economic growth, growth of the stock market, foreign direct investment (FDI), and oil consumption distrust the quality of the environment. Hence, the factual association between poverty, income inequality, and other economic variables with environmental degradation is not clear, and becomes the main objective and key motivation to conduct this study. The fulfillment of the study objective by resolving the problem of poverty and income inequality may help to mitigate environmental degradation and improve sustainability along with strong and green economic development. As discussed above, in less-developed regions, most people are poor and they may have to manipulate environmental resources blindly for their daily survival, which will degrade the quality of the environment (Grossman and Krueger, 1995). According to the work of Khan (2019), majority of emerging and developing economies in countries, such as Pakistan, Tajikistan, Bulgaria, Uzbekistan, and Costa Rica magnificently reduce their extreme poverty level without environmental degradation. Hence, the current study posits a few key research questions, such as what impact do poverty, income inequality, and population has on $CO_2e$ in the case of Pakistan? The current study tried to answer these questions by aiming to examine the nexus between poverty and income inequality along with population and economic growth on $CO_2e$. The main contribution and novelty of the current study are discussed below.

First, the current research not only focuses on the effect of absolute poverty and relative poverty on $CO_2e$, but also on the impact of population on $CO_2e$ as well. The empirical findings of the current study would provide valuable insight into how absolute as well as relative poverty influences environmental degradation in terms of $CO_2$ emissions. Secondly, the current study used population density and other control variables, in the environmental degradation model to avoid possible specification bias in the model. Thirdly, the current study has incorporated new evidence from NARDL cointegration to analyze long-run asymmetries as well as short-run dynamics between the variables used in the model (Shin et al., 2014). To do so, we will have to assess how the decrease and increase in the poverty HCR and income inequality along with population and GDP per capita (economic growth) have an impact on $CO_2e$. Besides, we have also incorporated a linear ARDL model to identify the long-run symmetries as well as the short-run association between variables (Pesaran et al., 2001).

The whole study is organized in five sections, such as Section I, Introduction and Section II, Review of the literature; similarly, Section III covers Materials and methods, while Section IV comprises the Empirical Finding along with explanation, and Section V consists of Conclusion of the current study.

## REVIEW OF THE LITERATURE

The current research study empirically analyzes the association between poverty and income inequality, along with the population and GDP per capita on $CO_2e$ in Pakistan. For a better understating, the Review of the literature was divided into three sub-sections, in which the first part contains the association between $CO_2e$ and poverty HCR; similarly, in the second and third parts of the literature review, we address the impact of income inequality along with population on $CO_2e$.

### Relationship Between Poverty and $CO_2e$

Alleviation of extreme poverty holds significance among government authorities and policymakers, as it is one of the key millennium development goals (MDGs). The impact of absolute poverty on environmental quality in the form of $CO_2e$ has been studied by many social and environmental scientists. This issue also has been acknowledged in previous research studies; however, empirical evidence backing the "poverty–$CO_2e$ relationship" is often not easy to get. There are progressive discussions in recent research studies regarding the relationship between poverty and the quality of the environment. Many researchers considered that extreme poverty is a key predictor of environmental degradation, particularly, in the LDCs. According to Koçak et al. (2019), extreme poverty along with a high population leads to feeble environmental safety strategies which raise pressure on natural means (means of production or environmental resources) and deteriorate the environment. Moreover, extreme poverty leads to deterioration in the quality of land and causes a higher amount of $CO_2e$ because poor and needy people waste natural resources (such as cutting of trees) for their economic well-being and survival. There is also a group of environmental economists who supports the two-way relationship of causality between environmental

**TABLE 1 |** Literature reviews in the tabulated form.

| References | Country | Period | Methodology | Result |
|---|---|---|---|---|
| **Recent research studies on poverty and environmental degradation** | | | | |
| Masron and Subramaniam (2019) | 50 developing countries | 2001–2014 | GMM | The findings of the study indicate that poverty is the main source of environmental degradation |
| Khan (2019) | ASEAN countries | 2007–2017 | GMM | The findings indicate that poverty has a positive and significant relationship with higher environmental degradation |
| Danish (2019) | 40 sub-Saharan African countries | 2010–2016 | DK regression and GMM | An increase in poverty has a positive effect on $CO_2$ emissions |
| Islam and Ghani (2018) | ASEAN countries | 1995–2014 | Johansen cointegration test and Granger causality test | Poverty has negative impacts on energy consumption for Malaysia |
| **Recent research studies on income inequality and environmental degradation** | | | | |
| Baloch et al. (2018) | Pakistan | 1966–2011 | ARDL bound testing approach | The study suggests that $CO_2$ emissions increase as income inequality expands |
| Grunewald et al. (2017) | 158 countries | 1980–2008 | OLS, FE, and GFE | The findings suggest that the relationship between $CO_2$ and income inequality is positive in middle-income developing countries |
| Masud et al. (2018) | ASEAN-5 | 1985–2015 | Granger causality and Panel regression tests | The results provide the bidirectional relationship between environmental sustainability and income inequality exists |
| Demir et al. (2019) | Turkey | 1963–2011 | ARDL | Environmental degradation rises as income inequality decreases. The findings also confirmed the EKC hypothesis. |

*GMM, Generalized method of moments; DK, Driscoll Kray (DK); OLS, ordinary least square; EKC, Environmental Kuznets curve; ARDL, autoregressive distributive lag; FE, fixed effect; GFE, grouped fixed effect.*

degradation and poverty. They suggested that both poor as well as rich people, contribute to the misuse of natural resources and environmental degradation, but the rich people get less affected than the impoverished because impoverished people are looked upon as victims as well as agents of environmental degradation. Few research studies had pointed out the association between environmental degradation and poverty, but these estimations are still inconclusive, while most of the studies failed to provide an inclusive and clear estimation on the nexus between poverty and environmental pollution. These most recent and related research studies on these issues are cited in **Table 1**. On behalf of the above-mentioned arguments and sparse literature, the current study provides the following four hypotheses.

**Hypothesis I**: *Poverty has an impact on $CO_2e$ emission in Pakistan.*

## Relationship Between Income Inequality and $CO_2e$

The analysis of income inequality along with economic growth and other economic indicators is extended and attracted the interest of the researchers, but the analysis of income inequality along with degradation of the environment has not been acknowledged with sufficient consideration and therefore it enters as an emerging and hot research phenomenon in the empirical literature. The literature on the income gap and environmental degradation can be classified into two parts. The first part of the literature deals with the Environmental Kuznets Curve (EKC), by analyzing the impact of the income gap on environmental degradation. Most of these studies documented that higher inequality in income degrades the environmental quality. For example, Boyce (1994) suggested that income inequality badly affects the quality of the environment. The estimation of Torras and Boyce (1998) supported the positive and direct impact of income inequality on environmental degradation. The most recent literature includes the studies of Easterlin and Angelescu (2012), Shahpouri et al. (2016), and Adelegan and Out (2020) who stated that the policies of the government regarding environmental issues deviate due to political power and income inequality; impoverished masses bear both the economic and terrible environmental costs, while the wealthy people only bear the economic costs. Moreover, the direct and positive effect of inequality on environmental degradation would further bring climate change and consequently, more polluted environment. On the other hand, various studies empirically proved that there is a negative impact of income inequality on the quality of the environment. For example, Hailemariam et al. (2020) proved that wider income inequality mitigates environmental degradation. This viewpoint relies on the marginal propensity to emit (MPE), in which environmental degradation varies as the distribution of income changes. Although these two viewpoints are different, they confirmed that environmental degradation in terms of $CO_2e$ and income inequality has a strong nexus. These most recent and related research studies on these issues are cited in **Table 1**. To verify this relationship, we hypothesized further.

**Hypothesis II**: *Income inequality has a relationship with $CO_2e$ emissions in Pakistan.*

## Relationship Between Population and $CO_2e$

The current study uses poverty and income inequality by incorporating population not just for the sensitive examination but also to evaluate its separate effect on $CO_2e$. Past research studies suggest that high population growth causes economic uncertainty, thereby negatively affecting the economic growth and development (Rit, 2014; Berger and Grabert, 2018). This type of socio-economic uncertainty also affects $CO_2e$ emissions through a multiplier effect; therefore, on one hand, it discourages investment-saving behavior, research and development (R&D), and energy consumption as well (Sun, 2014; Harting, 2019), whereas, on the other hand, it motivates the vulnerable people to find and use the cheaper source of energy, for example, fossil fuel (Khan, 2019). Thus, in light of the above discussion, we hypothesized that population has a statistically significant impact and vital role in the emission of $CO_2e$ in the case of Pakistan. For that reason, we further proposed the third hypothesis.

**Hypothesis III**: *Population has an impact on $CO_2e$ emission in Pakistan.*

# MATERIALS AND METHODS

## Model

From the discussions in the previous papers, this empirical research uses poverty, income inequality along with population, and GDP per capita to analyze the influence of poverty, income inequality, and population on environmental degradation in terms of $CO_2e$ emissions in Pakistan. While considering these relationships, we adopted the framework of empirical analysis similar to the empirical work of Hao et al. (2016) and Baek and Gweisah (2013). To verify the impact of poverty, income inequality, and population, we analyzed them with few control predictors so that we do not omit any important $CO_2e$ influencing factors. Besides, we transformed all variables into a logarithmic form to eliminate heteroscedasticity as shown in **Table 2**. Further, we estimated the following empirical model:

$$\begin{aligned} \mathrm{LnCO2et} &= \beta 0 + \beta 1\mathrm{LnPOVt} + \beta 2\mathrm{LnINEQt} + \beta 3\mathrm{POPDt} \\ &+ \beta 4\mathrm{LnGDPt} + \beta 5\mathrm{CONt} + \varepsilon t \end{aligned} \quad (1)$$

where, $\mathrm{LnCO_2et}$ represents the log form $CO_2e$ per capita, $\mathrm{LnPOV_t}$ represents poverty HCR; $\mathrm{LnINEQ_t}$ represents the Gini coefficient measure of income inequality, $\mathrm{POPD_t}$ represents population density, and $\mathrm{LnGDP_t}$ represents the GDP per capita. $\mathrm{CON_t}$ represents the control group of variables which are as follows: LnINFt indicates the logarithm of annual inflation rate, $\mathrm{LnIND_t}$ represents the density of population (per square kilometer), and $\mathrm{LnLSS_t}$ represents the logarithm of population using minimum elementary water sanitation services. Besides, small t (t) represents the years, while the error term or residual in Equation 1 is $\varepsilon_t$. The detailed list of variables along with their descriptions and definitions are elaborated in **Table 2**.

For the analysis of the current study, we used the yearly data from 1971 to 2015, where most of the data have been collected from World Development Indicators (WDI), while for the missing data, we used the methodology of linear imputation (LI). **Table 3** lists all the variables with their descriptive statistics.

**TABLE 2 |** Expected signs, description, definitions, and source of data of the variables.

| Variables | Description | Definitions | Expected Sign |
|---|---|---|---|
| *$CO_2e$ | Carbon emission | Carbon emission (measures in per capita metric tons) | |
| *POV | Poverty | Poverty HCR (percentage of total population below at $1.90 a day) | + |
| *INEQ | Income inequality | Gini index (World Bank estimate) | +/– |
| *PD | Population density | Population density (per square kilometer) | + |
| *GDP | GDP | GDP per capita (measured at fixed/constant 2010 US$) | + |
| **INF | Inflation | Inflation rate measured by CPI (yearly) | +/– |
| **IND | Industry | Industry (total percentage of value added including construction to GDP rate) | + |
| **LSS | Sanitation | Population with at least basic services of sanitation (total percentage of the population) | – |

*Data source: World Development Indictors (WDI).*
**Indicates main variables and **indicates control variables, respectively.*
*HCR, headcount ratio; GDP, gross domestic product; CPI, consumer price index.*

**TABLE 3 |** Descriptive statistics.

| Measures | $CO_2$ | POV | INEQ | GDP | POPD | INF | IND | LSS |
|---|---|---|---|---|---|---|---|---|
| Mean | −0.504 | 3.435 | 3.473 | 6.590 | 157.45 | 2.074 | 3.025 | 2.836 |
| Median | −0.406 | 3.628 | 3.487 | 6.668 | 151.88 | 2.069 | 3.027 | 3.043 |
| Maximum | −0.054 | 4.382 | 3.521 | 6.985 | 258.70 | 3.283 | 3.132 | 4.036 |
| Minimum | −1.176 | 1.386 | 3.356 | 6.110 | 77.488 | 0.927 | 2.864 | −0.878 |
| N | 45 | 45 | 45 | 45 | 45 | 45 | 45 | 45 |

## Unit Root Tests

Before assessing the long- and short-run dynamics of the model, we used unit root tests to check the level of stationarity in each series. Past studies provide the ideas of multiple tests regarding stationarity; nevertheless, in the present study, most commonly used unit root tests, namely "Phillips–Perron" (PP) along with "Augmented Dickey–Fuller" (ADF) has been incorporated. We have checked the level of stationarity at the "level" [I(0)] and the "first difference" [I(1)] of all the variables, as shown in **Table 4**. Both PP and ADF unit root testing approaches of stationarity suggest that all variables in the log form are stationary at I(1).

**TABLE 4 |** Unit root tests of stationarity.

| Methodology | | | Methodology | | |
|---|---|---|---|---|---|
| **Variables** | **ADF** | **PP** | **Variables** | **ADF** | **PP** |
| **At level** | **t-stat** | **t-state** | **First difference** | **t-stat** | **t-state** |
| $CO_2$ | −2.043 | −0.447 | *d*($CO_2$) | −10.436*** | −9.818*** |
| POV | −1.570 | −1.084 | *d(POV)* | −3.700** | −3.748** |
| INEQ | −2.788* | −1.990 | *d(INEQ)* | −5.136*** | −3.476*** |
| POPD | −2.332 | −7.066 | *d(POPD)* | −3.080** | −2.420* |
| GDP | −1.753 | −1.035 | *d(GDP)* | −5.736*** | −5.821*** |
| INF | −0.823 | −0.794 | *d(INF)* | −3.623*** | −6.092*** |
| IND | −2.761* | −2.770* | *d(IND)* | −7.597*** | −7.727*** |
| LSS | −0.934 | −1.133 | *d(LSS)* | −5.314*** | −5.317*** |

*Null hypothesis = H0; there is unit root or there is no stationarity. *, **, and *** indicate the significance level (10, 5, and, 1%, respectively).*

**TABLE 5 |** Optimal lag selection criteria.

| Lags | LogL | [a]LR | [a]FPE | [a]AIC | [a]SC | [a]HQ |
|---|---|---|---|---|---|---|
| Lag = 0 | −41.94594 | NA | 1.49e-09 | 2.378378 | 2.709363 | 2.499697 |
| Lag = 1 | 418.1538 | 723.0139 | 1.01e-17 | −16.48352 | −13.50465 | −15.39164 |
| Lag = 2 | 562.4540 | 171.7859 | 3.10e-19 | −20.30733 | −14.68059 | −18.24491 |
| Lag = 3 | 707.0174 | 117.0275* | 1.92e-20* | −24.14369* | −15.86907* | −21.11071* |

*aLR, aFPE, aAIC, aSC, and aHQ indicate lag selection criteria.*
**indicates optimal lag selection criterion.*

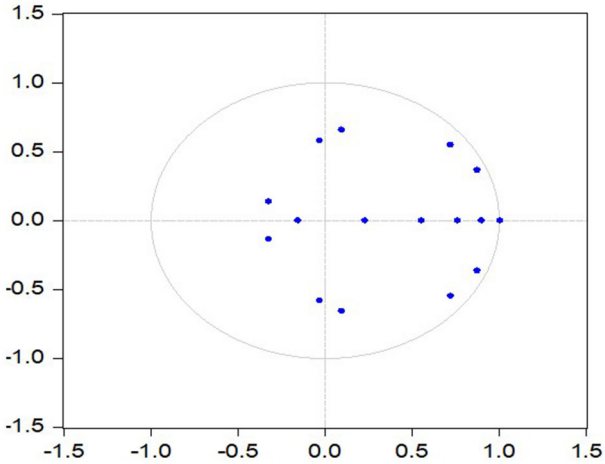



**FIGURE 1 |** Representation of polynomial graph for selection of optimal lags.

## Lag Selection Criteria Tests

For the analysis of cointegration, lag lengths are very essential. **Table 5** illustrates the findings for optimal lag selection criteria. All criteria of optimal lags selection (LR = sequential modified likelihood ratio, FPF = final prediction error, AIC = akaike information criterion, SC = Schwarz information criterion, HQ = Hannan–Quinn information criterion) provide the lag order length of 3 for the estimation of cointegration. Additionally, lag order selection criteria (under VAR = vector autoregressive) estimates have also been established as shown in **Figure 1**. All small dots within the circle of polynomial graph confirm the suitability of lag selection of 3 for the policy implications and decision making.

## Autoregressive Distributive Lag Under Cointegration Modeling

For the analysis of the long-run dynamic in the model, past research studies had proved many empirical modeling techniques. For instance, Johansen and Juselius (1990) and Johansen (1988) broadly used the full information of the maximum likelihood method for multivariate cointegration analysis, while Peter and Hansen (1990) and Engle and Yoo (1987) used the fully modified ordinary least square (OLS) and univariate cointegration analysis. Despite the fact, that Johansen techniques of cointegration are preferred and commonly used over other cointegration techniques, they can provide two or more cointegration linkage and can accommodate small (minimum 30) and biased sample, but it is essential that all variables of the model(s) should be integrated with the same direction. However, Pesaran et al. (2001) used the ARDL model to overcome the problem with Johansen cointegration. In the current study, the techniques of ARDL were applied to perform cointegration analysis among the variables of interest. The ARDL bounds testing approach has many benefits than any other method. For example, firstly, the ARDL provides robust estimates by effectively accommodating endogeneity among the variables, overcomes the serial correlation, and uses the appropriate selection of lags. Second, ARDL could be applicable for small sample size, as it is similar to Johansen's cointegration model (Odhiambo, 2009). Third, the ARDL technique has fewer requirements for integration, since the application of ARDL model does not require all variables integrated in the same direction (order), while it is different from the cointegration technique of Johansen. We can use ARDL methodology even when the variables are I (0), I (1), or a combination of both. Fourth and lastly, at the same time, ARDL bounds testing approach estimates the dynamics of both the short- and long-run in the model.

Based on the above mentioned selected time-series data and advantages of the ARDL model, in the current study, we proceeded with a similar method to inspect the influence of poverty and income inequality along with the population and economic growth on $CO_2e$. The ARDL model is specified as follows:

$$\begin{aligned}\Delta \mathrm{lnCO2et} &= \beta 0 + \sum_{\mathrm{i}=1}^{\mathrm{P}} \boldsymbol{\beta_1} \Delta \mathrm{lnCO2et} - \mathrm{i} + \sum_{\mathrm{i}=1}^{\mathrm{P}} \boldsymbol{\beta_2} \Delta \mathrm{lnPOVt} \\ &- \mathrm{i} + \sum_{\mathrm{i}=1}^{\mathrm{P}} \boldsymbol{\beta_3} \Delta \mathrm{lnINEQt} - \mathrm{i} + \sum_{\mathrm{i}=1}^{\mathrm{P}} \boldsymbol{\beta_4} \Delta \mathrm{POPDt} - \mathrm{i} \\ &+ \sum_{\mathrm{i}=1}^{\mathrm{P}} \boldsymbol{\beta_5} \Delta \mathrm{lnGDPt} - \mathrm{i} + \lambda 1 \ln \mathrm{CO2et} - 1 \\ &+ \lambda 2 \ln \mathrm{POVt} - 1 + \lambda 3 \ln \mathrm{INEQt} - 1 \\ &+ \lambda 4 \mathrm{POPDt} - 1 \lambda 5 \ln \mathrm{GDPt} - 1 + \varepsilon t \end{aligned} \quad (2)$$

Equation 2 is the ARDL bounds testing approach under the unrestricted error correction model (UECM), where $\Delta \ln CO_2e$, $\Delta$lnPOV, $\Delta$lnINEQ, $\Delta$POPD, and $\Delta$lnGDP represent their respective difference values. Similarly, $\beta_2$, $\beta_3$, $\beta_4$,and $\beta_5$represent the short-run dynamic relationship in the model, while $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, and $\lambda_5$ indicate the dynamical relationship in the long run. Likewise, $P$ explains the lag length of each variable (both dependent and independent variables). The ARDL bounds testing approach uses F-statistics or Wald for a joint (mutual) significance test to examine if there is any long-run or cointegration relationship.

In order to carry out the statistical diagnostic tests for the stability of the model and to estimate the short- and long-run coefficients, Equation 3 explains the short-run, where $\vartheta_i$ explains the adjustment speed of the equilibrium after some short-run economic shocks.

$$\begin{aligned}\Delta \text{lnCO2et} &= \beta 0 + \sum_{i=1}^{P} \boldsymbol{\beta_1} \Delta \text{lnCO2et} - i + \sum_{i=1}^{P} \boldsymbol{\beta_2} \Delta \text{lnPOVt} \\ &- i + \sum_{i=1}^{P} \boldsymbol{\beta_3} \Delta \text{lnINEQt} - i + \sum_{i=1}^{P} \boldsymbol{\beta_4} \Delta \text{POPDt} - i \\ &+ \sum_{i=1}^{P} \boldsymbol{\beta_5} \Delta \text{lnGDPt} - i + \theta i \, \text{ECTt} - I + \varepsilon t \qquad (3)\end{aligned}$$

### NARDL Methodology

Similar to the simple ARDL model, the applications of NARDL methodology do not have the need for all variables (included in the model) to be integrated in the same r direction. By proceeding with the method of Shin et al. (2014), we used the methodology of NARDL to capture the asymmetric relationship (non-linear relationship) among the variables; to do so, we used the current estimates in the following model:

$$\begin{aligned}\Delta \text{lnCO2et} &= \beta 0 + \beta 1 \Delta \boldsymbol{lnPOV}^{+} t + \beta 2 \Delta \boldsymbol{lnPOV}^{-} t \\ &+ \beta 3 \Delta \boldsymbol{lnINEQ}^{+} t + \beta 4 \Delta \boldsymbol{lnINEQ}^{-} t \\ &+ \beta 5 \Delta \boldsymbol{POPD}^{+} t + \beta 6 \Delta \boldsymbol{POPD}^{-} t \\ &+ \beta 7 \Delta \boldsymbol{lnGDP}^{+} t + \beta 8 \Delta \boldsymbol{lnGDP}^{-} t + \varepsilon t \qquad (4)\end{aligned}$$

In Equation 4, $\beta_1 \Delta lnPOV_t^{+}$, $\beta_3 \Delta lnINEQ_t^{+}$, $\beta_5 \Delta POPD^{+}{}_t$, and $\beta_7 \Delta lnGDP^{+}{}_t$ explain the positive variation in poverty, income inequality, population and GDP per capita, respectively. While, $\beta_2 \Delta lnPOV^{-} t$, $\beta_4 \Delta lnINEQ^{-} t$, $\beta_6 \Delta POPD^{-}\ t$ and $\beta_8 \Delta lnGDP^{-} t$ represent the negative variation in poverty, income inequality, population, and GDP per capita, respectively.

By following the same procedures as proceeded under simple methodology of ARDL, we have applied different stationarity tests (unit root tests) on $lnPOV^{+}$, $lnPOV^{-}$, $lnINEQ^{+}$, $lnINEQ^{-}$, $POPD^{+}$, $POPD^{-}$, and $lnGDP^{+} lnGDP^{-}$. The resulting probabilities of unit root tests found that all variables have no unit root and are stationary even at the I(1), while the lag order selection criteria suggested that the lag of 3 would be suitable for the NARDL methodology technique. After this, we used the NARDL model to examine the asymmetric association between the variables of interest.

**TABLE 6 |** NARDL cointegration bound test.

| Bound Test | H0: No cointegration | | | | Decision |
|---|---|---|---|---|---|
| | | Sig. | Level: I(0) | 1st Difference: I(1) | |
| F-stat | 11.814***** | 0.1 | 2.2 | 3.09 | |
| k | 4 | 0.05 | 2.56 | 3.49 | |
| | | 0.025 | 2.88 | 3.87 | |
| | | 0.01 | 3.29 | 4.37 | Cointegrated |

**, **, ***, and **** represent the level of significance at 10, 5, 2.5, and 1%, respectively.*

# RESULTS

## Estimation of ARDL Model

Before estimating the long- and short-run analysis, we inspected the bounds cointegration testing approach using the joint-F significance test to identify whether variables of the study are cointegrated (in the long-run). The calculated F-values and the level of significance (10, 5, 2.5, and 1%) are illustrated in **Table 6**, along with lower I(0) and lower I(1) critical values.

The tabulated (fixed) values of the upper bound I(1) are less than the computed F-values, and Narayan (2005) (with intercept and without trend) are highly significant even at 1% (0.1) level. The above-stated procedures confirmed that there is a long-run association of poverty (HCR), income inequality population density, and economic growth (GDP per-capita), and $CO_2e$.

This section provides the empirical findings of long-run dynamics as well as short-run dynamics; we used the ARDL approach to quantify the nexus between HCR of poverty, income inequality (Gini index), population (population density), GDP per capita, and $CO_2e$ for Pakistan. In this study, we have converted mostly variables (used in the model) into log form; for this reason, the probabilities of resulted coefficients can be interpreted as elasticity in the long term.

As reported in **Tables 7**, **8**, the resulting coefficients of ARDL estimates suggest that the association between poverty and $CO_2e$ is positive and statistically significant even at one percent. These findings indicate that the increasing trend of poverty in Pakistan positively influence $CO_2e$ which leads to environmental degradation. The empirical findings of the current study have verified that some serious steps are to be taken for the alleviation of extreme poverty which degrades the quality of environment in Pakistan. There might be different and large number of reasons for the positive impact of poverty (HCR) on environmental degradation in terms of $CO_2e$. One is the indirect mechanism of $CO_2e$ and poverty; for example, to accelerate the process of industrialization, reduction of poverty might be one of the reasons to raise the level of intensity of environmental degradation in terms of $CO_2e$. Encouraging and accelerating the process of industrialization is the key to enhance economic activities that will lead to higher economic growth, and therefore, reduce the intensity of extreme poverty (Khan M. I. et al., 2020). This process also worsens the quality of the environment (Jin et al., 2018). The empirical findings of our current study are consistent with the research findings of Dhrifi et al. (2020) and Masron and Subramaniam (2019) for developing countries, Lu (2017) for 24 Asian countries, and

**TABLE 7 |** Estimation of long-run ARDL with diagnostic and model stability.

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| LnPOV | 0.562*** (0.176) | 0.301*** (0.103) | 0.299*** (0.089) | 0.531*** (0.170) |
| LnINEQ | −0.283 (0.282) | −1.037*** (0.361) | −0.466** (0.211) | −0.271 (0.269) |
| PopD | 0.010*** (0.003) | 0.005** (0.001) | 0.003* (0.001) | 0.009*** (0.003) |
| LnGDP | 0.698*** (0.310) | 1.002*** (0.292) | 1.453*** (0.294) | 0.670** (0.300) |
| LnINF | | −0.068** (0.036) | | |
| LnIND | | | −0.277 (0.177) | |
| LnLSS | | | | 0.006 (0.012) |
| Constant | −8.232*** | −5.500*** | −9.436*** | −7.944*** |
| **ARDL bound test** | | | | |
| F-statistic | 11.814*** | 17.030*** | 11.668*** | 9.886*** |
| Model Sel. | 4, 0, 2, 0, 3 | 4, 0, 2, 2, 4, 4 | 4, 0, 0, 0, 2, 1 | 4, 0, 0, 2, 3, 0 |
| **Statistics** | | | | |
| R-Sq. | 0.997 | 0.999 | 0.997 | 0.997 |
| Adjusted R-Sq. | 0.995 | 0.997 | 0.996 | 0.995 |
| F-Stat. | 761.134*** | 917.890*** | 844.106*** | 687.437*** |
| **Diagnostic tests** | | | | |
| **Normality test** | | | | |
| Jarque–Bera test | 0.451 | 0.048 | 0.859 | 0.406 |
| **Heteroscedacity** | | | | |
| Breusch–Pagan test | 0.705 | 0.688 | 0.676 | 0.644 |
| ARCH | 0.690 | 0.123 | 0.776 | 0.559 |
| **Model stability** | | | | |
| CUSUM test | Yes | Yes | Yes | Yes |
| CUSUMSQ | Yes | Yes | Yes | Yes |

*Standard errors are in parenthesis. *, **, and *** represent the level of significance at 10, 5, and 1%, respectively.*

**TABLE 8 |** Short-run estimation of ARDL model.

| | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|
| LnPOV | 0.132*** (0.176) | 0.126** (0.056) | 0.173*** (0.047) | 0.128*** (0.063) |
| LnINEQ | −0.153 (0.282) | −0.447* (0.234) | −0.270 (0.159) | −0.157 (0.180) |
| PopD | 0.500** (0.003) | 0.899*** (0.312) | 0.565*** (0.135) | 0.487** (0.191) |
| LnGDP | 0.377*** (0.310) | 0.756** (0.281) | 0.842*** (0.141) | 0.388* (0.222) |
| LnINF | | 0.009 (0.009) | | |
| LnIND | | | 0.005 (0.078) | |
| LnLSS | | | | 0.003 (0.007) |
| Coint-Eq. | −4.452*** (1.059) | −4.154*** (0.904) | −5.468*** (0.885) | −4.602*** (1.113) |

*Standard errors are in parenthesis. *, **, and *** represent the level of significance at 10, 5, and 1%, respectively.*

Khan (2019) for the Association of Southeast Asian Nations (ASEAN) countries. Whereas, the result of the current study might differ from few studies, such as Abid (2016) for African countries, Koçak et al. (2019) for LDCs, Zaidi et al. (2019) for Sub-Saharan Africa, and Islam and Ghani (2018) for ASEAN countries, since the economic dynamics of these countries and regions are different from Pakistan. Furthermore, our study incorporated a different method to examine this relationship. In addition, the present study indicates a different scenario and existence of conflict of policy makers so as to pay attention to environmental protection events or to boost economic growth in Pakistan. Such inconsistent scenario grabs the policymakers and government authorities because it is not easy to focus on issues related to the quality of the environment to be of the top priority in devising a policy (Abid, 2016).

Concerning the income gap and $CO_2e$, the relationship is presented in **Tables 7**, **8**. The resulting coefficients of income inequality (Gini index) on $CO_2e$ suggest a negative and highly significant impact on environmental degradation, that is, the negative trend in inequality leads to increase in $CO_2e$. In other words, high-income inequality contributes to lower environmental degradation in Pakistan. One of the possible explanations might be that the lower value of Gini index in the least developed countries like Pakistan might decrease economic competition and the affordability to buy high-carbon emitting and energy-consuming sources among disadvantaged people. Therefore, lack of economic competition leads to a low environmental-friendly source of energy and thus high environmental degradation in terms of $CO_2e$. Another reason is that the narrow income gap in developing countries like Pakistan weakens the people to invest in new and high-emitting technology to increase their income and thereby lower the $CO_2e$ (Demir et al., 2019; Uddin et al., 2020). The result of the current study is interesting and contrary to the work of Boyce (1994), who suggested that the higher-income gap creates a power gap between the upper class and lower class in a society that most probably worsens the quality of the environment. Because of different economic statuses, rich people take advantage, but environmental costs are levied on poor people. According to Boyce (1994), increasing income inequality can increase $CO_2e$ and environmental degradation. By increasing the income gap between the rich and poor, high-privileged people could not exploit natural resources for the sake of their luxury life, indicating that high-income inequality is environmentally friendly but poses a big socioeconomic problem. The result of the current study is consistent with the empirical findings of Uddin et al. (2020) for the Group of Seven countries (G7), Wang and Ye (2017) for China, Kounetas (2018) for European countries, Grunewald et al. (2017) for Third World nations, and Demir et al. (2019) for Turkey. Based on the current study, we believe that this inconsistency in the result is not the same as in Pakistan, and this study incorporated different research techniques to get the results.

Regarding population, the result of ARDL shows that there is a direct and statistically significant relationship between population density and environmental degradation in both

the long- and short-runs, while the positive and significant coefficients for population density are varied within the models (1), (2), (3), and (4). From these results, it is inferred that population is also the main cause for the degrading environment in Pakistan. A possible explanation could be that the increasing rate of population leads to more people, while more people require more resources; consequently, the natural resources of earth deplete more rapidly. The main result of resource depletion is deforestation and loss of biodiversity as humans strip the earth off resources to accommodate and adjust the rising number of the population. The findings of the current study, supported by the work of Bakhsh et al. (2017), in the short-run as well as in the long-run are also consistent with the findings of Islam and Ghani (2018), for ASEAN countries. However, these findings are not in line with the work of Munir and Khan (2014) for the developing countries in Asia, and Hao et al. (2016) for China. The above researchers found an indirect relationship between population and environmental degradation. According to them, better-quality of infrastructure and more grounded ecological transmission bound higher population density areas, which lead to environmental sustainability.

As reported in **Tables 7**, **8**, the result of ARDL methodology show that there is a positive and statistically significant relationship between GDP per capita and environmental degradation in case of Pakistan, thus signifying that the GDP per capita rises $CO_2e$ in the short-run as well as in the long-run. Particularly, the findings of this study are comparable with the work of Adedoyin et al. (2020) for BRICS economies, Ali et al. (2020) for Nigeria, Khan (2019) for ASEAN countries, Malik et al. (2020) for Pakistan, Lu (2017) for developing countries, Usman et al. (2020) for upper middle income countries (UMICs), and Zhang et al. (2014) for China. From the empirical results of this study, we are able to conclude that the existing economic growth and economic activities in developing countries like Pakistan contribute to the reduction of extreme poverty but with high environmental costs in the shape of environmental degradation.

Regarding the control group, the impact is as follows: (i) in the long-run, inflation has a negative and statistically significant effect on CO2e, where our findings are consistent with the recent work of Khan (2019) and Malik et al. (2020) for Pakistan. (ii) We failed to find any statistically significant influence of industry, the value-added share of GDP, and least basic sanitation services (% of the total population) on environmental degradation in Pakistan. In other words, the industry, value-added share of GDP, and least basic sanitation services do not cast any impact on $CO_2e$ in the case of Pakistan.

## Diagnostic Tests and Stability of ARDL Model

Regarding model diagnostic and model stability tests, each model presented in **Table 7** provides the empirical results for the diagnostic tests. First, we incorporated the Jarque–Bera test for the normality of data, where the critical value of Chi-square test is higher than 5%, indicating statistically insignificance and failed to reject the H0 of normal distribution, meaning that the data is normally distributed. While, for heteroscedasticity, we incorporated Breusch–Pagan (BP) test along with ARCH test and found the resulting probability value of the Chi-square test to be statistically insignificant, suggesting that we failed to reject the null hypothesis of homoscedasticity. Besides, for the provision of the robust results of the models, we conducted and tested the stability of all models incorporating the cumulative sum of recursive residual (CUSUM) and cumulative sum of recursive residual square (CUSUMSQ) (Brown et al., 1975). **Figure 2** illustrates the graphical results for CUSUM and CUSUMSQ, which suggest the dynamic stability for all the models.

## Estimation of NARDL Model

As suggested in **Table 9**, the computed (critical value) F-values (6.727) are greater than the tabulated value (3.9) of the 0(1) (lower bound) as well as I(I) (upper bound) of constant (intercept) and no trend that are found to be statistically significant even at 1% level (Narayan, 2005). These results established a long-run

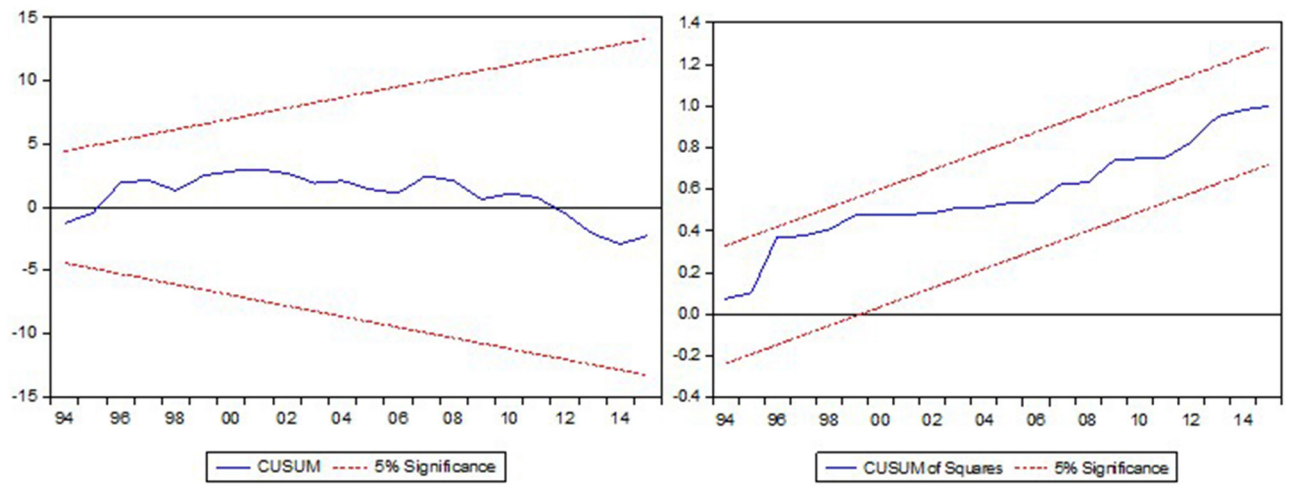



**FIGURE 2 |** ARDL CUSUM and ARDL CUSUM-square graphs.

**TABLE 9 |** NARDL cointegration bound test.

| Bound Test | | H0: No cointegration | | | Decision |
|---|---|---|---|---|---|
| | | Sig. | Level: I(0) | 1st Difference: I(1) | |
| F-stat | 6.727*** | 0.1 | 1.92 | 2.89 | |
| k | 7 | 0.05 | 2.17 | 3.21 | |
| | | 0.025 | 2.43 | 3.51 | |
| | | 0.01 | 2.73 | 3.9 | Cointegration |

**, **, ***, and **** represent the level of significance at 10, 5, 2.5, and 1%, respectively.*

**TABLE 10 |** NARDL long-run estimation with diagnostic tests and model stability tests.

| | Coefficient | t-statistics | Prob. |
|---|---|---|---|
| LnPOV_POS | 0.558*** | 2.682 | 0.0116 |
| LnPOV_NEG | −0.141* | −1.762 | 0.0879 |
| LnINEQ_POS | −1.647*** | −2.725 | 0.0105 |
| LnINEQ_NEG | 1.016 | 1.391 | 0.1738 |
| LnGDP_POS | 1.725*** | 9.844 | 0.0000 |
| LnGDP_NEG | −3.314 | −1.397 | 0.1722 |
| POPD_POS | −0.003* | −2.017 | 0.0524 |
| POPD_NEG | N/A | N/A | N/A |
| Constant | −0.995*** | −7.984 | 0.0000 |
| **Model statistics** | | | |
| R-squared | | | 0.997 |
| Adjusted R-squared | | | 0.996 |
| F-statistics | | | 1042.500 |
| Prob. (F-statistics) | | | 0.000 |
| **Diagnostic tests** | | | |
| **Heteroscedasticity tests** | | | |
| Breusch-Pagan-Godfrey | | | 0.954 |
| AECH test | | | 0.927 |
| Glejser test | | | 0.885 |
| Harvey test | | | 0.420 |
| **Serial correlation** | | | |
| Breusch–Godfrey LM Test | | | 0.080 |
| **Normality test** | | | |
| Jarque–Bera test | | | 0.340 |
| **Model stability** | | | |
| CUSUM | | | Stable |
| CUSUMSQ | | | Stable |
| Model selection | | | 1, 1, 0, 1, 0, 1, 0, 0 |

*Standard errors are in parenthesis. *, **, and *** represent the level of significance at 10, 5, and 1%, respectively.*

relationship among the CO2e, POPD, LnPOV_Pos, LnPOV_neg, LnGINI_Pos, LnGINI_neg, LnGDP_Pos, and LnGDP_neg.

The results of both the long-run non-linear ARDL and short-run NARDL long-run are presented in **Tables 10**, **11**. The resulted coefficients of NARDL and the relationship between poverty HCR are positive and statistically significant. The results of both the short- and long-run NARDLs are consistent with the results of the long and short-run ARDLs as mentioned in **Tables 7**, **8**. The asymmetric relationship between poverty and $CO_2$ emissions are direct (positive shocks in the partial sum

**TABLE 11 |** Short-run estimations of NARDL.

| | Coefficient | *t*-statistics | Prob. |
|---|---|---|---|
| d(LnPOV_POS) | 0.564*** | 2.867 | 0.007 |
| d(LnPOV_NEG) | −0.142* | −1.901 | 0.066 |
| d(LnINEQ_POS) | −1.667*** | −2.893 | 0.006 |
| d(LnINEQ_NEG) | 1.029 | 1.477 | 0.149 |
| d(nNGDP_POS) | 1.191*** | 4.106 | 0.000 |
| d(LnGDP_NEG) | −3.354 | −1.426 | 0.163 |
| d(POPD_POS) | −0.003* | −2.015 | 0.052 |
| d(POPD_NEG) | N/A | N/A | N/A |
| Coint Eq. | −1.006 | −4.706 | 0.000 |

*Standard errors are in parenthesis. *, **, and *** represent the level of significance at 10, 5, and 1%, respectively.*

of poverty HCR), while the decrease in poverty HCR (negative shocks in the partial sum of poverty HCR) does not cast a statistically significant impact on $CO_2$e. The result of negative shocks of poverty on $CO_2$e for both the short and long-runs are statistically insignificant even at the 5% level, while the findings of positive shocks of poverty on $CO_2$e are positive and statistically significant. Similarly, the results of GDP per capita (positive shocks) are statistically significant and positive in both the short- and long-run NARDLs, suggesting a positive impact of GDP per capita on environmental degradation. The income inequality (positive shocks) consistently suggests a negative and statistically significant impact on $CO_2$e. In the case of both the short- and long-run associations between population density and $CO_2$, an upsurge in the population (positive shocks of population density) increases the $CO_2$e, whereas a reduction in the population (negative shocks of population density) does not apply.

The diagnostic tests of the model are also presented in the lower part of **Table 10**, where the findings failed to reject the null hypothesis of normality, heteroscedasticity, and model stability. **Figure 3** illustrates the graphical results for CUSUM and CUSUMSQ for NARDL, which suggest the dynamic stability for all the models.

## DISCUSSION AND CONCLUSION

In order to achieve the study objective and SDG goals, the empirical findings of the current study deliver a significant insight by inspecting the nexus between poverty, income inequality, and $CO_2$e, incorporating population density and economic growth in the model for Pakistan using time-series data from 1971 to 2015. The findings of the current study were obtained from ARDL and NARDL methodologies to achieve unbiased and reliable regression results. The obtained results of the current study are summarized as follows: poverty HCR is the main predictor of $CO_2$e in the case of Pakistan, while, Gini coefficient (measures of income inequality) has an insignificant impact on $CO_2$e across the models. The population density and GDP per capita in Pakistan are also harmful to the quality of the environment.

The current study suggests that extreme poverty is the main predictor that degrades the quality of the environment in Pakistan; for that reason, the policymakers should plan the

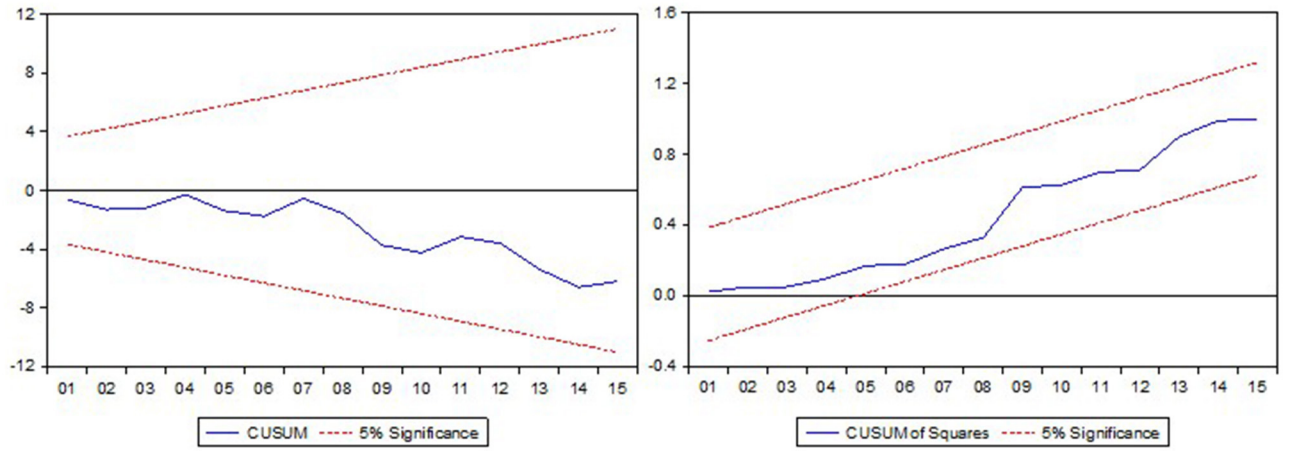



**FIGURE 3 |** NARDL CUSUM and NARDL CUSUM-square graphs.

policies to reduce extreme poverty level without environmental pollution. Regarding policy implementation, the results of this study considered few but important long, medium, and short-term implications for policymakers. The results of the current article confirmed that extreme poverty and high population density are detrimental factors for environmental degradation; moreover, increasing GDP per capita increases $CO_2e$ as well. From the result of this study, the policymakers might infer steps that can be taken to alleviate the intensity of extreme poverty that may have a rising effect on population and $CO_2e$. In terms of long-term policy implications, a universal and general holistic tactic that encourages pro-poverty economic growth and development protects the risk and vulnerability of environmental quality along with maximizing pro-poor opportunities. For that reason, the government and the concerned authorities should ensure a comprehensive policy that can make sure that gain from all productive and economic activities must reach the needy and the improvised public. The productive activities along with economic growth and development in Pakistan are detrimental to the quality of the environment; therefore, policymakers should not only just consider high economic growth but also focus on the environmental quality. To achieve sustainable economic growth and MDG goals, the government of Pakistan should implement environmental regulation policies and ensure the usage of environmental-friendly energy. Since the regulatory policy regarding environmental protection significantly encourages the process of innovation that further leads to an upsurge in the efficiency of energy, $CO_2e$ could be reduced.

For short-term policy implication, the prevailing situation could only be focused, if the concerned authorities of the government of Pakistan confirm that the low-income and impoverished group of people should not transform their needs into high $CO_2e$. For example, for short-term policy implication, the government of Pakistan should not only depend on economic growth but should also correspondingly handle the prevailing situation to create short-term employment opportunities for poor people. Furthermore, to provide immediate support to poor people, the policymakers should introduce micro-financial programs and also provide quick support through safety social nets. In addition, from the current study, we concluded that higher income inequality is not a significant predictor of $CO_2e$, but the policymaker and government should ensure distributive policy and make income distribution equally to achieve economic stability and make better the quality of environment and control $CO_2e$.

To conclude, there is a need to implement inclusive policy reforms that can control the population growth and make sure a situation that can alleviate extreme poverty level without degrading the quality of environment and economic growth. Additionally, the empirical results of the study emphasize economic growth and poverty reduction to achieve MDG goals. As mentioned earlier, the government of Pakistan should take few important steps on the effect of their policies regarding the reduction of extreme poverty and population density along with sustainable economic growth and stable income distribution policy to control environmental degradation. While this study has a very vital role for the policymaker, there are few limitations regarding policy implication. For example, the current study incorporating the most commonly used Gini index as an indicator for income inequality is insignificant; therefore, to extend the importance of the future study, it is important to use other indices for income inequalities, such as Theil index and Palma ratio (Palma and Stiglitz, 2016). By using other indices for income inequality, future researchers could verify whether the results obtained in this study are consistent and reliable or not. Finally, the current study inspected the relationship between poverty and income inequality along with population density and economic growth in the Islamic Republic of Pakistan. But, this study leaves the research gap for future researchers to examine this nexus for other developed and developing countries.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found at: https://databank.worldbank.org/source/world-development-indicators#.

## AUTHOR CONTRIBUTIONS

SK: software, validation, resources, data curation, writing—original draft, writing—review and editing, and visualisation. WY: conceptualisation, methodology, formal analysis, investigation, writing—review and editing, supervision, project administration, and funding acquisition. All authors: contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

The valuable comments and suggestions of editors and reviewers are gratefully acknowledged, while we also acknowledged the support of all funding sources. We are solely responsible for errors. The authors of this manuscript are thankful to the National Social Science Foundation of China (Grant No. 18BJY164), major projects on Philosophy and Social Science Research in colleges and universities in the Henan Province of China (Grand No: 2018-YYZD-16), the Education Department of Henan Province (Grant No. 19A790025), and the National Natural Science Foundation of China (Grant No: 41601566), for funding this work.

## REFERENCES

Abdullah, A., Doucouliagos, H., and Manning, E. (2015). Does education reduce income inequality? A meta-regression analysis. *J. Econ. Surveys* 29, 301–316. doi: 10.1111/joes.12056

Abid, M. (2016). Impact of economic, financial, and institutional factors on $CO_2$ emissions: evidence from sub-Saharan Africa economies. *Util. Policy* 41, 85–94. doi: 10.1016/j.jup.2016.06.009

Adedoyin, F. F., Gumede, M. I., Bekun, F. V., Etokakpan, M. U., and Balsalobre-Lorente, D. (2020). Modelling coal rent, economic growth and CO2 emissions: does regulatory quality matter in BRICS economies? *Sci. Tot. Environ*. 710:136284. doi: 10.1016/j.scitotenv.2019.136284

Adelegan, A. E., and Out, E. (2020). Environmental Quality, and economic development: is there a kuznets curve for air pollution in Nigeria? *South Asian Res. J. Human Soc. Sci*. 2. doi: 10.36346/sarjhss.2020.v02i03.001

Adusah-Poku, F. (2016). Carbon dioxide emissions, urbanization and population: empirical evidence from Sub-Sahran Africa. *Energy Econ. Lett*. 3, 1–16. doi: 10.18488/journal.82/2016.3.1/82.1.1.16

Akbar, M., Noor, F., Ahmad, I., and Sattar, A. (2018). Impact of energy consumption and CO2 emissions on food production in Pakistan: an econometric analysis. *Pak. J. Agric. Sci*. 55, 455–461. doi: 10.21162/PAKJAS/18.6703

Ali, H. S., Nathaniel, S. P., Uzuner, G., Bekun, F. V., and Sarkodie, S. A. (2020). Trivariate modelling of the nexus between electricity consumption, urbanization and economic growth in Nigeria: fresh insights from Maki Cointegration and causality tests. *Heliyon* 6:e03400. doi: 10.1016/j.heliyon.2020.e03400

Arshed, N., Hassan, M. S., and Aziz, O. (2020). Does income inequality lead to educational inequality? A cross-sectional study of Pakistan. *UMT J*. 3. doi: 10.32350/uer.31.03

Awodumi, O. B., and Adewuyi, A. O. (2020). The role of non-renewable energy consumption in economic growth and carbon emission: evidence from oil producing economies in Africa. *Energy Strategy Rev*. 27:100434. doi: 10.1016/j.esr.2019.100434

Baek, J., and Gweisah, G. (2013). Does income inequality harm the environment?: Empirical evidence from the United States. *Energy Policy* 62, 1434–1437. doi: 10.1016/j.enpol.2013.07.097

Bakhsh, K., Rose, S., Ali, M. F., Ahmad, N., and Shahbaz, M. (2017). Economic growth, CO2 emissions, renewable waste and FDI relation in Pakistan: new evidences from 3SLS. *J. Environ. Manage*. 196, 627–632. doi: 10.1016/j.jenvman.2017.03.029

Baloch, A., Shah, S. Z., Noor, Z. M., and Magsi, H. B. (2018). The nexus between income inequality, economic growth and environmental degradation in Pakistan. *GeoJournal* 83, 207–222. doi: 10.1007/s10708-016-9766-3

Baloch, M. A., Khan, S. U. D., Ulucak, Z. S., and Ahmad, A. (2020). Analyzing the relationship between poverty, income inequality, and CO2 emission in Sub-Saharan African countries. *Sci. Tot. Environ*. 740:139867. doi: 10.1016/j.scitotenv.2020.139867

Baloch, M. A., Ozturk, I., Bekun, F. V., and Khan, D. (2021). Modeling the dynamic linkage between financial development, energy innovation, and environmental quality: does globalization matter? *Bus. Strategy Environ*. 30, 176–184. doi: 10.1002/bse.2615

Berger, T., and Grabert, S. (2018). International output and inflation uncertainty and their impact on countries' macroeconomic performance. Evidence from a dynamic factor Garch-In-Mean model. *Macroecon. Dyn*. 22:1113. doi: 10.1017/S1365100516000596

Boyce, J. K. (1994). Inequality as a cause of environmental degradation. *Ecol. Econ*. 11, 169–178. doi: 10.1016/0921-8009(94)90198-8

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. R. Stat. Soc. Ser. B* 37, 149–163. doi: 10.1111/j.2517-6161.1975.tb01532.x

Brundtland, G. H. (1987). Our common future-Call for action. *Environ. Conserv*. 14, 291–294.

Butt, T. H., Abid, G., Arya, B., and Farooqi, S. (2020). Employee energy and subjective well-being: a moderated mediation model. *Serv. Indust. J*. 40, 133–157. doi: 10.1080/02642069.2018.1563072

Danish (2019). Moving toward sustainable development: the relationship between water productivity, natural resource rent, international trade, and carbon dioxide emissions. *Sustain. Dev*. 28, 540–549. doi: 10.1002/sd.2007

Demir, C., Cergibozan, R., and Gök, A. (2019). Income inequality and CO2 emissions: empirical evidence from Turkey. *Energy Environ*. 30, 444–461. doi: 10.1177/0958305X18793109

Dhrifi, A., Jaziri, R., and Alnahdi, S. (2020). Does foreign direct investment and environmental degradation matter for poverty? Evidence from developing countries. *Struct. Change Econ. Dyn*. 52, 13–21. doi: 10.1016/j.strueco.2019.09.008

Easterlin, R. A., and Angelescu, L. (2012). "Modern economic growth and quality of life: cross-sectional and time series evidence," in *Handbook of Social Indicators and Quality of Life Research* (Dordrecht: Springer), 113–136.

Engle, R. F., and Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *J. Econ*. 35, 143–159. doi: 10.1016/0304-4076(87)90085-6

Grossman, G. M., and Krueger, A. B. (1995). Economic growth and the environment. *Q. J. Econ*. 110, 353–377.

Grunewald, N., Klasen, S., Martínez-Zarzoso, I., and Muris, C. (2017). The trade-off between income inequality and carbon dioxide emissions. *Ecol. Econ*. 142, 249–256. doi: 10.1016/j.ecolecon.2017.06.034

Hailemariam, A., Dzhumashev, R., and Shahbaz, M. (2020). Carbon emissions, income inequality and economic development. *Empir. Econ*. 59, 1139–1159. doi: 10.1007/s00181-019-01664-x

Hao, Y., Chen, H., and Zhang, Q. (2016). Will income inequality affect environmental quality? Analysis based on China's provincial panel data. *Ecol. Indicat*. 67, 533–542. doi: 10.1016/j.ecolind.2016.03.025

Harting, P. (2019). "Macroeconomic stabilization and long-term growth: the role of policy design," in *Macroeconomic Dynamics*, (Cambridge University Press), 1–46. doi: 10.1017/S1365100519000488

Heubaum, H., and Biermann, F. (2015). Integrating global energy and climate governance: the changing role of the International Energy Agency. *Energy Policy* 87, 229–239. doi: 10.1016/j.enpol.2015.09.009

Islam, R., and Ghani, A. B. A. (2018). Link among energy consumption, carbon dioxide emission, economic growth, population, poverty, and forest area. *Int. J. Soc. Econ*. 45, 275–285. doi: 10.1108/IJSE-12-2016-0351

Jin, G., Fu, R., Li, Z., Wu, F., and Zhang, F. (2018). CO2 emissions and poverty alleviation in China: an empirical study based on municipal panel data. *J. Clean. Prod*. 202, 883–891. doi: 10.1016/j.jclepro.2018.08.221

Johansen, S. (1988). Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control* 12, 231–254. doi: 10.1016/0165-1889(88)90041-3

Johansen, S., and Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with appucations to the demand for money. *Oxf. Bull. Econ. Stat*. 52, 169–210. doi: 10.1111/j.1468-0084.1990.mp52002003.x

Khan, M. I., Teng, J. Z., and Khan, M. K. (2020). The impact of macroeconomic and financial development on carbon dioxide emissions in Pakistan: evidence with a novel dynamic simulated ARDL approach. *Environ. Sci. Pollut. Res*. 27, 39560–39571. doi: 10.1007/s11356-020-09304-z

Khan, M. K., Khan, M. I., and Rehan, M. (2020). The relationship between energy consumption, economic growth and carbon dioxide emissions in Pakistan. *Financial Innov*. 6, 1–13. doi: 10.1186/s40854-019-0162-0

Khan, S. A. R. (2019). The nexus between carbon emissions, poverty, economic growth, and logistics operations-empirical evidence from southeast Asian countries. *Environ. Sci. Pollut. Res*. 26, 13210–13220. doi: 10.1007/s11356-019-04829-4

Koçak, E., Ulucak, R., Dedeoglu, M., and Ulucak, Z. S. (2019). Is there a trade-off between sustainable society targets in Sub-Saharan Africa? *Sustain. Cities Soc*. 51:101705. doi: 10.1016/j.scs.2019.101705

Kounetas, K. E. (2018). Energy consumption and CO2 emissions convergence in European Union member countries. A tonneau des Danaides? *Energy Econ*. 69, 111–127. doi: 10.1016/j.eneco.2017.11.015

Lahoti, R., Jayadev, A., and Reddy, S. G. (2015). *The Global Consumption and Income Project (GCIP): An Introduction and Preliminary Findings*. United Nations.

Lu, W. C. (2017). Renewable energy, carbon emissions, and economic growth in 24 Asian countries: evidence from panel cointegration analysis. *Environ. Sci. Pollut. Res*. 24, 26006–26015. doi: 10.1007/s11356-017-0259-9

Maji, I. K. (2019). Impact of clean energy and inclusive development on CO2 emissions in sub-Saharan Africa. *J. Clean. Prod*. 240:118186. doi: 10.1016/j.jclepro.2019.118186

Malik, M. Y., Latif, K., Khan, Z., Butt, H. D., Hussain, M., and Nadeem, M. A. (2020). Symmetric and asymmetric impact of oil price, FDI and economic growth on carbon emission in Pakistan: evidence from ARDL and non-linear ARDL approach. *Sci. Tot. Environ*. 726:138421. doi: 10.1016/j.scitotenv.2020.138421

Masron, T. A., and Subramaniam, Y. (2019). Does poverty cause environmental degradation? Evidence from developing countries. *J. Poverty* 23, 44–64. doi: 10.1080/10875549.2018.1500969

Masud, M. M., Kari, F. B., Banna, H., and Saifullah, M. K. (2018). Does income inequality affect environmental sustainability? Evidence from the ASEAN-5. *J. Asia Pacific Econ*. 23, 213–228. doi: 10.1080/13547860.2018.1442146

Munir, S., and Khan, A. (2014). Impact of fossil fuel energy consumption on CO2 emissions: evidence from Pakistan (1980-2010). *Pak. Dev. Rev*. 53, 327–346. doi: 10.30541/v53i4IIpp.327-346

Narayan, P. K. (2005). The saving and investment nexus for China: evidence from cointegration tests. *Appl. Econ*. 37, 1979–1990. doi: 10.1080/00036840500278103

Odhiambo, N. M. (2009). Energy consumption and economic growth nexus in Tanzania: an ARDL bounds testing approach. *Energy Policy* 37, 617–622. doi: 10.1016/j.enpol.2008.09.077

Palma, J. G., and Stiglitz, J. E. (2016). "Do nations just get the inequality they deserve? The "Palma Ratio" re-examined," in *Inequality and Growth: Patterns Policy* (London: Palgrave Macmillan), 35–97.

Pesaran, M. H., Shin, Y., and Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *J. Appl. Econ*. 16, 289–326. doi: 10.1002/jae.616

Peter, C. B. P., and Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I (1) processes. *Rev. Econ. Stud*. 57, 99–125.

Qamruzzaman, M., and Jianguo, W. (2020). The asymmetric relationship between financial development, trade openness, foreign capital flows, and renewable energy consumption: fresh evidence from panel NARDL investigation. *Renew. Energy* 159, 827–842. doi: 10.1016/j.renene.2020.06.069

Qin, Y. (2020). China's experience of poverty alleviation in typical rural areas. *Adv. Appl. Sociol*. 10:313. doi: 10.4236/aasoci.2020.108019

Rai, J. (2019). Poverty, vulnerability, non-timber forest products and rural livelihood: an Indian experience. *J. Geogr. Environ. Earth Sci. Int*. 21, 1–10. doi: 10.9734/jgeesi/2019/v21i230124

Rit, B. (2014). The relationship between inflation, inflation uncertainty and output growth in India. *Econ. Aff*. 59, 465–477. doi: 10.5958/0976-4666.2014.00014.X

Sani, S., Mukhtar, S., and Gani, I. M. (2017). Relationship between electricity consumption, manufacturing output and financial development: a new evidence from Nigeria. *Energy Econ. Lett*. 4, 28–35. doi: 10.18488/journal.82.2017.43.28.35

Schleicher, J., Schaafsma, M., Burgess, N. D., Sandbrook, C., Danks, F., Cowie, C., and Vira, B. (2018). Poorer without it? The neglected role of the natural environment in poverty and wellbeing. *Sustain. Dev*. 26, 83–98. doi: 10.1002/sd.1692

Shahpouri, A., Biabi, H., and Abolhassani, L. (2016). Economic development and environmental quality: the environmental Kuznets curve for water pollution. *J. Appl. Sci. Environ. Manag*. 20, 161–169. doi: 10.4314/jasem.v20i1.19

Shin, Y., Yu, B., and Greenwood-Nimmo, M. (2014). "Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework," in *Festschrift in Honor of Peter Schmidt* (New York, NY: Springer), 281–314.

Sun, R. (2014). Nominal rigidity and some new evidence on the New Keynesian theory of the output-inflation tradeoff. *Int. Econ. Econ. Policy* 11, 575–597. doi: 10.1007/s10368-014-0267-x

Torras, M., and Boyce, J. K. (1998). Income, inequality, and pollution: a reassessment of the environmental Kuznets curve. *Ecol. Econ*. 25, 147–160. doi: 10.1016/S0921-8009(97)00177-8

Uddin, M. M., Mishra, V., and Smyth, R. (2020). Income inequality and CO2 emissions in the G7, 1870–2014: evidence from non-parametric modelling. *Energy Econ*. 88:104780. doi: 10.1016/j.eneco.2020.104780

Usman, M., Kousar, R., Yaseen, M. R., and Makhdum, M. S. A. (2020). An empirical nexus between economic growth, energy utilization, trade policy, and ecological footprint: a continent-wise comparison in upper-middle-income countries. *Environ. Sci. Pollut. Res*. 27, 38995–39018. doi: 10.1007/s11356-020-09772-3

Wang, Z., and Ye, X. (2017). Re-examining environmental Kuznets curve for China's city-level carbon dioxide (CO2) emissions. *Spatial Stat*. 21, 377–389. doi: 10.1016/j.spasta.2016.09.005

Yahong, W., and Khan, S. (2021). A cross-sectional analysis of employment returns to education and health status in China: moderating role of gender. *Front. Psychol*. 12:129. doi: 10.3389/fpsyg.2021.638599

Yang, L., Wang, J., and Shi, J. (2017). Can China meet its 2020 economic growth and carbon emissions reduction targets? *J. Clean. Prod*. 142, 993–1001. doi: 10.1016/j.jclepro.2016.08.018

Zaidi, S. A. H., Zafar, M. W., Shahbaz, M., and Hou, F. (2019). Dynamic linkages between globalization, financial development and carbon emissions: evidence from Asia Pacific Economic Cooperation countries. *J. Clean. Prod*. 228, 533–543. doi: 10.1016/j.jclepro.2019.04.210

Zhang, Y. J., Liu, Z., Zhang, H., and Tan, T. D. (2014). The impact of economic growth, industrial structure and urbanization on carbon emission intensity in China. *Nat. Hazards* 73, 579–595. doi: 10.1007/s11069-014-1091-x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Khan and Yahong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fenrg.2021.667163

# A Method to Deal With Inter-regional and Inter-provincial Transaction Settlement Deviation Quantity Based on Kernel Density–Entropy Weight

*Dongfa Wang*[1], *Weibin Ding*[1], *Feng Gao*[1], *Yang Xu*[2] *and Qin Miao*[3]*

[1] State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, China, [2] Strategy and Development Research Center, State Grid Zhejiang Economic and Technological Research Institute Co., Ltd., Hangzhou, China, [3] China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, China

The mismatch between energy distribution and power load in China can be alleviated by inter-regional and inter-provincial power transactions. However, it also brings challenges to transaction settlement. In the new round of electric power reform, the transaction settlement deviation quantity needs to be more standardized. Based on the analysis of the related work, this article used the analytical framework of the analytic hierarchy process to calculate the transaction type weight, the inter-regional, and the inter-provincial weight, respectively, and accordingly, quantifies the amount of deviation quantity allocated by the corresponding trading subjects. According to the size and volatility of the trading quantity, we further propose a comprehensive weighting method based on kernel density and entropy weight to quantify the deviation quantity of inter-regional and inter-provincial trading subjects of different trading types. Specifically, this article first used the kernel function weighting method to calculate the weights of different transaction types that measure the transaction quantity, and then the improved entropy weight method was used to calculate the weights of different transaction types that reflect the volatility of the transaction quantity. Then, the comprehensive weights were constructed by considering the influence of the above two dimensions on the distribution of the deviation electricity simultaneously. The deviation electricity responsibility determination model was used to clarify the transaction subject's deviated electricity responsibility, and the deviated electric quantity calculation model was used for quantification. At last, the validity and practicability of the method were verified through the analysis of examples using the inter-regional and inter-provincial power transaction data from China.

**Keywords: inter-provincial and inter-regional transaction, deviation electric quantity, electricity transaction, electricity settlement, deviated electricity responsibility**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Qingying Zheng,*
*Fujian Normal University, China*
*Die Hu,*
*Fuzhou University, China*

***Correspondence:**
*Qin Miao*
*miaoqin90s@stu.xmu.edu.cn*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *12 February 2021*
***Accepted:*** *17 March 2021*
***Published:*** *20 April 2021*

***Citation:***
*Wang D, Ding W, Gao F, Xu Y and Miao Q (2021) A Method to Deal With Inter-regional and Inter-provincial Transaction Settlement Deviation Quantity Based on Kernel Density–Entropy Weight. Front. Energy Res. 9:667163. doi: 10.3389/fenrg.2021.667163*

## INTRODUCTION

In 2015, China issued a document titled "Furthering Reform of the Electricity Market" to start a new round of electric power reform. In this new reform, great importance was attached to the price mechanism reform and competition will be gradually introduced into price determination. A scientific and reasonable inter-regional electricity trading price mechanism is conducive to the

optimization of energy allocation and sustainable resource management (Song et al., 2019). As an important part, electricity transaction settlement directly affects the market operation. Further, the gradual development of the electric power market has put forward higher requirements for the transaction settlement.

The focus of electricity marketization is to improve efficiency and reduce energy prices for other industries. Energy distribution in China does not match with the power load, which needs to be alleviated through inter-regional and inter-provincial power trading. Inter-regional and inter-provincial power trading in China is usually driven by the provincial government. Due to the homogeneity of electricity, it is impossible to derive the actual electricity of each seller from the total actual on-grid electricity. Besides, due to the loss in the transmission process and the difference between the actual loss and the power loss calculated according to the approved line loss rate, the cross-regional and cross-provincial power trading settlement has a deviation. Moreover, for new energy power plants, the fluctuation of clean energy generation tends to bring more deviated quantity than traditional thermal power generation. Therefore, how to deal with the deviated electricity in a standardized way has become an urgent problem to be solved. Different treatment methods will directly affect the settlement quantity. Additionally, the establishment of the assessment rules of the deviation electric quantity will directly affect the economic interests of both parties in the electric power transaction.

The settlement of electricity transactions includes electric quantity settlement and electric fee settlement. The current inter-regional and inter-provincial electricity transaction settlement in China is completed by the power transaction department. And the settlement amount is the sum of the planned amount and the distributed amount of deviated electricity. The electricity fee settlement is completed by the financial department. The electricity bill to be paid is determined based on the settled electricity quantity and the contract price. Because a real-time market for inter-regional and inter-provincial electricity transactions cannot be established in the short term, real-time balancing transaction cannot be realized. Figuring out how to deal with the deviated electric quantity is the key problem in the design of power trading mechanism in China.

The electric quantity deviation generally refers to the difference between the actual quantity and the contract quantity. In the medium- and long-term power market, many factors can lead to the electric quantity deviation, e.g., the sudden failure of the generator set. However, as for the spot market, the settlement period is shortened to less than an hour, and the electric quantity deviation, also known as the unbalanced quantity, refers to the difference between the planned transaction quantity and the actual generation quantity. The disputes on market construction target, assessment responsibility, and economic benefit caused by settlement deviation have brought great trouble to the trading subject. First, in the power market, the different market participants need to strictly follow the contract. The deviation quantity not only reflects the deviation between the generator and the dispatcher but also makes it difficult to confirm the deviation responsibility of different transaction parties in practice. Second, it is difficult to confirm each subcontract due to the grid structure and dispatching operations. Different off-grid power division schemes will lead to the difference in the current deviation quantity of each contract. And then, the contract settlement amount will change accordingly, which raises the problem of unfairness. Thus, the proper processing of deviation quantity has become one of the most important standards to measure the fairness and justice of the power market. If not handled properly, it will directly affect the economic interests of the participants in the power transaction, and may even have a negative impact on the power market order. In inter-regional power transactions, it is of great significance to reduce the influence of the deviation electric quantity on the participants starting from the settlement scheme, so as to establish a more perfect inter-regional power transaction mechanism.

The processing method of the deviated quantity can be divided into separate settlement and apportionment settlement. The separate settlement is to settle the deviated electricity quantity according to the formulated unified price. While the apportionment settlement is made according to the corresponding electricity price of the transaction type. The deviated electric quantity is allocated to each transaction according to the scheduling plan, contract electric quantity, and verified line loss rate. The difference between them is evident. In general, the operation of separate settlements is relatively simple, but the specific settlement electric quantity of each transaction cannot be measured. In addition, how to set a reasonable and fair uniform price for the deviation electric quantity is also a relatively difficult problem. The apportionment settlement can obtain the electric quantity of the on-grid end and the landing end of each transaction. It has more detailed processing of the deviation electricity, and the differentiated pricing and settlement of the deviation electricity according to the specific situation is more in line with the actual situation. At present, most of the electricity transaction settlements in China adopt the processing method that uses the annual base rather than other components to bear the deviation of electricity. With the advance of the reform of the power market, the annual base quantity of electricity will gradually decrease. At the same time, the market entities of power trading, such as power sales companies, will constantly increase. This promotes the gradual transition of the power market from the single buyer mode to the wholesale competition mode and enables power generation enterprises to conduct bilateral transactions with large power users, power sales companies, and other market members. The existing settlement method of deviation quantity transaction cannot meet the needs of power market reform and a new and more reasonable deviation electricity processing method needs to be proposed.

Therefore, this article proposes a settlement method based on kernel density and entropy weight for inter-regional and inter-provincial transaction electricity quantity, which is quantified in three steps. First, taking the transaction quantity as the target, we use the kernel density to assign the weight to the transaction type with a large transaction quantity. At the same time, entropy weight is used to describe the volatility of transaction quantity, and the transaction type is given weight

from the two dimensions of transaction: volume and volatility. Second, we clarify the responsibility judgment of deviation electric quantity. The responsibility of deviation electric quantity exists between the sending end and the receiving end. The contract electric quantity, actual electric quantity, and power deviation are used to determine to which side the responsibility of deviated electric quantity belongs. Third, we use the line loss rate and network loss rate to quantify the total amount of deviation in the transaction process. Then, the corresponding punishment strategy is set for the responsible party. The above three-step method comprehensively and objectively quantifies the responsibilities of power market participants for the deviated electric quantity. It breaks through the limitations of the existing methods of processing deviations in cross-regional and cross-provincial power transactions. And it provides a standard solution for the determination of deviated electric quantity among different departments.

## LITERATURE REVIEW

Electricity deviation refers to the difference between the actual measured electric quantity and the contract quantity. The factors influencing electricity deviation are complex. The reasonable handling of the electricity deviation is important to ensure the fairness and justice of the power trade. The cross-provincial and cross-regional development of the electric power trading market is an important breakthrough of the marketization reform in China. It not only conforms to the national energy development plan but also can realize the surplus and shortage adjustment and supply and demand complementarity, which is of great significance to the optimal allocation of electric power resources.

### The Treatment Method of Deviating Electric Quantity

Many scholars have carried out relevant research on the settlement mechanism and the treatment method of deviation electric quantity in the inter-regional and inter-provincial electric power transaction of China. With cross-regional trading mechanism and historical trading data, Li et al. (2016) proposed a deviation electric quantity processing method based on virtual time-of-use price. Zhao et al. (2016) proposed the reference electricity price based on the approved deviation electricity price of regional power generation cost, determined the deviation electricity price through the proposed logic, and finally settled the contract electricity quantity and deviation electricity quantity separately. Ding et al. (2017) proposed a responsibility determination model and pricing method for the deviation of cross-region alternating current (AC) power grid based on the average frequency deviation and constructed the "meter" region diagram for determining the liability for the deviation of the tie line. He et al. (2018) summarized the experience and characteristics of the efficient operation of the typical balanced market by analyzing the operation and transaction settlement mode of the PJM balanced market in northern Europe, the United Kingdom, and the United States. They proposed that the bivalent method should be adopted in the deviation settlement, and the capacity cost should be shared by the market members with deviations. Yan et al. (2019) proposed a deviation quantity treatment method based on fuzzy comprehensive evaluation theory. It uses the analytic hierarchy process (AHP) to set the weight of each index, the type of transaction, and the membership function of the transaction cycle, and finally determines the comprehensive evaluation score of each transaction and allocates the power deviation based on the score. By referring to the deviation assessment mechanism under the British balance mechanism, Ma et al. (2019) proposed the deviation quantity assessment mechanism in the domestic day-ahead market and constructed the profit model of the electricity selling companies. Xu et al. (2019) proposed three settlement methods of cardinal deviation electric quantity, namely, the expected revenue compensation method, the deviation electric quantity replacement matching method, and the average price difference compensation method.

### Deviation Electric Quantity Settlement of International Power Market

The research on deviation electric quantity settlement of international power market is more abundant. Zhang and Lo (2009) believed that deviation settlement is highly dependent on the real-time balance of electricity prices and market structure. Van Der Veen and Hakvoort (2009) showed that the power market balance responsibility and deviation settlement rules were different in the Nordic countries. Combined with the deviation settlement mechanism of the Belgian electricity market, De Vos et al. (2010) analyzed the effect of the deviation tolerance system of offshore wind power generation on the promotion of new energy grid connection. Haring et al. (2012) evaluated the impact of different prices on the energy market and deviation settlement under the two conditions of considering and ignoring transmission capacity limits. Haring et al. (2015) pointed out that the cost transparency of deviated electricity settlement in the European power market is insufficient, which may bring arbitrage opportunities and hinder the market integration of renewable energy. Mendes et al. (2016) compared the trading patterns of the UK and Brazilian electricity wholesale markets and pointed out that the deviation settlement design of the UK electricity market can provide a reference for the redesign of deviation settlement in Brazil. Zani and Rossi (2018) applied the node electricity price to the electricity deviation settlement and showed that the node electricity price was an effective method. Kermer (2019) studied the deviation settlement design considering demand-side management and thermoelectric combination and found that the deviation settlement design depended on the structure of the market's main body. Bottieau et al. (2020) pointed out that the single deviation pricing mechanism of positive and negative deviation electricity with a single price settlement has become a new mechanism in the European power market, but this mechanism will stimulate market participants to deviate from the plan to restore the balance of the power system, which brings risks.

## The Formation Characteristics of the Deviated Electric Quantity

The formation characteristics and corresponding solutions of deviated power from the subdivision of the power industry have also attracted much attention. Generally, the power industry is composed of four sub-industries, including thermal power generation, hydropower generation, photovoltaic power generation, and wind power generation. Chen et al. (2016) believed that the main problems in thermal power systems are power loss and voltage deviation, so a multi-objective enhanced particle swarm optimization (MOEPSO) algorithm is proposed to solve the above problems simultaneously. Compared with traditional thermal power, the proportion of renewable energy to participate in market transactions is increasing in recent years. However, the characteristics of renewable energy, such as high volatility and poor prediction accuracy, lead to the uneven distribution of base electricity among renewable energy power plants and low renewable energy consumption ratio. Therefore, the design of a deviation power model for renewable energy to participate in power market transactions is particularly critical. Moreover, Bo et al. (2019) also demonstrated that the accounting of deviation electricity has a significant impact on the economic benefit evaluation of wind power plants. The design of the deviation power model in the new energy industry generally has two ideas. One design method is to reduce the scale of the formation of deviation electricity in the new energy industry from the source and achieve it through technical and other means. For example, Limsakul et al. (2015) believed that the generator speed deviation caused by the fluctuation of photovoltaic power in the time domain and frequency domain can be indirectly reversed by changing the photovoltaic power, which is found through unit step function analysis. Similarly, Li et al. (2015) believed that with the increase of grid connection of large-scale wind farms, the fluctuation of wind power generation output would not only affect the power flow distribution of the power grid but would also affect the frequency of the power grid. Therefore, a quasi-steady state model of synchronous generator, a quasi-steady-state load model of asynchronous wind generator, and a simplified RX[1] model are proposed. Another design method is to make appropriate compensation to the new energy industry the electric power market transaction mechanism. For example, Liu et al. (2017) proposed a renewable energy alternative compensation method. First, the actual power generation of renewable energy power plants is subtracted from the prior contract power generation, and the basic power generation is determined based on the installed capacity of the machine. Then, according to the electricity quantity and electricity price in the market transaction, the comprehensive unit price of the renewable energy power generation settlement is determined, and fewer power plants are economically compensated.

Overall, the existing methods for dealing with deviation electricity have certain limitations, lacking a more mature deviation electricity liability determination mechanism and pricing mechanism. Especially in cross-regional and cross-province power transactions, there are problems such as the difficulty in determining the quantity of the deviated power and the inconsistency in the processing methods of different departments Therefore, this study will quantify the amount of electric deviation from the perspective of the power market transaction volume and transaction fluctuations, and propose the main accounting method for cross-regional and cross-provincial power transactions based on the idea that all parties involved in the transaction share the deviation electric quantity.

# PROPOSED METHOD

At present, electric quantity settlement plays a main part in the inter-regional and inter-provincial electric power transaction settlement in China, which is completed by the electric power transaction department. The accounting of transaction electricity consists of planned electricity quantity and apportioned deviation electricity quantity. Two issues need to be addressed to calculate the deviation electric quantity between transaction entities based on the transaction electric quantity. The first issue is about the transaction volume. If the transaction electric quantity is larger, the loss rate of the power line in the transaction process will be higher, thus the deviation power quantity (Wang et al., 2003; Gao and Song, 2012). The second issue is about the volatility of transaction volume. The greater the fluctuation of transaction electric quantity with time, the greater the inter-provincial and inter-regional power dispatching coordination cost, and thus the more the deviated power that should be borne (Shang et al., 2014). Therefore, based on the transaction electric quantity, this article considers the influence of the trading volume and the volatility of trading volume on the calculation of deviation electric quantity allocation.

The electricity market in China is divided into three regions: State Grid, Southern Power Grid, and West Mongolia Grid. And the power trade in each region can be divided into intra-provincial and inter-provincial transactions. Since the current electricity market transactions in China are mainly mid- and long-term transactions and spot transactions, intra-provincial electricity transactions can be subdivided into four categories: direct electricity transactions, power generation rights transactions, pumped storage transactions, and other transactions. Similarly, the inter-provincial electricity transactions can be subdivided into three types: inter-provincial direct electricity transactions, inter-provincial outbound transactions (network-to-point, network-to-network), and inter-provincial power generation rights transactions. To study the characteristics of inter-regional and inter-provincial transactions of different transaction types, this article adopted the AHP framework to calculate the weights of different transaction types and different regions, respectively, and then quantify the amount of deviated electricity allocated by corresponding transaction entities.

## Calculation of Transaction Type Weight

The weight analysis can adopt the methods of objective weighting, subjective weighting, or mixed weighting. From the mathematical perspective, kernel function weighting method,

[1]RX model is commonly used in electrical engineering. The RX model is based on the steady-state model of the induction machine, where R generally refers to resistance and X means reactance.

and entropy weighting method belong to the category of objective weighting method. How to assign different weights to different transaction types in the power market is the focus of this section.

If the transaction quantity of a certain transaction type $i$ within time $T$ is $X_{iT}$, we use linear kernel function $f(X_{iT})$ to weight transaction type $i$, that is,

$$w_i = \frac{1}{T} f\left(X_{iT}\right)$$

Where $f(X_{iT}) = \sum_{t=1}^{T} X_{it}$, and $f(X_{iT})$ represents the total transaction volume of the transaction type $i$ within the time $T$. This method is intuitive, but it ignores the influence of other transaction types on the weight $W_i$. Therefore, we have made the following corrections $f(X_{iT})$.

$$w_i = \frac{1}{T} \frac{\sum_{t=1}^{T} X_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} X_{it}}$$

It can be seen that the redefined $w_i$ reflects the transaction volume proportion of different transaction types $i$ in time $T$. The proportion of transaction volume reflects the degree of negative line loss of this transaction type. The greater the degree of loss, the greater the amount of deviation generated. In addition, due to the limitations of electric dispatching cost and the default conditions of the electricity contract, the deviation electric quantity will be larger if the volatility of transaction electric quantity is greater. The following will focus on how to assign weights to different transaction types according to the volatility of transaction electric quantity. We adopted the improved entropy weight method. The calculation steps were as follows:

Step 1: Establish panel data for different transaction types of intra-provincial and inter-provincial transactions:

$$X_k = \left[x_{it}^k\right] = \begin{bmatrix} x_{11}^k & \cdots & x_{1T}^k \\ \vdots & \ddots & \vdots \\ x_{n1}^k & \cdots & x_{nT}^k \end{bmatrix}$$

In the above equation, $i$ denotes the transaction type ($i = 1, ..., n$). $t$ denotes the trading time point ($t = 1, ..., T$). And superscript$k$indicates whether $x_{it}$ is an intra-provincial transaction or an inter-provincial transaction.

Step 2: Normalization. Due to differences in regions, transaction types, and transaction time, the transaction volume of electricity at each transaction time point fluctuates greatly. So non-dimensional processing is required. The specific expression is as follows:

$$\bar{x}_{it}^k = \frac{x_{it}^k - x_{min}^k}{x_{max}^k - x_{min}^k}$$

In the above formula, $x_{\max}^k$ and $x_{\min}^k$ denote the maximum and minimum trading volumes in period $T$, respectively.

Step 3: Calculate information entropy. Based on the standardized processing of transaction electric quantity, we calculate the information entropy according to the following formula:

$$H_i = -\beta \sum_{t=1}^{T} f_{it} \ln\left(f_{it}\right)$$

$$f_{it} = \frac{\bar{x}_{it}^k}{\sum_{i=1}^{n} \bar{x}_{it}^k}, \beta = \frac{1}{\ln(T)}$$

Information entropy is a description of the degree of uncertainty. $f_{it}$ denotes the proportion of the normalized electric quantity $\bar{x}_{it}^k$ of transaction type $i$ in the form of intra-province transaction or inter-province transaction. It measures the possibility of transaction type $i$, which is essentially a representation of probability. And $f_{it} \ln\left(f_{it}\right)$ indirectly reflects the volatility of the probability of transaction type $i$, which is essentially a description of variance. Therefore, with a minus sign, $H_i$ describes the degree of inverse fluctuation of transaction type $i$ based on trading quantity $\bar{x}_{it}^k$ in period $T$.

There may be some statistical problems in the actual calculation of $H_i$. Therefore, this article made the following assumptions: if $\bar{x}_{it}^k = 0$, then $f_{it} = 0$, and $f_{it} \ln\left(f_{it}\right) = 0$.

Step 4: Calculate the entropy weight. Generally, when the traditional entropy weight method is used to calculate the entropy weight $\omega_i$, the standardized processing method is adopted, as is shown in the following formula:

$$\omega_i = \frac{|1 - H_i|}{\sum_{i=1}^{n} |1 - H_i|}$$

Olendski (2016) and Kasolis and Clemens (2020) have shown that, when the entropy value is in a certain range, the slight difference in the entropy value may cause the entropy weight to change exponentially, which is inconsistent with the information conveyed by the entropy value. To avoid this problem, the improved entropy weight method is used to calculate the objective weight value of the *ith* transaction type:

$$w_i' = \frac{\sum_{i=1}^{n} H_i + 1 - 2H_i}{\sum_{i=1}^{n} \left(\sum_{i=1}^{n} H_i + 1 - 2H_i\right)}$$

$\sum_{i=1}^{n} H_i$ reflects the degree of reverse uncertainty of all transaction types. $\sum_{i=1}^{n} H_i + 1 - 2H_i$ measures the reverse volatility level of the remaining transaction types after removing the uncertainty degree of the transaction type $i$. The greater the inverse volatility of the transaction type $i$, the larger the value of $-2H_i$ will be, and thus the proportion of the transaction type $i$.

To sum up, this section, first, use the kernel function weighting method to calculate the weight of different transaction

types based on the transaction volume. And then, use the improved entropy weight method to calculate the weight of different transaction types that reflect the volatility of transaction volume. If the above two dimensions are considered at the same time, the comprehensive weight is constructed as follows:

$$W_i = \alpha w_i + (1-\alpha) w_i'$$

Where $\alpha$ reflects the importance of the power transaction volume to the allocation of deviated electric quantity. Correspondingly, $1-\alpha$ reflects the degree of influence of the volatility of transaction volume on the deviated electric quantity, which also reflects the composition of the comprehensive weight.

## Calculation of Inter-provincial and Inter-regional Weights

Since the electricity market in China is divided into three regions: State Grid, Southern Power Grid and West Mongolia. Each region involves intra-provincial and inter-provincial transactions. After quantifying the transaction type weight, the next step is to calculate the weight of each region according to the transaction quantity.

Similarly, according to the transaction quantity, the linear kernel function is used to assign weights to the transaction time point of a certain area in a period of time, so as to reflect the transaction quantity of the area at that time point. And the improved entropy weight method is used to measure the characteristics of the volatility of transaction quantity in different regions.

## Allocation of Deviation Electric Quantity

### Deviation Electricity Responsibility Judgment Model

Another core issue of establishing a reasonable settlement mechanism for inter-regional electricity trading is the allocation of deviation electricity. There are usually two ways to clarify the electric deviation responsibility of the two parties in an electricity transaction. One is based on rigorous theoretical calculations, such as the power flow tracking algorithm, which can intuitively describe the physical direction of the actual power grid flow and calculate the actual deviation electricity quantity of transaction parties at each time period (Huang et al., 2018; Zhang H. et al., 2019). This method can also ensure the accuracy and fairness of the calculation results. However, the calculation process is complicated and the amount of calculation is too large. Thus, it is not practical to use this method to settle the electricity quantity of grid transactions, and the responsible party of the deviated electricity quantity cannot be determined in this way either. The other method is based on the actual electricity trading market data. The deviation electricity price in the inter-regional electricity transaction is settled according to the original contract price without distinction, or allocated according to a certain proportion, or through a certain weighted average pricing (Wen et al., 2008; Zhang, 2008; Zhao et al., 2009). Although this method is simple and easy to use, it is unfair and lacks a rigorous scientific basis. In addition, the planned value of the next month is adjusted in a rolling way according to the current deviated electric quantity with more refunds and less compensations, which is not necessarily beneficial to the damaged party of the deviated electric quantity.

To sum up, the above two methods of determining the liability of deviated electric quantity adopt different ways and different perspectives to allocate the deviated electric quantity. But neither of them puts forward a clear criterion and method for determining the responsibility of deviated electric quantity. The trans-regional transmission in China is generally characterized by long-distance and large capacity. According to the power transaction settlement data of the state grid, the daily deviation of trans-regional ultrahigh voltage (UHV) power transmission can reach 4% of the planned electricity, which is a large amount. Therefore, it is urgent to establish a settlement mechanism based on the responsibility identification of the cross-region power grid deviation quantity. Therefore, this section will put forward the responsibility determination model of the cross-region power grid deviation quantity.

The responsibility determination of deviation power mainly revolves around the differentiation between the power grids at the receiving end. Two sub-regions $A$ and $B$ in the cross-region power system are set. The power sent from region $A$ to $B$ via the cross-region link line is $P_{AB}$, and $P_{AB} = -P_{BA}$. Each trading day is divided into $N$ periods of responsibility, and the duration of each period is $T_n$. From the power grid dispatch center dispatch management system (OMS), the contract electric quantity $W_{ABn}$ and the actual electric quantity $G_{ABn}$ at the gateway of $A$ and $B$, and in time period $n$ in the trading day are obtained, as well as the record data of frequency change curve, which can be calculated as follows:

$$\begin{cases} \Delta W_{ABn} = G_{ABn} - W_{ABn} \\ \Delta P_{ABn} = \frac{\Delta W_{ABn}}{T_n} \\ \Delta f_n = \frac{1}{T_n} \int_{t=0}^{T_n} (f_t - 50) dt \end{cases}$$

where $\Delta W_{ABn}$, $\Delta P_{ABn}$, $f_t$, and $\Delta f_n$ are cross-region transmission deviation of the period of time period $n$, the quantity, the average deviation of power, the instantaneous value of frequency and the average deviation of frequency, respectively. Especially, for $\Delta f_n$, 50 is the average value of instantaneous frequency between time 0 and $T_n$ (Ding et al., 2017).

The area control deviation $\lambda_n$ of area $A$ and $B$ in the calculation period $n$ is,

$$\lambda_n = K\Delta f_n + \Delta P_{ABn}$$

Where $K$ is the system power/frequency equivalent factor, which is called power frequency equivalent in short. According to the principle of controlling power frequency deviation of grid frequency tie line, under ideal operation state, when the power network load changes, automatic generation control of units in the region should be carried out through the dispatching center of the sending and receiving regions, respectively, so as to maintain the power balance in the region, that is, $\lambda_n = 0$. If the dispatching at the sending end or the receiving end and the output regulation of the generator set are delayed or deviated, which implies that $\Delta f_n \neq 0$and$\Delta P_{ABn} \neq 0$, the corresponding end be responsible for the deviation. And the

criteria for identifying the self-recognizing party of deviation electricity are proposed:

(a) When $\Delta f_n > 0, \Delta P_{ABn} > 0$ or $\Delta f_n < 0, \Delta P_{ABn} < 0$, the sending end shall be responsible for the electric deviation. When $\Delta f_n > 0, \Delta P_{ABn} < 0$or $\Delta f_n < 0, \Delta P_{ABn} > 0$, the receiving end shall be responsible for the electric deviation.
(b) When $|\Delta P_{ABn}| < \left|K\Delta f_n\right|$, the degree of responsibility is for a lighter level, when $|\Delta P_{ABn}| > \left|K\Delta f_n\right|$, the degree of responsibility is for the heavier grades.
(c) When $\Delta f_{\mathrm{n}} = 0$, no judgment on the responsibility for the electric deviation will be made, or the sending and receiving ends are equally responsible.

### Deviation Electricity Accounting Model

After clarifying the responsibility of deviation quantity of both parties in power transactions, it is necessary to quantify the total deviation quantity in the transaction process. The calculation of deviation quantity can be further divided into information deviation calculation and energy deviation calculation. Information deviation accounting refers to the process in which the trading center takes punitive measures against the market subject according to the degree of deviation, for there is a difference between the output/load curve approved by the dispatching agency and the output/load value actually measured by the market subject and the difference has exceeded the permissible range. The measure can be economic measures, such as the deviation degree according to the hierarchical fines, etc. And it can also be assessment measures, such as classifying the credit rating of the subject according to the degree of deviation and reducing the proportion of the transaction of the credit subject in the next cycle. The deviation rate $\beta_{\mathrm{i}}$ is the core to determine the trading main body *i*. Based on the deviation rate of trading subject *i*, the penalty $f(\beta_i)$ under different deviation levels is set. $f(\beta_i)$ is a dispersion function related to $\beta_{\mathrm{i}}$. Energy deviation accounting refers to the calculation of the actual measurement quantity of the trading subject and the deviation quantity of the agreed electricity quantity and the dispatching plan. It is the total deviation quantity shared by the trading subject *i* in each power market. The specific form of deviation electric quantity $Q_i^d$ is,

$$Q_i^d = f(\beta_i) + W_i Q_{2,i}$$

In the above formula, $f(\beta_i)$ is the information deviation accounting, and $W_i Q_{2,i}$ is the energy deviation accounting. To further discuss the composition of the total deviated electric quantity, it is assumed that $Q_{1,i}$ is the agreement quantity of market subject *i*, $Q_{2,i}$ is the total deviated electric quantity of market subject *i* in each market, and $Q_{3,i}$ is the actual measurement quantity of gateway data for the market subject *i*. Since the deviation of the measurement of electricity can be calculated by line loss rate and grid loss rate, the total deviation $Q_{2,i}$ of the subject *i* in each market power calculation is as follows:

$$Q_{2,i} = \sum_{i=1}^{n} Q_{1,i}(1-\eta_i) - Q_{3,i}\left(1-\eta_i^s\right)$$

where the $\eta_i$ is the line loss rate of trade type *i*, $Q_i^r$ is the displayed electric measured quantity, and the $\eta_i^s$ is the integrated network loss rate. The deviation rate $\beta$ can be calculated by the total deviation quantity $Q_{2,i}$ of the market agent *i* and the agreement quantity $Q_{1,i}$ of the market agent *i*:

$$\beta_i = Q_{2,i}/Q_{1,i}$$

The degree of deviation and the penalty for the responsible party are expressed by the function $f(\beta)$. The greater the number of grades of deviation rate, the easier it is to accurately determine the liability for deviation. This section takes the classification of five grades of deviation e.g., to analyze. $f(\beta)$ is a dispersion function related to and is divided into five deviation levels according to different deviation rates $\beta$, which are related to the performance of trading subjects in the previous three transaction periods. Each deviation level corresponds to a penalty gold function. The mathematical expression is,

$$f(\beta_i) = \begin{cases} 0, \beta_i < \beta_1 \\ a\left[(1+\beta_i) - \beta_1\right]\mu, \beta_1 < \beta_i \le \beta_2 \\ b\left[(1+\beta_i) - \beta_1\right]\mu, \beta_2 < \beta_i \le \beta_3 \\ c\left[(1+\beta_i) - \beta_1\right]\mu, \beta_3 < \beta_i \le \beta_4 \\ d\left[(1+\beta_i) - \beta_1\right]\mu, \beta_4 < \beta_i \end{cases}$$

where the coefficients *a*, *b*, *c*, and *d* are the basic fines for different deviation grades, respectively. In particular, $a < b < c < d$, which indicate that the higher the deviation rate is, the higher the level of base penalty is correspondingly. The penalty for subject *i* is determined by the deviation rate $\beta_i$ and the adjustment factor $\mu$. The grade of deviation is divided into five ranges from $\beta_1$ to $\beta_4$. $\mu$ is the adjustment coefficient, which is used to adjust the punishment according to the supply and demand. When the supply is tight, the value of $\mu$ will be increased, and when the market is surplus, the value of $\mu$ will be reduced.

## EXAMPLE ANALYSIS

### Example Data

To calculate the transaction type weight and the inter-provincial and inter-regional weight, this section selects the transaction type and inter-regional and inter-provincial transaction power data from the Electric Power Industry Statistics of China Electricity Council from July 2019 to December 2019 as the source of the calculation example data. The relevant data are summarized in **Tables 1**, **2**.

### Calculation of Weights

Using the calculation method and based on the transaction quantity data in **Table 1**, the kernel density weight and improved entropy weight of different transaction types within and outside the province were obtained, respectively. From the data distribution, the kernel density weight focused on the volume of transacted electric quantity, while the improved entropy weight focused on the fluctuation of transacted electric quantity, which is in line with the theoretical expectation. Where the kernel density weight and improved entropy weight of the four transaction types in the province are (0.8557, 0.1319, 0.0081, 0.0042) and

**TABLE 1 |** Transaction quantity of different transaction types between July and December 2019 (unit: TW·h).

| First-level indicators | Secondary indicators | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|
| Intra-provincial electricity transactions | Direct electricity transactions | 178.24 | 189.82 | 180.87 | 178.45 | 260.32 | 196.8 |
| | Power generation rights transactions | 18.55 | 35.2 | 21.01 | 26.33 | 49.12 | 32.47 |
| | Pumped storage transactions | 5.84 | 1.17 | 0.79 | 0.64 | 0.3 | 2.52 |
| | Other transactions | 1.19 | 1.76 | 0.96 | 0.01 | 1.82 | 0.07 |
| Inter-provincial electricity transactions | Inter-provincial direct electricity transactions | 20.26 | 24.01 | 20.05 | 17.25 | 13.53 | 9.79 |
| | Inter-provincial outbound transactions | 36.71 | 34.82 | 29.45 | 27.16 | 29.4 | 33.0 |
| | Inter-provincial power generation rights transactions | 3.41 | 3.16 | 3.13 | 3.37 | 5.58 | 2.51 |

**TABLE 2 |** Inter-regional and inter-provincial transaction quantity from July to December 2019 (Unit: TW·h).

| First indicators | Secondary indicators | Jul. | Aug. | Sep. | Oct. | Nov. to Dec. |
|---|---|---|---|---|---|---|
| State Grid | Trade in the province | 129.79 | 140.66 | 130.19 | 128.43 | 353.9 |
| | Trade outside the province | 20.26 | 24.02 | 20.05 | 17.25 | 23.32 |
| Southern Power Grid | Trade in the province | 36.63 | 36.82 | 38.51 | 37.48 | 76.96 |
| | Trade outside the province | 0 | 0 | 0 | 0 | 0 |
| West Mongolia Grid | Trade in the province | 11.82 | 12.33 | 12.17 | 12.54 | 26.26 |
| | Trade outside the province | 0 | 0 | 0 | 0 | 0 |

(0.1981, 0.2011, 0.3024, 0.2984), and the kernel density weight and improved entropy weight of the three transaction types in the province are (0.3313, 0.6019, 0.0668) and (0.2075, 0.1950, 0.5975), respectively. Based on the transaction quantity data in **Table 2**, the kernel density weight and improved entropy weight of the three regions can be obtained as: (0.7549, 0.1847, 0.0604) and (0.1998, 0.2004, 0.5998). On this basis, if the importance level $\alpha$ of the amount of electric transaction to the allocation of the deviation power is determined, then the allocation ratio of the four types of transactions in the province, the three types of transactions between the provinces, and the deviation of the three regions are determined accordingly. The specific calculation process is shown in **Table 3**.

## Contribution of Deviation Quantity

To allocate the deviated electricity quantity properly, we should first clarify the responsibility of the two sides of the electricity transaction, and then quantify the total deviated electricity quantity in the transaction process. Taking a 1,000 kV AC UHV test demonstration project to send power to Central China as an example for analysis, the existing UHV cross-regional power transaction method is organized by the State Grid Corporation of China, and the transaction centers at all levels cooperate. The power grid companies and local government departments of the purchase and sale of electricity determine the electricity and electricity prices through bilateral negotiations, and sign a medium and long-term framework agreement. Therefore, it can be determined that the transaction belongs to the direct electricity transaction type in the intra-provincial transaction of the State Grid area.

Based on the framework agreement, both the sending and receiving parties sign the annual contract for the next year through negotiation at the end of each year according to the agreement and the real-time power supply and demand situation. The annual contract plan electricity is broken down to the month, and the monthly deviation electricity is apportioned to the receiving end provinces. The electricity price is the same as the contract electricity and settled on a monthly basis. Because there is no responsibility assessment mechanism for the deviation quantity in the inter-regional transaction settlement, and there is no binding force for the fairness maintenance and execution of the transaction contract between the sending and receiving parties, it is not reasonable to allocate the deviation quantity to the receiving provinces in proportion.

According to the power transaction settlement management method of the State Grid Corporation, the current domestic cross-regional and cross-provincial connection line gateway electricity is measured once every 15 min in the United States, and the transmission plan is also formulated in units of 15 min. Therefore, the time period $T_n = 15\text{min}$ for deviation electricity metering is appropriate. In November of a year, the measured data at the gateway of receiving end power network are: contract (receiving end) 4,287.88 GW·h, and actual receiving end 4,256.55 GW·h. The difference between the contract electric quantity and the actual electric quantity can be calculated, so $\Delta W_{ABn}$ is −31.33 GW·h. Then we can figure out that the average deviation power$\Delta P_{ABn}$ equals to −125.32 GW according to the formula $\Delta P_{ABn} = \Delta W_{ABn}/T_n$ . In this example, we assume that the system power–frequency equivalent coefficient $K$ is 1,650 MW/Hz. The frequency instantaneous value $f_t$ is a fixed value in the unit $T_n$, which is set as 2,199.75 Hz in this example. By integrating the instantaneous frequency value $f_t$ and averaging it, we can obtain that $\Delta f_n$ is 96.65 Hz. So $K\Delta f_n$ is 159.47 GW. By the deviation of power liability judgment model of decision rule, when $\Delta f_n > 0, \Delta P_{ABn} < 0$ or $\Delta f_n < 0, \Delta P_{ABn} > 0$, the deviation power responsibility should be taken by the receiving side. Therefore, in this case, the receiving side should be responsible for the deviation of electricity, and because $|\Delta \mathrm{P_{ABn}}| < |\mathrm{K}\Delta f_{\mathrm{n}}|$, the degree of responsibility is of a lighter level.

**TABLE 3 |** The calculation process of kernel density weight and improved entropy weight of different transaction types.

| First indicators | Secondary indicators | Kernel density weight | | Improved entropy weight | |
|---|---|---|---|---|---|
| | | $\sum_{t=1}^{T} X_{it}$ | $\frac{1}{T}\frac{\sum_{t=1}^{T} X_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}}$ | $H_i$ | $\frac{\sum_{i=1}^{n} H_i+1-2H_i}{\sum_{i=1}^{n}\left(\sum_{i=1}^{n} H_i+1-2H_i\right)}$ |
| Intra-provincial electricity transactions | Direct electricity transactions | 11,845 | 0.8557 | 1 | 0.1981 |
| | Power generation rights transactions | 1,826.8 | 0.1319 | 0.9853 | 0.2011 |
| | Pumped storage transactions | 112.6 | 0.0081 | 0.4817 | 0.3024 |
| | Other transactions | 58.1 | 0.0042 | 0.502 | 0.2984 |
| Inter-provincial electricity transactions | Inter-provincial direct electricity transactions | 1,048.9 | 0.3313 | 0.9691 | 0.2075 |
| | Inter-provincial outbound transactions | 1,905.4 | 0.6019 | 1 | 0.1950 |
| | Inter-provincial power generation rights transactions | 211.6 | 0.0668 | 0 | 0.5975 |
| State Grid | Trade in and outside the province | 1,526.625 | 0.7549 | 1 | 0.1998 |
| Southern Power Grid | Trade in and outside the province | 373.6 | 0.1847 | 0.9985 | 0.2004 |
| West Mongolia Grid | Trade in and outside the province | 122.15 | 0.0604 | 0 | 0.5998 |

In this example, the comprehensive line loss rate of the cross-region UHV tie line is 1.52%, and the comprehensive network loss rate is 1.38%. From this, the deviation power $Q_2$ is calculated to be 24.89 GW·h, and the corresponding deviation rate $\beta$ is 0.058%. Since the $\beta$ here is relatively low, which lies in the first range of deviation rate, we have $f(\beta) = 0$. Therefore, the receiving side of the deviation power is not affected by the penalty factor. The transaction belongs to the direct electricity transaction type in the intra-provincial transaction of the State Grid region. If the importance level of the amount of electricity transaction to the distribution of the deviation power is set to 0.5, then the corresponding weight is 0.5269 accordingly. Therefore, it can be determined that the deviation electric quantity assumed by both parties in the direct transaction within the province of the State Grid region is 13.11 GW·h.

## CONCLUSIONS

China is gradually deepening the reform of electricity marketization. With the deepening of the degree of marketization, the demand for large-scale cross-regional electric power transmission and energy consumption becomes more urgent (Zhang D. et al., 2019). And thus, the processing requirements of deviation electric quantity in transaction settlement are also increasing. At present, the way to deal with the deviation electric quantity in the electricity settlement business is still under exploration. According to the idea that all participants in the transaction share the deviation electric quantity, this article quantifies the deviation electric quantity from the perspective of the transaction volume and transaction fluctuation degree of the electricity market and proposes the calculation method of the deviation electric quantity in the inter-regional and inter-provincial electricity transaction. And this article also establishes the responsibility judgment mechanism and pricing mechanism of electric quantity deviation to deal with the problem of electric quantity deviation sharing among different departments, which a new idea for quantifying the deviation electric quantity.

To be more specific, this article proposes a method for calculating the apportionment ratio of the deviation electric quantity of the four types of transactions within the province, the three types of transactions between the provinces and the three regions, and then constructs a comprehensive weight calculation method. On this basis, this article then adopts the deviation electricity responsibility judgment model and the deviation electricity accounting model, respectively, to clarify the deviation electricity responsibility and the specific deviation electricity that should be undertaken by the transaction subjects. From this perspective, this article expands and enriches the measurement methods of deviation electric quantity to a certain extent, which has strong practical value in actual work. In addition, this article makes an example analysis using the transaction data of the China Electric Power Industry, inter-regional and inter-provincial transaction electric quantity data, and the specific contract electric quantity transaction data, which proves the effectiveness and practicability of the proposed method. However, this method is mainly aimed at the initial stage of the power market in China. With the continuous promotion of the power market, the treatment method of deviation electric quantity will be improved and optimized in combination with the construction progress of the power market.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

Conceptualization, methodology, and writing—original draft were contributed by DW and WD. Software, data curation, and writing—review and editing were contributed by FG. Conceptualization, methodology, writing—original draft, writing—review and editing, and supervision were contributed by YX and QM. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was supported by the project Settlement Mechanism Optimization Strategy under the Further Promotion of Electric Power Reform (Grant no. SGZJJY00JJJS2000129), the National Natural Science Foundation of China (Grant no. 71874149), and the National Social Science Fund of China (Grant no. 20ZDA084).

## REFERENCES

Bo, S., Li, S., Fan, Y., and Li, Z. (2019). Research on deviation electricity assessment model of wind-storage hybrid power generators in electricity market environment. *Electr. Power Constr.* 40, 107–114. doi: 10.3969/j.issn.1000-7229.2019.07.014

Bottieau, J., Hubert, L., De Grève, Z., Vallée, F., and Toubeau, J. F. (2020). Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement. *IEEE Trans. Power Syst.* 35, 1218–1230. doi: 10.1109/TPWRS.2019.2940756

Chen, G., Liu, L., Guo, Y., and Huang, S. (2016). Multi-objective enhanced PSO algorithm for optimizing power losses and voltage deviation in power systems. *Compel* 35:23. doi: 10.1108/COMPEL-02-2015-0030

De Vos, K., Driesen, J., and Belmans, R. (2010). "Financial implications of imbalance settlement exemptions for wind power generators," in *2010 7th International Conference on the European Energy Market* (Madrid: EEM). doi: 10.1109/EEM.2010.5558668

Ding, Y., Zhang, Y., Yang, D., Chen, J., Bai, S., and Sheng, Y. (2017). Responsibility determination and pricing method of deviation electric quantity based on frequency deviation of cross-regional grids trading. *Autom. Electr. Power Syst.* 41, 105–110.doi: 10.7500 /AEPS20161127001

Gao, W., and Song, B. (2012). A quantitative calculation method for actual monthly line loss rate. *Autom. Electric Power Syst.* 36, 86–90. doi: 10.3969/j.issn.1000-1026.2012.02.016

Haring, T., Mountouri, D., and Andersson, G. (2012). "Ensuring energy-market liquidity through adequate imbalance settlement," in *Proceedings of the Universities Power Engineering Conference* (London, UK), 1–6. doi: 10.1109/UPEC.2012.6398420

Haring, T. W., Kirschen, D. S., and Andersson, G. (2015). Incentive compatible imbalance settlement. *IEEE Trans. Power Syst.* 30, 3338–3346. doi: 10.1109/TPWRS.2014.2387947

He, Y., Zhou, M., Wu, Z., Long, S., and Xu, J. (2018). Study on operation mechanism of foreign representative balancing markets and its enlightenment for China. *Power Grid Technol.* 42, 3520–3528. doi: 10.13335/j.1000-3673.pst.2018.1394

Huang, X., Ouyang, S., and Liang, W. (2018). Line loss allocation model considering influence of access of distributed generators on economic operation time for distribution network. *Autom. Electr. Power Syst.* 42, 127–133. doi: 10.7500/AEPS20171122005

Kasolis, F., and Clemens, M. (2020). Information-based model reduction for nonlinear electro-quasistatic problems. *J. Comput. Phys.* 404:109118. doi: 10.1016/j.jcp.2019.109118

Kermer, S. S. O. (2019). "Imbalance settlement design with demand side management and combined heat and power," in *2019 IEEE Texas Power and Energy Conference* (College Station, TX: TPEC), 1–6. doi: 10.1109/TPEC.2019.8662185

Li, C., Wei, Z., Sun, G., and Sun, Y., G. (2015). State estimation of power system considering frequency deviation caused by fluctuation of wind power generation. *Power Grid Technol.* 39, 1301–1306. doi: 10.13335/j.1000-3673.pst.2015.05.020

Li, Z., Zhao, B., and Yan, X. (2016). Deviation electric quantity processing method in cross-regional and cross-province electricity transaction based on virtual time-of-use power price. *Electr. Power Constr.* 37, 40–46. doi: 10.3969/j.issn.1000-7229.2016.07.006

Limsakul, C., Songprakorp, R., Sangswang, A., and Parinya, P. (2015). "Impact of photovoltaic grid-connected power fluctuation on system frequency deviation in contiguous power systems," in *IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society* (Yokohama), 003236–003241. doi: 10.1109/IECON.2015.7392599

Liu, D., Tang, H., Yang, M., Li, Q., and Xu, L. (2017). Settlement compensation mechanism of power trading deviation with promotion of renewable energy consumption. *Autom. Electr. Power Syst.* 41, 105–111. doi: 10.7500/AEPS20170615020

Ma, J., Tan, Z., and Yu, X. (2019). Optimal model for daily deviation penalty of electricity retailer considering real-time electricity trading. *Electric Power Constr.* 40, 11–19.doi: 10.3969/j.issn.1000-7229.2019.02.002

Mendes, A. L. S., De Castro, N., Brandao, R., Camara, L., and Moszkowicz, M. (2016). "The role of imbalance settlement mechanisms in electricity markets: A comparative analysis between UK and Brazil," in *International Conference on the European Energy Market* (Porto: EEM), 1–6. doi: 10.1109/EEM.2016.7521348

Olendski, O. (2016). Theory of the Robin quantum wall in a linear potential. I. Energy spectrum, polarization and quantum-information measures. *Ann. Der Physik* 528, 856–881. doi: 10.1002/andp.201600080

Shang, C., Ding, J., Xie, D., Yang, D., and Hu, T. (2014). Deviation electric quantity processing method in cross-regional grids transmission trading. *Electric Power Constr.* 35, 127–130. doi: 10.3969/j.issn.1000-7229.2014.12.022

Song, M., Fisher, R., and Kwoh, Y. (2019). Technological challenges of green innovation and sustainable resource management with large scale data. *Technol. Forecast. Soc. Change* 144, 361–368. doi: 10.1016/j.techfore.2018.07.055

Van Der Veen, R. A. C., and Hakvoort, R. A. (2009). "Balance responsibility and imbalance settlement in Northern Europe - An evaluation," in *2009 6th International Conference on the European Energy Market* (Leuven: EEM). doi: 10.1109/EEM.2009.5207168

Wang, T., Zhang, J., and Li, X. (2003). Calculation of scheduled loss ratio and its assessment. *Power System Technol.* 27, 40–42. doi: 10.13335/j.1000-3673.pst.2003.07.011

Wen, L., Liu, J., Wu, Z., Zhang, L., Xing, Z., and Pang, W. (2008). Study and application of resolution algorithm of contract volume based on monthly rolling amendment. *Modern Electric Power* 25, 82–87. doi: 10.3969/j.issn.1007-2322.2008.01.017

Xu, F., Xie, X., Shi, L., Tu, M., and Li, L. (2019). Analysis on settlement method for unbalanced base generated energy in medium- and long- term power market. *Autom. Electric Power Syst.* 43, 186–206. doi: 10.7500/AEPS20180727001

Yan, Y., Li, G., Zhou, M., Huang, C., Cao, Y., and Zhang, X. (2019). Processing method for transaction settlement bias power based on fuzzy comprehensive evaluation. *Autom. Electric Power Syst.* 43, 200–205.doi: 10.7500/AEPS20180612009

Zani, A., and Rossi, S. (2018). "Nodal pricing applied to imbalance settlement: Issues for implementation in zonal markets," in *International Conference on the European Energy Market* (Lodz: EEM), 1–5. doi: 10.1109/EEM.2018.8469842

Zhang, D., Li, M., Ji, X., Wu, J., and Dong, Y. (2019). Revealing potential of energy-saving behind emission reduction: a DEA-based empirical study. *Manag. Environ. Q. Int. J.* 30, 714–730. doi: 10.1108/MEQ-07-2018-0125

Zhang, H., Wang, Q., Wang, X., and Ge, Y. (2019). Line loss allocation based on economic operation analysis for distribution network with distributed generators. *Power Demand Side Manag.* 21, 26–30. doi: 10.3969/j.issn.1009-1831.2019.05.006

Zhang, M., and Lo, K. L. (2009). "A comparison of imbalance settlement methods of electricity markets," in *Proceedings of the Universities Power Engineering Conference* (Glasgow), 1–5.

Zhang, S. (2008). Practical research on settlement mode of transacted electricity quantity in regional electricity market. *Power System Technol.* 32, 56–61. doi: 10.3724/SP.J.1011.2008.00482

Zhao, B., Yan, Y., Liu, Y., Yan, X., and Hu, Z. (2016). Pricing method of deviation electric quantity based on regional generation cost of cross-regional and cross-province electricity transaction. *Power Grid Technol.* 40, 3334–3341. doi: 10.13335/j.1000-3673.pst.2016.11.009

Zhao, G., Li, F., Yuan, L., Cheng, J., Wang, Y., Zhang, Y., et al. (2009). A study on transacted electricity quantity settlement scheme of North China grid. *Electric Power* 42, 38–40. doi: 10.3969/j.issn.1004-9649.2009.06.008

**Conflict of Interest:** DW, WD, and FG were employed by the company State Grid Zhejiang Electric Power Co., Ltd. YX was employed by State Grid Zhejiang Economic and Technological Research Institute Co., Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Wang, Ding, Gao, Xu and Miao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 30 April 2021
doi: 10.3389/fenrg.2021.634556

# Does Anti-corruption Policy Influence Energy Efficiency in China?

*Jie Liu*[1,2], *Jianming Guo*[1,3]*, *Xiangfeng Liu*[1,2]*, *Xiao Bai*[3,4] *and Farhad Taghizadeh-Hesary*[5]*

[1] China Research Institute of Regulation and Public Policy, Hangzhou, China, [2] China Institute of Regulation Research, Zhejiang University of Finance and Economics, Hangzhou, China, [3] School of Public Administration, Zhejiang University of Finance and Economics, Hangzhou, China, [4] School of Finance, Zhejiang University of Finance and Economics, Hangzhou, China, [5] Social Science Research Institute, Tokai University, Hiratsuka, Japan

The Chinese government has stepped up its anti-corruption efforts since 2013; the National Energy Administration (NEA), an energy regulatory agency, has been hard hit by corruption. This paper analyzes the impact of government corruption on energy efficiency (EE) from a perspective of energy regulations and also a mechanism of the impact of corruption on EE, and then adopts statistical methods to estimate the impact of corruption on EE in China. The results show that government corruption generally has a negative effect on EE in China. The establishment of dispatched energy regulatory agencies in eastern China can significantly reduce the adverse effects of corruption on EE. Furthermore, a single threshold effect was found between corruption and EE. Finally, based on the results of an empirical analysis, suggestions for a constructive policy are put forward.

**Keywords: energy efficiency, energy regulation, energy policy, government corruption, energy regulatory agency**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Kangyin Dong,*
*University of International Business and Economics, China*
*Yang Zhang,*
*Hunan University of Technology and Business, China*

***Correspondence:**
*Jianming Guo*
*jmguo@zufe.edu.cn*
*Farhad Taghizadeh-Hesary*
*farhad@tsc.u-tokai.ac.jp*
*Xiangfeng Liu*
*liuxiangfeng85@zufe.edu.cn*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *28 November 2020*
***Accepted:*** *15 March 2021*
***Published:*** *30 April 2021*

***Citation:***
*Liu J, Guo J, Liu X, Bai X and Taghizadeh-Hesary F (2021) Does Anti-corruption Policy Influence Energy Efficiency in China? Front. Energy Res. 9:634556. doi: 10.3389/fenrg.2021.634556*

## INTRODUCTION

Since the reform and opening up of China, with a rapid development of urbanization and industrialization, the conflicting demands of energy, environment, and economic development have become increasingly prominent (Zhao et al., 2020). Energy efficiency (EE) is defined by the International Energy Agency (IEA) as "a way to manage and limit energy consumption, which has also become an important goal and basis for energy management and energy policy-making in various countries" (Proskuryakova and Kovalev, 2015). Improving EE is an inevitable choice to alleviate the contradiction between energy supply and energy demand and to guarantee sustainable development, which is also an important measure to protect the environment and realize a sustainable development strategy. In the past 20 years, given the rapid growth of Chinese economy, energy production and energy consumption have increased exponentially. In 2018, China consumed the energy equivalent to 3,422.96 million tons of oil, which was 4.3% more than in 2017; annual power generation reached 7.4 trillion kilowatt-hours, which was a 7.2% year-on-year increase (BP World Energy Statistical Yearbook, 2019). **Table 1** shows that China occupies the first position in energy production and energy consumption in the world. As shown in **Figure 1**, EE in China has increased from 0.66 in 2000 to 1.19 in 2017 while maintaining a high growth rate of 3.53% per year. EE was calculated based on Equation (1), and the data were collated from the National Bureau of Statistics (2020). Thus, from the chart, we find that energy production and energy consumption in China are among the most prominent ones in the world energy market.

**TABLE 1 |** Comparison of energy consumption and electricity generation between China and other areas of the world in 2018.

| Countries or regions | Energy consumption (Million Tons Oil Equivalent) | Electricity generation (Terawatt-hours) |
|---|---|---|
| China | 3422.9 | 7422.0 |
| North American | 2831.9 | 5447.3 |
| South and Central America | 702.0 | 1305.3 |
| Europe | 2050.7 | 4077.3 |
| Commonwealth of Independent States (CIS) | 930.5 | 1417.4 |
| Middle East | 902.2 | 1240.4 |
| Africa | 461.5 | 853.7 |
| Asia Pacific(Except Whole China) | 2562.8 | 4851.5 |

*BP World Energy Statistical Yearbook (2019).*

However, China is facing severe corruption issues especially in the energy industry; **Figure 2** shows that China has a low ranking in the Corruption Perceptions Index released by Transparency International (ranging from 0 to 100, with higher scores corresponding to less corruption). Statistics published by the Central Commission for Discipline Inspection (CCDI) show that since the National Energy Administration (NEA) was established 10 years ago, two of the four directors (at a ministerial level) have been investigated for corruption (source: http://www.ccdi.gov.cn/scdc/). In addition, two deputy directors (at a bureau level) and eight department-level officials were also investigated by the CCDI. As the investigations progressed, many officials at different levels in energy systems of China were dismissed.

With a deepening of the investigations, many corruption cases associated with the energy industry, such as oil, coal, and electricity, have gradually been exposed and 80 officials at different levels have been investigated (source: https://www.sohu.com/a/255202925_752796). The exposed corruption was shocking and present in different areas of the energy industry. For example, in 2013, executives in more than 40 different levels from the China National Petroleum Corporation were involved.

How does China create EE under high-level corruption in the energy industry? This interesting phenomenon inspired us to study the relationship between the two in China. This toxic culture of corruption interacts with energy regulators; the exposure can help to motivate officials and reduce administrative delays and thus becoming a lubricant for the improvement of EE (Wang and You, 2012). On the other hand, rent-seeking and other corrupt behaviors of government regulatory officials will lead to the failure of regulatory policies, a waste of social costs, and low government efficiency (Kaller et al., 2018). Political expediency related to utilities such as electricity and gas tends to undermine regulatory independence and rigor in political environments with high levels of corruption. Moreover, Chinese local government officials often control resources, and the collusion of officials provide opportunities for rent-seeking and corruption, further aggravating the problem of resource mismatch (Aidt, 2016). This study combines corruption, energy regulations, and EE to address the following four questions: (1) Under the current situation in China, will government corruption hinder or promote EE? (2) What is the impact of different levels of corruption on EE? (3) Will increased energy regulations promote EE? (4) Are there any measures to improve EE?

This study uses the panel data from 2000 to 2017 of 30 provinces and provincial municipalities in China to analyze the difference between corruption and EE, utilizing the models of individual fixed effects measurement and the panel threshold. This research examines the interaction between dispatched energy regulatory agencies in different areas of China and the degree of corruption. Moreover, the influence of dispatched energy agencies on the efficiency of energy regulations by curbing corruption was also analyzed. Based on an empirical analysis, countermeasures and suggestions for improving supervision systems of the government, the organizational structure of regulations, and the promotion of the market reforms of energy industries in China are proposed.

The rest of this paper is organized as follows: Section Literature Review and the Theoretical Influencing Mechanism shows the literature review and the theoretical influencing mechanism. Section Methodology and Data presents the methodology and data. Section Empirical Analysis provides an empirical analysis. Section Conclusions and Policy Recommendations concludes the paper and provides policy recommendations.

## LITERATURE REVIEW AND THE THEORETICAL INFLUENCING MECHANISM

Energy industry is a regulated industry in which the regulatory authorities issue permits, approvals, and exemptions for operation, development, and construction. Lovei and McKechnie (2000) analyzed and evaluated the results of corruption activities and identified various forms of corruption in the energy sector. However, the study does not propose the steps that the USA should take to mitigate the risk of corruption at the public utility level. China is in the stage of economic transformation, and the content of regulatory responsibilities, laws, and regulations is still being improved. This inevitably involves business negotiations with the government, which has a larger space for rent-seeking. In a political environment with a high degree of corruption, political expediency related to utilities such as electricity and natural gas tends to weaken the independence and rigor of regulations (Eberhard, 2007). Therefore, regulations are easily captured

**Abbreviations:** Cor, Degree of Corruption; $CO_2$, Carbon Dioxide; CCDI, Central Commission for Discipline Inspection; EP, Energy Prices; EX, Degree of Openness; EE, Energy Efficiency; GV, Government Intervention; IEA, International Energy Agency; IS, Changes in Economic Structure; NEA, Administration; $PM_{2.5}$, Suspended Particulate Matter; R&D, Research and Development; $SO_2$, Sulfur Dioxide; TP, Technological Progress; TEPI, Total-factor Energy Performance Indicator.

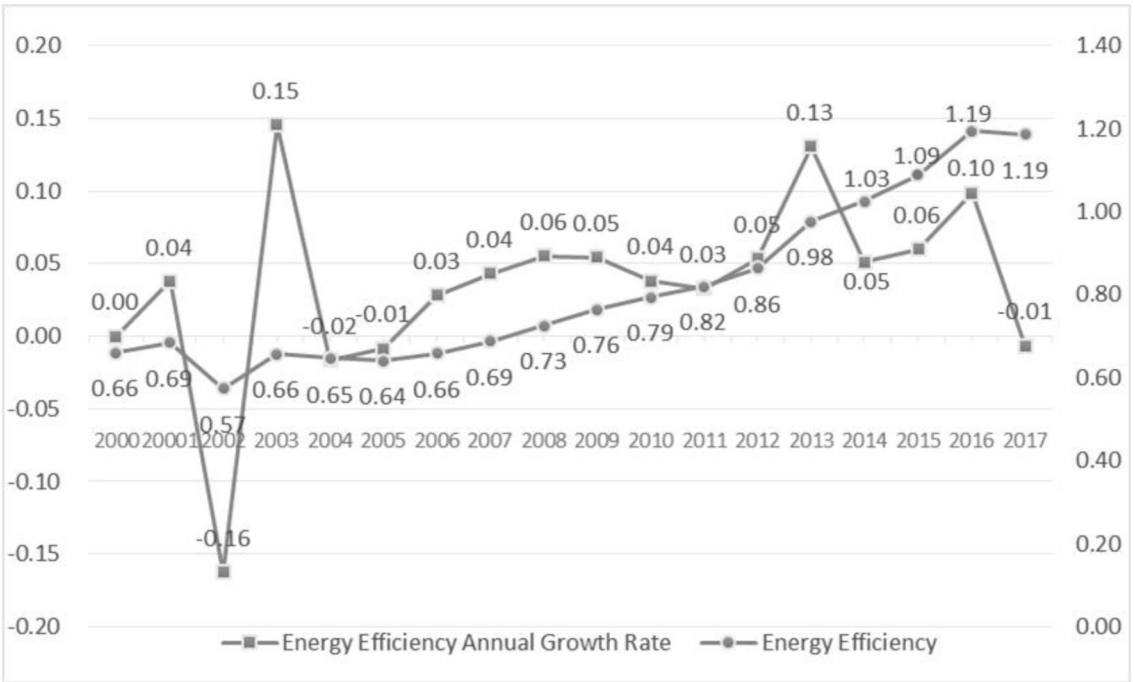



**FIGURE 1 |** Energy efficiency (EE) and its annual growth rate from 2000 to 2017 in China. Source: National bureau of statistic, People's Republic of China.

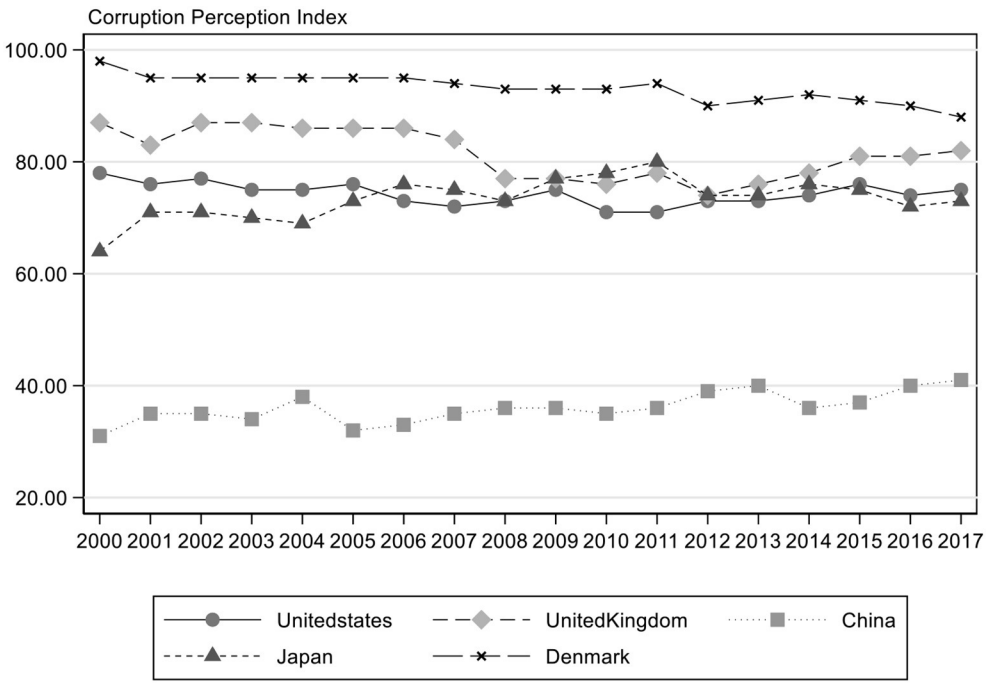



**FIGURE 2 |** Transparency international's corruption perception index from 2000 to 2017. Source: Transparency international.

by politics or become a rent-seeking tool for government officials. Many consumers also believe that utility operators have colluded with regulators to make excess profits (Stiglitz, 1998).

China is readjusting and reforming its economic structure and distribution model, but the problem of resource misallocation is still quite serious. Current research suggests that resource misallocation in China is greatly affected by local governments

that use administrative or economic means to influence the free flow of production factors to achieve the purpose of political execution and thus resulting in serious resource misallocation. From a perspective of resource allocation, some scholars believe that corruption further exacerbates the problem of resource misallocation (Liu et al., 2015; Restuccia and Rogerson, 2017; Pailler, 2018). The government regulatory officials' rent-seeking behavior will cause a failure of regulatory policies, a waste of social resources, and low government efficiency (Halkos and Tzeremes, 2013; Rose-Ackerman and Palifka, 2016; Kaller et al., 2018). In recent years, scholars have begun to study the relationship between corruption, resource misallocation, and EE. Corruption exacerbates resource misallocation and, in turn, impacts EE, leading to energy inefficiency (Wei and Li, 2017; Yang et al., 2018; Shibao et al., 2019, 2020, 2021). Local governments in China have concentrated on a large amount of capital and labor resources in some enterprises through administrative orders, which have artificially caused the misallocation of capital and labor among the industries. These enterprises tend to be traditional enterprises that cause high pollution and high energy consumption, so EE is greatly affected. Low EE due to resource misallocation leads to high emissions of sulfur dioxide ($SO_2$), soot, and suspended particulate matter ($PM_{2.5}$) (Bian et al., 2019). Wang et al. (2020) used the provincial panel data from 2006 to 2015 in China to empirically study the impact of corruption and resource misallocation on ecological efficiency. The research concludes that corruption directly reduces ecological efficiency and intensifies resource misallocation, leading to a further decline in ecological efficiency.

However, the "grease" theory of corruption has led some scholars to put forward a contrary research conclusion (Vial and Hanoteau, 2010; Wang and You, 2012). The risk of corruption in the energy sector is often linked to a significant influence of the government in the sector by organizing auctions and granting concessions. In legal systems with high regulatory interventions in business activities, the incidence of corruption is significantly higher. However, in the context of excessive regulation, rigid management, and a lack of market competition, corruption helps to overcome the resulting deeper distortions and becomes a suboptimal option (Leff, 1964; Leys, 1965; Huntington, 1968). Bribery or "refueling" reduces long administrative delays and advances through slow-moving public service queues. Kaufmann and Wei (2000) found that corruption cannot only save the management time wasted by the bureaucracy but also reduce the regulatory burden on enterprises, such as avoiding cumbersome regulations, reducing long waiting times, and speeding up the allocation of licenses and permits. Fan and Grossman (2001) found that those local officials who made greater contributions to the reform are likely to accept more bribes and have more valuable public property. Corruption in China can be regarded as a compensation or an incentive for the contributions of officials to improve their work enthusiasm. In this sense, corruption can also be seen as a new way to improve the effectiveness of supervision. Egger and Winner (2005) studied the data of 73 countries and found that corruption can bypass the regulations and speed up administrative processes. In addition, it improves the possibility of foreign investment to obtain public-funded projects and plays a positive role in foreign direct investment (FDI) both in the short and long term.

Some scholars also found that a positive or negative impact of corruption on energy regulations cannot be simply concluded because of its dependence on the complex national mechanism and various factors related to the specific economic, political, and social environment of each country (Vial and Hanoteau, 2010).

George and Tzeremes (2013) applied non-parametric estimators to study the impact of six governance measures (voice and accountability, political stability and lack of violence, government effectiveness, regulatory quality, rule of law, and corruption control) on $CO_2$ emissions by G20 governments. These results showed that the quality improvement of different governance factors in countries did not always reduce $CO_2$ emissions. Kaller et al. (2018) assessed the quality of supervision and the impact of corruption on electricity costs in European electricity market regulatory reforms. The results showed that improving regulatory quality and decreasing corruption could reduce electricity prices. However, when reforms were conducted in an institutional environment with severe corruption and poor regulatory quality, the reforms only limited prices.

Based on the previous research, we conclude that coordination costs of bribes may have two kinds of impact on energy regulatory policies and EE. Coordination costs can play a similar role as lubricants, promoting mutual cooperation between enterprises and governments and reducing rent-seeking costs. These factors are not always harmful in some periods and, thus, corruption levels can have both active and opposing sides on EE. However, in China, few studies have examined the relationship between the degree of corruption in government energy regulations or management and EE. Furthermore, almost no studies have analyzed the impact of government corruption and the structure of an energy regulatory organization on EE. Due to the complexity of the relationship between corruption and government regulation, resource allocation, and EE, further studies should be conducted to identify its impact on EE and provide effective recommendations for improving EE.

# METHODOLOGY AND DATA

In this section, 2000–2017 panel data from 30 provinces of China and provincial municipalities were obtained and analyzed. We adopted an individual fixed effects model and a panel threshold regression model to investigate the relevance of government corruption to EE. In the model, EE and the degree of corruption were utilized as an explanatory variable and a core explaining variable, respectively. Moreover, five indicators, technological progress (TP), energy prices (EP), changes in economic structure (IS), degree of openness (EX), and government intervention (GV), were selected as control variables in the model.

## Variables and Data

### Description of Variables

This study selects EE as an explanatory variable and the degree of corruption as a core explanatory variable. The interaction between government corruption and dispatched

energy regulatory agencies in different provinces and provincial municipalities was introduced in this study.

There are different indicators of EE particularly at a policy level. The economic outcomes of the energy investment ratio or the output-to-input ratio are the two most commonly used indicators of EE in a certain area. The former stands for "energy efficiency," which measures the macro EE level of a region or country in a certain phase by the service or output per unit of energy costs (Patterson, 1996); the latter measures the energy consumption of unit output and it is known as energy intensity. These two indicators are reciprocal and easy to evaluate (Kander et al., 2017).

In this thesis, the proportion of constant price gross domestic product (GDP; constant prices, base year 2000) to the total energy consumption is a measure of appraising comprehensive EE, which agrees with the indicators used in the literature to measure EE (Liu et al., 2015; Li and Lin, 2017). Researchers such as Dan (2006), Yang (2010), and Wang J. (2018) also used the same measure. Based on this, the following equation is developed for evaluating the regional EE for each province of China:

$$EE_{it} = \frac{GRP_{it}}{E_{it}} \tag{1}$$

where $EE_{it}$ represents EE for the province $i$ in the year $t$, $GRP_{it}$ stands for a constant price gross regional product (constant prices, base year 2000) for the province $i$ in the year $t$ measured by 108 Yuan, and $E_{it}$ is the regional energy consumption in the province $i$ in the year $t$ measured by an equivalent of 104 tons of coal.

The degree of government corruption in each region is a core explanatory variable in the proposed model. Because corruption is hidden, a determination is made how to measure corruption effectively is usually difficult in quantitative research. International organizations have raised several corruption indices, such as the Corruption Perception Index and Bribe Payers Index released by Transparency International and a set of indexes reflecting the degree of corruption of various countries published by Business International (Knack, 2006). Although the data are widely used in cross-border corruption research, their focus is quite different and no index can fully measure all kinds of corruption (Knack, 2006). Moreover, the validity of an index is difficult to guarantee because it is based on the survey information. Therefore, researchers have attempted to use objective indicators to characterize the degree of corruption. At present, the most commonly used objective indicators for corruption mainly include the number of corruption cases per million of the population (Del Monte and Papagni, 2007; Wu, 2008) and the number of corruption cases among the public officials (Fisman and Gatti, 2002; Yongzheng and Haibo, 2011). Because the data show only the government corruption that has been found or investigated, it may be just a tip of the iceberg of the actual amount of corruption. However, the results of the research of domestic and foreign scholars show that taking data as an indicator of the degree of government corruption is more consistent with reality (Apergis et al., 2010; Yongzheng and Haibo, 2011; Wang et al., 2020). The subject of corruption in Chinese law is defined as government staff, and this paper mainly studies the degree of corruption in the public sector. Therefore, the number of cases of corruption, bribery, and malfeasance in the province $i$ in the year $t$, denoted as $Case_{it}$, and the number of civil servants, noted as $CivilServants_{it}$, is used to evaluate the degree of corruption (Cor) in each province $i$ in the year $t$ as follows:

$$Cor_{it} = \frac{Case_{it}}{CivilServants_{it}} \tag{2}$$

The natural logarithm of the corruption degrees, lnCor, for the 30 Chinese provinces and provincial municipalities is calculated and is shown in **Table 4**, Section Data and Statistical Description.

China has implemented an administrative regulation unification mode so that dispatched energy regulatory agencies are government bureaus. There exists an interaction between government corruption and the existence of dispatched energy regulatory agencies. Therefore, this paper introduces a virtual variable ($Regu_{it}$) representing the setup of dispatched energy regulatory agencies in the province $i$ in the year $t$. The value $Regu_{it}$ equals 1 if a dispatched energy regulatory agency has been set up in the province $i$ in the year $t$; otherwise, it equals 0. **Table 2** illustrates provinces and provincial municipalities or regions setting up energy regulatory agencies and their corresponding times. The interaction term between the setup of the dispatched energy regulatory agency and regional corruption is used as an explanatory variable in the model, which is denoted as Regu*lnCor.

**Table 3** shows the selected variables and their brief explanation. This study also selected five representative indicators as control variables: TP, EP, IS, EX, and GV.

Technological progress: Li and Lin (2017) developed a total-factor energy performance indicator (TEPI) to evaluate the EE of 30 provinces and provincial municipalities in China. The results illustrate that TP is fundamental to improving the EE in China. Zhu et al. (2019) studied how the abovementioned factors relate to each other and believed that technological advancements could make energy use more efficient at an annual average rate of 7.1%. Indicators are productivity improvements (Aguilera and Ripple, 2012), patents (Albino et al., 2014; Li and Lin, 2017), research and development (R&D) investment (Lin and Zhang, 2013; Ouyang and Lin, 2015), financial development, and FDI (Li and Lin, 2014). The average growth rate of per capita GDP over the past 3 years (Antonietti and Fontini, 2019) is often used as a surrogate variable to measure changes in TP. In this paper, we use the R&D spending in each province, as a proportion of GDP for measuring regional R&D capacity (Lin and Zhang, 2013; Ouyang and Lin, 2015). The greater the investment in R&D, the stronger the ability to develop advanced technologies.

Energy prices (EP): EP are generally considered to be an essential motivation for saving energy and promoting EE (Huang et al., 2017). However, market regulation barriers caused by information dissymmetry, high transaction costs, and the contorted energy price resulting from inappropriate promotion policies adversely affect improvements in EE (Oikonomou et al., 2008). EP in different regions of the world have various impacts on EE (Antonietti and Fontini, 2019). In recent years, researchers

**TABLE 2 |** Year of establishment of local electricity dispatched agencies.

| Level | Name of organization | Location | Setup time |
|---|---|---|---|
| Six regional | Northeast electricity regulatory bureau | Jilin | 2004 |
| regulatory bureaus | East China electricity regulatory bureau | Shanghai | 2004 |
| | Central China electricity regulatory bureau | Hubei | 2004 |
| | South electricity regulatory bureau | Guangdong | 2004 |
| | North China electricity regulatory bureau | Beijing | 2005 |
| | Northwest electricity regulatory bureau | Shanxi | 2005 |
| 12 provincial | Sichuan electricity regulatory office | Sichuan | 2004 |
| regulatory offices | Shanxi electricity regulatory office | Shanxi | 2005 |
| | Gansu electricity regulatory office | Gansu | 2005 |
| | Zhejiang electricity regulatory office | Zhejiang | 2005 |
| | Jiangsu electricity regulatory office | Jiangsu | 2005 |
| | Fujian electricity regulatory office | Fujian | 2005 |
| | Henan electricity regulatory office | Henan | 2005 |
| | Yunnan electricity regulatory office | Yunnan | 2005 |
| | Guizhou electricity regulatory office | Guizhou | 2005 |
| | Shandong electricity regulatory office | Shandong | 2006 |
| | Hunan electricity regulatory office | Hunan | 2006 |
| | Xinjiang electricity regulatory office | Xinjiang | 2010 |

*Tan (2014).*

**TABLE 3 |** Selected variables and their brief explanation.

| Attribute | Variable | Abbreviation | Variable Description |
|---|---|---|---|
| Explained variable | energy efficiency | EE | The ratio of constant price GDP (the constant year 2000 prices) to the total energy consumption, see Equation (1) |
| Explanatory variable | degree of corruption (ln) | lnCor | The ratio of the number of cases including corruption, bribery and malfeasance to the number of civil servants, see Equation (2) |
| Virtual variable | administrative regulation | Regu | Indicator for the existence of related dispatched energy regulatory agencies, see **Table 2** |
| Control variables | Technological progress | TP | The proportion of research and development (R&D) expenditure to GDP |
| | Energy price | EP | Purchase price indexes of major raw materials, fuel, and electricity |
| | changes in the economic structure | IS | The proportion of the added value of the tertiary industry to GDP |
| | Degree of openness | EX | The proportion of total imports and exports to GDP |
| | Government intervention | GV | The ratio of government fiscal expenditure to GDP |

have conducted extensive and in-depth studies on how energy efficiency in China influences its prices. Some scholars (Fisher-Vanden et al., 2004; Wu, 2009) found that improving EP was conducive to improving EE, while others (Lin and Du, 2013; Ju et al., 2017) also proposed that the price distortions of power, natural gas, coal, and other energy markets caused by strict price regulations and imperfect market mechanisms in China significantly prevented EE from improving. In this study, the producer price indexes of primary raw materials, fuel, and electricity were adopted as proxy variables for EP (Shi and Shen, 2013).

Economic structural change (IS): due to the large differences in energy demand between various economic structures, economic structure can also play an essential role in EE. Moreover, structural changes may also cause tremendous changes in EE over time (Dong et al., 2018). Voigt et al. (2014) analyzed trends and driving factors of energy demand strength in about 40 major economies and concluded that economic structural changes do not significantly affect EE in most countries. However, in Japan, the USA, Australia, Taiwan, Mexico, Brazil, and other countries, an adjustment to industrial structure is regarded as a main driver of the decline in energy demand intensity. Evaluating the ratio of industrial added value to GDP has been widely used to measure economic structure (Dong et al., 2018). We use this method to calculate the proportion of the added value of the tertiary industry in the nominal GDP as the proxy variable of the economic structure.

Degree of openness: foreign commerce is not only an important source of TP but also drives improvements in EE (Lin and Liu, 2015; Roy and Yasar, 2015). However, foreign trade will make a difference in EE, but it varies from country to country. Cole and Tran (2011) analyzed the panel data of 32 developed and developing countries from 1975 to 1995 and found that imported or exported energy-intensive commodities influenced the trade impact on energy intensity. In this thesis, the appropriate volume of imports and exports to nominal GDP of each province is the measurement index.

Government intervention: Chinese GV has positive and negative effects on EE of a country. On one hand, GV may lead to a distorted resource allocation and reduce EE. On the other hand, due to market failures in environmental quality governance,

appropriate market and GV will benefit the improvement in EE (Shi and Shen, 2013). Relevant indicators of marketization level and the degree of the government interference are measured by government spending in proportion to GDP. Lower ratios indicate less GV and higher marketization.

### Data and Statistical Description

To ensure data validity, the author selected the data from 30 provinces and cities in China (except Tibet, Hong Kong, Taiwan, and Macao) from 2000 to 2017. The number of cases of corruption and malfeasance was obtained from the annual work reports of provincial procuratorates in the Procuratorial Yearbook of China. Due to the lack of data in the pro-curatorial yearbooks of some provinces and provincial municipalities in some years of some provinces and provincial municipalities in some years, we have supplemented the data with relevant local legal yearbooks, economic yearbooks, and work reports. The calculation of the degree of corruption (Cor) is given in Equation (2), and the natural logarithm value of the corruption degrees (lnCor) of each region in the sample period is shown in **Table 4**. In **Table 4**, Beijing, Gansu, and Hunan have low degrees of corruption and are ranked in the top three positions.

The energy expenses of each province over the studied period were obtained from the China Energy Statistical Yearbooks[1,2] of the provinces and provincial municipalities. The nominal R&D expenditures were obtained from the China Statistical Yearbook of Science and Technology. The energy price data used the producer price indexes of major raw materials, fuel, and power in the China Statistical Yearbook. The total volume of imports and exports was converted into RMB at the average exchange rate of RMB against the US dollar for the current year. In calculating EE, the CPI was used to reduce GDP to the actual value based on the year 2000 to eliminate price factors. Descriptive statistics of the variables are listed in **Table 5**.

As **Table 5** shows, the sample SD of the interaction term between the setup of dispatched energy regulatory agencies, and corruption and the EX are slightly larger than the means, and the sample SD of other variables are all smaller than the means; that is, the sample coefficient of variation is less than 1. This illustrates that the sample data structure are stable and that there are no extreme outliers; thus, subsequent empirical analysis can be performed on this basis.

## Statistical Model

### Fixed Effects Model

The framework of a static panel data model has been widely applied in empirical studies on the factors affecting EE. This method is useful for controlling the bias effect of individual effects in the sample, which are difficult to observe and do not change over time. Before regression, the model settings were subjected to the Hausman test. The Hausman test statistic value is 133.81, which is significant at the level of 1%; therefore, we preferred fixed effects to the alternative hypothesis of the random one.

We develop the following equation of individual fixed effects:

$$EE_{it} = \alpha_0 + \beta \ln Cor_{it} + \theta X_{it} + \eta_i + \lambda_t + \varepsilon_{it} \quad (3)$$

where $i$ is the province index ($i = 1,2,\ldots,30$), $t$ is the year index ($t = 2000, 2001,\ldots, 2017$), $EE_{\mathrm{it}}$ is the EE of province $i$ in year $t$, is the corruption degree of the province $i$ in the year $t$, $X_{\mathrm{it}}$ represents a set of control variables that may affect a province's EE, $\theta$ is the corresponding coefficient variable, $\eta_{\mathrm{i}}$ represents individual effects that are difficult to observe, $\lambda_{\mathrm{t}}$ is the time effect, and $\varepsilon_{\mathrm{it}}$ is the disturbance term.

Government structure and size are important factors that influence government corruption. Wang J. (2018) have suggested that the NEA power regulatory dispatch agency will improve EE. China has implemented an administrative regulation unification mode, so the energy regulatory department is a part of government agency; thus, government corruption and the establishment of energy regulatory dispatch agencies have a mutual influence. The model takes the interaction term $Regu \times \ln Cor$ as an explanatory variable to analyze its effects on EE, which can be represented by the following equation:

$$EE_{it} = \alpha_0 + \beta_1 \ln Cor_{it} + \beta_2 Regu_{it} \times \ln Cor_{it} + \theta X_{it} + \eta_i + \lambda_t + \varepsilon_{it} \quad (4)$$

### Panel Threshold Model

Corruption reduces government efficiency, damages social welfare, and inhibits economic growth (Mauro, 1995; Mo, 2001; Chen et al., 2008). However, in developing countries, some studies have indicated that because corruption can relax the inefficiencies and rigid regulations imposed by governments, welfare can be improved (Leff, 1964). Nye (1967) argued that there may exist a positive level of corruption that maximizes the output.

Méndez and Sepúlveda (2006) have suggested that economic growth affected by corruption may be non-linear. When corruption is at a low level, it is conducive to economic growth, while it would harm economic growth when the corruption levels are high. Wu and Rui (2010) have researched the connection between the abovementioned two factors in various regions of China indicated that there is a U-shaped relationship between the two. Corruption has a dual threshold effect on China's regional TP brought about by FDI technology spillover and a single threshold effect on R&D, foreign trade, and TP (Ma, 2015). Since corruption has different influences on economic growth and technology spillovers, we doubt that it exists in EE as well. This phenomenon is called the "threshold effect" by Borensztein et al. (1998); in other words, only when the extent of government corruption in a certain area exceeds a certain "threshold" level, it can have an impact on EE. To evaluate this threshold value, Hansen (1999) proposed a regression model for the non-dynamic panel threshold, which can also perform a significance test on the correctness of the value.

The idea aims to introduce the threshold, which is an unknown variable in the empirical model, to develop a segmented

[1]China Energy Statistical Yearbooks 1986-2018. Available online at: http://data.cnki.net/ Trade/yearbook/single/N 201 90800 25?z=Z024. (accessed December 15, 2019).

[2]China Statistical Yearbook 1999-2019. Available online at: http://data.stats.gov.cn/publish.htm? sort =14. (accessed December 15, 2019).

**TABLE 4 |** Ranking of corruption of 30 provinces of China and provincial municipalities 2000–2017.

| Province | lnCor | Rank | Province | lnCor | Rank | Province | lnCor | Rank |
|---|---|---|---|---|---|---|---|---|
| Beijing | 2.337 | 1 | Shanxi | 3.179 | 11 | Tianjin | 3.398 | 21 |
| Gansu | 2.945 | 2 | Zhejiang | 3.20 | 12 | Jiangsu | 3.429 | 22 |
| Hunan | 2.984 | 3 | Shaanxi | 3.203 | 13 | Henan | 3.454 | 23 |
| Xinjiang | 3.014 | 4 | Ningxia | 3.275 | 14 | Anhui | 3.479 | 24 |
| Guangdong | 3.054 | 5 | Hebei | 3.287 | 15 | Hubei | 3.483 | 25 |
| Inner Mongolia | 3.077 | 6 | Shandong | 3.311 | 16 | Liaoning | 3.506 | 26 |
| Qinghai | 3.090 | 7 | Guizhou | 3.319 | 17 | Guangxi | 3.509 | 27 |
| Hainan | 3.102 | 8 | Jiangxi | 3.359 | 18 | Fujian | 3.550 | 28 |
| Shanghai | 3.149 | 9 | Yunnan | 3.390 | 19 | Heilongjiang | 3.585 | 29 |
| Sichuan | 3.169 | 10 | Chong qing | 3.391 | 20 | Jilin | 3.702 | 30 |

*Data taken from the National Bureau of Statistics, collated by the author. Tibet, Hong Kong, Macao, and Taiwan were excluded.*

**TABLE 5 |** Descriptive statistics of variables.

| Variable | *N* | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| EE | 540 | 0.813566 | 0.397413 | 0.193191 | 2.90826 |
| lnCor | 540 | 3.264933 | 0.385547 | 0.550979 | 4.38794 |
| Regu* lnCor | 540 | 9.734171 | 12.9847 | 0 | 59.32627 |
| TP | 540 | 0.012868 | 0.010375 | 0.001381 | 0.060136 |
| IS | 540 | 0.423146 | 0.083241 | 0.283029 | 0.805562 |
| EX | 540 | 0.312913 | 0.387557 | 0.016868 | 1.7843 |
| GV | 540 | 0.1975567 | 0.091709 | 0.069126 | 0.626863 |
| EP | 540 | 1.0585 | 0.081186 | 0.835 | 1.261 |

function for the regression coefficient of an explanatory variable, and perform a series of estimates and significance tests on the threshold and threshold effects. The fundamental form of the fixed effects panel threshold model can be represented as follows:

$$\begin{aligned} Y_u &= u_i + \beta_1 X_{it}(q_{it} \le \gamma) + \varepsilon_{it} \\ Y_u &= u_i + \beta_2 X_{it}(q_{it} > \gamma) + \varepsilon_{it} \end{aligned} \tag{5}$$

where $i$ is the province index ($i = 1, 2, \ldots, 30$), $t$ is the year index ($t = 2000, 2001, \ldots, 2017$), $q_{it}$ is the threshold parameter, $\gamma$ is the threshold value to be estimated, $x_{it}$ represents an explanatory variable, which is independent of the disturbance item $\varepsilon_{it}$, and $u_i$ is the individual intercept term. Equation (5) is a segmented function model, when $q_{it} \le \gamma$, the coefficient of $x_{it}$ is $\beta_1$, $q_{it} > \gamma$, and the coefficient is $\beta_2$.

To further test the influence of corruption levels on EE, we have introduced the threshold value of corruption as an unknown variable into the fixed effects model and developed a piece-wise function of corruption levels, represented as follows:

$$\begin{aligned} EE_{it} &= \alpha_0 + \beta_1 \ln Cor_{it}(\ln Cor_{it} \le \gamma) + \beta_2 \ln Cor_{it}(\ln Cor_{it} > \gamma) \\ &\quad + \theta X_{it} + \eta_i + \lambda_t + \varepsilon_{it} \end{aligned} \tag{6}$$

where $\mathrm{InCor_{it}}$is the threshold variable and $\gamma$ is the threshold value to be estimated. Wald tests can be applied to test threshold effects (Hansen, 1999).

# EMPIRICAL ANALYSIS

## National Level

In the regression results at the national level, as the number of control variables increases (from experiments (1) to (5) in **Table 6**), the regression coefficient of the explanatory variable lnCor is negative and stable between −0.13 and −0.20, maintaining a significance level of 1%. In addition, other control variables showed a high level of significance. **Table 6** shows that the test results are robust.

Based on the results of experiment (4) in **Table 6**, the corruption level affecting EE in China was systematically analyzed. After controlling the variables including TP, IS, EX, EP, and GV and controlling the regional and time effects, the regression coefficient of an explanatory variable is −0.139, passing the significance inspection at the 1% level. This means that corruption increases by every 1%, and the EE decreases by 0.139%. Based on these results, it can be seen that corruption significantly suppresses EE in China.

Considering the impact of corruption on EP and GV, there may be endogeneic problems. Therefore, this study uses the endogenous test method, the Davidson–MacKinnon test, to test for endogeneity. The F-statistic of the Davidson–MacKinnon test is 16.81, and the value of $p$ is 0.000, which significantly rejects the original hypothesis, indicating that there is no endogenous problem in the model. Therefore, we believe that the model has no endogeneic problems.

**TABLE 6 |** National regression results.

| Variable | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | EE | EE | EE | EE | EE |
| lnCor | −0.203*** | −0.156*** | −0.140*** | −0.139*** | −0.132*** |
| | (0.0275) | (0.0257) | (0.0249) | (0.0250) | (0.0265) |
| Regu*lnCor | | | | | −0.000673 |
| | | | | | (0.000795) |
| TP | | 20.94*** | 15.50*** | 15.65*** | 15.60*** |
| | | (2.403) | (2.486) | (2.500) | (2.501) |
| IS | | 0.354* | 0.434** | 0.425** | 0.452** |
| | | (0.210) | (0.203) | (0.204) | (0.206) |
| EX | | | −0.111** | −0.108** | −0.105** |
| | | | (0.0510) | (0.0513) | (0.0515) |
| GV | | | −1.095*** | −1.104*** | −1.120*** |
| | | | (0.201) | (0.202) | (0.203) |
| EP | | | | 0.114 | 0.109 |
| | | | | (0.184) | (0.184) |
| Constant | 1.794*** | 1.117*** | 1.431*** | 1.309*** | 1.290*** |
| | (0.0859) | (0.137) | (0.143) | (0.242) | (0.243) |
| Regional effect | Controlled | Controlled | Controlled | Controlled | Controlled |
| Time effect | Controlled | Controlled | Controlled | Controlled | Controlled |
| Observations | 540 | 540 | 540 | 540 | 540 |
| R–squared | 0.706 | 0.756 | 0.773 | 0.773 | 0.774 |
| Number of id | 30 | 30 | 30 | 30 | 30 |

*Authors' calculation.*
*SE are in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1.*

The regression coefficient of the interaction term Regu × lnCor in experiment (5) was also negative. However, they failed to pass the significance test. Thus, from a national perspective, the interaction between the political environment of corruption and dispatched energy regulatory agencies has weakened government regulatory efficiency. In a corrupt environment, the contribution of dispatched energy regulatory agencies to the improvement of EE is limited and insignificant.

From the regression results of the control variables, an increase in the proportional ratio of R&D investment to GDP and tertiary industry can promote EE in China to varying degrees. The proportional ratio of R&D investment to GDP can obviously promote EE, which also shows that the TP in China has a prominent role in facilitating EE. GV has significantly inhibited the improvement of EE, which is consistent with a previous research showing that GV has distorted resource allocation and reduced EE.

The EX has a significantly negative correlation with EE at the level of 5%. On one hand, EE can be improved by foreign investment for advanced technology and managerial expertise; on the other hand, they can also reduce EE by introducing high energy consumption and pollution enterprises. From the negative coefficient of the level of openness, we can infer that the negative effect is larger than the positive effect.

The coefficient of EP is positive but fails to pass the significance test. This means that China has not yet established a market-based energy price mechanism. Prices do not have obvious effects on improving the EE. Therefore, the government should promote the market mechanism reform of EP.

## Eastern[3], Central[4], and Western[5] Regions of China

The regression results for the eastern, central, and western regions of China are shown in experiments (6) to (11) in **Table 7**. In the eastern and western regions, regional corruption negatively influences EE, and the negative effect is significant at 1 and 5% levels, respectively. Corruption is significant when it comes to EE in the eastern region, with a coefficient of −0.264, and a lesser impact on the western region, only −0.0552. Therefore, EE is more sensitive to changes in the level of corruption in the eastern region. For the central region, the

[3]Eastern region of China refers to 11 provinces and provincial municipalities including Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan.

[4]Central region of China refers to eight provinces including Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan.

[5]Western region of China in this paper refers to 11 provinces and provincial municipalities including Sichuan, Guizhou, Yunnan, Inner Mongolia, Guangxi, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang, and Chongqing. Tibet province is not included due to the missing data.

**TABLE 7 |** Regional regression results.

| Variable | Eastern region | | Central region | | Western region | |
|---|---|---|---|---|---|---|
| | (6) | (7) | (8) | (9) | (10) | (11) |
| lnCor | −0.264*** | −0.215*** | 0.0752** | 0.104** | −0.0552** | −0.0505** |
| | (0.0592) | (0.0605) | (0.0346) | (0.0399) | (0.0245) | (0.0252) |
| Regu* lnCor | | −0.00571*** | | −0.00164 | | 0.000348 |
| | | (0.00202) | | (0.00117) | | (0.000863) |
| TP | 15.70*** | 19.13*** | −10.10** | −12.36*** | 5.641 | 7.014** |
| | (5.907) | (5.911) | (4.152) | (4.438) | (3.430) | (3.263) |
| IS | 0.688 | 0.793 | −0.694** | −0.704*** | 0.4736** | 0.474** |
| | (0.654) | (0.642) | (0.267) | (0.266) | (0.207) | (0.229) |
| EX | −0.0647 | −0.000368 | 0.680** | 0.604** | 0.931*** | 0.929*** |
| | (0.0934) | (0.0943) | (0.288) | (0.292) | (0.129) | (0.129) |
| GV | −0.476 | −0.968 | −0.500 | −0.868 | −1.000*** | −1.059*** |
| | (0.675) | (0.684) | (0.621) | (0.672) | (0.151) | (0.147) |
| EP | −0.247 | −0.188 | 0.420* | 0.376 | 0.257 | 0.257 |
| | (0.419) | (0.411) | (0.252) | (0.253) | (0.182) | (0.185) |
| Constant | 1.982*** | 1.767*** | 0.960*** | 1.073*** | 0.581** | 0.596** |
| | (0.562) | (0.556) | (0.363) | (0.370) | (0.226) | (0.232) |
| Regional Effect | Controlled | Controlled | Controlled | Controlled | Controlled | Controlled |
| Time effect | Controlled | Controlled | Controlled | Controlled | Controlled | Controlled |
| Observations | 198 | 198 | 144 | 144 | 198 | 198 |
| R–squared | 0.798 | 0.808 | 0.909 | 0.910 | 0.811 | 0.811 |
| Number of id | 11 | 11 | 8 | 8 | 11 | 11 |

*Authors' calculation.*
*SE are in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1.*

corruption coefficient is 0.0752, which passes the significance test at 5% level, indicating that a positive effect of "corruption lubricant" in the middle part is larger than its impact on efficiency distortion. In contrast, the regression coefficients of the explanatory variable lnCor in the central and western regions are smaller than those for the eastern region, which shows that the reflection of EE on the change in corruption is weaker than that for these two regions in the eastern region. Therefore, corruption has different effects on EE, which can be positive or negative in different regions.

Setting up of dispatched regional and provincial energy regulatory agencies is an important measure to increase energy regulations, but corruption has affected the effectiveness of regulations. Experiments (7), (9), and (11) show the regression results after introducing the interaction between the establishment of dispatched regulatory agencies and corruption, Regu*lnCor.

The interaction coefficient in the eastern region is −0.00571 and is significant at the level of 1%, which means that corruption negatively affects government regulatory efficiency in the east. However, as the regression coefficient of "corruption" has decreased from −0.264 to −0.215, which shows that an increasing regulatory oversight can reduce the influence of corruption on EE. The coefficents of interaction terms of the central part and western part are not significant, which indicates that corruption has a little effect on government regulatory efficiency. However, when the interaction term was added, positive impact coefficient of the central region was increased, which caused corruption to increase EE, but that of the western region remained basically unchanged.

The correlation between the overall impact of EE and TP is significant and positive. However, the coefficient of TP in the central region is −10.10 and through a significance test at the level of 5%, and shows that blindly improving the technological level cannot completely settle the issue of low EE. EP in different regions have a little effect on EE, mainly because China's energy price market mechanism has not been fully developed, and the energy price cannot fully reflect the corresponding cost. The proportion of tertiary industry structures is positively correlated with overall EE. However, the positive impact is not significant in the east, and the effect direction is opposite in the middle and west, which shows that EE will be affected by the interaction between the degree of marketization and the industrial structure. In the eastern region, the influence of openness on EE is not apparent, but positive in the central and western regions. GV also showed regional heterogeneity. The effect of GV on EE in the eastern region and the central region is not conspicuous, but it significantly contributes to the control of the western region.

## An Empirical Analysis of the Threshold Effect Model and Discussion

From the abovementioned analysis, we find that the degree of corruption in EE varies in different regions. To determine whether there is a threshold effect value for the degree of

**TABLE 8 |** Panel threshold self–sampling model.

| Model | F value | *P*-value | BS times | Critical value | | |
|---|---|---|---|---|---|---|
| | | | | 1% | 5% | 10% |
| Single-threshold value | 114.54 | 0.0000 | 300 | 21.7178 | 25.7195 | 41.1447 |
| Double-threshold value | 5.72 | 0.8400 | 300 | 27.4541 | 86.7179 | 140.0782 |
| Triple-threshold value | 18.27 | 0.0725 | 200 | 15.9984 | 19.5583 | 29.1774 |

**TABLE 9 |** Threshold effect estimates and CIs.

| Model | Threshold estimate value | 95% CI |
|---|---|---|
| Single-threshold model | 2.4307 | [2.3218, 2.4788] |

**TABLE 10 |** Parameter estimation results.

| Variable | Coef. | Std. Err. | t | P>\|t\| |
|---|---|---|---|---|
| TP | 24.22135 | 2.277747 | 10.63 | 0.000 |
| IS | 1.36065 | 0.1749581 | 7.78 | 0.000 |
| EX | −0.2758207 | 0.0534638 | −5.16 | 0.000 |
| GV | 4248051 | 0.1439853 | 2.95 | 0.003 |
| EP | −0.2050109 | 0.0850848 | −2.41 | 0.016 |
| lnCor1 | 0.0512546 | 0.0393031 | 1.3 | 0.193 |
| lnCor2 | −0.1310521 | 0.0255351 | −5.13 | 0.000 |
| c | 0.5651986 | 0.1724586 | 3.28 | 0.001 |

corruption, the corruption threshold effect model proposed by Equation (6) was utilized for the analysis. **Tables 8**, **9** illustrate the estimated threshold values for corruption and their CIs. The results indicate that the model has a single threshold effect at a significance level of 1%, with a threshold of 2.4307. As shown in **Figure 3**, the likelihood ratio (LR) value corresponding to the threshold estimate of the degree of corruption is significantly less than the critical value, and thus the obtained threshold estimate is effective. According to the threshold value, the selected provinces and provincial municipalities are divided into low corruption regions with $\ln Cor \leq 2.4307$ and high corruption regions with $\ln Cor > 2.4307$.

As shown in **Table 9**, when the level of corruption is below the threshold of 2.4307, which means that the incidences of corruption, bribery, and malfeasance cases per 10,000 government officials are <11.37, the regression coefficient is positive but not significant. The result indicates that when the degree of corruption is lower than the threshold, corruption has a little effect on EE. However, when the level of corruption is greater than the threshold, it has a prominent negative influence on EE and has a regression coefficient of −0.13, which means that when the corruption level is 1%, the EE decreases by 0.13%. Judging from the average corruption level in the sample period, other regions in China, except Beijing, have passed the high threshold of corruption. In other words, government corruption in most regions of China has largely inhibited the improvement of EE. Therefore, the government should continue to strengthen anti-corruption efforts to improve EE.

## Robustness Test

To verify the robustness between corruption and EE, a sensitivity analysis was conducted.

First, this study utilizes another index to characterize the general corruption degree, that is, the number of corruption cases in a million rather than per 10,000 civil servants in Equation (2). The new index is employed as a proxy variable for the degree of corruption to retest how corruption and EE relate to each other (Dong and Torgler, 2013; Wang and Wang, 2016; Wang and Gao, 2017). The results of the robustness test are shown in **Table 10** shows the parameter estimation results. **Table 11** and experiments (12) to (14) will be discussed in the following sections.

Experiment (12) shows the regression results when applying a new index of the degree of corruption. In **Table 11** (12), the absolute value of the corruption coefficient on EE is 0.118, which is slightly smaller than 0.139 in the previous experiment (4) in **Table 6**, while the negative effect and coefficient significance remain the same. As for the other control variables, the sign of the coefficients and their significance are almost consistent with the previous results.

Second, corruption cases, as a main measure of the degree of corruption, have a significant time lag. Generally, corruption cases that occurred in a certain year will not be investigated and punished in that year. The lag period of cases normally varies from 1 to 6 years, even more than 10 years for certain special cases, although most corruption cases will be exposed within 3 years. Therefore, following Zhu (2016), the degree of corruption is recalculated based on the average of the number of post crimes in the current period and the next three periods. Moreover, a regression experiment (13) is performed based on a new degree of corruption to further validate the relation between corruption and EE. As shown in **Table 11** (13), the influence of corruption on EE is consistent with the previous conclusion. The absolute value of the corruption coefficient to an EE of 0.187 becomes slightly larger than the previous 0.139, which further emphasizes a negative influence of corruption on EE in China.

Third, considering a steady number of corruption cases in China in the past 10 years, the entire period from 2008 to 2017 was taken as the research object from the full sample to test the stability of the results. The regression results are presented in **Table 11** (14). Compared with the regression results in experiments (12) and (13), the signs of the explanatory variable

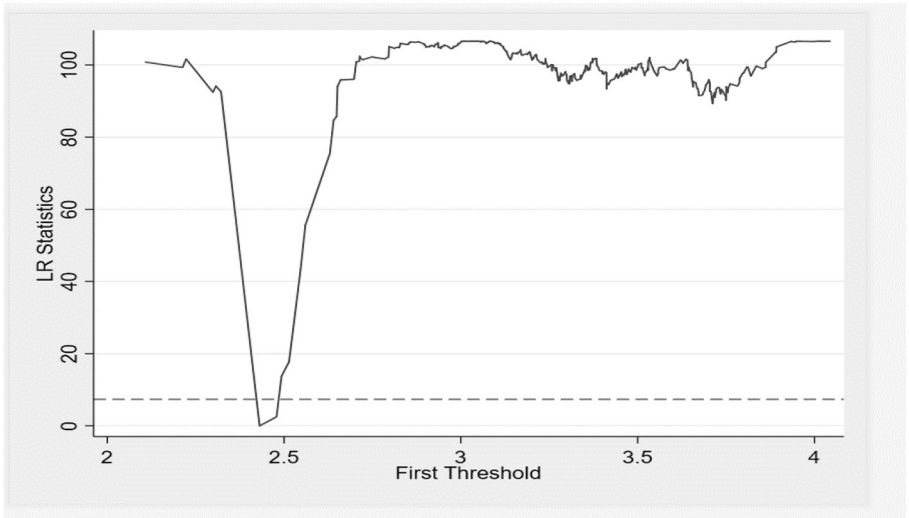



**FIGURE 3 |** Threshold effect.

**TABLE 11 |** Robustness test.

| Variable | (12) | (13) | (14) |
|---|---|---|---|
| | EE | EE | EE |
| lnCor | −0.118*** | −0.187*** | −0.155*** |
| | (0.0248) | (0.0257) | (0.0328) |
| TP | 16.75*** | 15.45*** | 24.93*** |
| | (2.510) | (2.447) | (4.934) |
| IS | 0.503** | 0.326 | 0.164 |
| | (0.205) | (0.201) | (0.342) |
| GV | −0.132** | −0.0995** | −0.261*** |
| | (0.0517) | (0.0503) | (0.0877) |
| EX | −1.003*** | −0.853*** | −1.525*** |
| | (0.207) | (0.203) | (0.387) |
| EP | 0.103 | 0.142 | −0.0721 |
| | (0.185) | (0.180) | (0.282) |
| Constant | 0.682*** | 1.410*** | 1.677*** |
| | (0.227) | (0.235) | (0.361) |
| Regional effect | Controlled | Controlled | Controlled |
| Time effect | Controlled | Controlled | Controlled |
| Observations | 540 | 540 | 540 |
| R-squared | 0.770 | 0.782 | 0.770 |
| Number of id | 30 | 30 | 30 |

*Authors' calculation.*
*SE are in parentheses; ***p < 0.01, **p < 0.05.*

and almost all control variables remain the same, only with the coefficient values slightly changed. The above conclusion is consistent with our expectations.

The robustness test confirms a strong negative impact of corruption on EE in China.

# CONCLUSIONS AND POLICY RECOMMENDATIONS

## Conclusions

This study utilizes the corruption and EE data of 30 provinces in China from 2000 to 2017 and uses a fixed effects model to analyze the relationship between corruption and EE. The study indicates that an increase in corruption generally decreases EE, while a negative impact of corruption on EE varies in different regions. A negative impact was observed in both eastern and western China, but a negative influence in eastern China is significantly higher than that in western China. Furthermore, a threshold-effect model was constructed to analyze the impact of different corruption levels on EE. When the corruption is below a certain threshold (11.37 cases per 10,000 public officials), its impact on EE is relatively weak, whereas when the corruption degree exceeds a certain threshold, a significant hindrance to EE can be found. Most of the provinces in China are currently above the threshold level of corruption, so anti-corruption efforts should continue to be intensified to ensure a steady improvement in EE.

In addition, this study analyzes the relationship between corruption and EE from a perspective of energy regulations. Although the dispatched provincial-level energy regulatory agencies have strengthened regulations, a negative impact of corruption can still be observed. In central and western China, where corruption is more severe, corruption harms energy regulations and weakens a positive effect of regulations on EE. However, in eastern China, where the degree of corruption is relatively low, the establishment of regulatory branches can significantly inhibit the negative effects of corruption.

## Policy Recommendations

This study not only uncovers the correlation between corruption and EE statistically but also provides practical references and

policy suggestions for an anti-corruption work in the energy industry of China. The following policy suggestions are proposed:

First, corruption can significantly hinder the improvement of EE, so the relationship between corruption and EE should be properly studied. Although China's anti-corruption work has achieved some results in recent years, the degree of corruption in most of the provinces of China is still serious, and the role of anti-corruption in the long-term development of the energy industry should be viewed properly. The centralization of administrative power and the lack of transparency in the Chinese government make it difficult for external supervision agencies to play an effective role in anti-corruption, which encourages the continuous growth of corruption. Therefore, an anti-corruption work requires a systemic reform and innovation. The state should effectively play legislative and supervisory roles in the people's congress, legally fill the loopholes in the congress's ability and procedures for a review and supervision of the public budget, and improve the transparency of government finances. In addition, the state should make the effective use of think tanks, strengthen the research on key fiscal and taxation issues, and promote the pluralistic participation of society.

Second, based on the findings of the cases of corruption investigated in the energy industry, government officials and enterprise executives form a chain of interest transmission. On one hand, the fixed asset investment projects of energy are often worth hundreds of millions, which are huge and can easily induce corruption. On the other hand, the power of examination and approval for energy market access and energy fixed asset investment is highly centralized, and the examination and approval processes are not very transparent, which provide a space for officials to seek rent. Therefore, for weak links and key areas of corruption in the energy industry, the government should use advanced technologies such as cloud computing and blockchain to innovate the means of anti-corruption work (Abeyratne and Monfared, 2016). Employing advanced technologies that are unable to tamper with and forge data, monitoring agencies and the public can track and verify each data block (Pellegrini et al., 2020) to suppress corruption effectively and improve the transparency of government public services.

Third, government regulations can improve the efficiency of resource allocation and ensure the effective implementation of regulatory policies, and thus improving EE. However, the existence of provincial energy regulatory agencies may lead to excessive GV and reduce EE. At the same time, government corruption will also affect energy regulatory agencies and reduce EE. Considering the different levels of corruption and marketization in the eastern, central, and western regions, the government should set up regulatory agencies according to the local corruption level, giving priority to regions with less corruption, such as the eastern provinces. In addition, regulators of the energy industry should be effectively supervised. The government should design a diversified external supervision system for energy regulations in China from the four dimensions of legislation, administration, justice, and society. In addition, to reduce the corruption of government regulations, the government should also advocate a rational participation of social organizations and citizens, and form an internal and external supervision mechanism of energy regulations with complementary functions and an effective coordination. Meanwhile, the government should improve the marketization mechanism of EP (Liu et al., 2020), play an effective role in improving resource allocation and EE, reducing GV in the economy, and gradually changing the positioning of government functions from an interventional government to a service one.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

JL and JG: conceptualization and methodology. JL: data curation and software. JL and XB: formal analysis. XL and XB: investigation. XL: resources. FT-H: validation. JG and JL: writing—original draft. XL, FT-H, and JL: writing—review and editing. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by funding from the National Natural Science Foundation of China (Grant Nos. 17AZZ017 and 71742001), the Natural Science Foundation of Zhejiang Province, China (Grant Nos. LY19G030016 and LQ18G010006), and the Grant-in-Aid for Excellent Young Researcher of the Ministry of Education of Japan (MEXT), of the Humanities and Social Sciences Research of the Ministry of education of the People's Republic of China.

## REFERENCES

Abeyratne, S., and Monfared, R. (2016). Blockchain ready manufacturing supply chain using distributed ledger. *Int. J. Renew. Energy Technol.* 5, 1–10. doi: 10.15623/ijret.2016.0509001

Aguilera, R. F., and Ripple, R. D. (2012). Technological progress and the availability of European oil and gas resources. *Appl. Energy* 96, 387–392. doi: 10.1016/j.apenergy.2012.02.069

Aidt, T. S. (2016). Rent seeking and the economics of corruption. *Constitute. Pol. Econ.* 27, 142–157. doi: 10.1007/s10602-016-9215-9

Albino, V., Ardito, L., Dangelico, R. M., and Petruzzelli, A. M. (2014). Understanding the development trends of low-carbon energy technologies: a patent analysis. *Appl. Energy* 135, 836–854. doi: 10.1016/j.apenergy.2014.08.012

Antonietti, R., and Fontini, F. (2019). Does energy price affect energy efficiency? Cross-country panel evidence. *Energy Policy* 129, 896–906. doi: 10.1016/j.enpol.2019.02.069

Apergis, N., Dincer, O. C., and Payne, J. E. (2010). The relationship between corruption and income inequality in US states: evidence from a panel cointegration and error correction model. *Public Choice* 145, 125–135. doi: 10.1007/s11127-009-9557-1

Bian, Y., Song, K., and Bai, J. (2019). Market segmentation, resource misallocation and environmental pollution. *J. Clean. Prod.* 228, 376–387. doi: 10.1016/j.jclepro.2019.04.286

Borensztein, E., De Gregorio, J., and Lee, J. W. (1998). How does foreign direct investment affect economic growth? *J. Int. Econ.* 45, 115–135. doi: 10.1016/S0022-1996(97)00033-0

BP World Energy Statistical Yearbook (2019). *Beijing: China Statistics Press*. Beijing.

Chen, G., Li, S., and Yin, X. (2008). Corruption and China's economic growth: an empirical perspective. *Compar. Econ. Soc. Syst.* 03, 59–68. (in Chinese)

Cole, S., and Tran, A. (2011). 14 Evidence from the firm: a new approach to understanding corruption. *International handbook on the economics of corruption*, 408(7).

Dan, S. H. I. (2006). Regional differences in China's energy efficiency and conservation potentials. *China Industr. Econ.* 10, 49–58. (in Chinese) Data Available online at: http://www.ccdi.gov.cn/scdc/ (accessed December 25, 2019).

Del Monte, A., and Papagni, E. (2007). The determinants of corruption in Italy: regional panel data analysis. *Eur. J. Polit. Econ.* 23, 379–396. doi: 10.1016/j.ejpoleco.2006.03.004

Dong, B., and Torgler, B. (2013). Causes of corruption: evidence from China. *China Econ. Rev.* 26, 152–169. doi: 10.1016/j.chieco.2012.09.005

Dong, K., Sun, R., Hochman, G., and Li, H. (2018). Energy intensity and energy conservation potential in china: a regional comparison perspective. *Energy* 155, 782–795. doi: 10.1016/j.energy.2018.05.053

Eberhard, A. (2007). *Infrastructure Regulation in Developing Countries; An Exploration of Hybrid and Transitional Models*. Washington, DC: Public-Private Infrastructure Advisory Facility (PPIAF), World Bank.

Egger, P., and Winner, H. (2005). Evidence on corruption as an incentive for foreign direct investment. *Eur. J. Polit. Econ.* 21, 932–952. doi: 10.1016/j.ejpoleco.2005.01.002

Fan, C. S., and Grossman, H. I. (2001). Incentives and corruption in Chinese economic reform. *Econ. Policy Reform*. 4, 195–206. doi: 10.2139/ssrn.261701

Fisher-Vanden, K., Jefferson, G. H., Liu, H., and Tao, Q. (2004). What is driving China's decline in energy intensity? *Resour. Energy Econ.* 26, 77–97. doi: 10.1016/j.reseneeco.2003.07.002

Fisman, R., and Gatti, R. (2002). Decentralization and corruption: Cross-country and cross-state evidence. *J. Public Econ.* 83, 325–345. doi: 10.1016/S0047-2727(00)00158-4

George, E. G., and Tzeremes, N. G. (2013). Carbon dioxide emissions and governance:a nonparametric analysis for the G-20. *Energy Econ.* 40, 110–118. doi: 10.1016/j.eneco.2013.06.010

Halkos, G. E., and Tzeremes, N. G. (2013). Carbon dioxide emissions and governance: a nonparametric analysis for the G-20. *Energy Econ.* 40, 110–118.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *J. Econom.* 93, 345–368. doi: 10.1016/S0304-4076(99)00025-1

Huang, J., Du, D., and Tao, Q. (2017). An analysis of technological factors and energy intensity in china. *Energy Policy* 109, 1–9. doi: 10.1016/j.enpol.2017.06.048

Huntington, S. (1968). *Political Order in Changing Societies*. New Haven: Yale University Press.

Ju, K., Su, B., Zhou, D., and Wu, J. (2017). Does energy-price regulation benefit China's economy and environment? Evidence from energy-price distortions. *Energy Policy* 105, 108–119. doi: 10.1016/j.enpol.2017.02.031

Kaller, A., Bielen, S., and Marneffe, W. (2018). The impact of regulatory quality and corruption on residential electricity prices in the context of electricity market reforms. *Energy Policy* 123, 514–524. doi: 10.1016/j.enpol.2018.09.008

Kander, A., Warde, P., Henriques, S. T., Nielsen, H., Kulionis, V., and Hagen, S. (2017). International trade and energy intensity during European industrialization, 1870–1935. *Ecol. Econ.* 139, 33–44. doi: 10.1016/j.ecolecon.2017.03.042

Kaufmann, D., and Wei, S. (2000). "Does 'grease money' speed up the wheels of commerce?" in *National Bureau of Economic Research Working, Paper No. 7093* (Cambridge, MA). doi: 10.3386/w7093

Knack, S. (2006). *Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators*. Washington, DC: The World Bank. doi: 10.1596/1813-9450-3968

Leff, N. H. (1964). Economic development through bureaucratic corruption. *Am. Behav. Sci.* 8, 8–14. doi: 10.1177/000276426400800303

Leys, C. (1965). What is the problem about corruption? *Modern Afr. Stud.* 3, 215–230. doi: 10.1017/S0022278X00023636

Li, K., and Lin, B. (2014). The nonlinear impacts of industrial structure on China's energy intensity. *Energy* 69, 258–265. doi: 10.1016/j.energy.2014.02.106

Li, K., and Lin, B. (2017). An application of a double bootstrap to investigate the effects of technological progress on total-factor energy consumption performance in China. *Energy* 128, 575–585. doi: 10.1016/j.energy.2017.04.044

Lin, B., and Du, K. (2013). The energy effect of factor market distortion in China. *Econ. Res. J.* 9, 125–136. (in Chinese)

Lin, B., and Zhang, G. (2013). Estimates of electricity saving potential in Chinese nonferrous metals industry. *Energy Policy* 60, 558–568. doi: 10.1016/j.enpol.2013.05.051

Lin, B. Q., and Liu, H. X. (2015). Do energy and environment efficiency benefit from foreign trade?—the case of China's industrial sectors. *Econ. Res. J.* 9, 127–141. (in Chinese)

Liu, Q., Lu, R. S., and Ma, X. J. (2015). Corruption, financial resources and exports. *Rev. Int. Econ.* 235, 1023–1043. doi: 10.1111/roie.12194

Liu, W., Zhang, X., Wu, Y., and Feng, S. (2020). Economic analysis of renewable energy in the electricity marketization framework: a case study in Guangdong, China. *Front. Energy Res.* 8:98. doi: 10.3389/fenrg.2020.00098

Lovei, L., and McKechnie, A. (2000). *The Costs of Corruption for the Poor-The Energy Sector*. Washington, DC: World Bank Other Operational Studies 11437, the World Bank.

Ma, L. (2015). Corruption and FDI technology spillover – research on threshold effect based on provincial panel data. *Commer. Econ. Res.* 18, 115–117. (in Chinese)

Mauro, P. (1995). Corruption and growth. *Q. J. Econ.* 110, 681–712. doi: 10.2307/2946696

Méndez, F., and Sepúlveda, F. (2006). Corruption, growth and political regimes: cross country evidence. *Eur. J. Polit. Econ.* 22, 82–98. doi: 10.1016/j.ejpoleco.2005.04.005

Mo, P. H. (2001). Corruption and economic growth. *J. Comp. Econ.* 29, 66–79. doi: 10.1006/jcec.2000.1703

National Bureau of Statistics (2020). Since the establishment of the National Bureau Energy, 2 directors and 8 directors have been injured! No. 30 in Nogen system, No. 80 in Nogen business (Appendix details). Available online at: https://www.sohu.com/a/255202925_752796 (accessed December 15, 2019).

Nye, J. S. (1967). Corruption and political development: a cost-benefit analysis. *Am. Polit. Sci. Rev.* 61, 417–427. doi: 10.2307/1953254

Oikonomou, V., Jepma, C., Becchis, F., and Russolillo, D. (2008). White Certificates for energy efficiency improvement with energy taxes: a theoretical economic model. *Energy Econ.* 30, 3044–3062. doi: 10.1016/j.eneco.2008.04.005

Ouyang, X., and Lin, B. (2015). An analysis of the driving forces of energy-related carbon dioxide emissions in China's industrial sector. *Renew. Sust. Energ. Rev.* 45, 838–849. doi: 10.1016/j.rser.2015.02.030

Pailler,S. (2018). Re-election incentives and deforestation cycles in the Brazilian Amazon. *J. Environ. Econ. Manag.* 88, 345–365.doi10.1016/j.jeem.2018.01.008 doi: 10.1016/j.jeem.2018.01.008

Patterson, M. G. (1996). What is energy efficiency? Concepts, indicators and methodological issues. *Energy Policy.* 24, 377–390. doi: 10.1016/0301-4215(96)00017-1

Pellegrini, L., Campi, S., Locatelli, M., Pattini, G., Di Giuda, G. M., and Tagliabue, L. C. (2020). Digital transition and waste management in architecture, engineering, construction, and operations industry. *Front. Energy Res.* 8:576462. doi: 10.3389/fenrg.2020.576462

Proskuryakova, L., and Kovalev, A. (2015). Measuring energy efficiency: is energy intensity a good evidence base? *Appl. Energy.* 138, 450–459. doi: 10.1016/j.apenergy.2014.10.060

Restuccia, D., and Rogerson, R. (2017). The causes and costs of misallocation. *J. Econ. Perspect.* 31, 151–174. doi: 10.1257/jep.31.3.151

Rose-Ackerman, S., and Palifka, B. J. (2016). *Corruption and Government: Causes, Consequences, and Reform.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139962933

Roy, J., and Yasar, M. (2015). Energy efficiency and exporting: evidence from firm-level data. *Energy Econ.* 52, 127–135. doi: 10.1016/j.eneco.2015.09.013

Shi, B., and Shen, K. (2013). The government intervention, the economic agglomeration and the energy efficiency. *Management World.* 10, 6–18. (in Chinese).

Shibao, L., Xiao, B., Wei, L., and Ning, W. (2019). Impacts of climate change on water resources and grain production. *Technol. Forecast. Soc. Change* 143, 76–84. doi: 10.1016/j.techfore.2019.01.015

Shibao, L., Xiaoling, Z., Huarong, P., Martin, S., Xiao, B., and Zhihong, Z. (2021). The energy-food-water nexus: water footprint of Henan-Hubei-Hunan in China. *Renew. Susta. Energy Rev*. 135:110417. doi: 10.1016/j.rser.2020.110417

Shibao, L., Xiaoling, Z., and Yao, T. (2020). Evolutionary analysis on structural characteristics of water resource system in basins of Northern China. *Sus. Dev*. 2020, 1–13. doi: 10.1002/sd.2031

Stiglitz, J. (1998). Distinguished lecture on economics in government: the private uses of public interests; incentives and institutions. *J. Econ. Perspect* 12, 3–22. doi: 10.1257/jep.12.2.3

Tan, R. (2014). *Ten-Year Overview of the State Electricity Regulatory Commission*. Beijing: China Electric Power Press, 279–323. (in Chinese).

Vial, V., and Hanoteau, J. (2010). Corruption, manufacturing plant growth, and the Asian paradox: Indonesian evidence. *World Dev*. 38, 693–705. doi: 10.1016/j.worlddev.2009.11.022

Voigt, S., De Cian, E., Schymura, M., and Verdolini, E. (2014). Energy intensity developments in 40 major economies: structural change or technology improvement? *Energy Econ*. 41, 47–62. doi: 10.1016/j.eneco.2013.10.015

Wang, J. (2018). *Research on China's Modern Energy Regulatory System and Regulatory Efficiency*. Beijing: China Social Sciences Press. (in Chinese).

Wang, J., and Gao, M.-H. (2017). Anticorruption, entrepreneurial competencies and enterprise innovation. *Busin. Manag. J*. 39, 36–52. (in Chinese).

Wang, S., Zhao, D., and Chen, H. (2020). Government corruption, resource misallocation, and ecological efficiency. *Energy Econ*. 85:104522. doi: 10.1016/j.eneco.2019.104573

Wang, X., and Wang, K.-Q. (2016). Impact of corruption on banks' non-performing loans – based on chinese provinical dynamic panel GMM method. *Collected Essays Fin. Econ.* 08, 36–43. (in Chinese).

Wang, Y., and You, J. (2012). Corruption and firm growth: evidence from China. *China Econ. Rev*. 23, 415–433. doi: 10.1016/j.chieco.2012.03.003

Wei, C., and Li, C. Z. (2017). Resource misallocation in Chinese manufacturing enterprises: evidence from firm-level data. *J. Clean. Prod*. 142, 837–845. doi: 10.1016/j.jclepro.2016.04.083

Wu, L. (2009). The Fluctuations of China' s energy efficiency: theoretical explains, numerical simulations and policy experiments. *Econ. Res.* 5, 130–142. (in Chinese)

Wu, Y. (2008). Fiscal decentralization, corruption and governance. *China Econ. Q.* 7, 1045–1060. (in Chinese).

Wu, Y., and Rui, M. (2010). Regional corruption, marketization and China's economic growth. *Manag. World* 11, 10–17. (in Chinese).

Yang, M. A., Yang, F. X., and Sun, C. W. (2018). Factor market distortion correction, resource reallocation and potential productivity gains: an empirical study on China's heavy industry sector. *Energ. Econ*. 69, 270–279. doi: 10.1016/j.eneco.2017.11.021

Yang, Z. (2010). Study on influence factors of china manufacturing energy efficiency: business circle and heavy industry. *Statist. Res.* 10, 33–39. (in Chinese).

Yongzheng, L., and Haibo, F. (2011). Corruption, public expenditure efficiency and economic growth. *Econ. Res. J.* 9, 17–28. (in Chinese).

Zhao, N., Xia, T., Yu, T., and Liu, C. (2020). Subsidy-related deception behavior in energy-saving products based on game theory. *Front. Energy Res*. 7:154. doi: 10.3389/fenrg.2019.00154

Zhu, L. (2016). *The influence of corruption and market of TFP on Chinese provinces – an empirical analysis based on 30 provincial panel data of china during 2001-2011* (Doctoral dissertation). Dongbei University Of Finance And Economics, China (in Chinese).

Zhu, W., Zhang, Z., and Li, X. (2019). Assessing theeffects of technological progress on energy efficiency in the construction industry: a case of China. *J. Cleaner Product.* 238, 1–12. (in Chinese). doi: 10.1016/j.jclepro.2019.117908

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Liu, Guo, Liu, Bai and Taghizadeh-Hesary. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 03 May 2021
doi: 10.3389/fenvs.2021.655080

# How to Promote Compliance Management in the Electricity Market? An Analysis Based on the Evolutionary Game Model

*Luosong Jin*[1], *Cheng Chen*[1], *Yun Li*[2], *Xiangyang Wang*[1] *and Yuanyuan Cheng*[3]*

[1] Zhejiang Electricity Trading Center Co., Ltd., Hangzhou, China, [2] Zhejiang Huayun Mdt Info Tech Ltd., Hangzhou, China, [3] China Center for Energy Economics Research, School of Economics, Xiamen University, Xiamen, China

In this paper, we aim to analyze how to effectively promote compliance management in the electricity market. We construct an evolutionary game model under the two different scenarios, i.e., the scenario without governmental supervision and the scenario with governmental supervision, and explicitly describes the strategic behaviors and dynamic evolution process of power enterprises and regulators in the power market. According to the results of the evolutionary stable strategy, we find that, in the absence of governmental supervision, the long-term stable equilibrium of power enterprises' choice of strategy "Compliance" and regulatory agencies' choice of strategy "Not bribery" is hard to be realized. Only if the government effectively supervises the compliance management of the electricity market can the ideal compliance behavior of the two stakeholders be achieved. Furthermore, we conduct a simulation study to analyze the impacts of the various model parameters on the dynamic evolution process. The specific results show that the lower compliance cost, the higher psychological cost, and the larger profit loss of the power enterprises, as well as the lower inspection cost, the higher psychological cost, and the larger rewards of the regulatory agencies, can promote the formation of compliance management. Besides, the larger penalty charged by the government is also conducive to the compliance management of the electricity market.

**Keywords: compliance management, electricity market, evolutionary game model, governmental supervision, strategy behavior**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Chang Liu,*
*Southwestern University of Finance and Economics, China*
*Jianghua Liu,*
*Shanghai University of Finance and Economics, China*

***Correspondence:**
*Yuanyuan Cheng*
*yycheng_xmu@sina.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *18 January 2021*
***Accepted:*** *23 February 2021*
***Published:*** *03 May 2021*

***Citation:***
*Jin L, Chen C, Li Y, Wang X and Cheng Y (2021) How to Promote Compliance Management in the Electricity Market? An Analysis Based on the Evolutionary Game Model. Front. Environ. Sci. 9:655080. doi: 10.3389/fenvs.2021.655080*

## INTRODUCTION

The electricity market is undergoing rapid and dramatic changes (Höhne and Tiberius, 2020). With the emergence of new business modes, the trading modes of the electricity market have been increasingly complex. Moreover, the rapid development and wide application of big data technology and machine learning in China have also brought opportunities and challenges to the power industry (Song et al., 2019). Besides, the electricity market is facing compliance risks brought about by the potential non-standard business operation of enterprises, which leads to additional difficulties and challenges in the management and development of the electricity market. In 2015, the Chinese government pointed out that "it is necessary to ameliorate power regulatory institutions, measures and methods, improve governmental regulatory means, raise the level of scientific supervision over technology, safety, transactions and operations, and further achieve

scientific supervision of power system" in the document named "Several Opinions on Further Deepening the Reform of the Power System." Moreover, to better promote corporate compliance management, the Chinese government has successively promulgated a series of policies such as "Guidelines for the Compliance Management System" and the "Guidelines for the Compliance Management of Central Enterprises" to comprehensively improve the level of compliance for enterprises, including power enterprises.

The requirements of compliance management include three aspects, namely national laws, corporate rules, and professional ethics. Through continuous improvement of market-related laws and rules, the government can discover the market entities' violations of trading rules and regulations timely, thereby safeguarding the overall interests of stakeholders, reducing market risks, and ensuring the sustainable development of the electricity market. The development of compliance management in the electricity market is carried out in conformity with the requirement of the Chinese reform in the field of the electricity market, and catering to the future development demands of the electricity market. Moreover, compliance management is an important guarantee for effectively preventing risks in the electricity market and is of great significance to the business expansion of the electricity market. However, the compliance management of China's electricity market is still in its infancy. At present, there is still a lack of systematic policies to ensure the effective implementation of compliance management in the electricity market. Only by building a complete compliance management system can the electricity market provide better services for society and provide solid and effective support for economic and social development. Therefore, it is very important and necessary to study the compliance management of the electricity market.

Many previous studies have discussed the possible risks in the electricity market from different aspects. Vehviläinen and Keppo (2003) apply the method of financial risk management to the price risk in the electricity market and prove that if the properties of the electricity market are fully considered, the financial method could also be useful for the risk management of the electricity market. Similarly, Deng and Oren (2006) review the applications of various financial derivatives in the electricity market, i.e., electricity derivatives, and analyze the effects of these derivatives in reducing market risks and optimizing hedging strategies. Taking CVaR as the risk measure, Kharrati et al. (2016) construct a model for the decision-making problem of the retailers in the competitive retail power market. The numerical results show that the decisions of rivals will affect the strategies selected by the retailers. Besides, based on the hourly electricity prices of Nord Pool, Uribe et al. (2020) find that the consumers and producers in the power market have asymmetric risks. The tail distribution of the electricity price plays an important role in the risk exposure. Coulon et al. (2013) construct a model for load and price dynamics in the electricity market which aims to reduce the risk of hedging load and price and further find out the optimal hedging strategies. Meunier (2013) tries to analyze the impacts of the risk aversion on the technology mixes of firms in the electricity market and figure out how the uncertainty of fossil fuel prices affects the investment decisions of power producers. The risks that exist in renewable power are also studied. Bahrami and Amini (2018) propose a decentralized energy trading model to reduce the generation risks in the renewable resources. At the same time, the goal of minimizing the cost of load aggregators and maximizing the profit of the generators could be also achieved. Taking wind power as an example, Klessmann et al. (2008) discuss the risks of integrating renewable energy generators into the electricity market in Germany, Spain, and the United Kingdom. The results show that when the market risk is higher, the development of renewable energy needs more financial support. However, higher market risks may also limit indirect costs to society. Aquila et al. (2017) construct a model to analyze the investment decision of wind power generators under different risks in a regulated electricity market and use the method of VaR to make the risk management.

However, the researches on the compliance management risks of the electricity market are limited. Therefore, this paper aims to study the strategic behavior of different stakeholders in the electricity market based on game theory, so as to specifically analyze the influence mechanism of various factors on the compliance management of the electricity market and effectively prevent the compliance management risks in the electricity market.

Compared with the static game, the dynamic game is more suitable to analyze the relationship between different stakeholders, because players' decision-making behavior will change with other participants' strategy to maximize the utility. Therefore, we may prefer the use of the dynamic game model when we want to study sensible countermeasures or propose several policy recommendations (Xie et al., 2017; Jiang et al., 2020). The evolutionary game model, as an important part of the dynamic game, has been widely applied in the analysis of the interactive strategies between different stakeholders. For example, Jiang et al. (2019) analyze the effects of implementing environmental regulation under fiscal decentralization in China; the strategies of three players, i.e., central government, local government, and polluting enterprises, are discussed through the evolutionary game model. Similarly, Sheng et al. (2020) explore the strategy choices of the central government, local government, and firms. The results illustrate that government supervision is important for the implementation of environmental regulation policies. da Silva Rocha and Salomão (2019) analyze the interplay between firms and auditors under the framework of the evolutionary game model as well as well-mixed and structured populations. The results show that high inspecting cost is not conducive to the auditing process. The study of Sun and Zhang (2019) discuss the problem of greenwashing and the role of governmental regulation. Two kinds of firms (dominant and inferior firms) are regarded as the players in the evolutionary game model. The results show that governmental punishment can inhibit the greenwashing behavior of both two kinds of firms, but the governmental tax subsidy can only curb the greenwashing behavior of dominant firms. Zhu et al. (2020) study the renewable portfolio standard through the evolutionary game model. The strategy selection of two different kinds of power sales enterprises and the regulators are analyzed. The results

show that tradable green certificate prices, net profit, reward, and punishment are important factors to make different kinds of power sales enterprises comply with quota obligations in the electricity market. Ji et al. (2015) discuss green purchasing in the supply chain based on the evolutionary game model. The cooperation tendency of two players, i.e., suppliers and manufacturers, is studied in detail. Furthermore, some policy suggestions on establishing the green supply chain are obtained. Tian et al. (2014) analyze how to effectively promote green supply chain management from the perspective of subsidy and use the evolutionary game model to study the strategies of three different stakeholders in the automotive manufacturing industry. Based on the evolutionary game model, the work of Gu et al. (2019) mainly focuses on reverse logistics. The interactive strategy between the natural resource-intensive and energy-intensive enterprises is analyzed to discuss how to achieve sustainable development in these industries.

The main contributions of this paper are three-fold. Firstly, in this paper, we construct an evolutionary game model to analyze the dynamic evolution process of stakeholders' strategies in the electricity market. Secondly, we compare and analyze the evolutionary stability strategy under two different scenarios, namely, the scenario without governmental supervision and the scenario with governmental supervision. Moreover, we analyze the specific impacts of different model parameter changes on the realization of compliance management in the electricity market. Thirdly, based on the model results, we further put forward relevant policy recommendations for the promotion of compliance management in the electricity market.

The structure of this paper is as follows. In section Construction of the Evolutionary Game Model, we discuss the evolutionary game model and the specific details. We conduct the simulation study in section Simulation Study. Section Further Discussion is about the discussion. Section Conclusion and Policy Suggestion concludes and proposes relevant policy recommendations.

# CONSTRUCTION OF THE EVOLUTIONARY GAME MODEL

## Model Description

To study the compliance management in the electricity market, we mainly focus on the strategic behavior of two stakeholders, namely, the power enterprises and regulatory agencies. Hence, the strategy choices and payoffs of these two players in the electricity market need to be fully considered when constructing the evolutionary game model. Moreover, the policy implementation of the central government may also affect the strategy selections and payoffs of the power enterprises and regulatory agencies. For instance, when the central government decides to carry out strict supervision over the compliance management in the electricity market, it will promote the regulatory agencies to strengthen the enforcement of compliance management, thereby forcing the power enterprises to choose compliance in the operation of the business with higher probability. On the contrary, if the central government decides not to conduct strict supervision, the regulatory agencies may not effectively perform their responsibilities, giving enterprises incentives to be non-compliant. Therefore, this paper will analyze the dynamic decision-making process between the two stakeholders in two scenarios: the existence and absence of governmental supervision.

The development of compliance management in China's electricity market is still in its infancy. This paper specifically studies how to effectively control compliance risks from the perspective of evolutionary game theory. The two assumptions, completely rational and complete information in the traditional game theory, are released in the evolutionary game theory. In the evolutionary game theory, decision-makers are assumed to be bounded rationality, which means that decision-makers do not always make decisions to maximize their utility, and their strategic choices may be affected by the environment, i.e., the actions of other players. In this paper, we follow the evolutionary game theory and make the assumptions that there exist many power enterprises and regulatory agencies in the electricity market, and in the process of decision-making, they will learn from others and adjust their choices to obtain a larger payoff.

### Power Enterprises

The power enterprises have two possible strategies {Compliance, Not compliance}. We use $x$ to denote the probability that the power enterprises select the strategy of "Compliance." When the power enterprises choose the strategy of "Compliance," it means that the power enterprises will make more effort to meet the external (including national and industrial levels) and internal (corporate level) requirements of various laws and regulations in the electricity market. We use $c$ to represent the costs of compliance effort. When the power enterprises choose the strategy of "Not Compliance," the compliance costs $c$ can be omitted. However, the power enterprises will bear psychological costs $p_e$ for fear of being discovered. At the same time, the power enterprises tend to collude with regulatory agencies to avoid exposure of their illegal actions. We use $b$ to represent the bribery costs of the power enterprises, which can also be considered as the amount of money that regulatory agencies earn from the power enterprises. If the power enterprises and the regulatory agencies reach a collusive agreement, both of them will face fines equal to $f$ once the central government finds out. However, if the power enterprises adopt unilateral bribery, they will receive larger fines $\lambda f$ ($\lambda > 1$) charged by the central government. Moreover, if the regulatory agencies discover the non-compliance behavior of the power enterprises, the power enterprises' reputation will be damaged. Thence, we assume that the profits of power enterprises will reduce $r$ due to the damaged reputation and the fierce competition in the electricity market.

### Regulatory Agencies

The regulatory agencies also have two possible strategies {Not bribery, Bribery}. $y$ denotes the probability that the regulatory agencies select the strategy of "Not bribery." The strategy of "Not bribery" means that the regulatory agencies will carry out the strict inspection of the behavior of the power enterprises. Moreover, we use $s$ to represent the costs of inspection effort.

If the non-compliant behavior of the power enterprises was detected, the regulatory agencies will receive a certain reward $e$. Conversely, if the regulatory agencies choose the strategy of "Bribery," they will not scrutinize the behavior of the power enterprises, and the inspection costs $s$ could be ignored. However, the regulatory agencies will bear psychological costs $p_r$ for fear of being discovered by the central government. Moreover, they can receive money equals to $b$ from enterprises if they conspire. However, if the central government executes supervision over the compliance management in the electricity market, the bribery behavior will incur fines $f$ charged by the central government. Similar to the power enterprises, the unilateral bribery of the regulatory agencies will be severely punished with fines equals to $\lambda f$. The notations of all variables used in the evolutionary game model are shown in **Table 1**.

## Payoff Matrix

### Scenario Without Governmental Supervision

In the scenario without governmental supervision, the power enterprises will get $-c$ and the regulatory agencies will get $-s$ in the ideal state, namely, the regulatory agencies will earnestly perform their regulatory responsibilities, and the power enterprises will also act in compliance. If the power enterprises choose the strategy "Not compliance" and the regulatory agencies choose the strategy "Not bribery," the non-compliance behaviors of power companies will be discovered by regulatory agencies. Hence, the power enterprises will get $-p_e - r$ and the regulatory agencies will get $-s + e$. If the power enterprises choose the strategy "Compliance" and the regulatory agencies choose the strategy "Bribery," the power enterprises will get $-c$ and the regulatory agencies will get $-p_r$. If these two stakeholders collude, the power enterprises will get $-p_e - b$ and the regulatory agencies will get $-p_r + b$. The payoff matrix in the scenario without governmental supervision is shown in **Table 2**.

**TABLE 1 |** Notation.

| Parameters | Descriptions | Range |
|---|---|---|
| $c$ | Compliance costs of the enterprises | $c > 0$ |
| $p_e$ | Psychological costs of the power enterprises | $0 < p_e < c$ |
| $b$ | Bribery costs of the power enterprises | $0 < b < c$ |
| $r$ | Profit loss of the power enterprises | $r > 0$ |
| $s$ | Inspection costs of the regulatory agencies | $s > 0$ |
| $p_r$ | Psychological costs of the regulatory agencies | $0 < p_r < s$ |
| $e$ | Reward of detecting the non-compliant behavior for the regulatory agencies | $e > 0$ |
| $f$ | Fines charged by the central government | $f > 0$ |
| $\lambda$ | Coefficient of fines in the unilateral bribery | $\lambda > 1$ |
| $x$ | Probability that the enterprises choose to "Compliance" | $0 \leq x \leq 1$ |
| $y$ | Probability that the regulatory agencies choose to "Not bribery" | $0 \leq y \leq 1$ |

### Scenario With Governmental Supervision

In the scenario with governmental supervision, the payoff matrix is a bit more complicated because we should take into account the fines charged by the central government. In the ideal state, the power enterprises will get $-c$ and the regulatory agencies will get $-s$, which is exactly the same as in the scenario without governmental supervision. If the power enterprises choose the strategy "Not compliance" and the regulatory agencies choose the strategy "Not bribery," the non-compliance behaviors of power companies will be discovered by regulatory agencies. Besides, the central government will also punish the non-compliance behaviors of power companies. Hence, the power enterprises will get $-p_e - \lambda f - r$ and the regulatory agencies will get $-s + e$. If the power enterprises choose the strategy "Compliance" and the regulatory agencies choose the strategy "Bribery," the central government will punish the regulatory agencies. So the power enterprises will get $-c$ and the regulatory agencies will get $-p_r - \lambda f$. If the collusion occurs, the power enterprises will get $-p_e - b - f$ and the regulatory agencies will get $-p_r + b - f$. The payoff matrix in the scenario with governmental supervision is shown in **Table 3**.

## Replicator Dynamic System

### Scenario Without Governmental Supervision

In the evolutionary game theory, if the payoffs of some players are less than the overall average payoff, then these players will change their strategies. Moreover, if the payoffs of some players are higher than the average, then in the population, the probability of choosing the same strategy will increase (Taylor and Jonker, 1978; Friedman, 1991). Therefore, based on the replicator dynamic system, it is possible to describe the dynamic changes in the probability of players choosing different strategies.

**TABLE 2 |** Payoff matrix of stakeholders without governmental supervision.

| | | Regulatory agencies | |
|---|---|---|---|
| | | **Not bribery $y$** | **Bribery $1-y$** |
| Enterprises | Compliance $x$ | $-c, -s$ | $-c, -p_r$ |
| | Not compliance $1-x$ | $-p_e - r, -s + e$ | $-p_e - b, -p_r + b$ |

**TABLE 3 |** Payoff matrix of stakeholders with governmental supervision.

| | | Regulatory agencies | |
|---|---|---|---|
| | | **Not bribery $y$** | **Bribery $1-y$** |
| Enterprises | Compliance $x$ | $-c, -s$ | $-c, -p_r - \lambda f$ |
| | Not compliance $1-x$ | $-p_e - \lambda f - r$, $-s + e$ | $-p_e - b - f$, $-p_r + b - f$ |

Firstly, we will calculate the expected payoffs under the different strategic behaviors of the power enterprises and regulatory agencies. Using $\pi_C^E$ and $\pi_{NC}^E$ to represent the strategies of "Compliance" and "Not compliance" of the power enterprises. Then, we have:

$$\pi_C^E = y(-c) + (1-y)(-c) \quad (1)$$
$$\pi_{NC}^E = y(-p_e - r) + (1-y)(-p_e - b) \quad (2)$$

Then, we can obtain the expected payoff of the power enterprises, which is shown in Equation (3).

$$\bar{\pi}^E = x\pi_C^E + (1-x)\pi_{NC}^E \quad (3)$$

Therefore, the replicator dynamic function of the strategy "Compliance" can be seen in Equation (4).

$$F(x) = \frac{dx}{dt} = x\left(\pi_C^E - \bar{\pi}^E\right) = x(1-x)(yr - c + p_e + b - by) \quad (4)$$

Similarly, we use the variables $\pi_{NB}^R$ and $\pi_B^R$ to represent the expected payoffs of the regulatory agencies for choosing the "Not bribery" and "Bribery." Then, we have:

$$\pi_{NB}^R = x(-s) + (1-x)(-s+e) \quad (5)$$
$$\pi_B^R = x(-p_r) + (1-x)(-p_r + b) \quad (6)$$

Then, the expected payoff of the regulatory agencies is shown in Equation (7).

$$\bar{\pi}^R = y\pi_{NB}^R + (1-y)\pi_B^R \quad (7)$$

Therefore, the replicator dynamic function of the strategy "Not bribery" is shown in Equation (8).

$$F(y) = \frac{dy}{dt} = y\left(\pi_{NB}^R - \bar{\pi}^R\right) = y(1-y)(-s+e-ex+p_r-b+bx) \quad (8)$$

### Scenario With Governmental Supervision

In the same way, we can get the expected payoffs of the power enterprises and regulatory agencies under the situation of governmental supervision. When the central government decides to take some measures to promote compliance management in the electricity market, the government will intensify the supervision on the behavior of power enterprises and regulatory agencies. The expected payoffs under the two different strategies of the power enterprises are:

$$\pi_C^E = y(-c) + (1-y)(-c) \quad (9)$$
$$\pi_{NC}^E = y(-p_e - \lambda f - r) + (1-y)(-p_e - b - f) \quad (10)$$

Then, the expected payoff of the enterprises is shown in Equation (11).

$$\bar{\pi}^E = x\pi_C^E + (1-x)\pi_{NC}^E \quad (11)$$

Therefore, the replicator dynamic function of the strategy "Compliance" is Equation (12).

$$F(x) = \frac{dx}{dt} = x\left(\pi_C^E - \bar{\pi}^E\right) = x(1-x)(\lambda fy + yr - c + p_e + b - by + f - fy) \quad (12)$$

Similarly, the expected payoffs under the two strategies of the regulatory agencies are:

$$\pi_{NB}^R = x(-s) + (1-x)(-s+e) \quad (13)$$
$$\pi_B^R = x(-p_r - \lambda f) + (1-x)(-p_r + b - f) \quad (14)$$

Then, the expected payoff of the regulatory agencies is shown in Equation (15).

$$\bar{\pi}^R = y\pi_{NB}^R + (1-y)\pi_B^R \quad (15)$$

Therefore, the replicator dynamic function of the strategy "Not bribery" is shown in Equation (16).

$$F(y) = \frac{dy}{dt} = y\left(\pi_{NB}^R - \bar{\pi}^R\right) = y(1-y)(\lambda fx - s + e - ex + p_r - b + bx + f - fx) \quad (16)$$

## Evolutionary Stable Strategy

### Scenario Without Governmental Supervision

Based on the replicator dynamic functions obtained in Equations (4) and (8), we can get the equilibrium points $(x, y) \in [0,1] \times [0,1]$ that satisfy the requirement that the replicator dynamic function should equal to zero due to the evolutionary game theory. Therefore, we have Proposition 1, as shown below.

**Proposition 1.** The replicator dynamic system in the scenario without governmental supervision has five equilibrium points, i.e., (0, 0), (0, 1), (1, 0), (1, 1), and $(x^*, y^*)$, where $x^* = \frac{s-e-p_r+b}{b-e}$, and $y^* = \frac{c-p_e-b}{r-b}$, if $0 \le \frac{s-e-p_r+b}{b-e} \le 1$ and $0 \le \frac{c-p_e-b}{r-b} \le 1$.

**Proof:**

(i) Obviously, the four fixed points (0, 0), (0, 1), (1, 0), and (1, 1) can satisfy the equilibrium conditions.

(ii) When we solve Equation (17) which is shown as follows:

$$\begin{cases} yr - c + p_e + b - by = 0 \\ -s + e - ex + p_r - b + bx = 0 \end{cases} \quad (17)$$

We can get the solution specified in Equation (18).

$$\begin{cases} x^* = \frac{s-e-p_r+b}{b-e} \\ y^* = \frac{c-p_e-b}{r-b} \end{cases} \quad (18)$$

Moreover, the conditions $0 \le \frac{s-e-p_r+b}{b-e} \le 1$ and $0 \le \frac{c-p_e-b}{r-b} \le 1$ should be satisfied to ensure $(x^*, y^*) \in [0,1] \times [0,1]$.

Next, we calculate the Jacobian matrix of the replicator dynamic system to obtain the evolutionary stability strategy (ESS). The Jacobian matrix is specified in Equation (19), and we can further obtain Proposition 2.

$$J_1 = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} \end{bmatrix} = \begin{bmatrix} (1-2x)(yr - c + p_e + b - by) & x(1-x)(r-b) \\ y(1-y)(b-e) & (1-2y)(e-s-ex+p_r-b+bx) \end{bmatrix} \tag{19}$$

**Proposition 2.**

(i) The equilibrium point (0, 0) is an ESS if $p_e + b < c$ and $e + p_r < s + b$.

(ii) The equilibrium point (0, 1) is an ESS if $r + p_e < c$ and $s + b < e + p_r$.

(iii) The equilibrium point (1, 0) is an ESS if $c < p_e + b$ and $p_r < s$.

(iv) The equilibrium point (1, 1) is an ESS if $c < r + p_e$ and $s < p_r$.

(v) The equilibrium point ($x^*$, $y^*$) is not a stable equilibrium point.

**Proof:**

(i) The Jacobian matrix at the equilibrium point (0, 0) is shown as follows:

$$J_1 = \begin{bmatrix} -c + p_e + b & 0 \\ 0 & e - s + p_r - b \end{bmatrix} \tag{20}$$

When $p_e + b < c$ and $e + p_r < s + b$, the two eigenvalues ($\lambda_1 = -c + p_e + b$, $\lambda_2 = e - s + p_r - b$) of the Jacobian

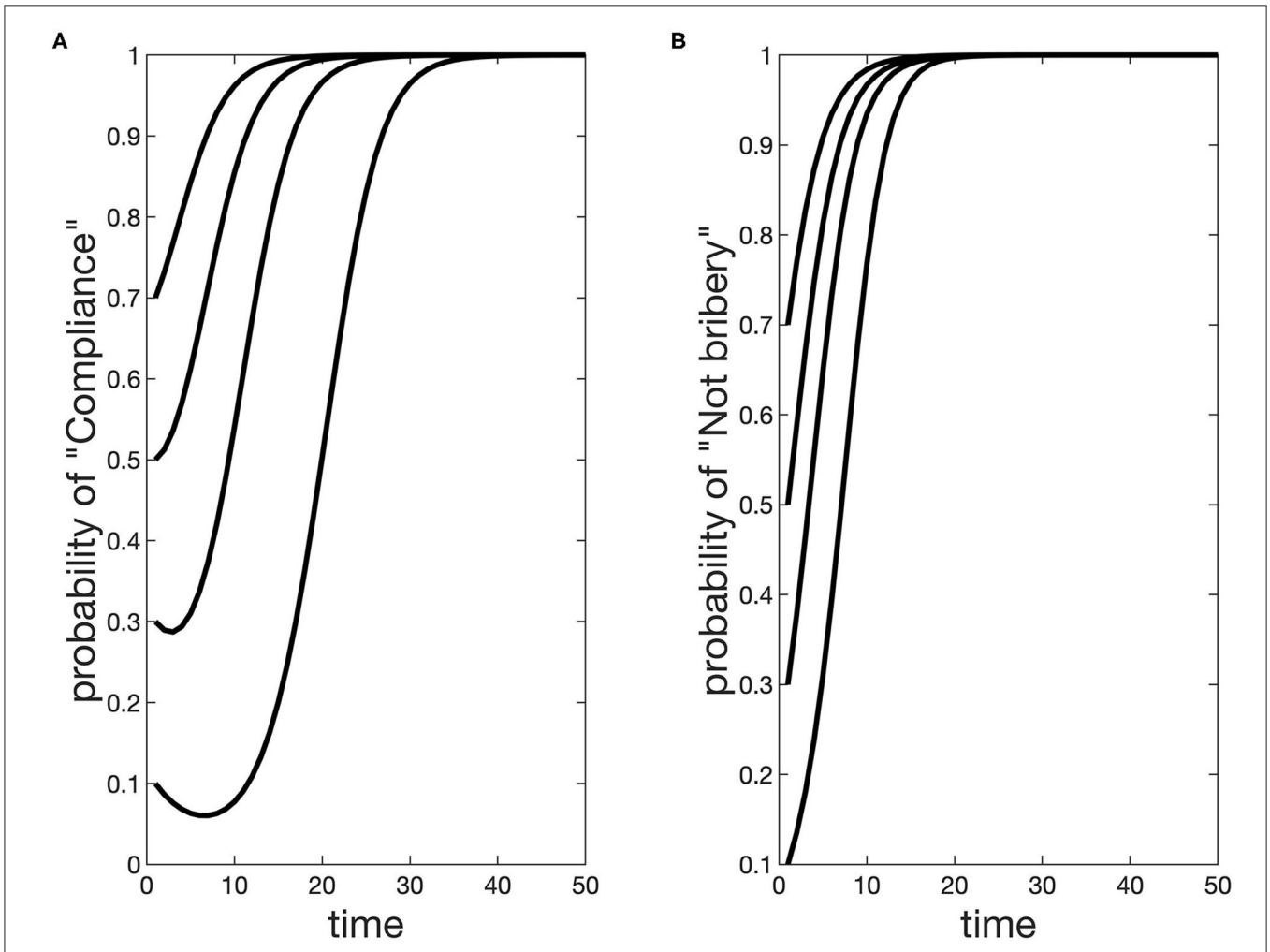



**FIGURE 1 |** Evolutionary process of the two stakeholders' strategies. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

**TABLE 4 |** The initial values of the parameters.

| Parameters | Descriptions | Values |
|---|---|---|
| $c$ | Compliance costs of the enterprises | 1 |
| $p_e$ | Psychological costs of the enterprises | 0.2 |
| $b$ | Bribery costs of the enterprises | 0.2 |
| $r$ | Profit loss of the enterprises | 0.5 |
| $s$ | Inspection costs of the regulatory agencies | 0.5 |
| $p_r$ | Psychological costs of the regulatory agencies | 0.2 |
| $e$ | Reward of detecting the non-compliant behavior of the enterprises | 0.5 |
| $f$ | Fines charged by the central government | 0.4 |
| $\lambda$ | Coefficient of fine in the unilateral bribery | 1.5 |
| $x$ | Probability that the enterprises choose to "Compliance" | 0.1/0.3/0.5/0.7 |
| $y$ | Probability that the regulatory agencies choose to "Not bribery" | 0.1/0.3/0.5/0.7 |

matrix $J_1$ are both negative. Also, we have $\det J_1 > 0$ and $\mathrm{tr} J_1 < 0$. Therefore, the equilibrium point (0, 0) is an ESS if $p_e + b < c$ and $e + p_r < s + b$.

(ii) The Jacobian matrix at the equilibrium point (0, 1) is shown as follows:

$$J_1 = \begin{bmatrix} r - c + p_e & 0 \\ 0 & s - e + b - p_r \end{bmatrix} \tag{21}$$

When $r + p_e < c$ and $s + b < e + p_r$, the two eigenvalues ($\lambda_1 = r - c + p_e$, $\lambda_2 = s - e + b - p_r$) of the Jacobian matrix $J_1$ are both negative. Therefore, the equilibrium point (0, 1) is an ESS if $r + p_e < c$ and $s + b < e + p_r$.

(iii) The Jacobian matrix at the equilibrium point (1, 0) is shown as follows:

$$J_1 = \begin{bmatrix} c - p_e - b & 0 \\ 0 & p_r - s \end{bmatrix} \tag{22}$$

When $c < p_e + b$ and $p_r < s$, the two eigenvalues ($\lambda_1 = c - p_e - b$, $\lambda_2 = p_r - s$) of the Jacobian matrix $J_1$ are both negative. Therefore, the equilibrium point (1, 0) is an ESS if $c < p_e + b$ and $p_r < s$.

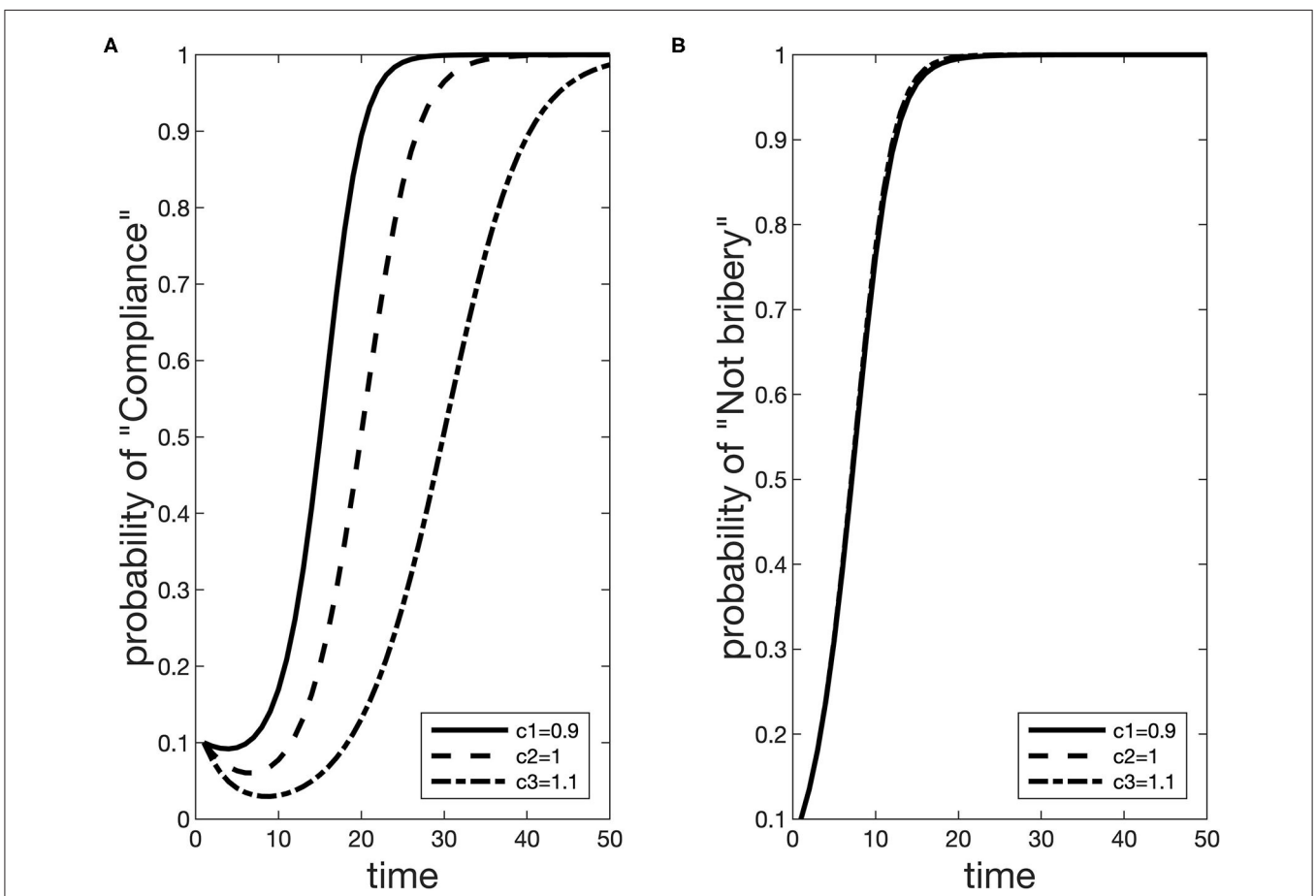



**FIGURE 2 |** Evolutionary process of the two stakeholders' strategies with different values of $c$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

(iv) The Jacobian matrix at the equilibrium point (1, 1) is shown as follows:

$$J_1 = \begin{bmatrix} c - r - p_e & 0 \\ 0 & s - p_r \end{bmatrix} \tag{23}$$

When $c < r + p_e$ and $s < p_r$, the two eigenvalues ($\lambda_1 = c - r - p_e$, $\lambda_2 = s - p_r$) of the Jacobian matrix $J_1$ are both negative. Therefore, the equilibrium point (1, 1) is an ESS if $c < r + p_e$ and $s < p_r$.

(v) In the evolutionary game model, an equilibrium point is stable when, and only when, it is a pure-strategy Nash equilibrium (Selten, 1980; Ritzberger and Weibull, 1995). Because the equilibrium point ($x^*$, $y^*$) is a hybrid-strategy Nash equilibrium, it is not a stable equilibrium point (Sheng et al., 2020).

## Scenario With Governmental Supervision

Similarly, we could obtain the equilibrium points under the scenario with governmental supervision. Furthermore, we propose Proposition 3 which is shown as follows.

**Proposition 3.** The replicator dynamic system in the scenario with governmental supervision has five equilibrium points, i.e., (0, 0), (0, 1), (1, 0), (1, 1) and ($x^*$, $y^*$), where $x^* = \frac{s-e-p_r+b-f}{\lambda f-e+b-f}$, and $y^* = \frac{c-p_e-b-f}{\lambda f+r-b-f}$, if $0 \leq \frac{s-e-p_r+b-f}{\lambda f-e+b-f} \leq 1$ and $0 \leq \frac{c-p_e-b-f}{\lambda f+r-b-f} \leq 1$.

**Proof:**

(i) Obviously, the four fixed points (0, 0), (0, 1), (1, 0), and (1, 1) can satisfy the equilibrium conditions.

(ii) When we solve Equation (24):

$$\begin{cases} \lambda fy + yr - c + p_e + b - by + f - fy = 0 \\ \lambda fx - s + e - ex + p_r - b + bx + f - fx = 0 \end{cases} \tag{24}$$

We have the solution which is shown in Equation (25).

$$\begin{cases} x^* = \frac{s-e-p_r+b-f}{\lambda f-e+b-f} \\ y^* = \frac{c-p_e-b-f}{\lambda f+r-b-f} \end{cases} \tag{25}$$

Moreover, the conditions $0 \leq \frac{s-e-p_r+b-f}{\lambda f-e+b-f} \leq 1$ and $0 \leq \frac{c-p_e-b-f}{\lambda f+r-b-f} \leq 1$ should be satisfied to ensure $(x^*, y^*) \in [0, 1] \times [0, 1]$.

Then, we calculate the Jacobian matrix of the replicator dynamic system to obtain the ESS in the scenario with governmental supervision. The Jacobian matrix is specified in Equation (26), and we can further obtain Proposition 4.

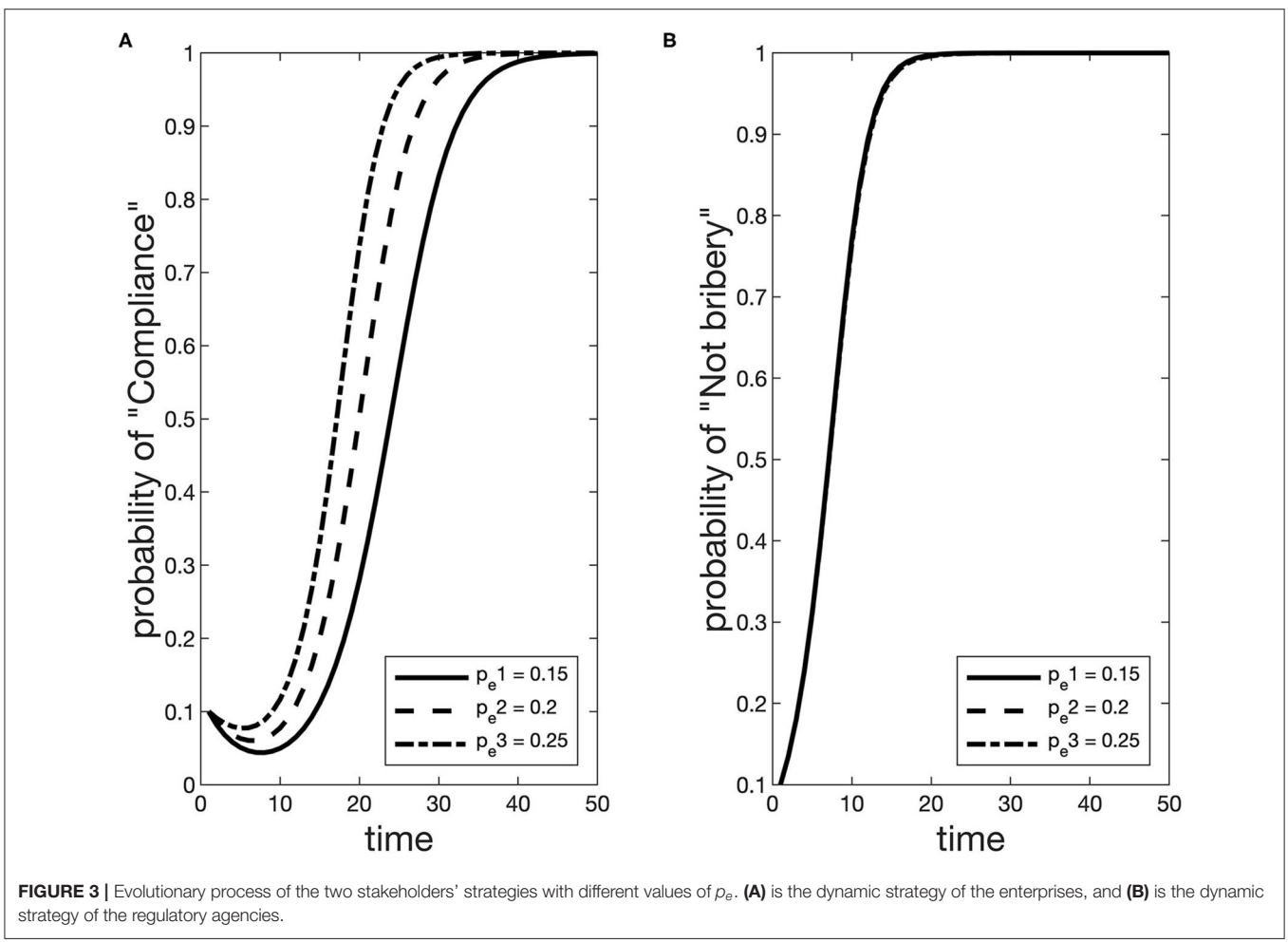



**FIGURE 3 |** Evolutionary process of the two stakeholders' strategies with different values of $p_e$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

$$J_2 = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} \end{bmatrix} = \begin{bmatrix} (1-2x)(\lambda fy + yr - c + p_e + b - by + f - fy) & x(1-x)(\lambda f + r - b - f) \\ y(1-y)(\lambda f - e + b - f) & (1-2y)(\lambda fx - s + e - ex + p_r - b + bx + f - fx) \end{bmatrix} \quad (26)$$

**Proposition 4.**

(i) The equilibrium point (0, 0) is an ESS if $p_e + b + f < c$ and $e + p_r + f < s + b$.
(ii) The equilibrium point (0, 1) is an ESS if $\lambda f + r + p_e < c$ and $s + b < e + p_r + f$.
(iii) The equilibrium point (1, 0) is an ESS if $c < p_e + b + f$ and $\lambda f + p_r < s$.
(iv) The equilibrium point (1, 1) is an ESS if $c < \lambda f + r + p_e$ and $s < \lambda f + p_r$.
(v) The equilibrium point ($x^*$, $y^*$) is not a stable equilibrium point.

**Proof:**

(i) The Jacobian matrix at the equilibrium point (0, 0) is shown as follows:

$$J_2 = \begin{bmatrix} p_e + b + f - c & 0 \\ 0 & e - s + p_r - b + f \end{bmatrix} \quad (27)$$

When $p_e + b + f < c$ and $e + p_r + f < s + b$, the two eigenvalues ($\lambda_1 = p_e + b + f - c$, $\lambda_2 = e - s + p_r - b + f$) of the Jacobian matrix $J_2$ are both negative. Moreover, we have $\det J_1 > 0$ and $\mathrm{tr} J_1 < 0$. Therefore, the equilibrium point (0, 0) is an ESS if $p_e + b + f < c$ and $e + p_r + f < s + b$.

(ii) The Jacobian matrix at the equilibrium point (0, 1) is shown as follows:

$$J_2 = \begin{bmatrix} \lambda f + r - c + p_e & 0 \\ 0 & s - e - p_r + b - f \end{bmatrix} \quad (28)$$

When $\lambda f + r + p_e < c$ and $s + b < e + p_r + f$, the two eigenvalues ($\lambda_1 = \lambda f + r - c + p_e$, $\lambda_2 = s - e - p_r + b - f$) of the Jacobian matrix $J_2$ are both negative. Therefore, the equilibrium point (0, 1) is an ESS if $\lambda f + r + p_e < c$ and $s + b < e + p_r + f$.

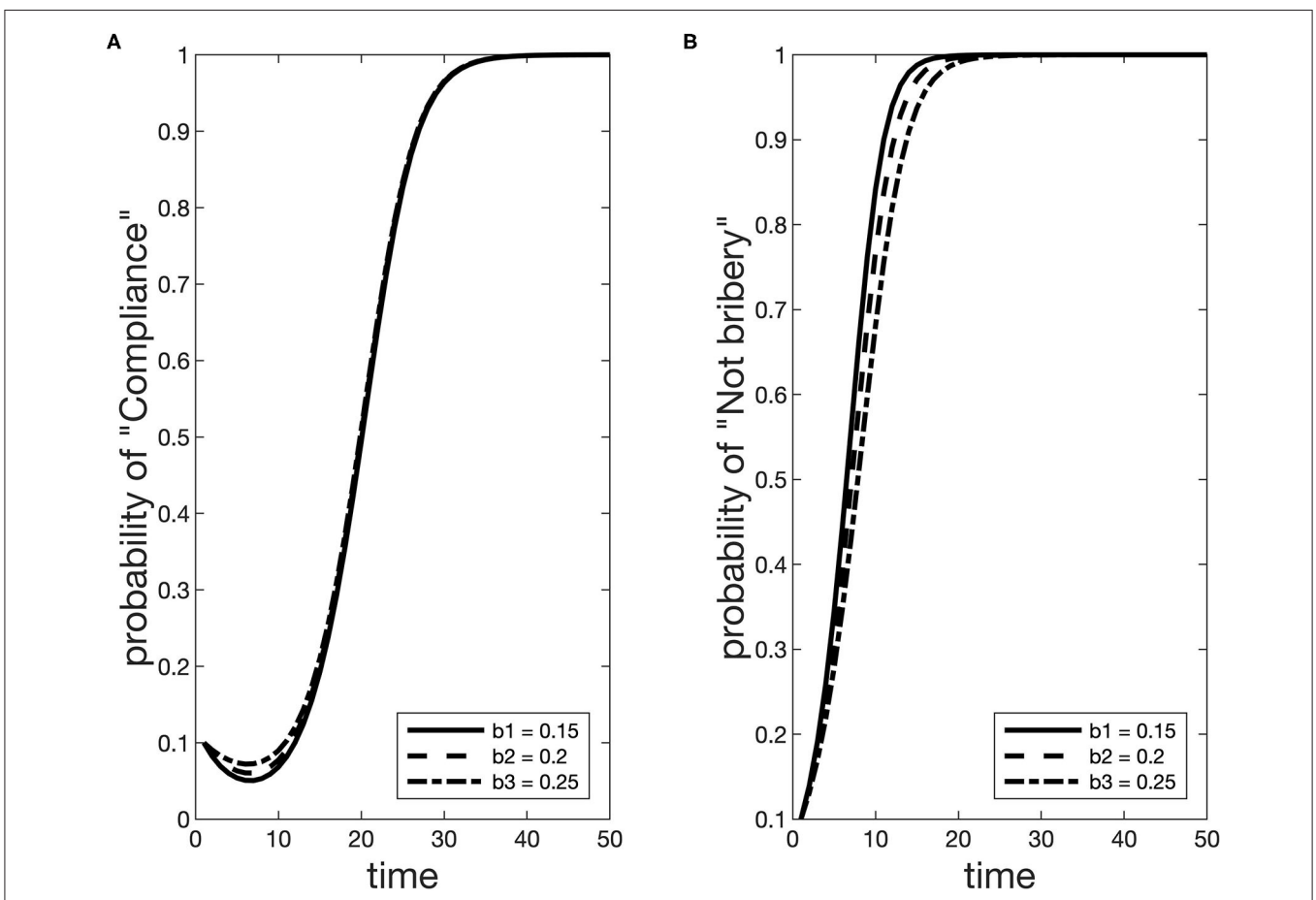



**FIGURE 4 |** Evolutionary process of the two stakeholders' strategies with different values of *b*. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

(iii) The Jacobian matrix at the equilibrium point (1, 0) is shown as follows:

$$J_2 = \begin{bmatrix} c - p_e - b - f & 0 \\ 0 & \lambda f - s + p_r \end{bmatrix} \tag{29}$$

When $c < p_e + b + f$ and $\lambda f + p_r < s$, the two eigenvalues ($\lambda_1 = c - p_e - b - f, \lambda_2 = \lambda f - s + p_r$) of the Jacobian matrix $J_2$ are both negative. Therefore, the equilibrium point (1, 0) is an ESS if $c < p_e + b + f$ and $\lambda f + p_r < s$.

(iv) The Jacobian matrix at the equilibrium point (1, 1) is shown as follows:

$$J_2 = \begin{bmatrix} c - \lambda f - r - p_e & 0 \\ 0 & s - \lambda f - p_r \end{bmatrix} \tag{30}$$

When $c < \lambda f + r + p_e$ and $s < \lambda f + p_r$, the two eigenvalues ($\lambda_1 = c - \lambda f - r - p_e, \lambda_2 = s - \lambda f - p_r$) of the Jacobian matrix $J_2$ are both negative. Therefore, the equilibrium point (1, 1) is an ESS if $c < \lambda f + r + p_e$ and $s < \lambda f + p_r$.

(v) Same as the proof in Proposition 2.

# SIMULATION STUDY

## Parameter Setting

In this paper, we aim to analyze how to effectively promote compliance management in the electricity market based on the evolutionary game model. Therefore, in the long-term stable equilibrium, the optimal equilibrium point is (1, 1). As stated in Proposition 2, in the scenario without governmental supervision, the equilibrium point (1, 1) is an ESS if $c < r + p_e$ and $s < p_r$. However, the condition $s < p_r$ will be hard to satisfy because the psychological costs of the regulatory agencies to select the strategy "Bribery" are usually less than the inspection costs in the reality. This is exactly the reason why the regulatory agencies have the incentive to choose the strategy "Bribery." Therefore, when the central government does not take effective measures to supervise compliance management in the electricity market, it is difficult for all the stakeholders to choose ideal strategies in long-term stable equilibrium.

Proposition 4 shows that in the scenario with governmental supervision, the equilibrium point (1, 1) is an ESS if $c < \lambda f + r + p_e$ and $s < \lambda f + p_r$. This condition can be realized in the reality. Therefore, we mainly focus on the dynamic evolution process

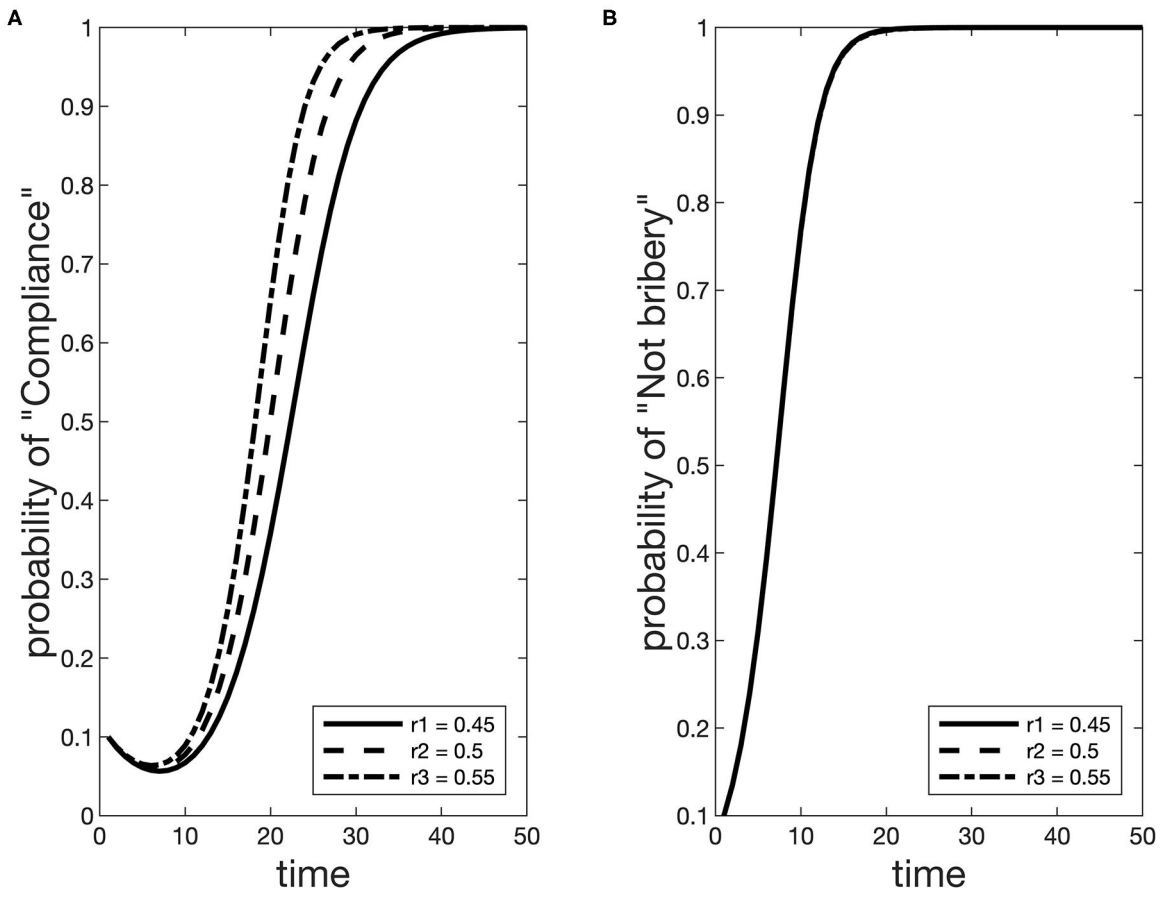



**FIGURE 5 |** Evolutionary process of the two stakeholders' strategies with different values of $r$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

in the scenario with governmental supervision and analyze the various influencing factors through simulation. According to the relevant conditions of the parameters in the ESS, we set the initial values of the parameters (randomly generated) in our evolutionary game model as shown in **Table 4**.

To check the robustness of the results, we set the initial values of $x$ and $y$ to 0.1, 0.3, 0.5, and 0.7, respectively. It can be seen from **Figure 1** that no matter what the initial value is set, the probability that the power enterprises choose "Compliance" and the probability that the regulatory agencies choose "Not bribery" both tend to be 1 in the long run. Therefore, we verify the result that the equilibrium point (1, 1) is an ESS if $c < \lambda f + r + p_e$ and $s < \lambda f + p_r$.

## Sensitivity Analysis

To analyze the impacts of various parameters on the dynamic evolutionary process specifically, we further make the sensitivity analysis for each model parameter.

The heterogeneous impacts of the compliance costs $c$ of the power enterprises are shown in **Figure 2**. It can be seen that changes in compliance costs $c$ exhibit greater impacts on the strategic choices of power enterprises. When the value of compliance costs increases, the probability evolution speed of power enterprises in choosing "Compliance" will decrease, which leads to the persistent choice of "Not compliance" in the electricity market. However, changes in compliance costs have little impact on the strategic choices of the regulatory agencies, and the influence is indirect according to the replicator dynamic system. In other words, the impacts of the changes in compliance costs on the strategic behavior of the regulatory agencies are totally based on the changes in the strategic behavior of power enterprises.

As shown in **Figure 3**, for power enterprises, the impact of psychological cost $p_e$ on the probability of choosing the strategy "Compliance" is just the opposite of the impact of compliance costs $c$. That is to say, the greater the psychological pressure that the power enterprises will face when choosing the strategy "Not compliance," the lower the probability that the power enterprises will choose the strategy "Not compliance." This result illustrates that higher psychological costs $p_e$ can force the power enterprises to be more likely to make decisions that comply with relevant laws and regulations. Similar to the costs of compliance $c$, the psychological costs of power enterprises $p_e$ have little influence on the decision-making process of the regulatory agencies. It mainly affects the strategic behavior of the power enterprises and then plays role on the selection of the regulatory agencies.

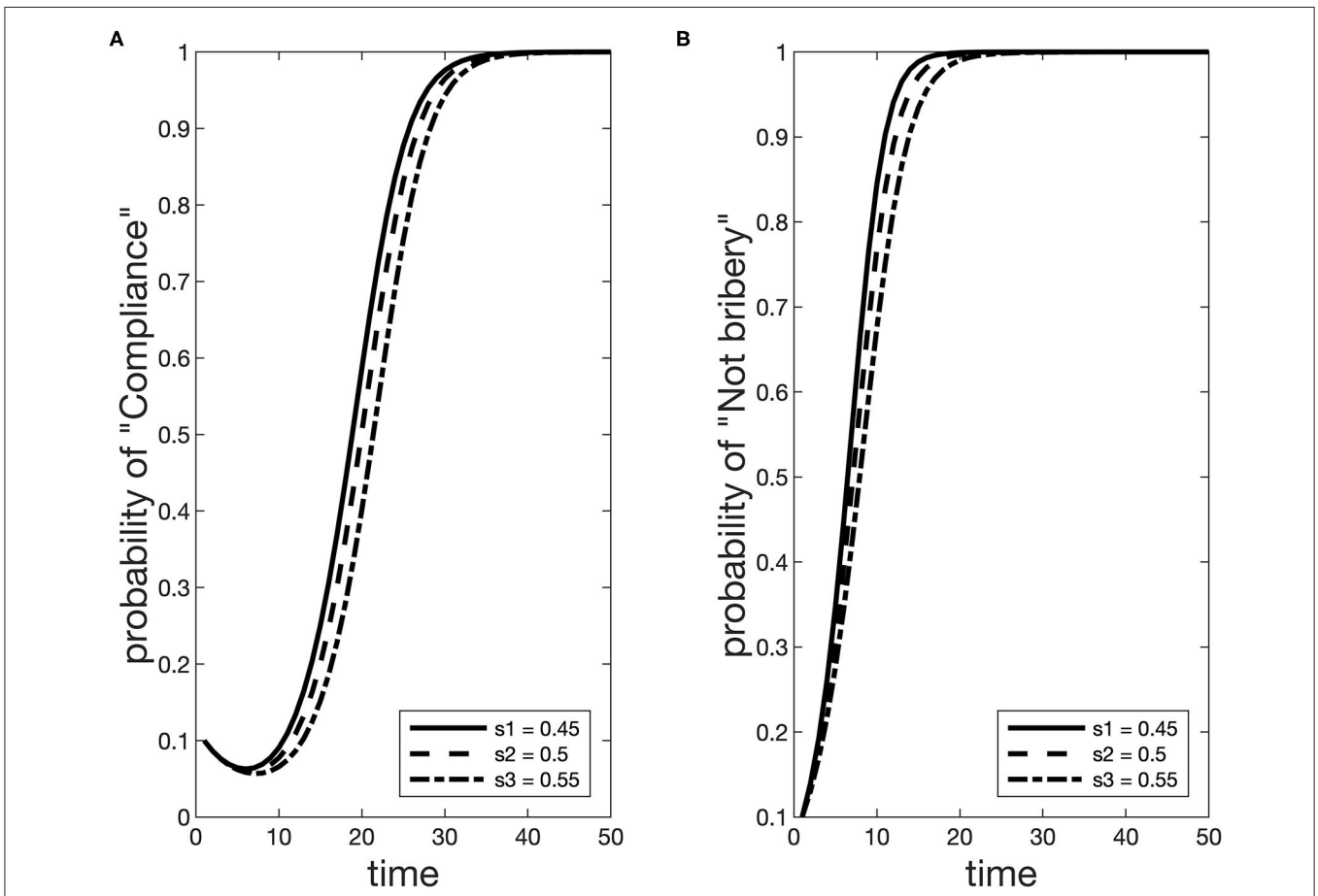



**FIGURE 6 |** Evolutionary process of the two stakeholders' strategies with different values of $s$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

It is seen from **Figure 4** that the bribery costs of the power enterprises $b$ have a certain impact on the strategy selections of the two stakeholders. When the bribery costs $b$ are greater, the power enterprises will choose the strategy "Compliance" instead of the strategy "Not compliance" early, because the power enterprises need to pay more money to the regulatory agencies. As far as the regulatory agencies are concerned, if the amount of bribes is larger, the regulatory agencies are more inclined to collude with power enterprises, which will slow down the dynamic changes of the regulatory agencies' decision to "Not bribery."

The impacts of profit loss $r$ on the strategic choices of the power enterprises are similar to the impacts of psychological cost $p_e$. When the non-compliance of the power enterprises is discovered, the corporate reputation will be damaged, thereby reducing the profits of the power enterprises. The greater the impact on corporate profits, the more likely the power enterprises are to choose strategy "Compliance," which makes the proportion of compliance companies in the electricity market higher. Besides, it is seen from **Figure 5** that the change in a profit loss of the power enterprises $r$ has less impact on the decision choices of the regulatory agencies.

The results in **Figure 6** show that the inspection costs of the regulatory agencies $s$ exhibit similar impacts on the decision-making of power enterprises and regulatory agencies. When the inspection costs of the regulatory agencies $s$ are higher, the probability of power enterprises choosing strategy "Compliance" and the probability of regulatory agencies choosing strategy "Not bribery" will approach 1 more slowly in the long run. This illustrates that higher costs of inspection are not conducive to implement compliance management and effective supervision in the electricity market.

When regulatory agencies choose not to implement effective inspection and tend to collude with power enterprises, they will bear a certain psychological cost $p_r$. As shown in **Figure 7**, the psychological costs of the regulatory agencies $p_r$ also display the same influences on the strategic choices of the power enterprises and the regulatory agencies, which is similar to the inspection costs $s$. When the psychological costs $p_r$ are higher, the possibility of the regulatory agencies choosing strategy "Bribery" will decrease. Moreover, the higher psychological costs $p_r$ can more effectively promote the regulatory agencies to perform the duties of inspection effectively. Furthermore, as the regulatory agencies are less likely to accept bribes, power enterprises are

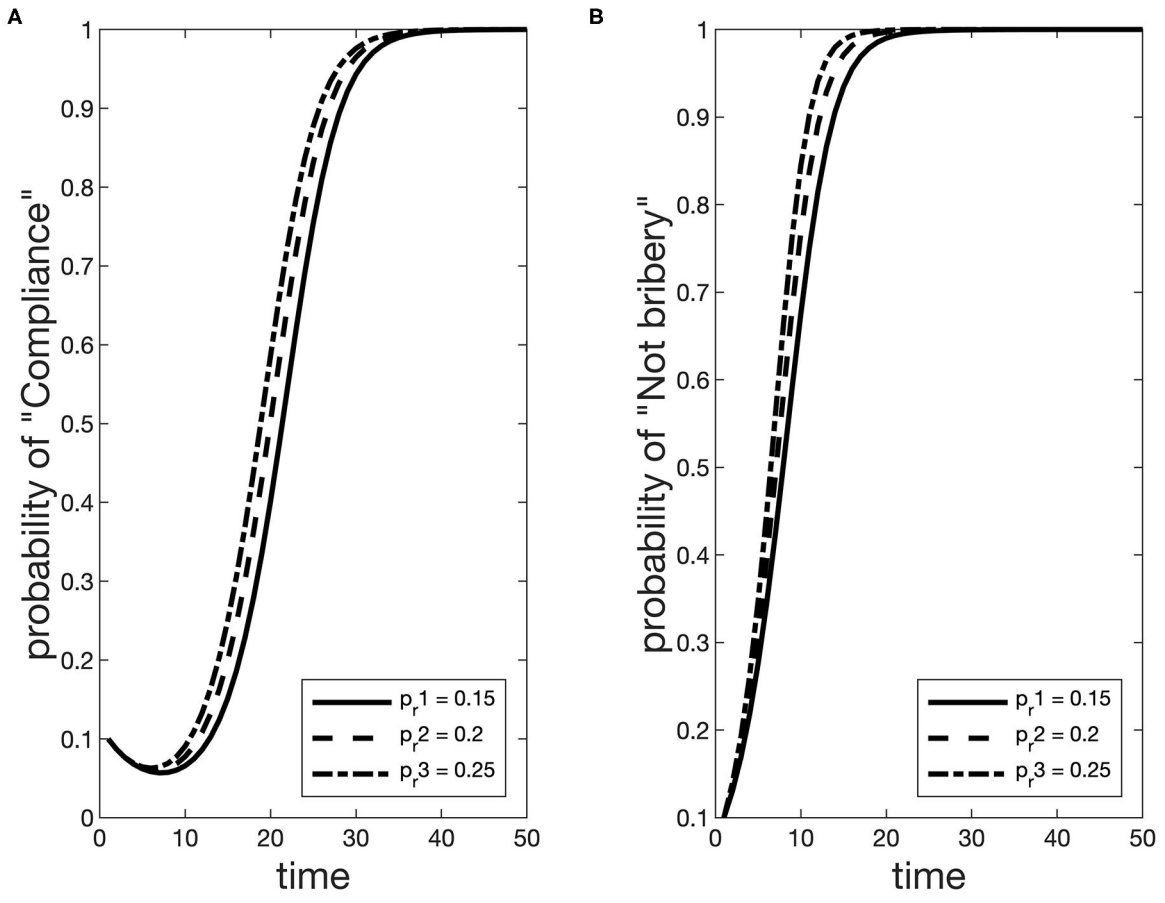



**FIGURE 7 |** Evolutionary process of the two stakeholders' strategies with different values of $p_r$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

also more inclined to conduct "Compliance." Therefore, the higher psychological cost $p_r$ of the regulatory agencies will better promote the compliance management of the electricity market.

As shown in **Figure 8**, the role of the rewards $e$ to the regulatory agencies in detecting the non-compliant behavior in the electricity market is similar to the role of the psychological cost of the regulatory agencies $p_r$. If the regulatory agencies will get higher rewards $e$ when they find the non-compliance behavior of the power enterprises, the regulatory agencies have more incentive to inspect the behavior of power enterprises strictly. As a result, the probability that regulatory agencies choose the strategy "Not bribery" will approach to 1 faster. Besides, as the power enterprises know that their actions will be strictly supervised by the regulatory agencies, the probability of choosing strategy "Compliance" for power enterprises will be higher.

The impact of the fines $f$ charged by the central government for non-compliance behaviors on the two stakeholders' strategy selection process is specifically shown in **Figure 9**. We can find that the change of the fines charged by the central government $f$ has a greater impact on the decision-making of power enterprises, but less on the strategic behavior of regulatory agencies. In addition, the higher the fine charged by the central government, the more effective it can be to promote the power enterprises and regulatory agencies to choose the strategy "Compliance" and the strategy "Not bribery," respectively.

In **Figure 10**, we describe the influence of the coefficient λ of fine in the unilateral bribery on the evolutionary process of the two stakeholders' strategies. It can be seen that compared with the impact on the regulatory agencies, the change of the coefficient λ of fines in the unilateral bribery has a greater impact on power enterprises. If the power enterprises carry out unilateral bribery, the punishment of the central government will be greater, and the power enterprises are more inclined to implement the strategy "Compliance" and avoid heavier fines.

## FURTHER DISCUSSION

From the results of sensitivity analysis, we can find that the impacts of different model parameters on the dynamic process of power enterprises and regulatory agencies' strategy selections are inconsistent. From the perspective of power enterprises, the compliance costs $c$, psychological cost $p_e$, and profit loss $r$ are vital to the dynamic evolutionary process. The increase in compliance costs is not conducive to the implementation of compliance

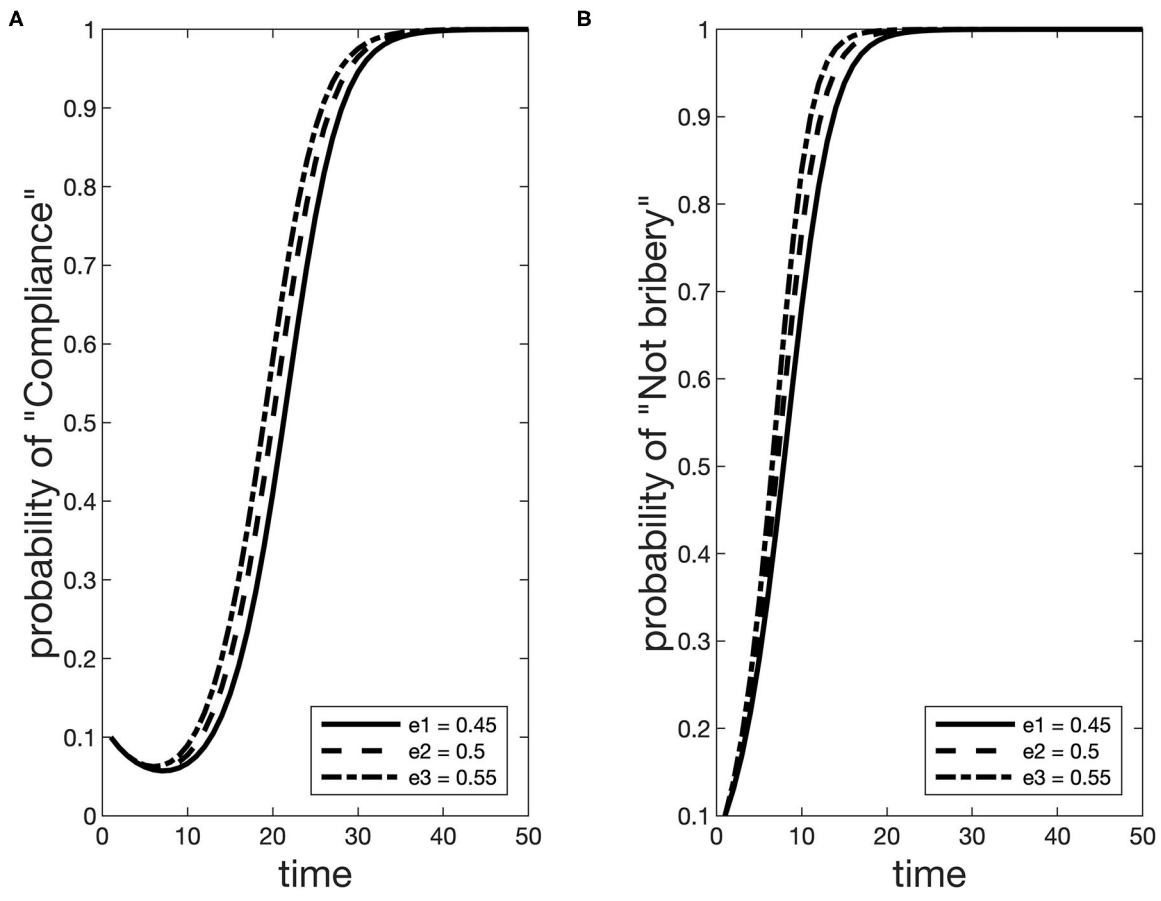



**FIGURE 8 |** Evolutionary process of the two stakeholders' strategies with different values of $e$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

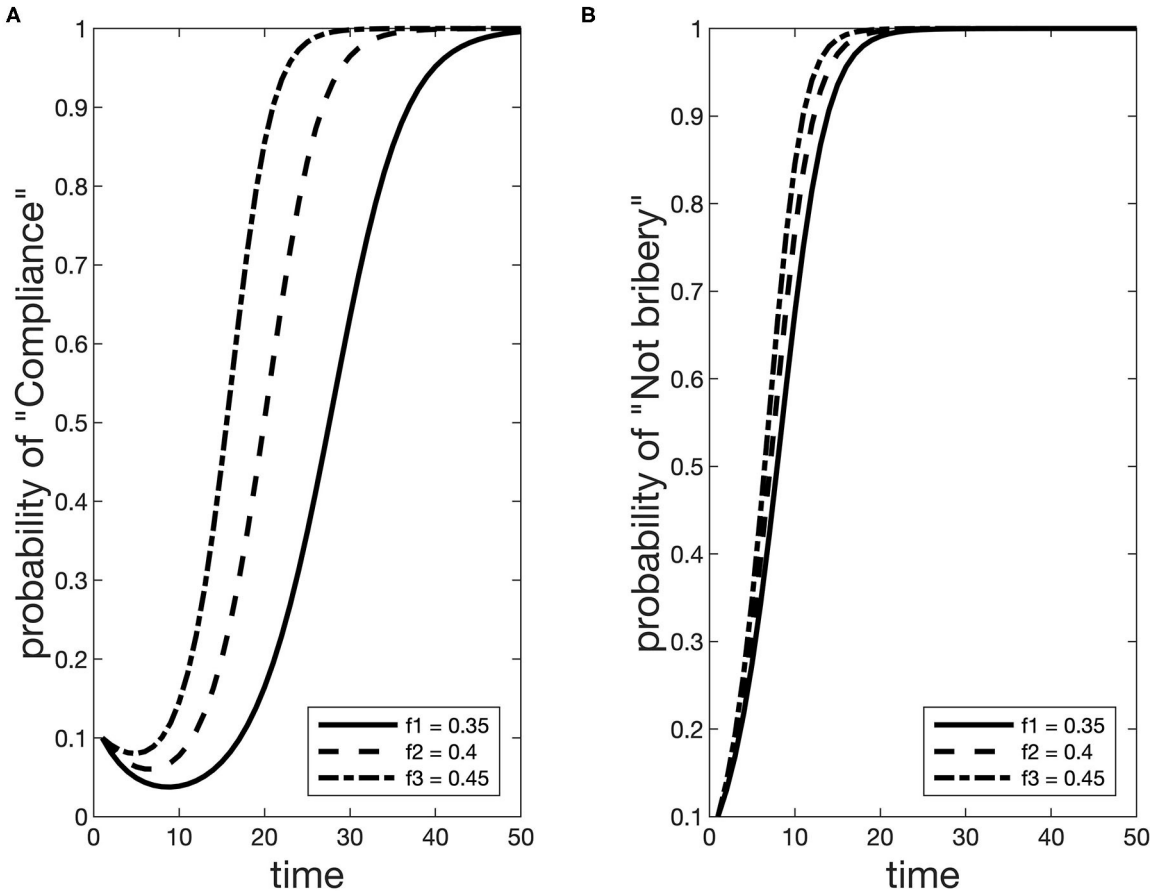



**FIGURE 9 |** Evolutionary process of the two stakeholders' strategies with different values of $f$. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

management in the electricity market. When the costs of compliance increase, power enterprises tend to choose the strategy "No compliance," which leads to the slow evolutionary process of power enterprises to the choice of compliance. Moreover, if the psychological cost of strategy "No compliance" is large enough, the power enterprises will tend to choose the strategy "Compliance." Besides, once the non-compliance behavior of power enterprises is disclosed, their reputation and profit will be seriously damaged. The higher the profit loss of power enterprises, the more likely they are to choose the strategy "Compliance." From the perspective of regulatory agencies, the bribery profits $b$, inspection costs $s$, psychological cost $p_r$, and rewards $e$ are important. When the regulatory agencies face the greater temptation of bribery, they are more inclined not to seriously perform their regulatory duties. The higher inspection costs are not conducive to the implementation of the compliance management of the power market. The regulatory agencies, as well as power enterprises, would decrease the probability of choosing the strategies of compliance management. The psychological costs of regulatory agencies when they choose the strategy "Bribery" will also influence their choices. The higher psychological costs can promote the compliance management of the electricity market. Moreover, once the regulatory agencies discover the non-compliance behavior of the power enterprises, the government will give rewards to the regulatory agencies, and the higher rewards are conducive to the compliance management of the electricity market. From the perspective of the central government, the fines $f$ and the coefficient $\lambda$ of fine will help to affect the dynamic evolutionary process. The amount of penalty imposed by the government on power enterprises and regulatory agencies can affect the decision-making of these two stakeholders. When the fines are higher, the power enterprises are more inclined to choose the strategy "Compliance," and the regulatory agencies will tend to choose the strategy "No bribery."

## CONCLUSION AND POLICY SUGGESTION

This paper attempts to analyze how to effectively promote compliance management in the electricity market. By constructing the evolutionary game model of two stakeholders (power enterprises and regulatory agencies) in the electricity market, the ESS under different scenarios is analyzed, i.e., the scenario without governmental supervision and the scenario with governmental supervision. Moreover, we further make a comparative analysis with different scenarios; the results show that, in the absence of government supervision, the long-term

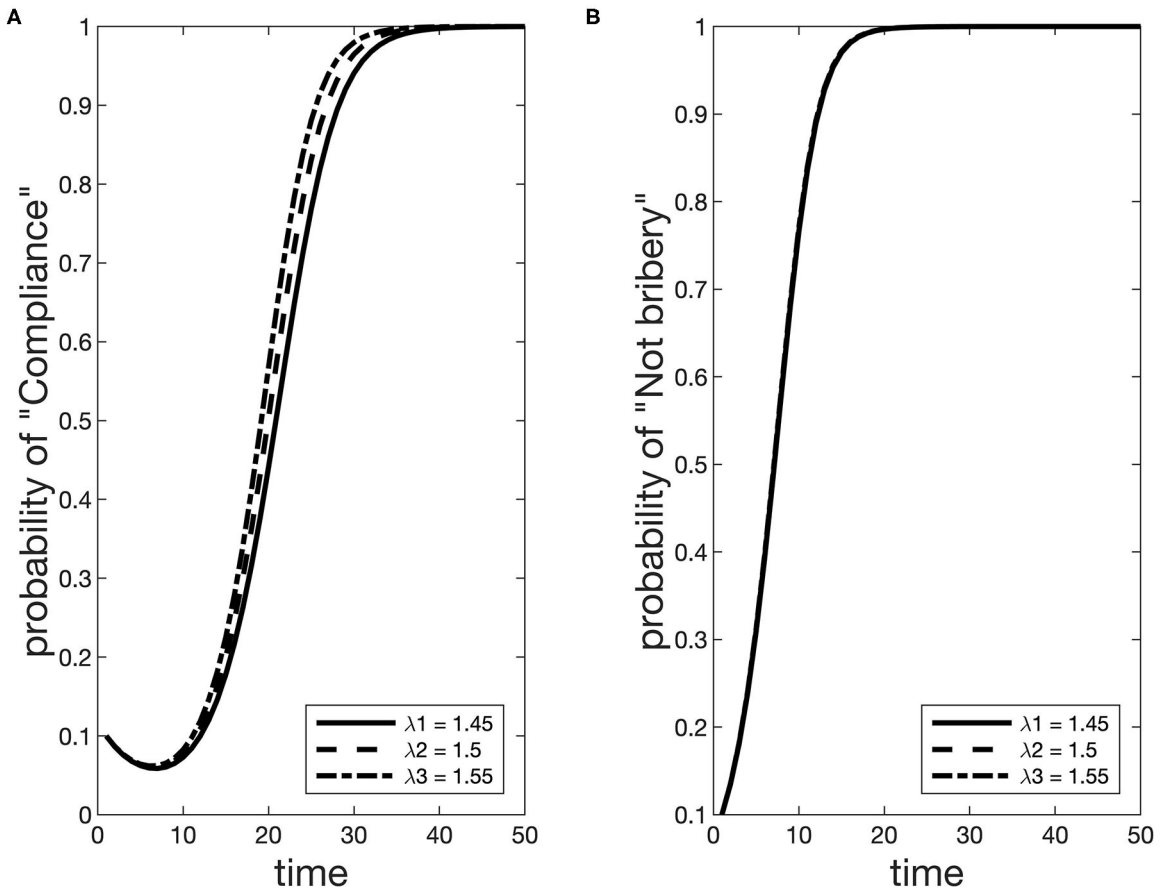



**FIGURE 10 |** Evolutionary process of the two stakeholders' strategies with different values of λ. **(A)** is the dynamic strategy of the enterprises, and **(B)** is the dynamic strategy of the regulatory agencies.

stable equilibrium of power enterprises' choice of strategy "Compliance" and regulatory agencies' choice of strategy "Not bribery" cannot be realized. Only if the government effectively supervises the compliance management of the electricity market can the ideal compliance management of the electricity market be realized. Moreover, we also analyze the roles of different model parameters in the dynamic evolutionary process. The relevant results show that the lower compliance cost, the higher psychological cost, and the larger profit loss of the power enterprises, the lower inspection cost, the higher psychological cost, and the larger rewards of the regulatory agencies, as well as the larger penalty charged by the government, are conducive to the compliance management of the electricity market.

Therefore, to promote the effective implementation of compliance management in the electricity market, the central government should focus on the following aspects. (1) The government should consider taking some measures to reduce the compliance costs of power enterprises and encourage the power enterprises to choose compliance management in the business operation. (2) The government should strengthen the recognition and implementation of compliance management in the electricity market, to make sure that there will be greater psychological pressure if the power enterprises choose the strategy "No compliance," so as to promote the compliance behavior of power enterprises more validly. (3) The government can disclose information of non-compliance power enterprises publicly, so as to give full play to the effective supervision role of the society on compliance management and jointly resist the occurrence of non-compliance in the electricity market. (4) The government should take some measures to reduce the inspection costs of the regulatory agencies to perform their duties and encourage the regulatory agencies to actively inspect the behavior of the power enterprises, rather than colluding with power enterprises. (5) The government should consider conducting stricter supervision on the works of the regulatory agencies, which can stimulate the regulatory agencies to perform their duties strictly. (6) Besides, the government should increase the incentive payments to the regulatory agencies to stimulate their motivation. (7) Last but not least, the government should increase the punishment of non-compliance behavior in the electricity market.

This paper also has some shortcomings. Due to the limitation of data, this paper uses randomly generated data rather than the real data in the simulation. Thus, it can only analyze the dynamic strategy changes of the compliance management in the electricity market, not the real situation of China's electricity market. With the increasing development of compliance management in

China's electricity market, we aim to make a detailed analysis of compliance management in the electricity market through real data in future research.

## DATA AVAILABILITY STATEMENT

The datasets generated in this study are available directly from the corresponding author.

## AUTHOR CONTRIBUTIONS

LJ and YL: conceptualization, methodology, and writing—original draft. XW: software, data curation, and writing—review and editing. CC and YC: conceptualization, methodology, writing—original draft, writing—review and editing, and supervision. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This work was supported by the project of Zhejiang Electricity Trading Center Co., Ltd. (Research and Design of Power Market Compliance Management System, Contract number SGZJDJ00JSJS1900024), National Natural Science Foundation of China (grant numbers 71874149, 71934001), and National Social Science Fund of China (grant number 20ZDA084).

## REFERENCES

Aquila, G., Junior, P. R., Pamplona, E. D. O., and de Queiroz, A. R. (2017). Wind power feasibility analysis under uncertainty in the Brazilian electricity market. *Energy Econ.* 65, 127–136. doi: 10.1016/j.eneco.2017.04.027

Bahrami, S., and Amini, M. H. (2018). A decentralized trading algorithm for an electricity market with generation uncertainty. *Appl. Energy* 218, 520–532. doi: 10.1016/j.apenergy.2018.02.157

Coulon, M., Powell, W. B., and Sircar, R. (2013). A model for hedging load and price risk in the Texas electricity market. *Energy Econ.* 40, 976–988. doi: 10.1016/j.eneco.2013.05.020

da Silva Rocha, A. B., and Salomão, G. M. (2019). Environmental policy regulation and corporate compliance in evolutionary game models with well-mixed and structured populations. *Eur. J. Oper. Res.* 279, 486–501. doi: 10.1016/j.ejor.2019.05.040

Deng, S. J., and Oren, S. S. (2006). Electricity derivatives and risk management. *Energy* 31, 940–953. doi: 10.1016/j.energy.2005.02.015

Friedman, D. (1991). Evolutionary games in economics. *Econometrica* 59, 637–666. doi: 10.2307/2938222

Gu, W., Wei, L., Zhang, W., and Yan, X. (2019). Evolutionary game analysis of cooperation between natural resource- and energy-intensive companies in reverse logistics operations. *Int. J. Prod. Econ.* 218, 159–169. doi: 10.1016/j.ijpe.2019.05.001

Höhne, S., and Tiberius, V. (2020). Powered by blockchain: forecasting blockchain use in the electricity market. *Int. J. Energy Sect. Manag.* 14, 1221–1238. doi: 10.1108/IJESM-10-2019-0002

Ji, P., Ma, X., and Li, G. (2015). Developing green purchasing relationships for the manufacturing industry: an evolutionary game theory perspective. *Int. J. Prod. Econ.* 166, 155–162. doi: 10.1016/j.ijpe.2014.10.009

Jiang, B., Wang, X., Xue, H., Li, J., and Gong, Y. (2020). An evolutionary game model analysis on emission control areas in China. *Mar. Policy* 118:104010. doi: 10.1016/j.marpol.2020.104010

Jiang, K., You, D., Merrill, R., and Li, Z. (2019). Implementation of a multi-agent environmental regulation strategy under Chinese fiscal decentralization: an evolutionary game theoretical approach. *J. Clean. Prod.* 214, 902–915. doi: 10.1016/j.jclepro.2018.12.252

Kharrati, S., Kazemi, M., and Ehsan, M. (2016). Equilibria in the competitive retail electricity market considering uncertainty and risk management. *Energy* 106, 315–328. doi: 10.1016/j.energy.2016.03.069

Klessmann, C., Nabe, C., and Burges, K. (2008). Pros and cons of exposing renewables to electricity market risks—a comparison of the market integration approaches in Germany, Spain, and the UK. *Energy Policy* 36, 3646–3661. doi: 10.1016/j.enpol.2008.06.022

Meunier, G. (2013). Risk aversion and technology mix in an electricity market. *Energy Econ.* 40, 866–874. doi: 10.1016/j.eneco.2013.10.010

Ritzberger, K., and Weibull, J. W. (1995). Evolutionary selection in normal-form games. *Econometrica* 63, 1371–1399. doi: 10.2307/2171774

Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *J. Theor. Biol.* 84, 93–101. doi: 10.1016/S0022-5193(80)81038-1

Sheng, J., Zhou, W., and Zhu, B. W. (2020). The coordination of stakeholder interests in environmental regulation: lessons from China's environmental regulation policies from the perspective of the evolutionary game theory. *J. Clean. Prod.* 249:119385. doi: 10.1016/j.jclepro.2019.119385

Song, M., Fisher, R., and Kwoh, Y. (2019). Technological challenges of green innovation and sustainable resource management with large scale data. *Technol. Forecast. Soc. Change* 144, 361–368. doi: 10.1016/j.techfore.2018.07.055

Sun, Z., and Zhang, W. (2019). Do government regulations prevent greenwashing? an evolutionary game analysis of heterogeneous enterprises. *J. Clean. Prod.* 231, 1489–1502. doi: 10.1016/j.jclepro.2019.05.335

Taylor, P. D., and Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Math. Biosci.* 40, 145–156. doi: 10.1016/0025-5564(78)90077-9

Tian, Y., Govindan, K., and Zhu, Q. (2014). A system dynamics model based on evolutionary game theory for green supply chain management diffusion among Chinese manufacturers. *J. Clean. Prod.* 80, 96–105. doi: 10.1016/j.jclepro.2014.05.076

Uribe, J. M., Mosquera-López, S., and Guillen, M. (2020). Characterizing electricity market integration in Nord Pool. *Energy* 208:118368. doi: 10.1016/j.energy.2020.118368

Vehviläinen, I., and Keppo, J. (2003). Managing electricity market price risk. *Eur. J. Oper. Res.* 145, 136–147. doi: 10.1016/S0377-2217(01)00399-X

Xie, H., Cheng, L., and Lv, T. (2017). Factors influencing farmer willingness to fallow winter wheat and ecological compensation standards in a groundwater funnel area in Hengshui, Hebei Province, China. *Sustainability* 9:839. doi: 10.3390/su9050839

Zhu, C., Fan, R., and Lin, J. (2020). The impact of renewable portfolio standard on retail electricity market: a system dynamics model of tripartite evolutionary game. *Energy Policy* 136:111072. doi: 10.1016/j.enpol.2019.111072

**Conflict of Interest:** LJ, CC, and XW were employed by the company Zhejiang Electricity Trading Center Co., Ltd. YL was employed by the company Zhejiang Huayun Mdt Info Tech Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Jin, Chen, Li, Wang and Cheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 20 May 2021
doi: 10.3389/fenvs.2021.696626

# Evaluating the Impact of Urban Blue Space Accessibility on Housing Price: A Spatial Quantile Regression Approach Applied in Changsha, China

*Huang Tuofu*[1], *He Qingyun*[1]*, *Yang Dongxiao*[2] *and Ouyang Xiao*[3]

[1]College of Geographic Science, Hunan Normal University, Changsha, China, [2]School of Economics and Trade, Hunan University, Changsha, China, [3]Hunan Institute of Economic Geography, Hunan University of Finance and Economics, Changsha, China

Urban nature spaces are increasingly recognized as essential urban features providing crucial amenities to the residents' health and well-being. While many studies have been conducted focusing on the influence of green spaces on house prices, very few have explored the impact of urban blue spaces. In this study, we analyzed the proximity effects of different types and sizes of urban blue spaces on property value in Changsha metropolis, China, and examined the spatial quantile effect across different housing prices. A two-stage instrumental method (2SLS) hedonic model was employed to evaluate the impact of different types of urban blue space: river (mainstream and tributary), wetland park, and lakes (large, medium, and small). Spatial quantile regression (SQR) was then used to measure the spatial effect of accessibility on various house price ranges. The 2SLS results show that, except for small-sized lakes, proximity to blue spaces significantly increases property value. Analysis of the SQR model reveals that proximity to major blue spaces increases the marginal willingness-to-pay among homebuyers of high-priced properties, while ordinary blue spaces are more attractive to buyers of low- and medium-priced houses. This may be broadly related to the level of education, utility, and sensitivity to ecosystem services across income groups. Based on these findings, we recommend that urban planners adopt different strategies to promote space utilization efficiency. This is one of the few studies that analyze the capitalization of blue space accessibility in house prices across different quantiles. By categorizing blue spaces and employing the SQR model, this study found the effect of blue spaces on housing prices to be heterogeneous, providing new perspectives to the existing literature.

**Keywords: blue space, housing price, spatial quantile regression, accessibility, urban China**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Han Li,*
*University of Miami, United States*
*He Nie,*
*National University of Singapore, Singapore*
*Chen Ziyue,*
*Norwegian University of Science and Technology, Norway*

****Correspondence:***
*He Qingyun*
*heqingyun1848@gmail.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *17 April 2021*
***Accepted:*** *03 May 2021*
***Published:*** *20 May 2021*

***Citation:***
*Tuofu H, Qingyun H, Dongxiao Y and Xiao O (2021) Evaluating the Impact of Urban Blue Space Accessibility on Housing Price: A Spatial Quantile Regression Approach Applied in Changsha, China. Front. Environ. Sci. 9:696626. doi: 10.3389/fenvs.2021.696626*

## INTRODUCTION

In the process of intensified urbanization across modern China, the well-being of residents faces numerous challenges. Deteriorating ecological environment, unhealthy lifestyle, and highly competitive employment market (Wu Q. et al., 2020, Wu et al., 2021), have put significant pressure and risks on the urban population's health and well-being. In response to these social

and environmental concerns, nature spaces have become increasingly popular not just for their aesthetic functions but also for their various ecosystem services such as regulating phytoclimate, reducing the heat island effect, and providing recreation and leisure spaces (Gómez-Baggethun and Barton, 2013; Roebeling et al., 2017).

In general, urban nature space can be divided into two categories: green spaces (urban vegetation areas) and blue spaces (urban surface waters) (Targino et al., 2019). Numerous articles suggest that attractive landscapes composed of green and blue spaces promote physical and mental health (Voelker and Kistemann, 2015). With the concept promoted by the government, media, and property developers, urban residents have started to put increased value on the ecological-environmental quality around their settlements. In many cities, housing with better ecological surroundings has become expensive. According to the hedonic price theory (Rosen, 1974), adjacent nature spaces have become key attributes to the value of a property. Empirical evidence from hedonic models has shown that open public green spaces (e.g., forests, grasslands, parks) can increase the price of the surrounding properties through visibility and accessibility (Kong et al., 2007). However, while urban blue spaces are extremely important landscape elements (Kaplan and Kaplan, 1989), their effect on property value has largely been overlooked. Studies on the benefits of visits and access to urban blue spaces found that these areas provide positive health effects as much as their terrestrial counterparts (Lianyong and Eagles, 2009). Despite the growing research trend on the health benefits of urban blue spaces, studies on the capitalization of blue attributes to the housing market have remained limited. Very few hedonic housing price studies have explicitly analyzed the value impact of urban blue spaces and instead often merged blue spaces with other urban open space types.

Proximity and access to water have been regarded as key factors for human settlement, providing numerous benefits. Some studies have tried to value the effect of urban blue spaces on the housing market across different cities (Sander and Zhao, 2015; Liebelt et al., 2019; Liu et al., 2020). However, these studies did not differentiate blue spaces. Empirical evidence on the impact caused by urban blue spaces on house prices remains vague and unclear. In addition to cities with coastlines, blue spaces in inland cities comprise different freshwater bodies, including rivers, wetlands, and lakes. More studies are needed to investigate how the different types and sizes of urban blue spaces may affect prices in the housing market.

The conventional hedonic price models using ordinary least square (OLS) regression may lead to inconsistencies between estimated results and reality (McMillen, 2012). Assuming housing is a normal good, wealthy buyers usually purchase high-priced houses that provide higher quality hedonic attributes, including natural amenities. This suggests that high-income families pay more at the margin for natural resources. Considering that the willingness-to-pay for blue amenities may vary for buyers of different income groups, estimated effects of blue spaces across the conditional distribution of housing prices could be biased. Previous studies evaluating hedonic attributes on house prices using quantile regression found significant differences among samples at different price regions (Liao and Wang, 2012). These results indicate that the proximity of some attributes (e.g., facilities or green spaces) has a positive effect on nearby housing prices and that the impact varies for high- and low-priced houses. Another study found that proximity to educational facilities positively correlates with house prices and that homebuyers of high-priced houses are willing to pay more for educational resources (Wen et al., 2014).

However, for blue spaces, their impact on house prices has not been sufficiently explored. Spatial dependence also needs to be considered in exploring the effect of blues spaces in hedonic studies. Adjacent residences share public facilities, natural resources, and other amenities, which means that their prices may correlate spatially (Zhang, 2016). Spatial econometrics has been broadly applied in hedonic models (LeSage and Pace, 2004) and on the capitalization of housing attributes (Huang et al., 2017), but not for hedonic studies on the capitalization of blue spaces. Variations in the spatial effects across the conditional distribution of property prices have also been largely overlooked in the existing literature.

To address this research gap, this study used a spatial quantile hedonic model to investigate the proximity effects of different types of blue spaces on property prices across different conditional distributions. The study area was Changsha city, an important regional central city in south-central China. To the best of our knowledge, this is the first study to use a spatial quantile hedonic model for blue space capitalization in urban China. This research can provide empirical evidence for the effective use of land resources for urban managers and planners.

This paper is organized as follows. The next section reviews the relevant literature and discusses the contributions of this study. *Data and Methodology* presents the research area, the data used in the study, and the main research methodology. *Results* provides the empirical results and robustness check analysis to avoid bias. *Discussion* examines how the results correspond to previous studies and explores their implications. *Conclusion and Recommendation* provides the conclusion.

## LITERATURE REVIEW

Given the clear relations between the natural environment and human health (Gascon et al., 2015), numerous articles have found a positive capitalization effect of green spaces using the hedonic approach (Kong et al., 2007; Conway et al., 2010). Different subcategories of green spaces have been analyzed to compare their effects on housing prices (Czembrowski and Kronenberg, 2016). While the studies found the impact of green spaces varies, the positive effect of most green spaces is clear. In addition, rich patches of green spaces have also been found to have significant influence on house prices (Kong et al., 2007). Visibility and accessibility are key variables indicating the impact of green spaces on property value (Jim and Chen, 2006; Brander and Koetse, 2011; Czembrowski and Kronenberg, 2016). Some studies have also explored the different effects of green spaces on

capitalization from different ecosystem services (Sander and Haight, 2012; Gómez-Baggethun and Barton, 2013).

Studies have found that urban blue spaces have a unique effect on the health and well-being of residents, different from the impact of green spaces. Some studies have found unique health-enhancing effects of urban blue in multiple conceptual therapeutic landscape dimensions, particularly experienced and symbolic (Voelker and Kistemann, 2015). Social interaction and psychological benefits have been found as the most important benefits obtained from visiting blue spaces, especially among the elderly (Garrett et al., 2019; Targino et al., 2019). The unique resistance of blue spaces to urban heat island effects cannot be ignored (Yu et al., 2020). Considering that the environmental amenity has a strong effect on housing prices, these previous studies suggest that urban blue as a residential environment-promoting factor needs a more comprehensive evaluation of its general and specific capitalization effects.

A growing number of studies have found a positive capitalization effect of blue spaces. One hedonic study used blue spaces as a subcategory of green spaces to value blue and green spaces in Ramsey and Dakota Counties in the Unites States (Sander and Zhao, 2015). Proximity to large parks, as well as water bodies, was found to increase house prices in the Phoenix Metropolitan Area (Larson and Perrings, 2013). Another study found a positive impact of lakes on property valuation (Anderson and West, 2006). While these studies were able to find significant impact of blue spaces on house prices, this was only one of their sub-objectives. Lakeside and riverside have been used as main subjects in previous hedonic studies on urban blue; however, the impact of blue spaces in these studies was found to have radically different magnitudes (Kelejian and Prucha, 1999; Sander and Zhao, 2015; Mei et al., 2018; Wu P.-I. et al., 2020). One possible reason is that many of these studies did not distinguish between the various types of blue spaces. In general, research on this subject matter remains limited, particularly compared to green spaces. More research is needed on the capitalization of blue spaces, particularly on assessing the impact of different kinds of urban blue spaces.

Another possible factor is the quantile effect, which had been rarely considered in previous studies. As a hedonic attribute, blue space would be valued differently by residents with varying income levels and social status. Housing price levels often correspond to households with different incomes, which may result in very different marginal buying strategies for the same characteristic attributes. In many cases, having segmented housing markets distinguished by price is even more reasonable than separating them from geographical boundaries (Rajapaksa et al., 2017). Numerous studies have found that the influence on housing prices of property attributes, such as house structure and neighborhood facilities, is heterogeneous across the whole distribution (Ebru and Eban, 2011). This heterogeneity caused by different quartile effects can be detected through the quantile model (QR), which can comprehensively evaluate the various distributions (Coulson and McMillen, 2007).

Employing the QR model can reveal more information about sub-markets, which OLS regression may result in biased estimates. QR has better robustness to outliers, can avoid biased estimation caused by OLS truncation of dependent variables, and better handle heteroscedasticity (McMillen, 2012). For these reasons, QR has been used in hedonic models. For example, some studies concluded that the quartile effect of green space accessibility on house price across distributions could not be ignored (Zietz et al., 2008; Mak et al., 2010).

QR is also applicable to blue space capitalization for the following reasons. First, the same blue attribute may be valued variedly by homebuyers of different income levels. Studies have found that geographically accessible blue spaces in local areas are more highly valued by the elderly or senior residents with restricted mobility. In general, mid-age and elderly residents have higher capabilities and willingness to pay for high-priced properties with better amenities. Correspondingly, the coefficient estimates of urban blue spaces based on the sample of high-priced houses may be different with conditional mean. Second, residents of high-priced properties may be more cognizant of ecological-environmental factors and are willing to pay higher prices for blue spaces. Herbert and Thomas (1998) found a strong positive correlation between education level and economic status. People with higher income levels are more likely to receive better education and better understand the relationship between blue spaces and personal well-being. Thus, homebuyers from different price regions may perceive blue spaces differently. Few studies have explored the hedonic pricing of urban blue spaces using the quantile perspective.

In the QR model, spatial effects should be considered, as studies have shown high spatial dependence in property prices. Ignoring spatial effect would likely lead to overestimation in attribute variables. Kang and Liu (2014) used a quantile regression model to study the impact of the 2008 financial crisis on Taiwan's real estate market. They found that the price impact on high-priced houses was more significant than on low-priced ones. Mueller and Loomis (2014) conducted an empirical study and found that the influence of fires in Southern California differs across the conditional distribution of housing prices. These two studies did not consider the spatial dependence of prices, which could explain the huge gap in estimated coefficients between price samples.

To address the spatial aspects, the SAR method is a commonly used modification to the hedonic model; however, this method still has major limitations. The lag term of the dependent variable as a lag factor may be related to the error term, which causes endogenous problems (Anselin, 2002). To deal with these concerns, the two-stage regression method (2SLS) is used, which is essentially an instrumental variable method, leading to a more consistent estimate (Chernozhukov and Hansen, 2006). At the first stage, the explanatory variable regresses the instrumental parameter to obtain the estimates. In the second stage, the obtained estimate of the explanatory variable is regressed on the explained variable, which is the final regression result of 2SLS.

In this study, we employed a 2SQR model, which combines the 2SLS and QR (Kim and Muller, 2004). Similar to previous studies (Zietz et al., 2008; Kuethe and Keeney, 2012), the instrumental variable adopts the lag term of the independent variable (WX),

which is used in adding a regressor in the second stage. To deal with weak instruments, we adopted the IVQR (Chernozhukov and Hansen, 2006). This approach is similar to the GMM method and has been widely used in numerous research fields (Kostov, 2009).

As far as we know, few studies have investigated the quantile effect of blue spaces on housing prices, especially for large inland cities in China. This study, which utilizes a quantile regression model of the two-stage instrumental variable method, explores the effect of various types and sizes of blue spaces and different categories of homebuyers, filling the gap of the detailed and accurate determination of urban blue spaces general and specific capitalization effects in the existing literature across modern urban China.

# DATA AND METHODOLOGY

## Study Overview

As a kind of ecological land, blue space has several benefits and costs to urban residents' well-being. A number of studies have analyzed ecosystem services provided by blue spaces, such as increasing recreation, reducing harm, supplying leisure places, and beautifying the environment (Gómez-Baggethun and Barton, 2013). On the other hand, blue spaces may also result in higher flood risks, negatively affecting house prices (Wu P.-I. et al., 2020). In this study, the net impact of blue spaces in house prices is equal to the positive minus the negative utilities. The accessibility of blue spaces (with different categories based on attribute and size) is used as the variable.

To calculate accessibility, the walking distance between transaction houses and nature blue in the year 2019 was calculated using ArcGIS subtract from house transaction and road network data. Since hedonic regression requires other house attributes, the housing structural features, community attributes, public facilities, and location are also obtained. Considering that businesses, living facilities, and other information may impact housing prices, network capture technology was used to obtain various points of interest (POI) in Changsha City for 2019.

In this study, a 2SLS regression model was constructed to quantify the mean effect of blue space accessibility across all distribution samples, which would prevent biased outcomes and avoid overestimating the effects of independent variables. The methodology used in this study strikes the right balance between unbiased and valid estimates. We employed the 2SQR model to investigate the spatial quantile effect of varying kinds of urban blue spaces. The 2SQR model eliminates endogeneity produced by the spatial heterogeneity of housing price and provides integrated and complete descriptions of various distributions. This model, developed by McMillan in 2012, greatly improves the performance of quantile regression OLS regression, and the densities of the different sub-samples were added to estimate the robustness of the test results.

## Model Description

For the robustness check, we used the reduced hedonic form for cross-checking. Considered the general research conclusions, the logarithm of the dependent variable can be taken to obtain a better model explanatory power. The reduced form of the model is as follows:

$$lnP = \beta_0 + \beta_i S_i + \gamma_j L_j + \delta_k N_k + \theta_m B_m + \rho_n G_n + \varepsilon \quad (1)$$

where P is the property price; S is the structure characteristics; L is the location attributes; N is the neighborhood facilities attributes; B is the blue space attributes; G is the green space without waterscape; α0, $\beta_0$, $\beta_i$, $\gamma_j$, $\delta_k$, $\theta_m$, and $\rho_n$ are the coefficients; and ε is the error term.

When considering the spatial dependence effect in house pricing caused by neighborhood houses, the spatial auto aggression model (SAR) is the preferred option. The lag term (Wlnp) is added to the reduced hedonic model in order to indicate the influence of adjacent housing prices. The evolved spatial hedonic form as follows:

$$\ln P = \beta_0 + \psi W \ln p + \beta_i S_i + \gamma_j L_j + \delta_k N_k + \theta_m B_m + \rho_n G_n + \varepsilon \quad (2)$$

where W is a spatial weight matrix, and ψ is the coefficient, which reveals the extent of spatial price dependence. The sign of ψ indicates the direction of spatial dependence. The value of ψ represents the average impact on house prices across conditional distributes. The greater the absolute value of ψ, the higher the extent of spatial dependence. W is a spatial weight matrix, which is the pivotal part of spatial econometrics and statistics. There are several methods to construct W. Considering that house price spatial effect decay is as long as the distance, we used the Inverse distance weight matrix, which takes the form

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}}, & d_{ij} \le h \\ 0, & d_{ij} > h \end{cases} \quad (3)$$

where $d_{ij}$ is the Euclidean distance between i and j, and h is the distance bandwidth, calculated using the minimum variance error method (MSE). When the distance between i and j is less than the bandwidth, the value of $W_{ij}$ is the reciprocal of the distance; when the distance between the two is greater than the bandwidth, $W_{ij}$ is 0. The value of the bandwidth in this study is equal to 3.71 km.

We also applied the 2SLS method based on hedonic SAR to further eliminate the endogeneity of the model (Kelejian and Prucha, 1999). Similar to the instrumental variable method, it has two steps. First, WlnP is regressed on instruments of X and WX. Then, the predicted value of WlnP is added to SAR as an independent variable. To examine the impact of blue space attribution on different house prices, we adopted a quantile form with a spatial lag term based on the 2SLS technique, which has been shown useful in previous studies (Kim and Muller, 2004; Liao and Wang, 2012; McMillen, 2012). The final form of the spatial quantile form as follows:

$$\ln P = \beta_0(\tau) + \psi(\tau) W \ln p + \beta_i(\tau) S_i + \gamma_j(\tau) L_j + \gamma_j(\tau) N_k + \theta_m(\tau) B_m + \rho_n(\tau) G_n + \varepsilon \quad (4)$$

where $\tau$ is the corresponding quantile of housing price; $\beta_0(\tau)$, αi(τ), $\beta_i(\tau)$, $\gamma_j(\tau)$, $\gamma_j(\tau)$, $\theta_m(\tau)$ and $\rho_n(\tau)$ are the τth quantile coefficients; and, $\psi(\tau)$ is the coefficients of the lag term.

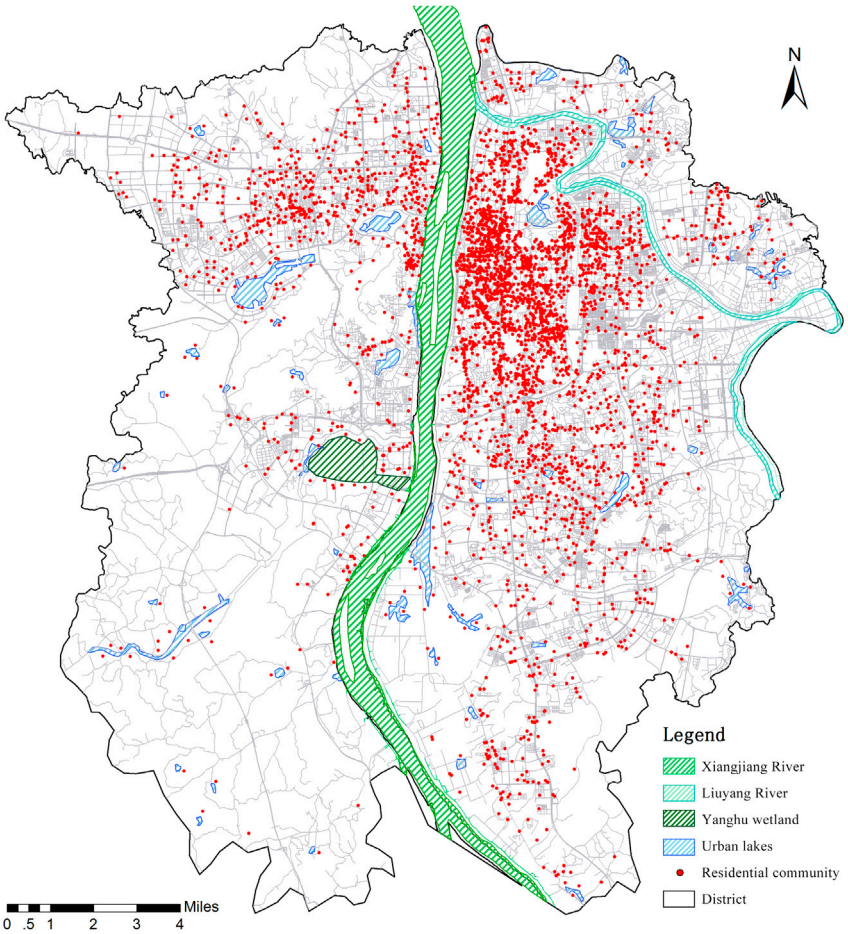



**FIGURE 1 |** Urban blue spaces and residential community in Changsha.

## Data and Variables

The study area is Changsha, a typical inland second-tier city in the middle region of China. This city has shown considerable economic growth in recent years. From 2014 to 2019, the city's average annual GDP growth was 10.4%, much higher than the national average of 6.8%. As China continues to shift its development focus to the Middle Region, Changsha's future is promising.

In terms of the housing market, as one of the pioneers restoring the Housing Purchasing Restriction Policy (HPR), Changsha has been described as the model city for housing market regulations by the central government and media. Although Changsha is considered a model for regulations, its average housing prices remain comparatively low compared with other capital cities in China, mainly due to its sufficient housing and land supply. Oversupply has kept Changsha's housing prices tepid for years. Changsha also has a large number of houses constructed by state companies or institutions solely for its employees. These houses usually are sold at low prices as incentive for their employees, lowering the average housing prices in this city. Despite strict regulations, the city's economy remains vibrant, and the local government does not directly intervene with housing prices. Changsha's HPR experience has been promoted nationally to encourage other mega-cities to apply similar management strategies. For these reasons, Changsha is an appropriate research object for housing prices, and to some extent, can be for the housing market in modern China.

**Figure 1** shows the study area. Focusing on the agglomeration of residential communities, the urban district of Changsha is used as the research area. Data employed in this study could be divided into three categories: housing data, the POI data, and the Land-Use Land-Cover Change dataset (LUCC).

There are three sources of housing data for Chinese cities: local official data, tax assessor data, and websites of real estate (Li et al., 2019a). With respect to data accuracy, the local official data is highly appropriate for housing price studies. We obtained the housing data from the Housing and Urban-rural Development of Changsha, including the property's location, size, and market value. We collected 62,187 residential transactions from January 2019 to December 2019 and merged the corresponding information for residential communities. We also obtained community data, such as location, plot ratio, house age for all 5,860 communities in the region from the same source.

The POI data were gathered based on the service facilities of Baidu Map (map.baidu.com) and complemented by digitizing satellite maps of Google Earth. This dataset includes private service amenities and public facilities, such as schools, shopping malls, and subway stations. Web open data (e.g., Baidu Map) has been used in previous studies because of their availability and operability. However, there are still concerns about the accuracy and usefulness of open-source web data (Li et al., 2019b). To ensure data accuracy, we carefully checked all POI locations and attributes., removed redundant and duplicate points, and corrected incorrect points. After data preprocessing, the POI dataset fulfilled the requirements of this study.

Using the latest 2019 Changsha Metropolitan Area LUCC database (provided by the Department of Environment and Natural Resources of Changsha), we mapped the distribution of water bodies in the study area. To explore the difference across various types of water bodies, we differentiated the four categories of blue spaces as follows.

Xiangjiang River runs through Changsha and serves as an important blue landscape in the city. Since Changsha Government implemented the River Upgrade and Reconstruction Plan in 1995, the riverbank, scenario square, jogging track, and bicycle paths have been constructed. Liuyang River is the largest tributary of Xiangjiang River, which was also constructed with riverfront amenities and riverbank facilities in later years. Both river systems have nearby areas experiencing high population growths. Although the risk of river flooding exists, we assume that the presence of the rivers still has an overall positive effect on housing prices.

According to the National Wetland Park Evaluation Criteria of China, the 141-ha Yanghu National Wetland Park is the largest wetland in central China and is the only wetland in the Changsha urban area. It is a multi-functional wetland park integrating ecology, climate, and recreation, which has also become an important stop for migratory birds. Because of the significant microclimate regulation function of the wetland park, the temperature and humidity in the surrounding area of the park are more comfortable than in other urban areas. These ecological functions of wetland combined with its unique ecological landscape distinguish from other urban blue space. However, research on the impact of wetlands on housing prices has been very limited. Although one study (Mei et al., 2018) found wetlands to have a positive effect on housing prices, their impact, particularly on Chinese cities, is still largely uncertain.

For this study, the lakes were used to define the urban waterscape. Similar to previous studies on green spaces (Czembrowski and Kronenberg, 2016), we categorized the lakes based on size: small (<18,000 $m^2$); medium (18,000–200,000 $m^2$); and large (>200,000 $m^2$). We assumed all lake sizes would have a positive impact on house prices. We set other control variables according to the model in *Model Description*. For control and comparison, we added non-water green space variables, including the accessibility to the Yuelu Mountain National Forest Park and proximity to green space parks. Neighborhood and location variables were obtained through the points of interest (POI) of Baidu Map. The location variable pertains to the accessibility from the sample house to the central business district (CBD). All the accessibility variables were determined based on the walking distance using the road network in Changsha.

Structure variables, such as plot ratio, were calculated using the values of the indicators. Neighborhood attributes, some of which were defined using dummy variables (1 or 0), included proximity parameters to vital community facilities or amenities. For instance, for the proximity to subway station variable, 1 means that a subway station is within 1 km from the property. Four structure attributes, one location parameter, six neighborhood features, five blue space parameters, and two green space parameters were adopted in this study. We used community level as the research scale because most of the variables used corresponded to community attributes. The descriptions of these variables are shown in **Table 1**, and the descriptive statistics for blue space categories are in **Table 2**. All data were processed and calculated in ArcGIS and STATA.

# RESULTS

## OLS Regression and Quantile Effect

The OLS conditional mean results were consistent with our expectations of the variable. As shown in **Table 3**, most variables were found to affect house prices, significant at the 10% level. The walking distance from the location center to the nature landscape is inversely proportional to the housing price. Several Structural factors, such as the plot ratio and the age of the house, have negative effects on house prices, while proximity to important public facilities, such as quality schools and subway stations, have a positive impact. The coefficient of determination (R-squared) is 0.52, indicating that the model has moderate explanatory power.

The results of quantile regression (QR) under the same set of variables are also presented in **Table 3**. To explore the heterogeneity across quantiles, the full regression estimation of landscape variables were visualized (**Figure 2**) using the quantile 0.05 steps. The regression results of 19 quantiles from 0.05 to 0.95 are in solid lines, and the shaded area indicates the 95% confidence interval. The OLS estimate is represented by a dashed line, and the two dashed lines indicate the 95% confidence interval. The results suggest QR explains more changes in real estate prices at different quantiles, revealing the hidden utility of key variables on housing at different price ranges (not observed in OLS). Factors with strong impact on house prices, such as location, neighborhood, and urban blue, have varying effects at different price ranges (some have increasing trends while others have decreasing trends). QR results also showed that some non-significant factors were significant at different quantiles. For example, conditional mean coefficients of some structural and neighborhood variables were not significant overall but were significant for high-priced houses above the 80th quantile. Some variables with significant conditional mean, such as large lakes and Xiangjiang River, were not significant below the 40th and 70th quartile, respectively. With additional information obtained, the QR model was able to provide more depth on the impact of different variables on housing at different price points.

**TABLE 1 |** Variables used in the model.

| Category | Variable | Variable definition and measuring methods | Expected sign |
|---|---|---|---|
| Structure | Floor area | Score of the floor area of a property divided into 5 grates, 1 for less than 60 $m^2$, 2 for 60–90 $m^2$, 3 for 90–120 $m^2$, 4 for 120–144 $m^2$, and 5 for more than 144 $m^2$ | ? |
| | Plot ratio | Community plot ratio (below 1 is 1, 1–2 is 2, 2–3 is 3, 3–5 is 4, above 5 is 5) | – |
| | Property management (PM) | The level of the property management (no property management is 0, no level property management is 1, three-level and below property management is 2, two-level property management is 3, one level property management is 4, super grade property management is 5) | + |
| | Building age | Age of a house (year), 2019 minus the building year of a house | – |
| Neighborhoods | Quality middle-school (QMS) | Whether the residential district belongs to the quality middle school district of changsha (yes, 1, no, 0) | + |
| | Quality primary-school (QPS) | Whether the residential district belongs to the quality primary school district of changsha (yes, 1, no, 0) | + |
| | Subway station | Subway station within 1 km of property (yes, 1, no, 0) | + |
| | Living facilities (LF) | Hospital, GYM, bank, food market, and supermarket within 1 km from the community; each item is scored 1, total is 5 | + |
| Location | Distance to CBD wuyi square | Walking distance of the property from changsha CBD. (km) | + |
| Urban blue | Distance to xiangjiang river | Walking distance of the property from xiangjiang river, liuyang river, yanghu wetland park, large medium and small lakes. (km) | – |
| | D.liuyang river | | – |
| | D.wetland | | ? |
| | D.large lakes (LL) | | ? |
| | D.mideum lakes (ML) | | ? |
| | D.small lakes (SL) | | ? |
| Urban green | D.Yuelu Forest park (YL) | Walking distance of the property from the yuelu national forests, parks, and other non-water green spaces, respectively. (km) | ? |
| | D.non-water green (NWG) | | ? |

The spatial dependence on house prices was also found to affect the results of the study (Xiao et al., 2020). We used the Moran index and Getis-Ord Gi*(GO) index to detect the spatial effect of housing prices. The Moran index was 0.103, while the GO index was 0.37, indicating pronounced positive spatial effects.

## Spatial Quantile Regression

In the OLS regression, we added instrumental variables to perform a spatial two-stage regression. As shown in **Table 4** (first column), the adjusted conditional mean 2SLS R2 was 0.53, slightly higher than the OLS model. Columns 2–10 of **Table 4** present the 2SQR results and show all pseudo r2 values increasing. These results show that the improved model performed significantly better, similar to the findings of Hui, Liang, Yip (Hui et al., 2018).

**Figure 3** shows the complete spatial quantile regression estimation of blue facilities with 0.05 quantile increments. In column 1 of **Table 4**, most of the variables were significant at 10%. Most parameter estimates in the two regression models were very similar. For example, the location, education, and subway parameters were found to significantly affect housing prices in both models. The main difference between the two models was due to the OLS not controlling for the spatial effect. This caused variable overestimation where the coefficients for most of the OLS variables were higher than in the 2SLS. Liao and Wang (2012) also reached a similar conclusion. Combined with the significant effect of the spatial lag factor, we found that the spatial dependence had

**TABLE 2 |** Urban blue space categories in Changsha identified in our study.

| Category | Variables | Characteristics and typical use | Quantity | Total area (ha) | Mean area (ha) |
|---|---|---|---|---|---|
| River | Xiangjiang | Flood bank, scenario platform, landmark, leisure square, slow urban racing or walking road along the river, many benches, well-developed in terms of recreational infrastructure | 1 | 5,843 | — |
| | Liuyang | Flood bank, urban slow racing and walking road along the river, many benches, recreational facilities | 1 | 383 | — |
| Wetland | Yanghu | It is a multi-functional wetland park integrating ecology, culture, recreation, and education. on December 25, 2019, it officially became the "national wetland park". | 1 | 141 | — |
| Lakes | Large | Many recreational opportunities, including sports, although some look more natural, and are not so well-developed in terms of recreational infrastructure | 14 | 622.11 | 44.44 |
| | Medium | Opportunities for a typical walk in a park, usually well-maintained, with well-developed recreational infrastructure (benches, paths, playgrounds) | 23 | 294.28 | 12.79 |
| | Small | An embodied waterscape surrounding with some slow track and benches for a minimum walk (such as passing by), walking a dog, and sometimes sitting and enjoying a view | 47 | 53.32 | 1.13 |

**TABLE 3 |** OLS and QR model results.

| Variable | OLS (1) | 0.1 (2) | 0.2 (3) | 0.3 (4) | 0.4 (5) | 0.5 (6) | 0.6 (7) | 0.7 (8) | 0.8 (9) | 0.9 (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Floor area | −0.015 | 0.0126 | 0.00578 | −0.0059 | −0.0127 | −0.0228* | −0.0292** | −0.0239** | −0.0390*** | −0.0469*** |
| | [0.0107] | [0.0170] | [0.0122] | [0.0128] | [0.0124] | [0.0122] | [0.0123] | [0.0117] | [0.0133] | [0.0160] |
| Plot ratio | −0.0056** | −0.0042* | −0.0051*** | −0.0066*** | −0.0072** | −0.0078** | −0.0072* | −0.0013 | −0.0035 | −0.0056*** |
| | [0.0023] | [0.0019] | [0.0019] | [0.0019] | [0.0026] | [0.0028] | [0.0043] | [0.0045] | [0.0042] | [0.0033] |
| PM | 0.0129*** | 0.0207*** | 0.0205*** | 0.0198*** | 0.0158*** | 0.0138*** | 0.0148*** | 0.00982* | 0.00575 | 0.00663 |
| | [0.0036] | [0.0069] | [0.0060] | [0.0036] | [0.0027] | [0.0038] | [0.0040] | [0.0050] | [0.0068] | [0.0060] |
| Building age | −0.0016*** | −0.0027*** | −0.0026*** | −0.0028*** | −0.0025*** | −0.0022*** | −0.0018*** | −0.0016*** | −0.00080*** | −0.00057*** |
| | [0.0001] | [0.0002] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] |
| QMS | 0.0834*** | 0.0495*** | 0.0510*** | 0.0644*** | 0.0750*** | 0.0739*** | 0.0781*** | 0.0827*** | 0.0845*** | 0.0979*** |
| | [0.0109] | [0.0189] | [0.0167] | [0.0150] | [0.0171] | [0.0171] | [0.0131] | [0.0188] | [0.0175] | [0.0167] |
| QPS | 0.0294** | 0.0427** | 0.0222 | 0.0235* | 0.0177 | 0.0157 | 0.0274*** | 0.0389*** | 0.0468*** | 0.0299* |
| | [0.0118] | [0.0192] | [0.0232] | [0.0164] | [0.0187] | [0.0170] | [0.0167] | [0.0158] | [0.0152] | [0.0174] |
| Subway | 0.0481*** | 0.0606*** | 0.0417*** | 0.0451*** | 0.0500** | 0.0612*** | 0.0553*** | 0.0690*** | 0.0480** | 0.00169 |
| | [0.0129] | [0.0264] | [0.0211] | [0.0184] | [0.0197] | [0.0226] | [0.0180] | [0.0178] | [0.0216] | [0.0256] |
| LF | 0.0081*** | 0.0097*** | 0.0031*** | 0.0035*** | 0.0041*** | 0.0045*** | 0.0034*** | 0.0039*** | 0.0083*** | 0.0159*** |
| | [0.0003] | [0.0005] | [0.0006] | [0.0005] | [0.0004] | [0.0004] | [0.0003] | [0.0003] | [0.0004] | [0.0005] |
| Dcbd | −0.0625** | −0.103*** | −0.0675*** | −0.060** | −0.069* | −0.0705** | −0.0615*** | −0.0489** | −0.03295 | −0.054 |
| | [0.0290] | [0.0331] | [0.0395] | [0.0302] | [0.0297] | [0.0333] | [0.0323] | [0.0195] | [0.0505] | [0.0813] |
| D.Xriver | −0.0168*** | −0.0125 | −0.0085 | 0.00265 | 0.00066 | −0.0064 | −0.0087 | −0.013* | −0.0252** | −0.0280** |
| | [0.0056] | [0.0111] | [0.0083] | [0.0072] | [0.0061] | [0.0078] | [0.0072] | [0.0071] | [0.0097] | [0.0083] |
| D.Lriver | −0.0245** | −0.0357 | −0.0318** | −0.0315*** | −0.0263*** | −0.0223*** | −0.01385 | −0.0075 | −0.0025 | −0.0171 |
| | [0.0120] | [0.0224] | [0.0158] | [0.0119] | [0.0083] | [0.0061] | [0.0371] | [0.0435] | [0.0446] | [0.0537] |
| D.YHwetland | −0.0251*** | −0.00795 | −0.0183** | −0.0166* | −0.0235** | −0.0273*** | −0.0299*** | −0.0312*** | −0.0405*** | −0.0420*** |
| | [0.0049] | [0.0097] | [0.0101] | [0.0092] | [0.0095] | [0.0087] | [0.0086] | [0.0069] | [0.0067] | [0.0084] |
| D.LL | −0.0173*** | 0.00134 | −0.0156 | −0.0164 | −0.0205* | −0.0207** | −0.0223** | −0.0243*** | −0.0209** | −0.0319*** |
| | [0.0066] | [0.0181] | [0.0131] | [0.0121] | [0.0089] | [0.0069] | [0.0052] | [0.0068] | [0.0077] | [0.0105] |
| D.ML | −0.0143*** | −0.0344*** | −0.0156 | −0.0039 | −0.0014 | −0.004 | −0.0082 | −0.0066 | −0.0053 | −0.0019 |
| | [0.0045] | [0.0108] | [0.0098] | [0.0048] | [0.0043] | [0.0045] | [0.0055] | [0.0052] | [0.0048] | [0.0058] |
| D.SL | −0.0124 | −0.0214 | 0.00524 | −0.0001 | 0.0044 | 0.00533 | −0.0113 | −0.0118 | −0.0098 | −0.0421 |
| | [0.0167] | [0.0260] | [0.0288] | [0.0212] | [0.0241] | [0.0216] | [0.0139] | [0.0183] | [0.0293] | [0.0378] |
| D.YL | −0.0234*** | −0.0277*** | −0.0323*** | −0.0349*** | −0.0311*** | −0.0281*** | −0.0251*** | −0.0190** | −0.0182** | −0.0212** |
| | [0.0045] | [0.0068] | [0.0068] | [0.0056] | [0.0037] | [0.0052] | [0.0044] | [0.0078] | [0.0061] | [0.0073] |
| D.NWG | −0.0124*** | −0.0082*** | −0.0108*** | −0.0861*** | 0.0089*** | −0.0138*** | −0.0135*** | −0.0131*** | −0.0095*** | −0.0054 |
| | [0.0047] | [0.0021] | [0.0022] | [0.0015] | [0.0022] | [0.0024] | [0.0022] | [0.0026] | [0.0026] | [0.0038] |
| Cons | 9.603*** | 9.136*** | 9.503*** | 9.502*** | 9.596*** | 9.734*** | 9.738*** | 9.701*** | 9.836*** | 10.14*** |
| | [0.1021] | [0.2057] | [0.2139] | [0.1399] | [0.1168] | [0.1379] | [0.1226] | [0.1027] | [0.1305] | [0.1737] |
| Adj.$R^2$ | 0.519 | | | | | | | | | |
| Pseudo-$R^2$ | | 0.321 | 0.307 | 0.313 | 0.346 | 0.331 | 0.358 | 0.329 | 0.343 | 0.341 |

*$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$*

been corrected to a certain extent in the 2sls model and that the use of instrumental variables in the spatial 2sls model is needed to circumvent endogeneity.

Based on the regression coefficients of the spatial lag term, we found that the 30–90 quantile is significant at the 10% level, and the spatial dependence coefficient increases significantly with the increase of the quantile (from 0.013 to 0.087). This indicates a more pronounced spatial dependence in high-priced housing. The confidence interval above the 50th quantile exceeds the 2SLS conditional mean of 0.022, indicating that the 2SLS results may overestimate the spatial dependence of low-priced houses and underestimate the high-priced houses.

## Blue Space of Changsha

The results in **Table 4** show that the Xiangjiang River has a significant impact on property value, with distance negatively correlated with price. Each kilometer increase in the distance to Xiangjiang River correlated with a decline of about 1.72% in property value (i.e., 210 RMB per $m^2$). The capitalization effect of the Xiangjiang River (less than Yanghu wetland) was not consistent with its status as the center of the nature landscape in Changsha, which may due to increased flood risk. In 2017, the Xiangjiang River experienced flooding, resulting in economic losses for residents. Recent flood events can significantly alter people's perception and behavior towards flood risks (Bin and Polasky, 2004). With respect to the different housing prices, proximity to the Xiangjiang only has a significant impact above the 70th quartile, peaking at 90th (from 2.36% to 3.14%, each km). This suggests that nearness to the Xiangjiang river affects mainly house prices for high-value properties. This heterogeneity in proximity effects may be caused by better river scenarios and increased safety found mainly in high-value properties, similar to the conclusions of Wu Q. et al. (2020).

The Liuyang River is a tributary of the Xiangjiang River, flowing through the entire eastern part of the city, and is an

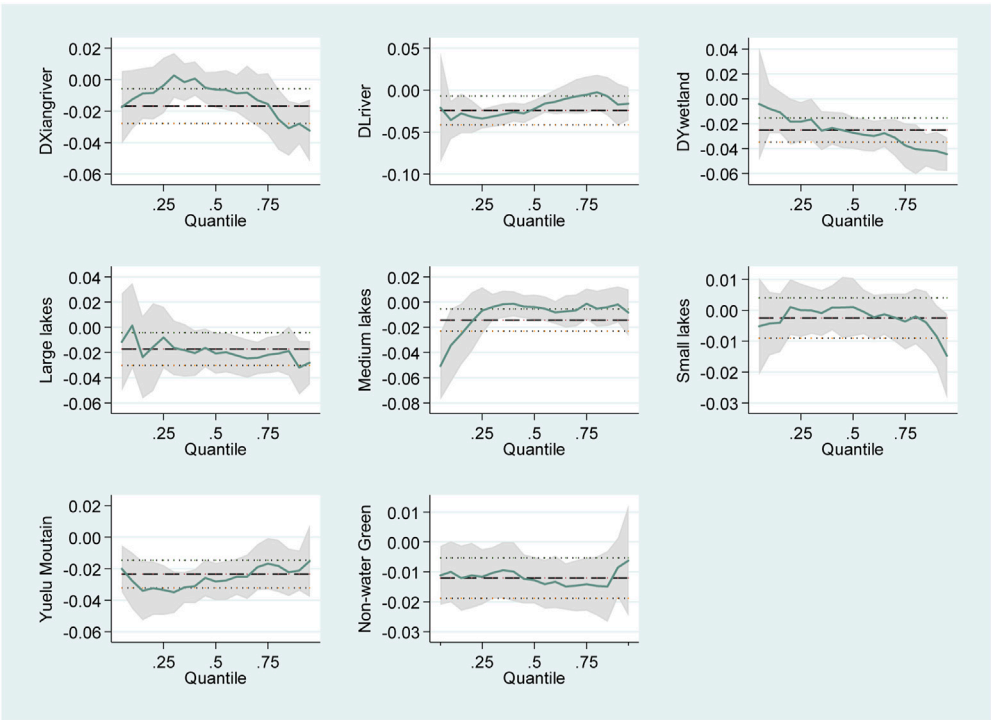



**FIGURE 2 |** Coefficients of blue spaces estimated by QR.

important blue space in Changsha. As shown in **Table 4**, proximity to the Liuyang River has a significant impact on housing prices (at the 10% level), with a 2SLS conditional mean regression coefficient of 2.36% value decline per km. The SQR results show that only low and medium-priced houses (i.e., 30th–50th quantile) are affected by proximity to this river.

Proximity to the Yanghu National Wetland Park was also found to significantly affect housing prices, similar to the findings of Mei et al. (2018), with every kilometer distance correlating with a 2.27% decline in housing prices (equivalent to 277 RMB per $m^2$). Similar to the effects of the Xiangjiang River, the SQR shows that the price effect caused by proximity to the Yanghu wetland increases as the quantile rises and is most pronounced for high-value properties (From 1.53% to 2.85%). Yanghu National Wetland Park was found to have a strong effect on housing prices among the different landscape factors in Changsha and has an estimated coefficient almost higher than any blue and green space. This may suggest wetlands are increasingly becoming more valued in cities.

While urban lake is the most common blue space in the city, our results found noticeable differences between varying sizes. The conditional mean coefficient of large lakes (LL) using 2sls was 1.66% value loss per km (i.e., 202 RMB per $m^2$), significant at the 1% level, and is higher than most landscape coefficients. LL has a significant impact on house prices above the 40th quantile. The estimated LL coefficients increased significantly, especially for mid-to-high-priced housing above the 60–70 range (2.31%, 2.35%, respectively), similar to the impact of wetland parks and the Xiangjiang River landscape. The impact of medium lakes (ML) was significant at the 10% level, resulting in a 1.45% value decline per km. The ML coefficient value was smaller than for large lakes and had a nonlinear trend across quantiles. According to the SQR, ML affected house prices below the 10th quantile but had no significant impact on those above the 10th quantile. Small lake (SL) is the commonest blue space in Changsha city. Although SL has a negative coefficient, the variable was not statistically significant. Housing prices seem to be unaffected by small water bodies, which is in sharp contrast to the impact on price of other kinds of blue space.

## Green Spaces in Changsha

For comparison, we used the non-water body green spaces as the control variables. As shown in **Table 4**, the conditional mean loss of property value for the Yuelu Forest Park was 2.03% per km, relatively similar to some previous studies of forest parks (Czembrowski and Kronenberg, 2016). The SQR results show that proximity to the Yuelu forest does not significantly influence house prices above the 80th quantile. For the other common green spaces, their impact was found to have a noticeable steady trend. The 2SLS results were significant at the 1% level, with a value loss of 1.21%. Most of the quantile coefficients for non-water green spaces were higher. These findings are consistent with existing studies (Liao and Wang, 2012), and the estimated accessibility coefficient is also comparatively similar.

**TABLE 4 |** 2Stage Least Squares and SQR model results.

| Variable | 2SLS (1) | 0.1 (2) | 0.2 (3) | 0.3 (4) | 0.4 (5) | 0.5 (6) | 0.6 (7) | 0.7 (8) | 0.8 (9) | 0.9 (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Floor area | −0.0153 | 0.0099 | 0.00319 | −0.00587 | −0.0131 | −0.0224* | −0.0273** | −0.0278** | −0.0436** | −0.0566** |
| | [0.0147] | [0.0138] | [0.0147] | [0.0120] | [0.0094] | [0.0098] | [0.0080] | [0.0125] | [0.0160] | [0.0219] |
| Plot ratio | −0.00549** | −0.00525** | −0.005** | −0.00677*** | −0.00704* | −0.00759 | −0.00684 | −0.00135 | −0.00263 | −0.0065 |
| | [0.0023] | [0.0021] | [0.0024] | [0.0027] | [0.0023] | [0.0041] | [0.0045] | [0.0048] | [0.0037] | [0.0040] |
| PM | 0.0129*** | 0.0212*** | 0.0211*** | 0.0199*** | 0.0159*** | 0.0143*** | 0.0143*** | 0.0103** | 0.00461 | −0.000734 |
| | [0.0036] | [0.0050] | [0.0045] | [0.0037] | [0.0044] | [0.0056] | [0.0053] | [0.0045] | [0.0072] | [0.0046] |
| Building age | −0.0015*** | −0.0026*** | −0.0024*** | −0.0026*** | −0.0024*** | −0.0021*** | −0.0020*** | −0.0017*** | −0.0009*** | −0.0006*** |
| | [0.0001] | [0.0002] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] | [0.0001] |
| QMS | 0.0817*** | 0.0623*** | 0.0478** | 0.0653*** | 0.0705*** | 0.0669*** | 0.0743*** | 0.0797*** | 0.0773*** | 0.0981*** |
| | [0.0111] | [0.0210] | [0.0145] | [0.0115] | [0.0136] | [0.0099] | [0.0127] | [0.0134] | [0.0124] | [0.0164] |
| QPS | 0.0280** | 0.0503*** | 0.0208 | 0.0244 | 0.0182 | 0.0104 | 0.0236* | 0.0348** | 0.0462** | 0.0342 |
| | [0.0119] | [0.0136] | [0.0140] | [0.0149] | [0.0134] | [0.0178] | [0.0138] | [0.0161] | [0.0156] | [0.0306] |
| Subway | 0.0472*** | 0.0534*** | 0.0470*** | 0.0417** | 0.0475*** | 0.0573*** | 0.0553*** | 0.0707*** | 0.0444* | 0.0165 |
| | [0.0129] | [0.0220] | [0.0178] | [0.0142] | [0.0196] | [0.0216] | [0.0183] | [0.0200] | [0.0205] | [0.0230] |
| LF | 0.0079*** | 0.0113*** | 0.0033*** | 0.0031*** | 0.0045*** | 0.0056*** | 0.0023*** | 0.0045*** | 0.0073*** | 0.0125*** |
| | [0.0003] | [0.0006] | [0.0007] | [0.0006] | [0.0005] | [0.0004] | [0.0003] | [0.0003] | [0.0004] | [0.0007] |
| D.CBD | −0.0575** | −0.1365*** | −0.079*** | −0.0595*** | −0.0672*** | −0.0557* | −0.0555** | −0.0381** | −0.00635 | −0.0417 |
| | [0.0267] | [0.0450] | [0.0271] | [0.0202] | [0.0266] | [0.0291] | [0.0274] | [0.0224] | [0.0507] | [0.0984] |
| D.Xriver | −0.0172*** | −0.0128 | −0.00706 | 0.0004 | −0.0007 | −0.0073 | −0.0066 | −0.0115 | −0.0236** | −0.0314*** |
| | [0.0057] | [0.0106] | [0.0084] | [0.0071] | [0.0086] | [0.0091] | [0.0097] | [0.0135] | [0.0132] | [0.0097] |
| D.Lriver | −0.0236** | −0.0312 | −0.0325 | −0.0221** | −0.0252*** | −0.0214*** | −0.0112 | −0.0062 | −0.0015 | −0.0083 |
| | [0.0118] | [0.0512] | [0.0529] | [0.0111] | [0.0104] | [0.0081] | [0.0261] | [0.0263] | [0.0363] | [0.0543] |
| D.YHwetland | −0.0227*** | −0.0125 | −0.0153* | −0.0171** | −0.0222*** | −0.0237*** | −0.0235*** | −0.0215* | −0.0260** | −0.0285*** |
| | [0.0056] | [0.0099] | [0.0081] | [0.0069] | [0.0067] | [0.0071] | [0.0095] | [0.0086] | [0.0071] | [0.0073] |
| D.LL | −0.0166** | −0.0074 | −0.0189 | −0.0152 | −0.0210** | −0.0179** | −0.0231*** | −0.0235*** | −0.0195*** | −0.0143* |
| | [0.0008] | [0.0011] | [0.0184] | [0.0181] | [0.0103] | [0.0087] | [0.0095] | [0.0093] | [0.0077] | [0.0081] |
| D.ML | −0.0145* | −0.0332** | −0.0159 | −0.0051 | −0.0013 | −0.0038 | −0.0071 | −0.0051 | −0.0049 | −0.0050 |
| | [0.0085] | [0.0137] | [0.0122] | [0.0066] | [0.0066] | [0.0047] | [0.0050] | [0.0048] | [0.0055] | [0.0041] |
| D.SL | −0.0128 | −0.0207 | 0.00136 | −0.0065 | 0.00368 | −0.0008 | −0.0107 | −0.0137 | −0.0218 | −0.029 |
| | [0.0167] | [0.0267] | [0.0216] | [0.0231] | [0.0220] | [0.0228] | [0.0227] | [0.0284] | [0.0284] | [0.0345] |
| D.YL | −0.0203*** | −0.0424*** | −0.0373*** | −0.0320*** | −0.0285*** | −0.0211*** | −0.0214*** | −0.0176* | −0.0066 | −0.00744 |
| | [0.0058] | [0.0084] | [0.0085] | [0.0073] | [0.0077] | [0.0072] | [0.0066] | [0.0103] | [0.0118] | [0.0111] |
| D.NWG | −0.0121*** | 0.0043 | −0.005 | −0.0082** | -0.0091** | −0.0155*** | −0.0164*** | −0.0183*** | −0.0114*** | −0.00907 |
| | [0.0017] | [0.0032] | [0.0034] | [0.0025] | [0.0014] | [0.0018] | [0.0021] | [0.0022] | [0.0030] | [0.0301] |
| WlnP | 0.0221** | −0.0028 | −0.0072 | 0.0136* | 0.0152* | 0.0443** | 0.0468** | 0.0447*** | 0.0809*** | 0.0874*** |
| | [0.0121] | [0.0076] | [0.0084] | [0.0081] | [0.0091] | [0.0223] | [0.0232] | [0.0192] | [0.0231] | [0.0157] |
| Cons | 9.364*** | 10.18*** | 9.799*** | 9.377*** | 9.444*** | 9.236*** | 9.273*** | 9.214*** | 8.895*** | 9.047*** |
| | [0.2961] | [0.3726] | [0.4683] | [0.3900] | [0.3734] | [0.3357] | [0.3232] | [0.3933] | [0.6152] | [0.6469] |
| Adj.$R^2$ | 0.531 | | | | | | | | | |
| Pseudo-$R^2$ | | 0.335 | 0.311 | 0.329 | 0.352 | 0.341 | 0.367 | 0.339 | 0.361 | 0.353 |

****p < 0.01; **p < 0.05; *p < 0.10.*

## Other Key Variables

The study also found that the housing price gradient can diversify with the distance from the CBD (Wen et al., 2014). By 2SLS, each kilometer away from the CBD leads to a 5.75% decay of house prices. The SQR results showed that the price for accessibility to the CBD decreases for higher quantiles (from 13.65% to 3.81%, each km). Low- and medium-priced properties were influenced more strongly by CBD than high-priced houses. Neighborhood variables, such as proximity to quality primary and middle education facilities (2.80% and 8.17%, respectively) and nearness to a subway station (4.72%), were also found to have significant impact on housing prices. These results suggest education-related factors are becoming important factors affecting property house prices in China.

## Density Test of Urban Blue Accessibility

As suggested by Coulson and McMillen (2007), the distribution of house prices varying to the explanatory variables could exam the heterogeneity across dependent variable quantiles. If a change in the value of the independent variable (distance) causes a change in the dependent variable's distribution, this would mean that different quantiles would have varying responses to the independent variable and that heteroscedasticity exists across different quantiles. To investigate this effect, we drew the kernel density curve to compare the distribution of prices among varying accessibility to various urban blue parameters. The results are presented in **Figure 4**. For the Xiangjiang River, the kernel curve changes its kurtosis as the distance increases. Similarly, the parameters Liuyang river, Yanghu wetland, large lakes, and middle lakes have a significant change in the distribution of housing prices when the value of accessibility independent variable changing. Most have changes in the kurtosis (right tail of the distribution), with some increase in variance distribution, coincident with the SQR results. In contrast, the kernel density curve for small lakes has little distributional change, which means no heteroscedasticity in this variable.

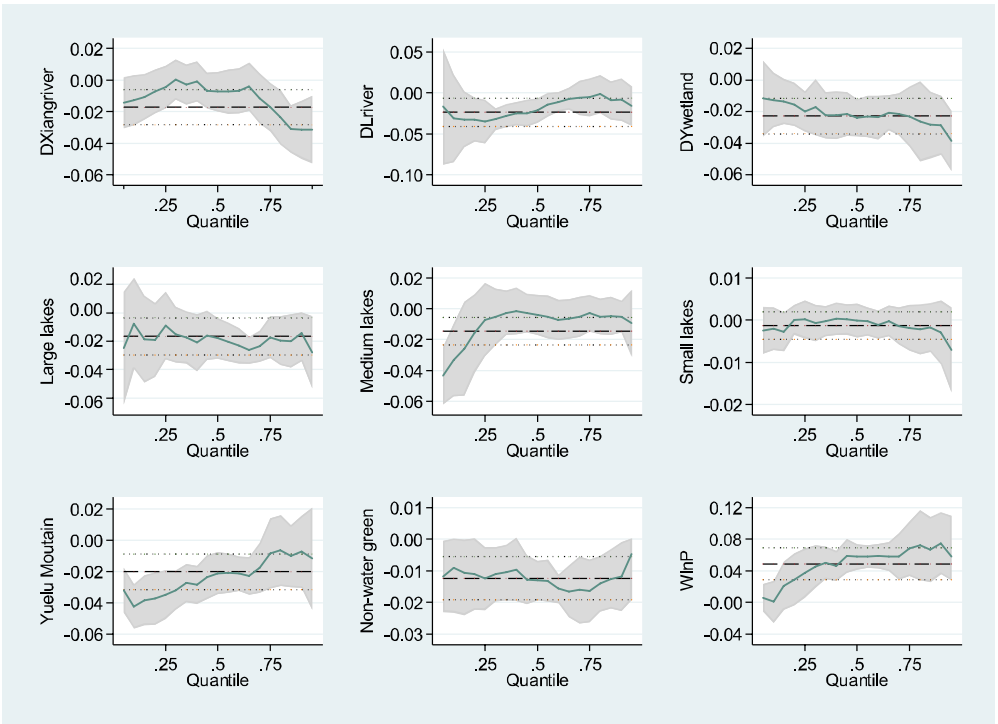



**FIGURE 3 |** Coefficients of blue spaces estimated by 2SQR.

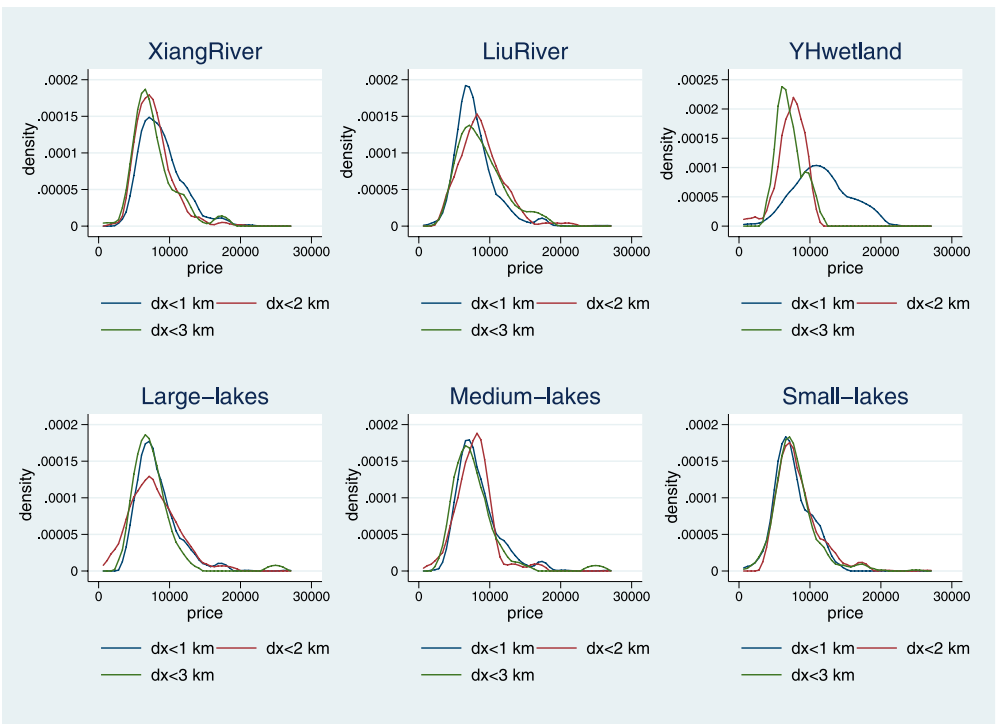



**FIGURE 4 |** Shifts in the distribution of house prices.

# DISCUSSION

## Effects of Blue Space Type and Size

Accessibility to wetlands, lakes, and rivers all have a positive impact on housing prices, which varies for different types and sizes of blue spaces. Accessibility to wetland parks was found to have the quite strong effect among blue space features, possibly because wetlands provide recreation, biodiversity, ecological conservation, and other ecosystem services that urban residents value. While Xiangjiang River is an important blue landscape in Changsha, its impact on housing prices was less than what we anticipated, coming second to wetlands, possibly due to increased flooding risks. The impact of the Xiangjiang River's tributary, the Liuyang River, was close to the YangHu wetland. For lakes, which are the blue spaces most frequently used by residents, their accessibility significantly affected housing prices, which varied considerably based on lake size. Accessibility to medium-sized lakes was found to have smaller impact than large-sized lakes, while small-sized lakes did not significantly affect property value.

The empirical results suggest that the larger the size of the water body, the more it affects the surrounding house prices. One possible reason is that larger water spaces are able to provide more (or more perceivable) ecosystem services. One clear example is its reduction of urban heat islands, which is directly related to the size and fragmentation of the water body (Yu et al., 2020). Another possible reason is the water body's frequency of use. Small urban blue spaces are usually embedded in residential communities, which often means lower transportation costs to use. According to the existing literature, a higher transportation cost to a nature space, a more sensitive of the residents to the perceived walking distance to arrive (Nielsen and Hansen, 2007; Voelker et al., 2018). This explained in part of the independent variable (accessibility of small lakes) has a non-significant effect on the house price because residents' perceived walking distance to these embedded blue spaces is not as sensitive as the large-sized lakes, which are usually located farther from the residential community with higher transportation costs. Moreover, most small blue spaces are located in the city's core, where highly dense surrounding communities (commonly high-rise buildings) obstruct their visibility and limit their effect, which further limits the effect of the little blue on house prices. Besides, awareness and familiarity with small and medium urban blue spaces are considerably lower than key water bodies, minimizing their impressions on homebuyers and limiting their effect on property value.

## Quantile Effect

In the SQR results, the effect of blue spaces varied considerably across different quantiles. The impact of accessibility to the Xiangjiang River, large lakes, and Yanghu wetland increased linearly with property value, such that blue spaces had greater effect on high-priced houses. In comparison, proximity effects to medium-sized lake parks and the Liuyang River were more pronounced on low- and middle-quantile properties and had a nonlinear trend. Various possible reasons may explain why the proximity effect of these three urban blue spaces intensifies as property value increases. First, high-priced properties generally have better viewshed access, often at higher altitudes, better orientation, and lower flood risks (Wu P.-I. et al., 2020). Second, buyers of high-priced houses are highly concerned about their physical and mental health. Large-scale urban blue spaces are able to address such concerns, providing more perceivable ecosystem services that generate higher marginal utility. Also, better marketing and advertising strategies are able to highlight the attributes of blue spaces, which are often used to entice high-income buyers. This makes higher income groups more aware and perceptive of amenities generated by blue spaces. According to a previous study, familiarity with a landscape positively correlates with the purchase premium (LaRiviere et al., 2014). Homebuyers may even be led to believe that properties around the famous water bodies have higher investment values. In contrast, medium-sized lakes and the Liuyang River have limited appeal to high-income homebuyers due to their smaller extent and less perceivable ecological amenities. High-priced houses are more likely to be located in better environments since high-income residents use their own cars to commute. In contrast, most low- and middle-income homebuyers who use public transport to commute do not choose to live near a large blue space far from the city area, preferring medium-sized blue spaces that are usually distributed in downtown areas, which are more economic for a commute. Therefore, the quantile effects of medium-size blue thus show a lower impact on the high-priced houses while it is significant to low-priced houses.

## Outlook for Further Research

Our research found that the type and size of blue spaces have varying influences on house prices. Our results also showed significant differences in the impact on property value across different housing price quantiles. Using the framework presented in *Data and Methodology*, we assumed that the effect on house prices depends mainly on the ecosystem services provided by the various blue spaces. However, the impact of specific ecosystem services can not be quantified in this study since buyers do not naturally have the cognition or willingness to distinguish between different ecosystem services (Czembrowski and Kronenberg, 2016).

While the SQR results showed that buyers of high-priced houses might be more concerned about the value of these ecosystem services, it does not allow us to explore the impact of individual ecosystem services. SQR also does not analyze how homebuyers consider different ecosystem services offered by different types and sizes of urban blue spaces. Rather, our analysis focused on evaluating the effects of different types and sizes of blue spaces at various quantiles. Hedonic pricing is rarely used to accurately quantify the price effects of individual ecosystem services. To value the impact of specific ecosystem services, more studies are needed using other valuation techniques that are able to explicitly link value effects to particular ecosystem services. Follow-up research can further adopt more metadata, such as questionnaire surveys, and use

other quantitative methods to further explore the economic value of different ecosystem serves of blue spaces.

## CONCLUSION AND RECOMMENDATION

In this era of rising awareness and demands for healthy living standards, access to urban blue spaces is now being demanded by more and more urban dwellers. However, research on how access and proximity of different urban blue spaces affect property value has been largely limited. To address this current research gap, this study employed a spatial quantile hedonic model to evaluate the proximity effects of different types of blue spaces on property prices across different conditional distributions.

Spatial dependence in property value was found to be significant across price quantiles. This suggests that using spatial lag term in hedonic research is needed in order to prevent biased outcomes. We also found that the positive effect of spatial dependence increases in higher price quantiles. High-priced houses were found to cluster together, which confirms the findings of other urban hedonic studies. The results suggest rich homebuyers concern more about scarce public facilities and ecological resources, with a feature of the housing market in China's big cities, which may be one of the cause led to the spatial aggregation of high-priced house.

The 2SLS results show that the type and size of urban blue spaces influence the capitalization effect on property value. Despite possible increases in the flood risks, proximity to rivers was found to have a positive impact on house prices. In terms of proximity effects of the Xiangjiang River, property value declined by about 1.72% per kilometer of walking distance. The proximity effect of the Liuyang River was also found to significantly impact house prices, each kilometer of increased walking distance from it results in about a 2.36% decline in house prices. Wetland park is a new kind of environmental amenity that has recently emerged across many Chinese cities. In our results, property values declined by about 2.27% per kilometer of increased distance from the Yanghu national wetland park. We also found similar proximity effects of large and medium-size urban lakes, with increased distance resulting in house price declines of 1.66% and 1.45% per kilometer, respectively. These proximity effects were found to generally increase with the size of the blue space, possibly because larger blue spaces provide increased and more perceivable ecosystem services.

Under the quantile setting, the SQR model provided more details on blue space effects than the 2SLS. Only proximity effects of the Yanghu wetland were significant almost across all house price ranges in QR and SQR. Since the values of most parameters of blue spaces in 2SLS are all significant, this means property values in different price ranges affected by blue spaces are overestimated. The SQR results suggest that the Xiangjiang River and large lakes have significant influence only with high-value properties and that the 2SLS overestimated the effect of these variables in low-priced houses. In contrast, only low- and medium-priced houses were found significant to the accessibility to medium-sized lakes and the Liuyang River, which means the effect on high-priced houses of these two variables have been overestimated in 2SLS. Different variables also showed varying SQR trends for different price quantiles. The proximity effects of Xiangjiang river, Yanghu wetland, and large lakes are not significant at low-value properties but are significant and increases considerably for higher-priced houses. To some extent, the results suggest that the greater wealth of the homebuyer, the stronger the marginal willingness to pay for accessibility to core blue spaces. In other blue spaces, like the Liuyang River and middle-size lakes, proximity effects had nonlinear trends across different quantiles. Residents in low-priced houses might pay more for the accessibility to these blue spaces.

Based on the findings of this study, we conclude some recommendations for future urban planning. First, more small blue spaces that are effectively embedded in residential communities should be developed. Small blue space should be encouraged owing to its convenience in daily life and no significant influence on house price, which means improving residential well-being without adding to the cost of buying a house. Second, improving the service efficiency of the key blue spaces through better transport network systems and more balanced residential development density planning. Transport departments should improve the accessibility of blue space by renovating some unreasonable road networks and reducing the interchange time. It is also important to construct subway stations near famous blue spaces, which greatly shortens the commuting distance. Besides, since high-priced houses are more affected by the key blue spaces, developers tend to build higher-height houses with better views closer to blue. Therefore, restricting the residential density and building height, setting reasonable prohibit development lines of key blue spaces is necessary to ensure the accessibility and visibility of the water landscapes. In short, the government should create a rational opportunity to enable more residents to enjoy the famous blue spaces amenity value, rather than the monopoly of urban blue by rich homebuyers. Last but not the least, It is important to regulate the land revenue distribution mechanism. As results revealed, the key urban blue has a significant effect on housing price, which most benefit the property developers. However, the increment value created by the key blue spaces is usually invested by the local government. Therefore, it is more reasonable not to gather land rent one-time charging but adjust incremental land revenue according to the pricing effect of key urban blue spaces, which improves the fiscal pattern of public services in cities.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

Conceptualization, methodology, software, and writing: HT and HQ; OX acquired the data; YD analyzed and interpreted the data for the work.

## FUNDING

This research was funded by the Major Program of National Social Science Foundation of China (18ZDAA040) and Natural Science Foundation of Hunan Province (2020JJ5111), Project of Hunan Social Science Achievement Appraisal Committee (XSP20YBZ081), Young Teacher Innovation Drive Project of Hunan University Technology and Business (19QD04).

## REFERENCES

Anderson, S. T., and West, S. E. (2006). Open Space, Residential Property Values, and Spatial Context. *Reg. Sci. Urban Econ.* 36 (6), 773–789. doi:10.1016/j.regsciurbeco.2006.03.007

Anselin, L. (2003). Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *Int. Reg. Sci. Rev.* 26, 153–166. doi:10.1177/0160017602250972

Bin, O., and Polasky, S. (2004). Effects of Flood Hazards on Property Values: Evidence before and after Hurricane Floyd. *Land Econ.* 80, 490–500. doi:10.2307/3655805

Brander, L. M., and Koetse, M. J. (2011). The Value of Urban Open Space: Meta-Analyses of Contingent Valuation and Hedonic Pricing Results. *J. Environ. Manage.* 92 (10), 2763–2773. doi:10.1016/j.jenvman.2011.06.019

Chernozhukov, V., and Hansen, C. (2006). Instrumental Quantile Regression Inference for Structural and Treatment Effect Models. *J. Econom.* 132 (2), 491–525. doi:10.1016/j.jeconom.2005.02.009

Conway, D., Li, C. Q., Wolch, J., Kahle, C., and Jerrett, M. (2010). A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace on Residential Property Values. *J. Real Estate Finan Econ.* 41 (2), 150–169. doi:10.1007/s11146-008-9159-6

Coulson, N. E., and McMillen, D. P. (2007). The Dynamics of Intraurban Quantile House Price Indexes. *Urban Stud.* 44 (8), 1517–1537. doi:10.1080/00420980701373446

Czembrowski, P., and Kronenberg, J. (2016). Hedonic Pricing and Different Urban Green Space Types and Sizes: Insights into the Discussion on Valuing Ecosystem Services. *Landscape Urban Plann.* 146, 11–19. doi:10.1016/j.landurbplan.2015.10.005

Ebru, Ç., and Eban, A. (2011). Determinants of House Prices in Istanbul: A Quantile Regression Approach. *Qual. Quant.* 45 (2), 305–317. doi:10.1007/s11135-009-9296-x

Garrett, J. K., White, M. P., Huang, J., Ng, S., Hui, Z., Leung, C., et al. (2019). Urban Blue Space and Health and Wellbeing in Hong Kong: Results from a Survey of Older Adults. *Health Place* 55, 100–110. doi:10.1016/j.healthplace.2018.11.003

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A., et al. (2015). Mental Health Benefits of Long-Term Exposure to Residential Green and Blue Spaces: A Systematic Review. *Int J Environ Res Public Health.* 12 (4), 4354–4379. doi:10.3390/ijerph120404354

Gómez-Baggethun, E., and Barton, D. N. (2013). Classifying and Valuing Ecosystem Services for Urban Planning. *Ecol. Econ.* 86, 235–245. doi:10.1016/j.ecolecon.2012.08.019

Herbert, D. T., and Thomas, C. J. (1998). School Performance, League Tables and Social Geography. *Appl. Geogr.* 18 (3), 199–223. doi:10.1016/S0143-6228(98)00015-0

Huang, Z., Chen, R., Xu, D., and Zhou, W. (2017). Spatial and Hedonic Analysis of Housing Prices in Shanghai. *Habitat Int.* 67, 69–78. doi:10.1016/j.habitatint.2017.07.002

Hui, E. C. M., Liang, C., and Yip, T. L. (2018). Impact of Semi-obnoxious Facilities and Urban Renewal Strategy on Subdivided Units. *Appl. Geogr.* 91, 144–155. doi:10.1016/j.apgeog.2017.12.020

Jim, C. Y., and Chen, W. Y. (2006). Impacts of Urban Environmental Elements on Residential Housing Prices in Guangzhou (China). *Landscape Urban Plann.* 78 (4), 422–434. doi:10.1016/j.landurbplan.2005.12.003

Kang, H.-H., and Liu, S.-B. (2014). The Impact of the 2008 Financial Crisis on Housing Prices in China and Taiwan: A Quantile Regression Analysis. *Econ. Model.* 42, 356–362. doi:10.1016/j.econmod.2014.07.018

Kaplan, R., and Kaplan, S. (1989). *The Experience of Nature: A Psychological Perspective*. New York: Cambridge University Press, xii-340. doi:10.1037/030621

Kelejian, H. H., and Prucha, I. R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *Int. Econ. Rev.* 40 (2), 509–533. doi:10.1111/1468-2354.00027

Kim, T. H., and Muller, C. (2004). Two-stage Quantile Regression when the First Stage Is Based on Quantile Regression. *Econom. J.* 7 (1), 218–231. doi:10.1111/j.1368-423x.2004.00128.x

Kong, F., Yin, H., and Nakagoshi, N. (2007). Using GIS and Landscape Metrics in the Hedonic Price Modeling of the Amenity Value of Urban Green Space: A Case Study in Jinan City, China. *Landscape Urban Plann.* 79 (3–4), 240–252. doi:10.1016/j.landurbplan.2006.02.013

Kostov, P. (2009). A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices. *Spat. Econ. Anal.* 4 (1), 53–72. doi:10.1080/17421770802625957

Kuethe, T. H., and Keeney, R. (2012). Environmental Externalities and Residential Property Values: Externalized Costs along the House Price Distribution. *Land Econ.* 88 (2), 241–250. doi:10.3368/le.88.2.241

LaRiviere, J., Czajkowski, M., Hanley, N., Aanesen, M., Falk-Petersen, J., and Tinch, D. (2014). The Value of Familiarity: Effects of Knowledge and Objective Signals on Willingness to Pay for a Public Good. *J. Environ. Econ. Management.* 68 (2), 376–389. doi:10.1016/j.jeem.2014.07.004

Larson, E. K., and Perrings, C. (2013). The Value of Water-Related Amenities in an Arid City: The Case of the Phoenix Metropolitan Area. *Landscape Urban Plann.* 109 (1), 45–55. doi:10.1016/j.landurbplan.2012.10.008

LeSage, J. P., and Pace, R. K. (2004). Models for Spatially Dependent Missing Data. *J. Real Estate Finance Econ.* 29 (2), 233–254. doi:10.1023/B:REAL.0000035312.82241.e4

Li, H., Wei, Y. D., Wu, Y., and Tian, G. (2019a). Analyzing Housing Prices in Shanghai with Open Data: Amenity, Accessibility and Urban Structure. *Cities* 91, 165–179. doi:10.1016/j.cities.2018.11.016

Li, H., Wei, Y. D., and Wu, Y. (2019b). Urban Amenity, Human Capital and Employment Distribution in Shanghai. *Habitat Int.* 91, 102025. doi:10.1016/j.habitatint.2019.102025

Lianyong, W., and Eagles, P. F. J. (2009). Some Theoretical Considerations: From Landscape Ecology to Waterscape Ecology. *Acta Ecologica Sinica.* 29 (3), 176–181. doi:10.1016/j.chnaes.2009.07.006

Liao, W.-C., and Wang, X. (2012). Hedonic House Prices and Spatial Quantile Regression. *J. Housing Econ.* 21 (1), 16–27. doi:10.1016/j.jhe.2011.11.001

Liebelt, V., Bartke, S., and Schwarz, N. (2019). Urban Green Spaces and Housing Prices: An Alternative Perspective. *Sustainability.* 11 (13), 3707. doi:10.3390/su11133707

Liu, T., Hu, W., Song, Y., and Zhang, A. (2020). Exploring Spillover Effects of Ecological Lands: A Spatial Multilevel Hedonic Price Model of the Housing Market in Wuhan, China. *Ecol. Econ.* 170, 106568. doi:10.1016/j.ecolecon.2019.106568

Mak, S., Choy, L., and Ho, W. (2010). Quantile Regression Estimates of Hong Kong Real Estate Prices. *Urban Stud.* 47 (11), 2461–2472. doi:10.1177/0042098009359032

McMillen, D. P. (2012). *Quantile Regression for Spatial Data*. Berlin: Springer Science & Business Media.

Mei, Y., Sohngen, B., and Babb, T. (2018). Valuing Urban Wetland Quality with Hedonic Price Model. *Ecol. Indicators* 84, 535–545. doi:10.1016/j.ecolind.2017.09.022

Mueller, J. M., and Loomis, J. B. (2014). Does the Estimated Impact of Wildfires Vary with the Housing Price Distribution? A Quantile Regression Approach. *Land Use Policy.* 41, 121–127. doi:10.1016/j.landusepol.2014.05.008

Nielsen, T. S., and Hansen, K. B. (2007). Do green Areas Affect Health? Results from a Danish Survey on the Use of Green Areas and Health Indicators. *Health Place* 13 (4), 839–850. doi:10.1016/j.healthplace.2007.02.001

Rajapaksa, D., Wilson, C., Hoang, V.-N., Lee, B., and Managi, S. (2017). Who Responds More to Environmental Amenities and Dis-Amenities?. *Land Use Policy.* 62, 151–158. doi:10.1016/j.landusepol.2016.12.029

Roebeling, P., Saraiva, M., Palla, A., Gnecco, I., Teotonio, C., Fidelis, T., et al. (2017). Assessing the Socio-Economic Impacts of Green/blue Space, Urban Residential and Road Infrastructure Projects in the Confluence (Lyon): A Hedonic Pricing Simulation Approach. *J. Environ. Plann. Management.* 60 (3), 482–499. doi:10.1080/09640568.2016.1162138

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Polit. Economy* 82 (1), 34–55. doi:10.1086/260169

Sander, H. A., and Haight, R. G. (2012). Estimating the Economic Value of Cultural Ecosystem Services in an Urbanizing Area Using Hedonic Pricing. *J. Environ. Manage.* 113, 194–205. doi:10.1016/j.jenvman.2012.08.031

Sander, H. A., and Zhao, C. (2015). Urban Green and Blue: Who Values what and where?. *Land Use Policy*. 42, 194–209. doi:10.1016/j.landusepol.2014.07.021

Targino, A. C., Coraiola, G. C., and Krecl, P. (2019). Green or Blue Spaces? Assessment of the Effectiveness and Costs to Mitigate the Urban Heat Island in a Latin American City. *Theor. Appl. Climatol.* 136 (3–4), 971–984. doi:10.1007/s00704-018-2534-1

Voelker, S., Heiler, A., Pollmann, T., Classen, T., Hornberg, C., and Kistemann, T. (2018). Do perceived Walking Distance to and Use of Urban Blue Spaces Affect Self-Reported Physical and Mental Health?. *Urban For. Urban Green.* 29, 1–9. doi:10.1016/j.ufug.2017.10.014

Voelker, S., and Kistemann, T. (2015). Developing the Urban Blue: Comparative Health Responses to Blue and Green Urban Open Spaces in Germany. *Health Place*. 35, 196–205. doi:10.1016/j.healthplace.2014.10.015

Wen, H., Zhang, Y., and Zhang, L. (2014). Do educational Facilities Affect Housing Price? an Empirical Study in Hangzhou, China. *Habitat Int.* 42, 155–163. doi:10.1016/j.habitatint.2013.12.004

Wu, P.-I., Chen, Y., and Liou, J.-L. (2020). Housing Property along Riverbanks in Taipei, Taiwan: A Spatial Quantile Modelling of Landscape Benefits and Flooding Losses. *Environ. Dev. Sustain.* 23, 2404–2438. doi:10.1007/s10668-020-00680-7

Wu, Q., Cheng, Z. G., and Yang, D. X. (2020). Social Stratification and Housing Inequality in Transitional Urban China. *Contemp. Soc. Sci.* 537, 1, 16 16. Available at: https://ideas.repec.org/a/eee/phsmap/v537y2020ics0378437119314773.html. doi:10.1080/21582041.2020.1797148

Wu, W-p., Zeng, W-k., Gong, S-w., and Chen, Z-g. (2021). Does Energy Poverty Reduce Rural Labor Wages? Evidence from China's Rural Household Survey. *Front. Energ. Res.* 9, 670026, 2021 . Available at: https://ideas.repec.org/a/eee/phsmap/v537y2020ics0378437119314773.html. doi:10.3389/fenrg.2021.670026

Xiao, Y., Hui, E. C. M., and Wen, H. (2020). The Housing Market Impacts of Human Activities in Public Spaces: The Case of the Square Dancing. *Urban For. Urban Green.* 54, 126769. doi:10.1016/j.ufug.2020.126769

Yu, Z., Yang, G., Zuo, S., Jørgensen, G., Koga, M., and Vejre, H. (2020). Critical Review on the Cooling Effect of Urban Blue-Green Space: A Threshold-Size Perspective. *Urban For. Urban Green.* 49, 126630. doi:10.1016/j.ufug.2020.126630

Zhang, L. (2016). Flood Hazards Impact on Neighborhood House Prices: A Spatial Quantile Regression Analysis. *Reg. Sci. Urban Econ.* 60, 12–19. doi:10.1016/j.regsciurbeco.2016.06.005

Zietz, J., Zietz, E. N., and Sirmans, G. S. (2008). Determinants of House Prices: A Quantile Regression Approach. *J. Real Estate Finance Econ.* 37 (4), 317–333. doi:10.1007/s11146-007-9053-7

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Tuofu, Qingyun, Dongxiao and Xiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 21 May 2021
doi: 10.3389/fenvs.2021.645010

# Does the New Urbanization Influence Air Quality in China?

*Yan Wu*[1], *Zhixiang Cui*[2] *and Cong Hu*[3]*

[1]School of Economics, Beijing Technology and Business University, Beijing, China, [2]School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China, [3]School of International Development and Cooperation, University of International Business and Economics, Beijing, China

Previous studies have empirically investigated the influence of China's urbanization on atmosphere pollution, the findings in the literature are however controversial and inconclusive across regions, data, and methodologies. This study uses the city-level panel data, 113 key cities of environment protection from the Ministry of Ecology and Environment, covering most of the provinces in China for the period 2013–2017 to investigate the different impacts of the new urbanization pilot policy on air quality and related air pollutants including six major pollutant sources, which are $PM_{2.5}$, $PM_{10}$, $SO_2$, CO, $NO_2$, and $O_3$. The study finds that, first, based on the difference-in-difference (DID) method, the new urbanization on average tends to improve the air quality in the pilot cities. Second, based on the quantile DID method, the new urbanization tends to improve the air quality in the lower air quality quantiles (0.1–0.6); however, it has no significant impact in the higher air quality quantiles (0.7–0.9). Third, the impacts of the new urbanization on the air quality vary among different energy-related air pollutants. The new urbanization pilot policy tends to restrain $SO_2$, $PM_{10}$, and $PM_{2.5}$, increase CO and $O_3$ and has no impact on $NO_2$. The results indicate that China should pay more attention to promote green consumption and new energy applications and increase urban construction efficiency to further reducing air pollutions in the new urbanization process.

**Keywords: air quality, air pollutants, difference in difference, new urbanization strategy, China**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, China*

***Reviewed by:***
*Yu Hao,*
*Beijing Institute of Technology, China*
*Zhen Qi,*
*Guangzhou University, China*

***Correspondence:**
*Cong Hu*
*bshida@126.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *22 December 2020*
***Accepted:*** *03 May 2021*
***Published:*** *21 May 2021*

***Citation:***
*Wu Y, Cui Z and Hu C (2021) Does the New Urbanization Influence Air Quality in China? Front. Environ. Sci. 9:645010. doi: 10.3389/fenvs.2021.645010*

## INTRODUCTION

As China entered a stage of rapid urbanization with a large increase in urban energy demand, urban air pollution, especially smog caused by $PM_{2.5}$ and $PM_{10}$, had become increasingly serious and attracted wide attention from the public. In 2014, the Chinese government announced the New Urbanization Plan (2014–2020), which aimed to promote human-oriented and sustainable urban development in China, thus distinguishing itself from the past urbanization model of excessive concentration of population, destruction of the environment and excessive consumption of resources.

The new urbanization is a national strategy of China. Chinese Premier Li Keqiang said that one of the core objectives of this strategy is to integrate the concept and principles of ecological civilization into the whole urbanization process. It is not at the expense of destroying the ecology and polluting the environment, but promotes the economic, social and environmental balanced development, and pursues a new path of intensive, intelligent, green and low-carbon urbanization. The greatest feature of this plan is that it seeks to address environmental and social issues in the process of urbanization, while also laying the foundation for economic sustainable development through environmental

constraints. Until now, the Chinese National Development and Reform Commission has implemented the new type of urbanization pilot policy in 61 cities at the prefecture-level in 2015 and added 19 more in December 2016, which covered most of the regions of China, involving 31 provinces and greatly boosting Chinese urbanization.

Whether China's new urbanization conform to the environmental Kuznets curve (Grossman and Krueger, 1991), causing pollution to the atmosphere? Some studies have empirically investigated the influence of China's urbanization on atmosphere pollution, the findings in the literature are however controversial and inconclusive across regions, data, and methodologies.

First, many studies proposed that China's urbanization might bring some environmental issues in addition to economic growth. On the one hand, some studies have focused on the impact of urbanization on air pollution in certain areas of China. For example, based on the $PM_{2.5}$ concentration data during 2000–2016 in the Chinese Yangtze River Delta region, Yang et al. found that there was a positive correlation between urbanization and $PM_{2.5}$ concentrations (Yang et al., 2020). Based on the case study in Pearl River Delta of Guangdong, China, Zhang et al. found that the ozone precursors, $NO_x$ and VOC emission from vehicles, industries and power plants increased region-wide, and in particular in the newly formed prefectural municipalities (Zhang et al., 2011). Based on high-resolution urban climate simulations for evaluating the impact of urbanization on air stagnation, Li et al. found that urbanization worsened the problem of air stagnation in Shenzhen of China (Li et al., 2019). Taking the Beijing-Tianjin-Hebei urban agglomeration of China as an example, Du et al. found that economic urbanization had a much stronger influence on air quality than land urbanization or population urbanization (Du et al., 2019). On the other hand, some studies investigated the relationship between urbanization and air pollution in China as a whole. For example, based on emission inventory data from 2014, gathered from 289 cities, Fang et al. found that urbanization has played an important negative role in determining air quality in Chinese cities. The population, urbanization rate, automobile density, and the proportion of secondary industry were all found to have had a significant influence on air quality (Fang et al., 2015). By examining and comparing $PM_{2.5}$ concentrations in urban and the surrounding regions, significant positive correlations were found between $PM_{2.5}$ and urban population, and between $PM_{2.5}$ and urban second industry fraction, suggesting that urbanization had a considerable impact on $PM_{2.5}$ concentrations (Han et al., 2014). Exploring the spatiotemporal profile of $PM_{2.5}$ concentrations in China from 1998 to 2016, Wang et al. verified that $PM_{2.5}$ concentrations increased along with urbanization and industry, but enhancing abatement investment and clean energy consumption can reverse the increasing trend (Wang et al., 2018). Employing atmospheric models with the higher-order decoupled direct method to simulate effects of urbanization on $O_3$, Yim et al. found that Chinese urbanization increased $O_3$, and $O_3$ has become less/more sensitive to unit changes in $NO_x$ and VOC emissions in various cities (Yim et al., 2019). Using full-coverage remotely sensed $PM_{2.5}$ and population density data, Han et al. found that both population density and $PM_{2.5}$ concentration increased rapidly from 2000 to 2014, especially in East and Central China, as well as China's high population density urban areas and the major cities (Han et al., 2018).

Second, some other studies considered that urbanization would benefit the environment, especially reducing air pollution. For example, a case study on 17 cities in Shandong Province of China showed that the values of urban environmental entropy were negative in most cities from 2001 to 2008, which implies that there is a positive correlation between the development of urbanization and air quality (Wang et al., 2012). Using a cross-sectional data of 282 cities, Lin and Zhu found that the urbanization process has a significant and negative effect on air pollutant concentration, which means that cities with higher urbanization rate tend to have lower air pollutant concentration, including $SO_2$, $PM_{2.5}$ and $PM_{10}$ (Lin and Zhu, 2018). Using China Census data on population, migration, and household fuel use for 2000 and 2010, Aunan and Wang found that the largest population-weighted exposure reduction was estimated for rural-to-urban migrants (Aunan and Wang, 2014). By extending the STIRPAT Model and using China's provincial panel data from 2005 to 2015, Xu et al. found that the higher levels of population and land urbanization in eastern China would promote the production of sulfur dioxide and nitrogen oxides, but reduce $PM_{2.5}$; the improvement of population and land urbanization in the central and western regions has reduced all air pollutants (Xu et al., 2019). Chen et al. examined the influence of land urbanization on meteorology and air quality in China's Yangtze River Delta, finding that land urbanization results in a decrease in the $PM_{2.5}$ concentration in most urban areas (Chen et al., 2020).

Third, a few others considered that in the process of urbanization, the air quality and the level of urbanization have different relations under different urbanization stages and types and in different regions. For example, based on panel data of 30 provinces in China during 2006–2015, Liang and Yang found that there is an environmental Kuznets inverted U curve between economic growth and environmental pollution, and between urbanization and environmental pollution (Liang and Yang, 2019). Using a prefecture-level panel dataset of Yangtze River Economic Belt over the period 2003–2014, Zhu et al.(2019b) found that there is no significant relationship between economic urbanization and $PM_{2.5}$ concentrations, but population urbanization exerts a significant and positive effect on $PM_{2.5}$ concentrations, while land urbanization shows a negative but insignificant influence (Zhu et al., 2019). Taking Beijing-Tianjin-Hebei urban agglomeration from 2000 to 2015 as an example and using a geographically and temporally weighted regression model, Liang et al. found that urbanization improved the environmental quality in mountainous areas, but aggravated the degree of local environmental pollution in plains and coastal areas (Liang et al., 2019).

Though the previous analyses show that China's urbanization has significant impacts on air pollution, there is no directly related study on the impact of the new urbanization pilot policy on air quality and energy-related air pollutants. Therefore, the direction

of the impact of the new urbanization pilot policy on air quality and pollutions is still uncertain and needs to be empirically tested further. To fill this gap, this paper is intended to empirically study the heterogeneous impact of the new urbanization pilot policy on air quality based on the difference-in-difference (DID) method and quantile DID method meanwhile investigating its impacts on different sources of related air pollutants in the pilot cities. The result of this study will help clarify the potential problems of the implementation of the new urbanization pilot policy from the perspective of air pollution sources and provide important empirical evidence and policy implications for the environmental improvement of China. The new urbanization pilot policy provides an effective quasi-natural experiment in assessing the impact of specific policy event on urban air quality.

The study has made two major contributions. First, it is the first study to access the impact of the New Urbanization pilot policy on air quality based on the DID method. This specific analysis can pinpoint whether the strategy made a significant impact. Second, it investigates that the impact of the strategy on the air varies among different energy-related air pollutants including $PM_{2.5}$, $PM_{10}$, $SO_2$, CO, $NO_2$, and $O_3$, and varies among different quantile levels of the air quality. Such heterogeneous effects are conducive to understanding the relationship between air quality and human economic activities and making air improvement policies based on different air pollutants under the context of the new urbanization.

There are three findings from this study. First, China's new urbanization strategy on average tends to improve the air quality in the pilot cities compared with the other key cities of environment protection where the new urbanization pilot policy has not been implemented. Second, based on the quantile DID method, the strategy tends to improve the air quality in the pilot cities at the lower quantile levels from 0.1 to 0.6, which have the higher degree of air pollution, and it has no significant impact in the pilot cities at the higher quantile levels from 0.7 to 0.9. Third, the impact of new urbanization pilot on air pollution varies among six major energy-related air pollutant sources. It tends to restrain $SO_2$, $PM_{10}$, and $PM_{2.5}$ concentrations, increase CO and $O_3$ concentrations and have no significant effect on $NO_2$ concentration.

In the next section, we will introduce the study areas and the air quality situation. *Results* analyses the relationship between urbanization and air pollutions and gives a literature review. *Discussion* estimates the regression models for air quality and energy-related pollutions based on the DID method. *Conclusion* empirically assesses the impact of the new urbanization pilot policy on air quality and discusses whether the impacts vary among different energy-related air pollution sources in the pilot cities. Finally, *Data Availability Statement* provides the conclusion and policy implications.

# MATERIALS AND METHODS

## Pilot Cities Distribution

For the number of cities, the sample range focused by this study is 113 key monitoring cities of environment protection from the Ministry of Ecology and Environment (previously Ministry of Environmental Protection). The problem of air pollution is more serious in these cities. In other words, both the new urbanization pilot cities of the treatment group and the non-pilot cities of the control group are within the scope of these cities. This sampling method can guarantee that the air qualities of the two groups have parallel trends before the policy event for the DID model. In addition, the Ministry of Ecology and Environment started to collect air pollutants data in key cities such as $PM_{2.5}$ and $PM_{10}$ from 2013.

**Figure 1** shows the geographical locations of 113 key cities and 37 new urbanization pilot cities out of 113 in 2015. We find that the pilot cities almost scattered in most of the regions in China and the control group regions and the treatment group regions are overlapped.

China's air quality index consists of six major pollutants, including $PM_{2.5}$, $PM_{10}$, $SO_2$, CO, $NO_2$, and $O_3$. The scatter plot of **Figure 2** shows the distribution of the six pollutants' levels f from 2013 to 2017 and gives a pollution standard with a horizontal line for each air pollutant. The point above the horizontal line indicates that the annual concentration of the city exceeds the pollution standard, while the point below the line indicates that the annual concentration is lower than the standard. We can find that: First, the annual average concentrations of pollutants $PM_{2.5}$, $PM_{10}$, and $SO_2$ fell obviously, transforming from most above the standard to most below the standard. Second, the annual average concentration of CO has decreased, but the pollution level is still relatively lower than the other air pollutants. In addition, $NO_2$ declined but exceeded the standard in most pilot cities, possibly due to the increased traffic and car use in China. Third, $O_3$ pollution is absolutely contrary to other pollutants. In recent years, the $O_3$ pollution level in pilot cities has gradually increased. By 2017, most cities have exceeded the annual average concentration. This may be related to the quick increase of Volatile Organic Compounds (VOC) in Chinese cities, and the increase of ultraviolet radiation caused by the decrease of $PM_{2.5}$ and $PM_{10}$ concentration, which is a necessary condition for forming the VOCs.

For depicting the changing trends, we also present the average change of six air pollutants in the pilot cities as shown in **Figure 3**. We can conclude that except for Ozone pollution, all the major air pollutants have been restrained so the air quality has been improved in the pilot cities.

## Relationship Between New Urbanization and Air Quality

In general, with the per capita income increases and economic growth, energy consumption by residents and industrial productions will increase, and related pollutions will follow up with an increase. As per capita income rises to a certain stage, people tend to use clean energy to protect the environment, and related pollution declines. Therefore, in the early stage of Chinese urbanization with the rapid growth of income, there are unavoidable air pollutions, such as CO and $NO_2$ from traffic development, $PM_{10}$, and $PM_{2.5}$ from the construction industry.

**FIGURE 1 |** Locations of the pilot cities.

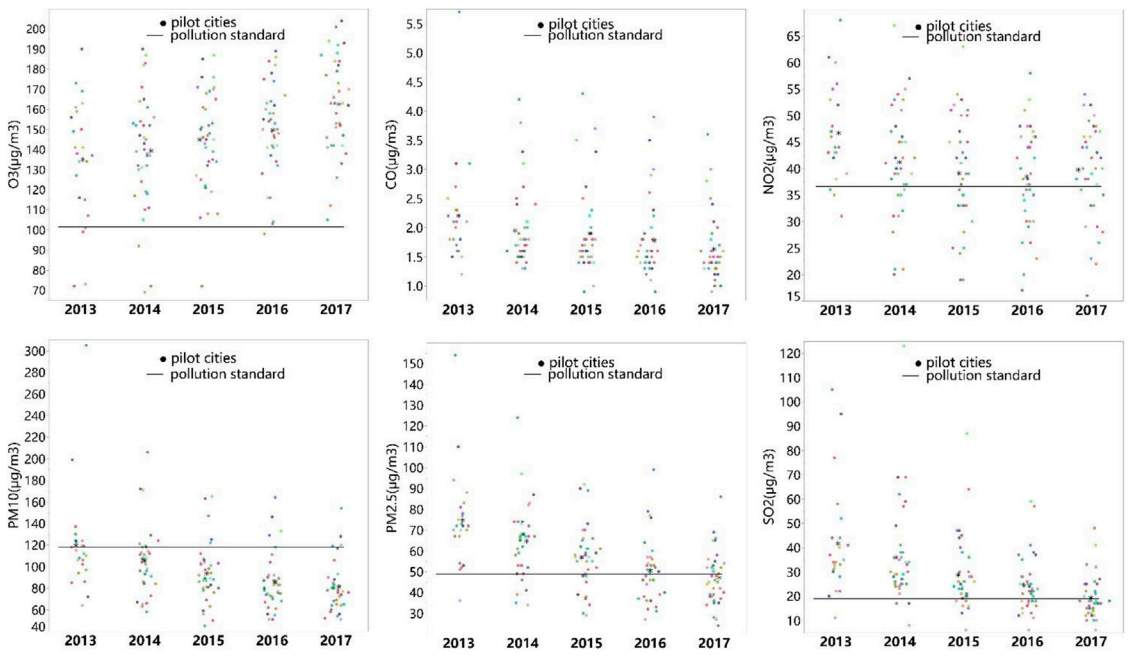



**FIGURE 2 |** Air pollutant distribution of the pilot cities.

After a certain stage of urban development, with the improvement of people's environmental awareness, the strictness of social regulation requirements on the environment, and the upgrading of the city's leading industries from the manufacturing industry to the service industry, the degree of air pollution may decline (Zhao et al., 2020).

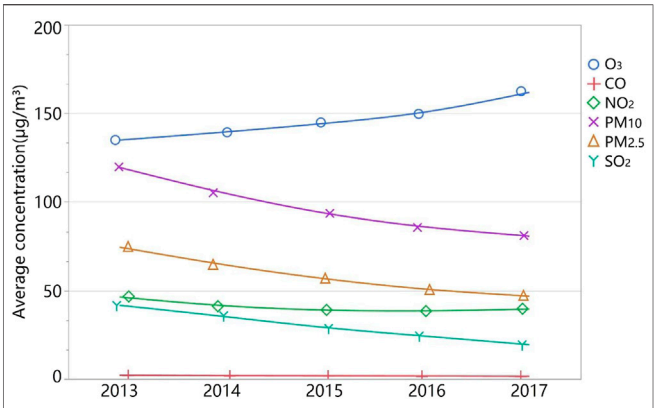



**FIGURE 3 |** Air pollutants in the pilot cities.

Chinese new urbanization strategy launched in 2014 is not a policy specifically designed to affect air quality but compared with the past urbanization, there are obvious differences in energy utilization and environmental protection because the new urbanization puts forward high requirements in terms of green and low-carbon transformation development and controlling the size of the urban population through four channels: intensive urbanization, intelligent urbanization, green urbanization, and low-carbon urbanization. The coordination development between urbanization and the ecological environment has become the direction of China's new urbanization (Wu, 2018). **Figure 4** shows the basic mechanism as follows.

First, intensive urbanization requires that urban development is intensive in land resource use and development scale. In the past, China's urbanization was always accompanied by the excessive development of land resources, resulting in land urbanization significantly faster than population urbanization. The overexpansion of some cities have not only resulted in the inefficient use of resources but also increased air pollution (Zhu et al., 2019; Han and Wu, 2020). In particular, the new urbanization pilot cities are required to promote urbanization on an appropriate scale. For example, the urban population density should be limited to about 100 people per hectare in the pilot cities. Therefore, intensive planning, construction and use of land and the limitation of population urbanization are highlighted, which lead to a more effectively and economically fossil energy use than other cities where the new urbanization pilot has not been implemented.

Second, intelligent urbanization related to green ecology requires that urban development should be accompanied by technological innovation. Nowadays, China's new urbanization takes intelligence and green ecology as the core theory. Increasing cities have joined the planning and construction of green ecology and intelligent urbanization (Su et al., 2020). The construction of intelligent cities under the background of new urbanization bears the important national mission of green development, environmental governance, ecological civilization, and sustainable development (Niu, 2014). Based on the intelligent urbanization development such as the internet of things, cloud computing and big data, the rate of technological upgrading is getting faster and faster. Science and technology progress not only enhance the efficiency of production and management but also break the limit of time and space, realize an organic combination of the factors between production and life. Owing to the technological innovation under intelligent urbanization, the industrial structure is upgraded and energy use is more efficient, which will promote energy conservation and pollution emission reduction, thus improving the urban air quality.

Third, green urbanization requires that urbanization space and layout should fully consider the carrying capacity of resources and environment and maintain the healthy and sustainable development of urbanization (Hao et al., 2020; Wang and Chen, 2021). For example, China is constructing the "sponge cities" in the new urbanization pilot cities, which will strengthen the soil and water conservation capacity based on large-scale afforestation. Another case is the sunken green space, roof greening and surface softening movement in the pilot cities for the purification of air pollution. However, based on the green

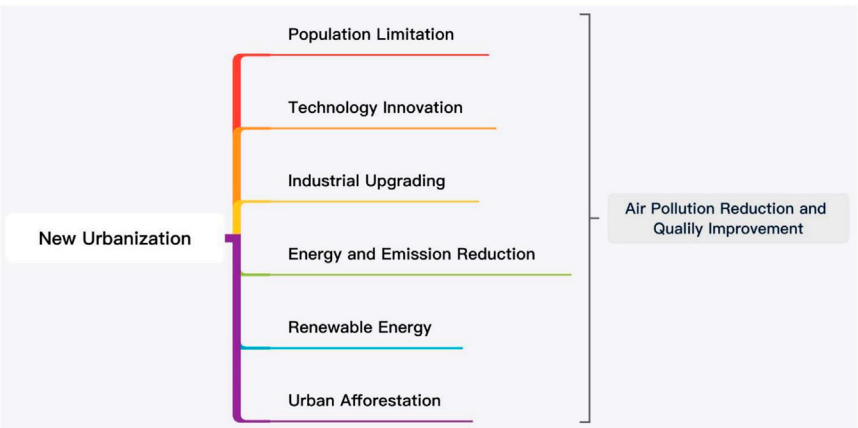



**FIGURE 4 |** Influencing mechanism.

urbanization level panel data of nine provinces in the Yellow River Basin from 2006 to 2018, Shi et al. (2020) measured the green urbanization level of each province through principal component analysis and factor analysis, finding that the overall level of green urbanization in the Yellow River Basin has steadily and rapidly increased, but there are significant spatial differences. The green urbanization level of eastern provinces is significantly higher than that of central and western provinces.

Fourth, low-carbon urbanization requires a clean and efficient utilization of fossil energy, especially for the development of renewable energy (Wu et al., 2019). In addition, it requires the recycling of resources in pilot urban areas, with the focus on promoting a low-carbon lifestyle among urban residents. Wang and Shi (2019) found that low-carbon city construction can significantly reduce urban haze pollution through technological effect and environmental regulation effect. However, Liu et al. (2016) considered that in China, the urban construction closely related to the field of energy, construction, transportation operation has not formed an effective mode of low carbon development. For example, urban areas have excessive energy density and the energy structure is dominated by coal, the energy-saving potential of urban construction is not effectively developed, low carbon mode of public transport and infrastructure construction is also imperfect.

In fact, many of the pilot cities have achieved significant improvements in air quality under the guidance of the new urbanization development strategy. For example, Shijiazhuang, as a pilot of the new urbanization, has continuously promoted the transformation of emission reduction in key industries in the city. First, Shijiazhuang has promoted low-emission upgrading in the steel, cement, glass, ceramic and other related industries, and ban open burning of straw and garbage as well as fireworks and crackers. Second, Shijiazhuang has strengthened the control of dust from mines and sand fields, and the control of construction and urban dust. Third, Shijiazhuang was actively promoting the application of new energy vehicles. By the end of 2020, 80 per cent of new postal, taxi, commuter and light logistics delivery vehicles in the main urban areas have used new energy vehicles or clean energy vehicles, and all buses were replaced with new energy vehicles. Fourth, Shijiangzhuang effectively promoted clean heating by strictly controlling total coal consumption. Construction and expansion of coal - consuming projects are prohibited in the city. Fifth, the city pushed forward the relocation and closure of heavily polluting enterprises. The data show that in Shijiazhuang in 2017, $PM_{2.5}$ concentration decreased by about 20% compared with that in 2016.

For another example, Guangzhou, as a pilot of the new urbanization, has continuously adjusted and optimized its industrial structure to accelerate the formation of a low-emission industrial system. First, Guangzhou attached importance to the construction of a low-carbon transportation system, and vigorously promote the construction of a network of bus lanes at more levels. Second, Guangzhou actively promoted green building and low-carbon development through the whole process of urban construction and strived to avoid large-scale demolition and construction. Third, the whole city has been designated as a coal-free zone. In addition, Guangzhou has also built an international low-carbon industrial park. The enterprise introduction standard is less than or equal to 0.398 tons of standard coal consumption per 10,000 yuan GDP. At the same time, the average annual growth rate of the industrial scale of the enterprise is more than 20%. The park has mainly attracted new energy industry, internet industry, new material industry and other low-carbon and environmental protection enterprises. The data show that Guangzhou has achieved six consecutive years of improvement in six air pollutant sources from 2015 to 2020.

**TABLE 1 |** The variables descriptive statistics.

| variable | Observation | Mean | Std. dev | Min | Max |
|---|---|---|---|---|---|
| *AQI* | 504 | 250.6726 | 63.2599 | 49 | 366 |
| $PM_{10}$ | 504 | 97.17857 | 34.4019 | 37 | 305 |
| $PM_{2.5}$ | 504 | 57.19444 | 19.91398 | 20 | 154 |
| *CO* | 504 | 2.140476 | 0.948195 | 0.8 | 5.9 |
| $SO_2$ | 504 | 29.34425 | 18.72913 | 2.5 | 123 |
| $NO_2$ | 504 | 38.61905 | 10.82656 | 12 | 69 |
| $O_3$ | 504 | 145.5258 | 27.40999 | 69 | 218 |

*Notes: Std. Dev. is the standard error. The unit is μg/m$^3$.*

## Variables and Methods

We use the panel data, containing 113 key cities of environment protection from the Chinese Ministry of Ecology and Environment from 2013 to 2017, which covers 37 new urbanization pilot and also key protection cities constituting the treatment group, and 76 other key but not new urbanization pilot cities constituting the control group. The data of the air quality and the related air pollutants come from the China Statistical Yearbooks on Environment NBS (2018a), and the data of all the control variables come from the China City Statistical Yearbooks (NBS, 2018b). The basic descriptive statistics of the above variables are shown in the following **Table 1**.

According to previous studies, we control for the following variables which are expected to have impacts on air quality and related pollutants.

Urbanization is defined as the proportion of the urban population in the total population and denoted as *UR*. Existing studies on the relationship between urbanization and air pollution are mainly divided into three categories. First, cities lead to increased energy consumption, which in turn worsens the air. But most of this situation happens in the early stages of urbanization. Second, urbanization improves the utilization rate of public facilities and transportation, forms an industrial agglomeration, and then reduces the cost of pollution treatment, which is conducive to reducing pollution emissions. Third, more studies believe that the relationship between urbanization and air pollution is nonlinear, such as inverted U, N-shaped or U-shaped, etc (e.g., Zhao et al., 2020). Therefore, we do not prejudge the effect of urbanization on air quality.

The industrial structure is defined as the ratio of service industrial value-added over GDP and denoted as *SER*. Promoting the service industry with low energy consumption and high added-value is an effective way to realize the transformation and upgrading of industrial structure and it is

conducive to reducing the level of urban pollution (e.g, Wu et al., 2019; Hao et al., 2020). We predicted that it has a positive effect on air quality in cities.

Per capita GDP is the natural logarithm itself and denoted as *PGDP*. Some scholars considered that the increase in per capita GDP will result in increased urban air pollution (e.g., Zhu et al., 2019). However, other scholars believed that there is a negative effect or inverted-U shaped relationship between per capita GDP and urban air pollution based on the Environmental Kuznets Curve (Hao et al., 2019). Therefore, we do not prejudge the effect of urbanization on air quality.

FDI is defined as the share of foreign direct investment (FDI) stock over GDP and denoted as *FDI/GDP*. Some scholars considered that FDI has a negative impact on urban air quality (e.g., Zhu et al., 2019). However, other scholars believed that FDI has significant spatial technological spillovers, improving air quality in China (Jiang et al., 2018). Therefore, we do not prejudge the effect of urbanization on air quality.

Technology spending is defined as the share of fiscal science and technology expenditure over GDP and denoted as *TEC/GDP*. Technology investment will promote technological progress, which will greatly improve production efficiency and reduce energy consumption, especially the progress of energy-saving and pollution reduction technology. We predicted that it has a positive effect on air quality in cities.

The simplest method for estimating the effects of a policy is to compare the difference before and after the policy implementation in the treatment group (the region or individual affected by the policy). This is to subtract the sample mean before the policy implementation from the sample means after the policy implementation in the treatment group. However, since the macroeconomic environment also changes over time (time effects), differences before and after the policy implementation are probably not the same as the policy effects (treatment effects). To solve this problem, we can find a control group that is not included in the policy region as a counterfactual reference, then taking the difference before and after the policy implementation of the control group as the time effects of the treatment group. We subtract the difference before and after the policy in the control group from the difference before and after the policy in the treatment group to obtain a more reliable estimate of the policy effect (Ashenfelter, 1978).

Therefore, the paper uses the above difference in difference (DID) method to investigate the impact of the new urbanization strategy on air quality and energy-related air pollutants in the new urbanization pilot cities (treatment group) by comparing the implementation of this pilot policy before and after, within and without the new urbanization pilot group in the range of key cities with air pollution from the Ministry of Ecology and Environment. Other cities that are not in the pilot range but within the key monitoring cities constitute the control group. The DID method can avoid the endogeneity problem existing in the policy as the independent variable, that is, it can effectively control the unobtainable individual heterogeneity and the influence of the unobtainable factors changing with time, and avoid the reciprocal causation between the dependent variable and independent variable. However, the difference between the regions before and after the implementation of policies may not be the treatment effects. Multiple tests, such as the parallel trend test and placebo test, are needed to judge whether the results meet the conditions for using the DID method. Also, several issues related to the validity of the method need to be clarified further.

First, this paper focuses on the new urbanization pilot cities within the key monitoring cities of environmental pollution from the Ministry of Environment Protection. In our sample period, 37 pilot cities out of 61 meet the conditions and they are almost scattered in 34 provinces of China (**Figure 1**). Therefore, the policy at the provincial level only affects one or two cities, it will not interfere with the overall effects of the new urbanization pilot policy. In addition, during this period there were no other major national policies that could reduce air pollution in these new urbanization pilot and also key monitoring cities with air pollution. For example, "China's "2 + 26" urban air improvement policy (Tough battle against urban air pollution) started in 2018 and there are not similar broad emission reduction policies overlapping in our data period 2013–2017."

Second, considering the selective bias issue, on the one hand for the DID method, selective bias generally occurs when improper grouping fails to satisfy the assumption of the parallel trend. If we choose all the other non-pilot cities in China as the control group, or we choose the non-key monitoring cities as the control group, the trends in air pollution in both groups may be completely inconsistent before the pilot policy. It may result in the biased estimation of the results and lead to the failure of the natural experiment characterized by randomness based on the DID method. However, in the range of key environmental monitoring cities, we choose the new urbanization pilot and also key monitoring cities with air pollution as the treatment group and other cities that are not in the pilot range but within the key monitoring cities as the control group, which can contribute to satisfying the parallel trend hypothesis before the pilot policy event. Therefore, based on the sampling and grouping of our study the selective bias in outcomes is inexistent. On the other hand, the choice of the new urbanization pilot cities is not concerned about the level of air quality in advance, so the sample selection is random. Its main purpose is to improve the urbanization level in China and it is not a specialized environment-improving policy. After the new urbanization pilot policy implemented, more considerations may be given to sustainable development and environmental friendliness, but there are also other key reforming areas such as the equalization of public services and optimization of economic structure in the new urbanization. Therefore, from the perspective of the pilot policy itself, there may is not sample pre-selection bias related to final air quality outcomes.

Finally, considering that the new urbanization pilot cities in the treatment group cover almost all provinces in China, the control group regions and the treatment group regions are overlapped. The propensity score matching (PSM) sampling method can only control unobservable inter-group differences that do not change with time, so the PSM method may be ineffective and not suitable for the sample in our study.

In this paper, the DID method will be adopted, and the new urbanization pilot policy in 2015 will be put forward as a quasi-natural experiment to investigate its effect on air quality and energy-related air pollutants, including $PM_{2.5}$, $PM_{10}$, $SO_2$, CO, $NO_2$, and $O_3$. Therefore, the following DID models are constructed:

$$\ln AQD_{it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (1)$$

$$\ln SO_{2it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (2)$$

$$\ln NO_{2it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (3)$$

$$\ln PM_{10it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (4)$$

$$\ln CO_{it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (5)$$

$$\ln O_{3it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (6)$$

$$\ln PM_{2.5it} = \beta NUR_i \cdot Post_{it} + \gamma X_{it} + \alpha_i + \psi_t + \varepsilon_{it}. \quad (7)$$

**Eqs 1–7** are the DID estimation models that take time and city fixed effects into account. In all the models, *i* stands for the city from 1 to 113, and *t* stands for the year from 2013 to 2017. The dependent variables $lnAQD_{it}$, $lnSO_{2it}$, $lnNO_{2it}$, $lnPM_{10it}$, $lnCO_{it}$, $lnO_{3it}$, and $lnPM_{2.5it}$ in this paper are respectively the days of air quality equal to or above grade II (Good), $SO_2$ concentration, $NO_2$ concentration, $PM_{10}$ concentration, CO concentration, $O_3$ concentration and $PM_{2.5}$ concentration of the new urbanization pilot cities. $Post_{it}$ is the dummy variable for the processing time effect of the new urbanization pilot. For $Post_{it}$, the years after the implementation of the new urbanization pilot are set as 1, and the previous years are set as 0. $NUR_i$ is the dummy variable for processing the treatment group, indicating whether the city belongs to the new urbanization pilot scope. If it is a city in the new urbanization pilot scope, set this variable as 1, otherwise it is 0. $NUR_i \cdot Post_{it}$ is the interaction term between $NUR_i$ and $Post_{it}$, which is the core variable concerned by the DID method. $X_{it}$ is a series of control variables that may cause changes in air quality and energy-related air pollutant emissions, including urbanization rate, industrial structure, per capita GDP, FDI, and Science and Technology expenditure. $\psi_t$ is the year fixed effect, $\alpha_i$ *is* the city fixed effects, $\varepsilon_{it}$ is the random disturbance term.

We focus on the coefficient $\beta$ of the variable $NUR_i \cdot Post_{it}$, whose economic implication can be explained by the impact of the new urbanization pilot on the dependent variables. Taking Model 1 as an example, if the coefficient of $\beta$ is positive and statistically significant, then the new urbanization pilot tends to improve the air quality in the pilot cities. On the contrary, if it is negative and statistically significant, the pilot policy event tends to deteriorate the air quality in the pilot cities.

# RESULTS

## New Urbanization Pilot and Urban Air Quality

We gradually add the city fixed effect, year fixed effect and city time trend effect. Standard errors are clustered at the city level. Compared with other regions, the air quality and energy-related air pollutants may have some inherent change trend, rather than the new urbanization development strategy effect. For example, the process of industrial upgrading to the service sector will lead to a gradual decrease in energy consumption and pollutant emissions. Ignoring the underlying trend change of dependent variables in the treatment group will produce the bias of missing variables and make the estimated results unreliable. The interaction item between city and time trends is added in column (2) of **Table 2**. This interaction item control for the likely trends in the treatment group. All subsequent regressions are controlled for city and time fixed effects, and due to space constraints, there may be no repeated reporting in the subsequent tables.

**TABLE 2 |** The DID results for the air quality.

| Dependent variable | Model 1: Air quality | |
|---|---|---|
| | (1) | (2) |
| *NUR·Post* | 0.1367*** (0.000) | 0.1378*** (0.000) |
| *PGDP* | 0.0061** (0.036) | 0.0055*** (0.000) |
| *SERT* | 0.011*** (0.000) | 0.005*** (0.000) |
| *TEC/GDP* | 0.154* (0.089) | 0.190* (0.082) |
| *FDI/GDP* | 0.428 (0.161) | 0.486* (0.058) |
| *UR* | 0.091*** (0.000) | 0.086*** (0.000) |
| R Square | 0.22 | 0.23 |
| Year FE | Y | Y |
| City FE | Y | Y |
| NUR·Time | | Y |
| Observations | 477 | 477 |

*Source: Authors' estimation.*

*Notes: p-values are in brackets and * refers to $p < 0.1$, ** refers to $p < 0.05$, and *** refers to $p < 0.01$.*

In **Table 2**, the coefficients of the interaction terms are significantly positive at the 1% level, indicating that the new urbanization strategy has a significantly positive impact on the air quality in the new urbanization pilot cities. There may be three main reasons for this finding. First, with the increase in per capita income along with urbanization, urban residents' awareness of environmental protection is rising. Second, under the background of new urbanization, the land is no longer excessively developed and resources are no longer excessively consumed. More attention is paid to the improvement of their utilization efficiency, and the pillar industry of the city shifts to the service industry with less pollution. Third, the progress and promotion of smart cities, low-carbon and energy-saving technologies also promote the improvement of air quality in the new urbanization process.

As to the control variables, the regression results of per capita GDP and urbanization rate are positive and statistically significant, indicating that the increases of per capita GDP and urbanization rate will promote the air quality in the pilot cities. Different from the Kuznets curve, urbanization and economic development of our sample cities with relatively low or medium development levels can promote the improvement of air quality, which is may due to the green and low-carbon urbanization modes and the optimization of industrial structure in recent years. The regression results of FDI stock over GDP are positive and statistically significant, indicating that FDI has promoted the air quality in the pilot cities. The possible reason is that through technology spillovers and industry-leading effects, FDI promotes

**TABLE 3 |** The Quantile DID results for the air quality.

| quantile | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|
| NUR·Post | 0.088** | 0.160* | 0.079** | 0.083*** | 0.079** | 0.063*** | 0.035 | 0.017 | 0.005 |
| | (0.019) | (0.054) | (0.013) | (0.002) | (0.016) | (0.006) | (0.385) | (0.602) | (0.852) |
| Control variables | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| R Square | 0.24 | 0.22 | 0.17 | 0.14 | 0.15 | 0.18 | 0.19 | 0.21 | 0.23 |

*Source: Authors' estimation.*
*Notes: p-values are in brackets and * refers to p < 0.1, ** refers to p < 0.05, and *** refers to p < 0.01.*

**TABLE 4 |** The DID results for different air pollutants.

| Dependent variable | $SO_2$ | $NO_2$ | $PM_{10}$ | CO | $O_3$ | $PM_{2.5}$ |
|---|---|---|---|---|---|---|
| | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
| *NUR·Post* | −0.191*** (0.000) | −0.053 (0.558) | −0.099*** (0.000) | 0.145* (0.089) | 0.045** (0.033) | −0.134*** (0.000) |
| *PGDP* | −0.089** (0.032) | −0.013 (0.152) | −0.043** (0.012) | 0.010* (0.071) | 0.056*** (0.007) | −0.016** (0.034) |
| *SERT* | −0.029*** (0.000) | −0.003** (0.011) | −0.012*** (0.000) | −0.0013*** (0.000) | 0.013*** (0.000) | −0.012*** (0.000) |
| *TEC/GDP* | −0.225*** (0.002) | −0.653** (0.036) | −0.709** (0.032) | −0.132*** (0.002) | 0.126*** (0.001) | −0.125*** (0.002) |
| *FDI/GDP* | −0.045 (0.252) | −0.028 (0.467) | −0.038*** (0.003) | −0.002* (0.055) | 0.010* (0.083) | −0.136*** (0.004) |
| *UR* | −0.072*** (0.000) | −0.029* (0.065) | −0.076** (0.033) | −0.032* (0.062) | 0.028** (0.049) | −0.085*** (0.000) |
| R Square | 0.22 | 0.12 | 0.25 | 0.18 | 0.19 | 0.24 |
| City FE | Y | Y | Y | Y | Y | Y |
| Year FE | Y | Y | Y | Y | Y | Y |
| NUR·Time | Y | Y | Y | Y | Y | Y |
| Observations | 477 | 477 | 477 | 477 | 477 | 477 |

*Source: Authors' estimation.*
*Notes: p-values are in brackets and * refers to p < 0.1, ** refers to p < 0.05, and *** refers to p < 0.01.*

local technological progress and industrial upgrading, and promotes the reduction of energy use and pollutant emissions. Besides, the regression results of other control variables are consistent with theory and expectation.

In **Table 3**, we divide the air quality into nine sub-points from low to high. Based on the quantile DID method, the new urbanization strategy tends to improve the air quality at the lower and medium quantile levels from 0.1 to 0.6. However, it has no significant impact at the higher quantile levels from 0.7 to 0.9. The results indicate that the new urbanization strategy tends to improve the air quality in the lower air quality cities and has no significant impact on the air quality in the higher air quality cities. The reason is perhaps that in the pilot cities with higher air quality, the new urbanization strategy is likely to focus more on other aspects of the new urbanization such as the equal acce ss to public services for both urban and rural populations, which is another core element of the new urbanization, but focus less on air quality because of its lower air pollution level.

## New Urbanization Pilot and Urban Air Pollutants

The effects of the new urbanization strategy on different sources of energy-related air pollution may vary because of the different sources. In this part, we will examine the impacts of the new urbanization pilot policy on six major air pollutants including $PM_{2.5}$, $PM_{10}$, $SO_2$, CO, $NO_2$, and $O_3$ based on the DID method.

In **Table 4**, the regression results show that the variable *NURPost* is significantly negative at the 1 per cent level in Model 2, Model 4, and Model 7, significantly positive at the 10 per cent level in Model 5 and significantly positive at the 5 per cent level in Model 6, but negative and insignificant in Model 3, indicating that the new urbanization development strategy tends to restrain energy-related $SO_2$, $PM_{10}$, and $PM_{2.5}$ concentrations, increase CO and $O_3$ concentrations and has no impact on $NO_2$ concentration.

There may be several reasons for these findings. First, 90% of $SO_2$ in China comes from industrial coal combustion. In recent years, along with the new urbanization process, the gradual reduction of coal use and the progress of emission filtration technology in urban industrial development are quite obvious, thus reducing the $SO_2$ emissions in the urban areas. Second, China has been eliminating a large number of outdated capacities in cities, such as the steel industry, and the proportion of coal in urban energy consumption continues to decline, which are also beneficial to $PM_{10}$ and $PM_{2.5}$ reduction. Third, low-carbon urbanization, as well as intelligent urbanization under the new urbanization strategy, can effectively promote the improvement of energy efficiency and the use of renewable energy, and promote fossil energy conservation and pollution reduction. Fourth, more than 80% of the CO in the atmosphere of Chinese cities is discharged by fuel vehicles. The pilot cities are mainly small and medium-sized cities with lower and middle levels of urban development, and the urban residents of these cities mainly purchase fuel cars. As income rises, private car ownership and

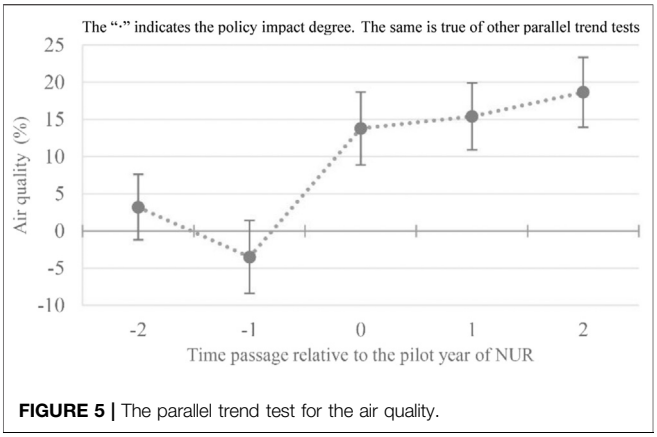



**FIGURE 5 |** The parallel trend test for the air quality.

car travel rise, thereby unavoidably driving CO emissions. Fifth, in 2019, $O_3$ has become the second-largest air pollutant in Chinese cities after $PM_{2.5}$. $O_3$ pollution occurs mostly in sunny summer, and it is more common in developed countries. For China, on the one hand, the average temperature increases because of the urban heat island effect and global warming, especially in the summer when extremely hot weather increases significantly the $O_3$ pollutions are severe; on the other hand, the reduction of haze during the new urbanization leads to more sufficient ultraviolet rays, which can promote the synthesis of $O_3$ between nitrogen oxides and organic volatiles. Sixth, unlike CO pollution, which mainly comes from human factors, although industrial production and automobile exhaust can cause $NO_2$, in general, it is more likely affected by natural factors, which can be also proved from the lowest coefficient of $R^2$ (0.12). For example, lightning can form a large amount of $NO_2$ in thunderstorms weather. Therefore, the impact of the new urbanization strategy on $NO_2$ is insignificant.

For the control variables, the impacts of FDI stock over GDP on CO, $PM_{10}$, and $PM_{2.5}$ pollutions are significantly negative, but it is significantly positive for $O_3$ and insignificant for $SO_2$ and $NO_2$. Foreign enterprises in China mainly invest in labour-intensive industries and tertiary industries. Heavy industries with high $SO_2$ and $NO_2$ pollution are mainly Chinese state-owned enterprises. Besides, FDI can promote the industrial upgrading, technology spillover, and economic growth of the local cities, thus tending to restrain $PM_{10}$, $PM_{2.5}$, and CO pollutions but increase $O_3$ concentration to some extent. The regression results of other control variables are consistent with theories and expectations.

# DISCUSSION

## Data Validity and Robustness Check

First, we test whether the mean values of the dependent variable air quality between the treatment group and the control group are equal after the new urbanization pilot. The results show that the *p*-value equals zero from the base period, which indicates that there is a significant difference in air quality between the two groups.

Second, this paper conducts a further parallel trend check before the new urbanization pilot. DID model does not require the mean values to be the same between the control group and treatment group but hypothesizes that the trends in the treatment group and control group should be the same before the policy implementation. To support this assumption, **Figure 5** shows that there is no obvious difference between the two groups in the 2 years before the implementation of the new urbanization pilot policy. It shows that before the implementation of the new urbanization pilot, the dummy variables are not statistically significant at the 5% level. Therefore, there is a parallel trend before the new urbanization pilot policy between the treatment group and control group, because the treatment effect is not distinguishable from zero before the policy event. Furthermore, from the year 2015, the impact of the policy on the air quality is positive, gradually increasing with time and is statistically significant each year. In the two years after the

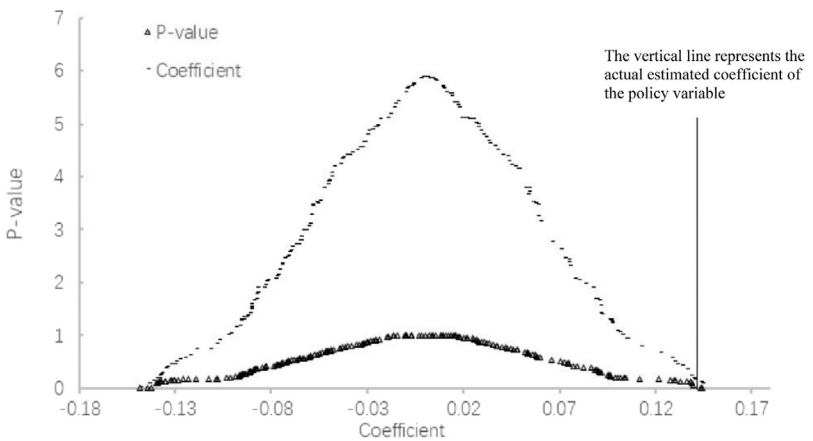



**FIGURE 6 |** Placebo test for the robustness.

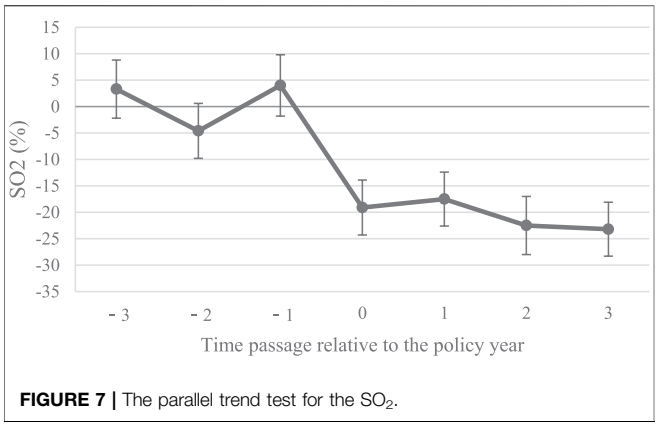



**FIGURE 7 |** The parallel trend test for the $SO_2$.

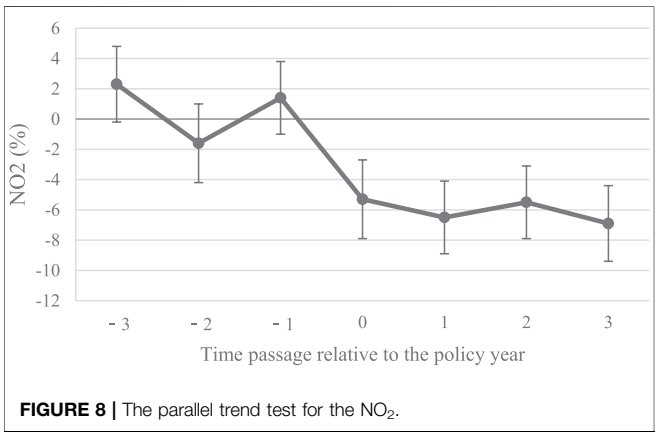



**FIGURE 8 |** The parallel trend test for the $NO_2$.

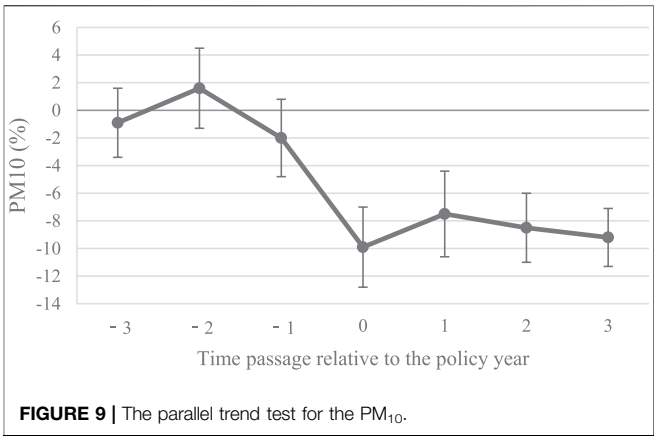



**FIGURE 9 |** The parallel trend test for the $PM_{10}$.

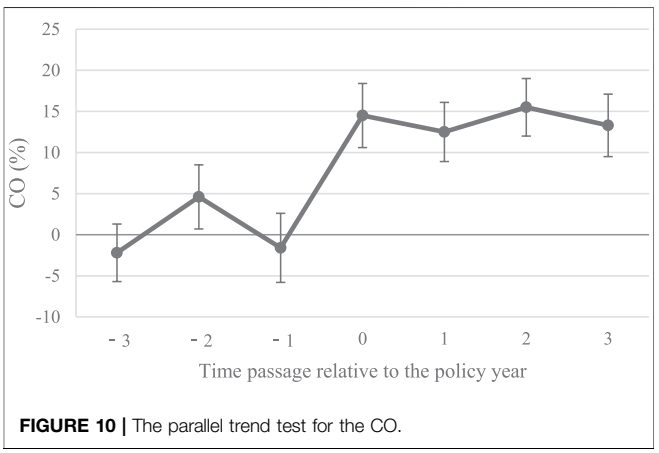



**FIGURE 10 |** The parallel trend test for the CO.

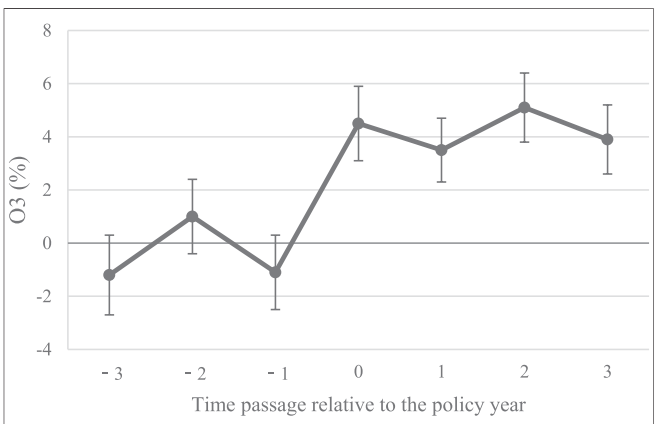



**FIGURE 11 |** The parallel trend test for the $O_3$.

implementation of the new urbanization pilot policy, the effects of the policy increase from 13.8% in the year of implementation to 19.2% 2 years after implementation.

Third, a possible cause of estimation bias is the omission of important variables. Similar to Cai et al. (Cai et al., 2016) and La Ferrara et al. (La Ferrara et al., 2012), by randomly selecting the new urbanization pilot cities from the sample, this paper conducts a further placebo test for the results. The sample of this paper includes 113 key cities of environment protection, of which 37 are the new urbanization pilot cities. Accordingly, 37 cities were randomly selected from 113 cities and set as the pseudo treatment group, and the remaining cities are set as the pseudo control group, to build a dummy cross term variable $NUR_i{}^{false}\ Post_{it}$ for the placebo test. Because the pseudo treatment group was randomly generated, the placebo cross-term should not have a significant effect on the dependent variable. In other words, the regression coefficients should be overall insignificant if there is no significant deviation of the missing variable. Furthermore, to avoid the interference of other low-probability events on the estimated results, the above process was repeated 200 times for regression analysis. **Figure 6** reports the estimated coefficient and the distribution of the corresponding *p*-value in the 200 random generation treatment group. It shows that the mean of the regression coefficients is close to zero (coefficient mean 0.0006), and most *p* values are greater than 0.1. Meanwhile, the vertical line in **Figure 6** represents the actual estimation coefficient, which is an outlier in the estimated coefficients of the placebo test. It is proved that the improvement of air quality is not caused by other unobserved factors. In other words, the estimated results without errors caused by missing variables are robust.

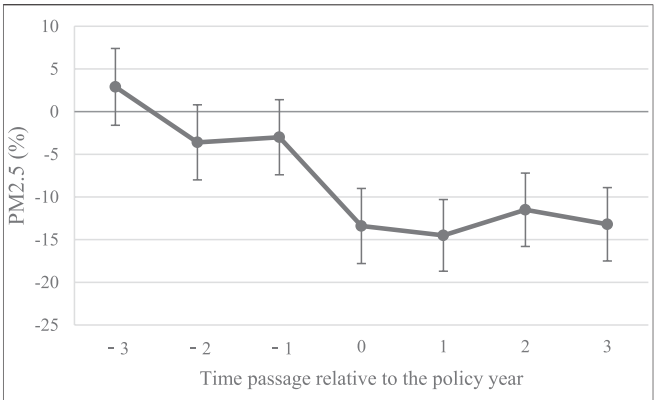



**FIGURE 12 |** The parallel trend test for the $PM_{2.5}$.

## Method Rationality and Research Limitations

First, it is true that air quality in non-pilot cities has been improved because of China's overall upgrading of industrial structure, advances in energy conservation and emission reduction technologies, and stronger environmental regulations. The parallel trend test does not require that the treatment group and the control group have opposite trends, but that the development trend in the treatment group is significantly faster than or lower than the control group after the policy. So even though the trend in both groups is upward, there may exist a faster and a slower trend in the treatment group after the policy, which can be clearly indicated in **Figures 7–12**.

Second, although our DID model only examined the effect of the policy in the year it was implemented, we think the effect of the policy on air quality will last for a long time. China's new-type urbanization pilot policy is not a one-year plan, but a long-term one. The goal is to promote sustainable urban development on the premise of efficient use of resources and environmentally friendly protection, and to raise the level of urbanization to the level of developed countries, with an urbanization rate of 75% by 2035.

Third, the above research has given some findings and policy implications on this topic below. However, it should be acknowledged that due to the time series is just till 2017, whether the effects of the new urbanization pilot policy on air quality and related air pollutants continue to this day is unknown. China's Ministry of Ecology and Environment has only started to collect air pollutants in some major polluting cities since 2013. Except for the DID method, by comparing the new urbanization development of different regions, we may find out something new between China's new urbanization and air pollution.

## CONCLUSION

This study uses the city-level panel data, 113 key cities of environment protection from the Ministry of Ecology and Environment, covering most of the provinces in China for the period 2013–2017 to investigate the different impacts of the new urbanization pilot policy on air quality and related air pollutants including six major pollutant sources based on the difference-in-difference (DID) method and quantile DID method.

The study finds that, first, based on the difference-in-difference (DID) method, the new urbanization on average tends to improve the air quality in pilot cities compared with other cities where the new urbanization pilot policy has not been implemented. Second, based on the quantile DID method, the new urbanization tends to improve the air quality in the lower air quality cities, however, it has no significant impact in the higher air quality cities. Third, the impacts of the new urbanization pilot policy on the air quality vary among different energy-related air pollutants. The new urbanization development strategy tends to restrain $SO_2$, $PM_{10}$, and $PM_{2.5}$ concentration, increase CO and $O_3$ concentration and has no impact on $NO_2$ concentration.

The findings of this study may generate some policy implications. In line with the United Nations Sustainable Development Goals (SDGs) of building sustainable consumption and production patterns and taking urgent action to address climate change, China should make more national efforts to control urban air pollution. For example, in the process of urban construction, China needs to promote clean energy application and increase construction efficiency to further reducing haze pollutions in big cities. Second, some small and medium-sized cities' governments can further intensify the vehicle emission standards and extend the tax and subsidy policies designed to encourage the purchase of electric vehicles like Beijing to reduce oil consumption. At the same time, the public transportation system like subways and electric buses can be more quickly improved. Third, the Chinese urban manufacturing industry can reduce the emissions of pollutants such as $NO_2$ and VOC that will form ozone pollution in cities. For example, from June to August when ozone pollution is high, the processes involving high VOC emissions of related industries can be stopped during the high-temperature period of the day, such as 1 PM–5 PM. Fourth, more government expenditure should be used in increasing the Urban forest coverage for improving the air pollution purifying capability in cities, which is also in line with China's recently proposed carbon-neutral plan. Fifth, China can further develop intelligent and intensive cities from new urbanization pilot cities to other cities for increasing energy efficiencies and reducing consumption thus decreasing air pollution in cities. Finally, China can pay more attention to promote the conception of green production and consumption among urban companies and residents, upgrading urban industrial structure, facilitating the advances in energy conservation and emission reduction technologies, and implementing stronger environmental regulations to reduce pollutant emissions for dealing with potential urban air pollution issues in the future.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

YW: Conceptualization, Methodology, Writing-Original draft preparation, Validation. ZC: Investigation, Visualization, Software. CH: Data curation, Writing-Reviewing and Editing, Supervision.

## FUNDING

This work was financially supported by the Beijing Municipal Commission of Education (PXM2019_014213_000007, SM201910011011).

## REFERENCES

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *Rev. Econ. Stat.* 60, 47–57. doi:10.2307/1924332

Aunan, K., and Wang, S. (2014). Internal Migration and Urbanization in China: Impacts on Population Exposure to Household Air Pollution (2000-2010). *Sci. Total Environ.* 481, 186–195. doi:10.1016/j.scitotenv.2014.02.073

Cai, X., Lu, Y., Wu, M., and Yu, L. (2016). Does Environmental Regulation Drive Away Inbound Foreign Direct Investment? Evidence from a Quasi-Natural Experiment in China. *J. Dev. Econ.* 123, 73–85. doi:10.1016/j.jdeveco.2016.08.003

Chen, T., Li, M., Luo, L., Deng, S., Zhou, R., and Chen, D. (2020). Simulating the Effects of Land Urbanization on Regional Meteorology and Air Quality in Yangtze River Delta, China. *Appl. Geogr.* 120, 102228. doi:10.1016/j.apgeog.2020.102228

Du, Y., Wan, Q., Liu, H., Liu, H., Kapsar, K., and Peng, J. (2019). How Does Urbanization Influence PM 2.5 Concentrations? Perspective of Spillover Effect of Multi-Dimensional Urbanization Impact. *J. Clean. Prod.* 220, 974–983. doi:10.1016/j.jclepro.2019.02.222

Fang, C., Liu, H., Li, G., Sun, D., and Miao, Z. (2015). Estimating the Impact of Urbanization on Air Quality in China Using Spatial Regression Models. *Sustainability* 7, 15570–15592. doi:10.3390/su71115570

Grossman, G., and Krueger, A. (1991). Environmental Impacts of a North American Free Trade Agreement. *Natl. Bur. Econ. Res.* doi:10.3386/w3914

Han, H., and Wu, W. (2020). Analysis of Interaction Effect between New Urbanization and Intensive Land Use: Comprehensive Evaluation of Beijing-Tianjin-Hebei Urban Agglomeration Based on PVAR Model. *Urban Issues* 07, 11–20. (In Chinese).

Han, L., Zhou, W., Li, W., and Li, L. (2014). Impact of Urbanization Level on Urban Air Quality: A Case of Fine Particles ($PM_{2.5}$) in Chinese Cities. *Environ. Pollut.* 194, 163–170. doi:10.1016/j.envpol.2014.07.022

Han, L., Zhou, W., Li, W., and Qian, Y. (2018). Urbanization Strategy and Environmental Changes: An Insight with Relationship between Population Change and Fine Particulate Pollution. *Sci. Total Environ.* 642, 789–799. doi:10.1016/j.scitotenv.2018.06.094

Hao, Y., ChenLiao, Y. F. H., and Wei, Y. M. (2019). China's Fiscal Decentralization and Environmental Quality: Theory and an Empirical Study. *Environ. Dev. Econ.* 25 (2), 159–181. doi:10.1017/S1355770X19000263

Hao, Y., Zheng, S. Q., Zhao, M. Y., Wu, H. T., Guo, Y. X., and Li, Y. W. (2020). Reexamining the Relationships Among Urbanization, Industrial Structure, and Environmental Pollution in China—New Evidence Using the Dynamic Threshold Panel Model. *Energ. Rep.* 6, 28–39. doi:10.1016/j.egyr.2019.11.029

Jiang, L., Zhou, H. F., Bai, L., and Zhou, P. (2018). Does Foreign Direct Investment Drive Environmental Degradation in China? an Empirical Study Based on Air Quality Index from a Spatial Perspective. *J. Clean. Prod.* 176, 864–872. doi:10.1016/j.jclepro.2017.12.048

La Ferrara, E., Chong, A., and Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *Am. Econ. J. Appl. Econ.* 4, 1–31. doi:10.1257/app.4.4.1

Li, Z., Zhou, Y., Wan, B., Chen, Q., Huang, B., Cui, Y., et al. (2019). The Impact of Urbanization on Air Stagnation: Shenzhen as Case Study. *Sci. Total Environ.* 664, 347–362. doi:10.1016/j.scitotenv.2019.01.232

Liang, L., Wang, Z., and Li, J. (2019). The Effect of Urbanization on Environmental Pollution in Rapidly Developing Urban Agglomerations. *J. Clean. Prod.* 237, 117649. doi:10.1016/j.jclepro.2019.117649

Liang, W., and Yang, M. (2019). Urbanization, Economic Growth and Environmental Pollution: Evidence from China. *Sustain. Comput. Inform. Syst.* 21, 1–9. doi:10.1016/j.suscom.2018.11.007

Lin, B., and Zhu, J. (2018). Changes in Urban Air Quality during Urbanization in China. *J. Clean. Prod.* 188, 312–321. doi:10.1016/j.jclepro.2018.03.293

Liu, Q., Li, G., Chen, Y., and Ouyang, H. (2016). Problems and Countermeasures of Low-Carbon Urbanization in China. *Chin. Popul. Resour. Environ.* 26, 42–46. (In Chinese).

NBS, NBS (National Bureau of Statistics of China) and Ministry of Ecology and Environment (2018a). *China Statistical Yearbooks on Environment*. (Beijing: China Statistics Press). (in Chinese).

NBS, NBS (National Bureau of Statistics of China) (2018b). *China City Statistical Yearbook*. (Beijing: China Statistics Press). (in Chinese).

Niu, W. Y. (2014). Smart Cities Are the Driving Force of New Urbanization. *Bull. Chin. Acad. Sci.* 29, 34–41. (In Chinese).

Shi, L., Cai, Z., Ding, X., Di, R., and Xiao, Q. (2020). What Factors Affect the Level of Green Urbanization in the Yellow River Basin in the Context of New-type Urbanization?. *Sustainability* 12, 2488. doi:10.3390/su12062488

Su, L., Fan, J., and Fu, L. (2020). Exploration of Smart City Construction under New Urbanization: A Case Study of Jinzhou-Huludao Coastal Area. *Sust. Comput. Inform. Syst.* 27, 100403. doi:10.1016/j.suscom.2020.100403

Wang, H., and Shi, D. (2019). A New Type of Urbanization Is Conducive to the Alleviation of Haze Pollution: Empirical Evidence from the Construction of Low-Carbon Cities. *J. Shanxi Univ. Finance Econ.* 41, 15–27. (In Chinese).

Wang, K., and Chen, M. (2021). Epistemology of Green Urbanization in China. *J. Urban Plann.* 01, 10–17. (In Chinese).

Wang, Q., Yuan, X., Ma, C., Zhang, Z., and Zuo, J. (2012). Research on the Impact Assessment of Urbanization on Air Environment with Urban Environmental Entropy Model: A Case Study. *Stoch. Environ. Res. Risk Assess.* 26, 443–450. doi:10.1007/s00477-011-0493-5

Wang, X., Tian, G., Yang, D., Zhang, W., Lu, D., and Liu, Z. (2018). Responses of $PM_{2.5}$ Pollution to Urbanization in China. *Energy Policy* 123, 602–610. doi:10.1016/j.enpol.2018.09.001

Wu, H. L. (2018). A Review of the Research on the Synergistic Development of New Urbanization and Environment under the Concept of Sustainable Development. *Ind. Tech. Econ.* 37, 02–105. (In Chinese).

Wu, H. T., Hao, Y., and Weng, J. (2019). How Does Energy Consumption Affect China's Urbanization? New Evidence from Dynamic Threshold Panel Models. *Energy Policy* 127, 24–38. doi:10.1016/j.enpol.2018.11.057

Xu, S. C., Miao, Y. M., Gao, C., Long, R. Y., Chen, H., Zhao, B., et al. (2019). Regional Differences in Impacts of Economic Growth and Urbanization on Air Pollutants in China Based on Provincial Panel Estimation. *J. Clean. Prod.* 208, 340–352. doi:10.1016/j.jclepro.2018.10.114

Yang, D., Chen, Y., Miao, C., and Liu, D. (2020). Spatiotemporal Variation of $PM_{2.5}$ Concentrations and its Relationship to Urbanization in the Yangtze River Delta Region, China. *Atmos. Pollut. Res.* 11, 491–498. doi:10.1016/j.apr.2019.11.021

Yim, S. H. L., Wang, M., Gu, Y., Yang, Y., Dong, G., and Li, Q. (2019). Effect of Urbanization on Ozone and Resultant Health Effects in the Pearl River Delta Region of China. *J. Geophys. Res. Atmos.* 124, 11568–11579. doi:10.1029/2019JD030562

Zhang, Y. N., Xiang, Y. R., Chan, L. Y., Chan, C. Y., Sang, X. F., Wang, R., et al. (2011). Procuring the Regional Urbanization and Industrialization Effect on Ozone Pollution in Pearl River Delta of Guangdong, China. *Atmos. Environ* 45, 4898–4906. doi:10.1016/j.atmosenv.2011.06.013

Zhao, Z. B., Yuan, T., Shi, X. P., and Zhao, L. D. (2020). Heterogeneity in the Relationship between Carbon Emission Performance and Urbanization: Evidence from China. *Mitigation Adaptation Strateg. Glob. Change* 25 (7), 1363–1380. doi:10.1007/s11027-020-09924-3

Zhu, L. Y., Hao, Y., Lu, Z. N., Wu, H. T., and Ran, Q. Y. (2019b). Do economic Activities Cause Air Pollution? Evidence from China's Major Cities. *Sust. Cities Soc.* 49, 101593. doi:10.1016/j.scs.2019.101593

Zhu, W., Wang, M., and Zhang, B. (2019). The Effects of Urbanization on $PM_{2.5}$ Concentrations in China's Yangtze River Economic Belt: New Evidence from Spatial Econometric Analysis. *J. Clean. Prod.* 239, 118065. doi:10.1016/j.jclepro.2019.118065

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Wu, Cui and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 28 June 2021
doi: 10.3389/fenvs.2021.685939

# The Impacts of Two Revisions of the China's Environmental Protection Fee on Firm Performance: Evidence From Stock Markets

*Huan Zheng*[1] *and Yu He*[2]*

[1]School of Economics and Management, Chongqing Jiaotong University, Chongqing, China, [2]Research Center for Economy of Upper Reaches of the Yangtse River, Chongqing Technology and Business University, Chongqing, China

This comparative study aims to investigate how China's environmental protection fee affected firm performance in two separate sample periods, namely 2001–2006 and 2012–2017, as this policy was revised twice in 2003 and 2014. With the Difference-in-Differences (DID) estimation, we find that the second revision of environmental protection fee had a negative impact on heavy polluting firms, while the influence of the first revision seemed to be insignificant. Moreover, the environmental protection fee had a limited impact on other firms, implying that such a policy had achieved its designed effect by precisely governing heavy polluters. Besides, our additional tests show that the environmental protection fee had stronger impacts on non-connected firms and non-SOEs than politically connected firms and SOEs. Our results are robust to various potential endogeneity issues.

**Keywords: environmental protection fee, firm performance, heavy polluting firms, environmental friendly firms, difference-in-differences estimation**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Firmin Doko Tchatoka,*
*University of Adelaide, Australia*
*Mathilde Maurel,*
*UMR8174 Centre d'économie de la Sorbonne (CES), France*

***Correspondence:**
*Yu He*
*yu.he@ctbu.edu.cn*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *26 March 2021*
***Accepted:*** *16 June 2021*
***Published:*** *28 June 2021*

***Citation:***
*Zheng H and He Y (2021) The Impacts of Two Revisions of the China's Environmental Protection Fee on Firm Performance: Evidence From Stock Markets. Front. Environ. Sci. 9:685939. doi: 10.3389/fenvs.2021.685939*

## INTRODUCTION

Over the last 4 decades, China has maintained rapid growth and created an economic miracle. However, such a miracle was at the expense of resource depletion and environmental pollution. This country became a "pollution haven" because of its lax environmental regulations (Dean et al., 2009). Following the developed countries, the Chinese authorities first introduced the environmental protection fee, namely the pollution discharge fee (PRC State Council 1982), to control firms' discharge of pollutants in 1982. Considered the national economy at that time, this policy had relatively low charge standards. Not surprisingly, such a policy did not achieve its design effect, as evidenced by the fact that China had become the largest $CO_2$ emitter in 2007 (Netherlands Environmental Assessment Agency). The pollution problems forced the authorities to reconsider their environmental regulations and thus largely increased the charge standards of environmental protection fee in 2003 and 2014.

Similar as other environmental regulations, the environmental protection fee can be a double-edged sword, which contributes to sustainable growth but also affects firm operation. There is a considerable debate on the effects of environmental regulation on corporate performance[1]. The

[1]See, for example, Porter (1991), Porter and van der Linde (1995a), Porter and van der Linde (1995b), Kahn and Knittel (2003), Lanoie et al. (2008), Qi et al. (2014), Zhao et al. (2018), Huang and Liu (2019) and Maurel et al. (2019).

Porter Hypothesis (Porter, 1991; Porter and van der Linde, 1995a; Porter and van der Linde, 1995b) states that a well-designed policy may lead to win-win situations in both social welfare and firm performance as such a policy may stimulate firms' innovative capabilities, increase their productivity, and thus partly or fully offset the cost of complying with environmental regulation. Nevertheless, other scholars such as Lanoie et al. (2008), Greenstone et al. (2012), and Huang and Liu (2019) argue that environmental regulations impose additional costs for firms and thus worsen corporate performance.

In this paper, we apply a comparative study in two separate sample periods, namely 2001–2006 and 2012–2017, in order to answer three major research questions. Since previous studies observe both positive and negative relationships between environmental regulations and firm performance, we are interested in how the environmental protection fee affects firm operation in China. This policy was designed to follow the polluter-pays principle. In other words, firms releasing pollutants into the environment will be charged, while others without releasing pollutants into the environment will not be affected by this policy. Thus, we apply heavy polluting firms as treated group and environmental friendly firms (namely, firms do not release pollutants into the environment) as control group, and explore the influence of the two revisions (in 2003 and 2014) of the charge standards of the environmental protection fee on corporate performance with Difference-in-Differences (DID) estimations. The second research question we aim to examine whether the two revisions of the environmental protection fee are efficient environmental regulations. We define an efficient environmental regulation as a policy has stronger negative impact on heavy polluters than other firms. To investigate, we further apply normal firms (neither heavy polluters nor environmental friendly firms) as treated group and environmental friendly firms as control group, and re-estimate the potential impacts of the environmental protection fee. The last research question we discuss the different impacts of the environmental protection fee on corporate performance for various firm characteristics, including politically connected and non-connected firms, as well as state owned enterprises (SOEs) and non-SOEs.

We reach several important findings in this study. A negative relationship is observed between environmental protection fee and firm performance for the sample period of 2012–2017. More specifically, heavy polluting firms have worse performance—in terms of lower development capacity, leverage conditions and profitability—than environmental friendly firms after 2014. These firms are more likely to fall into financial distress. Nevertheless, the first revision in 2003 seems to have insignificant influence on firm performance. The possible reason is that the second revision of this policy significantly increases heavy polluting firms' environmental costs and worsens firm performance. The first revision increases firms' costs to some extent, but the charge standards are still at relatively low levels and thus present insignificant impacts on firm performance for the sample period of 2001–2006. In addition, the environmental protection fee has insignificantly influence on normal firms for both of the sample periods, with a negative impact on firm development capacity during the sample period of 2012–2017 being the exception. Lastly, our additional tests show that the environmental protection fee seems to have worse impact on non-connected firms than that of connected firms for the sample period of 2012–2017. Also, the environmental protection fee is insignificantly related to the firm performance of SOEs, while negatively associated with non-SOEs over the sample period of 2012–2017. Our results are robust when controlling for potential endogeneity issues.

Our study makes several contributions. First, the impacts of environmental regulations on firm performance have been debated for more than 2 decades without conclusive results. Our study contributes to the literature by furthering the discussion and finding that the China's environmental protection fee has a negative influence on firm performance. Second, through a comparative study, we demonstrate that the second revision of environmental protection fee significantly affects firm performance, while the influence of the first revision seems to be insignificant. Third, our empirical results show that the environmental protection fee has stronger impacts on heavy polluters than other firms. These results provide robustness for our main tests and also indicate that the environmental protection fee is an efficient environmental regulation as such a policy precisely governs heavy polluters. Finally, this study provides practical suggestions for policymakers. Our sub-sample tests find that the environmental protection fee has negative impacts on non-connected firms and non-SOEs. The possible reason is that non-connected firms and non-SOEs have competitive disadvantage against politically connected firms and SOEs, as the later ones are more likely to acquire funding support from the government and relieve the negative influence of environment regulations. Since non-connected firms and non-SOEs are also important parts of the national economy, the Chinese government should provide more financial support to assist these firms in transformation and upgrading. Given the unique institutional background in China, the above results enhance our understanding of the impact of environmental protection fee on firm performance in the world's largest emerging market.

The remainder of the paper is organized as follows. In *Institutional Background and Literature Review*, we briefly discuss the institutional background and the literature on environmental policies. We then describe the empirical design (including data collection, sample description, variables, and regression methodology) of this study in *Empirical Design*. We report regression results in *Regression Results*. Finally we provide conclusions in *Conclusion*.

# INSTITUTIONAL BACKGROUND AND LITERATURE REVIEW

## Institutional Background-The Environmental Protection Fee

China initially promulgated its environmental protection fee in 1982, namely the Provisional Measures for Collecting Pollution

Fees (PRC State Council 1982). This policy was designed to follow the polluter-pays principle, in other words, firms were charged by this policy only if they released pollutants into the environment. The charge amounts is based on the amounts of pollutants firms released into the environment. Since the first version of the environmental protection fee was implemented at the start of the reform and opening-up policy, the Chinese authorities are more concerned about national economic development than environmental protection. Thus, this policy had relatively low charge standards and did not distinguish different kinds of pollutants.

Two decades later, based on 20 years of implementation experience, the Chinese authorities started to realize the limitations of such a policy and thus released the second version (i.e., the first revision) of the environmental protection fee in 2003, namely the Management Regulation on Collection and Use of Pollution Fees (PRC State Council 2003). In the second version, several new concepts were introduced. More specifically, the previous charge on single pollutants was replaced by multi-pollutants, and the notion of pollutant equivalent value was introduced to distinguish different harmful levels of various pollutants. Besides, the charge standards were also increased. For instance, the charge standard of $SO_2$ was 0.04 RMB (Renminbi) per kg in the first version. Through translating the units, we find that the charge standards of $SO_2$ were 0.038 RMB and 0.6 RMB per pollutant equivalent value before and after the first revision.

Although the revision in 2003 increased the fee rates of different pollutants, the charge standards were still at relatively low levels. As a result, the pollution behavior of enterprises was not effectively restrained, and the China's environment was getting worse. In response to the serious polluting problems, the National Development and Reform Commission released the third version (i.e., the second revision) of the environmental protection fee in 2014, namely the Circular on Relevant Issues Including Adjustment of Pollution Fee Collection Standard (National Development and Reform Commission 2014). Based on the second revision, local governments are entitled to set the local fee rates, which should be no less than 1.2 RMB and 1.4 RMB per pollution equivalent values of air and water pollutants.[2]

## Literature Review

The problem of environmental pollution has attracted public attention in recent decades. Many countries in the world released a number of environmental policies, including laws, administrative regulations, and standards, to control firms' pollution behavior. It is certain that environmental regulations contribute to environmental protection, while the impacts of these regulations on corporate values seem to be inconclusive. Currently, there are three groups of studies that can be identified based on their results.

One stream of the literature argues that environmental regulation has negative influence on firm performance, as a stringent regulation imposes additional costs and worsens firms' productivity (Barbera and McConnell, 1990; Gray and Shadbegian, 2003; Lanoie et al., 2008; Greenstone et al., 2012; Huang and Liu, 2019). Through investigating the United States polluting industries, Barbera and McConnell (1990) and Gray and Shadbegian (2003) find that stringent environmental regulation causes a decline in firms' productivity. Greenstone et al. (2012) further state that stricter air quality regulations are associated with a decline of roughly 2.6 per cent in total factor productivity (TFP) at manufacturing plants. Lanoie et al. (2008) observe a similar result by exploring Canadian manufacturing sector, where environmental regulation has adverse impacts on growth in TFP and a positive lagged effect on productivity. The negative relationship is also detected in China. Huang and Liu (2019) suggest that environmental regulation promotes firm productivity slightly with a lagged effect, and has a harmful impact on firm exports. Moreover, there is a U-shaped relationship with firm exports, and China is to the far left of the inflection point.

Another stream shows a positive relationship between environmental regulation and firm performance. According to the famous Porter Hypothesis (Porter, 1991; Porter and van der Linde, 1995a, Porter and van der Linde, 1995b), a well-designed environmental policy may stimulate firms' innovative capabilities, increase their productivity, and thus partly or fully offset the cost of complying with environmental regulation. As a result, win-win situations in both social welfare and firm performance can be achieved. Such an argument has been demonstrated in many countries, such as the United States (Jaffe and Palmer, 1997), European countries (Kneller and Manderson, 2012), China (Qi et al., 2014), and 25 member countries in the Organization for Economic Cooperation and Development (OECD) (Johnstone et al., 2010). More specifically, Jaffe and Palmer (1997) find that environmental regulation has a positive impact on R&D expenditures in the United States manufacturing industries. Besides, through investigating China's industrial firms over the time period of 1990–2010, Qi et al. (2014) observe a positive relationship between corporate or industrial-level environmental performance and firm performance. Ramanathan et al. (2017) further state that firms are more likely to reap the private benefits of sustainability, if they adopt a more dynamic approach to respond to environmental regulations innovatively and take a proactive approach to manage their environmental performance.

In addition to the previous studies, the last stream of literature finds that the impacts of environmental regulations can be different in various situations. For instance, Kahn and Knittel (2003) argue that stringent environmental policies impose firms' environmental costs and thus should have negative influence on firm performance. Nevertheless, the empirical evidence shows that the United States coal mining firms have negative abnormal returns on the one hand while electricity firms experience positive abnormal returns on the other. Similar results have also been detected when examining the price changes of European Union Emission Allowance (EUA) (Oberndorfer, 2009; Bushnell et al., 2013). Besides, Horváthová (2010) states that environmental regulations are more likely to have positive impacts on firm

[2]For the details of the fee rates, please refer to the website of the Chinese Ministry of Ecology and Environmental: http://www.mee.gov.cn/.

performance in common law countries than in civil law countries. Zhao et al. (2018) find that different kind of environmental regulations, such as legislative regulation, environmental information disclosure, and administrative regulation, have a variety of impacts on firm performance. In summary, after more than 2 decades' debate, the influence of environmental regulations on corporate values still remains inconclusive. Against this background, this study joins in the discussion and explores the impacts of the two revisions of China's environmental protection fee on firm performance.

# EMPIRICAL DESIGN

## Data Collection and Sample Description

The financial data used in this study are extracted from the China Stock Market and Accounting Research (CSMAR) database. As mentioned previously, we apply DID estimation to explore the impact of the two revisions (2014 and 2003) of environmental protection fee on firm performance. Based on the principle of the DID estimation, two separate sample periods, namely 2012–2017 and 2001–2006 (three years before and after each of the revisions), are chosen for four reasons: 1) we choose six years for each of our sample periods, as these sample periods provide us neither too large nor too small sample size; 2) if the sample periods are too long, our empirical analyses may be interfered by other events, such as the Global Financial Crisis in 2008, the China Environmental Protection Tax Law in 2018, the China Resource Tax Law in 2020, and other government policies; 3) if the sample periods are too short, the impact of this policy on sample firms may not be fully manifested; 4) several data are unavailability before 2000 and after 2018.

The original sample includes all the A shares listed on Shanghai Stock Exchange (SHSE) and Shenzhen Stock Exchange (SZSE). A number of firm-year observations were deleted based on the following four criteria: 1) firms in the financial sector, due to differences in regulations; 2) firms listed in the Growth Enterprise Market (GEM), as the lifecycle may significantly affect firm operation (Koh et al., 2015). Most of the listed firms in Chinese stock markets are in maturity stage except those firms listed in the GEM; 3) the first firm-year observation and those firms with less than two consecutive years of financial information, as we need two consecutive years' observations to calculate change variables; 4) firms with missing or extreme values. Finally, the sample periods of 2012–2017 and 2001–2006 include 2,248 and 1,317 firms, respectively.

## Variables

### Dependent Variables

We are interested in how environmental protection fee affects firm performance. Three dependent variables are adopted to capture three aspects of firm performance, including development capability (Assets growth, which is the change of total assets, defined as $(Assets_t - Assets_{t-1})/Assets_{t-1}$. Assets represent the total assets), leverage condition (Leverage growth, which is the change of leverage, defined as $Leverage_t - Leverage_{t-1}$. Leverage equals to total liabilities divided by total assets), and profitability (EPS, which is the earnings per share). In addition, Chan and Chen (1991) state that financially distressed firms are inefficient producers with poorer performance, higher financial leverage, and more cash flow problems. Thus, we further utilize financial distress to measure firm performance, as such a measure is a more comprehensive indicator which can capture firm performance, leverage, and cash flow conditions simultaneously. Previous studies mainly adopt three indicators to measure firm financial distress in China, including ST status (Wang and Deng, 2006; Wang and Li, 2007), Altman Z-score (Fan et al., 2013), and interest coverage ratio (Kam et al., 2008; Bhattacharjee and Han, 2014). However, we identify too few financial distress firms to complete empirical analyses with the ST status and Altman Z-score. Thus, we apply the interest coverage ratio to measure financial distress in this study, as this indicator provides us a suitable sample size. Follow previous studies (Kam et al., 2008; Bhattacharjee and Han, 2014), we define firms as financially distressed if their interest coverage ratios are below the cut-off point of 1 for two consecutive years. *Distress* is a dummy variable, which equals 1 if a firm falls into financial distress in a given year during the sample period and 0 otherwise.

### Independent Variables

This study applies the DID estimation to explore the impact of environmental protection fee on firm performance. As mentioned previously, this policy follows the polluter-pays principle. In other words, firms are affected only if they discharge pollutants into the environment. Thus, we select the treated and control groups as follows: 1) we apply heavy polluting firms as our treated group. These firms are strongly affected by this policy as they release a large amount of pollutants during their manufacturing process. We follow the China Corporate Environmental Information Disclosure Guidelines (government document) and categorize firms as heavy polluters if they belong to the following sectors: thermal power, steel, cement, electrolytic aluminum, coal, metallurgy, chemical, petrochemical, building materials, paper making, brewing, pharmaceutical, fermentation, textile, leather and mining; 2) besides, we apply environmental friendly firms as our control group. Since there is no official definition on these firms, thus we develop our own definition: firms are defined as environmental friendly if they do not use or release any specified pollutants (including Greenhouse Gas) and do not produce hazardous waste or pollutes water during their manufacturing process. Obviously, these firms are not affected by this policy. Based on the above criteria, we observe 675 (406) heavy polluting firms and 191 (88) environmental friendly firms for the sample periods of 2012–2017 (2001–2006), respectively. Moreover, we adopt two variables, namely Heavy Polluting (HP) Firms and After, to construct the DID estimation. HP Firms is a dummy variable, which equals 1 if a firm is a heavy polluter and 0 if it is an environmental friendly firm. After is also a dummy variable, which takes a value of 1 for firm-year observation after the revision of environmental protection fee and 0 otherwise. Since

**TABLE 1 |** Summary statistics.

This table presents the summary statistics of main variables. The financial data is extracted from the CSMAR database. The study includes two separate sample periods, namely 2012–2017 and 2001–2006. The statistics of most variables are based on firm-year observations, excepting that for Distress and HP Firms, which are based on firm-level observations. For the definitions of variables refer to **Appendix A**.

**Panel A: 2012–2017 vs. 2001–2006**

| Variables | 2012–2017 | | | | 2001–2006 | | | | *t*-test of difference | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Median | Std. Dev. | Obs | Mean | Median | Std. Dev. | Difference in mean | *t*-stat. |
| Development capacity | 4,870 | 0.2593 | 0.0852 | 1.2387 | 2,604 | 0.1308 | 0.0839 | 0.3021 | 0.1285*** | 5.211 |
| Leverage condition | 4,870 | −0.0006 | 0.0033 | 0.1185 | 2,604 | 0.0309 | 0.0204 | 0.1347 | −0.0315*** | −10.441 |
| Profitability | 4,870 | 0.3124 | 0.2237 | 0.6517 | 2,604 | 0.1773 | 0.1614 | 0.4348 | 0.1351*** | 9.509 |
| Distress | 866 | 0.3684 | 0.0000 | 0.4824 | 494 | 0.2571 | 0.0000 | 0.4370 | 0.1113*** | 4.228 |
| HP firms | 866 | 0.7794 | 1.0000 | 0.4146 | 494 | 0.8219 | 1.0000 | 0.3826 | −0.0424* | −1.864 |
| After | 4,870 | 0.5228 | 1.0000 | 0.4995 | 2,604 | 0.5511 | 1.0000 | 0.4975 | −0.0283** | −2.336 |
| Gas fee | 4,870 | 1.4623 | 1.2000 | 1.8520 | 2,604 | 0.3477 | 0.6000 | 0.2796 | 1.1146*** | 30.535 |
| Water fee | 4,870 | 1.4855 | 1.4000 | 1.7922 | 2,604 | 0.3858 | 0.7000 | 0.3481 | 1.0996*** | 30.998 |
| Board size | 4,870 | 2.1637 | 2.1972 | 0.2032 | 2,604 | 2.2547 | 2.1972 | 0.2302 | −0.0911*** | −17.615 |
| CEO duality | 4,870 | 0.2556 | 0.0000 | 0.4363 | 2,604 | 0.1137 | 0.0000 | 0.3175 | 0.1420*** | 14.660 |
| First | 4,870 | 0.3499 | 0.3323 | 0.1550 | 2,604 | 0.4279 | 0.4206 | 0.1704 | −0.0780*** | −20.015 |
| Sedtenth | 4,870 | 0.2142 | 0.1988 | 0.1288 | 2,604 | 0.1833 | 0.1554 | 0.1406 | 0.0309*** | 9.566 |
| Age | 4,870 | 2.3812 | 2.5649 | 0.6413 | 2,604 | 1.8214 | 1.9459 | 0.5150 | 0.5599*** | 38.415 |
| Size | 4,870 | 22.2833 | 22.1203 | 1.3554 | 2,604 | 21.2368 | 21.1422 | 1.0551 | 1.0465*** | 34.240 |
| Leverage | 4,870 | 0.4539 | 0.4436 | 0.2526 | 2,604 | 0.4902 | 0.4871 | 0.2188 | −0.0363*** | −6.193 |

**Panel B: 2012–2017**

| | Heavy polluting firms | | | | Environmental friendly firms | | | | *t*-test of difference | |
|---|---|---|---|---|---|---|---|---|---|---|
| Development capacity | 3,815 | 0.2165 | 0.0789 | 1.1739 | 1,055 | 0.4138 | 0.1294 | 1.4390 | −0.1973*** | −4.588 |
| Leverage condition | 3,815 | −0.0012 | 0.0024 | 0.1121 | 1,055 | 0.0014 | 0.0072 | 0.1394 | −0.0027 | −0.645 |
| Profitability | 3,815 | 0.3028 | 0.2132 | 0.6432 | 1,055 | 0.3473 | 0.2497 | 0.6809 | −0.0446** | −1.966 |
| Distress | 675 | 0.3319 | 0.0000 | 0.4712 | 191 | 0.4974 | 0.0000 | 0.5013 | −0.1655*** | −4.225 |
| After | 3,815 | 0.5208 | 1.0000 | 0.4996 | 1,055 | 0.5299 | 1.0000 | 0.4993 | −0.0090 | −0.519 |
| Gas fee | 3,815 | 1.3487 | 1.2000 | 1.5827 | 1,055 | 1.8729 | 1.2000 | 2.5621 | −0.5242*** | −8.192 |
| Water fee | 3,815 | 1.3642 | 1.4000 | 1.5099 | 1,055 | 1.9239 | 1.4000 | 2.5184 | −0.5597*** | −9.052 |
| Board size | 3,815 | 2.1687 | 2.1972 | 0.1999 | 1,055 | 2.1454 | 2.1972 | 0.2136 | 0.0233*** | 3.305 |
| CEO duality | 3,815 | 0.2372 | 0.0000 | 0.4254 | 1,055 | 0.3223 | 0.0000 | 0.4676 | −0.0850*** | −5.622 |
| First | 3,815 | 0.3589 | 0.3417 | 0.1549 | 1,055 | 0.3173 | 0.2955 | 0.1511 | 0.0416*** | −7.763 |
| Sedtenth | 3,815 | 0.2084 | 0.1921 | 0.1282 | 1,055 | 0.2350 | 0.2229 | 0.1287 | −0.0266*** | −5.955 |
| Age | 3,815 | 2.4004 | 2.6391 | 0.6306 | 1,055 | 2.3118 | 2.3979 | 0.6743 | 0.0886*** | 3.977 |
| Size | 3,815 | 22.4267 | 22.2198 | 1.3397 | 1,055 | 21.7649 | 21.7858 | 1.2841 | 0.6617*** | 14.326 |
| Leverage | 3,815 | 0.4673 | 0.4642 | 0.2294 | 1,055 | 0.4054 | 0.3864 | 0.3183 | 0.0619*** | 7.079 |

**Panel C: 2001–2006**

| | Heavy polluting firms | | | | Environmental friendly firms | | | | *t*-test of difference | |
|---|---|---|---|---|---|---|---|---|---|---|
| Development capacity | 2,124 | 0.1416 | 0.0917 | 0.3022 | 480 | 0.0829 | 0.0456 | 0.2972 | 0.0587*** | 3.853 |
| Leverage condition | 2,124 | 0.0305 | 0.0219 | 0.1096 | 480 | 0.0324 | 0.0140 | 0.2131 | −0.0019 | −0.280 |
| Profitability | 2,124 | 0.1990 | 0.1794 | 0.4303 | 480 | 0.0813 | 0.1050 | 0.4423 | 0.1177*** | 5.385 |
| Distress | 406 | 0.2389 | 0.0000 | 0.4269 | 88 | 0.3409 | 0.0000 | 0.4767 | −0.1020** | −1.989 |
| After | 2,124 | 0.5537 | 1.0000 | 0.4972 | 480 | 0.5396 | 1.0000 | 0.4990 | 0.0141 | 0.560 |
| Gas fee | 2,124 | 0.3492 | 0.6000 | 0.2794 | 480 | 0.3412 | 0.6000 | 0.2804 | 0.0079 | 0.560 |
| Water fee | 2,124 | 0.3877 | 0.7000 | 0.3480 | 480 | 0.3778 | 0.7000 | 0.3492 | 0.0099 | 0.560 |
| Board size | 2,124 | 2.2571 | 2.1972 | 0.2247 | 480 | 2.2442 | 2.1972 | 0.2532 | 0.0129 | −1.109 |
| CEO duality | 2,124 | 0.1177 | 0.0000 | 0.3223 | 480 | 0.0958 | 0.0000 | 0.2947 | 0.0219 | 1.363 |
| First | 2,124 | 0.4396 | 0.4409 | 0.1702 | 480 | 0.3763 | 0.3460 | 0.1616 | 0.0633*** | 7.423 |
| Sedtenth | 2,124 | 0.1748 | 0.1430 | 0.1383 | 480 | 0.2207 | 0.2171 | 0.1447 | −0.0459*** | −6.506 |
| Age | 2,124 | 1.7951 | 1.9459 | 0.5163 | 480 | 1.9378 | 2.0794 | 0.4933 | −0.1428*** | −5.516 |
| Size | 2,124 | 21.3460 | 21.2363 | 1.0357 | 480 | 20.7539 | 20.7117 | 1.0040 | 0.5921*** | 11.374 |
| Leverage | 2,124 | 0.4903 | 0.4908 | 0.1975 | 480 | 0.4894 | 0.4761 | 0.2952 | 0.0009 | 0.080 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 2 |** Correlation matrix.
This table presents the correlations between main variables in two separate sample periods. The DC, LC, and CD are short for Development capability, Leverage condition, and CEO duality, respectively. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

**Panel A: 2012–2017**

| | DC | LC | Profitability | HP firms | After | Board size | CD | First | Sedtenth | Age | Size |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LC | −0.0737*** | — | — | — | — | — | — | — | — | — | — |
| | (0.000) | — | — | — | — | — | — | — | — | — | — |
| Profitability | 0.0655*** | −0.1709*** | — | — | — | — | — | — | — | — | — |
| | (0.000) | (0.000) | — | — | — | — | — | — | — | — | — |
| HP firms | −0.0656*** | −0.0092 | −0.0282** | — | — | — | — | — | — | — | — |
| | (0.000) | (0.519) | (0.049) | — | — | — | — | — | — | — | — |
| After | 0.0183 | −0.0454*** | 0.0196 | −0.0074 | — | — | — | — | — | — | — |
| | (0.203) | (0.002) | (0.171) | (0.604) | — | — | — | — | — | — | — |
| Board size | −0.0359** | −0.0240* | 0.0349** | 0.0473*** | −0.0643*** | — | — | — | — | — | — |
| | (0.012) | (0.094) | (0.015) | (0.001) | (0.000) | — | — | — | — | — | — |
| CD | 0.0597*** | 0.0111 | 0.0132 | −0.0803*** | 0.0312** | −0.1872*** | — | — | — | — | — |
| | (0.000) | (0.439) | (0.359) | (0.000) | (0.029) | (0.000) | — | — | — | — | — |
| First | 0.0142 | 0.0068 | 0.0869*** | 0.1106*** | −0.0726*** | 0.0535*** | −0.01173*** | — | — | — | — |
| | (0.323) | (0.636) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | — | — | — | — |
| Sedtenth | 0.1467*** | −0.0663*** | 0.1533*** | −0.0850*** | 0.1093*** | −0.0010 | 0.0725*** | −0.3827*** | — | — | — |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.942) | (0.000) | (0.000) | — | — | — |
| Age | 0.0107 | −0.0874*** | −0.0795*** | 0.0569*** | 0.1322*** | 0.0911*** | −0.1372*** | −0.0663*** | −0.2890*** | — | — |
| | (0.455) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | — | — |
| Size | 0.0332** | 0.0096 | 0.1288*** | 0.2011*** | 0.1404*** | 0.2908*** | −0.1360*** | 0.3402*** | −0.0040 | 0.1888*** | — |
| | (0.021) | (0.505) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.783) | (0.000) | — |
| Leverage | −0.0238* | 0.2949*** | −0.2571*** | −0.1009*** | −0.0602*** | 0.1210*** | −0.0705*** | 0.0706*** | −0.1746*** | 0.2620*** | 0.2668*** |
| | (0.097) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

**Panel B: 2001–2006**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LC | 0.1774*** | — | — | — | — | — | — | — | — | — | — |
| | (0.000) | — | — | — | — | — | — | — | — | — | — |
| Profitability | 0.3452*** | −0.3297*** | — | — | — | — | — | — | — | — | — |
| | (0.000) | (0.000) | — | — | — | — | — | — | — | — | — |
| HP Firms | 0.0753*** | −0.0055 | 0.1050*** | — | — | — | — | — | — | — | — |
| | (0.000) | (0.779) | (0.000) | — | — | — | — | — | — | — | — |
| After | −0.0483** | 0.0097 | 0.0189 | 0.0110 | — | — | — | — | — | — | — |
| | (0.014) | (0.622) | (0.334) | (0.575) | — | — | — | — | — | — | — |
| Board size | 0.0743*** | −0.0539*** | 0.1108*** | 0.0217 | −0.0256 | — | — | — | — | — | — |
| | (0.000) | (0.006) | (0.000) | (0.267) | (0.191) | — | — | — | — | — | — |
| CD | −0.0147 | 0.0179 | −0.0492** | 0.0267 | 0.0167 | −0.0491*** | — | — | — | — | — |
| | (0.453) | (0.361) | (0.012) | (0.173) | (0.393) | (0.012) | — | — | — | — | — |
| First | 0.0759*** | −0.0188 | 0.1598*** | 0.1440*** | −0.1305*** | −0.0047 | −0.0688*** | — | — | — | — |
| | (0.000) | (0.337) | (0.000) | (0.000) | (0.000) | (0.812) | (0.000) | — | — | — | — |
| Sedtenth | −0.0113 | 0.0557*** | −0.0615*** | −0.1265*** | 0.0759*** | 0.0451** | 0.0589*** | −0.6670*** | — | — | — |
| | (0.564) | (0.004) | (0.002) | (0.000) | (0.000) | (0.021) | (0.003) | (0.000) | — | — | — |
| Age | −0.1075*** | −0.0639*** | −0.1666*** | −0.1075*** | 0.2667*** | −0.0246 | −0.0106 | −0.2257*** | −0.0150 | — | — |
| | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.210) | (0.589) | (0.000) | (0.445) | — | — |
| Size | 0.2406*** | −0.0573*** | 0.3475*** | 0.2176*** | 0.1113*** | 0.2618*** | −0.0794*** | 0.2563*** | −0.1932*** | 0.0263 | — |
| | (0.000) | (0.003) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.180) | — |
| Leverage | −0.0240 | 0.5394*** | −0.3347*** | 0.0016 | 0.1646*** | −0.0626*** | 0.0369* | −0.1477*** | 0.0882*** | 0.2079*** | 0.0046 |
| | (0.222) | (0.000) | (0.000) | (0.936) | (0.000) | (0.001) | (0.060) | (0.000) | (0.000) | (0.000) | (0.813) |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

we have two sample periods (2012–2017 and 2001–2006) in this study, After equals 1 for firm-year observations after 2015 and 2004, and 0 otherwise.

In order to control firms' corporate governance conditions, we employ four factors: 1) board characteristics, including Board size (the logarithm of number of directors employed in the corporate board) and CEO duality (dummy variable which equals 1 if an employee works as both chief executive officer, CEO, and chair of board, and 0 otherwise); 2) ownership concentration, including First (the percentage of shares held by the largest shareholder) and Sedtenth (the percentages of shares held by the second to 10th largest shareholders). Through investigating 2,980 American firms over the period of 1996–2004, Cheng (2008) states that firms with larger boards have lower variability of corporate performance. Yang and Zhao (2014) find that duality firms outperform non-duality firms by 3–4% when their competitive

environments change. In addition, Ma et al. (2010) find that ownership concentration has stronger impacts than any category of ownership in determining performance over Chinese listed firms. The significant influence of corporate governance is also demonstrated in other countries (Earle et al., 2005; Khan et al., 2021).

Besides, a large body of finance literature (e.g., O'Hara and Shaw, 1990; Opler and Titman, 1994; Lennox, 1999 and Poncet et al., 2010) demonstrates that firm financial fundamentals have a significant influence on firm performance. For instance, O'Hara and Shaw (1990) and Lennox (1999) find that fundamental conditions, such as age, size, leverage, and industry sector, are related to firm performance and financial distress. In this study, we adopt three factors to control firm fundamental conditions, which are firm age (Age, the logarithm of years since the firm went public), firm size (Size, the logarithm of firm's total assets), and leverage ratio (Leverage, the ratio of total debt on total assets).

## Regression Methodology

The major objective of this study is exploring how the revisions of environmental protection fee affect firm performance in two sample periods (2012–2017 and 2001–2006). To this end, we utilize the following Ordinary Least Squares (OLS) regression model and Probit regression model for empirical analyses. It is worthwhile mentioning that the OLS regression model is adopted to investigate the impacts of environmental protection fee and other proposed factors on firm development capability, leverage condition, and profitability. In addition, the Probit regression model is applied to explore how firms fall into financial distress with the distressed firm sub–sample. Based on our financial distress measure, we observe a total of 319 and 127 firms falling into distress during the sample periods of 2012–2017 and 2001–2006, respectively.

$$\begin{aligned} OLS(\text{Performance})_{i,t} = {} & \alpha^1_{i,t} + \beta^1_{i,t}\text{HPFirms}_{i,t} + \gamma^1_{i,t}\text{After}_{i,t} \\ & + \delta^1_{i,t}\text{HPFirms}_{i,t}*\text{After}_{i,t} \\ & + \zeta^1_{i,t}\text{CorporateGovernance}_{i,t} \\ & + \eta^1_{i,t}\text{Fundamental}_{i,t} + \theta^1_t\text{Year}_t \\ & + \tau^1_{i,t}\text{Industry}_{i,t} + \varepsilon^1_{i,t}, \end{aligned} \tag{1}$$

$$\begin{aligned} \text{Probit}(\text{Distress})_{i,t} = {} & \alpha^2_{i,t} + \beta^2_{i,t}\text{HPFirms}_{i,t} + \gamma^2_{i,t}\text{After}_{i,t} \\ & + \delta^2_{i,t}\text{HPFirms}_{i,t}*\text{After}_{i,t} \\ & + \zeta^2_{i,t}\text{CorporateGovernance}_{i,t} \\ & + \eta^2_{i,t}\text{Fundamental}_{i,t} + \theta^2_t\text{Year}_t \\ & + \tau^2_{i,t}\text{Industry}_{i,t} + \varepsilon^2_{i,t}. \end{aligned} \tag{2}$$

For dependent variables, we adopt four indicators to capture firm performance, including development capability (Assets growth, which is the change of total assets, defined as $(\text{Assets}_t - \text{Assets}_{t-1})/\text{Assets}_{t-1}$. Assets represent the total assets), leverage condition (Leverage growth, which is the change of leverage, defined as $\text{Leverage}_t - \text{Leverage}_{t-1}$. Leverage equals to total liabilities divided by total assets), profitability (EPS, which is the earnings per share), and financial distress (Distress, dummy variable which equals 1 if firm *i* falls into financial distress in a given year *t* during the sample period, and 0 otherwise). Regarding firm development capability and leverage condition, we apply the time-serial changes in variables rather than absolute levels, which enable us to mitigate the effects of firm heterogeneity (Opler and Titman, 1994; Lin et al., 2009; Fan et al., 2013).

For independent variables, we use HP Firms, After and the interaction term (HP Firms*After) to construct DID estimations, and explore the impacts of environmental protection fee on firm performance. HP Firms is a dummy variable, which equals 1 if a firm is a heavy polluter and 0 if it is an environmental friendly firm. Since we have two sample periods (2012–2017 and 2001–2006) in this study, After equals 1 for firm-year observations after 2015 and 2004, and 0 otherwise. Corporate Governance contains four variables: Board size (the logarithm of number of directors employed in the corporate board), CEO duality (dummy variable which equals 1 if an employee works as both CEO and chair of board, and 0 otherwise), First (the percentage of shares held by the largest shareholder), and Sedtenth (the percentages of shares held by the second to 10th largest shareholders). Fundamental includes three variables: Age (the logarithm of years since the firm went public), Size (the logarithm of firm's total assets), and Leverage (the ratio of total debt on total assets). Besides, Year and Industry fixed effects are also controlled in the empirical analyses. For the detail of variable definition, see **Appendix A**.

## Summary Statistics

Panel A of **Table 1** provide summary statistics for the variables during two separate sample periods (2012–2017 and 2001–2006). It is worthwhile mentioning that the statistics of most variables are based on firm-year observations, excepting Distress and HP Firms, which are based on firm-level observations. Firms during the period of 2012–2017 generally have better development capacity and profitability than the period of 2001–2006. Nevertheless, firms have worse leverage condition and are more likely to fall into financial distress during the period of 2012–2017 than the period of 2001–2006, where the *t*-tests of differences in mean are −0.0315 and 0.1113, respectively.

Regarding the independent variables, there are 675 (77.94 per cent of 866 firms) and 406 (82.19 per cent of 494 firms) heavy polluting firms during the sample periods of 2012–2017 and 2001–2006, respectively. The Environmental Protection Fee rates, including gas and water fee rates, are significantly larger during the sample period of 2012–2017 than those of the sample period of 2001–2006. Thus, we expect that the revision in 2014 may have stronger impacts on firm performance than that of the revision in 2003. Firms during the period of 2012–2017 have smaller board size and higher CEO duality than the period of 2001–2006. The *t*-tests of differences in mean of First and Sedtenth are −0.0780 and 0.0309, indicating that firms' ownership structure became more dispersed during the period of 2012–2017 than the period of 2001–2006. Beside, firms have longer time periods since going public, larger firm size, and lower leverage ratios during the period of 2012–2017 than 2001–2006.

Panel B and C of **Table 1** report the results of t-tests of differences in mean between heavy polluters and environmental friendly firms for the sample periods of 2012–2017 and 2001–2006. From Panel B of **Table 1**, we find that the Environmental Protection Fee rates for environmental friendly firms are larger than those for heavy polluting firms during the sample period of 2012–2017. It should

be notice that these results do not mean environmental friendly firms will be charged with higher fee rates. On the contrary, these data show that most environmental friendly (heavy polluting) firms are located in those provinces with higher (lower) fee rates. These results imply that the deterrent effect of the revision in 2014 prompts heavy polluters to locate to those provinces with lower fee rates. However, the revision in 2003 seems do not has such a deterrent effect, as the *t*-tests of differences of Environmental Protection Fee rates between heavy polluters and environmental friendly firms are statistically insignificant. Thus, these results suggest that the revision in 2014 may have stronger impacts on firm performance than that of the revision in 2003.

Since this study contains two separate sample periods, the Panel A and B of **Table 2** report the correlations between main variables during the period of 2012–2017 and 2001–2006, respectively. Although statistically significant in many cases, the correlations between our main variables are generally small. Therefore, the potential problems arising from collinearity between variables are unlikely to be an issue in this study.

## REGRESSION RESULTS

As mentioned in the introduction section, this study attempts to answer three major research questions: 1) How does the environmental protection fee affect firm operating performance? 2) Are the two revisions of the environmental protection fee efficient environmental regulations? 3) What are the different impacts of the environmental protection fee on corporate performance for various firm characteristics?

### Firm Performance

In order to answer the first research question, this study utilizes the OLS regression model and Probit regression model for empirical analysis. The Panel A and B of **Table 3** present the empirical results for the sample periods of 2012–2017 and 2001–2006, respectively. Regarding the sample period of 2012–2017, we find that the interaction between HP Firms and After is negatively related to firms' development capacity, leverage conditions and profitability, and positively associated with financial distress. For the sample period of 2001–2006, the impact on firm performance of the environmental protection fee is statistically insignificant, with a positive impact of the interaction term on firm profitability being the exception. These results indicate that the second revision of environmental protection fee increase firms' environmental costs—especially for heavy polluting firms—and worsens firm performance, while the first revision seems to have insignificant impacts on firm performance.

With respect to corporate governance conditions, Panel A of **Table 3** shows that firms with smaller board size and higher CEO duality have better development capacity. Also, Board size and CEO duality are negatively related to firm financial distress. Firms with higher percentage of shares held by the first to 10th largest

**TABLE 3 |** Firm performance.

This table presents the results of firm performance using DID estimations. Firm performance consists of four variables to track firm's development capability, leverage condition, profitability, and financial distress status. The interaction between HP Firms and After is used to measure the impacts of the two revisions of the environmental protection fee. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2017 | | | | Panel B: 2001–2006 | | | |
|---|---|---|---|---|---|---|---|---|
| | **DC** | **LC** | **Profitability** | **Distress** | **DC** | **LC** | **Profitability** | **Distress** |
| HP firms | 0.0613 | −0.0081 | 0.0497 | −0.2137** | 0.0178 | 0.0022 | −0.0293 | 0.2948** |
| | (0.418) | (0.420) | (0.187) | (0.012) | (0.488) | (0.822) | (0.283) | (0.037) |
| After | 0.4477*** | −0.0051 | −0.0556 | 0.2067 | −0.0691* | 0.0155 | −0.0867* | 0.2582 |
| | (0.000) | (0.583) | (0.256) | (0.242) | (0.051) | (0.236) | (0.058) | (0.293) |
| HP*After | −0.4136*** | −0.0143* | −0.0756* | 0.3033* | 0.0239 | −0.0142 | 0.0792** | 0.0710 |
| | (0.000) | (0.090) | (0.087) | (0.066) | (0.421) | (0.305) | (0.039) | (0.749) |
| Board size | −0.2625*** | −0.0213** | 0.0491 | −0.2799** | −0.0002 | −0.0046 | −0.0060 | 0.0481 |
| | (0.006) | (0.031) | (0.293) | (0.047) | (0.995) | (0.592) | (0.852) | (0.799) |
| CEO duality | 0.1713*** | −0.0018 | 0.0349 | −0.1004* | 0.0011 | −0.0024 | −0.0215 | 0.2778* |
| | (0.001) | (0.674) | (0.115) | (0.085) | (0.954) | (0.783) | (0.362) | (0.067) |
| First | 0.8839*** | −0.0411*** | 0.4693*** | −0.1419 | 0.0463 | 0.0508** | 0.1246** | 0.7740* |
| | (0.000) | (0.000) | (0.000) | (0.488) | (0.393) | (0.040) | (0.035) | (0.062) |
| Sedtenth | 1.9991*** | −0.0827*** | 0.7458*** | −0.1377 | 0.1164** | 0.0347 | 0.1862*** | 0.3532 |
| | (0.000) | (0.000) | (0.000) | (0.581) | (0.031) | (0.162) | (0.007) | (0.471) |
| Age | 0.1936*** | −0.0379*** | 0.0171 | −0.1358*** | −0.0584*** | −0.0412*** | −0.1018*** | −0.1560 |
| | (0.000) | (0.000) | (0.371) | (0.005) | (0.000) | (0.000) | (0.000) | (0.124) |
| Size | 0.0124 | −0.0014 | 0.0812*** | −0.0062 | 0.0708*** | −0.0070** | 0.1383*** | −0.2468*** |
| | (0.276) | (0.495) | (0.000) | (0.809) | (0.000) | (0.038) | (0.000) | (0.000) |
| Leverage | −0.0648 | 0.1651*** | −0.7509*** | −0.3159*** | −0.0015 | 0.3608*** | −0.6421*** | −0.1300 |
| | (0.246) | (0.000) | (0.000) | (0.003) | (0.972) | (0.000) | (0.000) | (0.400) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 986 | 2,604 | 2,604 | 2,604 | 413 |
| R–squared | 0.049 | 0.129 | 0.132 | 0.152 | 0.077 | 0.331 | 0.256 | 0.217 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 4 |** Quantile analyses.
This table presents the results of quantile analyses. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

**Panel A: 2012–2017**

| | Development capacity | | | Leverage condition | | | Profitability | | |
|---|---|---|---|---|---|---|---|---|---|
| | q (25) | q (50) | q (75) | q (25) | q (50) | q (75) | q (25) | q (50) | q (75) |
| HP firms | 0.0051 | −0.0102 | −0.0656*** | −0.0047 | −0.0131*** | −0.0146** | 0.0439** | 0.0503** | 0.0115 |
| | (0.579) | (0.396) | (0.007) | (0.440) | (0.002) | (0.010) | (0.011) | (0.014) | (0.768) |
| After | −0.0032 | 0.0331** | 0.1293*** | −0.0111 | −0.0075 | 0.0065 | −0.0311 | −0.0450* | −0.1030** |
| | (0.789) | (0.037) | (0.000) | (0.164) | (0.179) | (0.384) | (0.169) | (0.094) | (0.044) |
| HP*After | −0.0189* | −0.0468*** | −0.1064*** | −0.0031 | 0.0001 | −0.0126** | −0.0113 | −0.0235 | 0.0087 |
| | (0.068) | (0.001) | (0.000) | (0.649) | (0.986) | (0.049) | (0.563) | (0.310) | (0.844) |
| Board size | −0.0194* | −0.0433*** | −0.0746** | −0.0018 | −0.0112** | −0.0238*** | −0.0154 | −0.0204 | −0.0314 |
| | (0.082) | (0.003) | (0.012) | (0.812) | (0.029) | (0.001) | (0.462) | (0.413) | (0.508) |
| CEO duality | 0.0121** | 0.0251*** | 0.0452*** | −0.0025 | −0.0001 | 0.0039 | 0.0025 | 0.0044 | 0.0073 |
| | (0.016) | (0.000) | (0.001) | (0.448) | (0.952) | (0.202) | (0.789) | (0.697) | (0.733) |
| First | −0.0164 | −0.0114 | 0.0440 | −0.0389*** | −0.0188** | −0.0177* | 0.1442*** | 0.2403*** | 0.4198*** |
| | (0.328) | (0.605) | (0.323) | (0.000) | (0.015) | (0.086) | (0.000) | (0.000) | (0.000) |
| Sedtenth | 0.1107*** | 0.2289*** | 0.4942*** | −0.0827*** | −0.0272*** | 0.0013 | 0.2874*** | 0.4904*** | 0.6869*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.003) | (0.913) | (0.000) | (0.000) | (0.000) |
| Age | −0.0329*** | −0.0306*** | −0.0328*** | −0.0261*** | −0.0185*** | −0.0254*** | −0.0441*** | −0.0388*** | −0.0228 |
| | (0.000) | (0.000) | (0.001) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.161) |
| Size | 0.0228*** | 0.0198*** | 0.0283*** | 0.0038*** | −0.0007 | −0.0056*** | 0.0562*** | 0.0752*** | 0.1166*** |
| | (0.000) | (0.000) | (0.000) | (0.003) | (0.414) | (0.000) | (0.000) | (0.000) | (0.000) |
| Leverage | −0.0843*** | −0.0196 | −0.0138 | 0.0682*** | 0.0588*** | 0.1169*** | −0.4748*** | −0.4730*** | −0.5404*** |
| | (0.000) | (0.101) | (0.568) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 |
| R–squared | 0.031 | 0.026 | 0.034 | 0.033 | 0.026 | 0.062 | 0.075 | 0.097 | 0.100 |
| **Panel B: 2001–2006** | | | | | | | | | |
| HP firms | 0.0296* | 0.0216 | 0.0138 | 0.0039 | 0.0031 | 0.0040 | 0.0123 | −0.0251 | −0.0113 |
| | (0.071) | (0.183) | (0.574) | (0.551) | (0.634) | (0.659) | (0.591) | (0.236) | (0.712) |
| After | 0.0127 | −0.0076 | −0.0186 | 0.0236** | 0.0136 | 0.0060 | −0.0167 | −0.0313 | −0.0314 |
| | (0.588) | (0.742) | (0.596) | (0.012) | (0.146) | (0.646) | (0.611) | (0.300) | (0.474) |
| HP*After | −0.0214 | −0.0063 | 0.0068 | −0.0103 | −0.0029 | 0.0043 | 0.0191 | 0.0320 | 0.0472 |
| | (0.295) | (0.756) | (0.823) | (0.204) | (0.720) | (0.706) | (0.503) | (0.224) | (0.216) |
| Board size | 0.0175 | 0.0129 | 0.0177 | −0.0025 | 0.0044 | −0.0070 | −0.0317 | −0.0156 | −0.0427 |
| | (0.331) | (0.465) | (0.509) | (0.722) | (0.540) | (0.482) | (0.206) | (0.500) | (0.203) |
| CEO duality | −0.0135 | −0.0103 | −0.0037 | 0.0012 | 0.0035 | 0.0002 | −0.0081 | 0.0073 | 0.0172 |
| | (0.280) | (0.405) | (0.844) | (0.811) | (0.478) | (0.972) | (0.642) | (0.652) | (0.461) |
| First | −0.0014 | 0.0247 | 0.0728 | −0.0074 | 0.0056 | 0.0426** | 0.1296 | 0.1036** | 0.0848 |
| | (0.966) | (0.460) | (0.149) | (0.580) | (0.680) | (0.023) | (0.784) | (0.018) | (0.180) |
| Sedtenth | 0.0146 | 0.0939** | 0.1330** | −0.0256* | −0.0118 | 0.0468** | 0.0466 | 0.1896*** | 0.2320*** |
| | (0.708) | (0.014) | (0.022) | (0.098) | (0.447) | (0.030) | (0.391) | (0.000) | (0.001) |
| Age | −0.0489*** | −0.0637*** | −0.0816*** | −0.0282*** | −0.0241*** | −0.0182*** | −0.0970*** | −0.1011*** | −0.1364*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Size | 0.0431*** | 0.0462*** | 0.0569*** | 0.0001 | −0.0039** | −0.0090*** | 0.0935*** | 0.0926*** | 0.1299*** |
| | (0.000) | (0.000) | (0.000) | (0.932) | (0.020) | (0.000) | (0.000) | (0.000) | (0.000) |
| Leverage | −0.0872*** | 0.0312* | 0.2028*** | 0.1236*** | 0.1315*** | 0.2013*** | −0.4321*** | −0.3050*** | −0.2798*** |
| | (0.000) | (0.091) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 | 4,870 |
| R–squared | 0.065 | 0.051 | 0.059 | 0.050 | 0.052 | 0.090 | 0.104 | 0.122 | 0.156 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

shareholders generally experience better development capacity and profitability, and lower leverage conditions for the sample period 2012–2017. Meanwhile, ownership concentration factors also have significant impacts on firm performance for the sample period 2001–2006, where a one-unit increase in First (Sedtenth) leads to a 0.1246 (0.1862) unit increase in firm profitability.

Besides, financial fundamentals have different impacts on firm performance between the sample periods of 2012–2017 and 2001–2006. For the sample period of 2012–2017, firm age is positively associated with development capacity, and negatively related to firm leverage condition and financial distress, where a one-unit increase in Age leads to 19.36% increases in Assets

growth, as well as 3.79 and 13.58% decrease in Leverage growth and Distress. Larger firms are more likely to experience better profitability. Leverage is demonstrated to have negative influence on profitability and financial distress, where a one-unit increase in Leverage leads to 75.09 and 31.59% decrease in EPS and Distress. Regarding the sample period of 2001–2006, we find that firms with longer time periods of being listed have lower development capacity, leverage condition, and profitability. Larger firm generally have lower leverage condition, better profitability, and lower possibility of falling into financial distress. To be specific, a one-unit increase in Size leads to 13.83% increases in EPS, as well as 0.70 and 24.68% decrease in Leverage growth and Distress. Leverage is negatively related to firm profitability.

The empirical results in **Table 3** show that the interaction between HP Firms and After is negatively related to firm performance. However, it is possible that these effects are different at various quantiles of performance measures, while the mean effects are negative. To investigate, we further adopt quantile analyses and report the results in **Table 4**. We find that the second revision of environmental protection fee has worse impacts on those firms with higher development capacity and leverage condition (per Panel A), while the influence of the first revision still remains insignificant (per Panel B).

# Robustness Test

## Other Environmental Regulations

In the previous section, we adopt the DID estimation to investigate the impacts of two revisions of the China's Environmental Protection Fee on firm performance. It should be notice that other environmental policies that affect Chinese firms should also be controlled for. Since the authorities implemented a lot of environmental regulations (even not major ones) during these years, it is impossible to control for these policy one by one. Thus, we decide to adopt ER (the ratio of GDP on energy, in terms of ten thousand RMB/ton of standard coal equivalent) to measure the overall levels of Chinese environmental regulation during our sample periods. Higher ER value represents more stringent environmental regulation (Ben Kheder and Zugravu, 2008; Ma et al., 2019). After we control for the impacts of other environmental policies, the empirical results (per **Table 5**) remain the same as those in **Table 3**, which further confirm our previous findings that the second revision increase heavy polluting firms' environmental costs and worsens firm performance, while the first revision seems to have insignificant impacts on firm performance.

## Parallel Trend Assumption

Based on the principle of the DID estimation, there should be a common trend between treated and control groups before the

**TABLE 5 |** Other environmental regulations.

This table presents the results after we control for other environmental policies. ER is adopted to measure the overall levels of Chinese environmental regulation during our sample periods, which equals to the ratio of GDP on energy (in terms of ten thousand RMB/ton of standard coal equivalent). Higher ER value represents more stringent environmental regulation. In this table, we do not control for year fixed effect, as there is a multicollinearity issue between year fixed effect and ER. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2017 | | | | Panel B: 2001–2006 | | | |
|---|---|---|---|---|---|---|---|---|
| | DC | LC | Profitability | Distress | DC | LC | Profitability | Distress |
| HP firms | 0.0616 | −0.0082 | 0.0496 | −0.2141** | 0.0173 | 0.0022 | −0.0293 | 0.2512* |
| | (0.416) | (0.418) | (0.187) | (0.012) | (0.498) | (0.827) | (0.281) | (0.064) |
| After | 0.6625*** | −0.0290* | −0.1814** | −0.4790 | −0.0485* | 0.0035 | −0.0079 | 1.0669*** |
| | (0.000) | (0.054) | (0.026) | (0.103) | (0.075) | (0.772) | (0.821) | (0.000) |
| HP*After | −0.4135*** | −0.0143* | −0.0757* | 0.3111* | 0.0241 | −0.0140 | 0.0783** | 0.1253 |
| | (0.000) | (0.089) | (0.087) | (0.060) | (0.416) | (0.312) | (0.041) | (0.570) |
| ER | −0.8983*** | 0.0997*** | 0.4657** | 2.6732*** | −0.4971*** | −0.1280** | 0.1673 | 7.7324*** |
| | (0.006) | (0.002) | (0.010) | (0.000) | (0.001) | (0.037) | (0.452) | (0.000) |
| Board size | −0.2620*** | −0.0214** | 0.0502 | −0.2769* | −0.0035 | −0.0059 | −0.0029 | 0.1073 |
| | (0.006) | (0.030) | (0.282) | (0.050) | (0.892) | (0.496) | (0.929) | (0.557) |
| CEO duality | 0.1711*** | −0.0017 | 0.0348 | −0.1014* | 0.0029 | −0.0023 | −0.0204 | 0.3018** |
| | (0.001) | (0.678) | (0.116) | (0.083) | (0.875) | (0.795) | (0.386) | (0.043) |
| First | 0.8836*** | −0.0411*** | 0.4703*** | −0.1396 | 0.0225 | 0.0488* | 0.1123* | 0.7775* |
| | (0.000) | (0.000) | (0.000) | (0.496) | (0.676) | (0.053) | (0.057) | (0.064) |
| Sedtenth | 2.0010*** | −0.0829*** | 0.7431*** | −0.1424 | 0.1018* | 0.0327 | 0.1822*** | 0.3174 |
| | (0.000) | (0.000) | (0.000) | (0.569) | (0.060) | (0.193) | (0.009) | (0.511) |
| Age | 0.1934*** | −0.0379*** | 0.0164 | −0.1369*** | −0.0560*** | −0.0420*** | −0.0956*** | −0.0605 |
| | (0.000) | (0.000) | (0.391) | (0.004) | (0.000) | (0.000) | (0.000) | (0.538) |
| Size | 0.0124 | −0.0014 | 0.0808*** | −0.0073 | 0.0720*** | −0.0071** | 0.1399*** | −0.1926*** |
| | (0.274) | (0.493) | (0.000) | (0.778) | (0.000) | (0.036) | (0.000) | (0.003) |
| Leverage | −0.0653 | 0.1651*** | −0.7495*** | −0.3159*** | 0.0011 | 0.3604*** | −0.6377*** | -0.0501 |
| | (0.243) | (0.000) | (0.000) | (0.003) | (0.981) | (0.000) | (0.000) | (0.779) |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 986 | 2,604 | 2,604 | 2,604 | 413 |
| R-squared | 0.049 | 0.129 | 0.133 | 0.154 | 0.080 | 0.332 | 0.254 | 0.217 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 6 |** Robustness test-parallel trend assumption.

This table presents the results of robustness test-parallel trend assumption. After* is adopted to replace After and re-estimate the influence of environmental protection fee on firm performance for the sample periods of 2012–2014 and 2001–2003. After* is a dummy variable, which equals 1 for firm-year observations after 2013 and 2002, and 0 otherwise. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2014 | | | | Panel B: 2001–2003 | | | |
|---|---|---|---|---|---|---|---|---|
| | DC | LC | Profitability | Distress | DC | LC | Profitability | Distress |
| HP firms | 0.0814 | 0.0013 | 0.0090 | −0.3361** | 0.0352 | 0.0177 | −0.0001 | 0.2996 |
| | (0.323) | (0.949) | (0.863) | (0.018) | (0.367) | (0.401) | (0.997) | (0.307) |
| After* | 0.1517 | 0.0111 | 0.0050 | 0.0909 | −0.0755 | 0.0140 | −0.0764 | 0.4658 |
| | (0.171) | (0.522) | (0.924) | (0.609) | (0.138) | (0.533) | (0.123) | (0.220) |
| HP*After* | 0.0161 | −0.0143 | −0.0091 | 0.1769 | 0.0140 | −0.0256 | 0.0290 | −0.0095 |
| | (0.834) | (0.390) | (0.841) | (0.255) | (0.763) | (0.243) | (0.496) | (0.977) |
| Board size | −0.2216* | −0.0013 | 0.0616 | −0.3701** | −0.0580 | −0.0155 | −0.0071 | 0.0679 |
| | (0.092) | (0.928) | (0.328) | (0.049) | (0.129) | (0.153) | (0.831) | (0.818) |
| CEO duality | 0.1963** | −0.0078 | 0.0058 | −0.1150 | 0.0219 | 0.0032 | 0.0098 | 0.3923* |
| | (0.047) | (0.262) | (0.823) | (0.143) | (0.404) | (0.772) | (0.700) | (0.082) |
| First | 1.0604*** | −0.0485*** | 0.3986*** | −0.2437 | −0.0729 | 0.0026 | 0.0233 | 1.0370 |
| | (0.001) | (0.003) | (0.000) | (0.372) | (0.380) | (0.908) | (0.752) | (0.100) |
| Sedtenth | 2.2170*** | −0.1203*** | 0.8815*** | −0.2834 | −0.0009 | −0.0061 | 0.0877 | 0.5827 |
| | (0.000) | (0.000) | (0.000) | (0.406) | (0.992) | (0.825) | (0.324) | (0.430) |
| Age | 0.2510*** | −0.0518*** | 0.0218 | −0.1416** | −0.0510*** | −0.0337*** | −0.0789*** | −0.1664 |
| | (0.005) | (0.000) | (0.494) | (0.033) | (0.007) | (0.000) | (0.000) | (0.301) |
| Size | −0.0012 | −0.0001 | 0.0775*** | −0.0048 | 0.0627*** | −0.0004 | 0.1271*** | −0.2647*** |
| | (0.941) | (0.979) | (0.000) | (0.892) | (0.000) | (0.919) | (0.000) | (0.009) |
| Leverage | −0.0246 | 0.1528*** | −0.6221** | −0.4736*** | 0.1376 | 0.2147*** | −0.5764*** | −0.4518 |
| | (0.642) | (0.000) | (0.015) | (0.002) | (0.105) | (0.000) | (0.000) | (0.206) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 2,324 | 2,324 | 2,324 | 740 | 1,169 | 1,169 | 1,169 | 272 |
| R–squared | 0.040 | 0.142 | 0.164 | 0.051 | 0.057 | 0.141 | 0.252 | 0.074 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

implementation of a particular policy. To further verify the main findings in *Firm Performance*, we artificially assume the two revisions of the environmental protection fee were in 2013 and 2002 and replace the sample periods with 2012–2014 and 2001–2003. Another independent variable, namely After* (dummy variable, which equals 1 for firm-year observations after 2013 and 2002, and 0 otherwise), is adopted for empirical analyses. If the interaction between HP Firms and After* is still significantly related to firm performance, then there is no common trend between treated and control groups before the implementation of a particular policy. As per **Table 6**, the interaction between HP Firms and After* is insignificantly related to firm performance for both of the sample periods, which demonstrates a common trend between our treated and control groups during 2012–2014 and 2001–2003. Thus, the empirical results provide robustness for our main findings.

### Environmental Protection Fee Rate

In the previous section, we adopt HP Firms and After to construct DID estimations, and explore the impacts of environmental protection fee on firm performance. After is a dummy variable which takes a value of 1 for firm-year observation after the revisions of environmental protection fee and 0 otherwise. For robustness purpose, we further adopt the environmental protection fee rate to replace After and re-estimate the influence of this policy on firm performance[3].

As shown in **Table 7**, our main findings hold when we use the alternative indicator to estimate the impact of environmental protection fee on firm performance. Specifically, the interaction between HP Firms and Fee Rate is negatively related to firms' development capacity and positively associated with financial distress for the sample period of 2012–2017. The influence of the first revision of the environmental protection fee (2001–2006) still remains insignificant. These results further confirm our previous finding that the second revision of environmental protection fee increase firms' environmental costs—especially for heavy polluting firms—and worsens firm performance, while the first revision seems to have insignificant impacts on firm performance.

## Endogeneity Tests

In the previous section, we control for firm fundamental factors when we explore the impacts of environmental protection fee on the firm performance. Nevertheless, firms with better performance may due to their better fundamental conditions rather than other factors (Albornoz et al., 2009; Cole et al., 2013).

[3]As mentioned above, there are different fee rates for various water and gas pollutants. We apply water and gas fee rates to replace After and achieve similar results. For brevity purpose, we only report the results of water fee in **Table 7**.

**TABLE 7 |** Robustness test-environmental protection fee rate.

This table presents the results of robustness test-environmental protection fee rate. The environmental protection fee rate is adopted to replace After and re-estimate the influence of environmental protection fee on firm performance. There are different fee rates for various water and gas pollutants. We apply all water and gas fee rates to replace After and achieve similar results. For brevity purpose, we only report the results of water fee. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2017 | | | | Panel B: 2001–2006 | | | |
|---|---|---|---|---|---|---|---|---|
| | DC | LC | Profitability | Distress | DC | LC | Profitability | Distress |
| HP firms | −0.0986 | −0.0121 | 0.0114 | −0.4810*** | 0.0178 | 0.0023 | −0.0293 | 0.2947** |
| | (0.256) | (0.194) | (0.757) | (0.001) | (0.488) | (0.821) | (0.283) | (0.037) |
| Fee rate | 0.0042 | 0.0013 | −0.0006 | 0.0922 | −0.0988* | 0.0222 | −0.1239* | 0.3689 |
| | (0.648) | (0.344) | (0.915) | (0.261) | (0.051) | (0.236) | (0.058) | (0.293) |
| HP*Fee rate | −0.0267** | −0.0017 | 0.0015 | 0.3664*** | 0.0341 | −0.0204 | 0.1131** | 0.1015 |
| | (0.010) | (0.330) | (0.836) | (0.006) | (0.421) | (0.305) | (0.039) | (0.749) |
| Board size | −0.2649*** | −0.0215** | 0.0478 | −0.2798** | −0.0002 | −0.0046 | −0.0060 | 0.0481 |
| | (0.006) | (0.030) | (0.306) | (0.049) | (0.995) | (0.592) | (0.852) | (0.799) |
| CEO duality | 0.1753*** | −0.0015 | 0.0359 | −0.1004* | 0.0011 | −0.0024 | −0.0215 | 0.2778* |
| | (0.001) | (0.720) | (0.106) | (0.087) | (0.954) | (0.783) | (0.362) | (0.067) |
| First | 0.8788*** | −0.0408*** | 0.4664*** | −0.1529 | 0.0463 | 0.0508** | 0.1246** | 0.7740* |
| | (0.000) | (0.000) | (0.000) | (0.459) | (0.393) | (0.040) | (0.035) | (0.062) |
| Sedtenth | 2.0003*** | −0.0821*** | 0.7442*** | −0.1222 | 0.1164** | 0.0347 | 0.1862*** | 0.3532 |
| | (0.000) | (0.000) | (0.000) | (0.626) | (0.031) | (0.162) | (0.007) | (0.471) |
| Age | 0.1896*** | −0.0376*** | 0.0173 | −0.1282*** | −0.0584*** | −0.0412*** | −0.1018*** | −0.1560 |
| | (0.000) | (0.000) | (0.362) | (0.008) | (0.000) | (0.000) | (0.000) | (0.124) |
| Size | 0.0210* | −0.0013 | 0.0821*** | −0.0104 | 0.0708*** | −0.0070** | 0.1383*** | −0.2468*** |
| | (0.072) | (0.542) | (0.000) | (0.686) | (0.000) | (0.038) | (0.000) | (0.000) |
| Leverage | −0.0733 | 0.1650*** | −0.7512*** | −0.3258*** | −0.0015 | 0.3608*** | −0.6421*** | −0.1300 |
| | (0.190) | (0.000) | (0.000) | (0.002) | (0.972) | (0.000) | (0.000) | (0.400) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 986 | 2,604 | 2,604 | 2,604 | 413 |
| R–squared | 0.044 | 0.128 | 0.130 | 0.152 | 0.077 | 0.331 | 0.256 | 0.217 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 8 |** Endogeneity test–sector average.

This table presents the results of endogeneity test–sector average. We utilize sector-level variables (AveAge, AveSize, and AveLeverage) to replace firm-level variables and repeat our empirical analyses. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2017 | | | | Panel B: 2001–2006 | | | |
|---|---|---|---|---|---|---|---|---|
| | DC | LC | Profitability | Distress | DC | LC | Profitability | Distress |
| HP firms | 0.1126 | 0.0014 | 0.1392*** | −0.2091** | 0.0379 | 0.0122 | −0.0099 | 0.1790 |
| | (0.181) | (0.912) | (0.000) | (0.015) | (0.141) | (0.270) | (0.739) | (0.192) |
| After | 0.4693*** | −0.0045 | −0.0378 | 0.1443 | −0.0699* | 0.0134 | −0.0883* | 0.1985 |
| | (0.000) | (0.645) | (0.432) | (0.416) | (0.057) | (0.344) | (0.077) | (0.408) |
| HP*After | −0.4373*** | −0.0152* | −0.1034** | 0.3008* | 0.0264 | −0.0140 | 0.0839** | 0.0880 |
| | (0.000) | (0.081) | (0.015) | (0.072) | (0.394) | (0.385) | (0.045) | (0.690) |
| Board size | −0.2118** | −0.0088 | 0.1426*** | −0.3439** | 0.0573** | −0.0289** | 0.1350*** | −0.0521 |
| | (0.021) | (0.415) | (0.002) | (0.014) | (0.027) | (0.024) | (0.000) | (0.783) |
| CEO duality | 0.1373*** | 0.0023 | 0.0209 | −0.0793 | −0.0018 | 0.0058 | −0.0408 | 0.2775* |
| | (0.005) | (0.601) | (0.351) | (0.170) | (0.925) | (0.600) | (0.129) | (0.065) |
| First | 0.7662*** | −0.0112 | 0.8134*** | 0.0495 | 0.1553*** | 0.0419 | 0.3875*** | 0.7895** |
| | (0.000) | (0.341) | (0.000) | (0.784) | (0.004) | (0.119) | (0.000) | (0.047) |
| Sedtenth | 1.6769*** | −0.0641*** | 1.0177*** | 0.2295 | 0.1356** | 0.0913** | 0.1762** | 0.4972 |
| | (0.000) | (0.000) | (0.000) | (0.287) | (0.015) | (0.023) | (0.025) | (0.299) |
| AveAge | 0.2770** | −0.0309*** | 0.0169 | −0.1961** | −0.0157 | −0.0561** | −0.0295 | −0.3211** |
| | (0.021) | (0.001) | (0.669) | (0.017) | (0.672) | (0.015) | (0.662) | (0.043) |
| AveSize | −0.0010 | −0.0081 | 0.0141 | 0.0365 | 0.0573*** | −0.0053 | 0.1103*** | −0.1256 |
| | (0.972) | (0.209) | (0.422) | (0.407) | (0.000) | (0.306) | (0.000) | (0.170) |
| AveLeverage | −0.5186* | 0.1139 | −1.1292*** | −0.6320*** | −0.0123 | 0.3736** | −0.6640*** | −0.0406 |
| | (0.057) | (0.101) | (0.000) | (0.002) | (0.925) | (0.023) | (0.000) | (0.858) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4,870 | 4,870 | 4,870 | 986 | 2,604 | 2,604 | 2,604 | 413 |
| R–squared | 0.043 | 0.016 | 0.081 | 0.150 | 0.036 | 0.042 | 0.086 | 0.196 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 9 |** Normal firm performance.

This table presents the results of normal firm performance. We define normal firms as those neither belong to heavy polluters nor belong to environmental friendly firms. Normal Firms is a dummy variable which equals 1 if a firm is neither heavy polluting firms nor environmental friendly firms and 0 if it is an environmental friendly firm. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| | Panel A: 2012–2017 | | | | Panel B: 2001–2006 | | | |
|---|---|---|---|---|---|---|---|---|
| | **DC** | **LC** | **Profitability** | **Distress** | **DC** | **LC** | **Profitability** | **Distress** |
| Normal firms | −0.0438 | −0.0046 | −0.0078 | −0.0242 | −0.0149 | −0.0053 | −0.0154 | 0.1862 |
| | (0.261) | (0.508) | (0.718) | (0.653) | (0.503) | (0.565) | (0.500) | (0.106) |
| After | 0.4582*** | −0.0086*** | −0.0043 | 0.2241 | −0.0651** | 0.0147 | −0.0986** | 0.2901 |
| | (0.000) | (0.338) | (0.921) | (0.129) | (0.038) | (0.320) | (0.019) | (0.169) |
| Normal*After | −0.2804*** | −0.0041 | −0.0559 | 0.1681 | 0.0294 | −0.0098 | 0.0464 | 0.0512 |
| | (0.001) | (0.629) | (0.171) | (0.232) | (0.296) | (0.518) | (0.211) | (0.793) |
| Board size | −0.2339*** | 0.0033 | −0.0850*** | −0.0030 | −0.0169 | −0.0019 | 0.0441 | −0.0460 |
| | (0.003) | (0.578) | (0.004) | (0.974) | (0.392) | (0.830) | (0.103) | (0.706) |
| CEO duality | 0.1109*** | 0.0003 | 0.0063 | −0.0107 | 0.0070 | 0.0123* | −0.0276 | 0.1000 |
| | (0.003) | (0.907) | (0.656) | (0.784) | (0.610) | (0.060) | (0.228) | (0.280) |
| First | 0.6133*** | −0.0152 | 0.4796*** | 0.3277** | 0.0407 | 0.0106 | 0.2093*** | 0.2450 |
| | (0.000) | (0.114) | (0.000) | (0.012) | (0.311) | (0.642) | (0.000) | (0.341) |
| Sedtenth | 1.7840*** | −0.0610*** | 0.6090*** | 0.5140*** | 0.1417*** | −0.0090 | 0.2341*** | −0.4190 |
| | (0.000) | (0.000) | (0.000) | (0.002) | (0.002) | (0.693) | (0.000) | (0.160) |
| Age | 0.2031*** | −0.0267*** | −0.0013 | −0.1378*** | −0.0689*** | −0.0382*** | −0.0610*** | −0.1713*** |
| | (0.000) | (0.000) | (0.912) | (0.000) | (0.000) | (0.000) | (0.000) | (0.007) |
| Size | 0.0544*** | −0.0017 | 0.1509*** | 0.0163 | 0.0770*** | −0.0055* | 0.1205*** | −0.1285*** |
| | (0.000) | (0.676) | (0.000) | (0.310) | (0.000) | (0.077) | (0.000) | (0.000) |
| Leverage | −0.0742 | 0.1059** | −0.4183*** | −0.3428*** | −0.0398 | 0.2839*** | −0.5422*** | 0.0888 |
| | (0.318) | (0.031) | (0.000) | (0.000) | (0.031) | (0.000) | (0.000) | (0.309) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 8,813 | 8,813 | 8,813 | 1,937 | 4,891 | 4,891 | 4,891 | 1,097 |
| R–squared | 0.033 | 0.058 | 0.154 | 0.143 | 0.074 | 0.266 | 0.209 | 0.158 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 10 |** Subsample tests-Politically connected and non-connected firms.

This table presents the results of sub-sample tests: Politically connected and non-connected firms. We define a firm as politically connected if the CEO or Chair of Board is or was a government official. For the sample period of 2001–2006, we find that the environmental protection fee has insignificant influence on firm performance for various sub-samples. In this case, we only report the results of the sample period of 2012–2017 in this section for brevity purpose. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| **2012–2017** | Panel A: PC | | | | Panel B: Non-PC | | | |
|---|---|---|---|---|---|---|---|---|
| | **DC** | **LC** | **Profitability** | **Distress** | **DC** | **LC** | **Profitability** | **Distress** |
| HP firms | 0.1069 | −0.0171 | 0.0953** | −0.4141** | 0.0174 | −0.0032 | 0.0316 | −0.1523 |
| | (0.467) | (0.187) | (0.020) | (0.019) | (0.838) | (0.826) | (0.510) | (0.141) |
| After | 0.2191** | −0.0156 | −0.0434 | 0.2827 | 0.5403*** | 0.0018 | −0.0667 | 0.1785 |
| | (0.035) | (0.297) | (0.535) | (0.322) | (0.000) | (0.888) | (0.301) | (0.442) |
| HP*After | −0.2441** | −0.0050 | −0.0084 | 0.4244 | −0.4787*** | −0.0199* | −0.0958 | 0.2424 |
| | (0.011) | (0.667) | (0.887) | (0.135) | (0.000) | (0.096) | (0.104) | (0.265) |
| Board size | −0.1791 | 0.0013 | −0.0756 | −0.0849 | −0.2912** | −0.0322** | 0.1037* | −0.4000** |
| | (0.245) | (0.931) | (0.330) | (0.698) | (0.019) | (0.013) | (0.077) | (0.031) |
| CEO duality | 0.1818 | 0.0006 | −0.0222 | −0.1310 | 0.1568*** | −0.0028 | 0.0747** | −0.0912 |
| | (0.102) | (0.926) | (0.417) | (0.204) | (0.002) | (0.577) | (0.011) | (0.214) |
| First | 1.2563*** | −0.0366* | 0.3057*** | −0.0379 | 0.7400*** | −0.0386*** | 0.5525*** | −0.1368 |
| | (0.005) | (0.055) | (0.009) | (0.905) | (0.002) | (0.008) | (0.000) | (0.611) |
| Sedtenth | 2.3092*** | −0.0843** | 0.4488*** | −0.4470 | 1.8535*** | −0.0829*** | 0.8814*** | 0.0586 |
| | (0.004) | (0.020) | (0.001) | (0.262) | (0.000) | (0.000) | (0.000) | (0.860) |
| Age | 0.2228** | −0.0346*** | 0.0203 | −0.1796** | 0.1900*** | −0.0387*** | 0.0228 | −0.0757 |
| | (0.027) | (0.000) | (0.396) | (0.020) | (0.000) | (0.000) | (0.364) | (0.231) |
| Size | −0.0274 | −0.0041 | 0.1275*** | −0.0731 | 0.0299*** | −0.0004 | 0.0594*** | 0.0195 |
| | (0.313) | (0.258) | (0.000) | (0.144) | (0.009) | (0.889) | (0.002) | (0.536) |
| Leverage | 0.0319 | 0.1561*** | −1.1007*** | −0.2092 | −0.0934 | 0.1691*** | −0.6344*** | −0.3895*** |
| | (0.811) | (0.000) | (0.000) | (0.216) | (0.145) | (0.000) | (0.003) | (0.005) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,695 | 1,695 | 1,695 | 360 | 3,175 | 3,175 | 3,175 | 626 |
| R–squared | 0.044 | 0.107 | 0.200 | 0.161 | 0.056 | 0.141 | 0.116 | 0.159 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

**TABLE 11 |** Subsample tests-SOEs and non-SOEs.

This table presents the results of sub-sample tests: SOEs and non-SOEs. For the sample period of 2001–2006, we find that the environmental protection fee has insignificant influence on firm performance for various sub-samples. In this case, we only report the results of the sample period of 2012–2017 in this section for brevity purpose. For the probit regression (column Distress), we report the marginal effect rather than the coefficients. For the definitions of variables refer to **Appendix A**. *P*-values are in parentheses.

| 2012–2017 | Panel A: SOE | | | | Panel B: Non-SOE | | | |
|---|---|---|---|---|---|---|---|---|
| | DC | LC | Profitability | Distress | DC | LC | Profitability | Distress |
| HP firms | 0.0885 | −0.0159 | −0.0806 | −0.7657* | 0.0437 | −0.0044 | 0.0718* | −0.1792** |
| | (0.691) | (0.621) | (0.305) | (0.052) | (0.581) | (0.674) | (0.072) | (0.036) |
| After | 0.2895 | −0.0074 | 0.1063 | 0.4787 | 0.4740*** | −0.0047 | −0.0866 | 0.1475 |
| | (0.259) | (0.786) | (0.381) | (0.365) | (0.000) | (0.642) | (0.102) | (0.458) |
| HP*After | −0.0829 | −0.0313 | −0.2100* | 0.3714 | −0.4626*** | −0.0116 | −0.0513 | 0.3178* |
| | (0.672) | (0.208) | (0.065) | (0.421) | (0.000) | (0.192) | (0.278) | (0.095) |
| Board size | −0.4068 | 0.0050 | −0.0195 | 0.0972 | −0.2157** | −0.0244** | 0.0811 | −0.2739* |
| | (0.145) | (0.844) | (0.868) | (0.839) | (0.029) | (0.022) | (0.112) | (0.065) |
| CEO duality | −0.1217 | 0.0082 | −0.0561 | 0.3472 | 0.1985*** | −0.0041 | 0.0442* | −0.1496** |
| | (0.110) | (0.481) | (0.280) | (0.180) | (0.000) | (0.339) | (0.057) | (0.014) |
| First | 2.3492*** | −0.0704* | 0.6601*** | 1.1333 | 0.5686*** | −0.0374*** | 0.4581*** | −0.1961 |
| | (0.004) | (0.063) | (0.000) | (0.107) | (0.003) | (0.002) | (0.000) | (0.375) |
| Sedtenth | 1.8989*** | −0.1834*** | 0.5421*** | 1.2466 | 2.0020*** | −0.0644*** | 0.8208*** | −0.2858 |
| | (0.008) | (0.000) | (0.008) | (0.169) | (0.000) | (0.002) | (0.000) | (0.281) |
| Age | 0.2743*** | −0.0209*** | −0.0138 | 0.0115 | 0.1685*** | −0.0391*** | 0.0301 | −0.1479*** |
| | (0.001) | (0.004) | (0.709) | (0.937) | (0.002) | (0.000) | (0.175) | (0.005) |
| Size | −0.0719 | 0.0132* | 0.0624*** | −0.1281 | 0.0197* | −0.0024 | 0.0862*** | 0.0080 |
| | (0.155) | (0.058) | (0.005) | (0.148) | (0.081) | (0.253) | (0.000) | (0.777) |
| Leverage | 0.2894 | 0.0570 | −0.5784*** | 0.3841 | −0.0812 | 0.1785*** | −0.7560*** | −0.3889*** |
| | (0.224) | (0.274) | (0.000) | (0.267) | (0.176) | (0.000) | (0.000) | (0.002) |
| Year | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 688 | 688 | 688 | 145 | 4,182 | 4,182 | 4,182 | 841 |
| R–squared | 0.046 | 0.063 | 0.154 | 0.176 | 0.060 | 0.152 | 0.137 | 0.161 |

*Notes: ***, ** and * refer to the significance at the 1, 5 and 10% levels, respectively.*

To control for the potential endogeneity issue, we replace firm-level with sector-level variables (AveAge, AveSize, and AveLeverage) and repeat our empirical analyses. **Table 8** shows that our empirical results almost remain the same as those in **Table 3**, where the second revision of environmental protection fee increases firms' environmental costs and worsens firm performance (Panel A: 2012–2017), while the first revision seems to have insignificant impact on firm performance (Panel B: 2001–2006). Corporate governance conditions and firm financial fundamentals also have significant impacts on firm performance. These results suggest that our findings are robust to the potential endogeneity issue.

## Normal Firms

Regarding the second research question, we discuss whether the two revisions of the environmental protection fee are efficient environmental regulations. As mentioned in *Institutional Background-The Environmental Protection Fee*, the Environmental Protection Fee policy was designed to follow the polluter-pays principle, in other words, firms were charged by this policy only if they released pollutants into the environment. The charge amounts is based on the amounts of pollutants firms released into the environment. Thus, we define an efficient environmental regulation as a policy has stronger negative impact on heavy polluters than other firms. To investigate, we categorize sample firms into three groups: 1) heavy polluting firms; 2) normal firms; and 3) environmental friendly firms. We define normal firms as those neither belong to heavy polluters nor belong to environmental friendly firms. Obviously, heavy polluters release more pollutants than normal firms. In the previous empirical analyses, we apply heavy polluters and environmental friendly firms as our treated and control groups, and exclude normal firms. In this section, we further use normal firms to replace heavy polluters and re-estimate the impacts of environmental protection fee and other proposed factors on firm performance. Normal Firms (dummy variable, which equals 1 if a firm is neither heavy polluting firms nor environmental friendly firms and 0 if it is an environmental friendly firm) is adopted for empirical analyses, which is a. We observe 1,382 and 823 normal firms for the sample periods of 2012–2017 and 2001–2006, respectively.

From **Table 9**, we find that the interaction between Normal Firms and After is insignificantly related to firm performance for both of the sample periods, with a negative impact of the interaction term on firm development capacity during the sample period of 2012–2017 being the exception. Thus, the findings support our previous argument that the environmental protection fee is an efficient environmental regulation, which precisely governs heavy polluters rather than other firms. It is worthwhile mentioning that the above empirical analyses also provide robustness for our main tests in *Firm Performance*, as the results become insignificant after we replace the treated group.

## Additional Tests

For the last research question, we divide our firms into several sub-samples, including political connected and non-connected firms, as well as state owned enterprises (SOEs) and non-SOEs[4], and re-estimate the different impacts of the environmental protection fee on corporate performance for various firm characteristics.[5]

### Sub-Sample Tests: Politically Connected and Non-Connected Firms

Previous studies (Pargal and Wheeler, 1996; Fryxell et al., 2004) state that political connection plays an important role on corporate operation when environmental regulations come into force. For instance, Fryxell et al. (2004) suggest that compliance with environmental regulation is somewhat lower for firms with political connections than other firms. Moreover, politically connected firms are more likely to acquire funding support from the government to improve their production equipment, and thus achieve lower pollution intensity than those without connections (Pargal and Wheeler, 1996). In this section, we divide our sample into two sub-samples, including connected and non-connected firms, and re-estimate the impacts of environmental protection fee and other proposed factors on firm performance. Following Wu et al. (2012) and Pang and Wang (2020), we define a firm as politically connected if the CEO or Chair of Board is or was a government official[6].

As shown in **Table 10**, environmental protection fee seems to have worse impact on non-connected firms' development capacity than those of connected firms, as the coefficients of HP Firms*After are −0.4787 and −0.2441, respectively. Nevertheless, the environmental protection fee also worsens non-connected firms' leverage conditions. Thus, the empirical results suggest that the government should provide more support to relieve the negative impacts of environmental protection fee on non-connected firms.

### Sub-Sample Tests: State Owned Enterprises and Non-SOEs

Another firm characteristic that concerns us is state ownership. China experiences a national-wide economic restructuring since the reform and opening-up policy in 1978, shifting from a centrally planned to a market-oriented economy. The unique re-establishment of China's stock markets in the early 1990s assisted SOEs to undertake reforms to raise capital. As a result, SOEs obtain a competitive advantage against non-SOEs, as the government becoming both a major stockholder[7] (in terms of state ownership) and also the regulator of Chinese stock markets. To examine the different impacts of environmental protection fee, we separate our sample into two sub-samples, including SOEs and non-SOEs.

The empirical results show that the interaction term (HP Firms*After) is insignificantly related to the firm performance of SOEs over the sample period of 2012–2017 (Panel A of **Table 11**), with a negative impact of the interaction term on firm profitability being the exception. Meanwhile, the environmental protection fee is negatively associated with non-SOEs' development capacity, and positively related to the likelihood of financial distress (Panel B of **Table 11**). These results further confirm our argument that SOEs have competitive advantage against non-SOEs. Since non-SOEs are also an important part of the national economy, the Chinese government should provide more financial support and assist them to relieve the negative impacts of environmental regulations.

## CONCLUSION

There is a long-time debate about the relationship between environmental regulation and firm performance. Our study contributes to the literature by furthering the discussion and exploring the influence of environmental regulation in this important transition economy. With two separate sample periods (2001–2006 and 2012–2017) and DID estimation, we answer three major research questions as discussed above. Our empirical results show a negative relationship between the second revision of environmental protection fee and firm performance, as heavy polluting firms had worse development capacity, leverage conditions and profitability, and higher likelihood of falling into financial distress than environmental friendly firms after 2015. Nevertheless, the first revision in 2003 seems to have insignificant influence on firm performance. Besides, the environmental protection fee has insignificantly influence on normal firms for both of the sample periods, implying that the environmental protection fee is an efficient environmental regulation as such a policy precisely governs heavy polluters. Regarding our additional tests, we find that the environmental protection fee has worse impacts on non-connected firms and non-SOEs. Thus, the Chinese government should provide more financial support to assist these firms in transformation and upgrading.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data that support the findings of this study are available from the CSMAR database. Restrictions

[4]We also consider splitting sample firms into three sub–samples: three main industries, to further explore the different influences of the environmental protection fee on corporate performance. However, we observe too few observations, especially for sub–samples of primary (Agriculture) and tertiary (Service) industries, to complete empirical analyses.

[5]For the sample period of 2001–2006, we find that the environmental protection fee has insignificant influence on firm performance for various sub-samples, which further confirms our previous findings in *Firm Performance*. In this case, we only report the results of the sample period of 2012–2017 in this section for brevity purpose.

[6]Political connection refers to a type of relationship with Chinese authorities, such as Chinese government, China Communist Party Committee, Central Military Commission, Chinese People's Political Consultative Conference (CPPCC) and National People's Congress.

[7]More than 80% of listed firms were SOEs before 2004 and they were still the most powerful participants in markets after the Non-Tradable Share (NTS) reform in 2005. Details of NTS reform please refer to Liu and Tian (2012).

apply to the availability of these data, which were used under license for this study. Data are available from Huan Zheng (Chongqing Jiaotong University) and Yu He (Chongqing Technology and Business University) with the permission of the CSMAR database. Requests to access these datasets should be directed to Huan Zheng, huan.zheng@cqjtu.edu.cn; Yu He, yu.he@ctbu.edu.cn.

## AUTHOR CONTRIBUTIONS

All authors contributed to conception and design of the study, as well as organized the database. HZ performed the statistical analysis. YH wrote the first draft of the manuscript. All authors wrote sections of the manuscript. Also, all authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This work was supported by the National Social Science Fund of China (19CJY022).

## ACKNOWLEDGMENTS

The authors wish to thank the members of the "Innovation Team of Compound Ecosystem Management in Upper Reaches of Yangtze River" for their comments and support.

## REFERENCES

Albornoz, F., Cole, M. A., Elliott, R. J. R., and Ercolani, M. G. (2009). In Search of Environmental Spillovers. *The World Economy* 32 (1), 136–163. doi:10.1111/j.1467-9701.2009.01160.x

Barbera, A. J., and McConnell, V. D. (1990). The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects. *J. Environ. Econ. Manag.* 18 (1), 50–65. doi:10.1016/0095-0696(90)90051-y

Ben Kheder, S., and Zugravu, N. (2008). *The Pollution haven Hypothesis: a Geographic Economy Model in a Comparative Study*. FEEM Working Paper No 73, 2008.

Bhattacharjee, A., and Han, J. (2014). Financial Distress of Chinese Firms: Microeconomic, Macroeconomic and Institutional Influences. *China Econ. Rev.* 30, 244–262. doi:10.1016/j.chieco.2014.07.007

Bushnell, J. B., Chong, H., and Mansur, E. T. (2013). Profiting from Regulation: Evidence from the European Carbon Market. *Am. Econ. J. Econ. Pol.* 5 (4), 78–106. doi:10.1257/pol.5.4.78

Chan, K. C., and Chen, N.-F. (1991). Structural and Return Characteristics of Small and Large Firms. *J. Finance* 46 (4), 1467–1484. doi:10.1111/j.1540-6261.1991.tb04626.x

Cheng, S. (2008). Board Size and the Variability of Corporate Performance. *J. Financial Econ.* 87, 157–176. doi:10.1016/j.jfineco.2006.10.006

Cole, M. A., Elliott, R. J. R., Okubo, T., and Zhou, Y. (2013). The Carbon Dioxide Emissions of Firms: A Spatial Analysis. *J. Environ. Econ. Manag.* 65 (2), 290–309. doi:10.1016/j.jeem.2012.07.002

Dean, J. M., Lovely, M. E., and Wang, H. (2009). Are Foreign Investors Attracted to Weak Environmental Regulations? Evaluating the Evidence from China. *J. Develop. Econ.* 90, 1–13. doi:10.1016/j.jdeveco.2008.11.007

Earle, J. S., Kucsera, C., and Telegdy, A. (2005). Ownership Concentration and Corporate Performance on the Budapest Stock Exchange: Do Too many Cooks Spoil the Goulash? *Corporate Governance* 13 (2), 254–264. doi:10.1111/j.1467-8683.2005.00420.x

Fan, J. P. H., Huang, J., and Zhu, N. (2013). Institutions, Ownership Structures, and Distress Resolution in China. *J. Corporate Finance* 23, 71–87. doi:10.1016/j.jcorpfin.2013.07.005

Fryxell, G. E., Chung, S. S., and Lo, C. W. H. (2004). Does the Selection of ISO 14001 Registrars Matter? Registrar Reputation and Environmental Policy Statements in China. *J. Environ. Manage.* 71 (1), 45–57. doi:10.1016/j.jenvman.2004.01.005

Gray, W. B., and Shadbegian, R. J. (2003). Plant Vintage, Technology, and Environmental Regulation. *J. Environ. Econ. Manag.* 46 (3), 384–402. doi:10.1016/s0095-0696(03)00031-7

Greenstone, M., List, J. A., and Syverson, C. (2012). *The Effects of Environmental Regulation on the Competitiveness of U.S. Manufacturing*. NBER Working Paper no. 18392, September. doi:10.3386/w18392

Horváthová, E. (2010). Does Environmental Performance Affect Financial Performance? A Meta-Analysis. *Ecol. Econ.* 70 (1), 52–59. doi:10.1016/j.ecolecon.2010.04.004

Huang, X., and Liu, X. (2019). The Impact of Environmental Regulation on Productivity and Exports: A Firm-Level Evidence from China. *Emerging Markets Finance and Trade* 55 (11), 2589–2608. doi:10.1080/1540496x.2019.1584556

Jaffe, A. B., and Palmer, K. (1997). Environmental Regulation and Innovation: A Panel Data Study. *Rev. Econ. Stat.* 79, 610–619. doi:10.1162/003465397557196

Johnstone, N., Haščič, I., and Popp, D. (2010). Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts. *Environ. Resource Econ.* 45, 133–155. doi:10.1007/s10640-009-9309-1

Kahn, S., and Knittel, C. R. (2003). *The Impact of the Clean Air Act Amendments of 1990 on Electric Utilities and Coal Mines: Evidence from the Stock Market*. Working Paper 118. Berkeley: Centre for the Study of Energy Markets, University of California

Kam, A., Citron, D., and Muradoglu, G. (2008). Distress and Restructuring in China: Does Ownership Matter? *China Econ. Rev.* 19 (4), 567–579. doi:10.1016/j.chieco.2008.07.002

Khan, M. T., Al-Jabri, Q. M., and Saif, N. (2021). Dynamic Relationship between Corporate Board Structure and Firm Performance: Evidence from Malaysia. *Int. J. Finance Econ.* 26 (1), 644–661. doi:10.1002/ijfe.1808

Kneller, R., and Manderson, E. (2012). Environmental Regulations and Innovation Activity in UK Manufacturing Industries. *Resource Energ. Econ.* 34, 211–235. doi:10.1016/j.reseneeco.2011.12.001

Koh, S., Durand, R. B., Dai, L., and Chang, M. (2015). Financial Distress: Lifecycle and Corporate Restructuring. *J. Corporate Finance* 33, 19–33. doi:10.1016/j.jcorpfin.2015.04.004

Lanoie, P., Patry, M., and Lajeunesse, R. (2008). Environmental Regulation and Productivity: Testing the Porter Hypothesis. *J. Prod. Anal.* 30 (2), 121–128. doi:10.1007/s11123-008-0108-4

Lennox, C. (1999). Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches. *J. Econ. Business* 51 (4), 347–364. doi:10.1016/s0148-6195(99)00009-0

Lin, X., Zhang, Y., and Zhu, N. (2009). Does Bank Ownership Increase Firm Value? Evidence from China. *J. Int. Money Finance* 28 (4), 720–737. doi:10.1016/j.jimonfin.2008.12.015

Liu, Q., and Tian, G. (2012). Controlling Shareholder, Expropriations and Firm's Leverage Decision: Evidence from Chinese Non-tradable Share Reform. *J. Corporate Finance* 18 (4), 782–803. doi:10.1016/j.jcorpfin.2012.06.002

Ma, S., Dai, J., and Wen, H. (2019). The Influence of Trade Openness on the Level of Human Capital in China: on the Basis of Environmental Regulation. *J. Clean. Prod.* 225, 340–349. doi:10.1016/j.jclepro.2019.03.238

Ma, S., Naughton, T., and Tian, G. (2010). Ownership and Ownership Concentration: Which Is Important in Determining the Performance of China's Listed Firms? *Account. Finance* 50, 871–897. doi:10.1111/j.1467-629x.2010.00353.x

Maurel, M., Pernet, T., and Zhao, R. (2019). Financial Dependencies, Environmental Regulation, and Pollution Intensity: Evidence From China

(No. 19029). Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.

Oberndorfer, U. (2009). EU Emission Allowances and the Stock Market: Evidence from the Electricity Industry. *Ecol. Econ.* 68 (4), 1116–1126. doi:10.1016/j.ecolecon.2008.07.026

O'Hara, M., and Shaw, W. (1990). Deposit Insurance and Wealth Effects: The Value of Being "Too Big to Fail". *J. Finance* 45 (5), 1587–1600. doi:10.2307/2328751

Opler, T. C., and Titman, S. (1994). Financial Distress and Corporate Performance. *J. Finance* 49 (3), 1015–1040. doi:10.1111/j.1540-6261.1994.tb00086.x

Pang, C., and Wang, Y. (2020). Political Connections, Legal Environments and Firm Performance Around the World. *Int. J. Finance Econ*. forthcoming. doi:10.1002/ijfe.2021

Pargal, S., and Wheeler, D. (1996). Informal Regulation of Industrial Pollution in Developing Countries: Evidence from Indonesia. *J. Polit. Economy* 104 (6), 1314–1327. doi:10.1086/262061

Poncet, S., Steingress, W., and Vandenbussche, H. (2010). Financial Constraints in china: Firm-Level Evidence. *China Econ. Rev.* 21 (3), 411–422. doi:10.1016/j.chieco.2010.03.001

Porter, M. E. (1991). America's green Strategy. *Scientific Am.* 264 (4), 96. doi:10.1038/scientificamerican0491-168

Porter, M. E., and Linde, C. v. d. (1995b). Toward a New conception of the Environment-Competitiveness Relationship. *J. Econ. Perspect.* 9 (4), 97–118. doi:10.1257/jep.9.4.97

Porter, M. E., and van der Linde, C. (1995a). Green and Competitive: Ending the Stalemate. *Harv. Business Rev.* 73 (5), 120–134.

Qi, G. Y., Zeng, S. X., Shi, J. J., Meng, X. H., Lin, H., and Yang, Q. X. (2014). Revisiting the Relationship between Environmental and Financial Performance in Chinese Industry. *J. Environ. Manage.* 145, 349–356. doi:10.1016/j.jenvman.2014.07.010

Ramanathan, R., He, Q., Black, A., Ghobadian, A., and Gallear, D. (2017). Environmental Regulations, Innovation and Firm Performance: A Revisit of the Porter Hypothesis. *J. Clean. Prod.* 155, 79–92. doi:10.1016/j.jclepro.2016.08.116

Wang, Z.-J., and Deng, X.-L. (2006). Corporate Governance and Financial Distress: Evidence from Chinese Listed Companies. *The Chin. Economy* 39 (5), 5–27. doi:10.2753/ces1097-1475390501

Wang, Z., and Li, H. (2007). Financial Distress Prediction of Chinese Listed Companies: A Rough Set Methodology. *Chin. Manag. Stud.* 1 (2), 93–110. doi:10.1108/17506140710758008

Wu, W., Wu, C., Zhou, C., and Wu, J. (2012). Political Connections, Tax Benefits and Firm Performance: Evidence from China. *J. Account. Public Pol.* 31 (3), 277–300. doi:10.1016/j.jaccpubpol.2011.10.005

Yang, T., and Zhao, S. (2014). CEO Duality and Firm Performance: Evidence from an Exogenous Shock to the Competitive Environment. *J. Banking Finance* 49, 534–552. doi:10.1016/j.jbankfin.2014.04.008

Zhao, X., Fan, Y., Fang, M., and Hua, Z. (2018). Do environmental Regulations Undermine Energy Firm Performance? an Empirical Analysis from China's Stock Market. *Energ. Res. Soc. Sci.* 40, 220–231. doi:10.1016/j.erss.2018.02.014

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Zheng and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

# APPENDIX A | VARIABLE DEFINITION

| Variable | Definition |
|---|---|
| Dependent variables | |
| Development capacity | The change of total assets, defined as $(Assets_t - Assets_{t-1})/Assets_{t-1}$. Assets represent the total assets |
| Leverage condition | The change of leverage, defined as $Leverage_t - Leverage_{t-1}$. Leverage equals to total liabilities divided by total assets |
| Profitability | Earnings per share |
| Distress | Dummy variable which equals 1 if firm i falls into financial distress in a given year *t* during the sample period and 0 otherwise |
| Independent variables | |
| HP Firms | Dummy variable which equals 1 if a firm is a heavy polluter and 0 if it is an environmental friendly firm |
| Normal firms | Dummy variable which equals 1 if a firm is neither heavy polluting firms nor environmental friendly firms and 0 if it is an environmental friendly firm |
| After | For sample period 2012–2017 (2001–2006), after equals 1 for firm-year observations after 2015 (2004) and 0 otherwise |
| After* | For sample period 2012–2014 (2001–2003), after* equals 1 for firm-year observations after 2013 (2002) and 0 otherwise |
| Fee rate | The environmental protection fee rate |
| Board size | The logarithm of number of directors employed in the corporate board |
| CEO duality | Dummy variable which equals 1 if an employee works as both CEO and chair of board and 0 otherwise |
| First | The percentage of shares held by the largest shareholder |
| Sedtenth | The percentages of shares held by the second to 10th largest shareholders |
| Age | The logarithm of the number of years since the firm was listed |
| Size | The logarithm of the value of total assets |
| Leverage | Total liabilities divided by total assets |
| Year | Year dummy variables |
| Industry | Industry dummy variables |

frontiers in Energy Research

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fenrg.2021.704317

# How Does Target-Based Performance Evaluation Affect the Accuracy of Energy-Saving Data: Evidence From China

*Ping Zhang*[1,2], *Tiantian Yu*[1,2], *Shengfeng Lu*[3] *and Ruting Huang*[4]*

[1]Center for Social Security Studies, Wuhan University, Wuhan, China, [2]School of Political Science and Public Administration, Wuhan University, Wuhan, China, [3]Economics and Management School, Wuhan University, Wuhan, China, [4]School of Political Science and Public Administration, Wuhan University, Wuhan, China

Energy intensity measured by energy consumption per unit of GDP has always been the main assessment indicator for the design of energy-saving policies, but its accuracy is highly dependent on the reliability of GDP data. This paper finds that the indicator accuracy is improving after the central government has included energy intensity into the performance appraisal system for local officials. This means that the energy-saving target-based performance evaluation has restrained the data misrepresentation behavior of local officials. Further mechanism analysis shows that the pressure of energy saving restricts the development of the industrial sector, which weakens the ability of local governments to manipulate GDP data, thus improving the accuracy of energy intensity statistics. These findings provide some insightful references for China's future green development and policy design.

**Keywords: target-based performance evaluation, data accuracy, energy intensity, green development, policy design**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Guochang Zhao,*
*Southwestern University of Finance and Economics, China*
*Kevin Lo,*
*Hong Kong Baptist University, China*
*Houyin Long,*
*Fuzhou University, China*

***Correspondence:**
*Ruting Huang*
*rutinghuang@foxmail.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *02 May 2021*
***Accepted:*** *07 June 2021*
***Published:*** *02 July 2021*

***Citation:***
*Zhang P, Yu T, Lu S and Huang R (2021) How Does Target-Based Performance Evaluation Affect the Accuracy of Energy-Saving Data: Evidence From China. Front. Energy Res. 9:704317. doi: 10.3389/fenrg.2021.704317*

## INTRODUCTION

Energy intensity, calculated by energy consumption per unit of GDP, is an important measure of energy efficiency. This indicator takes both the economic growth and policy objectives of energy saving into account, which is widely used for public policy and academic research (Zhang et al., 2011; Shi, 2014). China, the world's largest energy consumer, has taken a series of measures to achieve the goal of reducing energy intensity, thus promoting green development. Among them, the 11th (2006–2010) and the 12th (2011–2015) Five-Year Plan (FYP) put forward a binding target of energy intensity reduction (Song and Zheng, 2012; Li H. et al., 2016). Moreover, the national target was allocated to the provincial level, and the energy-saving target was for the first time to be incorporated into the performance evaluation system for the local officials (Zhang et al., 2019). The political system in China is centralized, and thus, the local government is responsible for national policies (Xiong, 2018; Lo, 2020). The top–down target-based responsibility system ensures that the local officials are tied to satisfying the required mandates for career concerns as the energy intensity reduction has positive effects on the promotion probability (Zheng et al., 2014; Chen et al., 2018).

The previous studies comprehensively analyzed the influencing factors of China's energy intensity from the aspects of technological progress, industrial structure adjustment, and policy regulation (Zhang et al., 2020). However, most studies ignored that the calculation accuracy of energy intensity basically depends on data reliability. For a long time, there has been a dispute about authenticity and accuracy of local government GDP data. For example, at the end of 2015, the Xinhua News Agency

reported that data manipulation of GDP was rampant in Northeast China. Moreover, in 2017 and 2018, some local governments, such as Liaoning, Inner Mongolia, and Tianjin, admitted that there was fraud in the past GDP data, which reduced the historical economic growth significantly. Some scholars also believe that the official GDP data seriously overestimate the economic growth rate (Rawski, 2001; Young, 2003). In view of this, the data accuracy of energy intensity is doubtful. Zhang et al. (2019) demonstrated that the energy intensity of most provinces in China was overestimated during 2006–2010 after correcting the GDP growth rate.

As mentioned above, to achieve the policy goals of energy saving, the central government included energy intensity into the performance evaluation for local officials during the 11th and the 12th FYP, requiring the local government to meet corresponding targets of energy intensity reduction. So, on the one hand, it is obvious that the overestimation of GDP data cannot only provide excellent economic growth performance but also make it easier to achieve the goals of energy intensity reduction (Lo, 2014; Lo, 2020). This provides a higher incentive for local governments to falsely report economic growth data, which may further weaken the data accuracy. However, on the other hand, energy consumption is highly correlated with GDP growth in the short term, and it is relatively difficult to manipulate energy consumption data (Wallace, 2016). If GDP statistics are blindly improved without corresponding increase in energy consumption, it will raise a question and then increase the risk of false data being found (Rawski, 2001). From this point of view, the central government includes the energy intensity indicators into the performance evaluation for local officials, which may inhibit the GDP misrepresentation and improve the accuracy of energy-saving data. Therefore, the impact of target-based performance evaluation on the energy data accuracy is an effort worth studying. Moreover, in the era of big data, the utilization of information technologies may provide opportunities to strengthen the supervision of behaviors of the central government, thus helping to promote statistical data accuracy.

Based on the study of Henderson et al. (2012), this paper firstly corrects the GDP growth rate by using nighttime lighting data and then calculates the change rate of energy intensity more accurately on this basis. The accuracy of energy intensity is represented by the difference between the revised change rate in energy intensity and the official change rate. Finally, it is analyzed whether the target-based assessment enlarges or narrates the above gap, that is, it weakens or improves the accuracy of energy-saving data. Our research suggests that official data do generally overstate the decline in energy intensity across regions, thanks to widespread GDP misreporting. Further study on the effect of energy conservation policies shows that the inclusion of energy intensity in the performance appraisal system for local officials improves the data accuracy. Specifically, the main mechanism is the decline of the industrial share in economic output.

Our paper contributes to the literature in the following two aspects. First, although previous studies have pointed out that energy-efficiency policy is an important factor affecting China's energy intensity, they have not taken into account that policy design may also affect the energy data quality. Our findings provide direct evidence of how energy-efficiency target-based performance evaluation policies affect the energy data accuracy. Second, the study sheds light on a growing body of literature on government data manipulation. Previous studies have found that government target assessment can affect the quality of statistical data (Chen et al., 2012; Ghanem and Zhang, 2014; Wallace, 2016), but our results reveal that the statistical data quality has improved.

This paper is organized as follows. *Literature Review* reviews the related research. *Estimation Strategy* introduces the research methods, variable construction, and data source. Specifically, we recalculate the change rate in energy intensity to infer the official statistical data accuracy by using the nighttime lighting data. *Main Results* mainly empirically discusses the impact of target-based performance evaluation on the accuracy of energy-saving data. Finally, *Conclusion and Policy Implications* concludes this article.

## LITERATURE REVIEW

The existing studies comprehensively analyze the influencing factors of China's energy intensity change from various aspects. For example, Zhang et al. (2020) found that better access to credit can improve energy efficiency measured by energy intensity. Xue and Wang (2021) demonstrated that mitigation of financial pressure makes a significant contribution to the energy intensity reduction. Lin and Zhu (2021) confirmed the new-type urbanization or human-centered urbanization has a positive effect on energy intensity. Overall, energy intensity is an important indicator for evaluating the performance of local officials. Especially since the 11th FYP, this indicator has been included in the performance evaluation for local officials. However, most of the studies measure energy intensity using the official data without considering data reliability.

Since data misreporting is nothing new, several studies have shown that many governments deliberately manipulate official statistics for specific purposes (Li P. et al., 2016). For example, in order to meet the economic conditions of joining the EU, Italy and Greece have deliberately understated their budget deficits (Barber and hope, 2010). Michalski and Stoltz (2013) found that the balance of payments data sometimes does not conform to the statistical distribution law and thus inferred that the government might make strategic data misrepresentation. The opposite data manipulation also exists, and not only for the bad economic situation. Kerner et al. (2014) revealed that some countries would deliberately lower their per capita national income in order to obtain concessional loans from the World Bank. Research on China's official data manipulation is mainly related to the top–down official assessment system (Edin, 1998). Kung and Chen (2011) found that, during 1958–1960, local officials falsely reported grain output to cater to the goals of their superiors, resulting in excessive grain purchase and serious difficulties in the national economy. Chen et al. (2012) and Ghanem and Zhang (2014) analyzed the air pollution data at the city level and revealed

that there is a discontinuous breakpoint when the air pollution concentration is at the critical point of the "blue sky" standard, which indicates that the local government may satisfy the air quality assessment standard through data fraud. Wallace (2016) measured the promotion-oriented data fraud by the difference between the official GDP growth rate and the electricity consumption growth rate and found that the gap is larger than that in other periods during the change of government officials.

Given that the Chinese central government has set a series of mandatory targets for energy conservation and environmental protection, many studies also have focused on the issues on the target-based responsibility scheme. For example, Lo (2014) proved that the energy conservation target responsibility system has been constrained by problems including weak targets and lack of reliable local energy statistics. Li H. et al. (2016) highlighted that the national energy intensity target cannot be fully disaggregated without omissions. Chen et al. (2018) estimated how local officials respond to an emission reduction target–based performance evaluation system and found that the performance evaluation system leads to the decrease in $SO_2$ emission with the decline of cost of GDP growth rate. Lo (2020) examined the key aspects and limitation of the energy conservation target responsibility system in China.

In general, although the existing literature has done a more detailed study on energy intensity, few literature works discuss whether the target-based performance evaluation of local government improves or worsens the quality of energy-saving statistics. In this paper, the distortion of energy intensity statistical index caused by the false report of the GDP data is eliminated by using nighttime lighting data. Based on this, the impact of the assessment of energy conservation policies on the accuracy of energy data during the 11th and the 12th FYP is discussed, thus making up for the deficiencies of the existing research.

# ESTIMATION STRATEGY

## Estimation Framework

The main objective of our empirical analysis is to test the effect of energy intensity reduction regulation on data accuracy. The time and spatial variation of the reduction target of the 11th and the 12th FYP since 2006 provides an opportunity to implement the difference in difference (DID) method. Specifically, before 2006, namely, the first year of the 11th FYP, the regulation stringency of energy intensity reduction ratio was almost uniform across the provinces. However, some provinces face a stringent energy intensity reduction mandate assigned by the central government. Thus, the DID estimation compares the provincial data accuracy of energy intensity before and after 2006 with more stringent targets relative to lax targets. The DID estimation specification is set as follows (Beck et al., 2010; Li P. et al., 2016; Chen et al., 2018; Shi and Xu, 2018):

$$gap_{i,t} = c + \beta target_{i,t} * post + \gamma controls_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t}, \quad (1)$$

where $gap_{i,t}$ is the difference in the revised ratio of energy intensity change and official data in province $i$ at year $t$. We use this indicator to represent data distortion, and a higher value means lower accuracy of official energy intensity data. $target_{i,t}$ is the planned reduction ratio of energy intensity, namely, the target assessment ratio required by the central government. The higher planned reduction rate of energy intensity means that the local government faces greater assessment pressure. That is, the assessment intensity of energy conservation (*target*) will be greater. As 2006 is the beginning year of the 11th FYP, *post* takes a value of one if $t \geq 2006$ or otherwise 0.

$Controls_{i,t}$ denotes control variables, including some provincial economic indicators and the characteristics of the provincial party secretary and governor. Specifically, provincial economic indicators mainly refer to the percentage of second industry (*Second*), the percentage of third industry (*Third*), the proportion of fixed investment to GDP (*Invest*), the proportion of total consumption sales to GDP (*Consume*), and the total value of foreign trade (*Foreign*). Furthermore, as the personal characteristics and behavior of local officials will have a significant impact on the economic development of the jurisdiction, we further control the characteristics of provincial officials, mainly including age, gender, the experience of serving in the central government, the experience of serving in his hometown or not, the term of office, and the quadratic term of age to control the possible non-linear relationship. $\delta_i$ and $\mu_t$ are the province fixed effect and year fixed effect, respectively. $\varepsilon_{i,t}$ is the error term, and $c$ is the constant.

The positive value of $\beta$ means the higher requirement of energy intensity reduction will lead to the worse accuracy of energy intensity. That is, faced with the assessment pressure, local governments tend to distort the statistical indicators. In contrast, the negative value of $\beta$means the higher requirement results in the higher data accuracy. In other words, in view of the great impact of the central target assessment on a local official's career, the incorporation of energy-saving indicators into the assessment system may inhibit the fraudulent behavior of local officials.

## Variable Construction

As mentioned, the difference in the revised energy intensity change rate and official data is the explained variable in our paper. However, this indicator cannot be obtained directly. The energy intensity indicator is usually measured by energy consumption per unit of GDP. Especially, energy consumption data involve the specific performance in terms of power grid companies and other units, which does not interfere with the GDP pursuit of local officials, and thus, there is limited room and motivation for manipulation. Therefore, it is usually recognized that energy consumption data are more real and exact as it is relatively difficult to manipulate energy consumption data (Wallace, 2016). However, the past evidence shows the GDP data in China have likely been manipulated (Xu et al., 2015). To calculate this indictor, our basic idea is to revise the GDP growth rate with nighttime lighting intensity and then measure data accuracy of energy intensity by using the revised GDP growth rate. The specific setting and calculation are described in detail as follows.

**TABLE 1 |** Relationship between satellite lighting and the GDP growth rate.

| | (1) Lngdp | (2) Lngdp |
|---|---|---|
| Lnlight | 0.427*** | 0.267** |
| | (4.31) | (2.12) |
| Constant | 8.049*** | 7.774*** |
| | (43.91) | (49.94) |
| Year fixed effect | No | Yes |
| Province fixed effect | No | Yes |
| Observations | 290 | 290 |
| R-squared | 0.350 | 0.992 |

*Note:*
***p < *0.01*.
**p < *0.05*.
*p < *0.1*.***, **, * *denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

## Estimating GDP Growth Rate Based on Nighttime Lighting Intensity

To revise the GDP growth rate based on nighttime lighting intensity, we first need to accurately measure the correlation between nighttime lighting intensity and economic growth. Thus, we have the following regression specification:

$$y_{i,t} = c + \alpha l_{i,t} + \omega_i + \mu_t + \varepsilon_{i,t}, \tag{2}$$

where $y_{i,t}$ is the logarithm of official GDP in province $i$ at year $t$ with 2000 as the base year. $l_{i,t}$ is the logarithmic value of nighttime lighting intensity. $\alpha$ captures the relationship between real GDP and light intensity as well as the relationship between the official GDP growth rate and the light density growth rate. $\omega_i$, $\mu_t$, and $\varepsilon_{i,t}$ are the province fixed effect, the year fixed effect, and the error term, respectively.

Based on Henderson et al. (2012), we estimate the growth rate of real GDP with the following ideas. We assume that the statistical relationship among the real GDP growth rate, official GDP growth rate, and light intensity growth rate satisfies the following conditions:

$$y_i = \beta l_i + \varepsilon_i, \tag{3}$$

$$l_i = \gamma y_i^* + \varepsilon_{l,i}, \tag{4}$$

$$y_i = \varnothing y_i^* + \varepsilon_{y,i}, \tag{5}$$

$$E(\varepsilon_{l,i}) = E(\varepsilon_{y,i}) = 0, \tag{6}$$

$$E(\varepsilon_{l,i}\varepsilon_{y,l}) = 0, \tag{7}$$

where $y^*$ represents the real GDP growth rate, y is the official GDP growth rate, $l$ is the nighttime lighting intensity growth rate, $\beta$ reflects the degree of correlation, and $i$ represents the province. Then, we have the following equations:

$$\widehat{y}_i^* = \lambda y_i + (1-\lambda)\beta l_i, \tag{8}$$

$$\lambda = \frac{\theta - \rho^2}{1 - \rho^2}, \tag{9}$$

$$\theta = \frac{\sigma^2_{y_i^*}}{var(y_i)}. \tag{10}$$

The correlation coefficient between $y$and $l$ is $\rho$, and the predicted value of $\widehat{y}_i^*$is more accurate than that of $y_i$. Among them, $\lambda$ describes the accuracy of official statistics. In the following analysis, based on the World Bank (2002) evaluation of China's statistical capacity and quality, we give a more objective value of λ.

Based on the method mentioned above, we empirically test the relationship between GDP and nighttime lighting intensity. **Table 1** shows the regression results from **Eq. 2**. The results show that there is a significant positive relationship between nighttime lighting intensity and GDP, which is very close to the result of Henderson et al. (2012). They found that the correlation coefficient was about 0.28.

Furthermore, to correct the GDP growth rate, we need to assign an appropriate value to **Eq. 8**, that is, to determine the respective weight of the official GDP growth rate and the nighttime lighting intensity growth rate. Based on the findings in **Table 1**, we choose the results in column (2) with $\beta = 0.267$. According to the research of the World Bank (2002), China is a developing country with a relatively sound statistical system and better statistical quality. Henderson et al. (2012) assigned the official data weight of such countries to be 0.85. We also use this value, and thus, $\lambda$ is set as 0.85. Therefore, we can calculate the revised GDP growth rate. That is, $cogdpgrow = 0.85 * GDP + 0.267 * 0.15 * lggrow$, where *cogdpgrow* is the revised GDP growth rate, *GDP* is the official GDP growth rate, and *lggrow* is the growth rate of nighttime lighting intensity.

## Measuring Data Accuracy of Energy Intensity

The energy intensity indicator is usually measured by energy consumption diving GDP. Thus, the change rate of energy intensity is equal to the difference between the change rate of energy consumption and the GDP growth rate. Given the unchanged energy consumption growth rate, different measures on the GDP growth rate will lead to different change rates of energy intensity. This paper applies the difference in the

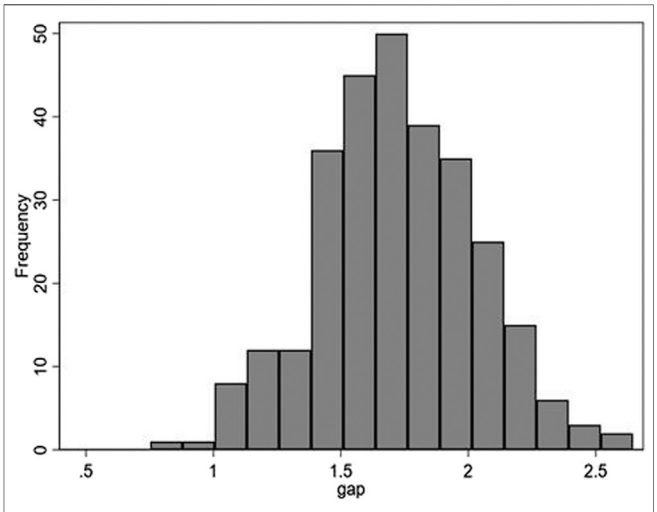



**FIGURE 1 |** Distribution of data accuracy of energy intensity (gap).

**TABLE 2 |** Descriptive statistics.

| Variable | Definition | Mean | SD | OBS |
|---|---|---|---|---|
| Lngdp | Gross domestic product | 8.655 | 0.967 | 290 |
| Lnlight | Nighttime light density | 1.418 | 1.338 | 290 |
| lnEc | Change rate of energy consumption | 9.387 | 5.528 | 290 |
| Gap | Data accuracy of energy intensity reduction ratio | 1.717 | 0.310 | 290 |
| Target | Planned reduction rate of energy intensity | 18.65 | 2.568 | 290 |
| Actualtarget | Actual reduction rate of energy intensity | 19.11 | 2.967 | 290 |
| Second | Percentage of second industry | 47.03 | 7.879 | 290 |
| Third | Percentage of third industry | 40.48 | 8.027 | 290 |
| Invest | Proportion of fixed investment to GDP | 55.79 | 16.48 | 290 |
| Consume | Proportion of total consumption sales to GDP | 0.344 | 0.0470 | 290 |
| Foreign | Total value of foreign trade | 14.46 | 1.733 | 290 |
| sz_gender | Gender of the province's governor | 0.972 | 0.164 | 290 |
| sz_center | Governor's experience of working in the central government | 0.324 | 0.469 | 290 |
| sz_home | Governor serving in home or not | 0.334 | 0.473 | 290 |
| sz_term | Governor's term of office | 3.097 | 1.887 | 290 |
| sz_age | Governor's age | 57.39 | 3.898 | 290 |
| sj_gender | Provincial party secretary's gender | 0.990 | 0.101 | 290 |
| sj_center | Provincial party secretary's experience of working in the central government | 0.590 | 0.493 | 290 |
| sj_home | Provincial party secretary serving in home or not | 0.110 | 0.314 | 290 |
| sj_term | Office term of provincial party secretary | 3.376 | 2.127 | 290 |
| sj_age | Age of provincial party secretary | 58.86 | 4.209 | 290 |

revised ratio of energy intensity change and official data to measure the data accuracy (*gap*). The higher value always means lower data accuracy of energy intensity due to false statistical data issued by the local officials.

Based on the revised GDP growth rate, we also calculate the change rate of energy intensity with a constant rate of energy consumption. Then, the difference between the revised and the official energy intensity change rates can be obtained, namely, the degree of data distortion (*gap*). As mentioned, this indicator is applied to measure the data accuracy of energy intensity. **Figure 1** shows the distribution frequency of these data. We find that the value of *gap*is always greater than 0. Combined with **Table 2**, the average value and standard deviation of the gap in the sample are 1.717 and 0.31, which indicates the overestimated energy intensity reduction from 2003 to 2012.

## Data Source

Our sample contains 29 provinces, cities, and autonomous regions over 2003 to 2012 in China. Though the sample period is limited to 2012 as the nighttime light data are not available since then, the empirical findings are still valuable and relevant (Li P. et al., 2016; Chen et al., 2018; Jia et al., 2021) because the indicator assessment applied by the central government in the 11th and the 12th FYP is rather similar. Due to data limitation, Xinjiang, Tibet, Taiwan, Hong Kong, and Macao are not included in this paper. The original data of global nighttime lighting intensity come from the US Defense Meteorological Satellite Project (DMSP-OLS), which includes global nighttime lighting images from 1992 to 2012 obtained by several satellites. The saturation value of the pixel gray value (DN value) is 63, with the problem that the DN value obtained by different sensors in the same year is different. Especially, the existing raw data of satellite nighttime lighting face the problems of discontinuity and saturation (Donaldson and Storeygard, 2016). Therefore, we use the lighting image correction method including inter-calibration and intra-annual composition based on the invariant target area to correct the original lighting image data and get the provincial lighting intensity value (*light*) from 2003 to 2012 (Liu et al., 2012), which can more reasonably reflect the regional economic development differences. This method is widely used in the literature to correct the long time series nighttime lighting image data (Henderson et al., 2012; Wu et al., 2013; Jia et al., 2021).

Provincial energy data mainly include the change rate of energy consumption index (*lnEc*), the planned rate of energy intensity reduction (*target*), and the actual reduction rate of energy intensity (*actualtarget*). The data are mainly from China Statistical Yearbook, China Energy Statistical Yearbook. Other provincial economic indicators are all derived from China Statistical Yearbook. In addition, we have manually collected the personal information of provincial officials (secretary of the provincial party committee and governor) from government websites and Baidu Encyclopedia. The descriptive statistics are shown in **Table 2**.

# MAIN RESULTS

This section mainly analyzes the impact of energy-saving policy target assessment on the accuracy of energy statistics. In addition to baseline estimates, we also conduct a series of robust tests, including a parallel trend test, a placebo test, and replacement of the explained variable. Furthermore, we also study the regional heterogeneity and verify the influence mechanism.

## Baseline Estimation

**Table 3** reports the main results of the effect of target assessment on the data accuracy of energy intensity from **Eq. 1**. Only

**TABLE 3 |** Baseline estimates.

| | (1) Gap | (2) Gap | (3) Gap | (4) Gap |
|---|---|---|---|---|
| target×post | −0.027** | −0.025** | −0.029** | −0.027** |
| — | (−2.10) | (−2.18) | (−2.40) | (−2.17) |
| Second | — | 0.012 | 0.013 | 0.006 |
| — | — | (0.92) | (1.18) | (0.47) |
| Third | — | 0.023 | 0.020 | 0.017 |
| — | — | (1.63) | (1.52) | (1.20) |
| Investment | — | 0.011*** | 0.009*** | 0.008** |
| — | — | (4.06) | (3.53) | (2.65) |
| Consume | — | −1.675 | −1.416 | −1.150 |
| — | — | (−1.39) | (−1.16) | (−0.86) |
| Foreign | — | 0.091 | 0.069 | 0.039 |
| — | — | (0.96) | (0.85) | (0.42) |
| Constant | 1.506*** | −1.045 | −3.990 | −1.047 |
| — | (30.83) | (−0.71) | (−0.76) | (−0.22) |
| Provincial party secretary characteristics | No | No | Yes | Yes |
| Governor characteristics | No | No | Yes | Yes |
| Year fixed effect | Yes | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes | Yes |
| Year* regional fixed effect | No | No | No | Yes |
| Observations | 290 | 290 | 290 | 290 |
| R-squared | 0.384 | 0.499 | 0.552 | 0.592 |

*Note:*
***p < *0.1.*
**p < *0.05.*
*p < *0.01.*, **, * denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

province and year fixed effects are included in the first column, and the estimated results show that the assessment policy significantly improves the data accuracy. This means that the energy conservation assessment started in 2006 has a significant constraint or regulation effect on the behavior of local governments, making them less prone to fraud in energy intensity data statistics.

Provincial economic characteristics are introduced as control variables in column (2). We find that the provinces with a higher proportion of fixed asset investment to GDP have lower data accuracy of energy intensity. In column (3), the personal characteristics of provincial officials are controlled, and the results are still robust. In addition, in order to control the influence of time-varying unobservable factors at the regional level on the dependent variable, the interaction term of year and regional fixed effects in eastern, central, and western regions is further added in column (4), and the results are highly consistent with the findings in columns (1)–(3).

As shown in **Table 3**, the estimation coefficient between the target assessment intensity and the energy intensity data accuracy is always statistically significantly negative. This shows that the energy-saving target-based performance evaluation since 2006 has led to a significant improvement in data accuracy. In an economic sense, if energy intensity target assessment intensity changes by a standard deviation, the degree of data distortion will decrease by about 13.7% (0.137 = 0.027*8.71/1.72) according to the regression coefficient in column (4).

## Robustness Checks

To further verify our results are robust, we report a series of robustness checks, including a parallel trend test, a placebo test, and replacement of the explained variable.

### Parallel Trend Test

A necessary condition for supporting our identification is that provinces with different target assessment have similar time trends on the outcomes before and after treatment. This means that the differences of assessment requirements faced by different provinces before 2006 will not affect the accuracy of the previous data. Thus, we use the event study approach (Chen et al., 2018). The estimation specification is as follows:

$$gap_{i,t} = c + \sum_{t=2004}^{t=212} \beta_t * target_{i,t} * yeardummy_t + \gamma controls_{i,t} + \delta_i + \mu_t + \varepsilon_{i,t}, \quad (11)$$

where $yeardummy_t$ is the dummy variable taking a value of one if $t = year$ or 0 otherwise. $\beta_t$ captures the difference in outcomes between different provinces before and after treatment. As shown in **Figure 2**, the regression coefficients before 2006 are not significant, which indicates that there is no significant difference in data accuracy among provinces before treatment. Therefore, it is reasonable and feasible to use the DID method to test the impact of target assessment on the accuracy of energy intensity statistics. Further analysis of the

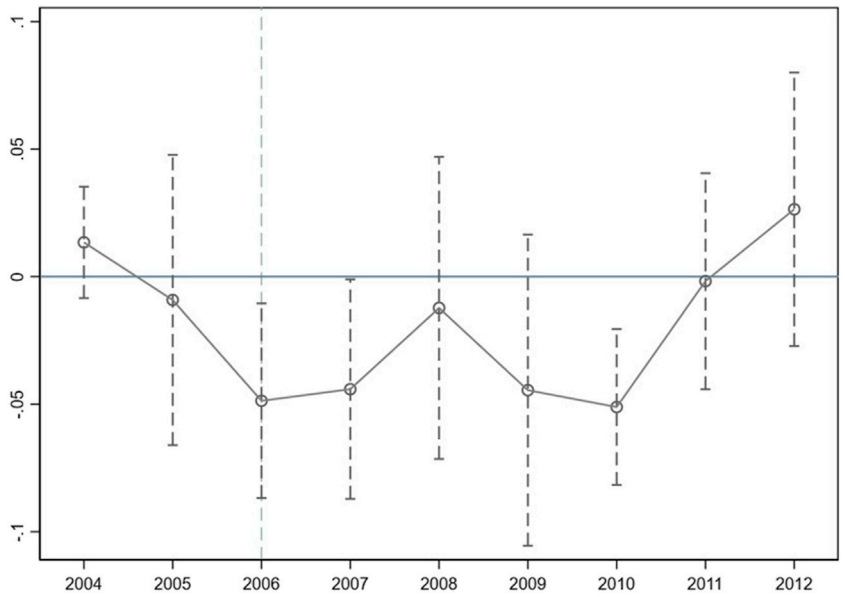



**FIGURE 2 |** Parallel trend test.

**TABLE 4 |** Placebo test.

| | (1) Gap | (2) Gap | (3) Gap |
|---|---|---|---|
| target×pre | 0.024 | 0.019 | 0.013 |
| — | (1.28) | (1.04) | (0.70) |
| Provincial economic indicators | No | Yes | Yes |
| Provincial party secretary characteristics | No | No | Yes |
| Governor characteristics | No | No | Yes |
| Year fixed effect | No | Yes | Yes |
| Province fixed effect | No | Yes | Yes |
| Observations | 290 | 290 | 290 |
| R-squared | 0.374 | 0.490 | 0.539 |

*Note:*
***p < 0.01.
**p < 0.05.
*p < 0.1.***, **, * denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

dynamic policy impact shows that the coefficient of the first evaluation year is negative, which indicates that the policy has a positive impact on data accuracy. The estimates are negative and stale until 2011. Particularly, the estimates in 2010 are the most prominent in both the numerical and the statistical sense, which is completely consistent with time when the local government submitted the final energy intensity assessment document to the central government in the 11th FYP. This has also deepened our understanding of the policy operation mechanism.

### Placebo Test

To further verify the reliability, we set the policy time for the implementation of the 11th FYP in 2005. Assuming that the energy intensity assessment is only implemented in 2005, we conduct regression analysis again and assign $pre = 1$ if $year = 2005$or 0 otherwise. In other words, if the accuracy improvement in energy intensity statistics is really caused by target assessment, the estimate should not be significant. According to the results in **Table 4**, the estimated coefficient is not significant, so the influence of other unobservable potential factors on the previous findings can be excluded.

### Replacement of the Explained Variable

To avoid one-sided understanding of the assessment indicators, we replace the core explanatory variable in **Table 5**. The planned reduction rate of energy intensity is replaced by the actual reduction rate. According to the regression results, we find that the coefficient is slightly smaller, but still statistically significantly negative. In an economic sense, the actual reduction rate of energy intensity is reduced by one standard deviation, and the data accuracy is improved by about 10%. Therefore, after changing the explanatory variables, the results still show that the higher requirements of energy-saving assessment will result in higher statistical accuracy of energy data.

## Regional Heterogeneity

Considering that economic development and policy orientation differ greatly in different regions, we divide all provinces into eastern, central, and western regions to investigate the heterogeneous impacts[1]. Columns (1)–(3) in **Table 6**, respectively, report the comparative results by controlling provincial economic variables, official characteristics, province fixed effect, and year fixed effect. Generally speaking, the

[1]According to the common division methods, the eastern region includes Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan; the central region includes Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan; and the western region includes Guangxi, Inner Mongolia, Shaanxi, Gansu, Qinghai, Ningxia, Sichuan, Chongqing, Yunnan, and Guizhou.

**TABLE 5 |** Replacement of the explained variable.

| | (1) Gap | (2) Gap | (3) Gap | (4) Gap |
|---|---|---|---|---|
| actualtarget×post | −0.023** | −0.020* | −0.022** | −0.019* |
| — | (−2.36) | (−2.00) | (−2.11) | (−1.77) |
| Provincial economic indicators | No | Yes | Yes | Yes |
| Provincial party secretary characteristics | No | No | Yes | Yes |
| Governor characteristics | No | No | Yes | Yes |
| Year fixed effect | Yes | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes | Yes |
| Year* regional fixed effect | No | No | No | Yes |
| Observations | 290 | 290 | 290 | 290 |
| R-squared | 0.383 | 0.497 | 0.549 | 0.588 |

*Note:*
***p < *0.01.*
**p < *0.05.*
*p < *0.1.****, **, * denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

**TABLE 6 |** Regional heterogeneity.

| | (1) East Gap | (2) Central Gap | (3) West Gap |
|---|---|---|---|
| target×post | −0.032*** | −0.153* | 0.003 |
| — | (−3.94) | (−1.95) | (0.20) |
| Provincial economic indicators | Yes | Yes | Yes |
| Provincial party secretary characteristics | Yes | Yes | Yes |
| Governor characteristics | Yes | Yes | Yes |
| Year fixed effect | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes |
| Observations | 110 | 80 | 100 |
| R-squared | 0.852 | 0.671 | 0.645 |

*Note:*
***p < *0.01.*
**p < *0.05.*
*p < *0.1.****, **, * denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

eastern and central provinces are stronger than the western provinces in both economic strength and political game ability, which may lead to the heterogeneity in the response of assessment policy. The analysis results show that the accuracy of energy intensity data is significantly improved in both eastern and central regions. The relationship between energy target assessment and data distortion is positive, but not significant in the western region. This means that local governments with greater economic power and political influence will weigh the potential risks and benefit more carefully when faced with appraisal pressure and reduce the tendency to falsify data.

## Mechanisms

We hope to better understand the specific mechanism how this policy effect can be achieved. Since the re-estimation of the change rate of energy intensity is mainly based on different estimates of the economic growth rate, we actually need to find the influence channels on the misreporting of economic growth rate. Chen et al. (2019) found that the local government mainly distorts statistical data by adjusting the industrial output on the output side and reduces statistical accuracy by adjusting the investment data on the expenditure side. Therefore, energy-saving assessment may also affect the economic growth rate through industrial output and fixed asset investment, finally influencing the accuracy of energy intensity data. Later, we will analyze the impact mechanism of energy intensity assessment on data accuracy from the perspective of industrial output and fixed asset investment.

### Industrial Output

Energy-saving assessment encourages local governments to adjust industrial output, thus affecting the accuracy of energy intensity data. To test this mechanism, this paper takes the proportion of industrial output to GDP as the explained variable. The corresponding regression results are reported in columns (1) and (2) of **Table 7**. The results show that the target assessment of energy intensity reduces the proportion of industrial output, which reduces the potential possibility of false reporting of output data and improves the accuracy of energy data.

### Fixed Asset Investment

Energy-saving assessment may encourage local governments to adjust fixed asset investment, thus affecting the data accuracy of energy intensity. Similarly, this paper takes the proportion of fixed asset investment to GDP as the explained variable. As shown in columns (3) and (4) of **Table 7**, there is a negative relationship between energy intensity assessment and fixed asset investment, but it is not statistically significant.

Overall, we find that the energy-saving target assessment policy is mainly to reduce the proportion of local industrial output to improve the accuracy of statistical data. It also has a certain impact on investment, but it is not significant.

# CONCLUSION AND POLICY IMPLICATIONS

Although energy intensity has always been the main measurement for the energy-saving policy design, its accuracy is affected by the false report of the GDP data. At the same time, few literature works discuss whether it will weaken or improve the

**TABLE 7 |** Mechanisms.

| | (1) Industry | (2) Industry | (3) Invest | (4) Invest |
|---|---|---|---|---|
| target×post | −0.448** | −0.425** | −0.335 | −0.269 |
| — | (−2.59) | (−2.61) | (−0.68) | (−0.72) |
| Provincial party secretary characteristics | No | Yes | No | Yes |
| Governor characteristics | No | Yes | No | Yes |
| Provincial economic indicators | Yes | Yes | Yes | Yes |
| Year fixed effect | Yes | Yes | Yes | Yes |
| Province fixed effect | Yes | Yes | Yes | Yes |
| Observations | 290 | 290 | 290 | 290 |
| R-squared | 0.395 | 0.444 | 0.762 | 0.824 |

*Note:*
***p < 0.01.
**p < 0.05.
*p < 0.1.***, **, * denote significance at 1%, 5% and 10%, respectively. The t-statistics are reported in parentheses and clustered at the provincial level.*

data accuracy when energy intensity is included in the performance appraisal system for the local government. The understanding of this issue affects how the government designs effective energy conservation and emission reduction policies to achieve the important strategic goal of green development.

This paper analyzes the impact of energy-saving target assessment on data accuracy by using satellite lighting data to re-calculate energy intensity. We find that, after controlling the impact of false report of the economic growth rate, the official data generally overestimate the decline rate of energy intensity. Furthermore, data accuracy has significantly improved after the central government included the energy intensity reduction target in the performance appraisal for local governments. The higher assessment requirements faced by local governments will lead to the improvement of data accuracy. The reason for the above phenomenon is that the pressure of energy saving has inhibited the development of the industrial sector. As the output data of the industrial sector contain a large amount of moisture, energy-saving target assessment weakens the ability of local governments to manipulate GDP data and improves the accuracy of energy intensity indicators.

The above conclusions have some practical policy implications on how to effectively reduce energy consumption and promote green development in China.

First, although the simplified single policy objective is easy to observe, it is vulnerable to be impacted by external interference in practice, resulting in insufficient accuracy. This study finds that the accuracy of energy-saving data is improved when facing the task of economic growth and energy-saving target assessment at the same time. Therefore, in future policy design for energy conservation and emission reduction, we should try to include cross multiple assessment indicators, so as to promote the effective implementation of policy objectives. Moreover, setting a concrete target-based responsibility system for energy saving or environmental policies may have positive effects. In the information age, how to use big data technology and machine learning methods to supervise the behavior of local officials and improve the performance of energy saving is also needed to be further discussed.

Second, it is found that the introduction of energy-saving index into local government performance assessment significantly improves the data accuracy. Though the fulfillment degree of policy objectives is reduced on the surface, it still contains positive factors such as the reduction of data fraud. Therefore, in the performance evaluation of energy saving, some appropriate flexibility is needed and the tolerance can be given if the energy saving does not perform as well as expected. This will provide reasonable incentives for local governments to take the initiative to reduce data distortions and thus to increase their real efforts to achieve energy conservation, emission reduction, and green development goals.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, and further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

PZ, SL, and RH conceptualized the idea and wrote the original draft. PZ, TY, SL, and RH performed the methodology. TY ran the software and curated the data. SL and RH supervised the work and reviewed and edited the paper. All authors have read and agreed to the published version of the article.

## FUNDING

This work was supported by the National Social Science Fund of China on “Study of Measurement of Statistical Misreporting and Correction Mechanism: Based on Incentive of Local Government” (Grant no. 17BJL070), the Humanity and Social Science Youth Foundation of Ministry of Education of China on “GDP Misreporting and Formation Mechanism under Official Evaluation: Study Based on

Satellite Light Data" (Grant no. 16YJC790136), the Projects of the National Natural Science Foundation of China (Grant no. 71773084), and the major projects of the National Social Science Fund of China on "Institutional Dividend of the Transformation of Government Functions" (Grant no. 18ZDA113).

## REFERENCES

Barber, T., and Hope, K. (2010). *Brussels Attacks Greece over False Data*. London: Financial Times, 13.

Beck, T., Levine, R., and Levkov, A. (2010). Big Bad banks? the Winners and Losers from Bank Deregulation in the United States. *J. Finance* 65 (5), 1637–1667. doi:10.1111/j.1540-6261.2010.01589.x

Chen, W., Chen, X., Hsieh, C. T., and Song, Z. (2019). *A Forensic Examination of China's National Accounts*. Massachusetts: National Bureau of Economic Research. doi:10.3386/w25754

Chen, Y., Jin, G. Z., Kumar, N., and Shi, G. (2012). Gaming in Air Pollution Data? Lessons from China. *BE J. Econ. Anal. Pol.* 13 (3). doi:10.1515/1935-1682.3227

Chen, Y. J., Li, P., and Lu, Y. (2018). Career Concerns and Multitasking Local Bureaucrats: Evidence of a Target-Based Performance Evaluation System in China. *J. Develop. Econ.* 133, 84–101. doi:10.1016/j.jdeveco.2018.02.001

Donaldson, D., and Storeygard, A. (2016). The View from above: Applications of Satellite Data in Economics. *J. Econ. Perspect.* 30 (4), 171–198. doi:10.1257/jep.30.4.171

Edin, M. (1998). Why Do Chinese Local Cadres Promote Growth? Institutional Incentives and Constraints of Local Cadres. *Forum Develop. Stud.* 25 (1), 97–127. doi:10.1080/08039410.1998.9666077

Ghanem, D., and Zhang, J. (2014). 'Effortless Perfection:' Do Chinese Cities Manipulate Air Pollution Data? *J. Environ. Econ. Manage.* 68 (2), 203–225. doi:10.1016/j.jeem.2014.05.003

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *Am. Econ. Rev.* 102 (2), 994–1028. doi:10.1257/aer.102.2.994

Jia, J., Liang, X., and Ma, G. (2021). Political Hierarchy and Regional Economic Development: Evidence from a Spatial Discontinuity in China. *J. Public Econ.* 194, 104352. doi:10.1016/j.jpubeco.2020.104352

Kerner, A., Jerven, M., and Beatty, A. (2014). *Are Development Statistics Manipulable?* Burnaby: Simon Fraser University, Simons Papers in Security and Development.37.

Kung, J. K.-S., and Chen, S. (2011). The Tragedy of the Nomenklatura: Career Incentives and Political Radicalism during China's Great Leap Famine. *Am. Polit. Sci. Rev.* 105, 27–45. doi:10.1017/s0003055410000626

Li, H., Zhao, X., Yu, Y., Wu, T., and Qi, Y. (2016a). China's Numerical Management System for Reducing National Energy Intensity. *Energy Policy* 94, 64–76. doi:10.1016/j.enpol.2016.03.037

Li, P., Lu, Y., and Wang, J. (2016b). Does Flattening Government Improve Economic Performance? Evidence from China. *J. Develop. Econ.* 123, 18–37. doi:10.1016/j.jdeveco.2016.07.002

Lin, B., and Zhu, J. (2021). *Impact of China's New-type Urbanization on Energy Intensity: A City-Level Analysis*. Amsterdam: Energy Economics. 105292. doi:10.1016/j.eneco.2021.105292

Liu, Z., He, C., Zhang, Q., Huang, Q., and Yang, Y. (2012). Extracting the Dynamics of Urban Expansion in China Using DMSP-OLS Nighttime Light Data from 1992 to 2008. *Landscape Urban Plann.* 106 (1), 62–72. doi:10.1016/j.landurbplan.2012.02.013

Lo, K. (2014). China's Low-Carbon City Initiatives: the Implementation gap and the Limits of the Target Responsibility System. *Habitat Int.* 42, 236–244. doi:10.1016/j.habitatint.2014.01.007

Lo, K. (2020). Governing Energy Consumption in China: a Comprehensive Assessment of the Energy Conservation Target Responsibility System. *Energy Transit* 4 (1), 57–67. doi:10.1007/s41825-020-00023-0

Michalski, T., and Stoltz, G. (2013). Do countries Falsify Economic Data Strategically? Some Evidence that They Might. *Rev. Econ. Stat.* 95 (2), 591–616. doi:10.1162/rest_a_00274

Rawski, T. G. (2001). What Is Happening to China's GDP Statistics? *China Econ. Rev.* 12 (4), 347–354. doi:10.1016/s1043-951x(01)00062-1

Shi, X. (2014). Setting Effective Mandatory Energy Efficiency Standards and Labelling Regulations: A Review of Best Practices in the Asia Pacific Region. *Appl. Energ.* 133, 135–143. doi:10.1016/j.apenergy.2014.07.084

Shi, X., and Xu, Z. (2018). Environmental Regulation and Firm Exports: Evidence from the Eleventh Five-Year Plan in China. *J. Environ. Econ. Manage.* 89, 187–200. doi:10.1016/j.jeem.2018.03.003

Song, F., and Zheng, X. (2012). What Drives the Change in China's Energy Intensity: Combining Decomposition Analysis and Econometric Analysis at the Provincial Level. *Energy policy* 51, 445–453. doi:10.1016/j.enpol.2012.08.044

Wallace, J. L. (2016). Juking the Stats? Authoritarian Information Problems in China. *Br. J. Polit. Sci.* 46 (1), 11–29. doi:10.1017/s0007123414000106

World Bank (2002). *Building Statistical Capacity to Monitor Development Progress*. Washington, DC: World Bank.

Wu, J., He, S., Peng, J., Li, W., and Zhong, X. (2013). Intercalibration of DMSP-OLS Night-Time Light Data by the Invariant Region Method. *Int. J. remote sensing* 34 (20), 7356–7368. doi:10.1080/01431161.2013.820365

Xiong, W. (2018). "The Mandarin Model of Growth (No. W25296)," in *National Bureau of Economic Research*. Massachusetts: Working Paper. doi:10.3386/w25296

Xu, K., Chen, F., and Liu, X. (2015). The Truth of China Economic Growth: Evidence from Global Night-Time Light Data. *Econ. Res. J.* 9, 17–29.

Xue, X., and Wang, Z. (2021). Impact of Finance Pressure on Energy Intensity: Evidence from China's Manufacturing Sectors. *Energy* 226, 120220. doi:10.1016/j.energy.2021.120220

Young, A. (2003). Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period. *J. Polit. economy* 111 (6), 1220–1261. doi:10.1086/378532

Zhang, D., Aunan, K., Martin Seip, H., and Vennemo, H. (2011). The Energy Intensity Target in China's 11th Five-Year Plan Period-Local Implementation and Achievements in Shanxi Province. *Energy Policy* 39 (7), 4115–4124. doi:10.1016/j.enpol.2011.03.085

Zhang, D., Li, J., and Ji, Q. (2020). Does Better Access to Credit Help Reduce Energy Intensity in China? Evidence from Manufacturing Firms. *Energy Policy* 145, 111710. doi:10.1016/j.enpol.2020.111710

Zhang, P., Shi, X., Sun, Y., Cui, J., and Shao, S. (2019). Have China's Provinces Achieved Their Targets of Energy Intensity Reduction? Reassessment Based on Nighttime Lighting Data. *Energy Policy* 128, 276–283. doi:10.1016/j.enpol.2019.01.014

Zheng, S., Kahn, M. E., Sun, W., and Luo, D. (2014). Incentives for China's Urban Mayors to Mitigate Pollution Externalities: the Role of the central Government and Public Environmentalism. *Reg. Sci. Urban Econ.* 47, 61–71. doi:10.1016/j.regsciurbeco.2013.09.003

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 Zhang, Yu, Lu and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 23 July 2021
doi: 10.3389/fenvs.2021.692478

# Analysis of the Upgrading Effect of the Industrial Structure of Environmental Regulation: Evidence From 113 Cities in China

*Wei Shao*[1], *Yufei Yin*[1,2], *Xiao Bai*[3]* *and Farhad Taghizadeh-Hesary*[4]*

[1]School of Economics, Zhejiang University of Finance and Economics, Hangzhou, China, [2]Business School, Huzhou Normal University, Huzhou, China, [3]School of Finance, Zhejiang University of Finance and Economics, Hang Zhou, China, [4]Social Science Research Institute, Tokai University, Hiratsuka, Japan

At present, China's economy is in a crucial period of economic structural transformation. To alleviate the downward pressure on the economy and explore sustainable growth paths, the Chinese government has issued several environmental regulations. However, the impact of environmental regulation on industrial structure upgrading has not been carefully examined yet. This study utilizes the Pollution Information Transparency Index (PITI) to measure environmental regulation (ER) and examines the impact of ER on industrial structure upgrading (ISU). The sample cities are divided into 36 resource-based cities (RBCs) and 77 non–resource-based cities (NRBCs). The panel data containing 113 cities during 2008–2017 are used in this study. The empirical results show that ER has a significant impact on ISU of RBCs and NRBCs, and robust tests proved the reliability of this result. Analysis of heterogeneity shows ER has a more substantial role in promoting ISU in RBCs and the eastern region. Meanwhile, inside RBCs, ER has a more substantial impact on ISU in growth-RBCs than on that in other RBCs. The mechanism test shows that the mediation effect of technological innovation in RBCs and NRBCs is significant. At last, the impact of ER on ISU has a double-threshold effect in RBCs and a single-threshold effect in NRBCs. With the technological innovation progress, ER produces an increasing effect on ISU of RBCs and NRBCs.

**Keywords: environmental regulation, industrial structure upgrading, technological innovation, resource-based cities, Chinese cities**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University Of Economics, China*

***Reviewed by:***
*Solomon Prince Nathaniel,*
*University of Lagos, Nigeria*
*Jingjing Jiang,*
*Harbin Institute of Technology, Shenzhen, China*

***Correspondence:**
*Xiao Bai*
*baixiao@zufe.edu.cn*
*Farhad Taghizadeh-Hesary*
*farhad@tsc.u-tokai.ac.jp*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *08 April 2021*
***Accepted:*** *11 June 2021*
***Published:*** *23 July 2021*

***Citation:***
*Shao W, Yin Y, Bai X and Taghizadeh-Hesary F (2021) Analysis of the Upgrading Effect of the Industrial Structure of Environmental Regulation: Evidence From 113 Cities in China. Front. Environ. Sci. 9:692478. doi: 10.3389/fenvs.2021.692478*

## INTRODUCTION

Since the reform and opening-up, China has achieved remarkable industrial development, but this is at the cost of environmental pollution. Industrial production emissions are one of the leading causes of environmental pollution (Duzgoren, 2007). Developing countries such as China have realized the importance of achieving coordinated economic growth and environmental protection (Liu and Wang, 2016; Lu et al., 2020a). Hence, the Chinese government has adopted a series of environmental regulation policies to realize industrial structure upgrading and control environmental pollution problems caused by industrial development. The 2020 Environmental Performance Index jointly released by Yale University and Columbia University in the United States shows that China's air quality and many other indicators have improved. China ranks 120$^{th}$ among 180 countries, with an

environmental performance index of 37.3, which has increased by 8.4 percentage points over the past ten years.

As we all know, environmental pollution is a social problem with strong negative externalities and requires local governments' efforts to overcome the market failure in the treatment of environmental pollution. However, local officials in China are appointed by the higher level government or the central government, and the evaluation system is mainly GDP-oriented (Chen, 2015; Rochlitz et al., 2015). This has led to fierce GDP competition and different efforts of controlling environmental pollution among local governments (Yin et al., 2014). Especially for resource-based cities (RBCs), industrial development relies heavily on natural resources and faces a greater challenge of controlling environmental pollution emissions than non–resource-based cities (NRBCs). To what extent does China's environmental regulatory policies help promote industrial structure upgrading (ISU), and whether there are different policy effects among cities with different resource endowments? What is the mechanism? These are the major questions addressed in this study.

In addition, with the popularization and development of the Internet and information technology, the public's awareness of environmental protection has gradually increased. The number of environmental nongovernmental organizations (ENGOs) in China has multiplied since 1994 (the first ENGO, named "Friends of Nature," was established). The IPEA, as a representative ENGO in China, is committed to promoting information openness and public participation to raise environmental awareness of the public by promoting environmental information disclosure and public environmental information supervision. In 2008, the IPEA and NRDC (an international ENGO) jointly released the PITI[1] to evaluate China's government environmental information disclosure. If a city has a higher PITI score, this city discloses more pollution information, and the public has stronger awareness of environmental information supervision. The PITI reflects the efforts of local governments in environmental governance. Hence, the PITI is classified as an informal environmental regulation. Even though the PITI has been used by some scholars (Tian et al., 2016; Yao and Liang, 2017; Li, 2019), using the PITI to measure environmental regulation is still relatively rare. This article uses the PITI to measure urban-level environmental regulations from the perspective of environmental information disclosure and study the impact of ER on industrial structure upgrading (ISU).

[1]http://www.ipe.org.cn/reports/reports.aspx?cid=18336&year=0&key=. It includes "contamination information of daily violations of pollution sources," "cleaning production audit information," "concentration information on pollution sources," "information on sewage charges," "integrated evaluation information of corporate environmental behavior," and "acceptance of environmental impact assessment documents for construction projects" The construction project completion environmental protection acceptance results information, "the investigation and verification of the public on environmental issues or corporate pollution of the petition complaints and their processing results," and "according to the application disclosure" eight aspects.

There have been several related studies in the literature about ER and ISU. First is the theoretical debate between "Following cost" hypothesis and "The Porter Hypothesis (PH)." According to "Following cost" hypothesis, ER increases the cost of enterprises' operation and government management, leads to the decline of production efficiency, and is not beneficial to economic growth (Greenstone, 2001; Blackman and Kildegaard, 2010). The PH posits that reasonable stringency of environmental regulatory policies can increase enterprises' competitiveness and produce more innovations. Ultimately, it helps enterprises meet the standard of ER and reduce production costs (Porter et al., 1995; Franco and Marin, 2013; Ramanathan et al., 2016; Qiu et al., 2018). Yang et al. (2012) found that ER is conducive to promoting innovations in Taiwanese manufacturing sectors. However, Zhao et al. (2018) used the data of China's carbon industries as a research sample and found no strong PH effect, and high-intensity ER produces adverse effects on TFP in the long run. Other scholars have found similar findings (Rassier and Earnhart, 2015; Cohen and Tubb, 2018; Lu et al., 2020b).

Second is the examination of the pollution haven hypothesis (PHH). Sun et al. (2017) found that pollution enterprises in developing countries move to countries or regions with a low ER standard. Taking 16 manufacturing industries of 13 European countries as a research sample, Mulatu et al. (2010) confirmed the existence of PHH. Martin (2014) found that if the EU implements a distinction carbon dioxide emission policy, pollution-intensive enterprises will make regional transfers in the EU economy. Liu et al. (2012) and Wu et al. (2019) found that with the ER intensity increasing in China's developed regions, pollution industries have significant characteristics of moving to the midwest area in China. The third factor includes ER and ISU. Zhu et al. (2019) found that ERs of the steel industry are conducive to improving steel efficiency, and outputs and inputs, based on China's steel industry data from 2006 to 2013. Taking the data of 30 provinces in China as the sample, Zhang et al. (2020) found that ER is beneficial to ISU in the long run. Meanwhile, other scholars believed that ER produces no obvious impacts on industrial upgrading since research samples' selection affected the reliability of conclusions (Rubashkina et al., 2015; Zhang et al., 2019).

The related literature has analyzed the relationship between ER and ISU, but some areas still need to be further expanded. On the one hand, whether ER can promote ISU has not yet formed a unified conclusion. Meanwhile, the existing literature uses national-, provincial-, or industry-level data, while the city-level data for research are relatively rare. The cities are the main policy objects of ER in China, and the regional or industrial heterogeneity of ER has been verified (Yuan et al., 2017; Feng and Li, 2020). It is necessary to study the policy effect of ER from the perspective of urban resource endowment.

On the other hand, it is more difficult to accurately measure ER with a single indicator such as enterprise pollution investment, pollution penalty, or their ratio in all corporate investment (Gray et al., 2014; Liao and Shi, 2018). Some studies are also trying to make comprehensive indicators (including $SO_2$ removal rate and solid waste utilization rate) to measure ER (Dam and Scholens, 2012; Zhou et al., 2017). Even

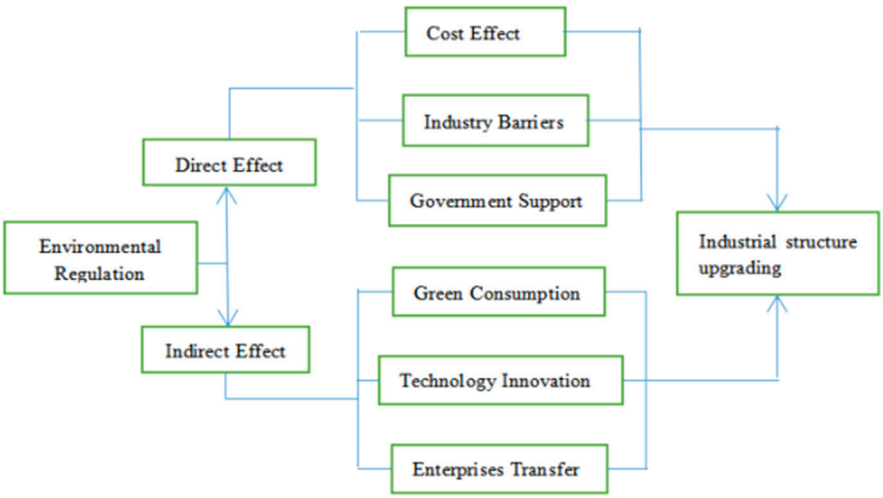



**FIGURE 1 |** The mechanism of ER affecting ISU.

though these indicators could be used as a proxy variable to some degree, they only reflect the consequences of ER. In addition, changes in indicator weights will reduce the reliability of the results even if the same comprehensive indicator is used. Solving environmental pollution requires joint efforts of governments, enterprises, and the public. In particular, public participation in environmental governance can effectively address the asymmetry of governments and enterprises in environmental information and supervision, and it has attracted more and more attention in recent years. Hence, using comprehensive indicators including government environmental supervision, corporate emission data, and public evaluation to measure ER is necessary.

Compared to the existing literature, this study makes the following contributions: 1) this study provides a new stream from the perspective of urban resource endowments to study the impact of ER on ISU of RBCs and NRBCs[2] in China, to expand the existing literature on economic consequences of ER. 2) The PITI of 113 Chinese cities, including government environmental supervision, enterprise emission data, and public evolution, is chosen to measure ER. As a more comprehensive indicator, the PITI contains more aspects of ER and provides empirical evidence for the positive effect of environmental information disclosure on ISU. 3) The results of heterogeneity analysis show that there is not only a heterogeneous effect between RBCs and NRBCs or in different regions but also a heterogeneous effect inside RBCs. This provides evidence for differentiated environmental regulatory policies. 4) The mediation effect and the threshold effect of technological innovation are examined in this study, and it helps to deepen the understanding of the positive role of technological innovation in the impact of ER on ISU.

[2]The National Plan for Sustainable Development of Resource-Based Cities (2013–2020) promulgated by the State Council on December 3, 2013. http://www.gov.cn/zfwj/2013-12/03/content_2540070.htm (Accessed on May 26, 2021)

The rest of this article is arranged as follows: Section *The Mechanism of Environmental Regulation Affecting Industrial Structure Upgrading* is analyzing the mechanism of ER affecting ISU; Section *Sample Data and Variables* is the model design, including variables and data; Section *Model and Results Analysis* is the empirical analysis; Section *Further Discussion* is further discussion. Section *Conclusion and Policy Recommendations* is the conclusions and policy recommendations.

## THE MECHANISM OF ENVIRONMENTAL REGULATION AFFECTING INDUSTRIAL STRUCTURE UPGRADING

This section mainly discusses the direct and indirect effects of ER on ISU, as shown in **Figure 1**.

### Direct Effect

According to "Following cost" hypothesis, the negative externality of environmental pollution would be transferred into enterprises' production costs by ER (Millimet et al., 2009). Enterprises are forced to change the original production and management strategies to meet environmental regulatory standards by adjusting the product structure, rationalizing resources' allocation, and entering or exiting the market. As the intensity of ER increases, it is easier for large-scale pollution-intensive enterprises to achieve internalization of environmental governance cost than small ones because large-scale enterprises can meet the environmental protection standards through buying pollution control equipment and increasing investment in green production technology. However, it is much harder for medium and small enterprises to do what large-scale enterprises have done. The increasing environmental governance cost is more likely to break their break-even point and make profits harder and harder. Eventually, they are forced to quit (Yin et al., 2014). It can be

inferred that environmental regulatory policies produce the effects of survival of the fittest on polluting enterprises.

In addition, enterprises' marginal cost and sunk cost will increase for the sake of green environmental barriers made by ER, when they enter a polluting industry (Ryan, 2012). Green environmental barriers will stop pollution industries from expanding their scale, reduce the ratio of new enterprise entering pollution industries, and produce positive effects on ISU (Cui and Ji, 2011). Meanwhile, governments prefer to make environmental regulation policies to improve the level of green development. These policies such as taxation and subsidies are used to enhance the competitiveness of clean industries and promote the popularization and application of the environmental protection production style in the whole society. As the ratio of clean industries improves, the industry structure will also develop to a higher level.

### Indirect Effect

ER improves the public awareness of green consumption and environmental protection. Environmental label products such as energy efficiency and green certification help consumers eliminate the asymmetry of different enterprises' product information, buy green products, and increase consumers' green consumption (Lyon and Maxwell, 2003; Bjorner et al., 2004; Lu et al., 2016). The increasing market demand for green products will affect enterprises' production strategies, promote enterprises to provide more green products, and promote the green transformation of the industrial structure. In addition, there are two methods for pollution enterprises to solve the increasing production cost caused by ER when facing increasingly strict environmental supervision policies in the long run. On the one hand, pollution enterprises improve the production efficiency by green technology innovation to reduce the cost. Technology innovation promotes industrial division and can significantly reduce enterprises' product cost and utilize sustained power into high-technology industries (Borghesi et al., 2015), thus driving the development of the clean industries and ISU (Chakraborty and Chatterjee, 2017).

On the other hand, moving to regions with lower ER levels could be the second way for pollution enterprises. For economic growth, lower standard environmental regulatory policies are generally made in developing countries. Strict environmental regulatory policies in developed countries lead to the transfer of pollution enterprises to developing countries (Du and Li, 2020). Even though pollution enterprises' multinational transfer will promote the economic growth of developing countries quickly, it is not favorable to upgrading the industry structure of developing countries in the long run. In addition, this phenomenon also occurs in different areas of a country. Polluting enterprises' location transfer within a country will not benefit to ISU of economically backward cities.

## SAMPLE DATA AND VARIABLES

### Sample Data

In 2008, IPEA and NRDC jointly developed the pollution information transparency index (PITI), including 113 cities in China. Among these 113 cities, 52 cities belong to Eastern China, 31 cities to Central China, and 30 cities to the Western China. The lowest GDP is 12.9 billion RMB (city of Tongchuan). In comparison, the highest GDP is 1,406.9 billion RMB (city of Shanghai), indicating that sample cities are not limited to large-scale cities or small cities and can represent the basic situation of Chinese cities. Meanwhile, there may be heterogeneity for RBCs and NRBCs. According to the National Plan for Sustainable Development of Resource-Based Cities, we divided sample cities into RBCs (36 cities) and NRBCs (77 cities) to analyze the impact of ER on ISU. Hence, annual data of 113 cities are used in this study during the period 2008–2017. The sample data are from the China Statistical Yearbook (2009–2018), the China City Statistical Yearbook (2009–2018), the China Statistical Yearbook of Science and Technology (2009–2018), and CSMAR database.

### Dependent Variables

Industrial structural upgrade (ISU) is the dependent variable. Industry structure upgrading is a dynamic process that changes the value of the primary, secondary, and tertiary industries. The critical feature of the industrial structure upgrade is that the growth rate of the tertiary industry is faster than the growth rate of the secondary industry (Wu, 2013). We use the proportion of the output value of the third industry to the second industry to measure ISU because it reflects whether the industrial structure is developing to service-oriented (Zhou et al., 2017).

### Explanatory Variable

Environmental regulation (ER): single indicator such as investment in industrial pollution control, total energy consumption, and treatment rate of industrial wastewater and gas (Domazlicky and Weber, 2004; Lanoie et al., 2008; Zhang et al., 2010; Qian and Liu, 2014) can be used. Alternatively, comprehensive indicators (Zhang et al., 2019) can be used as a proxy variable of ER, but it is still more challenging to measure ER. The PITI conducts eight indicators that reflect the information of government environmental supervision, enterprises' emission, and public participation. Suppose a city has a high PITI score, in that case, it means this city has released more environmental information and more public participation in environmental protection, and made more effects to protect environment. Based on this, this study uses the PITI to measure ER.

### Control Variables

The process of industry structure upgrading is more complex, affected by many economic and social factors. In this study, the major influencing factors are introduced in the method of control variable. 1) Investment in fixed assets (INV): According to Xie et al. (2021), the change in investment structure is closely related to green transformation and green development. The ratio of investment in fixed assets to GDP is used to the level of investment in fixed assets. 2) Foreign direct investment (FDI): It is controversial that FDI promotes ISU in the existing literature. FDI brings advanced management mode and technology, and

**TABLE 1 |** Major variables' statistics.

| Variable | Observations | Mean | Standard error | Min | Max |
|---|---|---|---|---|---|
| ISU | 1,130 | 0.901 | 0.481 | 0.055 | 4.237 |
| ER | 1,130 | 41.693 | 16.717 | 8.300 | 85.300 |
| INV | 1,130 | 0.678 | 0.252 | 0.000 | 2.381 |
| FDI | 1,130 | 0.062 | 0.407 | 0.000 | 7.269 |
| GOV | 1,130 | 0.135 | 0.055 | 0.003 | 0.503 |
| UR | 1,130 | 0.499 | 0.193 | 0.025 | 0.995 |

*Source: Authors' compilation.*

improves the environmental quality of inflow areas (Eskeland and Harrison, 2003; Zhao and Niu, 2013). However, according to the "PHH" hypothesis, List and Co (2000) found that FDI also decreases the environmental quality of inflow areas by transferring pollution enterprises. In view of this, the ratio of FDI to GDP is used to measure the level of FDI. 3) Government intervention (GOV): The local industry structure has largely been affected by governments' policies (Lin, 2012). According to Pan et al. (2008), the government's fiscal expenditure ratio to GDP is used to measure GOV. D) Urbanization (UR): Urbanization can promote population flowing and resource agglomeration, and achieve an external economy. There exists a positive correlation between urbanization and ISU (Han et al., 2012). Urbanization is measured by the ratio of the urban population to the total population. **Table 1** shows the descriptive statistics of the above variables.

# MODEL AND RESULTS ANALYSIS

## Model

The panel data model has cross-sectional and time dimensions. Its data capacity is usually large, which can effectively increase the reliability of the regression model and overcome missing variables to some degree. A panel model containing major variables is established in this study. The empirical model is as follows:

$$LnISU_{it} = \beta_0 + \beta_1 LnER_{it} + \beta_i LnX_{it} + \varphi_i + \delta_i + \varepsilon_{it} \quad (1)$$

where $i$ means the sample city, $t$ means the time, and $ISU$ is the dependent variable, which means industry structure upgrading. $X$ means control variables, and it includes investment in fixed assets (INV): government intervention ($GOV$), foreign direct investment ($FDI$), and urbanization ($UR$); $\varphi$ means the individual fixation effect; $\delta$ means the time fixation effect; and $\varepsilon$ means the random error term.

## The Test of Multiple Collinearities

Variance inflation is used to detect the collinearity problem of the linear regression model. **Table 2** shows the result of variance inflation, and it indicates that the maximum numerical value of variance inflation is 1.95 and the minimum is 1.03. The numerical values of variance inflation of all variables are less than 10. Therefore, it can be inferred that the problem of multiple collinearities among variables is likely to be smaller.

**TABLE 2 |** Variance inflation (VIF).

| | ER | INV | FDI | GOV | UR | MEAN |
|---|---|---|---|---|---|---|
| *VIF* | 1.949 | 1.664 | 1.074 | 1.023 | 1.026 | 1.271 |
| *1/VIF* | 0.513 | 0.601 | 0.931 | 0.974 | 0.975 | 0.787 |

## Empirical Results Analysis

Before examining the impact of ER on industrial structure, we make the LM test and Hausman test for the sample data. The $p$-values of the LM test and Hausman test are 0.00, and this indicates that the fixed-effects model is available in this study. **Table 3** shows the regression results. First, the regression coefficient of ER for ISU in the OLS model is 0.209, and the level of significance is 1% in the whole sample. The OLS model ignores individual effects, which may cause bias in the regression results. Then, we use the fixed-effects model to study the impact of ER on ISU. The regression coefficient of the fixed-effects model is 0.148, and the level of significance is 1%. It shows that ER can produce a positive effect on ISU for sample cities. There exists a possibility of a win-win situation for environmental protection and economic growth.

In terms of control variables, the regression coefficient of investment in fixed assets (INV) is −0.75, obviously negative. It shows that the increasing ratio of investment in fixed assets will make the capital and investment harder to flow into the third industry, and then it will be difficult to promote industry structure upgrading. Since the reforming and opening-up, many industries have gotten rapid growth and formed a much larger scale in China. However, the overcapacity and low technology level of industries have produced serious pollution problems and set obstacles to upgrade the industrial structure for these industries. The coefficient of FDI is 0.093, and the level of significance is 1%, which shows that FDI plays a positive role in ISU. The strict ER in developed countries drives pollution enterprises to developing countries with low ER standards and reduces environmental quality. In recent years, Chinese governments have realized the importance of environment protection and prefer to attract clean FDI. This is conducive to promoting the spillover effects of technology, management, and capital brought about by FDI (Howell, 2019), and produced a positive effect on ISU. The regression coefficient of government intervention (GOV) is generally positive. Fiscal expenditures have a sustainable influence on industrial structural upgrade by increasing the industry's capital stock and improving labor productivity. The regression coefficient of UR is −0.049, which is statistically significant. It shows that as the speed of UR slowed down in China, the role of UR in promoting ISU has become less and less, and even hindered.

## Robustness Test

This article studies the impact of ER on ISU in RBCs and NRBCs, but other factors may affect the reliability of the empirical conclusions. In order to enhance the credibility, the robustness tests are made as follows: 1) Change the regression model: The unobservable variables affecting ISU or the correlation between

**TABLE 3 |** ER's influence on the ISU.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Variable** | **OLS** | **FE** | **RBCs** | **NRBCs** |
| LnER | 0.209** (0.000) | 0.148*** (0.000) | 0.142*** (0.000) | 0.124*** (0.000) |
| *Ln*INV | −0.202** (0.000) | −0.750** (0.000) | −1.340*** (0.000) | −0.590*** (0.000) |
| *Ln*GOV | −0.031*** (0.004) | 0.029** (0.018) | 0.036 (0.134) | 0.026*** (0.005) |
| *Ln*FDI | 0.144*** (0.001) | 0.093*** (0.001) | 0.033* (0.071) | 0.153*** (0.000) |
| *Ln*UR | 0.006 (0.657) | −0.049*** (0.001) | −0.140*** (0.001) | −0.001 (0.954) |
| Obs | 1,130 | 1,130 | 360 | 770 |
| R-square | 0.076 | 0.352 | 0.44 | 0.359 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

**TABLE 4 |** Robustness test 1.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Model | SYS-GMM | SYS-GMM | SYS-GMM | SYS-GMM |
| Variable | ISU | STRU | ISU | STRU |
| L.ISU | 0.914*** (0.000) | | 0.908*** (0.000) | |
| L.STRU | | 1.033*** (0.000) | | 1.038*** (0.000) |
| LnER | 0.058*** (0.000) | 0.006*** (0.000) | | |
| LnER2 | | | −0.125*** (0.000) | −0.005*** (0.000) |
| Control variables | Yes | Yes | yes | Yes |
| Sargen | 70.904*** (0.005) | 92.376*** (0.000) | 65.524*** (0.000) | 90.893*** (0.000) |
| AR (1) | 0.585 | 0.000 | 0.004 | 0.000 |
| AR (2) | 0.000 | 0.497 | 0.755 | 0.648 |
| Wald-test | 0.000 | 0.000 | 0.000 | 0.000 |
| Obs | 1,017 | 1,015 | 980 | 978 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

the dependence variable's lagged term and individual effects of stochastic disturbance can easily lead to endogenous problems. According to Blundell and bond (1998) and Sebastian (2017), SYS-GMM is more effective as it contains horizontal and difference estimation in a system and uses the lagging variable as an instrument variable. Hence, the SYS-GMM is used as a method of robustness test. 2) Replacing variables of ISU: Scholars have not established a uniformed indicator to measure ISU. Based on this, we remeasured ISU again. According to Yu and Wang (2021), the ratios of the first, second, and third industries are used to measure ISU. This measurement indicator, giving the third industry a higher weight, can more sensitively test the changes in the ratio of the third industry. The measurement indicator is as follows:

$$STRU = y_1 + 2y_2 + 3y_3 \quad (2)$$

where $y_i$ means the ratio of $i$ industry in local GDP. The higher the value of STRU, the higher the level of industrial structure is. 3) Replace the variable of ER: The impact of ER on enterprises is reflected in the changes in the discharge of various pollutants. Based on the existing literature and data availability, the annual sulfur dioxide ($SO_2$) emission data of sample cities are used to measure ER in this study. The data are divided by one million tons and then processed by logarithm, named LnER2. The results are shown in **Tables 4**, **5**. The SYS-GMM is used in **Table 4**. The coefficients of LnER and LnER2 are significant, and they have positive impacts on ISU or STRU. When sample cities were divided into RBCs and NRBCs, ER has a significant positive effect on ISU or STRU of RBCs and NRBCs. To sum up, the results of **Tables 4**, **5** show that the previous conclusions are stable.

## Heterogeneity Analysis

The previous empirical result has found the positive effect of ER on ISU, but there might exist the heterogeneity effects of ER on ISU in cities with different resource endowment, regions, and growth stages.

Hence, we conduct the heterogeneity analysis from the perspectives of RBCs and NRBCs, spatial heterogeneity, and growth-stage heterogeneity.

### RBCs and NRBCs

A resource-based city is a city with mining and processing of natural resources such as minerals and forest as its leading industry ((Martinez-Fernandez et al., 2012). The overexploitation of some RBCs leads to the problems of high energy consumption, high pollution, and high emission projects. Promoting the sustainable development of RBCs is of great significance for building a resource-saving and environment-friendly society. Depending on the National Plan for Sustainable Development of Resource-Based Cities (2013–2020), we divided the sample cities into 36 RBCs and 77 NRBCs. The results are shown in the third and fourth columns of **Table 3**. The regression coefficient of RBCs and NRBCs are 0.142 and 0.124, respectively, and the significance is 1%. It shows that ER's positive influence on ISU and the influence on the RBCs

**TABLE 5 |** Robustness test 2.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | STRU | STRU | ISU | ISU | STRU | STRU |
| Variable | RBCs | NRBCs | RBCs | NRBCs | RBCs | NRBCs |
| LnER | 0.024*** (0.000) | 0.018*** (0.000) | | | | |
| LnER2 | | | −0.076*** (0.000) | −0.066*** (0.000) | −0.008*** (0.000) | −0.006*** (0.000) |
| Control variables | Yes | Yes | Yes | Yes | Yes | Yes |
| Obs | 360 | 770 | 344 | 747 | 343 | 747 |
| R-square | 0.344 | 0.347 | 0.429 | 0.432 | 0.277 | 0.521 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

**TABLE 6 |** The impact of ER on ISU from different regions.

| Variable | RBCs | | | NRBCS | | |
|---|---|---|---|---|---|---|
| | Eastern | Central | Western | Eastern | Central | Western |
| *Ln*ER | 0.268*** (0.000) | 0.174*** (0.000) | 0.254*** (0.001) | 0.169*** (0.000) | 0.0800** (0.037) | 0.00653 (0.873) |
| Control variables | Yes | Yes | Yes | Yes | Yes | Yes |
| *Obs* | 140 | 90 | 130 | 450 | 160 | 160 |
| *R-square* | 0.216 | 0.65 | 0.388 | 0.517 | 0.428 | 0.227 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

are much stronger than those on NRBCs. Compared to NRBCs, RBCs' economic growth heavily relies on resource-type industries, and the industry structure is relatively simple. The high-energy, high-pollution, and high-emission industries of RBCs produce serious environmental pollution problems. ER has significantly forced the withdrawal of outdated production capacity of RBCs, promoted the flow of production factors from the industrial sector to the tertiary industry, and promoted ISU.

### Spatial Heterogeneity

Since there might exist heterogeneity in the impact of ER on ISU in different regions, we divided sample cities into eastern, central, and western regions.[3] The empirical results are shown in **Table 6**. The regression results indicate that regional heterogeneity exists. According to **Table 6**, the impact of ER on ISU of eastern regions is much stronger than those of ER on ISU of the central and western regions in resource-based cites and NRBCs. The central and western regions with backward economy depend on high-pollution and energy-consumption industries for economic development, which leads local governments prefer economic growth to environmental protection (Yasmeen et al., 2020). Local governments in eastern regions make more efforts to control environmental pollution. Therefore, the impact of ER on ISU is more significant in the eastern region.

### Growth-Stage Heterogeneity

According to the National Resource-Based City Sustainable Development Plan (2013–2020), the RBCs are divided into growth, mature, declining, and regenerative cities.[4] Resource development is at an early stage in growth cities, but in mature cities, resource exploitation is at a stable level. In declining cities, resources are depleting, but regenerative cities are no longer dependent on natural resources. The difference between these cities may produce different impacts of ER on ISU. The regression result is revealed in **Table 7**.

**Table 7** shows that the regression results of growth, mature, and declining cites are significantly positive and that of regenerative cities is not significant. It indicates that there exists growth-stage heterogeneity in RBCs for the impact of ER on ISU. Facing ER, growth cities can actively introduce green production equipment and technologies, formulate environment-friendly and scientific long-term plans for resource extraction, and pay attention to the coordinated development of different industries. In mature cities, resource industries usually occupy the major position in economic development, and the third industry lacks enough attention and development space, which produces the least positive impact of ER on the industrial structure. In declining cities, resource industries have gradually shrunk, environment pollution pressure has risen significantly, and local government has a stronger motivation to upgrade the industrial structure in order to achieve sustainable growth. The impact of ER on declining cities is obvious. Regenerative cities are no longer dependent on natural resources and have undergone green industry transformation through the development of artificial intelligence, big data, new energy, and other industries. Therefore, ER has little impact on ISU in regenerative cities.

## Mechanism Test

The previous findings show that ISU of RBCs and NRBCs can be significantly promoted by ER. Then, how does ER produce an effect on ISU? According to Porter et al. (1995), reasonable

[3]The region division standard is based on the 2017 National Bureau of Statistics.
[4]Refer to the resource-based city classification in the National Resource-Based City Sustainable Development Plan (2013–2020) published by the State Council of China in 2013.

**TABLE 7 |** Discussion on RBCs in different city life cycles.

| Variable | Growth | Mature | Declining | Regenerative |
|---|---|---|---|---|
| *Ln*ER | 0.258* (0.052) | 0.083* (0.064) | 0.192*** (0.003) | 0.065 (0.256) |
| Control variables | Yes | Yes | Yes | Yes |
| *Obs* | 30 | 200 | 70 | 60 |
| *R-square* | 0.732 | 0.590 | 0.378 | 0.652 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

**TABLE 8 |** Estimation of the influence of ER on technological innovation.

| | RBCs | | | NRBCs | | |
|---|---|---|---|---|---|---|
| **Variable** | **ISU** | **LnTECH** | **ISU** | **ISU** | **LnTECH** | **ISU** |
| *LnER* | 0.142*** (0.000) | 0.347*** (0.000) | 0.105*** (0.002) | 0.124*** (0.000) | 0.328*** (0.000) | 0.110*** (0.000) |
| *LnTECH* | | | 0.236*** (0.000) | | | 0.126*** (0.000) |
| Control variables | Yes | Yes | Yes | Yes | Yes | Yes |
| *Obs* | 360 | 360 | 360 | 770 | 770 | 770 |
| *R-square* | 0.44 | 0.184 | 0.487 | 0.359 | 0.084 | 0.413 |

*Note: Significance:*p < 10%, **p < 5%, and ***p < 1%.*

stringency of environmental regulatory policies is helpful to technological innovations. Taking the expense of pollution control as proxy variable of ER, Rubashkina et al. (2015) found that ER promotes the enterprises' R&D activities. Turken et al. (2020) found that ER impels enterprises to focus more on green technology innovation in order to reduce emissions. Even though ER will significantly increase production costs in the short time, technological innovation enables enterprises to gain long-term competitive advantage. In addition, technological innovation can promote ISU through upgrading traditional industries, eliminating backward production capacity, and strengthening the coordination in different industries (Debnath, 2015).

The mediation effect model could be used to examine whether explanatory variable affects explained variable through mediation variable (Gonzalez and Mackinnon, 2016). To measure the mediation effect of technological innovation, the mediation effect model is made as follows:

$$LnISU_{it} = \beta_0 + \beta_1 LnER_{it} + \beta_i LnX_{it} + \varphi_i + \delta_i + \varepsilon_{it} \quad (3)$$

$$LnTECH_{it} = \alpha_0 + \alpha_1 LnER_{it} + \alpha_i LnX_{it} + \varphi_i + \delta_i + \varepsilon_{it} \quad (4)$$

$$LnISU_{it} = \gamma_0 + \gamma_1 LnER_{it} + \gamma_2 LnTECH_{it} + \gamma_i LnX_{it} + \varphi_i + \delta_i + \varepsilon_{it} \quad (5)$$

where *LnTECH* represents technology innovation, *LnER* represents ER, and *LnISU* represents ISU. If the coefficients of $\alpha$, $\beta$, and $\gamma$ are all significant, the mediation effect of technological innovation exists. A number of patent applications and patent inventions, R&D investment, and number of researchers (Fabrizi et al., 2018; Plank and Doblinger, 2018; Dong and Wang, 2019) have been used to measure technology innovation in the existing literature. The full-time equivalent of R&D personnel is suitable for measuring technological innovation in cities due to statistical differences and large data gaps. According to Wang et al. (2018), the number of prefecture-level cities engaged in scientific research and technical service personnel is approximately selected to measure technological innovation.

**Table 8** shows that the coefficients of the impact of ER on technological innovation in RBCs and NRBCs are 0.347 and 0.328, respectively, which is statistically significant. Both ER and technological innovation have positive and significant effects on ISU in resource-based cites and NRBCs. Hence, there exists the significant mediation effect of technological innovation between ER and ISU. It also indicates that the innovation compensation effect of ER is more significant than the "follow cost effect," and it is conducive to stimulating the innovation activities of enterprises and promoting ISU. The ratio of the mediating effect of technological innovation to the total effect is 25.7% for RBCs and 29.1 for NRBCs. It might be caused by the reason that compared to NRBCs, ER of RBCs not only promoted high-pollution enterprises and overcapacity enterprises, which do not meet the environmental standards and quit from the market, but also promoted enterprises' technological innovation.

## FURTHER DISCUSSION

The previous empirical results indicate that ER has a significant impact on ISU and there exists the significant heterogeneity between ER and ISU in RBCs and NRBCs. The mediation effect of technological innovation is further confirmed. Meanwhile, there are extensive nonlinear relationships between economic variables (Lu et al., 2019; Serdar and Ismet, 2019). Due to the difference in technological innovation of RBCs and NRBCs, there might be a nonlinear relationship between ER and ISU based on technological innovation. The threshold effect model divides the variables into multiple threshold values according to the variables' characteristics and studies the relationship between the variables in different intervals. To further study whether the nonlinear relationship between ER

**TABLE 9 |** Test for the existence of threshold.

| Independent variable | Sample | Threshold test | F | *p*-value | Critical values | | |
|---|---|---|---|---|---|---|---|
| | | | | | 90% | 95% | 99% |
| LnISU | RBCs | Single threshold | 22.60 | 0.060 | 19.801 | 27.013 | 36.104 |
| | | Double thresholds | 20.82 | 0.030 | 15.412 | 17.895 | 23.617 |
| | | Three thresholds | 10.00 | 0.580 | 21.556 | 32.974 | 38.714 |
| | NRBCs | Single threshold | 27.61 | 0.070 | 22.662 | 28.593 | 37.353 |
| | | Double threshold | 21.69 | 0.110 | 21.708 | 26.374 | 35.808 |

*Note:* p-*value and threshold values are from repeating bootstrap 1,000 times.*

**TABLE 10 |** Panel threshold regression.

| Independent variable | Sample | Threshold values | Variable | Coefficients | Std.err. | Control variable |
|---|---|---|---|---|---|---|
| LnISU | RBCs | LnTECH ≤ −3.507 | LnER | 0.0112* | 0.080 | Yes |
| | | −3.507 < LnTECH ≤ −1.020 | LnER | 0.0930** | 0.037 | Yes |
| | | LnTECH > −1.020 | LnER | 0.135*** | 0.034 | Yes |
| | NRBCs | LnTECH ≤ −0.3442 | LnER | 0.101*** | 0.018 | Yes |
| | | LnTECH > −0.3442 | LnER | 0.125*** | 0.017 | Yes |

*Note: Significance:* *p < *10%,* **p < *5%, and* ***p < *1%.*

and ISU exists, a fixed panel threshold effect model is established as follows:

$$LnISU_{it} = \vartheta_0 + \vartheta_1 LnER_{it} * I\left(LnTECH_{it} \le \mu\right) + \vartheta_2 LnER_{it} * I\left(LnTECH_{it} > \mu\right) + \vartheta_i LnX_{it} + \varepsilon_{it} \quad (6)$$

where LnTECH is the threshold variable, μis is the threshold value, and $I$ (*) is the indicator function.

To test the threshold effects between ER and ISU based on technological innovation, the threshold numbers should be examined first. After taking bootstrap estimation 1,000 times, the results of threshold numbers in RBCs and NRBCs are shown in **Table 9**. In RBCs, the tests for the single-threshold effect and double-threshold effect are significant, and *p*-values are 0.06 and 0.03, respectively. However, the test for the three-threshold effect is not significant, and *p*-value is 0.58. In NRBCs, the test for the single-threshold effect is significant, and *p*-value is 0.07, while the test for the double-threshold effect is insignificant and *p*-value is 0.11. The results in **Table 9** show that there exist the threshold effects between ER and ISU in RBCs and NRBCs.

**Table 10** shows the results of the panel threshold effect model for RBCs and NRBCs. In RBCS, when LnTECH ≤ −3.507 (TECH≤0.03), the coefficient of ER is 0.0112, and it indicates that increasing technological innovation can promote the impact of ER on ISU. When −3.507 < LnTECH ≤ −1.020 (0.03 < TECH ≤ 0.3606), the coefficient of ER is 0.0930. When LnTECH ≥ −1.020 (TECH ≥ 0.3606), the coefficient of ER is 0.135. It indicates that with continuous increasing of technological innovation, ER will produce increasing effects on ISU. Compared to NRBCs, RBCs' economic growth heavily relies on natural resource industries, and other industries develop slowly and have lower proportions in the industrial structure. When technological innovation reaches a certain level, ER can significantly stimulate the compensation effect of technological innovation, promote the development of green and clean industries, and thus promote ISU. Hence, taking more technological innovation can help RBCs to achieve a win-win situation in environment protection and ISU as soon as possible. In NRBCs, when LnTECH ≤ −0.3442 (TECH ≤ 0.7088), the coefficient of ER is 0.101. When LnTECH ≥ −0.3442 (TECH ≥ 0.7088), the coefficient of ER is 0.125. It indicates that with the increasing of technological innovation, ER will produce increasing positive effect on ISU. The threshold value of NRBCs is higher than that of RBCs, which indicates that NRBCs need more technological innovations to release the positive effect of ER on ISU.

## CONCLUSION AND POLICY RECOMMENDATIONS

In this study, we use the PITI to measure environmental regulation (ER) of 113 cites during 2008–2017 in China. The sample cities are divided into RBCs and NRBCs to study the impact of ER on ISU. Then, we analyze the heterogeneity of spatial and growth-stage heterogeneities in RBCs and NRBCs. Furthermore, the mediation and the threshold effect of technological innovation are examined. The main conclusions are as follows: first, ER has a significant impact on ISU, and robust tests proved the reliability of this result. Second, analysis of heterogeneity shows ER has a stronger role in promoting ISU in RBCs and the eastern region. Meanwhile, inside RBCs, ER has a stronger impact on ISU in growth RBCs than other RBCs. Third, the mechanism test shows that the mediation effect of technological innovation is significant in RBCs and NRBCs. At last, the impact of ER on ISU has a double-threshold effect in RBCs and a single-threshold effect in NRBCs. With the increasing technological innovation, ER produces increasing effect on ISU of RBCs and NRBCs.

Based on the conclusion of this study, some recommendations are discussed as follows. First, local governments should pay more attention to drive ISU by ER. Meanwhile, there exists significant ER difference in Chinese cities. In 2017, the highest PITI score was 81(Wenzhou), while the lowest score was only 19 (Datong). Local governments should make differentiated ER to release compelling effect of ER on ISU. Second, RBCs are facing stronger conflicts between environmental protection and economic growth. RBCs should continue ER and support green and low-carbon industries, promote ISU to get rid of dependence on natural resources, and achieve a win-win situation for economic growth, ISU, and environmental protection. Third, local governments of RBCs and NRBCs should encourage more technological innovation activities by increasing R&D investment, tax incentives, and subsidies, and guide more funds and R&D personnel to green industries. At last, the establishment of the PITI is based on environmental information disclosure, and this study verifies its positive effect on ISU. Local governments should release more environmental information to the public and guide the public to actively participate in environmental protection and improve their environmental awareness.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

Conceptualization, WS; Formal analysis, WS; Investigation, XB; Methodology, YY; Software, YY; Validation, FT-H; Writing - original draft, WS and YY; Writing - review & editing, XB, FT-H. All authors have read and agreed to the published version of the manuscript.

## REFERENCES


Bjorner, T. B., Hansen, L. G., and Russell, C. S. (2004). Environmental Labeling and Consumers' Choice-An Empirical Analysis of the Effect of the Nordic Swan. *J. Environ. Econ. Manag.* 47 (3), 411–434. doi:10.1016/j.jeem.2003.06.002

Blackman, A., and Kildegaard, A. (2010). Clean Technological Change in Developing-Country Industrial Clusters: Mexican Leather Tanning. *Environ. Econ. Pol. Stud.* 12 (3), 115–132. doi:10.1007/s10018-010-0164-7

Blundell, R., and Bond, S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *J. Econom.* 87 (1), 115e143. doi:10.1016/s0304-4076(98)00009-8

Borghesi, S., Cainelli, G., and Mazzanti, M. (2015). Linking Emission Trading to Environmental Innovation: Evidence from the Italian Manufacturing Industry. *Res. Pol.* 44 (3), 669–683. doi:10.1016/j.respol.2014.10.014

Chakraborty, P., and Chatterjee, C. (2017). Does Environmental Regulation Indirectly Induce Upstream Innovation? New Evidence from India. *Res. Pol.* 46 (5), 939–955. doi:10.1016/j.respol.2017.03.004

Chen, S. (2015). Environmental Pollution Emissions, Regional Productivity Growth and Ecological Economic Development in China. *China Econ. Rev.* 35, 171–182. doi:10.1016/j.chieco.2014.08.005

Cohen, M. A., and Tubb, A. (2018). The Impact of Environmental Regulation on Firm and Country Competitiveness: a Meta-Analysis of the porter Hypothesis. *J. Assoc. Environ. Resource Economists* 5 (2), 371–399. doi:10.1086/695613

Cui, J., and Ji, Y. (2011). "The Environment, Trade and Innovation with Heterogeneous Firms, a Numerical Analysis." in AAEA Working Paper, Pittsburgh, Pennsylvania, July 24-26, 2011, 103478.

Dam, L., and Scholtens, B. (2012). The Curse of the haven: the Impact of Multinational enterprise on Environmental Regulation. *Ecol. Econ.* 78 (12), 148e156. doi:10.1016/j.ecolecon.2012.04.011

Debnath, S. C. (2015). Environmental Regulations Become Restriction or a Cause for Innovation - A Case Study of Toyota Prius and Nissan Leaf. *Proced. - Soc. Behav. Sci.* 195 (3), 324–333. doi:10.1016/j.sbspro.2015.06.360

Domazlicky, B. R., and Weber, W. L. (2004). Does Environmental Protection Lead to Slower Productivity Growth in the Chemical Industry?. *Environ. Resource Econ.* 28 (3), 301–324. doi:10.1023/b:eare.0000031056.93333.3a

Dong, Z. Q., and Wang, H. (2019). The Effect of Environmental Regulation on green Technology Progress of "local Neighborhood. *China's Ind. Econ.* 1 (1), 100–118.

Du, W., and Li, M. (2020). Influence of Environmental Regulation on Promoting the Low-Carbon Transformation of China's Foreign Trade: Based on the Dual Margin of export enterprise. *J. Clean. Prod.* 244, 118687. doi:10.1016/j.jclepro.2019.118687

Duzgorenaydin, N. (2007). Sources and Characteristics of lead Pollution in the Urban Environment of Guangzhou. *Sci. Total Environ.* 385 (1), 182–195. doi:10.1016/j.scitotenv.2007.06.047

Eskeland, G. S., and Harrison, A. E. (2003). Moving to Greener Pastures? Multinationals and the Pollution haven Hypothesis. *J. Develop. Econ.* 70 (1), 1–23. doi:10.1016/s0304-3878(02)00084-6

Fabrizi, A., Guarini, G., and Meliciani, V. (2018). Green Patents, Regulatory Policies and Research Network Policies. *Res. Pol.* 47, 1018–1031. doi:10.1016/j.respol.2018.03.005

Feng, M., and Li, X. (2020). Evaluating the Efficiency of Industrial Environmental Regulation in China:A Three-Stage Data Envelopment Analysis Approach. *J. Clean. Prod.* 242, 118535. doi:10.1016/j.jclepro.2019.118535

Franco, C., and Marin, G. (2013). The Effect of Within-Sector, Upstream and Downstream Energy Taxes on Innovation and Productivity. *Environ. Resour. Econ.* 97 (9), 1–31. doi:10.2139/ssrn.2378884

Gonzalez, O., and Mackinnon, D. P. (2016). A Bifactor Approach to Model Multifaceted Constructs in Statistical Mediation Analysis. *Educ. Psychol. Meas.* 78 (1), 5–31. doi:10.1177/0013164416673689

Gray, W. B., Shadbegian, R. J., Wang, C., and Meral, M. (2014). Do EPA Regulations Affect Labor Demand? Evidence from the Pulp and Paper Industry. *J. Environ. Econ. Manag.* 68 (1), 188e202. doi:10.1016/j.jeem.2014.06.002

Greenstone, M. (2001). The Impacts of Environmental Regulations on Industrial Activity: Evidence from the Evidence from the 1970&1977. Clean Air Act Amendments and the Census of Manufactures. Available at: http://www.nber.org/papers/w8484.

Han, X., Wu, P. L., and Dong, W. L. (2012). An Analysis on Interaction Mechanism of Urbanization and Industrial Structure Evolution in Shandong, China. *Proced. Environ. Sci.* 13 (1), 1291–1300. doi:10.1016/j.proenv.2012.01.122

Howell, A. (2019). Industry Relatedness, FDI Liberalization and the Indigenous Innovation Process in China. *Reg. Stud.* 54, 1–15. doi:10.1080/00343404.2019.1623871

Lanoie, P., Patry, M., and Lajeunesse, R. (2008). Environmental Regulation and Productivity: Testing the Porter Hypothesis. *J. Prod. Anal.* 30 (2), 121–128. doi:10.1007/s11123-008-0108-4

Li, C. (2019). How Does Environmental Regulation Affect Different Approaches of Technical Progress?-Evidence from China's Industrial Sectors from 2005 to 2015. *J. Clean. Prod.* 209, 572–580. doi:10.1016/j.jclepro.2018.10.235

Liao, X., and Shi, X. (2018). Public Appeal, Environmental Regulation and green Investment: Evidence from China. *Energy Policy* 119, 554–562. doi:10.1016/j.enpol.2018.05.020

Lin, Y. F. (2012). New Structural Economics and the Way of China's Development. *China Market* 50 (1), 3–8.

List, J. A., and Co, C. Y. (2000). The Effects of Environmental Regulations on Foreign Direct Investment. *J. Environ. Econ. Manage.* 40 (1), 1–20. doi:10.1006/jeem.1999.1095

Liu, X., and Wang, Q. (2016). An Optimal Emission Mechanism of Sustainability of China: How to Achieve a Win-Win Solution between Economy and Environment?. *J. Syst. Sci. Inf.* 4 (6), 534–546. doi:10.21078/jssi-2016-534-13

Liu, Q. L., Wang, Q., and Liu, Y. (2012). Study on the Relationship between Economic Growth, International Trade and Pollution Emissions: Empirical Analysis on the SO2 Emissions in the United States and China. China Popul. *Resour. Environ.* 5 (5), 170–176.

Lu, S., Zhang, X., Wang, J., and Pei, L. (2016). Impacts of Different media on Constructed Wetlands for Rural Household Sewage Treatment. *J. Clean. Prod.* 127, 325–330. doi:10.1016/j.jclepro.2016.03.166

Lu, S., Bai, X., Li, W., and Wang, N. (2019). Impacts of Climate Change on Water Resources and Grain Production. *Technol. Forecast. Soc. Change* 143, 76–84. doi:10.1016/j.techfore.2019.01.015

Lu, S., Dai, W., Tang, Y., and Guo, M. (2020a). A Review of the Impact of Hydropower Reservoirs on Global Climate Change. *Sci. Total Environ.* 711, 134996. doi:10.1016/j.scitotenv.2019.134996

Lu, S., Li, J., Xiao, B., and Guo, M. (2020b). Analysis of Standard Accounting Method of Economic Compensation for Ecological Pollution in Watershed. *Sci. Total Environ.* 737, 138157. doi:10.1016/j.scitotenv.2020.138157

Lyon, T. P., and Maxwell, J. W. (2003). Self-regulation, Taxation, and Public Voluntary Environmental Agreements. *J. Public Econ.* 87 (7), 1453–1486. doi:10.1016/s0047-2727(01)00221-3

Martin, R., Muûls, M., PREUX, D. E., Wagner, U. J., and Wagner, Ulrich, J. (2014). Industry Compensation under Relocation Risk: a Firm-Level Analysis of the EU Emissions Trading Scheme. *Am. Econ. Rev.* 104 (8), 2482–2508. doi:10.1257/aer.104.8.2482

Martinez-Fernandez, C., Wu, C.-T., Schatz, L. K., Taira, N., and Vargas-hernández, J. G. (2012). The Shrinking Mining City: Urban Dynamics and Contested Territory. *Int. J. Urban Reg. Res.* 36 (2), 245–260. doi:10.1111/j.1468-2427.2011.01094.x

Millimet, D., Roy, S., and Sengupta, A. (2009). Environmental Regulations and Economic Activity: Influence on Market Structure. *Annu. Rev. Resour. Econ.* 1 (1), 99–118. doi:10.1146/annurev.resource.050708.144100

Mulatu, A., Gerlagh, R., Rigby, D., and Wossink, A. (2010). Environmental Regulation and Industry Location in Europe. *Environ. Resource Econ.* 45, 459–479. doi:10.1007/s10640-009-9323-3

Pan, H. B., Xia, X. P., and Yu, M. G. (2008). Government Intervention, Political Connection and Mergers and Acquisitions of Local State-Owned Enterprises. *Econ. Res.* 4 (04), 41–52.

Plank, J., and Doblinger, C. (2018). The Firm-Level Innovation Impact of Public R&D Funding: Evidence from the German Renewable Energy Sector. *Energy Policy* 113, 430–438. doi:10.1016/j.enpol.2017.11.031

Porter, M. E., Van, Der., and Linde, C. (1995). Toward a New conception of the Environment-Competitiveness Relationship. *J. Econ. Perspect.* 9 (4), 91–118. doi:10.1257/jep.9.4.97

Qian, z., and Liu, X. C. (2014). Environmental Regulation, Industrial Structure Adjustment and Regional Economic Development. *Economist* 7 (05), 73–81.

Qiu, L. D., Zhou, M., and Wei, X. (2018). Regulation, Innovation, and Firm Selection: the porter Hypothesis under Monopolistic Competition. *J. Environ. Econ. Manage.* 92, 638–658. doi:10.1016/j.jeem.2017.08.012

Ramanathan, R., He, Q., Black, A., Ghobadian, A., and Gallear, D. (2016). Environmental Regulations, Innovation and Firm Performance: A Revisit of the Porter Hypothesis. *J. Clean. Prod.* 155 (2), 79–92. doi:10.1016/j.jclepro.2016.08.116

Rassier, D. G., and Earnhart, D. (2015). Effects of Environmental Regulation on Actual and Expected Profitability. *Ecol. Econ.* 112, 129–140. doi:10.1016/j.ecolecon.2015.02.011

Rochlitz, M., Kulpina, V., Remington, T., and Yakovlev, A. (2015). Performance Incentives and Economic Growth: Regional Officials in Russia and China. *Eurasian Geogr. Econ.* 56, 421–445. doi:10.1080/15387216.2015.1089411

Rubashkina, Y., Galeotti, M., and Verdolini, E. (2015). Environmental Regulation and Competitiveness: Empirical Evidence on the Porter Hypothesis from European Manufacturing Sectors. *Energy Policy* 83 (35), 288–300. doi:10.1016/j.enpol.2015.02.014

Ryan, S. P. (2012). The Costs of Environmental Regulation in a Concentrated Industry. *Econometrica* 80 (3), 1019–1061. doi:10.3982/ecta6750

Sebastian, Kripfganz. (2017). XTDPDGMM: Stata Module to Perform Generalized Method of Moments Estimation of Linear Dynamic Panel Data Models. Statistical Software Components S458395, Boston College Department of Economics, Revised 08 October 2020. https://ideas.repec.org/c/boc/bocode/s458395.html January 18, 2021).Accessed

Serdar, O., and Ismet, G. (2019). Re-Considering the Fisher Equation for South Korea in the Application of Nonlinear and Linear ARDL Models. *Quant. Finance Econom.* 3 (1), 75–87. doi:10.3934/QFE.2019.1.75

Sun, C., Zhang, F., and Xu, M. (2017). Investigation of Pollution haven Hypothesis for China: an ARDL Approach with Breakpoint Unit Root Tests. *J. Clean. Prod.* 161, 153–164. doi:10.1016/j.jclepro.2017.05.119

Tian, X. L., Guo, Q. G., Han, C., and Ahmad, N. (2016). Different Extent of Environmental Information Disclosure Across Chinese Cities: Contributing Factors and Correlation with Local Pollution. *Glob. Environ. Change* 39, 244–257. doi:10.1016/j.gloenvcha.2016.05.014

Turken, N., Carrillo, J., and Verter, V. (2020). Strategic Supply Chain Decisions under Environmental Regulations: when to Invest in End-Of-Pipe and green Technology. *Eur. J. Oper. Res.* 283, 601–613. doi:10.1016/j.ejor.2019.11.022

Wang, F. Z., Jiang, T., and Guo, X. C. (2018). Government Quality, Environmental Regulation and Enterprise Green Technological Innovation. *Scientific Res. Manage.* 39 (01), 26–33.

Wu, J., Wei, Y. D., Chen, W., and Yuan, F. (2019). Environmental Regulations and Redistribution of Polluting Industries in Transitional China: Understanding Regional and Industrial Differences. *J. Clean. Prod.* 206, 142–155. doi:10.1016/j.jclepro.2018.09.042

Wu, Jinglian. (2013). *The Choice of China's Growth Model*. Fourth Edition. Shanghai: Shanghai Far East Publishers.

Xie, L., Li, Z., Ye, X., and Jiang, Y. (2021). Environmental Regulation and Energy Investment Structure: Empirical Evidence from China's Power Industry. *Technol. Forecast. Soc. Change* 167, 120690. doi:10.1016/j.techfore.2021.120690

Yang, C.-H., Tseng, Y.-H., and Chen, C.-P. (2012). Environmental Regulations, Induced R&D, and Productivity: Evidence from Taiwan's Manufacturing Industries. *Resource Energ. Econ.* 34 (4), 514–532. doi:10.1016/j.reseneeco.2012.05.001

Yao, S., and Liang, H. (2017). Firm Location, Political Geography, and Environmental Information Disclosure. *Appl. Econ.* 49, 251–262. doi:10.1080/00036846.2016.1194966

Yasmeen, H., Tan, Q., Zameer, H., Tan, J., and Nawaz, K. (2020). Manuscript Title: Exploring the Impact of Technological Innovation, Environmental Regulations, and Urbanization on Ecological Efficiency of China in the Context of COP21. *J. Environ. Manage.* 274, 1–9. doi:10.1016/j.jenvman.2020.111210

Yin, K., Wang, R., An, Q., Yao, L., and Liang, J. (2014). Using Eco-Efficiency as an Indicator for Sustainable Urban Development: a Case Study of Chinese Provincial Capital Cities. *Ecol. Indicators* 36, 665–671. doi:10.1016/j.ecolind.2013.09.003

Yu, X., and Wang, P. (2021). Economic Effects Analysis of Environmental Regulation Policy in the Process of Industrial Structure Upgrading: Evidence from Chinese Provincial Panel Data. *Sci. total Environ.* 753, 142004. doi:10.1016/j.scitotenv.2020.142004

Yuan, B., Ren, S., and Chen, X. (2017). Can Environmental Regulation Promote the Coordinated Development of Economy and Environment in China's Manufacturing Industry?-A Panel Data Analysis of 28 Sub-sectors. *J. Clean. Prod.* 149, 11–24. doi:10.1016/j.jclepro.2017.02.065

Zhang, W. B., Zhang, L. Q., and Zhang, K. Y. (2010). Interprovincial Competition Patterns and Evolution of Environmental Regulation Intensity in China: Based on Durbin Fixed Effect Model of Two-Zone System Space. *Manage. World* 12 (1), 34–44.

Zhang, G. X., Zhang, P. D., Zhang, Z. G., and Li, J. X. (2019). Impact of Environmental Regulations on Industrial Structure Upgrading: an Empirical Study on Beijing-Tianjin-Hebei Region in China. *J. Clean. Prod.* 238, 1–11. doi:10.1016/j.jclepro.2019.117848

Zhang, M., Sun, X., and Wang, W. (2020). Study on the Effect of Environmental Regulations and Industrial Structure on Haze Pollution in China from the Dual

Perspective of independence and Linkage. *J. Clean. Prod.* 256, 120748. doi:10.1016/j.jclepro.2020.120748

Zhao, Q., and Niu, M. Y. (2013). Influence Analysis of FDI on China's Industrial Structure Optimization. *Proced. Comput. Sci.* 17 (1), 1015–1022.

Zhao, X., Liu, C., and Yang, M. (2018). The Effects of Environmental Regulation on China's Total Factor Productivity: an Empirical Study of Carbon-Intensive Industries. *J. Clean. Prod.* 179, 325–334. doi:10.1016/j.jclepro.2018.01.100

Zhou, P., Delmas, M. A., and Kohli, A. (2017). Constructing Meaningful Environmental Indices: a Nonparametric Frontier Approach. *J. Environ. Econ. Manage.* 85, 21–34. doi:10.1016/j.jeem.2017.04.003

Zhu, X., Zeng, A., Zhong, M., Huang, J., and Qu, H. (2019). Multiple Impacts of Environmental Regulation on the Steel Industry in China: a Recursive Dynamic Steel Industry Chain CGE Analysis. *J. Clean. Prod.* 210, 490–504. doi:10.1016/j.jclepro.2018.10.350

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Shao, Yin, Bai and Taghizadeh-Hesary. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fenvs.2021.704387

# The Main Research Characteristics of the Development of the Concept of the Circular Economy Concept: A Global Analysis and the Future Agenda

*Víctor Meseguer-Sánchez*[1], *Francisco Jesús Gálvez-Sánchez*[2]*, *Valentin Molina-Moreno*[3] *and Gonzalo Wandosell-Fernández-de-Bobadilla*[2]

[1]International Chair of Social Responsibility, Catholic University of Murcia, Guadalupe, Spain, [2]Department Business Organization, Catholic University of Murcia, Murcia, Spain, [3]Department of Management-1, University of Granada, Granada, Spain

OPEN ACCESS

**Edited by:**
*Xunpeng Roc Shi,*
*University of Technology Sydney, Australia*

**Reviewed by:**
*Nedumaran Swamikannu,*
*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India*
*María De La Cruz Del Río-Rama,*
*University of Vigo, Spain*

***Correspondence:**
*Francisco Jesús Gálvez-Sánchez*
*fjgalvez3@ucam.edu*

**Specialty section:**
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

**Received:** *02 May 2021*
**Accepted:** *16 July 2021*
**Published:** *27 July 2021*

**Citation:**
*Meseguer-Sánchez V, Gálvez-Sánchez FJ, Molina-Moreno V and Wandosell-Fernández-de-Bobadilla G (2021) The Main Research Characteristics of the Development of the Concept of the Circular Economy Concept: A Global Analysis and the Future Agenda. Front. Environ. Sci. 9:704387. doi: 10.3389/fenvs.2021.704387*

In recent years, the concept of the Circular Economy has acquired greater traction, both in the research community and in wider society, as an instrument that provides innumerable methodologies with which to face the environmental, social, and environmental challenges caused by the linear production model. This study focuses on analyzing the degree of advancement of the Circular Economy in the scientific field through a bibliometric analysis (or scientometrics) of 5,007 research articles available in Scopus database and published during the period 2005–2007. The main research characteristics are presented, and the Paris Climate Conference (COP21), held on December 12, 2015, is highlighted as a key driver of this particular line of research. The analysis revealed the main authors, journals, subject area, countries and current and future lines of research, finding that the Circular Economy is a paradigm with environmental, economic and social implications, both for the productive system and for consumers.

**Keywords: circular economy, environmental sustainability, economic sustainability, social sustainability, bibliometrics**

## INTRODUCTION

The traditional linear economy model of "take, make and throw away" is unsustainable (Frosch and Gallopoulos, 1989; Ness, 2008), which establishes the need to transition to more sustainable sociotechnical systems (Seiffert and Loch, 2005; Markard et al., 2012). The externalities of the linear production model are threatening the economic and environmental sustainability of our planet, and the natural ecosystems thereon (EMF, 2012; Geng et al., 2012; Stiehl and Hirth, 2012; Su et al., 2013; European Commission 2014a; European Commission, 2014b; Park and Chertow, 2014). Similarly, society faces high unemployment rates and poor working conditions, causing social vulnerability, which is conceptualized through poverty and increasing inequalities (Sen, 2001; Banerjee et al., 2011). Sustainability requires the development of a balanced production system that takes into consideration economic, social, environmental, and technological aspects (Ren et al., 2013). To this end, the Circular Economy (CE) is a new paradigm that contributes to the positive reconciliation of all these elements (EMF, 2012; Birat, 2015).

The concept of the Circular Economy has gradually gained prominence in political agendas (Brennan et al., 2015), for example, in the European Commission, through the comprehensive package of the European Circular Economy (European Commission, 2015) and, in China, through

the Law for the Promotion of the Circular Economy (Lieder and Rashid, 2016). From a scientific point of view, the Circular Economy (CE) is also gaining increasing attention, which has led to the publication of a large number of case studies, reviews, scientific reports, and research articles, etc., aimed at developing the concept of the Circular Economy (CE) (Yap, 2005; Andersen, 2007; Charonis, 2012; EMF, 2012; Lett, 2014; Naustdalslid, 2014; Prendeville et al., 2014; Club of Rome, 2015; Argudo-Garcia et al., 2017; Molina-Moreno et al., 2019).

The concept of the Circular Economy arises as a set of ideas drawn from different economic and environmental paradigms whose main objective is to improve the efficiency of resources in order to balance the relationship between the economy, the environment, and society (Ghisellini et al., 2016; Murray et al., 2017; Giampietro, 2019). Stahel and Reday (1976) introduced the concept of industrial economics, and Stahel (1982) introduced the need for a shift from property to use thinking. New features have been incorporated into the concept in recent years, such as the cradle-to-cradle theoretical concept (McDonough and Braungart, 2010), ecology (Commoner, 2020), loop economy and performance (Stahel, 2010), regenerative design (Lyle, 1996), biomimicry (Benyus, 1997), and the blue economy (Pauli, 2010).

According to Korhonen et al. (2018), the origin of the Circular Economy is in the ecological economy, while for Andersen (2007) and Su et al. (2013), it was adopted for the first time in 1989 by Pearce and Turner when they explored how natural resources provide the inputs for production and consumption, and eventually become waste, in their investigation of the characteristics of traditional economic systems (Pearce and Turner, 1990).

However, at present, the most recognized definition is that provided by the Ellen MacArthur Foundation, which defines the Circular Economy as "an industrial economy that is restorative or regenerative by intention and design" (Fundación Ellen MacArthur EMF, 2013). In this way, companies need to develop production models based on the 10r philosophy (reduce, reuse, reject, rethink, redistribute, repair, restore, reuse, recycle, and recover) (Bag et al., 2021). Under this philosophy, the concept of redesign is based on the ability of companies to use products and services more intensively and develop awareness in interested parties, focusing on the minimum use of natural resources (Gupta et al., 2019; Jabbour et al., 2020). This encourages the reuse of products that still have operational functionalities, favoring the development of repair systems (Goyal et al., 2018). At the same time, companies must take into account remanufacturing and repair, lengthening the useful life of products, both for the same use and to develop different functionalities (Gurita et al., 2018). Similarly, a more efficient use of recycling, considering circular principles, makes it possible to recover energy and materials, in order to either incorporate them back into the cycle or promote their natural absorption (Fischer and Pascucci, 2017; Bag et al., 2021).

The tension between the excessive use of natural resources and the limits of circular flows is a reality (Sehnem et al., 2019); the Circular Economy presents a new paradigm: "waste is food", which assumes that all materials and products are a temporary deposit of materials or nutrients that will become new inputs and will be incorporated back into the production chain, thus extending their useful life (Tukker, 2015). Therefore, the concept of waste is eliminated, so that the nutrients flow permanently (Smol et al., 2015), thus maintaining the status of waste as a productive resource (Braungart, 2007).

The Circular Economy not only has important implications from an environmental point of view. From an economic point of view, the European Commission considers the Circular Economy to be able to generate transactions that produce annual profits in Europe worth € 600,000 for the manufacturing sector (Fundación Ellen MacArthur EMF, 2013; European Commission, 2014a). According to the Finland's Independence Celebration Fund, the Circular Economy would generate annual profits of 2,500 million euros, and the world economy would generate profits of 1,000 million dollars annually. However, there is a certain relationship between the transition towards the Circular Economy model and the level of development of the territories. According to Singh and Ordoñez (2016), the principles of circularity appear more intensively in developed countries than less developed ones.

Nevertheless, for Korhonen et al. (2018), the circular economy has been developed and led mainly by experts, while the empirical content remains largely unexplored. This limited theoretical development leads to unclear practical implications for sustainability (Giampietro, 2019), resulting in some ambiguity regarding the circular economy. Other authors question the suitability of the concept in terms of social inclusion and climate change (Sehnem et al., 2019). Therefore, this paper attempts to analyze the level of theoretical and scientific development of the circular economy concept and fill the gap about its implications and theoretical benefits on social, economic, and environmental sustainability.

As far as we know, some theoretical reviews have been presented on the concept of circular economy, but we have not identified any bibliometric analysis, which is an appropriate method to uncover key attributes and research themes (Donthu et al., 2021; Kumar et al., 2021). The main objective of this study was to analyze the degree of the scientific or theoretical development of the concept of Circular Economy, based on the productivity of research articles, authors, journals, institutions, and countries considering their international cooperation networks, in order to establish current and future lines of research.

To this end, this research demonstrates a mapping of the circular economy concept through the bibliometric methodology, in which current knowledge and future research trends occupy a central place. Therefore, the following research questions are explored in this review paper:

Q1. What is the trend of CE publications?
Q2. What are the main topic areas in CE?
Q3. Who are the most prolific contributors (authors, journals, countries, and institutions)?
Q4. What are the main international collaborative networks (authors, countries, and institutions)?
Q5. What are the main research topics in CE?

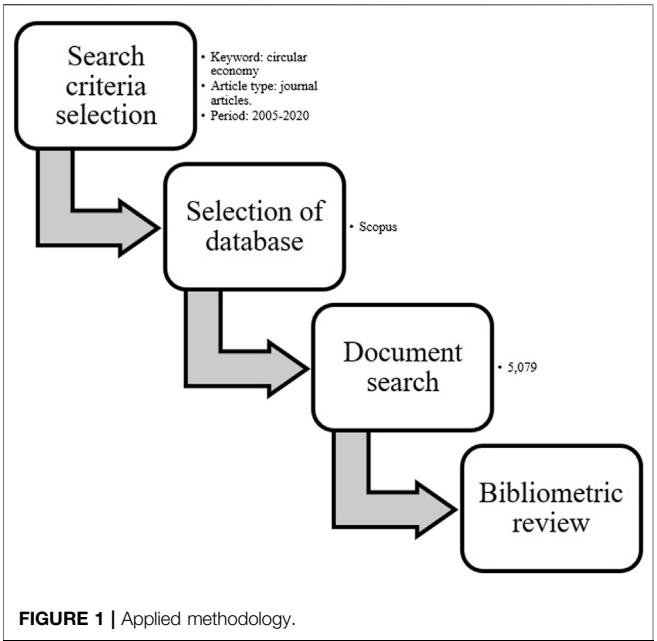



**FIGURE 1 |** Applied methodology.

## MATERIALS AND METHODS

The search for these studies took place in February 2021, and the study was carried out in four phases (**Figure 1** shows the logical sequence applied to obtain the data to be studied). The methodology of this research is based on bibliometric analysis or scientometry, which is a technique applied to publications and allows the extraction of metadata to examine the evolution of an area of knowledge in a certain period of time (Lievrouw, 1989; Cronin, 2001; Zhu and Guan, 2013).

First, the search criteria were selected: (a) keywords: Circular Economy (CE); (b) article type: journal articles. For Paul et al. (2021) bibliometric review should be performed on journal articles because 1) they are evaluated on the basis of novelty and 2) they undergo rigorous peer review, which indicates higher quality; (c) period: the previous 16 years were covered, i.e., the period of time between 2005 and 2020, as is done in other research studies related to bibliometrics (Veer and Khiste -Ganajan, 2017; Liao et al., 2018).

Second, the database was selected. The main databases closely related to the field were consulted, such as Web Of Science, Scopus, PubMed, and Google Scholar (Harzing and Alakangas, 2016; Mongeon and Paul-Hus, 2016). Finally, the research articles from Scopus were selected, as it is the scientific database that has a high number of articles, authors, and journals that meet the scientific quality requirements of peer review (Ackerson and Chapman, 2003; Mingers and Lipitakis, 2010). In addition, this repository provides the largest volume of information in terms of authors, institutions, and countries (Zhang and Eichmann-Kalwara, 2019). Indeed, Scopus is a recognized database to conduct bibliometric reviews (Donthu et al., 2021), as it is recognized as a high-quality repository (Baas et al., 2020), and the correlation with its measurements with those available, for example, in the Web of Science are extremely high (Archambault et al., 2009), although the coverage of the latter is lower (Paul et al., 2021), making it a complete database and high quality for your review.

Finally, we exported the database, carried out the analysis, and established the main conclusions. The sample consisted of 5,007 articles that met the search requirements. This methodology allowed us to analyze the year of publication, the journal, the subject area, the author and co-authors, the institution, the country, the keywords that were included in the research work, and the citations count, the H index, and the Scimago Journal Rank (SJR) impact factor of the main works (Durieux and Gevenois, 2010), establishing the degree of interest in the subject investigated.

We used the VOSwiever tool (Sedighi, 2016; Gálvez-Sánchez et al., 2021; Meseguer-Sánchez et al., 2021) to generate network maps for each of the variables used, allowing us to group and process words.

## RESULTS

### Evolution of Scientific Literature (Q1)

This section shows the results related to the main characteristics of the scientific literature. The 2005–2020 time horizon was divided into four quadrennia to facilitate comparative analysis and understanding.

**Table 1** presents the main results obtained concerning the evaluation of the number of articles published, the authors, the countries, the total and average number of citations, and the number of journals in which the research articles on Circular Economy had been published.

It is clear to see how the scientific literature of each period is at least double that of the previous 4-year period. While in the first quadrennium (2005–2008), 69 articles were published, in the 2017–2020 period, 4,436 were published, representing an increase of 6,429%. The number of publications in the last quadrennium is especially relevant, since it represents 88.6% of the total production of scientific articles on Circular Economy during the 2005–2020 period.

As can be seen in **Figure 2**, although the growth in articles was constant during each period, from the 2013–2016 quadrennium, this growth was exponential. This could have been driven by the Paris Climate Conference (COP21) on December 12, 2015, which was the first universal and legally binding agreement adopted by the European Union and its member states on climate change (COP21, 2015). Although the objective of this work was not to present the agreements dealt with at COP21, its importance should be noted, as it was ratified by the 55 countries that represented at least 55% of world emissions and is considered the first major Summit on environmental protection.

Paris Climate Conference (COP21) seems to have had an important impact, not only at the institutional level, but also in the research community, since the number of authors who delved into this line of research increased considerably throughout the period, especially in the 2017–2020 quadrennium. For example, during the 2005–2008 period, 162 researchers published on the concept, instruments, and methodologies of the Circular

**TABLE 1 |** Characteristics of the scientific literature on Circular Economy.

| Year | Articles | Authors | Countries | Citations | TC/A | Journals |
|---|---|---|---|---|---|---|
| 2005–2008 | 69 | 162 | 10 | 0 | 0.00 | 39 |
| 2009–2012 | 117 | 294 | 17 | 159 | 1.36 | 62 |
| 2013–2016 | 385 | 1,080 | 47 | 1,309 | 3.40 | 189 |
| 2017–2020 | 4,436 | 13,469 | 135 | 28,009 | 8.02 | 1,028 |

*(TC/A): average number of citations per article.*

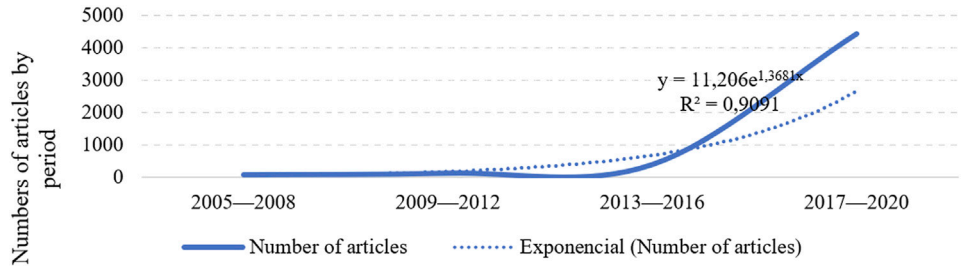



**FIGURE 2 |** Evolution of the number of articles published on Circular Economy.

Economy, and in the quadrennium 2017–2020, 13,469 did so, which represents an increase of 8,314%. It is especially relevant that the number of authors in this last quadrennium represent 90% of all authors. This clearly shows the increasing interest from the scientific community in the development of the Circular Economy as a new paradigm that provides valid solutions to the issues addressed at COP21. Obviously, this growth in the number of authors, which was comparatively greater than the number of articles published, led to an increase in the average number of authors per research article, which went from 2.3 in 2005–2008 to 3 in 2017–2020.

Moreover, the international magnitude of the development of the Circular Economy is clearly observed in the number of countries participating in the research field. For example, in 2005–2008, only 10 countries paid attention to the development of the research line, and in the 2017–2020 quadrennium, a total of 135 did so, which represents an increase of 1,250%.

Regarding the number of citations, the growth was also exponential. In the first period analyzed (2005–2008), no citations were made to other Circular Economy articles in the 69 articles published; however, in the other analyzed periods, a total of 29,477 citations were obtained, of which the last analyzed period (2017–2020) accounted for 28,009, representing 95% of the total citations in the research area. Thus, the average number of citations per article reached 6.31. Together, these two indicators show the degree of interest and development in the Circular Economy in the research community.

Finally, in the 2005–2008 period, 39 journals published research articles on the Circular Economy; in 2017–2020, 1,028 did so, with a total of 1,318 journals interested in this line of investigation. As with all of the indicators discussed above, there was a very significant increase here, further demonstrating the strength of the current trend in this line of research.

## Distribution of Publications by Subject Area (Q2)

This section shows the results of the main subject areas in which research articles on the Circular Economy were published.

For the 2005–2020 period, 27 thematic areas of knowledge related to scientific literature on the Circular Economy were identified in the Scopus database, focusing on articles that could be classified into more than one thematic area (Meseguer-Sánchez et al., 2020). **Figure 3** shows the five subject areas with the highest volume of published research articles.

It is observed how Environmental Science is the thematic area with the highest volume of research articles during the entire 2005–2020 period, with a total of 2,979, which represents 26.33% of the total scientific literature. In volume, this is followed by Energy ($n$ = 1,513, 13.37%), Engineering ($n$ = 1,409, 12.45%), Social Sciences ($n$ = 1,048, 9.26%), and Business, Management, and Accounting ($n$ = 980, 8.66%). The other thematic areas individually represent less than 5%. This diversity of thematic areas shows the transversal nature of the Circular Economy as a paradigm that not only encompasses environmental needs, but also social and economic ones.

## Journals and Authors Productivity (Q3 and Q4)

This section shows the results related to the productivity of authors and their main characteristics, such as international cooperation networks, and the most productive journals in this line of research.

**Table 2** lists the 20 most productive journals on the Circular Economy, and their main characteristics in the 2005–2020 period. Firstly, it is clear to see that 45% are positioned in the first quartile (Q1), while the remaining 40% are in the second quartile (Q2).

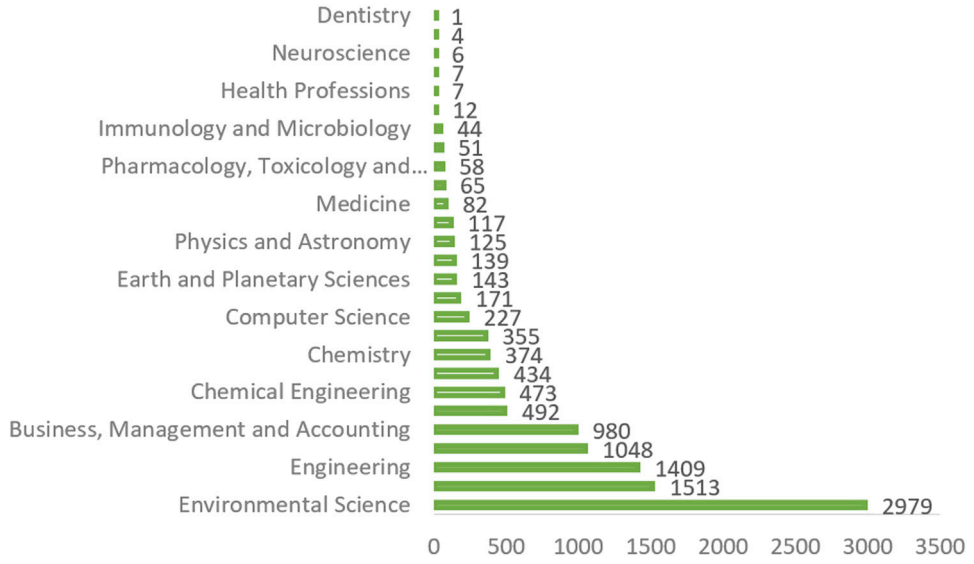



**FIGURE 3 |** Evolution of the number of articles published on the Circular Economy according to subject area.

This indicates that this line of research, despite being relatively immature, is having a very high impact on the international scientific community.

These journals received 2,124 research articles, representing 42% of all articles published on the Circular Economy. The analysis of the origin of the journals shows that 80% were European, while the remaining 20% were of American origin.

The Journal Of Cleaner Production was the journal with the highest volume of articles, with a total of 559 research articles published in the 2005–2020 period, followed by Sustainability Switzerland (422) and Resources Conservation And Recycling (249). The Journal Of Cleaner Production was the first journal to publish research articles on the Circular Economy, with the first publication in 2006, followed by Resources Conservation And Recycling and Journal Of Industrial Ecology, both with their first publications in 2017. These were the only three journals that published research articles on the Circular Economy throughout the entire period analyzed (2005–2020). Moreover, Applied Sciences Switzerland, Detritus, and the International Journal Of Production Research were the latest to publish in this line of research.

The Journal Of Cleaner Production was also the journal with the highest total number of citations, with 13,932 throughout the period, followed by Resources Conservation And Recycling, with 5,551. However, the publication of a significantly lower volume of articles by the Journal Of Industrial Ecology (80), with a high total number of citations (3,266), makes it the journal with the highest average number of citations per article, with 40; i.e., 83 citations were received for each published article. This was followed by the Journal Of Cleaner Production, with 24.92.

The H index of the articles (the number of articles that have at least the same number of citations as that many articles the journal has published in that line of research) shows that, again, the Journal Of Cleaner Production and Sustainability Switzerland demonstrated the best results (144 and 119, respectively). However, the journal's H index was higher in Science Of The Total Environment (224). As for the Scimago Journal Rank (SJR), the journal with the greatest influence was Resources Conservation And Recycling (2,215; Q1), followed by the Journal Of Cleaner Production (1886; Q1).

**Table 3** shows the 10 most productive authors in terms of Circular Economy publications in the 2005–2020 period and their main characteristics. Firstly, it is clear to see that 60% of the authors were European, followed by 20% who were American and 20% who were Asian.

Geng, Y., from Tongji University, China, was the most productive author, with a total of 25 research articles on the Circular Economy published in the 2005–2020 period, followed by Lundström, from Aalto University, Finland, with 20 published articles, who also had the highest H index in the field.

Moreover, Sarkis, J. was the earliest author, first publishing in 2008, followed by Geng, Y. in 2009: these are the only two authors, among the most productive, who began to develop the concept of the Circular Economy in the first decade of the 20th century. For this reason, they are the authors with the highest total numbers of citations and the highest average numbers of citations: Geng, Y. had 2,128 citations, with an average of 85.12 citations per article, which is more than Sarkis, J. with 774 citations and a 59.54 average number of citations per article. However, Geng, Y. did not publish in 2020, and so is the only author from the most productive list to interrupt their scientific literature on the Circular Economy this year.

**Figure 4** shows the international cooperation networks between the main authors on the Circular Economy based on

**TABLE 2 |** Ranking of the 20 most productive journals on the Circular Economy.

| Journal | A | TC | TC/a | H index articles | H index journal | SJR | C | First article | Last article | A | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | 2005–2008 | 2009–2012 | 2013–2016 | 2017–2020 |
| Journal of cleaner production | 559 | 13,932 | 24.92 | 144 | 173 | 1,886 (Q1) | Netherlands | 2006 | 2020 | 2 | 6 | 53 | 498 |
| Sustainability Switzerland | 422 | 3,456 | 8.19 | 119 | 68 | 0.581 (Q2) | Switzerland | 2013 | 2020 | 0 | 0 | 20 | 402 |
| Resources conservation and recycling | 249 | 5,551 | 22.29 | 74 | 119 | 2,215 (Q1) | Netherlands | 2007 | 2020 | 2 | 3 | 15 | 229 |
| Waste management | 125 | 1751 | 14.01 | 39 | 145 | 1,634 (Q1) | United Kingdom | 2009 | 2020 | 0 | 1 | 11 | 113 |
| Science of the total environment | 97 | 1,057 | 10.90 | 37 | 224 | 1,661 (Q1) | Netherlands | 2015 | 2020 | 0 | 0 | 3 | 94 |
| Journal of industrial ecology | 80 | 3,266 | 40.83 | 21 | 95 | 1,808 (Q1) | United States | 2007 | 2020 | 1 | 7 | 4 | 68 |
| ACS sustainable chemistry and engineering | 56 | 311 | 5.55 | 2.3 | 85 | 1,766 (Q1) | United States | 2016 | 2020 | 0 | 0 | 2 | 54 |
| Energies | 55 | 355 | 6.45 | 17 | 78 | 0.635 (Q2) | Switzerland | 2015 | 2020 | 0 | 0 | 2 | 53 |
| Environmental science and pollution research | 53 | 338 | 6.38 | 14 | 98 | 0.788 (Q2) | Germany | 2015 | 2020 | 0 | 0 | 3 | 50 |
| Journal of environmental management | 53 | 951 | 17.94 | 14 | 161 | 1,321 (Q1) | United States | 2010 | 2020 | 0 | 1 | 2 | 50 |
| Procedia environmental science engineering and management | 49 | 66 | 1.35 | 10 | 5 | 0.164 (Q3) | Romania | 2014 | 2019 | 0 | 0 | 4 | 45 |
| Business strategy and the environment | 43 | 745 | 17.33 | 11 | 94 | 1,828 (Q1) | United States | 2017 | 2020 | 0 | 0 | 0 | 43 |
| Waste and biomass valorization | 42 | 241 | 5.74 | 14 | 35 | 0.569 (Q2) | Netherlands | 2015 | 2020 | 0 | 0 | 3 | 39 |
| Resources | 41 | 438 | 10.68 | 14 | 24 | 0.722 (Q2) | Switzerland | 2014 | 2020 | 0 | 0 | 6 | 35 |
| Applied sciences Switzerland | 40 | 114 | 2.85 | 8 | 35 | 0.418 (Q2) | Switzerland | 2018 | 2020 | 0 | 0 | 0 | 40 |
| Materials | 39 | 207 | 5.31 | 12 | 98 | 0.647 (Q2) | Switzerland | 2016 | 2020 | 0 | 0 | 1 | 38 |
| Chemical engineering transactions | 31 | 92 | 2.97 | 14 | 32 | 0.316 (Q3) | Italy | 2015 | 2020 | 0 | 0 | 3 | 28 |
| Detritus | 30 | 43 | 1.43 | 10 | - | - | Italy | 2018 | 2020 | 0 | 0 | 0 | 30 |
| International journal of production research | 30 | 490 | 16.33 | 6 | 125 | 1,176 (Q1) | United Kingdom | 2018 | 2020 | 0 | 0 | 0 | 30 |
| Waste management and research | 30 | 148 | 4.93 | 10 | 73 | 0.650 (Q2) | United Kingdom | 2015 | 2020 | 0 | 0 | 6 | 24 |

*(J): journals (A): number of articles; (TC): number of citations (TC/A): an average number of citations per article; (SJR); Scimago Journal Rank (quartile); (C): country.*

**TABLE 3 |** Ranking of the 10 most productive authors in terms of Circular Economy publications.

| Author | A | TC | TC/a | Institution | C | 1st A | Last A | H index |
|---|---|---|---|---|---|---|---|---|
| Geng, Y. | 25 | 2,128 | 85.12 | Tongji university | China | 2009 | 2019 | 7 |
| Lundström, M. | 20 | 205 | 10.25 | Aalto university | Finland | 2017 | 2020 | 8 |
| Hou, H. | 15 | 46 | 3.07 | Kunming university of science and technology | China | 2017 | 2020 | 3 |
| Torretta, V. | 14 | 110 | 7.86 | Università degli studi dell'Insubria | Italy | 2018 | 2020 | 4 |
| Liu, X. | 13 | 3.4 | 2.62 | Kunming university of science and technology | China | 2018 | 2020 | 3 |
| Sarkis, J. | 13 | 774 | 59.54 | Worcester polytechnic institute | United States | 2008 | 2020 | 3 |
| Smol, M. | 13 | 148 | 11.38 | Mineral and energy economy research institute of the polish academy of sciences | Poland | 2016 | 2020 | 3 |
| Azapagic, A. | 12 | 271 | 22.58 | The university of manchester | United Kingdom | 2017 | 2020 | 4 |
| Bocken, N. | 12 | 425 | 35.42 | The international institute for industrial environmental economics | Sweden | 2017 | 2020 | 5 |
| Dewulf, J. | 12 | 292 | 24.33 | ETH zürich | Switzerland | 2017 | 2020 | 2 |

*(A): number of articles; (TC) number of citations; (C): country; (TC/A): an average number of citations per article; (First A): first article published; (Last A): last article publicated; (H-index): Hirsch index in this research area.*

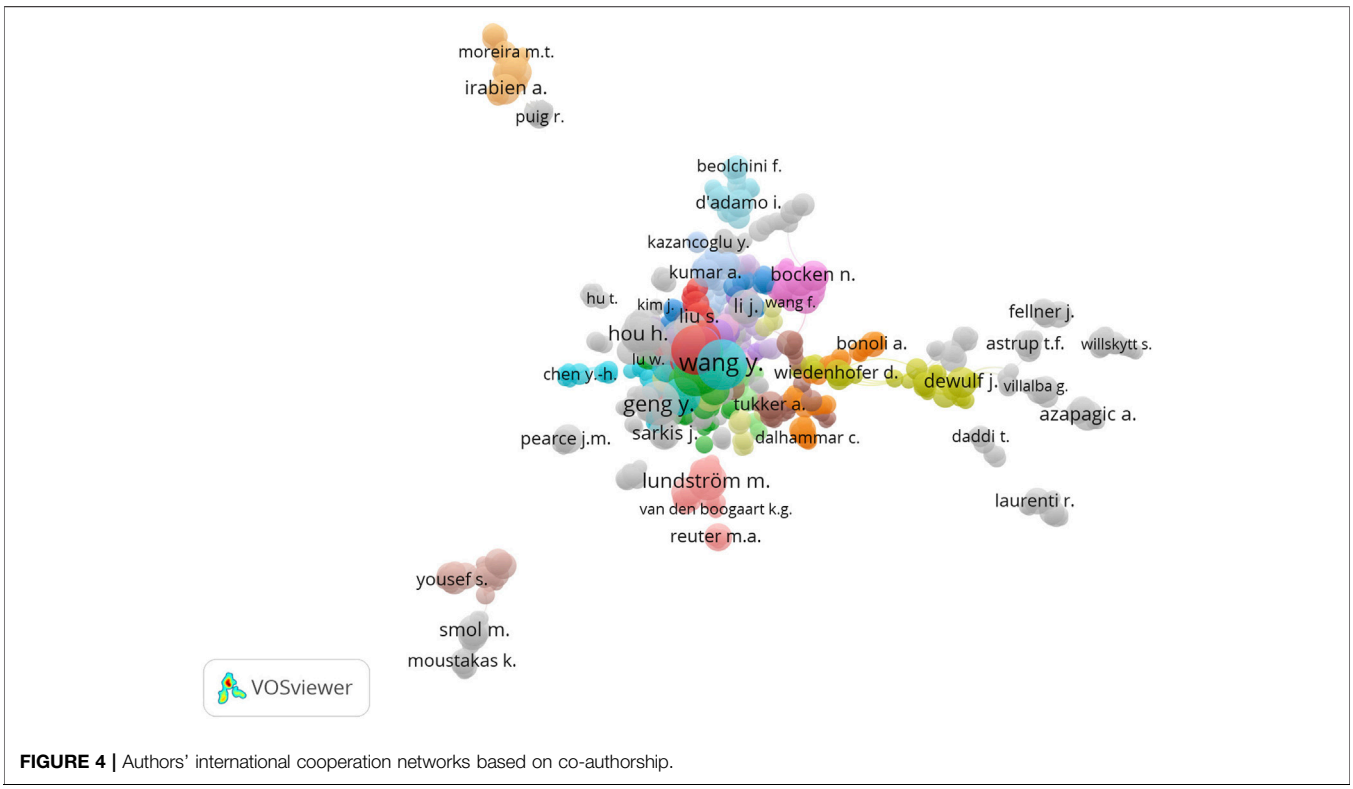



**FIGURE 4 |** Authors' international cooperation networks based on co-authorship.

the analysis of co-authorship. The analysis of cooperation networks allows us to understand the relationships between researchers and the dissemination of knowledge (Chen, 2006). Collaborations can contribute to the exchange of ideas that generate new research, with synergies increasing opportunities for publication in quality journals (Acedo et al., 2006). The colors show the working groups, and the sizes of the circles shows the volumes of the scientific literature developed.

Selecting an interaction of at least three co-authored published research papers, a total of 36 clusters were obtained, representing 709 authors. Thus, with the exception of Torretta, V. from Italy, all the authors within the 10 most productive belonged to different international cooperation clusters. This shows, on the one hand, that there is strong international cooperation in scientific literature on the Circular Economy and, on the other, that these international cooperation networks have very strong and productive ties.

## Productivity of Institutions and Countries (Q3 and Q4)

Here, we show the results of the main characteristics of the most productive institutions and countries in terms of the Circular Economy production, their cooperation rates, and their international cooperation networks over the 2005–2020 period.

**Table 4** shows the 10 most productive institutions in the field of the Circular Economy, in which 80% were European and the remaining 20% were Asian.

**TABLE 4 |** Ranking of the 10 most productive institutions in the field of Circular Economy.

| Institution | C | A | TC | TC/a | H index | CI (%) | TC/a | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | IC | NIC |
| Delft university of technology | Netherlands | 112 | 3,368 | 30.07 | 26 | 36.6 | 38.93 | 24.96 |
| Aalto university | Finland | 73 | 853 | 11.68 | 29 | 42.5 | 11.10 | 12.12 |
| Chinese academy of sciences | China | 68 | 2,282 | 33.56 | 16 | 51.5 | 53.29 | 12.64 |
| Consiglio nazionale delle ricerche | Italy | 59 | 289 | 4.90 | 2.3 | 40.7 | 6.08 | 4.09 |
| Lunds universitet | Sweden | 57 | 1,236 | 21.68 | 21 | 63.2 | 22.22 | 20.76 |
| Tsinghua university | China | 54 | 1,008 | 18.67 | 17 | 27.8 | 26.60 | 15.62 |
| Danmarks tekniske universitet | Denmark | 52 | 811 | 15.60 | 19 | 50.0 | 10.92 | 20.27 |
| Politecnico di Milano | Italy | 51 | 870 | 17.06 | 18 | 41.2 | 20.90 | 14.37 |
| CNRS center national de la recherche scientifique | France | 49 | 313 | 6.39 | 15 | 59.2 | 7.52 | 4.75 |
| Alma mater studiorum university of bologna | Italy | 49 | 1,667 | 34.02 | 6 | 44.9 | 69.73 | 4.93 |

*(C): country (A): number of articles; (TC) number of citations (TC/A): an average number of citations per article; (H-index): Hirsch index in this research area (CI%): percentage of articles produced with international collaboration; (IC): number of citations per article made with international collaboration; (NIC): number of citations per article made without international collaboration.*

**TABLE 5 |** Ranking of the 10 most productive countries in terms of Circular Economy publications.

| Country | A | TC | TC/a | H index | A | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2005–2008 | 2009–2012 | 2013–2016 | 2017–2020 |
| Italy | 721 | 7,983 | 11.07 | 171 | 0 | 3 | 39 | 679 |
| China | 683 | 10,182 | 14.91 | 135 | 59 | 101 | 121 | 402 |
| United Kingdom | 582 | 11,680 | 20.07 | 127 | 4 | 1 | 46 | 531 |
| Spain | 556 | 4,727 | 8.50 | 156 | 0 | 0 | 20 | 536 |
| United States | 321 | 5,487 | 17.09 | 68 | 3 | 7 | 19 | 292 |
| Germany | 314 | 4,135 | 13.17 | 76 | 0 | 1 | 33 | 280 |
| Netherlands | 311 | 6,936 | 22.30 | 73 | 0 | 1 | 33 | 277 |
| Sweden | 249 | 5,772 | 23.18 | 52 | 1 | 1 | 28 | 219 |
| France | 225 | 2,622 | 11.65 | 65 | 0 | 1 | 18 | 206 |
| Finland | 208 | 3,317 | 15.95 | 54 | 0 | 1 | 12 | 195 |

*(A): number of articles; (R): position rank by number of articles in the 4-year period; (TC): number of citations; (TC/A): number of citations per article; (H-index): Hirsch index on the research topic.*

Delft University of Technology was the most productive institution, with a total of 112 research articles on Circular Economy published, followed by Aalto University and Chinese Academy of Sciences, with 73 and 68 articles published, respectively. Regarding the total number of citations received, Delft University of Technology had the highest volume, with a total of 3,368, followed by the Chinese Academy of Sciences with 2,282, the latter also having the second highest average number of citations per published article, with 33.56, behind the Alma Mater Studiorum Università di Bologna, with an average of 34.02.

Despite the fact that Aalto University and Delft University of Technology were the two institutions from the most productive list with a higher H index in the research area (with 29 and 26, respectively), they were among the institutions that presented the worst results in the index international cooperation (with 42.5 and 36.6%, respectively).

In this sense, Lunds Universitet was the institution with the highest rate of international cooperation (with 63.2% of published articles), followed by CNRS Center National de la Recherche Scientifique, with 59.2% of articles published on the Circular Economy. Together with the Chinese Academy of Sciences and Danmarks Tekniske Universitet, they were the four institutions that published most articles with international co-authorship as opposed to domestic authors.

However, despite these data, 80% of the most productive institutions received a higher average number of citations per article when their articles were published with international co-authors. Alma Mater Studiorum Università di Bologna had the highest average number of citations in terms of international cooperation, with 69.73 citations for each published article. It is noteworthy that Alma Mater Studiorum Università di Bologna only received an average of 4.93 citations for articles published with only domestic authors. On the contrary, Danmarks Tekniske Universitet, which published 50% of its articles with international cooperation, received half the average number of citations as compared to articles in which only local authors contributed.

**Table 5** shows the main characteristics of the countries that contributed to the development of the Circular Economy concept in the 2005–2020 period.

Italy had the highest volume of research articles on the Circular Economy published, with a total of 721, followed by China, with 683. Moreover, they were the two countries with the highest H index in this line of research (with 171 and 135, respectively). However, the United Kingdom was the country

**TABLE 6 |** International cooperation networks of the 10 most productive countries in terms of Circular Economy publications.

| Country | NC | Main collaborators | CI (%) | TC/a | |
|---|---|---|---|---|---|
| | | | | IC | NIC |
| Italy | 59 | Spain, United Kingdom, Netherlands, Germany, Sweden | 37.2 | 16.35 | 7.95 |
| China | 57 | United States, United Kingdom, Netherlands, Italy, Australia | 34.0 | 27.55 | 8.41 |
| United Kingdom | 67 | China, United States, Italy, Netherlands, Spain | 52.9 | 21.31 | 18.68 |
| Spain | 47 | Italy, United Kingdom, Portugal, United States, Germany | 39.4 | 10.90 | 6.94 |
| United States | 54 | China, United Kingdom, France, Finland, Germany | 62.0 | 18.93 | 14.10 |
| Germany | 60 | Italy, United Kingdom, Netherlands, Finland, United States | 52.2 | 17.83 | 8.07 |
| Netherlands | 47 | United Kingdom, Italy, China, Germany, Belgium | 56.9 | 27.27 | 15.75 |
| Sweden | 45 | Italy, Finland, United Kingdom, Netherlands, United States | 54.6 | 26.92 | 18.68 |
| France | 61 | United Kingdom, United States, Italy, Netherlands, Canada | 62.2 | 13.99 | 7.81 |
| Finland | 37 | Sweden, United States, Germany, Netherlands, China | 48.6 | 22.67 | 9.60 |

*(NC): number of collaborative countries; (CI): percentage of countries in collaborative papers; (TC/A): number of citations per article; (IC): international cooperation; (NIC): no international cooperation.*

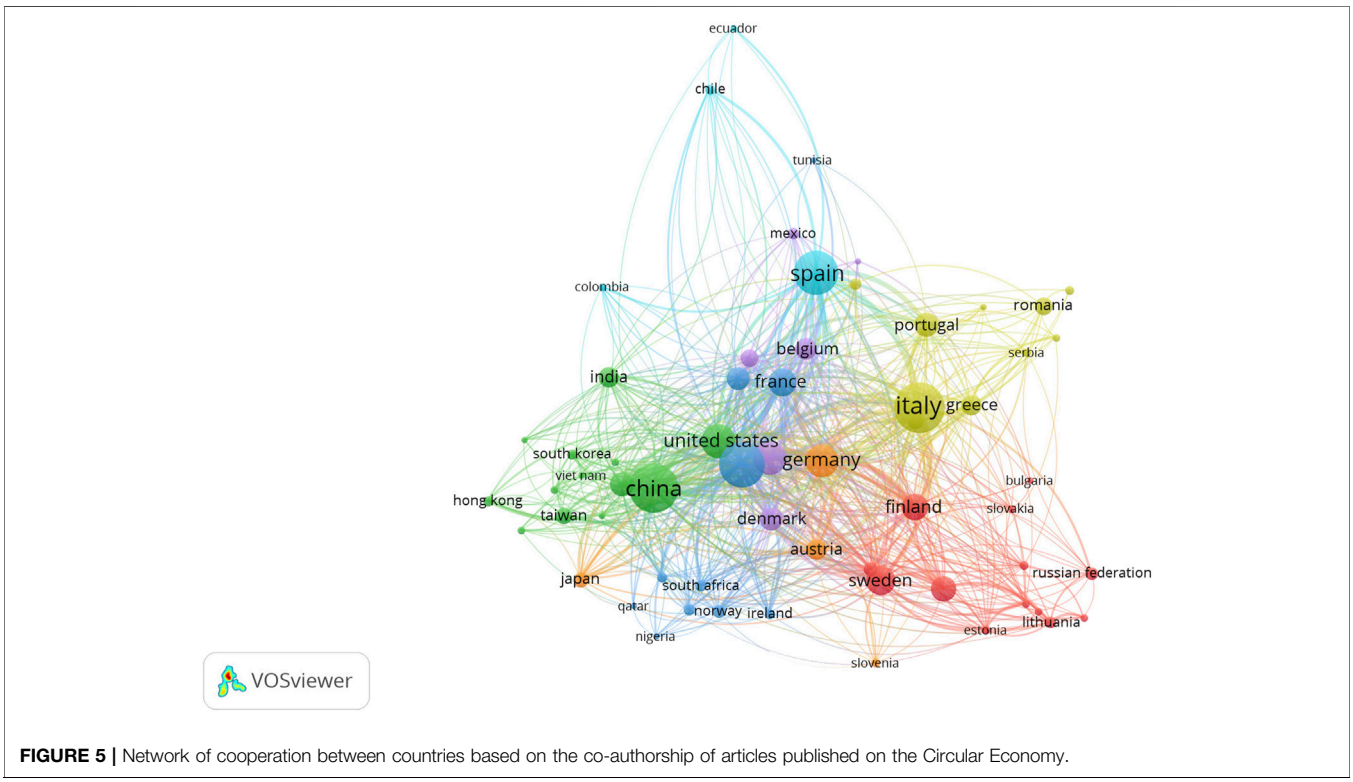



**FIGURE 5 |** Network of cooperation between countries based on the co-authorship of articles published on the Circular Economy.

with the highest number of total citations, with 11,680 for its 582 published articles; while China had a total of 10,182 citations. Despite this, Sweden and the Netherlands received the highest average number of citations per article, with 23.18 and 22.30 for each published article, respectively.

Finally, it should be noted that China, the United Kingdom, the United States, and Sweden were the only countries that published articles on the Circular Economy throughout the 16 years analyzed. In fact, if it were not for the strong growth in publications in the final 4 years in Italy, China would have been the most productive country, since it was in first position in all the periods analyzed, except for the final one.

**Table 6** shows the results of the analysis of the international cooperation networks of the most productive countries. We see that United Kingdom had the highest number of international collaborators, with a total of 67, followed by France and Germany, with 61 and 60, respectively.

France and the United States were the countries with the highest rates of cooperation, with 62.2 and 62% of their articles published with international co-authors, respectively. Only Italy, China, Spain, and Finland published more articles with domestic authors as opposed to international co-authors. However, all countries received a higher average number of citations per published article when they published through international cooperation networks, with China, Sweden, and the

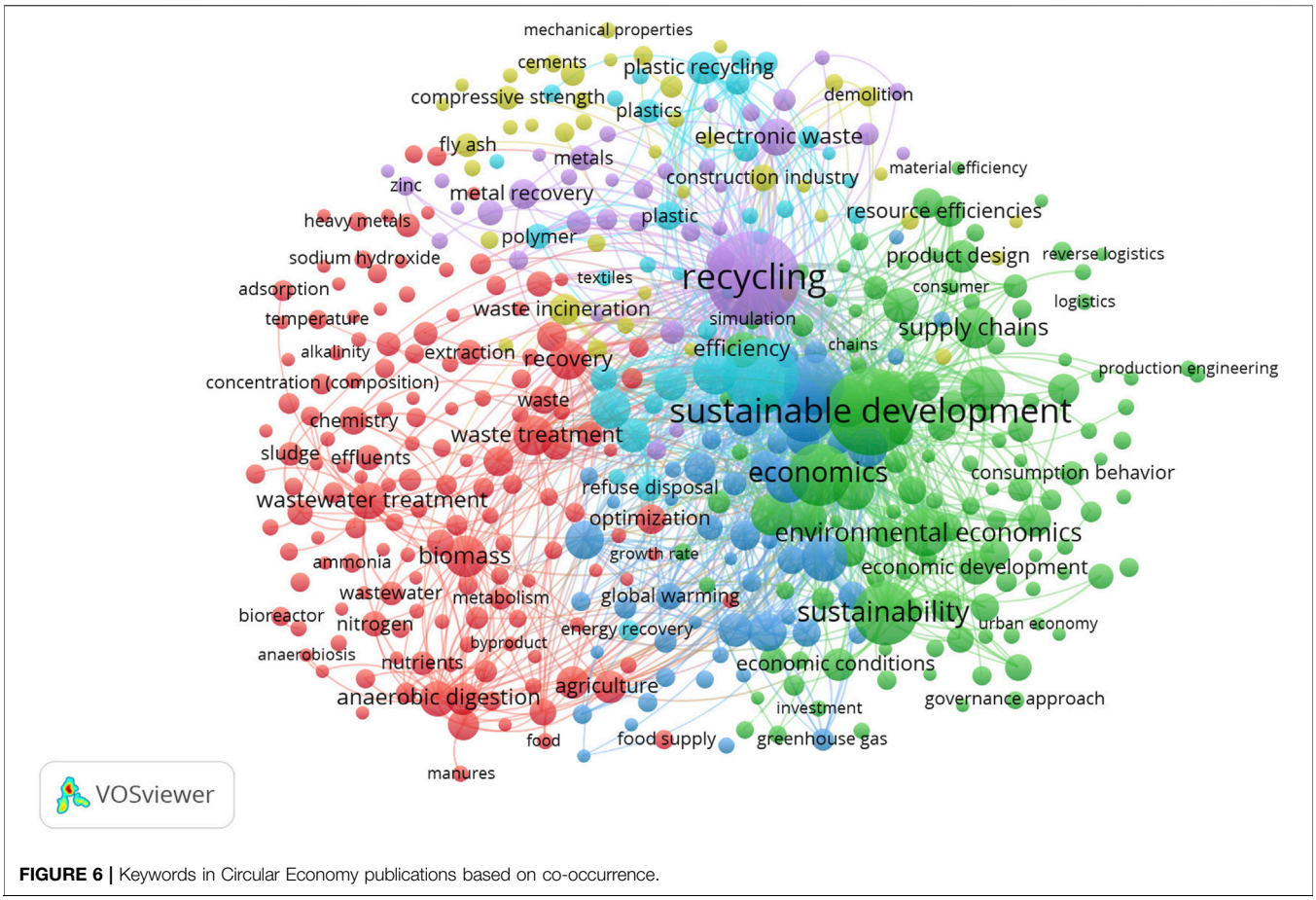



**FIGURE 6 |** Keywords in Circular Economy publications based on co-occurrence.

Netherlands demonstrating the highest average numbers of citations in the publication of articles through international cooperation (with 27, 55, 27.27, and 26.92, respectively).

Finally, **Figure 5** shows the collaboration map between the main countries based on the co-authorship of published research articles on the Circular Economy. The colors show the networks and the size of the circles indicates the productivity of the networks. Given the high number of articles published in this field, it was established that there was an interaction of at least 10 studies that were published with international co-authorship.

The results show an international cooperation network made up of 61 countries, grouped into seven international cooperation networks. The first cluster, red in color, brings together the largest number of countries, with a total of 13, led by Sweden and Finland. With the exception of Egypt, all its members are European.

The green international cooperation network is also made up of 13 countries, led by China and the United States, and is mainly made up of countries in Asia. On the other hand, the dark blue cluster, formed by 11 countries, and led by the United Kingdom, has a greater continental diversity, with African, European, American, and Asian countries.

Italy leads the yellow cluster, which forms an international network of 10 countries, which, excepting *Argentina*, are all European. The purple cluster, led by Netherlands, made up of six European, American, and African countries, is of a very similar composition to the dark blue cluster.

Finally, the light blue cluster, led by Spain and made up of only four countries, shows a strong cultural link, as it is mainly made up of South American countries. The orange cluster shows the same trend, also being made up of four countries led by Germany, and demonstrating a strong relationship with the Asian continent through the presence of Japan.

## KEYWORD ANALYSIS

This section provides the results of the analysis of keywords used by the authors with a relevant link to this field, the network map of these keywords based on co-occurrence, and their evolution during the period of time analyzed (2005–2020). The premise is that keywords are representative of the content of the article (Comerio and Strozzi, 2019). The co-occurrence of keywords helps to create an understanding of a research field (Ding et al., 2001). From a total of 5,007 research articles, 170,048 keywords were obtained.

On the basis of the co-occurrence method, **Figure 6** shows the main keywords used over the 16 year period, selecting those with at least 20 interactions. The analysis of the co-occurrence of these keywords is a very effective method to

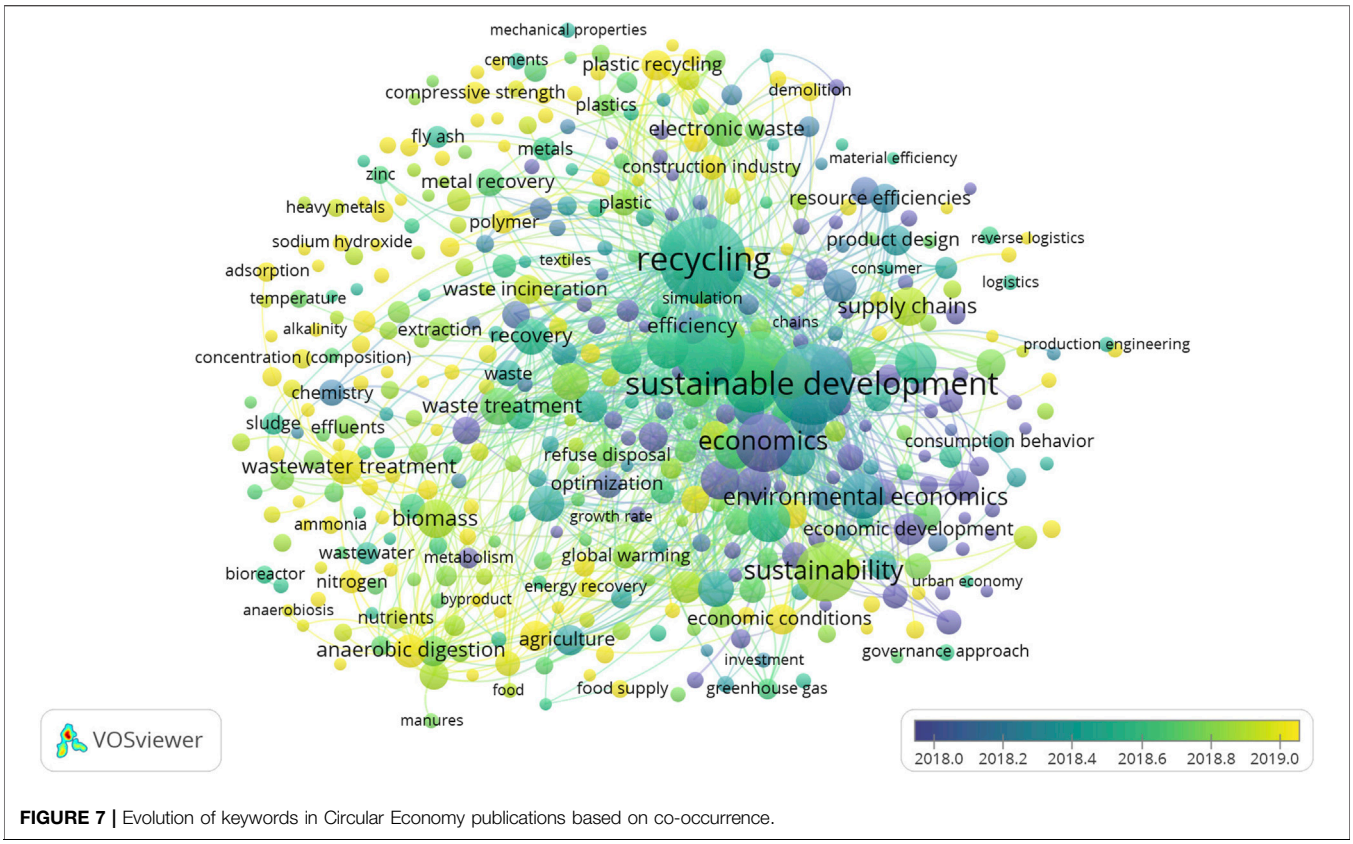



**FIGURE 7 |** Evolution of keywords in Circular Economy publications based on co-occurrence.

explore the main research topics and new research trends (Fang et al., 2018). In this way, Voswiever extracts the keywords from the documents, calculates the frequency, and finally acquires the matrix of keywords (Park and Nagy, 2018).

Excluding the concept of the Circular Economy to avoid erroneous conclusions, 438 keywords were obtained that were grouped around six clusters. The color of the circles shows the keyword groupings, and the size refers to the number of times it was used.

The first cluster, in red, which brings together the greatest number of keywords (147), refers to recuperative technologies, the main objective of which is to reduce the use of energy in the production process through reuse, thus reducing the amount of gases emitted. Numerous studies address the development of recuperative technology in mineral resources, such as natural gas (Nikpey et al., 2014; Abdul-Qyyum et al., 2018), biogas (Cobbledick et al., 2016; Yang et al., 2017), coal (Kuchonthara et al., 2005), and oil (Kansha et al., 2012), among others.

On the other hand, the second cluster, in green, is made up of 127 keywords and defines the Circular Economy as a transversal paradigm that provides solutions aimed at sustainable development in all its dimensions, such as the social and the social economy. In this way, the Triple Bottom concept argues that companies adopting environmentally responsible behavior will also obtain significant economic and social benefits (Alhaddi, 2015; Gimenez et al., 2012; McWilliams et al., 2016, among others). This has given rise to new businesses that favor the labor inclusion of vulnerable groups (Martínez-Medina et al., 2021).

The third cluster, in dark blue, is composed of 52 keywords and focuses on the paradigm shift to the Circular Economy, i.e., transitioning from the linear economy model to a waste management system in response to the need for urban spaces to effectively and efficiently manage local waste (Sharholy et al., 2008; Guerrero et al., 2013).

The yellow cluster is composed of 42 keywords focused on industry 4.0 and materials, and the main objective of integrating information and communication technologies (ICT) into manufacturing systems to promote industrial business performance. It is considered that industry 4.0 could represent a new industrial era (Lasi et al., 2014; Parlanti, 2017; Reischauer, 2018).

The purple cluster, composed of 36 keywords on critical raw materials, refers to understanding that mineral resources, whose reserves are limited and distribution is uneven, are extremely necessary for the development of certain functions in technology and production (Hofmann et al., 2018), thus establishing important economic, social, environmental, and political challenges.

Finally, the light blue cluster is made up of 34 by-products and waste items, and its main objective is the development of techniques for waste management, recovery, reprocessing, and commercialization, thus reducing environmental impact by increasing the shelf life of original products.

**Figure 7** shows the degree of maturity of the Circular Economy keywords.

The legend in the figure shows a time horizon of 2 years, which demonstrates that practically all keywords recently acquired importance, indicating a strong degree of evolution in the Circular Economy field. The development of recuperative technologies, the management of the supply chain, and the economic aspects derived from the Circular Economy and industry 4.0 are particularly relevant today for the research community.

## CONCLUSION AND DISCUSSIONS

This study aimed to analyze the degree of development of the concept of the Circular Economy from a scientific or theoretical perspective in the 2005–2020 period, using a bibliometric analysis (or the scientometry methodology) of 5,007 research articles located in the Scopus database. Although for some authors the scientific basis of the Circular Economy dates from the 18th and 19th centuries (Desrochers, 2002; Desrochers, 2004), the first research article available in Scopus that includes the term "Circular Economy" in the title of the article, abstract, or keywords is from 2004 (Dajian, 2004; Lei, and Yi, 2004; Strebel and Posch, 2004; Xu et al., 2004).

From reading the most relevant research articles in CE and analyzing the co-occurrence of the keywords, we extract a theoretical understanding and the proposed lines of research in this trend of research.

The theoretical concept of circular economy is derived from the combination of ideas from previous currents of thought whose main objective was to reduce the consumption of natural capital, cleaner production, or reuse of inputs (e.g., industrial ecology, cradle to cradle theory, performance economics, natural capitalism and in recent years new trends such as the concept of zero net emissions, among others). The original principles of the circular economy were reduction, reuse, and recycling (Feng and Yan, 2007; Lett, 2014). **Figure 6** shows a high number of keywords associated with these principles, highlighting that currently "recycling" appears in 20.9% of published research articles on the circular economy. This principle is very attractive from an environmental point of view, as it requires less use of natural resources, less energy and less labor (Castellani et al., 2015), while the principles of reuse and recycling aim to extend the life cycle of the product and reintroduce it into the production cycle at the end.

Internationally, the circular economy is considered an essential strategy for cleaner production (Bilitewski, 2012), as it introduces more environmentally friendly products and processes that reduce the flow of non-renewable and harmful inputs (Van Berkel et al., 1997; Van Berkel, 2007). The new research trends presented in **Figure 7** demonstrate that recuperative technologies, supply chain management, and Industry 4.0 will contribute to the development of environmental benefits in the coming years. Therefore, we propose the following research question for future research:

Q1. Is CE the paradigm that will help achieve net zero emissions by 2050?

With the theoretical development of the circular economy, its principles have been gradually expanded and currently, there are 10 (refuse, rethink, reduce, reuse, repair, refurbish, remanufacture, repurpose, recycle, and recover) (Bag et al., 2021). This has led to a larger dimension that includes important economic and social implications.

Redistribution represents a rethinking of the economy by giving a greater role to use than to ownership. This undoubtedly allows for a strengthening of the social base and creates a "safer and more equitable space for society" (Raworth, 2017). Along with the redesign, companies can offer products and services with high social and environmental implications to their customers, generating shared value (Dyllick and Muff, 2015). Indeed, Belmonte-Ureña et al. (2021) find that the circular economy contributes to the fulfillment of Sustainable Development Goals with high social implications, such as SDGs 1, 2, 3, 4, 5, 8, and 10. However, in practical application, this requires governments to value the benefits in resource efficiency, greenhouse gas emissions reduction, economic growth, and job creation (Wijkman and Skånberg, 2015; Genovese et al., 2017).

One of the major global challenges of our time is the inclusion of vulnerable groups. In this way, we propose the following research question:

Q2. Is CE a new paradigm that creates opportunities for social and professional inclusion for vulnerable groups?

There are five main conclusions of the bibliometric review:

a) Since 2004 there has been a significant and exponential increase in the number of research papers on the Circular Economy, especially in the 2017–2020 period. This was a consequence of the celebration of COP21 in Paris in 2015, in which the governments of 24 countries and the European Commission released 4.6 billion dollars annually for research into clean energies (Gates, 2021) (Q1).
b) The analysis of the subject areas showed that Environmental Science is the subject area with the highest volume of research articles; however, the five main subject areas also include Energy, Engineering, Social Sciences, Business, Management, and Accounting, which indicates both the high degree of transversality in the field and its multidisciplinary nature (Q2).
c) The authors' analysis shows an 8.314% increase in publications in just 16 years, with the Chinese Geng, Y. being the most prolific author in the field with 25 published research articles. the Journal Of Cleaner Production and Sustainability Switzerland were the two most productive journals in the period of time analyzed, with 559 and 422 published research articles, respectively. Delft University of Technology and Aalto University were the most productive, with 112 and 73 research articles published. With regard to countries, Italy and China were the most productive, with 721 and 683 articles published (Q3). Althoug China is the first nation with specific Circular Economy legislation (Beaulieu, 2015), in all of the indicators analyzed, the clear predominance of Europe is observed, probably motivated by the fact that a significant number of the countries participating in the COP21 in Paris were European, which, together with the consideration that more developed countries have a greater degree of implementation of the Circular Economy (Singh

and Ordoñez, 2016; Kirkman and Voulvoulis, 2017), shows the European Union to be the engine driving the transition to a productive model based on circularity. This study aimed at providing an accurate image of the evolution of the concept of the Circular Economy, showing the references within the research line and the main trends today.

d) There is a strong international collaborative network in EC. On the one hand, 706 authors in this line of research are grouped in 36 clusters with at least 3 co-authorship publications. On the other hand, as many as 61 countries are grouped in 7 clusters with at least 10 collaborations in publications. The United Kingdom and France are the countries with the highest number of international collaborators (67 and 61 respectively, while Alma Mater Studiorum Università di Bologna and the Chinese Academy of Sciences had the highest rates of international collaborations.

e) The analysis of the keywords shows six main lines of research within the Circular Economy field, such as recuperative technologies, sustainability in a transversal orientation, waste management, industry 4.0, critical raw materials, and the treatment of by-products and waste. The temporal analysis shows how all the keywords are relatively recent, which is why new concepts, techniques, tools, and methodologies are emerging that are rapidly attracting the attention of the research community. The hottest lines of research at present are as follows: the development of recuperative technologies, the management of the supply chain, the economic aspects derived from the Circular Economy, and Industry 4.0. (Q5).

Finally, this work has certain research limitations that could form the basis of future research. For instance, the methodology used is quantitative in nature, and does not consider qualitative aspects of the question, but further research could use systematic literature reviews to delve deeper in the areas for aggregation of the knowledge here outlined. Finally, the use of other keywords, a different study period, research materials other than the articles considered or performance of the search in other databases could all influence the results obtained.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

VM-S and FJG-S: conceptualization and investigation. VM-S, FJG-S, VM-M, and GW-F-B: methodology and writing—original draft preparation. VM-S and FJG-S: software and data curation. VM-S, FJG-S, VM-M, and GW-F-B: validation. VM-S, FJG-S, and VM-M: formal analysis. VM-S and GW-F-B: resources. VM-S and FJG-S: writing—review and editing. VM-S, FJG-S, VM-M and GW-F-B: supervision and project administration. All authors have read and agreed to the published version of the manuscript.

## REFERENCES


Abdul Qyyum, M., Qadeer, K., and Lee, M. (2018). Closed-loop Self-Cooling Recuperative N2 Expander Cycle for the Energy Efficient and Ecological Natural Gas Liquefaction Process. *ACS Sustainable Chem. Eng.* 6 (4), 5021–5033. doi:10.1021/acssuschemeng.7b04679

Acedo, F. J., Barroso, C., Casanueva, C., and Galán, J. L. (2006). Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis*. *J. Management Stud.* 43 (5), 957–983. doi:10.1111/j.1467-6486.2006.00625.x

Ackerson, L. G., and Chapman, K. (2003). Identifying the Role of Multidisciplinary Journals in Scientific Research. *Crl* 64 (6), 468–478. doi:10.5860/crl.64.6.468

Alhaddi, H. (2015). Triple Bottom Line and Sustainability: A Literature Review. *Bms* 1 (2), 6–10. doi:10.11114/bms.v1i2.752

Andersen, M. S. (2007). An Introductory Note on the Environmental Economics of the Circular Economy. *Sustain. Sci.* 2 (1), 133–140. doi:10.1007/s11625-006-0013-6

Archambault, É., Campbell, D., Gingras, Y., and Larivière, V. (2009). Comparing Bibliometric Statistics Obtained from the Web of Science and Scopus. *J. Am. Soc. Inf. Sci.* 60 (7), 1320–1326. doi:10.1002/asi.21062

Argudo Garcia, J. J., Molina Moreno, V., and Leyva Diaz, J. C. (2017). Valorización De Lodos Procedentes De Plantas De Tratamiento De Agua Potable. Una Apuesta Por La Economía Circular Y Sostenibilidad. *Dynaii* 92 (1), 71–75. doi:10.6036/8024

Baas, J., Schotten, M., Plume, A., Côté, G., and Karimi, R. (2020). Scopus as a Curated, High-Quality Bibliometric Data Source for Academic Research in Quantitative Science Studies. *Quantitative Sci. Stud.* 1 (1), 377–386. doi:10.1162/qss_a_00019

Bag, S., Gupta, S., and Kumar, S. (2021). Industry 4.0 Adoption and 10R advance Manufacturing Capabilities for Sustainable Development. *Int. J. Prod. Econ.* 231, 107844. doi:10.1016/j.ijpe.2020.107844

Banerjee, A. V., Banerjee, A., and Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

Beaulieu, L. (2015). Circular Economy: A Critical Literature Review of Concepts. *International Reference Centre for the Life Cycle of Products, Processes and Services (CIRAIG), Polytechnique Montréal*. Canada: Canadian Electronic Library.

Belmonte-Ureña, L. J., Plaza-Úbeda, J. A., Vazquez-Brust, D., and Yakovleva, N. (2021). Circular economy, degrowth and green growth as pathways for research on sustainable development goals: A global analysis and future agenda. *Ecol. Econ.* 185, 107050. doi:10.1016/j.ecolecon.2021.107050

Benyus, J. M. (1997). *Biomimicry: Innovation Inspired by Nature*. New York: Morrow.

Berkel, R., Willems, E., and Lafleur, M. (1997). The Relationship between Cleaner Production and Industrial Ecology. *J. Ind. Ecol.* 1 (1), 51–66. doi:10.1162/jiec.1997.1.1.51

Bilitewski, B. (2012). The Circular Economy and its Risks. *Waste Manag.* 32, 1–2. doi:10.1016/j.wasman.2011.10.004

Birat, J.-P. (2015). Life-cycle Assessment, Resource Efficiency and Recycling. *Metall. Res. Technol.* 112 (2), 206. doi:10.1051/metal/2015009

Braungart, M., McDonough, W., and Bollinger, A. (2007). Cradle-to-cradle Design: Creating Healthy Emissions - a Strategy for Eco-Effective Product and System Design. *J. Clean. Prod.* 15 (13-14), 1337–1348. doi:10.1016/j.jclepro.2006.08.003

Brennan, G., Tennant, M., and Blomsma, F. (2015). in *"Business and Production Solutions: Closing the Loop" in Sustainability: Key Issues*. Editors H. Kopnina and Shoreman-Ouimet. (London: Routledge).

Castellani, V., Sala, S., and Mirabella, N. (2015). Beyond the Throwaway Society: A Life Cycle-Based Assessment of the Environmental Benefit of Reuse. *Integr. Environ. Assess. Manag.* 11 (3), 373–382. doi:10.1002/ieam.1614

Charonis, G. K. (2012). Degrowth, Steady State Economics and the Circular Economy: Three Distinct yet Increasingly Converging Alternative Discourses to Economic Growth for Achieving Environmental Sustainability and Social Equity. World Economic Association Sustainability Conference, 24th September to 21st October, 2012.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Infor. Sci.* 57 (3), 359–377. doi:10.1002/asi.20317

Club of Rome (2015). The Circular Economy and Benefits for Society, Swedish Case Study Shows Jobs and Climate as Clear Winners. Available at: http://www.clubofrome.org/cms/wp-content/uploads/2015/04/Final-version-Swedish-Study-13-04-15-till-tryck-ny.pdf. (Accesed February 25, 2021).

Cobbledick, J., Aubry, N., Zhang, V., Rollings-Scattergood, S., and Latulippe, D. R. (2016). Lab-scale Demonstration of Recuperative Thickening Technology for Enhanced Biogas Production and Dewaterability in Anaerobic Digestion Processes. *Water Res.* 95, 39–47. doi:10.1016/j.watres.2016.02.051

Comerio, N., and Strozzi, F. (2019). Tourism and its economic impact: A literature review using bibliometric tools. *Tourism Econ.* 25 (1), 109–131. doi:10.1177/1354816618793762

Commoner, B. (2020). *The Closing circle: Nature, Man, and Technology*. Mineola: Courier Dover Publications.

Cop21 (2015). Paris Agreement. Available at: https://ec.europa.eu/clima/policies/international/negotiations/paris_es. (Accesed February 18, 2021).

Cronin, B. (2001). Bibliometrics and beyond: Some Thoughts on Web-Based Citation Analysis. *J. Inf. Sci.* 27 (1), 1–7. doi:10.1177/016555150102700101

Dajian, Z. (2004). Towards a Closed-Loop Materials Economy. *Chin. J. Popul. Resour. Environ.* 2 (1), 9–12. doi:10.1080/10042857.2004.10677342

Desrochers, P. (2004). Industrial Symbiosis: the Case for Market Coordination. *J. Clean. Prod.* 12 (8-10), 1099–1110. doi:10.1016/j.jclepro.2004.02.008

Desrochers, P. (2002). Regional Development and Inter-industry Recycling Linkages: Some Historical Perspectives. *Entrepreneurship Reg. Development* 14 (1), 49–65. doi:10.1080/08985620110096627

Ding, Y., Chowdhury, G. G., and Foo, S. (2001). Bibliometric Cartography of Information Retrieval Research by Using Co-word Analysis. *Inf. Process. Management* 37 (6), 817–842. doi:10.1016/S0306-4573(00)00051-0

Donthu, N., KumarMukherjeePandey, S. D. N., Mukherjee, D., Pandey, N., and Lim, W. M. (2021). How to Conduct a Bibliometric Analysis: An Overview and Guidelines. *J. Business Res.* 133, 285–296. doi:10.1016/j.jbusres.2021.04.070

Durieux, V., and Gevenois, P. A. (2010). Bibliometric Indicators: Quality Measurements of Scientific Publication. *Radiology* 255 (2), 342–351. doi:10.1148/radiol.09090626

Dyllick, T., and Muff, K. (2016)., 29:2. Clarifying the Meaning of Sustainable Business. *Organ. Environ*. Organization & Environment, 156–174. doi:10.1177/1086026615575176

Ellen Macarthur Foundation (2012). Towards the Circular Economy. Availableat: http://www.ellenmacarthurfoundation.org/business/reports.Accesed January 15, 2021).

European Commission (2015). *Closing the Loop - an EU Action Plan for the Circular Economy*. Brussels: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions European Commission.

European Commission (2014b). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a Circular Economy: A Zero Waste Programme for Europe*.

European Commission (2014a). MEMO, Questions and Answers on the Commission Communication "Towards a Circular Economy" and the Waste Targets Review. Available at: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_14_450. (Accesed February 22, 2021).

Fang, Y., Yin, J., and Wu, B. (2018). Climate Change and Tourism: A Scientometric Analysis Using CiteSpace. *J. Sustainable Tourism* 26 (1), 108–126. doi:10.1080/09669582.2017.1329310

Feng, Z., and Yang, N. (2007). Putting a Circular Economy into Practice in China. *Sustainability Sci.* 2, 95–101.

Fischer, A., and Pascucci, S. (2017). Institutional Incentives in Circular Economy Transition: The Case of Material Use in the Dutch Textile Industry. *J. Clean. Prod.* 155, 17–32. doi:10.1016/j.jclepro.2016.12.038

Frosch, R. A., and Gallopoulos, N. E. (1989). Strategies for Manufacturing. *Sci. Am.* 261 (3), 144–152. doi:10.1038/scientificamerican0989-144

Fundación Ellen MacArthur (Emf) (2013). Circular Economy. Available at: https://www.ellenmacarthurfoundation.org/es/economia-circular/concepto. (Accesed February 21, 2021).

Gálvez-Sánchez, F. J., Lara-Rubio, J., Verdú-Jóver, A. J., and Meseguer-Sánchez, V. (2021). Research Advances on Financial Inclusion: A Bibliometric Analysis. *Sustainability* 13, 3156. doi:10.3390/su13063156

Gates, B. (2021). *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs*, New York: We Need. Knopf.

Geng, Y., Fu, J., Sarkis, J., and Xue, B. (2012). Towards a National Circular Economy Indicator System in China: an Evaluation and Critical Analysis. *J. Clean. Prod.* 23 (1), 216–224. doi:10.1016/j.jclepro.2011.07.005

Genovese, A., Acquaye, A. A., Figueroa, A., and Koh, S. C. L. (2017). Sustainable Supply Chain Management and the Transition towards a Circular Economy: Evidence and Some Applications. *Omega* 66, 344–357. doi:10.1016/j.omega.2015.05.015

Ghisellini, P., Cialani, C., and Ulgiati, S. (2016). A Review on Circular Economy: the Expected Transition to a Balanced Interplay of Environmental and Economic Systems. *J. Clean. Prod.* 114, 11–32. doi:10.1016/j.jclepro.2015.09.007

Giampietro, M. (2019). On the Circular Bioeconomy and Decoupling: Implications for Sustainable Growth. *Ecol. Econ.* 162, 143–156. doi:10.1016/j.ecolecon.2019.05.001

Gimenez, C., Sierra, V., and Rodon, J. (2012). Sustainable Operations: Their Impact on the Triple Bottom Line. *Int. J. Prod. Econ.* 140 (1), 149–159. doi:10.1016/j.ijpe.2012.01.035

Goyal, S., Esposito, M., and Kapoor, A. (2018). Circular Economy Business Models in Developing Economies: Lessons from India on Reduce, Recycle, and Reuse Paradigms. *Thunderbird Int. Bus Rev.* 60 (5), 729–740. doi:10.1002/tie.21883

Guerrero, L. A., Maas, G., and Hogland, W. (2013). Solid Waste Management Challenges for Cities in Developing Countries. *Waste Management* 33 (1), 220–232. doi:10.1016/j.wasman.2012.09.008

Gupta, S., Chen, H., Hazen, B. T., Kaur, S., and Santibañez Gonzalez, E. D. R. (2019). Circular Economy and Big Data Analytics: A Stakeholder Perspective. *Technol. Forecast. Soc. Change* 144, 466–474. doi:10.1016/j.techfore.2018.06.030

Gurita, N., Fröhling, M., and Bongaerts, J. (2018). Assessing Potentials for mobile/smartphone Reuse/remanufacture and Recycling in Germany for a Closed Loop of Secondary Precious and Critical Metals. *Jnl Remanufactur* 8 (1), 1–22. doi:10.1007/s13243-018-0042-1

Harzing, A.-W., and Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a Longitudinal and Cross-Disciplinary Comparison. *Scientometrics* 106 (2), 787–804. doi:10.1007/s11192-015-1798-9

Hofmann, M., Hofmann, H., Hagelüken, C., and Hool, A. (2018). Critical Raw Materials: A Perspective from the Materials Science Community. *Sustainable Mater. Tech.* 17, e00074. doi:10.1016/j.susmat.2018.e00074

Jabbour, C. J. C., Seuring, S., de Sousa Jabbour, A. B. L., Jugend, D., Fiorini, P. D. C., Latan, H., et al. (2020). Stakeholders, Innovative Business Models for the Circular Economy and Sustainable Performance of Firms in an Emerging Economy Facing Institutional Voids. *J. Environ. Manage.* 264, 110416. doi:10.1016/j.jenvman.2020.110416

Kansha, Y., Kishimoto, A., and Tsutsumi, A. (2012). Application of the Self-Heat Recuperation Technology to Crude Oil Distillation. *Appl. Therm. Eng.* 43, 153–157. doi:10.1016/j.applthermaleng.2011.10.022

Kirkman, R., and Voulvoulis, N. (2017). The Role of Public Communication in Decision Making for Waste Management Infrastructure. *J. Environ. Manage.* 203, 640–647. doi:10.1016/j.jenvman.2016.06.002

Korhonen, J., Honkasalo, A., and Seppälä, J. (2018). Circular Economy: the Concept and its Limitations. *Ecol. Econ.* 143, 37–46. doi:10.1016/j.ecolecon.2017.06.041

Kuchonthara, P., Bhattacharya, S., and Tsutsumi, A. (2005). Combination of Thermochemical Recuperative Coal Gasification Cycle and Fuel Cell for Power Generation. *Fuel* 84 (7-8), 1019–1021. doi:10.1016/j.fuel.2004.08.024

Kumar, S., Lim, W. M., Pandey, N., and Christopher Westland, J. (2021). 20 Years of Electronic Commerce Research. *Electron. Commer Res.* 21 (1), 1–40. doi:10.1007/s10660-021-09464-1

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. (2014). Industry 4.0. *Bus Inf. Syst. Eng.* 6 (4), 239–242. doi:10.1007/s12599-014-0334-4

Lei, S., and Yi, Q. (2004). Strategy and Mechanism Study for Promotion of Circular Economy in China. *Chin. J. Popul. Resour. Environ.* 2 (1), 5–8. doi:10.1080/10042857.2004.10677341

Lett, L. A. (2014). Las amenazas globales, el reciclaje de residuos y el concepto de economía circular. *Revista Argentina de Microbiología* 46 (1), 1–2. doi:10.1016/s0325-7541(14)70039-2

Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., and Zeng, X.-J. (2018). A Bibliometric Analysis and Visualization of Medical Big Data Research. *Sustainability* 10:1, 166. doi:10.3390/su10010166

Lieder, M., and Rashid, A. (2016). Towards Circular Economy Implementation: a Comprehensive Review in Context of Manufacturing Industry. *J. Clean. Prod.* 115, 36–51. doi:10.1016/j.jclepro.2015.12.042

Lievrouw, L. A. (1989). The Invisible College Reconsidered. *Commun. Res.* 16 (5), 615–628. doi:10.1177/009365089016005004

Lyle, J. T. (1996). *Regenerative Design for Sustainable Development*. New Jersey: John Wiley & Sons.

Markard, J., Raven, R., and Truffer, B. (2012). Sustainability Transitions: An Emerging Field of Research and its Prospects. *Res. Pol.* 41 (6), 955–967. doi:10.1016/j.respol.2012.02.013

Martínez-Medina, A., Gálvez-Sánchez, F. J., and Silva-Santos, B. (2020). "Economía Circular Inclusiva: Abono Café," in *Economía Circular: Fundamentos Y Aplicaciones. (Navarra: Aranzadi)*. Editors V. Meseguer-Sánchez, V. Molina-Moreno, and P. Núñez-Cacho-Utrilla, 309–322.

McDonough, W., and Braungart, M. (2010). *Cradle to Cradle: Remaking the Way We Make Things*. New York: North Point Press.

McWilliams, A., Parhankangas, A., Coupet, J., Welch, E., and Barnum, D. T. (2016). Strategic Decision Making for the Triple Bottom Line. *Bus. Strat. Env.* 25 (3), 193–204. doi:10.1002/bse.1867

Meseguer-Sánchez, V., Abad-Segura, E., Belmonte-Ureña, L. J., and Molina-Moreno, V. (2020). Examining the Research Evolution on the Socio-Economic and Environmental Dimensions on university Social Responsibility. *Ijerph* 17, 4729. doi:10.3390/ijerph17134729

Meseguer-Sánchez, V., Gálvez-Sánchez, F. J., López-Martínez, G., and Molina-Moreno, V. (2021). Corporate Social Responsibility and Sustainability. A Bibliometric Analysis of Their Interrelations. *Sustainability* 13, 1636. doi:10.3390/su13041636

Mingers, J., and Lipitakis, E. A. E. C. G. (2010). Counting the Citations: a Comparison of Web of Science and Google Scholar in the Field of Business and Management. *Scientometrics* 85 (2), 613–625. doi:10.1007/s11192-010-0270-0

Molina-Moreno, V., Núñez-Cacho, P., and Gálvez-Sánchez, F. J. (2019). Transición hacia la economía circular y sostenibilidad de la industria de defensa. Estudio de los casos de Navantia y Airbus Military. *Economía Ind.* 412, 149–156.

Mongeon, P., and Paul-Hus, A. (2016). The Journal Coverage of Web of Science and Scopus: a Comparative Analysis. *Scientometrics* 106, 213–228. doi:10.1007/s11192-015-1765-5

Murray, A., Skene, K., and Haynes, K. (2017). The Circular Economy: an Interdisciplinary Exploration of the Concept and Application in a Global Context. *J. Bus Ethics* 140 (3), 369–380. doi:10.1007/s10551-015-2693-2

Naustdalslid, J. (2014). Circular Economy in China - the Environmental Dimension of the Harmonious Society. *Int. J. Sustainable Development World Ecol.* 21 (4), 303–313. doi:10.1080/13504509.2014.914599

Ness, D. (2008). Sustainable Urban Infrastructure in China: Towards a Factor 10 Improvement in Resource Productivity through Integrated Infrastructure Systems. *Int. J. Sustainable Development World Ecol.* 15 (4), 288–301. doi:10.3843/SusDev.15.4:2a

Nikpey, H., Assadi, M., Breuhaus, P., and Mørkved, P. T. (2014). Experimental Evaluation and ANN Modeling of a Recuperative Micro Gas Turbine Burning Mixtures of Natural Gas and Biogas. *Appl. Energ.* 117, 30–41. doi:10.1016/j.apenergy.2013.11.074

Park, J. Y., and Chertow, M. R. (2014). Establishing and Testing the "Reuse Potential" Indicator for Managing Wastes as Resources. *J. Environ. Manage.* 137, 45–53. doi:10.1016/j.jenvman.2013.11.053

Park, J. Y., and Nagy, Z. (2018). Data on the Interaction between thermal comfort and Building Control Research. *Data in Brief* 17, 529–532. doi:10.1016/j.dib.2018.01.033

Parlanti, R. (2017). Smart Shopfloors and Connected Platforms in Industry 4.0. Electron. *World* 123, 26–28.

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., and Bresciani, S. (2021). Scientific Procedures and Rationales for Systematic Literature Reviews (SPAR-4-SLR). *Int. J. Consum Stud.* 45. doi:10.1111/ijcs.12695

Pauli, G. A. (2010). *The Blue Economy: 10 years, 100 Innovations, 100 Million Jobs*. New Mexico: Paradigm Publications.

Pearce, D. W., and Turner, R. K. (1990). *Economics of Natural Resources and the Environment*. Pennsylvania: JHU Press.

Prendeville, S., Sanders, C., Sherry, J., and Costa, F. (2014). Circular Economy: Is it Enough? Available at: http://www.edcw.org/sites/default/files/resources/Circular%20Ecomomy-%20Is%20it%20enough.pdf. (Accesed February 25, 2021).

Raworth, K. (2017). *Doughnut Economics: Seven Ways to Think like a 21st-century Economist*. London: Chelsea Green Publishing.

Reischauer, G. (2018). Industry 4.0 as Policy-Driven Discourse to Institutionalize Innovation Systems in Manufacturing. *Technol. Forecast. Soc. Change* 132, 26–33. doi:10.1016/j.techfore.2018.02.012

Ren, J., Manzardo, A., Toniolo, S., and Scipioni, A. (2013). Sustainability of Hydrogen Supply Chain. Part I: Identification of Critical Criteria and Cause-Effect Analysis for Enhancing the Sustainability Using DEMATEL. *Int. J. Hydrogen Energ.* 38 (33), 14159–14171. doi:10.1016/j.ijhydene.2013.08.126

Sedighi, M. (2016). Application of Word Co-occurrence Analysis Method in Mapping of the Scientific fields (Case Study: the Field of Informetrics). *Libr. Rev.* 65 (13), 52–64. doi:10.1108/LR-07-2015-0075

Sehnem, S., Chiappetta Jabbour, C. J., Farias Pereira, S. C., and de Sousa Jabbour, A. B. L. (2019). Improving Sustainable Supply Chains Performance through Operational Excellence: Circular Economy Approach. *Resour. Conservation Recycling* 149, 236–248. doi:10.1016/j.resconrec.2019.05.021

Seiffert, M. E. B., and Loch, C. (2005). Systemic Thinking in Environmental Management: Support for Sustainable Development. *J. Clean. Prod.* 13 (12), 1197–1202. doi:10.1016/j.jclepro.2004.07.004

Sen, A. (2001). *Development as freedom*. Oxford: Oxford Paperbacks.

Sharholy, M., Ahmad, K., Mahmood, G., and Trivedi, R. C. (2008). Municipal Solid Waste Management in Indian Cities - A Review. *Waste Management* 28 (2), 459–467. doi:10.1016/j.wasman.2007.02.008

Singh, J., and Ordoñez, I. (2016). Resource Recovery from post-consumer Waste: Important Lessons for the Upcoming Circular Economy. *J. Clean. Prod.* 134, 342–353. doi:10.1016/j.jclepro.2015.12.020

Smol, M., Kulczycka, J., Henclik, A., Gorazda, K., and Wzorek, Z. (2015). The Possible Use of Sewage Sludge Ash (SSA) in the Construction Industry as a Way towards a Circular Economy. *J. Clean. Prod.* 95, 45–54. doi:10.1016/j.jclepro.2015.02.051

Stahel, W. R., and Reday, G. (1976). *The Potential for Substituting Manpower for Energy*. Brussels: Report to the Commission of the European Communities.

Stahel, W. R. (1982). *The Product Life Factor. An Inquiry into the Nature of Sustainable Societies: The Role of the Private Sector*. Texas: Houston Area Research Center, 72–105.

Stahel, W. (2010). *The Performance Economy*. Berlin: Springer. doi:10.1057/9780230288843

Stiehl, C., and Hirth, T. (2012). Vom additiven Umweltschutz zur nachhaltigen Produktion. *Chem. Ingenieur Technik* 84 (84), 963–968. doi:10.1002/cite.201200008

Strebel, H., and Posch, A. (2004). Interorganisational Cooperation for Sustainable Management in Industry: on Industrial Recycling Networks and Sustainability Networks. *Pie* 1 (4), 348–362. doi:10.1504/PIE.2004.005841

Su, B., Heshmati, A., Geng, Y., and Yu, X. (2013). A Review of the Circular Economy in China: Moving from Rhetoric to Implementation. *J. Clean. Prod.* 42, 215–227. doi:10.1016/j.jclepro.2012.11.020

Tukker, A. (2015). Product Services for a Resource-Efficient and Circular Economy - a Review. *J. Clean. Prod.* 97, 76–91. doi:10.1016/j.jclepro.2013.11.049

Van Berkel, R. (2007). Cleaner Production and Eco-Efficiency Initiatives in Western Australia 1996-2004. *J. Clean. Prod.* 15 (8-9), 741–755. doi:10.1016/j.jclepro.2006.06.012

Veer, D. K., and Khiste Gajanan, P. (2017). Digital Library Output in Scopus during 1995-2016: A Bibliometric Analysis. *Int. J. Scientific Res. Computer Sci. Eng. Inf. Technology* 2 (5), 779–784.

Wijkman, A., and Skånberg, K. (2015). *The Circular Economy and Benefits for Society*. Switzerland: Club of Rome.

Xu, Y. J., Zhang, T. Z., Shi, L., and Chen, J. N. (2004). Material Flow Analysis in Guiyang. *Journal of Tsinghua*, 12.

Yang, S., Phan, H. V., Bustamante, H., Guo, W., Ngo, H. H., and Nghiem, L. D. (2017). Effects of Shearing on Biogas Production and Microbial Community Structure during Anaerobic Digestion with Recuperative Thickening. *Bioresour. Technology* 234, 439–447. doi:10.1016/j.biortech.2017.03.051

Yap, N. T. (2005). Towards a Circular Economy: Progress and Challenges. *Greener Management Int.* 50, 11–24.

Zhang, L., and Eichmann-Kalwara, N. (2019). Mapping the Scholarly Literature Found in Scopus on "Research Data Management": A Bibliometric and Data Visualization Approach. *J. Librarianship Scholarly Commun.* 7. doi:10.7710/2162-3309.2266

Zhu, W., and Guan, J. (2013). A Bibliometric Study of Service Innovation Research: Based on Complex Network Analysis. *Scientometrics* 94 (3), 1195–1216. doi:10.1007/s11192-012-0888-1

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Meseguer-Sánchez, Gálvez-Sánchez, Molina-Moreno and Wandosell-Fernández-de-Bobadilla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 05 August 2021
doi: 10.3389/fenvs.2021.721408

# The Spatial Network Contagion of Environmental Risks Among Countries Along the Belt and Road Initiative

*Ling Lu*[1,2]*, *Kai Fang*[3], *Chuan Ming Liu*[4] *and Cheng Sun*[1,2]

[1]School of Low Carbon Economics, Hubei University of Economics, Wuhan, China, [2]Center of Hubei Cooperative Innovation for Emissions Trading System, Hubei University of Economics, Wuhan, China, [3]School of Public Affairs, Zhejiang University, Hangzhou, China, [4]School of Economics, Shandong University of Finance and Economics, Jinan, China

Under the tide of global economic integration with aggravated environmental risks caused by intensive human activities, the spatial network correlation of environmental risks has become intensified. The close exchange of activities among the countries under the Belt and Road Initiative (BRI) will especially feel this contagion of environmental risks. Given this situation, this article analyzes the environmental risk contagion relations among the BRI countries and the characteristics of their network structure by using social network analysis (SNA). A block model is used to analyze the spatial clustering characteristics of the environmental risk contagion. Specifically, the driving factors of environmental risk contagion are analyzed through the quadratic assignment procedure (QAP) of SNA. The results of this article provide some references for the BRI to reduce the environmental risks and jointly control the environmental risk contagion, so as to assist in the promotion of a green silk road.

**Keywords: green silk road, environmental risk contagion, SNA, QAP, the gravity model, climate investment, energy transformation**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, China*

***Reviewed by:***
*Ya Zhou,*
*Guangdong University of Technology, China*
*Xianchun Liao,*
*University of Jinan, China*

***Correspondence:**
*Ling Lu*
*lynnlu699@126.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *07 June 2021*
***Accepted:*** *07 July 2021*
***Published:*** *05 August 2021*

***Citation:***
*Lu L, Fang K, Liu CM and Sun C (2021) The Spatial Network Contagion of Environmental Risks Among Countries Along the Belt and Road Initiative. Front. Environ. Sci. 9:721408. doi: 10.3389/fenvs.2021.721408*

## INTRODUCTION

In September and October of 2013, China's President Xi Jinping put forward the cooperation initiatives, the "New Silk Road Economic Belt" and the "21st Century Maritime Silk Road" (known together as "One Belt and One Road"), intending to seek common development and win-win cooperation with China's neighboring countries. The Belt and Road Initiative (BRI) countries have rich mineral resources but complex geographical conditions and fragile ecological environments. They are mainly developing countries still in the stage of promoting economic growth through large-scale resource consumption and concomitant pollution emissions, and faced with increasingly severe ecological and environmental problems. The environmental risk of the BRI is the primary concern due to the presence of many ecologically sensitive areas within the regions. These locations with higher ecological sensitivity include arid regions, biodiversity hotspots, and protected areas (Tracy et al., 2017).

The Belt and Road Initiative will greatly influence the future of global trade. However, it may also promote permanent environmental degradation (Ascensão et al., 2018). As globalization is typically accompanied by the transfer of industries with high carbon emissions from developed countries to emerging economies, many of the regions along the BRI are already experiencing resource and environmental problems along with their economic growth (Dong et al., 2015; Han et al., 2018).

However, disturbances to the BRI regional environmental risk could rise dramatically because of these intense human activities and productive activities. Moreover, environmental risks generated by implementing these projects may result in wastes of human and material resources, and misunderstanding and conflicts between outsiders and locals (Xiao, 2019; Yang and Li, 2019). To help address this problem, Xi Jinping called for a "green, healthy, intelligent, and peaceful" Silk Road (Ma, 2016). He suggested that the participating countries "deepen cooperation in environmental protection, intensify ecological preservation, and build a green Silk Road" (MOFCOM, 2017). While strengthening its ecological progress, China is committed to building a "green silk road" with its BRI copartners. Therefore, a comprehensive analysis of the spatial network contagion of the environmental risks in the BRI is of great significance for countries aiming to coordinate the control of environmental risk levels, reduce the environmental risk contagion, and build a green silk road.

## LITERATURE REVIEW

Environmental risk is the probability and consequences of unfortunate events caused by spontaneous natural events and human activities spread through environmental media. It can have destructive effects on human society and the natural environment (Zhong et al., 1996; Torinelli and Silva Júnior, 2021). The dimensions of the environmental physical risks are climatic, geologic, and ecosystemic (Leadership, 2017).

Mainly, studies on environmental risk always focus on the assessments of environmental risk. Greenpeace (2017) advocated that environmental impact assessments should be integrated into the BRI and that the results of these assessments be made publicly available. Torinelli and Silva Júnior (2021) used the environmental physical risks and associated transition risks to construct an environmental risk measurement index to assess the impact of environmental risks on financial assets and investments. Environment-related risks can cause "stranded assets," or ones impacted by unanticipated or premature write-downs, devaluations, or conversion to liabilities (Caldecott et al., 2014). Huang (2019) took three dimensions, including environmental quality, climate change, and ecological vulnerability, into account when assessing environmental risks, and analyzed the environmental opportunities of the BRI.

The level of environmental risk in a country may not only be determined by local environmental conditions but also by affected other regions. On the one hand, increased environmental risk in a country will affect the local environmental policy standard. If environmental regulation standards are too high, companies will locate their production activities in countries with lax environmental regulation and high emission intensity due to the "pollution paradise effect" (Copeland and Taylor, 2003), thereby affecting the environmental risks of other countries.

On the other hand, as an important link among macroeconomic activities, trade plays an irreplaceable guiding role in producing and consuming goods and services (Hu et al., 2018). In the global value chain, each country and department faces a potential environmental cost for each unit of GDP. The environmental cost will be transferred among different countries and departments through trade, which is closely related to the degree, mode, and position of a country's participation in the value chain (Meng et al., 2016). As more than 95% of the world's net carbon sequestration occurs in regions along the Belt and Road (Yao et al., 2018), the BRI bears considerable carbon emission pressures. International trade changes a country's technological progress and carbon productivity (Li et al., 2016), thereby affecting each country's environmental risk changes. Global trade can lead to a "carbon leak" of greenhouse gas emissions, with rich countries relocating their emission-intensive production activities to poorer countries (Kuik and Hofkes, 2010; Li et al., 2020). Hence, environmental risks are reduced in rich countries and increased in poor countries.

International trade also affects global biodiversity. Consumers in developed countries threaten some plant and animal species through their demand for commodities that are mainly produced in developing countries (Lenzen et al., 2012). Some environmental risks caused by consumers in developed countries have been transferred to developing countries through trade. Moreover, some consumers derive benefits from various services provided by nature, referred to as "natural capital," and cause plant and animal extinctions. The loss of biodiversity poses significant risks to the global economy and society (Dasgupta and McKenzie, 2020). The environmental risk caused by such biodiversity destruction increases the contagiousness of environmental risk among countries through economic and social systems.

As there are multiple activities taking place in the BRI countries under the ambit of common development, the contagion of environmental risks may be significant. With such concern, this article mainly targets on the countries along the BRI, to study the spatial network contagion of environmental risks among the countries.

The contributions of this article are as follows: first, the research object of this article is the environmental risk contagion relationship. Environmental risks are not only determined by local environmental risks but also be affected by environmental risks in other regions. Second, this article employs the social network analysis method to study the environmental risk contagion of the BRI. This method is mainly applied to the research of attribute data relations, which can not only examine the spatial relations of neighboring regions but also present the spatial relations between distant regions.

## METHODOLOGY AND DATA

### The Social Network Analysis Model

In this article, social network analysis (SNA) is used to investigate the spatial correlation network contagion of environmental risk among the BRI countries. SNA is an interdisciplinary analysis method for relational data which has been widely used in many

fields (Wasserman and Faust, 1994; Benítez-Andrades, et al., 2020). It takes the relationship between individuals in the structure as the analysis object and describes the relationship pattern among individuals. Owing to the advantages of iconic expression and precise calculation, SNA has been extensively adopted in an army of disciplines, including economics, sociology, management, geography, and tourism (Scott and Carrington, 2011). This method was chosen for two main reasons. First, the network analysis method has the characteristics of global analysis, which can avoid the limitation of the spatial econometric analysis method, that can only analyze "adjacent" and "close" areas (Wolfe, 1995; Borgatti et al., 2009). Second, the network analysis method mainly analyzes the structural relationship data, which often determines the attribute data and has more analytical value (Yin et al., 2020). Therefore, this article uses the social network analysis method to analyze the environmental risk contagion among the BRI countries. Reilly (1929) first introduced the gravity model to demographic geography, and it has since been incorporated into social science studies (Cantore and Cheng, 2018; Yu and Di, 2020). The modified gravity model was based on the law of universal gravitation (Wu et al., 2016) that was used to determine the spatial correlation of SNA (Liu and Xiao, 2021).

Environmental risk–related influenced factors include air quality, natural resources, and biodiversity (Tracy et al., 2017; Ascensão et al., 2018). Air pollutants dominated by particulate matter with a diameter <2.5 μm ($PM_{2.5}$) cause economic losses equal to 1% of the world's GDP (Sheehan et al., 2014; Silver et al., 2018). Natural resource depletion is a concern for sustainable development as it has the ability to degrade current environments and the potential to impact the needs of future generations (Salvati and Marco, 2008). Threatened species is one of the indicators of ecological vulnerability that reflects the degree of ecological fragility and sensitivity, as well as the degree to which regional ecological and environmental damage may occur (Huang, 2019). In gravity, the attraction between two objects depends on their masses and distance. Environmental footprints, including water, land, carbon, nitrogen, and phosphorus footprints, show strong spatial heterogeneity within the BRI (Fang et al., 2021). Therefore, the contagion paths should be affected by the distance between the nodes in the spatial network. Based on the above considerations, this article modifies the gravity model by taking into account the following key environmental risk contagion indicators, including distance between countries, threatened species, natural resource loss, and particulate matter emission damage. The formula is as follows:

$$R_{ij} = k_{ij} \times \frac{\sqrt[3]{P \times N \times S}\sqrt[3]{P \times N \times S}}{D^2}, \quad (1)$$

$$K_{ij} = \frac{GDP_i}{GDP_i + GDP_j},$$

where $R_{ij}$ refers to the gravitational force between the environmental risks of country $i$ and country $j$, $P$ is the particulate matter emission damage, $N$ is the natural resource loss, and $S$ is the threatened species; $k_{ij}$ is the adjustment coefficient and $GDP$ is taken as the adjustment coefficient of environmental risk. $D$ is the geographical distance between country $i$ and country $j$. The gravity matrix is calculated according to the formula, and the average value of each row of the gravity matrix is taken as the critical value. $R_{ij}$ represents the spatial contagion relationship among the BRI countries. If the gravity value is higher than the average value, the environmental risk contagion effect between the countries exists; otherwise, there is no environmental risk contagion effect.

## Spatial Correlation Index

### Overall Network Characteristics

Density indicates the closeness of the spatial correlation of environmental risks among the countries in the network. Network density is applied to evaluate the level of closeness among different nodes in a network (Bai et al., 2020). The value of this measure ranges from 0 to 1: the higher the density, the closer the contagion of environmental risks among the BRI countries. Network density can be defined as the ratio of the number of relationships owned by the node to the maximum possible relationships in the entire network (Scott and Carrington, 2011). Assume that the number of countries in the network is N and the relationship number of environmental risk contagion is $L$, then the calculation formula of network density ($Den$) is expressed as follows:

$$Den = \frac{L}{N \times (N-1)}. \quad (2)$$

Connectedness represents the robustness and vulnerability of the spatial association network of environmental risks among countries, and the value range is 0–1. If there is a direct or indirect path between any two countries in the environmental risk contagion network that connects the environmental risks as a whole, the network has intense correlation. Let the number of nodes in the network be N and the logarithm of unreachable points in the network be V, then the calculation formula of connectedness ($CD$) is expressed as follows:

$$CD = 1 - \frac{V}{N \times (N-1) \div 2}. \quad (3)$$

The hierarchy represents the extent to which countries are asymmetrically accessible and whether there is a rigid hierarchical structure in the network, indicating the dominant position of countries in the network. Let the symmetrically reachable point logarithm in the network be $K$ and the maximum possible symmetrically reachable point logarithm be max ($K$), then the hierarchy ($H$) is expressed as follows:

$$H = 1 - \frac{K}{\max(K)}. \quad (4)$$

Efficiency refers to the extent that redundant connections exist in the spatial network contagion of environmental risks among countries. Moreover, the greater the efficiency, the more intensive the network structure's stability. Let the number of redundant lines in the network be $M$ and the maximum possible number of redundant lines be max ($M$), then the efficiency ($E$) is expressed as follows:

$$E = 1 - \frac{M}{\max(M)}. \quad (5)$$

### Individual Centrality

In addition to describing the overall network structure characteristics, social network analysis can also analyze the network structure characteristics of each node through individual centrality so as to describe the status and role of each country in the network (Freeman, 2004; Liu et al., 2015). Degree centrality measures the degree to which a node is in the center of the network according to the number of connections in the network. The country with a higher degree is in the central position of the network. The degree is also divided into point out degree (Out) and point in degree (In). The point out degree of country *i* represents the number of special environmental risk that is transferred from one country to others; otherwise, it is the point in degree. Degree centrality (*DC*) is expressed as follows:

$$DC = \frac{(In_i + Out_i)}{2(N-1)}. \quad (6)$$

Closeness indicates the degree to which other countries do not control a node. The higher closeness indicates that the country belongs to the central actor. The shortcut distance between countries is denoted as $d_{ij}$, and closeness *C* is denoted as follows:

$$C = \frac{\sum_{j=1}^{N} d_{ij}}{N-1}. \quad (7)$$

Betweeness is used to measure the degree where a country controls the contagion of environmental risk. If a region is on the shortest path of multi-pair regions, it is likely to play an important role of "intermediary" or "bridge" in the network. Betweeness *B* is calculated by the following formula:

$$B = \frac{2\sum_{j}^{N}\sum_{k}^{N} g_{jk}(i)}{3N^2 - 3N + 2}, \quad (8)$$

where $g_{jk}(i)$ denotes the number of the shortest associated paths through node *i* between country *j* and country *k*, and $g_{jk}(i)$ denotes the number of all shortest associated paths between countries *k* and *j*, $k \neq j \neq i$, $j < k$.

## Block Model

Block model analysis means dividing all of the nodes into different modules and investigating the role and status of different modules in the environmental risk contagion. In this article, the BRI countries are divided into four types by the block model: main beneficiary, two-way spillover, broker, and net spillover.

## The Quadratic Assignment Procedure Method

There is a premise in the traditional multiple regression model; that is, the explained variables are independent, and there can be no multicollinearity. Otherwise, serious errors will occur in the regression. Therefore, the "relation" variables in this article are highly correlated and do not fit the traditional multiple linear regression. QAP (quadratic assignment procedure) analysis is a nonparametric test method based on random permutation. By comparing the differences between two matrices, the relation number between matrices is determined.

According to the gravity model of environmental risk mentioned constructed above, the most direct influencing factor of the environmental risk contagion between countries includes geographical distance, loss of natural resources, damage from particulate matter emissions, and threatened species. These indicators are presented in **Table 1**.

Climate investment is used mainly for national mitigation and adaptation projects to address environmental vulnerability caused by climate change, which will affect the contagious relationship of environmental risks between countries to a certain extent. Energy transition difference represented by differences in installed capacity of renewable energy indicates that the degree of energy transformation and upgrading in different countries may affect the environmental risk level and the contagion effect. The model is established as follows:

$$R = f\left(D, N', P', S', F', E'\right). \quad (9)$$

## The Data Source

The BRI countries were taken as the research objects. All data were extracted from the World Bank database. We selected 44 countries from 65 commonly accepted BRI countries due to lack of data in some countries. The selected 44 countries are presented in **Table 2**. The distance between the countries (*D*) was represented by spherical distance, which was calculated by ArcGIS. The World Bank estimated the damage from particulate emissions (*P*) and the depletion of natural resources (dollar as the unit) based on sources and methods described in the World Bank's *The Changing Wealth of Nations* (2011). Natural resource depletion (*N*) is the sum of net forest depletion, energy depletion, and mineral depletion. Energy depletion is highly related to fossil fuel energy consumption, which results in climate change risk. Net forest depletion is unit resource rent times the excess of roundwood harvest over natural growth. Energy depletion is the ratio of the value of the stock of energy resources to the remaining reserve lifetime; it covers coal, crude oil, and natural gas. Mineral depletion is the ratio of the value of the stock of mineral resources to the remaining reserve lifetime; it covers tin, gold, lead, zinc, iron, copper, nickel, silver, bauxite, and phosphate. Threatened species (*S*) is the number of species classified by the International Union for Conservation of Nature (IUCN) as endangered, vulnerable, rare, indeterminate, out of danger, or insufficiently known. The data of threatened species are only available in 2018. In order to ensure the data neatly in the paper, all data of variables are selected in the same year.

The difference in percentage of natural resource loss in GNI ($N'$), the difference in particulate emission damage as a percentage of GNI($P'$), and the difference in threatened species ($S'$) are all from World Bank database with matrix operations. In addition, it includes the climate investment difference ($F'$), which is represented by the Asian

**TABLE 1 |** Classification, variable, indicator, and reference.

| Classification | Variable | Indicator | References |
|---|---|---|---|
| Dependent variable | *R* | Environmental risk contagion | Mudakkar et al. (2013) |
| Independent variable | *D* | Geographical distance | Fang et al. (2021) |
| | *N′* | Difference in the natural resource loss | Salvati and Marco, (2008) |
| | *P′* | Difference in the particulate matter emission damage | Sheehan et al. (2014), Silver et al. (2018) |
| | *S′* | Difference in the threatened species | McClure et al. (2018), Huang (2019) |
| | *F′* | Difference in climate investment | D'Orazio and Popoyan (2019) |
| | *E′* | Difference in energy transition | Shahbaz et al. (2013), Farhani and Shahbaz (2014) |

**TABLE 2 |** Income situation of the BRI.

| | |
|---|---|
| Low-income country | Afghanistan, Nepal |
| Middle-low–income country | Egypt, Pakistan, Philippines, Kyrgyzstan, Mongolia, Bangladesh, Myanmar, Moldova, Ukraine, Uzbekistan, India, Indonesia, Vietnam, Georgia |
| Middle-high–income country | China, Albania, Azerbaijan, Belarus, Bulgaria, Russian Federation, Kazakhstan, Romania, Malaysia, Turkey, Thailand, Jordan, Iraq |
| High-income country | Oman, Kuwait, Bahrain, Qatar, Croatia, Lithuania, Estonia, Saudi Arabia, Slovak Republic, Slovenia, Poland, Hungary, Brunei, Israel, Greece |

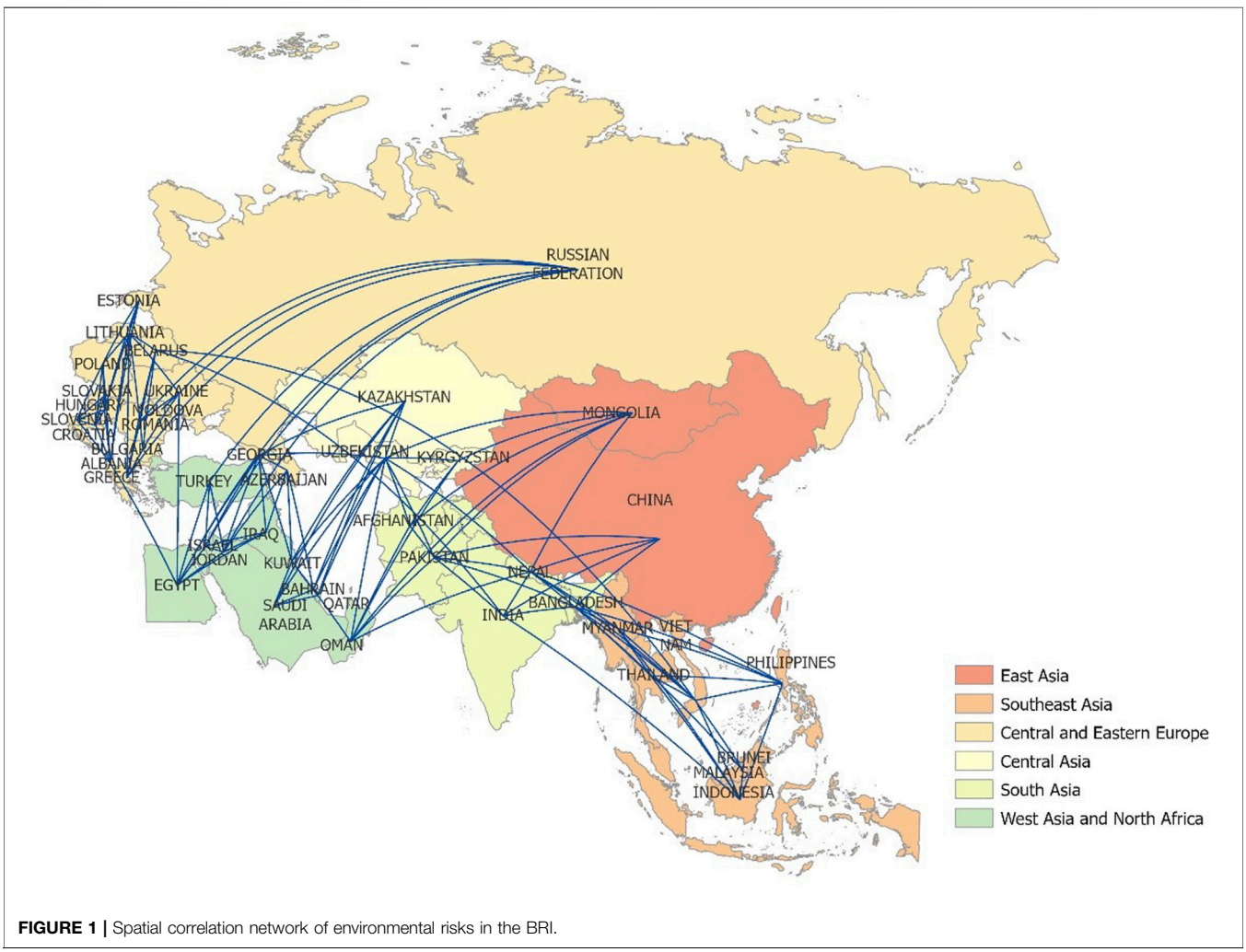



**FIGURE 1 |** Spatial correlation network of environmental risks in the BRI.

Infrastructure Investment Bank (AIIB) climate investment in the BRI. The AIIB's climate investment in the BRI was funded until the end of December 2018. The data come from the project information compilation on the official website of the AIIB, and the unit is 100 million dollars. The installed capacity of renewable energy data was derived from the International Renewable Energy Agency (IRENA) database in 2018, and the unit is MW.

# RESULTS AND DISCUSSION

## Overal Network Stucture Analysis

The spatial correlation network diagram of environmental risks in the BRI was drawn by ArcGIS (**Figure 1**). The environmental risks of the BRI countries present a complex and multi-threaded network structure, and present obvious spatial contagion effects along the "One Belt and One Road" economic corridor.

This article used SNA to analyze the network relevance of environmental risks in the BRI. The network correlation degree was 1, indicating that the correlation degree between the countries is strong, the environmental risk contagion exists among countries, and the network structure is relatively stable. As can be seen from **Figure 1**, the two important main lines of environmental risk network transmission are "Belarus–Vietnam–other Southeast Asian countries" and "Belarus–India–Indonesia," which are important paths to form the "One Belt and One Road" environmental risk network contagion structure, running through the whole Eurasian continent. In addition, the spatial network transmission routes of environmental risks on the "Silk Road Economic Belt" (China-Central Asia–Europe) and the "21st Century Maritime Silk Road" (China–Africa–Europe) are obvious. They show that the Belt and Road Initiative has played a decisive role in forming the spatial network pattern of environmental risk transmission among countries along the Silk Road. This article uses the network density to measure the tightness of the environmental risks of the BRI. With a network density of 0.093, 44 countries showed that environmental risks among the countries had certain compactness, but the compactness was low. The network rating is 0.7674, indicating the existence of a rigid hierarchical network structure. To a large extent, the environmental risk contagion is concentrated in a few countries so that these countries need to focus on the control of environmental risks. The network efficiency is 0.9048, which is relatively high with many redundant connections, indicating that the network of spatial association of environmental risks between countries is stable.

## Network Centralization Analysis

### Degree

The higher the degree, the higher the environmental risk contagion of these countries that will affect the surrounding countries to a greater extent. As shown in **Figure 2**, Uzbekistan, Nepal, Bangladesh, and Lithuania are in the highest range of degree for the following reasons.

First, in terms of economic development, these countries are all low-income or middle-low–income countries, except Lithuania (**Table 2**). They have a single economic structure and an extensive economic growth mode, and the investment in environmental governance is not enough. In addition, their economic development is relatively dependent on the exploitation and utilization of water, oil, and gas, and mineral resources. The insufficient environmental protection in the process of exploitation and utilization further exacerbates the ecological environment vulnerability, thereby seriously affecting the sustainable development of the region. Therefore, the countries in the highest range of degree centers have greater environmental risk contagiousness that may transmit environmental risk to other countries through various channels. Second, in terms of geographical location, these countries are the transportation hubs of the BRI and have a strong possibility to transmit environmental risks to the surrounding countries. Uzbekistan is at the center of the BRI, and Nepal and Bangladesh are adjacent. They are located in South Asia, connecting East and Southeast Asia. Lithuania is a developed country in eastern Europe, located in the heart of Europe. They are important individuals contributing to forming the spatial contagion network pattern of environmental risks along the "One Belt and One Road."

Similarly, the countries of lower degree have lower environmental risk contagiousness in the network and will not significantly impact neighboring countries. The countries with the lowest degree are Moldova, Brunei, Malaysia, Slovak Republic, China, Belarus, Bulgaria, Estonia, and Turkey. With the exception of Moldova which is categorized as a middle-low–income country, the other countries are high-income or middle-high–income countries. High-income countries invest more in environmental governance, optimizing their industrial structure, and have a stronger ability to control environmental risks. Hence, the contagion of environmental risks is lower. Moldova is in central and eastern European with relatively poor natural resources. It is a typical energy importer which lacks hard coal, iron ore, oil, and natural gas. Therefore, its potential environmental risks are relatively small, and the possibility of environmental risk contagion is relatively low.

### Closeness

As shown in **Figure 3**, countries at the highest range of closeness include Georgia, Uzbekistan, and Egypt, indicating that they have more direct connections with others and are prone to transmitting environmental risks to other countries while being less susceptible to be controlled and contagious from other countries. Countries with closeness at the lowest closeness include Brunei in Southeast Asia, Croatia, Slovakia, and Slovenia in central and eastern Europe, which are located at the edge of the BRI and have fewer direct connections with other countries. Thus, their environmental risks have fewer external contagion paths, and they are more vulnerable to be contagious from other countries.

### Betweeness

Countries with higher betweeness have a stronger ability to influence other countries in the spatial correlation network of environmental risks and act as intermediaries and bridges. As shown in **Figure 4**, Uzbekistan, Georgia, Egypt, Ukraine, Lithuania, and Nepal have the highest betweeness, indicating that they play strong intermediary roles in the environmental risk

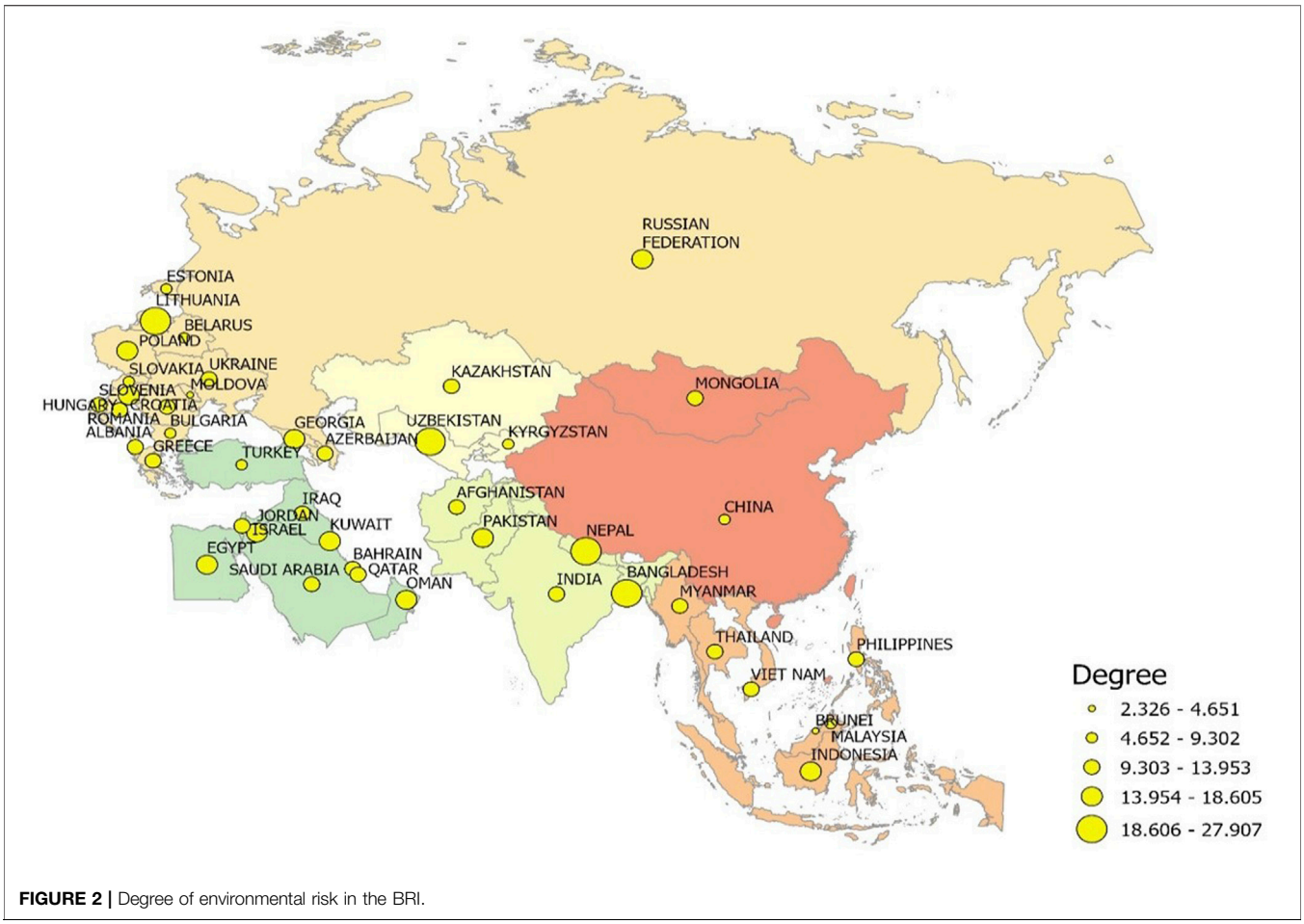



**FIGURE 2 |** Degree of environmental risk in the BRI.

contagion. Among them, Lithuania, Ukraine, and Georgia are located in central and eastern Europe, Uzbekistan is located in Central Asia, Egypt is located in West Asia and North Africa, and Nepal is located in South Asia. These countries with high betweeness are scattered in various regions and play the roles of intermediary contagion.

Countries with betweeness of 0 include China, Belarus, Brunei, Estonia, Kyrgyzstan, Moldova, Saudi Arabia, Slovak Republic, and Turkey, and they have almost no mediating effect in the network contagion of environmental risks. This indicates that they have relatively strict environmental governance and strong defense against environmental risks, and will hardly transmit environmental risks to other countries.

## Block Model Analysis

The block model was used to analyze the spatial clustering characteristics of the spatial correlation network of environmental risks in the BRI. The CONCOR method was used to select the maximum depth as 2 and the concentration standard as 2. The BRI countries were divided into four sections. The results are shown in **Figure 5**. There are seven members in the first plate, namely, Afghanistan, Nepal, Oman, Pakistan, Mongolia, Kyrgyzstan, and Uzbekistan, which are mainly situated in South Asia, West Asia, and Central Asia. The second plate has seven members, including Qatar, Saudi Arabia, Kuwait, Bahrain, India, China, and Kazakhstan, distributed mainly in East Asia and West Asia. The third plate has 10 members, including Croatia, Lithuania, Albania, Hungary, Slovak Republic, Estonia, Belarus, Slovenia, Bulgaria, and Poland, distributed mainly in central and eastern Europe. The fourth plate has 20 members, including Azerbaijan, Greece, Israel, Bangladesh, Moldova, Brunei Darussalam, Iraq, Egypt, Jordan, Georgia, Romania, Russian Federation, Myanmar, Indonesia, Malaysia, Thailand, the Philippines, Vietnam, Turkey, and Ukraine, distributed mainly in Southeast Asia, West Asia, and central and eastern Europe. The total number of contagion relationships in the BRI is 352, the number of intra-plate relations is 139, and the number of inter-plate relations is 213, indicating that the spillover effect between plates was noticeable.

According to the block theory model (White et al., 1976; Snyder and Kick, 1979), we divided the BRI into four plates (**Table 3**). The number of outgoing relations of plate I is 39, but the number of receiving relations is 32, and the number of internal relations is 26. The expected internal relation proportion is 14%, but the actual internal relation proportion is 69%. The members of this plate send more connections to other plates than they receive, which is the "net overflow" plate. The sending relation number of plate II is 19, the receiving relation

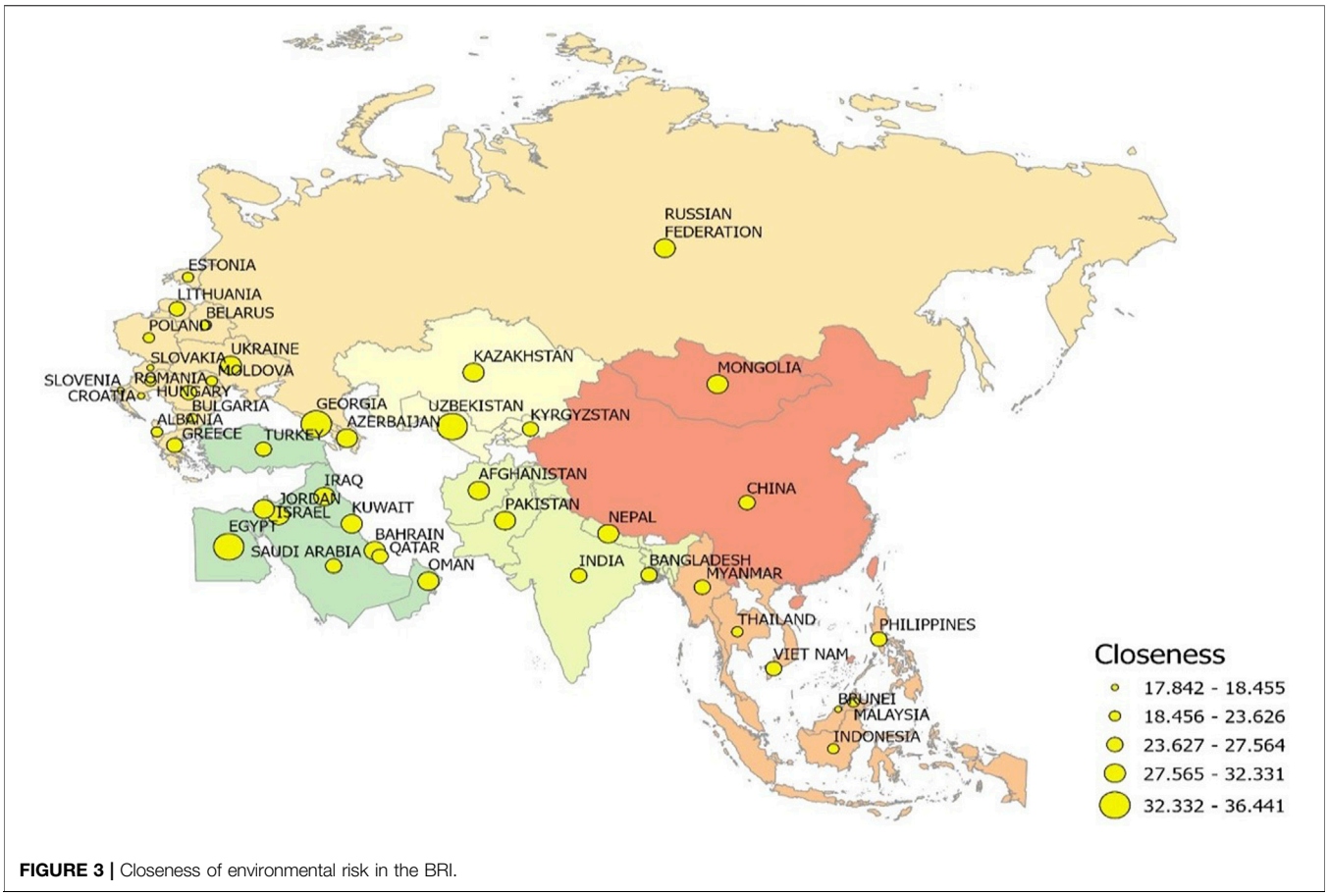



**FIGURE 3 |** Closeness of environmental risk in the BRI.

number is 28, and the internal relation number is 14. The number of internal relations of this plate is less than that of external relations, so it is the "broker" plate. The sending relation number of plate III is 39, the receiving relation number is 36, and the internal relation number is 32. The sending relation number of members of this plate is also higher than the receiving relation number of other plates, and it is the "main beneficiary" plate. The sending relation number of plate IV is 79, the receiving relation number is 80, and the internal relation number is 67. The expected internal relation ratio is 44%, and the actual internal relation ratio is 85%. The actual internal relation ratio is greater than the expected internal relation ratio, so it is a "two-way overflow" plate.

In addition, according to the distribution of the contagion relationship among environmental risk plates, this article calculates the network density matrix of each plate. If the plate density is greater than the network density which is 0.093, the value is assigned 1; otherwise, it is 0. In this way, the multivalued density matrix is transformed into an image matrix (**Table 4**). The probability value on the main diagonal of the image matrix is 1, indicating that the environmental risk contagion of the BRI within the plate has a significant correlation, showing an obvious "club" effect. As shown in **Table 4**, the spillover effect of the first plate is mainly reflected in the first and second plates, while the spillover effect on the other environmental risk plates is mainly reflected in the interior of each plate and has no significant impact on the other plates. **Figure 5** shows that the engine of environmental risk is the third plate, which transmits environmental risk to the fourth plate, and the fourth plate acts as a bridge and hub.

## Quadratic Assignment Procedure Analysis

### Quadratic Assignment Procedure Correlation Analysis

Environmental risk contagion represents a kind of "relationship" between countries. All variables in the model are "relational data" and appear in the form of a matrix. Therefore, it is not possible to test whether there are mutual effects among the relational data through the conventional statistical test method. The QAP method is one of the most common methods in network analysis (Li et al., 2014). As it does not require the assumption that independent variables are independent, it is more robust than the parametric method. Based on the substitution of the matrix, QAP correlation analysis generates the correlation coefficient by comparing the similarity of each lattice value in the two square matrices and then carries out a nonparametric test on the correlation coefficient.

The UCINET software is used herein to randomly permute the matrix data 5,000 times, thus obtaining the correlation coefficient between the spatial association matrix of environmental risk contagion and other driving factors with selective random permutation 5,000 times. Minimum value and maximum value, respectively, represent the minimum and the maximum

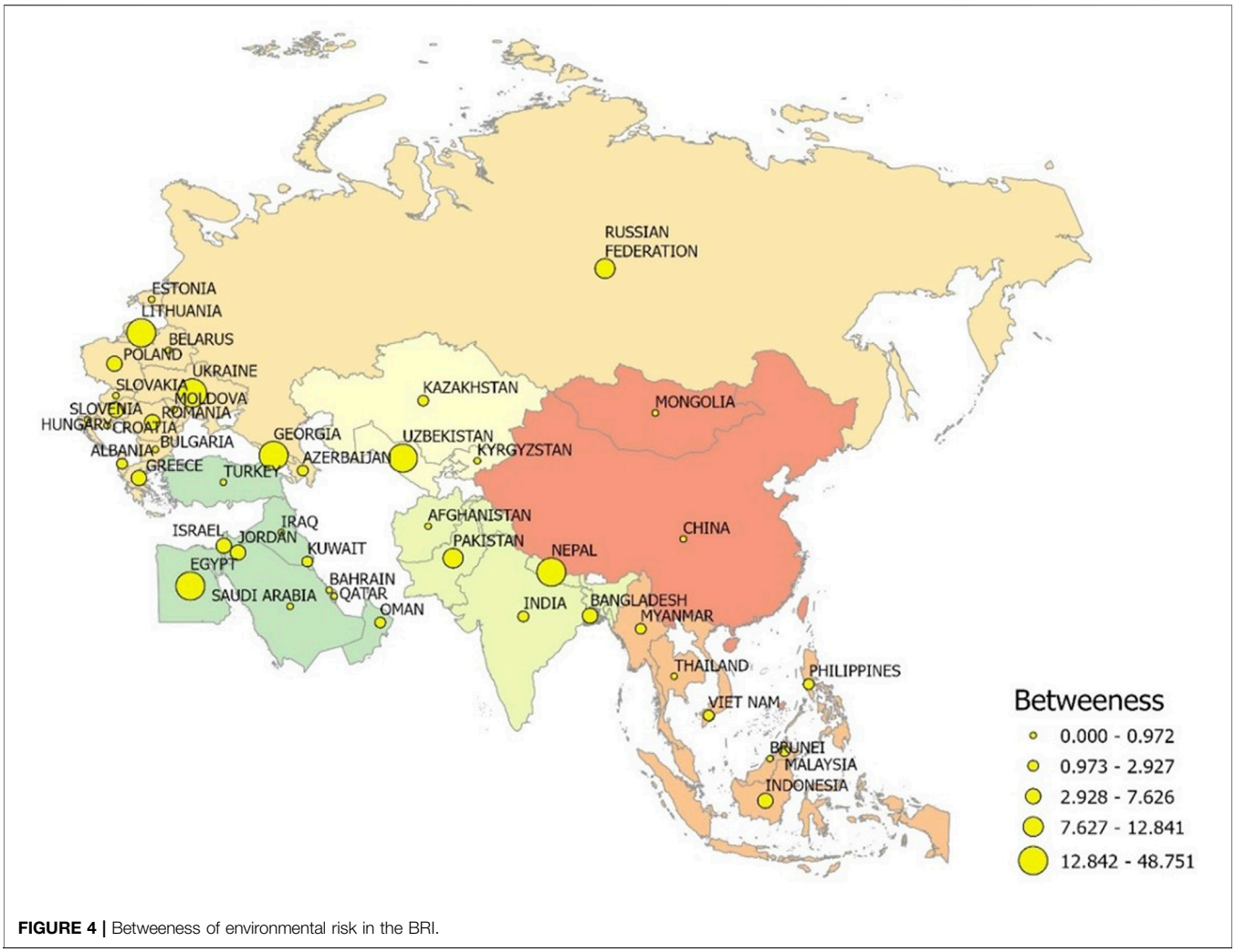



**FIGURE 4 |** Betweeness of environmental risk in the BRI.

correlation coefficients obtained by the matrix permutation, and $p \geq 0$ and $p \leq 0$, respectively, indicate the proportion of the correlation coefficients obtained by the matrix permutation greater than or equal to and less than or equal to the actual correlation coefficient. As can be seen from **Table 5**, all the driving factors including geographical distance $D$, difference in the proportion of natural resource loss $N'$, difference in the proportion of particulate matter emission damage $P'$, difference in threatened species $S'$, difference in AIIB's climate investment $F'$, and difference in renewable energy investment $E'$ are significantly correlated with the spatial contagion of environmental risks.

**Table 6** shows the correlation between the different variables in this part. Other variables have relationships with spatial adjacency relation under different significant levels; $D$ is significantly correlated with $S'$ at the 1% level, $P'$ is also significantly correlated with F′ at the 10% level, and $N'$ is significantly correlated with $S'$ at the 10% level and $E'$ at the 1% level. There exists a multicollinearity problem between independent variables. To address this issue, we have applied the QAP method to study the following regression analysis.

## Quadratic Assignment Procedure Regression Analysis

QAP regression analysis is used to investigate the regression relationship between multiple independent variable matrices and one dependent variable matrix. According to the QAP correlation analysis results, the five driving factors with significant correlation coefficients are selected as the independent variables for QAP regression analysis. Then, the UCINET software is used to randomly permute the data 2000 times, and the QAP regression results are obtained, as shown in **Table 7**. Since the QAP regression is based on matrix permutation to perform the test, normal standard errors cannot be obtained (Peoples and Sutton, 2015). The regression coefficients of the relation matrix variables and their significance test results are analyzed as follows. The absolute regression coefficients are significant except $P'$; the significance levels of $D$, $N'$, $S'$, and $E'$ are at 1% level, and the significance level of F is at 5% level. The regression coefficients of $D$, $N'$, $F'$, and $E'$ were negatively significant.

Geographical distance significantly decreases the contagion of environmental risk among countries. The farther the

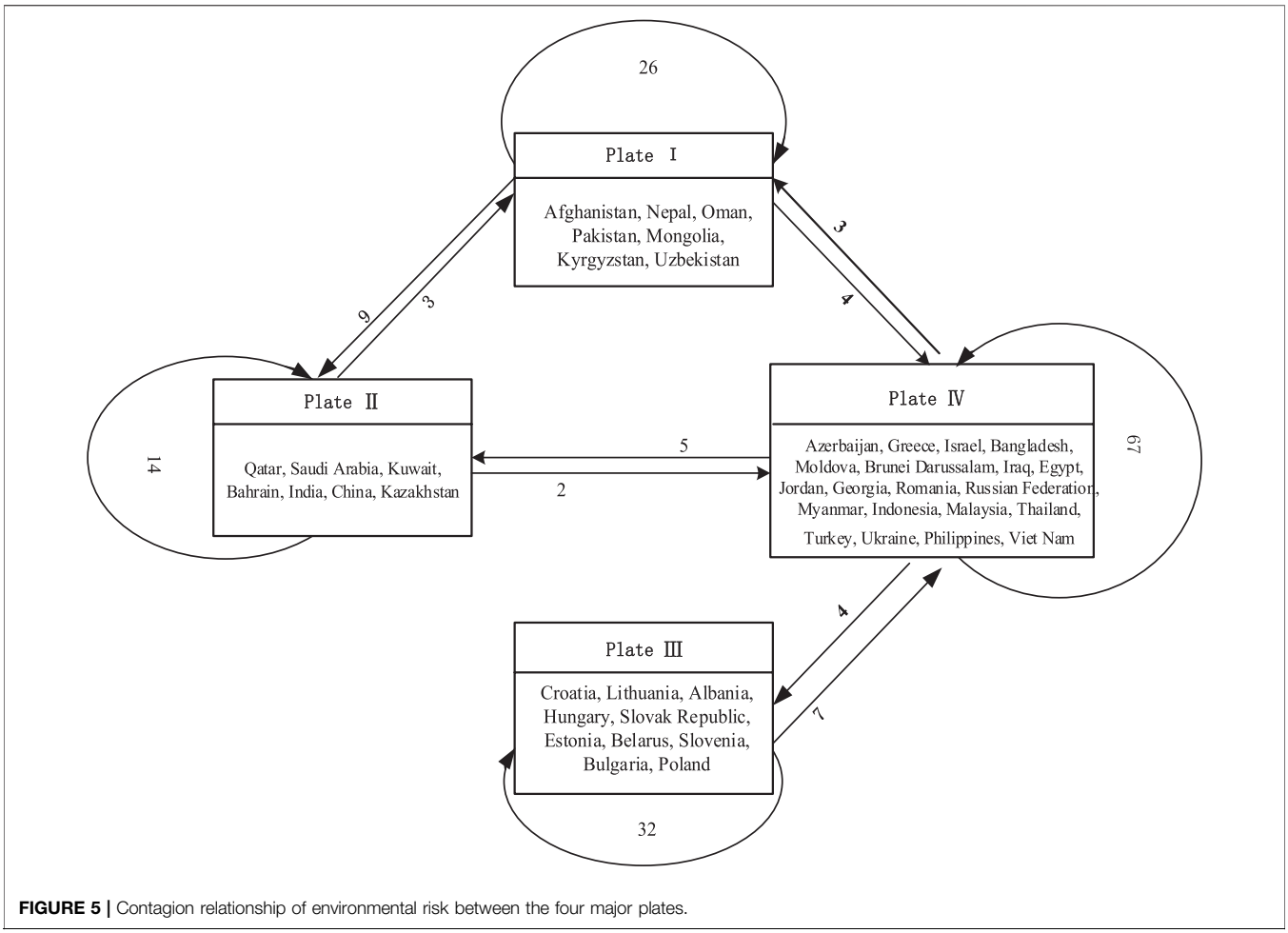



**FIGURE 5 |** Contagion relationship of environmental risk between the four major plates.

**TABLE 3 |** Spillover effects of environmental risk contagion among four plates.

| | Plate I | Plate II | Plate III | Plate IV | Number of plate members | Expected internal relationship ratio (%) | Actual internal relationship ratio (%) | Number of sending relations | Number of receiving relations |
|---|---|---|---|---|---|---|---|---|---|
| Plate I | 26 | 9 | 0 | 4 | 7 | 14 | 67 | 39 | 32 |
| Plate II | 3 | 14 | 0 | 2 | 7 | 14 | 74 | 19 | 28 |
| Plate III | 0 | 0 | 32 | 7 | 10 | 21 | 82 | 39 | 36 |
| Plate IV | 3 | 5 | 4 | 67 | 20 | 44 | 85 | 79 | 80 |

***Note:*** *The expected internal relationship ratio = (the number of members in the plate -1)/(the number of members in the whole network -1), and the actual internal relationship ratio = the number of internal relationships in the plate/the total number of overflowing relationships in the plate.*

**TABLE 4 |** Density matrix and image matrix.

| | Density matrix | | | | Image matrix | | | |
|---|---|---|---|---|---|---|---|---|
| | Plate I | Plate II | Plate III | Plate IV | Plate I | Plate II | Plate III | Plate IV |
| Plate I | 0.619 | 0.184 | 0.000 | 0.029 | 1 | 1 | 0 | 0 |
| Plate II | 0.061 | 0.333 | 0.000 | 0.014 | 0 | 1 | 0 | 0 |
| Plate III | 0.000 | 0.000 | 0.356 | 0.035 | 0 | 0 | 1 | 0 |
| Plate IV | 0.021 | 0.036 | 0.020 | 0.176 | 0 | 0 | 0 | 1 |

**TABLE 5 |** QAP analysis results of environmental risk contagion and other driving factors.

| Independent | Actual correlation coefficient | Significance | Correlation coefficient | Mean standard deviation | Minimum value | Maximum value | $p \geq 0$ | $p \leq 0$ |
|---|---|---|---|---|---|---|---|---|
| *D* | −0.356[a] | 0.000 | −0.000 | 0.027 | −0.119 | 0.084 | 1.000 | 0.000 |
| *N′* | −0.080[a] | 0.001 | 0.001 | 0.025 | −0.088 | 0.091 | 1.000 | 0.001 |
| *P′* | −0.037[b] | 0.055 | 0.000 | 0.024 | −0.082 | 0.102 | 0.945 | 0.055 |
| *S′* | −0.118[a] | 0.000 | −0.000 | 0.024 | −0.087 | 0.093 | 1.000 | 0.000 |
| *F′* | −0.037[b] | 0.055 | −0.000 | 0.023 | −0.084 | 0.091 | 0.945 | 0.055 |
| *E′* | −0.035[b] | 0.057 | −0.000 | 0.023 | −0.060 | 0.086 | 0.943 | 0.057 |

*Notes.*
[a]*Significance at 1% level.*
[b]*Significance at 10% level.*

**TABLE 6 |** Data on the relationships of driving factors using the QAP method.

| | *D* | *P′* | *N′* | *S′* | *F′* | *E′* |
|---|---|---|---|---|---|---|
| *D* | 1[a] | 0.007 | −0.002 | 0.5[a] | −0.001 | 0.029 |
| *P′* | 0.007 | 1[a] | −0.093 | −0.004 | 0.146[b] | −0.089 |
| *N′* | −0.002 | −0.093 | 1[a] | −0.116[b] | −0.002 | 0.843[a] |
| *S′* | 0.5[a] | −0.004 | −0.116[b] | 1[a] | 0.392[a] | −0.126[a] |
| *F′* | −0.001 | 0.146[b] | −0.002 | 0.392[c] | 1[a] | −0.032 |
| *E′* | 0.029 | −0.089 | 0.843[a] | −0.126[c] | −0.032 | 1[a] |

*Notes.*
[a]*Significance at 1% level.*
[b]*Significance at 10% level.*
[c]*Significance at 5% level.*

geographical distance, the lower the possibility of the environmental risk contagion among countries.

The greater the difference in loss of natural resources, the smaller the environmental risk contagion. If the loss of natural resources in a country is lower than that in other countries, the stronger ability will the country own to control environmental risks, and the less likely it is to transmit risks to other countries or be transmitted from other countries.

The greater the difference in threatened species, the greater the environmental risk contagion, that is, the more the threatened species in one country relative to other countries, the more serious the destruction of biodiversity, and the more likely it is to transmit the risks to other countries.

Climate investment difference is significantly negatively correlated with environmental risk contagion. The greater the climate investment difference, that is, the more the climate investment one country receives relative to other countries, the more the funds will be used to tackle climate change, the less likely it is to transmit risks to other countries or be transmitted from other countries.

Particulate matter emission damage difference is insignificant for environmental risk contagiousness. The possible reason is that the environmental pollutants such as $PM_{2.5}$ show significant spatial autocorrelation and spatial spillover effects in the BRI countries (Fang et al., 2020). Particulate matter emissions as environmental pollution source have produced spatial spillover effect before they resulted in environmental risk, so the contagion effect of environmental risk is weakened.

Energy transformation difference is significantly negatively correlated with environmental risk contagiousness. The greater the difference in the energy transformation, the greater the degree of energy transformation in the country than that in other countries, and the smaller the environmental risk contagion between the two countries. This indicates that the countries with the greater energy transformation and upgrading have the stronger ability to resist the environmental risk and will not be affected by other countries. Moreover, the lower the level of environmental risk within a country, the less likely it is to transmit the risks to other countries.

# CONCLUSION AND POLICY RECOMMENDATIONS

## Conclusion

The environmental risks among the BRI countries present a complex and multi-threaded network structure with strong spatial correlation and a stable network structure. The Belt and Road Initiative has played an important role in the formation of the spatial network pattern of the environmental risk contagion in countries along the Silk Road Economic Belt (China–Central Asia–Europe) and the 21st Century Maritime Silk Road (China–Africa–Europe). The spatial network contagion paths of environmental risk are obvious.

The environmental risk of the BRI has a strict hierarchical network structure, and the environmental risk contagion was concentrated in a few countries to a large extent that these countries need to focus on controlling environmental risk. The degree of Uzbekistan, Nepal, Bangladesh, and Lithuania is in the highest range. Compared with other countries in the network, the environmental risk contagious of these countries is higher, which will affect the surrounding countries to a large extent. Uzbekistan, Georgia, Egypt, Ukraine, Lithuania, and Nepal have the highest degree of betweeness, which played intermediary roles in the environmental risk contagion and have strong ability to influence the environmental risks of other countries.

According to the block model, the spatial clustering characteristics of environmental risk contagiousness of the countries along the "One Belt and One Road" route were analyzed and divided into four plates. The countries of different plates play different roles in the spacial network contagion of environmental risks. The spillover effect of the first plate is reflected mainly in the first plate and the second plate, while the spillover effect of other environmental risk plates

**TABLE 7 |** QAP regression analysis of each variable matrix.

| Independent | Unitized coefficient | Studied coefficient | Significance | Proportion as large | Proportion as small |
|---|---|---|---|---|---|
| Intercept | 0.271013 | 0 | | | |
| D | −0.000049 | −0.430356[a] | 0 | 1 | 0 |
| *N′* | −0.003501 | −0.079448[a] | 0.008 | 0.992 | 0.008 |
| *P′* | −0.021103 | −0.025639 | 0.219 | 0.782 | 0.219 |
| *S′* | 0.000156 | 0.154015[a] | 0 | 0 | 1 |
| *F′* | −0.006378 | −0.07022[b] | 0.028 | 0.972 | 0.028 |
| *E′* | 0 | −0.103788[a] | 0.002 | 0.999 | 0.002 |

*Notes.*
[a]*Significance at 1% level.*
[b]*Significance at 5% level.*
[c]*Significance at 10% level.*

is reflected mainly in the interior of each plate. The engine of environmental risk is the third plate, which transmits environmental risk to the fourth plate, and the fourth plate acts as an obvious bridge and hub.

The environmental risk contagion relationship of the BRI is significantly negatively correlated with geographical distance, differences in natural resource loss, differences in particular matter emission damage, differences in climate investment, and differences in energy transition. It was significantly positively correlated with differences in threatened species.

The limitations and future works of the article are as follows. First, the data in this article are cross-sectional data, which makes it impossible to do time series analysis. Second, this paper only aims to analyze the overall situation of "environmental risk contagion" between countries, and lack of country-by-country analysis. It's difficult to analyze all aspects in a single study, the following research may further analyze the influence mechanism of the environmental risk level of the BRI countries.

## Policy Recommendations

The Belt and Road Initiative has played an important role in the formation of the spatial network structure of the environmental risk contagion. While strengthening environmental governance and striving to reduce their environmental risks, the countries should strengthen collaborative governance, reduce ecological and environmental risks, prevent environmental risk contagion, so as to achieve the green silk road. From the analysis of this article, it can be concluded that reducing a country's environmental risk loss and strengthening the ability of environmental risk defense can effectively prevent the environmental risk contagion from other countries. Therefore, we put forward the following suggestions for reducing the contagion of environmental risks among the BRI countries.

Strict environmental systems and strong environmental regulation can reduce the environmental risk and improve the ability to resist the environmental risk contagion effectively. Stricter environmental regulations should be formulated, especially for the countries in the BRI's transport hub, with high betweeness that play the role of the intermediary contagion. Green trade barriers should be strengthened, such as adding "green labels" to products, to achieve green production and consumption from the upstream and downstream in the trade.

The loss of natural resources is one of the driving factors of the environmental risk contagion that include forest loss, energy loss, and mineral resource loss. Therefore, the countries should expand the forest area to enhance the ecological environment's self-repair ability and carbon storage ability, and provide a suitable living environment for biodiversity. In the process of mining mineral resources, attention should be paid to protect the environment and take rational exploitation. The BRI should further increase investment in developing renewable energy by taking advantages of the geographical resources of various countries to reduce greenhouse gas emissions.

The BRI should attach great importance to protect biodiversity. As the BRI includes 27 of the 35 globally recognized biodiversity hotspots, biodiversity conservation needs to be considered as an important aspect of high-quality development in the BRI. The investment in biodiversity hotspots must take green development into consideration with the active implementation of the Convention on Biological Diversity (CBD) to avoid increasing environmental risks due to excessive investment and consumption.

Climate investment from AIIB should be increased more, especially for low-income countries. The BRI should leverage the role of the assistance fund for South-South cooperation, strengthen climate change projects, and promote energy transformation and upgrading. The BRI also can improve the green financial system and encourage more financial institutions to participate in green investments such as renewable energy.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

LL: conceptualization, methodology, formal analysis, writing original draft and editing. KF: writing - review and editing. CL: writing - review and editing. CS: draw graph by ArcGis and editing.

## FUNDING

This work has received support from the UK-China Cooperation on Climate Change Risk Assessment (PF3051_CH-WS3HBUE) and the Project of Excellent Young and Middle-aged Scientific and Technological Innovation Team in colleges and universities of Hubei Province (T201917).

## REFERENCES

Ascensão, F., D'Amico, M., and Barrientos, R. (2019). Validation data is needed to support modelling in road ecology. *Biol. Conservation* 230, 199–200. doi:10.1016/j.biocon.2018.12.023

Bai, C., Zhou, L., Xia, M., and Feng, C. (2020). Analysis of the spatial association network structure of China's transportation carbon emissions and its driving factors. *J. Environ. Manage.* 253, 109765. doi:10.1016/j.jenvman.2019.109765

Benítez-Andrades, J. A., García-Rodríguez, I., Benavides, C., Alaiz-Moretón, H., and Labra Gayo, J. E. (2020). An ontology-based multi-domain model in social network analysis: Experimental validation and case study. *Inf. Sci.* 540, 390–413. doi:10.1016/j.ins.2020.06.008

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network Analysis in the Social Sciences. *Science* 323 (323), 892–895. doi:10.1126/science.1165821

Caldecott, B., Tilbury, J., and Carey, C. (2014). *Stranded Assets and Scenarios*. Oxford, United Kingdom: University of Oxford.

Cantore, N., and Cheng, C. F. C. (2018). International trade of environmental goods in gravity models. *J. Environ. Manage.* 223, 1047–1060. doi:10.1016/j.jenvman.2018.05.036

Copeland, B. R., and Taylor, M. S. (2003). Trade, Growth and the Environment. *J. Econ. Lit.* 42 (1), 7–71. Working papers. doi:10.3386/w9823

D'Orazio, P., and Popoyan, L. (2019). Fostering green investments and tackling climate-related financial risks: which role for macroprudential policies? *SSRN J.* 160 (JUN), 25–37. doi:10.2139/ssrn.3106350

Dasgupta, P., and McKenzie, E. (2020). *The Dasgupta Review - Independent Review on the Economics of Biodiversity Interim Report*. Available at: https://www.researchgate.net/publication/346629601_The_Dasgupta_Review__Independent_Review_on_the_Economics_of_Biodiversity_Interim_Report (Accessed January 30, 2021).

Dong, S., Li, Z., Li, Y., Shi, G., Yu, H., Wang, J., et al. (2015). Resources, Environment and Economic Patterns and Key Scientific Issues of the Silk Road Economic Belt. *Am. J. Polit. Sci.* 47 (3), 389–402. doi:10.1101/lm.1480309

Fang, K., Wang, S., He, J., Song, J., Fang, C., and Jia, X. (2021). Mapping the environmental footprints of nations partnering the belt and road initiative. *Resour. Conservation Recycling* 164, 105068. doi:10.1016/j.resconrec.2020.105068

Fang, K., Wang, T., He, J., Wang, T., Xie, X., Tang, Y., et al. (2020). The distribution and drivers of PM2.5 in a rapidly urbanizing region: The Belt and Road Initiative in focus. *Sci. Total Environ.* 716, 137010. doi:10.1016/j.scitotenv.2020.137010

Farhani, S., and Shahbaz, M. (2014). What role of renewable and non-renewable electricity consumption and output is needed to initially mitigate CO2 emissions in MENA region?. *Renew. Sust. Energ. Rev.* 40, 80–90. doi:10.1016/j.rser.2014.07.17010.1016/j.rser.2014.07.170

Freeman, L. C. (2004). *The Development of social network Aanalysis: A study in the social networks*. Vancouver, BC Canada: Empirical Press.

Greenpeace (2017). Sustainability should be priority for China's Belt and Road-Greenpeace. Available online. http://www.greenpeace.org/eastasia/press/releases/others/2017-/Sustainability-should-be-priority-forChinas-Belt-and-Road–Greenpeace.

Han, M., Yao, Q., Liu, W., and Dunford, M. (2018). Tracking Embodied Carbon Flows in the Belt and Road Regions. *J. Geogr. Sci.* 28 (9), 1263–1274. doi:10.1007/s11442-018-1524-7

Hu, T., and Yang, D.WWF-US. (2018). Constructing sustainable trade under the background of a community of shared future for mankind: Promoting UN SDGs by sustainable trade. *J. Environ. Econ.* 3 (1), 139–149. doi:10.19511/j.cnki.jee.2018.01.009

Huang, Y. (2019). Environmental risks and opportunities for countries along the belt and road: location choice of China's investment. *J. Clean. Prod.* 211 (FEB.20), 14–26. doi:10.1016/j.jclepro.2018.11.093

Kuik, O., and Hofkes, M. (2010). Border adjustment for European emissions trading: competitiveness and carbon leakage. *Energy Policy* 38 (4), 1741–1748. doi:10.1016/j.enpol.2009.11.048

Leadership (2017). 'Enhancing environmental risk assessment in financial decision-making.' Background Paper for the G20 Green Finance Study Group. Available online at: http://www.pbc.gov.cn/goutongjiaoliu/113456/113469/3344238/2017071318341876206.pdf.

Lenzen, M., Moran, D., Kanemoto, K., Foran, B., Lobefaro, L., and Geschke, A. (2012). International trade drives biodiversity threats in developing nations. *Nature* 486 (7401), 109–112. doi:10.1038/nature11145

Li, H., Jiang, W., and Tang, Z. (2020). Net embodied carbon flow network in global trade and community finding analysis. *Resour. Sci.* 42 (6), 1027–1039. doi:10.18402/resci.2020.06.02

Li, J., Chen, S., Wan, G., and Fu, C. (2014). Study on the spatial correlation and explanation of regional economic growth in China: Based on analytic network process. *Econ. Res. J.* 11, 4–16. CNKI:SUN:JJYJ.0.2014-11-002.

Li, X., Yang, X., and Wang, Y. (2016). Does international trade improve the carbon productivity in China's manufacturing?. *J. Environ. Econ.* 1 (2), 8–24.

Liu, H. J., Liu, C. M., Sun, Y. N., and Economics, S. O. (2015). Spatial correlation network structure of energy consumption and its effect in China. *China Ind. Econ.* 326 (5), 86–95. doi:10.19581/j.cnki.ciejournal.2015.05.008

Liu, S., and Xiao, Q. (2021). An empirical analysis on spatial correlation investigation of industrial carbon emissions using SNA-ICE model. *Energy* 224, 120183. doi:10.1016/j.energy.2021.120183

Ma, Z. (2016). *President Xi Calls for Building 'green, Healthy, Intelligent and Peaceful Silk Road*. Available online http://www.scio.gov.cn/32618/Document/1481477/htm.

McClure, C. J. W., Westrip, J. R. S., Johnson, J. A., Schulwitz, S. E., Virani, M. Z., Davies, R., et al. (2018). State of the world's raptors: Distributions, threats, and conservation recommendations. *Biol. Conservation* 227, 390–402. doi:10.1016/j.biocon.2018.08.012

Meng, B., Peters, G., and Wang, Z. (2016). Tracing China's $CO_2$ emissions in global value chains. *Environ. Economy Res.* 1 (1), 139–149. doi:10.19511/j.cnki.jee.2016.01.002

MOFCOM (2017). Guidance on Promoting Green Belt and Road. (Ministry of Commerce of the People's Republic of China). Available online at: https://www.yidaiyilu.gov.cn/wcm.files/upload/CMSydylgw/201705/201705161104041.pdf.

Mudakkar, S. R., Zaman, K., Khan, M. M., and Ahmad, M. (2013). Energy for Economic Growth, Industrialization, Environment and Natural Resources: Living with just Enough. *Renew. Sust. Energ. Rev.* 25, 580–595. doi:10.1016/j.rser.2013.05.024

Peoples, C. D., and Sutton, J. E. (2015). Congressional bribery as state-corporate crime: a social network analysis. *Crime L. Soc Change* 64 (2-3), 103–125. doi:10.1007/s10611-015-9584-4

Reilly, W. J. (1929). *Methods for the Study of Retail Relationships*. Austin, TX: University of Texas.

Salvati, L., and Marco, Z. (2008). Natural Resource Depletion and Economic Performance ofLocal Districts: Suggestions from Within Country Analysis. *J. Sust. Dev. World Ecol.* 15 (6), 518–523. doi:10.1080/13504500809469847

Scott, J. P., and Carrington, P. J. (2011). *The sage handbook of social network analysis*. New York: Sage Publication.

Shahbaz, M., Hye, Q. M. A., Tiwari, A. K., and Leitão, N. C. (2013). *Economic Growth, Energy Consumption, Financial Development, International Trade and $CO_2$ Emissions in Indonesia*. Germany: MPRA Paper.

Sheehan, P., Cheng, E., English, A., and Sun, F. (2014). China's response to the air pollution shock. *Nat. Clim Change* 4 (5), 306–309. doi:10.1038/nclimate2197

Silver, B., Reddington, C. L., Arnold, S. R., and Spracklen, D. V. (2018). Substantial changes in air pollution across China during 2015-2017. *Environ. Res. Lett.* 13 (11), 114012. doi:10.1088/1748-9326/aae718

Snyder, D., and Kick, E. L. (1979). Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions. *Am. J. Sociol.* 84 (5), 1096–1126. doi:10.1086/226902

Torinelli, V. H., and Silva Júnior, A. F. d. A. d. (2021). Environmental risk analysis (era) in the strategic asset allocation (saa) of the international reserves (irs) managed by central banks (cbs). *Latin Am. J. Cent. Banking* 2 (3), 100021. doi:10.1016/j.latcb.2021.100021

Tracy, E. F., Shvarts, E., Simonov, E., and Babenko, M. (2017). China's new Eurasian ambitions: the environmental risks of the Silk Road Economic Belt. *Eurasian Geogr. Econ.* 58, 56–88. doi:10.1080/15387216.2017.1295876

Wasserman, S., and Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge, UK: Cambridge University Press. doi:10.1017/cbo9780511815478

White, H. C., Boorman, S. A., and Breiger, R. L. (1976). *Social Structure from Multiple Networks*. United Kingdom: Oxford University Press.

Wolfe, A. W. (1995). Social network analysis: methods and applications. *Contemp. Sociol.* 91 (435), 219–220. doi:10.2307/2077235

Wu, P.-C., Liu, S.-Y., Hsiao, J.-M., and Huang, T.-Y. (2016). Nonlinear and time-varying growth-tourism causality. *Ann. Tourism Res.* 59 (Jul), 45–59. doi:10.1016/j.annals.2016.04.005

Xiao, B. (2019). Environmental risks of outward investment in BRI countries and China's legal countermeasures. *Int. Forum.* doi:10.7176/JESD/11-16-04

Yang, Y., and Li, C. (2019). Green governance innovation against eco-environmental risks of the Belt and Road Initiative - A case study of Lancang-Mekong River Subregion. *Probe* 35, 184–193. doi:10.16501/j.cnki.50-1019/d.2019.05.020

Yao, Q. H., Han, M. Y., and Liu, W. D. (2018). Tracking embodied carbon flows in the belt and road regions. *Acta Geographica Sinica* 28 (9), 1263–1274. doi:10.1007/s11442-018-1524-7

Yin, C., Gu, H., and Zhang, S. (2020). Measuring technological collaborations on carbon capture and storage based on patents: A social network analysis approach. *J. Clean. Prod.* 274, 122867. doi:10.1016/j.jclepro.2020.122867

Yu, Z., and Di, Q. (2020). The coordination between maritime economies and marine carrying capacity and their spatiotemporal evolution in the cities of the bohai rim in China. *Ecol. Model.* 438, 109192. doi:10.1016/j.ecolmodel.2020.109192

Zhong, Z. L., Zeng, G. M., and Yang, C. P. (1996). The development of environmental risk assessment research. *Adv. Environ. Sci.* 4 (6). CNKI: SUN:HJJZ.0.1996-06-002.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Lu, Fang, Liu and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 06 August 2021
doi: 10.3389/fenrg.2021.726127

# The Major Driving Factors of Carbon Emissions in China and Their Relative Importance: An Application of the LASSO Model

*Wai Yan Shum*[1], *Ning Ma*[2]*, *Xiaomei Lin*[2] *and Tingting Han*[2]

[1]Hang Seng University of Hong Kong, Hong Kong, SAR China, [2]Hainan College of Economics and Business, Haikou, China

China is one of the biggest energy consumers and carbon emitters in the world. Understanding the factors affecting carbon emissions is critical for policymakers to control the rising trend of carbon emissions. This paper investigates the relative importance of carbon emissions drivers in China. Literature review has been carried out to determine a set of predominant independent variables; the LASSO model is then introduced to rank the relative importance among the set of independent variables. The results find that 1) carbon emissions were mainly driven by economic growth and energy consumption followed by population size and industrialization; and 2) income growth slowed down carbon emissions during the studied period, but it is the least significant factor among the other factors. The ranking allows policy makers to focus on the most critical contributors to carbon emissions and gives policymakers more flexibility in determining policy interventions.

**Keywords: energy consumption, carbon emissions, least absolute shrinkage and selection operator (LASSO), China, City'**

OPEN ACCESS

***Edited by:***
*Xunpeng Shi,*
*University of Technology Sydney, Australia*

***Reviewed by:***
*Chi Lau,*
*Teesside University, United Kingdom*
*Xing Shi,*
*Hefei University of Technology, China*

***Correspondence:**
*Ning Ma*
*china_ma_ning@hotmail.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *16 June 2021*
***Accepted:*** *27 July 2021*
***Published:*** *06 August 2021*

***Citation:***
*Shum WY, Ma N, Lin X and Han T (2021) The Major Driving Factors of Carbon Emissions in China and Their Relative Importance: An Application of the LASSO Model. Front. Energy Res. 9:726127. doi: 10.3389/fenrg.2021.726127*

## INTRODUCTION

Climate change is now a global challenge. The increase of carbon emissions, which is now at the highest level in history, is the main cause for global warming and climate change. Given that the environmental deterioration became a serious issue, studies on the driving factors of carbon emissions has become a subject incredibly significant both at the international and domestic level. Over the past few decades since 1978, China has experienced remarkable economic growth, which has been associated with deteriorating environmental conditions in the country (Ma et al., 2018). Taking measures to control the rising trend of carbon emissions is increasingly recognized as critical in the Chinese government's efforts to mitigating climate change (Liu et al., 2011). As the world's largest carbon emitter, China's effort to achieve its reduction target assumes significance for global climate change control. Understanding the driving factors of carbon emissions in Chinese cities is critical for policymakers to control the rising trend of carbon emissions.

Countries have signed agreements of economic cooperation, which have increased globalization throughout the world (Zaidi et al., 2019). Globalization has promoted worldwide growth in trade, increase in energy consumption and intensity, industrial expansion and people's income. Whether the developing countries can increase growth rates with the help of globalization without damaging environmental is a critical question.

Energy intensity and carbon emissions have a long-term linkage (Shahbaz et al., 2016). When globalization comes with innovative technology and increases the country's energy intensity and its increases the level of carbon emissions.

The relationship between carbon emissions and energy consumption and economic growth has gained little attention in the literature. From the methodology perspective, most of previous studies have used traditional panel regression models, but we introduce the Least Absolute Shrinkage and Selection Operator (LASSO) regression model to investigates the relative importance of the impact factors influencing carbon emissions in China. Compare with traditional regression method, LASSO obtains a refined regression model which provides the smallest possible forecast error with a minimum number of regressors, and it offers important information on the relative importance of the variables. The benefits of using the LASSO regression model can be summarised as follows: 1) LASSO adds first order penalty to the regressors and this allows LASSO to select out the relevant predictors for dependent variables (Hastie et al., 2019); 2) The importance of independent variables in terms of the change of parameters of the LASSO model can be ranked and, this gives policymakers more flexibility in determining policy interventions (Shi et al., 202).

The remainder of this paper is structured as follows: *Literature Review* presents the literature review linked to energy consumption, carbon emission, income, population size, industrialization, and economic growth. *Methodology* introduces the LASSO regression model. *LASSO Regression Results* represents the empirical results and discusses the key drivers of carbon emissions. *Conclusion and Policy Implications* comprises the conclusion of the study and policy implications.

## LITERATURE REVIEW

Numerous studies have proved that economic growth and energy consumption are important driving factors for carbon emissions. Shabaz et al. (2013) examined the relationship between economic growth, energy consumption, financial development, trade openness, and carbon emissions from 1975 to 2011 in Indonesia. Their study applied Vector Error Correction Model (VECM) causality analysis, finding that economic growth and energy consumption increases carbon emissions. Munir et al. (2020) examined the relationship between carbon emissions, energy consumption, and economic growth for the five main Association of Southeast Asian Nations (ASEAN) countries. They have applied Granger non-causality test and found that there was causality between GDP and energy consumption for all countries. Meanwhile, Shahbaz et al. (2016) examined the direction of causality among carbon emissions, energy consumption, and economic growth in 11 countries for the period 1972–2013. The authors used a novel approach and found that economic growth caused carbon emissions in Bangladesh and Egypt. Hu et al. (2020) used the Tapio decoupling model and Kaya-LMDI (Logarithmic Mean Divisia Index) model to investigate the spatiotemporal evolution of decoupling and driving factors of carbon emissions of 57 Belt and Road Initiative countries from 1991 to 2016. According to the results, most countries' carbon emissions significantly increased due to economic growth. Liu and Hao (2018) examined the relationship between energy consumption and economic growth of 69 countries along the Belt and Road between 1970 and 2013. This paper applied the VECM model, fully modified OLS, and dynamic OLS approaches. According to the findings, there was a bidirectional causality between carbon emissions, energy use, and GDP per capita for the energy-exporting countries. Zaman and Moemen (2017) examined the interrelationship between energy consumption, economic growth, and carbon emissions under the six alternative and plausible hypotheses in the context of low and middle-income countries, high-income countries, and in the aggregated panel, from 1975 to 2015. The results supported energy-induced emissions in different regions of the world. Kahouli (2018) examined the linkages between electricity consumption, carbon emissions, R&D stocks, and economic growth for Mediterranean countries over 1990–2016. The findings revealed that there were strong feedback effects between electricity consumption, carbon emissions, R&D stocks, and economic growth. Awodumi and Adewuyi (2020) adopted a non-linear autoregressive distributed lag (ARDL) technique to examine the role of non-renewable energy in economic growth and carbon emissions of oil-producing countries in Africa during 1980–2015. The result has confirmed that there is an asymmetric effect of per capita consumption of petroleum and natural gas on economic growth and carbon emissions per capita in all the selected countries except Algeria. Correspondingly, Mohmand et al. (2020) investigated the causal relationship between transport infrastructure, economic growth, and transport emissions in Pakistan from 1971–2017. The results found a short-term causality running from transport infrastructure, economic growth, and fuel consumption to carbon emissions. In the long run, a bidirectional relationship exists between economic growth and infrastructure.

In addition to economic growth and energy consumption, other factors like industrialization, population size, and income level are also important driving forces for carbon emissions. Nasir et al. (2021) examined the relationship between carbon emissions, economic growth, energy consumption, industrialization, and other factors in Australia from 1980 to 2014. The study found that all variables have affected carbon emissions. Li et al. (2021) discussed the effect of economic growth, economic structure, and other factors on per capita carbon emissions in 147 countries from 1990 to 2015. The results showed that at the global level, economic growth and economic structure were respectively the most significant positive and negative factors affecting carbon emissions. Luo et al. (2021) investigated the factors influencing carbon emissions in Shanghai during 1995–2017. They applied the LMDI method and Granger causality test, and they found that motor vehicle amount, disposable personal income, and carbon intensity are the top three driving factors of carbon emissions. Zhang et al. (2020) analysed the decoupling elasticity between carbon emissions, GDP, and energy consumption in China and ASEAN countries throughout 1990–2014. Based on the LMDI method, the authors found that carbon density, energy intensity,

GDP, and population have a positive relationship with carbon emissions.

Based on the above literatures, it can be concluded that economic growth, energy consumption, population size, industrialization and income can be classified as the predominant primary driving factors for carbon emissions. Other factors, such as foreign direct investment (Essandoh et al., 2020; Le et al., 2020; Khan and Rana, 2021), financial development (Shahbaz et al., 2013; Bhattacharya et al., 2017; Zaidi et al., 2019; Wang et al., 2020), oil price (Brini et al., 2017; Shahbaz et al., 2021), renewable energy consumption (Vo et al., 2020; Assi et al., 2021; Magazzino et al., 2021) and innovation (Nguyen et al., 2020) can be largely classified as the secondary driving factors which affect the primary factors. For example, it has already been proven that foreign direct investment, financial development and innovation are all positively related to growth (Jones, 1995; Muhammad and Khan, 2019), while oil price and renewable energy consumption will affect energy consumption as a whole. Instead of looking at the detailed components of each driving factors, this study limits the focus on ranking the relative importance of the primary driving factors.

## METHODOLOGY

The Least Absolute Shrinkage and Selection Operator (LASSO) is a reduction and selection method for the linear regression model which enhances its prediction accuracy and limits the regression coefficients within a certain range at the same time. It was originally proposed by Robert Tibshirani (Tibshirani, 1996) based on Leo Breiman's non-negative parameter inference (Breiman, 1995). The main objective of LASSO is to obtain a refined regression model which provides the smallest possible forecast error with a minimum number of regressors. Given a set of regressors $x_1, x_2, \cdots, x_n$ and the regressand $y$, LASSO fits the linear model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$. The selection criterion is to minimize the objection function:

$$\frac{1}{2n}\sum_{i=1}^{n}(y-\hat{y})^2 + \lambda\sum_{j=1}^{p}\left|\beta_j\right|$$

where $\hat{y}$ denotes the fitted value of the linear model, $n$ is the number of observations, $p$ signifies the number of regressors and $\lambda$ represents a non-negative regularization parameter.

The first sum of the objective function is the usual sum of squared errors of the multiple linear least squares regression while the second sum is a LASSO regularization term. The regularization term has no effect when $\lambda$ is small enough and the LASSO regression method is the same as the least squares' method. However, when $\lambda$ is large enough, all the regression coefficients are forced to be zero; for this reason, the LASSO solutions are reduced versions of the least squares estimates. It follows that, the coefficients will change from zero to nonzero one after another by continuously adjusting the value of $\lambda$ in descending order. The quality of the LASSO estimator can be measured by the mean squared error (MSE) defined as $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. MSE is a reliable quality measurement for LASSO regression model selection. The smaller the MSE, the higher the quality and vice versa. Hence, it is apparent that the LASSO estimation method can be used to derive a model which provides the smallest possible forecast error with a minimum number of regressors.

The LASSO regression method has been applied by Shi et al. (2020) to study carbon emissions in China at household level. The present study follows the approach given in their paper to find the optimal model at city level. The full details of the approach can be found in their paper whereas a summary is given below: In Shi et al. (2020), the value of $\lambda$ is adjusted as a descending geometric sequence. The maximum value of $\lambda$ is set in such a manner that all coefficients except the intercept are forced to be zero; the value of $\lambda$ is then adjusted downward geometrically such that the minimum value of $\lambda$ is 1.00E-04 times the maximum value. A hundred specifications are constructed and estimated within the range of the maximum and minimum values of $\lambda$. Specification one is the specification of the minimum value of $\lambda$, and specification (100) is the specification of the maximum value of $\lambda$.

According to the order in which the coefficients appear, it is possible to identify the regressor which is the most important for model prediction. The more important factors appear earlier, whereas the less important factors appear later. Thereby, unlike the traditional regression method which signifies a set of significant factors, LASSO offers important information on the relative importance of the variables. In essence, LASSO provides a ranking for the significance of the variables. The ranking has important policy implications for policy makers when resources are scarce; it allows them to focus on the most critical areas and allocate resources effectively on those areas.

Apart from the relative importance of the variables, LASSO also provides important information on the robustness of the coefficients. Unlike the traditional regression method which estimates the model only once, LASSO estimates the model repeatedly for different values of $\lambda$. Thus, the values and the significance of the coefficients can be observed when the value of $\lambda$ is adjusted downward. Therefore, the model's robustness can also be studied by observing the significance of the variables for different specifications.

## LASSO REGRESSION RESULTS

Six main factors of carbon emissions have been considered in this study. The details of the variables of the linear model are illustrated in **Table 1**. There were 286 prefecture and above-prefecture levels (PAA) cities in China by the end of 2014. However, there are two cities in Tibet for which most of the data are not available (Lasa and Rikaze). Therefore, we use a panel data set encompassing across 284 cities across the period 2002 to 2014. All variables are logarithmic transformed except *IND*. Our datasets are comprised mainly from four main components: 1) China City Statistical Yearbook (CCSY); 2) China Environmental Statistical Yearbook (CEST); 3) China Urban Construction Statistical Yearbook (CUCSY); China Statistical Yearbook (CSY). The summary of our variables is well represented in

**TABLE 1 |** Variables description.

| Variables | Abbreviations | Units |
|---|---|---|
| Carbon emissions | $CO_2$ | Ten thousand tons |
| GDP | *GDP* | (100) million Chinese Yuan |
| Population | *POPULATION* | 10 thousand people |
| GDP per Capita | *GDP_PC* | Chinese Yuan |
| Per Capita Energy Consumption | *ENERGY* | Tons of coal |
| Percentage of Secondary Industry | *IND* | % |
| Average Annual Salary | *INCOME* | Chinese Yuan |

**TABLE 2 |** Descriptive statistics.

| Variable | Mean | Std. Dev. |
|---|---|---|
| Log Carbon emissions | 2.73 | 0.47 |
| Log Per Capita Energy Consumption | –0.44 | 0.44 |
| Log GDP | 2.72 | 0.44 |
| Log GDP per Capita | 4.18 | 0.35 |
| Log Average Annual Salary | 4.28 | 0.21 |
| Log Population | 2.53 | 0.30 |
| Percentage of Secondary Industry | 0.49 | 0.28 |

**TABLE 3 |** LASSO regression results for specification (99) and (100).

| | Specification | |
|---|---|---|
| | (99) | (100) |
| Regressors | Estimates | Estimates |
| Log GDP | 0.070089 | 0 |
| Log POPULATION | 0 | 0 |
| Log GDP_PC | 0 | 0 |
| Log ENERGY | 0.01255 | 0 |
| IND | 0 | 0 |
| Log INCOME | 0 | 0 |
| Intercept | 2.54587 | 2.731,028 |
| d.f. | 2 | 0 |
| λ | 0.343,431 | 0.376,915 |
| MSE | 0.193,211 | 0.219,535 |

*Notes: 1) d.f. is the number of non-zero estimates. 2) λ is the LASSO regularization parameter. 3) MSE is the mean squared error of the model defined as* $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

**TABLE 4 |** LASSO regression results for specification (82) and (83).

| | Specification | |
|---|---|---|
| | (82) | (83) |
| Regressor | Estimates | Estimates |
| Log GDP | 0.492,542 | 0.488,897 |
| Log POPULATION | 0.012009 | 0 |
| Log GDP_PC | 0 | 0 |
| Log ENERGY | 0.446,819 | 0.430,176 |
| IND | 0 | 0 |
| Log INCOME | 0 | 0 |
| Intercept | 1.555,549 | 1.588,681 |
| d.f. | 3 | 2 |
| λ | 0.070627 | 0.077513 |
| MSE | 0.035029 | 0.036883 |

*Notes: 1) d.f. is the number of non-zero estimates. 2) λ is the LASSO regularization parameter. 3) MSE is the mean squared error of the model defined as* $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

**Table 2** below and this is a strongly balanced panel dataset, which will show us more robust results.

**Table 3** presents the LASSO regression results for specifications (99) and (100). Specification (100) uses the maximum $\lambda$ with a value of 0.376,915; as such, all coefficients are forced to be zero with the exception of the intercept. When the value of $\lambda$ reduced by one step from 0.376,915 in specification (100) to 0.343,431 in specification (99), the variables Log *GDP* and Log *ENERGY* become nonzero. Moreover, the MSE decreased by 12% from 0.219,535 to 0.193,211. Thus, it can be concluded that GDP and per capita energy consumption are the two most important influencing factors for carbon emissions in China. The results are in line with the previous studies. Please see, for example Shabaz et al. (2013), Munir et al. (2020), Hu et al. (2020, Mohmand et al. (2020, Zaman and Moemen (2017), and Shahbaz et al. (2016) for a detail discussion on the effects of GDP growth and energy consumption on carbon emission. The effects of GDP and per capita energy consumption on carbon emissions can be summarized as follow: $CO_2$ is expected to be positively related to Log *GDP*; without major policy interventions, improving energy efficiency alone is unlikely to cope with the negative environmental impacts attributed to economic growth. Meanwhile, with an increase in energy consumption per head, the carbon emission level is expected to increase unless the increase in energy consumption per head is driven mainly by the decrease in population.

**Table 4** unveils the third factor affecting carbon emissions in China. From **Table 4**, it is apparent that when the value of $\lambda$ reduces from 0.077513 in specification (83) to 0.070627 in specification (82), Log *POPULATION* becomes nonzero and the MSE decreases by 82% from 0.193,211 in specification (99) to 0.035029 in specification (82). It indicates that Log *POPULATION* is an important variable since its inclusion as an additional variable in the regression model significantly reduced the model's MSE. Unsurprisingly, Shahbaz et al. (2016), and Hu et al. (2020 also found the similar results. This finding is alarming because given a fixed carbon emission target, the larger the population, the lower the allowable per capita emission level. Put differently, as the population in China grows, everyone's pollution rights are declining.

Note that GDP per capita (*GDP_PC*) is calculated by dividing the GDP of a country by its population. Having all three variables (Log *GDP,* Log *POPULATION,* and Log *GDP_PC*)) in the model

**TABLE 5 |** LASSO regression results for specification (66) and (67).

| | Specification | |
|---|---|---|
| | (66) | (67) |
| **Regressor** | **Estimates** | **Estimates** |
| Log GDP | 0.158,274 | 0.169,883 |
| Log POPULATION | 0.695,022 | 0.672,686 |
| Log GDP_PC | 0 | 0 |
| Log ENERGY | 0.867,245 | 0.855,307 |
| IND | 0.000145 | 0 |
| Log INCOME | 0 | 0 |
| Intercept | 0.907,977 | 0.934,912 |
| d.f. | 4 | 3 |
| λ | 0.015941 | 0.017495 |
| MSE | 0.013801 | 0.014087 |

*Notes: 1) d.f. is the number of non-zero estimates. 2) λ is the LASSO regularization parameter. 3) MSE is the mean squared error of the model defined as* $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

**TABLE 6 |** LASSO regression results for specification (63) and (64).

| | Specification | |
|---|---|---|
| **Regressor** | **(63)** | **(64)** |
| Log GDP | 0.129,156 | 0.136,991 |
| Log POPULATION | 0.752,109 | 0.73538 |
| Log GDP_PC | 0 | 0 |
| Log ENERGY | 0.897,685 | 0.888,075 |
| IND | 0.000598 | 0.000461 |
| Log INCOME | −0.00535 | 0 |
| Intercept | 0.856,457 | 0.857,175 |
| d.f. | 5 | 4 |
| λ | 0.012059 | 0.013234 |
| MSE | 0.013131 | 0.013346 |

*Notes: 1) d.f. is the number of non-zero estimates. 2) λ is the LASSO regularization parameter. 3) MSE is the mean squared error of the model defined as* $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

will introduce perfect multicollinearity, which will prevent the least squares method from solving the system of equations. Evidently, it is feasible to drop one variable to avoid perfect multicollinearity. The challenge is, no regression model can automatically identify the set of linearly dependent variables and correctly pick the most significant variables out of the set. With LASSO, the solutions can be found by omitting the most insignificant variable and including the most important ones. Based on the result presented in **Table 3**, it can be inferred that, out of the three linearly related variables, Log *GDP* is the most significant factor affecting carbon emissions. Together with the evidence in **Table 4**, it can be concluded that Log *GDP* and Log *POPULATION* are more important factors affecting carbon emissions in China when compared with Log *GDP_PC* = Log*(GDP/POPULATION)*. In other words, GDP per capita which shows the manner in which the economy grows with the population is not as important as GDP and population by themselves.

**Table 5** presents the LASSO regression results for specifications (66) and (67). It can be observed that when the value of $\lambda$ reduced from 0.017495 in specification (67) to 0.907,977 in specification (66), the variable *IND* becomes nonzero. Moreover, the MSE decreased from 0.035029 in specification (82) to 0.013801 in specification (66). Therefore, it can be concluded that *IND* is the fourth influencing factor for carbon emissions in China. The results were in consonance with the expectations due to Bhattacharya et al. (2017), Zaman and Moemen (2017) and Liu and Hao (2018) found the similar results. Over the past few decades, China has focused on using its cheap labour force and coal to expand its manufacturing sector. Thus, its energy-intensive manufacturing sector is expected to be one of the main drivers of carbon emissions in China.

**Table 6** shows the LASSO regression results for specifications (63) and (64). It reveals the fifth significant factors affecting carbon emissions in China. According to **Table 6**, when the value of $\lambda$ reduced from 0.013234 in specification (64) to 0.012059 in specification (63), the variable Log *INCOME* becomes nonzero. Moreover, the MSE decreased from 0.013801 in specification (66) to 0.013131 in specification (63). Thus, it can be concluded that Log *INCOME* is the fifth influencing factor for carbon emissions in China. For previous studies, Le et al. (2020, Awodumi and Adewuyi (2020), Apergis et al. (2018) and Nasir et al. (2021) have concluded that there was a negative relationship between income and carbon emissions. The variable *INCOME* is a measure of the average annual salary of Chinese citizens. As the Chinese citizens' income level increases, carbon emissions reduce - although their effect of it is not that important. Notably, the variable Log *GDP_PC* remains zero and indeed it remains zero for the rest of the 62 specifications (i.e., specifications 1–62). This finding is consistent with what has been presented in **Table 3** and **Table 4**. In essence, the LASSO regression method identified the two most important variables out of the three perfectly linearly related variables and ignored the least important one.

Lastly, it can be observed the significance of the variables remains stable throughout the regression process. Put succinctly, once the variables turn to nonzero, they will not turn to zero again later. Therefore, it implies that the LASSO estimation results presented in this study are robust.

## CONCLUSION AND POLICY IMPLICATIONS

In this paper, the main driving factors of China's carbon emissions have been analysed. The evidence shows that, firstly, carbon emissions were mainly driven by economic growth (*GDP*) and Per Capita Energy Consumption (*ENERGY*) followed by population size (*POPULATION*) and Percentage of Secondary Industry (*IND*). Secondly, annual average salary (*INCOME*) slowed down the growth of carbon emissions during the studied period, but it is the least significant factor among the five factors. Thirdly, out of the three perfectly linearly related variables (Log *GDP,* Log *POPULATION,* and Log *GDP_PC*), Log *GDP* is the most significant factor affecting carbon emissions followed by population size (Log *POPULATION*) whereas Log *GDP_PC* is the least important factor and is dropped automatically. In other words, GDP per capita which shows the manner in which the economy grows with the population

is not as important as economic growth and population by themselves. Lastly, since the significance of the variables remains stable throughout the regression process, the above results are robust.

This study contributes to the existing literature in the following ways. The LASSO regression method not only addressed the problem of perfect multicollinearity and identified a set of significant factors influencing carbon emissions in China but also provided important information on the ranking for the variables' significance. It allows policymakers to identify the main area of focus, allocate resources to those prioritized areas, and reduce carbon emissions more effectively by having more flexible policy interventions.

Based on the relevant importance of the driving factors, this research has the following policy recommendations. The first is to promote green manufacturing to achieve sustainable economic development. As the largest developing country in the world, China is facing an extremely difficult task in maintaining sustainable growth. China's energy structure is still dominated by coal, and coal occupies an important share in energy conversion. The large amount of coal consumption, especially the direct combustion of coal in the terminal production process, puts tremendous pressure on carbon emission reduction. Green manufacturing can be enhanced in two directions: On the one hand, advanced and applicable energy-saving, low-carbon, and environmentally friendly technologies can be used to transform traditional industries into low-carbon factories to reduce process-related emissions. One the other hand, efficient and effective polices can be implemented to promote the use of renewable energy, such as solar and wind energy, instead of fossil fuels to reduce energy-related emissions.

Secondly, population is also an important factor affecting carbon emissions in China. The larger the population, the higher the total emissions and the lower the allowable per capita emission level given a fixed carbon emission target. Thus, considering the relative importance of the variable and the huge population size in China, lowering energy consumption per head is the second most important tasks to reduce carbon emissions in China. It is essential to promote the green life movement extensively to encourage residents to adopt a green, low carbon, civilized and healthy lifestyle in terms of clothing, food, housing, transport, and travel.

In this study, the focus is limited to the relative importance of the primary driving factors (i.e., economic growth, energy consumption, population size, percentage of secondary industry and average annual salary) on carbon emissions in China. These effects are worth investigating for the future research on a global perspective. In addition, future research could incorporate other non-linear analytical techniques into the analysis to complement existing studies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

WS: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing" Original draft preparation, Writing–Review and Editing. NM: Conceptualization, Data Curation, Methodology, Software, Visualization, Writing" Original draft preparation, Writing–Review and Editing. XL: Conceptualization, Project administration, Data Curation, Writing–Review and Editing. TH: Conceptualization, Project administration, Data Curation, Writing–Review and Editing.

## FUNDING

1. This work was supported by the Hainan College of Economics and Business (Project Reference Number: hnjmk2021301). 2. This work was supported by the Faculty Development Scheme of the Research Grants Council of Hong Kong (Project Reference Number: UGC/FDS14/B19/16).

## REFERENCES

Apergis, N., Can, M., Gozgor, G., and Lau, C. K. M. (2018). Effects of export Concentration on CO2 Emissions in Developed Countries: an Empirical Analysis. *Environ. Sci. Pollut. Res.* 25 (14), 14106–14116. doi:10.1007/s11356-018-1634-x

Assi, A. F., Zhakanova Isiksal, A., and Tursoy, T. (2021). Renewable Energy Consumption, Financial Development, Environmental Pollution, and Innovations in the ASEAN + 3 Group: Evidence from (P-ARDL) Model. *Renew. Energ.* 165, 689–700. doi:10.1016/j.renene.2020.11.052

Awodumi, O. B., and Adewuyi, A. O. (2020). The Role of Non-renewable Energy Consumption in Economic Growth and Carbon Emission: Evidence from Oil Producing Economies in Africa. *Energ. Strategy Rev.* 27, 100434. doi:10.1016/j.esr.2019.100434

Bhattacharya, M., Rafiq, S., Lean, H. H., and Bhattacharya, S. (2017). The Regulated Coal Sector and CO2 Emissions in Indian Growth Process: Empirical Evidence over Half a century and Policy Suggestions. *Appl. Energ.* 204, 667–678. doi:10.1016/j.apenergy.2017.07.061

Breiman, L. (1995). Better Subset Regression Using the Nonnegative Garrote. *Technometrics* 37 (4), 373–384. doi:10.1080/00401706.1995.10484371

Brini, R., Amara, M., and Jemmali, H. (2017). Renewable Energy Consumption, International Trade, Oil price and Economic Growth Inter-linkages: The Case of Tunisia. *Renew. Sust. Energ. Rev.* 76, 620–627. doi:10.1016/j.rser.2017.03.067

Essandoh, O. K., Islam, M., and Kakinaka, M. (2020). Linking International Trade and Foreign Direct Investment to CO2 Emissions: Any Differences between Developed and Developing Countries? *Sci. Total Environ.* 712, 136437. doi:10.1016/j.scitotenv.2019.136437

Hastie, T., Tibshirani, R., and Wainwright, M. (2019). *Statistical Learning with Sparsity: The Lasso and Generalizations.* Boca Raton, FL: Chapman and Hall/CRC.

Hu, M., Li, R., You, W., Liu, Y., and Lee, C.-C. (2020). Spatiotemporal Evolution of Decoupling and Driving Forces of CO2 Emissions on Economic Growth along the Belt and Road. *J. Clean. Prod.* 277, 123272. doi:10.1016/j.jclepro.2020.123272

Jones, C. I. (1995). R & D-Based Models of Economic Growth. *J. Polit. Economy* 103, 759–784. doi:10.1086/262002

Kahouli, B. (2018). The Causality Link between Energy Electricity Consumption, CO2 Emissions, R&D Stocks and Economic Growth in Mediterranean Countries (MCs). *Energy* 145, 388–399. doi:10.1016/j.energy.2017.12.136

Khan, M., and Rana, A. T. (2021). Institutional Quality and CO2 Emission-Output Relations: The Case of Asian Countries. *J. Environ. Manage.* 279, 111569. doi:10.1016/j.jenvman.2020.111569

Le, T.-H., Le, H.-C., and Taghizadeh-Hesary, F. (2020). Does Financial Inclusion Impact CO2 Emissions? Evidence from Asia. *Finance Res. Lett.* 34, 101451. doi:10.1016/j.frl.2020.101451

Li, R., Wang, Q., Liu, Y., and Jiang, R. (2021). Per-capita Carbon Emissions in 147 Countries: The Effect of Economic, Energy, Social, and Trade Structural Changes. *Sustainable Prod. Consumption* 27, 1149–1164. doi:10.1016/j.spc.2021.02.031

Liu, L.-C., Wu, G., Wang, J.-N., and Wei, Y.-M. (2011). China's Carbon Emissions from Urban and Rural Households during 1992-2007. *J. Clean. Prod.* 19 (15), 1754–1762. doi:10.1016/j.jclepro.2011.06.011

Liu, Y., and Hao, Y. (2018). The Dynamic Links between CO2 Emissions, Energy Consumption and Economic Development in the Countries along "the Belt and Road". *Sci. total Environ.* 645, 674–683. doi:10.1016/j.scitotenv.2018.07.062

Luo, Y., Zeng, W., Hu, X., Yang, H., and Shao, L. (2021). Coupling the Driving Forces of Urban CO2 Emission in Shanghai with Logarithmic Mean Divisia index Method and Granger Causality Inference. *J. Clean. Prod.* 298, 126843. doi:10.1016/j.jclepro.2021.126843

Ma, N., Wu, Y., and Wu, J. (2018). Environmental Efficiency and its Distribution Dynamics in Chinese Cities. *J. Chin. Econ. Business Stud.* 16 (4), 417–445. doi:10.1080/14765284.2018.1512817

Magazzino, C., Mele, M., and Schneider, N. (2021). A Machine Learning Approach on the Relationship Among Solar and Wind Energy Production, Coal Consumption, GDP, and CO2 Emissions. *Renew. Energ.* 167, 99–115. doi:10.1016/j.renene.2020.11.050

Mohmand, Y. T., Mehmood, F., Mughal, K. S., and Aslam, F. (2020). Investigating the Causal Relationship between Transport Infrastructure, Economic Growth and Transport Emissions in Pakistan. *Res. Transportation Econ.*, 100972. doi:10.1016/j.retrec.2020.100972

Muhammad, B., and Khan, S. (2019). Effect of Bilateral FDI, Energy Consumption, CO2 Emission and Capital on Economic Growth of Asia Countries. *Energ. Rep.* 5, 1305–1315. doi:10.1016/j.egyr.2019.09.004

Munir, Q., Lean, H. H., and Smyth, R. (2020). CO2 Emissions, Energy Consumption and Economic Growth in the ASEAN-5 Countries: A Cross-Sectional Dependence Approach. *Energ. Econ.* 85, 104571. doi:10.1016/j.eneco.2019.104571

Nasir, M. A., Canh, N. P., and Lan Le, T. N. (2021). Environmental Degradation & Role of Financialisation, Economic Development, Industrialisation and Trade Liberalisation. *J. Environ. Manage.* 277, 111471. doi:10.1016/j.jenvman.2020.111471

Nguyen, T. T., Pham, T. A. T., and Tram, H. T. X. (2020). Role of Information and Communication Technologies and Innovation in Driving Carbon Emissions and Economic Growth in Selected G-20 Countries. *J. Environ. Manag.* 261, 110162. doi:10.1016/j.jenvman.2020.110162

Shahbaz, M., Hye, Q. M. A., Tiwari, A. K., and Leitão, N. C. (2013). Economic Growth, Energy Consumption, Financial Development, International Trade and CO2 Emissions in Indonesia. *Renew. Sust. Energ. Rev.* 25, 109–121. doi:10.1016/j.rser.2013.04.009

Shahbaz, M., Mahalik, M. K., Shah, S. H., and Sato, J. R. (2016). Time-varying Analysis of CO2 Emissions, Energy Consumption, and Economic Growth Nexus: Statistical Experience in Next 11 Countries. *Energy Policy* 98, 33–48. doi:10.1016/j.enpol.2016.08.011

Shahbaz, M., Sharma, R., Sinha, A., and Jiao, Z. (2021). Analyzing Nonlinear Impact of Economic Growth Drivers on CO2 Emissions: Designing an SDG Framework for India. *Energy Policy* 148, 111965. doi:10.1016/j.enpol.2020.111965

Shahbaz, M., Solarin, S. A., Mahmood, H., and Arouri, M. (2013). Does Financial Development Reduce CO2 Emissions in Malaysian Economy? A Time Series Analysis. *Econ. Model.* 35, 145–152. doi:10.1016/j.econmod.2013.06.037

Shi, X., Wang, K., Cheong, T. S., and Zhang, H. (2020). Prioritizing Driving Factors of Household Carbon Emissions: An Application of the LASSO Model with Survey Data. *Energ. Econ.* 92, 104942. doi:10.1016/j.eneco.2020.104942

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Vo, D. H., Vo, A. T., Ho, C. M., and Nguyen, H. M. (2020). The Role of Renewable Energy, Alternative and Nuclear Energy in Mitigating Carbon Emissions in the CPTPP Countries. *Renew. Energ.* 161, 278–292. doi:10.1016/j.renene.2020.07.093

Wang, R., Mirza, N., Vasbieva, D. G., Abbas, Q., and Xiong, D. (2020). The Nexus of Carbon Emissions, Financial Development, Renewable Energy Consumption, and Technological Innovation: What Should Be the Priorities in Light of COP 21 Agreements? *J. Environ. Manage.* 271, 111027. doi:10.1016/j.jenvman.2020.111027

Zaidi, S. A. H., Zafar, M. W., Shahbaz, M., and Hou, F. (2019). Dynamic Linkages between Globalization, Financial Development and Carbon Emissions: Evidence from Asia Pacific Economic Cooperation Countries. *J. Clean. Prod.* 228, 533–543. doi:10.1016/j.jclepro.2019.04.210

Zaman, K., and Moemen, M. A.-e. (2017). Energy Consumption, Carbon Dioxide Emissions and Economic Development: Evaluating Alternative and Plausible Environmental Hypothesis for Sustainable Growth. *Renew. Sust. Energ. Rev.* 74, 1119–1130. doi:10.1016/j.rser.2017.02.072

Zhang, J., Fan, Z., Chen, Y., Gao, J., and Liu, W. (2020). Decomposition and Decoupling Analysis of Carbon Dioxide Emissions from Economic Growth in the Context of China and the ASEAN Countries. *Sci. Total Environ.* 714, 136649. doi:10.1016/j.scitotenv.2020.136649

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Shum, Ma, Lin and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 06 August 2021
doi: 10.3389/fenvs.2021.715149

# Time-Varying Impact of Economic Growth on Carbon Emission in BRICS Countries: New Evidence From Wavelet Analysis

*Lijin Xiang*, Xiao Chen, Shuling Su and Zhichao Yin*

*School of Finance, Shandong University of Finance and Economics, Jinan, China*

Carbon emission leads to environmental and social consequences, which could be severe in the emerging economies. Owing to the dilemma of emission and economic expansion, it is necessary to achieve a more comprehensive understanding of the dynamic relationship between economic growth and carbon emission. Multivariate Wavelet analysis is introduced in addition to the decoupling analysis for BRICS countries. The decoupling analysis detects an obvious trend of economic growth decoupling from carbon emission in China, and generates mixed results for the other countries. Estimates of wavelet coherency suggest that BRICS countries have experienced different kinds of structural changes in growth–emission nexus. Results of partial phase-difference and wavelet gain imply that different resource endowments and growth paths lead to varied impact of economic growth on carbon emission and time-varying characteristics of the causality relationship over different frequencies. Energy structure and trade openness matter for anatomizing this time-varying relationship. To succeed in the fight against climate change, the policy makers need to pay serious attention to the dynamic impact of economic growth, energy structure, and trade openness on carbon emission.

**Keywords: economic growth, carbon emission, time-varying effect, wavelet analysis, decoupling**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Partha Gangopadhyay,*
*Western Sydney University, Australia*
*Muntasir Murshed,*
*North South University, Bangladesh*

****Correspondence:***
*Lijin Xiang*
*xianglijin@sdufe.edu.cn*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *26 May 2021*
***Accepted:*** *07 July 2021*
***Published:*** *06 August 2021*

***Citation:***
*Xiang L, Chen X, Su S and Yin Z (2021) Time-Varying Impact of Economic Growth on Carbon Emission in BRICS Countries: New Evidence From Wavelet Analysis. Front. Environ. Sci. 9:715149. doi: 10.3389/fenvs.2021.715149*

## INTRODUCTION

The climate change has been deemed as one of the most urgent issues that are confronted by the humankind. It results in far-reaching and complex socioeconomic consequences (Dai et al., 2016; Dottori et al., 2018; Pinsky et al., 2019). So far, a global consensus has been reached that the greenhouse gas, especially CO2, causes climate change. Facing the threat of global warming to economic prosperity and sustainability, the carbon emission reduction measures have been proposed as a response.

World economy is fueled largely by the fossil energy sources. The dilemma of environmental degradation and economic growth needs to be handled to optimize the human welfare. The decoupling indexes reflect a variety of different coupling states of carbon emission and economic growth in the literature (Gao et al., 2015). Although this approach provides real-time measure for the relationship between carbon emission and economic growth, the influence from other important factors is left out of consideration. As a result, the decoupling analysis fails to reveal the time-varying real effect of economic growth on carbon emission and to identify the evolving impact from other crucial factors.

Besides the studies that have investigated the relationship between economic growth and carbon emission, much attention has been paid on the impact of energy transition and international trade on

emission emerges. That is because not only the quantity of production output (GDP), but also the change in energy sources and production distribution worldwide, which can be proxied by energy structure and trade openness respectively, pose a significant impact on carbon emission. Energy is a critical input to economic development and an essential part of human activity. The current world energy consumption is mainly based on fossil fuel energy. Thus, the research on energy transition is of practical importance (Wang and Wang, 2020). In the process of globalization, there is no doubt that the change in one country's level of trade openness will reshape its production distribution and industrial emission structure (Wang and Zhang, 2021). Therefore, it is necessary to incorporate variables that represent both the economic output itself and the way to achieve it into the model.

The BRICS countries (Brazil, Russia, India, China, and South Africa) are experiencing drastic economic transition and structural change. The shifts in their energy structure and level of trade openness are profound and influencing, which makes them excellent samples for empirical investigation. This article attaches importance to the effects of energy structure and trade openness on carbon emission along with the economic growth. Based on results from decoupling analysis, the continuous wavelet analysis framework is applied to dig deeper into the complex relationships among carbon emission and economic growth.

The contribution of this article to the extant literature is as follows: 1) the application of wavelet transform allows us to estimate partial phase-difference to depict both the short-term (high-frequency) and long-term (low-frequency) causality between economic growth and carbon emission. The relationship changes over time and across frequencies. Our analysis advances the analysis of the decoupling states to the identification of causality; 2) the wavelet analysis framework is extended to estimate the magnitude of marginal impact of economic growth on carbon emission by introducing the (partial) wavelet gain. Compared with previous studies, the interference from other factors is suppressed. And the applied wavelet model generates much richer empirical results including correlation, causality, and statistics that measure magnitude of marginal impact; 3) The physical inputs, including capital and labor, and the proxy of economic development have been introduced into the estimation to examine whether the newly developed wavelet analysis approach is able to overcome the omitted variable bias documented in the literature; 4) The wavelet analysis approach captures the nonlinear dynamics among economic variables. In addition to the decoupling states, more dynamic characteristics and evolving patterns can be inferred from our estimated statistics such that the conclusions drawn from the empirical results provide a deeper understanding of the time-varying impact of economic growth on carbon emission.

The remainder of the article is organized as follows. *Literature Review* Section reviews the relevant literature and presents the motivation and contribution. In *Methodology* Section the methodology is introduced in detail. *Data and Empirical Results* Section presents the data and empirical result. Then in the last section, we draw conclusion and provide corresponding policy implication.

## LITERATURE REVIEW

Along with economic development, environmental degradation problems gradually emerge. Many researchers believe that there is a close relationship between economic development and carbon emissions (e.g., Selden and Song, 1994; Galeotti et al., 2009; Saboori et al., 2012). Extensive studies have begun to focus on the relationship between GDP growth and carbon dioxide emissions based on testing the effectiveness of environmental Kuznets curve (EKC) in different countries and regions (Apergis, 2016; Apergis et al., 2017; Murshed, 2020; Murshed et al., 2021a; Murshed et al., 2021b).

Scholars draw different conclusions on the relationship between GDP growth and energy consumption using Granger causality. The results from the Granger causality tests suggest that in the long run there is unidirectional Granger causality running from electricity consumption and emissions to economic growth (Lean and Smyth, 2010). Apergis and Tang (2013) reexamine the validity of the energy-led growth hypothesis using different model specifications. Compared with the Granger causality model containing only two variables, the Granger causality model containing three and four variables is more likely to support the assumption. In addition, both developed and developing countries are more likely to support the energy-led growth assumption than less developed or low-income countries.

There are many factors affecting carbon emissions. Scholars have conducted different analyses and found some major factors. Banerjee and Murshed (2020) confirmed the pollution paradise hypothesis by examining cross-sectional dependencies between national teams in 2005–2015 and determining the long-term equilibrium relationship between net export emissions and real GDP, FDI, trade openness, energy consumption, and financial development. Murshed (2020) drew a conclusion that in South Asia ICT trade reduces CO2 emissions by directly increasing renewable energy consumption, increasing the share of renewable energy, reducing energy intensity, promoting cleaner cooking fuels, indirectly increasing the level of renewable energy consumption, improving energy efficiency, and strengthening cleaner cooking fuel channels. In addition, environmental policy is also an important influencing factor. From an environmental perspective there is potential to use pricing policies in the G7 countries to curtail residential electricity demand, and thus curb carbon emissions, in the long run (Narayan et al., 2007).

In the case of China, Ma et al. (2021) find that provincial growth and the development of the tertiary industry are the reasons for the deterioration of China's carbon dioxide emission trend. The survey results also show that emission taxes, investment in research and development, technological innovation, and the use of renewable energy together further reduce carbon dioxide emission. Abdul et al. (2021) indicated that the positive impact on grain crop production will only increase carbon dioxide emissions in the long term, the impact on forestry will not have any significant impact on China's carbon dioxide emissions level, and the negative impact on livestock production will only increase carbon dioxide emissions in the short term.

## Decoupling of Economic Growth and Carbon Emission

The decoupling analysis framework was proposed to confirm the disconnection between economic growth and environmental degradation (Tapio, 2005). It has been widely applied to various levels (Wang and Zhang, 2020). Compared with the Environmental Kuznets Curve framework, decoupling analysis is straightforward and able to reveal the real-time relationship in different horizons (Dong et al., 2020). The rationale of decoupling analysis is to dissociate economic growth from environmental degradation to achieve sustainable development (Zhang and Zhang, 2020). It provides a way to perform decomposition on the complex relationship between two variables. Many scholars address the dilemma of economic growth and environmental degradation using the decoupling framework and make significant progresses (Csereklyei and Stern, 2015; Kovanda and Hak, 2007; Loo and Banister, 2016).

Dai et al. (2016) identified decoupling state of BRICS countries and further explored driving factors on decoupling. Some scholars conduct decomposition and decoupling analysis of carbon dioxide emissions from economic growth in China and the ASEAN countries (Zhang et al., 2020). Another study took the Belt and Road as the starting point to analyze spatiotem poral evolution of decoupling and driving forces of CO2 emissions on economic growth (Hu et al., 2020). Adedoyin et al. (2020) introduced coal rents vs coal consumption and other energy sources as determinants of CO2 emissions in BRICS economies. Wang and Zhang (2020) studied the influence of increasing investment in research and development on economic growth decoupling from carbon emission. Wang and Wang (2020) discovered the promoting effect of energy transition on the decoupling economic growth from emission. Nandini and Joyashree (2020) decomposed energy-related CO2 emissions, using LMDI approach, to quantify drivers of climate change. And decoupling elasticity is estimated to identify historical attainment in decoupling economic growth from emissions in India.

## Energy Structure and Carbon Emission

Energy is essential material input for production. Economic growth cannot sustain without enough energy consumption. Hence, intuitively, energy consumption should maintain a long-term equilibrium with economic growth (Zhang, 2011). Adedoyin et al. (2020) find that the BRICS countries are heavily dependent on energy-intensive sectors such as construction, mining, and manufacturing, because of a rapid increase in population, lifestyle changes, and urbanization. It is a natural choice to achieve emission reduction by improving energy efficiency. Guan et al. (2018) find that the most important factor in reducing carbon emissions in China is the improvement of energy efficiency, during the sample period from 2007 to 2010.

Among the factors affecting carbon emissions, energy consumption is one of the most important factors and energy consumption has different effects in countries with different development status. The results show that the United States strongly supports the hypothesis of neutrality. While a developing economy panel (90 countries) favors the conservative hypothesis, and a panel of 32 lower middle-income countries suggests that energy consumption per capita predicts real GDP per capita (Narayan, 2016). Narayan and Smyth (2007) suggest that the demand for oil in the Middle East is being driven largely by strong economic growth. In the near future, Malaysia's energy import dependency will rise. Carbon emissions will triple by 2030 (Gan and Li, 2008). New energy and renewable energy have become the focus in the energy consumption–growth relationship research (Apergis and Payne, 2010a; Apergis and Payne, 2010b; Apergis et al., 2010; Apergis and Payne, 2011a; Apergis and Payne, 2011b; Apergis and Payne, 2012; Murshed et al., 2021). Thus, it is necessary to incorporate an energy structure in the empirical study of growth–carbon emission relationship.

When the marginal carbon reduction effect of energy efficiency improvement gradually reaches its ceiling, renewable energy turns out to be a more promising alternative. Energy transition has caught much attention and has been regarded as an efficient approach to reduce carbon emission and achieve low-carbon development (Obama, 2017; Wang and Su, 2020). International Energy Agency (IEA) carried out the Clean Energy Transitions Program (CETP) to facilitate global energy transition[1]. IEA demonstrated that energy transition can enhance energy security, robust energy system at the same time thrive economy in its annual report2018[2].

For the developing countries, scholars find that increasing the share of renewable energy use could reduce emission (Liu et al., 2017; Hu et al., 2018). The recent research found that renewable energy decreases CO2 emissions in African countries (Dauda et al., 2021). Fragkos et al. (2021) investigate Energy system transitions and low-carbon pathways. Their results imply the technical-economic feasibility of achieving large emission reductions by 2050.

## Trade Openness and Carbon Emission

Trade liberalization has significantly reshaped industries and production network worldwide. Since the seminal work by Grossman and Krueger (1991), many scholars have attempted to examine the effect of trade openness on the environment.

Since countries experience different levels of income and foreign trade depending on their level of economic development, the relationship between emissions and trade depends on where an economy is currently placed in its development trajectory (Baek et al., 2009). Antweiler et al. (2001) have systematically described three major categories of the impact on the environment, namely the scale, technique, and composition effects. They point out that the technique effect overshadows the scale effect.

The debate regarding the relationship between trade openness and environmental degradation has not reached a consensus yet. The findings based on the different methodology and data are quite discrepant (Cole and Elliott, 2003; Frankel and Rose, 2005;

[1]https://www.iea.org/cetp/.

[2]Clean Energy Transitions Program (CETP) Annual Report 2018.

Managi et al., 2008). Copeland and Taylor (2003), Tiba et al. (2015), and Tiba and Frikha (2018), among others, revealed the existence of a significant positive impact of trade openness on environmental quality. In constant, Salman et al. (2019) found that export increases CO2 emissions in some countries in Asia after disaggregating trade into export and import. The relationship between trade openness and carbon emissions still merits further investigation (Wang and Zhang, 2021).

This article uses data from the BRICS countries to explore the dynamic relationship between carbon emission and economic growth. Decoupling approach is introduced to reflect the states of growth–emission nexus. And based on that, the wavelet analysis is applied to provide deeper insight and extensive discussion on the impact of economic growth on carbon emission.

# METHODOLOGY

## Tapio Decoupling Model

In this article, we adopt the decoupling approach that is being widely used in the literature to provide a portrait for the complex growth–emission relationship. The OECD decoupling index (de Freitas and Kaneko, 2011; Yu et al., 2013; Zhao et al., 2017), the Tapio decoupling model (Climent and Pardo, 2007; Dong et al., 2016; Kan et al., 2019; Roman-Collado et al., 2018; Sorrell et al., 2012; Tapio, 2005; Zhang et al., 2015), the IGTX decoupling method (Ma et al., 2016), and the EA method (Enevoldsen et al., 2007; Mielnik and Goldemberg, 2002) are four alternative approaches for analyzing economic growth decoupling from environmental pollution. Since the Tapio (2005) method is most widely applied in the literature, we follow this approach to define the decoupling elasticity as:

$$e(C) = \frac{\Delta C/C_0}{\Delta G/G_0} \tag{1}$$

where $e(C)$ indicates the decoupling elasticity coefficient between economic activities and carbon emissions, $\Delta C$ and $\Delta G$ respectively indicate the total carbon emission change and the total GDP change from the base period to the end period. $C_0$ indicates the carbon emissions at the base period, and $G_0$ indicates the base period GDP.

As shown in **Figure 1**, in the first quadrant, both carbon emission and GDP increase simultaneously, represent expansive negative decoupling; $0.8 \le e \le 1.2$ represents expansive coupling; $0 \le e < 0.8$ represents weak decoupling. While in the third quadrant, both carbon emission and GDP decrease simultaneously, $e > 1.2$ represents recessive decoupling; $0.8 \le e \le 1.2$ represents recessive coupling; $0 \le e < 0.8$ represents weak negative decoupling. In the second and fourth quadrants, $e < 0$ represent that carbon emission and GDP change asynchronously. In the fourth quadrant, there is strong decoupling, the best decoupling state, indicating economy increasing while carbon emission decreasing. In the second quadrant, it is strong negative decoupling, the worst decoupling state, showing economy decreasing while carbon emission increasing.

## Wavelet Model

To carry out further in-depth research, we introduce the wavelet analysis framework. The wavelet model is an extension of spectral analysis (Mandler and Scharnagl, 2014). Wavelet analysis can reveal the frequency components of variables just like the Fourier transform, in addition to extraction of series characteristics in the time domain. It is a sophisticated time-frequency analysis technique that outperforms pure time series or frequency domain methods. As a result, it has been widely used for data processing and econometric analysis. While the Fourier transform breaks down a time series into constituent sinusoids of different frequencies and infinite duration in time, the wavelet transform expands the time series into shifted and scaled versions of a function, the so-called mother wavelet, that has limited spectral band and limited duration in time (Aguiar-Conraria et al., 2018).

### Continuous Wavelet Transform

Discrete wavelet transform (DWT) and continuous wavelet transform (CWT) are two classes of wavelet transform. The DWT is useful for noise reduction and data compression, while the CWT is more helpful for feature extraction and data self-similarity detection (Grinsted et al., 2004; Loh, 2013). In application, continuous wavelet transformation is often chosen to extract the characteristics of economic variables and perform correlation and causality analysis. This approach utilizes information from two dimensions, namely time and frequency. Thus, it is suffered less estimation bias other than empirical methods developed to identify economic relationship solely in the time domain.

Assuming the mother wavelet function is $\psi(t)$, the corresponding base wavelet function is

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{2}$$

where $\tau$ is the translation factor, which represents the magnitude of the mother wavelet function's translation, and different values will produce different window positions. $s$ is the expansion factor, which represents the degree of expansion of the mother wavelet function. When the $s$ is less than 1, the function is compressed, which is beneficial to capture the high-frequency part of original sequence, and when the $s$ is greater than 1, the function is stretched, which is beneficial to capture the low-frequency part. By changing the two parameters, one can construct a picture showing how the amplitudes of $x(t)$ vary across scales and how such amplitudes change over time (Torrence and Compo, 1998).

A continuous wavelet transformation can be expressed as follows:

$$W_x(\tau, s) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-\tau}{s}\right)d_t \tag{3}$$

Because of the need for both amplitude and phase information, it is necessary to select complex wavelet, and the Morlet wavelet satisfies this condition. Based on the Morlet continuous wavelet

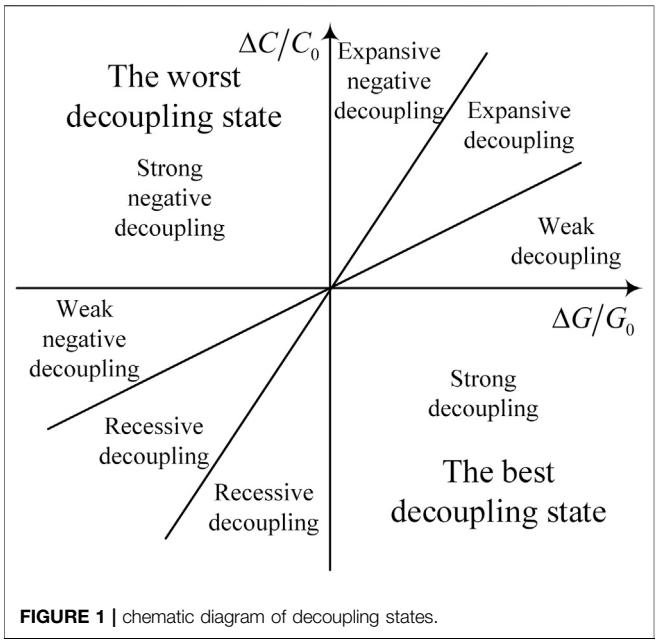



**FIGURE 1 |** chematic diagram of decoupling states.

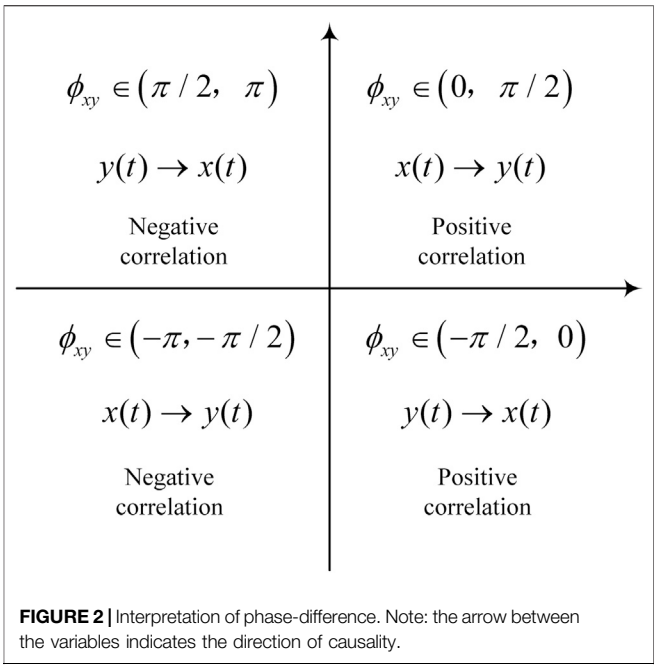



**FIGURE 2 |** Interpretation of phase-difference. Note: the arrow between the variables indicates the direction of causality.

transform, we can calculate (partial) wavelet coherency, (partial) phase-difference, and (partial) wavelet gain to analyze the correlation and causality relationship between two or more time series and the magnitude of the impact.

## Wavelet Coherency and Phase-Difference

In analogy with the terminology used in the Fourier case, the local wavelet power spectrum of the time series is defined as $|W_x(\tau, s)|^2$, which measures the magnitude of the time series fluctuation in the time-frequency space. The wavelet power spectrum may be averaged over time for comparison with classical spectral methods. When the average is taken over all times, we obtain the global wavelet power spectrum as follows:

$$GWPS_x(s) = \int_{-\infty}^{\infty} |W_x(\tau, s)|^2 d\tau \tag{4}$$

For two time series $x(t)$ and $y(t)$, we can obtain the cross-wavelet transform corresponding to$(\tau, s)$:

$$W_{xy}(\tau, s) = W_x(\tau, s) W_y^*(\tau, s) \tag{5}$$

And the corresponding cross-wavelet power spectrum is as follows:

$$|W_{xy}(\tau, s)|^2 = |W_x(\tau, s)|^2 |W_y^*(\tau, s)|^2 \tag{6}$$

where $W_y^*(\tau, s)$ represents the complex conjugate of the continuous wavelet transform function.

Following Torrence & Webster (1999), the Wavelet Coherence of two time series is defined as:

$$R_n^2(s) = \frac{|S(s^{-1} W_n^{XY}(s))|^2}{S(s^{-1}|W_n^X(s)|^2) \cdot S(s^{-1}|W_n^Y(s)|^2)} \tag{7}$$

where$S$is the smoothing operator.

According to Bloomfield et al. (2004), the wavelet phase-difference enables us to calculate the phase of the wavelet transform of each time series. And wavelet phase-difference is defined as follows:

$$\phi_n(s) = \arctan\left(\frac{\mathfrak{I}\{S(s^{-1} W_n^{XY}(s))\}}{\mathfrak{R}\{S(s^{-1} W_n^{XY}(s))\}}\right) \tag{8}$$

where $\mathfrak{I}$ and $\mathfrak{R}$ are the imaginary and real parts of the smoothed cross-wavelet transform, respectively. Using the phase-difference, the time lag or time difference, which gives the lead or lag of the series in the time domain, can be calculated as $\Delta T_n(s) = \varphi_n(s)/\omega$, where $\omega$ is the angular frequency with respect to the time scale.

When the phase-difference falls in different quadrants, it represents different types of correlations and causality relationships between variables, as shown in **Figure 2**. As referred to partial phase-difference, it indicates the time-varying correlations and causality relationships between two variables after controlling other factors. By its nature, this indicator can precisely capture the sign of correlation and the direction of causality over time and across different frequencies.

## Partial Phase-Difference and Partial Wavelet Gain

Partial phase-difference is introduced to incorporate energy transition and trade openness into the estimation of dynamic causality relationship between economic growth and carbon emission. And partial wavelet gain can be estimated to identify the real effect of economic growth on carbon emission, after controlling for influence from other factors.

The smoothed cross-wavelet transform is denoted as $S_{ij} = S(W_{ij})$. We can take these smoothed wavelet transforms as elements to construct a matrix $\mathbb{S}$ *and* $\mathbb{S}_{ij}^d$ denotes the cofactor of the element in position $(i, j)$. The multivariate complex wavelet coherency for $x_1$ and $x_2, \cdots, x_p$ is given by:

$$R^c_{(1,\mathbb{K}^1)} = 1 - \frac{|\mathbb{S}|}{S_{11}\mathbb{S}^d_{11}} \tag{9}$$

where $\mathbb{k}^i = \{1, 2, \cdots, p\}\backslash\{i\}$.

The complex partial wavelet coherency for $x_1$ and $x_i (2 \le i \le p)$ is calculated according to the following equation:

$$R^c_{(1,i \mid \mathbb{k}^i)} = -\frac{\mathbb{S}^d_{i1}}{\sqrt{\mathbb{S}^d_{11}\mathbb{S}^d_{ii}}} \tag{10}$$

Taking the modulus of complex coherency, the partial wavelet coherency $R_{(1,i \mid \mathbb{k}^i)} = |R^c_{(1,i \mid \mathbb{k}^i)}|$ is obtained.

Similar to the definition of the wavelet phase-difference in the case of two variables, we can define the partial wavelet phase-difference between $x_1$ and $x_i (2 \le i \le p)$ in the multivariate case after controlling the influence of other variables:

$$\phi_{(1,i \mid \mathbb{k}^i)} = \tan^{-1}\left(\frac{\mathfrak{T}\left(\rho^c_{(1,i \mid \mathbb{k}^i)}\right)}{\mathfrak{R}\left(\rho^c_{(1,i \mid \mathbb{k}^i)}\right)}\right) \tag{11}$$

where $\mathfrak{T}(\rho^c_{(1,i \mid \mathbb{k}^i)})$ and $\mathfrak{R}(\rho^c_{(1,i \mid \mathbb{k}^i)})$ represent the imaginary and real parts of the complex wavelet coherency $\rho^c_{(1,i \mid \mathbb{k}^i)}$. The cross wavelet gain is defined as follows:

$$G_{y,x}(s) = \frac{\left|S\left(s^{-1}W_{xy}(s)\right)\right|}{S\left(s^{-1}|W_x|^2\right)} \tag{12}$$

which can be interpreted as the absolute value of regression coefficient in the regression of $y$ on $x$. Taking account of other variables, the formula for the partial wavelet gain is given by

$$G_{(1,i \mid \mathbb{k}^i)} = \frac{\left|\mathbb{S}^d_{i1}\right|}{\mathbb{S}^d_{11}} \tag{13}$$

$G_{(1,i \mid \mathbb{k}^i)}$ can be interpreted as the absolute value of coefficients (in modulus) of multivariate regression of $x_1$ on $x_2, ..., x_p$ at each point of time and frequency. According to Aguiar-Conraria et al. (2018), wavelet gain should be interpreted simultaneously with the phase-difference, which reveals the sign of correlation and the direction of causality.

# DATA AND EMPIRICAL RESULTS

## Data

The data used in this study are obtained from three sources, including the World Bank WDI database[3], the Global Carbon Project[4], and the BP Statistical Review of World Energy[5]. The definition of variables and data sources are presented in **Table 1**. The decoupling analysis is based on the level series of emission and GDP, while the wavelet analysis is conducted after taking a first-order difference of these two variables. In the robustness examination, we also take the first-order difference of labor force.

**TABLE 1 |** Data sources and variable definition.

| Variable | Definition | Source |
|---|---|---|
| GDP | Gross domestic production (in constant 2010 US$ billions) | World Bank WDI database |
| Carbon emission | CO2 emissions (million tonnes) | Global carbon project |
| Energy structure | Renewable energy ratio (% of total primary energy consumption) | BP statistical review of world energy |
| Trade openness | Total trade value as % of GDP | World Bank WDI database |
| Capital formation | Capital investment as % of GDP | World Bank WDI database |
| Labor force | Total labor force (million) | World Bank WDI database |
| Industrialization | Industry value added growth rate (%) | World Bank WDI database |

*Note: Primary energy is calculated using the "substitution method" that takes account of the inefficiencies energy production from fossil fuels.*

Compared to previous studies, we gather emission and energy structure data from additional sources other than the WDI database to obtain original real GDP and CO2 emission series. Inputs such as capital and labor are also included in the estimation to perform a robustness check for omitted variable. And we use industry value added to proxy the level of economic development. In this way, we are able to alleviate concerns about the impact of economic development on the growth–emission relationship. The descriptive statistics are provided in **Table 2**. There exists significant heterogeneity in country-specific carbon emission and energy structure among the BRICS countries. It is natural that the relationship between emission and economic growth will vary across these countries.

## Decoupling Analysis

Following the Tapio approach, we calculate the decoupling elastic coefficients between emission and economic growth during the period of 1960–2019 for the BRICS countries, as shown in **Table 3**. The decoupling states are marked by different colors. As it can be observed, the decoupling indexes of different countries in different periods show great differences.

In Brazil, the expansive negative decoupling state appears the most frequently during the period of 1960–2019, which demonstrates that the growth rate of carbon emission is higher than that of economic development. The expected strong decoupling state appears during 1964–1965, 1979–1980, 2001–2003, 2005–2006, and 2017–2019, showing that economy grew while carbon emission decreasing at the same time, and the state of strong decoupling and weak decoupling appears more frequently in the period of 2000–2019. Evaluations of the period 2004 to 2009 are compared with the period 1980 to 1994 when Brazil also experienced an apparent decoupling (Freitas and Kaneko, 2011).

For Russia, we focus on analyzing the results after 1989. The strong decoupling state and weak decoupling state are two of the most frequent decoupling states, especially from 1998 to 2019. There also exist recessive coupling in 2008–2009, expansive

[3]World Development Indicators (WDI) database from World Bank, available at: http://wdi.worldbank.org.

[4]Global Carbon Project. (2020). Supplemental data of Global Carbon Budget 2020 (Version 1.0) (Data set). Available at: https://doi.org/10.18160/gcp-2020.

[5]BP Statistical Review of World Energy. Available at: http://www.bp.com/statisticalreview.

**TABLE 2 |** Descriptive statistics.

| Variables | | Mean | Sd | min | Median | Max | Sample period |
|---|---|---|---|---|---|---|---|
| Brazil | GDP | 1,252.93 | 680.90 | 246.66 | 1,195.94 | 2,423.27 | 1960–2019 |
| | Carbon emission | 243.06 | 138.20 | 46.85 | 209.84 | 523.89 | 1960–2019 |
| | Energy structure | 36.77 | 6.88 | 23.90 | 38.63 | 45.02 | 1965–2019 |
| | Trade openness | 19.54 | 5.32 | 9.06 | 19.13 | 29.68 | 1960–2019 |
| | Capital formation | 18.61 | 2.30 | 14.63 | 18.16 | 22.99 | 1990–2019 |
| | Labor force | 86.30 | 14.00 | 61.50 | 89.70 | 107.00 | 1990–2019 |
| | Industrialization | 1.40 | 4.36 | −5.76 | 0.54 | 10.17 | 1990–2019 |
| Russia | GDP | 1,316.68 | 323.06 | 813.03 | 1,386.61 | 1762.46 | 1989–2019 |
| | Carbon emission | 1,690.87 | 390.28 | 885.86 | 1,620.76 | 2,525.29 | 1960–2019 |
| | Energy structure | 5.63 | 0.65 | 4.34 | 5.81 | 6.59 | 1985–2019 |
| | Trade openness | 53.65 | 13.75 | 26.26 | 50.95 | 110.58 | 1989–2019 |
| | Capital formation | 23.02 | 4.44 | 14.83 | 22.62 | 36.27 | 1990–2019 |
| | Labor force | 73.80 | 1.96 | 68.90 | 74.10 | 76.00 | 1990–2019 |
| | Industrialization | 0.05 | 8.26 | −21.60 | 1.81 | 12.14 | 1990–2019 |
| India | GDP | 812.65 | 761.03 | 148.77 | 494.26 | 2,940.16 | 1960–2019 |
| | Carbon emission | 820.20 | 733.92 | 111.49 | 559.99 | 2,616.45 | 1960–2019 |
| | Energy structure | 8.41 | 2.25 | 5.03 | 7.77 | 14.05 | 1965–2019 |
| | Trade openness | 23.39 | 15.18 | 7.66 | 15.45 | 55.79 | 1960–2019 |
| | Capital formation | 31.82 | 5.56 | 23.97 | 30.17 | 41.93 | 1990–2019 |
| | Labor force | 425.00 | 53.40 | 324.00 | 449.00 | 495.00 | 1990–2019 |
| | Industrialization | 6.22 | 3.00 | 0.59 | 6.01 | 13.24 | 1990–2019 |
| China | GDP | 2,439.28 | 3,184.53 | 87.93 | 812.26 | 11,537.48 | 1960–2019 |
| | Carbon emission | 3,594.74 | 3,231.24 | 432.22 | 2,403.84 | 10,174.68 | 1960–2019 |
| | Energy structure | 5.29 | 2.46 | 2.60 | 4.45 | 12.67 | 1965–2019 |
| | Trade openness | 27.12 | 17.92 | 4.92 | 25.53 | 64.48 | 1960–2019 |
| | Capital formation | 40.73 | 4.11 | 33.57 | 40.48 | 46.66 | 1990–2019 |
| | Labor force | 746.00 | 42.30 | 650.00 | 764.00 | 787.00 | 1990–2019 |
| | Industrialization | 10.83 | 4.10 | 4.87 | 10.29 | 21.03 | 1990–2019 |
| South Africa | GDP | 238.63 | 103.87 | 79.07 | 218.58 | 430.17 | 1960–2019 |
| | Carbon emission | 308.30 | 128.17 | 97.84 | 327.38 | 502.26 | 1960–2019 |
| | Energy structure | 0.53 | 0.49 | 0.01 | 0.43 | 2.21 | 1965–2019 |
| | Trade openness | 52.36 | 7.14 | 37.49 | 51.72 | 72.87 | 1960–2019 |
| | Capital formation | 18.66 | 1.85 | 15.16 | 18.47 | 23.15 | 1990–2019 |
| | Labor force | 18.00 | 3.00 | 12.80 | 18.30 | 23.30 | 1990–2019 |
| | Industrialization | 1.08 | 2.40 | -6.00 | 1.36 | 4.72 | 1990–2019 |

negative decoupling in 2009–2010, strong negative decoupling in 2014–2015, and expansive coupling in 2016–2018.

The three major types of the decoupling state that appeared in India each account for about 30% of the sample during the whole period: weak decoupling, expansive coupling, and expansive negative decoupling. And we observed that weak decoupling has become a trend since 2000 in India. The expected strong decoupling state appears in 1963–1964, and strong negative decoupling appears in 1964–1966 and 1971–1972.

Weak decoupling state appears the most frequently in China from 1960 to 2019, especially after 1978. This result indicates aggressive efforts in the field of energy conservation and emission reduction made by China. There emerged six different decoupling states during the sample period: recessive coupling in 1960–1961, recessive decoupling in 1961–1962 and 1966–1967, strong decoupling in 1962–1963, weak decoupling in 1963–1965 and 1972–1973, expansive coupling in 1965–1966 and 1973–1974, strong negative decoupling in 1967–1968 and 1975–1976, and expansive negative decoupling in the remaining time period. The expansive coupling state that appeared from 2009 to 2011 may be caused by the recovery after the world economic crisis in 2008.

The results for South Africa are similar to those of Brazil. Expansive negative decoupling state appears the most frequently, which is followed by weak decoupling state and strong decoupling state. It is obvious that the decoupling index of South Africa fluctuates significantly. The expected strong decoupling state appears more frequently after 1997.

In general, the relationship between economic growth and carbon emission is variable and reflects different patterns of economic development. The results from the Tapio decoupling model only provide decoupling state from an elastic perspective in a certain period. In the following section, wavelet transform is introduced to perform an in-depth analysis in the time-frequency domain for richer results.

## Wavelet Analysis

The continuous wavelet transform method makes it possible to analyze the complex and evolving impact of economic growth on

**TABLE 3 |** Decoupling states of carbon emission and economic growth.

| Period | Brazil | | Russia | | India | | China | | South Africa | |
|---|---|---|---|---|---|---|---|---|---|---|
| | e | State | e | State | e | State | e | State | e | State |
| 60–61 | 0.48 | WD | — | — | 2.19 | END | 1.07 | RC | 1.14 | EC |
| 61–62 | 1.75 | END | — | — | 3.45 | END | 3.63 | RD | 0.56 | WD |
| 62–63 | 4.09 | END | — | — | 1.24 | END | −0.08 | SD | 0.52 | WD |
| 63–64 | 0.58 | WD | — | — | −0.28 | SD | 0.00 | WD | 1.13 | EC |
| 64–65 | −0.19 | SD | — | — | −3.87 | SND | 0.53 | WD | 1.17 | EC |
| 65–66 | 3.37 | END | — | — | −66.61 | SND | 0.93 | EC | 0.02 | WD |
| 66–67 | 0.59 | WD | — | — | 0.01 | WD | 2.97 | RD | 0.60 | WD |
| 67–68 | 1.48 | END | — | — | 2.68 | END | −2.01 | SND | 0.75 | WD |
| 68–69 | 0.91 | EC | — | — | 0.29 | WD | 1.36 | END | 0.80 | WD |
| 69–70 | 1.27 | END | — | — | 0.47 | WD | 1.75 | END | 0.86 | EC |
| 70–71 | 0.84 | EC | — | — | 3.43 | END | 1.91 | END | 2.93 | END |
| 71–72 | 0.94 | EC | — | — | −10.42 | SND | 1.65 | END | 1.13 | EC |
| 72–73 | 1.13 | EC | — | — | 0.91 | EC | 0.51 | WD | 0.23 | WD |
| 73–74 | 0.91 | EC | — | — | 2.73 | END | 0.87 | EC | 0.30 | WD |
| 74–75 | 1.02 | EC | — | — | 0.93 | EC | 1.83 | END | 2.83 | END |
| 75–76 | 0.26 | WD | — | — | 2.71 | END | −2.73 | SND | 1.90 | END |
| 76–77 | 1.08 | EC | — | — | 0.80 | EC | 1.26 | END | −37.65 | SND |
| 77–78 | 2.64 | END | — | — | 0.28 | WD | 1.02 | EC | 0.36 | WD |
| 78–79 | 0.94 | EC | — | — | −0.95 | SND | 0.29 | WD | 2.19 | END |
| 79–80 | −0.08 | SD | — | — | 0.83 | EC | −0.24 | SD | 0.66 | WD |
| 80–81 | 1.87 | RD | — | — | 1.33 | END | −0.22 | SD | 2.36 | END |
| 81–82 | 0.32 | WD | — | — | 0.95 | EC | 0.98 | EC | −23.70 | SND |
| 82–83 | 0.92 | RC | — | — | 1.13 | EC | 0.50 | WD | −2.21 | SND |
| 83–84 | 0.25 | WD | — | — | 0.69 | WD | 0.58 | WD | 1.59 | END |
| 84–85 | 0.93 | EC | — | — | 1.90 | END | 0.62 | WD | −2.16 | SND |
| 85–86 | 1.20 | END | — | — | 1.51 | END | 0.58 | WD | 114.88 | END |
| 86–87 | 1.21 | END | — | — | 1.72 | END | 0.58 | WD | −0.26 | SD |
| 87–88 | −8.40 | SND | — | — | 0.83 | EC | 0.64 | WD | 1.01 | EC |
| 88–89 | 0.67 | WD | — | — | 1.67 | END | 0.40 | WD | −0.24 | SD |
| 89–90 | 0.79 | WND | −2.39 | SND | 1.25 | END | 0.36 | WD | 25.65 | RD |
| 90–91 | 3.29 | END | 1.01 | RC | 6.12 | END | 0.52 | WD | −4.11 | SND |
| 91–92 | −1.14 | SND | 1.26 | RD | 1.19 | EC | 0.32 | WD | 3.54 | RD |
| 92–93 | 0.98 | EC | 0.58 | WND | 0.70 | WD | 0.50 | WD | 5.12 | END |
| 93–94 | 0.91 | EC | 0.93 | RC | 0.86 | EC | 0.47 | WD | 1.78 | END |
| 94–95 | 1.52 | END | 0.42 | WND | 0.83 | EC | 0.77 | WD | 2.19 | END |
| 95–96 | 4.54 | END | 0.54 | WND | 1.11 | EC | 0.44 | WD | 0.13 | WD |
| 96–97 | 1.61 | END | −4.75 | SD | 1.01 | EC | 0.02 | WD | 2.30 | END |
| 97–98 | 11.69 | END | 0.22 | WND | 0.34 | WD | −0.55 | SD | −4.23 | SD |
| 98–99 | 5.53 | END | 0.30 | WD | 0.95 | EC | −0.03 | SD | −0.27 | SD |
| 99–00 | 0.57 | WD | −0.10 | SD | 0.73 | WD | 0.33 | WD | 0.22 | WD |
| 00–01 | 1.94 | END | 0.49 | WD | 0.30 | WD | 0.28 | WD | −0.66 | SD |
| 01–02 | −0.55 | SD | −0.17 | SD | 0.80 | EC | 1.14 | EC | −1.10 | SD |
| 02–03 | −2.66 | SD | 0.27 | WD | 0.45 | WD | 1.76 | END | 4.55 | END |
| 03–04 | 0.88 | EC | 0.05 | WD | 0.78 | WD | 1.50 | END | 2.44 | END |
| 04–05 | 0.79 | WD | 0.17 | WD | 0.68 | WD | 1.10 | EC | −1.40 | SD |
| 05–06 | −0.01 | SD | 0.46 | WD | 0.77 | WD | 0.83 | EC | 1.32 | END |
| 06–07 | 0.72 | WD | −0.01 | SD | 1.02 | EC | 0.53 | WD | 0.76 | WD |
| 07–08 | 1.29 | END | 0.39 | WD | 2.50 | END | 0.78 | WD | 2.01 | END |
| 08–09 | 42.33 | RD | 0.85 | RC | 1.30 | END | 0.55 | WD | −1.00 | SND |
| 09–10 | 1.89 | END | 1.22 | END | 0.48 | WD | 0.90 | EC | −2.31 | SD |
| 10–11 | 1.15 | EC | 0.75 | WD | 1.00 | EC | 1.09 | EC | 0.46 | WD |
| 11–12 | 3.65 | END | 0.22 | WD | 1.82 | END | 0.33 | WD | −1.22 | SD |
| 12–13 | 2.53 | END | −2.06 | SD | 0.74 | WD | 0.22 | WD | −0.46 | SD |
| 13–14 | 11.56 | END | 0.27 | WD | 1.00 | EC | 0.03 | WD | 3.09 | END |
| 14–15 | 1.54 | RD | −0.02 | SND | 0.40 | WD | −0.20 | SD | −5.27 | SD |
| 15–16 | 1.03 | RC | −1.45 | SD | 0.75 | WD | −0.20 | SD | 4.69 | END |
| 16–17 | 0.97 | EC | 0.94 | EC | 0.38 | WD | 0.30 | WD | 0.93 | EC |
| 17–18 | −2.81 | SD | 1.08 | EC | 0.89 | EC | 0.31 | WD | 1.61 | END |
| 18–19 | −0.18 | SD | −0.57 | SD | 0.23 | WD | 0.36 | WD | 9.17 | END |

*Note: WD: weak decoupling; EC: expansive coupling; END: expansive negative decoupling; SND: strong negative decoupling; WND: weak negative decoupling; RC: recessive coupling; RD: recessive decoupling; SD: strong decoupling.*

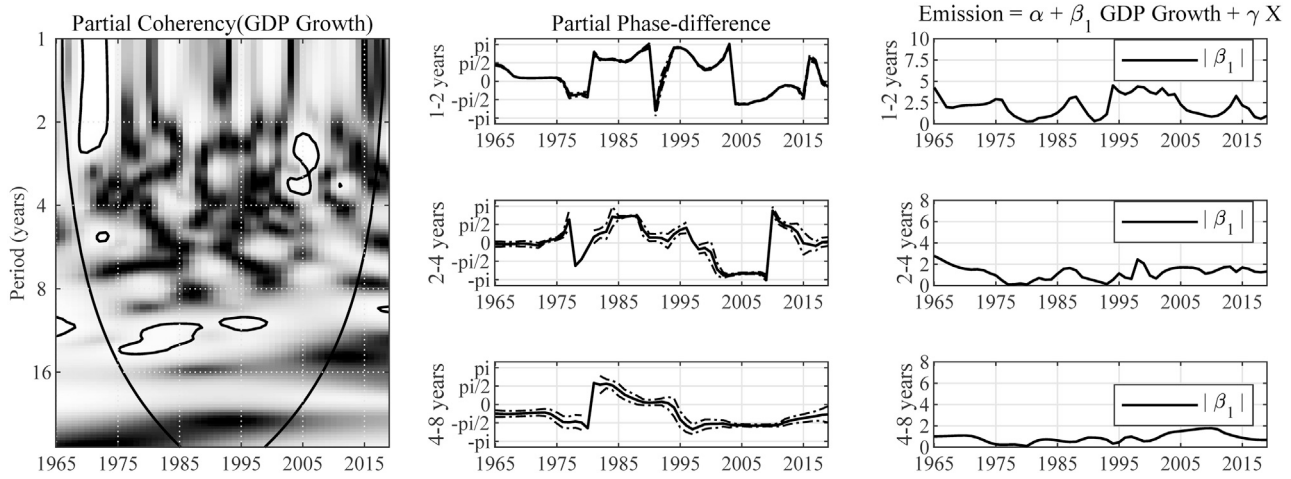



**FIGURE 3 |** Impact of GDP growth on carbon emission in China. Note: The expression "Partial Coherency (GDP Growth)" denotes the partial coherency between GDP growth and carbon emission, controlled for the influence from energy structure and trade openness. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

carbon emission that varies over time and across different frequencies. We estimate partial coherency, partial phase-difference, and partial wavelet gain at different frequencies and over time. Not only the time-varying impact of economic growth on carbon emission, but also the influences from the covariates are well identified. When the partial phase-difference and wavelet gain are estimated regarding a given variable, other variables are served as covariates. The results for each BRICS country are interpreted as follows.

China has been experiencing booming economy and significant shift of growth mode. We analyze the impact of China's GDP growth on carbon emission in short-, medium-, and long-term horizons. As shown in **Figure 3**, the partial wavelet coherency between carbon emission and GDP growth suggests more consistent correlation between these two variables in the medium- and long-term. And in the early stage of China's economic development, partial phase-difference indicator is near zero, indicating non-causality between economic growth and carbon emission.

In the short term (with a period of 1–2 years), relationship between emission and growth is fluid. In the period of 1975–1980, the partial phase-difference is estimated to be between $-\pi/2$ and 0, which indicates that the increase in emission rate leads to GDP growth. In this period, China's GDP growth depends on heavy industry. The extensive growth mode implies expansion in energy-consuming industries, along with emission surge, will result in higher growth rate. As for 1980–2004, the partial phase-difference is estimated to be between $\pi/2$ and $\pi$ with occasional large deviations around the years around 1991. It indicates negative correlation and causality from emission to GDP growth in general, which suggests that the increasing carbon emission restrains the growth of economy. The reason behind this is that after the rapidly growing phase of the economy development, carbon emission starts to inhibit economic growth to a certain degree; at the same time, the dramatic increase in emission pushes the government to balance the goal of economic expanding and the environment. In the period of 2004–2010, the partial phase-difference is estimated to be between $-\pi$ and $-\pi/2$, which indicates that reduced carbon emission leads to increased GDP growth rate during this period. It is largely the result of the transition of economic growth mode and the enhanced awareness of environmental protection. For the time after 2010, the results are quite mixed, reflecting impact of cross currents in the economy after the shock of the financial crisis.

In the medium-term (with a period of 2–4 years), it can be observed that the pattern of variation in partial phase-difference and wavelet gain is very similar to that in the short-term. So, more attention is paid to the results from the perspective of long-term.

In the long run (with a period of 4–8 years), the partial wavelet gain of carbon emission over GDP growth is around 1. It is smaller than that in the short- and medium-term due to the fact that influences from temporary factors fade out eventually. During the period of 1970–1979, the partial phase-difference is estimated to be between $-\pi/2$ and 0, which indicates that the increase in emission rate poses positive impact on GDP growth rate. From 1980 to 1994, the partial phase-difference is estimated to be between 0 and $\pi/2$, suggesting that increase in GDP growth rate causes higher carbon emission. During this period, China was experiencing an unprecedented rapid economic growth, and the environmental consideration did not feature prominently in the government's policy goals yet. The partial phase-difference lies in the interval $(-\pi, -\pi/2)$ from 1995 to 2013, which means with the increasing of GDP growth, the growth rate of carbon emissions decreased. The significant decoupling of emission from economic growth attributes to the effective environmental

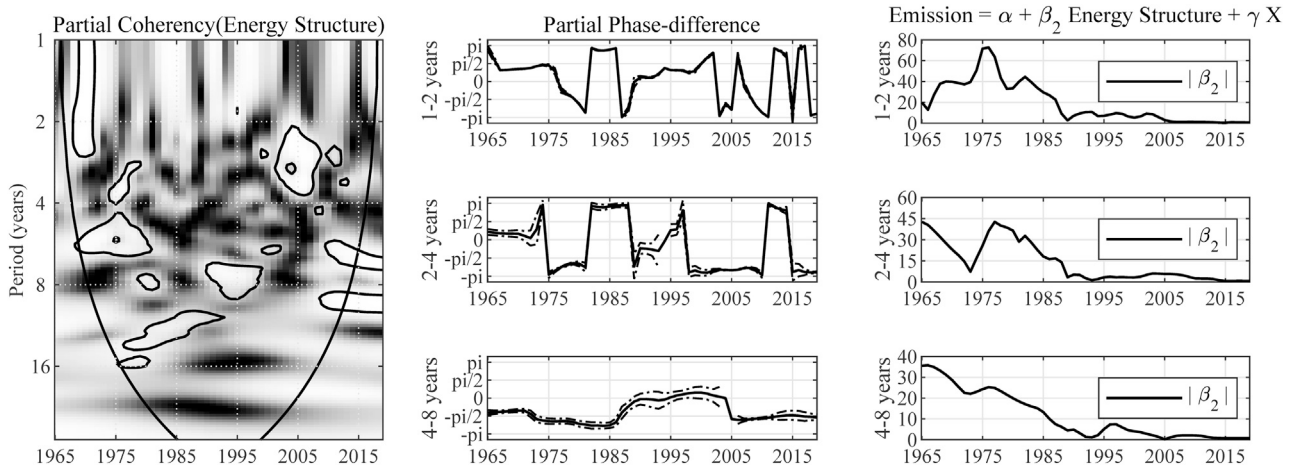



**FIGURE 4 |** Impact of energy structure on carbon emission in China. Note: The expression "Partial Coherency (Energy Structure)" denotes the partial coherency between energy structure and carbon emission, controlled for the influence from economic growth and trade openness. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

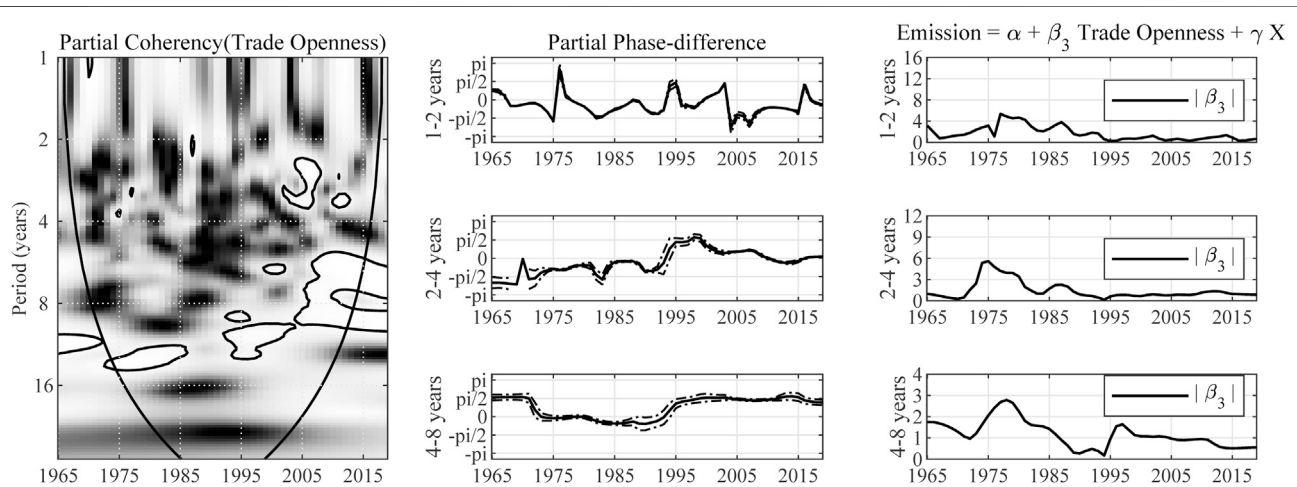



**FIGURE 5 |** Impact of trade openness on carbon emission in China. Note: The expression "Partial Coherency (Trade Openness)" denotes the partial coherency between trade openness and carbon emission, controlled for the influence from energy structure and economic growth. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

protection measures and the promoting usage of renewable energy in China.

The partial wavelet coherency between energy structure and carbon emission in China indicates that there is a significant correlation between proportion of renewable energy consumption and carbon emissions (see **Figure 4**). And the partial wavelet gain results show a decreasing trend of marginal impact of energy structure on carbon emission throughout the sample period, regardless of the perspectives of time-spans.

In the short-term and medium-term, the partial phase-difference falls into the range of $(0, \pi/2)$ from 1967 to 1975, indicating that the increase in renewable energy share is accompanied by higher level of carbon emission. The main reason behind this unusual result is that China was hungered for energy from all sources to support its economic development during this period. In the following period of 1975 to 1980, the partial phase-difference falls into the interval of $(-\pi, -\pi/2)$, showing that the increasing proportion of renewable energy leads to lowering the growth rate of emission. From 1981 to

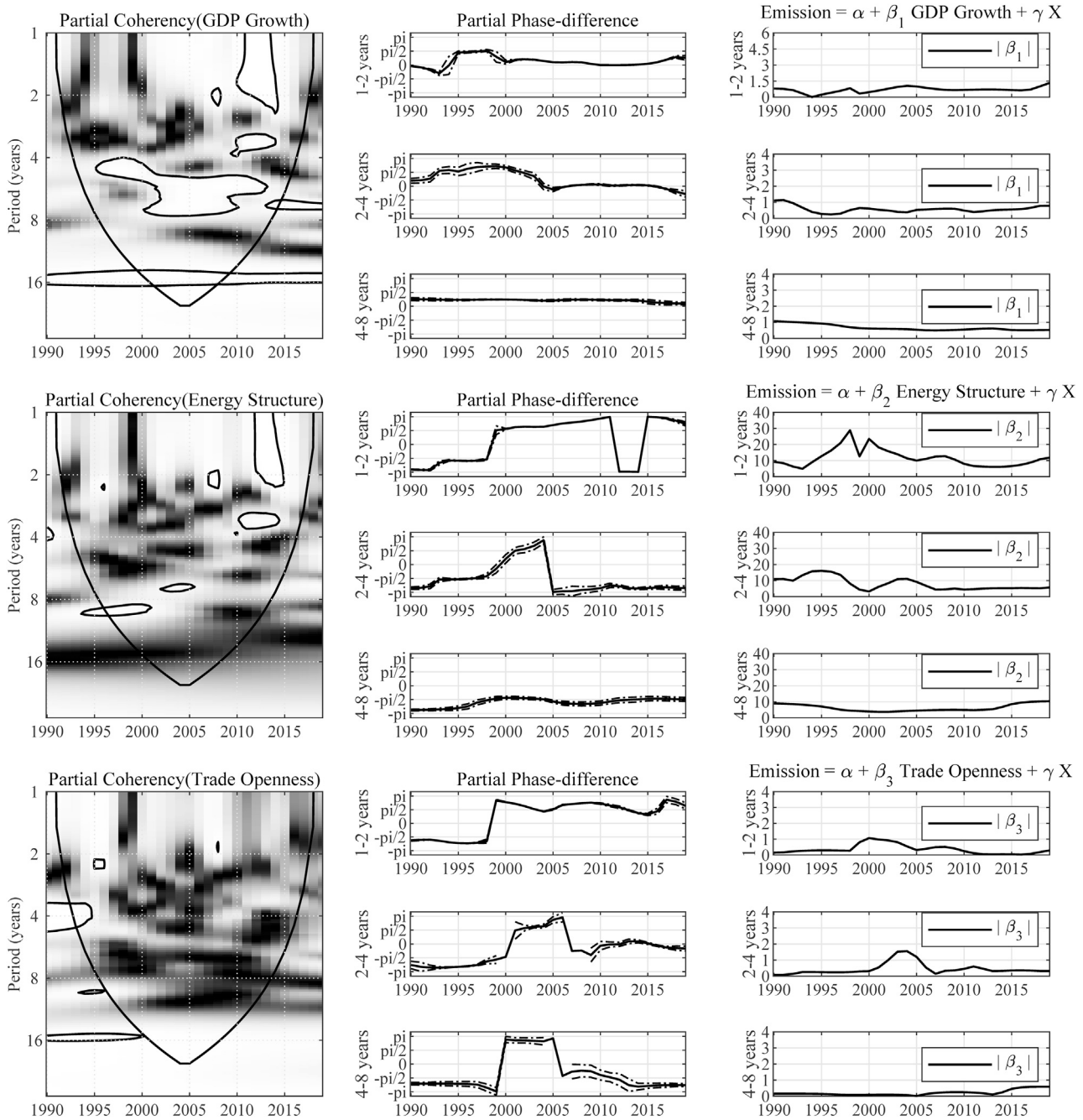



**FIGURE 6 |** Multivariate wavelet analysis results for Russia. Note: The expression "Partial Coherency (Z)" denotes the partial coherency between variable Z and carbon emission, controlled for the influence from other factors. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

1986, the partial phase-difference falls into the range of $(\pi/2, \pi)$. This means that the increasing emission is a cause of higher proportion of renewable energy consumption, which implies that economic growth has once again become the country's top priority. In the following sample period, the partial phase-difference fluctuates across all four quadrants frequently, reflecting ambiguous causality relationship.

The results from the perspective of medium-term are similar to those in the short-term, except for the sharp dive of partial phase-difference in 1996. This sudden drop may reflect the

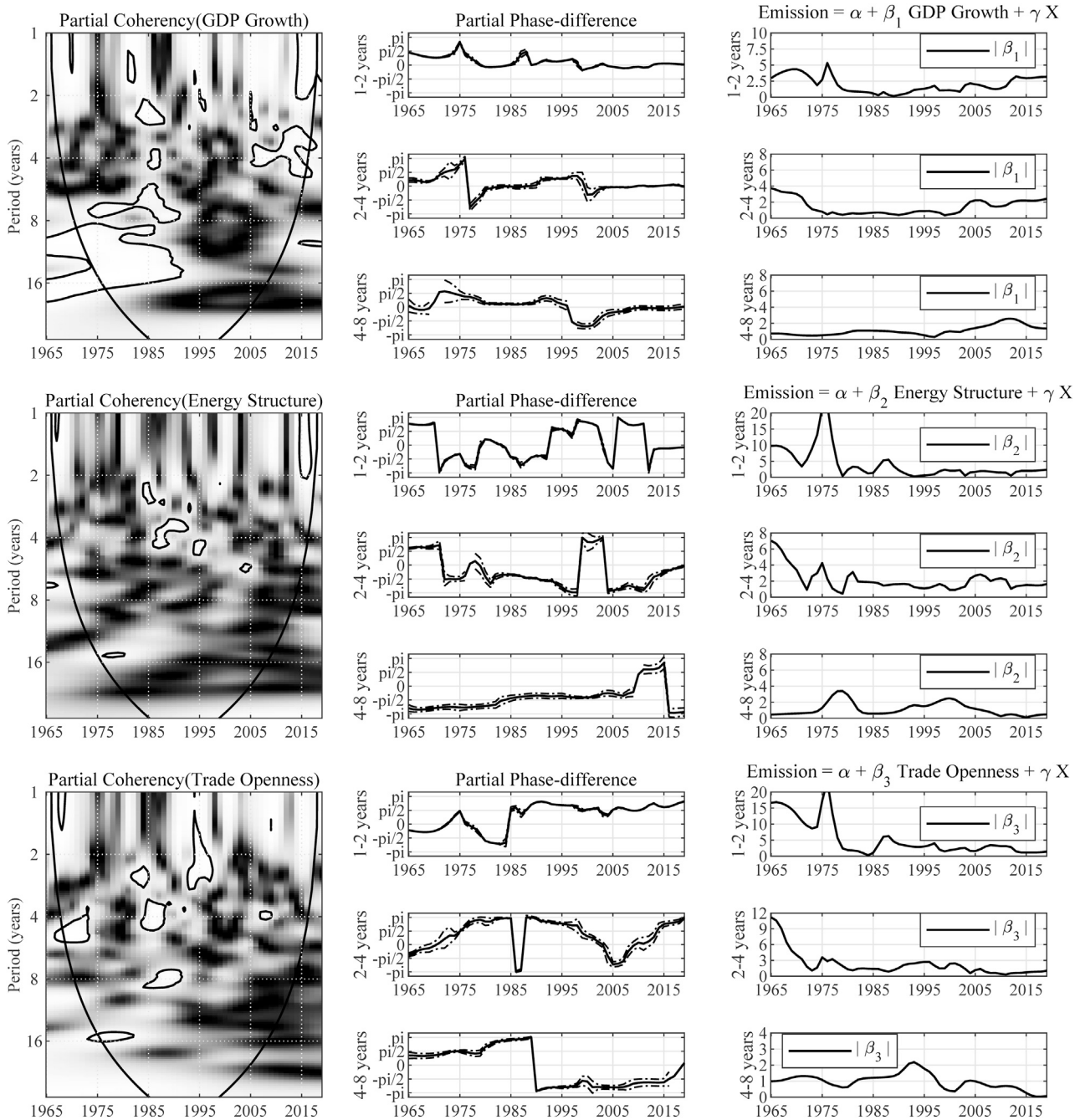



**FIGURE 7 |** Multivariate wavelet analysis results for Brazil. Note: The expression "Partial Coherency (Z)" denotes the partial coherency between variable Z and carbon emission, controlled for the influence from other factors. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

impact of changes in national policy which emphasizes the importance of development of new and renewable energy[6].

In the long-term, the partial phase-difference is estimated to fall into the range of $(-\pi, -\pi/2)$ from 1970 to 1987, suggesting an emission reduction effect of increasing share of renewable energy consumption. From 1990 to 2004, the partial phase-difference falls into the interval of $(0, \pi/2)$. After the reform and opening up,

[6]New and renewable energy development plan: 1996–2010, issued in 1995.

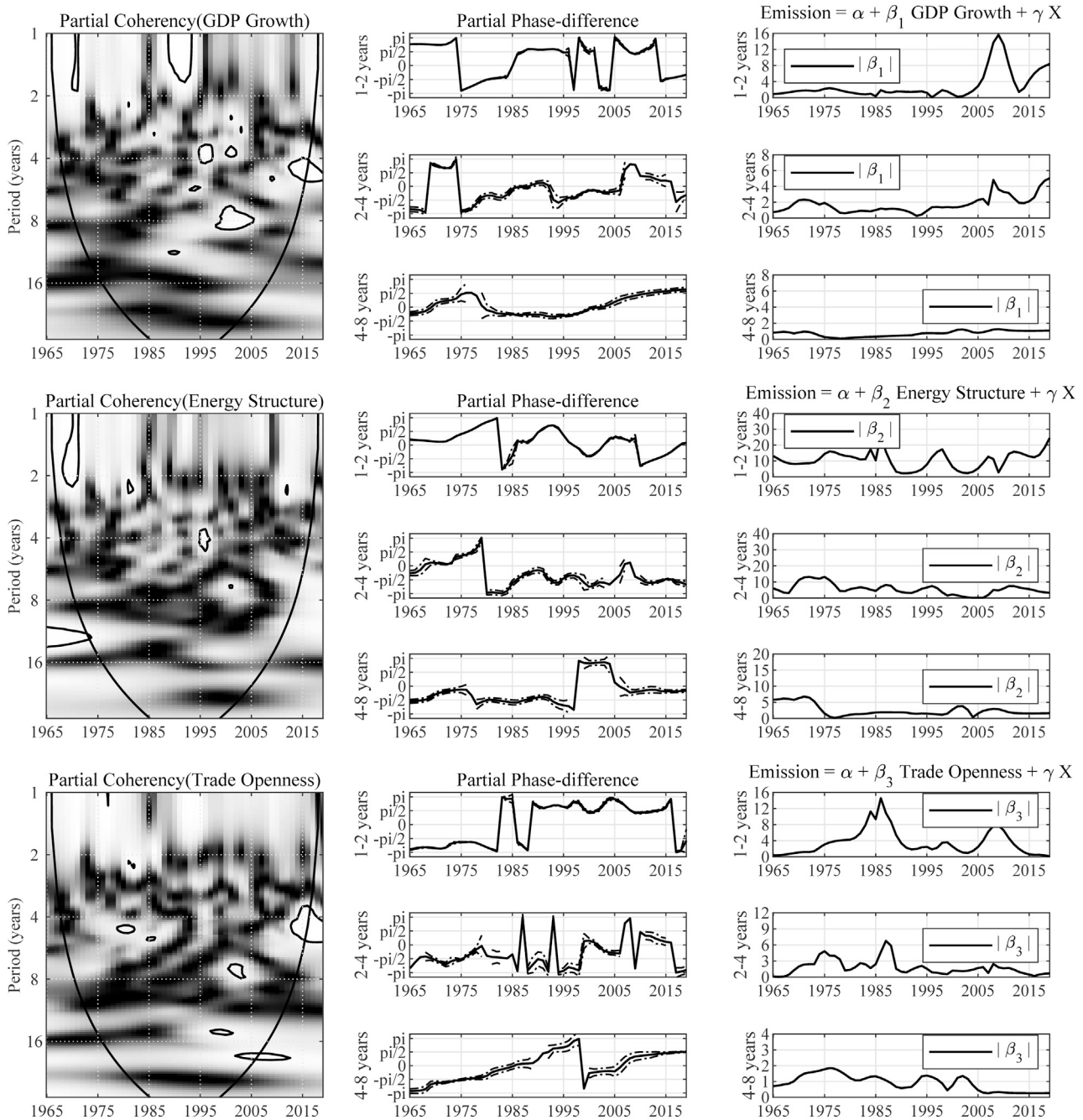



**FIGURE 8 |** Multivariate wavelet analysis results for India. Note: The expression "Partial Coherency (Z)" denotes the partial coherency between variable Z and carbon emission, controlled for the influence from other factors. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

China's economic growth was largely driven by the introduction of energy-inefficient technologies, which had depressed the carbon reduction effect of renewable energy consumption. From 2005 to 2013, the partial phase-difference falls into the range of $(-\pi, -\pi/2)$, indicating that production energy efficiency has begun to improve and a greater proportion of renewable energy has been a significant cause of the decline in the growth rate of carbon emissions.

According to the wavelet coherency estimates in **Figure 5**, there exists significant and consistent correlation between emission and trade openness. In the short-term, the partial phase-difference lies in $(-\pi/2, 0)$ from 1967 to 1994 and the

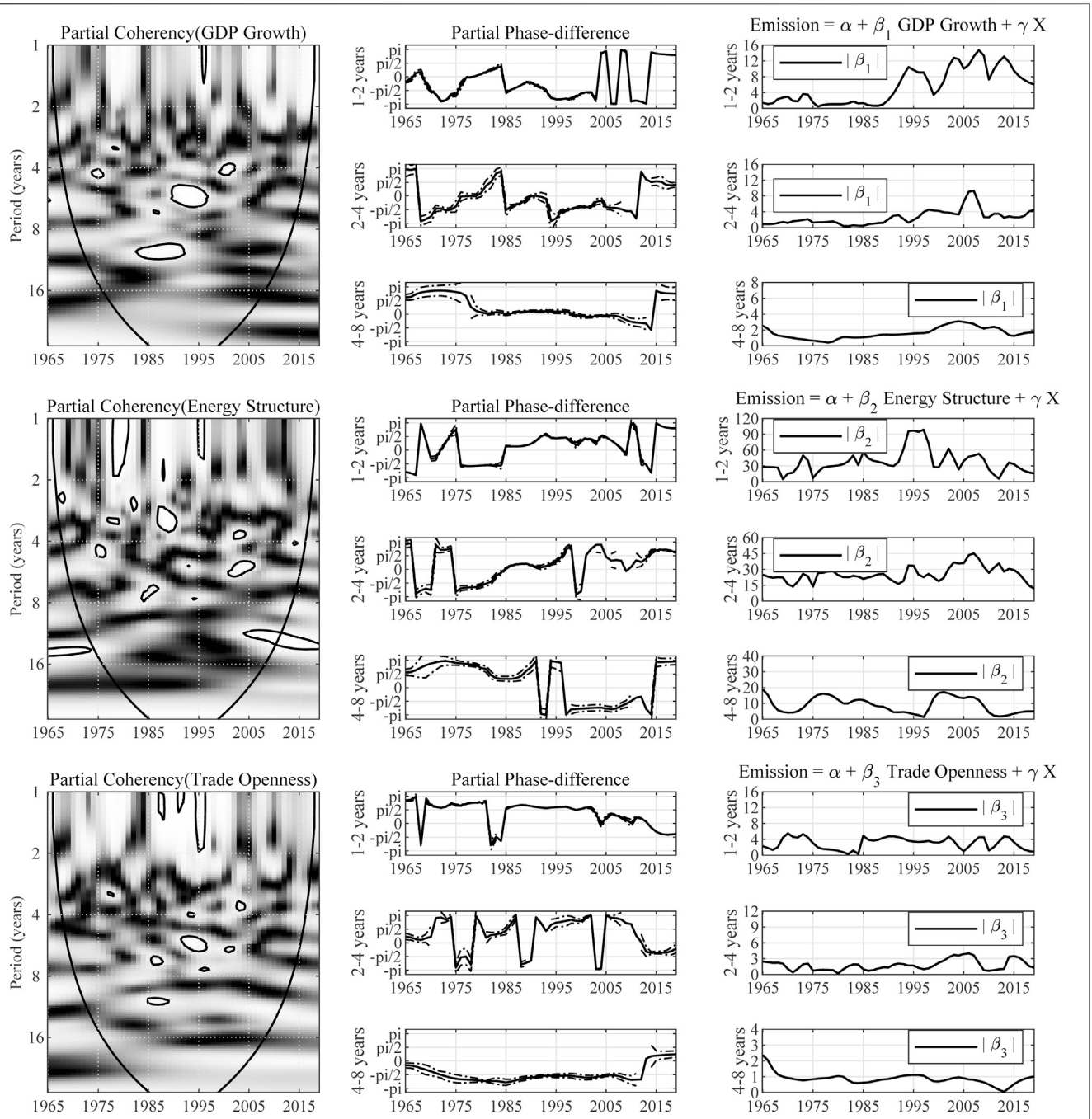



**FIGURE 9 |** Multivariate wavelet analysis results for South Africa. Note: The expression "Partial Coherency (Z)" denotes the partial coherency between variable Z and carbon emission, controlled for the influence from other factors. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

period of 2004–2015, indicating positive impact of carbon emission on trade openness. That is because China was a major producer of industrial semifinished products in the period from 1967 to 1994. The export of these products has resulted in a significant increase in China's trade volume, such that China's increasing carbon emission rate was followed by higher proportion of trade to GDP. And this effect was evident again after China's entry into the World Trade Organization. The result for the period from 1994 to 2004 shows the positive impact of trade openness on the emission rate. In the long run, the partial

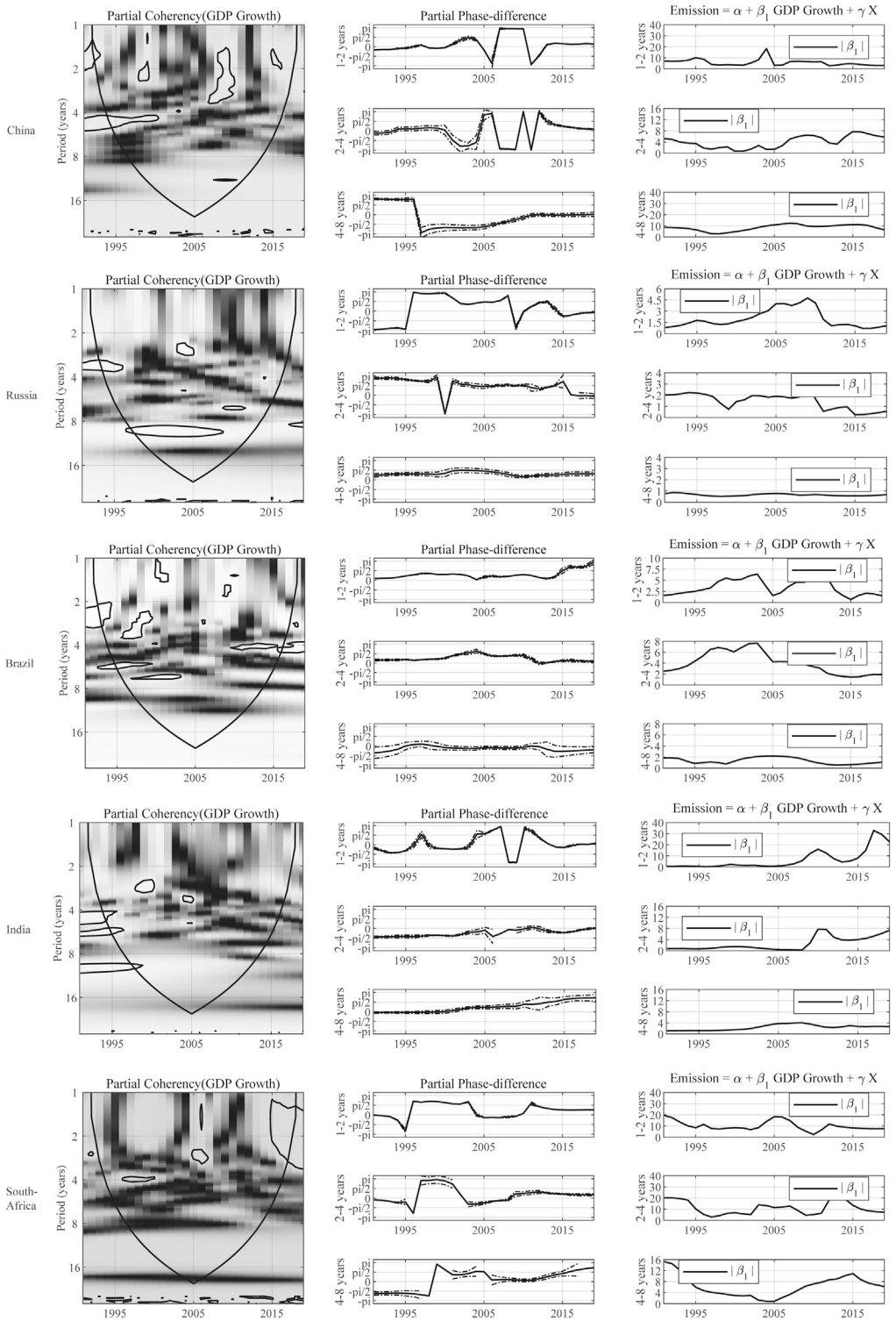



**FIGURE 10 |** Multivariate wavelet analysis results controlled for additional factors. Note: The expression "Partial Coherency (Z)" denotes the partial coherency between variable Z and carbon emission, controlled for the influence from other factors. The areas below the symmetric black convex curve that appears in the figure of coherency are called the "Cone of Influence (COI)," in which the edge effects are profound. Inside the COI, the results tend to be unreliable. The contours in the coherency diagrams indicate 5% significance level, bootstrapped for 10,000 replications. And, the brighter colors correspond to higher significance level, and darker colors correspond to lower significance level. The dotted lines accompanying the partial phase-difference indicate the 95% confidence intervals.

phase-difference lies in $(-\pi/2, 0)$ and $(0, \pi/2)$ respectively during the periods 1972–1994 and 1994–2013. The relationship between trade openness and carbon emission is quite consistent with the results in the short-term.

The results of multivariate wavelet analysis for Russian are presented in **Figure 6**. According to partial coherency between GDP growth and emission, the correlation is much more significant in the short-term. This result implies a temporary relationship between these two variables. The estimated partial wavelet gain of emission over GDP growth is approximately 0.5, which is quite constant across different frequencies.

In the short-term, the partial phase-difference lies in $(-\pi/2, 0)$ from 1992 to 1994, indicating that the increasing of emission rate causes higher rate of GDP growth. From 1994 to 2018, the partial phase-difference is greater than 0 and less than $\pi/2$ all along. As a typical industrial economy, the Russian production system is dominated by heavy industries, resulting in the causality from GDP growth to higher level of carbon emission. In the long run, the partial phase-difference falls into the range of $(0, \pi/2)$ throughout the whole sample period. This result reveals the dependence of Russia's economic growth on high-emission heavy industry in the long-term and the persistent impact of growth on emission.

As for other factors, the partial phase-difference between emission and energy structure is estimated to be between $-\pi$ and $-\pi/2$ in the long-term. And the partial wavelet gain is estimated to be approximately 10. These results indicate that the rising proportion of renewable energy has effectively restrained the growth of carbon emission rate since the 1990s. The partial wavelet gain of emission over trade openness is approximately 0.2 in the long-term, implying a much weaker impact on carbon emission other than energy structure. And the partial coherency also provides evidence of insignificant correlation between trade openness and carbon emission.

In **Figure 7** we observe a significant correlation between Brazil's economic growth and carbon emission. The partial wavelet gain of emission over GDP growth is approximately 1. In both short-term and long-term, the partial phase-difference roughly falls into the range of $(-\pi, -\pi/2)$ from 1996 to 2004, reflecting the inhibitory effect of the increase of GDP growth rate on the growth rate of carbon emissions. Brazil currently possesses one of the most sustainable energy structures among the emerging economies. There is 44.1% of energy supply coming from renewable sources, and approximately 81.7% of all electricity supply coming from hydropower (Hsiao-Tien and Hsin-Chia, 2013).

As for other factors, the partial coherency between energy structure and carbon emission seems insignificant in most areas. And the long-term partial phase-difference between emission and trade openness falls into the range of $(\pi/2, \pi)$ from 1972 to 1989, and into the range of $(-\pi, -\pi/2)$ from 1990 to 2015. When the correlation remains negative, the leading variable has changed from carbon emissions to trade openness after 1990, indicating a reversion in the direction of causality. The relationship between the two variables has shifted from that carbon emission restricts scale of international trade to that increased trade openness mitigating the growth of carbon emission.

A careful look at India's NDC reveals that India's climate commitment is grounded on various macroeconomic indicators such as GDP growth, emission intensity, mix of energy sources, and changing structure of the economy, which are also drivers of carbon emission (Nandini and Joyashree, 2020). As presented in **Figure 8**, the partial wavelet gain of emission over GDP growth is approximately 1 and the partial coherency indicates a significant correlation between the two variables only in a small area. Compared with countries previously discussed, the correlation between India's economic growth and carbon emissions is relatively weak. The long-run partial phase-difference falls into the range of $(-\pi/2, \pi/2)$ from 1965 to 2015, and implies a positive correlation between the two variables in the long-term. But, the low significant level of partial coherency makes it meaningless for the correlation and causality inference. As for the energy structure and trade openness, estimated partial coherency suggests low significant level of correlation and causality.

South Africa is one of the world's most carbon-intensive economies (Marcel, 2013). As presented in **Figure 9**, partial phase-difference is estimated to be between $-\pi/2$ and $\pi/2$ during 1976–2015. It is easy to observe that there is a strong positive correlation between the emission and GDP growth in the long run. Given that most of South Africa's energy needs are met by fossil fuels, and considering its status as a non-oil producer, a strong link between energy structure and CO2 emissions is to be expected. The partial phase-difference is estimated to be between $\pi/2$ and $\pi$ and between $-\pi$ and $-\pi/2$ in the most time, indicating a strong negative correlation in the long run. The result also suggests causality from increase in GDP growth to slower adoption of renewable energy, before 1990. And the improving energy structure starts to mitigate incremental carbon emission ever since 1996.

## Robustness

There is little consensus on that the causality relationship has been reached in the field of the energy consumption–economic growth nexus due to the omitted variable bias. So, we have reasons to concern that our empirical results also suffer from the omitted variable problem. We follow the approach from prior researches to examine whether the empirical results suffer from this problem (see Lean and Smyth, 2010; Apergis and Payne, 2010a; Apergis and Payne, 2010b; Apergis and Payne, 2011a; Apergis and Payne, 2011b). The physical inputs and a proxy of economic development are introduced and the multivariate wavelet analysis is conducted as a robustness check. The multivariate wavelet analysis results with these additional control variables are presented in **Figure 10**.

As **Figure 10** demonstrates, the results are consistent with those reported in previous sections, with certain difference in statistic values. The reason may be the shortened sample period due to the inclusion of more variables. We believe the adoption of wavelet analysis not only makes it possible to extract more abundant information from the perspective of time and frequency, but also alleviates the interference caused by omitted variable problem to a certain extent.

## CONCLUSION

The decoupling relationship between economic growth and carbon emissions has received extensive academic discussion. But the dynamic time-varying impact of growth on carbon emission has not received enough attention. Besides the decoupling analysis, this paper further introduces the wavelet analysis method that estimates the dynamic relationship in time-frequency space to fill this gap.

The results of the decoupling analysis show that only China exhibits an obvious decoupling trend among the BRICS countries. This result indicates aggressive efforts in the field of energy conservation and emission reduction made by China. The decoupling indies of other four emerging economies are relatively mixed, and part of them are more of an irregular list of decoupling states at different time points.

The wavelet analysis provides more extensive results, fully considering the influence of covariates and the time-varying characteristics of the relationship between variables. Among the BRICS countries, coherency analysis indicates a significant impact of economic growth on carbon emission only in China, Russia, and Brazil.

In the early stage of China's economic development, the near-zero partial phase-difference indicates a non-causality relationship between the economic growth and carbon emission. When the expansion of heavy industries gets started, the accompanying emission leads to rise in the growth rate. And after the country enters a path of rapid growth, the increase in growth rate continues to push up carbon emission. Until 1995, the economic growth starts to suppress carbon emission growth. This is largely the result of transition of economic growth mode, change in energy structure, and enhanced awareness of environmental protection. Trade openness generally has a positive impact on emission, due to China's long-standing trade surplus.

Results from wavelet analysis reveal the dependence of Russia's economic growth on high-emission heavy industry in the long-term and the persistent positive causality from economic growth to carbon emission. Meanwhile, the improvement of energy structure helps in effectively restraining the growing carbon emission ever since the 1990s. And there is no obvious evidence that trade openness has a certain effect on carbon emission.

Estimation based on data from Brazil implies a significant inhibitory effect of the increase of GDP growth rate on the rise of carbon emissions. It attributes to the abundant renewable energy resources such that improvement in energy structure does not play a significant role in reducing emission. And increased openness helps in mitigating the growth of carbon emissions in Brazil since 1990.

In conclusion, different resource endowments and growth patterns lead to different impacts of economic growth on carbon emissions and the time-varying characteristics of the causality relationship between them. In addition to economic growth, changes in a country's energy structure and trade openness can also significantly affect carbon emissions. When the government formulates relevant carbon emission targets and emission reduction policies, it is necessary to pay full attention to the dynamic impact of economic growth, energy structure, and trade openness to carbon emission.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

LX: software, validation, resources, data curation, writing original draft. XC: writing—review and editing, and visualization. SS: conceptualization, methodology, formal analysis, investigation. ZY: writing—review and editing, supervision, project administration, and funding acquisition. All authors: contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We show our gratitude to participants at the Taishan Financial Forum for their helpful comments. We acknowledge the financial support from Shandong Provincial Natural Science Foundation (Grant Number: ZR2020QG032), Shandong Provincial Social Science Planning Office (Grant Numbers: 21DGLJ12; 21DJJJ02), Taishan Scholars Program of Shandong Province, China (Grant Numbers: ts201712059; tsqn201909135) and Youth Innovative Talent Technology Program of Shandong Province, China (Grant Number: 2019RWE004). All errors remain our own.

## REFERENCES

Adedoyin, F. F., Gumede, M. I., and Bekun, F. V. (2020). Modelling Coal Rent, Economic Growth and CO2 Emissions: Does Regulatory Quality Matter in BRICS Economies?. *Sci. Total Environ.* 710, 136284. doi:10.1016/j.scitotenv.2019.136284

Aguiar-Conraria, L., Martins, M. M. F., and Joana Soares, M. (2018). Estimating the Taylor Rule in the Time-Frequency Domain. *J. Macroeconomics* 57, 122–137. doi:10.1016/j.jmacro.2018.05.008

Antweiler, W., Copeland, R. B., and Taylor, M. S. (2001). Is Free Trade Good for the Environment?. *Am. Econ. Rev.* 80, 15–27. doi:10.1257/aer.91.4.87

Apergis, N., Christou, C., and Gupta, R. (2017). Are There Environmental Kuznets Curves for US State Level CO2 Emissions?. *Renew. Sustain. Energ. Rev.* 69, 551–558. doi:10.1016/j.rser.2016.11.219

Apergis, N. (2016). Environmental Kuznets Curves: New Evidence on Both Panel and Country-Level CO2 Emissions. *Energ. Econ.* 54, 263–271. doi:10.1016/j.eneco.2015.12.007

Apergis, N., Payne, J. E., Menyah, K., and Wolde-Rufael, Y. (2010). On the Causal Dynamics between Emissions, Nuclear Energy, Renewable Energy, and Economic Growth. *Ecol. Econ.* 69 (11), 2255–2260. doi:10.1016/j.ecolecon.2010.06.014

Apergis, N., and Payne, J. E. (2011a). Renewable and Non-renewable Electricity Consumption Growth Nexus: Evidence from Emerging

Market Economies. *Appl. Energ.* 88 (12), 5226–5230. doi:10.1016/j.apenergy.2011.06.041

Apergis, N., and Payne, J. E. (2012). Renewable and Non-renewable Energy Consumption-Growth Nexus: Evidence from a Panel Error Correction Model. *Energ. Econ.* 34 (3), 733–738. doi:10.1016/j.eneco.2011.04.007

Apergis, N., and Payne, J. E. (2010a). Renewable Energy Consumption and Growth in Eurasia. *Energy Economics 32* (6), 1392–1397. doi:10.1016/j.eneco.2010.06.001

Apergis, N., and Payne, J. E. (2011b). The Renewable Energy Consumption–Growth Nexus in Central America. *Appl. Energ.* 88 (1), 343–347. doi:10.1016/j.apenergy.2010.07.013

Apergis, N., and Payne, J. E. (2010b). The Emissions, Energy Consumption, and Growth Nexus: Evidence from the Commonwealth of Independent States. *Energy Policy* 38 (1), 650–655. doi:10.1016/j.enpol.2009.08.029

Apergis, N., and Tang, C. F. (2013). Is the Energy-Led Growth Hypothesis Valid? New Evidence from a Sample of 85 Countries. *Energ. Econ.* 38, 24–31. doi:10.1016/j.eneco.2013.02.007

Baek, J., Cho, Y., and Koo, W. W. (2009). The Environmental Consequences of Globalization: a Country-specific Time-Series Analysis. *Ecol. Econ.* 68, 2255–2264. doi:10.1016/j.ecolecon.2009.02.021

Banerjee, S., and Murshed, M. (2020). Do emissions Implied in Net export Validate the Pollution haven Conjecture? Analysis of G7 and BRICS Countries. *Int. J. Sustain. Economy*. 12, 297. doi:10.1504/IJSE.2020.111539

Bloomfield, D., McAteer, R., Lites, B., Judge, P., Mathioudakis, M., and Keena, F. (2004). Wavelet Phase Coherence Analysis: Application to a Quiet-Sun Magnetic Element. *Astrophysical J.* 617, 623–632. Available at: https://iopscience.iop.org/article/10.1086/425300. doi:10.1086/425300

Climent, F., and Pardo, A. (2007). Decoupling Factors on the Energy-Output Linkage: the Spanish Case. *Energy Pol* 35, 522–528. doi:10.1016/j.enpol.2005.12.022

Cole, M. A., and Elliott, R. J. (2003). Determining the Trade–Environment Composition Effect: the Role of Capital, Labor and Environmental Regulations. *J. Environ. Econ. Manage.* 46 (3), 363–383. doi:10.1016/S0095-0696(03)00021-4

Copeland, B., and Taylor, M. S. (2003). *Trade and the Environment: Theory and Evidence*. Princeton, NJ: Princeton University Press. Available at: http://refhub.elsevier.com/S0140-9883(20)30306-6/rf0110.

Csereklyei, Z., and Stern, D. I. (2015). Global Energy Use: Decoupling or Convergence?. *Energ. Econ.* 51, 633–641. doi:10.1016/j.eneco.2015.08.029

Dai, S., Zhang, M., and Huang, W. (2016). Decomposing the Decoupling of CO2 Emission from Economic Growth in BRICS Countries. *Nat. Hazards* 84 (2), 1055–1073. doi:10.1007/s11069-016-2472-0

Dauda, L., Xingle, L., Mensah, C. N., Salman, M., Boamah, K. B., Ampon-Wireko, S., et al. (2021). Innovation, Trade Openness and CO2 Emissions in Selected Countries in Africa. *J. Clean. Prod.* 281, 125143. doi:10.1016/j.jclepro.2020.125143

Dong, B., Zhang, M., Mu, H. L., and Su, X. M. (2016). Study on Decoupling Analysis between Energy Consumption and Economic Growth in Liaoning Province. *Energ. Pol* 97, 414–420. doi:10.1016/j.enpol.2016.07.054

Dong, B., Ma, X., Zhang, Z., Zhang, H., Chen, R., Song, Y., et al. (2020). Carbon Emissions, the Industrial Structure and Economic Growth: Evidence from Heterogeneous Industries in China. *Environ. Pollut.* 262, 114322. doi:10.1016/j.envpol.2020.114322

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., et al. (2018). Increased Human and Economic Losses from River Flooding with Anthropogenic Warming. *Nat. Clim. Change* 8 (9), 781–786. doi:10.1038/s41558-018-0257-z

Enevoldsen, M. K., Ryelund, A. V., and Andersen, M. S. (2007). Decoupling of Industrial Energy Consumption and CO2-emissions in Energy-Intensive Industries in Scandinavia. *Energy Econ* 29, 665–692. doi:10.1016/j.eneco.2007.01.016

Fragkos, P., Laura van Soest, H., Schaeffer, R., Reedman, L., Koberle, A. C., Macaluso, N., et al. (2021). Energy System Transitions and Low-Carbon Pathways in Australia, Brazil, Canada, China, EU-28, India, Indonesia, Japan, Republic of Korea, Russia and the United States. *Energy* 216, 119385. doi:10.1016/j.energy.2020.119385

Frankel, J., and Rose, A. (2005). Is Trade Good or Bad for the Environment? Sorting Out the Causality. *Rev. Econ. Stat.* 87 (1), 85–91. doi:10.1162/0034653053327577

Freitas, L. C., and Kaneko, S. (2011). Decomposing the Decoupling of CO2 Emissions and Economic Growth in Brazil. *Ecol. Econ.* 70, 1459–1469. doi:10.1016/j.ecolecon.2011.02.011

Galeotti, M., Manera, M., and Lanza, A. (2009). On the Robustness of Robustness Checks of the Environmental Kuznets Curve Hypothesis. *Environ. Resource Econ.* 42 (4), 551–574. doi:10.1007/s10640-008-9224-x

Gan, P. Y., and Li, Z. D. (2008). An Econometric Study on Long-Term Energy Outlook and the Implications of Renewable Energy Utilization in Malaysia. *Energy Policy* 2, 890–899. doi:10.1016/j.enpol.2007.11.003

Gao, D., Zhou, H., Wang, J., Miao, S., Yang, F., Wang, G., et al. (2015). Size-Dependent Electrocatalytic Reduction of CO2 over Pd Nanoparticles. *J. Am. Chem. Soc.* 137 (13), 4288–4291. doi:10.1021/jacs.5b00046

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the Cross Wavelet Transform and Wavelet Coherence to Geophysical Time Series. *Nonlinear Process. Geophys.* 11, 561–566. doi:10.5194/npg-11-561-2004

Grossman, G., and Krueger, A. (1991). Environmental Impacts of the North American Free Trade Agreement. NBER Working Paper. 3914. doi:10.3386/w3914

Guan, D., Meng, J., Reiner, D. M., Zhang, N., Shan, Y., Mi, Z., et al. (2018). Structural Decline in China's CO2 Emissions through Transitions in Industry and Energy Systems. *Nat. Geosci.* 11, 551–555. doi:10.1038/s41561-018-0161-1

Hsiao-Tien, P., and Hsin-Chia, F. (2013). Renewable Energy, Non-renewable Energy and Economic Growth in Brazil. *Renew. Sustain. Energ. Rev.* 25, 381–392. doi:10.1016/j.rser.2013.05.004

Hu, H., Xie, N., Fang, D., and Zhang, X. (2018). The Role of Renewable Energy Consumption and Commercial Services Trade in Carbon Dioxide Reduction: Evidence from 25 Developing Countries. *Appl. Energ.* 211, 1229–2124. doi:10.1016/j.apenergy.2017.12.019

Hu, M., Li, R., You, W., Liu, Y., and Lee, C. (2020). Spatiotemporal Evolution of Decoupling and Driving Forces of CO2 Emissions on Economic Growth along the Belt and Road. *J. Clean. Prod.* 277, 123272. doi:10.1016/j.jclepro.2020.123272

Kan, S., Chen, B., and Chen, G. (2019). Worldwide Energy Use across Global Supply Chains: Decoupled from Economic Growth?. *Appl. Energ.* 250, 1235–1245. doi:10.1016/j.apenergy.2019.05.104

Kovanda, J., and Hak, T. (2007). What Are the Possibilities for Graphical Presentation of Decoupling? an Example of Economy-wide Material Flow Indicators in the Czech Republic. *Ecol. Indicators* 7, 123–132. doi:10.1016/j.ecolind.2005.11.002

Lean, H. H., and Smyth, R. (2010). CO2 Emissions, Electricity Consumption and Output in ASEAN. *Appl. Energ.* 87 (6), 1858–1864. doi:10.1016/j.apenergy.2010.02.003

Liu, X., Zhang, S., and Bae, J. (2017). The Nexus of Renewable Energy-Agriculture-Environment in BRICS. *Appl. Energ.* 204, 489–496. doi:10.1016/j.apenergy.2017.07.077

Loh, L. (2013). Co-movement of Asia–Pacific with European and US Stock Market Returns: a Cross-Time-Frequency Analysis. *Res. Int. Business Finance* 29, 1–13. doi:10.1016/j.ribaf.2013.01.001

Loo, B. P. Y., and Banister, D. (2016). Decoupling Transport from Economic Growth: Extending the Debate to Include Environmental and Social Externalities. *J. Transport Geogr.* 57, 134–144. doi:10.1016/j.jtrangeo.2016.10.006

Ma, Q., Murshed, M., and Khan, Z. (2021). The Nexuses between Energy Investments, Technological Innovations, Emission Taxes, and Carbon Emissions in China. *Energy Policy* 155, 112345. doi:10.1016/j.enpol.2021.112345

Ma, X., Ye, Y., Shi, X., and Zou, L. (2016). Decoupling Economic Growth from CO2 Emissions: a Decomposition Analysis of China's Household Energy Consumption. *Adv. Clim. Change Res.* 7, 192–200. doi:10.1016/j.accre.2016.09.004

Managi, S., Hibiki, A., and Tsurumi, T. (2008). Does Trade Liberalization Reduce Pollution Emissions. Discussion papers, 8013. Available at: https://EconPapers.repec.org/RePEc:eti:dpaper:08013 (Accessed May 17, 2021).

Mandler, M., and Scharnagl, M. (2014). Money Growth and Consumer price Inflation in the Euro Area: A Wavelet Analysis. No 33/2014. Berlin: Deutsche Bundesbank. Available at: http://www.bundesbank.de (Accessed May 17, 2021).

Marcel, K. (2013). CO2 Emissions, Energy Consumption, Income and Foreign Trade: A South African Perspective. *Energy Policy* 63, 1042–1050. doi:10.1016/j.enpol.2013.09.022

Mielnik, O., and Goldemberg, J. (2002). Foreign Direct Investment and Decoupling between Energy and Gross Domestic Product in Developing Countries. *Energ. Pol* 30, 87–89. doi:10.1016/S0301-4215(01)00080-5

Murshed, M., Alam, R., and Ansarin, A. (2021a). The Environmental Kuznets Curve Hypothesis for Bangladesh: the Importance of Natural Gas, Liquefied Petroleum Gas, and Hydropower Consumption. *Environ. Sci. Pollut. Res.* 28, 17208–17227. doi:10.1007/s11356-020-11976-6

Murshed, M., Ali, S. R., and Banerjee, S. (2021b). Consumption of Liquefied Petroleum Gas and the EKC Hypothesis in South Asia: Evidence from Cross-Sectionally Dependent Heterogeneous Panel Data with Structural Breaks. *Energ. Ecol. Environ.* 6, 353. doi:10.1007/s40974-020-00185-z

Murshed, M. (2020). An Empirical Analysis of the Non-linear Impacts of ICTTrade Openness on Renewable Energy Transition, Energy Efficiency, Clean Cooking Fuel Access and Environmental Sustainability in South Asia. *Environ. Sci. Pollut. Res.* 27, 36254. doi:10.1007/s11356-020-09497-3

Nandini, D., and Joyashree, R. (2020). India Can Increase its Mitigation Ambition: An Analysis Based on Historical Evidence of Decoupling between Emission and Economic Growth. *Energ. Sustain. Develop.* 57, 189–199. doi:10.1016/j.esd.2020.06.003

Narayan, P. K., and Smyth, R. (2007). A Panel Cointegration Analysis of the Demand for Oil in the Middle East. *Energy Policy* 35, 6258–6265. doi:10.1016/j.enpol.2007.07.011

Narayan, P. K., Smyth, R., and Prasad, A. (2007). Electricity Consumption in G7 Countries: A Panel Cointegrating Analysis of Residential Demand Elasticities. *Energy Policy* 35, 4485–4494. doi:10.1016/j.enpol.2007.03.018

Narayan, S. (2016). Predictability within the Energy Consumption-Economic Growth Nexus: Some Evidence from Income and Regional Groups. *Econ. Model.* 54, 515–521. doi:10.1016/j.econmod.2015.12.037

Obama, B. (2017). The Irreversible Momentum of Clean Energy. *Science* 355 (6321), 126. doi:10.1126/science.aam6284

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L., and Sunday, J. M. (2019). Greater Vulnerability to Warming of marine versus Terrestrial Ectotherms. *Nature* 569 (7754), 108e111. doi:10.5281/zenodo.257619710.1038/s41586-019-1132-4

Rehman, A., Ulucak, R., Murshed, M., Ma, H., and Işık, C. (2021). Carbonization and Atmospheric Pollution in China: The Asymmetric Impacts of Forests, Livestock Production, and Economic Progress on CO2 Emissions. *J. Environ. Manage.* 294, 113059. doi:10.1016/j.jenvman.2021.113059

Roman-Collado, R., Cansino, J. M., and Botia, C. (2018). How Far Is Colombia from Decoupling? Two-Level Decomposition Analysis of Energy Consumption Changes. *Energy* 148, 687–700. doi:10.1016/j.energy.2018.01.141

Saboori, B., Sulaiman, J., and Mohd, S. (2012). Economic Growth and CO2 Emissions in Malaysia: a Cointegration Analysis of the Environmental Kuznets Curve. *Energy Policy* 51, 184–191. doi:10.1016/j.enpol.2012.08.065

Salman, M., Long, X., Dauda, L., Mensah, C. N., and Muhammad, S. (2019). Different Impacts of export and Import on Carbon Emissions across 7 ASEAN Countries: A Panel Quantile Regression Approach. *Sci. Total Environ.* 686, 1019–1029. doi:10.1016/j.scitotenv.2019.06.019

Selden, T. M., and Song, D. (1994). Environmental Quality and Development: Is There a Kuznets Curve for Air Pollution Emissions?. *J. Environ. Econ. Manage.* 27 (2), 147–162. doi:10.1006/jeem.1994.1031

Sorrell, S., Lehtonen, M., Stapleton, L., Pujol, J., and Champion, T. (2012). Decoupling of Road Freight Energy Use from Economic Growth in the United Kingdom. *Energ. Pol* 41, 84–97. doi:10.1016/j.enpol.2010.07.007

Tapio, P. (2005). Towards a Theory of Decoupling: Degrees of Decoupling in the Eu and the Case of Road Traffic in Finland between 1970 and 2001. *Transport Pol.* 12, 137–151. doi:10.1016/j.tranpol.2005.01.001

Tiba, S., and Frikha, M. (2018). Income, Trade Openness and Energy Interactions: Evidence from Simultaneous Equation Modeling. *Energy* 147, 799–811. doi:10.1016/j.energy.2018.01.013

Tiba, S., Omri, A., and Frikha, M. (2015). The Four-Way Linkages between Renewable Energy, Environmental Quality, Trade and Economic Growth: a Comparative Analysis between High and Middle-Income Countries. *Energy Syst* 7, 103–144. Available at: http://refhub.elsevier.com/S0140-9883(20) 30306-6/rf0425. doi:10.1007/s12667-015-0171-7

Torrence, C., and Compo, G. (1998). A Practical Guide to Wavelet Analysis. *Bull. Am. Meteorol. Soc.* 79 (61-78), 605–618. doi:10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

Torrence, C., and Webster, P. (1999). Interdecadal Changes in the Esnom on Soon System. *J. Clim.* 12, 2679–2690. Available at: http://refhub.elsevier.com/S0264-9993(13)00419-7/rf1010. doi:10.1175/1520-0442(1999)012<2679:icitem>2.0.co;2

Wang, Q., and Su, M. (2020). Drivers of decoupling economic growth from carbon emission e an empirical analysis of 192 countries using decoupling model and decomposition method. *Environ. Impact Assess. Rev.* 81, 106356. doi:10.1016/j.eiar.2019.106356

Wang, Q., and Wang, S. (2020). Is Energy Transition Promoting the Decoupling Economic Growth from Emission Growth? Evidence from the 186 Countries. *J. Clean. Prod.* 260, 120768. doi:10.1016/j.jclepro.2020.120768

Wang, Q., and Zhang, F. (2020). Does Increasing Investment in Research and Development Promote Economic Growth Decoupling from Carbon Emission Growth? an Empirical Analysis of BRICS Countries. *J. Clean. Prod.* 252, 119853. doi:10.1016/j.jclepro.2019.119853

Wang, Q., and Zhang, F. (2021). The Effects of Trade Openness on Decoupling Carbon Emissions from Economic Growth - Evidence from 182 Countries. *J. Clean. Prod.* 279, 123838. doi:10.1016/j.jclepro.2020.123838

Yu, Y., Chen, D., Zhu, B., and Hu, S. (2013). Eco-efficiency Trends in China, 1978–2010: Decoupling Environmental Pressure from Economic Growth. *Ecol. Indicators.* 24, 177–184. doi:10.1016/j.ecolind.2012.06.007

Zhang, J., Fan, Z., Chen, Y., Gao, J., and Liu, W. (2020). Decomposition and Decoupling Analysis of Carbon Dioxide Emissions from Economic Growth in the Context of China and the ASEAN Countries. *Sci. Total Environ.* 714, 136649. doi:10.1016/j.scitotenv.2020.136649

Zhang, J. K., and Zhang, Y. (2020). Chinese Tourism Economic Change under Carbon Tax Scenarios. *Curr. Issues Tourism* 23 (7), 836–851. doi:10.1080/13683500.2018.1551339

Zhang, M., Song, Y., Su, B., and Sun, X. M. (2015). Decomposing the Decoupling Indicator between the Economic Growth and Energy Consumption in China. *Energ Effic* 8, 1231–1239. Available at: https://link.springer.com/article/10.1007/s12053-015-9348-0. doi:10.1007/s12053-015-9348-0

Zhang, Y. (2011). Interpreting the Dynamic Nexus between Energy Consumption and Economic Growth: Empirical Evidence from Russia. *Energy Policy* 39, 2265–2272. Available at: https://10.1016/j.enpol.2011.01.024. doi:10.1016/j.enpol.2011.01.024

Zhao, X., Zhang, X., Li, N., Shao, S., and Geng, Y. (2017). Decoupling Economic Growth from Carbon Dioxide Emissions in China: a Sectoral Factor Decomposition Analysis. *J. Clean. Prod.* 142, 3500–3516. doi:10.1016/j.jclepro.2016.10.117

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Xiang, Chen, Su and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 23 August 2021
doi: 10.3389/fenrg.2021.730640

# Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption

*Yang Xu*[1,2], *Weijun Gao*[1,2]*, *Fanyue Qian*[2] *and Yanxue Li*[1]

[1]Innovation Institute for Sustainable Maritime Architecture Research and Technology, Qingdao University of Technology, Qingdao, China, [2]Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan

Predicting system energy consumption accurately and adjusting dynamic operating parameters of the HVAC system in advance is the basis of realizing the model predictive control (MPC). In recent years, the LSTM network had made remarkable achievements in the field of load forecasting. This paper aimed to evaluate the potential of using an attentional-based LSTM network (A-LSTM) to predict HVAC energy consumption in practical applications. To evaluate the application potential of the A-LSTM model in real cases, the training set and test set used in experiments are the real energy consumption data collected by Kitakyushu Science Research Park in Japan. Pearce analysis was first carried out on the source data set and built the target database. Then five baseline models (A-LSTM, LSTM, RNN, DNN, and SVR) were built. Besides, to optimize the super parameters of the model, the Tree-structured of Parzen Estimators (TPE) algorithm was introduced. Finally, the applications are performed on the target database, and the results are analyzed from multiple perspectives, including model comparisons on different sizes of the training set, model comparisons on different system operation modes, graphical examination, etc. The results showed that the performance of the A-LSTM model was better than other baseline models, it could provide accurate and reliable hourly forecasting for HVAC energy consumption.

**Keywords: energy consumption prediction, ultra-short-term forecast, deep learning, LSTM network, attention mechanism**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University of Economics, China*

***Reviewed by:***
*Yi Zong,*
*Technical University of Denmark, Denmark*
*Muhammad Mohsin,*
*Jiangsu University, China*

***Correspondence:**
*Weijun Gao*
*gaoweijun@me.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *25 June 2021*
***Accepted:*** *04 August 2021*
***Published:*** *23 August 2021*

***Citation:***
*Xu Y, Gao W, Qian F and Li Y (2021) Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption. Front. Energy Res. 9:730640. doi: 10.3389/fenrg.2021.730640*

## INTRODUCTION

According to statistics, building energy consumption accounts for about 40% of the total energy consumption (Mohsin et al., 2020), and the proportion of building carbon dioxide emissions is as high as 36% of the total emissions (Jradi et al., 2017; Mohsin et al., 2021). Heating, ventilation, and air-conditioning (HVAC) systems account for 40% (or even higher) of the commercial building energy consumption (Kim et al., 2019; Wang et al., 2021). Higher energy consumption usually means greater energy saving potential, so the research of energy-saving technology combining big data and artificial intelligence has become one of the hot spots in recent years (Sun et al., 2021). Simulation modeling is mainly used to predict energy consumption in the design stage of HVAC. The factors considered include building physical parameters, outdoor meteorological parameters, indoor environmental parameters, and room usage (Ma et al., 2017). Since there are different degrees of

assumptions and simplifications in the process of establishing the model, there may be a large error between the predicted energy consumption in the design stage and the actual operation time. The data-driven forecasting model is mainly used to predict energy consumption in the operation stage of HVAC. The HVAC energy consumption is affected by many factors, not only by the building itself, but also by meteorological conditions (such as outdoor temperature and illumination), internal personnel activities (such as occupancy and power consumption), time lag, and the actual use of air conditioning (such as control deviation and operation scheme adjustment) (Wang et al., 2016). Under the influence of these factors, the HVAC load curve has the characteristics of strong fluctuation, large randomness, and not obvious periodicity compared with the electric load curve. This presents great challenges in designing data-driven HVAC models. To solve this problem, Wasim Iqbal et al. proposed the method of negative Binomial regression (NBR) model analysis (Iqbal et al., 2021), BinbinYu proposed a method based on the Dynamic Spatial Panel Model (DSPM) Model (Yu, 2021), and WeiqingLi et al. proposed the data envelopment analysis (DEA) and entropy method (Li et al., 2021a) to analyze the interaction between various factors. At present, the operating parameters of the HVAC system in buildings are mainly set according to the load prediction in the design stage. It will lead to the HVAC system working in a state of low efficiency, resulting in a large amount of energy consumption.

To solve this problem, Model predictive control (MPC) was proposed and had been widely used (Mayne, 2014; Sultana et al., 2017; Zhan and Chong, 2021). There have been many examples of load forecasting applied to MPC in other energy fields. For example, Xueyuan Zhao proposed a short-term household load forecasting model to guide users' power consumption behavior and adjust power grid load (Zhao et al., 2021); Based on the forecast of wind power generation and electricity price, J.J. Yang proposed a charge-discharge control strategy for energy storage equipment based on the data drive, which realized the maximization of income of energy storage equipment (Yang et al., 2020a). Fangyuan Chang proposed a charging cost control solution based on reinforcement learning by using long and short-term memory networks (LSTM) to predict real-time electricity prices (Chang et al., 2020). Jin Li proposed an adaptive genetic algorithm based on neighborhood search to optimize the total cost related to energy consumption and running time of electric vehicles. (Li et al., 2020). This shows that predicting system energy consumption accurately and adjusting dynamic operating parameters of the HVAC system in advance is the basis of realizing MPC (Hazyuk et al., 2012).

Load forecasting can be divided into ultra-short-term, short-term, medium-term, and long-term according to different purposes, and the ultra-short-term load forecast refers to the load forecast within 1 h in the future (Guo et al., 2018; Qian et al., 2020). Due to the time delay of building thermal load, the change of influencing factors in the short term cannot immediately change the HVAC energy consumption (Huang and Chow, 2011). Therefore, the prediction of ultra-short-term load can minimize the impact of the uncertainty of input variables on the prediction results. The behavior of HVAC can be expressed as time-series data with a certain periodically. The prediction model can learn the load mode of the system from the time-series data, and use these modes to make load predictions. After the prediction range is determined, the appropriate algorithm needs to be selected. The algorithms in the field of time series prediction can be divided into two categories: traditional machine learning algorithms and deep learning algorithms. Traditional machine learning algorithms are mostly based on statistical models, Current popular algorithms include Autoregressive Moving Average (ARIMA) (Liu et al., 2021), support vector machine (SVM) (Ma et al., 2018), Regression Tree (Yang et al., 2020b), Random Forest (Li et al., 2021b), and artificial neural networks (ANN) (Wei et al., 2019; Bui et al., 2020).

In recent years, with the rapid development of deep learning, the deep neural network has been more and more applied in the field of load prediction. Deep learning is a series of new structures and new methods evolved based on multi-layer neural networks (Lv et al., 2021a). Deep learning models have obvious advantages over traditional machine learning models in predicting multivariable time series problems. In the real world, time series prediction presents multiple challenges, such as having multiple input variables, the need to predict multiple time steps, and the need to perform the same type of prediction for multiple actual observation stations (Askari et al., 2015). In particular, a deep learning model can support any number but a fixed number of inputs and outputs. Multivariable time series have multiple time-varying variables, each of which depends not only on its past value but also on other variables. These characteristics are correlated with each other, and in this case, multiple variables need to be considered to give the best-predicted energy consumption.

The most basic Deep learning model is Deep Neural Networks (DNN), also known as multi-layer perceptron (MLP). DNN has more hidden layers than ordinary artificial neural networks, which gives it the ability to learn complex patterns. Deb proposed a DNN - based model for predicting the daily cooling load of buildings (Deb et al., 2016). Massana proposed a DNN model-based method for predicting short-term electrical loads in non-residential buildings (Massana et al., 2015). Zhihan Lv et al. proposed a layered DAE support vector machine (SDAE-SVM) model based on a three-layer neural network and achieved good prediction results (Lv et al., 2021b). However, the DNN model cannot retain time-series information. It can only be predicted according to the current input and output values, and cannot learn the time dependence of data, which limits the accuracy of its prediction time series.

Recursive neural network (RNN), as a special deep neural network, can retain and consider the time variation of time series in the training process (Hochreiter and Schmidhuber, 1997), which makes it very suitable for time series data with periodicity. The time series of HVAC energy consumption is influenced by environmental variables and human living habits and has a strong periodicity. However, due to the problems of gradient explosion and gradient disappearance in the RNN network, the long-term dependence in time series cannot be retained, which limits the prediction accuracy of the RNN network. The long-short term memory (LSTM) network adds

a series of multi-threshold gates based on the RNN network, which can deal with a long-term dependency relationship to a certain extent. LSTM networks were first used in natural language processing (Verwimp et al., 2020), machine translation (Su et al., 2020) and video recognition (Lv et al., 2021c), etc. In recent years, LSTM networks have attracted more and more attention in the field of load prediction. The authors of (Wang et al., 2019a) used the LSTM network to build a regional scale building energy consumption prediction model. Sendra-Arranz proposed a variety of multi-step prediction models based on the LSTM network to predict residential HVAC consumption (Sendra-Arranz and Gutiérrez, 2020). Zhe Wang proposed a new method for predicting plug loads using the LSTM network. Through the data collected from a real office building in Berkeley, California, the prediction accuracy of this method was verified to be better than that of the traditional machine learning algorithm (Wang et al., 2019b; Wang et al., 2020).

Since the LSTM network adopts the code-decoding framework, the limitations of the code-decoding framework will lead to information loss when processing long time series. Bahdanau first introduced the attention mechanism into the code-decoding framework in 2014 (Bahdanau et al., 2014). The attention mechanism can quantitatively assign a weight to each specific time step in the time series feature, which improves the attention distraction defect of traditional LSTM (Li et al., 2019). As far as the author knows, the LSTM model based on the ATTENTION mechanism has not been applied in the field of HVAC energy consumption prediction, but some researchers have started experiments in other load fields and achieved some results (Lu et al., 2017; Yu et al., 2017; Liang, et al, 2018). Such as Heidari used an attention-based LSTM (A-LSTM) model to predict the load of the solar-assisted hot water system and proved that the prediction accuracy of the A-LSTM model was better than that of the traditional LSTM model (Heidari and Khovalyg, 2020). Jince Li proposed an improved attention-based LSTM (A-LSTM) model for multivariate time series of predictions of two process industry cases (Li, et al, 2021c). Tongguang Yang proposed an attention-based LSTM model to predict the day-ahead PV power output (Yang et al., 2019). All these cases show that the A-LSTM model has significant advantages over the traditional LSTM model in dealing with time series problems.

This paper aims to develop a new HVAC ultra-short-term energy consumption prediction model. To achieve this goal, this paper first conducted potential rule analysis and feature engineering for 9 years' operation data of Kitakyushu Science and Research Park's (KSRP) Energy Center and constructed a data set for modeling. Then, we developed the LSTM model based on this data set and developed the A-LSTM model by adding the attention layer to the LSTM model. Besides, the RNN model, DNN model, and SVR model were also developed to compare performance. The hyper-parameters of the above models were optimized by the TPE algorithm to ensure prediction accuracy. Next, we used the data from the Energy Center from 2002 to 2009 as the training set and the data from 2010 as the test set to conduct experiments, and gradually reduced the size of data sets to evaluate the performance of the above five models in different training sets. Finally, we also evaluated the small-scale prediction effects of the above five models under four typical operating modes.

Based on the literature review the main novelty of the paper can be summarised as follows:

- This paper studies and verifies that the deep learning method with the attention mechanism has advantages in the memory and feature selection of time series information of air conditioning load prediction. As far as the author knows, the LSTM model based on the ATTENTION mechanism has not been applied in the field of HVAC energy consumption prediction.
- In this paper, we use the Tree-Structured of Parzen Estimators (TPE) algorithm to optimize the parameters of each baseline model, to ensure that all the baseline models involved in horizontal comparison are in the best prediction performance, and to ensure the authenticity of the prediction performance comparison
- In this paper, we used the real building data set instead of the standard data set to verify the load prediction potential of the A-LSTM model, which verified the value of this model in actual engineering. The results show that the A-LSTM model has advantages over the traditional machine learning algorithm in various operating modes, especially in the heating high load period, and has great potential in the field of ultra-short-term load prediction of HVAC.

# METHODOLOGY

## LSTM Neural Network

RNN network is a kind of neural network used to process time series. Compared with the traditional DNN network and CNN network, RNN adopts a cyclic structure to replace the hidden layer of the feedforward neural network. In the process of information transmission, there will be a part of the information left in the current neuron during each cycle, and the retained information will be used as the input of the next neural unit with the new information. In this way, the RNN network implements "memory". However, when the input time series is too long, it is difficult to retain the information in the RNN network, this is also known as the phenomenon of gradient disappearance and gradient explosion (Hochreiter and Schmidhuber, 1997).

Based on the RNN unit, the input gate $i_t$, output gate $o_t$, forgetting gate $f_t$ and cell state $C_t$ are added to the LSTM unit to control the inheritance and abandonment of information. There are three inputs of the LSTM unit: the input vector $x_t$ at the current time slot t, the unit state $C_{t-1}$ at the time slot t−1, and the state of the hidden layer $h_{t-1}$ at the time slot t−1. The final output of the LSTM unit is the cell state $C_t$ at the current time slot t and the state of the hidden layer at the current time $h_t$ To figure out the $h_t$, we first let the $W_i$, $W_o$ and $W_f$ be the weight matrix of the input gate, the output gate, and the forgetting gate, and let the $[h_{t-1}, x_t]$ represent the combination of the hidden state at the moment t−1 and the input of the unit at the time slot t into a new vector. Besides, let the $b_i$, $b_o$, and $b_f$ be their bias vectors, and let

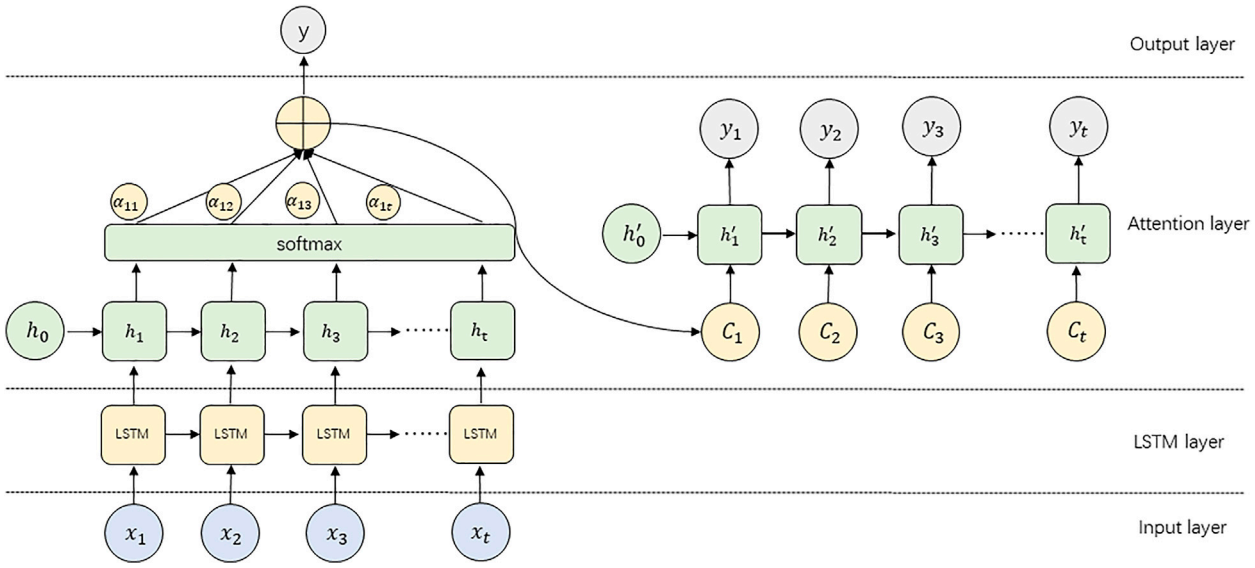



**FIGURE 1 |** The structure of the A-LSTM model.

the $\sigma$ represent the Sigmoid activation function. The formulas of the $i_t$, $o_t$, and $f_t$ are shown below:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{1}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2}$$

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

Finally, let tan$h$ represent the activation function, * represent the Hadamard product, and let the $\widetilde{C}_t$ be the state of the intermediate unit input at time slot t, then we can calculate the $h_t$ as follows:

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

## Attention Mechanism

Although the LSTM model has a memory function, and it can save some time-series information, but because the standard LSTM model uses the traditional encoder-decoder structure, it still has some limitations. When processing the time series $x_t$, the Encoder will first encode the input sequence into a fixed-length implicit vector $h$ and give the same weight to the implicit vector. However, when the length of $x_t$ increases, the average weight distribution will reduce the discrimination of $x_t$, and some important time-series information will be ignored in the process of training the model, thus affecting the prediction accuracy of the model.

The Attention mechanism is a mechanism used to optimize the Encoder-Decoder structural model, It can be combined with a variety of models depending on the actual situation. An encoder-decoder model with an Attention mechanism first learns the weight of each element from the sequence and then recombines the elements by weight. By assigning different weight parameters to each input element, the Attention mechanism can focus more on the parts that are relevant to the input element, thereby suppressing other useless information. Its biggest advantage is that it can consider global connection and local connection in one step, and can realize the parallel computation, which is particularly important for big data computation. In this paper, the Bahdanau algorithm (Bahdanau et al., 2014) is adopted to realize the Attention mechanism, and the structure of the A-LSTM model adopted is shown in **Figure 1**:

The calculation process of the Attention layer is shown in **Eqs 7–9**. i denotes the moment; $j$ denotes the j element in the sequence; $T$ denotes the length of the sequence; $\alpha_{tj}$ denotes the weight value; $e_{tj}$ is the matching degree between the element to be encoded and other elements.

$$e_{tj} = a\left(h_{i-1}, h_j\right) \tag{7}$$

$$\alpha_{tj} = \frac{\exp\left(e_{tj}\right)}{\Sigma_{k=1}^{t} \exp\left(e_{tj}\right)} \tag{8}$$

$$C_t = \sum_{j=1}^{T} \alpha_{tj} h_i \tag{9}$$

## Tree-Structured of Parzen Estimators

The performance of machine learning models largely depends on the selection of hyperparameters. With the increase of model complexity and the amount of training data, automatic hyperparameter optimization plays an increasingly important role in the developments (Nguyen et al., 2020). Compared with traditional manual parameter adjustment, automatic parameter optimization has the following advantages: 1) It reduces the manpower of development work. 2) Improve the performance of machine learning models. 3) Improve the

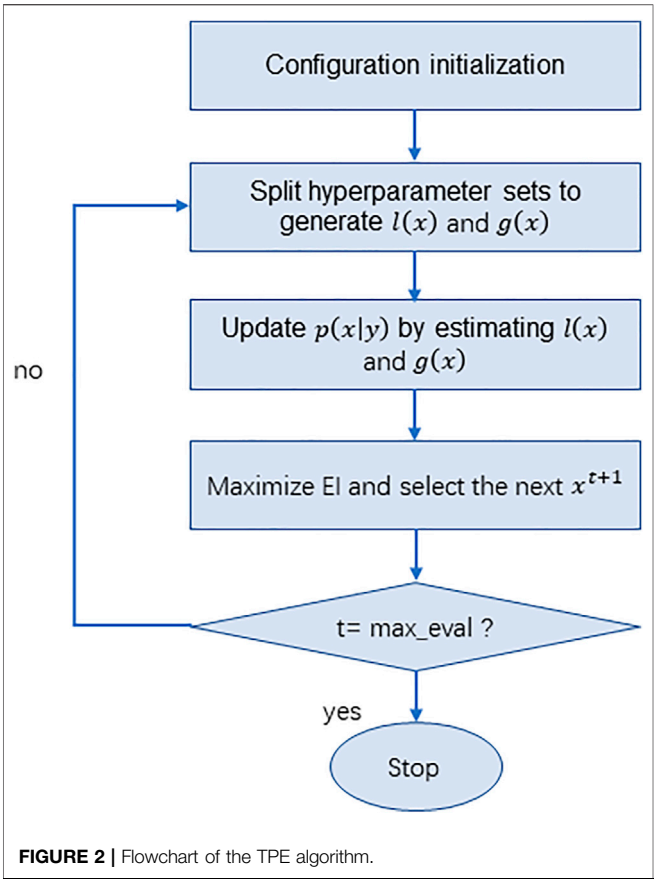



**FIGURE 2 |** Flowchart of the TPE algorithm.

reproducibility of the results (Luo et al., 2021). In this paper, the Tree-Structured of Parzen Estimators (TPE) algorithm was used to achieve automatic optimization of the model's hyperparameters. TPE algorithm is an improved algorithm of Bayesian optimization algorithm (BO). It solves the limitation of the traditional BO algorithm to deal with classification parameters and conditional parameters, so it has higher efficiency.

The main process of the TPE algorithm is to convert the hyperparametric space into the nonparametric density distribution first, and then model the process $p(x|y)$. As shown in **Eq. 10**, TPE uses two density distributions of Equation to define $p(x|y)$, $y < y^*$ indicates that the value of the objective function is less than the threshold, and $y \geq y^*$ denotes that the value of the objective function is greater than or equal to the threshold.

$$p(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases} \tag{10}$$

The calculation of Expected Improvement (EI) is shown in **Eqs 11–13**.

$$E(x) = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} \mathrm{d}y \tag{11}$$

$$\gamma = p(y < y^*) \tag{12}$$

$$P(x) = \int_R P(x|y)P(y)\mathrm{d}y \tag{13}$$

Substitute **Eqs 12**, **13** into **Eq. 11** to get the final **Eq. 14**.

$$EI_{y^*}(x) = \left(\gamma + \frac{g(x)}{l(x)}(1-\gamma)\right)^{-1} \tag{14}$$

It can be seen from **Eq. 13** that point $x^*$ with the largest Ei is the point with the smallest $g(x)/l(x)$. The TPE algorithm evaluates the improvement points according to $g(x)/l(x)$ in each iteration, and finally returns a point $x^*$ with the largest EI. The corresponding process is shown in **Figure 2**.

# CASE STUDY

## Case Introduction

In this paper, data collected by the CCHP system of Kitakyushu Science Research Park (KSRP) in Japan were used as the research object. The KSRP system was a distributed energy system consisting of a gas engine (160 kW), a fuel cell (200 kW), and a photovoltaic system (150 kW). The system mainly supplied energy to the main teaching building of The University of Kitakyushu, which can meet the teaching and office needs of more than 3,000 people. The target building was divided into four floors, the first floor included the student center, meeting rooms and classrooms. The second to fourth floors were teachers' offices and student research rooms.

The system was powered by the gas engine, fuel cells, photovoltaic system and the utility grid. The cooling load, heating load and hot water load were mainly provided by the absorption chiller, while the gas engine and fuel cell also provided part of the cooling and heating load when generating electricity. The basic schematic diagram of CCHP system at KSRP was shown in **Figure 3**. The system also included a detailed data acquisition system that recorded not only detailed operational data for each device, but also environmental data related to the target building. Using the temperature and flow data collected by the data acquisition system for hot and cold water supply and recovery, we could calculate the hot and cold load requirements of the target building in real-time. The KSRP cogeneration system was established in 2001. To make the model reflect the most real operating state of the system, we selected the data from 2002 to 2010 as the research object (78,820 pieces of data), because only 3 days of system failure occurred during this period, which could reduce the impact of data missing on the modeling.

## Potential Analysis of Input Data Set

We first calculated the cooling and heating output of the equipment from January 1, 2002, to December 31, 2010, based on the gas consumption of the equipment and the annual average COP(cooling 1.00, heating 0.85). To verify the authenticity of the data, we also calculated the cooling and heating output based on the temperature and flow rate of cold and hot water supply and

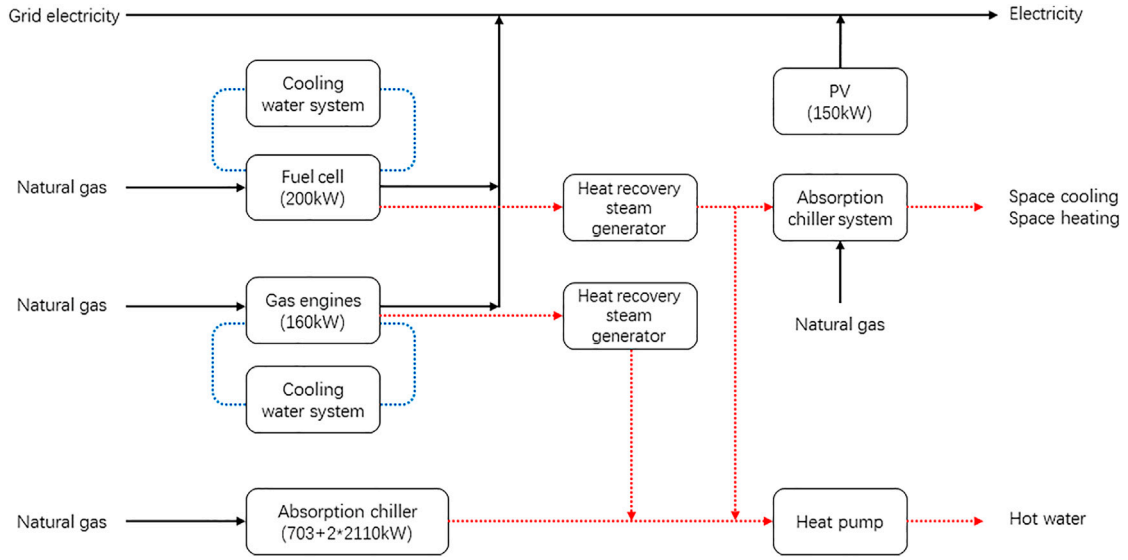



**FIGURE 3 |** The basic schematic diagram of the CCHP system at KSRP.

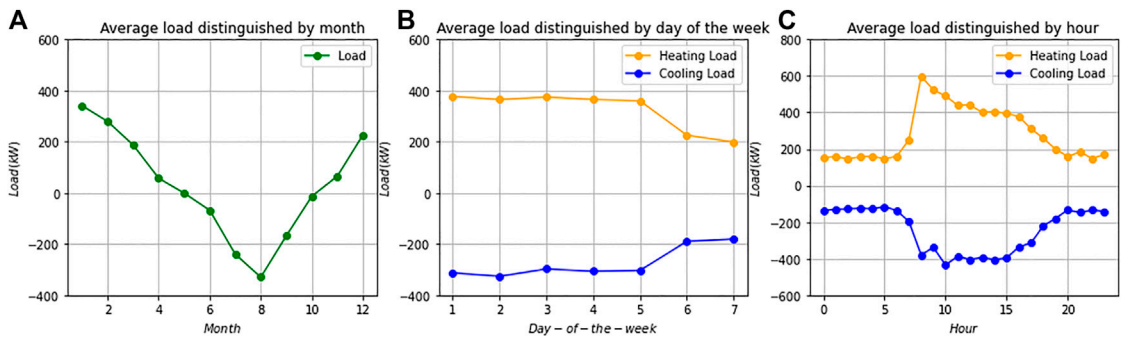



**FIGURE 4 |** The diagram of average load distributed by the time: **(A)** average load distinguished by month, **(B)** average load distinguished by day of the week, **(C)** average load distinguished by hour.

recovery collected by the system. To explore the distribution rule of these data in time series, we calculated the mean value of load in units of the month, week and hour respectively, and the results are shown in **Figure 4**. As can be seen from **Figure 4A**, the average load varies greatly each month. The annual peak value of total heating load output occurs in January, and that of total cooling load output occurs in August. Therefore, December, January, February, and March were defined as the heating season; July, August, and September as the cooling season; April, May, June, October, and November as the low-load season. The prediction effect of the model will be evaluated respectively according to this division. It can be seen from **Figure 4B** that the cooling and heating loads are higher on weekdays than on weekends; It also can be seen from **Figure 4C** that the average daily load distribution in the heating season and the cooling season is significantly different. All of the above time information can reflect the impact of human activities on load, so they can be used as characteristic factors for database construction.

We also selected other environmental factors that might affect the heat and cold output to build the initial database, including data collected by the Energy Center every hour from January 1, 2002, to December 31, 2010, a total of 78,820 pieces of data. Each group of data includes time information, outdoor temperature (°C), relative humidity (%), irradiance ($W/m^2$), wind speed (m/s), wind direction, and load output (The positive load indicates the heating load, and the negative load indicates the refrigeration load). In addition, the index "trend" indicates the ordinal number of the data in the time series. Since features with small correlation will provide unnecessary information in the training of the model, which will affect the robustness of the model, Pearson correlation analysis was conducted for all features, and the results are shown in **Figure 5**. It can be seen that the biggest factor affecting the load is outdoor temperature, followed by the time serial number. The correlation between wind direction and load was too small (−0.0087), and we had deleted this feature in later modeling. Examples of these data were shown in **Table 1**. In the following experiments, we used the data from 2002 to 2009 as the training

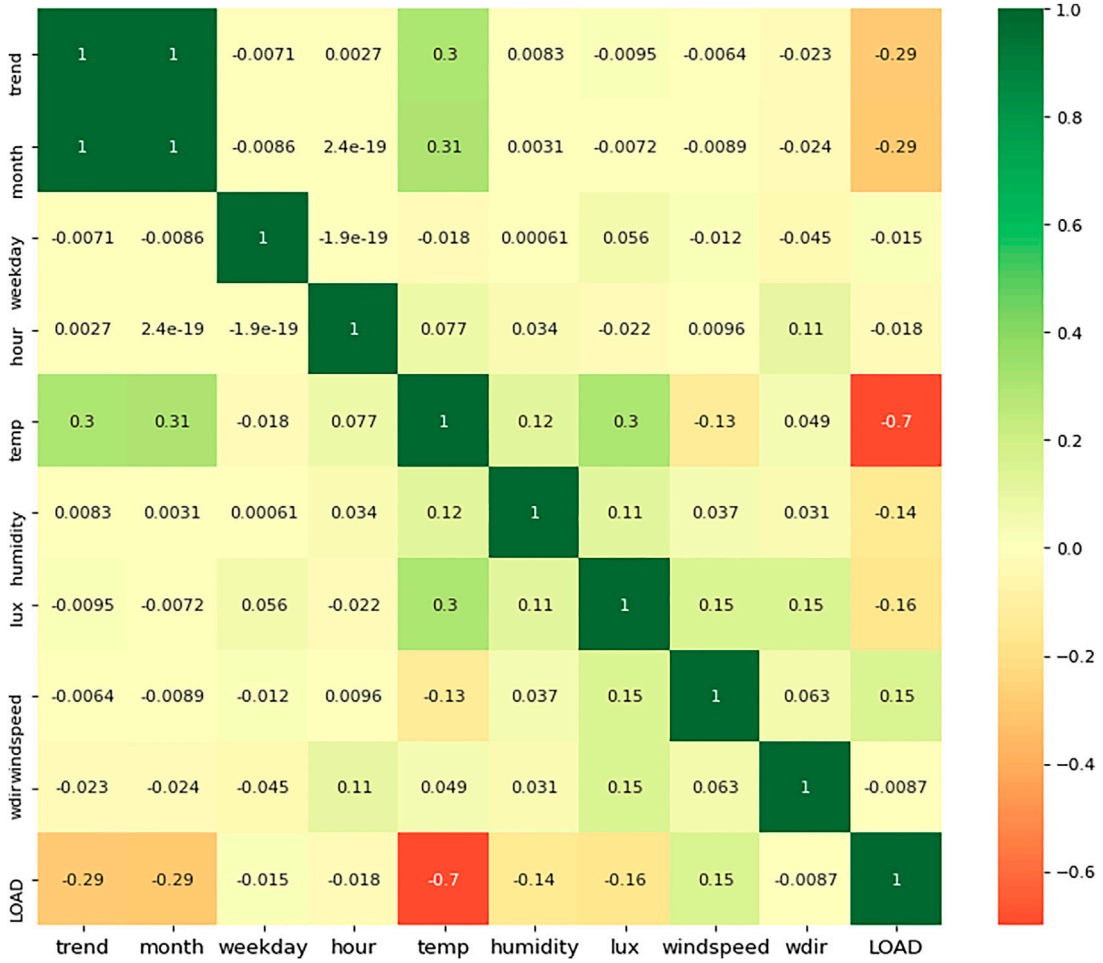



**FIGURE 5 |** Correlation analysis between the available features.

**TABLE 1 |** Example of the database.

| Trend | Month | Weekday | Hour | Temperature (°C) | Humidity (%) | Illuminance (W/m$_2$) | Windspeed (m/s) | Load (kW) |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 5 | 1 | 1.7 | 18 | 0 | 3.6 | 0 |
| 1 | 1 | 5 | 2 | 1.4 | 19 | 0 | 9.1 | 0 |
| 2 | 1 | 5 | 3 | 1.4 | 19 | 0 | 5.5 | 147.208 |
| 3 | 1 | 5 | 4 | 1.2 | 20 | 0 | 6.9 | 165.156 |
| 4 | 1 | 5 | 5 | 1.1 | 20 | 0 | 2.2 | 111.105 |
| 5 | 1 | 5 | 6 | 1.3 | 19 | 0 | 5.7 | 79.517 |
| 6 | 1 | 5 | 7 | 1.6 | 19 | 0 | 6.4 | 127.376 |

set, the data from 2010 as the test set, and then randomly extracted 20% data from the training set as the verification set. Since the numerical dimensions of different variables are greatly different, it is necessary to normalize the data so that the data can be uniformly mapped to the interval [0,1]. In the next section, based on this data set, we will build a prediction model that can predict the next hour's load according to the input data of the first N hours.

# RESULT AND DISCUSSION

## Model Parameter Setting

In this study, we used the Hyperopt framework to implement the TPE algorithm and automatically optimized the hyperparameters of all baseline models. The programming language is Python and the deep learning framework is TensorFlow2.0. Hyperopt is a Python library for hyperparametric optimization based on

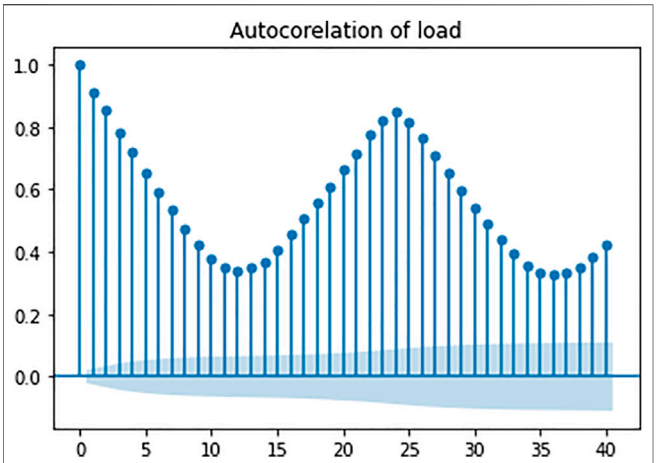



**FIGURE 6 |** Load autocorrelation analysis results.

Bayesian optimization. It supports the optimization of continuous, discrete, and condition variables. Using the Hyperopt framework requires setting four parameters: specifying the objective function to be optimized, defining the search space with super parameters, Trails Database, and the search algorithm. This section will take the LSTM model as an example to outline the method of constructing the model. After the parameter optimization of the LSTM baseline model was completed, we added an attention layer after the hidden layer of the LSTM model to build the A-LSTM model.

The LSTM baseline model needs to optimize four parameters, which are the time step L of each layer in LSTM (using the length of the previous data), the size of the hidden unit m of each layer, the size of the batch processing b in the training process (we used the two-layer LSTM structure, and set the same hidden unit for each layer by default), and the drop rate of the Dropout layer. To determine the range of L, we first performed autocorrelation analysis on load data to identify data cycle patterns, and the results are shown in **Figure 6**. In **Figure 6**, the X-axis represents "hours" and the Y-axis represents the autocorrelation coefficient. We found that the overall autocorrelation of the load is in the form of cycle decline, and the autocorrelation of the load is a cycle of 24 h, which means the autocorrelation peak occurs every 24 h. Therefore, we define the conditional parameters of L as (12, 24, 36, 48). To avoid overfitting, we added a dropout layer after each LSTM layer, and the conditional parameters of drop rate are (0.2, 0.3, 0.4, 0.5). Due to the limited computational force, based on ensuring the prediction accuracy, we set the conditional parameter sets of m and b based on the empirical method: m $\in${32,64,128,256} and b $\in${32,64,128,256} (Wang et al., 2019b). We input the above conditional parameters into the Hyperopt framework and use the TPE algorithm to optimize the model's super parameters. **Figure 7** shows the optimized RNN, LSTM, and A-LSTM model structure, the hyperparameters of these models are determined by the TPE algorithm.

Since the above three models are all recurrent neural networks, we also set up the DNN model and the SVR model for horizontal comparison. These two models also input all the data 24 h before the time slot t and the time data at the time slot t to predict and finally output the load data at the time slot t. The optimal hyperparameters of the DNN model optimized by the TPE algorithm are shown in **Table 2**. The optimal hyperparameters of the SVR model optimized by the TPE algorithm are shown in **Table 3**. The topology of the above five models depends on the characteristics of the KSRP dataset, so for the other datasets, the structure and hyperparameters of the model should be adjusted according to the data. Since the focus of this study was to explore the potential of the LSTM model with the attention mechanism in the field of load prediction. Therefore, on the premise of ensuring the prediction accuracy, the topological structure and input characteristics of the model were simplified as far as possible, to improve the generalization ability of the model and reduce the required computational force.

## Annual Prediction Performance Comparison

To evaluate the time series prediction effect of the A-LSTM model on this data set, we compared it with the same type of RNN, LSTM model, and DNN model without memory function in this experiment. All models have been trained and tested 5 times, and the final data used for comparison is the average of the five test results to reduce the errors caused by random numbers. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-Square Value (R2_SCORE) were used as indicators of the evaluation model, which were calculated according to **Eqs 15–17**. The $y_i$ denotes the real observations, $\overline{y}_i$ denotes the average of the observed value, $\widetilde{y}_i$ denotes the predicted value, N denotes the number of test samples.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2} \tag{15}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widetilde{y}_i| \tag{16}$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}\left(y_i - \widetilde{y}_i\right)^2}{\Sigma_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{17}$$

We first used the data of 8 years (from 2002 to 2009) as the training set to train the models and evaluated the effect of the load forecast in 2010. The results of the five models respectively predicting the annual data of 2010 are shown in **Table 4**. It could be seen that although the prediction results of each model were close to, the prediction accuracy of the A-LSTM model was the highest. Compared with the second-best predicted LSTM, A-LSTM's RMSE decreased by 3.06%, MSE decreased by 6.54%, and $R^2$ value increased by 0.43%. The

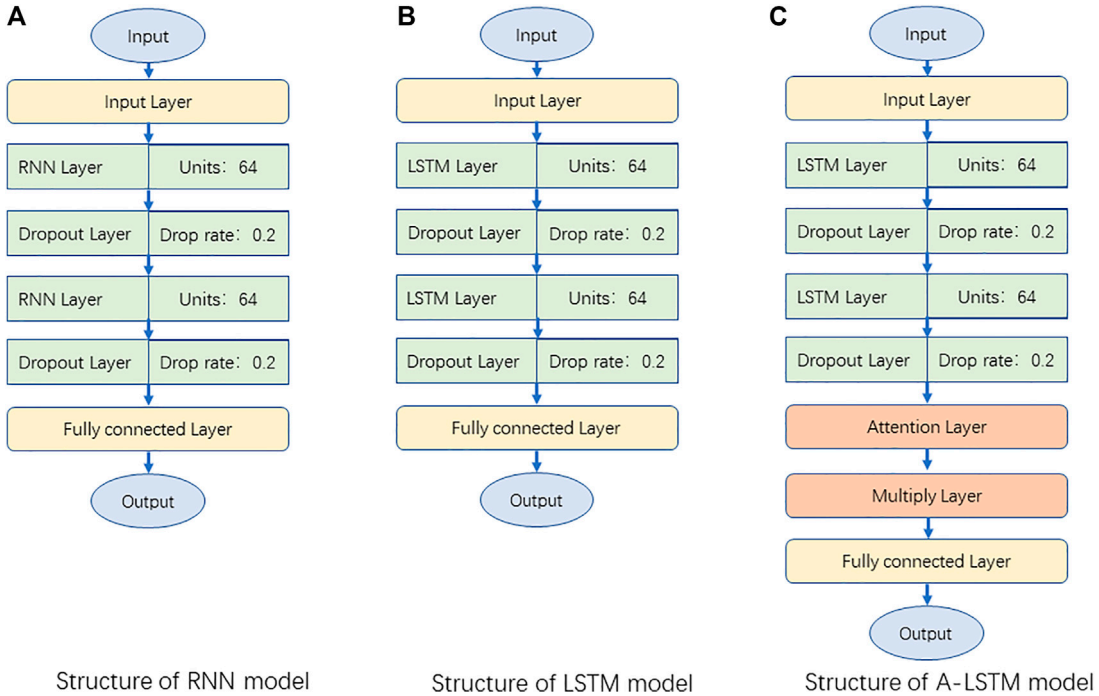



**FIGURE 7 |** Structure of RNN, LSTM, and A-LSTM model.

**TABLE 2 |** Hyperparameters for DNN model.

| Model | Layer1 Units | Layer2 Units | Layer3 Units | Batch size | Drop rate |
|---|---|---|---|---|---|
| DNN | 128 | 64 | 32 | 64 | 0.2 |

**TABLE 3 |** Hyperparamers for SVR model.

| Model | Kernel | C | Gamma |
|---|---|---|---|
| SVR | Rbf | 97.227588 | 0.001032 |

**TABLE 4 |** Comparison of prediction errors between different models.

| | SVR | DNN | RNN | LSTM | A-LSTM |
|---|---|---|---|---|---|
| RMSE (kW) | 106.490 | 78.788 | 80.638 | 77.340 | 74.977 |
| MAE (kW) | 85.632 | 48.752 | 48.979 | 47.929 | 44.793 |
| $R^2$ | 0.854 | 0.922 | 0.918 | 0.925 | 0.929 |

reason why the evaluation results are close is that the system operates under low or zero load for a large amount of time in a year, and the prediction error during these periods is very small, which may reduce the overall average prediction error. We will explore this phenomenon in the next section.

To explore the influence of the size of the training set on the prediction accuracy of the model, we also conducted the following experiments: keeping the topological structure of the above five models unchanged, gradually reducing the training set in a unit of 2 years, and the 2010 data were used as the test set to evaluate each model separately. The experimental results are shown in **Figure 8**. We found that the prediction accuracy of each model decreases with the reduction of the training set. The experiment shows that the prediction accuracy of the A-LSTM model was the best when the data of 8, 6, and 4 years were used as the training set. Compared with the suboptimal LSTM model, its RMSE decreased by 3.06, 10.86, and 11.29%, respectively. $R^2$ value increased by 0.43, 2.21, and 2.57%, respectively. However, when 2 years' data were used as the training set, the prediction accuracy of the A-LSTM model decreased significantly, and its prediction accuracy was only better than that of the SVR model. This indicates that the prediction accuracy of the A-LSTM model will increase with the length of training set, and the prediction accuracy of 4-years or 6-years data sets of the A-LSTM model has obvious advantages compared with other models. This means that when the length of training set is greater than a certain threshold (6–8 years), the advantage of its prediction accuracy will gradually decrease compared with other cyclic neural network models. Besides, when the length of the

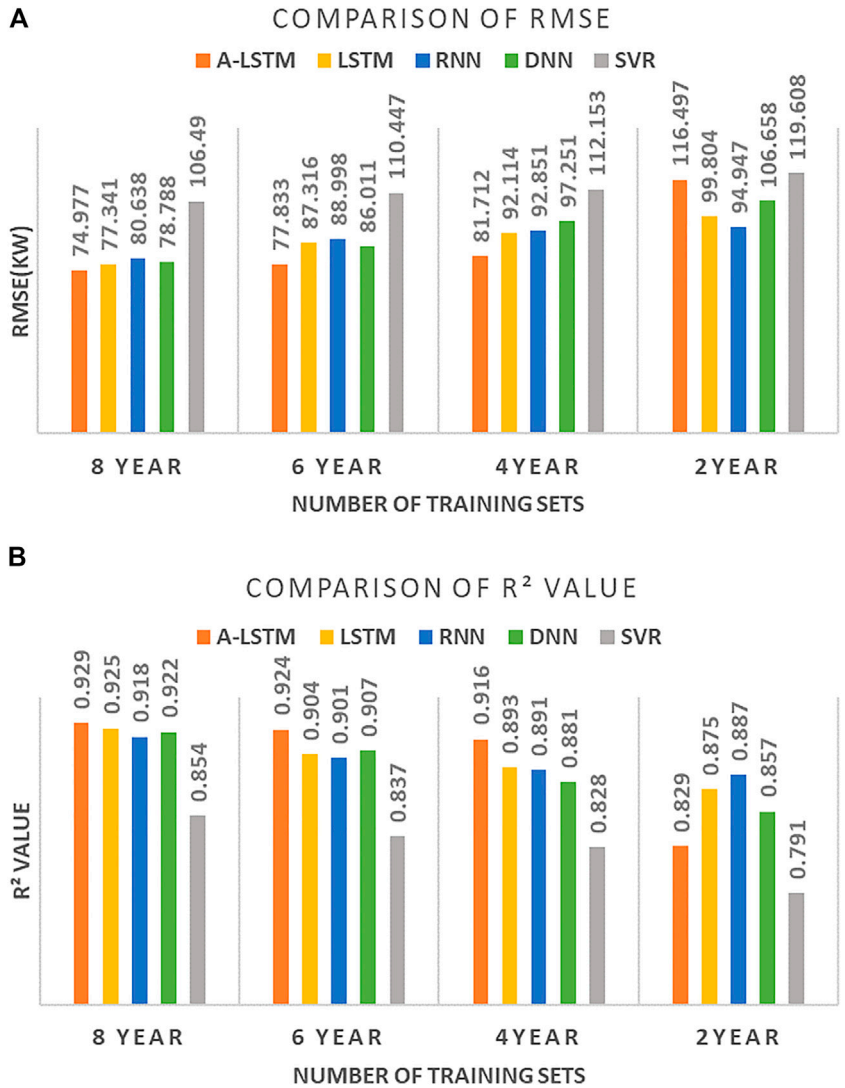



**FIGURE 8 |** The prediction accuracy of each model under different lengths of training set: **(A)** the comparison of RMSE, **(B)** the comparison of $R^2$ value.

**TABLE 5 |** Performance of A-LSTM models compared to the baseline model.

| | 2010.1.1~2010.1.14 | | | | | 2010.8.1~2010.8.14 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **SVR** | **DNN** | **RNN** | **LSTM** | **A-LSTM** | **SVR** | **DNN** | **RNN** | **LSTM** | **A-LSTM** |
| RMSE (kW) | 123.508 | 103.035 | 93.796 | 96.558 | 86.876 | 108.849 | 78.129 | 79.959 | 75.294 | 68.361 |
| MAE (kW) | 102.083 | 72.996 | 66.898 | 66.192 | 62.262 | 88.199 | 57.575 | 54.966 | 50.127 | 46.529 |
| $R^2$ | 0.737 | 0.817 | 0.839 | 0.848 | 0.870 | 0.800 | 0.896 | 0.892 | 0.904 | 0.921 |
| | **2010.3.1~2010.3.14** | | | | | **2010.5.16~2010.5.30** | | | | |
| | **SVR** | **DNN** | **RNN** | **LSTM** | **A-LSTM** | **SVR** | **DNN** | **RNN** | **LSTM** | **A-LSTM** |
| RMSE (kW) | 99.075 | 79.973 | 81.230 | 78.325 | 73.427 | 68.277 | 54.021 | 50.452 | 53.515 | 47.805 |
| MAE (kW) | 78.165 | 48.027 | 53.816 | 49.141 | 47.206 | 56.184 | 35.185 | 31.627 | 38.351 | 29.855 |
| $R^2$ | 0.521 | 0.687 | 0.678 | 0.701 | 0.737 | 0.642 | 0.776 | 0.805 | 0.781 | 0.824 |

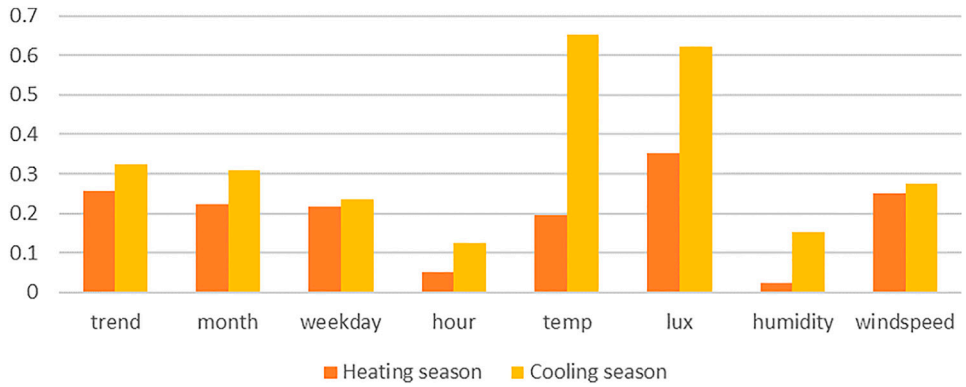



**FIGURE 9 |** Absolute value of correlation coefficients of load and other features in the database between heating season and cooling season.

training set is less than A certain threshold value (4–2 years), the prediction accuracy of the A-LSTM model will decrease significantly.

## Prediction Performance Comparison at High and Low Loads

In the previous section, we discussed how a large number of zero-load and low-load forecasts over a year may affect the average error. To more intuitively evaluate the prediction effect of the A-LSTM model, we selected data of 2 weeks in each of four periods in the 2010 years for comparison. Among the data selected for the experiment, two groups were high-load period data (2010.1.1 to 2010.1.14 and 2010.8.1 to 2010.1.14), and the other two groups were low-load period data (2010.3.1 to 2010.3.14 and 2010.5.16 to 2010.5.30). the results are shown in **Table 5**. As can be seen from **Table 4**, the prediction accuracy of the A-LSTM model was significantly higher than that of other models. In the high heating load stage, compared with the second-best predicted LSTM, A-LSTM's RMSE decreased by 10.02%, MSE decreased by 5.93%, and $R^2$ value increased by 2.59%; In the high cooling load stage, RMSE of A-LSTM decreased by 9.21%, MSE decreased by 8.80%, and $R^2$ value increased by 1.88%, compared with that of the second-best predicted LSTM. In the low heating load stage, compared with the second-best predicted LSTM, A-LSTM's RMSE decreased by 6.25%, MSE decreased by 3.94%, and $R^2$ value increased by 5.14%; In the low cooling load stage, RMSE of A-LSTM decreased by 5.24%, MSE decreased by 4.47%, and $R^2$ value increased by 2.36%, compared with that of the second-best predicted RNN. This indicates that compared with the low-load stage, A-LSTM in the high-load stage has an obvious improvement compared with other models, which also indicates that A-LSTM has more potential in peak prediction. It can be seen from **Table 5** that the prediction accuracy of the five models for the cooling load is higher than that for the heating load. Taking the A-LSTM model as an example, RMSE decreased by 18.515 (kW), MAE decreased by 15.733 (kW), and $R^2$ value increased by 0.051 in the peak cooling period compared with the peak heating period. This change was also evident during periods of low load.

To explain this phenomenon, the characteristic correlation coefficients of the cooling season and heating season were statistically analyzed. **Figure 9** can explain the reasons for the above phenomena from one perspective. It can be seen from **Figure 9** that the correlation coefficients between the load and other characteristics in the refrigeration season are higher than those in the heating season, especially the temperature, illumination, and humidity. This indicates that the output of cooling load is more affected by environmental factors, while the output of heat load is more affected by the laws of human production and life. The existing data cannot fully reflect the laws of human production and life, but it reflects the environmental factors more comprehensively, so this phenomenon occurs.

The actual prediction curves corresponding to **Table 5** are shown in **Figure 10**. As can be seen from the fitting curve results, the load output of the KSRP system in the high load stage was distributed discretely, and it fluctuated greatly in the short term, as the peaks and troughs often appear alternately in the time series. Except for the SVR model, other models had a good fitting effect. By comparing $R^2$ values in **Table 5**, we also found that the prediction curve fitting rate of all models, including the A-LSTM model, was higher in the high load period than in the low load period. It could also be seen from **Figure 10** that the curve fitting effect in the period of the high load was better than that in the period of low load. This indicates that the load in the low load stage is more affected by random factors and is more difficult to predict.

There are two limitations in the current study. First, due to the limited computational force, the search method adopted in the hyperparameter optimization in this paper is based on the conditional parameters, rather than the search based on the assignment interval. Although the conditional parameters based on the empirical method can ensure the accuracy of the prediction, it is undeniable that there is room for further optimization of the super parameters of the models. Secondly, the Bahdanau algorithm adopted is the classical gradient-based method to obtain the

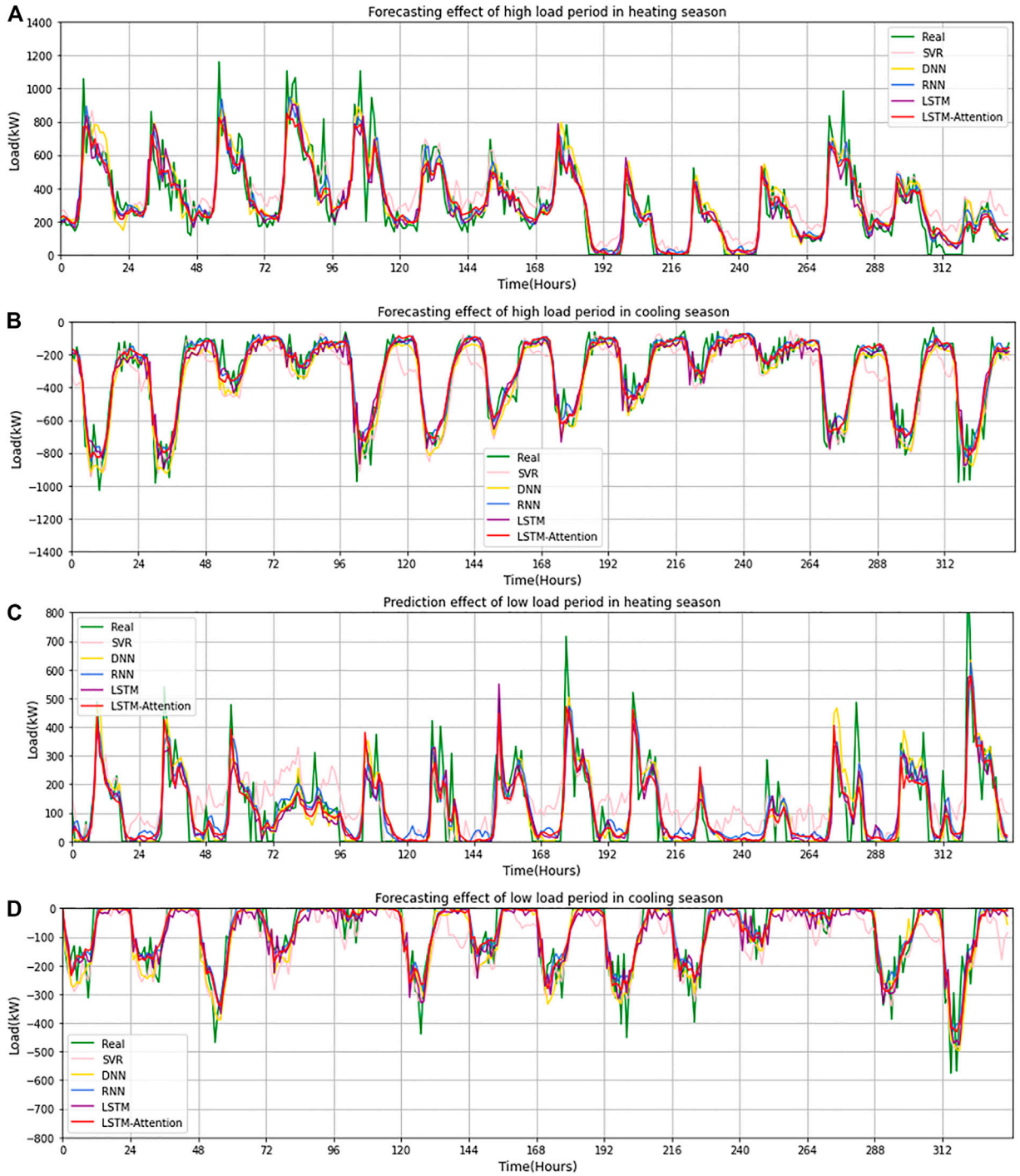



**FIGURE 10 |** Predicted load use *versus* measured load use by different models for 2 weeks as a test period: **(A)** Forecasting effect of high load period in the heating season, **(B)** Forecasting effect of high load period in the cooling season, **(C)** Forecasting effect of low load period in the heating season, **(D)** Forecasting effect of low load period in the cooling season.

optimal solution. The gradient-based method has the advantage of easy implementation, but at the same time, it will bring premature convergence and the problem of falling into a locally optimal solution. Therefore, there is room for further optimization at the algorithm level of this study.

## CONCLUSION

Predictive control had attracted more and more attention in building energy efficiency. Previous studies had shown that the HVAC system of large buildings was complex in structure, and its

operation was affected by random environmental factors and human activities, so it was very challenging to predict its short-term HVAC load. In this paper, we first analyzed the underlying patterns in the data based on the actual operation data of KSRP Energy Center in 9 years and then determined the factors used to establish the model according to the results of the Pearce correlation calculation. The results showed that the cooling and heating load of HVAC was most affected by the outdoor temperature, and the time of daily peak load was concentrated in a specific period.

Therefore, this paper proposed a new model combining the attention mechanism with the LSTM neural network, which was implemented by the following steps: First, according to the autocorrelation analysis results of HVAC load, we determined the data of the previous 24 h as the time step to predict the load of the next hour. In the second step, we used the TPE optimization method to optimize the hyperparameters of the baseline LSTM model. The test results showed that the LSTM model with two layers of 64 neurons had the best prediction effect. Thirdly, we added the attention layer to the baseline LSTM model to build the A-LSTM model. Finally, we also set up RNN, DNN, and SVR models as horizontal comparison objects.

Finally, we took the data from KSRP Energy Center from 2002 to 2009 as the training set and the data from 2010 as the test set to test the above five models respectively. The results showed that the prediction accuracy of the A-LSTM model was the best. Compared with the LSTM model, the overall RMSE decreased by 3.06%, MSE decreased by 6.54%, and $R^2$ value increased by 0.43%. By progressively reducing the size of the training set, we found that the performance advantage of the A-LSTM model was most significant when the length of training set was between 4 and 6 years. Besides, when the size of the training set dropped to 2 years, the prediction accuracy of the A-LSTM model declined sharply, which indicates that it has limitations in predicting small sample data. To verify the impact of low-load and zero-load samples on the experimental results, we respectively selected four typical operating mode samples in 2010 for evaluation and drew the effect chart of the predicted results. The results showed that the prediction effect of the A-LSTM model for refrigeration load was better than that for heating load, and the prediction effect for the high load period is better than that for the low load period.

In conclusion, for the cooling and heating load prediction of large buildings, the introduction of the attention mechanism can not only effectively improve the prediction accuracy of the traditional LSTM model, but also improve the accuracy of peak prediction. However, in practical application, the prediction effect of this model for different operating modes is different, which in-depth influence mechanism and solutions need to be further analyzed. Besides, there is still room for optimization in the algorithm of attention mechanism. In future work, we will try to apply the A-LSTM model to real-time HVAC energy-saving control. Since the traditional MPC system is a model-based control system, it needs to model the controlled objects accurately, which may affect the generality of the model. Therefore, we are more inclined to adopt a model-free deep reinforcement learning (RL) algorithm to solve this problem, such as taking the predicted value as the observed state of the agent to improve the control accuracy of the RL model.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

WG, FQ, and YL contributed to the conception and design of the study. YX organized the database. YX performed the statistical analysis. YX wrote the first draft of the manuscript. WG, FQ, and YL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## REFERENCES


Askari, S., Montazerin, N., and Zarandi, M. H. F. (2015). A Clustering Based Forecasting Algorithm for Multivariable Fuzzy Time Series Using Linear Combinations of Independent Variables[J]. *Appl. Soft Comput.* 35, 151–160.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate[J]. *arXiv*.

Bui, D-K., Nguyen, T. N., Ngo, T. D., and Nguyen-Xuan, H. (2020). An Artificial Neural Network (ANN) Expert System Enhanced with the Electromagnetism-Based Firefly Algorithm (EFA) for Predicting the Energy Consumption in Buildings[J]. *Energy* 190, 116370.

Chang, F., Chen, T., Su, W., and Alsafasfeh, Q. (2020). Control of Battery Charging Based on Reinforcement Learning and Long Short-Term Memory Networks[J]. *Comput. Electr. Eng.* 85, 106670.

Deb, C., Eang, L. S., Yang, J., and Santamouris, M. (2016). Forecasting Diurnal Cooling Energy Load for Institutional Buildings Using Artificial Neural Networks[J]. *Energy and Buildings* 121, 284–297.

Guo, Z., Zhou, K., Zhang, X., and Yang, S. (2018). A Deep Learning Model for Short-Term Power Load and Probability Density Forecasting[J]. *Energy* 160, 1186–1200.

Hazyuk, I., Ghiaus, C., and Penhouet, D. (2012). Optimal Temperature Control of Intermittently Heated Buildings Using Model Predictive Control: Part II – Control Algorithm[J]. *Building Environ.* 51, 388–394.

Heidari, A., and Khovalyg, D. (2020). Short-term Energy Use Prediction of Solar-Assisted Water Heating System: Application Case of Combined Attention-Based LSTM and Time-Series Decomposition[J]. *Solar Energy* 207, 626–639.

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory[J]. *Neural Comput.* 9 (8), 1735–1780.

Huang, G., and Chow, T-T. (2011). Uncertainty Shift in Robust Predictive Control Design for Application in CAV Air-Conditioning Systems[J]. *Building Serv. Eng. Res. Tech.* 32 (4), 329–343.

Iqbal, W., Tang, Y. M., Chau, K. Y., Irfan, M., and Mohsin, M. (2021). Nexus between Air Pollution and NCOV-2019 in China: Application of Negative Binomial Regression Analysis[J]. *Process Saf. Environ. Prot.* 150, 557–565.

Jradi, M., Veje, C., and Jørgensen, B. N. (2017). Deep Energy Renovation of the Mærsk Office Building in Denmark Using a Holistic Design Approach[J]. *Energy and Buildings* 151, 306–319.

Kim, B., Yamaguchi, Y., Kimura, S., Ko, Y., Ikeda, K., and Shimoda, Y. (2019). Urban Building Energy Modeling Considering the Heterogeneity of HVAC System Stock: A Case Study on Japanese Office Building Stock[J]. *Energy and Buildings* 199, 547–561.

Li, J., Wang, F., and He, Y. (2020). Electric Vehicle Routing Problem with Battery Swapping Considering Energy Consumption and Carbon Emissions[J]. *Sustainability* 12 (24).

Li, J., Yang, B., Li, H., Wang, Y., Qi, C., and Liu, Y. (2021A), DTDR–ALSTM: Extracting Dynamic Time-Delays to Reconstruct Multivariate Data for

Improving Attention-Based LSTM Industrial Time Series Prediction Models[J], *Knowledge-Based Systems*, 211, 106508. doi:10.1016/j.knosys.2020.106508

Li, R., Xu, M., Chen, Z., Gao, B., Cai, J., Shen, F., et al. (2021B). Phenology-based Classification of Crop Species and Rotation Types Using Fused MODIS and Landsat Data: The Comparison of a random-forest-based Model and a Decision-Rule-Based Model[J]. *Soil Tillage Res.* 206, 104838.

Li, W., Chien, F., Hsu, C-C., Zhang, Y., Nawaz, M. A., Iqbal, S., et al. (2021C). Nexus between Energy Poverty and Energy Efficiency: Estimating the Long-Run Dynamics[J]. *Resour. Pol.* 72, 102063.

Li, Y., Zhu, Z., Kong, D., Han, H., and Zhao, Y. (2019). EA-LSTM: Evolutionary Attention-Based LSTM for Time Series Prediction[J]. *Knowledge-Based Syst.* 181 (Oct.1), 1047851–1047858.

Liang, Y., Ke, S., Zhang, J., Yi, X., and Zheng, Y. (2018), GeoMAN: Multi-Level Attention Networks for Geo-Sensory Time Series Prediction[C], Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18.Stockholm, Sweden, July 13-19, 2018.

Liu, M-D., Ding, L., and Bai, Y-L. (2021). Application of Hybrid Model Based on Empirical Mode Decomposition, Novel Recurrent Neural Networks and the ARIMA to Wind Speed Prediction[J]. *Energ. Convers. Manag.* 233, 113917.

Lu, J., Xiong, C., Parikh, D., and Socher, R. (2017). Knowing when to Look: Adaptive Attention *via* A Visual Sentinel for Image Captioning[C], Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA, 21-26 July 2017, doi:10.1109/CVPR.2017.345

Luo, H., Cai, J., Zhang, K., Xie, R., and Zheng, L. (2021). A Multi-Task Deep Learning Model for Short-Term Taxi Demand Forecasting Considering Spatiotemporal Dependences[J]. *J. Traffic Transportation Eng. (English Edition)* 8 (1), 83–94.

Lv, Z., Qiao, L., Li, J., and Song, H. (2021). Deep-Learning-Enabled Security Issues in the Internet of Things[J]. *IEEE Internet Things J.* 8 (12), 9531–9538.

Lv, Z., Qiao, L., Singh, A. K., and Wang, Q. (2021). Fine-grained Visual Computing Based on Deep Learning[J]. *ACM Transactions on Multimidia Computing Communications and Applications*, 17. doi:10.1145/3418215

Lv, Z., Singh, A., and Li, J. (2021). Deep Learning for Security Problems in 5G Heterogeneous Networks[J]. *IEEE Netw.* 35, 67–73.

Ma, W., Fang, S., Liu, G., and Zhou, R. (2017). Modeling of District Load Forecasting for Distributed Energy System[J]. *Appl. Energ.* 204, 181–205.

Ma, Z., Ye, C., Li, H., and Ma, W. (2018). Applying Support Vector Machines to Predict Building Energy Consumption in China[J]. *Clean. Energ. Clean. Cities* 152, 780–786.

Massana, J., Pous, C., Burgas, L., Melendez, J., and Colomer, J. (2015). Short-term Load Forecasting in a Non-residential Building Contrasting Models and Attributes[J]. *Energy and Buildings* 92, 322–330.

Mayne, D. Q. (2014). Model Predictive Control: Recent Developments and Future Promise[J]. *Automatica* 50 (12), 2967–2986.

Mohsin, M., Hanif, I., Taghizadeh-Hesary, F., Abbas, Q., and Iqbal, W. (2021). Nexus between Energy Efficiency and Electricity Reforms: A DEA-Based Way Forward for Clean Power Development[J]. *Energy Policy* 149, 112052.

Mohsin, M., Taghizadeh-Hesary, F., Panthamit, N., Anwar, S., Abbas, Q., and Vo, X. V. (2020). Developing Low Carbon Finance Index: Evidence from Developed and Developing Economies[J]. *Finance Res. Lett.*, 101520.

Nguyen, H-P., Liu, J., and Zio, E. (2020). A Long-Term Prediction Approach Based on Long Short-Term Memory Neural Networks with Automatic Parameter Optimization by Tree-Structured Parzen Estimator and Applied to Time-Series Data of NPP Steam Generators[J]. *Appl. Soft Comput.* 89, 106116.

Qian, F., Gao, W., Yang, Y., and Yu, D. (2020). Potential Analysis of the Transfer Learning Model in Short and Medium-Term Forecasting of Building HVAC Energy Consumption[J]. *Energy* 193, 116724.

Sendra-Arranz, R., and Gutiérrez, A. (2020). A Long Short-Term Memory Artificial Neural Network to Predict Daily HVAC Consumption in Buildings[J]. *Energy and Buildings* 216, 109952.

Su, C., Huang, H., Shi, S., Jian, P., and Shi, X. (2020). Neural Machine Translation with Gumbel Tree-LSTM Based Encoder[J]. *J. Vis. Commun. Image Representation* 71, 102811.

Sultana, W. R., Sahoo, S. K., Sukchai, S., Yamuna, S., and Venkatesh, D. (2017). A Review on State of Art Development of Model Predictive Control for Renewable Energy Applications[J]. *Renew. Sust. Energ. Rev.* 76, 391–406.

Sun, H., Awan, R. U., Nawaz, M. A., Mohsin, M., Rasheed, A. K., and Iqbal, N. (2021). Assessing the Socio-Economic Viability of Solar Commercialization and Electrification in South Asian Countries, *Environment, Dev. Sustainability*, 23, 9875–9897.

Verwimp, L., Van Hamme, H., and Wambacq, P. (2020). State Gradients for Analyzing Memory in LSTM Language Models[J]. *Comp. Speech Lang.* 61, 101034.

Wang, J., Huang, G., Sun, Y., and Liu, X. (2016). Event-driven Optimization of Complex HVAC Systems[J]. *Energy and Buildings* 133, 79–87.

Wang, W., Hong, T., Xu, X., Chen, J., Liu, Z., and Xu, N. (2019). Forecasting District-Scale Energy Dynamics through Integrating Building Network and Long Short-Term Memory Learning Algorithm[J]. *Appl. Energ.* 248, 217–230.

Wang, Z., Hong, T., and Piette, M. A. (2020). Building thermal Load Prediction through Shallow Machine Learning and Deep Learning[J]. *Appl. Energ.* 263, 114683.

Wang, Z., Hong, T., and Piette, M. A. (2019). Predicting Plug Loads with Occupant Count Data through a Deep Learning Approach[J]. *Energy* 181, 29–42.

Wang, Z., Liu, J., Zhang, Y., Yuan, H., Zhang, R., and Srinivasan, R. S. (2021). Practical Issues in Implementing Machine-Learning Models for Building Energy Efficiency: Moving beyond Obstacles[J]. *Renew. Sust. Energ. Rev.* 143, 110929.

Wei, Y., Xia, L., Pan, S., Wu, J., Zhang, X., Han, M., et al. (2019). Prediction of Occupancy Level and Energy Consumption in Office Building Using Blind System Identification and Neural Networks[J]. *Appl. Energ.* 240, 276–294.

Yang, F., Wang, D., Xu, F., Huang, Z., and Tsui, K-L. (2020). Lifespan Prediction of Lithium-Ion Batteries Based on Various Extracted Features and Gradient Boosting Regression Tree Model[J]. *J. Power Sourc.* 476, 228654.

Yang, J. J., Yang, M., Wang, M. X., Du, P. J., and Yu, Y. X. (2020). A Deep Reinforcement Learning Method for Managing Wind Farm Uncertainties through Energy Storage System Control and External reserve Purchasing[J]. *Int. J. Electr. Power Energ. Syst.* 119, 105928.

Yang, T., Li, B., and Xun, Q. (2019), LSTM-Attention-Embedding Model-Based Day-Ahead Prediction of Photovoltaic Power Output Us-Ing Bayesian Optimization[J]. *IEEE Access*, 7, 171471-171484, 1-1.PP(99, doi:10.1109/ACCESS.2019.2954290

Yu, B. (2021). Urban Spatial Structure and Total-Factor Energy Efficiency in Chinese Provinces[J]. *Ecol. Indicators* 126, 107662.

Yu, Z., Yu, J., Fan, J., and Tao, D. (2017), Multi-modal Factorized Bilinear Pooling with Co-attention Learning for Visual Question Answering[C], 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22-29 Oct. 2017, doi:10.1109/ICCV.2017.202

Zhan, S., and Chong, A. (2021). Data Requirements and Performance Evaluation of Model Predictive Control in Buildings: A Modeling Perspective[J]. *Renew. Sust. Energ. Rev.* 142, 110835.

Zhao, X., Gao, W., Qian, F., and Ge, J. (2021). Electricity Cost Comparison of Dynamic Pricing Model Based on Load Forecasting in home Energy Management System[J]. *Energy* 229, 120538.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Xu, Gao, Qian and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 23 August 2021
doi: 10.3389/fenvs.2021.721517

# Driving Factors of $CO_2$ Emissions: Further Study Based on Machine Learning

*Shanshan Li*[1], *Yam Wing Siu*[2]* *and Guoqin Zhao*[1]

[1]Institute for Finance and Economics, Central University of Finance and Economics, Beijing, China, [2]Department of Economics and Finance, The Hang Seng University of Hong Kong, Hong Kong, China

Greenhouse gases, especially carbon dioxide ($CO_2$) emissions, are viewed as one of the core causes of climate change, and it has become one of the most important environmental problems in the world. This paper attempts to investigate the relation between $CO_2$ emissions and economic growth, industry structure, urbanization, research and development (R&D) investment, actual use of foreign capital, and growth rate of energy consumption in China between 2000 and 2018. This study is important for China as it has pledged to peak its carbon dioxide emissions ($CO_2$) by 2030 and achieve carbon neutrality by 2060. We apply a suite of machine learning algorithms on the training set of data, 2000–2015, and predict the levels of $CO_2$ emissions for the testing set, 2016–2018. Employing rmse for model selection, results show that the nonlinear model of k-nearest neighbors (KNN) model performs the best among linear models, nonlinear models, ensemble models, and artificial neural networks for the present dataset. Using KNN model, sensitivity analysis of $CO_2$ emissions around its centroid position was conducted. The findings indicate that not all provinces should develop its industrialization. Some provinces should stay at relatively mild industrialization stage while selected others should develop theirs as quickly as possible. It is because $CO_2$ emissions will eventually decrease after saturation point. In terms of urbanization, there is an optimal range for a province. At the optimal range, the $CO_2$ emissions would be at a minimum, and it is likely a result of technological innovation in energy usage and efficiency. Moreover, China should increase its R&D investment intensity from the present level as it will decrease $CO_2$ emissions. If R&D reinvestment is associated with actual use of foreign capital, policy makers should prioritize the use of foreign capital for R&D investment on green technology. Last, economic growth requires consuming energy. However, policy makers must refrain from consuming energy beyond a certain optimal growth rate. The above findings provide a guide to policy makers to achieve dual-carbon strategy while sustaining economic development.

**Keywords: Machine learning, $CO_2$ emissions, economic growth, industry structure, forecasting**

OPEN ACCESS

***Edited by:***
*Xunpeng (Roc) Shi,*
*University of Technology Sydney, Australia*

***Reviwed by:***
*Muntasir Murshed,*
*North South University, Bangladesh*
*Jianfu Shen,*
*Hong Kong Polytechnic University, Hong Kong, China*
*Feng Dong,*
*China University of Mining and Technology, China*

***Correspondence:**
*Yam Wing Siu*
*ywsiu@hsu.edu.hk*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *07 June 2021*
***Accepted:*** *22 July 2021*
***Published:*** *23 August 2021*

***Citation:***
*Li S, Siu YW and Zhao G (2021) Driving Factors of $CO_2$ Emissions: Further Study Based on Machine Learning. Front. Environ. Sci. 9:721517. doi: 10.3389/fenvs.2021.721517*

## INTRODUCTION

Greenhouse gases, especially carbon dioxide ($CO_2$) emissions, are viewed as one of the core causes of climate change, and it has become one of the most important environmental problems in the world (Rehman et al. (2021a)). At the press conference on WMO State of the Climate 2019 Report, António Guterres, UN Chief, reported that 2019 was the second hottest year on record during his opening

remarks. According to the World Meteorological Organization's (WMO) flagship State of the Global Climate report, the global average temperature in 2020 was about 1.2°C above preindustrial level.

To mitigate the threat of runaway climate change, the Paris Agreement calls for limiting global warming to well below 2 and preferably to 1.5°C, compared to preindustrial levels. This requires global emissions to peak as soon as possible, with a rapid fall of 45 percent from 2010 levels by 2030, and to continue to drop off steeply to achieve net zero emissions by 2050 (Bertram et al., 2021). The world is way off track in meeting this target at the current level of nationally determined contributions. Global greenhouse gas emissions of developed countries and economies in transition have declined by 6.5 percent over the period 2000–2018. Meanwhile, the emissions of developing countries are up by 43.2 percent from 2000 to 2013. The rise is largely attributable to increased industrialization and enhanced economic output measured in terms of GDP.

Carbon dioxide emissions have been the primary source of extreme environmental pollution (Rehman et al. (2021a)). With the rapidly growing agriculture and farm mechanization, agricultural sector has become a factor in the surge in $CO_2$ emissions and other greenhouse gases in the globe (Rehman et al. (2021b)).

Economic, social, and environmental suitability are the three core pillars of the UN's Sustainable Development Goals (SDG) declarations (Rehman et al. (2021a)). In September 2019, Heads of State and Government gathered in the SDG Summit at the United Nations Headquarters in New York to follow up and comprehensively review progress in the implementation of the 2030 Agenda for Sustainable Development and the 17 Sustainable Development Goals (SDGs). The summit resulted in the adoption of the Political Declaration and its core message is to take action to respond to climate emergencies. Relevant research shows that if economic growth and climate and environmental sustainability are achieved at the same time, emission reduction policies need to be incorporated into the economic growth policies of various countries (Murshed et al. (2020), Li et al. (2021), Rehman et al. (2021a)).

As the world's second largest economy, the Chinese government strives to achieve environmental sustainability through a series of policies and measures. At the General Debate of the 75th session of the United Nations General Assembly on 22nd September 2020, President Xi Jinping of China announced that China will scale up its Intended Nationally Determined Contributions by adopting more vigorous policies and measures. It also aims to have $CO_2$ emissions peak before 2030 and achieve carbon neutrality before 2060.

Generally speaking, various economic activities will affect carbon emissions, Liu et al. (2021). They include industrial structure (Shen et al., 2021), energy consumption, trade, and urbanization (Kasman and Duman, 2015), consumption structure of fossil fuel and cleaner fuel (Murshed et al., 2020), foreign investment (Elliott and Sun, 2013), and technology advancement (Yu and Du, 2018).

Based on this background, this paper studies the relation between China's economic growth, industrial structure, urbanization, R&D investment, foreign investment, energy consumption growth, and $CO_2$ emissions from 2000 to 2018 and predicts it.

Choosing China as an ideal case to study driving factors on $CO_2$ emissions is because China has accounted for the highest level of $CO_2$ emissions across the globe in 2017 (Ma et al., 2021). President Xi Jinping addressed the General Assembly of United Nations and declared China's national goal of turning carbon neutral by 2060. China is an important country to play a key role in achieving the 2030 Sustainable Development Agenda of the United Nations. In order to achieve the 2030 Sustainable Development Agenda of the United Nations and the Paris Agreement at the same time, China must achieve the carbon emissions peak by 2030 and the carbon neutrality by 2060 while sustaining a certain economic growth. To this end, China has formulated a "dual-carbon" strategy. Therefore, it is vital to study the drivers that influence $CO_2$ emissions. Economic growth and $CO_2$ emissions go hand in hand as economic activities give rise to $CO_2$ emissions. Therefore, economic growth is the core factor affecting $CO_2$ emissions. Industrialization and urbanization are the two main lines of China's economic and social development that includes the $CO_2$ emissions of the production side and the consumption side, respectively (Cao et al, 2016; Han et al., 2019). Industrialization and urbanization are compound factors affecting carbon emissions. It is because the process of industrialization and urbanization includes the factors driving $CO_2$ emissions and limiting $CO_2$ emissions. Industrialization has brought the change of industrial structure, and the $CO_2$ emissions of different industries are different. On the one hand, urbanization has an impact on the $CO_2$ emissions caused by residents' consumption, which is quite different between urban residents and rural residents. On the other hand, urbanization is the movement of industries and population in different areas. Therefore, urbanization also reflects the different performance of carbon emissions in urban area and rural area. Technological progress, foreign investment, and energy consumption are the specific factors of $CO_2$ emissions, technological progress reduce $CO_2$ emissions by exploring and usage of clean energy, foreign investment reflects the pollution haven (tax environmental regulation, good market access to high-income countries, and corruption opportunities) (Candau and Dienesch, 2017), and energy consumption determines the quantity of $CO_2$ emissions.

This paper contributes to the literature in two ways. 1) This is a comprehensive research; we try to build a framework which includes three levels of six driving factors on $CO_2$ emissions as shown in **Figure 1**. The most important factors include economic growth, industrialization, urbanization, technology progress, foreign direct investment, and energy consumption. 2) Most of the existing studies are based on OLS framework to explore the relation between carbon emissions and related factors. It is difficult to avoid the omission of variables or endogeneity issues, Kasman and Duman (2015). An increasing number of recent studies (Li et al., 2021; Liu et al., 2021) have been using cross-sectionally augmented autoregressive distributed lag (CS-ARDL) approach developed by Chudik and Pesaran (2015) for

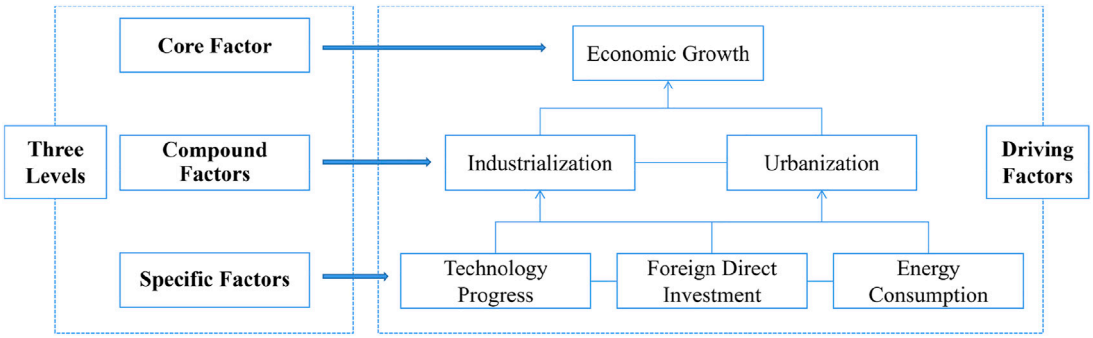



**FIGURE 1 |** The framework of driving factors on $CO_2$ emissions.

short- and long-term $CO_2$ emissions forecast. This research applies a suite of machine learning algorithms in predicting $CO_2$ emissions using the factors discussed. Machine learning avoids omission of variables and endogeneity issues. In addition, the trends and relation between $CO_2$ emissions and various factors are predicted.

The rest of this paper is organized as follows. *Literature Review* provides a literature review on $CO_2$ emissions. *Data and the Variables* describes the data and variables under study. *Methodology* describes the machine learning algorithms deployed for predicting the level of $CO_2$ emissions. *Results* compares the accuracy of predictions among various machine learning algorithms. *Discussions* discusses the results using the best performing model, while *Conclusion and Policy Implications* concludes the paper.

## LITERATURE REVIEW

Economic scale, economic structure, and technological level are the three major factors affecting the environment (Grossman and Krueger, 1995). Economic scale is the output of the economy; more economic output means more pollution. It is because that economic growth needs more resources investment and more energy consumption. Economic structure is industry structure. The change of industry structure will reduce the pollution. With economic developing, percentage of secondary industry, especially energy-intensive industry, will reduce percentage of tertiary industry, and energy consumption will increase, so the pollution will be reducing. Technology progress will realize the usage of resource efficiency and reduce the energy consumption. So, technological level is an important factor which influences the energy intensity and pollution. Many research studies are based on these three environment factors and extend them accordingly. The research can be classified into relation between economic growth and $CO_2$ emissions, industry structure, technology, and $CO_2$ emissions, and urbanization and $CO_2$ emissions. But results differ from research focus, theories, and methods. There are three parallel literatures on factors what will influence $CO_2$ emissions.

The first group of studies has investigated the relation between $CO_2$ emission, economic growth, and energy consumption. Environmental Kuznets Curve (EKC) is often used to discuss the relation between environmental pollution and economic growth, which is also the main method to analyze the relation between $CO_2$ emissions and economic growth (Lin and Jiang, 2009). Grossman and Krueger (1991) found the U-shaped relation between economic growth and $CO_2$ emissions. But the result is opposite if $CO_2$ is used as the environmental indicator. Holtz-Eakin and Selden (1995), Sachs et al. (1999), Friedl and Getzner (2003), and Galeotti et al. (2006) found that the relation between $CO_2$ emission and economic growth is inverted U-shape. It is opposite in the study of Shafik (1994), Martin (2008) and Murshed and Dao (2020) which find that per capita $CO_2$ emission increased in parallel with per capita income, and there is no turning point. Moomaw and Unruh (1997), Martinez-Zarzoso and Bengochea-Morancho (2004), Friedl and Getzner (2003), and Akpan and Chuku (2011) found that the relation between $CO_2$ emission and economic growth is N-shape. Saidi and Hammami (2015) examined the effect of energy use and the $CO_2$ emissions on economic growth for 58 countries, and their empirical results showed that $CO_2$ emissions negatively affected economic growth. Rahman et al. (2020), Liu et al. (2012), and Lantz and Feng (2006) found that per capita GDP has no relation with $CO_2$ emission.

Environmental Kuznets Curve describes the economic growth in developed countries and the inverted U-shaped relation between environmental pollution, consciously or unconsciously, as for the developed countries to adjust economic structure and the energy consumption structure and achieve a faster pace of the inverted U-shaped path, the overall environmental quality as economic growth accumulation showed a trend of deterioration before improvement (Lin and Jiang, 2009). Acheampong (2018) found that energy consumption has a negative impact on economic growth in global level, economic growth has a negative impact on $CO_2$ emission, and $CO_2$ emission has positive impact on economic growth. In the Asia-Pacific region, economic growth does not cause $CO_2$ emissions. But in Caribbean-Latin America, there is a feedback causality between economic growth and carbon emissions.

The second group of studies has investigated the relation between $CO_2$ emissions, industry structure, and technology progress. Bernardini and Galli (1993) found that the decline in energy intensity shows a decline trend with the increase in

income. The three reasons behind the relationship descent are the following. First of all, with the development of the economy, the final demand structure changes with changes in the stage of industrialization. In the preindustrial stage, agriculture is the leading industry in economic development, and economic growth is driven by basic needs, which can be met with low energy intensity. In the stage of industrialization, the infrastructure network needs to be built up to facilitate large-scale production and consumption. The primitive accumulation of capital stock related to industrialization can increase energy intensity, but it eventually reached the saturation point. At this time, the consumption of materials tended to replace durables rather than create durables. In the postindustrialized stage, the decline of manufacturing industry in relation between services and energy intensity in service-based economies is smaller than that in manufacturing-oriented economies. Shahbaz et al. (2018) and Khan et al. (2019) found that financial development helps control $CO_2$ emissions in both France and China. However, Liu et al. (2021) found that with 1% financial development, $CO_2$ emissions increased 0.17–0.52%.

Technology progress is the dominant factor of long-run economic growth with scarce resources. Technology change has a positive influence on energy efficiency and negative influence on energy intensity (Lin and Du, 2014; Sadorsky, 2013; Yu et al., 2021). Ang (2009) used the framework to combine modern growth theoretically, which can analyze the role of R&D activity and technology progress in reducing pollution. Technology progress is the result of R&D investment, which contributes to energy intensity reduction (Young, 1998). Wei et al. (2010) extended Antweiler's model (Antweiler et al., 2001) to analyze the influence factors of $CO_2$ emissions. The study found that GDP, industrialization, and free trade have positive influence on $CO_2$ emissions, but independent research and development and technology import contribute to reducing $CO_2$ emissions.

One source of technology progress is independent innovation; another source is FDI and trade. FDI and trade are latecomer advantage of countries, which develops later. Elliott and Sun (2013) found that FDI has negative influence on energy intensity. The last study (Khan et al., 2021) investigates the roles of export diversification and composite country risks in carbon emissions abatement. The researchers found that lowering country risks, undergoing renewable energy transition, and enhancing environmental-related technological innovations assist in reducing $CO_2$ emissions in the long run.

The third group of studies has investigated the relation between $CO_2$ emissions and urbanization. At present, there are a large number of literatures on urbanization and its impact on carbon dioxide ($CO_2$) emissions for reference. A lot of research have directly investigated the positive impact of urbanization on carbon dioxide emissions (Behera and Dash, 2017; York et al., 2003; Zhang and Lin, 2012). Shahbaz et al. (2017) provided evidence showing that the development of urbanization leads to higher demands for food, housing, transportation, land usage, and energy consumption and causes serious environmental degradation problems. For instance, traffic congestion, waste management, and poor sanitation could cause pollution and health problems in most urban areas.

A number of studies have tested the linear impact of urbanization on global carbon dioxide emissions. You can find contributions that support it, such as those by York et al. (2003), Cole and Neumayer (2004), Liddle and Lung (2010), Wang et al. (2012), and Behera and Dash (2017), or those that refute it, such as those by Hossain (2011) and Liu and Bae (2018). Specifically, York et al. (2003) used panel data from 143 countries to record the positive impact of urbanization on $CO_2$. Cole and Neumayer (2004) and Liddle and Lung (2010) reached similar findings using panel data and Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT). Wang et al. (2012) applied PLS with STIRPAT model in Beijing, China, and concluded that urbanization is the most influential factor that has adverse impact on environmental quality. Subsequently, Wang et al. (2013) found that urbanization, industrial growth, income levels, and population stimulate $CO_2$ emissions during a provincial study. Behera and Dash (2017) used panel cointegration test to study the positive impact of urbanization on carbon emissions in South and Southeast Asian countries. Conversely, several studies proposed uncertain results and reported the negligible impact of urbanization on $CO_2$ emissions (Hossain, 2011; Liu and Bae, 2018).

To a large extent, the level of economic development of the country may alleviate the nature of the relation between urbanization and pollution (Fan et al., 2006; Li and Lin, 2015; Poumanyvong and Kaneko, 2010). However, higher urbanization growth rates and development rates can improve the environment by promoting technological innovation in energy usage and efficiency, increasing awareness of environmental issues, and using green technologies, Bekhet and Othman (2017). Urbanization has an inverted U-shaped relation with $CO_2$ emissions in Asia (Fan et al., 2020). However, Zhu et al. (2012) found there is limited support for inverted U-shaped relation between $CO_2$ emissions and urbanization in 20 emerging economies. There was a long-run bidirectional positive relation between $CO_2$ emissions, urbanization, and energy consumption in MENA countries. However, the long-run relation is based on the countries' income and development (Al-mulalia et al., 2013). Urbanization has a positive influence on $CO_2$ emissions; in the stage of urbanizing, it needs more energy consumption which will increase $CO_2$ emissions (Lin and Du, 2013). $CO_2$ emissions are higher in big cities or urban agglomeration areas, because of the high energy consumption on residential electricity consumption, residential gas consumption, residential heating consumption, and residential transportation energy consumption (Bai et al., 2019). In east and central China, the center and surroundings featured high levels (high-high cluster) of total $CO_2$ emissions and low levels (low-low-cluster) of per-unit-GDP $CO_2$ emission in urban agglomerations. The Yangtze-River-Delta, the Beibu-Gulf, and the Guangdong-Hong Kong-Macao UAs were more efficient at emission reduction with the cities' rising scales, while cities of the Beijing-Tianjin-Hebei UA and the Chengdu-Chongqing UA performed less efficiently (Cui et al., 2020).

Two main methodologies are used by three groups of studies. One of the methodologies is econometrics. Econometric methods

include spatial autocorrelation analysis, semiparametric fixed effect (Zhu et al., 2012), panel threshold regression (Zi et al., 2016), panel threshold regression (Du and Xia, 2018), autoregressive distributed lag model and vector-error correction model (Bekhet and Othman, 2017), two-stage least squares (2SLS) and augmented Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model (Bai et al., 2019), and autoregressive distributed lag (ARDL) (Ang, 2009). Econometric methods have been used to estimate the long-run relationship and the short-run dynamics for environmental pollution and its determinants. To address the issues of multicollinearity and overfitting, a recent study introduced the least absolute shrinkage and selection operator (LASSO) regression model which can pinpoint the most important determinants to investigate the driving factors influencing household carbon emissions (Shi et al., 2020). Another study on methods called cross-sectionally augmented autoregressive distributed lags (CS-ARDL) can account for cross-sectional dependency, slope heterogeneity, and structural break issues in the data (Li et al., 2021; Ma et al., 2021). The other methodology is calculating the quantity of $CO_2$ emissions. Many research studies are based on Kaya identity and Logarithmic Mean Divisa Index (LMDI) (Ang and Zhang, 2000). Using these methods, researchers calculate the industrial $CO_2$ emissions, regional $CO_2$ emissions, and national $CO_2$ emissions (Yang and Li, 2017). Based on LMDI, index decomposition analysis (IDA) is developed and becomes one of the most popular methods. However, IDA calculates the technology efficiency of economy system, not the efficiency of energy usage (Lin and Du, 2013). Wang (2011) developed the method based on production-theory decomposition approach (PDA), which is based on output-oriented distance function to decompose the energy production to technology efficiency, technology program, and input alternative. Lin and Du (2014) gave a complex framework (L-D framework) of index decomposition and production theory. Then, Yang et al. (2019) used L-D framework to calculate $CO_2$ emissions of major industries.

There are two gaps in the above literature. Factor choice is confused by economic methods which do not support all factors (Shi et al., 2020). So, the studies always try to select one or two important factors. Actually, factors framework is a hierarchical structure, and they inevitably influence each other. Methodologies reviewed above are very useful and have been adopted with many successes. However, there are many restrictions such as collinearity and causality issues of variables. On the other hand, it is not necessary to consider these issues in machine learning. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention. Machine learning aims to develop algorithms that can learn and create statistical models for data analysis and prediction. The ML algorithms should be able to learn by themselves, based on data provided, and make accurate predictions, without having been specifically programmed for a given task.

# DATA AND THE VARIABLES

## $CO_2$ Emissions

The International Panel on Climate Change (IPCC) had introduced three methods of calculating $CO_2$ emissions (Y) from fossil fuel combustion in both stationary and mobile sources. "Method 1" is based on the amount of fuel burned and the emission factor, and it is achievable (Wang et al., 2010). Thus, this method is adopted by this paper accordingly. The method is specified as follows:

$$CO_2 = \sum_{i=1}^{14} CO_{2,i} = \sum_{i=1}^{14} E_i \cdot NCV_i \cdot CEF_i \quad (1)$$

In **Eq. 1**, $CO_2$ represents the amount of carbon dioxide emissions to be estimated; $i$ represents various energy fuels, including coal, coke, coke oven gas, blast furnace gas, converter gas, other gas, crude oil, gasoline, kerosene, diesel, fuel oil, and liquefied petroleum, natural gas, and liquefied natural gas; $E_i$ represents the combustion consumption of various energy sources; $NCV_i$ is the average low calorific value of various energy sources, used to convert various energy consumption into energy units ($TJ$); $CEF_i$ represents carbon dioxide emission factor of the energy consumption, which is calculated by **Eq. 2**:

$$CEF_i = CC_i \cdot COF_i \cdot (44/12) \quad (2)$$

In **Eq. 2**, $CC_i$ is the carbon content of energy sources. $COF_i$ is the carbon oxidation factor of energy sources; usually, the value is 1, which means that the energy is completely oxidized. In this paper, coal and coke are set to 0.99 and the rest is 1 (Chen, 2011). (44/12) is the molecular weight ratio of carbon dioxide to carbon. The $CO_2$ emissions related data are derived from China Energy Statistical Yearbook (2001–2019) and Report of IPCC (2006).

## Industrial Structure Rationalization Index

Industrial structure rationalization (X1) reflects the coordination of different industries; moreover, it reflects the efficiency of energy usage (Gan et al., 2011). The Theil index measures the industrial structure rationalization (Gan et al., 2011). The Theil index is defined as the equation below:

$$\mathrm{TL} = \sum_{i=1}^{n} \left(\frac{\mathrm{Y_i}}{\mathrm{Y}}\right) \ln\left(\frac{\mathrm{Y_i}}{\mathrm{Y}} \Big/ \frac{\mathrm{Y}}{\mathrm{L}}\right) \quad (3)$$

TL is the Theil index, Y is GDP, L is employment, $i$ represents industries, and $n$ represents industry sectors. When economy is equilibrium, TL = 0, industrial structure is rational. The industrial structure rationalization index related data are derived from Chinese Statistical Yearbook (2001–2019).

## Other Variables and Data

This paper also includes other important variables. They are GDP, urbanization, research and development (R&D)

investment, actual use of foreign capital, and growth rate of energy consumption. Data for GDP (X2) and actual use of foreign capital (X5) are derived from Chinese Statistical Yearbook (2001–2019) and statistical yearbook of 30 provinces from 2000 to 2018. Data on urbanization (X3), the share of urban population, is derived from Statistical Yearbooks (2001–2019) for the 30 provinces. Data for R&D reinvestment intensity (X4) are from Statistical Communique on National Science and Technology Expenditures (2000–2018). Data for growth rate of energy consumption (X6) are from China Energy Statistical Yearbook (2001–2019). In summary, there are one output variable and six input variables and annual data of these seven variables of 30 provinces are obtained for 2000–2018 from the sources stated above.

## METHODOLOGY

This study uses a number of machine learning algorithms, or function, f, to map the output variable (Y) from input variables (X1, X2, . . . , X6) so that Y = f(X1, X2, . . . , X6). Several types of algorithms have been adopted in this study, and they are briefly described here.

### Linear Models–Linear Regression, Lasso, and ElasticNet

When we make assumptions to the learning process, we can simplify the process a lot. However, they can also limit what can be learned. Algorithms that simplify the function to a known form are called linear models. Examples of this class include linear regression and logistic regression. In this study, we tested three linear models, and they are linear regression (LR), least absolute shrinkage and selection operator (LASSO), and Elastic Net (EN) that adds regularization penalties to the loss function during training. Linear models provide the benchmark to measure other machine learning algorithms. However, it is expected that linear models would not provide good prediction because $CO_2$ emissions are complicated and depend on many factors. Furthermore, it is not expected that those factors relate to $CO_2$ emissions linearly.

### Nonlinear Models: Classification and Regression Tree, Support Vector Regression, and k-Nearest Neighbors Regression

When we do not make strong assumptions about the form of the mapping function, the algorithms are called nonlinear models. Examples of this class include Classification and Regression Tree (CART), Support Vector Regression (SVR), and k-Nearest Neighbors (KNN). These models are useful for problems involving datasets with large number of features, many of which may be correlated. As the name implied, CART works for both classification and regression problems. For SVR, as the name suggests, it is a regression algorithm, and it should not be confused with Support Vector Machine (SVM) which is for classification. The major difference between the two is there is only one slack variable in SVM and there are two slack variables in SVR during its optimization for locating the hyperplane. For KNN algorithm, it can be applied for both classification and regression problems. In classification, the algorithm tries to predict the class to which the output variable belongs by computing the local probability, while it tries to predict the values of the output variable by using a local average in regression. One of the strengths of machine learning is that it can work with nonlinear data. If a system is nonlinear (i.e., a system that contains $CO_2$ emissions and its six input variables), nonlinear models would be more appropriate.

### Ensemble Methods

Traditionally, machine learning application consisted of a single learner (say, a Decision Tree). Then, ensemble methods were born, which involve using many learners to enhance the performance of any single one of them individually.

#### Bagging Methods: Random Forest and Extra Trees

Bagging is a method of merging the same type of predictions. The idea of bagging is then simple: we want to fit several independent models and "average" their predictions in order to obtain a model with a lower variance. In bagging, weak learners are trained in parallel using randomness, and each model receives an equal

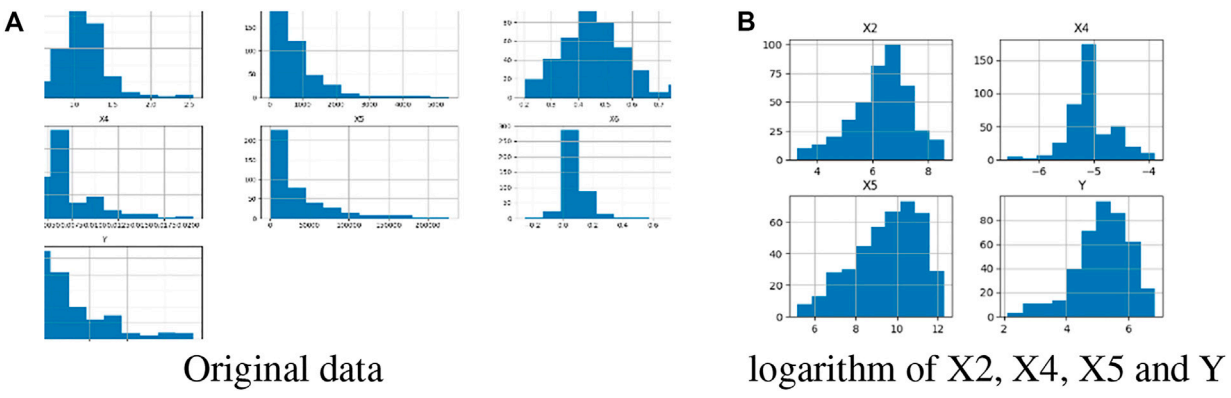



**FIGURE 2 |** Histogram of data between 2000 and 2015.

**TABLE 1 |** Descriptive statistics for original data between 2000 and 2015.

| | X1 | X2 | X3 | X4 | X5 | X6 | Y |
|---|---|---|---|---|---|---|---|
| Min | 0.2332 | 2.6368 | 0.2035 | 0.0014 | 167.44 | −0.2503 | 8.1842 |
| First quartile | 0.8301 | 30.0985 | 0.4012 | 0.0052 | 4,483.90 | 0.0280 | 100.8598 |
| Average | 1.0510 | 87.9108 | 0.4962 | 0.0070 | 34,974.14 | 0.0703 | 242.2765 |
| Third quartile | 1.2636 | 109.9589 | 0.5686 | 0.0081 | 47,477.80 | 0.1008 | 322.9348 |
| Max | 2.5528 | 536.8455 | 0.8650 | 0.0206 | 225,732.22 | 0.9356 | 948.0496 |
| Std. Dev. | 0.3345 | 88.6889 | 0.1387 | 0.0031 | 43,286.63 | 0.0798 | 193.4006 |
| Median | 1.0437 | 62.6338 | 0.4850 | 0.0061 | 17,956.95 | 0.0644 | 188.0715 |
| Skew | 0.6604 | 2.2506 | 0.5403 | 1.6093 | 1.8672 | 3.1862 | 1.4709 |
| Kurtosis | 2.3140 | 6.0578 | 0.2066 | 3.1208 | 3.3609 | 30.1929 | 2.1163 |

**TABLE 2 |** Descriptive statistics for Ln(X2), Ln(X4), Ln(X5), and Ln(Y) between 2000 and 2015.

| | Ln(X2) | Ln(X4) | Ln(X5) | Ln(Y) |
|---|---|---|---|---|
| Min | 3.2722 | −6.5673 | 5.1206 | 2.1022 |
| First quartile | 5.7070 | −5.2645 | 8.4082 | 4.6137 |
| Average | 6.3101 | −5.0470 | 9.5258 | 5.1545 |
| Third quartile | 7.0027 | −4.8216 | 10.7680 | 5.7775 |
| Max | 8.5883 | −3.8839 | 12.3271 | 6.8544 |
| Std. Dev. | 1.0522 | 0.4131 | 1.6125 | 0.8935 |
| Median | 6.4399 | −5.1012 | 9.7957 | 5.2368 |
| Skew | −0.5121 | −0.0282 | −0.4972 | −0.6515 |
| Kurtosis | 0.1566 | 1.5317 | −0.5375 | 0.5463 |

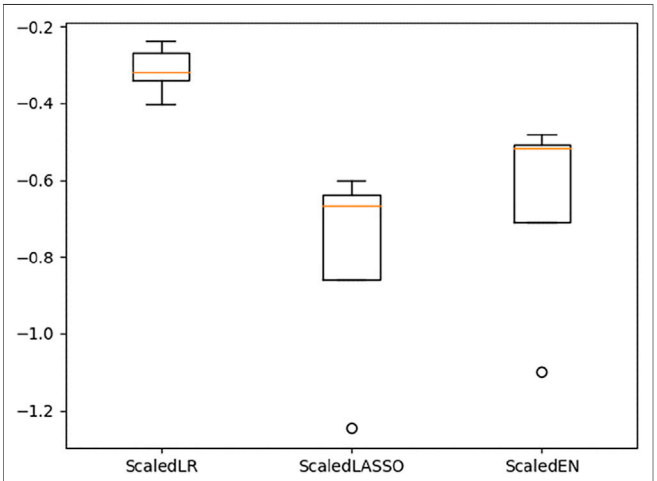



**FIGURE 3 |** Linear models comparison.

weight. Bagging decreases variance, not bias, and solves overfitting issues in a model.

### Boosting Methods: XGBoost, AdaBoost, and Gradient Boost

Boosting models fall inside this family of ensemble methods. Boosting is a method of merging different types of predictions. Boosting decreases bias, not variance. In boosting, models are weighted based on their performance. Boosting should not be confused with bagging. In boosting, the weak learners are trained sequentially.

AdaBoost is a specific boosting algorithm developed for classification problems. The weakness is identified by the weak estimator's error rate.

Gradient boosting approaches the problem a bit differently. Instead of adjusting weights of data points, Gradient boosting focuses on the difference between the prediction and the ground truth.

XGBoost builds the model by calculating similarity scores between the observations that end up in a node. Also, XGBoost allows for regularization, reducing the possible overfitting of individual trees and therefore of the ensemble model.

## Artificial Neural Networks

Neural networks consist of nodes connected by links. They have three types of layers: an input layer with a node for each input, hidden layers where learning occurs in training and inputs are processed on trained nets, and an output layer with a node for each target variable, which passes information outside the network. Learning takes place in the hidden layer nodes, each of which consists of a summation operator and an activation function. Note that for neural networks, the inputs should be scaled (i.e., standardized) to account for differences in the units of the data. This is important as scaling could improve the performance by a considerable margin (Chaudhari, 2019).

In recent times, ANNs have become popular and helpful model for classification, clustering, and pattern recognition in many disciplines (Abiodun et al., 2018). With its versatility, one would expect it will work well. However, neural networks usually require much more data than traditional machine learning algorithms. In fact, the amount of data required depends both on the complexity of the problem and on the complexity of chosen algorithm. Given that the present study has only 400 + rows of panel data, whether this would impose any limitation on the accuracy of this method remains to be seen.

# RESULTS

To get the best results, it is necessary to understand the data first by inspecting their descriptive statistics and plotting their histograms. The descriptive statistics and histogram of the

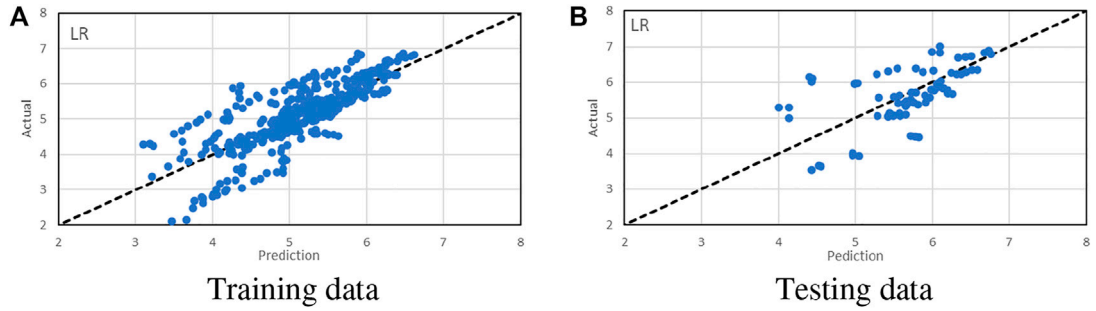



**FIGURE 4 |** Actual vs. prediction of Linear models comparison.

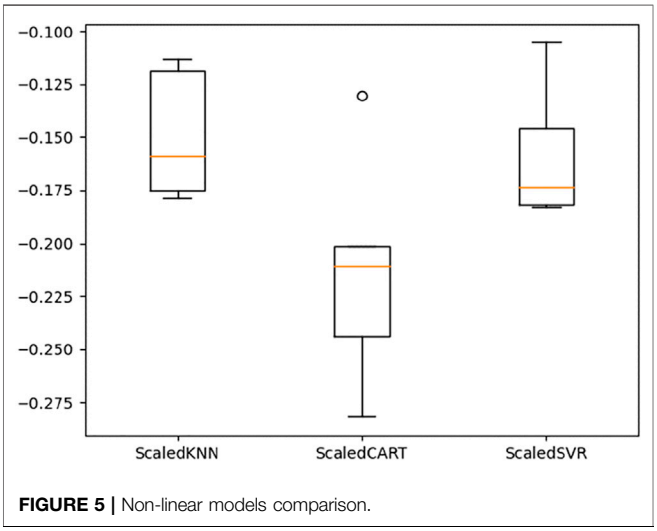



**FIGURE 5 |** Non-linear models comparison.

original data between 2000 and 2015 are shown in **Figure 2A**; **Table 1**, respectively.

Looking at the data, it is revealed that better results could be obtained by taking the logarithm of X2, X4, X5, and Y. The descriptive statistics and histograms of the logarithms of X2, X4, X5, and Y are shown in **Figure 2B**; **Table 2**, respectively.

Data scaling is important for some machine learning algorithms, e.g., KNN and ANN, and less critical for some others such as linear regression. For consistency and easy comparison, the second step of data preparation is standardization of data with its mean and standard deviation rather than normalization of data with its maximum and minimum vales. It is because the data are Gaussian-like than bounded by a maximum and minimum as shown in the histograms.

## Linear Models: Linear Regression, Lasso, and ElasticNet

Three linear models have been applied to the scaled dataset using k-fold cross validation. There is no formal rule for the choice of k. In the present study, we set k = 5 so that the length of the validation data match that of the testing set (i.e., 2016-2018). The box-and-whisker plot of mean and standard deviation of each validation for the three models are shown in **Figure 3**. It can be seen that the mean and standard deviation for linear regression model is tighter than the other two linear models. However, after Lasso and ElasticNet models are tuned for their hyperparameters and used to fit the whole set of training data (i.e., without k-fold cross validation), the rmse between the prediction and the actual data of the training set (2000–2015) are all the same at 0.5482. Furthermore, when they are applied to the testing set (2016–2018), the rmse among the three models are practically the same at 0.6732. To show how good the models are, we plot the actual against the prediction in **Figure 4**. For a good fit, the points should be close to the dotted line. As it can be seen, we can hardly describe that linear models are able to predict $CO_2$ emissions. This prompts us to apply non-linear models accordingly.

## Nonlinear Models: Classification and Regression Tree, Support Vector Regression, and k-Nearest Neighbors Regression

Similar to linear models, we applied k-fold cross validation to the three nonlinear models. The box-and-whisker plot of the three models is shown in **Figure 5**. It can be seen that the performance of SVR and KNN is better than that of CART. The graphs of actual against prediction for SVR and KNN are shown in **Figures 6**, **7**, respectively. It can be seen in **Figures 6**, **7** that nonlinear models, especially KNN, have done much better in predicting $CO_2$ emissions. In particular, if you compare the actual against prediction graph, you can see the points are much tighter and closer to the dotted lines. The rmse are 0.1750 and 0.3641 for the training set and testing set of data when the number of neighbors is set to 2. This is a remarkable improvement over the linear models.

## Ensemble Methods

In this study, five ensemble methods, 2 bagging and 3 boosting algorithms, are applied. As k-fold cross validation randomly divided the dataset, the box-and-whisker plots change every time we run. **Figure 8** shows the typical results for four runs. It can be seen that Extra Trees consistently

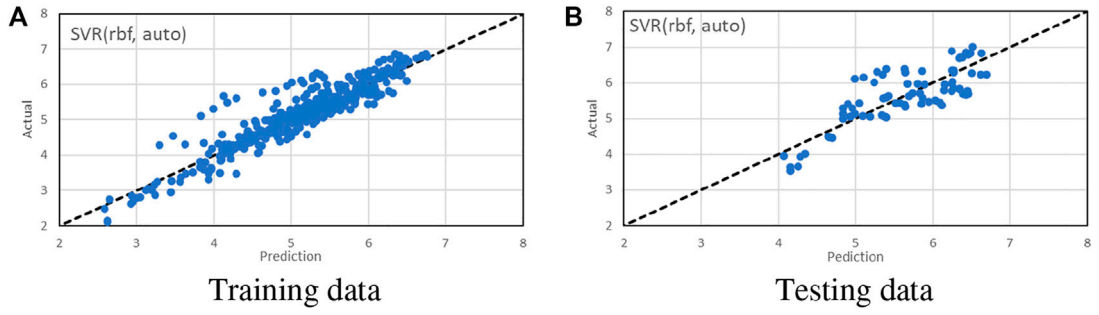



**FIGURE 6 |** Actual vs. prediction of SVR (rbf, auto) model.

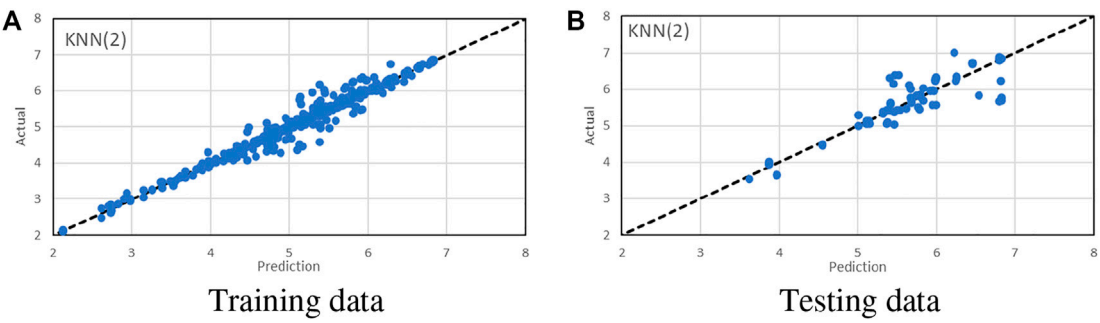



**FIGURE 7 |** Actual vs. prediction of KNN (2) model.

outperformed the other four models in the present study. If we apply Extra Trees algorithm to fit the combined training and validation dataset, we can see it can fit the prediction almost perfectly with the actual data as shown in **Figure 9A**. However, when it is applied to the testing data in **Figure 9B**, it gives a rmse of 0.4128 when the number of trees (or estimators) is 20. One thing to note for ET model is that the rmse are relatively stable with respect to the number of trees: the values of rmse are 0.4370, 0.4128, and 0.4155 when the number of trees is 10, 20, and 50, respectively. Though ET model the best among the five ensemble models under study, its performance is not as good as the KNN model discussed above.

## Artificial Neural Network

As mentioned in *Artificial Neural Networks*, there are three types of layers in ANN. To apply ANN, one needs to determine the number of layers and number of neurons used in each layer. On top of the k-fold cross validation that introduces randomness, the stochastic nature of the model results in different output every time we run the model. Therefore, it is necessary to experiment the combination of these parameters to get the best results. As the number of instances of our dataset is only slightly over 400, one hidden layer is sufficient after experimentation. After random search, it is found that the number of neurons should be between 6 and 15 in both input layer and hidden layer. Then, we run the model at least 10 times for each combination of neurons in the input layer and hidden layer. It is found that the best combination is six neurons in the input layer and 10 neurons in the hidden layer. With this configuration of the network, we run the model 30 times. Then, we take the average of the results, which is shown in **Figure 10**. The values of the rmse of the training and testing data are 0.2430 and 0.4849, respectively. It can be seen that though ANN model performs better than linear models, it is not as good as nonlinear models. Comparatively, its accuracy is only a distant second to KNN model.

Based on rmse for model selection, the results presented above shows that KNN model performed the best, ANN model achieved a distant second and ET came third in predicting $CO_2$ emissions with the dataset described in *Data and the Variables*. In the next section, we shall make use of KNN model and perform sensitivity analysis that would enable policy makers in setting policies to reduce $CO_2$ emissions.

# DISCUSSIONS

Having established that KNN model performs the best in the dataset, we attempt to use KNN model to perform sensitivity analysis of independent variables on $CO_2$ emissions. We would like to determine how the target variable, $CO_2$ emissions, is affected based on changes in other input variables. As there are six input variables, we need to select a base case before we conduct sensitivity analysis. The base case consists of the input

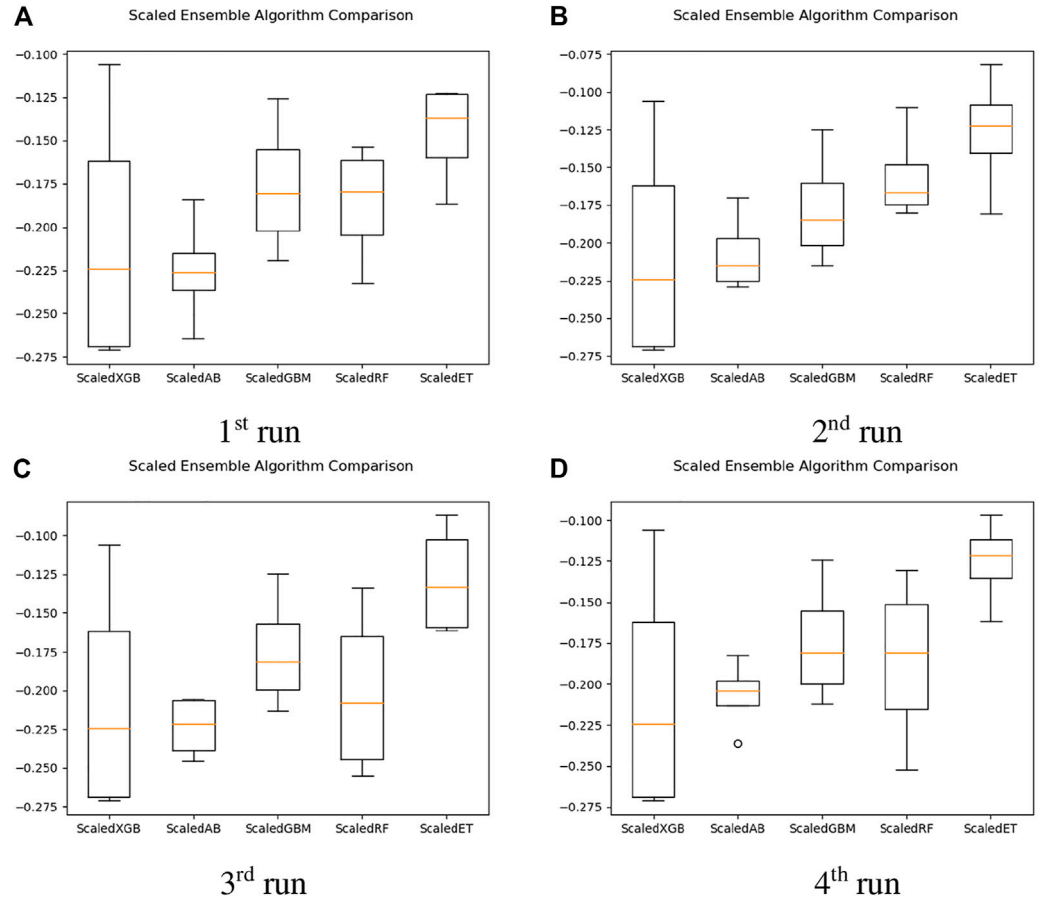



**FIGURE 8 |** Ensemble models comparison–four runs.

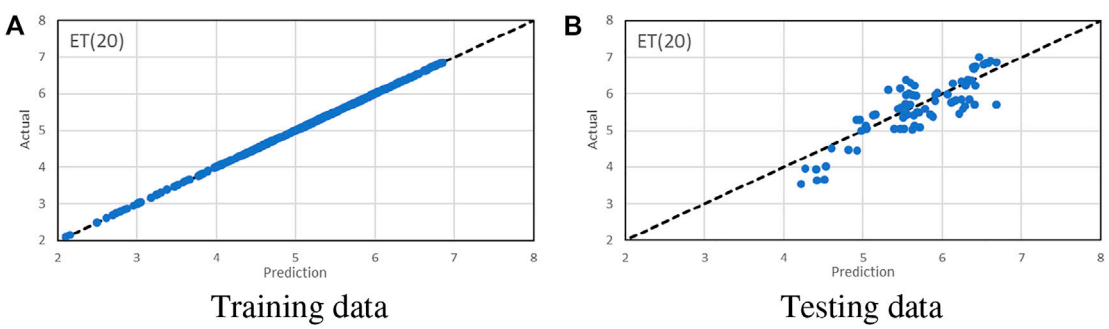



**FIGURE 9 |** Actual vs. prediction of ET (20) model.

variables with the most common values. The procedure is described below.

From the histograms, we have divided each variable into 10 bins of equal width that cover the minimum and maximum. For each variable, say X1–Industrial Structure Rationalization, we pick the midpoint value, $X_{1M}$, of the bin that contains the highest number of data. With six variables, we have $X_{1M}$, $X_{2M}$, . . . and $X_{6M}$ accordingly. Let us call this the "centroid" of input variables.

Now, we can vary the value of one variable, say X1—Industrial Structure Rationalization, from minimum to maximum while keeping the values of other five variables constant at their midpoint values of the highest bin. In this way, we can inspect the sensitivity of variable, X1—Industrial Structure

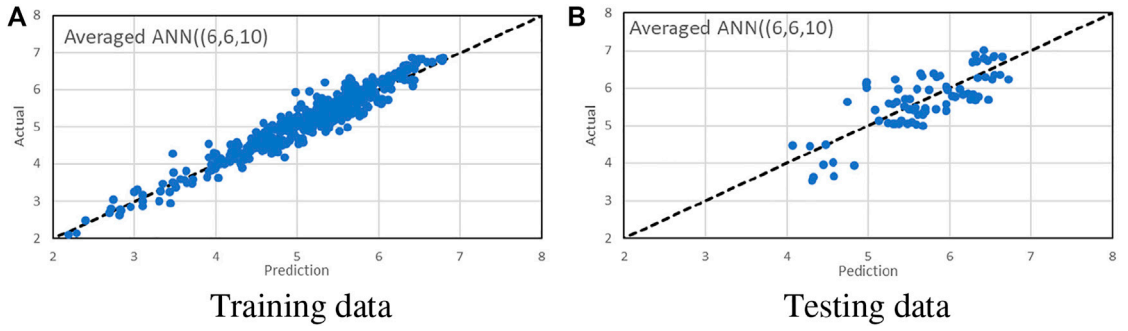



**FIGURE 10 |** Actual vs. prediction of ANN (first layer: six inputs, six neurons; second layer: 10 neurons).

Rationalization around the centroid. We can repeat this analysis to other input variables and form a more complete picture about the six variables affect the CO2 emission around the centroid. The result is shown in **Figure 11**.

First, when all the six variables are at centroid, the predicted Ln($CO_2$ emissions),Y, is 5.4960 (or equivalent to 543.70 million tonnes CO2 emissions), shown with symbol ○ in **Figure 11**. Then, when we adjust one of the variables, the change of Ln($CO_2$ emissions), Y, is summarized in the following.

Industrial Structure Rationalization, X1: The effect on industrial structure rationalization on $CO_2$ emissions is shown in **Figure 11A**. It can be seen that their relationship is nonlinear and nonmonotonic. It exactly demonstrates the strength of machine learning is able to pick up the nonlinearity of the variables and make better predictions. It can be seen that when the industrial structure rationalization increases from 0.7486 to 1.6507, $CO_2$ emissions increase. Beyond that range, its effect is the opposite. It can be interpreted that industrial structure is not the only target for the policy makers. Industrial rationalization index is the equilibrium in economy; it includes output value, sectors of branch of industry employment, and industrial rationalization. If the industrial structure is rationalized, the industry, especially the output of second and third industry, should be in equilibrium, and the regional disparity should be continuously decreased. But the negative influence of industrialization will lead to the increase of $CO_2$ emissions. It means that the $CO_2$ emissions decreasing not only need economic equilibrium but also need the balance between the industrialization and harmful gas emission. Therefore, policy makers should develop economy of each province as rapid as possible. It is because the $CO_2$ emission will eventually decrease after the saturation point at the postindustrialization stage as explained by Bernardini and Galli (1993). On the other hand, the industrial structure rationalization should stay at 0.7486 for some provinces as their $CO_2$ emissions would be at minimum.

GDP on a natural log scale, X2: The effect on GDP on $CO_2$ emissions is shown in **Figure 11B**. It can be seen that $CO_2$ emissions are the most sensitive when the range of GDP is from *exp*(5.0442) to *exp*(5.6349) (or equivalent to 155.12 billion dollars to 280.03 billion dollars). As mentioned at the beginning of *Methodology*, X2 is taken logarithmically. Each interval of the *x*-axis represents 1.8 times of the previous level. Every country would like to develop their economy. Therefore, it would be unlikely that a country would sacrifice economic growth to curb $CO_2$ emissions. **Figure 11B** shows that $CO_2$ emissions will increase when economy grows. It will definitely harm the environment. Furthermore, China cannot simply grow its economy without considering $CO_2$ emissions. It is because one of the pledges China has committed in Paris Accord is to peak $CO_2$ emissions by 2030. However, with the advancement of technology, it is possible to reduce emissions without economic sacrifice. One thing that must be noted is that in **Figure 11B**, there is no inverted U-shaped relation between $CO_2$ emissions and economic growth as found by Galeotti et al. (2006). However, we can see that when GDP grows beyond *exp*(8.5883) (or equivalent to 5,368 billion dollars), $CO_2$ emissions would level off and they could even come down. It means China can fulfill its Paris Accord's pledges.

Urbanization, X3: The impact of urbanization on $CO_2$ emissions is very mixed and complicated as shown in literature reviewed in *Literature Review*. While most of the previous studies indicate a positive relationship between urbanization and $CO_2$ emissions, in this study, it is found that a flanged U-shape is observed as shown in **Figure 11C**. Given that the urbanization is 0.4975 at centroid now, policy makers of China can aim to reduce its $CO_2$ emissions by increasing urbanization to the trough region of 0.571 and 0.6445. This decrease is likely a result of technological innovation in energy usage and efficiency, increasing awareness of environmental issues, and using green technologies (Bekhet and Othman, 2017).

R&D Reinvestment Intensity on a Natural Log Scale, X4: R&D reinvestment intensity stimulates technological advancement, and it also affects economic growth. It can be seen from **Figure 11D** that $CO_2$ emissions increase mildly when reinvestment intensity increases from −6.5673 to its centroid position of −4.4802. Afterwards, it decreases mildly. Looking at the figure, the reinvestment intensity is at critical moment now at its centroid position. If it decreases from its current position, $CO_2$ emissions would decrease too. But it is likely to be accompanied by a decrease of economic growth. The implication is that China should increase its reinvestment intensity so that it could advance technology more rapidly, increase energy usage and efficiency, and make contribution in reducing $CO_2$ emissions accordingly.

Actual Use of Foreign Capital on a Natural Log Scale, X5: The impact of actual use of foreign capital on $CO_2$ emissions is shown in **Figure 11E**. It is observed that $CO_2$ emissions increase rapidly

**FIGURE 11 |** Sensitivity analysis of variables around the centroid of KNN (2). ●represents the centroid of the data – the most populated bin of the data Only one variable varies while keeping the other variables unchanged at centroid values.

when actual use of foreign capital increases from 5.1206 to 5.9213. Then, it levels and even decreases gradually, afterwards. According to pollution haven hypothesis, foreign firms in dirty sectors are more likely to relocate pollution activities from developed countries to poorly regulated developing countries to avoid domestic environmental control cost, which directly undermines the environmental interests of recipient countries like China. This implies that the higher the actual use of foreign capital (FDI), the higher the $CO_2$ emissions, termed "direct" mechanism. However, there is "indirect" mechanism that affect the $CO_2$ emissions. Foreign capital could act as a channel for environmentally friendly technologies, more stringent environmental regulations can be designed and implemented in low-emissions provinces to attract clean foreign capital. So the two mechanisms have the opposite effect on $CO_2$ emissions. In China, actual use of foreign capital of most of the provinces has well passed 5.9212 and reached its centroid position of 10.7257 already. The implication is that the impact of actual use of foreign capital is not significant as the $CO_2$ emissions is quite stable around that position.

Growth Rate of Energy Consumption, X6: When the economy is robust and growing, more energy is consumed. Therefore, it will result in higher $CO_2$ emissions. Energy consumption is in an interesting situation now. It is because $CO_2$ emissions are at a local maximum when the growth rate of energy consumption is at 0.0132 as shown in **Figure 11F**. Interestingly, when the growth rate was higher than 0.0132 during the period under study, $CO_2$ emissions decreased unless the growth rate was too rapid beyond 0.145 level. It could be explained that China has made good use of foreign capital and R&D investment. Therefore, it is expected that cleaner and greener energy such as hydropower and nuclear power have been used when the growth rate increases from 0.0132 to 0.145. Last, but not least, policy makers should refrain from consuming energy beyond a growth rate of 0.145. It is because it can be seen that $CO_2$ emissions increases sharply beyond 0.145 level. Also, the results between 0.5403 and 0.9356 can be ignored as there are only one or two (or even zero) pieces of data of that range.

Noting that the above analysis applies to the centroid position, it provides an overall picture for China as a whole. This approach can also be applied to other position that might be more relevant to individual province.

## CONCLUSION AND POLICY IMPLICATIONS

Following the pledges China has committed in Paris Accord is to peak CO2 emissions by 2030 and the declaration of the 2060 carbon-neutrality goal of Chinese government; it requires proactive measures to be undertaken to reduce carbon emissions while maintaining continuous economic growth and improving in living standards. Against this background, this paper analyzed the effects of industrial structure rationalization index, GDP, urbanization, R&D reinvestment, actual use of foreign capital, and growth rate of energy consumption on forecasting $CO_2$ emissions.

Data across 30 provincial administrative regions of China from 2000 to 2018 are used for the study. Data from 2000 to 2015 are used as training set, and data from 2016 to 2018 are used as testing set. We apply a suite of machine learning algorithms on the testing set and predict the levels of $CO_2$ emissions for the testing set. Machine learning algorithms include linear and nonlinear models, ensemble methods with boosting and bagging, and artificial neural networks. Employing rmse for model selection, results show that k-nearest neigbors (KNN) model performs the best when the number of neighbors is set to two for the present dataset.

Using KNN model, we conducted a sensitivity analysis of $CO_2$ emissions around its centroid position on its dependent variables. The overall findings revealed that economic growth measured by GDP, X2, contribute to higher $CO_2$ emissions. As China needs to maintain its economic growth to continuously improve living standards, it brings several implications for policy makers when setting policies concerning other variables. First, in terms of industrial structure rationalization, X1, not all provinces should develop its industrialization. Some provinces should stay at relatively mild industrialization stage that their $CO_2$ emissions would be at minimum. For other provinces, they should develop their economy as rapidly as possible. It is because $CO_2$ emissions will eventually decrease after saturation point. Therefore, the duration of high $CO_2$ emissions that comes with industrialization would be as short as possible. Second, in terms of urbanization, X3, there is an optimal range for a province. To minimize $CO_2$ emissions, provinces should try to achieve urbanization around 0.571 and 0.6445. With the range, the $CO_2$ emissions would be at minimum and the decrease is likely a result of technological innovation in energy usage and efficiency. It also suggests that a province should not be too densely populated. Third, the result of R&D reinvestment intensity, X4, suggests that China should increase its reinvestment intensity further. At present, there is a positive relationship between $CO_2$ emissions and reinvestment intensity. Therefore, it seems that monies for R&D reinvestment have not been put, or not enough, into green technology yet. Only when there is a further increase of R&D reinvestment intensity into green technology, there will be a decrease of $CO_2$ emissions. Fourth, it is found that the impact of the actual use of foreign capital, X5, on $CO_2$ emissions is insignificant, relatively speaking, when compared with other variables. If we assume that R&D reinvestment is associated with actual use of foreign capital, policy makers should prioritize the use of foreign capital for R&D investment on green technology. That would reduce $CO_2$ emissions while maintaining economic growth. Last, it is possible to increase the growth rate of energy consumption, X6, gradually if R&D reinvestment and use of foreign capital are directed towards cleaner and green energy sources such as hydropower and nuclear power. Policy makers must refrain from consuming energy beyond a growth rate of 0.1450 for economic growth. Otherwise, $CO_2$ emissions would increase rapidly and may jeopardize the pledges China committed in Paris Accord and the 2060 carbon-neutrality declaration. In summary, the above policy implications provide a blueprint for policy makers for ensuing environmentally sustainability economic development in China.

It is worth noting that the approach applied in this study can easily be replicated for other countries to make better forecasting

of $CO_2$ emissions for the future. The major constraint of this approach is the data limitation. For successful application of machine learning, the number of data required is usually more than traditional econometric models. With more data, more advanced machine learning algorithms can be applied to further check the robustness of the findings.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were used and analyzed in this study. The sources of data are listed in Data and the Variables section of this paper in details.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

The paper is supported by Program for Innovation Research in Central University of Finance and Economics and by General Project of Beijing Social Science Foundation Research Base (18JDLJB001), Research on Beijing Urban Governance Path Based on the Optimization of Urban Multi-dimensional Spatial Structure.

## REFERENCES

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., and Arshad, H. (2018). State-Of-The-Art In Artificial Neural Network Applications: A Survey. *Heliyon* 4, e00938. doi:10.1016/j.heliyon.2018.e00938

Acheampong, A. O. (2018). Economic Growth, CO2 Emissions and Energy Consumption: What Causes what and where? *Energ. Econ.* 74, 677–692. doi:10.1016/j.eneco.2018.07.022

Akpan, U. F., and Chuku, A. (2011). Economic Growth and Environmental Degradation in Nigeria: Beyond the Environmental Kuznets Curve. *Mpra Paper* 8 (5), 568–577. doi:10.1108/14777830710778328

Al-mulali, U., Fereidouni, H. G., Lee, J. Y. M., and Sab, C. N. B. C. (2013). Exploring the Relationship between Urbanization, Energy Consumption, and CO2 Emission in MENA Countries. *Renew. Sustain. Energ. Rev.* 23, 107–112. doi:10.1016/j.rser.2013.02.041

Ang, B. W., and Zhang, F. Q. (2000). A Survey of index Decomposition Analysis in Energy and Environmental Studies. *Energy* 25 (12), 1149–1176. doi:10.1016/s0360-5442(00)00039-6

Ang, J. B. (2009). CO2 Emissions, Research and Technology Transfer in China. *Ecol. Econ.* 68 (10), 2658–2665. doi:10.1016/j.ecolecon.2009.05.002

Antweiler, W., Copeland, B. R., and Taylor, M. S. (2001). Is Free Trade Good for the Environment? *Am. Econ. Rev.* 91, 877–908. doi:10.1257/aer.91.4.877

Bai, Y., Deng, X., Gibson, J., Zhao, Z., and Xu, H. (2019). How Does Urbanization Affect Residential CO2 Emissions? an Analysis on Urban Agglomerations of China. *J. Clean. Prod.* 209, 876–885. doi:10.1016/j.jclepro.2018.10.248

Behera, S. R., and Dash, D. P. (2017). The Effect of Urbanization, Energy Consumption, and Foreign Direct Investment on the Carbon Dioxide Emission in the SSEA (South and Southeast Asian) Region. *Renew. Sustain. Energ. Rev.* 70, 96–106. doi:10.1016/j.rser.2016.11.201

Bekhet, H. A., and Othman, N. S. (2017). Impact of Urbanization Growth on Malaysia CO2 Emissions: Evidence from the Dynamic Relationship. *J. Clean. Prod.* 154, 374–388. doi:10.1016/j.jclepro.2017.03.174

Bernardini, O., and Galli, R. (1993). Dematerialization: Long-Term Trends in the Intensity of Use of Materials and Energy. *Futures* 25, 431–448. doi:10.1016/0016-3287(93)90005-e

Bertram, C., Keywan Riahi, K., Hilaire, J., Bosetti, V., Drouet, L., Fricko, O., et al. (2021). Energy System Developments and Investments in the Decisive Decade for the Paris Agreement Goals. *Environ. Res. Lett.* 16 (7). doi:10.1088/1748-9326/ac09ae

Candau, F., and Dienesch, E. (2017). Pollution haven and Corruption paradise. *J. Environ. Econ. Manage.* 85, 171–192. doi:10.1016/j.jeem.2017.05.005

Cao, Z., Wei, J., and Chen, H. B. (2016). CO2 Emissions and Urbanization Correlation in China Based on Threshold Analysis. *Ecol. Indicators* 61, 193–201. doi:10.1016/j.ecolind.2015.09.025

Chaudhari, P. (2019). Importance of Feature Scaling for Artificial Neural Networks and K-Nearest Neighbors. Available at: https://medium.com/@piyush.kailash.chaudhari/importance-of-feature-scaling-for-artificial-neural-networks-and-k-nearest-neighbors-4b7aa618d5ea. (Accessed June 5, 2021).

Chen, S. (2011). The Fluctuation and Decrease Mode of China's Carbon Emission Intensity and its Economic Explanation. *The J. World Economy* 34 (04), 124–143. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2011&filename=SJJJ201104009&v=PKVIHlLP1q%25mmd2FhF%25mmd2FkK6z80YIy5msaiQr1ZOQco3gj65bo4cCpLmeL1cPGQf9VKfY9q. (Accessed June 5, 2021). doi:10.1111/j.1467-9701.2011.01370.x

Chudik, A., and Pesaran, M. H. (2015). Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *J. Econom.* 188 (2), 393–420. doi:10.1016/j.jeconom.2015.03.007

Cole, M. A., and Neumayer, E. (2004). Examining the Impact of Demographic Factors on Air Pollution. *Popul. Environ.* 26, 5–21. doi:10.1023/b:poen.0000039950.85422.eb

Cui, C., Cai, B., and Wang, G. Z. (2020). Decennary Spatial Pattern Changes and Scaling Effects of CO2 Emissions of Urban Agglomerations in China. *Cities* 105, 102818. doi:10.1016/j.cities.2020.102818

Du, W. C., and Xia, X. H. (2018). How Does Urbanization Affect GHG Emissions? A Cross-Country Panel Threshold Data Analysis. *Appl. Energ.* 229, 872–883. doi:10.1016/j.apenergy.2018.08.050

Elliott, R. J. R., Sun, P. S., and Chen, S. (2013). Energy Intensity and Foreign Direct Investment: A Chinese City-Level Study. *Energ. Econ.* 40, 484–494. doi:10.1016/j.eneco.2013.08.004

Fan, H., Hashmi, S. H., Habib, Y., and Ali, M. (2020). How Do Urbanization and Urban Agglomeration Affect CO2 Emissions in South Asia? Testing Non-linearity Puzzle with Dynamic STIRPAT Model. *Chin. J. Urban Environ. Stud.* 08 (01), 205000308. doi:10.1142/s2345748120500037

Fan, Y., Liu, L.-C., Wu, G., and Wei, Y.-M. (2006). Analyzing Impact Factors of CO2 Emissions Using the STIRPAT Model. *Environ. Impact Assess. Rev.* 26 (4), 377–395. doi:10.1016/j.eiar.2005.11.007

Friedl, B., and Getzner, M. (2003). Determinants of CO2 Emissions in a Small Open Economy. *Ecol. Econ.* 45, 133–148. doi:10.1016/s0921-8009(03)00008-9

Galeotti, M., Lanza, A., and Pauli, F. (2006). Reassessing the Environmental Kuznets Curve for CO2 Emissions: a Robustness Exercise. *Ecol. Econ.* 57 (1), 152–163. doi:10.1016/j.ecolecon.2005.03.031

Gan, C., Zheng, R., and Yu, D. (2011). An Empirical Study on the Effects of Industrial Structure on Economic Growth and Fluctuations in China. *Econ. Res. J.* 46 (05), 4–16+31. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2011&filename=JJYJ201105002&v=jXi17UDWTaaqRHWJh8aSKBVxDK77Pz%25mmd2FC3DruT4vnekzOQ8DZnDbFnuQFv3dcUcCh. (Accessed June 5, 2021).

Grossman, G. M., and Krueger, A. B. (1991). *Environmental Impacts of a North American Free Trade Agreement*. National Bureau of Economic Research Working Paper. No. 3914. doi:10.3386/w3914 Available at: https://www.nber.org/papers/w3914 (Accessed August 15, 2021).

Grossman, G. M., and Krueger, A. B. (1995). Economic Growth and the Environment. *Q. J. Econ.* 110 (2), 353–377. doi:10.2307/2118443

Han, X., Cao, T., and Sun, T. (2019). Analysis on the Variation Rule and Influencing Factors of Energy Consumption Carbon Emission Intensity in China's Urbanization Construction. *J. Clean. Prod.* 238, 124605. doi:10.1016/j.jclepro.2019.117958

Holtz-Eakin, D., and Selden, T. M. (1995). Stoking the Fires? CO2 Emissions and Economic Growth. *J. Public Econ.* 57, 85–101. doi:10.1016/0047-2727(94)01449-x

Hossain, M. S. (2011). Panel Estimation for CO2 Emissions, Energy Consumption, Economic Growth, Trade Openness and Urbanization of Newly Industrialized Countries. *Energy Policy* 39 (11), 6991–6999. doi:10.1016/j.enpol.2011.07.042

Kasman, A., and Duman, Y. S. (2015). CO2 Emissions, Economic Growth, Energy Consumption, Trade and Urbanization in New EU Member and Candidate Countries: A Panel Data Analysis. *Econ. Model.* 44, 97–103. doi:10.1016/j.econmod.2014.10.022

Khan, Z., Murshed, M., Dong, K., and Yang, S. (2021). The Roles of export Diversification and Composite Country Risks in Carbon Emissions Abatement: Evidence from the Signatories of the Regional Comprehensive Economic Partnership Agreement. *Appl. Econ.*, 1–19. doi:10.1080/00036846.2021.1907289

Khan, Z., Sisi, Z., and Siqun, Y. (2019). Environmental Regulations an Option: Asymmetry Effect of Environmental Regulations on Carbon Emissions Using Non-linear ARDL. *Energy Sourc. A: Recovery, Utilization, Environ. Effects* 41, 137–155. doi:10.1080/15567036.2018.1504145

Lantz, V., and Feng, Q. (2006). Assessing Income, Population, and Technology Impacts on CO2 Emissions in Canada: Where's the EKC? *Ecol. Econ.* 57, 229–238. doi:10.1016/j.ecolecon.2005.04.006

Li, K., and Lin, B. (2015). Impacts of Urbanization and Industrialization on Energy Consumption/CO2 Emissions: Does the Level of Development Matter? *Renew. Sustain. Energ. Rev.* 52, 1107–1122. doi:10.1016/j.rser.2015.07.185

Li, Z.-Z., Li, R. Y. M., Malik, M. Y., Murshed, M., Khan, Z., and Umar, M. (2021). Determinants of Carbon Emission in China: How Good Is Green Investment? *Sustainable Prod. Consumption* 27, 392–401. doi:10.1016/j.spc.2020.11.008

Liddle, B., and Lung, S. (2010). Age-structure, Urbanization, and Climate Change in Developed Countries: Revisiting STIRPAT for Disaggregated Population and Consumption-Related Environmental Impacts. *Popul. Environ.* 31 (5), 317–343. doi:10.1007/s11111-010-0101-5

Lin, B., and Du, K. (2014). Understanding the Changes in China's Energy Intensity: a Comprehensive Decomposition Framework. *J. World Economy* 37 (04), 69–87. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?filename=SJJJ201404005&dbcode=CJFQ&dbname=CJFD2014&v=7ZdISdpBS7Ocxj7t_Kly5sUetQA6bDAudZuqf7tsBLFTt4L-55mxiTThz9tEW3Jy. (Accessed June 5, 2021).

Lin, B., and Du, K. (2013). What Is the Driving Force of My Country's Energy Productivity Growth: Based on the Decomposition of Distance Function. *J. Financial Res.* (09), 84–96. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?filename=JRYJ201309007&dbcode=CJFQ&dbname=CJFD2013&v=gazb_TmM57RsImpyvmwXBe67SvuxDMQJVNSt68dYl5hNcsOm3uxVUvRoF-W2Cott. (Accessed June 5, 2021).

Lin, B., and Jiang, L. (2009). Prediction of Environmental Kuznets Curve of Carbon Dioxide in China and Analysis of Influencing Factors. *Manage. World* 4, 2736, 2009. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2009&filename=GLSJ200904005&v=hnvCX%25mmd2Bnv89Jb8ZEhkoTGxX8Ufj8SzOC2I8xzaqyApuknmDeyvWbifRExmNIGtEQn. (Accessed June 5, 2021).

Liu, J., Murshed, M., Chen, F., Shahbaz, M., Kirikkaleli, D., and Khan, Z. (2021). An Empirical Analysis of the Household Consumption-Induced Carbon Emissions in China. *Sustain. Prod. Consumption* 26, 943–957. doi:10.1016/j.spc.2021.01.006

Liu, X., and Bae, J. (2018). Urbanization and Industrialization Impact of CO2 Emissions in China. *J. Clean. Prod.* 172, 178–186. doi:10.1016/j.jclepro.2017.10.156

Liu, Z., Geng, Y., Lindner, S., and Guan, D. (2012). Uncovering China's Greenhouse Gas Emission from Regional and Sectoral Perspectives. *Energy* 45 (1), 1059–1068. doi:10.1016/j.energy.2012.06.007

Ma, Q., Murshed, M., and Khan, Z. (2021). The Nexuses between Energy Investments, Technological Innovations, Emission Taxes, and Carbon Emissions in china. *Energy Policy* 155 (30), 112345. doi:10.1016/j.enpol.2021.112345

Martin, W. (2008). The Carbon Kuznets Curve: A Cloudy Picture Emitted by Bad Econometrics? *Resource Energ. Econ.* v30, 388–408.

Martinez-Zarzoso, I., and Bengochea-Morancho, A. (2004). Pooled Mean Group Estimation for an Environmental Kuznets Curve for CO2. *Econ. Lett.* 82, 121–126.

Moomaw, W. R., and Unruh, G. C. (1997). Are Environmental Kuznets Curves Misleading Us? the Case of CO2 Emissions. *Envir. Dev. Econ.* 2, 451–463. doi:10.1017/s1355770x97000247

Murshed, M., and Dao, N. T. T. (2020). Revisiting the CO2 Emission-Induced EKC Hypothesis in South Asia: the Role of Export Quality Improvement. *GeoJournal* (8). doi:10.1007/s10708-020-10270-9

Murshed, M., Ali, S. R., and Banerjee, S. (2020). Consumption of Liquefied Petroleum Gas and the EKC Hypothesis in South Asia: Evidence from Cross-Sectionally Dependent Heterogeneous Panel Data with Structural Breaks. *Energ. Ecol. Environ.* 6 (4), 353–377. doi:10.1007/s40974-020-00185-z

Poumanyvong, P., and Kaneko, S. (2010). Does Urbanization lead to Less Energy Use and Lower CO2 Emissions? A Cross-Country Analysis. *Ecol. Econ.* 70 (2), 434–444. doi:10.1016/j.ecolecon.2010.09.029

Rahman, M. M., Saidi, K., and Mbarek, M. B. (2020). Economic Growth in South Asia: the Role of CO2 Emissions, Population Density and Trade Openness. *Heliyon* 6 (5), e03903. doi:10.1016/j.heliyon.2020.e03903

Rehman, A., Ulucak, R., Murshed, M., Ma, H., and Isik, C. (2021b). Carbonization and Atmospheric Pollution in China: The Asymmetric Impacts of Forests, Livestock Production, and Economic Progress on CO2 Emissions. *J. Environ. Manage.* 294, 1–10. doi:10.1016/j.jenvman.2021.113059

Rehman, A., Ma, H., Ozturk, I., Murshed, M., and Dagar, V. (2021a). The Dynamic Impacts of CO2 Emissions from Different Sources on Pakistan's Economic Progress: a Roadmap to Sustainable Development. *Environ. Dev. Sustain.*. doi:10.1007/s10668-021-01418-9

Sachs, J., Panayotou, T., and Peterson, A. (1999). *Developing Countries and the Control of Climate Change: A Theoretical Perspective and Policy Implications*. CAER II Discussion Paper. No. 44. Available at: https://www.semanticscholar.org/paper/Developing-Countries-and-the-Control-of-Climate-%3A-A-Sachs-Panayotou/4a052fa713252eadb020ced77323cbf5c0d7a360 and https://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/1999/HIIDCAERII_DevelopingCountriesandtheControlofClimateChange_Nov1999.pdf. (Accessed August 15, 2021).

Sadorsky, P. (2013). Do urbanization and Industrialization Affect Energy Intensity in Developing Countries? *Energ. Econ.* 37, 52–59. doi:10.1016/j.eneco.2013.01.009

Saidi, K., and Hammami, S. (2015). The Impact of CO2 Emissions and Economic Growth on Energy Consumption in 58 Countries. *Energ. Rep.* 1, 62–70. doi:10.1016/j.egyr.2015.01.003

Shafik, N. (1994). Economic Development and Environmental Quality: an Econometric Analysis. *Oxford Econ. Pap.* 46, 757–773. doi:10.1093/oep/46.supplement_1.757

Shahbaz, M., Chaudhary, A. R., and Ozturk, I. (2017). Does Urbanization Cause Increasing Energy Demand in Pakistan? Empirical Evidence from STIRPAT Model. *Energy* 122, 83–93. doi:10.1016/j.energy.2017.01.080

Shahbaz, M., Nasir, M. A., and Roubaud, D. (2018). Environmental Degradation in France: the Effects of FDI, Financial Development, and Energy Innovations. *Energ. Econ.* 74, 843–857. doi:10.1016/j.eneco.2018.07.020

Shen, X., Chen, Y., and Lin, B. (2021). The Impacts of Technological Progress and Industrial Structure Distortion on China's Energy Intensity. *Econ. Res. J.* (02), 157–173. Available at: http://kns.cnki.net/kcms/detail/11...F.20210406.1417.020.html.

Shi, X, K. Wang., Cheong, T. S., and Zhang, H. (2020). Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey da, 104ta. *Energ. Econ.* 92942, 1–13. doi:10.1016/j.eneco.2020.104942

Wang, C. (2011). Sources of Energy Productivity Growth and its Distribution Dynamics in China. *Resource Energ. Econ.* 33 (1), 279–292. doi:10.1016/j.reseneeco.2010.06.005

Wang, F., Wu, L., and Yang, C. (2010). Driving Factors for Growth of Carbon Dioxide Emissions during Economic Development in China. *Econ. Res. J.* 45 (02), 123–136. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2010&filename=JJYJ201002011&v=O6IUlfr0xi3lU0SXtaGeZhrF%25mmd2FUmFUiFJAc%25mmd2BkKLVVy2pyXHnkjBYtLEQTuMQiPFLT. (Accessed August 15, 2021).

Wang, P., Wu, W., Zhu, B., and Wei, Y. (2013). Examining the Impact Factors of Energy-Related CO2 Emissions Using the STIRPAT Model in Guangdong Province, China. *Appl. Energ.* 106, 65–71. doi:10.1016/j.apenergy.2013.01.036

Wang, Z., Yin, F., Zhang, Y., and Zhang, X. (2012). An Empirical Research on the Influencing Factors of Regional CO2 Emissions: Evidence from Beijing City, China. *Appl. Energ.* 100, 277–284. doi:10.1016/j.apenergy.2012.05.038

Wei, W. X., Yang, F., and Iang, F. (2010). Impact of Technology advance on Carbon Dioxide Emission in China. *Stat. Res.* 27 (07), 36–44. doi:10.19343/j.cnki.11-1302/c.2010.07.006

Yang, L., Zhu, J., and Jia, Z. (2019). Influencing Factors and Current Challenges of CO2 Emission Reduction in China: a Perspective Based on Technological Progress. *Econ. Res. J.* 54, 118–132. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2020&filename=JJYJ201911009&v=ffsbFCXZ%25mmd2BXHaBhuX5uzpMjUQQJLF2u2ENdBwXz9hFkKQ1VS5NSK9k258gY9lErsg. (Accessed June 5, 2021).

Yang, L., and Li, Z. (2017). Technology advance and the Carbon Dioxide Emission in China - Empirical Research Based on the Rebound Effect. *Energy Policy* 101 (FEB.), 150–161. doi:10.1016/j.enpol.2016.11.020

York, R., Rosa, E. A., and Dietz, T. (2003). STIRPAT, IPAT and ImPACT: Analytic Tools for Unpacking the Driving Forces of Environmental Impacts. *Ecol. Econ.* 46 (3), 351–365. doi:10.1016/s0921-8009(03)00188-5

Young, A. (1998). Growth without Scale Effects. *J. Polit. Economy* 106 (1), 41–63. doi:10.1086/250002

Yu, Y., and Du, Y. (2018). Impact of Technological Innovation on CO2 Emissions and Emissions Trend Prediction on 'new normal' Economy in China. *Atmos. Pollut. Res.* 10, 152–161. doi:10.1016/j.apr.2018.07.005/

Yu, J., Shi, X., Guo, D., and Yang, L. (2021). Economic Policy Uncertainty (EPU) and Firm Carbon Emissions: Evidence Using a China Provincial EPU Index. *Energy Economics* 94, 105071.

Zhang, C., and Lin, Y. (2012). Panel Estimation for Urbanization, Energy Consumption and CO2 Emissions: A Regional Analysis in China. *Energy Policy* 49, 488–498. doi:10.1016/j.enpol.2012.06.048

Zhu, H.-M., You, W.-H., and Zeng, Z.-f. (2012). Urbanization and CO2 Emissions: A Semi-parametric Panel Data Analysis. *Econ. Lett.* 117 (3), 848–850. doi:10.1016/j.econlet.2012.09.001

Zi, C., Jie, W., and Jie, C. (2016). CO2 Emissions and Urbanization Correlation in China Based on Threshold Analysis. *Ecol. Indicators* 61, 193–201. doi:10.1016/j.ecolind.2015.09.013

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Li, Siu and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fenrg.2021.749065

# Effects of Age Dependency and Urbanization on Energy Demand in BRICS: Evidence From the Machine Learning Estimator

*Zhou Lu*[1,2], *Mantu Kumar Mahalik*[3], *Hemachandra Padhan*[4], *Monika Gupta*[5] *and Giray Gozgor*[6]*

[1]School of Hospitality Administration, Zhejiang Yuexiu University, Shaoxing, China, [2]School of Economics, Tianjin University of Commerce, Tianjin, China, [3]Department of Humanities and Social Sciences, Indian Institute of Technology Kharagpur, Kharagpur, India, [4]Department of Humanities and Social Sciences, Indian Institute of Technology Madras, Chennai, India, [5]Economics Area, S. P. Jain Institute of Management and Research (SPJIMR), Mumbai, India, [6]Faculty of Political Sciences, Istanbul Medeniyet University, Istanbul, Turkey

This paper examines the effects of age dependency ratio (the young age, old-age and overall age) and urbanization on renewable and non-renewable energy consumption in Brazil, India, China, and South Africa, considering the panel data from 1990 to 2019. We control economic growth and foreign direct investment inflows as key factors in the energy demand function using the Stochastic Impacts by Regression on Population, Affluence and Technology approach. Empirical analysis has been implemented using the Kernel Regularized Least Squares machine learning method to solve possible classification problems in the traditional regressions without relying on the linearity assumption. It is observed that the young age dependency, overall age dependency, and urbanization negatively affect both renewable and non-renewable energy demand. On the contrary, old-age dependency and economic growth are positively associated with renewable and non-renewable energy demand. The mixed effects of foreign direct investment inflows on renewable and non-renewable energy demand patterns are also found. Thus, the findings suggest that environment policymakers in the BRICS economies should prioritize urbanization, young age, and overall age population to improve energy efficiency.

**Keywords: renewable energy demand, non-renewable energy demand, machine learning estimator, age dependency, urbanization, STIRPAT**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, Hong Kong, SAR China*

***Reviewed by:***
*Pabitra Kumar Jena,*
*Shri Mata Vaishno Devi University, India*
*Yuhua Song,*
*Zhejiang University, China*

***Correspondence:**
*Giray Gozgor*
*giray.gozgor@medeniyet.edu.tr*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *28 July 2021*
***Accepted:*** *16 August 2021*
***Published:*** *26 August 2021*

***Citation:***
*Lu Z, Mahalik MK, Padhan H, Gupta M and Gozgor G (2021) Effects of Age Dependency and Urbanization on Energy Demand in BRICS: Evidence From the Machine Learning Estimator. Front. Energy Res. 9:749065. doi: 10.3389/fenrg.2021.749065*

## INTRODUCTION

Since the early 1990s, renewable energy consumption and production have risen in developing and developed economies (British Petroleum, 2020). The rise of renewable energy may have four important issues. The first issue is the potential problems of climate change and renewable energy usage in decreasing $CO_2$ emissions. Therefore, renewable energy can mitigate the adverse effects of climate change on economic performance (Payne, 2009). The second issue is technological progress and declining investment costs on renewable energy facilities (Apergis and Payne, 2010). The third is that governments have provided various policy support for renewable energy consumption and production, e.g., they have implemented credit easing and tax deduction policies on renewable energy investments. There are also rising production standards in renewable energy facilities (Apergis and Payne, 2012). Finally, the volatility and rise in crude oil prices have promoted renewable energy consumption (Gozgor, 2018a). However, non-renewable energy consumption

**TABLE 1 |** Summary of representative literature.

| Study | Countries covered | Period | Determinants | Major findings |
|---|---|---|---|---|
| A. Impact of Urbanization on Energy Demand | | | | |
| Jones (1989) | 59 developing countries | 1980 | Urbanization, economic growth, industrialization and population density | The elasticity of energy consumption for urbanization is found between 0.35 and 0.48 |
| Jones (1991) | 59 developing countries | 1980 | Urbanization, per capita income, industrialization | A 10% increase in the population living in an urban area would drive a 4.5% increase in energy consumption per capita GDP. |
| Parikh and Shukla (1995) | Around 78 developed and developing countries of the World | 1965–87 | Urbanization, GNI per capita, population density, the share of agriculture in GDP | Urbanization leads to an increase in aggregate energy use |
| Pachauri and Jiang (2008) | India and China | 1980–81-2004–5 | Urbanization, income, population size | Less than proportionate change in energy use concerning urbanization changes; rural energy consumption exceeds urban energy consumption because of inefficient solid fuels used by households |
| Liu (2009) | China | 1978–2008 | Population growth, urbanization and economic growth | Unidirectional causality from urbanization to total energy consumption in the long run and the short run. Significant relationship among all variables |
| Poumanyvong and Kaneko (2010) | 99 countries | 1975–2005 | Population size, economic growth, energy intensity, industrialization and share of services in GDP, $CO_2$ emissions and emission intensity and urbanization | The effect varies in different development stages—a negative relationship between urbanization in low-income countries where the positive relationship is in the middle and high-income group |
| O'Neill et al. (2012) | India and China | 2004 as the base year | Urbanization, economic growth and carbon emissions | Less than a proportionate change of urbanization on energy use, whereas the income effect is strong |
| Shahbaz and Lean (2012) | Tunisia | 1971–2008 | Industrialization, economic growth, financial development and urbanization | Long-run bidirectional causality between financial development and energy consumption as well as between industrialization and energy consumption |
| Sadorsky (2014) | 18 emerging economies which including BRICS | 1971–2008 | Income, urbanization and industrialization | Urbanization depresses energy consumption, whereas industrialization and income increase it in the long run |
| Ghosh and Kanjilal (2014) | India | 1971–2008 | Economic activity, urbanization | Unidirectional causality from energy consumption to economic activity and from economic activity to urbanization |
| Li and Lin (2015) | 73 countries, including BRICS except for Russia | 1971–2010 | Urbanization, GDP per capita, industrialization and energy intensity | Urbanization decreases energy demand in low-income countries, whereas, in middle-income countries, it increases energy consumption. In the case of high-income countries, urbanization has a significant effect on energy demand |
| Sheng et al. (2017) | 78 countries | 1995–2012 | Urbanization, GDP, industrial structure, population size and energy efficiency | Actual energy consumption increases by 0.495%, and energy efficiency decrease by 0.201%, with a 1% increase in the average urbanization index |
| Zhao and Zhang (2018) | China | 1980–2010 | Urbanization, GDP and industrialization rate | A 1% increase in the urban population will increase national energy consumption by 1.4%, and the urban population has 50% more energy consumption than rural households |
| Mrabet et al. (2019) | 28 developed and emerging economies, including India, China and South Africa | 1980–2014 | Income growth, urbanization, industrialization, $CO_2$ emissions, energy price | A 1% increase in urbanization leads to a 0.72% increase in energy consumption in the long run |
| B. Impact of Demographic Factors on Energy Demand | | | | |
| O'Neill and Chen (2002) | US (cross-section data of households) | 1993–94 | Age structure, income, household size and composition | There is a direct and positive relationship between household energy use and age. In contrast, transportation energy use shows an inverted U-shaped pattern and increased continuously with age reaching the peak of 51–55 years' energy use deciles at an older age; the substantial |

(Continued on following page)

**TABLE 1 |** (*Continued*) Summary of representative literature.

| Study | Countries covered | Period | Determinants | Major findings |
|---|---|---|---|---|
| | | | | influence of demographic factor on energy use |
| York (2007) | European Union countries | 1960–2000 | Urbanization, GDP per capita, age structure and population size | A substantial positive role of population size and age structure in energy demand. This study also showed that the old age population has an increasing role in energy demand |
| Liddle and Lung (2010) | 17 developed countries | 1960–2005 | Urbanization, different age groups of 20–34, 35–49 and 50–64 years of age, GDP and other variables | People in the 35–49 age cohort have negative, whereas people in the 50–64 age cohort positively affect energy demand and the positive role of urbanization |
| Kim and Seo (2012) | 53 countries | 1976–2009 | Energy price, aging, GDP, government consumption share of GDP, investment share of GDP and openness | The inverted U-shaped relationship shows positive and negative relationships after the inflection point, 18–23% of the elderly over the working age |
| Liddle (2014) | Review of macro-level cross country studies | — | Population size, population density, age structure, household size and urbanization | Review establishes the positive role of urbanization on energy use; higher population density is lower energy demand |
| Liu et al. (2015) | China | 1990–2012 | Population density, GDP, the proportion of industrial output and tertiary industry output and industry energy intensity | Population density harms energy consumption, whereas industrial output has a significant positive impact on energy consumption |
| Hasanov and Mikayilov (2017) | Azerbaijan | 2000–2012 | Age groups in three categories of 0–14, 15–65 and above 65 years and GDP | The middle age group, the working-age group, is the most significant contributor to residential energy demand |
| Estiri and Zagheni (2019) | US | Household data of residential survey for 4 years (1987, 1990, 2005, 2009) | Different age groups, heating and cooling degree days, household income, type of housing unit, housing size and age of housing unit | Household size has a significant impact on energy consumption; residential energy consumption increases with the life course |

*Source: Authors' works.*

has steadily risen due to the increasing demand for household consumption and production in different sectors (Belaid and Youssef, 2017).

This paper examines non-renewable and renewable energy consumption determinants in four emerging economies, i.e., Brazil, India, China and South Africa, within a panel data framework primarily focusing on urbanization and age dependency factors. Unlike previous papers, the current study controls the effects of different age-dependency measures with urbanization to capture the role of demographics on the energy demand. Given that there is vast literature investigating the relationship between urbanization and energy demand, previous papers have not adequately examined the effects of urbanization combined with age-dependency on renewable and non-renewable energy demand (see **Table 1**). To the best of our knowledge, a limited number of studies have investigated population dynamics and age dependency as determinants of energy demand, particularly in developing and emerging economies. Population dynamics, for example, a higher young population, can create additional demand for energy in developing countries. There can be a significant role of aging in the structural transformation of economies. It is important to note that age-dependency affects energy consumption patterns and can shift towards more relevant and energy-saving products for the elderly. For instance, the energy sector, housekeeping, and health services create a higher demand in countries with an elderly population (Aiyar et al., 2016). Therefore, economies' structural transformation will occur on the supply-side and translate the manufacturing sector's economy to services (Siliverstovs et al., 2011; Gozgor, 2018b).

The paper focuses on the cases of Brazil, China, India, and South Africa as a part of the BRICS region for empirical testing within a panel data framework. These countries are important in terms of energy demand. According to World Bank (2021), China, India, and Brazil are the first, the third, and the tenth-largest economies in 2019 when we consider the Gross Domestic Product (GDP) based on Purchasing Power Parity. These countries have also experienced a consistent increase in renewable energy demand in the 2010s (Gozgor, 2016). For example, China became the largest renewable energy consumption in the World in 2019. Brazil and India are also the top ten economies globally in terms of GDP, but they are also the largest renewable energy consumers (British Petroleum, 2020).

Furthermore, these countries also have an increasing middle-class, which will create demand for different aspects of energy, including renewable energy. Therefore, they will have a growing demand in the domestic market in the forthcoming years, and hence they can be attractive markets for renewable energy

sources. At this stage, investors from abroad are very keen to invest in these developing countries. Expanding the domestic market and foreign investments will increase demand for all energy segments due to rising globalization (Gozgor et al., 2020). Therefore, we control the GDP and Foreign Direct Investments (FDI) to build a theoretical framework using the Stochastic Impacts by Regression on Population, Affluence and Technology (STIRPAT) model. We suggest that urbanization, economic growth, FDI attraction, and the different demographics will translate into large energy demand in renewable and non-renewable energy.

A limited number of existing studies analyzed the effects of age-dependency, an indicator of demographic change, on energy demand. At this stage, we aim to use the Kernel Regularized Least Squares (KRLS) machine learning method of Hainmueller and Hazlett (2014). Besides, previous studies show a nonlinear relationship between aging and economic growth (see, e.g., Hirono, 2021). Therefore, there can be a possible nonlinear relationship between aging and energy demand. In short, it is better to use the novel machine learning method of Hainmueller and Hazlett (2014) to model the nonlinearities and instabilities in determinants of renewable and non-renewable energy demand in the BRICS region except Russia.

A higher dependency ratio might lead to less overall energy demand and a higher share of non-renewable sources due to income distribution. It is challenging to decode the complex relations of the environment with population dynamics. High population creates pressure on limited energy sources, migration of population living in the city creates a problem on infrastructure, congestions and high energy demand. Therefore, analyzing the impact of age dependency and increasing urban areas would be important in sustainable energy sources. Moreover, such issues have not been addressed, especially in the catalyst group of emerging economies of BRICS. Further empirical research is required to prove this relationship, and accordingly, policymakers can use more renewable energy sources to meet the additional demand of the urban sector in developing countries.

Furthermore, the increasing population of young and working-age people will generate substantial additional demand for energy. Many studies proved that population and urbanization have an essential role in energy demand. These factors also triggered the demand for non-renewable resources due to low per capita income and difficulty affording clean energy sources. Given these countries' demographic and economic structure, analyzing the population dynamics such as age dependency may have an essential role in guiding the energy policies and renewable energy demand in the right direction. Most studies have taken only one dimension of population, i.e., population size, when analyzing the demographic effect on energy use and the environment. In contrast, other demographic dimensions such as age dependency also have an essential role in analyzing the environmental effect.

To the best of our knowledge, this study is the first research in the empirical literature to examine the effects of age dependency ratio and urbanization on renewable energy and non-renewable energy consumption in the BRICS region by considering the KRLS machine learning method for the panel analysis. We find that the young age dependency, overall age dependency, and urbanization negatively affect renewable and non-renewable energy demand. On the contrary, the old-age dependency, economic growth and FDI are positively associated with renewable and non-renewable energy consumption.

The rest of the paper is organized as follows. *Literature Review* provides a brief review of the studies. In *Model, Estimation Methods, and Data* model, methodology and data have been discussed. *Discussion of the Results and Policy Implications* presents the results, and finally, *Conclusion* concludes the study with the policy implications.

## LITERATURE REVIEW

With increasing energy demand and consumption in emerging economies, many empirical studies (see **Table 1**) have analyzed the link between urbanization, energy consumption, and economic growth in recent decades. Some of these studies have shown a significant positive effect of urbanization on energy consumption (Jones, 1989; Jones, 1991, Parikh and Shukla, 1995; Pachauri and Jiang, 2008; Liu, 2009; Zhao and Zhang, 2018), whereas other researchers found a negative relationship between these two variables (Ghosh and Kanjilal, 2014; Sadorsky, 2014). Few other studies have shown a mixed relationship which varies with the various income group of countries and can be explained by the inverted U shaped relationship (Poumanyvong and Kaneko, 2010; Li and Lin, 2015), where at first urbanization has a positive role in energy consumption after that it shows a negative effect on energy consumption. Precisely, most of the research has found out that urbanization increases energy demand, especially in middle and high-income countries. The reason behind this relationship is that growth causes an increase in urban population density. High levels of population concentration in urban areas are often associated with energy demand in various sectors such as transport mobility (O'Neill and Chen, 2002), industrial production and employment (Jones, 1989; Sadorsky, 2014; Liu et al., 2015), residential energy use and changing lifestyle (Poumanyvong and Kaneko, 2010; Estiri and Zagheni, 2019). These sectors and income are detrimental to the relationship between energy consumption and urbanization in many countries. The recent development and sustainability concern has also started looking at this relationship from the lens of energy sources, broadly segregated into renewable and non-renewable. However, not enough research has been carried out to highlight the impact of urbanization on renewable energy demand.

Similarly, the role of age dependency has been ignored in probing its relationship with renewable and non-renewable energy consumption given the demographic dividend, energy demand, economic growth and concern for sustainability (Sinha et al., 2019) in BRICS countries. The literature review critically analyses the role of urbanization and age dependency in total energy demand and discusses these two on renewable energy

demand. **Table 1** presents a summary of the significant studies in this connection.

The early studies done by Jones (1989, 1991) found that the proportion increase in the population living in an urban area would increase energy consumption per capita. Madlener and Sunak (2011) recommend urban energy planning for better energy use. Salim and Shafiei (2014) and Mrabet et al. (2019) opined that urbanization has a massive effect on non-renewable energy consumption compared to any other factor. Sadorsky (2014) segregated the long-term and short-term effects of urbanization and industrialization on energy demand in emerging economies, including BRICS, except for Russia. The study stated that in the long run, urbanization decreases energy demand, whereas industrialization increases it.

Though numerous studies (Liddle, 2000; Liddle and Lung, 2010; Liu et al., 2015; Sheng et al., 2017) directly probed the connection between the urbanization, population and energy demand using the STIRPAT model in case of various developed and developing countries; the literature is scarce which has analyzed the role of age dependency on energy demand in BRICS economies. This phenomenon can play a significant role in the renewable and non-renewable energy consumption pattern and help achieve sustainable development, given the vast population and demographic dividend in the fastest-growing emerging BRICS economies. Most of the existing literature has taken population size and growth as main drivers and analyzed its role in an Environmental Impact = Population, Affluence and Technology (IPAT) framework to see the demographic impact on energy demand (such as Liddle, 2013; Liddle, 2014). These studies proved that energy demand grows as the population rises. Moreover, in recent years, few researchers (Liddle and Lung, 2010; Hasanov and Mikayilov, 2017) segregated the population into various age groups and analyzed the role of people's working age and young age effect on energy demand.

Notwithstanding, the age dependency aspect has mostly been missed in various studies (see **Table 1**). Because of the immense potential of development and high working-age population, middle-income countries have exerted high pressure on environmental resources and energy demand in manifold ways. For example, Kim and Seo (2012) did a dynamic panel analysis for 53 countries over 35 years and found an inverted U-shaped relationship between aging and energy demand in the long run. This effect is more extensive in residential energy use than industrial energy use because older people use more energy due to health reasons and breakaway in the labor force. Unlike the previous studies, a recent study done by Estiri and Zagheni (2019) has taken the Bayesian Generalized Linear model instead of IPAT or STIRPAT model and found that residential energy consumption increases with life course in the Unites States.

The above studies have not discussed the impact of increasing urbanization and economic growth on renewable energy demand. Few studies (Pachauri and Jiang, 2008; O'Neill et al., 2012) found a strong income effect on household consumption leading to a switch on renewable energy sources such as electricity and natural gas. Salim and Rafiq (2012) found income and pollutant emission as major determinants and have long-term bidirectional causality with renewable energy in Brazil, India, China and a few other emerging economies. They claimed that a 1% increase in GDP causes a 1.228% increase in renewable energy consumption. Yang et al. (2016) have shown that the energy mix effect, i.e., the ratio of renewable energy to total energy consumption, economic growth, and population, significantly affects renewable energy consumption.

In contrast, urbanization has a different role during various growth stages of renewable energy. The authors defined the three stages of renewable energy consumption as slow, fluctuant and accelerated growth stage. Urbanization has the highest contribution (76%) in the accelerated growth stage of renewable energy consumption.

In BRICS countries, a recent study by Banday and Aneja (2020) segregated renewable and non-renewable energy consumption. It revealed a unidirectional causality from non-renewable energy to GDP growth and bidirectional causality from renewable energy to GDP growth. They suggested that all these emerging nations have the potential to develop sustainable energy sources. A similar feedback hypothesis between economic growth and renewable energy has been found by Sebri and Ben-Salha (2014). Zakarya et al. (2015) have explored the connection between energy demand, FDI and economic growth in BRICS countries. They mention that FDI may help invest in renewable energy sources, technology transfer and increasing energy efficiency. The same connection has been found by Doytch and Narayan (2016), who suggested that sectoral FDI reduces non-renewable energy consumption.

In contrast, an augmenting effect of FDI has been found in the case of renewable energy consumption. Amri (2016) found a 1% increase in renewable energy drives FDI by 0.185%; on the other hand, FDI enhances renewable energy by 0.292% in BRICS countries, excluding Russia. The above studies show that urbanization, economic growth, population and FDI may have a catalyst role in the penetration of renewable energy. Therefore, more empirical analyses are required to shed light on the potential and specific ways urbanization and age dependency may affect BRICS countries' renewable energy.

Though diverse literature is available in probing the link between urbanization and energy demand, studies have not segregated urbanization in renewable and non-renewable energy demand. No study rarely talks about the population dynamics and age dependency on energy demand. The role of age dependency and urbanization on renewable and non-renewable energy demand is neglected in the BRICS nations to the best of our knowledge. To fill the existing literature's cavity and provide a comprehensive analysis of the role of age dependency and urbanization while considering renewable and non-renewable energy resources are the primary focus of this research. Given these research gaps, this work differs from existing studies and contributes to the literature in various ways. First, it will analyses urbanization in renewable and non-renewable energy demand, which has not been discussed much in BRICS countries. Second, In the case of demographic characteristics and age structure, most research has been carried out for China or developed countries. However, the age

dependency factor has been neglected. No study rarely segregated this effect from population age distribution, which may give significant insights for making an effective energy policy. Third, it will also analyze other renewable and non-renewable energy demand determinants such as industrialization, FDI, and economic growth within a panel framework.

## Model, Estimation Methods, and Data

### STIRPAT Model

The empirical farmworker draws upon the STIRPAT framework developed) to study the effects of age dependency and urbanization on energy demand of renewable and non-renewable in selected BRICS economies. The STIRPAT model presents unique features. First, it does not impose a prior functional form between variables. Second, the STIRPAT framework avoids the quadratic transformation of non-stationary variables. Finally, the STIRPAT framework model relaxes the underlying assumption that environmental impact population elasticity is united by not scaling environmental and other covariates by population (Fang et al., 2021; Liddle, 2014). Accordingly, we utilize the STIRPAT model, which takes the following form:

$$I_{it} = \alpha P_{it}^{\beta} A_{it}^{\lambda} T_{it}^{\delta} \xi_{it} \quad (1)$$

Where $I_{it}$signifies the renewable and non-renewable energy demand impact, $P_{it}$, $A_{it}$, and $T_{it}$ are population, affluence, and technology, respectively. The parameter a is the constant term, while, $\gamma$, δ, and ξ elasticities of the respective variables that relate to $I_{it}$. Since the objective of this study is to assess the effect of age dependence and urbanization on renewable and non-renewable energy demand via nonparametric methods, we first conduct benchmark exercises by estimating the parametric version of the STIRPAT model. Along with a set of covariates to minimize omitted variables bias, the specification of the same is given below:

$$OIL_{i,t} = \alpha_0 + \beta_1 P_{i,t} + \beta_2 A_{i,t} + \beta_3 T_{i,t} + \varepsilon_{i,t} + \nu_{i,t} + \xi_{i,t} \quad (2)$$

$$COAL_{i,t} = \alpha_1 + \beta_4 P_{i,t} + \beta_5 A_{i,t} + \beta_6 T_{i,t} + \varepsilon_{i,t} + \nu_{i,t} + \xi_{i,t} \quad (3)$$

$$GAS_{i,t} = \alpha_2 + \beta_7 P_{i,t} + \beta_8 A_{i,t} + \beta_9 T_{i,t} + \varepsilon_{i,t} + \nu_{i,t} + \xi_{i,t} \quad (4)$$

$$RE_{it} = \alpha_3 + \beta_{10} P_{i,t} + \beta_{11} A_{i,t} + \beta_{12} T_{i,t} + \varepsilon_{i,t} + \nu_{i,t} + \xi_{i,t} \quad (5)$$

Where *i* represents countries, *t* is the time, $P_{it}$ shows urbanization, young-age, old-age and overall age dependence $A_{it}$is real GDP and $T_{it}$is used for FDI. Further, $OIL_{it}$ it shows that oil consumption, $COAL_{it}$ representing coal consumption,$GAS_{it}$ is the natural gas consumption and $RE_{it}$ represents renewable energy consumption. The parameters $\alpha_0$, $\alpha_1$, $\alpha_2$, and $\alpha_3$ are the constant term and $\beta_1$, $\beta_2$, $\beta_3 \ldots \beta_{12}$ are the slope coefficient concerning independent variables. The term$\varepsilon_{it}$is country-specific fixed-effect, $\nu_{it}$ is the time fixed-effect, and$\xi_{it}$is the random error term.

## Estimation Method: Kernel-Based Regularized Least Squares

The Kernel-based Regularized Least Squares (KRLS) is a nonparametric machine learning method developed by Hainmueller and Hazlett (2014). The KRLS estimation technique has the advantage over classical regression estimators for solving classification and efficiency problems without relying on the linearity assumption. The KRLS estimator is a very flexible estimator to minimize the errors in the estimators. Therefore, this approach can successfully model nonlinearities and instabilities (Hainmueller and Hazlett, 2014). We think capturing the possible presence of nonlinearities and instabilities in the series is essential since renewable energy investments are long-term projects requiring complicated technology. This long-run process can create possible nonlinearities and instabilities in the renewable and non-renewable energy variables, especially in developing economies. This technique allows tackling regression problems without manual specification search while retaining ease-of-use and interpretability and considering the possible nonlinear interaction among variables. The KRLS operates a wider space of possible functions based on the observations with similar expected covariate values to have similar average outcomes.

Furthermore, KRLS employs regularization that gives a prior preference for smoother functions over erratic ones. Indeed, it reduces the variance and fragility of estimates over-fitting and diminishing the influence of "bad leverage" points. This novel method of KRLS has manifold advantages as compared to other methods. First, given its combination of flexibility and interpretability, KRLS can be used for a wide variety of modeling tasks. Second, it is suitable for modeling problems whenever the correct functional form is unknown, including exploratory analysis, model-based causal inference, prediction problems, propensity score estimation, or other regression classification problems. Finally, the KRLS is superior to many other machine learning approaches for both (continuous) regression and (binary) classification tasks.

## Data

The study uses the annual data from 1990 to 2019 for the BRICS economies except for Russia. Therefore, the study is targeting Brazil, India, China and South Africa. Both the annual data and country selection are decided based on the data available for countries. The data brief is explained in **Table 2**. The renewable and non-renewable energy consumption statistics are sourced from the British Petroleum (BP) (2020), Statistical Review of the World Energy Database.[1] The rest of the variables are imported from the World Development Indicators (WDI) of the World Bank (2021).[2]

**Table 3** shows that the average mean is the highest for GDP, which takes a value of 5168.849, followed by coal consumption (379.440), oil consumption (144.357), overall age (54.0844), urbanization (52.945), young age (45.386), gas consumption (27.804), renewable energy consumption (12.438), old age

[1]https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html

[2]https://databank.worldbank.org/data/source/world-development-indicators/preview/on

**TABLE 2 |** Variables used in the study.

| Variables | Description | Unit | Sources |
|---|---|---|---|
| Overall Age | Age dependency ratio (% of working-age population) | % of the working-age population | World Bank (2021) |
| Old Age | The age dependency ratio, old (% of working-age population) | % of the working-age population | |
| Young Age | The age dependency ratio, young (% of working-age population) | % of the working-age population | |
| FDI | Foreign Direct Investment (net inflows) | % of GDP | |
| GDP | Real GDP per capita | Log Constant US$ (2010 prices) | |
| Urbanization | Urban population | % of the total population | |
| OIL | Oil consumption | million ton equlivant oil | British Petroleum (2020) |
| COAL | Coal consumption | million ton equlivant oil | |
| GAS | Natural gas consumption | million ton equlivant oil | |
| Renewable Energy (RE) | Solar energy consumption | million ton equlivant oil | |
| | Wind energy consumption | million ton equlivant oil | |
| | Geothermal and Biomass energy consumption | million ton equlivant oil | |
| | Other renewable energy consumption | million ton equlivant oil | |

*Note: Coal, oil and gas are included in non-renewable energy. Source: British Petroleum (2020) and World Bank (2021).*

**TABLE 3 |** Summary statistics.

| Variable | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|
| GDP | 5168.84 | 3561.815 | 575.501 | 11993.49 |
| Urbanization | 52.945 | 20.645 | 25.547 | 86.569 |
| Young Age | 45.386 | 11.992 | 24.862 | 71.754 |
| Old Age | 8.697 | 1.820 | 6.533 | 15.337 |
| Overall Age | 54.084 | 10.443 | 36.489 | 78.849 |
| FDI | 2.120 | 1.571 | –0.065 | 6.186 |
| OIL | 144.357 | 137.900 | 17.205 | 608.396 |
| GAS | 27.804 | 38.887 | 0.799 | 206.739 |
| COAL | 379.440 | 557.334 | 9.597 | 1969.073 |
| RE | 12.438 | 30.294 | 0 | 213.461 |

*Source: Authors' Works.*

(8.697), and FDI (2.120). In the case of variances, the variance of GDP has the highest value (3561.815) followed by coal consumption (557.334), oil consumption (137.900), gas consumption (38.887), renewable energy consumption (30.294), urbanization (20.645), young age (11.992), overall age (10.443), old age (1.820), and FDI (1.571). The standard deviation for all the variables except economic growth (GDP) and coal is small and predictable. A similar trend is also observed in the appendix of **Supplementary Figure S1**. Since all variables show a stable trend over time, it is fine with moving for the panel estimation for the BRICS region.

# DISCUSSION OF THE RESULTS AND POLICY IMPLICATIONS

## Empirical Results

The lower section of **Table 4** describes the correlation between the dependent and independent variables. Economic growth, old age people, and FDI inflows positively and significantly correlate with non-renewable energy consumption patterns (i.e., coal, oil and natural gas) and renewable energy consumption. In contrast, urbanization, young age and overall age people are negatively and significantly interlinked with it. Moreover, the pattern of non-renewable energy consumption and renewable energy consumption are positively correlated with each other. The low correlation between the independent variables does not produce any threat of multi-collinearity for the estimated models. Interestingly, both positive and negative significant correlations between the independent variables provide us with an expected sign of effects on the BRICS region's primary energy pattern.

**Tables 5**, **6** show the significant results and Hainmueller and Hazlett Kernel-based Regularized Least Squares (2014) models to capture the possible nonlinearity interactions, rich explanation, etc., heterogeneous effects (Hainmueller and Hazlett, 2014). In **Table 5**, oil consumption, gas consumption, coal consumption, and renewable energy consumption are the dependent variables, while economic growth, FDI and urbanization are the independent variables. Moreover, we take urbanization, young age, old age, and overall age dependency as independent variables across all the models.

In Model 1, the results show that economic growth positively impacts renewable energy consumption and non-renewable energy consumption (i.e., oil, coal, gas) while urbanization negatively impacts it. Moreover, in Model 2, while we use a young age in the absence of urbanization, results show that economic growth positively influences and young age negatively affect both renewable and non-renewable energy demand. We also find the mixed effects of foreign direct investment on the pattern of renewable and non-renewable energy.

**Table 6** again reports the Hainmueller and Hazlett Kernel-based Regularized Least Squares (2014) results. In Model 3, while we use overall age in the absence of young and old age, we get consistent findings with Model 2. Moreover, economic growth has positively connected, while overall age dependence has negatively influenced renewable energy consumption and non-renewable energy demand. Simultaneously, in Model 4, all the variables such as economic growth, foreign direct investment, and old-age dependency positively impact it. Overall, we find that the young age dependency, overall age dependency, and urbanization negatively affect renewable- and non-renewable energy demand. On the contrary, old-age dependency and economic growth are positively associated with renewable- and non-renewable energy demand. We also find the mixed effects of foreign direct

**TABLE 4 |** Correlation matrix.

| | GDP | Urbanization | Young age | Old age | Overall age | FDI | OIL | GAS | COAL | RE |
|---|---|---|---|---|---|---|---|---|---|---|
| GDP | 1 | | | | | | | | | |
| Urbanization | 0.886* | 1 | | | | | | | | |
| Young Age | –0.224* | –0.211* | 1 | | | | | | | |
| Old Age | 0.243* | 0.210* | –0.883* | 1 | | | | | | |
| Overall Age | –0.216* | –0.207* | 0.897* | –0.845* | 1 | | | | | |
| FDI | 0.099 | 0.124 | –0.651* | 0.538* | –0.655* | 1 | | | | |
| OIL | 0.147 | –0.164* | –0.757* | 0.834* | –0.728* | 0.381* | 1 | | | |
| GAS | 0.049* | –0.100 | –0.632* | 0.772* | -0.596* | 0.188* | 0.927* | 1 | | |
| COAL | 0.258* | –0.279* | –0.684* | 0.719* | –0.663* | 0.348* | 0.944* | 0.842* | 1 | |
| RE | 0.141* | –0.076* | –0.513* | 0.722* | –0.468* | 0.071* | 0.753* | 0.908* | 0.626* | 1 |

*Source: Authors' Works.*

**TABLE 5 |** Hainmueller and Hazlett (2014) Kernel-based regularized least squares.

| | Model 1 (Urbanization) | | | | Model 2 (Young age) | | | |
|---|---|---|---|---|---|---|---|---|
| | OIL | GAS | COAL | RE | OIL | GAS | COAL | RE |
| GDP | 0.084* [0.011] | 0.030* [0.003] | 0.251* [0.039] | 0.021* [0.003] | 0.007* [0.001] | 0.003* [0.0004] | 0.026* [0.002] | 0.0001 [0.0003] |
| FDI | 23.96* [9.003] | 5.456*** [2.881] | 107.3* [30.74] | 2.839 [2.701] | –0.252 [2.501] | –0.802 [0.983] | –0.011 [6.201] | –3.285** [1.420] |
| Urbanization | –9.692* [1.756] | –3.948* [0.563] | –36.18* [5.984] | –2.970* [0.543] | — | — | — | — |
| Young Age | — | — | — | — | –5.095* [0.237] | –1.304* [0.093] | –23.17* [0.587] | –1.043* [0.158] |

*Note: *p < 0.01, **p < 0.05, ***p < 0.1. [] denotes the robust standard errors. Source: Authors' Works.*

**TABLE 6 |** Hainmueller and Hazlett (2014) kernel-based regularized least squares.

| | Model 3 (Overall age) | | | | Model 4 (Old age) | | | |
|---|---|---|---|---|---|---|---|---|
| | OIL | GAS | COAL | RE | OIL | GAS | COAL | RE |
| GDP | 0.007* [0.001] | 0.003* [0.0004] | 0.026 [0.002] | 0.005 [0.0004] | 0.007* [0.0008] | 0.001* [0.0003] | 0.023* [0.002] | 0.00009 [0.0002] |
| FDI | –1.322 [2.911] | –0.987 [1.125] | –3.304 [7.243] | –3.208*** [1.742] | 3.789*** [1.963] | 1.302*** [0.728] | 8.800 [6.555] | 2.057*** [1.123] |
| Overall Age | –6.795* [0.319] | –1.811* [0.123] | –30.33* [0.784] | –1.444* [0.200] | — | — | — | — |
| Old Age | — | — | — | — | 35.33* [2.661] | 8.477* [0.993] | 133.94* [8.805] | 4.479* [1.143] |

*Note: *p < 0.01, **p < 0.05, ***p < 0.1. [] denotes the robust standard errors. Source: Authors' Works.*

investment inflows on the pattern of primary energy demand for both renewable and non-renewable energy sources.

## Policy Implications

The results show that economic growth is one of the drivers in raising the demand for renewable and non-renewable energy in the BRICS economies. This evidence is in line with previous studies (e.g., Gozgor et al., 2018). This evidence implies that as economic conditions rise or the people's living standard increases due to the income raised from rising employment and inclusive growth process. Therefore, people mainly focus more on non-renewable energy consumption than renewable energy consumption. This evidence can be related to costlier renewable energy consumption and lack of knowledge about clean energy utilization. They also have more comfortable following the traditional pattern of energy consumption.

Similarly, older people demand renewable and non-renewable energy, and their demand for fossil fuels (coal, oil, and natural gas) is higher than renewable energy. This evidence may be related to the fact that old-age people are also comfortable using non-renewable energy, lacking clean energy knowledge. At this stage, clean energy is also expensive for them to choose more fossil fuels in their consumption and production activities.

On the other hand, urbanization, young age and overall age reduce renewable and non-renewable energy consumption. However, old age depends more on renewable and the pattern of non-renewable energy consumption. Simultaneously, foreign direct investment, young age and overall age depend on non-renewable energy consumption. It means that both young and overall are inelastic with foreign direct investment while impacting the renewable and pattern of non-renewable energy consumption. Nevertheless, urbanization and old age dependence are elastic with FDI while using renewable and non-renewable energy consumption.

The study's empirical findings have important policy implications for energy efficiency and quality protection of the natural environment. We find that urbanization, young age, and overall age dependency reduce renewable and non-renewable energy usage in the BRICS economies. It can be argued that growing urbanization has proved to be beneficial now as migrated

people living in urban cities of this region require less energy in their consumption and production activities. Moreover, young people also reduce energy usage in their consumption and production activities because they want to save the electricity bills. Reducing renewable and non-renewable energy consumption can save the households and green environment ships for the selected BRICS economies. The more significant will be households' savings if young people use less energy while driving a car and using the utilities. This issue can improve the natural environment and increase output while not compromising its profit levels. This can be possible if they use energy-saving and product-enhancing technologies because young age people in the selected BRICS economies are more careful about the beneficial effects of increased savings and a green environment in the long run. A similar attitude is also coming from the overall age population while improving energy efficiency. Thus, the findings suggest that environment policymakers in the BRICS economies should prioritize urbanization, young age, and overall age population to improve energy efficiency.

## CONCLUSION

This paper investigated the effects of age dependency ratio (the young age, old-age and overall age) and urbanization on renewable and non-renewable energy consumption (coal, oil and natural gas) in the selected BRICS economies, considering the panel data from 1990 to 2019. We included economic growth and foreign direct investment inflows to build a theoretical framework for the energy demand function using the STIRPAT model. We used the KRLS machine learning technique in the empirical analysis to solve regression and classification problems without relying on the linearity assumption. According to the findings, the young age dependency, overall age dependency, and urbanization negatively impact renewable and non-renewable energy demand. However, old-age dependency and economic growth are positively associated with renewable and non-renewable energy consumption. The mixed effects of FDI inflows on renewable and non-renewable energy are found.

We also find the increasing effects of old age people and economic growth on the BRICS region's renewable and non-renewable energy demand. This evidence implies that people consume more fossil fuels as part of non-renewable energy in driving their consumption activity. They lack the knowledge of operating electricity/using clean energy at their home. Similarly, rising economic growth increases the demand for primary energy (renewable and non-renewable) in the BRICS region, but non-renewable energy usage was higher than renewable energy. It may be possible because people with increased income levels and employment opportunities demand and consume fossil fuels, primarily in the transport sector. This evidence is a sign of fuel inefficiency if people drive more powerful petrol and diesel usage. Fuel inefficiency can also occur if business firms continue to grow more and more output using higher energy amounts. Lastly, foreign direct investment inflows in renewable and non-renewable energy can have both positive and negative effects. On the positive side, foreign investors' inflows can demand more energy if they do not use energy-saving technology in their business activities. In contrast, foreign investors' inflows can also stimulate a green environment using energy-saving technology in the BRICS region. In line with these findings, we can also suggest that climate policymakers should not ignore the importance of old age people, economic growth and FDI in the dynamics of renewable and non-renewable energy demand in the selected BRICS economies as long as renewable energy demand is in association with old age people and economic growth, it is beneficial for the health of the natural environment as it is clean energy in nature.

Nevertheless, higher consumption of non-renewable energy sources is harmful to the natural environment. However, our evidence is limited to the BRICS economies. Future research can use a global data set within a panel framework to generalize findings across all countries and benefit climate policymakers from effective climate mitigation and energy efficiency policies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

ZL: Methodology, Writing-Original Draft Preparation, Supervision MM: Data Curation, Writing-Original Draft Preparation HP: Investigation, Software, Visualization MG: Conceptualization, Writing-Original Draft Preparation GG: Supervision, Writing-Original Draft Preparation.

## FUNDING

The authors acknowledge the financial supports from the Philosophy and Social Science Fund of Tianjin City, China. Project #: TJYJ20-012 ("Prompting the Market Power of Tianjin City's E-commerce Firms in Belt and Road Countries: A Home Market Effect Approach").

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2021.749065/full#supplementary-material

## REFERENCES


Aiyar, S., Ebeke, C., and Shao, X. (2016). *The impact of workforce aging on European productivity*. International Monetary Fund (IMF) Working Paper. No. 16/238. Washington, DC: IMF.

Amri, F. (2016). The relationship amongst energy consumption, foreign direct investment and output in developed and developing countries. *Renew. Sust. Energ. Rev.* 64, 694–702. doi:10.1016/j.rser.2016.06.065

Apergis, N., and Payne, J. E. (2012). Renewable and non-renewable energy consumption-growth nexus: Evidence from a panel error correction model. *Energ. Econ.* 34 (3), 733–738. doi:10.1016/j.eneco.2011.04.007

Apergis, N., and Payne, J. E. (2010). Renewable energy consumption and economic growth: Evidence from a panel of OECD countries. *Energy Policy* 38 (1), 656–660. doi:10.1016/j.enpol.2009.09.002

Banday, U. J., and Aneja, R. (2020). Renewable and non-renewable energy consumption, economic growth and carbon emission in BRICS. *Ijesm* 14 (1), 248–260. doi:10.1108/ijesm-02-2019-0007

Bélaïd, F., and Youssef, M. (2017). Environmental degradation, renewable and non-renewable electricity consumption, and economic growth: Assessing the evidence from Algeria. *Energy Policy* 102, 277–287. doi:10.1016/j.enpol.2016.12.012

British Petroleum (BP) (2020). *Statistical Review of World Energy*. London: BP.

Doytch, N., and Narayan, S. (2016). Does FDI influence renewable energy consumption? an analysis of sectoral FDI impact on renewable and non-renewable industrial energy consumption. *Energ. Econ.* 54, 291–301. doi:10.1016/j.eneco.2015.12.010

Estiri, H., and Zagheni, E. (2019). Age matters: Ageing and household energy demand in the United States. *Energ. Res. Soc. Sci.* 55, 62–70. doi:10.1016/j.erss.2019.05.006

Fang, J., Gozgor, G., Mahalik, M. K., Padhan, H., and Xu, R. (2021). The impact of economic complexity on energy demand in OECD countries. *Environ. Sci. Pollut. Res.* 28, 33771–33780. doi:10.1007/s11356-020-12089-w

Ghosh, S., and Kanjilal, K. (2014). Long-term equilibrium relationship between urbanization, energy consumption and economic activity: Empirical evidence from India. *Energy* 66, 324–331. doi:10.1016/j.energy.2013.12.052

Gozgor, G. (2018a). A new approach to the renewable energy-growth nexus: Evidence from the USA. *Environ. Sci. Pollut. Res.* 25 (17), 16590–16600. doi:10.1007/s11356-018-1858-9

Gozgor, G. (2016). Are shocks to renewable energy consumption permanent or transitory? an empirical investigation for Brazil, China, and India. *Renew. Sust. Energ. Rev.* 66, 913–919. doi:10.1016/j.rser.2016.08.055

Gozgor, G. (2018b). Does the structure of employment affect the external imbalances? Theory and evidence. *Struct. Change Econ. Dyn.* 45, 77–83. doi:10.1016/j.strueco.2018.02.004

Gozgor, G., Lau, C. K. M., and Lu, Z. (2018). Energy consumption and economic growth: New evidence from the OECD countries. *Energy* 153, 27–34. doi:10.1016/j.energy.2018.03.158

Gozgor, G., Mahalik, M. K., Demir, E., and Padhan, H. (2020). The impact of economic globalization on renewable energy in the OECD countries. *Energy Policy* 139, 111365. doi:10.1016/j.enpol.2020.111365

Hainmueller, J., and Hazlett, C. (2014). Kernel regularized least squares: Reducing misspecification bias with a flexible and interpretable machine learning approach. *Polit. Anal.* 22, 143–168. doi:10.1093/pan/mpt019

Hasanov, F. J., and Mikayilov, J. I. (2017). The impact of age groups on consumption of residential electricity in Azerbaijan. *Communist Post-Communist Stud.* 50 (4), 339–351. doi:10.1016/j.postcomstud.2017.09.005

Hirono, M. (2021). Demographic change, human capital accumulation, and sectoral employment. *J. Econ.* 132, 165–185. doi:10.1007/s00712-020-00716-6

Jones, C. (1989). Energy Policies and Strategies in the European Community. *Energ. J.* 10 (4), 29–43. doi:10.1007/978-1-349-10370-6_3

Jones, D. W. (1991). How urbanization affects energy-use in developing countries. *Energy Policy* 19 (7), 621–630. doi:10.1016/0301-4215(91)90094-5

Kim, J., and Seo, B. (2012). Aging in population and energy demand. In Proceedings of the 3rd IAEE Asian Conference, Kyoto, Japan (pp. 20–22).

Li, K., and Lin, B. (2015). Impacts of urbanization and industrialization on energy consumption/CO 2 emissions: Does the level of development matter? *Renew. Sust. Energ. Rev.* 52, 1107–1122. doi:10.1016/j.rser.2015.07.185

Liddle, B. (2014). Impact of population, age structure, and urbanization on carbon emissions/energy consumption: evidence from macro-level, cross-country analyses. *Popul. Environ.* 35 (3), 286–304. doi:10.1007/s11111-013-0198-4

Liddle, B., and Lung, S. (2010). Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. *Popul. Environ.* 31 (5), 317–343. doi:10.1007/s11111-010-0101-5

Liddle, B. (2000). Population growth, age structure, and environmental impact. *Popul. Environ.* 21 (4), 385–411. doi:10.1007/bf02436747

Liddle, B. (2013). Population, affluence, and environmental impact across development: Evidence from panel cointegration modeling. *Environ. Model. Softw.* 40, 255–266. doi:10.1016/j.envsoft.2012.10.002

Liu, Y. (2009). Exploring the relationship between urbanization and energy consumption in China using ARDL (autoregressive distributed lag) and FDM (factor decomposition model). *Energy* 34 (11), 1846–1854. doi:10.1016/j.energy.2009.07.029

Liu, Y., Zhou, Y., and Wu, W. (2015). Assessing the impact of population, income and technology on energy consumption and industrial pollutant emissions in China. *Appl. Energ.* 155, 904–917. doi:10.1016/j.apenergy.2015.06.051

Madlener, R., and Sunak, Y. (2011). Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sust. Cities Soc.* 1 (1), 45–53. doi:10.1016/j.scs.2010.08.006

Mrabet, Z., Alsamara, M., Saleh, A. S., and Anwar, S. (2019). Urbanization and non-renewable energy demand: A comparison of developed and emerging countries. *Energy* 170, 832–839. doi:10.1016/j.energy.2018.12.198

O'Neill, B. C., and Chen, B. S. (2002). Demographic determinants of household energy use in the United States. *Popul. Dev. Rev.* 28, 53–88.

O'Neill, B. C., Ren, X., Jiang, L., and Dalton, M. (2012). The effect of urbanization on energy use in India and China in the iPETS model. *Energ. Econ.* 34, S339–S345. doi:10.1016/j.eneco.2012.04.004

Pachauri, S., and Jiang, L. (2008). The household energy transition in India and China. *Energy Policy* 36 (11), 4022–4035. doi:10.1016/j.enpol.2008.06.016

Parikh, J., and Shukla, V. (1995). Urbanization, energy use and greenhouse effects in economic development. *Glob. Environ. Change* 5 (2), 87–103. doi:10.1016/0959-3780(95)00015-g

Payne, J. E. (2009). On the dynamics of energy consumption and output in the US. *Appl. Energ.* 86 (4), 575–577. doi:10.1016/j.apenergy.2008.07.003

Poumanyvong, P., and Kaneko, S. (2010). Does urbanization lead to less energy use and lower CO2 emissions? A cross-country analysis. *Ecol. Econ.* 70 (2), 434–444. doi:10.1016/j.ecolecon.2010.09.029

Sadorsky, P. (2014). The Effect of Urbanization and Industrialization on Energy Use in Emerging Economies: Implications for Sustainable Development. *Am. J. Econ. Sociol.* 73 (2), 392–409. doi:10.1111/ajes.12072

Salim, R. A., and Rafiq, S. (2012). Why do some emerging economies proactively accelerate the adoption of renewable energy?. *Energ. Econ.* 34 (4), 1051–1057. doi:10.1016/j.eneco.2011.08.015

Salim, R. A., and Shafiei, S. (2014). Urbanization and renewable and non-renewable energy consumption in OECD countries: An empirical analysis. *Econ. Model.* 38, 581–591. doi:10.1016/j.econmod.2014.02.008

Sebri, M., and Ben-Salha, O. (2014). On the causal dynamics between economic growth, renewable energy consumption, CO 2 emissions and trade openness: Fresh evidence from BRICS countries. *Renew. Sust. Energ. Rev.* 39, 14–23. doi:10.1016/j.rser.2014.07.033

Shahbaz, M., and Lean, H. H. (2012). Does financial development increase energy consumption? the role of industrialization and urbanization in Tunisia. *Energy Policy* 40, 473–479. doi:10.1016/j.enpol.2011.10.050

Sheng, P., He, Y., and Guo, X. (2017). The impact of urbanization on energy consumption and efficiency. *Energ. Environ.* 28 (7), 673–686. doi:10.1177/0958305x17723893

Siliverstovs, B., Kholodilin, K. A., and Thiessen, U. (2011). Does aging influence structural change? Evidence from panel data. *Econ. Syst.* 35 (2), 244–260. doi:10.1016/j.ecosys.2010.05.004

Sinha, A., Gupta, M., Shahbaz, M., and Sengupta, T. (2019). Impact of corruption in public sector on environmental quality: Implications for sustainability in BRICS and next 11 countries. *J. Clean. Prod.* 232, 1379–1393. doi:10.1016/j.jclepro.2019.06.066

World Bank (2021). *World Development Indicators*. Washington, DC: World Bank.

Yang, J., Zhang, W., and Zhang, Z. (2016). Impacts of urbanization on renewable energy consumption in China. *J. Clean. Prod.* 114, 443–451. doi:10.1016/j.jclepro.2015.07.158
York, R. (2007). Demographic trends and energy consumption in European Union Nations, 1960-2025. *Soc. Sci. Res.* 36 (3), 855–872. doi:10.1016/j.ssresearch.2006.06.007
Zakarya, G. Y., Mostefa, B., Abbes, S. M., and Seghir, G. M. (2015). Factors Affecting CO2 Emissions in the BRICS Countries: A Panel Data Analysis. *Proced. Econ. Finance* 26, 114–125. doi:10.1016/s2212-5671(15)00890-4
Zhao, P., and Zhang, M. (2018). The impact of urbanisation on energy consumption: A 30-year review in China. *Urban Clim.* 24, 940–953. doi:10.1016/j.uclim.2017.11.005

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Lu, Mahalik, Padhan, Gupta and Gozgor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 15 September 2021
doi: 10.3389/fenvs.2021.728787

# Tracking the Domestic Carbon Emission Intensity of China's Construction Industry: A Global Value Chain Perspective

*Xiaoping Li[1], Yuan Yu[1]*, Xunpeng Shi[2] and Xin Hu[3]*

*[1]School of Economics, Zhongnan University of Economics and Law, Wuhan, China, [2]Australia-China Relations Institute, University of Technology Sydney, Sydney, NSW, Australia, [3]School of Architecture and Built Environment, Deakin University, Geelong, VIC, Australia*

China is the largest producer of carbon in the world. China's construction industry has received widespread attention in recent years due to its environmental issues. However, little research has been conducted to investigate the environmental efficiency of the domestic part of this industry. As the foreign contribution is beyond China's control, identification of domestic carbon emissions is necessary to formulate effective policy interventions. Based on a multi-regional input-output model, this study attempts to reduce the statistical bias associated with international trade, thereby obtaining a more accurate indicator of domestic carbon emission intensity. This study aims to reveal the change in the domestic carbon emission intensity of China's construction industry during 2000–2014 and analyze the reason behind it. The results show that, first, both the constructed intensity indicator and commonly used measures of carbon emission intensity have exhibited a decreasing trend over the study period. However, the former has been consistently larger than the latter. Moreover, this difference first increased and then suddenly decreased after a particular year. Second, although the domestic carbon emission intensity shows a gradually declining trend, it has moved from second to first in global rankings, implying that China's domestic construction industry's carbon emission efficiency, while falling, lags behind other major economies. Third, the structural decomposition results reveal that changes in direct production emission intensity are the leading causes of the decline in domestic carbon emission intensity. In contrast, a change in the intermediate input structure led to an increase in the emission intensity in China's construction industry. In addition, the enormous gaps of domestic carbon emission intensity in the construction industry between China and the selected countries are mainly attributable to the difference in the intermediate input structure. The study suggests that China's construction industry needs to promote high value-added output, optimize intermediate input structure, and improve energy and emission efficiency.

**Keywords: domestic carbon emission intensity, cross-country comparisons, construction, China, global value chain**

OPEN ACCESS

***Edited by:***
*Yongping Sun,*
*Hubei University Of Economics, China*

***Reviewed by:***
*Ye Cao,*
*Beijing Institute of Technology, China*
*Solomon Prince Nathaniel,*
*University of Lagos, Nigeria*

****Correspondence:***
*Yuan Yu*
*sofia_yuyuan@126.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *22 June 2021*
***Accepted:*** *16 August 2021*
***Published:*** *15 September 2021*

***Citation:***
*Li X, Yu Y, Shi X and Hu X (2021) Tracking the Domestic Carbon Emission Intensity of China's Construction Industry: A Global Value Chain Perspective. Front. Environ. Sci. 9:728787. doi: 10.3389/fenvs.2021.728787*

## INTRODUCTION

The global climate crisis necessitates more effective reductions in carbon emissions. China is the largest $CO_2$ emitter in the world and has received increasing attention from the international community in this regard. In response, the Chinese government has specified the goal of reaching peak carbon emissions by 2030and committed to reducing the emission intensity to a value that is 60–65% lower than the 2005 levels by the year 2030 (Chen et al., 2019a). China has recently made an announcement in the UN General Assembly describing its vision to achieve carbon neutrality by 2060 (The Chinese Government, 2020). The attainment of such ambitious climate goals would require the development and implementation of vigorous policies, as well as the strengthening of existing measures.

The construction industry is a typical example of an energy-intensive sector that is particularly becoming a cause of great concern in China. On the one hand, it produces the second-highest amount of carbon, with the record that its embodied carbon input contributed 32% to the increase in total carbon emissions during 1992–2012 (Hou et al., 2020). On the other hand, it represents a backbone industry in China's economic development, accounting for approximately 7% of the gross domestic product (GDP) and creating over 55 million jobs (NBSC, 2019). Given the vast population, massive urbanization, and the ongoing modernization process, the ensuring excessive construction projects would make the emissions mitigation tasks even more challenging (Wu et al., 2018). Therefore, to better balance the emission reduction tasks and output demands, targeted measures should be formulated to lower the carbon emission intensity of China's construction industry.

Accurately measuring the carbon emission intensity of domestic production has been particularly challenging due to the proliferation of the global value chain (GVC), usually characterized by extensive intra-product division across countries. With the deepening of trade, production networks often extend to multiple countries. As Johnson and Noguera (2012) pointed out, intermediate goods transactions could explain about 67% of the global trade volume. In China's construction industry, 27% of the inputs are imported products (Huang et al., 2018). Another noticeable phenomenon is the transfer of pollution across the boundaries (Han et al., 2018). This implies that China's construction industry not only induces value-added and pollution emissions from domestic upstream production but also induces value-added and pollution emissions from abroad through import channels.

Considering that traditional emission intensity does not distinguish between country sources, it is difficult to measure the emissions per unit of domestic value-added, which could easily result in misleading policy design. However, the reconsideration of carbon emission intensity in the context of GVC has not yet been investigated in the previous literature.

The present attempts to fill this gap in research by constructing a new indicator called domestic carbon emission intensity (DCEI), which excludes foreign intermediate inputs and the associated emissions. In particular, this study uses the multi-regional inputoutput (MRIO) model and the structural decomposition analysis (SDA) method to analyze the changes in the DCEI of China's construction industry during 2000–2014. Furthermore, some cross-country comparative analyses of DCEI were carried out to provide a more thorough picture of the Chinese construction industry. The considerable references and policy implications that may follow from the empirical results of this study are expected to enhance the carbon emission efficiency of China's construction industry.

This study extends previous studies in this field from the following three aspects: 1) by separating domestic and foreign contents, this study proposes an analytical framework for measuring domestic carbon emission intensity that is applicable to various sectors; 2) this study uses the most recent data from the World InputOutput Database (WIOD) to measure the DCEI of China's construction industry during 2000–2014; 3) it analyzes the drivers of DECI variation and the gap between China and other global Frontier economies, thus expanding this type of research to 2014.

The remainder of this study is organized as follows. *Literature Review* provides a brief review of the relevant literature. The methodology and data are presented in *Methodology and Data*. *Results and Discussion* presents the results and discussion of the study. The final section presents the main conclusions of the study.

## LITERATURE REVIEW

As reported by Zhou et al. (2020), industry-related changes in China's carbon emission intensities are mostly shaped by the construction industry. A growing body of literature has focused on carbon mitigation in China's construction industry. One noticeable feature of this industry is its intensive industrial linkages, which describe the dependence of different economic sectors on each other's supply and demand (Zhang L. et al., 2019). With the prevalence of GVC, these complicated industrial linkages have broadened to involve various sectors in countries, resulting in misleading traditional measures (Xu et al., 2021). Therefore, it is essential to reevaluate China's construction industry's environmental performance from a more integrated production chain perspective. The literature relevant to this study covers the following two main categories.

### Carbon Emission Evaluation and Decomposition Analysis

Considerable research efforts have been devoted to evaluating the construction industry's carbon emissions with varying research scopes. At the macro level, Huang et al. (2018) compared the construction industry's carbon emissions performance in 40 countries and observed that China is the world's largest contributor to construction carbon emissions. In China, the carbon emissions in the construction sector have been investigated in different ways, such as process-based life cycle assessment (Chen et al., 2017), inputoutput analysis (Zhang Y. et al., 2019), and hybrid analysis (Zhang and Wang, 2016; Chen et al., 2017; Zhang Y. et al., 2019). At the megaregion level, Qi

et al. (2020) estimated the induced pollution emissions in the export of the Yangtze River Delta region through a four-level analytical framework and found that the environmental impact in the construction sector had the largest growth rate and contributed the most to the increase in cross-regional environmental emissions.

In addition, some researchers have identified the factors contributing to carbon emission changes and quantified their contributions. For example, Shi et al. (2017) examined the driving factors of carbon emission changes in China's construction industry during 1995–2009 using the SDA method. The total final demand has been the leading cause of the increase in the emissions while the energy intensity has led to the highest decrease in the emissions associated with China's construction industry. Following a life cycle perspective, Wu et al. (2019) applied the logarithmic mean divisia index (LMDI) method to evaluate the drivers of carbon emissions within China's construction industry during 2000–2015. It was observed that the extraction and manufacturing of raw materials accounted for 58% of the life cycle carbon emissions. Thus, improving energy efficiency as well as lowering emission factors was proposed.

Most existing literature on carbon emission intensity conducted the decomposition analysis, with the index decomposition analysis (IDA) being one of the most intensively used methods (Zhao et al., 2021). However, one major limitation of the IDA approach is that it only considers the driver's direct impact and disregards the indirect impact of the drivers. In contrast, the SDA method addresses this issue by capturing both direct and indirect effects (Hoekstra and Van den Bergh, 2003). As mentioned above, many studies have utilized the SDA method for total carbon emission decomposition analysis. For example, Wang et al. (2017a) and Wang et al. (2017b) explored the changes in carbon emission intensity at national and sectoral levels, respectively, using the SDA method. Dong et al. (2018) combined SDA and quantile regression to analyze the drivers of changes in China's carbon emission intensities. Su and Ang (2020) employed the SDA and attribution analysis methods to investigate Singapore's aggregate carbon intensity at different levels, including the final demand, sector, and transmission levels.

Although the above literature has enriched our understanding of decomposition analysis on carbon intensity by covering many levels, few designs enable both temporal and cross-country comparisons of carbon intensities.

## Value-Added Accounting Perspective

With the prevalence of GVC, the misleading quality of traditional statistics has been verified universally with the expansion of intermediate product trade (Johnson and Noguera, 2012). As mentioned before, intermediate inputs and value-added in a certain stage are exported and counted as the cost of the next stage of production, resulting in double-counting problems and statistical biases. This is especially true for countries located downstream of the GVC, such as China (Koopman et al., 2014). A typical example is the iPhone production case. China, the largest exporter of the iPhone, completes the final step of production as a processing hub. However, China gains only 1.8% of the price of each iPhone. In terms of gross calculations, China's actual economic profits from export trade are far less than those reported using traditional statistics (Kremer et al., 2011).

As Wang et al. (2018) suggested, the deepening GVC system has challenged conventional wisdom on value creation and emissions generation in a given industry in any country. However, studies on carbon emission intensity against the background of GVC are relatively rare (Xiao et al., 2020), particularly in the construction industry. In terms of value measurement, domestic value-added has become a reliable and popular measure of a country's true economic profits (Xu et al., 2021). In the area of pollution generation, the complexity of GVC makes it difficult to attribute responsibility for domestic and foreign emissions, thus exacerbating the difficulty of formulating sound environmental governance policies. More recent literature has shown that the energy consumption or carbon emissions resulting from a country's demand are closely associated with the extent, manner, and location of the importing country's engagement in GVC in addition to its own level of green production (Meng et al., 2018; Wang et al., 2020).

Nevertheless, these limited studies have focused almost exclusively on aggregate indicators rather than intensity or efficiency indicators. Analysis of trends in intensity indicators within the context of GVC has been lacking. One exception is the study conducted by Xiao et al. (2020), which measured the carbon intensities in aggregated demand and further differentiated the demand routes within the context of GVC. However, Xiao et al. (2020) decomposed aggregate emissions instead of intensity indicators for the impact mechanism analysis.

A brief review of the extant literature shows that there are still some gaps in the decomposition analysis of carbon emission intensity from a GVC perspective. As reported by Hung et al. (2019), ongoing vertical specialization and globalization have made the construction industry's supply chain span an increasing number of economies and have accelerated the displacement of environmental impacts. Therefore, it is important to gain a comprehensive understanding of the carbon emission intensity of China's construction industry from the GVC trade perspective.

# METHODOLOGY AND DATA

## Domestic Emissions per Domestic Value-added

As mentioned above, three common approaches are widely used to estimate embodied emissions in the construction field: process-based analysis, hybrid analysis, and inputoutput analysis (Zhang and Wang, 2016). Owing to the distinct research processes and characteristics of these methods, they have been adopted in various studies. Moreover, the construction industry has a strong link with the rest of the economy, particularly its input suppliers. Therefore, the inputoutput approach is more feasible when analyzing both direct and indirect environmental impacts (Acquaye and Duffy, 2010).

The single-region inputoutput (SRIO) model has been commonly adopted to evaluate environmental impacts for quite

**TABLE 1 |** Illustration of inter-regional inputoutput table (two countries and two sectors).

| | | | Intermediate demand | | | | Final demand | | Total output |
|---|---|---|---|---|---|---|---|---|---|
| | | | *Region s* | | *Region r* | | *Region s* | *Region r* | |
| | | | Sector 1 | Sector 2 | Sector 1 | Sector 2 | | | |
| Intermediate input | *Region s* | Sector 1 | – | – | – | – | – | – | – |
| | | Sector 2 | – | – | – | – | – | – | – |
| | *Region r* | Sector 1 | – | – | – | – | – | – | – |
| | | Sector 2 | – | – | – | – | – | – | – |
| Value-added | | | – | – | – | – | – | – | – |
| Pollutant emissions | | | – | – | – | – | – | – | – |
| Total input | | | – | – | – | – | – | – | – |

some time (Nansai et al., 2009; Shan et al., 2017). However, considering the chains of intermediate inputs and final products, the MRIO model would be more effective from a global perspective (Wiedmann, 2009). This method combines bilateral trade data at the sectoral level, which facilitates the identification of relationships between different countries/regions as well as between various sectors in different countries/regions. Therefore, it is a powerful tool for tracking value-added or pollution emissions amidst the surge in international trade (Zhao et al., 2021).

To facilitate intuitive illustration, this section assumes a simplified MRIO model that includes only two countries and two sectors throughout the world (**Table 1**). This assumption and its findings can be easily extended to a real-world case where there are *n* countries and/or regions and *m* sectors. The countries and sectors are denoted using superscripts (*s* and *r*) and subscripts (1 and 2), respectively. Letters in the upper case denote the matrix, and lowercase letters denote an element of the corresponding matrix or some numerical value.

Additionally, this study defines the intermediate input coefficient matrix as $\boldsymbol{A}$, the industry output vector as $\boldsymbol{X}$, the final demand matrix as $\boldsymbol{Y}$, and the pollutant emissions vector as $\boldsymbol{FT}$. Based on the equilibrium attribute of the inputoutput table, the row balance condition holds:

$$\boldsymbol{X} = \boldsymbol{AX} + \boldsymbol{Y} \tag{1}$$

Let the parameter $\boldsymbol{B}$ denotes the Leontief inverse matrix. Then the total output vector can be computed using **Eq. 2** and can be further extended to a blocked matrix form, as shown in **Eq. 3**.

$$\boldsymbol{X} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{Y} = \boldsymbol{BY} \tag{2}$$

$$\begin{bmatrix} x_1^s \\ x_2^s \\ x_1^r \\ x_2^r \end{bmatrix} = \begin{bmatrix} b_{11}^{ss} & b_{12}^{ss} & b_{11}^{sr} & b_{12}^{sr} \\ b_{21}^{ss} & b_{22}^{ss} & b_{21}^{sr} & b_{22}^{sr} \\ b_{11}^{rs} & b_{12}^{rs} & b_{11}^{rr} & b_{12}^{rr} \\ b_{21}^{rs} & b_{22}^{rs} & b_{11}^{rr} & b_{22}^{rr} \end{bmatrix} \begin{bmatrix} y_1^s \\ y_2^s \\ y_1^r \\ y_2^r \end{bmatrix} \tag{3}$$

where $\boldsymbol{V}$ and $\boldsymbol{F}$ represent the value-added coefficient vectors and the emissions coefficient vectors, respectively. Two identities hold for sector $i$ in country $s$ are:

$$v_s^i = va_s^i / x_s^i \tag{4}$$

$$f_s^i = ft_s^i / x_s^i \tag{5}$$

Since the value-added (or pollutant emissions) can be expressed as the product of the industry output and the value-added coefficient (or emissions coefficient), this relationship can be used to quantitatively compute value-added (or pollutant emissions). Taking the value-added as an example, it can be expressed as follows[1]:

$$\boldsymbol{VA} = \boldsymbol{V} \cdot \boldsymbol{X} = \boldsymbol{V} \cdot \left[ (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{Y} \right] = \boldsymbol{V} \cdot (\boldsymbol{BY}) \tag{6}$$

$$\begin{bmatrix} va_1^s \\ va_2^s \\ va_1^r \\ va_2^r \end{bmatrix} = \begin{bmatrix} v_1^s \\ v_2^s \\ v_1^r \\ v_2^r \end{bmatrix} \cdot \begin{bmatrix} x_1^s \\ x_2^s \\ x_1^r \\ x_2^r \end{bmatrix} = \begin{bmatrix} v_1^s \\ v_2^s \\ v_1^r \\ v_2^r \end{bmatrix} \cdot \left( \begin{bmatrix} b_{11}^{ss} & b_{12}^{ss} & b_{11}^{sr} & b_{12}^{sr} \\ b_{21}^{ss} & b_{22}^{ss} & b_{21}^{sr} & b_{22}^{sr} \\ b_{11}^{rs} & b_{12}^{rs} & b_{11}^{rr} & b_{12}^{rr} \\ b_{21}^{rs} & b_{22}^{rs} & b_{11}^{rr} & b_{22}^{rr} \end{bmatrix} \begin{bmatrix} y_1^s \\ y_2^s \\ y_1^r \\ y_2^r \end{bmatrix} \right) \tag{7}$$

To further track the value creation induced by the final demand in different sectors, **Eq. 7** can be translated into a 4 × 4 square form:

$$\hat{\boldsymbol{V}}\boldsymbol{B}\hat{\boldsymbol{Y}} = \begin{bmatrix} v_1^s & 0 & 0 & 0 \\ 0 & v_2^s & 0 & 0 \\ 0 & 0 & v_1^r & 0 \\ 0 & 0 & 0 & v_2^r \end{bmatrix} \begin{bmatrix} b_{11}^{ss} & b_{12}^{ss} & b_{11}^{sr} & b_{12}^{sr} \\ b_{21}^{ss} & b_{22}^{ss} & b_{21}^{sr} & b_{22}^{sr} \\ b_{11}^{rs} & b_{12}^{rs} & b_{11}^{rr} & b_{12}^{rr} \\ b_{21}^{rs} & b_{22}^{rs} & b_{11}^{rr} & b_{22}^{rr} \end{bmatrix} \begin{bmatrix} y_1^s & 0 & 0 & 0 \\ 0 & y_2^s & 0 & 0 \\ 0 & 0 & y_1^r & 0 \\ 0 & 0 & 0 & y_2^r \end{bmatrix}$$

$$= \begin{bmatrix} v_1^s b_{11}^{ss} y_1^s & v_1^s b_{12}^{ss} y_2^s & v_1^s b_{11}^{sr} y_1^r & v_1^s b_{12}^{sr} y_2^r \\ v_2^s b_{21}^{ss} y_1^s & v_2^s b_{22}^{ss} y_2^s & v_2^s b_{21}^{sr} y_1^r & v_2^s b_{22}^{sr} y_2^r \\ v_1^r b_{11}^{rs} y_1^s & v_1^r b_{12}^{rs} y_2^s & v_1^r b_{11}^{rr} y_1^r & v_1^r b_{12}^{rr} y_2^r \\ v_2^r b_{21}^{rs} y_1^s & v_2^r b_{22}^{rs} y_2^s & v_2^r b_{11}^{rr} y_1^r & v_2^r b_{22}^{rr} y_2^r \end{bmatrix} \tag{8}$$

In addition, the value creation in each sector can be derived using the row summation of **Eq. 8**, and the value creation induced by the final demand can be obtained through the column summation of **Eq. 8**. Since final demand induces domestic value creation through the use of domestic inputs and foreign inputs, it accordingly induces foreign value creation as well. To isolate the domestic value-added component of the final product, we take the example of the final demand in sector 1 of country *s*. Assuming that the superscript *u* denotes any given country and the subscript *i* represents any given

[1]The dot product between the two matrices here represents the corresponding elements of the two for multiplication.

sector, the value-added induced by the final demand in sector 1of country $s$ can be expressed as follows.

$$\sum_{u=1}^{2}\sum_{i=1}^{2} v_i^u b_{i1}^{us} y_1^s = \underbrace{v_1^s b_{11}^{ss} y_1^s + v_2^s b_{21}^{ss} y_1^s}_{domestic\ value\ added} + \underbrace{v_1^r b_{11}^{rs} y_1^s + v_2^r b_{21}^{rs} y_1^s}_{foreign\ value\ added}$$
$$= \underbrace{\sum_{i=1}^{2} v_i^s b_{i1}^{ss} y_1^s}_{domestic\ value\ aded} + \underbrace{\sum_{i=1}^{2} v_i^r b_{i1}^{rs} y_1^s}_{foreign\ value\ added} \quad (9)$$

It can be observed that the value in the final demand of sector 1 for country $s$ can be divided into domestic value-added in final demand (DVA) and foreign value-added in final demand (FVA). FVA is the only part created by using foreign intermediate inputs imported from abroad. Thus, the value-added due to the final demand in country $s$ is not all derived from the home country, and only the DVA part represents the pure value from country $s$.

Analogously, the pollutant emissions caused by the demand for final demand can be obtained by substituting the value-added coefficient vector $\boldsymbol{V}$ with the emissions coefficient vector $\boldsymbol{F}$.

$$\hat{\boldsymbol{F}}\boldsymbol{B}\hat{\boldsymbol{Y}} = \begin{bmatrix} f_1^s & 0 & 0 & 0 \\ 0 & f_2^s & 0 & 0 \\ 0 & 0 & f_1^r & 0 \\ 0 & 0 & 0 & f_2^r \end{bmatrix} \begin{bmatrix} b_{11}^{ss} & b_{12}^{ss} & b_{11}^{sr} & b_{12}^{sr} \\ b_{21}^{ss} & b_{22}^{ss} & b_{21}^{sr} & b_{22}^{sr} \\ b_{11}^{rs} & b_{12}^{rs} & b_{11}^{rr} & b_{12}^{rr} \\ b_{21}^{rs} & b_{22}^{rs} & b_{21}^{rr} & b_{22}^{rr} \end{bmatrix} \begin{bmatrix} y_1^s & 0 & 0 & 0 \\ 0 & y_2^s & 0 & 0 \\ 0 & 0 & y_1^r & 0 \\ 0 & 0 & 0 & y_2^r \end{bmatrix}$$
$$= \begin{bmatrix} f_1^s b_{11}^{ss} y_1^s & f_1^s b_{12}^{ss} y_2^s & f_1^s b_{11}^{sr} y_1^r & f_1^s b_{12}^{sr} y_2^r \\ f_2^s b_{21}^{ss} y_1^s & f_2^s b_{22}^{ss} y_2^s & f_2^s b_{21}^{sr} y_1^r & f_2^s b_{22}^{sr} y_2^r \\ f_1^r b_{11}^{rs} y_1^s & f_1^r b_{12}^{rs} y_2^s & f_1^r b_{11}^{rr} y_1^r & f_1^r b_{12}^{rr} y_2^r \\ f_2^r b_{21}^{rs} y_1^s & f_2^r b_{22}^{rs} y_2^s & f_2^r b_{21}^{rr} y_1^r & f_2^r b_{22}^{rr} y_2^r \end{bmatrix} \quad (10)$$

Thus, for sector 1 of country $s$, the pollutant emissions resulting from the demand for final demand can be expressed as follows.

$$\sum_{u=1}^{2}\sum_{i=1}^{2} f_i^u b_{i1}^{us} y_1^s = \underbrace{f_1^s b_{11}^{ss} y_1^s + f_2^s b_{21}^{ss} y_1^s}_{domestic\ CO2\ emissions} + \underbrace{f_1^r b_{11}^{rs} y_1^s + f_2^r b_{21}^{rs} y_1^s}_{foreign\ CO2\ emissions}$$
$$= \underbrace{\sum_{i=1}^{2} f_i^s b_{i1}^{ss} y_1^s}_{domestic\ CO2\ \ emissions} + \underbrace{\sum_{i=1}^{2} f_i^r b_{i1}^{rs} y_1^s}_{foreign\ CO2\ \ emissions} \quad (11)$$

Given that country $s$ only bears the burden of domestic $CO_2$ emissions in the final demand, attention should be paid to the $f_1^s b_{11}^{ss} y_1^s + f_2^s b_{21}^{ss} y_1^s$ part of the equation. Using the ratio of domestic $CO_2$ emissions in final demand to embodied domestic value-added, the carbon emission intensity index DCEI can be expressed as follows:

$$\mathrm{DCEI}_1^s = \sum_{i=1}^{2} f_i^s b_{i1}^{ss} y_1^s / \sum_{i=1}^{2} v_i^s b_{i1}^{ss} y_1^s \quad (12)$$

This implies that the higher the value of $\mathrm{DCEI}_1^s$ is, the more pollutant emissions are generated per unit DVA. Therefore, the DCEI index can be used to measure the domestic carbon emission intensity of sector 1 in country s.

## Structural Decomposition Analysis Method

### Changes in DCEI During Different Periods

Considering that static measurements of DCEI cannot capture the effects of drivers on changes in emission intensity, this study further applies the SDA method to elaborate the inter-temporal changes of sector 1 in country s. To depict this dynamic process, this study defines the indicator $dC_1^s$, which represents the change in DCEI between two periods. When $dC_1^s > 0$, the value of DCEI increases from period $t$–1 to period $t$; whereas when $dC_1^s < 0$, the value of DCEI decreases from period $t$–1 to period $t$.

$$dC_1^s = DCEI_{1,t}^s - DCEI_{1,t-1}^s \quad (13)$$

For the sake of convenience, the four matrices are detailed as follows:

$$\boldsymbol{F}^s = \begin{bmatrix} f_1^s \\ f_1^s \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{V}^s = \begin{bmatrix} v_1^s \\ v_2^s \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{B}^{ss} = \begin{bmatrix} b_{11}^{ss} & b_{12}^{ss} & 0 & 0 \\ b_{21}^{ss} & b_{22}^{ss} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \boldsymbol{Y}_1^s = \begin{bmatrix} y_1^s \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (14)$$

where $\boldsymbol{F}^s$ and $\boldsymbol{V}^s$ are the emission coefficient vector and the value-added coefficient vector of country $s$, respectively. $\boldsymbol{B}^{ss}$ is Leontief's inverse matrix of country $s$, and $\boldsymbol{Y}^s$ is the final demand vector of sector 1 in country $s$. Thus, **Eq. 13** can be expressed as follows.[2]

$$d\widehat{C}_1^s = \frac{\left[(\boldsymbol{F}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right]} - \frac{\left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}{\left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]} =$$
$$\frac{\left[(\boldsymbol{F}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] - \left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]} =$$
$$\underbrace{\frac{\left[(\boldsymbol{F}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] - \left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}}_{d\widehat{F}} +$$
$$\underbrace{\frac{\left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] - \left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}}_{d\widehat{V}} +$$
$$\underbrace{\frac{\left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] - \left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}}_{d\widehat{B}} +$$
$$\underbrace{\frac{\left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t-1}^s\right] - \left[(\boldsymbol{F}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right] \times \left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right]}{\left[(\boldsymbol{V}_t^s)^T \boldsymbol{B}_t^{ss} \boldsymbol{Y}_{1,t}^s\right] \times \left[(\boldsymbol{V}_{t-1}^s)^T \boldsymbol{B}_{t-1}^{ss} \boldsymbol{Y}_{1,t-1}^s\right]}}_{d\widehat{Y}}$$
$$= d\widehat{F} + d\widehat{V} + d\widehat{B} + 0 \quad (15)$$

Consequently, the change in DCEI ($d\widehat{C}_1^s$) can be decomposed into four distinct parts $d\widehat{F}$, $d\widehat{V}$, $d\widehat{B}$, and $d\widehat{Y}$, which reflect the direct production emission intensity effect, the value-added coefficient effect, the intermediate input structure effect, and the final demand structure effect on DCEI, respectively. Because there is no change in the final demand product structure for a certain single sector ($d\widehat{Y} \equiv 0$), the change in DCEI is primarily shaped by $d\widehat{F}$, $d\widehat{V}$, and $d\widehat{B}$.

[2]The superscript T represents the transpose of the matrix. Since the SDA decomposition form is not unique, this paper defines $d\widehat{C}_1^s$ as the first pole decomposition and $d\overline{C}_1^s$ as the second pole decomposition.

To eliminate the limitation that the decomposition results are not unique because of the different ordering of elements in the SDA process, this study uses the geometric average method suggested by Xu and Dietzenbacher (2014) to estimate the size of each part. First, the symmetrical form of **Eq. 15** can be constructed as below.

$$d\overline{C}_1^s = \frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]-\left[(F_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]} =$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]-\left[(F_t^s)^T B_t^{ss} Y_{1,t-1}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]}}_{d\overline{Y}}\times$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t-1}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t}^s\right]-\left[(F_t^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_{t-1}^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]}}_{d\overline{B}}\times$$
$$\underbrace{\frac{\left[(F_t^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_{t-1}^s)^T B_t^{ss} Y_{1,t}^s\right]-\left[(F_t^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]}}_{d\overline{V}}\times$$
$$\underbrace{\frac{\left[(F_t^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]-\left[(F_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_{t-1}^s)^T B_{t-1}^{ss} Y_{1,t-1}^s\right]}}_{d\overline{F}}$$
$$= 0 + d\overline{B} + d\overline{V} + d\overline{F} \tag{16}$$

Second, the geometric average of these two parts is calculated as below.

$$dC_1^s = \left(d\widehat{C}_1^s + d\overline{C}_1^s\right)/2 \tag{17}$$

Finally, the index $dC_1^s$ is converted into the geometric average to accurately analyze the contribution of the four factors to the DCEI changes.

### Changes in Differences in DCEI Between Various Countries

The SDA method can also be used to capture the variance in the DCEI of the two countries for a given period. For example, during period *t*, the difference in DCEI between country *s* and country *r* can be specified in two forms.

$$d\widehat{C}_1^{sr} = DCEI_{1,t}^s - DCEI_{1,t}^r = \frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]-\left[(F_t^r)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y T_{1,t}^r\right]} =$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]-\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\widehat{F}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]-\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\widehat{V}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{ss} Y_{1,t}^s\right]-\left[(F_t^s)^T B_t^{ss} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\widehat{B}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^s\right]-\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\widehat{Y}}$$
$$= d\widehat{F} + d\widehat{V} + d\widehat{B} + 0 \tag{18}$$

$$d\overline{C}_1^{sr} = DCEI_{1,t}^s - DCEI_{1,t}^r = \underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]-\left[(F_t^s)^T B_t^{ss} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\overline{Y}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{ss} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]-\left[(F_t^r)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\overline{B}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{ss} Y_{1,t}^s\right]-\left[(F_t^r)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\overline{V}} +$$
$$\underbrace{\frac{\left[(F_t^s)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^s\right]-\left[(F_t^r)^T B_t^{rr} Y_{1,t}^r\right]\times\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]}{\left[(V_t^s)^T B_t^{ss} Y_{1,t}^s\right]\times\left[(V_t^r)^T B_t^{rr} Y_{1,t}^r\right]}}_{d\overline{F}}$$
$$= 0 + d\overline{B} + d\overline{V} + d\overline{F} \tag{19}$$

Furthermore, taking the geometric average of **Eqs 18**, **19**, this study also facilitates the traceability of the drivers of the difference in DCEI across various countries/regions based on the bipolar decomposition that has been described in **Eq. 17**.

## Data

This study investigates $CO_2$ emissions and value-added induced by the final demand for China's construction industry during 2000–2014. In addition, four developed countries (the United States, Japan, France, and Germany) were selected for further cross-country comparisons. These countries have been among the world's leaders in terms of the final demand scale, as well as their carbon emission efficiency in the construction industry. In particular, the United States, although not very efficient, is included in this analysis on the basis that its final demand scale is merely lower than that of China.

The SRIO model can only explain the total emissions of an economy. It fails to measure the foreign value-added and total emissions caused by the construction industry. Moreover, due to the fact that the data for each country has been published during different years, single-region inputoutput tables cannot be used for consistent cross-country comparisons. With the improvement of inputoutput data, many studies have started to use multi-regional inputoutput tables such as WIOD or Eora to address environmental issues in the context of global value chains. However, the Eora data have a negative final output scenario, contrary to our economic intuition. Therefore, we chose the WIOD database, a more widely used database in the prevailing literature, for our research analysis.

The core database of WIOD is the World InputOutput Tables (WIOT), which includes 28 member countries of the European Union and 15 other important economies, including the United States and China. All other countries are classified into the rest of the world (RoW) region. The 43 economies directly accounted for 90.8 and 85.8% of the world's GDP in 2000 and 2014, respectively. Moreover, their production outputs accounted for 89.9 and 84.1% of the global production output in 2000 and 2014, respectively (WIOD, 2016). Accordingly, the WIOT can effectively depict the patterns of production technology and domestic (or international) trade. These economies accounted for 82.1 and 78.4% of the world's carbon emissions in 2000 and 2014, respectively (WIOD, 2016). Therefore, it is adequate to adopt these WIOT sequences and environmental accounts provided by the WIOD to study the impacts of economic operations on natural resources and the environment in different countries or regions.

# RESULTS AND DISCUSSION

This section first provides an overview of DVA and DCE of China's construction industry over the study period, followed by a comparison between DCEI and total carbon emission intensity

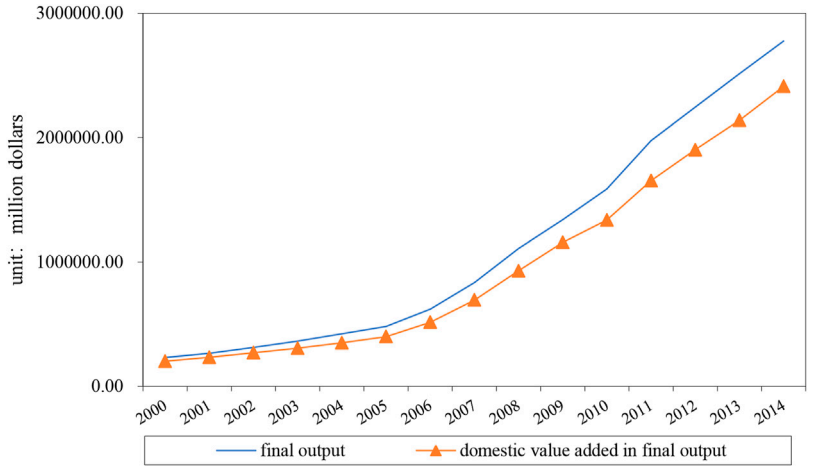



**FIGURE 1 |** The final output and domestic value-added in the final output of China's construction industry during 2000–2014.

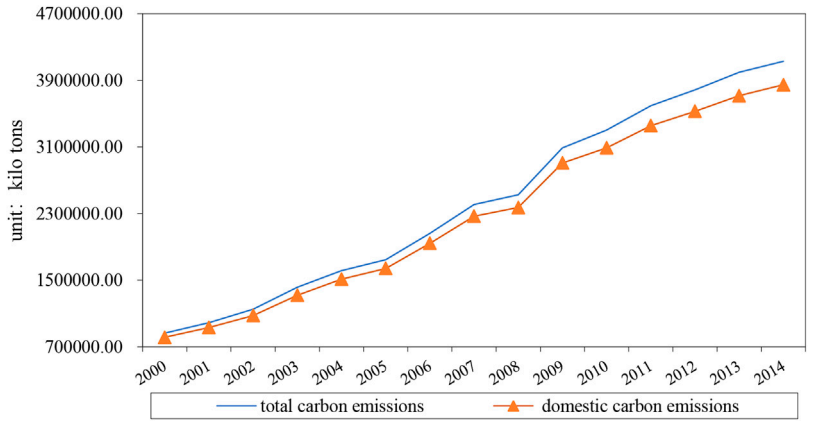



**FIGURE 2 |** The total carbon emissions and the domestic carbon emissions of China's construction industry during 2000–2014.

(TCEI) of this sector. This section then analyzes the factors influencing the changes in China's construction DCEI by utilizing the SDA technique. Furthermore, this section discusses the differences in DCEI between China and other major construction countries (the United States, Japan, France, and Germany) to analyze and bridge the gap between international and domestic carbon efficiency.

## DVA and DCE of China's Construction Industry

By plugging the data into **Eqs 8**, **9**, the final output of China's construction industry and its domestic part during 2000–2014 were measured (**Figure 1**). The total final output of China's construction industry increased from 231.5 billion United States$ in 2000–2775.6 billion United States$ in 2014, with its DVA part increasing from 202.2 billion United States$ in 2000–2410.6 billion United States$ in 2014. Both parameters maintained an increasing trend throughout the study period, with nearly identical growth rates. By plugging the data into **Eqs 10**, **11**, the values of the total induced carbon emissions from China's construction industry and its domestic part during 2000–2014 were calculated (**Figure 2**). The total carbon emissions induced by the final output increased from 864.1 million tons in 2000 to 4.1 billion tons in 2014, with its DCE part increasing from 814.3 million tons in 2000 to 3.8 billion tons in 2014. These two parameters also maintained consistent growth over the entire study period and exhibited similar growth rates.

Combining the two results, it can be observed that regardless of the year, the estimation results under the gross value method were higher than those under the domestic-only accounting method. As mentioned before, this deviation is caused by the ever-increasing decentralization of global production and trade liberalization. Without the capability to distinguish value-added sources, traditional estimation methods tend to overestimate domestic value creation, as

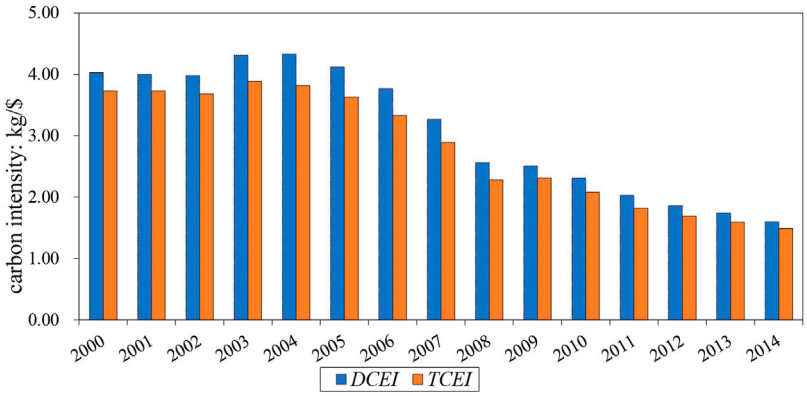



**FIGURE 3 |** DCEI and TCEI of China's construction industry during 2000–2014.

well as domestic-induced carbon emissions. Another interesting point that can be intuitively understood from the graph is that the gross final output overestimates the domestic component to a greater extent than total carbon emissions. This may imply a persistent bias between traditional carbon intensity and pure domestic carbon intensity. Accordingly, the constructed DCEI can serve as an essential basis for depicting China's pure-domestic construction market's economic and environmental performance.

## DCEI of China's Construction Industry

Total carbon emission intensity is a commonly used measure in existing studies, and it is usually defined as the amount of carbon emissions per unit of output. As highlighted earlier, both the emissions and output measurements have become complicated in the context of GVC, and this study attempts to reduce such statistical bias. The indicator DCEI constructed in *Domestic Emissions Per Domestic Value Added* enables us to better understand the true carbon intensity of domestic production in China's construction industry than the traditional indicator TCEI.

**Figure 3** provides an overview of the DCEI and TCEI of China's construction industry during 2000–2014. On the whole, both indicators have exhibited a significant decline, except for a brief rise during 2002–2004 and a short stagnation during 2008–2009. The TCEI fell from 3.7 kg/$ in 2000 to 1.5 kg/$ in 2014, which represents a decrease of 59.5%. At the domestic level, the carbon intensity fell from 4.0 kg/$ in 2000 to 1.6 kg/$ in 2014, which represents a decrease of 60%. Moreover, the value of DCEI has been consistently larger than that of TCEI, suggesting that the domestic carbon emission efficiency of China's construction industry is actually worse than its global environmental performance.

This is in accordance with the conclusions implied by Huang et al. (2018), who showed that China's construction industry imports 27% of its intermediate inputs, which only release 5% of its total carbon emissions. Combined with the fact that a tiny percentage of the industry's final products are exported, the vast majority of the construction carbon emissions would eventually be absorbed domestically. Therefore, the domestic production of China's construction industry exhibits more extensive characteristics. This implies that the TCEI underestimates the carbon intensity of domestic production to some extent and overestimates its environmental efficiency.

On the positive side, however, the gap between these two indicators was observed to be narrowing during the study's later phases. The numerical difference peaked at 0.5 kg/$ in 2005 and then continued to decrease until it reached 0.1 kg/$ in 2014. This may indicate that China's construction industry's domestic carbon emission efficiency has gradually improved recently, considerably bridging the gap with the rest of the world.

**Table 2** reports the final output and DCEI in China's construction industry over the study period. Generally, as mentioned in the previous subsection, the final output in China's construction industry has been expanding, while its DCEI has been diminishing. However, these two parameters have one feature in common: China has consistently ranked the highest in the world in terms of both these parameters. Besides, the global ranking of DCEI surpassed that of the final output. Taken together, China's construction industry has been big in economic output, but not excellent in carbon efficiency.

Furthermore, to better visualize the carbon emission efficiency of China's construction industry, the DCEI of China and the seven major construction countries are depicted in **Figure 4**. Overall, the construction DCEI in all the countries declined over the study period, indicating an overall improvement in clean technology at the global level. In these samples, all developed countries have a lower DCEI than developing countries. France had the lowest DCEI, with a value of 0.1 kg/$ in 2014, which represents only 6.7% of China's construction DCEI. In terms of changes in DCEI, the decrease in DCEI was also greater in developing countries than in

**TABLE 2 |** Final output and DCEI of China's construction industry during 2000–2014.

| | Final output (million dollars) | | DCEI (kilogram/dollar) | |
|---|---|---|---|---|
| | **Absolute value** | **Global ranking** | **Absolute value** | **Global ranking** |
| 2000 | 231,453.29 | 4 | 4.03 | 2 |
| 2001 | 266,023.82 | 4 | 4.00 | 1 |
| 2002 | 312,967.95 | 4 | 3.98 | 1 |
| 2003 | 364,043.14 | 4 | 4.31 | 1 |
| 2004 | 422,895.37 | 4 | 4.33 | 1 |
| 2005 | 480,976.28 | 4 | 4.12 | 1 |
| 2006 | 619,901.15 | 3 | 3.77 | 1 |
| 2007 | 833,746.95 | 3 | 3.27 | 1 |
| 2008 | 1,108,227.34 | 2 | 2.56 | 1 |
| 2009 | 1,338,163.29 | 1 | 2.51 | 1 |
| 2010 | 1,585,449.10 | 1 | 2.31 | 1 |
| 2011 | 1,973,944.19 | 1 | 2.03 | 1 |
| 2012 | 2,242,936.96 | 1 | 1.86 | 1 |
| 2013 | 2,513,099.6 | 1 | 1.74 | 1 |
| 2014 | 2,775,619.87 | 1 | 1.60 | 1 |

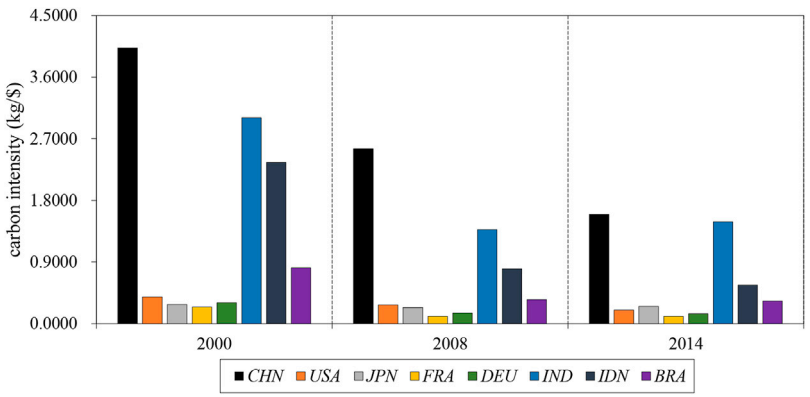



**FIGURE 4 |** DCEI of the world's major construction countries in the years 2000, 2008, and 2014.

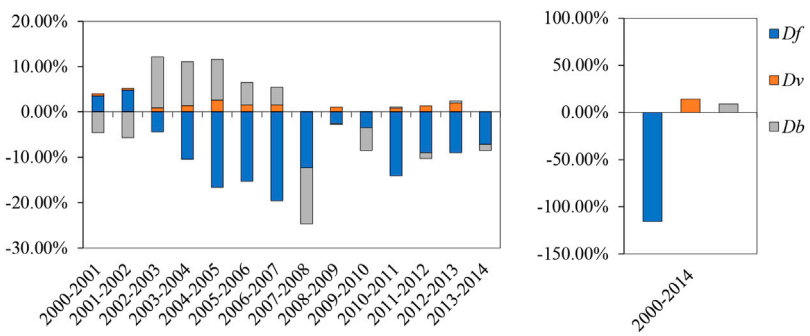



**FIGURE 5 |** SDA results of changes in DCEI of China's construction industry during 2000–2014.

developed countries. In fact, China's construction industry has made the most progress in carbon efficiency, reducing DCEI by 2.4 kg/$. Despite this remarkable improvement, it has still maintained its top ranking throughout the world. Therefore, the green development of China's construction industry is not optimal.

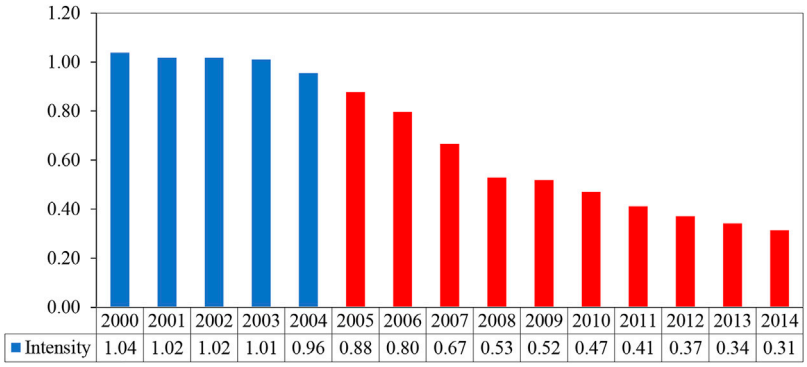



**FIGURE 6 |** Weighted direct production emission intensity of China's domestic industry during 2000–2014.

## Decomposition of DCEI of China's Construction Industry

As discussed above, the DCEI of China's construction industry has been ranked first globally, indicating significant room for the improvement of carbon efficiency through targeted measures. Based on **Eqs 13–17**, the changes in DCEI during the entire study period and subperiods can be decomposed into the effects of three factors: the direct production emission intensity, the value-added coefficient, and the intermediate input structure, as displayed in **Figure 5**.

Throughout the study period, the direct production emission intensity effect contributed the most to the decrease in the DCEI of China's construction industry, driving a drop of approximately 115.6%. High direct production emission intensity adversely alters the environment, meaning that this factor's decrease promotes a reduction in DCEI. According to **Figure 5**, the effect of direct production emission intensity on the decline of DCEI was negative before 2002 and positive thereafter. This phenomenon is possibly closely related to a sharp improvement in cleanliness in many industries in China. With the implementation of environmental regulations such as the Law of the People's Republic of China on Promoting Cleaner Production in 2003 (Hicks and Dietmar, 2007), the emission intensities of the sectors located upstream of the construction industry have decreased significantly.

Additionally, the construction industry is typically in great need of intermediate inputs from domestic upstream industries. For example, the construction industry accounts for approximately 70% of cement products and 25% of steel products in China (Shi et al., 2017). Consequently, through the transmission along with the production chains, the domestic emissions caused by one unit of final output would reduce drastically. Consequently, the domestic emissions caused by one unit of final output reduce drastically via transmission along the production chains.

In fact, many initiatives have shown that China has endeavored to improve energy efficiency to save energy and reduce emissions. For example, the Chinese government set the goal of a 20% reduction in national energy intensity against the 2006 levels in the 11th Five-Year Plan (Liu et al., 2018). From **Figure 6**, it is evident that the direct production emission intensity of all sectors maintained a downward trend, especially during the 11th Five-Year Plan period. Such a recent decline shows the effective management of energy efficiency, which should be sustained. For the construction industry itself, China implemented a national energy-efficient design standard for commercial buildings by setting the goal of reducing energy use by up to 50% compared to preexisting buildings. In addition, for large-scale public buildings, a range of targeted regulations concerning energy statistics, energy consumption information disclosure, and energy monitoring systems have been developed and implemented since 2007.

The impact of the intermediate input structure was slightly more complicated. Overall, this factor contributed to a 9.0% increase in DCEI. Although this value was much smaller than that of the direct production emission intensity, it did not mean that it had a weak influence on the carbon intensity. It can be observed from **Figure 5** that during each subperiod, the impact of the intermediate input structure was slightly weaker than that of the direct production emission factor but much stronger than that of the value-added rate factor. In fact, the relatively low total effect arose from the offset of the different effects at different subperiods. Basically, the intermediate input structure effect contributed negatively towards the decrease in DCEI from 2002 to 2007; however, its contribution was positive during other time periods. That implies that it played a significant role in lowering the carbon efficiency of China's construction industry in the period after China's accession to the WTO and before the 2008 subprime mortgage crisis.

From the perspective of historical development, the overall competitiveness of China's construction industry was relatively inadequate during 2002–2007. In the face of fierce competition brought about by China's opening up, the construction sector has exhibited characteristics of high energy consumption and low value-added acquisition capacity along the production chain, eventually resulting in high carbon emission intensity. However, after the subprime mortgage crisis, with the decay of foreign competitors and the implementation of the four trillion

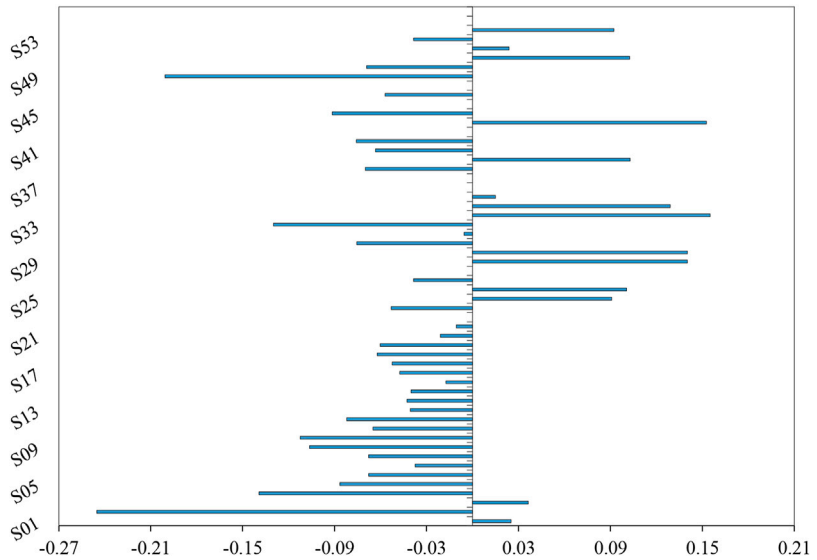



**FIGURE 7 |** The change in the value-added coefficients of China's domestic industries during 2000–2014.

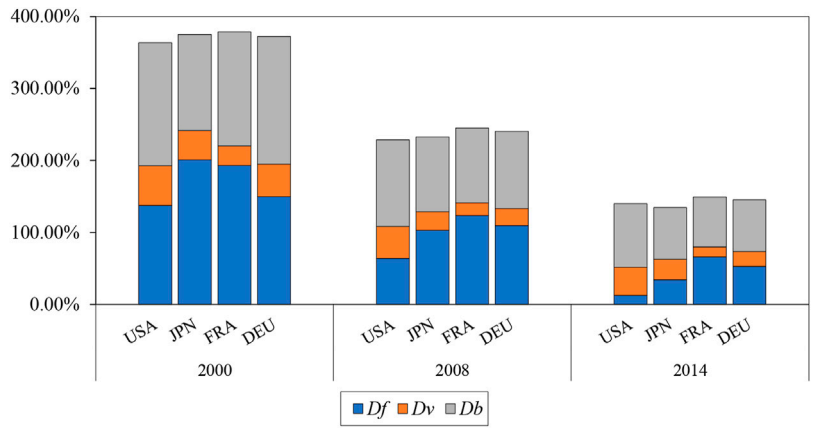



**FIGURE 8 |** SDA results of DCEI differences between China and selected countries in the years 2000, 2008, and 2014.

stimulus plan, the Chinese construction industry, which had become increasingly competitive, switched to the development mode of low energy consumption and high profitability acquisition. Therefore, subject to contrasting effects at each stage, the overall impact is relatively weak.

The role of the value-added coefficient is of particular note, as it adversely affects the reduction of DCEI in China's construction industry. As shown in **Figure 5**, the change in the value-added rate resulted in a total increase of 14.0% in DCEI relative to the base period. The explanation for this may lie in the following points: On the one hand, according to the inputoutput theory, the input of one industry is mainly composed of intermediate input and primary input (i.e., value-added). With China's opening up to the outside world, a large influx of foreign capital has intensified competition between domestic industries. To fulfill the demands of discerning customers and improve the quality of goods and services, the proportion of foreign intermediate inputs has continued to increase during the domestic industrial development process, which has squeezed the primary input (value-added) space (Kim and Xin, 2021). Accordingly, there has been a substantial reduction in the value-added coefficients of most domestic sectors.

As shown in **Figure 7**, only 13 sectors' value-added coefficients achieved a positive shift out of 56 domestic departments. On the other hand, with the decline in value-added coefficient of the domestic upstream sector, the extent of domestic value-added by one unit towards the final output of the construction sector was bound to decline. Therefore, specific policies are required to prevent further decline in the domestic value-added coefficient.

## Comparison in Construction DCEI of China and Selected Countries

Based on **Eqs 18**, **19**, this section employs the SDA approach to further identify the drivers for the large gaps in the DCEI of the construction industry between China and other major construction countries. The results are presented in **Figure 8**.

The decomposition results of cross-country comparisons show that the construction intermediate input structure was the primary cause of the variation in DCEI between China and the selected countries. On average, this factor index could explain 43.0, 46.0, and 52.9% of the inter-country gaps in DCEI in 2000, 2008, and 2014, respectively.

One critical reason is that there has been a much greater demand for new buildings and infrastructure projects in China due to the ongoing modernization and urbanization wave. These new construction activities would inevitably induce massive upstream carbon emissions through the increasing demand for basic raw materials, which are typically carbon-intensive (Huang et al., 2018). In contrast, developed economies generally focus more on maintenance and repair (M&R) activities in the construction industry, which are relatively low carbon and high value-added. For example, Chen et al. (2019b) observed that the United States construction industry absorbs intermediate inputs mostly from repairing works such as the renting of M&R activities. However, China will maintain this considerable demand for new and sophisticated buildings for the foreseeable future. For instance, as a multitrillion-dollar infrastructure program, China's BeltRoad Initiative will definitely trigger more demand for new construction activities.

In addition, China's urbanization rate reached 60.3% in 2019, which is still much smaller than 82.5 and 91.7% in the United States and Japan, respectively (World Bank, 2020). Therefore, it might not be practical to reduce the negative impact of intermediate input structures on DCEI gaps by limiting the demand for new construction activities (Zhang et al., 2021). Given that upstream suppliers play a tremendously important role, it will be useful to improve the environmental performance of the construction industry by promoting low-carbon production technology within upstream sectors and optimizing the industrial structure. Regarding the construction industry itself, innovative construction technologies, such as off-site construction and low-carbon building materials such as wood, should be given higher priority.

The second most influential factor was the difference in direct production emission intensities. On average, it accounted for 45.7, 42.3, and 29.2% of the international gaps in the DCEI in 2000, 2008, and 2014, respectively. According to Peng et al. (2016), China's carbon emissions intensities were estimated to be 2.4 times greater than the world's average level in 2007, primarily due to its coal-dominated energy consumption pattern. In 2014, coal consumption constituted 66.0% of China's primary energy consumption, while for the United States, Japan, France, and Germany, it accounted for 19.7, 27.7, 3.8, and 24.9%, respectively (BP, 2015). Although the share of coal use in China fell to 58.2% in 2018 (BP, 2019), the effect of such a slight decline on environmental efficiency would be quite limited. Coal causes more pollution emissions under the same energy demand, given that coal has a higher pollutant emission factor than other fossil fuels such as oil and natural gas (Jiang et al., 2019).

To optimize the energy structure, the Chinese government imposed a cap on coal consumption at 4.2 billion tons before 2020, and at the same time, the use of nonfossil fuels was set to account for 15% (Wu et al., 2018). However, compared with developed countries, clean energy still accounts for a small proportion of total energy in China, making it difficult to completely replace coal when meeting the growing demand for energy consumption in the short term. To control coal consumption without negatively affecting economic development, policy-makers should take active measures, such as reducing the coal share in the energy mix, encouraging the use of clean energy sources, and improving energy efficiency.

Another factor was the difference in value-added coefficients, which could only explain 11.3, 11.8, and 18.0% of these regional DCEI gaps in 2000, 2008, and 2014, respectively. The low value-added rate in China's construction industry has been expanding the international gap in environmental efficiency, reinforcing the above finding. Policy guidance is needed to close the gap in value-added rates between China and advanced economies. For example, although outsourcing and importing intermediate inputs are conducive to the national economy through the "learning by doing" effect, encouraging deep processing after import trade is equally essential. In addition, enhancing the labor force's skill training and promoting technological innovation will help to upgrade China's capacity for supplying high-value-added products.

## Comparisons of Construction Intermediate Input Structure Across Countries

As discussed in the previous subsection, the intermediate input structure was the most critical factor for China's higher DCEI than other economies, and its effect has grown over time. This subsection comparatively examines the construction intermediate input structures of China and the selected countries. **Figure 9** depicts the distinct differences in the construction intermediate input structure between China and the four selected countries. Compared with other countries, China's construction industry relies more on primary products, manufacturing, electricity, gas, and water collection industries, but less on the service industry when absorbing intermediate inputs.

Consider the difference between China and the United States as an example. In 2000, the proportions of China's construction intermediate inputs transformed from the primary product, manufacturing, electricity, gas, and water collection sectors, and the service industry in China was 3.9% higher, 7.5% higher, 3.0% higher, and 14.1% lower than those of the United States, respectively. These gaps have widened. In 2014, the ratios of construction intermediate inputs converted from the primary product, manufacturing, electricity, gas, and water collection sectors, and the service industry in China was 3.6%

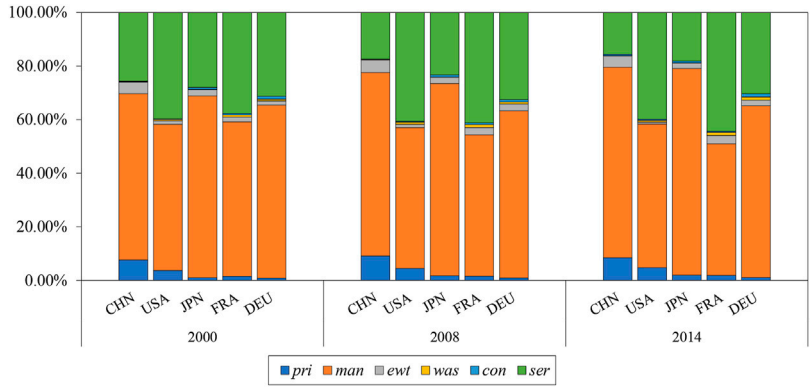



**FIGURE 9 |** Differences in the construction intermediate input structure between China and selected countries in the years 2000, 2008, and 2014. Note: pri represents the primary product industry including four sectors (S01-S04); man represents the manufacturing industry including 19 sectors (S05-S23); ewt represents electricity, gas, and water collection sectors (S24-S25); was represents the waste collection sector (S26); con represents construction industry (S27); ser represents the service industry including 29 sectors (S28-S56).

higher, 17.7% higher, 3.4% higher, and 24.2% lower than those in the United States

From the perspective of industry characteristics, in general, the service industry has the advantages of higher value-added rate and lower pollution emissions compared with the primary product, manufacturing, electricity, gas, and water collection sectors. Consequently, the development mode of construction intermediate input structures has seriously hindered the reduction of its DCEI in China. More emphasis should be placed on transforming and optimizing the construction intermediate input structure to narrow the specific gap with other economies. In addition, as a highly profitable and environmentally friendly industry, the service industry has great potential for improving the domestic environmental performance of China's construction industry. Therefore, there is an urgent need for China's construction industry to raise the share of input from the service sector by utilizing more research and development (R&D) products and brand operation services, rather than high pollution production links.

## CONCLUSIONS AND POLICY IMPLICATIONS

Given the tremendous amount of carbon emissions from the construction industry, reducing the carbon intensity is imperative to achieve China's ambitious carbon mitigation targets. In the context of GVC, it is easy to disambiguate the value-added source and destination of carbon emissions when studying the carbon intensity of an industry in a country. This may create misleading information for policy interventions, which can only be effective for domestic production.

In this study, we rigorously extracted the domestic value-added and domestic carbon emission components from the total carbon emissions. Furthermore, using the SDA technique, the changes in DCEI during different periods, as well as the changes in differences of DCEI across various countries, can be decomposed into the effects of three influencing factors: direct production emission intensity, value-added coefficient, and intermediate input structure.

Several key findings are provided. First, for China's construction industry, the commonly used TCEI and the constructed DCEI have been declining over the entire research period. Nevertheless, the values of TCEI have always been smaller than those of DCEI, revealing that China's domestic construction production has worse emission performance compared to its international suppliers.

Second, although the DCEI of China's construction industry has decreased significantly, it has moved from second to first in terms of global rankings. That is, the carbon efficiency of China's construction industry, while improving, lags behind other major economies of the world. Therefore, China's construction industry shows the typical characteristic of being large in economic outputs, but relatively poor in environmental efficiency.

Third, the cross-year decomposition analysis reveals that changes in direct production emission intensity (principally from upstream sectors) and intermediate input structure are the key factors for reducing DCEI in China's construction industry. Specifically, the former has a promoting effect, while the latter has an inhibitory effect. Finally, compared with other major construction countries, China's poor environmental performance is chiefly attributable to its extensive mode of intermediate input structure, which is too biased towards high pollution and low value-added products.

In light of the above findings, some useful recommendations for policy-makers are provided. First, emphasis should be placed on optimizing the construction intermediate input structure. This can be achieved through two policy instruments. One involves promoting innovative and low-carbon building structures. With the massive demand for new and complex buildings in China, there are many opportunities to promote innovative building structures. For example, compared with the conventional brick concrete structure or shear-wall structure, the steel structure, encouraged in recent years, consumes less cement while meeting

the needs of high-rise, long-span buildings. The other instrument involves promoting the development and utilization of low-carbon and recyclable building materials. The intermediate input structure transformation could also be strengthened by reducing the dependence on high energy consumption and carbon-intensive products and enhancing construction services.

Second, there is an urgent need for the construction industry and its upstream sectors to employ clean production technologies. From the perspective of China's energy consumption structure, limiting coal consumption, improving energy efficiency, and encouraging sustainable energy sources should be prioritized. In addition, to promote clean innovation and technological development, the government should provide sufficient support for capital investment, such as financial subsidies and tax reductions, to relevant enterprises.

Finally, raising the value-added coefficients of the involved industries is also an important part of the efforts directed towards the mitigation of emissions. In other words, the construction industry should upgrade the capacity to provide high value-added products. To fulfill this purpose, the government should improve the talent training mechanism, enabling lower-level workers to improve vocational skills. Additionally, policies should strengthen intellectual property protection, encourage the transformation of R&D results into products, and promote cooperation between industries and research institutes.

Due to the constraints regarding data availability, the main limitation of this study is that the data used were confined between 2000 and 2014. Therefore, future studies should work on strengthening the research effort by revising these data-related issues. In addition, this study can be extended to analyze other pollutant emissions or other important industries in various countries.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

XL and YY: writing original draft preparation, conceptualization, methodology, software. XS: writing review and editing. YY and XH: validation. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors gratefully acknowledge the financial support provided by the Major Project of National Social Science Foundation of China (grant number 18ZDA038).

## REFERENCES

Acquaye, A. A., and Duffy, A. P. (2010). Input-output Analysis of Irish Construction Sector Greenhouse Gas Emissions. *Building Environ.* 45, 784–791. doi:10.1016/j.buildenv.2009.08.022

BP (2015). *BP Statistical Review of World Energy*.

BP (2019). *BP Statistical Review of World Energy*.

Chen, J., Shen, L., Shi, Q., Hong, J., and Ochoa, J. J. (2019a). The Effect of Production Structure on the Total CO2 Emissions Intensity in the Chinese Construction Industry. *J. Clean. Prod.* 213, 1087–1095. doi:10.1016/j.jclepro.2018.12.203

Chen, J., Shen, L., Song, X., Shi, Q., and Li, S. (2017). An Empirical Study on the CO2 Emissions in the Chinese Construction Industry. *J. Clean. Prod.* 168, 645–654. doi:10.1016/j.jclepro.2017.09.072

Chen, J., Shi, Q., Shen, L., Huang, Y., and Wu, Y. (2019b). What Makes the Difference in Construction Carbon Emissions between China and USA? *Sust. Cities Soc.* 44, 604–613. doi:10.1016/j.scs.2018.10.017

Dong, F., Yu, B., Hadachin, T., Dai, Y., Wang, Y., Zhang, S., et al. (2018). Drivers of Carbon Emission Intensity Change in China. *Resour. Conservation Recycling* 129, 187–201. doi:10.1016/j.resconrec.2017.10.035

Han, L., Han, B., Shi, X., Su, B., Lv, X., and Lei, X. (2018). Energy Efficiency Convergence across Countries in the Context of China's Belt and Road Initiative. *Appl. Energ.* 213, 112–122. doi:10.1016/j.apenergy.2018.01.030

Hicks, C., and Dietmar, R. (2007). Improving Cleaner Production through the Application of Environmental Management Tools in China. *J. Clean. Prod.* 15, 395–408. doi:10.1016/j.jclepro.2005.11.008

Hou, H. M., Bai, H. T., Ji, Y. J., Wang, Y., and Xu, H. (2020). A Historical Time Series for Inter-industrial Embodied Carbon Transfers within China. *J. Clean. Prod.* 264, 121738. doi:10.1016/j.jclepro.2020.121738

Huang, L., Krigsvoll, G., Johansen, F., Liu, Y., and Zhang, X. (2018). Carbon Emission of Global Construction Sector. *Renew. Sust. Energ. Rev.* 81, 1906–1916. doi:10.1016/j.rser.2017.06.001

Hung, C. C. W., Hsu, S.-C., and Cheng, K.-L. (2019). Quantifying City-Scale Carbon Emissions of the Construction Sector Based on Multi-Regional Input-Output Analysis. *Resour. Conservation Recycling* 149, 75–85. doi:10.1016/j.resconrec.2019.05.013

Jiang, T., Huang, S., and Yang, J. (2019). Structural Carbon Emissions from Industry and Energy Systems in China: An Input-Output Analysis. *J. Clean. Prod.* 240, 118116. doi:10.1016/j.jclepro.2019.118116

Johnson, R. C., and Noguera, G. (2012). Accounting for Intermediates: Production Sharing and Trade in Value Added. *J. Int. Econ.* 86, 224–236. doi:10.1016/j.jinteco.2011.10.003

Kim, M., and Xin, D. (2021). Export Spillover from Foreign Direct Investment in China during Pre- and post-WTO Accession. *J. Asian Econ.* 75, 101337. doi:10.1016/j.asieco.2021.101337

Koopman, R., Wang, Z., and Wei, S.-J. (2014). Tracing Value-Added and Double Counting in Gross Exports. *Am. Econ. Rev.* 104, 459–494. doi:10.1257/aer.104.2.459

Liu, B., Wang, D., Xu, Y., Liu, C., and Luther, M. (2018). A Multi-Regional Input-Output Analysis of Energy Embodied in International Trade of Construction Goods and Services. *J. Clean. Prod.* 201, 439–451. doi:10.1016/j.jclepro.2018.08.029

Meng, B., Peters, G. P., Wang, Z., and Li, M. (2018). Tracing CO2 Emissions in Global Value Chains. *Energ. Econ.* 73, 24–42. doi:10.1016/j.eneco.2018.05.013

Nansai, K., Kagawa, S., Suh, S., Fujii, M., Inaba, R., and Hashimoto, S. (2009). Material and Energy Dependence of Services and its Implications for Climate Change. *Environ. Sci. Technol.* 43, 4241–4246. doi:10.1021/es8025775

Peng, S., Zhang, W., and Sun, C. (2016). 'Environmental Load Displacement' from the North to the South: A Consumption-Based Perspective with a Focus on China. *Ecol. Econ.* 128, 147–158. doi:10.1016/j.ecolecon.2016.04.020

Qi, X., Mao, X., Huang, X., Wang, D., Zhao, H., and Yang, H. (2020). Tracing the Sources of Air Pollutant Emissions Embodied in Exports in the Yangtze River Delta, China: A Four-Level Perspective. *J. Clean. Prod.* 254, 120155. doi:10.1016/j.jclepro.2020.120155

Shan, Y., Guan, D., Liu, J., Mi, Z., Liu, Z., Liu, J., et al. (2017). Methodology and Applications of City Level CO 2 Emission Accounts in China. *J. Clean. Prod.* 161, 1215–1225. doi:10.1016/j.jclepro.2017.06.075

Shi, Q., Chen, J., and Shen, L. (2017). Driving Factors of the Changes in the Carbon Emissions in the Chinese Construction Industry. *J. Clean. Prod.* 166, 615–627. doi:10.1016/j.jclepro.2017.08.056

The Chinese Government (2020). *China Sets Tough Targets for Reducing CO2 Emissions.*

Wang, H., Ang, B. W., and Su, B. (2017a). A Multi-Region Structural Decomposition Analysis of Global CO 2 Emission Intensity. *Ecol. Econ.* 142, 163–176. doi:10.1016/j.ecolecon.2017.06.023

Wang, H., Ang, B. W., and Su, B. (2017b). Multiplicative Structural Decomposition Analysis of Energy and Emission Intensities: Some Methodological Issues. *Energy* 123, 47–63. doi:10.1016/j.energy.2017.01.141

Wang, L., Yue, Y., Xie, R., and Wang, S. (2020). How Global Value Chain Participation Affects China's Energy Intensity. *J. Environ. Manage.* 260, 110041. doi:10.1016/j.jenvman.2019.110041

Wang, Q., Hang, Y., Su, B., and Zhou, P. (2018). Contributions to Sector-Level Carbon Intensity Change: An Integrated Decomposition Analysis. *Energ. Econ.* 70, 12–25. doi:10.1016/j.eneco.2017.12.014

Wiedmann, T. (2009). A Review of Recent Multi-Region Input-Output Models Used for Consumption-Based Emission and Resource Accounting. *Ecol. Econ.* 69, 211–222. doi:10.1016/j.ecolecon.2009.08.026

WIOD (2016). *World Input-Output Database*.

World Bank (2020). *Urban Population (% of Total Population)*.

Wu, P., Song, Y., Zhu, J., and Chang, R. (2019). Analyzing the Influence Factors of the Carbon Emissions from China's Building and Construction Industry from 2000 to 2015. *J. Clean. Prod.* 221, 552–566. doi:10.1016/j.jclepro.2019.02.200

Wu, Y., Chau, K. W., Lu, W., Shen, L., Shuai, C., and Chen, J. (2018). Decoupling Relationship between Economic Output and Carbon Emission in the Chinese Construction Industry. *Environ. Impact Assess. Rev.* 71, 60–69. doi:10.1016/j.eiar.2018.04.001

Xiao, H., Sun, K., Tu, X., Bi, H., and Wen, M. (2020). Diversified Carbon Intensity under Global Value Chains: A Measurement and Decomposition Analysis. *J. Environ. Manage.* 272, 111076. doi:10.1016/j.jenvman.2020.111076

Xu, X., Wang, Q., Ran, C., and Mu, M. (2021). Is burden Responsibility More Effective? A Value-Added Method for Tracing Worldwide Carbon Emissions. *Ecol. Econ.* 1, 181.

Xu, Y., and Dietzenbacher, E. (2014). A Structural Decomposition Analysis of the Emissions Embodied in Trade. *Ecol. Econ.* 101, 10–20. doi:10.1016/j.ecolecon.2014.02.015

Zhang, L., Liu, B., Du, J., Liu, C., and Wang, S. (2019a). CO2 Emission Linkage Analysis in Global Construction Sectors: Alarming Trends from 1995 to 2009 and Possible Repercussions. *J. Clean. Prod.* 221, 863–877. doi:10.1016/j.jclepro.2019.02.231

Zhang, S., Li, Z., Ning, X., and Li, L. (2021). Gauging the Impacts of Urbanization on CO2 Emissions from the Construction Industry: Evidence from China. *J. Environ. Manage.* 288, 112440. doi:10.1016/j.jenvman.2021.112440

Zhang, Y., Yan, D., Hu, S., and Guo, S. Y. (2019b). Modelling of Energy Consumption and Carbon Emission from the Building Construction Sector in China, a Process-Based LCA Approach. *Energy Policy* 134, 110949. doi:10.1016/j.enpol.2019.110949

Zhang, Z., and Wang, B. (2016). Research on the Life-Cycle CO 2 Emission of China's Construction Sector. *Energy and Buildings* 112, 244–255. doi:10.1016/j.enbuild.2015.12.026

Zhao, Y., Cao, Y., Shi, X., Zhang, Z., and Zhang, W. (2021). Structural and Technological Determinants of Carbon Intensity Reduction of China's Electricity Generation. *Environ. Sci. Pollut. Res.* 28, 13469–13486. doi:10.1007/s11356-020-11429-0

Zhou, X., Zhou, D., Wang, Q., and Su, B. (2020). Who Shapes China's Carbon Intensity and How? A Demand-Side Decomposition Analysis. *Energ. Econ.* 1, 85. doi:10.1016/j.eneco.2019.104600

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors XS.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Li, Yu, Shi and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fenrg.2021.756311

# Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the $CO_2$ Emissions Analysis Using Gaussian Process Regression

*Ning Ma*[1], *Wai Yan Shum*[2], *Tingting Han*[1] *and Fujun Lai*[3]*

[1]School of Financial Management, Hainan College of Economics and Business, Haikou, China, [2]Department of Economics and Finance, The Hang Seng University of Hong Kong, Hong Kong, SAR China, [3]School of Finance, Yunnan University of Finance and Economics, Kunming, China

In this paper, a nonparametric kernel prediction algorithm in machine learning is applied to predict $CO_2$ emissions. A literature review has been conducted so that proper independent variables can be identified. Traditional parametric modeling approaches and the Gaussian Process Regression (GPR) algorithms were introduced, and their prediction performance was summarized. The reliability and efficiency of the proposed algorithms were then demonstrated through the comparison of the actual and the predicted results. The results showed that the GPR method can give the most accurate predictions on $CO_2$ emissions.

**Keywords: Gaussian process regression, $CO_2$ emissions, energy consumption, economics growth, industralization**

OPEN ACCESS

***Edited by:***
*Xunpeng (Roc) Shi,*
*University of Technology Sydney, Australia*

***Reviewed by:***
*Manman Yuan,*
*University of Science and Technology Beijing, China*
*JunCheng Li,*
*Chinese Academy of Social Sciences, China*

***Correspondence:**
*Fujun Lai*
*lfjlfj999@163.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *10 August 2021*
***Accepted:*** *08 September 2021*
***Published:*** *24 September 2021*

***Citation:***
*Ma N, Shum WY, Han T and Lai F (2021) Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the $CO_2$ Emissions Analysis Using Gaussian Process Regression. Front. Energy Res. 9:756311. doi: 10.3389/fenrg.2021.756311*

## INTRODUCTION

As the population of the earth is being exponentially increasing, the exhaustion of carbon dioxide is increasing day by day resulting in the extreme overheating of the environment, which has become a significant reason for climate change. Global efforts to mitigate climate change were focused on the reduction of future days with extreme overheating of the environment. There are many research and surveys that have been conducted by various scientists, students, and other officials which were about the reasons for the high emission of $CO_2$ in different countries. Most of the empirical studies took the parametric modeling approach to analyze the factors that initiate and support the emission of $CO_2$. However, the traditional parametric approach optimizes a function to a known form with a set of finite and pre-determined parameters. This rigidity limits the predictive power of the parametric models. In recent years, nonparametric machine learning techniques have played dominantly with the enhancement of the forecast.

In this paper, a Bayesian nonparametric kernel prediction algorithm in machine learning is applied to predict $CO_2$ emissions. A literature review has been conducted so that the proper independent variables have been identified. Classical least squares, robust least squares, and algorithms of the GPR were introduced and their prediction performance, including the evaluation criteria that are effective in the measurements for model performance, were summarized. The reliability and efficiency of the proposed algorithms were then demonstrated through the comparison between the actual data and the predicted results. It is found that GPR can give the most accurate predictions on $CO_2$ emissions.

## LITERATURE REVIEW

The growth of the economy, energy utilization, and $CO_2$ emissions are deeply related to each other. Kolstad and Krautkraemer (1993) point out that while the use of resources like the energy has a bright side on growth, it has negative environmental impacts. Traditional growth theories like the Solow growth model failed to consider the environmental impacts of growth (Solow, 1956). More modern growth theories study the interrelationship among energy, the environment, and economic growth (see, for example, Kolstad and Krautkraemer (1993), Jorgenson and Wilcoxen (1993) or Xepapadeas (2005) for a brief review).

Empirical studies depict that the growth of the economy and the ingestion of energy incorporates with the process of $CO_2$ emission. Recently, Hu et al. (2020) study the dominant reasons for carbon emission among the Belt and Road countries and find that $CO_2$ emissions have increased significantly due to economic growth. Similarly, Shabaz et al. (2013) found that in Indonesia, the emission of $CO_2$ increased for the extreme boost of the economic zone, while Shahbaz et al. (2016) found that economic growth led to $CO_2$ emissions in Bangladesh and Egypt. Meanwhile, other studies discovered a bilateral causal relationship among the three variables. Munir et al. (2020) prove the fact that there is a relation of aftermaths and economy between GDP and energy ingestion in the major countries of the ASEAN (Association of Southeast Asian Nations), while Liu and Hao (2018) find that in energy-exporting countries, there is a bilateral relationship which may be a full-duplex connection between $CO_2$ emissions, energy utilization, and GDP per capita. Similarly, a repeating loop effect is observed between energy ingestion, $CO_2$ emission, and the advancement of the economy by Kahouli (2018). Accordingly, Mohmannd et al. (2020) observed the working principle of the causal relationship among transportation infrastructure, economic growth, and transportation emissions from 1971 to 2017 in Pakistan. The results show short-term causality from transportation infrastructure, economic growth, fuel consumption to $CO_2$ emissions, and the long-run relationship between economic advancement and infrastructure.

Apart from the growth and energy consumption, industrialization, population growth, and income level also contributed a great share in global carbon emissions. Minx et al. (2011) found that "industrialization" can be taken into consideration for the rapid increase of carbon dioxide emission in China from 2002 to 2007 while Zhang et al. (2014) found that the growth of the tertiary industry can decrease the $CO_2$ exhalation intensity. Nasir et al. (2021) examined the connection between the factors which are the exhalation of $CO_2$, industrialization, growth of the economy, energy ingestion, and several connecting factors from 1980 to 2014 in Australia. The observations of those involved say that all variables affect $CO_2$ emissions. Li et al. (2021) discussed the effect of the growth and structure of the economy on per capita $CO_2$ emissions in 147 countries from 1990 to 2015. The results show that at the global level, economic growth and economic structure are the most significant positive and positive effects, respectively.

Studies on population have thus far concentrated on the relationship between population growth and emission increase. The effect of population growth on $CO_2$ emissions can be summarized as follows (Birdsall, 1992): On one side, the energy demand was increased for power generation, industry, and transport. On the contrary, it increased deforestation emissions due to population growth. Empirically, Knapp and Mookerjee (1996) conducted a Granger causality test on annual data from 1880 to 1989 to determine the connecting clauses between global population expansion and carbon dioxide exhaustion. The results show there is a short-term dynamic relationship between the exhaustion of carbon dioxide and population growth. Very recently, Zhang et al. (2020) analyzed the knot between $CO_2$ emissions, GDP, and fuel ingestion in China and ASEAN countries. It was found that carbon density, energy intensity, GDP, and population are positively correlated with $CO_2$ emissions. Empirical findings also show that the developing countries are facing the effect of overpopulation and that's why, they are facing more of a carbon emissions record per year other than the developed countries (Shi, 2003).

In the past decade, the theory and methodology of the Environmental Kuznets Curve (EKC) have been used to analyze the relationship between the net income and exhaustion of carbon of an area (Dinda, 2004; Williams and Rasmussen, 1996). According to the EKC, at relatively low-income levels, emissions increase as income increases. After a certain point, emissions will decline with income. Thus, the emission of $CO_2$ varies concerning the level of income. Luo et al. (2021) investigated the influencing factors of Shanghai's $CO_2$ emissions from 1995 to 2017. They found that personal disposable income is one of the top drivers of $CO_2$ emissions. Yuan et al. (2014) examined the long-term relationship between China's per capita income, ingestion of energy, and the emission of $CO_2$ from 1953 to 2008. They found out, there is a unilateral Granger inter-relation between the gross national income and the emission of $CO_2$.

Based on the literature above, it concludes that economic up-gradation, energy utilization, manpower density, industrialization, and income can be classified as the predominant factors affecting $CO_2$ emissions. Other factors might also affect $CO_2$ emissions in China. For example, R&D (Nguyen et al., 2020; Jones, 1995), financial development (Bhattacharya et al., 2017; Zaidi et al., 2019; Wang et al., 2020), the degree of foreign direct investment (Essandoh et al., 2020; Le et al., 2020; Khan and Rana, 2021. etc). This paper limits the focus on how well the different prediction models perform based on the information set which includes only the most predominant driver of $CO_2$ emissions and excludes those unimportant ones to be captured by the stochastic terms in the models.

## METHODOLOGY

Gaussian Process Regression (GPR) method can be introduced as a non-parametric Bayesian regression method (Gershman and Blei, 2012). It captures a wide variety of relations between inputs

and outputs and lets the data determine the complexity of the underlying functions through the means of Bayesian inference (Williams, 1998). Considering the output $y$ of a function $w$ at input $x$ with independent and identically distributed random noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$. The function accompanied with the distributed random noise can be presented as:

$$y = w(x) + \varepsilon \tag{1}$$

In classical linear regression, $w(x)$ is deterministic whereas the noise term is random. In Gaussian process regression, however, $w(x)$ is assumed to be random and follows a Gaussian process. A Gaussian process is an extension of multivariant Gaussian distribution to infinite dimensions; any finite subset sampled from the Gaussian process follows multiple Gaussian distributions (MacKay, 1998). The distribution over functions can be described with the help of the Gaussian process,

$$w(x) = \mathcal{GP}\big(m(x), k\big(x, x^{'}\big)\big) \tag{2}$$

where $x$ is applied as the input variable, $m(x)$ is denoted as the mean function, finally, $k(x, x^{'})$ is known as covariance function. These two functions are defined respectively as:

$$\begin{aligned} m(x) &= \mathbb{E}[w(x)] \\ k\big(x, x^{'}\big) &= cov\big[w(x), w\big(x^{'}\big)\big]. \end{aligned} \tag{3}$$

A finite collection of function values sampled from the Gaussian process follows multiple Gaussian distributions:

$$\big[w\big(x^{(1)}\big), w\big(x^{(2)}\big), \ldots, w\big(x^{(n)}\big)\big]^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, K) \tag{4}$$

where $K$ is a $n \times n$ (n by n) matrix with the entries $K_{ij} = k\big(x^{(i)}, x^{(j)}\big)$ and $\mu$ has entries $\mu_i = m\big(x^{(i)}\big)$. Given a training set that contains observation points $\boldsymbol{y} = [y\big(x^{(1)}\big), y\big(x^{(2)}\big), \ldots, y\big(x^{(n)}\big)]^{\mathrm{T}}$ and function values $\boldsymbol{w} = [w\big(x^{(1)}\big), w\big(x^{(2)}\big), \ldots, w\big(x^{(n)}\big)]^{\mathrm{T}}$, it follows that the conditional distribution $p(y|w)$ and the Gaussian prior $p(w)$ are $\mathcal{N}(w, \sigma_\varepsilon^2 I)$ and $\mathcal{N}(\mu, K)$, respectively. By definition, the set of observations $y$ and the set of function values $w$ follow a joint multivariate Gaussian distribution. The join distribution $p(w(x^*), y)$ is defined as

$$\begin{bmatrix} w(x^*) \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu^* \\ \mu \end{bmatrix}, \begin{bmatrix} k(x^*, x^*) & k^{*\mathrm{T}} \\ k^* & K + \sigma_\varepsilon^2 I \end{bmatrix}\right) \tag{5}$$

Here, $I$ will be considered as the identity matrix, $\sigma_\varepsilon^2$ is the unknown variance of the random noise and $(\boldsymbol{k}^*)_i = k(x^*, x^{(i)})$ for $i = 1, 2, \ldots, N$. Using the Bayesian rule, the predictive posterior, $p(w(x^*)|\boldsymbol{y}) \sim \mathcal{N}(\overline{w}^*, \Sigma^*)$, can be obtained, and the mean $\overline{w}^*$ and variance $\Sigma^*$ are defined by

$$\overline{w}^* = \mu^* + (k^*)^T\big[K + \sigma_\varepsilon^2 I\big]^{-1}(y - \mu)\Sigma^* = k(x^*, x^*) - (k^*)^T\big[K + \sigma_\varepsilon^2 I\big]^{-1}(k^*) \tag{6}$$

The covariance function determines the characteristics of the Gaussian method that can be expressed as $k\big(x^{(i)}, x^{(j)}\big)$. The covariance function models the dependence between the function values at different input points $x^{(i)}$ and $x^{(j)}$. The covariance function is often called the kernel of the Gaussian process. There are many possible options for the prior covariance function. A popular kernel is the exponential covariance function which allows the model to general a non-negative definite covariance matrix for any set of input points (Williams and Rasmussen, 1996). The exponential covariance function is defined as

$$k\big(x^{(i)}, x^{(j)}\big) = \sigma_f^2 exp\left(-\frac{\| x^{(i)} - x^{(j)}\|^2}{2I^2}\right) + \sigma_\varepsilon^2 \delta_{ij} \tag{7}$$

where $I$ is the characteristic length scale, $\sigma_f^2$ is the signal variance, and $\delta_{ij}$ is a Kronecker delta. The Gaussian process regression employs a set of hyperparameters $\theta$ including $I$, $\sigma_f^2$ and $\sigma_n^2$ to increase or reduce the priority correlation between points and consequentially the variability of the resulting function. The hyperparameters $\theta$ can be optimized based on the log-likelihood framework:

$$\begin{aligned} L &= \log p(y|w, \theta) = -\frac{1}{2} y^T C^{-1} y - \frac{1}{2}\log|C| - \frac{n}{2}\log 2\pi, \quad C \\ &= K + \sigma_\varepsilon^2 I \end{aligned} \tag{8}$$

More details about the regression process of Gaussian can be researched and acknowledged in the book of Williams and Rasmussen (2006), available free online and is accessible via the link: www.GaussianProcess.org/gpml.

## EMPIRICAL RESULTS

A literature review has been conducted so that five independent variables; namely: economic growth, energy consumption, population, industrialization, and income, have been identified. In this study, the GPR method and the other proposed algorithms are applied to study carbon emissions in China. Economic growth is approximated by GDP (100 million RMB), energy consumption is approximated by per capita energy consumption (tons of standard coal), the population is approximated by population size (10,000 people), industrialization is weighted by the percentage of secondary industry in China, and income is measure by the average annual salary (RMB).

The data of GDP, population size, energy consumption, percentage of secondary industry, and average annual salary are collected from the China City Statistical Yearbook. $CO_2$ emissions data come from four main sources of energy consumption. These are electricity, fuel, heating, and transportation. Those data can be obtained and calculated through the China Urban Construction Statistical Yearbook, the China City Statistical Yearbook, and the submerged government Panel on the change of weather and climate. Since some of those data is not available after 2014, the data in this paper range from the year 2002–2014.

### Statistical Analysis of Prediction Results

The commonly used criteria in prediction performance are used in this study to evaluate the validity of the fitting. In **Table 1**, the root means squared error (RMSE), the mean squared error (MSE), the R-square, and the mean absolute error (MSE) are shown, where a well-fitted model should have R-square close to 1, whereas the RMSE, the MSE, and the MAE should be as small as possible. As per the observation from

**TABLE 1 |** Statistical Analysis of Prediction Results.

| | Classical least squares | Robust least squares | Exponential GPR |
|---|---|---|---|
| RMSE | 0.10964 | 0.11045 | 0.083734 |
| R-Squared | 0.95 | 0.94 | 0.97 |
| MSE | 0.012022 | 0.012198 | 0.0070114 |
| MAE | 0.08049 | 0.079101 | 0.059303 |

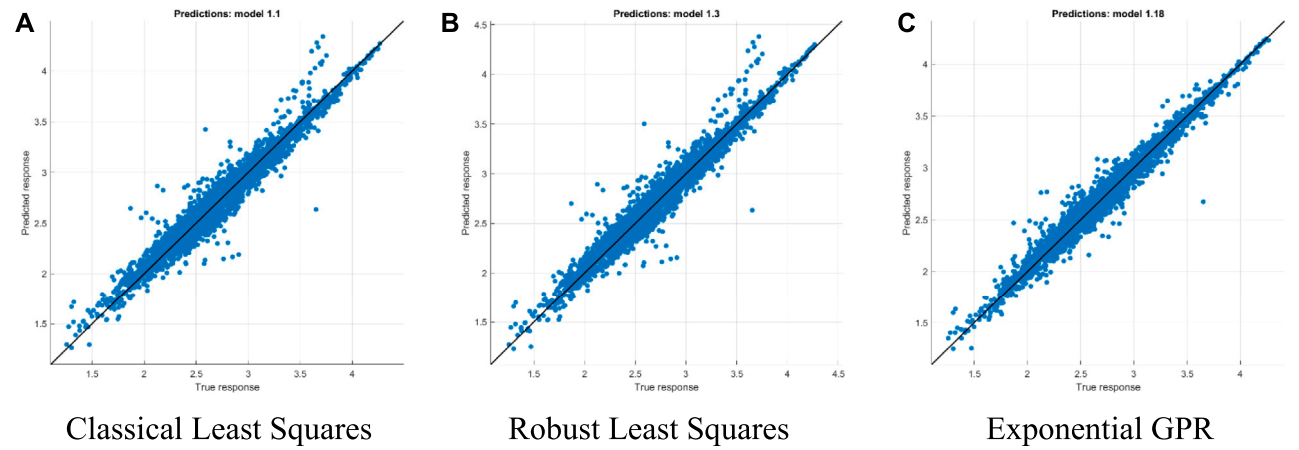



**FIGURE 1 |** Comparison of actual value and prediction of $CO_2$ emissions between the selected models. **(A)** Classical Least Squares. **(B)** Robust Least Squares. **(C)** Exponential GPR. Notes: 1) The horizontal axis represents actual $CO_2$ emissions in logarithm, and the vertical axis represents predicted $CO_2$ emissions in logarithm. 2) $CO_2$ emissions are measured in ten thousand tons of standard coal.

**Table 1**, Exponential GPR provides the best fit data as it has the smallest RMSE, MSE, and MAE, and an R-square closest to 1.

## Data Visualization

Since the data set is large, which made it difficult to demonstrate and view the whole set of data, visualization methods are typically needed especially for representative scenarios. The prediction results were analyzed at the model level to see the allover authenticity of the three models and at the individual component level to get a picture of the estimates produced by the three models over the range of some particular variable.

At the overall level, the comparison and deviation of the actual value and the predicted dimension of the emission of carbon dioxide are determined. **Figure 1** demonstrates the comparison of actual value and prediction of $CO_2$ emissions predicted by the three models; for each model, the predicted value is plotted against the actual value. To have a good fit, each plot should resemble a straight line at 45°. However, compare with the exponential GPR model, for the classical least squares model and the robust least-squares model, the predicted values are larger than the actual values over the range of 3.5–4 logarithm units of $CO_2$ emissions. This means that the classical least squares model and the robust least-squares model are overestimating $CO_2$ emissions over a particular range compare with the exponential GPR model. The same issue can be observed from **Figure 2** which shows the deviation of actual value and prediction of $CO_2$ emissions for the three models. **Figure 2** shows that, compare with the other two models, the deviations for the exponential GPR model cluster more closely around the horizontal line which represents no deviations. It suggests that the exponential GPR model provides a much better fit than the other two models.

Apart from analyzing the prediction results at the overall model level, the all over performance of the three models is also be evaluated at an individual component level. At the individual component level, the estimates produced by the selected models are analyzed over the extended range of some particular variables. **Figures 3–7** below plot the actual and predicted values of $CO_2$ emissions against each of the most predominant factors of the models.

**Figure 3** plots the predicted values of $CO_2$ emissions against the logarithm of the GDP measured in 10,000 Chinese Yuen. Ideally, it's convenient if the predicted values are as much closer possible to the actual values for all conducted observations. As shown in **Figure 3C** predicted $CO_2$ emissions are quite close to the actual values predicted by using the logarithm of GDP. Even though a small number of deviations can be observed. On the contrary, **Figures 3A,B** revealed that the classical least squares and the robust least-squares overestimate the $CO_2$ emissions over the range of 3.2–3.7 logarithm units of GDP. It implies that conditioning on GDP, the Exponential GPR model provides more accurate $CO_2$ emissions predictions compare with the other two models.

**Figures 4–6** show similar results. The predicted $CO_2$ emissions by using the exponential GPR model are tensed to the actual values over the entire range of population size (see **Figure 4**), the energy consumption (see **Figure 5**), and the level of industrialization (see **Figure 6**). However, when the classical least squares and the robust

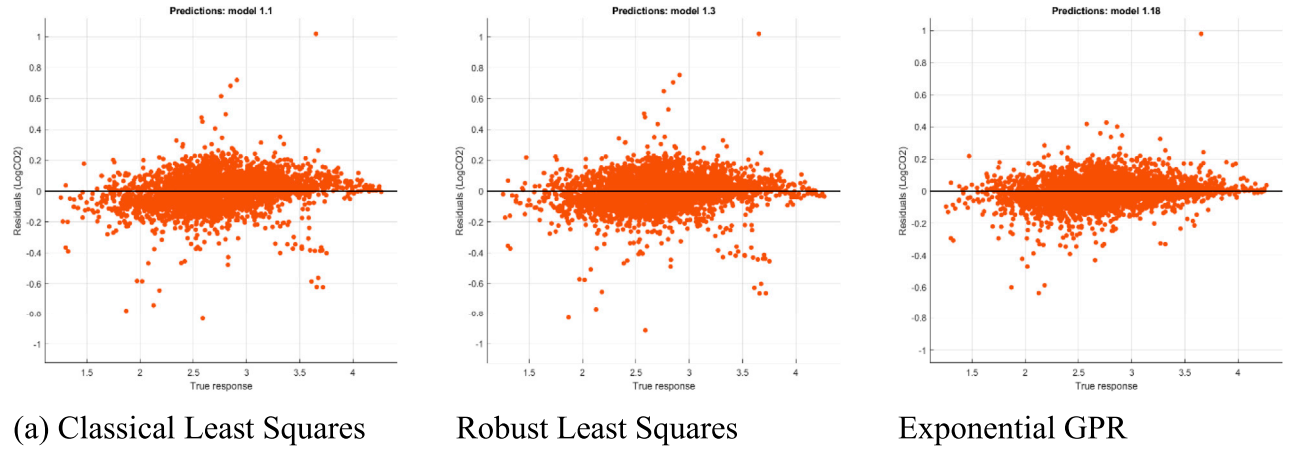



**FIGURE 2 |** The deviation of actual value and prediction of $CO_2$ emissions between the selected models. **(A)** Classical Least Squares, Robust Least Squares, Exponential GPR. Notes: 1) The horizontal axis represents actual $CO_2$ emissions in logarithm, and the deviation of the actual extremity of the $CO_2$ emission from the predicted value is represented by the vertical axis. 2) $CO_2$ emissions are measured in ten thousand tons of

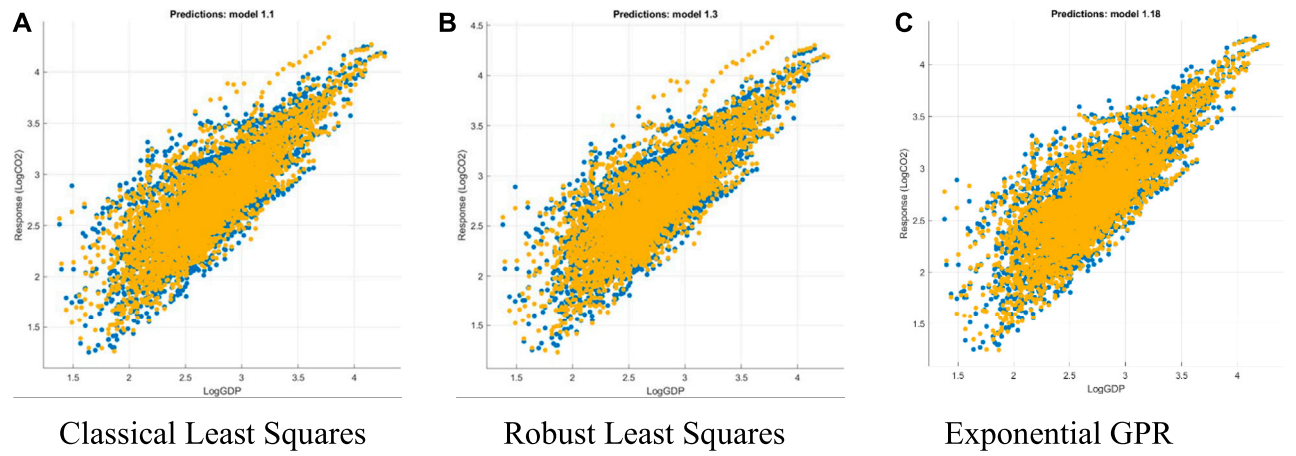



**FIGURE 3 |** Comparing actual and predicted $CO_2$ emissions against GDP of the selected models. **(A)** Classical Least Squares. **(B)** Robust Least Squares. **(C)** Exponential GPR. Notes: 1) The vertical axis denotes the value of CO2 emissions in the logarithm. 2) The horizontal axis denotes the value of GDP in logarithm. 3) $CO_2$ emissions are measured in 10,000 tons of standard coal. 4) The blue dots represent actual values whereas the yellow dots represent the predicted values. (5) GDP is measured in 100 million Chinese Yuan.

least-squares model are used, extreme deviations between the actual value and predicted value can be observed. In **Figures 4A,B**, it is determined that the classical least squares and the robust least-squares model overestimate the $CO_2$ emissions over the range of 2.7–2.8 logarithm units of population size. Similarly, in **Figures 5A,B**, $CO_2$ emissions are overestimated by the classical least squares and the robust least squares models over the range of 0.5–1 logarithm units of per capita energy consumption. In **Figure 6**, although not obvious, $CO_2$ emissions are overestimated by the classical least squares and the robust least squares models over the range of 40–50% of secondary industry in China.

**Figure 7** shows how the predicted values deviate from the actual values when the independent variable is non-Gaussian for the presence of threshold data points from extreme references. As with the evidence presented above, extreme upward bias over a particular range can be observed when the classical least squares and the robust least squares are used; the models overestimated $CO_2$ emissions over the range of 4.5– to 4.75 logarithm units of average annual salary. The extreme bias disappears when the exponential GPR model is used. Moreover, when the exponential GPR model is used, the deviations between the actual values and the predicted values are smaller for the extreme data values observed over the range of 1– to 1.5 and 5 to 5.5 logarithm units of average annual salary.

In summary, **Figures 3–7** show that predicted $CO_2$ emissions conditional on individual components (i.e., GDP, population size, energy consumption, and industrialization) are quite close to the actual values predicted using the exponential GPR model. Even the underlying distribution of

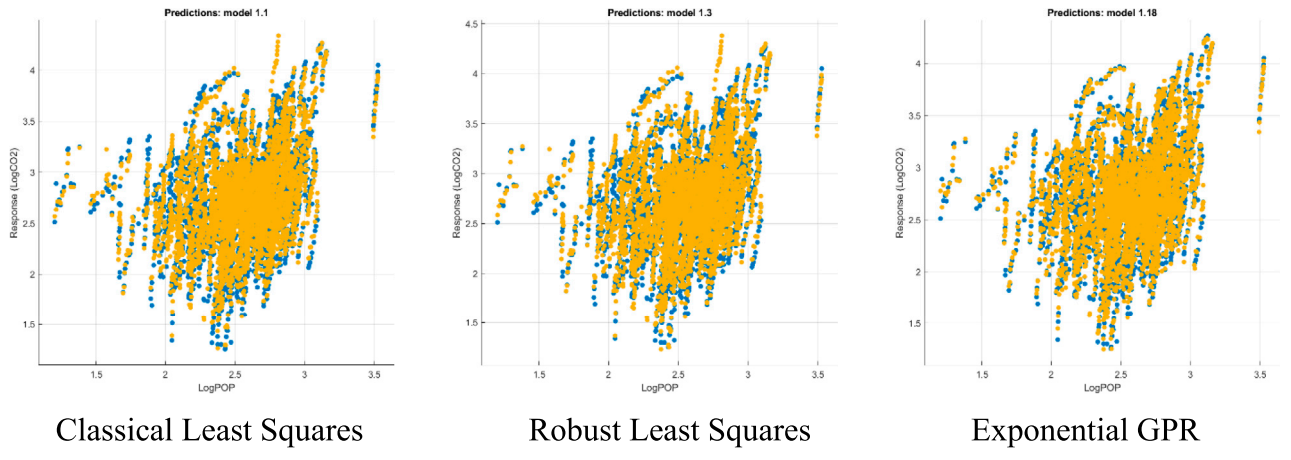



**FIGURE 4 |** Comparing actual and predicted $CO_2$ emissions against population size of the selected models. Classical Least Squares, Robust Least Squares, Exponential GPR. Notes: 1) The vertical axis represents the value of $CO_2$ emissions in logarithm. 2) The horizontal axis represents the population size in logarithm. 3) $CO_2$ emissions are measured in 10,000 tons of standard coal. 4) The blue dots represent actual values whereas the yellow dots represent the predicted values. 5) The unit of population is 10,000 people.

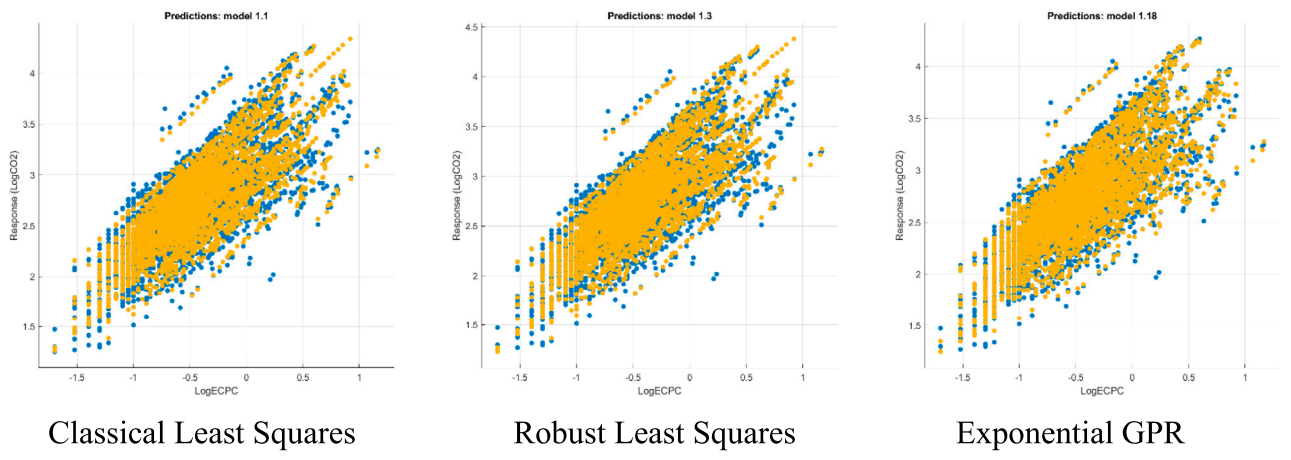



**FIGURE 5 |** Comparing actual and predicted $CO_2$ emissions against energy consumption of the selected models. Classical Least Squares, Robust Least Squares, Exponential GPR. Notes: 1) The vertical axis represents the value of $CO_2$ emissions in logarithm. 2) The horizontal axis represents the value of per capita energy consumption in a logarithm. 3) $CO_2$ emissions are measured in 10,000 tons of standard coal. 4) The blue dots represent actual values whereas the yellow dots represent the predicted values. 5) Per capita energy consumption is measured in tons of standard coal.

the independent variable is non-Gaussian. Meanwhile, extreme upward bias per component in the technique can be observed when the classical least squares and the robust least squares models are used. Thus, a conclusion may be drawn upon the study that the exponential GPR model gives the most accurate predictions on $CO_2$ emissions compared with the remaining models.

## CONCLUSION AND FUTURE WORKS

In this paper, the Gaussian process regression method is proposed for $CO_2$ emissions analysis in China. The traditional linear regression approach is limited by its rigid functional form and the approach often encounters an over-fitting problem. The Gaussian progress regression approach relaxes the parametric assumption by applying the Bayesian nonparametric inference approach. The preciseness and exactitude of the prediction of the exponential GPR were compared and discussed with the classical least squares and the robust least-squares model. Based on the outcome of the whole study, it is proved that the Gaussian progress regression algorithms can give the most accurate predictions on $CO_2$ emissions compared with the other two traditional models and thus is applicable for $CO_2$ emissions prediction analysis to enhance forecast performance.

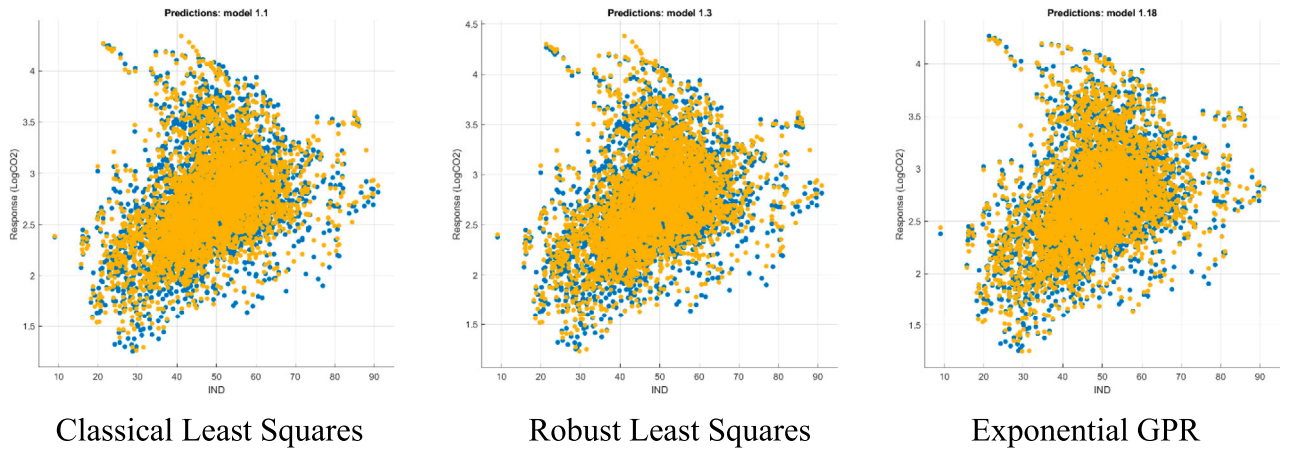



**FIGURE 6 |** Comparing actual and predicted $CO_2$ emissions against the industrialization of the selected models. Classical Least Squares, Robust Least Squares, Exponential GPR. Notes: 1) The vertical axis represents the value of $CO_2$ emissions in logarithm. 2) The horizontal axis represents the percentage of secondary industry in China. 3) The blue dots represent actual values whereas the yellow dots represent the predicted values. 4) $CO_2$ emissions are measured in 10,000 tons of standard coal.

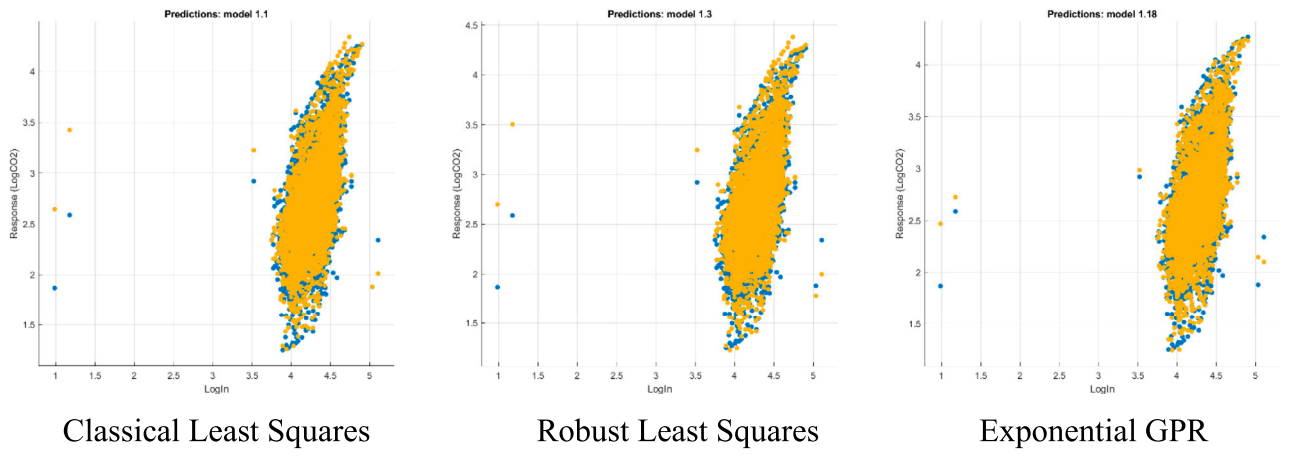



**FIGURE 7 |** Comparing actual and predicted $CO_2$ emissions against average annual salary of the selected models. Classical Least Squares Robust Least Squares Exponential GPR Notes: 1) The vertical axis represents the value of $CO_2$ emissions in logarithm. 2) The horizontal axis represents the average annual salary in logarithm. 3) $CO_2$ emissions are measured in 10,000 tons of standard coal. 4) The blue dots represent actual values whereas the yellow dots represent the predicted values. 5) Average annual salary is measured in the Chinese Yuan.

The prediction performances of the selected methods discussed only focus on the six predominant factors affecting carbon emissions. Future research should focus on further reviewing the completeness of the set of driving factors and the effectiveness of model predictions, compared them with other commonly used models.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

NM proposed the conceptualization, methodology and funding acquisition. WS gave the formal data analysis and wrote the original formal draft. TH performed the data collection and original arrangement. FL gave formal methodology, writing-review and editing.

## FUNDING

This study was supported by Hainan College of Economics and Business (Project Reference Number: hnjmk2021301) and the Scientific Research Foundation of Yunnan University of Finance and Economics (No. 2020B03).

## REFERENCES


Bhattacharya, M., Rafiq, S., Lean, H. H., and Bhattacharya, S. (2017). The Regulated Coal Sector and CO2 Emissions in Indian Growth Process: Empirical Evidence over Half a century and Policy Suggestions. *Appl. Energ.* 204, 667–678. doi:10.1016/j.apenergy.2017.07.061

Birdsall, N. (1992). Another Look at Population and Global Warming. In Policy Research Working Papers ; No. WPS 1020. Population, Health, and Nutrition, 1020. Retrieved from http://documents.worldbank.org/curated/en/985961468766195689/Another-look-at-population-and-global-warming.

Dinda, S. (2004). Environmental Kuznets Curve Hypothesis: A Survey. *Ecol. Econ.* 49 (4), 431–455. doi:10.1016/j.ecolecon.2004.02.011

Essandoh, O. K., Islam, M., and Kakinaka, M. (2020). Linking International Trade and Foreign Direct Investment to CO2 Emissions: Any Differences between Developed and Developing Countries?. *Sci. Total Environ.* 712, 136437. doi:10.1016/j.scitotenv.2019.136437

Gershman, S. J., and Blei, D. M. (2012). A Tutorial on Bayesian Nonparametric Models. *J. Math. Psychol.* 56 (1), 1–12. doi:10.1016/j.jmp.2011.08.004

Hu, M., Li, R., You, W., Liu, Y., and Lee, C.-C. (2020). Spatiotemporal Evolution of Decoupling and Driving Forces of CO2 Emissions on Economic Growth along the Belt and Road. *J. Clean. Prod.* 277, 123272. doi:10.1016/j.jclepro.2020.123272

Jones, C. I. (1995). R & D-Based Models of Economic Growth. *J. Polit. Economy* 103 (4), 759–784. doi:10.1086/262002

Jorgenson, D. W., and Wilcoxen, P. J. (1993). Reducing U.S. Carbon Dioxide Emissions: an Assessment of Different Instruments. *J. Pol. Model.* 15 (5-6), 491–520. doi:10.1016/0161-8938(93)90003-9

Kahouli, B. (2018). The Causality Link between Energy Electricity Consumption, CO2 Emissions, R&D Stocks and Economic Growth in Mediterranean Countries (MCs). *Energy* 145, 388–399. doi:10.1016/j.energy.2017.12.136

Khan, M., and Rana, A. T. (2021). Institutional Quality and CO2 Emission-Output Relations: The Case of Asian Countries. *J. Environ. Manage.* 279, 111569. doi:10.1016/j.jenvman.2020.111569

Knapp, T., and Mookerjee, R. (1996). Population Growth and Global CO2 Emissions. *Energy Policy* 24 (1), 31–37. doi:10.1016/0301-4215(95)00130-1

Kolstad, C. D., and Krautkraemer, J. A. (1993). "Natural Resource Use and the Environment,". In *Handbook of Natural Resource and Energy Economics*. Editors A. V. Kneese and J. L. Sweeney. 1st ed. Amsterdam North Holland, 3, 1219–1265. doi:10.1016/s1573-4439(05)80013-2

Li, R., Wang, Q., Liu, Y., and Jiang, R. (2021). Per-capita Carbon Emissions in 147 Countries: The Effect of Economic, Energy, Social, and Trade Structural Changes. *Sustainable Prod. Consumption* 27, 1149–1164. doi:10.1016/j.spc.2021.02.031

Liu, Y., and Hao, Y. (2018). The Dynamic Links between CO2 Emissions, Energy Consumption and Economic Development in the Countries along "the Belt and Road". *Sci. Total Environ.* 645, 674–683. doi:10.1016/j.scitotenv.2018.07.062

Luo, Y., Zeng, W., Hu, X., Yang, H., and Shao, L. (2021). Coupling the Driving Forces of Urban CO2 Emission in Shanghai with Logarithmic Mean Divisia index Method and Granger Causality Inference. *J. Clean. Prod.* 298, 126843. doi:10.1016/j.jclepro.2021.126843

Mackay, D. J. C. (1998). Introduction to Gaussian Processes. In Neural Networks and Machine Learning. Berlin: Springer.

Minx, J. C., Baiocchi, G., Peters, G. P., Weber, C. L., Guan, D., and Hubacek, K. (2011). A "Carbonizing Dragon": China's Fast Growing CO2 Emissions Revisited. *Environ. Sci. Technol.* 45 (21), 9144–9153. doi:10.1021/es201497m

Mohmand, Y. T., Mehmood, F., Mughal, K. S., and Aslam, F. (2020). Investigating the Causal Relationship between Transport Infrastructure, Economic Growth and Transport Emissions in Pakistan. *Res. Transportation Econ.*, 100972. doi:10.1016/j.retrec.2020.100972

Munir, Q., Lean, H. H., and Smyth, R. (2020). CO2 Emissions, Energy Consumption and Economic Growth in the ASEAN-5 Countries: A Cross-Sectional Dependence Approach. *Energ. Econ.* 85, 104571. doi:10.1016/j.eneco.2019.104571

Nasir, M. A., Canh, N. P., and Lan Le, T. N. (2021). Environmental Degradation & Role of Financialisation, Economic Development, Industrialisation and Trade Liberalisation. *J. Environ. Manage.* 277, 111471. doi:10.1016/j.jenvman.2020.111471

Nguyen, T. T., Pham, T. A. T., and Tram, H. T. X. (2020). Role of Information and Communication Technologies and Innovation in Driving Carbon Emissions and Economic Growth in Selected G-20 Countries. *J. Environ. Manage.* 261, 110–162. doi:10.1016/j.jenvman.2020.110162

Shahbaz, M., Hye, Q. M. A., Tiwari, A. K., and Leitão, N. C. (2013). Economic Growth, Energy Consumption, Financial Development, International Trade and CO2 Emissions in Indonesia. *Renew. Sustain. Energ. Rev.* 25, 109–121. doi:10.1016/j.rser.2013.04.009

Shahbaz, M., Mahalik, M. K., Shah, S. H., and Sato, J. R. (2016). Time-varying Analysis of CO2 Emissions, Energy Consumption, and Economic Growth Nexus: Statistical Experience in Next 11 Countries. *Energy Policy* 98, 33–48. doi:10.1016/j.enpol.2016.08.011

Shi, A. (2003). The Impact of Population Pressure on Global Carbon Dioxide Emissions, 1975-1996: Evidence from Pooled Cross-Country Data. *Ecol. Econ.* 44 (1), 29–42. doi:10.1016/s0921-8009(02)00223-9

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Q. J. Econ.* 70 (1), 65–94. doi:10.2307/1884513

Wang, R., Mirza, N., Vasbieva, D. G., Abbas, Q., and Xiong, D. (2020). The Nexus of Carbon Emissions, Financial Development, Renewable Energy Consumption, and Technological Innovation: What Should Be the Priorities in Light of COP 21 Agreements?. *J. Environ. Manage.* 271, 111027. doi:10.1016/j.jenvman.2020.111027

Williams, C. K. (1998). *Learning in Graphical Models*. Dordrecht ; Boston: Kluwer Academic Publishers. doi:10.1007/978-94-011-5014-9_23Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and beyond

Williams, C. K., and Rasmussen, C. E. (2006). *Gaussian Processes for Machine Learning*. Cambridge, Mass: Mit Press.

Williams, C. K., and Rasmussen, C. E. (1996). Gaussian Processes for Regression. In Advances in Neural Information Processing Systems. Cambridge, MassLondon: Mit Press.

Xepapadeas, A. (2005). "Chapter 23 Economic Growth and the Environment,". *Economywide and International Environmental Issues*. Editors Karl-GöranM. and J. R. Vincent (Amsterdam: North Holland), 3, 1219–1271. doi:10.1016/s1574-0099(05)03023-8

Yuan, J., Xu, Y., and Zhang, X. (2014). Income Growth, Energy Consumption, and Carbon Emissions: The Case of China. *Emerging Markets Finance and Trade* 50, 169–181.

Zaidi, S. A. H., Zafar, M. W., Shahbaz, M., and Hou, F. (2019). Dynamic Linkages between Globalization, Financial Development and Carbon Emissions: Evidence from Asia Pacific Economic Cooperation Countries. *J. Clean. Prod.* 228, 533–543. doi:10.1016/j.jclepro.2019.04.210

Zhang, J., Fan, Z., Chen, Y., Gao, J., and Liu, W. (2020). Decomposition and Decoupling Analysis of Carbon Dioxide Emissions from Economic Growth in the Context of China and the ASEAN Countries. *Sci. Total Environ.* 714, 136649. doi:10.1016/j.scitotenv.2020.136649

Zhang, Y.-J., Liu, Z., Zhang, H., and Tan, T.-D. (2014). The Impact of Economic Growth, Industrial Structure and Urbanization on Carbon Emission Intensity in China. *Nat. Hazards* 73 (2), 579–595. doi:10.1007/s11069-014-1091-x


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Ma, Shum, Han and Lai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

BRIEF RESEARCH REPORT
published: 24 September 2021
doi: 10.3389/fenrg.2021.750652

# The Role of Economic Policy Uncertainty in Renewable Energy-Growth Nexus: Evidence From the Rossi-Wang Causality Test

*Zhou Lu[1], Linchuang Zhu[1]*, Chi Keung Marco Lau[2]*, Aliyu Buhari Isah[3] and Xiaoxian Zhu[2]*

*[1]School of Economics, Tianjin University of Commerce, Tianjin, China, [2]Teesside University Business School, Teesside University, Middlesbrough, United Kingdom, [3]Huddersfield Business School, University of Huddersfield, Huddersfield, United Kingdom*

This paper examines the causal relationship between renewable energy consumption and economic growth in four countries: Brazil, Germany, Japan, and the United States. Unlike previous papers, we control economic policy uncertainty's effects to capture the role of capabilities on the renewable energy-growth nexus. The recent Vector Autoregression (VAR)-based Granger-causality test of Rossi-Wang shows a bidirectional causal relationship between renewable energy and the economic growth in Brazil and Germany. There is also a significant causality from renewables to economic growth in the United States.

**Keywords: renewable energy consumption, economic growth, renewable energy, economic policy uncertainty, Granger-causality test, instabilities**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong, Hang Seng University of Hong Kong, China*

***Reviewed by:***
*Jianmin Sun, Nanjing University of Posts and Telecommunications, China*
*Hua Zhang, Zhejiang University, China*

****Correspondence:***
*Linchuang Zhu*
*linchuang_zhu2021@163.com*
*Chi Keung Marco Lau*
*c.lau@tees.ac.uk*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *30 July 2021*
***Accepted:*** *25 August 2021*
***Published:*** *24 September 2021*

***Citation:***
*Lu Z, Zhu L, Lau CKM, Isah AB and Zhu X (2021) The Role of Economic Policy Uncertainty in Renewable Energy-Growth Nexus: Evidence From the Rossi-Wang Causality Test. Front. Energy Res. 9:750652. doi: 10.3389/fenrg.2021.750652*

## INTRODUCTION

Both developing and developed economies have increased their renewable energy use and production since the 1980's. There are four aspects to the emergence of renewable energy. The first is technological advancement, which has reduced the cost of investing in renewable energy installations (Apergis and Payne, 2010a; Apergis and Payne, 2010b; LuqmanAhmad and Bakhsh, 2019). The second aspect relates to governments regulations which provided supportive policy implications for renewable energy investments such as the establishment of credit easing and tax deductions for green energy by most governments, which in turn, has increased certificate and portfolio standards in renewable energy investments (Apergis and Payne, 2012; Asiedu et al., 2021). The third point has to do with climate change issues. It is argued that increased use of renewable energy reduces CO2 emissions, allowing renewable energy to mitigate the harmful effects of climate change (Bowden and Payne, 2009; Payne, 2009, Ali, 2021). Finally, the increase in the price of fossil fuels has encouraged the use of renewable energy (Gozgor, 2018). With these four factors in mind, renewable energy has the potential to bring long-term economic growth.

The role of renewable energy in bringing long-tern economic growth seems to be challenged, recently, by policy uncertainties and instabilities currently facing the global economy. One of such emerging uncertainties/instabilities is the economic policy uncertainty (EPU). The EPU is policy uncertainty associated with changes in government policy direction (e.g., monetary, or fiscal policy changes, tax regulations, etc.), which tends to cause delays in spending and investment by individuals and businesses until such uncertainty is resolved. Policy changes and uncertainty relating to energy consumptions have been argued to affect the concerned country's overall economic growth (Aizenman and Marion, 1993; Tiwari, 2011; Gulen and Ion, 2016). For example, uncertainty

about energy prices can force profit-optimizing firms to either suspend or quit investment decisions, which ultimately causes a decline in the aggregate output (Bernanke, 1983; Elder and Serletis, 2010). The empirical evidence further suggests that policy uncertainty about energy use, such as reducing CO2 emissions and fossil fuels, is harmful to investment in renewable energy, which can slow down the pace of economic growth (Adedoyin and Zakari, 2020).

Although the effect of the EPU on the number of variables has been widely explored in the literature (see, e.g., Pao and Fu, 2013; Solarin and Ozturk, 2015; Bhattacharya et al., 2016); however, there are very few papers for analyzing the effect of the EPU on energy indications, the only exceptions being Adedoyin and Zakari (2020) and Ulucak and Khan (2020). For instance, Adedoyin and Zakari (2020) attempt to provide evidence on the role of the EPU in the United Kingdom's energy consumption. By applying the Autoregressive Distributed Lag (ARDL) bound and Granger Causality tests, they show that the EPU matters to the United Kingdom most in the short run for two reasons. First, it reduces the growth of CO2 emissions. Second, with the prolonged use of the EPU, there will be a continuous rise in CO2 emissions. Using the same methodology (i.e., the ARDL bound test) for the case of the United States, Ulucak, and Khan, 2020) show that higher energy intensity contributes to pollution and that the EPU does not only adversely affects environmental quality but also strengthens the detrimental effect of energy intensity on CO2 emissions.

One major takeaway from these studies is their limited application to either the United Kingdom (Adedoyin and Zakari, 2020) or the United States (Ulucak and Khan, 2020). Besides, neither of these studies captures the effect of the EPU in the growth model. A broader understanding of how the EPU affects energy use across countries will provide a clearer insight into the global policy focus regarding uncertainty and energy use. Therefore, differing from existing studies, our paper examines the causal relationship between renewable energy consumption and economic growth in four countries: Brazil, Germany, Japan, and the United States by controlling for the effects of the EPU to capture the role of capabilities on renewable energy-growth nexus.

At this stage, we implement the recent Vector Autoregression (VAR)-based Granger-causality test of Rossi and Wang (2019). The authors argue that the traditional Granger causality test that assumes stationarity of variables is not reliable in modelling relationship between variables with high uncertainty/ instabilities as it can lead to misleading inferences. They extend (Rossi's, 2005) methodology and include robust versions of the Mean and Exponential Wald, Nyblom, and Andrew's Quasi-Likelihood Ratio tests. Rossi and Wang (2019) illustrate how to test the Granger causality to produce robust results in the presence of instabilities.

Our paper focuses on the cases of Brazil, Germany, Japan, and the United States because they have one of the world's major renewable energy matrices, with high renewable energy generation and use. A study by the International Renewable Energy Agency (IRENA) (2016) estimated that, by 2030, renewable energy would contribute to the overall global economic growth by values between 0.6 and 1.1%, amounting to $706 billion to $1.3 trillion. A further study estimate that renewable energy will be worth $ T2 trillion in 2050, with the United States, Germany, Japan, and Brazil, among other nations, experiencing the most of these increases (IRENA, 2018).

On the other hand, the United States is the biggest economy globally and one of the largest renewable energy consumers. Of the world's total primary energy consumption, the United States is the largest energy consumer country with a share of 16.2% annual growth rate and $CO_2$ emission of 14.5% as of 2019.[1] Japan, Germany, and Brazil had a share of 3.2, 2.3, and 2.1%, and the $CO_2$ emission rate of 3.3, 2 and 1.3%, respectively (British Petroleum (BP), 2020). While the electric power sector accounted for 56% of total renewable energy consumption in the United States in 2019, about 17% of the country's electricity generation was from renewable energy sources (U.S. Energy Information Administration, 2021). EIA 2021 estimates that renewable energy consumption will continue to rise in the country up to 2050.

Being a technology giant country, Japan has recently put forward new legislation in 2017 with plans to increase its reliance on and use renewable energy to 24% of its energy mix, representing more than double of its current production. In 2019, it was reported that Japan needs to import 90% of its energy. Due to the Fukushima disaster in 2011, the country's share from nuclear sources plummeted significantly from 30% to a production target of just 20% by 2030 (Casey, 2017). Consequently, Japan is now turning to other energy sources, including wind, solar, and tidal power, to reduce its reliance on foreign sources and further enhance innovation in its domestic energy sector.

In Germany, renewable energy is mainly based on solar, wind, and biomass. The country has one of the world's largest photovoltaic installed capacities, making it the world's major renewable energy economy with a powerful wind power capacity. German's share of renewable electricity reached 10% in 2005 from just 3.4% in 1990, and by 2019, it recorded 42.1% of gross electricity consumption (EIA, 2020). Recently, the German government is focusing particularly on the offshore wind farm, following the European Union (EU)'s mandatory energy plan requirement of reducing $CO_2$ emission by 20% before the end of 2020 and consumption of renewable energies of 20% of the total EU consumption.

Brazil is also another country with a solid renewable energy matrix. A large chunk of Brazil's renewable energy is from hydroelectric power plants. With the recently introducing public policies (e.g., the 10-years Energy Expansion Plan 2026) that aim to promote other non-traditional renewables (e.g., ethanol, made from sugarcane), the share of Brazil's renewable energy is now rising reasonably. Renewable sources supplied about 44% of the Brazilian energy consumption as of 2016 and 79.3% in 2018 (IEA Bioenergy, 2018).

[1]Primary energy here is composed of commercially traded fuels and modern renewables that are used to generate electricity (BP, 2020: 8).

To the best of our knowledge, this study is the first research in the empirical literature to examine the causal relationship between renewable energy and economic growth by controlling the role of economic policy uncertainty. For this purpose, we utilize the recent VAR-based Granger-causality test of Rossi and Wang (2019) in instabilities, which is another novelty of the study. We observe a bidirectional causal relationship between renewable energy and economic growth in Germany. There is also a significant causality from renewables to economic growth in Brazil and the United States. However, there is no significant causality between renewable energy and economic growth in Japan.

The remainder of the study is organized as follows. *Literature Review* provides a literature review of the renewable energy-growth nexus. *Data, Model, and Methodology* defines the data, sets the empirical model, and explains the Granger-causality test. *Empirical Results* discusses the empirical findings. *Conclusion* concludes.

## LITERATURE REVIEW

There are four views to explain the causality between economic growth and renewable energy. First, the "growth hypothesis" indicates that renewable consumption causes economic growth. Second, the "conservation hypothesis" states that economic growth causes renewable energy. Third, the "feedback hypothesis" suggests a bidirectional causal relationship between renewable energy and economic growth. Finally, the "neutrality hypothesis" proposes no causality between renewable energy and economic growth (Tiba and Omri, 2017).

There are previous papers to examine these hypotheses. For instance, Sadorsky (2009a) argues that energy use in all forms (such as industrial and residential, among others) is a primary determinant of economic growth and prosperity. As countries prosper, energy demand also increases. Sadorsky (2009b) also models the relationship between energy consumption and income for a panel of 18 emerging market economies from 1994 to 2003. The panel estimates confirm, yet again, that income per capita exert a statistically significant positive impact on per capita renewable energy consumption, thus providing evidence in support of the *conservation hypothesis*.

Using the panel dataset of 20 OECD countries, Apergis and Payne (2010a) investigate the causal link between renewable energy consumption and economic growth within a multivariate framework over 1985–2005. The Granger-causality test results show a two-way causal relationship between renewable energy consumption and economic growth in both the short-run and the long run. However, considering Eurasia's dependency on fossil fuels, Apergis and Payne (2010b) further extend their analyses to examine the validity of the feedback hypothesis for the panel of 13 countries within Eurasia, this time using a short data set covering the period of 1992–2007 within a multivariate framework. The heterogeneous panel cointegration test suggests a long-run equilibrium relationship between real GDP and renewable energy consumption, among other variables. Similarly, for a panel of six Central American countries over the period 1980–2006, Apergis and Payne (2011) find evidence of a bidirectional causal relationship between renewable energy consumption and economic growth in both the short-run and long run. A unidirectional causality from economic growth to renewable energy use is also found by Rahman and Velayutham (2020). Similar results also hold for the panel of 80 countries throughout 1999–2004 are obtained by Apergis and Payne (2012); thus, reiterating the validity of the feedback hypothesis that renewable energy consumption and economic growth are interdependently related.

Motivated by the role of renewable and non-renewable energy in Brazil's economic growth and how, in turn, economic growth experienced in Brazil played a crucial role the growth and development as well as the recent massive investment in the country's renewable energy sector, Pao and Fu (2013) attempted to obtain if there was evidence for the "feedback hypothesis" in Brazil throughout 1980–2010 by exploring the causal relationship between economic growth (real GDP) and four types of energy consumption: non-hydroelectric renewable energy consumption, total renewable energy consumption (TREC), non-renewable energy consumption, and the total primary energy consumption. Their study's long-run relationship indicates that a 1% increase in the TREC increases real GDP by 0.20%. An error correction model application further indicates a bidirectional causal relationship between the TREC and economic growth, among other findings. However, this outcome is fascinating because recently, development suggests that Brazil could change dramatically from a large oil importer to a self-sufficient energy economy. The pioneer works of Bhattacharya et al. (2016), Carmona et al. (2017), and Inglesi-Lotz (2016) also find support for the "feedback hypothesis". Similar results are obtained by Solarin and Ozturk (2015) for a panel of Latin American countries.

To conclude the literature review, various studies have analyzed the relationship between renewable energy consumption and economic growth in developing and developed countries. Even the most recent studies exist (e.g., Aydin, 2019; Ajmi and Inglesi-Lotz, 2020; Gao and Zhang, 2021) merely examine the causal relationship between renewable energy consumption and economic growth of some number of countries. None of these studies capture the effect of economic policy uncertainty in a growth model. Therefore, our paper contributes to this literature by introducing a new growth model that includes economic policy uncertainty on economic growth.

## DATA, MODEL, AND METHODOLOGY

Our sample covers the annual data in Brazil (1970–2019), Germany (1971–2019), Japan (1970–2019), and the United States (1965–2019). The begging date of the sample is based on the data availability. We use renewable energy

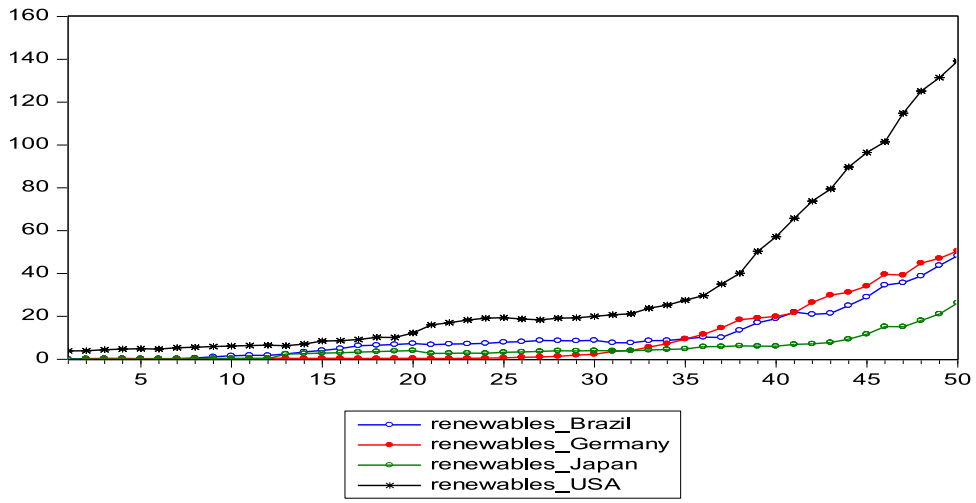



**FIGURE 1 |** Renewable energy consumption.

**TABLE 1 |** Traditional Granger causality test.

| Renewables → Growth | F-Statistics | *p*-Value |
|---|---|---|
| Brazil | 1.7565 | 0.1916 |
| Germany | 1.2371 | 0.2719 |
| Japan | 2.7058 | 0.1068 |
| United States | 0.4187 | 0.5205 |
| **Growth→Renewables** | | |
| Brazil | 0.0304 | 0.8623 |
| Germany | 0.3461 | 0.5593 |
| Japan | 0.0241 | 0.8774 |
| United States | 2.2196 | 0.1424 |

consumption (mtoe), and the data is obtained from British Petroleum Statistical Review. We also consider economic growth, based on logarithmic changes of per capita gross domestic product with the constant 2010 $ prices. We also use the World Uncertainty Index (WUI) to capture the role of economic policy uncertainty, and the data are obtained from https://worlduncertaintyindex.com/. A greater value of the WUI means that there is a greater level of economic policy uncertainty. At this stage, we test the following causal relationships:

$$\mathrm{GROWTH}_t = \alpha_0 + \alpha_1 RENEWABLES_{t-k} + \alpha_2 \mathrm{WUI}_{t-k} + \varepsilon_{t1} \quad (1)$$

$$RENEWABLES_t = \beta_0 + \beta_1 \mathrm{GROWTH}_{t-k} + \beta_2 \mathrm{WUI}_{t-k} + \varepsilon_{t2} \quad (2)$$

In **Eqs. 1**, **2**, $\mathrm{GROWTH}_t$ is economic growth, $RENEWABLES_t$ is renewable energy consumption, and $\mathrm{WUI}_{t-k}$ is the World Uncertainty Index. At this stage, we obtain four alternative test statistics: Exponential Wald (ExpW), Mean Wald (MeanW), Nyblom (Nybolm), and Quandt Likelihood Ratio (SupLR) to estimate the causal relationships. The lag selection is based on the Akaike Information Criterion (AIC). The VAR model is estimated with one lag. Here, all four test statistics must exceed the critical values of these tests and then we can conclude that there is a statistically significant causality between the related variables. Otherwise, there is no significant causal relationship. For details, refer to Rossi (2005) and Rossi and Wang (2019) and *gcrobustvar* Stata command.

## EMPIRICAL RESULTS

Following Gozgor (2016), we implement various unit root tests and confirm the renewable energy, growth, and the EPU series are stationary. **Figure 1** shows the renewable energy consumption for four countries. United States consumes the most of renewable energy among four countries. **Table 1** shows the results for traditional Granger Causality test of Granger (1969), and it indicates that no Causality relationships found between renewable energy consumption and economic growth for all countries at the 5% significant level.

Then we implement the Rossi-Wang causality test and report the results in **Table 2**. Regarding the causality from renewables to growth, we find that all test statistics (ExpW, MeanW, Nyblom, and SupLR) are statistically significant at the 1% level in Brazil, Germany, and the United States.

Regarding the causality from growth to renewables, we observe that all test statistics (ExpW, MeanW, Nyblom, and SupLR) are statistically significant at the 1% level in Brazil and Germany.

Therefore, we conclude a bidirectional causal relationship between renewable energy and the economic growth in Brazil and Germany. There is also a significant causality from renewables to economic growth in the United States. However, there is no significant causal relationship in Japan.

We check the robustness of the results. Specifically, we exclude outlier observations and rerun the causality analyses. The results

**TABLE 2 |** Results of Rossi-Wang causality test.

| Renewables → growth | ExpW | MeanW | Nyblom | SupLR |
|---|---|---|---|---|
| Brazil | 77.3306*** | 67.2236*** | 26.3675*** | 161.6903*** |
| Germany | 163.4795*** | 92.2268*** | 11.0431*** | 333.9459*** |
| Japan | 118.1134*** | 79.4802*** | 0.6625 | 243.2561*** |
| United States | 30.9121*** | 14.8006*** | 4.88341*** | 69.0515*** |
| **Growth→Renewables** | **ExpW** | **MeanW** | **Nyblom** | **SupLR** |
| Brazil | 31.6875*** | 4.76007 | 9.7408*** | 70.404*** |
| Germany | 27.9398*** | 17.0675*** | 12.6300*** | 62.8645*** |
| Japan | 39.8567*** | 27.8097*** | 2.7757 | 86.7362*** |
| United States | 86.5835*** | 26.2080*** | 5.1459 | 179.8434*** |

*Notes: Four test statistics are reported: Exponential Wald (ExpW), Mean Wald (MeanW), Nyblom (Nybolm) and Quandt Likelihood Ratio (SupLR) tests. Lag selection is based on the AIC. The VAR model is estimated with one lag.* ***p < 0.01.

align with the benchmark results, but we do not report them to save space.

## CONCLUSION

This paper examines the causal relationship between renewable energy consumption and economic growth in four countries: Brazil, Germany, Japan, and the United States. We control the impact of economic policy uncertainty on renewable energy-growth nexus. The findings from the VAR-based Granger-causality test of Rossi and Wang (2019) show a bidirectional causal relationship between renewable energy and economic growth in Brazil and Germany. There is also a significant causality from renewables to economic growth in the United States. However, there is no robust causality between renewable energy and economic growth in Japan.

Future studies can test for the relationship between different macroeconomic indicators and energy variables in other developed and developing economies. Finally, future studies can use disaggregated level renewable energy data (e.g., biomass, solar, and wind).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

ZL: Methodology, Writing-Original Draft Preparation, Supervision. LZ: Writing-Revision Draft Preparation. CL: Investigation, Software, Visualization, Writing-Original Draft Preparation. AI: Data Curation, Writing-Original Draft Preparation. XZ: Supervision, Editing, Writing-Original Draft Preparation.

## ACKNOWLEDGMENTS

The authors acknowledge the funding from the Philosophy and Social Science Fund of Tianjin City, China. Project #: TJYJ20-012 ("Prompting the Market Power of Tianjin City's E-commerce Firms in Belt and Road Countries: A Home Market Effect Approach").

## REFERENCES

Adedoyin, F. F., and Zakari, A. (2020). Energy Consumption, Economic Expansion, and CO2 Emission in the UK: The Role of Economic Policy Uncertainty. *Sci. Total Environ.* 738, 140014. doi:10.1016/j.scitotenv.2020.140014

Aizenman, J., and Marion, N. P. (1993). Policy Uncertainty, Persistence and Growth. *Rev. Int. Econ.* 1 (2), 145–163. doi:10.1111/j.1467-9396.1993.tb00012.x

Ajmi, A. N., and Inglesi-Lotz, R. (2020). Biomass Energy Consumption and Economic Growth Nexus in OECD Countries: A Panel Analysis. *Renew. Energ.* 162, 1649–1654. doi:10.1016/j.renene.2020.10.002

Apergis, N., and Payne, J. E. (2012). Renewable and Non-renewable Energy Consumption-Growth Nexus: Evidence from a Panel Error Correction Model. *Energ. Econ.* 34 (3), 733–738. doi:10.1016/j.eneco.2011.04.007

Apergis, N., and Payne, J. E. (2010a). Renewable Energy Consumption and Economic Growth: Evidence from a Panel of OECD Countries. *Energy Policy* 38 (1), 656–660. doi:10.1016/j.enpol.2009.09.002

Apergis, N., and Payne, J. E. (2010b). Renewable Energy Consumption and Growth in Eurasia. *Energ. Econ.* 32 (6), 1392–1397. doi:10.1016/j.eneco.2010.06.001

Apergis, N., and Payne, J. E. (2011). The Renewable Energy Consumption-Growth Nexus in Central America. *Appl. Energ.* 88 (1), 343–347. doi:10.1016/j.apenergy.2010.07.013

Asiedu, B. A., Hassan, A. A., and Bein, M. A. (2021). Renewable Energy, Non-renewable Energy, and Economic Growth: Evidence from 26 European Countries. *Environ. Sci. Pollut. Res.* 28 (9), 11119–11128. doi:10.1007/s11356-020-11186-0

Aydin, M. (2019). The Effect of Biomass Energy Consumption on Economic Growth in BRICS Countries: A Country-specific Panel Data Analysis. *Renew. Energ.* 138, 620–627. doi:10.1016/j.renene.2019.02.001

Bernanke, B. S. (1983). Irreversibility, Uncertainty, and Cyclical Investment. *Q. J. Econ.* 98 (1), 85–106. doi:10.2307/1885568

Bhattacharya, M., Paramati, S. R., Ozturk, I., and Bhattacharya, S. (2016). The Effect of Renewable Energy Consumption on Economic Growth: Evidence from Top 38 Countries. *Appl. Energ.* 162, 733–741. doi:10.1016/j.apenergy.2015.10.104

Bowden, N., and Payne, J. E. (2009). The Causal Relationship between U.S. Energy Consumption and Real Output: A Disaggregated Analysis. *J. Pol. Model.* 31 (2), 180–188. doi:10.1016/j.jpolmod.2008.09.001

British Petroleum (BP) (2020). *BP Statistical Review of World Energy*. June 2020. London: BP.

Carmona, M., Congregado, E., Feria, J., and Iglesias, J. (2017). The Energy-Growth Nexus Reconsidered: Persistence and Causality. *Renew. Sust. Energ. Rev.* 71, 342–347. doi:10.1016/j.rser.2016.12.060

Casey, J. P. (2017). *New Laws and New Targets: Renewable Power in Japan*. London: Power Technology.

Elder, J., and Serletis, A. (2010). Oil Price Uncertainty. *J. Money, Credit Banking* 42 (6), 1137–1159. doi:10.1111/j.1538-4616.2010.00323.x

Energy Information Administration (2020). *Energy Policy Review: Germany 2020*. Washington, DC: EIA.

Gao, J., and Zhang, L. (2021). Does Biomass Energy Consumption Mitigate CO2 Emissions? the Role of Economic Growth and Urbanization: Evidence from Developing Asia. *J. Asia Pac. Economy* 26 (1), 96–115. doi:10.1080/13547860.2020.1717902

Gozgor, G. (2018). A New Approach to the Renewable Energy-Growth Nexus: Evidence from the USA. *Environ. Sci. Pollut. Res.* 25 (17), 16590–16600. doi:10.1007/s11356-018-1858-9

Gozgor, G. (2016). Are Shocks to Renewable Energy Consumption Permanent or Transitory? an Empirical Investigation for Brazil, China, and India. *Renew. Sust. Energ. Rev.* 66, 913–919. doi:10.1016/j.rser.2016.08.055

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37, 424–438. doi:10.2307/1912791

Gulen, H., and Ion, M. (2016). Policy Uncertainty and Corporate Investment. *Rev. Financial Stud.* 29 (3), 523–564.

Inglesi-Lotz, R. (2016). The Impact of Renewable Energy Consumption to Economic Growth: A Panel Data Application. *Energ. Econ.* 53, 58–63. doi:10.1016/j.eneco.2015.01.003

International Energy Agency (2018). *Country Reports: Brazil – 2018 Update*. Paris: IEA.

International Energy Agency (2020). *Energy Policy Review: Germany 2020*. Paris: IEA.

International Renewable Energy Agency (2018). *Global Energy Transformation: A Road Map to 2050*. Abu Dhabi: IRENA.

International Renewable Energy Agency (2016). *Renewable Energy Benefits: Measuring the Economics*. Abu Dhabi: IRENA.

LuqmanAhmad, M. N., Ahmad, N., and Bakhsh, K. (2019). Nuclear Energy, Renewable Energy and Economic Growth in Pakistan: Evidence from Non-linear Autoregressive Distributed Lag Model. *Renew. Energ.* 139, 1299–1309. doi:10.1016/j.renene.2019.03.008

Pao, H.-T., and Fu, H.-C. (2013). Renewable Energy, Non-renewable Energy and Economic Growth in Brazil. *Renew. Sust. Energ. Rev.* 25, 381–392. doi:10.1016/j.rser.2013.05.004

Payne, J. E. (2009). On the Dynamics of Energy Consumption and Output in the US. *Appl. Energ.* 86 (4), 575–577. doi:10.1016/j.apenergy.2008.07.003

Rahman, M. M., and Velayutham, E. (2020). Renewable and Non-renewable Energy Consumption-Economic Growth Nexus: New Evidence from South Asia. *Renew. Energ.* 147, 399–408. doi:10.1016/j.renene.2019.09.007

Rossi, B. (2005). Optimal Tests for Nested Model Selection with Underlying Parameter Instability. *Econometric Theor.* 21 (5), 962–990. doi:10.1017/s0266466605050486

Rossi, B., and Wang, Y. (2019). Vector Autoregressive-Based Granger Causality Test in the Presence of Instabilities. *Stata J.* 19 (4), 883–899. doi:10.1177/1536867x19893631

Sadorsky, P. (2009b). Renewable Energy Consumption and Income in Emerging Economies. *Energy Policy* 37 (10), 4021–4028. doi:10.1016/j.enpol.2009.05.003

Sadorsky, P. (2009a). Renewable Energy Consumption, CO2 Emissions and Oil Prices in the G7 Countries. *Energ. Econ.* 31 (3), 456–462. doi:10.1016/j.eneco.2008.12.010

Sayigh, A. (2021). Up-date: Renewable Energy and Climate Change. *Renew. Energ. Environ. Sustain.* 6, 13. doi:10.1051/rees/2021004

Solarin, S. A., and Ozturk, I. (2015). On the Causal Dynamics between Hydroelectricity Consumption and Economic Growth in Latin America Countries. *Renew. Sust. Energ. Rev.* 52, 1857–1868. doi:10.1016/j.rser.2015.08.003

Tiba, S., and Omri, A. (2017). Literature Survey on the Relationships between Energy, Environment and Economic Growth. *Renew. Sust. Energ. Rev.* 69, 1129–1146. doi:10.1016/j.rser.2016.09.113

Tiwari, A. K. (2011). A Structural VAR Analysis of Renewable Energy Consumption, Real GDP and $CO_2$ Emissions: Evidence from India. *Econ. Bull.* 31 (2), 1793–1806.

Ulucak, R., and Khan, S. U. D. (2020). Relationship between Energy Intensity and $CO_2$ Emissions: Does Economic Policy Matter? *Sust. Dev.* 28 (5), 1457–1464. doi:10.1002/sd.2041

U.S. Energy Information Administration (2021). *Renewable Energy Explained*. Washington, DC: EIA.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Lu, Zhu, Lau, Isah and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 29 September 2021
doi: 10.3389/fenrg.2021.729690

# Stability Analysis of the World Energy Trade Structure by Multiscale Embedding

*Leyang Xue*[1,2], *Feier Chen*[1]*, *Guiyuan Fu*[3], *Qiliang Xia*[1] *and Luhui Du*[4]

[1]State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Engineering, Shanghai Jiao Tong University, Shanghai, China, [2]Grainger College of Engineering,University of Illinois Urbana-Champaign, Champaign, IL, United States, [3]Institute of Fintech, Shanghai University of Finance and Economics, Shanghai, China, [4]Shanghai Merchant Ship Design and Research Institute, CSSC, Shanghai, China

This study investigates the dynamic trading network structure of the international crude oil and gas market from year 2012 to 2017. We employed the dynamical similarity analysis at different time scales by inducing a multiscale embedding for dimensionality reduction. This analysis quantifies the effect of a global event on the dependencies and correlation stability at both the country and world level, which covers the top 53 countries. The response of China's trading structure toward events after the unexpected 2014 price drop is compared with other major traders. China, as the world's largest importing country, lacks strong stability under global events and could be greatly affected by a supply shortage, especially in the gas market. The trend of multi-polarization on the market share gives a chance for China to construct closer relationships with more stable exporters and join in the trade loop of major countries to improve its position in the energy trading networks. The hidden features of trade correlation may provide a deeper understanding of the robustness of relationship and risk resistance.

**Keywords: dynamical network, complex networks, energy market, heat kernel embedding, multiscale embedding**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Chau Yuen,*
*Singapore University of Technology and Design, Singapore*
*Vidya C. T,*
*Centre for Economic and Social Studies (CESS), India*

***Correspondence:**
*Feier Chen*
*chenfeier@sjtu.edu.cn*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *23 June 2021*
***Accepted:*** *09 August 2021*
***Published:*** *29 September 2021*

***Citation:***
*Xue L, Chen F, Fu G, Xia Q and Du L (2021) Stability Analysis of the World Energy Trade Structure by Multiscale Embedding. Front. Energy Res. 9:729690. doi: 10.3389/fenrg.2021.729690*

## 1 INTRODUCTION

It has long been the interest of academy and industry to research on the world market of energy, especially the crude oil and natural gas. Literature has investigated the economic impact of oil price shocks, including the response of economic growth, inflation, and stock markets in oil-importing and -exporting countries (Zhao and Chen, 2014; An et al., 2018; Gong and Lin, 2018; Youssef and Mokni, 2019). Recent developments in the energy market have discussed the role of oil prices in determining trade balances, international risk sharing, and long-standing country policy (Filis and Chatziantoniou, 2013; Aydoğan et al., 2017; Youssef and Mokni, 2019). Many studies focus on analyzing how the rising and falling of international oil prices affect the relationship between supply and demand in the oil market (Axon and Darton, 2021; Esfahani et al., 2021). In the past decades, significant changes have occurred in the world market. The U.S. declines at a rate of 2.9% per year in terms of imports, while grows at a yearly rate of 15.1% in export. China's oil consumption reached 13% of the world total volume in year 2017, more than 67% depending on import (BP, 2019). Consequently, literature investigating the dynamic interconnections among energy markets has emerged over the last two decades for the changing world energy market (Badel and McGillicuddy, 2015; Baumeister and Kilian, 2016a; Xue et al., 2019).

However, the energy price fluctuations affected by the global or regional events are very complex and have further motivated the evolution of the world energy trade network research by comparing

and measuring the structural changes over time. For example, Hamilton (2010) reviews the connection between historical events and oil price shocks. He states that the crude oil price jump and drop in the 2008 global financial crisis present very high amplitudes in the absence of critical geopolitical events. Meanwhile, geopolitical events, such as the civil war in Syria and chronic crises in oil-producing countries have only a minor impact on oil prices. The dynamic effects of different shock types on the price of oil are partly solved by researchers (AlbertJeong and Barabási, 2000; Delvenne et al., 2015; An et al., 2018). Understanding the recent evolution of the price of oil is important in assessing the macroeconomic outlook (Baumeister and Kilian, 2016b). However, the response of energy-exporting and -importing countries to heterogeneous events varies a lot (Blanchard and Gali, 2007; Kilian et al., 2009; Inoue and Kilian, 2016; Gong and Lin, 2018). They find that the overall effect of oil demand and supply shocks on the trade balance of oil importers (and exporters) depends critically on the response of the non-oil trade balance (Kilian, 2009). Countries' energy trade relationships are significant in analyzing trade network patterns (Zhao and Chen, 2014; Kremer et al., 2019; Mealy et al., 2019). A national energy trade network would affect a country's GDP and its centrality on international trade significantly (Xi et al., 2019). The identification and assessment of energy security risks are also important for the behavior of exporters and importers that changed the energy industry. So the potential effect of shock estimation for a specific country and the new insights on dynamic trade relationship together remain open and need a deeper understanding.

Given this context, the energy trade relationships of big countries are of growing interest in analyzing trade network patterns. Trade dependency including trade concentration, import or export dependency, and trade volume is useful to analyze the relationship between nations. There are two different types of features, intrinsic and induced (Li et al., 2018; Comin et al., 2020). Induced features can be used to characterize the immediate neighborhood of a node and the scaled features (Frey et al., 2019; Fu et al., 2019). With the methodology development and data availability, hidden information on international trade has been deployed and sheds light on the opportunities and threats of trade dependencies (Stavroglou et al., 2019; Wang et al., 2019). However, it is hard to do a system-level analysis for its limitation in a comparison between different countries (Yuan et al., 2020; Ren et al., 2021). The intrinsic features of a node cannot be obtained from the connectivity of the network. One special feature of a node is the spacial position of a node, which is frequently represented as a vector. So we will explore the nodes in networks by distance in a feature space and try to discuss this topic both from the feature similarity of countries and structural properties of the world market over time.

To study complex systems composed of discrete elements with observed connections, complex network theory is an applicable framework (De Domenico et al., 2015; Dehmamy et al., 2018; Tushar et al., 2020; Tushar et al., 2021). The world energy market comprising a large number of interacting dynamical nodes display a rich repertoire of behaviors across different time and length scale (Hamilton, 2010; Aydoğan et al., 2017; Lang and Auer, 2019; Comin et al., 2020). However, it is unclear how such a detailed description of all elements could be interpreted (Newman, 2018; Tushar et al., 2018; Cui et al., 2019) or it is necessary to understand the phenomenon (Casadiego et al., 2017). Accordingly, many studies try to reduce the complexity of the system more only with fewer, aggregated variables to explain the behavior of interest (Pecora et al., 2014; Sorrentino et al., 2016). Davies (1989) proposed heat kernels and spectral theory to investigate spectral properties and obtain pointwise bounds on eigenfunctions by studying the heat kernel. Embedding techniques, which define the network and its nodes in a metric vector-space and allow us to use a plethora of computational techniques, have emerged in analyzing network data (Grover and Leskovec, 2016). However, it is not clear how to define an appropriate diffusion for the networks with more general dynamics than diffusion or sighed and directed edges (Newman and Clauset, 2016). Consequently, metric embedding involves finding an appropriate feature space in which the distance between objects provides a good description of the known dissimilarity between them (Abraham et al., 2006; Borg and Groenen, 2010). By extracting lower-dimensional or coarse-grained descriptions at the aggregated level of markets, an effective subspace may be sufficient for the dynamical process (Ravasz et al., 2002; Battiston et al., 2014). However, many economic characters, such as cycles and higher-order dynamical couplings cannot be captured (Banisch and Djurdjevac Conrad, 2015). A dimensionality reduction of the system in continuous space can be exploited to uncover dynamical modules in the system of the energy trade market (Lambiotte et al., 2014; Schaub, 2014; Rozemberczki et al., 2019). Recently, dynamic topology of complex networks and dimensionality reduction of multiscale systems are hot topics with the development of network theory and embedding techniques (Song et al., 2016; Schaub et al., 2019; Comin et al., 2020). It provides us with a dynamics-based and geometric representation of the system, associating similarity and distance measures. Given this context, to investigate the underlining information of the global energy market, different scales in a long time window should be considered. Here, we express the dynamical energy trading system as a function of spatial location, exploit this vector space representation, and analyze the associated similarity and dual distance measures.

The main contribution of our research reflects in the following dimensions. Different from a multivariate statistical analysis of trade volume, we focus on the structural changes and dynamic processes of the global crude oil and gas market from a macro-view over time by investigating the peak and tail areas. Not only considering the connectivity between countries, we also look into intrinsic and induced features of the individual countries in the world trade system. Firstly, to observe the evolution of the world trade network from a macro-level perspective, we construct a low-dimensional embedding of the system to provide a dimensionality reduction of the world energy trading system in continuous space. This enables us to ignore the differences in various aspects of countries and master the structural features and functions of the trade networks. Secondly, to uncover

dynamical modules in the trading system, we exploit the dimensionality reduction ideas from a meso-level perspective. So importing or exporting countries that act approximately as a dynamical unit over a given time scale could be represented by groups of nodes. Thirdly, from a micro-lever view, we explore the important nodes in the network and quantify the feature of different countries, such as the United States, China, and Russia, to analyze energy security, including the changes in their energy trade structure over time, their stability under the events effect, and ability to resist risks. The dynamic structure analysis could track the movement of major traders and give insight into potential relationship development. Events' effect on the network and nodes help in a deeper understanding of the robustness of relationship and risk resistance.

The remainder of the article is organized as follows. **Section 2** describes the methodology used in this study. **Section 3** introduces the data. **Section 4** reports the results and discusses the balances of the energy market driven by events. **Section 5** contains our conclusions.

## 2 METHODOLOGY

Our objective is to capture the dynamics (changes) in trading networks by examining the relationships after global or local events. When events occur, the impact will diffuse along the trading cross correlation networks. In terms of trading networks topology, this can be observed by changes in trade flow in respect of time.

We note that simple statistics on trade volume and node degree can reveal market characteristics, such as power-law (Gabaix, 2016). But it is not enough for a deeper insight for processes observation. Naturally, the world market has geographical relations between countries (Hadfield, 2008; Hu and Cheng, 2008). When geo-location is under consideration, the market becomes more complex. It is not purely economics, rather politics, security, and national relationships. So it is better to embed geographical relations into the model of the world market. It is reasonable to employ heat kernel embedding (Davies, 1989) as a fundamental framework.

The world energy trading market contains more features than node degree and adjacency relationships, more features could be added to the model to find more hidden information. Here, we consider trade volume as another dominant factor in the world energy trading market, where countries with major trading volumes have a dominant impact on energy economics. As for geographical factors, the Laplacian matrix can take geographical relations into consideration. The degree and adjacency could capture the pair-wise relation between countries. The common fact we are aware of is that the energy transportation cost is proportional to the distance (pipeline construction and freight shipping). It can be a hidden factor implied by world trading relations. Hence, it is reasonable to employ the Laplacian matrix in our network modeling.

Based on the above statement, we model and analyze the world energy trade network by an extended heat kernel embedding. The world energy trading network can be viewed as a directed graph $G = \langle V, L \rangle$, where the number of vertices $|V|$ stands for the number of countries in the world, and the number of edges $|L|$ represents the trade flow. If vertex $u$ is an outgoing edge toward vertex $v$, then it represents an export from $u$ or import by $v$. The degree of a vertex represents the numbers of countries which it trades with, and the vertex strength is the trade volume. The edge weight is the trade volume in a single relationship.

Hence, we construct a dynamical network modeling of the data according to Schaub et al. (2019), which is an extended model of Davies (1989).

$$\dot{x} = \mathcal{A}x + \mathcal{B}u \quad \mathcal{A} \in \mathbb{R}^{m\times m}, \mathcal{B} \in \mathbb{R}^{m\times p}, \tag{1}$$

$$y = \mathcal{C}x \quad \mathcal{C} \in \mathbb{R}^{n\times m}, \tag{2}$$

where $x \in \mathbb{R}^m, y \in \mathbb{R}^n, u \in \mathbb{R}^p$ are the state, the observed state, and the input vectors for the dynamical network. Since we are going to work on the given data which covers all months and countries, we may consider all states we plan to analyze as they are observed. For **Eq. 2**, we will have $n = m$ and $y = x$, where $\mathcal{C}$ is an identical matrix. Meanwhile, considering the independence of every country, we may place the individual control signal on each country, leaving $\mathcal{B}$ an identical matrix. Finally, $\mathcal{A}$ is to be determined by a specific model we would like to achieve, where $n = m = p$. A unique aspect of dynamical embedding is that it captures the evolution of the network. The model assumes that there is an initial impulse to every node in the network. From the time scale of the dataset, **u** contains the financial crisis as a major background. After the major events occur, we observe the impulse response in each year. **u** can be quantified as changes in trade volume and cross correlations. The similarity of impulse response between each country may reveal the underlying connection and stability under major events.

The similarity between two nodes in a network can be regarded as a scalar value indicating how close the two nodes are according to some criterion. Here, we use node similarity to measure the level of dependence between variables associated to countries. This dependence can be across time, over the feature space, or both.

From **Eq. 1** and **Eq. 2**, the dynamical network could be regard as a system where each node $i$ has an impulse at time $t = 0$. We can have a mapping $i \mapsto y_i(t), \quad y_i(t) \in \mathbb{R}^n$, in which we map each node to its impulse response at time $t$. Conventionally, the node similarity can be defined as the similarity measured on an impulse response. As a result, similarity between node i and node j can be defined as

$$\psi_{ij}(t) = \langle y_i(t), y_j(t) \rangle_{\mathcal{W}} = y_i(t)^\top \mathcal{W} y_j(t), \tag{3}$$

where $\mathcal{W}$ is a positive (semi-)definite weighting matrix contains the features that we plan to enhance. Collecting the (zero state) impulse responses $y_i$ into the matrix $Y(t) = [y_1, \ldots, y_n] = \mathcal{C}exp(\mathcal{A}t)\mathcal{B} = exp(\mathcal{A}t)$, we could calculate the inner product matrix as follows:

$$\Psi(t) = Y^\top \mathcal{W} Y = exp(\mathcal{A}t)^\top \mathcal{W} exp(\mathcal{A}t). \tag{4}$$

We further employ a dynamical network modeling of the data from the general framework of multi-scale embedding (Schaub

**TABLE 1 |** Top 10 trade country by subject, for each category ranking, is presented in a descending order.

| Natural Gas Exporter | Crude oil exporter |
|---|---|
| Qatar (QAT) | Saudi Arabia (SAU) |
| Russia (RUS) | Russia (RUS) |
| Australia (AUS) | Iraq (IRQ) |
| Saudi Arabia (SAU) | Canada (CAN) |
| Malaysia (MYS) | United Arab Emirates (UAE) |
| Canada (CAN) | Kuwait (KUW) |
| Mexico (MEX) | Iran (IRI) |
| Norway (NOR) | Venezuela (VEN) |
| Nigeria (NGA) | Nigeria (NGA) |
| Algeria (DZA) | Angola (ANG) |
| **Natural Gas Importer** | **Crude Oil Importer** |
| Japan (JPN) | China (CHN) |
| China (CHN) | United Stated (USA) |
| Rep. of Korea (KOR) | India (IND) |
| France (FRA) | Japan (JPN) |
| Belgium (BEL) | Rep. of Korea (KOR) |
| Germany (DEU) | Germany (DEU) |
| Spain (ESP) | New Zealand (NZL) |
| India (IND) | Italy (ITA) |
| Italy (ITA) | Spain (ESP) |
| United Kingdom (GBR) | France (FRA) |

et al., 2019) to include trade volume as a factor. The resulting formula for modified heat kernel embedding is as follow:

$$\dot{x} = L_s x + \mathbf{u}, \tag{5}$$

$$\Psi(t) = \exp(L_s t)^{\top} \left(\mathbb{I} + SS^{\top}\right) \exp(L_s t). \tag{6}$$

Here, $L_s$ is the normalized signed Laplacian matrix. $S$ represents the strength of each node in the network, where we subtract the sum of exports by the sum of imports. The weighting matrix with form of $\mathbb{I} + SS^{\top}$ is not necessarily positive-(semi) definite. We can approximate a positive semi-definite form by minimizing the error under Frobenius norm. The values in the weighting matrix can be both positive and negative. Positive weighting represents bigger export than import, while the negative value represents bigger import than export. After the calculation of **Eq. 6**, we may encounter similarity values from the interval of [0,1]. This is due to the scaling of weighting matrix on import/export relationships.

We aim to map the result of embedding to be more understandable. A two-dimensional Euclidean space is most commonly used to present the result of embedding. From the two-dimensional space, we can point out the correlation between the decomposition value and its meaning in trading network more easily. We consider the spectral decomposition of $\Psi(t)$ into its eigenvectors $\mathbf{v}_1(t)$, $\mathbf{v}_2(t)$, ..., $\mathbf{v}_n(t)$ with associated eigenvalues $\mu_1(t) \geq \mu_2(t) \geq \ldots \geq \mu_n(t)$. Then we define the mapping $i \mapsto \phi_i(t)$ as follows:

$$\phi_i(t) = [\sqrt{\mu_1}\, v_{1,i}, \sqrt{\mu_2}\, v_{2,i}, \ldots, \sqrt{\mu_n}\, v_{n,i}]^{\top}. \tag{7}$$

Particularly, we are mostly interested in the first two components of this decomposition, namely, $\phi_{i,1}(t) = \sqrt{\mu_1} v_{1,i}$ and $\phi_{i,2}(t) = \sqrt{\mu_2} v_{2,i}$. These two components are the most significant factors in the result of embedding. By analyzing these two components, we may easily tell the dynamics of the network and the impact from major events. This also provides us with a concise plot of the behavior of the network and its impulse response at time $t$. The first value of the decomposition can represent the ordering in a similarity value under the reference of the largest entity (in our case is the whole world), while the second component is related to the import and export ratio.

**TABLE 2 |** Value distribution of different country type.

| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|
| Minor Country Pairs (mm-Pairs) | | | | | | |
| Tail count | 6,162 | 5,430 | 5,206 | 7,094 | 6,776 | 6,458 |
| Peak count | 19,598 | 14,030 | 19,286 | 18,666 | 17,404 | 20,602 |
| Major-Minor Pairs (Mm-Pairs) | | | | | | |
| Tail count | 3,476 | 3,444 | 3,358 | 4,080 | 3,982 | 4,050 |
| Peak count | 5,218 | 4,116 | 5,120 | 4,614 | 4,442 | 4,860 |
| Major Country Pairs (MM-Pairs) | | | | | | |
| Tail count | 462 | 506 | 506 | 552 | 552 | 600 |
| Peak count | 240 | 196 | 196 | 150 | 150 | 102 |

## 3 DATA DESCRIPTION

In this study, we mainly focus on the import and export annual data of natural gas and crude oil from the year 2012 to 2017. We download detailed monthly data from the UN Comtrade database (UN, 2019) for this research. Particularly for the goods we investigate, the commodity numbers are 271111 (petroleum gases and other gaseous hydrocarbons; liquefied natural gas) and 271121 (petroleum gases and other gaseous hydrocarbons; in the gaseous state, natural gas) for natural gas and 2,709 (petroleum oils and oils obtained from bituminous minerals; crude) and 270900 (oils; petroleum oils and oils obtained from bituminous minerals, crude) for crude oil.

To be brief in the following discussion, we introduce the ISO 3166-1 alpha-3 abbreviation for country names (ISO, 2019). We are going to use the abbreviations in the following discussion and in labeling the figures. To focus on the subject of discussion, we select a major trader according to the sum of the trade value across all data. In **Table 1**, we list the resulting top importer and exporter of natural gas and crude oil in full names and abbreviations.

Due to the limited functionality of the UN database and its limitation in the download size, it is impossible to download full data for all countries with man power, where full data need 60 thousands times of downloads. We first use a Bash script to crawl data from the UN database with simple Linux command line tools. Since full data is split into 60 thousands files after crawl, we combine them into one dataset and aggregate data by each month, reporter, partner, and commodity number. To avoid numeric underflow or overflow in latter computation, we

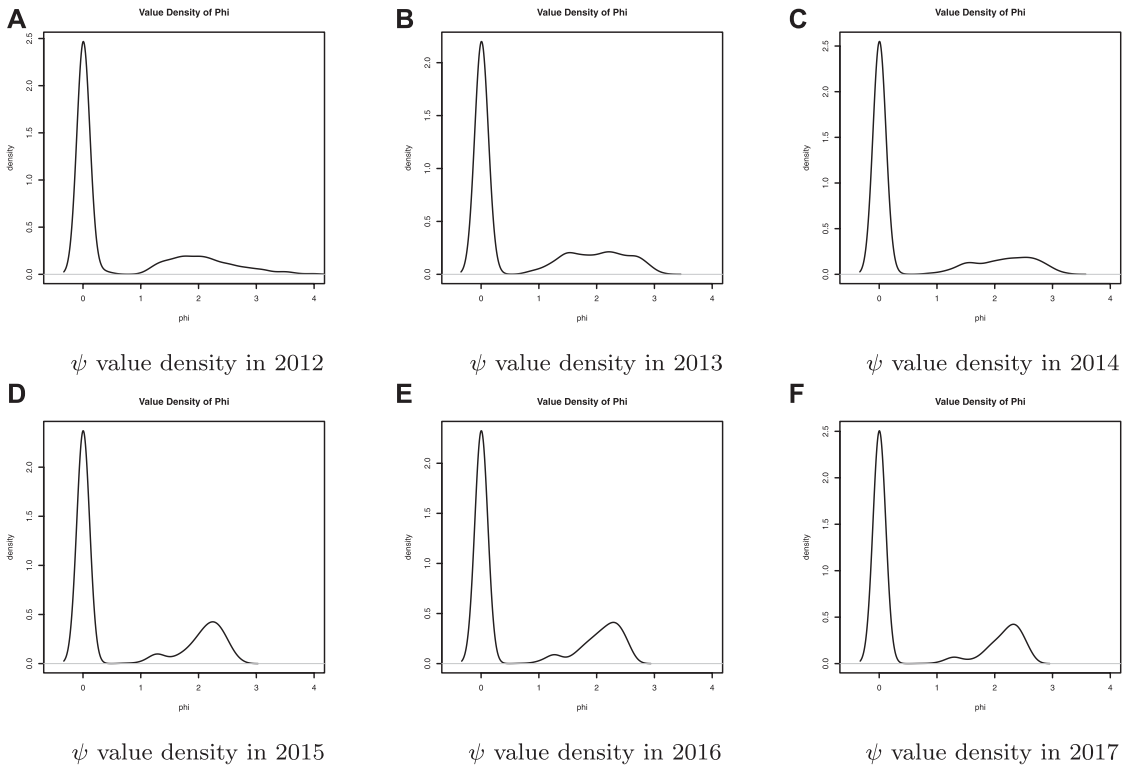



**FIGURE 1 |** $\psi$ value density from 2012 to 2017.

normalized the dataset so that its trade volume lies between 0 and 1. Since the trade volume is highly skewed among the world, we only take the top 53 countries with the largest overall trade volume, which produce around 800 edges and major hubs can be observed. As we plan to measure trade relation, we set weight of the graph to normalized trade volume.

Since the model evaluates the network impulse response after an initial input, the time unit of $t$ must fit the time unit of the dataset. As we obtain monthly data from the database, the time unit in our modeling is month. During our analysis, we construct a trading network each year. So the impact of major events can be implied by the change of trading volume and network structures during the observed years. We analyze the impulse response after 8 months of that year since the impact of most recent large event—financial crisis has reached its peak after about 8 months. We may assume impact of events can be observed after 8 months after initial impulse.

# 4 RESULTS AND DISCUSSION

In this section, we first explore the characteristics and statistics of the similarity matrix $\psi$, deriving from which we further investigate the networks embedding. We analyze the overall behavior of density and the dynamical structure of the market over the years by investigating the peak and tail area. The peak area refers to the area around zero, while the tail area is the small area on the right. To have a close look at the structure and behavior of major traders listed in **Table 1**, we visualize and analyze the two major components of network embedding.

## 4.1 Dynamical Structure of the Energy Market

Here, we investigate the similarity matrix derived from **Eq. 6**. In our setting, this similarity depends on number of countries to trade with and the import or export volume. Unlike the conventional similarity value, the similarity value after scaling exceeds the range of [0,1]. A node pair with either export or import dominant characteristics produces positive scaling, while a node pair with different characteristics produces negative scaling. The weighting matrix enhances the similarity where the hub node is involved and degrades the similarity when countries with small trading volume are involved. The density of these values can help us to uncover the pattern of trading pairs. There are three major categories we would like to examine, namely, pair among countries in **Table 1** (MM-Pair), pair among countries not in **Table 1** (mm-Pair), and pair among countries in **Table 1** and not in that table (Mm-Pair). According to the methodology above, we expect that majority of the similarity of mm-Pair lies around zero while majority MM-Pair in the tail area.

A common pattern can be observed in the density plot from 2012 to 2017 in **Figure 1** and **Table 2**. Each density plot has a high density around zero after the value is standardized, which means that the impulse response from the world around are quite

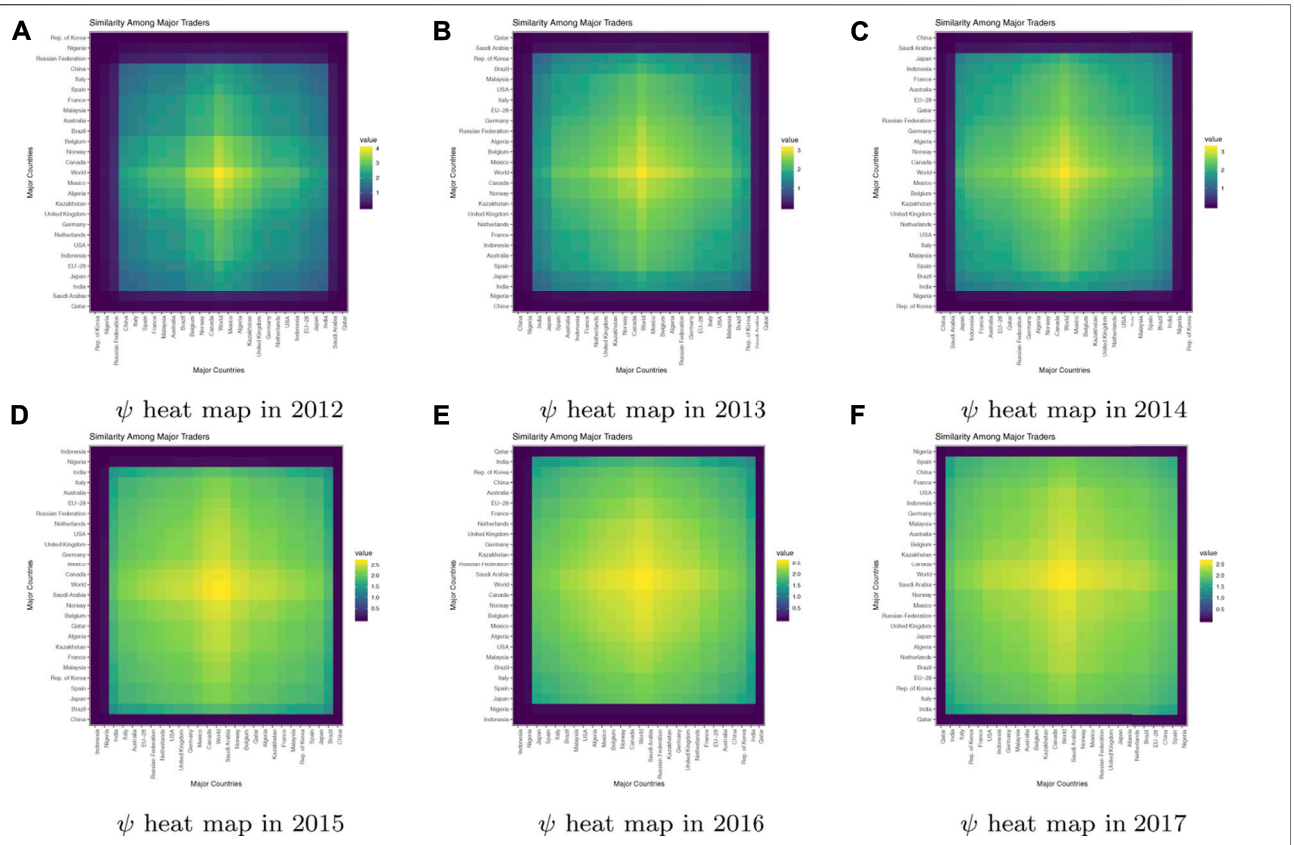



**FIGURE 2 |** $\psi$ heat map from year 2012 to 2017.

different from each other. This phenomenon is expected since the major importers and exporters of energy are very concentrated in a few countries among approximately 200 countries. Rest countries, except for the major traders, have weaker trading relations with each other and can be more easily disturbed by an external impulse (major events). These countries are connected to major traders and regional exporters and importers. With the trading network, there surely exist some nodes with similar impulse response, such as smaller countries next to a major exporter or major traders share the same market. Our research results are in accordance with An et al. (2018), whose study show that countries such as the United States and the United Kingdom have higher ability to control the relationships among other nodes.

We can also observe a trend of concentration in the tail area of the density. The tail area is relatively flat around 2012 to 2014 (**Figures 1A–C**), while there are clear gathering peaks around 2015 to 2017 (**Figures 1D–F**). We examine three major categories of similarity during different years. The similarity of MM-Pairs is strengthened throughout the years, as more values lies in tail area. Meanwhile, mm-Pairs and Mm-pairs are rather unstable. The major proportion of mm-Pairs lie in the peak area and Mm-Pairs are roughly equal in the two areas. The major importers and exporters are acting more and more similar on both trade volume and the relationship with other minor countries, which is also stated in (Filis and Chatziantoniou, 2013; Baumeister and Kilian, 2016b). The Mm-pairs similarity has a little growth, which implies that there's minor hub forming within countries of small trade volume. It could be the emerging of new energy exporters or the increasing energy consumption of developing countries.

## 4.2 Movement of Major Traders

By plotting the heat map of similarity among major traders in **Figure 2**, we can clearly observe a trend of concentration in the value. Here, we use accumulated data of the whole world as a reference point. So the world will always have the greatest similarity value to itself. If a country has a larger similarity value to the world, it is more likely to follow the global trend under events. We can see that these countries are either leader in great economic entities or countries geographically nearby. Notably in **Figure 2**, Canada and Mexico generally have a large similarity value with the world. These two countries both have simple energy trading relationships with several major traders, such as China and the United States. They also have stable trading relationships with each other, which strengthen their similarity.

As major traders could be hubs in the world energy market, similarity could be brought to the neighborhood by the weight on the edge. These countries are generally unchanged in their topology position, which may indicate the stability in world energy trading relationships. The United States, Canada,

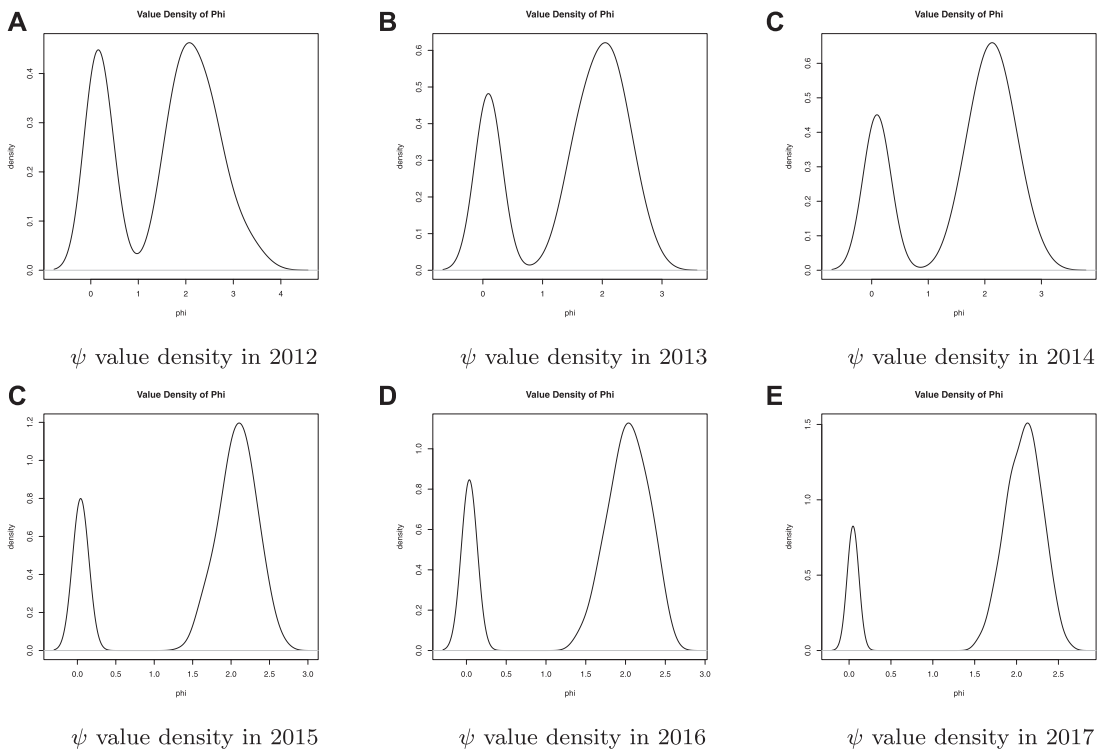



**FIGURE 3 |** $\psi$ value density within major traders.

Mexico, and the EU are all stable upon their relative similarity value. For huge economic entity like the EU, the similarity value is smaller due to its trade flow and risk sharing with countries which are net importers or exporters, such as Greece or Spain. This consequence is consistent with Youssef and Mokni (2019), whose research states that the response of market to oil price changes in oil-importing countries is more pronounced than that of oil-exporting countries during periods of turmoil.

The heat map becomes blurry after time to time when the region all expands. The major traders are becoming a more internally related entity. Changes in the market can also be enhanced by the increasing trade volume. When external events ever happen, those countries with large trade volume are likely to have similar responses. From the point of efficient market, any disturbance in a local region is more likely to propagate throughout the world regardless of geographical distance (Gabaix, 2016; Lang and Auer, 2019). So it always has the largest similarity with itself. We notice that most major traders have self similarity to some degree. Generally, self-similarity means a trading loop in the market as we scale self-weighting to zero for most cases. The loop means that the countries either reexport or reimport the energy, or a loop trading relation exists among countries. For instance, Canada, Spain, Mexico, and the United States can form a trade loop. Generally, these countries are major global or regional economic entities that establish both a significant import and export relationship with the entire world around. Mexico is a major energy provider for mid-south America, while it also has close import relationship with the United States and Canada.

A more clear view of this phenomenon could be observed by the density plot in **Figure 3**. A similar type of separation and aggregation can also be found when we zoom into the values only between the major traders. From **Figure 3**, we can clearly observe two peaks in each plot, while values are gathering towards the peak on the right side. Not only the global market has the enhancement of relation, this also appears among the major traders. This may suggest multi-polarization on market sharing. If the world market is emerging under single-polarization, **Figure 3** would then go to unify as a whole, since the response from the market will all depend on a single entity. However, we could observe peak aggregation in the other direction. During the process of multi-polarization, system risk is also shared among all major traders. This can force them to cooperate and act in a same way after events happen. As a result, it is inevitable that the values will gather towards one peak since cooperation among large entities is the best way to make greater benefits.

## 4.3 Events Effects on the Network and Nodes

Besides the major traders, we consider below the events around the world that affects energy production, transportation, and consumption. Here, the events we consider are namely financial crisis, the development of shale gas, and oil technology.

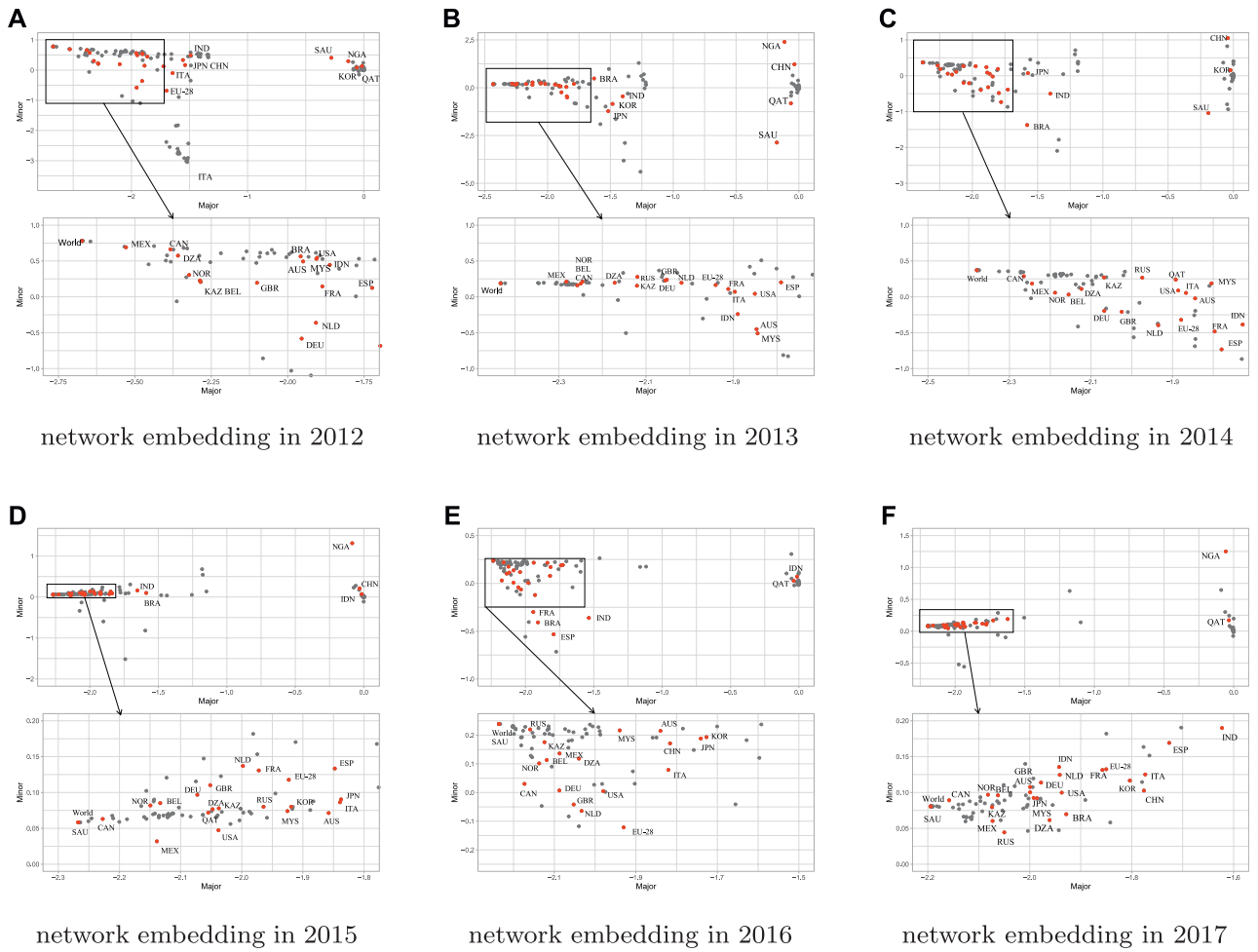



**FIGURE 4 |** First two components of network embedding.

### 4.3.1 Effects of Worldwide Events

There is a timely manner for the world to recover from the financial crisis. The time spends on recovery can be different for each country. A large economic entity can have the power and money to fill the gap and loss of the crisis in a short time. While those rather small economic entities might be more heavily damaged and wait for the proper regulation and relief to come back to the normal state (Aydoğan et al., 2017; An et al., 2018). This can be viewed as a process of flux from the most stable countries and organization to the others. This agrees with the phenomena in **Figure 1** where the outliers are reduced over the years. Furthermore, the development of shale gas technology can help some developed or developing countries to find a supplement energy source, which can also help their recovery and further development.

### 4.3.2 Effects of the Shale Gas Technology

The change in impulse response could be observed by looking at the change in the point position. We can observe this on the plot of embedding result in **Figure 4C**, recalling that *x*-axis is proportional to the similarity value and *y*-axis is proportional to node strength. After the quick development of shale gas technology in year 2014, we find that RUS is brought into the picture (from eight months after 2013 in **Figure 4B**). This also helps the US to be an energy-independent country, and certainly changes the network structure, leading to a world-wide effecting of clustering. In **Figure 4**, points of countries are gathering towards the left side and stacking in the vertical direction.

Meanwhile, Asian and European countries stay in the same scope in **Figure 4**, as Russia and Norway has long been major exporters of crude oil and natural gas. Malaysia and Indonesia are the major source area of shale gas in Asia. The energy market of these countries is as stable as those developed countries since natural gas is one of the major supplement energy other than crude oil. In recent decades, the world is not able to shift from fossil fuel to reproducible energy. As a result, one holding the major source of energy will be stable. For example, Norway is the top natural gas exporter, Germany is the second largest natural gas exporter, and Canada is the major oil exporter.

### 4.3.3 Effects on an Individual Country

If we look through **Figure 4**, we can observe that major traders (Japan and Russian) at the origin have the trend of moving towards the other end. Moreover, to examine the regional effect

and overall behavior, we add two reference points to the plot, which are "world" and "EU-28." Since the world overall could be regarded as a stable market, we suggest that the more a point deviates away from the origin on a major axis, the more stable it can be under major events. Each zoomed area of the plot largely consists of developed countries (European countries and North American countries) and major exporters of natural gas (**Table 1**). During all of the investigated years, no country moves out of the scope though their relative topology position may have changed. We notice that Mexico and Canada often take the leading position while the United States lags. Meanwhile, the United States, being the leading position of world economic entity, also has a lot of trades with other poles in the world. Also, the dependency on allies in all continents makes the United States to share and distribute the risk of the market. Mexico and Canada have rather simple energy trading relations and route with the United States. As a result, they gain more benefits of stability and less risk from trading relations.

From **Figure 4**, we can observe that position of China has constantly changed. In **Figure 4C**, it stays close to the developed country and outermost region in the plot. While in year 2015, China still stays away from the top and stable areas of countries. In the next following years (**Figures 4E,F**), it behaves even better than all previous years. The zigzag behavior of China indicates that it lacks strong stability for the increasing demand of gas. As a growing developing country that depends on energy heavily, China is greatly affected by the global market, especially for shortage of supply. The process that makes China a stable entity has both internal and external reasons. The internal reason might be the energy market opening strategy in 2015, introducing more freedom of trading and exporting to the market. This leads to a more open and efficient market, thus exposes China more close to the developed countries. Also, the natural gas pipeline between China and Russia begins construction in 2015. This gives a positive signal that China has even more demand for energy than before. The stimulation on the market related to China gives it more priority and interaction with the global market.

## 4.4 Discussion

The decrease of extreme outliers in **Figure 2** may attribute to a more efficient energy market. Meanwhile, there might not exist simple reasoning behind this phenomenon (Badel and McGillicuddy, 2015; Baumeister and Kilian, 2016a). One possible explanation is the increment of trade volume in each year. As we analyze the network considering degree and edge weight, the increment in the trade value will certainly enhance the network structure in terms of the node to node dynamics. The more commodity trades between a pair of countries, the closer connection can exist (Ravasz et al., 2002; Comin et al., 2020). This also holds for the major traders. Countries with small capital and regional and political situations can override the power of trade volume (Filis and Chatziantoniou, 2013). On the other hand, a country can be more dependent on another on energy with the growth of trade volume. Then the national economy might be more correlated, which is also stated in (Hu and Cheng, 2008; Gabaix, 2016; Gong and Lin, 2018). Generally, the more population we have, the greater energy consumption will be. So developing countries may show a sharp increase in energy trade volume as energy consumption grows. A more efficient market can also increase the dynamics of the network. So the consequence of the event may be better revealed in the structure. Consequently, population growth, technology, and the development of a nation can set the ground for further analysis.

The changing topology position of China in **Figure 4** from the year 2013 to 2017 indicates that the energy import system of China is not stable. The total demand for energy increases and leaves pressure on the energy import system. It could be affected by the unexpected demand increase and the effects last for years. For energy-importing countries, increasing reliance on imports exacerbates a country's vulnerability under supply shocks. The development of unconventional resources and multi-import sources can be solutions to mitigate risks. For energy exporters, domestic economic development is highly reliant on international energy markets with the high volatility of energy prices. Energy trade correlation stability is important both for importers and exporters. A deeper investigation of energy consumption and energy trade networks of individual countries is worthy of further research and could shed light on diplomatic policies.

# 5 CONCLUSION

We presented a framework of the dynamic structure analysis over time by the dynamical embedding of complex networks, which is an extension of the heat kernel embedding. We focused on its applications to dimensional reduction and the detection of dynamical modules to highlight both intrinsic and induced features of world crude oil and gas trade networks. The similarity measure provides integrated description of networks and multi-scale statistics of dynamic blocks in terms of coarse-grained representations, which are different from the correlation analysis of time-series data. It suggests that the energy market system can be effectively described by a small set of major traders that govern the dynamics over some time scale. So the complexity of the dynamics could be reduced just keeping track of a few aggregated "meta-nodes", whose state is governed by the major traders. By obtaining the dynamics characters of the world energy trade networks and localized features of the network structure, we assessed the effects of events on the trade correlation structure and importing or exporting countries. The visualized topological positions of the nodes could be understood as new features, which would give insights for long term decisions of countries' stability under events. Multi-scale dynamical embedding of complex networks could be applied to deploy the resilience ability from shocks both for nodes and systems. The trends of multi-polarization on market sharing and stability of the trade loop among major traders is also helpful for developing countries, for example, China in making diplomatic decisions.

Because the industrial structure of the developed western economies reached maturity for decades, most of the econometric research on economic structural evolution is for developed states. Less-developed regions with their specific features have not been deeply studied. For instance, some

external economy parameters, for example, GDP or economy growth of the countries of the year and the exchange rate in currency, play important roles in the results and should be considered. For further research, capturing information about a node, for example, for China, from the local distribution over node attributes around it is needed. The circumstance of population growth, technology, and economic development of the nation could be considered in a deeper investigation for China's energy import security. Energy consumption and energy trade networks of individual countries are important factors affecting energy system stability. Both the macro and micro level research can open up new views for diplomatic policies.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

FC is responsible for the whole manuscript and managed the research process. LX focuses on model construction and an analysis of the results. GF focuses on part of writing. QX focuses on part of writing. LD contributes on some of the results analysis.

## FUNDING

This article is sponsored by the Natural Science Foundation of Shanghai (18ZR1420200), the National Natural Science Foundation of China (61603253), China Postdoctoral Science Foundation funded project (2016M601598), and the Fundamental Research Funds for the Central Universities.

## ACKNOWLEDGMENTS

We thank the support from the State Key Laboratory of Ocean Engineering and the Key Laboratory of Marine Intelligent Equipment and System, Ministry of Education.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2021.729690/full#supplementary-material

## REFERENCES

Abraham, I., Bartal, Y., and Neiman, O. (2006). Advances in Metric Embedding Theory. *Adv. Maths.* 228 (6), 3026–3126. doi:10.1145/1132516.1132557

An, Q., Wang, L., Qu, D., and Zhang, H. (2018). Dependency Network of International Oil Trade before and after Oil price Drop. *Energy* 165, 1021–1033. doi:10.1016/j.energy.2018.09.098

Axon, C. J., and Darton, R. C. (2021). Sustainability and Risk - a Review of Energy Security. *Sustainable Prod. Consumption* 27, 1195–1204. doi:10.1016/j.spc.2021.01.018

Aydoğan, B., Tunç, G., and Yelkenci, T. (2017). The Impact of Oil price Volatility on Net-Oil Exporter and Importer Countries' Stock Markets. *Eurasian Econ. Rev.* 7, 231–253. doi:10.1007/s40822-017-0065-1

Badel, A., and McGillicuddy, J. (2015). Oil Prices: Is Supply or Demand behind the Slump?. *Econ. Synopses*. doi:10.20955/es.2015.8

Banisch, R., and Djurdjevac Conrad, N. (2015). Cycle-flow-based Module Detection in Directed Recurrence Networks. *Epl* 108, 68008. doi:10.1209/0295-5075/108/68008

Battiston, F., Nicosia, V., and Latora, V. (2014). Structural Measures for Multiplex Networks. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* 89, 032804. doi:10.1103/PhysRevE.89.032804

Baumeister, C., and Kilian, L. (2016a). Forty Years of Oil price Fluctuations: Why the price of Oil May Still surprise Us. *J. Econ. Perspect.* 30, 139–160. doi:10.1257/jep.30.1.139

Baumeister, C., and Kilian, L. (2016b). Understanding the Decline in the price of Oil since June 2014. *J. Assoc. Environ. resource economists* 3, 131–158. doi:10.1086/684160

Blanchard, O. J., and Gali, J. (2007). The Macroeconomic Effects of Oil Shocks: Why Are the 2000s So Different from the 1970s? Technical Report. *Natl. Bur. Econ. Res.*. doi:10.3386/w13368

Borg, I., and Groenen, P. J. F. (2010). Modern Multidimensional Scaling: Theory and Applications. *J. Educ. Meas.* 40, 277–280. doi:10.1111/j.1745-3984.2003.tb01108.x

BP (2019). Statistical Review of World Energy. Available at: https://www.bp.com/.

Casadiego, J., Nitzan, M., Hallerberg, S., and Timme, M. (2017). Model-free Inference of Direct Network Interactions from Nonlinear Collective Dynamics. *Nat. Commun.* 8, 2192–2210. doi:10.1038/s41467-017-02288-4

Comin, C. H., Peron, T., Silva, F. N., Amancio, D. R., Rodrigues, F. A., and Costa, L. d. F. (2020). Complex Systems: Features, Similarity and Connectivity. *Phys. Rep.* 861, 1–41. doi:10.1016/j.physrep.2020.03.002

Cui, P., Wang, X., Pei, J., and Zhu, W. (2019). A Survey on Network Embedding. *IEEE Trans. Knowl. Data Eng.* 31, 833–852. doi:10.1109/tkde.2018.2849727

Davies, E. B. (1989). *Heat Kernels and Spectral Theory*, Vol. 92. Cambridge University Press.

De Domenico, M., Nicosia, V., Arenas, A., and Latora, V. (2015). Structural Reducibility of Multilayer Networks. *Nat. Commun.* 6, 6864. doi:10.1038/ncomms7864

Dehmamy, N., Milanlouei, S., and Barabási, A. L. (2018). A Structural Transition in Physical Networks. *Nature* 563, 676–680. doi:10.1038/s41586-018-0726-6

Delvenne, J. C., Lambiotte, R., and Rocha, L. E. (2015). Diffusion on Networked Systems Is a Question of Time or Structure. *Nat. Commun.* 6, 7366. doi:10.1038/ncomms8366

Esfahani, A. N., Moghaddam, N. B., Maleki, A., and Nazemi, A. (2021). The Knowledge Map of Energy Security. *Energ. Rep.* 7, 3570–3589. doi:10.1016/j.egyr.2021.06.001

Filis, G., and Chatziantoniou, I. (2013). Financial and Monetary Policy Responses to Oil price Shocks: Evidence from Oil-Importing and Oil-Exporting Countries. *Rev. Quantitative Finance Account.* 42 (4), 709–729. doi:10.1007/s11156-013-0359-7

Frey, N. C., Matin, S., Stanley, H. E., and Salinger, M. A. (2019). Universal Fluctuations in Growth Dynamics of Economic Systems. *Sci. Rep.* 9, 713–715. doi:10.1038/s41598-018-38088-z

Fu, G., Chen, F., Liu, J., and Han, J. (2019). Analysis of Competitive Information Diffusion in a Group-Based Population over Social Networks. *Physica A: Stat. Mech. its Appl.* 525, 409–419. doi:10.1016/j.physa.2019.03.035

Gabaix, X. (2016). Power Laws in Economics: An Introduction. *J. Econ. Perspect.* 30, 185–206. doi:10.1257/jep.30.1.185

Gong, X., and Lin, B. (2018). Time-varying Effects of Oil Supply and Demand Shocks on China's Macro-Economy. *Energy* 149, 424–437. doi:10.1016/j.energy.2018.02.035

Grover, A., and Leskovec, J. (2016). "node2vec: Scalable Feature Learning for Networks," in Proceedings of the 22nd ACM SIGKDD international conference

on Knowledge discovery and data mining, New York, NY, August 24–27, 2016, 855–864.

Hadfield, A. (2008). EU-Russia Energy Relations: Aggregation and Aggravation1. *J. Contemp. Eur. Stud.* 16, 231–248. doi:10.1080/14782800802309953

Hamilton, J. D., (2010). Historical Oil Shocks. In *Handbook of Major Events in Economic History* (New York, NY: Routledge Taylor and Francis Group). doi:10.4324/9780203067871.ch21

Hu, L., and Cheng, T. S. (2008). China's Energy Security and Geo-Economic Interests in central Asia. *Cent. Eur. J. Int. Security Stud.* 2, 52, 2008 . Available at: webmail2.cejiss.org/articles/vol2–2/hu-cheng/

Inoue, A., and Kilian, L. (2016). Joint Confidence Sets for Structural Impulse Responses. *J. Econom.* 192, 421–432. doi:10.1016/j.jeconom.2016.02.008

ISO (2019). Country Codes. Available at: https://www.iso.org/obp/ui/#search/code/.

Kilian, L. (2009). Not all Oil price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *Am. Econ. Rev.* 99, 1053–1069. doi:10.1257/aer.99.3.1053

Kilian, L., Rebucci, A., and Spatafora, N. (2009). Oil Shocks and External Balances. *J. Int. Econ.* 77, 181–194. doi:10.1016/j.jinteco.2009.01.001

Kremer, M., Becker, A. P., Vodenska, I., Stanley, H. E., and Schäfer, R. (2019). Economic and Political Effects on Currency Clustering Dynamics. *Quantitative Finance* 19, 705–716. doi:10.1080/14697688.2018.1532101

Lambiotte, R., Delvenne, J. C., and Barahona, M. (2014). Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. *IEEE Trans. Netw. Sci. Eng.* 1, 76–90. doi:10.1109/tnse.2015.2391998

Lang, K., and Auer, B. R. (2019). The Economic and Financial Properties of Crude Oil: A Review. *North Am. J. Econ. Finance* 52, 100914. doi:10.1016/j.najef.2019.01.011

Li, T., Xue, L., Chen, Y., Chen, F., Miao, Y., Shao, X., et al. (2018). Insights from Multifractality Analysis of Tanker Freight Market Volatility with Common External Factor of Crude Oil price. *Physica A: Stat. Mech. its Appl.* 505, 374–384. doi:10.1016/j.physa.2018.02.107

Mealy, P., Farmer, J. D., and Teytelboym, A. (2019). Interpreting Economic Complexity. *Int. J. Agric. Stat. Sci.* 14, 1–9. doi:10.1126/sciadv.aau1705

Newman, M. E., and Clauset, A. (2016). Structure and Inference in Annotated Networks. *Nat. Commun.* 7, 1–11. doi:10.1038/ncomms11863

Newman, M. E. (2018). Network Structure from Rich but Noisy Data. *Nat. Phys.* 14 (6), 542–545. doi:10.1038/s41567-018-0076-1

Pecora, L. M., Sorrentino, F., Hagerstrom, A. M., Murphy, T. E., and Roy, R. (2014). Cluster Synchronization and Isolated Desynchronization in Complex Networks with Symmetries. *Nat. Commun.* 5, 1–8. doi:10.1038/ncomms5079

Albert, R., Jeong, H., and Barabási, A. L., (2000). Error and Attack Tolerance of Complex Networks. *Nature* 406, 378 – 382. doi:10.1038/35019019

Ravasz, E., Somera, A. L., Mongru, D. A., and Oltvai, Z. N. (2002). Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297, 1551–1555. doi:10.1126/science.1073374

Ren, S., Li, H., Wang, Y., Guo, C., Feng, S., and Wang, X. (2021). Comparative Study of the china and u.S. Import Trade Structure Based on the Global Chromium Ore Trade Network. *Resour. Pol.* 73, 102198. doi:10.1016/j.resourpol.2021.102198

Rozemberczki, B., Allen, C., and Sarkar, R., 2019. Multi-scale Attributed Node Embedding. arXiv preprint arXiv:1909.13021.

Schaub, M. T., Delvenne, J. C., Lambiotte, R., and Barahona, M. (2019). Multiscale Dynamical Embeddings of Complex Networks. *Phys. Rev. E* 99, 062308. doi:10.1103/PhysRevE.99.062308

Schaub, M. T. (2014). "Unraveling Complex Networks under the Prism of Dynamical Processes: Relations between Structure and Dynamics,". Ph.D. thesis.

Song, X., Niu, D., and Zhang, Y. (2016). The Chaotic Attractor Analysis of Djia Based on Manifold Embedding and Laplacian Eigenmaps. *Math. Probl. Eng.* 2016, 1–10. doi:10.1155/2016/8087178

Sorrentino, F., Pecora, L. M., Hagerstrom, A. M., Murphy, T. E., and Roy, R. (2016). Complete Characterization of the Stability of Cluster Synchronization in Complex Dynamical Networks. *Sci. Adv.* 2, e1501737. doi:10.1126/sciadv.1501737

Stavroglou, S. K., Pantelous, A. A., Stanley, H. E., and Zuev, K. M. (2019). Hidden Interactions in Financial Markets. *Proc. Natl. Acad. Sci. U S A.* 166, 10646–10651. doi:10.1073/pnas.1819449116

Tushar, W., Saha, T. K., Yuen, C., Smith, D., and Poor, H. V. (2020). Peer-to-peer Trading in Electricity Networks: An Overview. *IEEE Trans. Smart Grid* 11, 3185–3200. doi:10.1109/tsg.2020.2969657

Tushar, W., Yuen, C., Mohsenian-Rad, H., Saha, T., Poor, H. V., and Wood, K. L. (2018). Transforming Energy Networks via Peer-To-Peer Energy Trading: The Potential of Game-Theoretic Approaches. *IEEE Signal. Process. Mag.* 35, 90–111. doi:10.1109/msp.2018.2818327

Tushar, W., Yuen, C., Saha, T. K., Morstyn, T., Chapman, A. C., Alam, M. J. E., et al. (2021). Peer-to-peer Energy Systems for Connected Communities: A Review of Recent Advances and Emerging Challenges. *Appl. Energ.* 282, 116131. doi:10.1016/j.apenergy.2020.116131

UN (2019). Un Comtrade Database. Available at: https://comtrade.un.org/.

Wang, C., Zhang, X., Vilela, A. L., Liu, C., and Stanley, H. E. (2019). Industrial Structure Upgrading and the Impact of the Capital Market from 1998 to 2015: A Spatial Econometric Analysis in Chinese Regions. *Physica A: Stat. Mech. its Appl.* 513, 189–201. doi:10.1016/j.physa.2018.08.168

Xi, X., Zhou, J., Gao, X., Liu, D., Zheng, H., and Sun, Q. (2019). Impact of Changes in Crude Oil Trade Network Patterns on National Economy. *Energ. Econ.* 84, 104490. doi:10.1016/j.eneco.2019.104490

Xue, L., Chen, F., Guo, S., Fu, G., Li, T., and Yang, Y. (2019). Time Varying Correlation Structure of Chinese Stock Market of Crude Oil Related Companies Greatly Influenced by External Factors. *Physica A: Stat. Mech. its Appl.* 530, 121086. doi:10.1016/j.physa.2019.121086

Youssef, M., and Mokni, K. (2019). Do crude Oil Prices Drive the Relationship between Stock Markets of Oil-Importing and Oil-Exporting Countries? *Economies* 7, 70. doi:10.3390/economies7030070

Yuan, M., Zhang, H., Wang, B., Huang, L., Fang, K., and Liang, Y. (2020). Downstream Oil Supply Security in china: Policy Implications from Quantifying the Impact of Oil Import Disruption. *Energy Policy* 136, 111077. doi:10.1016/j.enpol.2019.111077

Zhao, C., and Chen, B. (2014). China's Oil Security from the Supply Chain Perspective: A Review. *Appl. Energ.* 136, 269–279. doi:10.1016/j.apenergy.2014.09.016

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Xue, Chen, Fu, Xia and Du. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fenvs.2021.724716

# Assessments of Air Pollution Control Effectiveness Based on a Sharp Regression Discontinuity Design —Evidence From China's Environmental Big Data

*Ren Wang*[1], *Jiaqi Huang*[2], *Lizhi Zhang*[1], *Yu Xia*[1], *Xu Xu*[3]* *and Tongli Nong*[1]

[1]School of Finance, Hunan University of Technology and Business, Changsha, China, [2]School of Mathematics and Statistics, Hunan University of Technology and Business, Changsha, China, [3]School of Economics, Beijing Technology and Business University, Beijing, China

The *Assessment and Appraisal Method for Ecological Construction Targets* (the *Method*) was promulgated in 2016, which provided a concrete instruction for China's air pollution control and established an explicit standard for reducing air pollutant concentration. This study implements a sharp regression discontinuity (RD) design and makes an assessment on air quality control effectiveness of the *Method* based on the high-volume big data acquired from 173 cities in China. The results show that the *Method* has significantly improved air pollution control on the overall air quality index (AQI) and reducing concentrations of $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, and CO across the country in the observation periods. However, no reduction effect was observed for $O_3$. The robustness tests support the conclusion as well. Besides, the heterogeneity analysis illustrates that the policy had a significant short-term treatment effect in East, South, Central, North, Northwest, Southwest, and Northeast China. However, the *Method*'s effect is found to decline over time either nationwide or regionally according to the persistence analysis. Therefore, this article puts forward several suggestions regarding the formulation of long-term regulations for air pollution control, the transformation of the growth model for sustainable development, and optimization of the incentive system for improved pollution control and prevention.

**Keywords: policy assessments, big data, air quality, pollutant concentration, sharp regression discontinuity design**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, China*

***Reviewed by:***
*Jun Wang,*
*University of Iowa, United States*
*Yongran Lin,*
*Henan Administration Institute, China*

***Correspondence:**
*Xu Xu*
*xux@th.btbu.edu.cn*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *14 June 2021*
***Accepted:*** *09 August 2021*
***Published:*** *30 September 2021*

***Citation:***
*Wang R, Huang J, Zhang L, Xia Y, Xu X and Nong T (2021) Assessments of Air Pollution Control Effectiveness Based on a Sharp Regression Discontinuity Design —Evidence From China's Environmental Big Data . Front. Environ. Sci. 9:724716. doi: 10.3389/fenvs.2021.724716*

## INTRODUCTION

With China's rapid economic growth and ever-accelerating increase in resource consumption, air pollution has become the primary problem threatening the country's sustainable development. Data showed that in 2016, 81% of the Chinese population were exposed to ambient fine particulate matter ($PM_{2.5}$) higher than the WHO IT1 (or CAAQS) standard (35 $\mu g/m^3$), and 50% of the population were exposed to an annual average $PM_{2.5}$ mass concentration of 48.2 $\mu g/m^3$ or above (Song et al., 2017). In China, air pollution is estimated to cause 1.6 million deaths per year, accounting for roughly 17% of all deaths nationwide (Rohde and Muller, 2015). Air pollution is also identified as a cause of mental illnesses for which the associated health expenditures are USD 22.9 billion annually in total

(Chen S. et al., 2018). The disease burden and health economic loss which resulted from ambient air pollution in China are greater than those in the remaining parts of the world as the country with one of the highest levels of air pollution and the largest size of the affected population (Niu et al., 2017). With increasing severity requiring government response, China has implemented a series of environmental protection policies and regulations to tackle air pollution.

In December 2016, the State Council of the People's Republic of China launched the *Assessment and Appraisal Method for Ecological Construction Targets* (the *Method*). The *Method*, which specified higher requirements on air pollution control in its special assessment of eco-environmental protection, was proposed to center on the completion of critical targets of ecological civilization construction in each region, strengthen the awareness of responsibility in ecological civilization construction of provincial Communist Party of China (CPC) committees and governments, and supervise and urge regions to promote ecological civilization construction on their own initiatives.

Specifically, the *Method*, combined with the *Green Development Index* as a quantitative basis, stipulated a comprehensive assessment from many aspects, including regional resource utilization, environmental improvement, ecological conservation, the level of high-quality development, green life, public satisfaction, etc. The goal of the *Method* is to ensure that all provinces and cities strictly follow the guidelines of the central government when implementing these pollution control measures. Especially, the assessment and appraisal results of the *Method* would be an important index of general assessment, appraisal, awarding, promotion, and dismissal of leading carders and other local officials. The region with good performance of the appraisal would be commended, while the region with poor performance would be named and shamed, and its officials would be accountable.

In general, the *Method* impacts environment protection through the below mechanism. It sets ecological civilization construction targets as appraisal indexes for the environmental supervision system. Hence, the leaders of the Party who rule the local bureaucracy and the local government are stimulated to make cooperated efforts and take measures to address the pollution threat.

Although the *Method* targeted many aspects of the ecological civilization, this study will try to concentrate on air quality. We would like to verify whether the *Method* has provided the necessary incentives for the local governments to achieve the appraisal goal. Especially, to what degree has the mechanism worked to improve the air quality? Are the improvements made temporary or sustainable? An assessment of the *Method* and an exploration of its effect will provide necessary suggestions for policymakers on how to improve and optimize the air pollution control and regulations in the next period. From this perspective, this article attempts to use a sharp regression discontinuity (RD) design to accurately evaluate the effectiveness of the *Method*'s implementation based on the environmental big data collected among 173 cities in China.

Existing research studies evaluating the effect of air pollution regulation policies tend to focus on the following areas: first, research studies have been carried out to explain the reasons to implement air pollution regulations and illustrate their models. The market mechanisms cannot solely solve the environmental pollution problems because of the externalities of environmental pollution and the characteristics of the environment as a public good. Environmental regulations have therefore been crucial for remedying market failures and solving environmental problems (Fan and Sun, 2020). Environmental regulations on market entities' pollutant discharge behavior can be roughly classified into three subtypes: government-oriented environmental regulations, market-oriented environmental regulations, and voluntary environmental regulations. To cope with growing resource and environmental pressures, the current environmental regulation approach in China is developing from the government-oriented only to the coordination of the government-oriented and the market-oriented (Zhao et al., 2009). Other environmental regulations may indirectly affect air pollution through three possible paths: adjusting the industrial structure (Chen S. et al., 2019; Yu and Wang, 2021), enhancing technological innovation (Jebaraj and Iniyan, 2006; Li et al., 2017; Chen X. et al., 2018), or attracting foreign direct investment (Wu and He, 2017). A second research focus is to evaluate the effect of environmental protection policies and regulations. It is confirmed that environmental regulations are effective in reducing harmful gas emissions and alleviating air pollution according to the studies of environmental regulations' effect on the ecological environment and on the energy resource protection including the effect of implementing renewable energy policies, coal substitution policies, and other environmental policies (Neves et al., 2020; Ouyang et al., 2019; Shou et al., 2020; Wang et al., 2019). Some scholars argue that environmental inspection at the behest of central authorities has a significant effect, but the effect lasts for a short time (Tan and Mao, 2021). A third research focus is on the effects of regional environmental protection policies. For example, Yang et al. (2019) found that a unified air control policy was more effective in Shanghai, Jiangsu, and Zhejiang, while one with various types of emission control was more effective in the Pearl River Delta region (Hou et al., 2019). As for the Beijing–Tianjin–Hebei region, research has found that air control policies can reduce the region's annual gross domestic product (GDP) growth rate by 1.4% and that the annual GDP growth rate will drop by 2.3% when the policies are rigidly enforced (Li et al., 2019). In addition, many scholars have started to use big data for environmental assessment research. Zhang et al. (2019), for example, assume that big data analysis (BDA) is helpful to provide decision makers with a sound scientific advice on solving global sustainable development problems. Therefore, BDA is gradually used in forecasting air pollutant concentrations (Alaoui et al., 2019; Al_Janabi et al., 2019; Xu et al., 2020) and other research fields.

Policy evaluation uses a range of research methods to systematically investigate the effectiveness of policy interventions, implementation, and processes and to determine their merit, worth, or value. There are four major methods to evaluate the public policies, including instrumental variable (IV), difference in difference (DID), propensity score matching (PSM), and regression discontinuity design (RDD) (Chang et al., 2018; Jia

and Chen, 2019; List et al., 2003; Tu et al., 2020). The IV method could easily generate estimate results, but a suitable instrumental variable is difficult to define. Moreover, the method has a precondition that the heterogeneity of a chosen policy's effect should be independent of the policy's overall effect, which is a rigorous hypothesis. Therefore, Heckman, 1997 argues that a heterogeneity result would be invalid when the IV method is adopted. DID has many strengths such as focusing on changes of factors instead of noises, but it is based on strong assumptions. In particular, the conventional DID estimator requires that there should be no time-varying distinction between the treated group and the control group. Besides, the *Method* is a nationwide policy, and it is impossible to set the control group under such a situation; thus, a DID method cannot be applied to the estimate based on nationwide city-level data.

The RD design possesses several strengths compared with DID and propensity score matching (PSM). The RD design requires relatively less strict assumptions which permit a nonlinear time trend. It yields an unbiased estimate of the local treatment effect. As a quasi-experiment, RD does not require ex-ante randomization and circumvents theoretical issues of random assignment. It can be seen as a randomized experiment in measuring a treatment effect for the purpose of a policy evaluation methodology.

Different from most previous studies, there are three main contributions of this article. First, the evaluation of the air pollution policy is improved. The environmental big data across the country are used and the national factors are considered, which enhance the applicability and credibility of the assessment. Second, the introduced model based on the sharp RD design is capable of reducing the endogeneity problems, which makes the conclusion more robust. Finally, this study is of great practical importance. The heterogeneity analysis is conducted to evaluate the regional effect of the *Method*'s release on air pollution control nationwide.

The structure of this article is presented as follows: after the Introduction, the second section discusses the research materials and methods, including the big data acquisition and processing, the description analysis, and model setting. The third section presents the results, including an analysis of baseline regression results, the robustness test, persistence analysis, and heterogeneity analysis. The final section consists of the article's conclusion and suggestions.

# MATERIALS AND METHODS

## Big Data Acquisition and Processing

The composite air quality index (AQI) is a rating indicator for reporting the ambient air pollution recorded at monitoring sites on a particular time scale like daily (Kai-guang et al., 2019). The statistic shows that the air pollutants in Chinese cities mainly include six pollutant items such as fine particulate matter ($PM_{2.5}$), inhalable particles ($PM_{10}$), sulfur dioxide ($SO_2$), nitrogen dioxide ($NO_2$), carbon monoxide (CO), and ozone ($O_3$) (Liu et al., 2013). Although it is observed by satellite monitoring that the concentration of $SO_2$ showed a decreasing trend (Calkins et al., 2016; Wang and Wang 2020), satellite observations are not seeing ground-based levels of air pollution. The AQI is defined according to the six pollutants and simplifies separately their concentration values into a conceptual index value (Individual Air Quality Index, IAQI), which is the maximum value of the conceptual index values (Zhang et al., 2019). The main objectives of AQI are to inform and caution the public about the risks of exposure to pollution levels and to enforce required regulatory measures. The overall ambient air quality of a specified area can be assessed in a better way and quantified in terms of AQI since it represents the collective effect of all the pollutants. The Chinese Government has collected mass air pollutant data across the country to monitor real-time air quality since 2013.

As mentioned, the *Method* was launched in 2016 and the first round of annual assessment on ecological environment protection was completed by August, 2017. In order to provide an up-to-date policy effectiveness evaluation research, the high-volume data collected for the period ranging from November 2015 to February 2018 have been processed to explain the change of air quality based on the daily measured indicators including AQIind, $hPM_{2.5avg}$, $hPM_{10}avg$, $hSO_2avg$, $hNO_2avg$, $hCO_{avg}$, and $hO_{3avg}$, each of which represents its daily average concentration. A total of 1,066,898 observations are collected from selected locations of residential areas. Among the indicators, AQIind is denoted as a description of the overall air quality, and the other indicators of the six pollutants are used to analyze how air quality changes specifically and respectively.

To focus on the impact of the policy on air quality, weather conditions must be controlled in the analysis (Wu et al., 2020). We take the highest temperature (Htmpt), the lowest temperature (Ltmpt), weather conditions (Wthcdt), and wind force (Wndfrc) as the control variables, amounting to 609,656 observations. For weather conditions, we construct a dummy variable: rainy and snowy weathers are counted as "1", while other conditions are "0". Wind forces are recorded as "1.5" when lower than 3, "3.5" for the range of 3–4, and "4" for changing from 3–4 to 4–5, and so on. Besides, rapid industrial development and the dominance of economic growth in political decision-making have threatened the environment and air quality in China. Apparently, the economic indicators including the gross regional product (Grorgpdut), per capita gross regional product (PcptlGRP), and gross regional product growth rate (GRPgwrate), amounting to 457,242 observations, should also be controlled for their potential influences on air pollution (Wu et al., 2021). The annual economic indicators are transformed into daily data, and all the variables except dummy ones are standardized.

In summary, in this study, 173 cities at or above the prefecture level are eventually covered because the data of the other cities are excluded from the sample for their data being unavailable or unable to match. Some missing values are filled by the linear interpolation method, and 3,647,669 valid observations in total are acquired as the big data sample. All the data are collected from Chinese Research Data Services (CNRDS), and the analysis is conducted through Stata.

## Data Description and Analysis

This article focuses on the impact of the *Method* on improving air quality based on air quality data from 173 cities at or above the

**TABLE 1 |** Comparison of air quality indexes before and after (mean).

| Sample time | Before the release | After the release | Mean difference |
|---|---|---|---|
| AQIind | 75.747 | 74.404 | −1.343 |
| h$PM_{2.5}$avg | 47.686 | 44.805 | −2.881 |
| h$PM_{10}$avg | 81.379 | 78.202 | −3.177 |
| h$SO_2$avg | 21.063 | 17.175 | −3.888 |
| h$NO_2$avg | 31.816 | 32.541 | 0.725 |
| hCOavg | 1.055 | 0.988 | −0.067 |
| h$O_3$avg | 56.479 | 61.847 | 5.368 |
| observations | 443,240 | 440,818 | 884,058 |

*Note: Inconsistencies between columns 2 and 3 are due to the fact that the cut point is not the 50th percentile point and the missing data differ between the two periods. Data source: Chinese Research Data Services (CNRDS).*

prefecture level across China during 1 year before and after the *Method*'s implementation (2015–2017). First, by comparing the mean of each air quality indicator data in the different sample time periods, the result is that the value of AQI decreased by 1.343 units 1 year after the *Method*'s implementation, which reached the expectation. **Table 1** summarizes the mean of the sample cities' AQI and air pollutant indicator data. The results of 1 year before and after the *Method*'s release periods are presented, respectively, in columns 1 and 2. The mean difference of the comparison between columns 1 and 2 is presented in column 3.

In the light of the policy's effects of continuity and hysteresis, this article extends the sample time span. A line chart (**Figure 1**) plots the changes of AQI for 3 years before and after the *Method* implementation. It can be seen in **Figure 1** that in the next 2 months after the *Method*, the value of AQI decreased significantly and kept at a low level. Also, its value in the year after the release was significantly lower than that in the year before the release from the overall trend of the year. Therefore, it can be preliminarily concluded that the AQI declined because of the implementation of the *Method*. However, it should be noted that national AQI dropped sharply on December 22, 2016, and then rebounded immediately. This is likely because it was the time for the public heating system in China to operate, which twisted the effects of the *Method*'s implementation temporarily.

To further find out the effectiveness of the *Method* on air quality improvement, the mean values of AQI of November, December, and January were compared for the same period among the year before the release (2015), the year of release (2016), and the year after the release (2017). The results are presented in **Table 2**. The data show that the AQI declined immediately in the next month after the *Method*'s release. However, the year-on-year growth rates of AQI were all positive in December and January in the pre-*Method* period and the year of release. It can be inferred that the *Method* is effective in improving air quality in a short time, but its continuous effect is relatively nominal. However, the effectiveness of the *Method* cannot be clearly estimated from the raw descriptive statistics, and a more comprehensive analysis needs to be done.

## Model Specification

Studies designed as a quasi-experiment to assess the public policy always encounter the problem of endogeneity. In other words, individual behavior is usually affected by other factors before and after the introduction of the policy, which does not meet the exogenous conditions of a natural experiment. The resulting problem of endogeneity is difficult to completely overcome in public policy evaluation. The so-called policy endogeneity problem is usually a case of "manipulation" and "heap" of data (Chen L. et al., 2019). The biggest difficulty in evaluating the effect of the *Method*'s implementation lies in evaluating its own effect excluding the impact of other policies, especially due to the *Method* requiring ecological progress evaluated annually and assessed every 5 years. Big data's role in the policy evaluation process is essentially the same as that of the corresponding traditional sample data. Thus, common policy evaluation methods can be adopted when big data are used to evaluate public policy evaluation. Lee (2008) believes that the method of RD design can effectively avoid the endogeneity problem in a parameter estimation to truly reflect the causal link between two variables when a randomized experiment cannot be carried out. Since the *Method* start date can be regarded as a mutation variable

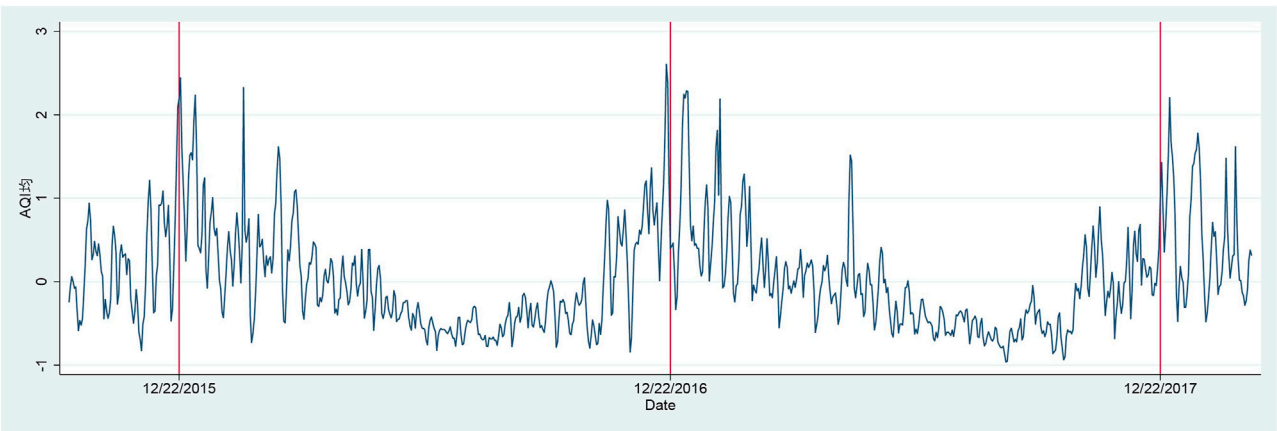



**FIGURE 1 |** Changes in national AQI (mean).

**TABLE 2 |** AQI comparison for the same period of the month (2015–2017).

| | 2015 | 2016 (release time) | 2017 | Growth rate between 2015 and 2016 (%) | Growth rate between 2017 and 2016 (%) |
|---|---|---|---|---|---|
| November | 86.013 | 106.279 | 81.818 | 23.562 | –23.015 |
| December | 113.198 | 105.536 | 106.553 | –6.769 | 0.954 |
| January | 81.613 | 100.184 | 106.552 | 22.755 | 6.356 |

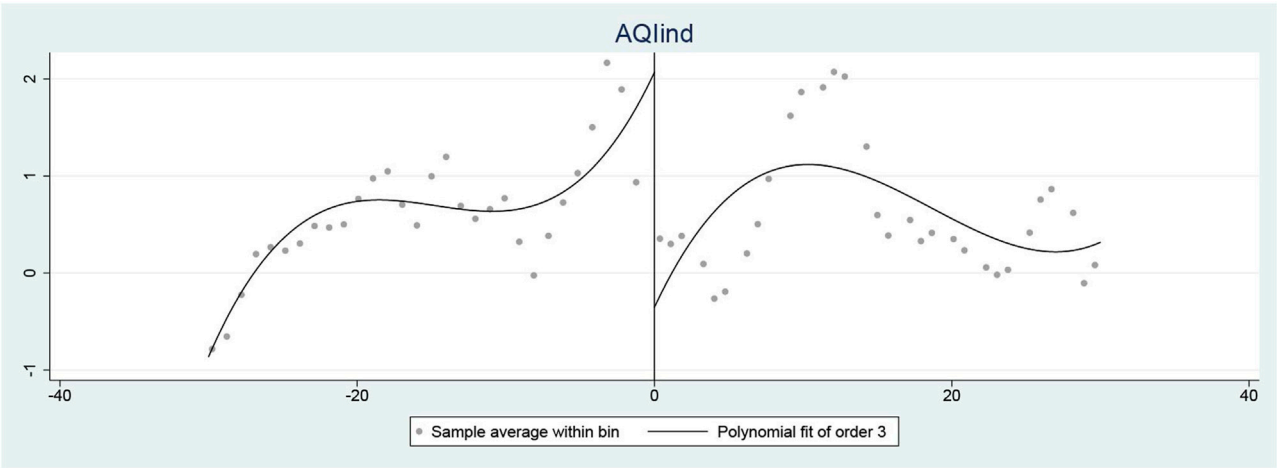



**FIGURE 2 | (A)** Changes in the concentration of AQI 30 days before and after the Method was released. **(B)** Changes in the concentration of various pollutants 30 days before and after the Method was released.

**TABLE 3 |** Results of the sharp RD design for AQI and six pollutant indicators.

| Variables | AQIind | $hPM_{2.5}avg$ | $hPM_{10}avg$ | $hSO_2avg$ | $hNO_2avg$ | hCOavg | $hO_3avg$ |
|---|---|---|---|---|---|---|---|
| RD_Estimate | –0.683*** | –0.762*** | –0.574*** | –0.268*** | –0.488*** | –0.378*** | 0.0553*** |
| — | (0.0476) | (0.0553) | (0.0441) | (0.0393) | (0.0332) | (0.0646) | (0.0182) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1. At this point, policy variables are used as dummy variables to cross and multiply with temporal trend variables to ensure the differentiated trend of air quality before and after policies.*

of influence, the method of RD design can distinguish the interaction and effect of the *Method* from other continuous variables. In this way, we can analyze the causal impact of the *Method* implementation on the jump cut point of outcome variables. In this article, if we can accurately identify whether there are abrupt changes in various pollutant concentration indicators before and after the *Method* start date while the control variables are continuously changing by using RD, it can be concluded that the *Method*'s implementation led to the changes in air quality, thereby proving the *Method*'s effectiveness. Otherwise, the *Method* is inefficient.

To analyze the influence of the air pollution control policy on the air quality, it is assumed that the improvement of air quality relies solely on the implementation of the relevant policy that could be set as a binary treatment variable, for which treatment is assigned to units with the consistent variable exceeding the cut point c. Let the receipt of treatment be denoted by the binary variable $D$, and let $D = 1$ if treatment is received and $D = 0$ otherwise. When $c = 0$,

$$D_i = \begin{cases} 1, x_i \geq c \\ 0, x_i < c \end{cases} \quad (1)$$

The cities affected by the policies or not are denoted, respectively, by ($y_{0i}$,$y_{0i}$). Then, we need to compare whether average outcome gains resulting from treatment differ from both groups. Considering the time period (c−ε, c+ε), it could be assumed that all other factors determining $Y$ of observations

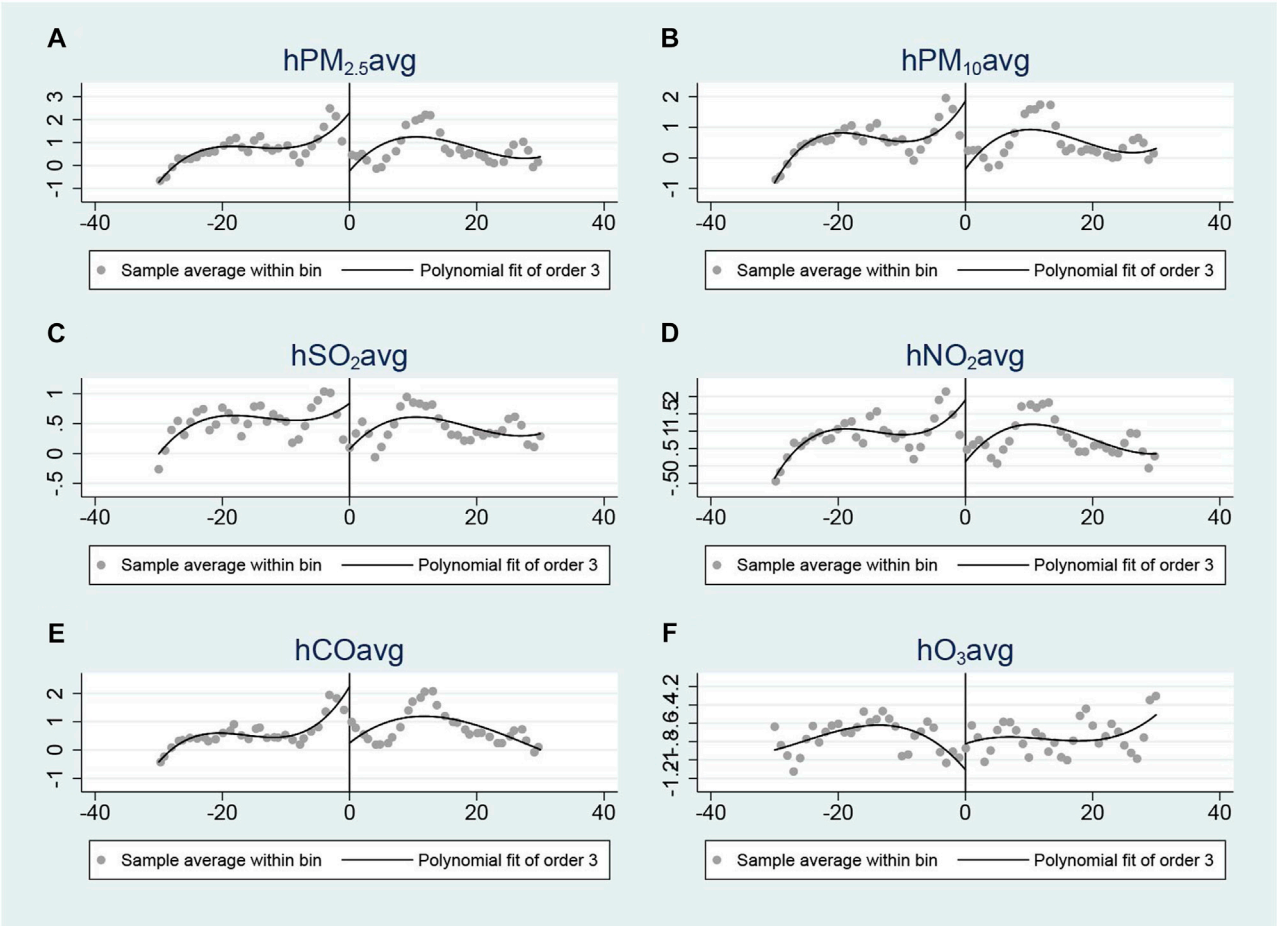



**FIGURE 2 | Continued.**

are evolving smoothly with respect to $x$ close to the cutoff. The observations after the cut point can be divided into the treatment group, while those before the cut point can be divided into the control group.

Since the *Method* was released on a specific date, we use sharp RD design to study the impact of the *Method*'s implementation on air quality referring to RD design. The regression estimation equation is given below:

$$AIR_{it} = \alpha_0 + \alpha_1 Policy + \alpha_2 D(t)_\tau + \alpha_3 PolicyD(t)_\tau + \lambda X_{it} + \mu_i + \delta_t + \varepsilon_{it}, \quad (2)$$

where $AIR_{it}$ is the outcome variable representing the air quality index (AQI) and the daily average concentrations of six pollutants (including $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, CO, and $O_3$) at city $i$ on date $t$; *Policy* represents a dummy variable for the *Method*, whose value is 0 before the start date; otherwise, it is 1; $D(t)_\tau$ is a time trend variable, which is used to fit the trend in air quality changes in sample cities before and after the *Method* start date. We multiply the dummy variable *Policy* and the time trend variable $D(t)_\tau$ to ensure the differentiation trend of air quality before and after the *Method* start date. $\tau$ is the time indicator variable. Limited to the range of the bandwidth, the value of $\tau$ of 0 represents the *Method*'s start date, 1 represents the following day, −1 represents the previous day, −2 represents the date 2 days before the release, and the like. $X_{it}$ is denoted as the control variable, meteorological and economic factors influencing air quality, including the highest temperature (Htmpt), the lowest temperature (Ltmpt), weather conditions (Wthcdt), wind force (Wndfrc), gross regional products (Grorgpdut), per capita gross regional products (PcptlGRP), and GRP growth rate (GRPgwrate). $\mu_i$ represents the regional fixed effect, including public transportation systems, green areas, and populations. $\delta_t$ represents the time fixed effect, and $\varepsilon_{it}$ represents the random disturbances.

# RESULTS

## Discussion on the Validity of the Cut Point

Normally, after selecting a certain bandwidth, if the outcome variable presents a sudden change at the cut point, the effectiveness of adopting RD design is considered to be verified and vice versa. As for determining the bandwidth, considering the fact that local officials might pursue the

**TABLE 4 |** Results of the sharp RD design with control covariates.

| Variables | AQIind | $hPM_{2.5avg}$ | $hPM_{10avg}$ | $hSO_{2avg}$ | $hNO_{2avg}$ | $hCO_{avg}$ | $hO_{3avg}$ |
|---|---|---|---|---|---|---|---|
| RD_Estimate | −0.580*** | −0.658*** | −0.482*** | −0.211*** | −0.356*** | −0.297*** | 0.0139 |
| | (0.0489) | (0.0566) | (0.0453) | (0.0388) | (0.0328) | (0.0647) | (0.0181) |
| Covariables | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

**TABLE 5 |** Results of the control variable continuity test.

| Variables | Htmpt | Ltmpt | Wthcdt | Wndfrc | Grorgpdut | PcptlGRP | GRPgwrate |
|---|---|---|---|---|---|---|---|
| RD_Estimate | −0.0937*** | −0.115*** | −0.0667*** | 0.0913*** | 0.000301 | 3.72e−05 | 0.000187 |
| — | (0.00902) | (0.00815) | (0.0127) | (0.0337) | (0.00451) | (0.000796) | (0.00206) |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

**TABLE 6 |** Results of different bandwidths.

| Variables | AQIind | $hPM_{2.5}avg$ | $hPM_{10}avg$ | $hSO_2avg$ | $hNO_2avg$ | hCOavg | $hO_3avg$ |
|---|---|---|---|---|---|---|---|
| RD_Estimate | −1.628*** | −1.750*** | −1.458*** | −0.549*** | −1.196*** | −1.188*** | 0.159*** |
| — | (0.0716) | (0.0833) | (0.0675) | (0.0558) | (0.0493) | (0.105) | (0.0260) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |
| **VARIABLES** | **AQIind** | **$hPM_{2.5avg}$** | **$hPM_{10avg}$** | **$hSO_{2avg}$** | **$hNO_{2avg}$** | **$hCO_{avg}$** | **$hO_{3avg}$** |
| RD_Estimate | −0.684*** | −0.716*** | −0.632*** | −0.351*** | −0.478*** | −0.305*** | −0.0606*** |
| — | (0.0322) | (0.0365) | (0.0291) | (0.0279) | (0.0230) | (0.0403) | (0.0129) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

short-term performance and referring to the existing literature (Tu et al., 2020), this study picks the bandwidth of 30 days. After controlling for the effects of meteorological and economic variables on the national air quality, air pollutant concentration trends were fitted for the 30 days before and after the *Method*'s release. To estimate the model's effectiveness, whether the *Method's* start date (the treatment date) can be seen as a cut point should be examined. All the results are shown in **Figure 2**. It is illustrated that except $O_3$, the concentrations of $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, and CO demonstrate obvious sudden changes at the start date, which suggests that the decrease in AQI might be owing to the sharp decrease in the five pollutants. On the other hand, $O_3$ is an indirect emission that is not directly emitted by the residents or factories. Its generation process is sophisticated, and no sudden change was observed. Therefore, it is reasonable to conduct the empirical analysis based on a sharp RD design.

## Baseline Regression

The selection of an optimal bandwidth involves a trade-off between accuracy and bias in linear hypotheses. On the one hand, using a larger bandwidth yields more precise estimates as more observations are available to estimate the regression. On the other hand, the linear specification is less likely to be accurate when a larger bandwidth is used, which can biased in the estimation of the treatment effect (Lee and Lemieux, 2010).

**TABLE 7 |** "Donut hole" test results.

| Variables | AQIind | $hPM_{2.5avg}$ | $hPM_{10avg}$ | $hSO_{2avg}$ | $hNO_{2avg}$ | $hCO_{avg}$ | $hO_{3avg}$ |
|---|---|---|---|---|---|---|---|
| RD_Estimate | –0.686*** | –0.766*** | –0.578*** | –0.273*** | –0.476*** | –0.370*** | 0.0582*** |
| (Excluding first-tier cities) | (0.0476) | (0.0554) | (0.0445) | (0.0402) | (0.0329) | (0.0645) | (0.0184) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 138,918 | 138,918 | 138,918 | 138,918 | 138,918 | 138,918 | 138,918 |
| **VARIABLES** | **AQIind** | **$hPM_{2.5avg}$** | **$hPM_{10avg}$** | **$hSO_{2avg}$** | **$hNO_{2avg}$** | **$hCO_{avg}$** | **$hO_{3avg}$** |
| RD_Estimate | –8.339*** | –7.717*** | –7.544*** | –2.018*** | –7.218*** | –5.957*** | 0.571*** |
| (Excluding 5% of samples near the cut point) | (0.406) | (0.410) | (0.364) | (0.393) | (0.305) | (0.364) | (0.217) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 134,421 | 134,421 | 134,421 | 134,421 | 134,421 | 134,421 | 134,421 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

**TABLE 8 |** Persistence analysis of the Method.

| Variables | AQIind | $hPM_{2.5avg}$ | $hPM_{10avg}$ | $hSO_{2avg}$ | $hNO_{2avg}$ | $hCO_{avg}$ | $hO_{3avg}$ |
|---|---|---|---|---|---|---|---|
| RD_Estimate | –0.191*** | –0.204*** | –0.298*** | –0.0754*** | –0.262*** | 0.0415** | 0.173*** |
| (Bandwidth:180) | (0.0181) | (0.0195) | (0.0163) | (0.0186) | (0.0131) | (0.0202) | (0.00921) |
| RD_Estimate | 0.299*** | 0.320*** | 0.196*** | 0.311*** | 0.305*** | 0.451*** | –0.266*** |
| (Bandwidth:360) | (0.0119) | (0.0126) | (0.0108) | (0.0130) | (0.00901) | (0.0127) | (0.00799) |
| RD_Estimate | 0.433*** | 0.481*** | 0.326*** | 0.439*** | 0.469*** | 0.595*** | –0.471*** |
| (Bandwidth:540) | (0.01000) | (0.0105) | (0.00920) | (0.0108) | (0.00773) | (0.0104) | (0.00731) |
| RD_Estimate | 0.468*** | 0.522*** | 0.360*** | 0.470*** | 0.512*** | 0.630*** | -0.523*** |
| (Bandwidth:720) | (0.00964) | (0.0101) | (0.00892) | (0.0104) | (0.00751) | (0.00991) | (0.00729) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 | 142,206 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

There are multiple outcome variables in this article, so there is a different optimal bandwidth for each one. Therefore, a bandwidth of 30 days is selected for unified regression estimation to better evaluate the short-term effect of the *Method*.

**Table 3** illustrates the baseline sharp RD design results of the AQI and six pollutant indicators with a bandwidth of 30 days prior to and following the cut point. Except for $O_3$, the estimated coefficients for AQI and five pollutant indicators are significantly negative, which proves that the *Method* implementation can significantly improve the air quality. Due to the local governments' quick response to the *Method* through the strong executive power, the significant improvement effect of air quality was achieved in a short period of time. For one thing, they strengthened the implementation of air pollution control. For another, they supervised enterprises' pollution emissions strictly. However, the *Method* implementation has not been able to significantly reduce the $O_3$ concentration because lower atmospheric $O_3$ is a secondary product of automobile pollutant emission under the irradiation of sunlight.

## Robustness Test

The estimator of RD design is considered to be biased if the assignment variable is manipulated or a large observation heap is found (Barreca et al., 2011). To ensure that the estimation results are unbiased and valid, the following test approaches are employed:

1) Examine the influence of incorporating the covariates;
2) Examine the continuity of pretreatment covariates, which means to examine whether the distribution of covariates also exhibits discontinuity at the cut point;
3) Examine the influence of different bandwidth choices;
4) Examine sample selection sensibility, also known as the "donut hole" test, which reveals the results of omitting the observations that are vulnerable to be selected for certain reasons.

The robustness tests are conducted as follows.

First, the influence of incorporating the covariates is examined. Since the influence of the covariates is ignored in the baseline regression, according to the article contributed by Lee and Lemieux (2010), and the inclusion of covariates in the regression should reduce the residual square of the model and increase the interpretation efficiency but should not exhibit

**TABLE 9 |** Sharp RD design results for the seven regions of China.

| Regions | Eastern China | Southern China | Central China | Northern China | Northwest of China | Southwest of China | Northeast of China |
|---|---|---|---|---|---|---|---|
| AQIind | −0.378*** | −0.117*** | −0.881*** | −1.608*** | −1.385*** | −1.312*** | −0.934*** |
| — | (0.0744) | (0.0413) | (0.119) | (0.188) | (0.211) | (0.195) | (0.109) |
| h$PM_{2.5}$avg | −0.389*** | −0.100** | −1.085*** | −1.926*** | −1.428*** | −1.349*** | −0.957*** |
| — | (0.0771) | (0.0407) | (0.158) | (0.232) | (0.239) | (0.222) | (0.108) |
| h$PM_{10}$avg | −0.344*** | −0.136*** | −0.698*** | −1.466*** | −1.017*** | −0.962*** | −0.746*** |
| | (0.0652) | (0.0341) | (0.116) | (0.191) | (0.173) | (0.160) | (0.0924) |
| h$SO_2$avg | −0.179*** | −0.0212 | −0.161*** | −1.025*** | −0.208*** | −0.197*** | −0.520*** |
| — | (0.0308) | (0.0220) | (0.0443) | (0.280) | (0.0797) | (0.0749) | (0.122) |
| h$NO_2$avg | −0.441*** | −0.401*** | −0.455*** | −1.115*** | −0.427*** | −0.419*** | −0.514*** |
| — | (0.0500) | (0.0649) | (0.0622) | (0.144) | (0.116) | (0.107) | (0.0753) |
| hCOavg | 0.0510 | 0.0736** | −0.282*** | −2.135*** | −0.658*** | −0.630*** | −0.523*** |
| — | (0.0660) | (0.0343) | (0.100) | (0.412) | (0.142) | (0.133) | (0.124) |
| h$O_3$avg | 0.00722 | −0.0605 | 0.105*** | 0.207*** | −0.00145 | 0.00462 | 0.244*** |
| — | (0.0302) | (0.0492) | (0.0380) | (0.0496) | (0.0397) | (0.0382) | (0.0450) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Bandwidth | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Observations | 43,566 | 22,194 | 25,482 | 16,440 | 9,042 | 9,864 | 13,974 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

**TABLE 10 |** Persistence analysis for the seven regions of China.

| Regions | Eastern China | Southern China | Central China | Northern China | Northwest of China | Southwest of China | Northeast of China |
|---|---|---|---|---|---|---|---|
| Variables | AQIind | AQIind | AQIind | AQIind | AQIind | AQIind | AQIind |
| RD_Estimate | −0.0301 | 0.378*** | −0.201*** | −0.368*** | −0.654*** | −0.622*** | −0.312*** |
| (Bandwidth:180) | (0.0253) | −0.104*** | (0.0398) | (0.0703) | (0.0907) | (0.0845) | (0.0426) |
| RD_Estimate | 0.412*** | 0.0840*** | 0.337*** | 0.446*** | 0.171*** | 0.151*** | 0.175*** |
| (Bandwidth:360) | (0.0159) | (0.0124) | (0.0254) | (0.0457) | (0.0620) | (0.0580) | (0.0281) |
| RD_Estimate | 0.532*** | 0.144*** | 0.525*** | 0.624*** | 0.397*** | 0.366*** | 0.277*** |
| (Bandwidth:540) | (0.0134) | (0.0107) | (0.0211) | (0.0378) | (0.0526) | (0.0492) | (0.0241) |
| RD_Estimate | 0.562*** | 0.162*** | 0.575*** | 0.667*** | 0.455*** | 0.421*** | 0.301*** |
| (Bandwidth:720) | (0.0129) | (0.0103) | (0.0203) | (0.0361) | (0.0507) | (0.0475) | (0.0234) |
| Covariables | No | No | No | No | No | No | No |
| Urban fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time fixed effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 43,566 | 22,194 | 25,482 | 16,440 | 9,042 | 9,864 | 13,974 |

*Note: Numbers in parentheses are standard errors. ***p < 0.01, **p < 0.05, and *p < 0.1.*

significant differences in estimation coefficients. Therefore, the covariates are included in the RD design as a part of the robustness test. The experimental results in **Table 4** show that the results remain roughly constant compared to the baseline regression results, which supports the validity of the results.

Second, the continuity of pretreatment covariates should be checked, which means that the baseline covariates should be continuous at the threshold. It is necessary to examine the possibility that other changes are occurring at cutoff, which could confound our analysis. That is to say, changes in the air pollutant concentration might be caused by other factors, and the assumption that the *Method*'s release significantly affects the air pollutant concentration could not be proved. It is reasonable to assume that the treatment effect illustrated in the model is caused by the assignment variable when the control variables are continuous and stable around the cut point.

The results are shown in **Table 5**. It exhibits the significant influence of the highest temperature, the lowest temperature, weather conditions, and wind force on air quality before and after the *Method*'s release, while the other variables did not, such as Grorgpdut, PcptlGRP, and GRPgwrate. Considering the lowest temperature, its coefficient is negative, which suggests that air pollution became worse as the temperature dropped. The change in the lowest temperature should have nothing to do with the remarkable improvement in air quality. Although weather conditions and wind force had significant coefficients, their values are −0.0667 and 0.0913, respectively, which should be relatively too small to contribute to the notable improvement of air quality. In conclusion, none of the control variables jumps at the cut point generally and should not affect the results, which further convinces that the release of the *Method* improved the air quality.

Third, various bandwidth selections are tested. When bandwidths are narrow, they tend to guarantee the unbiasedness of the local linear regression. However, narrow bandwidths would also enlarge estimation variance, and the

regression results might remain inconsistent as bandwidths expand. Therefore, the study has re-estimated, respectively, with bandwidths 0.5 times and 2 times as wide as that in the baseline regression (Lee and Lemieux, 2010; Tu et al., 2020)—that is, 15 and 60 days before and after the cut point. The results of the treatment effect under the two bandwidths are shown in **Table 6**. The results under two different bandwidths are essentially consistent with those in the baseline regression, which shows that the result is not sensitive to the bandwidth choices and confirms the robustness of the results.

Fourth, the sample selection sensitivity, which is of great importance to guarantee the validity of RD, has been tested. The present study tries to verify it from both the perspective of the economic factors and the release time of the *Method*. First, compared to other regions, economically developed cities could lower the AQI more effectively, relying on their better industrial structure, stronger public awareness of environmental protection, and more advanced pollution treatment technologies rather than policy incentives. However, the other cities may count more on policy promotion to achieve the purpose of air pollution control. Considering the gap in environmental governance at the city level, samples of the first-tier cities are excluded to examine the robustness. Second, the closer the time is to the *Method*'s release, the higher possibility the air quality is to be manipulated during the period close to the cut point. Therefore, 5% of the observations close to the cut point are omitted in order to avoid a very steep change in air quality. The test results are shown in **Table 7**. The results remain generally consistent compared with the baseline regression results, which support their validity.

### Policy Effect Persistence

Since the baseline regression model limits the sample to a 30-day bandwidth, only the short-term effects of the *Method* can be shown. The persistence or time variation of the *Method* effectiveness is also of particular interest.

**Table 8** shows regression results of the AQI and six air pollutants in the preceding and following half-year (180 days), year (360 days), one and a half years (540 days), and 2 years (720 days) from the release of the *Method*. It is suggested that the *Method* can effectively improve air quality in the next half a year after its release, but the coefficients turn positive despite the treatment effect being significant 1 year after the release. In addition, the coefficient of AQI increases by 0.134 from 1 year to one and a half years but increases only 0.035 from one and a half years to 2 years. Although the coefficient takes on a slightly declining trend, the findings still indicate the concern that the air quality improvement effects have weakened after the assessment ends.

## Heterogeneity Effects

It is known that China has a vast territory with discrepancies of air pollution control among different regions, which might result in varied effects from the *Method*. Regional disparity should also be taken into consideration (Wu et al., 2021). To do so, we have classified the country into seven regions as East, South, Central, North, Northwest, Southwest, and Northeast China according to the locations. Then, the AQI and the six air pollution indicators by the sharp RD design on these groups are estimated, and the policy persistence analysis is performed as well.

**Table 9** and **Table 10** illustrate the evidence for the heterogeneity of the policy effect on AQIind and the other indicators along with the results of the policy effect persistence analysis across regions with different air pollution levels. In the model with a bandwidth of 30, the treatment effect in the seven regions is significant overall, which indicates that the remarkable and sudden reduction in air pollutant concentration should be attributed to the fact that the local government of different regions actively responded to the *Method* and strengthened the regulations. Especially, the *Method* allows local governments to decompose the assessment targets scientifically and reasonably according to the regional discrepancy factors including the economic development level, resources, natural endowment, and other factors of all regions.

However, the result presented in **Table 10** also reveals that 1 year after the *Method*'s release, the coefficients of AQIind in each region have become positive, which reaffirms that the effect of the *Method* decayed throughout the months and years afterward. Although the coefficient has decreased relatively a little, the necessities for a more sustainable air pollution control system are still suggested.

# CONCLUSION AND DISCUSSION

This study uses environmental big data on seven air quality indicators from 173 Chinese cities at the prefecture level to construct a Sharp RD design and evaluates the effectiveness of the top-down environmental governance instrument, namely, *Assessment and Appraisal Method for Ecological Civilization Construction Targets* (the *Method*) on air quality across the country and the regions. The persistence effect is tested, and heterogeneity effects are analyzed further. The empirical analysis results include the following:

1) In the short term, the *Method* has had an overall significant treatment effect on various indicators nationally, which is consistent with the findings from many other relevant studies that one policy released by the central government will be very effective in the short term. This article examines the changes of air pollutant concentration 30 days before and after the release of the *Method*, and almost all single-pollutant indicators had significantly negative coefficients, except for the concentration of $O_3$ having a positive one. In addition, the national AQI decreased significantly by 0.683 units, indicating that the *Method* meets the expectations of improving overall national air quality in the short term.
2) In the long term, the *Method* has a less significant effect of improvement. The treatment effect remains significant nationwide half a year after the *Method* promulgated. The regression coefficients of all indicators become positive when taking intervals of a year, a year and a half, and 2 years after the *Method* though. Although the coefficient has declined slightly in the long run, the findings still illustrate that the air quality improvement effects lessen as time goes on.

3) The *Method*'s effect on air quality improvement and its persistency are generally similar across different regions. Especially, compared with some previous studies which did not show significant impacts on all the regions, in a short period, East, South, Central, North, Northwest, Southwest, and Northeast China all show a significant improvement in air quality in this study. However, the *Method*'s effect is also found to decline over time according to the regional persistence analysis.

Based on this study's empirical results, the following suggestions are proposed for the further improvement of air quality: first, considering effects of the environmental policies, which are effective at the beginning but decay as time goes by, it is necessary to formulate continuous and dynamic regulations for long-term air pollution control. The air pollution treatment is a vast but precise systematic project, and its success requires top-level institutions, effective execution, and good coordination with other relative regulations. Second, local governments of some regions should support innovations and optimizations in the technology, equipment, and service modes and improve related policies and mechanisms so as to facilitate the sustainable development of energy-efficient and environmentally friendly industries. Third, an optimized system of incentives for improved pollution control should be upgraded. The Government should encourage research and development on the industrial application of new technologies and equipment, such as those for low-grade waste heat power generation, fine particular matter control, and coordinated multi-pollutant treatment.

This study expounds on the effectiveness of the *Method* at the first stage and implies that a more sustainable air pollution governance system is necessary. Some recommendations on the formulation and improvement of environmental policies for governments are suggested. Several future research directions for this study are likewise presented. The continuous effect can be analyzed through empirical research based on a longer period, which may provide more precise and practical results. How to establish an effective and robust long-term environmental protection mechanism replacing traditional campaign-style policy enforcement could also be a challenging and meaningful direction for future research.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: all data generated or analyzed during this study will be made available by the authors. Further inquiries can be directed to the corresponding author. Requests to access these data sets should be directed to xux@th.btbu.edu.cn.

## AUTHOR CONTRIBUTIONS

RW conceived of the presented idea and supervised the findings. JH and LZ collected data and performed the computations. YX, XX, and TN revised them critically. All authors have made a substantial and intellectual contribution to this article and approved it for publication.

## FUNDING

This research was funded by the National Social Science Foundation of China (Grant No. 19BJY161), the Key program Provincial Education Department of Hunan Province (Grant No. 20A122 and No.15A191), and the Natural Science Foundation of Hunan Province Information (Grant No.2021JJ30197).

## ACKNOWLEDGMENTS

We gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the article.

## REFERENCES

Al_Janabi, S., Yaqoob, A., and Mohammad, M. (2019). "Pragmatic Method Based on Intelligent Big Data Analytics to Prediction Air Pollution," in *Big Data and Networks Technologies. BDNT 2019. Lecture Notes in Networks and Systems*. Editor Y. Farhaoui (Cham: Springer), Vol. 81, 84–109. doi:10.1007/978-3-030-23672-4_8

Barreca, A. I., Guldi, M., Lindo, J. M., and Waddell, G. R. (2011). Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification. *Q. J. Econ.* 126 (4), 2117–2123. doi:10.1093/qje/qjr042

Calkins, C., Ge, C., Wang, J., Anderson, M., and Yang, K. (2016). Effects of Meteorological Conditions on Sulfur Dioxide Air Pollution in the North China plain during winters of 2006-2015. *Atmos. Environ.* 147, 296–309. doi:10.1016/j.atmosenv.2016.10.005

Chang, C.-P., Wen, J., Dong, M., and Hao, Y. (2018). Does Government Ideology Affect Environmental Pollutions? New Evidence from Instrumental Variable Quantile Regression Estimations. *Energy Policy* 113, 386–400. doi:10.1016/j.enpol.2017.11.021

Chen, L., Wan, P. B., and Xu, Y. Y. (2019). The Ownership Structure and Innovation Behavior of Mixed Ownership Enterprises—Based on the Empirical Test of Natural experiment and Regression Discontinuity. *Management World* 35 (10), 186–205. doi:10.19744/j.cnki.11-1235/f.2019.0141

Chen, S., Oliva, P., and Zhang, P. (2018). Air Pollution and Mental Health: Evidence from china. *Nber Working Pap*. doi:10.3386/w24686

Chen, S., Zhang, Y., Zhang, Y., and Liu, Z. (2019). The Relationship Between Industrial Restructuring and China's Regional Haze Pollution: A Spatial Spillover Perspective. *J. Clean. Prod.* 239, 115808. doi:10.1016/j.jclepro.2019.02.078

Chen, X., Yi, N., Zhang, L., and Li, D. (2018). Does Institutional Pressure Foster Corporate Green Innovation? Evidence From China's Top 100 Companies. *J. Clean. Prod.* 188, 304–311. doi:10.1016/j.jclepro.2018.03.257

Fan, D., and Sun, X. T. (2020). Environmental Regulation, green Technological Innovation and green Economic Growth. *China Popul. Resour. Environ.* 30 (6), 105–115. doi:10.12062/cpre.20200123

Farhaoui, Y., Sossi Alaoui, S., and Aksasse, B. (2019). Air Pollution Prediction through Internet of Things Technology and Big Data Analytics. *Ijcistudies* 8 (3), 177–191. doi:10.1504/IJCISTUDIES.2019.10024282

Heckman, J. (1997). Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *J. Hum. Resour.* 32, 441–462. doi:10.2307/146178

Hou, X., Chan, C. K., Dong, G. H., and Yim, S. H. L. (2019). Impacts of Transboundary Air Pollution and Local Emissions on PM2. 5 Pollution in the Pearl River Delta Region of China and the Public Health, and the Policy Implications. *Environ. Res. Lett.* 14 (3), 034005. doi:10.1088/1748-9326/aaf493

Jebaraj, S., and Iniyan, S. (2006). A Review of Energy Models. *Renew. Energ.* 10 (4), 281–311. doi:10.1016/j.rser.2004.09.004

Jia, K., and Chen, S. (2019). Could Campaign-Style Enforcement Improve Environmental Performance? Evidence from China's central Environmental protection Inspection. *J. Environ. Manage.* 245, 282–290. doi:10.1016/j.jenvman.2019.05.114

Kai-guang, Z., Ming-Ting, B., Hong-ling, M., and Yan-min, S. (2019). An Air Quality Composite index Based on Pollutant Concentration Factor Analysis. *Afr. J. Math. Comput. Sci. Res.* 12 (2), 10–16. doi:10.5897/AJMCSR2019.0788

Lee, D. S., and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *J. Econ. Lit.* 48 (2), 281–355. doi:10.1257/jel.48.2.281

Lee, D. S. (2008). Randomized Experiments from Non-random Selection in US House Elections. *J. Econom.* 142 (2), 675–697. doi:10.1016/j.jeconom.2007.05.004

Li, D., Zheng, M., Cao, C., Chen, X., Ren, S., and Huang, M. (2017). The Impact of Legitimacy Pressure and Corporate Profitability on Green Innovation: Evidence From China Top 100. *J. Clean. Prod.* 141, 41–49. doi:10.1016/j.jclepro.2016.08.123

Li, N., Zhang, X., Shi, M., and Hewings, G. J. D. (2019). Does China's Air Pollution Abatement Policy Matter? an Assessment of the Beijing-Tianjin-Hebei Region Based on a Multi-Regional CGE Model. *Energy Policy* 127, 213–227. doi:10.1016/j.enpol.2018.12.019

List, J. A., Millimet, D. L., Fredriksson, P. G., and McHone, W. W. (2003). Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator. *Rev. Econ. Stat.* 85 (4), 944–952. doi:10.1162/003465303772815844

Neves, S. A., Marques, A. C., and Patrício, M. (2020). Determinants of CO2 Emissions in European Union Countries: Does Environmental Regulation Reduce Environmental Pollution? *Econ. Anal. Pol.* 68, 114–125. doi:10.1016/j.eap.2020.09.005

Niu, Y., Chen, R., and Kan, H. (2017). "Air Pollution, Disease burden, and Health Economic Loss in China," in *Ambient Air Pollution and Health Impact in China. Advances in Experimental Medicine and Biology*. Editor G. H. Dong (Singapore: Springer), Vol. 1017. doi:10.1007/978-981-10-5657-4_10

Ouyang, X., Shao, Q., Zhu, X., He, Q., Xiang, C., and Wei, G. (2019). Environmental Regulation, Economic Growth and Air Pollution: Panel Threshold Analysis for OECD Countries. *Sci. Total Environ.* 657, 234–241. doi:10.1016/j.scitotenv.2018.12.056

Rohde, R. A., and Muller, R. A. (2015). Air Pollution in China: Mapping of Concentrations and Sources. *PLoS One* 10 (8), e0135749. doi:10.1371/journal.pone.0135749

Shou, M. H., Wang, Z. X., Li, D. D., and Wang, Y. (2020). Assessment of the Air Pollution Emission Reduction Effect of the Coal Substitution Policy in China: an Improved Grey Modelling Approach. *Environ. Sci. Pollut. Res. Int.* 27 (27), 34357–34368. doi:10.1016/j.envpol.2017.01.060

Song, C., He, J., Wu, L., Jin, T., Chen, X., Li, R., et al. (2017). Health burden Attributable to Ambient PM2.5 in China. *Environ. Pollut.* 223, 575–586. doi:10.1016/j.envpol.2017.01.060

Tan, Y., and Mao, X. (2021). Assessment of the Policy Effectiveness of Central Inspections of Environmental Protection on Improving Air Quality in China. *J. Clean. Prod.* 288, 125100. doi:10.1111/j.1477-2574.2009.00087.x

Tu, Z. G., Deng, H., Shen, R. J., and Gan, T. Q. (2020). The Environmental and Economic Benefits of the central Environmental protection Supervision: Evidence from Policy Pilot in Hebei. *Econ. Rev.* 221 (01), 5–18. doi:10.19361/j.er.2020.01.01

Wang, L., Liu, X. F., and Xiong, Y. (2019). Central Environmental protection Inspector and Air Pollution Governance—An Empirical Analysis Based on Micro-panel Data of Prefecture-Level Cities. *China Ind. Economy* 379 (10), 7–24.

Wang, Y., and Wang, J. (2020). Tropospheric SO2 and NO2 in 2012–2018: Contrasting Views of Two Sensors (OMI and OMPS) from Space. *Atmospheric Environment* 223, 117214. doi:10.1016/j.atmosenv.2019.117214

Wu, W.-p., Chen, Z.-g., and Yang, D.-x. (2020). Do internal Migrants Crowd Out Employment Opportunities for Urban Locals in China?-Reexamining under the Skill Stratification. *Physica A: Stat. Mech. its Appl.* 537, 122580. doi:10.1016/j.physa.2019.122580

Wu, W.-p., Zeng, W.-k., Gong, S.-w., and Chen, Z.-g. (2021). Does Energy Poverty Reduce Rural Labor Wages? Evidence from China's Rural Household Survey. *Front. Energ. Res.* 9, 163. doi:10.3389/fenrg.2021.670026

Wu, W. P., and He, Q. (2017). "Forced Pollution Reduction" or "Regressive"? —Threshold Characteristic and Spatial Spillover of Pollution Reduction Effect from Environmental Regulation. *Econ. Management* 39 (02), 20–34. doi:10.19616/j.cnki.bmj.2017.02.002

XinChun, L., YuTing, Z., Qing, H., XingHua, Y., Ali, M., Wen, H., et al. (2013). Vertical Distribution Characteristics of Dust Aerosol Mass Concentration in the Taklimakan Desert Hinterland. *Sci. Cold Arid Regions* 5 (6), 745–0754. doi:10.3724/SP.J.1226.2013.00745

Xu, Y., Liu, H., and Duan, Z. (2020). A Novel Hybrid Model for Multi-step Daily AQI Forecasting Driven by Air Pollution Big Data. *Air Qual. Atmos. Health* 13 (2), 197–207. doi:10.1007/s11869-020-00795-w

Yang, W., Yuan, G., and Han, J. (2019). Is China's Air Pollution Control Policy Effective? Evidence from Yangtze River Delta Cities. *J. Clean. Prod.* 220, 110–133. doi:10.1016/j.jclepro.2019.01.287

Yu, X., and Wang, P. (2021). Economic Effects Analysis of Environmental Regulation Policy in the Process of Industrial Structure Upgrading: Evidence from Chinese Provincial Panel Data. *Sci. Total Environ.* 753, 142004. doi:10.1016/j.scitotenv.2020.142004

Zhang, D., Pan, S. L., Yu, J., and Liu, W. (2019). Orchestrating Big Data Analytics Capability for Sustainability: A Study of Air Pollution Management in China. *Inf. Management* 103231, 103231. doi:10.1016/j.im.2019.103231

Zhao, Y. M., Zhu, F. M., and He, L. L. (2009). Definition, Classification and Evolution of Environmental Regulations. *China Population. Resour. Environ.* 019 (006), 85–90.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Wang, Huang, Zhang, Xia, Xu and Nong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fenvs.2021.733488

# The Spatio-Temporal Evolution of China's Hydrogen Fuel Cell Vehicle Innovation Network: Evidence From Patent Citation at Provincial Level

*Pei Yu*[1], *Feng Jiang*[1], *Zhengfang Cai*[1] *and Yongping Sun*[2,3]*

[1]School of Economics, Wuhan University of Technology, Wuhan, China, [2]Center of Hubei Cooperative Innovation for Emissions Trading System, Hubei University of Economics, Wuhan, China, [3]School of Low Carbon Economics, Hubei University of Economics, Wuhan, China

OPEN ACCESS

**Edited by:**
Cosimo Magazzino,
Roma Tre University, Italy

**Reviewed by:**
Felix Haifeng Liao,
University of Idaho, United States
Fei Ma,
Chang'an University, China

***Correspondence:**
Yongping Sun
sunyp@hbue.edu.cn

**Specialty section:**
This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science

**Received:** 30 June 2021
**Accepted:** 27 August 2021
**Published:** 01 October 2021

**Citation:**
Yu P, Jiang F, Cai Z and Sun Y (2021) The Spatio-Temporal Evolution of China's Hydrogen Fuel Cell Vehicle Innovation Network: Evidence From Patent Citation at Provincial Level. Front. Environ. Sci. 9:733488. doi: 10.3389/fenvs.2021.733488

Hydrogen fuel cell vehicle industry is in a rapid development stage. Studying the domestic spatial distribution of hydrogen fuel cell vehicle industry across a country, especially the spatio-temporal evolution of the innovation level and position of each region in innovation network, will help to understand the industry's development trends and characteristics, and avoid repeated construction. This article uses social network analysis and patent citation information of 2,971 hydrogen fuel cell vehicle related invention patents, owned by 218 micro-innovators across 25 provinces of China from 2001 to 2020, to construct China's hydrogen fuel cell vehicle innovation network. Based on the dimensions of knowledge production, knowledge consumption, and network broker, the network positions of sample provinces in three periods divided by four main national policies are classified. The main findings are as follows. 1) In China, the total sales of hydrogen fuel cell vehicle and the development of supporting infrastructure are balanced, and a series of national and local industrial development polices have been issued. 2) China's hydrogen fuel cell vehicle innovation network density, the proportion of universities and research institutes among the innovators, and the active degree of the eastern provinces are all becoming higher. 3) The provinces in optimal network position are all from the eastern region. Shanghai and Liaoning are gradually replaced by Beijing and Jiangsu. 4) Sichuan in the western region is the only network broker based on knowledge consumption. 5) Although Zhejiang, Tianjin, Hebei, Guangdong, and Hubei are not yet in the optimal position, they are outstanding knowledge producers. Specifically, Guangdong is likely to climb to the optimal network position in the next period. The conclusions will help China's provinces to formulate relevant development policies to optimize industry layout and enhance collaborative innovation in the hydrogen fuel cell vehicle industry.

**Keywords: hydrogen fuel cell vehicle, patent citation network, spatio-temporal evolution, network position, innovation network density**

## 1 HIGHLIGHTS

1) China's hydrogen fuel cell vehicle innovation network density becomes higher.
2) Universities and research institutes are the main innovators in HFCV industry.
3) Provinces as knowledge producer are mainly from the east area.
4) Guangdong is climbing to optimal network position.
5) Hubei and Sichuan are outstanding in central and western areas respectively.

## 2 INTRODUCTION

Hydrogen fuel cell vehicle (HFCV) combines the advantages of short filling time, long continue voyage course, and truly achieving a zero-carbon emission goal during driving. Thus, HFCV is one of the important trends in the future of the vehicle industry. In the 1970s, the United States became the first country to use hydrogen energy as strategic energy. And then in the 1990s, Japan and South Korea started the R&D of HFCV. In recent years, more and more countries, including China, have involved in the HFCV industry. Nevertheless, due to the immaturity of commercialization in the HFCV industry, the countrie's focuses are still the industrial guidance and layout. For instance, the United States plans to form a hydrogen energy network based on 55 hydrogen-filling stations across 35 states, and various countries have successively introduced large-scale subsidies to support the R&D and demonstration applications of HFCV. Hence, studying the spatial layout of HFCV industry across a country will help to optimize the resources allocation and then enhance industrial international competitiveness.

As a major vehicle producer and consumer, China emits $CO_2$ in transportation sector with rapid growth. In 2018, the carbon emission in transportation sector accounts for 10.7% of the total emission in China (Yuan et al., 2021). Thus, the promotion of HFCV will be helpful for China to achieve the goal of carbon emissions peaking by 2030 and carbon neutrality by 2060 and will directly affect China's position in the new round of international division in manufacturing. China has participated in the global competition of HFCV industry over 20 years since the first HFCV successfully designed at the beginning of the 2000s by local firms, Shanghai Automotive Industry Corporation and Dongfeng Motor Corporation. In recent years, China makes further efforts to accelerate its development pace by releasing a series of promotion policies through national and local governments. For instance, Energy Saving and New Energy Vehicle Technology Roadmap 2.0 released by the State Council of China in 2020 proposes that the number of HFCVs in China will reach 1 million in 2030, which will exceed Japan and become the second-largest country on the total number of HFCVs in the world, just behind South Korea.

Compared with the traditional oil-fueled vehicle industry, China's HFCV industry has special characteristics. Firms in China's HFCV industry are still highly dependent on firms in traditional oil-fueled vehicle industry for technology integration and components production. Therefore, the geographic distribution of these two types of firms has similarity. To promote local economic growth, many provinces in China have vigorously developed the oil-fueled vehicle industry since the 1990s. However, it causes a lot of problems such as repeated construction, low resource allocation efficiency, and environmental pollution. Thus, in the background that China's national and local governments aim to formulate reasonable industrial development plan for HFCV industry, how to optimize the domestic layout of HFCV industry chain, for avoiding repeated construction, blind competition, and forming inter-region coordination, is one of the key issues needed to be considered in depth. Besides, HFCV industry is typically knowledge- and technology-intensive, which makes the innovation highly complex. In addition to the firms, universities and research institutes are the important knowledge producers and innovators as well in China. Therefore, how do these different organizations work together efficiently, and which provinces play the leading innovative role in China's HFCV industry are the main questions needed to be focused on as well. Hence, to work out the solutions for the above-mentioned key issues, it is necessary to have insights on the spatio-temporal evolution of China's HFCV industry in the past 20 years, specifically focusing on the regional innovation capability.

Previous related studies on the HFCV industry mainly focus on the qualitative description such as HFCV technology, HFCV industry development, and government policies, especially in Japan, South Korea, and ASEAN (Khan et al., 2021). In the case of China, the risk of blind competition among different provinces and industrial surplus due to the lack of national overall industrial plan are the main topics concerned by some researchers (Meng et al., 2020). However, the research on the spatio-temporal evolution of HFCV industry across provinces in China is nearly blank.

HFCV industry belongs to the strategic emerging industry, in which mastering core technologies independently is one of the effective ways of improving competitiveness. Owing to the "follower" role in the global traditional oil-fueled vehicles industry, Chinese local firms are in a passive position in international competition. So, it is necessary and valuable to identify the innovation capabilities and positions of the main players of HFCV industry in China. In related research on industrial innovation activities, patent is regarded as an important output indicator for innovation. The citing and cited situations of patents are often used to reflect knowledge flow among innovators and detect who plays the leading role in the innovation network. From the perspective of social network analysis in the strategic emerging industry, the previous related literature has analyzed the structure of innovation network and its relationship with innovation performance (Ritter et al., 2004; Ahmad Abuzaid, 2014; Semrau and Werner, 2014). There is no research on current status of China's HFCV innovation activities by focusing on patent citation network, which characterizes technological innovation and knowledge flow, especially the spatial-temporal evolution of innovation activities across provinces.

Therefore, this article aims to fulfil the research gap on HFCV industry in China by answering the following questions.

First, what is the present situation of HFCV industry development in China? We want to understand in depth the HFCV's industry spatial distribution across the provinces in China, and sort out the industrial development policies issued by Chinese national government.

Second, which provinces play the leading role of innovation in China's HFCV industry during various periods of time? We construct innovation network based on patent citation information of micro-innovators locating in different provinces.

Third, what is the specific network position of each province in different periods? Specifically, we want to distinguish the provinces in optimal position or having the potential to be in optimal position in the next period. By analyzing the main reasons pushing forward the province's evolution of the network position, we can verify the policy effect, and hope to provide implications for improving the competitiveness of China's HFCV industry.

For answering the above-mentioned questions, in this article, first, we collect and screen out the patent data on China's HFCV from 2001 to 2020, and sort out the relevant policies at the national and local levels. And then, we construct the innovation network of China's HFCV industry based on 2,971 HFCV-related invention patents owned by 218 micro-innovators across 25 provinces from 2001 to 2020. Our focus is the knowledge flow direction based on patents, which can help to reveal the evolution of provinces' innovation capabilities in HFCV. The possible contributions of this article are as follows.

First, by analyzing the authoritative statistical data, this article objectively evaluates the development process and current situation of China's HFCV industry, as well as the industrial layout of each province. More importantly, our analysis can provide reference for optimizing the cooperation between provinces and forming the coordinated development of the whole industrial chain. In addition, by sorting out the important HFCV industrial policies at the national and local levels, this article identifies the time nodes of HFCV industrial development and builds the foundation for observing the innovation network's evolution trend and evaluating the policy implication effect.

Second, according to patent citation information among micro-innovators locating in different provinces, this article innovatively constructs patent citation networks of HFCV in China. Compared to the number of patent applications, patent citation data can track the direction of knowledge flow and identify knowledge producers as well. Based on various statistical indicators and topological graphs of the patent citation network, we can intuitively understand the spatial layout characteristics of the innovation activities, the evolution trend of the innovation status, and the contributions made by multi-types of innovators in different time periods. In the process of HFCV patent identification, we decompose the core technical keywords of each link in the HFCV industry chain, screen these keywords in the authoritative patent database, and build a provincial innovation network by the SNA method through the patent citation relationship.

Third, based on the outdegree and indegree indictors in the HFCV patent citation network within different time periods, this article employs the cluster analysis method to identify three network statuses, namely, the knowledge producers, knowledge consumers, and network brokers. This article furtherly quantifies the status of the provinces and tracks their spatial evolution in the network. In particular, this article identifies which provinces are in the optimal network position and have great potential, that is, in the status both of knowledge producer and of network broker. Our findings can provide reference for in-depth understanding of the core competitiveness of HFCV industry.

This article obtains the main findings as follows. First, the total sales of HFCV and the development of supporting infrastructure are balanced in China, and a series of national and local industrial development polices have been issued. Second, the density of China's HFCV innovation network is increasing, and the provinces playing the leading roles are mainly from the east area. In terms of the micro-innovators, the proportion of universities and research institutes has been gradually larger than firms. Third, the provinces in the optimal network position, namely, the network broker based on knowledge production, are changing. Specifically, Shanghai and Liaoning are gradually replaced by Beijing and Jiangsu, and Guangdong is climbing to the optimal position. Fourth, Sichuan is the only one network broker based on knowledge consumption at the current stage. Fifth, among the provinces temporarily located in non-broker position, Zhejiang, Hubei, Tianjin, and Hebei are prominent in the characteristics of knowledge production with strong development potential.

The subsequent sections of this article are organized as follows. **Section 3** reviews the related literature. **Section 4** presents the panorama of HFCV industry development in China in the period of 2001–2020. **Section 5** describes the data, methodology, and the main characteristics of innovation network based on patent citation information among the provinces in China. **Section 6** analyzes the temporal evolution of China's HFCV innovation network by using topological graph. **Section 7** explores the spatial evolution of network, by analyzing the network position of the provinces. Finally, **Section 8** concludes and proposes policy implications.

## 3 LITERATURE REVIEW

Understanding the spatio-temporal evolution of China's HFCV innovation network is helpful for identifying each province's position and attributes, guiding the optimization of industrial layout, and forming future industrial development plan at both national and local levels. Therefore, we review the related literature from four aspects, including the development and innovation of HFCV industry, patent citation in innovation network, measurement for network position, and spatio-temporal evolution of network.

## 3.1 Development and Innovation of HFCV Industry

Nowadays, most nations are predominately preoccupied with the need to increase economic growth amidst pressure for increased energy consumption (Udemba et al., 2020). However, higher energy consumption from fossil fuel aggravates climatic deterioration, and global sustainable development is facing serious obstacles such as global warming caused by greenhouse gas emissions. Therefore, energy portfolio diversification is more urgently necessary than ever (Udemba et al., 2020). The most widely recognized clean energy source is hydrogen (Liu et al., 2018; Nowotny et al., 2018), which makes developing HFCV one of the most effective solutions for countries to reduce emissions of greenhouse gases and improve energy security (Moliner et al., 2016; Moriarty and Honnery, 2019). Some advanced countries have already started to develop HFCV. As one of the global leaders in hydrogen energy and fuel cell technology, Japan reduces the cost of fuel cell production and closes the price gap with oil-fueled vehicles by increasing R&D expenditure (Khan et al., 2021). South Korea aims to expand the supply of HFCV through government policy such as subsidies to cover production cost and lower selling price (Kim et al., 2019). When HFCV becomes widely used in ASEAN, hydrogen will be the sticking point to the coupling of renewable and the transport sector (Li and Kimura, 2021). For China, since the reduction of polluting emissions is linked to the principle of sustainable development (Mele and Magazzino, 2020), the development of hydrogen energy and HFCV industry is highly related to China's energy development strategy and ecological civilization construction. Nevertheless, China is facing the risk of blind competition and industrial surplus due to the lack of national overall plan for industry development and continuity of related policies (Meng et al., 2020). In addition, due to the imbalance of regional development, the contribution to the growth of transport investments is different from region to region in China (Magazzino and Mele, 2020). From the technology side, although HFCV is still in the initial technology development phase over a period of time (Hardman et al., 2013; Hacking et al., 2019), it is radical innovation with the potential to significantly disrupt conventional combustion engine technology (Zapata and Nieuwenhuis, 2010). Among the core technologies of HFCV, fuel cell is a typical representative. In response to increasing pressure from stakeholders, many automakers would have struggled to develop inter-organizational fuel cell innovations without the involvement of suppliers with specialist capabilities in fuel cell technologies (Borgstedt et al., 2017). Besides, the lack of infrastructure maintenance eliminates the positive effects of transport investments over time in the medium term in China (Magazzino and Mele, 2020), hence, the well development of hydrogen infrastructure is a guarantee of HFCV industry.

## 3.2 Patent Citation in Innovation Network

Owing to the standardized information related to the development of new ideas and technologies, patent is considered the most valuable output index of innovation activities (Frietsch and Grupp, 2006). The patent citation analysis method reveals the internal information and structure of patent by identifying the citing-cited relationship among patents, and this method has been applied in the network-related research (Abraham and Moitra, 2001; Breitzman and Mogee, 2002; Zhou et al., 2014). In addition, it has been used in the research on knowledge flow (Ho et al., 2014), constructing knowledge network (Jaffe et al., 1993), technology diffusion (Xiang et al., 2013), and the structure among innovators participating in knowledge creation (Belderbos, 2001; lo Storto, 2006; Ma et al., 2009). Moreover, the citation relationship between patents or applicants is the focus of patent citation analysis as well (Narin, 1994). Citations to prior patents indicate that knowledge in the cited patent is beneficial to innovate new knowledge described in the citing patent (Verspagen, 2007), which provides good evidence of the links between technological antecedents and descendants.

However, patent citation information cannot reveal the whole characteristics of knowledge flow (Li et al., 2007). Social network analysis (SNA) is widely used to understand the interaction and connection among innovators. It is a quantitative technology developed by sociologists based on mathematical methods and graph theory. Therefore, a methodology combining social network theory with patent citation information has emerged. For example, through patent citation analysis, Verspagen (2007) maps the technological trajectories of fuel cell industry. Ji et al. (2019) construct global networks of genetically modified crops technology by patent citation relations. Watanabe and Takagi (2021) examine how technology has evolved within the technological field of computer graphic processing systems based on patent citation network.

## 3.3 Measurement for Network Position

Node's position in network determines its quantity and quality of resource (Koene, 1984). The center position in the innovation network means the control of competitive advantage (Foerster, 2016; Chai et al., 2017). Hence, many scholars have identified the position of node in the network according to some indicators and classified nodes into different groups. By the patent-cited intensity, Bekkers and Martinelli (2012) divide the position of node in innovation network into knowledge producer and knowledge consumer. Kim and Song (2013) construct a patent lawsuit network, and use three dimensions, namely eigenvector, indegree, and outdegree centrality, to divide firms into four groups, including key players, patent troll, victim, and bystander. Based on outdegree and indegree centrality, Choe et al. (2016) design Outdegree-Indegree (OI) index and betweenness centrality to describe a node's ability of producing knowledge, and then divide innovators into 4 categories, namely, broker based on producing knowledge, broker based on absorbing knowledge, knowledge producer, and knowledge absorber.

## 3.4 Spatio-Temporal Evolution of Network

Many scholars have focused on the spatial and temporal development of a sector based on the SNA method. Jiang et al. (2017) collect the low-cost carrier's data from 2005 to

2013 to study the development trends of air transport network in China. They divide the temporal evolution into three periods, namely, funding period, developing period, and stable period, while from the spatial evolution aspect, the connectivity of network is increasingly close. Similarly, Liu et al. (2021) find that the green innovation network evolves from forming, boom to stable period, and the network based on provincial level are increasingly related closely by constructing the green innovation network in China from 2007 to 2017. Some researches combine the spatial and temporal evolution of network with knowledge spillover. For instance, Li et al. (2015) analyze the development of scientific knowledge networks and technological knowledge networks in biotechnology field by using published papers and applied patents from 2001 to 2012, and they find that knowledge spillovers in spatio-temporal evolution of the networks have features of both hierarchical and contagious diffusion. They also find that knowledge spillovers are shifting from inter-cities to inter-regions. More recently, some researches focus on the key nodes in the network. Liu and Yu (2020) analyze the spatio-temporal evolution of key nodes in the network, intra-regional, and cross-regional cooperative innovation relationships in China's new energy vehicle industry. The eastern regions including Beijing, Shanghai, Guangdong, Jiangsu, and Zhejiang are key innovators and they are active in intra and inter-regional cooperative innovation. Li et al. (2021) study the evolution of the patent collaboration network of China's intelligent manufacturing equipment industry. They find that Jiangsu mainly innovates through intra-provincial collaboration while Beijing, Guangdong, and Shandong tend to inter-provincial collaboration.

## 3.5 Comments

Previous researches on HFCV industry mainly focus on the qualitative description such as HFCV technology, industry development, and policies. There is no study on the construction of HFCV innovation network, the quantitative identification of innovator's position in the network, and the spatio-temporal evolution of the network. The detailed limits are as follows.

First, the studies, which decompose HFCV core technologies according to the industry chain and construct HFCV innovation network based on patents of specific technologies, are still very limited. Focusing on China, the decomposition of the core technologies based on industry chain is conducive to accurately locate the competitiveness of different provinces in HFCV industry, and helpful to formulate effective regional and inter-firm cooperative development strategies, for avoiding repeated construction and waste of resources.

Second, the SNA method has not been used to measure innovator's position in the HFCV innovation network. From the perspective of data visualization, we can more clearly and intuitively describe the development process of innovation in China's HFCV industry, knowledge flow among provinces, and the evolution of the roles played by provinces in the innovation network.

Third, there is still no study focusing on China's HFCV innovation network. Although China, as a late comer in global HFCV industry, has a short history for HFCV development, it has the characteristics of rapid development, large market potential, and active innovation. In 2020, China is the world's third biggest producers of HFCV, even overpassing Japan[1]. In terms of industry development potential, China and the U.S. will tie for the first place of HFCV producing in 2030. Therefore, it is important and significant to investigate the characteristics of spatio-temporal evolution of China's HFCV innovation network.

# 4 DEVELOPMENT OF THE HFCV INDUSTRY IN CHINA

## 4.1 Status of China's HFCV Industry in the World

The global HFCV industry has entered a new phase since Toyota launched the world's first HFCV in 2014. **Table 1** compares the HFCV industry development of the world's top five countries in HFCV total sales by 2020, including China, Japan, South Korea, the United States, and Germany, which together account for about 98% of HFCV world total sales. Driven by high subsidies from national government, South Korea sells 10,906 HFCVs in 2020 with 114.6% y-o-y growth rate, ranking first in the world, and it has set a target of 1.8 million HFCVs for 2030 according to the data from IEA. Japan is the world leader on the construction of hydrogen infrastructure with the greatest number of hydrogen filling stations at present. The United States also has a great number of HFCVs but lacks hydrogen-filling stations. In contrast with the United States, Germany has built 100 hydrogen-filling stations by 2020, but with low level of HFCVs total sales. China's HFCV industry development is more balanced compared with the United States and Germany. China has the fastest growth among the five countries in the number of hydrogen-filling stations, with y-o-y growth rate of 109.84% in 2020. In addition, China sells 7,352 HFCVs in 2020 ranking third in the world.

## 4.2 Development Characteristics of HFCV Industry in China

China has always attached great importance to the development of HFCVs. **Figure 1** shows the number of HFCV's production and sales volume in China from 2016 to 2020. Overall, China's production and sales volume of HFCV increased first and then decreased, both peaked in 2019 at 3,018 units and 2,737 units respectively. As a result of COVID-19 outbreak and changes of national subsidy policies in 2020, China's production and sales volume of HFCVs fell sharply in 2020, decreased to 1,199 units of production and 1,177 units of sales.

[1]Data source: China Automobile Industry Association; Japan Automobile Dealers Association.

**TABLE 1 |** Comparison on development of HFCV industry in representative countries.

| Country | HFCV total sales (2020)[a] | HFCV total sales y-o-y growth rate (2020) (%)[a] | HFCV total sales Target (2030)[b] | Number of built hydrogen-filling stations (2020)[c] | Annual growth rate of built hydrogen-filling station number (2020) (%)[c] |
|---|---|---|---|---|---|
| China | 7,352 | 19.1 | 1,000,000 | 124 | 109.84 |
| Japan | 4,456 | 20.6 | 811,200 | 142 | 24.56 |
| South Korea | 10,906 | 114.6 | 1,800,000 | 60 | 81.82 |
| United States | 8,931 | 11.7 | 1,000,000 | 49 | 2.08 |
| Germany | 890 | 52.9 | — | 100 | 3.09 |

*Data source:*

[a]*China Automobile Industry Association; Japan Automobile Dealers Association; Ministry of Land, Infrastructure and Transport of the Republic of Korea; The International Organization of Motor Vehicle Manufacturers; German Ministry of transport;*

[b]*IEA, Fuel cell EV deployment, 2017–2019, and national targets for selected countries, https://www.iea.org/data-and-statistics/charts/fuel-cell-ev-deployment-2017-2019-and-national-targets-for-selected-countries;*

[c]*H2 stations; Hydrogen Industrial Technology Innovation Alliance of China.*

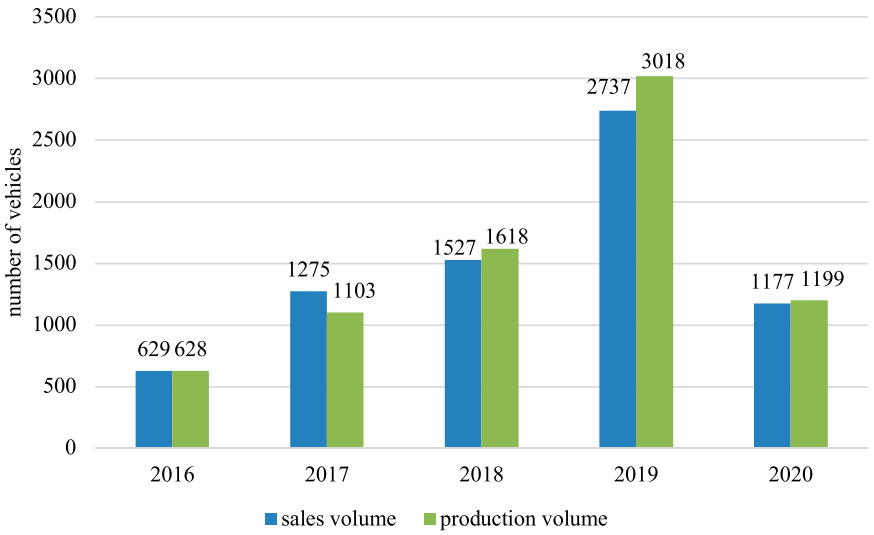



**FIGURE 1 |** Comparison on production and sales of HFCV in China (2016–2020). Data source: China Auto Parts industry Association.

HFCV industry chain is long and complex. Specifically, the upstream of the industry chain mainly includes hydrogen and other raw materials, fuel cell, and related core parts for HFCV. The midstream mainly includes HFCV manufacturing and integration. In contrast, the downstream mainly includes hydrogen infrastructure such as hydrogen-filling station and its supporting facilities, which is a very important link in the HFCV industry chain. According to the data from H2 stations, China's total number of built hydrogen-filling stations ranked second in the world by 2020, only behind Japan. **Figure 2** further shows the specific total number of hydrogen-filling stations built and under construction in various provinces of China by 2020. Guangdong is in the leading position with 31 built stations. Jiangsu, Shandong, and Shanghai all have more than 10 built stations. In terms of the number of hydrogen-filling station under construction, there are 57 hydrogen-filling stations under construction in China by 2020. The top three provinces are Guangdong, Shandong, and Hebei respectively.

## 4.3 National and Local Supporting Policies for HFCV Industry Development in China

Although China's HFCV industry has developed rapidly in recent years, it is still in the initial stage of development from the perspective of product life cycle. Therefore, it needs the support from national and local governments. **Table 2** presents China's main national supporting policies for the development of China's HFCV industry from 2001 to 2020, of which 2001, 2006, 2016, and 2020 are four important time nodes. To better analyze the China's HFCV innovation network on the temporal dimension, we use four time-nodes of 2001, 2006, 2016, and 2020 mentioned above as the basis of the evolution of China's HFCV innovation network. Therefore, we ultimately construct

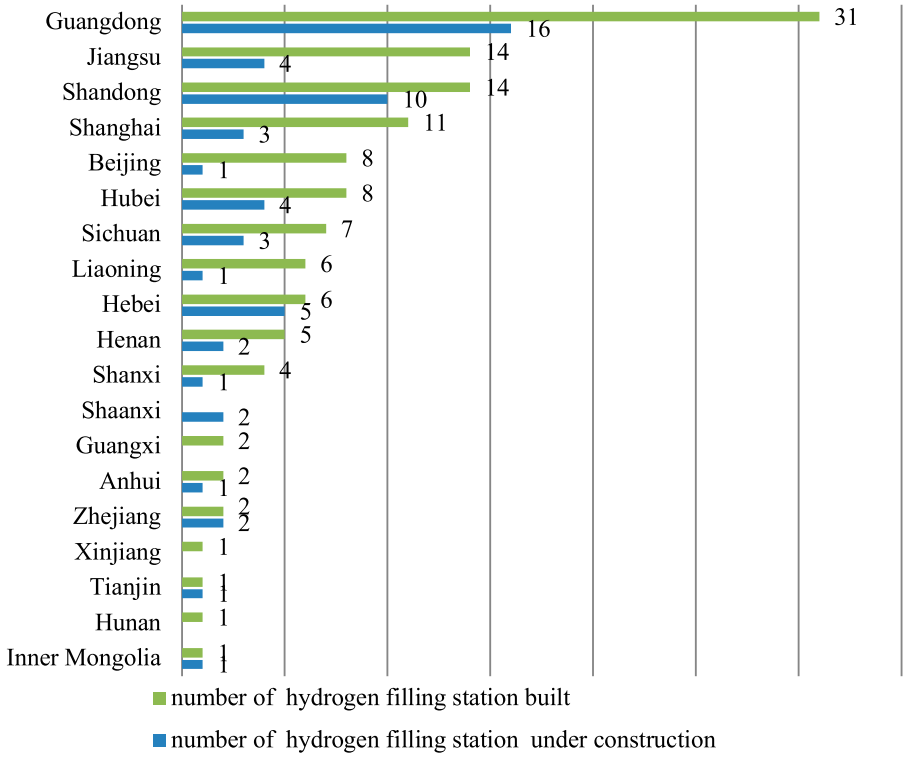



**FIGURE 2 |** The number of hydrogen-filling stations built and under construction of China's provinces by 2020. Data source: GGII. Note: For simplifying, "Province" in this article includes municipalities directly under the Central Government.

**TABLE 2 |** List of main national policies supporting the development of HFCV industry in China (2001–2020).

| Year | Department | National policy | Highlights |
|---|---|---|---|
| 2001 | The State Council | Electric Vehicle Heavy Truck Technology Project | Government funding to R&D of HFCVs |
| 2006 | The State Council | National Medium- and Long-term Scientific and Technological Development Plan (2006–2020) | Hydrogen and fuel cell as development direction of energy technology in the future |
| 2016 | The State Council | The 13th Five-Year National Plan for the Development of Strategic Emerging Industries | Develop the featured industry cluster and realize fuel cell vehicle mass production and demonstration application by 2020 |
| 2020 | Ministry of Finance; Ministry of Science and Technology; Ministry of Information Industry; Development and Reform Commission; National Energy Administration | Notice on Developing Demonstration Applications of Fuel Cell Vehicles | The HFCVs demonstration application is rewarded by the way of "Reward instead of subsidy" and the scope of reward is adjusted from the whole country to part of urban agglomeration |

*Data source: Each official national government department website of issued policies.*

three different multi-valued and directed networks to analyze the network evolution, each starting in 2001 and ending in 2006, 2016, and 2020 respectively.

As the intersection of traditional industry and strategic emerging industry, the development of HFCV industry has also been listed as the key field by China's local governments. Since Shanghai released the Shanghai Fuel Cell Vehicle Industry Development Plan in 2017, some major vehicle production provinces have also formulated relevant industry development plans for HFCVs. **Figure 3** shows the number of local policies related to HFCVs issued by 22 provinces in China between 2017 and 2020. The total number of related local policies is 105. Among them, the largest number of policies is issued in Guangdong with 20 policies, accounting for 19.05% of the total. The followed provinces are Jiangsu and Shandong both with 10 policies. In terms of geographical distribution, the provinces of Yangtze River Delta and Pearl River Delta account for 44.76% of the total number of policies, indicating that the eastern coastal regions are currently the key places of the development of China's HFCV industry.

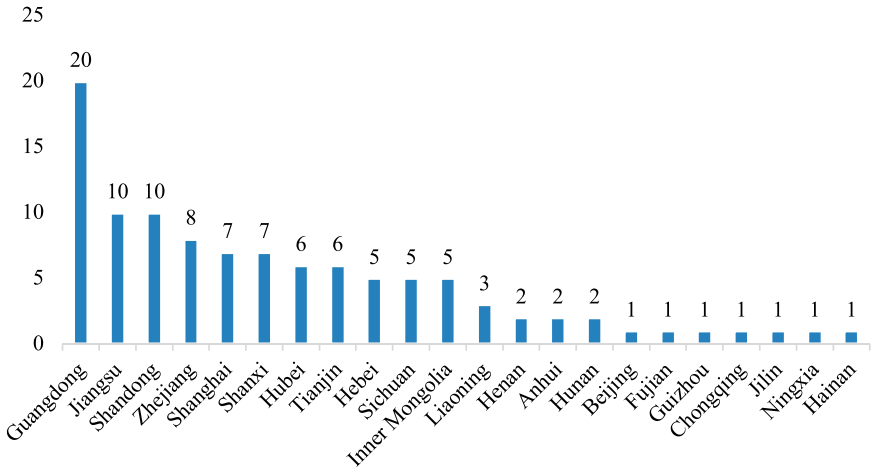



**FIGURE 3 |** The number of HFCV-related local supporting policies issued by China's province (2017–2020). Data source: Each official local government department website of issued policies.

# 5 DATA AND METHODOLOGY

## 5.1 Data

### 5.1.1 Patent Database

Based on the core technologies and key products of HFCV industry, we search the patents in the Incopat patent database. Incopat is a leading provider of IP information worldwide, focusing on deep integration and value mining of IP data. There are more than 2,000 users worldwide, including well-known firms, universities, and research institutions, such as Tsinghua University. Incopat is a database that provides relevant information related to the prior art for each patent in State Intellectual Property Office (SIPO) in China (Howell, 2019). In addition to the China's patent data, Incopat collects official and verified patent data from more than 130 countries, regions, and organizations in the world, and leads the world in the rate of updating patent information.

### 5.1.2 Data Collection Step

The steps of our data collection in Incopat and screening are as follows.

First, the key words for patent retrieval are determined, which is "Title or Abstract = Metal Bipolar Plate or Graphite Bipolar Plate or Composite Bipolar Plate or Proton Exchange Membrane or Platinum Catalyst or Carbon Paper or Gas Diffusion Layer or Membrane Electrode Assembly or Integration Stacks or DC/DC Transformation or Water and Heat Management or Air Compressor or Hydrogen Circulation Pump or Fuel cell System Integration or Hydrogen Production or Hydrogen Storage or Hydrogen Transportation or Hydrogen Station or Hydrogen Fueling Receptacle or Hydrogen Fueling Nozzle or Hydrogen Fuel cell or Hydrogen Fuel cell Vehicle."

Second, the patents are further screened after preliminary searching. Since invention patent represents the higher innovation level and ability, we only keep invention patents, excluding the utility model patents and design patents. Besides, we restrict the country in China and delete the patents without patent citation relationships or unrelated with HFCV industry. We also set the time range from 2001 to 2020. Therefore, we preliminarily obtain 7,336 invention patents with patent citation relationships and related with HFCV industry. Furthermore, we obtain 4,845 invention patents of 1,252 innovators after duplicate removal.

Third, based on the Pareto's rule, also known as the law of the vital few and the principle of factor sparsity, which states that for many phenomena 80% of consequences stem from 20% of the causes, we set the threshold as 4 for innovator's number of invention patents. Specifically, we exclude the innovators whose invention patent number is less than 4, and then obtain 3,528 invention patents of 254 innovators.

Besides, data of Taiwan, Hong Kong, and Macao are excluded in our research due to the missing of data. We further sort out the innovators based on their provincial identity of China. Firms are classified based on their headquarters, while universities and research institutes are based on their geographical locations. Therefore, finally, we obtain 2,971 invention patents of 218 innovators across 25 provinces in China from 2001 to 2020.

## 5.2 Methodology

### 5.2.1 Methods for Network Construction

Patent citation represents the process of knowledge flow. Backward citations indicate the degree of knowledge inflow, and forward citations show the inventive quality from the technological and economic perspectives (Henderson et al., 1998). Based on the final sample after patent searching and screening, and the patent citation relationships among sample provinces, we construct China's HFCV patent citation innovation network in spatial dimension. Since the patent citation relation strength is represented by the frequency of patent citation between innovators, our patent citation innovation network is multi-valued and directed.

**TABLE 3 |** Indicators for network topological analysis.

| Indicators | Definition |
|---|---|
| Number of nodes | The total number of nodes in a network |
| Number of core nodes | The total number of core nodes in a network |
| Number of periphery nodes | The total number of periphery nodes in a network |
| Network density | The ratio of actual links to all possible links in a network |
| Average distance | The average value of the distance between any pair of nodes in a network |
| Distance-based cohesion | The harmonic mean of the entries in the distance matrix, also called compactness, which has a value of 1 when the network is a clique and zero when the network is entirely made up of isolates. Larger value indicates greater cohesiveness of a network |
| Average degree | The degree is the number of links that a node has to other nodes. The average degree is calculated by dividing the sum of all node degrees by the total number of nodes in a network |
| Clustering coefficient | A node's clustering coefficient is the ratio of the number of actual links between the node's neighbors, to the number of the maximum possible links between those neighbors. The network's clustering coefficient is the average of the clustering coefficients for all the nodes |

### 5.2.2 Measures for Network Analysis and Network Position

For analyzing the structure and characteristics of China's HFCV innovation networks in different periods, we use various statistical indicators to conduct network topology analysis, including network density, average distance, distance-based cohesion, average degree, and clustering coefficient. These indicators have been widely used in network literature for analyzing the structure and properties of the network (Okamura and Vonortas, 2006). **Table 3** concludes the indicators we use and their definitions for the network evolution topological analysis.

Based on the patent citation relationships between nodes, we visualize the network evolution through UCINET6 to present knowledge flow and node's properties. And then we carry out the centrality analysis by using the indicators proposed by Freeman (1979). Degree centrality and betweenness centrality are used to estimate the value and significance of top 10 provinces in China's HFCV innovation network in different periods. The detail calculation process of the key index is explained as follows.

Degree centrality is measured by using the sum of nodes directly connected to one node, so it has a meaning of local centrality. The degree centrality of Node $i\left(C(i)_d\right)$ is calculated by **Eq. 1**, where $a(N_i, N_k)$ is 1 if, and only if, Node $i(N_i)$ and Node $k(N_k)$ are connected by a line, otherwise 0; n is the total number of nodes in a network. Since the networks are directed and multivalued, we have two types of degree centrality, namely indegree centrality and outdegree centrality.

$$C(i)_d = \frac{\sum_{k=1}^{n} a(N_i, N_k)}{n-1} \tag{1}$$

We also use betweenness centrality to measure the degree to which one node as the broker in a network. Betweenness centrality represents the function of a province in bridging information flow between nodes innovation network. The betweenness centrality of Node $i(C(i)_b)$ is calculated by **Eq. 2**, where $d_{jk}$ denotes the number of geodesics linking Node $j$ and Node $k$; $d_{jk}(i)$ denotes the number of geodesics linking Node $j$ and Node $k(j \neq k)$ that go through Node $i$; $n$ is the total number of nodes in a network.

$$C(i)_b = \frac{\sum_{j<k} d_{jk}(i)/d_{jk}}{[(n-2)(n-1)/2]} \tag{2}$$

Choe et al. (2016) calculate the network broker and knowledge production capability based on betweenness centrality and Outdegree-Indegree (OI) index. We refer this method in our analysis. Furthermore, we divide the sample provinces in China's HFCV innovation network into four groups according to these two indicators by using k-means cluster analysis. The detailed method and formula are introduced in **Section 6**.

# 6 EVOLUTION OF PATENT CITATION INNOVATION NETWORK

## 6.1 Features of Network Evolution

For discovering the features of the formation and evolution of China's HFCV patent citation network, we analyze the network evolution from three perspectives. First, from the micro perspective, we analyze the composition of innovators and the proportion of their patent applications in each period. Second, from the provincial perspective, we compare the number of provinces' patents in each period to find out the main patent application regions of the HFCV industry in China. Last, we analyze the performance of the provinces on indegree, outdegree, and betweenness centrality for each period to identify the most dynamic 10 provinces in China's HFCV patent citation innovation network.

**Table 4** shows the number and proportion of innovators of HFCV for the three periods[2]. The innovators are divided into two types, namely firm and university & research institute. We find that the universities and research institutes are the main innovators in China's HFCV industry. Compared with the period from 2001 to 2006, the gap between the two types of

[2]According to the important plans and policies for the development of China's HFCV industry at the national level, this article selects four time-nodes of 2001, 2006, 2016, and 2020. The details of the policies are presented in **Section 4**.

**TABLE 4 |** The information of innovators in China's HFCV industry.

| Innovator's type | 2001–2006 | | 2001–2016 | | 2001–2020 | |
|---|---|---|---|---|---|---|
| | Number | Proportion (%) | Number | Proportion (%) | Number | Proportion (%) |
| Firm | 4 | 57.14 | 48 | 29.63 | 58 | 26.61 |
| University & research institute | 3 | 42.86 | 114 | 70.37 | 160 | 73.39 |

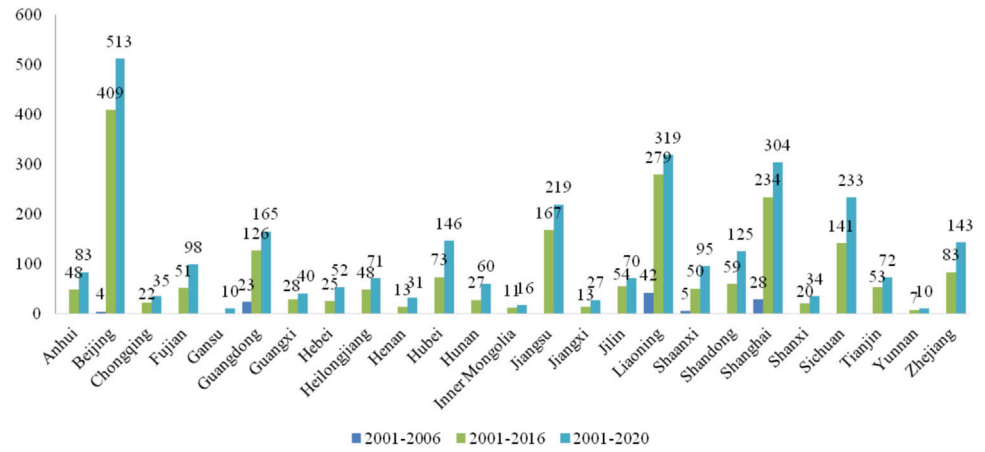



**FIGURE 4 |** Province's number of invention patents in China's HFCV industry.

**TABLE 5 |** Top 10 provinces with the highest centrality in China's HFCV innovation network.

| Rank | 2001–2006 | | | 2001–2016 | | | 2001–2020 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Degree centrality | | Betweenness | Degree centrality | | Betweenness | Degree centrality | | Betweenness |
| | In | Out | | In | Out | | In | Out | |
| 1 | Guangdong (10) | Liaoning (7) | Shanghai (3.5) | Beijing (330) | Beijing (340) | Beijing (27.624) | Beijing (513) | Beijing (604) | Beijing (21.914) |
| 2 | Shanghai (5) | Shanghai (6) | Liaoning (2.5) | Liaoning (242) | Shanghai (301) | Liaoning (19.965) | Shanghai (302) | Shanghai (482) | Jiangsu (20.529) |
| 3 | Liaoning (3) | Shaanxi (4) | Beijing (1) | Shanghai (193) | Liaoning (234) | Jiangsu (15.692) | Liaoning (298) | Liaoning (362) | Guangdong (10.687) |
| 4 | Beijing (1) | Beijing (2) | Shaanxi (0) | Jiangsu (178) | Jiangsu (163) | Shanghai (12.901) | Sichuan (269) | Jiangsu (331) | Liaoning (10.645) |
| 5 | Shaanxi (0) | Guangdong (0) | Guangdong (0) | Guangdong (138) | Guangdong (152) | Guangdong (12.309) | Jiangsu (265) | Guangdong (242) | Sichuan (10.638) |
| 6 | — | — | — | Sichuan (120) | Zhejiang (130) | Jilin (11.948) | Guangdong (196) | Zhejiang (230) | Shanghai (9.808) |
| 7 | — | — | — | Zhejiang (81) | Hubei (121) | Tianjin (11.945) | Zhejiang (191) | Hubei (205) | Shandong (7.045) |
| 8 | — | — | — | Hubei (78) | Tianjin (99) | Anhui (11.436) | Hubei (187) | Tianjin (175) | Anhui (7.017) |
| 9 | — | — | — | Jilin (72) | Heilongjiang (59) | Sichuan (5.676) | Shandong (179) | Shaanxi (106) | Hubei (6.799) |
| 10 | — | — | — | Tianjin (63) | Jilin (53) | Heilongjiang (5.584) | Fujian (139) | Shandong (100) | Zhejiang (6.208) |

*Note: numbers in parentheses indicate the specific value of each centrality.*

innovators is continually widen from 2001 to 2020, namely, firm's patent accounts for 26.61%, while that of university and research institute's patent accounts for 73.39%.

**Figure 4** presents the sample province's number of invention patents in HFCV industry in three distinct periods. From 2001 to 2006, the total number of patents at provincial level is 102. In this period, Liaoning has the largest number of invention patents, followed by Shanghai and Guangdong. The number of patents in Liaoning is 42, which accounts for 41.18% of the total. From 2001 to 2016, the total number of patents is 2,041. It is worth noting

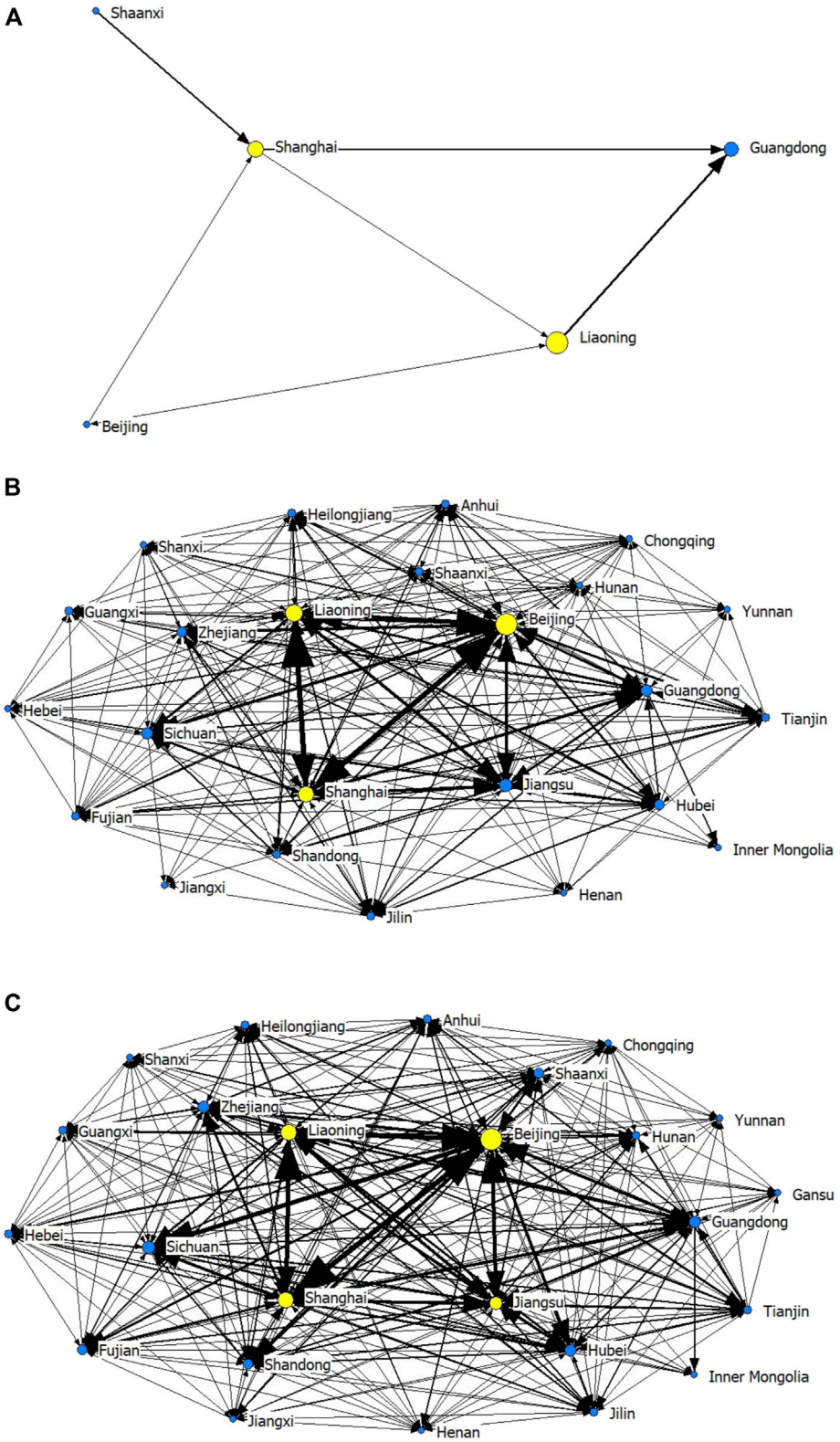



**FIGURE 5 |** Evolution of China's HFCV patent citation network for the three periods. **(A)** (2001–2006), **(B)** (2001–2016), **(C)** (2001–2020). Note: Node represents the province. Node's color represents node's activity level. Yellow node means core province. Blue node means periphery province. The arrowed line between nodes represents the patent citation relationship between provinces. The direction of the arrowed line represents the flow of the knowledge.

that Beijing shows a strong innovation ability, whose number of patents comes to the first-place and accounts for 20.4% of the total. From 2001 to 2020, the total number of province's patents is 2,971. Beijing still has the biggest number of patents among all the provinces with 513 patents, and followed by Liaoning and Shanghai. Therefore, we find that the eastern provinces are more dynamic for patent application in China's HFCV industry. What is noteworthy is that Sichuan, the only province from the west, has overpassed Jiangsu with 233 patents.

**Table 5** furtherly shows the top 10 provinces ranked by centrality of indegree, outdegree, and betweenness in three periods. Owing to the limitation of data, the period from 2001 to 2006 only includes five provinces, with Guangdong, Liaoning, and Shanghai as the top 3. Guangdong has the highest indegree, while its outdegree and betweenness are the lowest among all the provinces, which indicates that Guangdong tends to cite patents from other provinces in this period. Liaoning and Shanghai are relatively stable, ranking in the top 3 among all the three centralities. From 2001 to 2016, Beijing shows the most active performance among all the sample provinces, with indegree, outdegree, and betweenness centrality ranking the first. It is also interesting to note that the top 5 provinces on the three centralities indicators all include Beijing, Liaoning, Shanghai, Jiangsu, and Guangdong, which indicates that these five provinces are the most active in this period. From 2001 to 2020, Beijing still occupies the first place in all three indicators. What's the difference from the previous period is that Sichuan shows a good performance with ranking fourth place in indegree centrality and fifth place in betweenness centrality, which breaks the stable situation that the top 5 provinces are all from the east. We deduce the reasons why the eastern regions in leading position are caused by the high level of economic development and innovation ability, and better vehicle industry foundation. While for Sichuan province, the local industry development policy for HFCV industry plays the certain impetus function. Sichuan has issued 5 related policies on HFCVs and built 7 hydrogen-filling stations with 220 HFCVs in demonstration application, which makes it rank top in the western regions of China. Besides, it has carried out deep cooperation with Chongqing province. The southwest industrial cluster of the HFCV industry in China has been formed.

## 6.2 Evolution of Network Topology

Based on 2,971 patents of 218 micro-innovators from 25 provinces from 2001 to 2020, we construct three multi-valued and directed patent citation innovation networks of China's HFCV in three periods. In addition, we also conduct core-periphery analysis on these three networks and then find out the specific geographical features of core and periphery nodes. In **Figures 5A–C**, the node represents the province. Node's color represents node's activity level. Yellow node means core province. Blue node means periphery province. The arrowed line between nodes represents the patent citation relationship between provinces. The direction of the arrowed line represents the flow of the knowledge, namely, starting node of the arrowed line is the knowledge consumer and the ending node of the arrowed line is the knowledge producer.

**TABLE 6 |** Results of network topological analysis of China's HFCV industry.

| Measures | 2001–2006 | 2001–2016 | 2001–2020 |
|---|---|---|---|
| Number of nodes | 5 | 24 | 25 |
| Number of core nodes | 2 | 3 | 4 |
| Number of periphery nodes | 3 | 21 | 21 |
| Network density | 0.950 | 3.598 | 6.010 |
| Average distance | 1.538 | 1.332 | 1.242 |
| Distance-based cohesion | 0.492 | 0.765 | 0.879 |
| Average degree | 1.4 | 14.125 | 18.2 |
| Clustering coefficient | 0.875 | 5.927 | 8.218 |

**Table 6** shows the results of the network topology analysis for three periods. In detail, the density of network from 2001 to 2006 is 0.950, which is relatively sparse with only 5 provinces participating. The density of network from 2001 to 2016 reaches 3.598 with 24 provinces participating, and that from 2001 to 2020 nearly double to 6.010 with 25 provinces. This result reveals that China's HFCV patent citation network has changed from sparse to denser and the number of participating provinces has increased over time. In addition, the clustering coefficient of the three networks is increasing, which indicates that China's HFCV patent citation networks present high cluster feature. Besides, from the perspective of the connection between nodes, we find that the average distance between provinces decreases, the value of distance-based cohesion increases, and the average degree increases. Therefore, the patent citation relationship between provinces shows an increasingly close trend. Combined with network topological graphs in **Figure 5**, we conclude that Shanghai and Liaoning are the core provinces in the period of 2001–2006, while Beijing joins the core province family in the period of 2001–2016, and then followed by Jiangsu in the period of 2001–2020. We deduce that Beijing is the center of science and technology innovation of China with the largest number of high-quality universities and research institutions, which provides strong foundation for R&D of HFCVs. In contrast Jiangsu province is one of the most competitive provinces in the development of China's HFCV industry. Not only does the local government attach great importance to the development of HFCV industry, with the number of related policies issued for HFCVs ranking second among all the provinces of China, but also has formed a complete HFCV industry chain layout inside the Jiangsu province. The industry chain includes the upstream of fuel cell core parts R&D, the midstream of the whole vehicle manufacturing, as well as the downstream of hydrogen-filling stations and its supporting facilities construction.

# 7 ANALYSIS OF PROVINCE POSITION IN PATENT CITATION NETWORK

## 7.1 The Method for Identifying Province Position

To understand the position of each province in China's HFCV patent citation network, and according to Choe et al. (2016), we

**TABLE 7 |** Division of province position according to OI index and betweenness centrality.

| | Low OI index | High OI index |
|---|---|---|
| High betweenness centrality | Group 2: broker based on knowledge consumption | Group 1: broker based on knowledge production (optimal network position) |
| Low betweenness centrality | Group 4: knowledge consumer (unfavorable network position) | Group 3: knowledge producer |

use OI index and betweenness centrality as two indicators for identifying province position in the China's HFCV innovation network.

The OI index compares the degree centrality to which a province cites other provinces (indegree) with the degree to which a province is cited by other provinces (outdegree). The calculation method is shown in **Eq. 3**:

$$\text{OI index} = \frac{(\text{outdeg} - \text{indeg})}{(\text{outdeg} + \text{indeg})} \tag{3}$$

If the OI index is greater than 0, the number of cited patents in a province is greater than the number of citing patents, indicating that the province is a knowledge producer. If the index is less than 0, the number of cited patents in the province is less than the number of citing patents, indicating that the province belongs to knowledge consumer.

Betweenness centrality represents the function of a province in bridging information flow between nodes in innovation network. The greater the value of betweenness centrality, the stronger the function of the province as a broker in the network, and the greater the possibility of the province as a network center.

According to the values of OI index and betweenness centrality, we divide the position of the provinces into four groups, as shown in **Table 7**.

In **Table 7**, we define that the provinces in Group 1 are both network brokers and knowledge producers, which are in the optimal network position. The provinces in Group 2 are network brokers based on knowledge consumption. The provinces in Group 3 are not network brokers, but they are knowledge producers. The provinces in Group 4 are non-brokers with low knowledge production capacity and in an unfavorable network position.

To unify the scale of the OI index and betweenness centrality, we standardize the two indicators. Then, we use the k-mean cluster analysis method to classify the province position in three periods, as shown in **Figure 6**. In the first and second periods, due to the lack of provinces in Group 2, we set the k value to 3, which means that the provinces are divided into three groups. In the third period, we successfully and completely identify the provinces in Group 2; therefore, the k value is 4. In addition, we use dotted circles in **Figure 6** to distinguish and classify the province position, and insert a straight broken line where the value of OI index is 1 according to the Pareto's rule, treating provinces with an OI index greater than or equal to 1 as provinces with high knowledge production capacity.

## 7.2 Identification of the Optimal Network Position

According to the division of province position in **Table 7**, the provinces in Group 1 play the role of the knowledge bridge in the innovation network and are also the knowledge producers with the optimal network position. **Figure 6** shows the position changes of each province in China's HFCV innovation network. In the first period, Shanghai and Liaoning are in Group 1, but Beijing and Jiangsu replace them climbing to the optimal network position in the second and third periods. The results of the core-periphery analysis also show the characteristics of Beijing and Jiangsu from late comers to first movers. Shanghai and Liaoning are the first batch of provinces in China to develop HFCV industry, promoting them to become the brokers in the HFCV innovation network in the first period. For instance, in 2001, Shanghai Automotive Industry Corporate (SAIC) with General Motor and the Pan Asia Technical Automotive Center jointly successfully developed China's first hybrid fuel cell vehicle powered by hydrogen, namely, the Phoenix Fuel Cell Vehicle. Dalian Institute of Chemical Physics (DICP) in Liaoning made great progress in the field of proton exchange membrane fuel cell technology in the 1990s and jointly developed the first 30-kilowatt fuel cell bus with Dongfeng Motor in 2001. However, as China attaches more importance to the HFCV industry, many provinces have been involved in the HFCV industry. From 2017 to 2020, China's 22 local governments have issued 105 policies for the HFCV industry. The vehicle firms in many provinces start to develop HFCVs. For instance, Chinese old brands, such as Beiqi Foton Auto, Chery Auto, and China First Automotive Works (FAW) Corporate, have gradually started the R&D and production of HFCVs. Meanwhile, new competitors, such as Grove, have emerged in recent years. In this context, Beijing and Jiangsu stand out from many provinces and gradually replace Shanghai and Liaoning at the optimal position in the network. We deduce that this phenomenon may be also related to the number of universities and research institutes participating in the HFCV innovation network as well. As the HFCV industry is in its infancy in the world and China, its core technologies are still in the transition stage from innovation to commerce from the perspective of product life cycle. Therefore, universities and research institutes are undertaking the task of innovation and generating technology. The number of universities and research institutes participating in the HFCV innovation network in Beijing and Jiangsu are the top 2 in the second and third periods. Specifically, in the second period, based on our data, Beijing had 21 universities and research institutes, and Jiangsu had 14 universities and research institutes conducing HFCV-related innovation activities. In the third period, Beijing's number of university

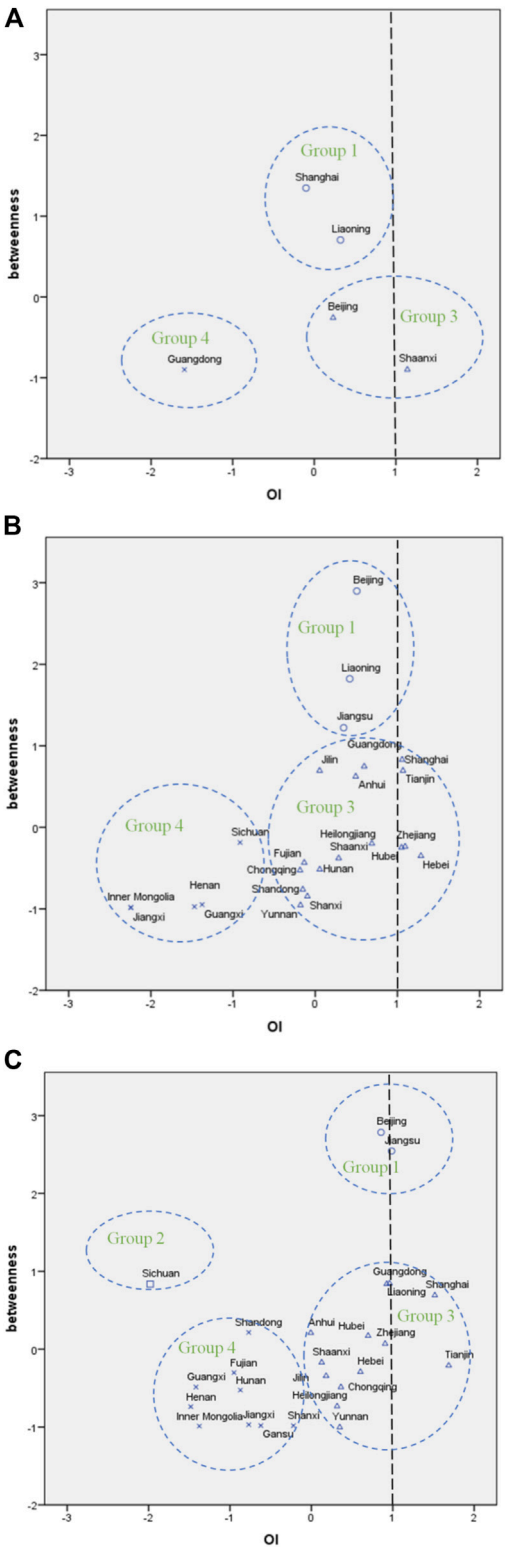



**FIGURE 6 |** Position of Chinese provinces in China's HFCV innovation network. Note: circle: Group 1, square: Group 2, triangle: Group 3, cross: Group 4. **(A)** (2001–2006), **(B)** (2001–2016), **(C)** (2001–2020).

and research institute has increased to 25, and Jiangsu has increased to 17, which has facilitated these two provinces to become the innovation network centers with the optimal position in China's HFCV industry.

## 7.3 Identification of Provinces With Strong Knowledge Production Capacity

In **Table 7**, except for the provinces in Group 1, the provinces in Group 3 play the role of knowledge producers in the network as well. We regard provinces with OI index greater than or equal to 1 as provinces with high knowledge production capacity. **Figure 6** shows that Beijing and Shaanxi have high knowledge production capacity in the first period. While in the second period, the OI indexes of Tianjin, Zhejiang, Hubei, and Hebei are all greater than 1, showing strong knowledge productivity. In the third period, the OI indexes of Hubei, Hebei, and Zhejiang decrease, but they are still close to 1. In addition, Guangdong's OI index rises to 1 in the third period, becoming an outstanding knowledge producer. In the meanwhile, its betweenness centrality is the highest among all the knowledge producers in Group 3. Therefore, we conclude that Zhejiang, Hubei, Hebei, Tianjin, and Guangdong are outstanding knowledge producers in China's HFCV industry, and Guangdong is expected to climb to Group 1, the optimal network position, in the next period. According to the arrangement of relevant industrial policies in **Section 4**, the number of HFCV industry policies in these five knowledge-producing provinces are all among the top 10 in China as shown in **Figure 3**, confirming that local industrial policy is an important factor for promoting the technological production in China's HFCV industry.

It is worth noting that Zhejiang, Hebei, Tianjin, and Guangdong are all from the east of China. Specifically, Zhejiang has issued 8 policies related to the HFCV industry from 2017 to 2020, the most representative is the Development of Hydrogen Fuel Cell Vehicle Industry in Zhejiang in 2020. The plan details how to build a demonstration zone for the HFCV industry in Zhejiang, which is inseparable from the transportation industry foundation of Zhejiang. Besides, in 2019, the Ministry of Transport of China released the Opinions on Zhejiang Province's Development of Constructing a Modern Comprehensive Three-dimensional Transportation Network and Other Pilot Works for Country with Strong Transportation, facilitating Zhejiang to design the demonstration scene of HFCV.

Tianjin is also a key place for China's HFCV industry. Tianjin has issued 6 related policies in 2017–2020. Tianjin Hydrogen Energy Industry Development Action Plan (2019–2022) proposes the necessity to introduce international advanced HFCV model and encourage the mass production of domestic HFCV. In terms of vehicle firm, Toyota, the international pioneer firm in HFCV, plans to cooperate with China FAW Corporation to invest 1.22 billion dollars to build a new energy vehicle plant in Tianjin, providing chance to cooperate in the field of HFCV.

Affected by the radiation effect of the Beijing-Tianjin-Hebei cluster, Hebei has issued 5 relevant policies from the period of

2017 to 2020, contributing to building the whole HFCV industry chain. Besides, Hebei has been included in the first batch of national HFCV demonstration cities in 2021, which will further strengthen Hebei's capacity in HFCVs technology. The most typical example of demonstration is that Zhangjiakou city in Hebei province as the main place for 2022 Winter Olympics will provide 2,000 HFCVs for spectators, tourists, and athletes.

Since 2013, Guangdong has formed the whole HFCV industrial chain. Guangdong also creates Yunfu Hydrogen Energy Town in 2018, which is a HFCV cluster of R&D, production, hydrogen energy infrastructure supporting services, and demonstration zone. By the end of 2020, Guangdong has owned 2,415 HFCVs and 31 hydrogen-filling stations, both ranking first among all the China's provinces since its HFCV industrial policies covering the development of fuel cell core components, hydrogen infrastructure, HFCV demonstration application, and industry standard setting.

Industrial foundation may be another reason for Guangdong becoming a knowledge producer. Guangdong has been the province with the largest vehicle production in China, reaching 3.132 million vehicles in 2020. The experience and technology of traditional vehicles, especially vehicle integration technology, have become an industrial foundation for HFCV.

In contrast, the number of university and research institutes in Guangdong participating in China's HFCV innovation network is only 4 and 7 in 2016 and 2020 respectively, which may be one of the main reasons for Guangdong's lack of betweenness to climb to optimal position. To improve the R&D strength and capability from local universities and research institutes, Guangdong is actively introducing universities and research institutes from other provinces. Therefore, we take the view that Guangdong has the potential to be in the optimal position of China's HFCV innovation network in the near future.

It is a remarkable fact that Hubei is the only province located in the central region with strong knowledge production capacity. Hubei has issued a total number of 6 related policies between 2017 and 2020, ranking fifth in China. At the same time, Hubei masters core technologies in the field of fuel cells. For instance, relying on local university, namely Wuhan University of Technology, WUT HyPower Technology Corporation in Hubei has become the largest membrane electrode supplier in domestic market and the fifth in international market. With the independent R&D and production technology of membrane electrode, this firm exports products to the United States, Germany, South Korea, and other advanced countries in large quantities, and becomes the main membrane electrode supplier of SinoHytech, leading to revenue exceeding 100 million yuan in 2019. In terms of industry foundation, Hubei, as a large vehicle province, has vehicle integration technology and brand advantages. For instance, Dongfeng Motor in Hubei has also been committed to the production of HFCVs since 2001. Besides, Hubei is a large hydropower province with abundant water resources, which will promote the development of hydrogen energy and HFCV industry.

## 7.4 Identification of Broker Based on Knowledge Consumption

According to **Table 7**, the provinces in Group 2 and Group 4 both are knowledge consumers in the network. The provinces in Group 4 are in an unfavorable network position, and the provinces in Group 2 are brokers based on knowledge consumption. It is worth noting that in the third period of China's HFCV innovation networks, Sichuan has become the first and only broker based on knowledge consumption among sample provinces. Specifically, Sichuan has issued five policies in 2017–2020 to support the development of the local HFCV industry. In 2019, Sichuan issued the Implementation Plan for Dealing the Pollution Control of Diesel Trucks in Sichuan to encourage the demonstration applications of fuel cell trucks and the construction of hydrogen-filling stations. Besides, this plan supports technological breakthroughs in the field of alternative fuels, hybrid power, blade electric, and fuel cell, strengthening the focus on the HFCV industry. At the same time, a few related firms have emerged in Sichuan. In 2018, Dongfang Electric and Chengdu Bus jointly developed hydrogen fuel cell bus. In the same year, Sichuan Natural Gas Investment Corporation and Sichuan Jinxing Corporation jointly constructed a skid-mounted hydrogen station with 400 kg per day. In terms of industry foundation, according to the Research Report on the Comprehensive Development Index of China's Electronic Information Manufacturing Industry (2018) issued by the Ministry of Industry and Information Technology of China, Sichuan's electronic information manufacturing industry development index ranks seventh in China, and ranks first in the central and western regions, providing industry foundation for the breakthrough in fuel cell technology. Besides, Sichuan has abundant water resources, which can provide resource support for the development of the hydrogen energy industry. In short, the support of government policies, the emergence of related firms, the foundation of industry, and the abundance of hydrogen resources have jointly helped Sichuan gradually moving toward the center of the innovation network. However, from the perspective of patent citation, Sichuan has been in the role of a knowledge consumer, indicating that Sichuan becomes a network broker by technological importing.

# 8 CONCLUSION AND POLICY IMPLICATIONS

The development of the HFCV industry is one of the important paths to achieve the upgrade of China's vehicle industry and the goal of carbon neutrality in 2060. To avoid repeated construction and blind competition existing in the China's traditional oil-fueled vehicle industry, it is important to optimize the layout of the China's HFCV industry among provinces and to form the coordinated development of different regions based on the advantages and positions of provinces in the industry. Therefore, we use patent citation information of 2,971 HFCV patents, applied by 218 innovators across 25 provinces of China from 2001 to 2020, to construct China's HFCV innovation network through social network analysis. And then the

temporal evolution of network has been in-depth analyzed. Furthermore, to study the spatial evolution of the innovation network, the provinces position in innovation network is divided into four groups based on the knowledge production, knowledge consumption, and network broker by the cluster analysis method.

The main findings are as follows. First, the total sales of hydrogen fuel cell vehicle and the development of supporting infrastructure are balanced in China, and a series of national and local industrial development polices have been issued. Second, the proportion of universities and research institutes in China's HFCV innovation network is increasing. Third, from the first period (2001–2006) to the third period (2001–2020), more and more provinces have participated in the construction of China's HFCV innovation network, and the eastern provinces have become increasingly active. At the same time, the density of innovation network is increasing. Fourth, in the three periods, the provinces in the optimal network position in China's HFCV innovation network are all from the eastern regions. Beijing and Jiangsu gradually replaced Shanghai and Liaoning, which may be related to the number of universities and research institutes participating in the innovation network. Guangdong is likely to climb to the optimal network position in the next period with great potential. Fifth, the network broker based on knowledge consumption in the innovation network only appeared in the third period, that is Sichuan from the western region. Besides, in the third period, the non-broker provinces with strong knowledge production capabilities are Zhejiang, Tianjin, Hebei, Guangdong, and Hubei. Based on the main findings, we propose the following policy implications for the reasonable spatial layout and competitiveness promotion of China's HFCV industry.

First, in terms of industry layout, China should prioritize provinces in the leading position of HFCV industry or with industry foundation and resource abundance. Specifically, Beijing can make full use of the hydrogen energy resource from Fangshan district in Beijing, Tianjin, and Hebei, and hydrogen production ability from Sinopec to form Beijing-Tianjin-Hebei demonstration cluster in hydrogen industry. Shanghai, Jiangsu, and Zhejiang can construct HFCV demonstration cluster based on traditional vehicle industry, equipment manufacturing industry, and transportation industry. Liaoning and Hubei can focus on the advantages of core components in the fuel cell field to drive the development of the fuel cell industry in the northeast and central regions of China, respectively. Guangdong can contribute to designing the HFCV standard system based on its demonstration experience of HFCV and hydrogen-filling station.

Second, in terms of industry-university-research collaboration, China should encourage firms, universities, and research institutes to establish strategic alliances or jointly set up firms to accelerate the industrialization of research findings. For example, Guangdong can introduce some high-class universities from other provinces to collaborate with local HFCVs innovators, in order to climb to the network center in the next period. At the same time, Sichuan also needs to enhance the knowledge production capacity with the help of universities and research institutes to transfer from the network center based on knowledge consumption to the network center based on knowledge production.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

PY: Conceptualization, Validation, Formal analysis, Writing—original draft. FJ: Writing-review & editing, Data curation ZC: Methodology, Data curation, Formal analysis. YS: Validation, Formal analysis, Writing—original draft.

## FUNDING

This work was supported by National Natural Science Foundation of China (Grant No. 72074069), and National Social Science Foundation of China (Grant No. 19BJY107). All remaining errors are on our own.

## REFERENCES

Abraham, B. P., and Moitra, S. D. (2001). Innovation Assessment through Patent Analysis. *Technovation* 21, 245–252. doi:10.1016/S0166-4972(00)00040-7

Ahmad Abuzaid, A. N. (2014). The Impact of Strategic Alliance Partner Characteristics on Firm's Innovation: Evidence from Jordan. *Int. J. Bus. Manag. Sci.* 9, 77–87. doi:10.5539/ijbm.v9n3p77

Bekkers, R., and Martinelli, A. (2012). Knowledge Positions in High-Tech Markets: Trajectories, Standards, Strategies and True Innovators. *Technol. Forecast. Soc. Change* 79, 1192–1216. doi:10.1016/j.techfore.2012.01.009

Belderbos, R. (2001). Overseas Innovations by Japanese Firms: An Analysis of Patent and Subsidiary Data. *Res. Pol.* 30, 313–332. doi:10.1016/S0048-7333(99)00120-1

Borgstedt, P., Neyer, B., and Schewe, G. (2017). Paving the Road to Electric Vehicles - A Patent Analysis of the Automotive Supply Industry. *J. Clean. Prod.* 167, 75–87. doi:10.1016/j.jclepro.2017.08.161

Breitzman, A. F., and Mogee, M. E. (2002). The many Applications of Patent Analysis. *J. Inf. Sci.* 28, 187–205. doi:10.1177/016555150202800302

Chai, R., Chen, M., Chen, Q., and Gao, Y. (2017). An Optimal Joint User Association and Power Allocation Algorithm for Secrecy Information Transmission in Heterogeneous Networks. *Wireless Commun. Mobile Comput.* 2017, 1–13. doi:10.1155/2017/5120538

Choe, H., Lee, D. H., Kim, H. D., and Seo, I. W. (2016). Structural Properties and Inter-Organizational Knowledge Flows of Patent Citation Network: The Case of Organic Solar Cells. *Renew. Sustain. Energ. Rev.* 55, 361–370. doi:10.1016/j.rser.2015.10.150

Foerster, M. (2016). Dynamics of Strategic Information Transmission in Social Networks. *SSRN J.* doi:10.2139/ssrn.2888492

Freeman, L. C. (1979). Centrality in Social Networks. *Soc. Networks* 1, 215–239. doi:10.1016/0378-8733(78)90021-7

Frietsch, R., and Grupp, H. (2006). There's a New Man in Town: The Paradigm Shift in Optical Technology. *Technovation* 26, 13–29. doi:10.1016/j.technovation.2004.07.007

Hacking, N., Pearson, P., and Eames, M. (2019). Mapping Innovation and Diffusion of Hydrogen Fuel Cell Technologies: Evidence from the UK's Hydrogen Fuel Cell Technological Innovation System, 1954-2012. *Int. J. Hydrogen Energ.* 44, 29805–29848. doi:10.1016/j.ijhydene.2019.09.137

Hardman, S., Steinberger-Wilckens, R., and Van Der Horst, D. (2013). Disruptive Innovations: The Case for Hydrogen Fuel Cells and Battery Electric Vehicles. *Int. J. Hydrogen Energ.* 38, 15438–15451. doi:10.1016/j.ijhydene.2013.09.088

Henderson, R., Jaffe, A. B., and Trajtenberg, M. (1998). Universities as a Source of Commercial Technology: A Detailed Analysis of university Patenting, 1965-1988. *Rev. Econ. Stat.* 80, 119–127. doi:10.1162/003465398557221

Ho, M. H.-C., Lin, V. H., and Liu, J. S. (2014). Exploring Knowledge Diffusion Among Nations: A Study of Core Technologies in Fuel Cells. *Scientometrics* 100, 149–171. doi:10.1007/s11192-014-1265-z

Howell, A. (2019). Heterogeneous Impacts of China's Economic and Development Zone Program. *J. Reg. Sci* 59, 797–818. doi:10.1111/jors.12465

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Q. J. Econ.* 108, 577–598. doi:10.2307/2118401

Ji, J., Barnett, G. A., and Chu, J. (2019). Global Networks of Genetically Modified Crops Technology: A Patent Citation Network Analysis. *Scientometrics* 118, 737–762. doi:10.1007/s11192-019-03006-1

Jiang, Y., Yao, B., Wang, L., Feng, T., and Kong, L. (2017). Evolution Trends of the Network Structure of Spring Airlines in China: A Temporal and Spatial Analysis. *J. Air Transport Manage.* 60, 18–30. doi:10.1016/j.jairtraman.2016.12.009

Khan, U., Yamamoto, T., and Sato, H. (2021). An Insight into Potential Early Adopters of Hydrogen Fuel-Cell Vehicles in Japan. *Int. J. Hydrogen Energ.* 46, 10589–10607. doi:10.1016/j.ijhydene.2020.12.173

Kim, H., and Song, J. (2013). Social Network Analysis of Patent Infringement Lawsuits. *Technol. Forecast. Soc. Change* 80, 944–955. doi:10.1016/j.techfore.2012.10.014

Kim, J.-H., Kim, H.-J., and Yoo, S.-H. (2019). Willingness to Pay for Fuel-Cell Electric Vehicles in South Korea. *Energy* 174, 497–502. doi:10.1016/j.energy.2019.02.185

Koene, J. (1984). Applied Network Analysis: A Methodological Introduction. *Eur. J. Oper. Res.* 17 (3), 422–423. doi:10.1016/0377-2217(84)90146-2

Li, C., Bai, L., Liu, W., Yao, L., and Travis Waller, S. (2021). Urban Mobility Analytics: A Deep Spatial-Temporal Product Neural Network for Traveler Attributes Inference. *Transportation Res. C: Emerging Tech.* 124, 102921. doi:10.1016/j.trc.2020.102921

Li, D., Wei, Y. D., and Wang, T. (2015). Spatial and Temporal Evolution of Urban Innovation Network in China. *Habitat Int.* 49, 484–496. doi:10.1016/j.habitatint.2015.05.031

Li, Y., and Kimura, S. (2021). Economic Competitiveness and Environmental Implications of Hydrogen Energy and Fuel Cell Electric Vehicles in ASEAN Countries: The Current and Future Scenarios. *Energy Policy* 148, 111980. doi:10.1016/j.enpol.2020.111980

Liu, F., Zhao, F., Liu, Z., and Hao, H. (2018). The Impact of Fuel Cell Vehicle Deployment on Road Transport Greenhouse Gas Emissions: The China Case. *Int. J. Hydrogen Energ.* 43, 22604–22621. doi:10.1016/j.ijhydene.2018.10.088

Liu, Y., Shao, X., Tang, M., and Lan, H. (2021). Spatio-Temporal Evolution of Green Innovation Network and its Multidimensional Proximity Analysis: Empirical Evidence from China. *J. Clean. Prod.* 283, 124649. doi:10.1016/j.jclepro.2020.124649

Liu, Y., and Yu, Q. (2020). Temporal-Spatial Evolution and Innovation Agglomeration of Technological Innovation Network in New Energy Vehicle Industry. *J. Dalian Univ. Technol. Soc. Sci.* 41, 40–48. doi:10.19525/j.issn1008-407x.2020.06.005

lo Storto, C. (2006). A Method Based on Patent Analysis for the Investigation of Technological Innovation Strategies: The European Medical Prostheses Industry. *Technovation* 26, 932–942. doi:10.1016/j.technovation.2005.10.005

Ma, Z., Lee, Y., and Chen, C.-F. P. (2009). Booming or Emerging? China's Technological Capability and International Collaboration in Patent Activities. *Technol. Forecast. Soc. Change* 76, 787–796. doi:10.1016/j.techfore.2008.11.003

Magazzino, C., and Mele, M. (2020). On the Relationship between Transportation Infrastructure and Economic Development in China. *Res. Transportation Econ.*, 100947. doi:10.1016/j.retrec.2020.100947

Mele, M., and Magazzino, C. (2020). A Machine Learning Analysis of the Relationship Among Iron and Steel Industries, Air Pollution, and Economic Growth in China. *J. Clean. Prod.* 277, 123293. doi:10.1016/j.jclepro.2020.123293

Meng, X., Gu, A., Wu, X., Zhou, L., Zhou, J., Liu, B., et al. (2020). Status Quo of China Hydrogen Strategy in the Field of Transportation and International Comparisons. *Int. J. Hydrogen Energ.* 46, 28887–28899. doi:10.1016/j.ijhydene.2020.11.049

Moliner, R., Lázaro, M. J., and Suelves, I. (2016). Analysis of the Strategies for Bridging the Gap Towards the Hydrogen Economy. *Int. J. Hydrogen Energ.* 41, 19500–19508. doi:10.1016/j.ijhydene.2016.06.202

Moriarty, P., and Honnery, D. (2019). Prospects for Hydrogen as a Transport Fuel. *Int. J. Hydrogen Energ.* 44, 16029–16037. doi:10.1016/j.ijhydene.2019.04.278

Narin, F. (1994). Patent Bibliometrics. *Scientometrics* 30, 147–155. doi:10.1007/BF02017219

Nowotny, J., Dodson, J., Fiechter, S., Gür, T. M., Kennedy, B., Macyk, W., et al. (2018). Towards Global Sustainability: Education on Environmentally Clean Energy Technologies. *Renew. Sustain. Energ. Rev.* 81, 2541–2551. doi:10.1016/j.rser.2017.06.060

Okamura, K., and Vonortas, N. S. (2006). European Alliance and Knowledge Networks1. *Technol. Anal. Strateg. Manage.* 18, 535–560. doi:10.1080/09537320601019677

Ritter, T., Wilkinson, I. F., and Johnston, W. J. (2004). Managing in Complex Business Networks. *Ind. Marketing Manage.* 33, 175–183. doi:10.1016/j.indmarman.2003.10.016

Semrau, T., and Werner, A. (2014). How Exactly Do Network Relationships Pay off? The Effects of Network Size and Relationship Quality on Access to Start-Up Resources. *Entrepreneurship Theor. Pract.* 38, 501–525. doi:10.1111/etap.12011

Udemba, E. N., Magazzino, C., and Bekun, F. V. (2020). Modeling the Nexus between Pollutant Emission, Energy Consumption, Foreign Direct Investment, and Economic Growth: New Insights from China. *Environ. Sci. Pollut. Res.* 27, 17831–17842. doi:10.1007/s11356-020-08180-x

Verspagen, B. (2007). Mapping Technological Trajectories as Patent Citation Networks: A Study on the History of Fuel Cell Research. *Advs. Complex Syst.* 10, 93–115. doi:10.1142/S0219525907000945

Watanabe, I., and Takagi, S. (2021). Technological Trajectory Analysis of Patent Citation Networks: Examining the Technological Evolution of Computer Graphic Processing Systems. *Rev. Socionetwork Strat* 15, 1–25. doi:10.1007/s12626-020-00066-1

Xiang, X.-Y., Cai, H., Lam, S., and Pei, Y.-L. (2013). International Knowledge Spillover through Co-Inventors: An Empirical Study Using Chinese Assignee's Patent Data. *Technol. Forecast. Soc. Change* 80, 161–174. doi:10.1016/j.techfore.2012.07.003

Yuan, Z., Li, Z., Kang, L., Tan, X., Zhou, X., Li, X., et al. (2021). A Review of Low-Carbon Measurements and Transition Pathway of Transport Sector in China. *Clim. Chang. Res.* 17, 27–35. doi:10.12006/j.issn.1673-1719.2020.202

Zapata, C., and Nieuwenhuis, P. (2010). Exploring Innovation in the Automotive Industry: New Technologies for Cleaner Cars. *J. Clean. Prod.* 18, 14–20. doi:10.1016/j.jclepro.2009.09.009

Zhou, X., Zhang, Y., Porter, A. L., Guo, Y., and Zhu, D. (2014). A Patent Analysis Method to Trace Technology Evolutionary Pathways. *Scientometrics* 100, 705–721. doi:10.1007/s11192-014-1317-4

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Yu, Jiang, Cai and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fenvs.2021.761736

# Carbon Neutrality Policies and Technologies: A Scientometric Analysis of Social Science Disciplines

*Yuan Zhang*[1], *Chung-Lien Pan*[1,2] *and Han-Teng Liao*[2,3,4]*

[1]Accounting School, Nanfang College, Guangzhou, Guangzhou, China, [2]Higher Education Impact Assessment Center, Nanfang College, Guangzhou, Guangzhou, China, [3]School of Public Administration, Nanfang College, Guangzhou, Guangzhou, China, [4]School of Literature and Communication, Nanfang College, Guangzhou, Guangzhou, China

Carbon neutrality, or "net zero", has become the impact assessment project of human impact on Earth, increasingly structured to examine the implications, for the environment and people, of proposed actions and consequences of inaction. International and local collaboration efforts have been made with the aim of achieving carbon neutrality or "net-zero" emissions; thus, policies and technological innovations have been developed. Such impact-oriented risk assessment and control efforts amount to carbon-neutral pathways. Although such pathways may diverge in terms of energy, resources, and cost, it is critical to summarize essential and promising preparatory work on related policies and technologies to inform both policy-makers and social scientists to take actions. Through a scientometric analysis and systematic review of the latest social science literature, the study identified the size, scope and exemplar work for each social science discipline on carbon neutrality, based on 907 articles collected in early 2021 from the Web of Science database. This study reveals a set of disciplines focusing on certain common and distinct aspects of carbon neutrality. By outlining the possibilities and application areas for future research and policy development for socio-technical transition towards a net-zero or post-carbon future, this study has contributed to the understanding of the global efforts to achieve a clearer and viable carbon-neutral pathway. In conclusion, as many aspects of the planet and humans have become datafied, digitized, and networked, carbon neutrality, as the ecological indicator that guide human production and consumption patterns, must take on a central role in guiding our conscious green digital transformation of many political, economic, social and psychological aspects of our societies according to the existing and emerging social science knowledge.

**Keywords: decarbonization, carbon information, sustainable consumption and production, carbon accounting, social science, just transition, socio-technical transition, carbon management**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, Hong Kong, SAR China*

***Reviewed by:***
*Benson Lam,*
*Hang Seng University of Hong Kong, Hong Kong, SAR China*
*Yongping Sun,*
*Hubei University of Economics, China*

***Correspondence:**
*Han-Teng Liao*
*h.liao@oxon.org*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *20 August 2021*
***Accepted:*** *20 September 2021*
***Published:*** *18 October 2021*

***Citation:***
*Zhang Y, Pan C-L and Liao H-T (2021) Carbon Neutrality Policies and Technologies: A Scientometric Analysis of Social Science Disciplines. Front. Environ. Sci. 9:761736. doi: 10.3389/fenvs.2021.761736*

## INTRODUCTION

To prevent global warming, an urgent threat to life on Earth, achieving carbon neutrality has become an important goal that is pursued by various organizations across the world, with the ultimate aim of measuring and controlling direct and indirect emissions of carbon dioxide (CO2) and other greenhouse gasses (GHG), such as methane (CH4) and nitrous oxide (N2O) (Smith and Fletcher, 2020). When governments adopted the Paris Agreement in 2015 at the 21st

Conference of the Parties (COP21) in the United Nations Framework Convention on Climate Change (UNFCCC), the Intergovernmental Panel on Climate Change (IPCC) was requested to publish a report. The report was later released in 2018 and was titled, "Global Warming of 1.5°C, an IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty" (IPCC, 2018). In this report, scientific findings regarding the potential impacts and associated risks of global warming were summarized, with language phrases such as "to neutralize emissions," "carbon neutrality," etc. In short, neutralizing the emissions of greenhouse gases caused by human activities ["the root cause of global warming" (IPCC, 2018)] requires both robust scientific understanding of the impacts of global warming on both natural and human systems and collective action on strengthening and implementing the global response. In 2020, China and the European Union (EU), two of the world's top emitters, pledged to become carbon neutral by 2060 and 2050, respectively (Harvey, 2020), by invoking policies to move the economy towards sustainability (Mallapaty, 2020). Japan also has made a similar pledge for the date of 2050 (METI, 2020). Major companies such as Apple, Amazon, IBM, Microsoft, BP, Ford, General Motors, and FedEx have also pledged to become carbon neutral. Companies' disclosures of their Greenhouse Gas (GHG) emissions also highlight the vital role of companies in mitigating the effects of global warming, which requires cost-effective and verifiable accounting for emissions (Downie and Stubbs, 2013; Diniz et al., 2021). These carbon neutrality pledges made by major political and economic entities demonstrate the widespread commitment to galvanize action on the climate crisis. Since the pathways used to achieve carbon neutrality require a viable socio-technical transition from industrial processes based on petroleum, coal, and gasoline towards a more environmentally friendly sustainable future (Wang et al., 2020), it is thus essential to focus on the political, economic, social, and psychological aspects of carbon neutrality.

Indeed, the multidisciplinary nature of carbon neutrality goes beyond science and technology and involves issues such as policies, investment, communication, behavior change and adaption, and international relations (Hawken, 2017; Farnworth, 2018). For instance, combining social science knowledge and artificial intelligence techniques can advance innovations, as exemplified by a promising research agenda on transnational industry cooperation (TIC) and transnational university cooperation (TUC) innovations, especially in the context of EU–China science, technology, and innovation cooperation (Cai et al., 2019). In another example, efforts to contribute to a sustainable transition have encountered reactions from entities with vested interests, such as the petroleum and coal industry, often in political lobby activities and misinformation campaigns or "green washing" efforts (Smith and Fletcher, 2020). Ten social, economic and legal insights have also been summarized in climate science research since mid-2019, highlighting the need for advancing knowledge about the interactions between the human world and planetary systems (Pihl et al., 2021). New methods and data are needed to support monitoring and decision-making systems, such as using remote sensing big data for more precise and efficient forest resource management and policy-making (Wang et al., 2021). Thus, to achieve carbon neutrality, social science knowledge must be developed to understand the multiple aspects of preparing human societies for a sustainable transition for the mitigation of climate change. By informing decision-makers and collective actions, such knowledge will help researchers and policy-makers design and implement policies and technologies that work in initiating and implementing individual, organizational, and social changes. This explains the rationales behind global initiatives such as bringing transparency and rigor to the voluntary carbon market (Twidale, 2021) and research that connects mortality with carbon emissions under the notion of "the mortality cost of carbon" (Bressler, 2021; Schwartz, 2021). Social science knowledge matters.

To the best knowledge of the authors, based on our research and analysis of related literature, there has so far been no systematic review of social science knowledge on carbon neutrality using scientometric or bibliometric methods, whereas such bibliometric analysis on the wider and more developed topic such as circular economy have be conducted (e.g., Cui and Zhang, 2018; Meseguer-Sánchez et al., 2021). An updated understanding of carbon neutrality social science literature focusing on the relationship between policies and technologies is expected to be useful in informing the debates and decisions surrounding the goal of achieving carbon neutrality, along with diverse stakeholders across different industrial sectors and knowledge domains. Such an understanding will not only help to guide the urgent policy, research, and technology innovations with an initial road map, but also identify the challenges and opportunities in the design and application of relevant policies and technologies. To gain both general and specific social knowledge, it is important to provide a systematic review of carbon neutrality literature to summarize how social science knowledge can contribute to the worldwide collective efforts to achieve carbon neutrality.

The purpose of this paper is to provide a review of social science literature on the topic of carbon neutrality, asking the following research questions:

1) When has the intellectual structure of the knowledge base developed, and by whom?
2) How and when has the conceptual structure of the knowledge base developed?
3) Where and how has the social structure of the knowledge base developed?
4) Where and what has been contributed from the social science disciplines?

To answer the above questions regarding the overall contribution of social sciences to the topic of carbon neutrality or decarbonization, a scientometric and systematic review was conducted based on Clarivate Analytics' Web of Science (WoS) database using science mapping tools, as detailed in the next

section. By answering these descriptive, analytical, and reflective questions, this paper aims to contribute to the existing human knowledge on the topic of carbon neutrality by focusing on the socio-technical pathways that social science knowledge has already contributed and ought to contribute.

## MATERIALS AND METHODS

Science mapping (Cobo et al., 2011) and scoping review mapping (Fridell et al., 2019) aim to use the visualization (Small, 1999) of "research fronts" (Garfield, 1994), knowledge structure, or intellectual bases (Persson, 1994) to gain insight. For such visualization outcomes to be meaningful, especially for the research questions raised above, a careful research design is needed to analyze social science knowledge in various "research fronts" that are accessible for both policy makers and researchers.

### Science Mapping Approach

Our science mapping research, which aims to facilitate theoretical and empirical development across the main and emerging social sciences disciplines, has conducted a scientometric analysis on the topic of carbon neutrality. The ultimate objective of this paper is to provide a holistic and systematic overview of the social science research of carbon neutrality worldwide over the years to shed light on carbon neutrality policies and technologies. Instead of examining a specific policy or technology, the scientometric approach, along with its bibliometric methods and data, can provide a valuable overall picture that allows researchers across different disciplines to understand the composition and structure of the intellectual knowledge base of carbon neutrality. To this end, this paper used quantitative science mapping techniques to categorize and visualize the disciplinary venues and topical focus of carbon neutrality research and then complemented it with brief traditional research reviews of selected articles that are relevant to the overall discussions of carbon neutrality policies and technologies.

### Raw Materials

Clarivate Analytics' Web of Science (WoS) database was used, along with the WoS Core Collection of SCI-Expanded, SSCI, A&HCI, and ESCI indices, to collect the bibliometric data for analysis in this review. It should be noted that SCI-Expanded and ESCI journals were included. Some journals are indexed in more than one index, for instance, one of the top journals, *Energy Policy,* is indexed both by SCI-Expanded and SSCI.

Web of Science offers a subject categorization scheme called "research areas" that covers five broad disciplinary categories (Clarivate Analytics, 2020):

- Arts and Humanities,
- Life Sciences and Biomedicine,
- Physical Sciences,
- Social Sciences,
- Technology.

As one important dimension of the research fronts is "a posteriori constructs that provide both highly specific and broad access to the subject matter" (Garfield, 1994), the WoS research area's social sciences category contains 25 disciplines ranging from archaeology to women's studies. Such categorization has been assigned by the WoS experts at the levels of journals and books. Thus, the WoS research areas are features at the level of journals. Given the multi-disciplinary nature of carbon neutrality, it is important to examine the disciplinary features of the literature by not limiting our scope of search to the social-science-only index such as SSCI.

In terms of the document type, this review examined articles, book chapters, and reviews. Although the WoS does provide conference proceedings such as CPCI, this review does not include them because the quality and quantity of such proceedings may differ widely across various social science disciplines. By focusing on journal articles and books, this review should provide a more succinct picture of the overall knowledge base of the carbon neutrality literature.

In terms of topics, a series of synonyms of carbon neutrality, along with its related topics of carbon accounting and decarbonization, were chosen. The assumption is that the use of these terms in the title, abstract, or keywords of work indicates the authors' intention to discuss, at the very least, topics such as carbon footprints or decarbonization, which have implications on reaching carbon neutrality.

The overall research design consists of a query design, a data process, and a taxonomy design to explore the policies and technologies that contribute to the goal of carbon neutrality, with a focus on the work across social science disciplines.

### Query Design

The query design consists mainly of two parts: topics and disciplines. As shown in the query below, the first part aims to cover the relevant literature, which is expressed by the field tag of "Topic" (i.e., TS) and a series of synonyms and underlying topics of carbon neutrality that are joined by "OR" Boolean operators.

- TS = ("carbon neutrality" OR "carbon neutral*" OR "carbon-neutral*" OR "carbon positive*" OR "carbon-positive*" OR "carbon negative*" OR "carbon-negative*" OR "carbon accounting" OR "net-zero" OR "decarboni?ation") AND SU = ("Social Sciences" OR "Archaeology" OR "Area Studies" OR "Biomedical Social Sciences" OR "Business and Economics" OR "Communication" OR "Criminology and Penology" OR "Cultural Studies" OR "Demography" OR "Development Studies" OR "Education and Educational Research" OR "Ethnic Studies" OR "Family Studies" OR "Geography" OR "Government and Law" OR "International Relations" OR "Linguistics" OR "Mathematical Methods In Social Sciences" OR "Psychology" OR "Public Administration" OR "Social Issues" OR "Social Sciences Other Topics" OR "Social Work" OR "Sociology" OR "Urban Studies" OR "Women's Studies")

Similarly, the second part consists of the field tag of "Research Area" (i.e., SU) and a set of disciplines under the WoS research area's social sciences category. The two parts of the "Topic" query and "Research Area" query were joined by an "AND" Boolean operator, which should give us initial database search results containing social science literature on carbon neutrality.

It should be noted that the series of synonyms have been gathered through an iterative snowballing process that begins with the baseline terms of decarbonization and carbon neutrality. Based on the top keywords of the initial datasets, closely-related terms such as "net-zero" and "carbon accounting" terminologies (e.g., those containing negative and positive sides) have been included in the search query. Other more detailed methods or terms were not included because the purpose of research aims to focus first on the comprehensive overview of the carbon neutrality literature, and thus adding terms from specific subfields or subdomains may render such comprehensive coverage unbalanced. This component of research design will be further discussed later in the Conclusion section as the trade-off between comprehensiveness and depth.

## Data Processes

In accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al., 2009), the screened resources for the study consist of searched articles indexed by Clarivate Analytics' SCI-EXPANDED, SSCI, A&HCI, and ESCI indices for articles published in reliable academic sources. Our initial database searching resulted in 918 articles, which was collected on February 1, 2021. Bibliographic data (the "Full Record and Citation References" option was selected) of the documents were downloaded for further analysis, along with "meta data" such as the article title, author name(s), organization affiliations, keywords, abstracts, publication source, reference citations, and various data.

Second, during the eligibility phase, we first decided to examine the literature written in the English language, leaving out 11 articles. The 907 articles found constitute the basis of this scientometric analysis. We then decided to examine the full-text articles of the most relevant ones, for further qualitative synthesis, in terms of 1) the number of citations received and 2) the most recently published. The selection criteria reflect the research inquiries on 1) knowledge bases and 2) emerging research. For each discipline, we identified the most cited and most recent documents, then the documents were examined based on the relevance of their contribution to the carbon neutrality social science literature. More classic literature review work was conducted on the selected work across different social science disciplines.

Third, during the final "included phase" of the PRISMA, data preparation work was conducted to provide more meaningful insights. *Via* iterative processes of three experts' work, two thesaurus files were compiled for both keywords and publication sources to disambiguate concepts and entities. For example, author keywords such as "climate change policy" and "climate policy frameworks" were replaced by the more frequently used term "climate policy". Different volumes of book sources (e.g., *Annual Review of Resource Economics*) were replaced by the common name without volume numbers. Such data preparation work facilitated more accurate science mapping (van Eck and Waltman, 2014). These thesaurus files are essential ground work for scientometric methods such as keyword co-occurrence analysis because the preprocessing work can remove duplicate terms and expressions in a more standardized consistent fashion.

To implement scientometric analysis, the data sets were processed by VOSviewer (Yu et al., 2020) Bibliometrix (Aria and Cuccurullo, 2017) and Python data science packages to gain insight. In particular, the use of the Python Data Analysis Library (a.k.a. pandas) allowed the "split-apply combine strategy" (Wickham, 2011) to be implemented to break up the searched outcomes into manageable pieces (e.g., the WoS disciplines of "research areas", topics, sources, etc.) and then explore how these pieces constitute the whole. Based on the "split-apply-combine strategy", the largely quantitative data analysis based on a series of combined descriptive statistics, scientometric analyses, or social network analyses of scientometric relationships aimed to provide the empirical structure of the knowledge base.

The statistical network analysis was conducted using the Network Analysis in Python (a.k.a. NetworkX) library, resulting in network centrality metrics (i.e., closeness centrality, eigenvector centrality, and betweenness centrality) and clustering coefficient (Onnela et al., 2005). Although these advanced metrics are not often reported in full by bibliometric analysis (e.g. Cui and Zhang, 2018; Meseguer-Sánchez et al., 2021), the study will report them as complementary verification of the visualization outcomes. Based on the network analysis conventions (Freeman, 1978; Hansen et al., 2020), the centrality measures consists of closeness centrality (how close a node is to the rest of the nodes), eigenvector centrality (how central a node is based on its neighbors), and betweenness centrality (how central a node is in bridging different segments of the network). In addition, different from these centrality measures, the clustering coefficient metric measures the density of the 1.5-degree egocentric network for each node, indicating how connected the connected nodes are. These metrics can then complement and verify the observations of network visualization. Following the bibliometric analysis conventions, the number of clusters are determined by human trial-and-error for meaningful and succinct algorithm-based grouping outcomes for the authors. Such clustering outcomes are reproducible with a set of parameters, but other outcomes can be produced for fewer or more groups.

CiteSpace (Chen, 2016) was also used—mainly for creating the citation burst tables—with the aim to see which topics had gained significantly more attention in recent years or over certain periods of time. This can complement the time-related findings as part of a more holistic understanding over time (Chen, 2004).

## Scientometric Methods to Show the Structure of Knowledge Base

This study implements several scientometric methods to determine the overall intellectual, conceptual, and social structure of the knowledge base (Donthu et al., 2021),

**TABLE 1 |** Basic information about the bibliometric dataset.

| | Description | Results |
|---|---|---|
| Size | Timespan | 1983–2021 |
| | Sources (Journals, Books, etc) | 266 |
| | Documents | 907 |
| | References | 42,244 |
| Descriptive statistics | Average years from publication | 5.15 |
| | Average citations per documents | 16.94 |
| | Average citations per year per doc | 2.248 |
| Contents | Keywords Plus (ID) | 1,685 |
| | Author's Keywords (DE) | 2,114 |
| Author | Authors | 2,247 |
| | Author Appearances | 2,716 |
| | Authors of single-authored documents | 206 |
| | Authors of multi-authored documents | 2,041 |
| Author collaboration | Single-authored documents | 223 |
| | Documents per Author | 0.404 |
| | Authors per Document | 2.48 |
| | Co-Authors per Documents | 2.99 |
| | Collaboration Index | 2.98 |

complemented by systematic review of the top-cited work from the social science disciplinary perspective.

To answer the first question regarding the intellectual structure, the scientometric methods of historiograph and author-co-citation analysis have been implemented to reveal the clusters of organizational and individual authors have been contributed to the main knowledge base. Since published work can accumulate citations over time, the citation approach (Small, 1999) can provide an accurate mapping of clusters of items (e.g., documents, sources, etc.). Based on the most prominent direct citations, the historiograph maps out the intellectual links in a chronological order, revealing the basic works in the dataset. The author co-citation analysis visualizes the relationship network among cited work as the fundamental sources.

To answer the second question on the conceptual structure, the scientometric methods of keyword co-occurrence, keywords with strongest citation bursts, and a series of thematic maps have been conducted to reveal the clusters of ideas and their dynamic evolutions over different time periods. Keyword co-occurrence maps show the existing relationship among topics, and the advanced methods of thematic maps and keywords with citation bursts provide additional temporal insights to the evolution of conceptual structure.

To answer the third question on the social structure, the scientometric methods of bibliometric coupling analysis (Jarneving, 2007; Boyack and Klavans, 2010) have been applied to the journal, discipline, organization, and country information of the bibliometric data. Previous work has also shown that insights into clusters of disciplines can be derived based on WoS disciplinary information (Chen and Liao, 2020).

To answer the last question on the social science contribution, the study uses the bibliometric data provided by the WoS' subject categorization scheme's social sciences category. Thus, the disciplinary landscape will be shown according to, and based upon, such a categorization assignment.

Altogether, the scientometric analysis of carbon neutrality social science literature is expected to show the knowledge base as documented and cited in the empirical evidence of academic publications.

# RESULTS

The findings of the study are organized into four sections, moving from the main structural work to the emerging peripheral work, with the purpose to cover both the common and distinct features of the social science knowledge on the topic. The first three sections demonstrate the intellectual, conceptual, and social structure of the knowledge base, and the last section samples the main work across various social science sub-disciplines.

**Table 1** lists basic information about the dataset, with its size, descriptive statistics, contents, author and author collaboration numbers. It has 907 documents over the timespan of 1983–2021.

For the purpose of examining policy implications, the years of 2015 and 2020 were chosen based on the Paris Agreement in December 2015. The first time period (1983–2015, $N$ = 310) marks the early period before the Paris Agreement which includes the Kyoto Protocol adopted in December 2012. The second time period (2016–2019, $N$ = 382) covers the most of the 5-years circle of the global climate action. The main reason for the third time period (2020–2021, $N$ = 215) is to explore the emerging knowledge by comparisons. Note that the number of publications is relatively even for each period, ranging from 215 to 382.

## Intellectual Structure

To explore the intellectual structure of the knowledge base, the study begins with a historiography and an author co-citation network, thereby describing the most cited work and its authors (including the organization authors). The findings should provide some descriptions that may amount to the summary of the main schools of thought, and their shared and divergent concepts/issues. Indeed, for instance, the main individual and organization authors have evolved mostly from European and international organization actors, as evidenced by findings below.

### Historiograph

**Figure 1** shows the bibliometric historiograph by using science mapping tool bibliometrix (Aria and Cuccurullo, 2017). Such mapping first creates the historical direct citation network from the most-cited work and then visualizes the network in a chronological order (Garfield, 2004). The following subsections will describe these networks, from bottom to the top.

#### *Systems and Technologies Cluster*

The two networks at the bottom of **Figure 1** can be described as the "systems and technologies" cluster of networks for its focus on technology and systems thinking. The concerns on technological forecasting and socio-economic change have shown in the relatively more early work by Nakićenović (1996) on decarbonization and later by Tapio et al. (2007) on European energy and transport sectorial development on both decarbonization and dematerialization. Both work highlights the strong link between decarbonization and dematerialization,

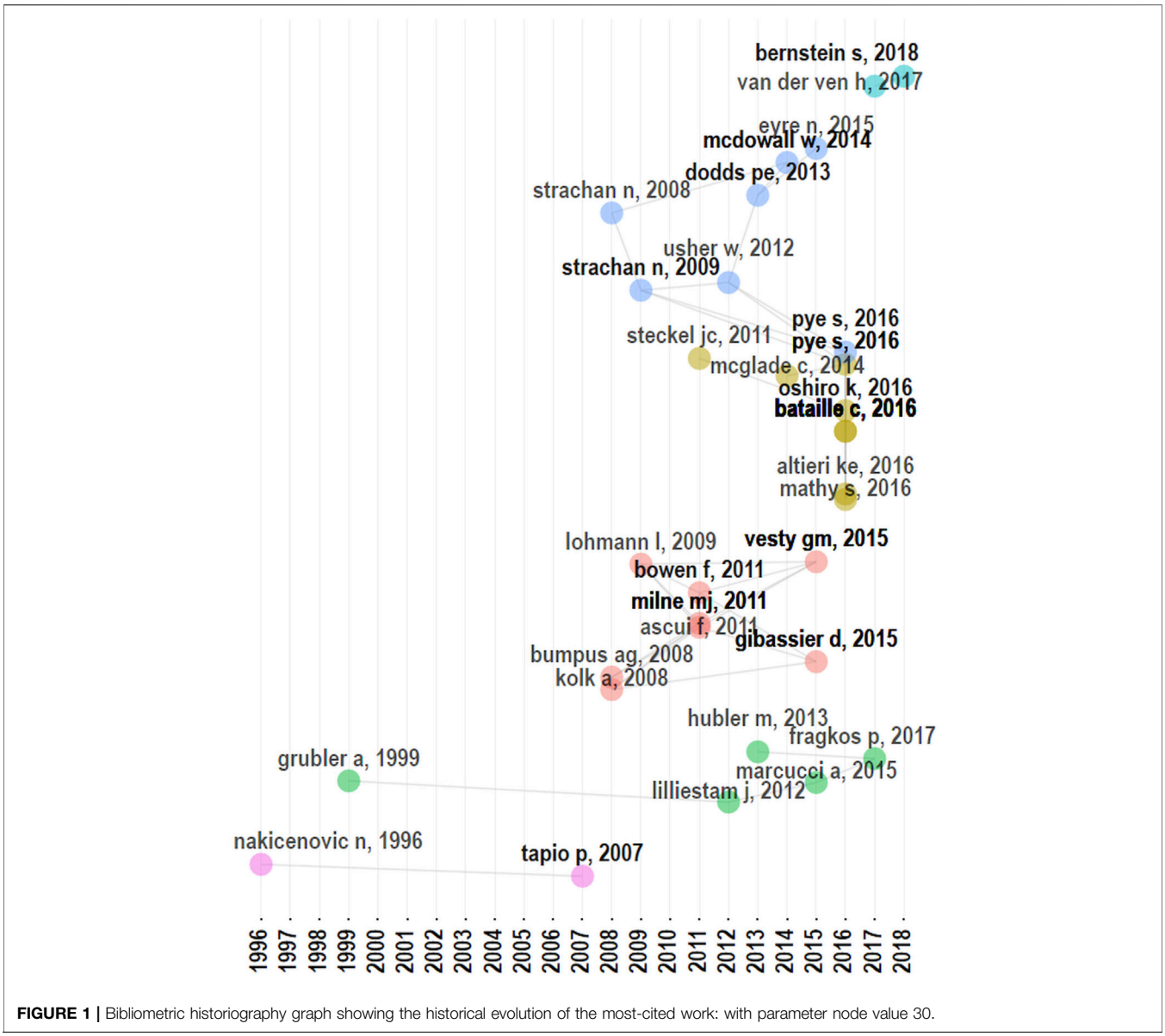



**FIGURE 1 |** Bibliometric historiography graph showing the historical evolution of the most-cited work: with parameter node value 30.

arguing that carbon emission represents one of the largest mass flows in the human activities on Earth. Most cited work is the article by Grübler et al. (1999) on energy technologies and global change, of which Nakićenović is the second author. The main research question is framed as technological choices that have long-term deterministic impacts on industrial societies and the natural environment, with an analytical focus on energy technologies that have been creating problems in the atmospheric environment. Such a research agenda has assumed the historical trend of decarbonization must be accelerated by the conscious technological choice and social change. As shown in the second line of work from the bottom of **Figure 1**, the article by Grübler et al. (1999) is cited by Lilliestam et al. (2012) on the costs and risks analysis of carbon capture and storage (CCS) versus concentrating solar power (CSP) comparison. In turn, the work is cited by Marcucci and Fragkos (2015) on the exploratory multi-model decomposition analysis of the decarbonization pathways in four major economies of China, India, Europe and U.S., suggesting that while energy efficiency remains the main strategy, the decarbonization strategy after 2030, and the negative carbon emissions after 2050 are critical. The work informs the computable general equilibrium (CGE) analysis of the EU Decarbonisation Roadmap 2050 by Hübler and Löschel (2013). It proposes a framework that integrates a forest sector model, a carbon accounting model, an ecosystem model, and climate change scenarios. The computable general equilibrium (CGE) is also used earlier by Fragkos et al. (2017) to draw policy implications on European Intended Nationally Determined Contribution (INDC). Altogether, such a line of work, as presented in the two networks at the bottom of **Figure 1**, can thus be called "systems and technologies" for their focus on the technologies and systems perspective.

### *Accounting and Accountability Cluster*

At the middle of **Figure 1**, the network begins by Bumpus and Liverman (2008) and Kolk et al. (2008) can be described as the "accounting and accountability" cluster. Although Bumpus and Liverman (2008) focus more on international governance of carbon offsets (i.e., Kyoto Protocol's Clean Development Mechanism), and Bumpus and Liverman (2008) focus more on corporate responses in the form of carbon emission reporting and disclosure mechanisms, both have informed the ensuing carbon accounting research (Lohmann, 2009; Ascui and Lovell, 2011; Bowen and Wittneben, 2011; Milne and Grubnic, 2011; Gibassier and Schaltegger, 2015; Vesty et al., 2015) with the main publication venues such as *Accounting, Auditing and Accountability Journal*. The premise of such work can be seen as based upon the calculative (or calculative agency) and persuasive features of the social need to perform (i.e., the performativity in the sociology of quantification). As such accounting performativity is applied in the international governance and corporate reporting and disclosure efforts, the notion of accountability via calculative agency becomes central to the cluster.

### *Transition and Scenario Cluster*

At the upper-side of **Figure 1**, the network begins by Strachan and Kannan. (2008), Strachan et al. (2009) and Steckel et al. (2011) can be described as the "transition and scenario" cluster. Although the work by Strachan and Kannan. (2008); Strachan et al. (2009) appears to be focusing on systems modelling work like the work in the "systems and technologies" cluster, the focus on specific policy-making insights for decarbonization scenarios (United Kingdom), highlights the interface between modeling and policy. Similarly, the work by Steckel et al. (2011) also provides scenarios-modeling work, on the case of China as indispensable role in the global goals of decarbonization, which in turn informed the research on "un-burnable oil" (McGlade and Ekins, 2014).

The cluster features a line of research focusing on pathways, especially on the global and national energy systems' transition towards decarbonization. It features a global collaborative initiative to share pathway-modeling findings and capacities, under the term "deep decarbonization" (Bataille et al., 2016a; Bataille et al., 2016b). Based on scenario-modelling findings, the decarbonization pathway findings have been published for Japan (Oshiro et al., 2016), South Africa (Altieri et al., 2016), and France and Germany (Mathy et al., 2016), right after the Paris Agreement in December 2015. An exploratory analysis has been conducted on how such decarbonization pathways may impact on global energy trade flows (Pye et al., 2016). An approach for "dynamic adaptive policy pathway" has been proposed (Mathy et al., 2016). In particular, how different modelling approaches can be applied in different national or regional contexts for deep decarbonization pathways have be discussed for both developed and developing countries (Pye and Bataille, 2016).

Although it is beyond the scope of the study to produce a complete stand-alone bibliometric historiography analysis, the three clusters, as identified and discussed above, demonstrate the common fundamental cross-cluster concerns on decarbonization (especially energy-related issues as related to carbon emissions) and different focuses thus their main approaches (i.e., the system-modelling for the systems and technologies cluster, calculative accountability for the accounting and accountability cluster, the pathway research for the transition and scenario cluster).

## Author Co-citation Analysis

Providing another perspective into the intellectual structure, the author-co-citation analysis is a bibliometric method commonly used to show the authors' influence and impact through co-citation networks (Boyack and Klavans, 2010; Aria and Cuccurullo, 2017). At the level of authors commonly cited in the dataset, **Figure 2** shows the visualization of the co-citation network.

At the center to the right, the presence of organization authors such as International Energy Agency (IEA), European Commission, United Nations Framework Convention on Climate Change (UNFCCC), Intergovernmental Panel on Climate Change (IPCC), World Bank, the Organisation for Economic Co-operation and Development (OECD), United Nations, the US Energy Information Administration (EIA), and International Renewable Energy Agency (IRENA), clearly demonstrates the significance of these organizations as important sources on the carbon emissions, climate change and energy. The work by individual lead authors consists of Beng Wah Ang's work on decomposition analysis for energy policymaking (e.g., Ang, 2004); Rogelj et al.'s perspective published in *Nature* (Rogelj et al., 2016) on the need for additional enhancement of national, sub-national and non-state actions beyond the Intended Nationally Determined Contributions (INDCs) required by the Paris Agreement, and Böhringer et al.'s major work on carbon tariffs (Böhringer et al., 2015). The work also includes an opinion piece on fairness measures (Peters et al., 2015), a major 700-page report for the United Kingdom government (Stern, 2007), and a body of work initiated by Gössling et al. on international tourism (e.g., Gössling et al., 2006).

At the lower left of **Figure 2**, the blue cluster consists of non-organizational authors' work, mainly on socio-technical transition. It features the main work by Frank W. Geels and Benjamin K. Sovacool, that culminates to the article called "Sociotechnical transitions for deep decarbonization" (Geels et al., 2017), based on their past and ongoing work such as socio-technical transitions (e.g., Geels, 2005; Geels, 2010; Geels, 2018; Geels, 2019) and transition pathways (e.g., Sovacool, 2017; Köhler et al., 2019). It also presents work by Unruh (2000) on "carbon lock-in" and by Bulkeley and Betsill (2003) on the role of global cities in climate change. The cluster also features work by the aforementioned work by Nakićenović (1996) and the seminal work on energy technology and global environmental changes by Grübler et al. (1999), which will be further discussed in **section 3.4.1**.

Finally at the upper left of **Figure 2**, the green cluster consist of major United Kingdom bodies such as Department of Energy and Climate Change (DECC), Office of Gas and Electricity Markets (Ofgem), Department for Business, Energy and Industrial Strategy (BEIS), and Climate Change Committee (CCC), an independent, statutory body established under the

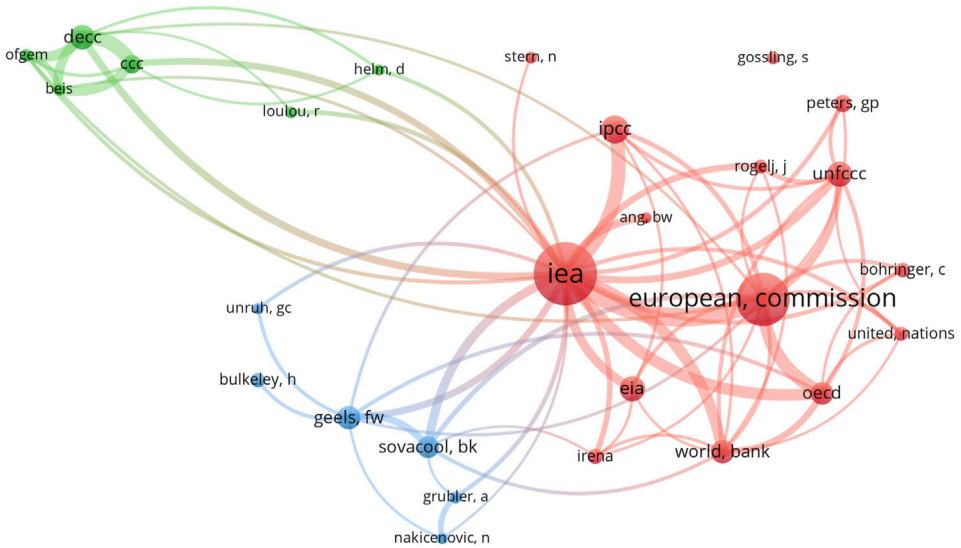



**FIGURE 2 |** Author co-citation analysis.

United Kingdom Climate Change Act 2008. It also features the work on the analysis of government failure and rent-seeking in climate change policy (Helm, 2010) and the main energy modeling documentation (Loulou et al., 2004).

In short, the historiography (**section 3.1.1**) and author co-citation network (**section 3.1.2**) provide several insights on the intellectual structure of the knowledge base, revealing the fundamental concerns, authoritative organizational and individual sources, and divergent focuses.

## Conceptual Structure and Its Evolution

To explore the conceptual structure, the study begins with a keyword co-occurrence analysis, followed by a selected list of keywords with the strongest citation bursts, and a set of thematic evolution maps. Together they will provide the relationship and the temporal dynamics among the keywords as the empirical bases to understand the structure of concepts and its evolution. Indeed, for instance, politics and just transition emerge as the most recent component contributed by mostly European actors, as evidenced by multiple findings as follows.

### Keyword Co-occurrence Analysis

To detect the main topics and evolving research frontiers, the study has conducted keyword co-occurrence analysis based on the author keywords. Because author keywords represent the succinct concepts and topics of the articles, the keyword co-occurrence analysis can reveal the conceptual structure of knowledge base, as these keywords can be further clustered into categories of concepts based on the network relationship.

**Figure 3** and **Table 2** show the keyword co-occurrence analysis outcomes of the top 30 keywords in four clusters. The first digit of the ID represents the membership of four clusters: two large-size clusters (A, and B), a mid-size cluster (C) and a smaller cluster (D). The clustering coefficient values are higher for the peripheral and often low-ranking nodes with low-level of betweenness centrality. The measurements confirm the core-periphery structure. Although the top nodes for each cluster (A-2, B-1, C-13, and D-16) can be close near the core of the network, each cluster contains other peripheral nodes with consistently high clustering coefficient values, confirming the clustering outcomes.

Note that the second part of the ID corresponds to the ranking number based on the metrics of "total links strength" first and then "occurrences," as shown to the right of the column ID in **Table 2**. The metric "total link strength" is the derived indicator provided by VOSviewer to attribute the total strength of the links of a node with other nodes in the network, and such an attribution suggests how nodes are closely related in terms of distance (MacDonald and Dressler, 2018; Yu et al., 2020; Elisha and Viljoen, 2021).

The top keyword nodes in terms of the number of occurrences (as visualized in **Figure 3** as the size of the nodes and listed in **Table 7**) are "decarbonization" (B-1), with 148 occurrences, "climate change" (A-2), with 128 occurrences. The study describes each cluster with the top keyword accordingly. Cluster B (decarbonization, in green) and cluster A (climate change, in red) occupy the central positions of the network diagram, taking up at least two-thirds of the space. The two clusters also connect to clusters C (low-carbon transition, in blue) and D (politics, in yellow).

Overall, the conceptual structure is dominated by two major themes of climate change and decarbonization, represented by clusters A and B respectively, while the themes of transition and politics appear to emerge, represented by clusters C and D respectively. The key themes such as carbon accounting, carbon neutrality and carbon trading in cluster A are closely related to the sectors of transport and forestry in the same cluster. The key themes such as decarbonization, climate policy, governance, and net zero in cluster B are closely related to the sector of energy and the entity of European Union in the same

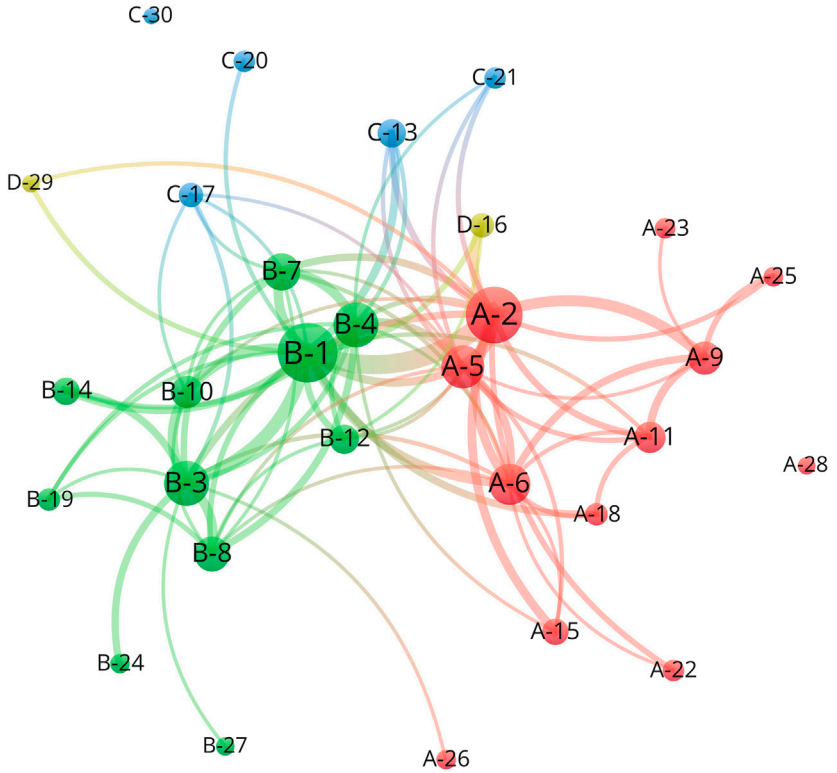



**FIGURE 3 |** Keyword co-occurrence network visualization, with each node representing a keyword (**Table 2**).

cluster. The main conceptual components of knowledge base can thus be summarized as two main set of themes: one is carbon accounting efforts for climate change mitigation (especially in relation to the sectors of transport and forestry), another is decarbonization governance for energy transition (especially in relation to the energy sector and European Union).

The smaller clusters indicate perhaps emerging or declining set of themes. Clusters C and D represents two sets of themes: one is low-carbon and sustainability transition (especially in relation to the Paris Agreement), and another is about politics and just transition.

## Keywords With Strongest Citation Bursts

To further clarify the evolution of research themes (i.e., emerging or declining), the findings detailed in **Table 3** further reveal the temporal dimensions of the top keywords with strongest citation bursts.

Indicating the intensity of attention over certain time periods, topics with the strongest citation burst provide additional insights, showing the topical transitions of the knowledge base over time (Chen, 2004). For instance, the keyword "global warming" receives relatively long-lasting, intensive attention from 1999 to 2011, indicated by the relative strength (3.05), beginning and end in **Table 3**.

**Table 3** further enriches our understanding of the knowledge base so far as follows. First the concepts of carbon sequestration, afforestation, and the "Reducing Emissions from Deforestation and Degradation" (REDD) mechanism occur intensively around 2006 and 2010, and they are closely related to the forestry sector and its critical role in carbon neutrality discussion. Then, the concepts of bioenergy and carbon accounting occur intensively around 2013, indicating the calculative agency in analyzing technological choices at that time. After the Paris Agreement signed in 2015, the themes of Germany, electricity market and transport emerge. The last 3 years of 2019–2020 show the intensive occurrences of European Union, just transition, energy storage, city, net-zero and energy transition, indicating these themes are indeed emerging, not declining ones.

## Thematic Evolution and Maps

To advance further understanding of the conceptual structure, the study has conducted a thematic evolution analysis, following the thematic analysis conventions (Cobo et al., 2011). Using the aforementioned time periods, **Figure 4** shows the thematic evolution graph generated from Bibliometrix. It reveals the continued focus on renewable energy and climate change, and the recent rise of carbon accounting, decarbonization and sustainability transitions, especially after the Paris Agreement in 2015.

Using network centrality measures, thematic maps distinguish the centrality (x-axis) and density (y-axis) of themes (topics or keywords here), as shown in **Figures 4B–D** for each periods. Applying the common interpretations of the four quadrants of themes, the maps show the flowing four themes with distinct features based on the centrality measures:

**TABLE 2 |** Main keywords: key scientometric metrics based on keyword co-occurrence network.

| Keywords | ID | Total links strength | Occurrences | Clustering coefficient | Centrality | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Eigenvector | Degree | Closeness | Betweenness |
| Climate change | A-2 | 186 | 128 | 0.651 | 0.402 | 0.966 | 0.967 | 0.032 |
| Climate change mitigation | A-5 | 114 | 69 | 0.661 | 0.268 | 0.931 | 0.935 | 0.029 |
| Carbon emissions | A-6 | 104 | 85 | 0.675 | 0.241 | 0.931 | 0.935 | 0.026 |
| Carbon accounting | A-9 | 73 | 64 | 0.737 | 0.159 | 0.690 | 0.763 | 0.011 |
| GHG emissions | A-11 | 66 | 37 | 0.727 | 0.151 | 0.759 | 0.806 | 0.014 |
| Carbon neutrality | A-15 | 50 | 32 | 0.749 | 0.119 | 0.655 | 0.744 | 0.010 |
| Transport | A-18 | 37 | 35 | 0.813 | 0.103 | 0.483 | 0.659 | 0.004 |
| Global warming | A-22 | 33 | 15 | 0.859 | 0.085 | 0.448 | 0.644 | 0.002 |
| Carbon trading | A-23 | 32 | 20 | 0.778 | 0.077 | 0.621 | 0.725 | 0.007 |
| Forestry | A-25 | 29 | 25 | 0.781 | 0.068 | 0.517 | 0.674 | 0.005 |
| Sustainability | A-26 | 28 | 23 | 0.735 | 0.052 | 0.586 | 0.707 | 0.008 |
| Emissions reduction | A-28 | 25 | 16 | 0.842 | 0.057 | 0.552 | 0.690 | 0.004 |
| Decarbonization | B-1 | 203 | 148 | 0.643 | 0.450 | 1.000 | 1.000 | 0.036 |
| Renewable energy | B-3 | 125 | 91 | 0.667 | 0.279 | 0.931 | 0.935 | 0.028 |
| Climate policy | B-4 | 123 | 74 | 0.681 | 0.292 | 0.931 | 0.935 | 0.025 |
| Policy | B-7 | 88 | 49 | 0.763 | 0.230 | 0.690 | 0.763 | 0.009 |
| Energy policy | B-8 | 79 | 43 | 0.727 | 0.184 | 0.759 | 0.806 | 0.014 |
| Electricity | B-10 | 70 | 38 | 0.723 | 0.186 | 0.759 | 0.806 | 0.015 |
| Governance | B-12 | 57 | 32 | 0.779 | 0.141 | 0.690 | 0.763 | 0.009 |
| Energy transition | B-14 | 53 | 29 | 0.743 | 0.138 | 0.655 | 0.744 | 0.010 |
| European Union | B-19 | 36 | 16 | 0.817 | 0.087 | 0.552 | 0.690 | 0.004 |
| Energy efficiency | B-24 | 30 | 23 | 0.825 | 0.070 | 0.552 | 0.690 | 0.004 |
| Net zero | B-27 | 26 | 24 | 0.857 | 0.058 | 0.483 | 0.659 | 0.002 |
| Low-carbon transition | C-13 | 55 | 31 | 0.716 | 0.148 | 0.690 | 0.763 | 0.012 |
| Technology | C-17 | 42 | 37 | 0.749 | 0.100 | 0.655 | 0.744 | 0.010 |
| Sustainability transition | C-20 | 35 | 22 | 0.743 | 0.085 | 0.586 | 0.707 | 0.008 |
| Paris Agreement | C-21 | 34 | 19 | 0.824 | 0.093 | 0.586 | 0.707 | 0.005 |
| Energy | C-30 | 19 | 19 | 0.846 | 0.043 | 0.448 | 0.644 | 0.002 |
| Politics | D-16 | 43 | 22 | 0.758 | 0.127 | 0.552 | 0.690 | 0.006 |
| Just transition | D-29 | 23 | 15 | 0.733 | 0.068 | 0.345 | 0.604 | 0.003 |

- Quadrant I: motor-theme, at the upper-right, indicating the well-established structure;
- Quadrant II: niche-theme, at the upper-left, representing well-established but less important structure;
- Quadrant III: emerging-declining-theme, at the lower-left, neither important nor developed;
- Quadrant IV: basic-theme, at the lower-right, important but under-developed;

It is revealed by **Figure 4B**, for the first period, themes such as carbon accounting, REDD, and forestry constitute the more well-established conceptual structure, whereas the themes such as climate change mitigation, climate policy and carbon neutrality remain in the basic-theme quadrant. For the second period, as shown in **Figure 4C**, themes such as carbon accounting and forestry remains well-established, joined by the theme of carbon trading. Themes such as carbon neutrality, GHG emissions, transport, climate change mitigation appear in the niche-theme quadrant, whereas renewable energy, climate policy and policy appear in the basic-theme quadrant. Confirming the previous findings based on citation bursts, the theme of electricity markets appears in the emerging-declining-theme quadrant. For the last and most recent period of 2019–2021, **Figure 4D** shows the new concept of just transition enters the well-established motor-theme quadrant, confirming the findings in **section 3.2.2**. The policy-related themes remain in the basic-theme quadrant, along with themes such as net zero and carbon footprint. The themes of sustainability transition, scenarios, renewable energy, energy transition and policy appear in the niche-theme quadrant.

Altogether, the conceptual structure of the knowledge base reveals the major and emerging themes such as decarbonization, climate change, carbon accounting, net-zero, and just transition, the most discussed sectors such as energy, forestry and transport, and entities such as the European Union. After first answering the "when" questions, the study now turns to the "where" questions. (Where have such social sciences activities on carbon neutrality occurred?)

## Social Structure

To explore the social structure of knowledge production, the study continues to examine the main journal sources and top-performing countries or regions. Together they will provide the contextual understanding of the knowledge productions in terms of their geographic and disciplinary contexts. Indeed, the body of knowledge has spanned across multiple geographic and disciplinary contexts, as evidenced by multiple findings as follows.

Since the distribution of publications across journal sources can provide initial answers to the "where" question, it is then helpful to examine the features of top journal sources.

**TABLE 3 |** Top keywords with the strongest citation bursts.

| Keywords | Strength | Begin | End | 1983–2022 |
|---|---|---|---|---|
| Global warming | 3.05 | 1999 | 2011 | |
| Energy | 2.71 | 2001 | 2008 | |
| Clean development mechanism | 2.55 | 2004 | 2008 | |
| Carbon sequestration | 3.53 | 2005 | 2013 | |
| REDD | 4.28 | 2006 | 2014 | |
| Afforestation | 2.54 | 2006 | 2010 | |
| Carbon neutrality | 4.24 | 2009 | 2012 | |
| Carbon | 4.17 | 2010 | 2015 | |
| Bioenergy | 2.6 | 2010 | 2016 | |
| Carbon accounting | 2.67 | 2013 | 2013 | |
| Carbon emission | 3.03 | 2015 | 2016 | |
| Germany | 2.78 | 2016 | 2017 | |
| Electricity market | 2.52 | 2016 | 2016 | |
| Transport | 2.95 | 2018 | 2018 | |
| European Union | 5.26 | 2019 | 2019 | |
| Just transition | 4.49 | 2019 | 2020 | |
| Energy storage | 2.99 | 2019 | 2019 | |
| City | 2.49 | 2019 | 2019 | |
| Net zero | 4.36 | 2020 | 2020 | |
| Energy transition | 2.96 | 2020 | 2022 | |

## Top Journal Sources and Their Multi- and Cross-Disciplinary Features

As mentioned earlier, the study employed the "split-apply-combine strategy" (Wickham, 2011) of data analysis. Thus, **Table 4** shows the cross-tabulation results by breaking up the bibliographic records according to the publication sources (see the first column) and then exploring their respective disciplines and top topics (see the second and third columns).

The top journal sources are *Energy Policy*, *Climate Policy*, *Energy Economics*, *Ecological Economics*, *Global Environmental Change-Human and Policy Dimensions*, and *Technological Forecasting and Social Change*, as shown in the first column of **Table 4**. The ranking numbers are shown in parentheses before the journal title names, and the numbers of the publications are shown after the colon following the titles. The titles of these journals reveal the clear presence of policies, economics, and environmental and social changes, indicating dynamic policy and economics concerns regarding such changes.

The dominant social science disciplines are Business and Economics, Public Administration, and Geography, as shown in the second column of **Table 4**. The four top journal sources also belong to Environmental Sciences and Ecology, within the broader category of "Life Sciences and Biomedicine", suggesting the significant cross-disciplinary feature across the disciplines of Environmental Sciences and Ecology and the social science disciplines of Business and Economics, Public Administration, and Geography. Additionally, the top journal *Energy Policy* belongs to the discipline of Energy and Fuels, which belongs to the broader category of "Technology". The topical features are mostly dominated by the topics of "decarbonization", "climate change mitigation", "renewable energy", "carbon accounting", and "sustainability transition" for the top journal sources.

Based on the initial answers above on the features of the top five journal sources, this section advances the analysis by examining the bibliographic coupling of all journal sources that have five or more articles. Bibliographic coupling analysis can allow us to map and cluster the journals based on the similarities of the bibliographic references (Jarneving, 2007; Boyack and Klavans, 2010).

In total, 26 journal sources out of the 887 articles of the second time period met the criteria. The bibliographic coupling analysis outcomes resulted in five clusters, as visualized in **Figure 5**. At the center is the core cluster of journal sources with label initial A in color red, such as *Energy Policy* (A-1), *Climate Policy* (A-2), *Energy Economics* (A-3), *Global Environmental Change-Human and Policy Dimensions* (A-4), and *Technological Forecasting and Social Change* (A-6). Since **Figure 5** and **Table 5** together show the bibliographic coupling relationship among journal sources, the "total link strength" indicates the total strength of such a bibliographic coupling relationship of a given journal source with other sources, thereby substantiating the empirical claims of research fronts as networks of nodes based on links of bibliometric metrics. At the upper right of **Figure 5** is the peripheral cluster of journal sources with label initial B in color green, including *Accounting Auditing and Accountability Journal* (B-8) and *Environment and Planning C-Politics and Space* (B-12). At the lower right of **Figure 5** is the peripheral cluster of sources labeled with the initial C in blue, including *Ecological Economics* (C-5), and *Journal of Forest Economics* (C-10). At the lower left of **Figure 5** is the peripheral cluster of sources labeled with the initial D in the color yellow, including *Transportation Research Part A-Policy and Practice* (D-7) and *Transport Policy* (D-9). Revealing the main research fronts while confirming the validity of bibliometric methods, **Figure 5** and **Table 5** show the topical and disciplinary closeness of journal sources. Core cluster A not only consists of the top journals, but also indicates the interdisciplinary nature of policy, economics, and social change, especially for energy and climate. The cluster B largely contains cross-disciplinary research fronts across the topics of

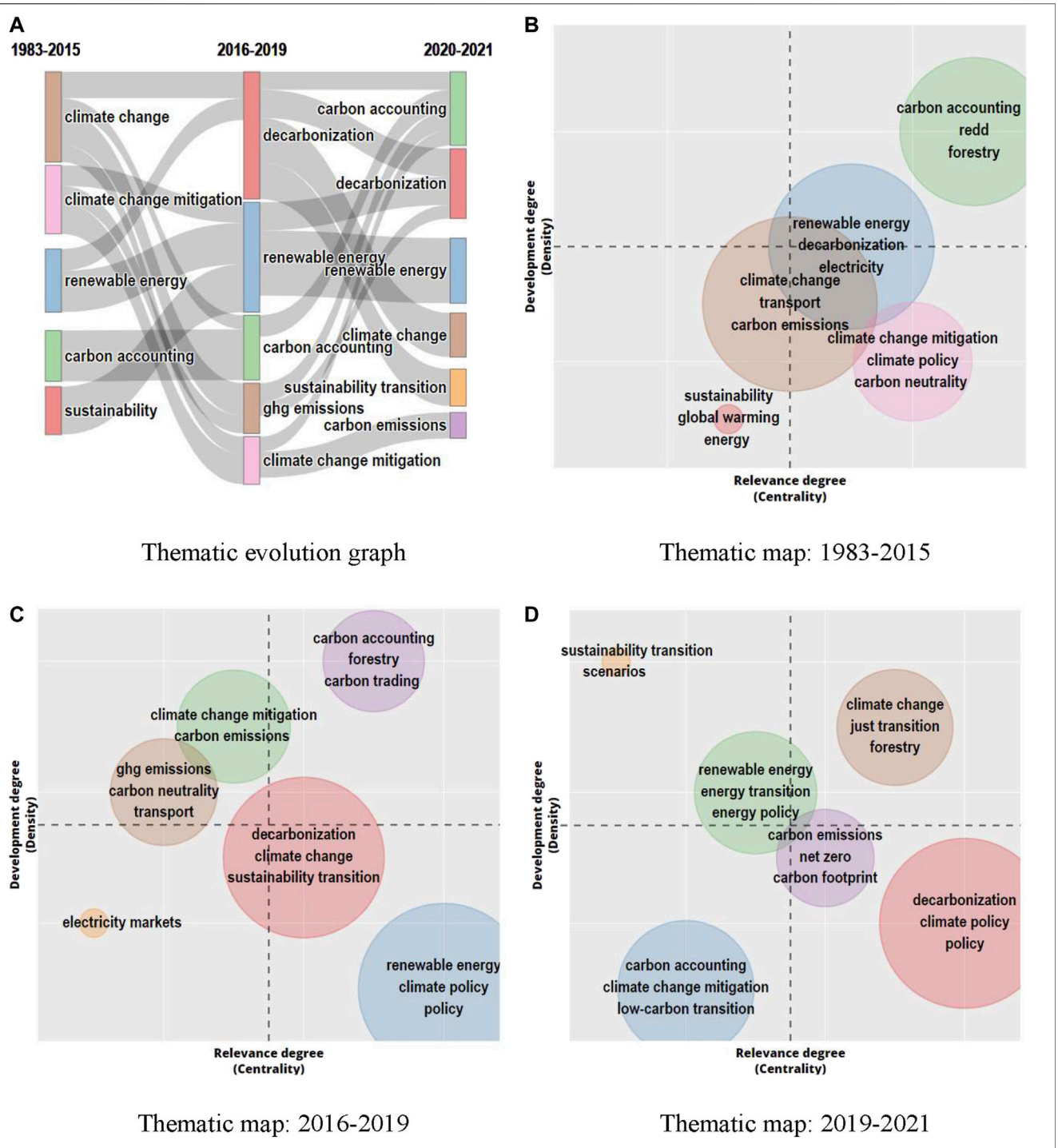



**FIGURE 4 |** Thematic evolutions: the continued focus on renewable energy and the rise of carbon accounting, decarbonization and sustainability transition. **(A)** Thematic evolution graph, **(B)** Thematic map: 1983-2015, **(C)** Thematic map: 2016-2019, **(D)** Thematic map: 2019-2021.

accounting management, geography, and politics. Cluster C concerns the research fronts of ecological, forest, and resource economics. Cluster D relates to transportation. Cluster E, with one node E-20, concerns sustainable tourism. Additionally, **Table 5** shows the citation attributes (i.e., the absolute and normalized numbers) and the network centrality metrics (i.e., closeness centrality, eigenvector centrality, and betweenness centrality) and clustering coefficients.

Overall, clusters A, B, C, D, and E show that the core clusters A and B are multi-disciplinary, with C, D, and E focusing on the cross-disciplinary topics of forestry, transportation, and sustainable tourism, respectively.

**TABLE 4 |** Top journal sources and their respective disciplines and topics.

| Top journal sources | Disciplines[a] | Top topics |
|---|---|---|
| (1) Energy policy: 252 | Business and Economics[b]; Energy and Fuels; Environmental Sciences andand Ecology | decarbonization: 56; renewable energy: 38; climate change mitigation: 21; electricity: 19; climate policy: 17; transport: 17; climate change: 14; carbon emissions: 13; energy transition: 13; energy policy: 12; GHG emissions: 11; energy efficiency: 10; carbon accounting: 6; energy system modelling: 6; just transition: 6; Paris agreement: 6; carbon trading: 5; emissions reduction: 5; energy: 5; sustainability: 5; sustainability transition: 5; bioenergy: 4; coal phase-out: 4; energy systems modelling: 4; flexibility: 4; heat pumps: 4; hydrogen: 4; net zero: 4; policy: 4; power sector: 4; United Kingdom: 4; uncertainty: 4 |
| (2) Climate policy: 84 | Environmental Sciences and Ecology; Public Administration[b] | climate policy: 23; climate change mitigation: 21; decarbonization: 14; carbon accounting: 10; carbon trading: 8; climate change: 8; energy policy: 8; Paris agreement: 8; renewable energy: 8; emissions reduction: 7; sustainability transition: 6; carbon emissions: 5; electricity: 5; energy models: 5; energy scenarios: 4; carbon leakage: 3; climate finance: 3; Kyoto protocol: 3; NDCS: 3; transport: 3 |
| (3) Energy economics: 38 | Business and Economics[b] | climate change mitigation: 8; renewable energy: 8; decarbonization: 5; carbon emissions: 4; climate policy: 4; electricity: 4; renewable integration: 4; China: 3; climate change: 3; energy modeling: 3; energy policy: 3; integrated assessment modeling: 3 |
| (4) Ecological economics: 23 | Business and Economics[b]; Environmental Sciences and Ecology | carbon accounting: 7; climate change mitigation: 5; climate change: 4; bioenergy: 3; carbon sequestration: 3 |
| (5–1) Global environmental change-human and policy dimensions: 21 | Environmental Sciences and Ecology; Geography[b] | climate change: 6; sustainability transition: 3 |
| (5–2) Technological forecasting and social change: 21 | Business and Economics[2]; Public Administration[b] | carbon emissions: 4; climate change: 3; climate policy: 3; scenarios: 3 |

[a]*Data ("Research Areas") as assigned by the Clarivate Analytics' WoS for each journal or book source.*
[b]*Disciplines ("Research Areas") categorized by Clarivate Analytics' WoS as belonging to social sciences.*

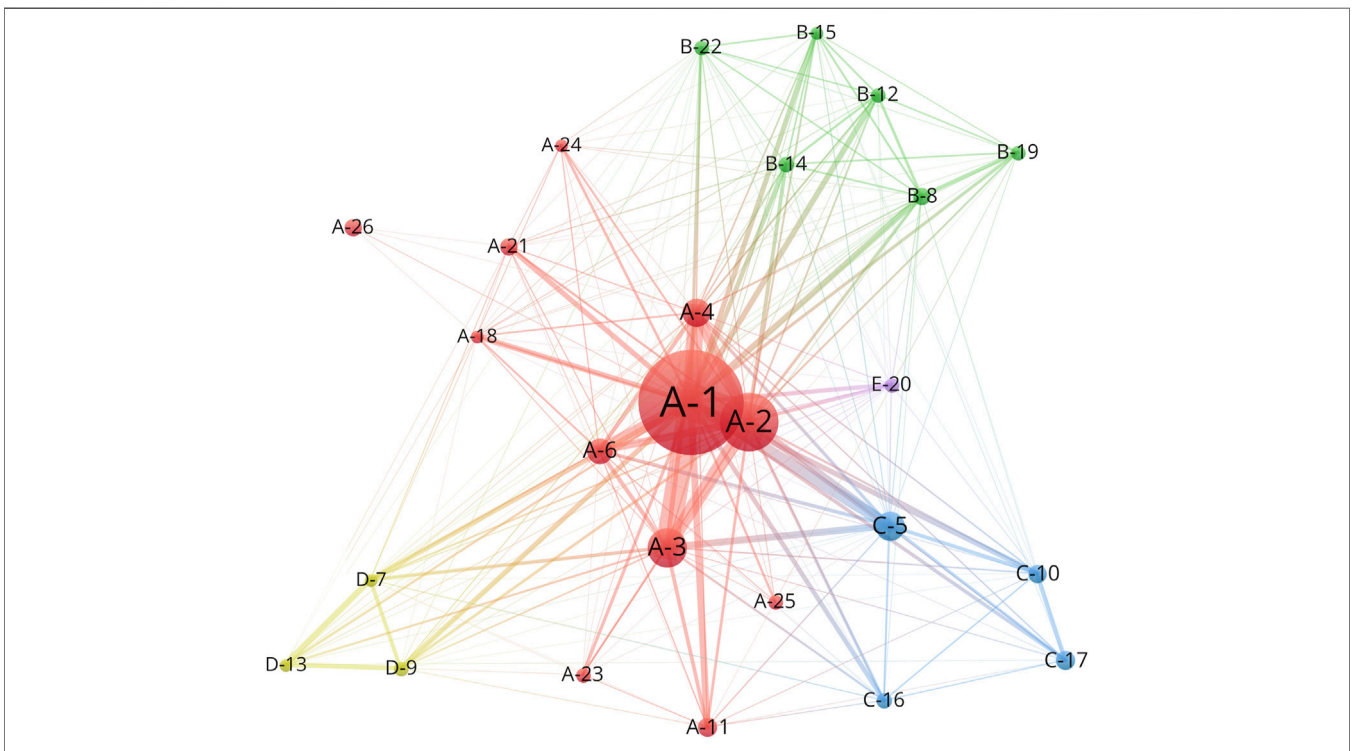



**FIGURE 5 |** Main journal sources: a bibliographic coupling network visualization, with each node representing a journal source (**Table 5**).

**TABLE 5 |** Main journal sources: key scientometric metrics.

| Journal source | ID | Total links strength | Documents | Citations | Clustering coefficient | Centrality | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Closeness | Eigenvector | Betweenness |
| Energy policy | A-1 | 3,936 | 241 | 5,969 | 0.653 | 1.000 | 0.645 | 0.053 |
| Climate policy | A-2 | 2,414 | 79 | 807 | 0.699 | 0.962 | 0.566 | 0.030 |
| Energy economics | A-3 | 1,392 | 38 | 558 | 0.699 | 0.962 | 0.358 | 0.030 |
| Global environmental change-human and policy dimensions | A-4 | 905 | 21 | 716 | 0.653 | 1.000 | 0.232 | 0.053 |
| Technological forecasting and social change | A-6 | 629 | 17 | 385 | 0.708 | 0.926 | 0.159 | 0.027 |
| Energy journal | A-11 | 248 | 10 | 100 | 0.802 | 0.694 | 0.064 | 0.007 |
| Futures | A-18 | 172 | 5 | 89 | 0.801 | 0.806 | 0.042 | 0.018 |
| Politics and governance | A-21 | 153 | 8 | 20 | 0.779 | 0.758 | 0.040 | 0.016 |
| Economics of energy and environmental policy | A-23 | 83 | 6 | 80 | 0.945 | 0.641 | 0.023 | 0.001 |
| Environmental politics | A-24 | 83 | 5 | 19 | 0.890 | 0.694 | 0.019 | 0.003 |
| Bulletin of the atomic scientists | A-25 | 47 | 6 | 46 | 0.929 | 0.595 | 0.013 | 0.001 |
| Open house international | A-26 | 7 | 8 | 28 | 1.000 | 0.543 | 0.002 | 0.000 |
| Accounting auditing and accountability journal | B-8 | 307 | 8 | 257 | 0.786 | 0.862 | 0.050 | 0.014 |
| Environment and planning C-politics and space | B-12 | 235 | 6 | 18 | 0.908 | 0.735 | 0.042 | 0.003 |
| Global environmental politics | B-14 | 227 | 7 | 119 | 0.925 | 0.735 | 0.044 | 0.002 |
| Geoforum | B-15 | 198 | 5 | 95 | 0.867 | 0.735 | 0.037 | 0.005 |
| Sustainability accounting management and policy journal | B-19 | 158 | 6 | 27 | 0.867 | 0.714 | 0.022 | 0.005 |
| Environment and planning A-economy and space | B-22 | 123 | 6 | 50 | 0.807 | 0.806 | 0.019 | 0.010 |
| Ecological economics | C-5 | 724 | 22 | 458 | 0.711 | 0.926 | 0.168 | 0.026 |
| Journal of forest economics | C-10 | 262 | 9 | 67 | 0.867 | 0.714 | 0.046 | 0.004 |
| Environmental and resource economics | C-16 | 187 | 7 | 43 | 0.885 | 0.676 | 0.043 | 0.003 |
| Forest policy and economics | C-17 | 181 | 10 | 153 | 0.894 | 0.658 | 0.026 | 0.002 |
| Transportation research Part A-policy and practice | D-7 | 318 | 5 | 41 | 0.766 | 0.806 | 0.044 | 0.013 |
| Transport policy | D-9 | 264 | 6 | 127 | 0.858 | 0.735 | 0.043 | 0.005 |
| Journal of environmental planning and management | D-13 | 229 | 5 | 68 | 0.901 | 0.694 | 0.019 | 0.003 |
| Journal of sustainable tourism | E-20 | 154 | 6 | 251 | 0.807 | 0.806 | 0.038 | 0.010 |

From the top 5 to the top 26 journals, the multi- and cross-disciplinary features were analyzed and visualized, showing how knowledge publication is distributed across sources with their associated disciplines and topics. Business and Economics, Public Administration, and Geography were shown to be the most dominant disciplines.

### Top-Performing Countries or Regions

This sub-section advances the analysis of the features of the countries, with the aim to explore national and regional contributions as well as implicit collaboration (in the form of citing the same knowledge sources) across national and regional boundaries.

At the level of countries, **Figure 6** and **Table 6** show visualizations of the bibliographic coupling network and key scientometric indicators. The first digit of the ID represents the membership of five clusters: four mid-size clusters (A, B, C, and E) and a smaller cluster (D). The most important node in terms of the number of documents (as visualized in **Figure 6** as the size of the nodes and listed in the column titled "Documents" in **Table 7**) is England (E-1), with 220 documents, as part of cluster E with other United Kingdom regions such as Scotland (E-15) and Wales (E-25), as well as Denmark (E-16). The second-ranking node is the United States (B-2), along with Australia (B-4), Canada (B-5), South Africa (B-27), and Russia (B-30). Shown in the lower part of **Figure 6** is cluster C, which contains four European countries: Germany, Italy, the Netherlands, and Austria. Clusters B and C occupy the central positions of the network diagram, connecting to the main clusters A (red) on the left and clusters D (yellow) and E (purple) on the right. With the highest number of nodes, cluster A on the left consists of many more Asian and South American countries, such as China (A-10), Japan (A-21), Brazil (A-23), India (A-2), and Singapore (A-26). Finally, the smaller cluster D (in yellow) is shown to be at the periphery on the lower right of **Figure 6**, consisting of three Nordic countries—Sweden (D-7), Norway (D-11), and Finland (D-18)—along with New Zealand (D-19). Based on the network visualization (**Figure 6**) and its centrality indicators (**Table 6**), the most central nodes are England (E-1), the United States (B-2), and Germany (C-3).

## Main Research: Most-Cited Work Across Social Science Disciplines

The previous section indicates that the body of related social science knowledge has three core clusters and other scattered peripheral work. To explore the specific details across disciplinary (and often thus geographic) contexts, a total 55 articles were selected for further review based on the top two most-cited articles in each social science discipline. Together they will provide a mosaic of disciplinary research with specific disciplinary and geographic details.

To determine the measurement of "most-cited" work, this research has used the data of the "Total Times Cited Count" (with

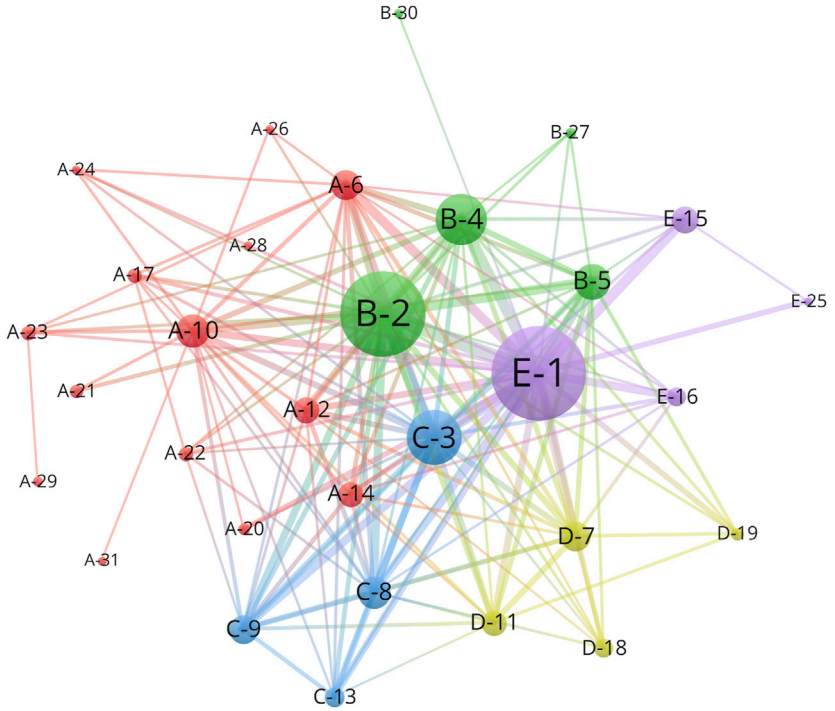



**FIGURE 6 |** Main countries: a bibliographic coupling network visualization, with each node representing a country (**Table 6**).

the field tag Z9), including citation data from the following database: Web of Science Core Collection, BIOSIS Citation Index, Chinese Science Citation Database, Data Citation Index, Russian Science Citation Index, and SciELO Citation Index.

**Table 7** lists the most-cited work across different social science disciplines, and it is sorted based on the number of publications (see the column headed "Counts"), which are unevenly distributed. The following discussions were systematically organized based on the clusters found in **Table 8**, with the aim to highlight the disciplinary features.

## Business and Economics Cross-Disciplinary Core Cluster

Based on the business and economics cross-disciplinary core cluster shown in **Table 8**, this subsection will summarize the major works in the only social science discipline in the business and economics cluster.

First, business and economics research examples include the energy technology and global environmental changes (Grübler et al., 1999), as well as the sustainability of sugarcane ethanol (Goldemberg et al., 2008). Grübler et al. (1999) proposed a useful technical analysis typology for analyzing the impact of energy technology on the global environment, especially global warming (Grübler et al., 1999). Goldemberg et al. (2008) found that the rapid expansion of sugarcane ethanol production in Brazil has caused sustainability issues, including the positive impact of air quality improvement and negative impacts on food security and field working conditions (Goldemberg et al., 2008). Both studies point to the central policy concerns of energy policies and their environmental impact (Grübler et al., 1999; Goldemberg et al., 2008).

Both articles were published in *Energy Policy*, an international journal for the political, economic, planning, environmental, and social aspects of energy. This exemplar research provides a deep analysis of the intricate relationship between energy and the environment, as well as the impacts of specific technologies and policies. The human energy technology shift from coal to oil to natural gas requires a better understanding of the impact on global environmental changes over time, including measurements such as unit energy of carbon pollution (Grübler et al., 1999). The systematic approach is also required to gain a holistic view of both the positive and negative impacts, as illustrated in the analysis of Brazil's sugarcane ethanol (Goldemberg et al., 2008). Although carbon dioxide emissions have also been reduced, this has led to worsening the conditions of high-biodiversity areas—for example, through deforestation, water contamination, soil degradation, and soil damage. Indeed, the sustainability aspects of the energy–climate (or even energy–environment) relationship require unpacking and clarity.

## Public Policy Cross-Disciplinary Core Cluster

As **Table 8** suggests that several disciplines are in close relationship with public administration (i.e., the public policy cross-disciplinary core cluster), this subsection will summarize the major work in these disciplines, as listed in **Table 7**.

Public administration work examples include climate policy (Hohne et al., 2017) and smart-sustainable cities (Yigitcanlar and

**TABLE 6 |** Main countries: key scientometric metrics.

| Country or region | ID | Total links strength | Documents | Citations | Clustering coefficient | Centrality | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Closeness | Eigenvector | Betweenness |
| France | A-6 | 7,572 | 42 | 771 | 0.966 | 1.000 | 0.217 | 0.001 |
| Peoples R China | A-10 | 6,592 | 51 | 544 | 0.966 | 1.000 | 0.177 | 0.001 |
| Switzerland | A-12 | 5,127 | 35 | 554 | 0.966 | 1.000 | 0.151 | 0.001 |
| Spain | A-14 | 3,658 | 33 | 318 | 0.975 | 0.968 | 0.095 | 0.001 |
| Greece | A-17 | 2,477 | 15 | 306 | 0.966 | 1.000 | 0.060 | 0.001 |
| Ireland | A-20 | 1945 | 12 | 100 | 0.966 | 1.000 | 0.058 | 0.001 |
| Japan | A-21 | 1934 | 14 | 237 | 0.966 | 1.000 | 0.045 | 0.001 |
| Belgium | A-22 | 1920 | 16 | 252 | 0.966 | 1.000 | 0.048 | 0.001 |
| Brazil | A-23 | 1854 | 15 | 682 | 0.966 | 1.000 | 0.045 | 0.001 |
| India | A-24 | 1,434 | 6 | 186 | 0.989 | 0.909 | 0.039 | 0.000 |
| Singapore | A-26 | 1,097 | 8 | 153 | 0.973 | 0.968 | 0.029 | 0.001 |
| Poland | A-28 | 689 | 8 | 96 | 0.984 | 0.938 | 0.019 | 0.001 |
| Portugal | A-29 | 678 | 9 | 102 | 0.993 | 0.857 | 0.015 | 0.000 |
| Lithuania | A-31 | 496 | 5 | 55 | 0.993 | 0.857 | 0.010 | 0.000 |
| United States | B-2 | 16,422 | 191 | 3,608 | 0.966 | 1.000 | 0.420 | 0.001 |
| Australia | B-4 | 8,614 | 93 | 1,686 | 0.966 | 1.000 | 0.245 | 0.001 |
| Canada | B-5 | 7,586 | 55 | 839 | 0.966 | 1.000 | 0.237 | 0.001 |
| South Africa | B-27 | 994 | 11 | 125 | 0.975 | 0.968 | 0.030 | 0.001 |
| Russia | B-30 | 630 | 6 | 24 | 0.997 | 0.857 | 0.020 | 0.000 |
| Germany | C-3 | 13,129 | 102 | 1,584 | 0.966 | 1.000 | 0.358 | 0.001 |
| Italy | C-8 | 7,240 | 48 | 570 | 0.966 | 1.000 | 0.206 | 0.001 |
| Netherlands | C-9 | 7,031 | 41 | 1,036 | 0.966 | 1.000 | 0.200 | 0.001 |
| Austria | C-13 | 4,232 | 24 | 319 | 0.975 | 0.968 | 0.130 | 0.001 |
| Sweden | D-7 | 7,307 | 42 | 920 | 0.966 | 1.000 | 0.207 | 0.001 |
| Norway | D-11 | 5,319 | 33 | 409 | 0.966 | 1.000 | 0.152 | 0.001 |
| Finland | D-18 | 2,353 | 22 | 331 | 0.966 | 1.000 | 0.069 | 0.001 |
| New Zealand | D-19 | 2,314 | 13 | 297 | 0.975 | 0.968 | 0.071 | 0.001 |
| England | E-1 | 22,783 | 220 | 4,458 | 0.966 | 1.000 | 0.498 | 0.001 |
| Scotland | E-15 | 3,571 | 38 | 776 | 0.966 | 1.000 | 0.130 | 0.001 |
| Denmark | E-16 | 2,968 | 22 | 318 | 0.970 | 0.968 | 0.101 | 0.001 |
| Wales | E-25 | 1,142 | 8 | 167 | 0.997 | 0.882 | 0.043 | 0.000 |

Lee, 2014). In specific, the importance and implications of the 2015 Paris Agreement were discussed and examined by Hohne et al. (2017), with the analytical focus on the "individual intended nationally determined contributions" (INDCs); they found that the progress of decarbonization has been made faster than expected in sectors such as solar photovoltaics, onshore wind, and electric cars, and they argued that policies must be implemented in order to meet the INDC pledges made by national governments to resolve the inconsistency between INDCs and global goals such as the 2015 Paris Agreement (Hohne et al., 2017). At the city level, eco-city initiatives were studied by Yigitcanlar and Lee (2014), focusing on the Korean case of ubiquitous-eco-city (u-eco-city), with the goal being to distinguish whether this contributes to an ideal model of sustainable and smart urban forms beyond being a mere branding event (Yigitcanlar and Lee, 2014). Both studies hint at the political, financial, technological, and other forms of support in implementing such policy initiatives (Yigitcanlar and Lee, 2014; Hohne et al., 2017).

Published in the journal *Habitat International* (Zuo et al., 2012; Kibwami and Tutesigensi, 2016), these exemplar works in urban studies recognize the importance of sustainability agenda in the building sector as one of the biggest contributors of carbon emissions and energy consumption. With the aim of examining the factors that hinder or help carbon neutrality goals regarding reducing emissions in commercial developments, Zuo et al. (2012) conducted semi-structured interviews of practitioners and found that a clear definition of carbon-neutral building is needed, along with other facilitating factors such as material choice, managers' knowledge, market demand, government support, and leadership (Zuo et al., 2012). Kibwami and Tutesigensi (2016) investigated whether and how the integration of embodied carbon in the development approval process by regulators can facilitate sustainability in construction, based on a set of structured interviews regarding policy proposals in Uganda (Kibwami and Tutesigensi, 2016). In the development approval process, the authors argue, both environmental and social sustainability should be considered, even in developing countries. Similarly, an exemplar work in education discussed sustainability initiatives at a Canadian university and proposed a multi-bottom line approach based on a set of tools using a driving-force–pressure–state–exposure–effect–action framework (Waheed et al., 2011). Stadel et al. (2011) developed engineering education that integrates carbon accounting and building information modeling as the basis for intelligent sustainable design (Stadel et al., 2011).

## Social Science Disciplinary Core Cluster

Based on the social science disciplinary core cluster shown in **Table 8**, this subsection will summarize the most-cited works in these disciplines, as listed in **Table 7**.

**TABLE 7 |** Main research: social science work.

| Disciplines | Counts | Main work | Key ideas |
|---|---|---|---|
| Business and economics | 557 | Grübler et al. (1999), Goldemberg et al. (2008) | Energy technology and global environmental changes Grübler et al.(1999); sustainability of sugarcane ethanol Goldemberg et al. (2008). |
| Public administration | 172 | Hohne et al. (2017), Yigitcanlar and Lee (2014) | Individual intended nationally determined contributions (INDCs) Hohne et al.(2017); Eco-city initiatives, Korean u-eco-city initiatives Yigitcanlar and Lee (2014) |
| Government and law | 75 | Vandenbergh and Steinemann (2007), Green (2013) | The carbon-neutral individual, carbon neutrality norms Vandenbergh and Steinemann (2007), the institutional complexity of carbon neutrality, and carbon management Green (2013) |
| Geography | 70 | Bumpus and Liverman (2008), Byers et al. (2014) | International governance of carbon offsets Bumpus and Liverman (2008); United Kingdom decarbonization pathways for the United Kingdom Byers et al. (2014) |
| Urban studies | 33 | Zuo et al. (2012), Kibwami and Tutesigensi (2016) | Sustainability agenda in the building sector Zuo et al. (2012) and related regulatory policies Kibwami and Tutesigensi (2016) |
| International relations | 32 | Green (2013), van der Ven et al. (2017) | Institutional complexity under the Kyoto Protocol (Green, 2013); contributions by nonstate and subnational actors on climate governance van der Ven et al. (2017) |
| Development studies | 31 | Zuo et al. (2012), Andrews (2008) | Sustainability agenda in the building sector Zuo et al. (2012); land-use pattern changes Andrews (2008) |
| Social sciences-other topics | 29 | Higham and Cohen (2011), Gössling (2009) | Consumers' attitudes regarding the impact of tourism and travel on climate change Higham and Cohen (2011); industrial actors' role in greenhouse gas emissions Gössling (2009) |
| Sociology | 15 | Scott et al. (2019), Marres (2011) | Public consent and acceptability Scott et al. (2019); the role of technology in carbon accounting Marres (2011) |
| Social issues | 12 | Pidgeon and Demski (2012), Boehmer-Christiansen (2003) | Public consent and acceptability Pidgeon and Demski (2012); the role of technology in carbon accounting Boehmer-Christiansen (2003) |
| Education and educational research | 11 | Waheed et al. (2011), Stadel et al. (2011) | University sustainability initiatives, using a driving force-pressure-state-exposure-effect-action framework Waheed et al. (2011); intelligent sustainable design education that integrates carbon accounting for engineers Stadel et al. (2011) |
| Area studies | 4 | Benjaminsen (2017), Gilley (2017) | The legitimacy of the local adoption of global policy frameworks such as REDD+ Benjaminsen (2017); integrated governance solutions and central-local relations on environmental issues Gilley (2017) |
| Psychology | 4 | Quimby and Angelique (2011), Payne (2020) | Pro-environmental behavior research and community psychology knowledge aiming for a carbon neutral future Quimby and Angelique (2011); psychologically vulnerable groups impacted by decarbonization Payne (2020) |
| Communication | 2 | Hopke and Hestres (2018), Koteyko (2012) | Visual narratives, the media and climate stakeholder relationship Hopke and Hestres (2018); critical discourse analysis work on carbon neutrality Koteyko (2012) |
| Linguistics | 2 | Dury (2008), Koteyko (2012) | The English term "carbon neutral" evolution Dury (2008); critical discourse analysis work on carbon neutrality Koteyko (2012) |
| Social work | 2 | Quimby and Angelique (2011), Chapman and Boston (2007) | Community psychology knowledge relating to a carbon-neutral future Quimby and Angelique (2011); public attitudes towards and behaviors concerning carbon emission reduction Chapman and Boston (2007) |
| Archaeology | 1 | Pigliautile et al. (2019) | Environmental sustainability and energy efficiency in buildings Pigliautile et al. (2019) |
| Cultural studies | 1 | Pallesen (2016) | Cultural and framing politics of pricing Pallesen (2016) |

Government and law discipline works include the seminal work by Vandenbergh and Steinemann (2007) on the legal and regulatory basis on the notion of the carbon-neutral individual (Vandenbergh and Steinemann, 2007) and the analysis of carbon management standards by Green (2013). Drawing upon empirical studies and norms theory, the notion of the carbon-neutral individual suggests the possibility of advancing legal reforms that combine the existing norm of personal responsibility with the emerging norm of carbon neutrality (Vandenbergh and Steinemann, 2007). Focusing on the role of private authority, the network analysis of public and private standards for carbon management reveals an emerging climate change governance order that is complex, suggesting the importance of private authority in institutional complexity (Green, 2013). Both articles have contributed to the legal and regulatory imagination to address the governance and management issues surrounding carbon neutrality and carbon management.

Works on major international relations includes the article by Green (2013), which was not only published in *Global Environmental Politics* but also touches upon the institutional complexity under the Kyoto Protocol (Green, 2013). Additionally, the article by van der Ven et al. (2017) that was published in the same journal examined the effect of the valuable contributions of non-state and subnational actors on climate governance. It proposes an approach that is more comprehensive in identifying how such contributions can become far-reaching and durable (Van der ven et al., 2017). Such an approach is applied to an analysis of the Carbon Trust's initiative on the creation of product carbon footprints, indicating some initial

**TABLE 8 |** Main disciplines: key scientometric metrics.

| Disciplines | ID | Total links strength | Documents | Citations | Clustering coefficient | Centrality | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Closeness | Eigenvector | Betweenness |
| Public administration[a] | A-4 | 26,649 | 161 | 1762 | 0.783 | 1.000 | 0.216 | 0.023 |
| Science and technology-other topic | A-9 | 4,767 | 28 | 486 | 0.871 | 0.920 | 0.031 | 0.008 |
| Urban studies[a] | A-11 | 3,837 | 32 | 196 | 0.831 | 0.958 | 0.023 | 0.014 |
| Education and educational research[a] | A-19 | 598 | 11 | 54 | 0.895 | 0.852 | 0.003 | 0.007 |
| Operations research and management science | A-20 | 502 | 7 | 90 | 0.895 | 0.742 | 0.005 | 0.004 |
| Engineering | A-21 | 427 | 5 | 57 | 0.912 | 0.719 | 0.002 | 0.003 |
| Architecture | A-22 | 401 | 9 | 30 | 0.956 | 0.639 | 0.002 | 0.001 |
| Government and law[a] | B-6 | 10,623 | 74 | 517 | 0.852 | 0.920 | 0.075 | 0.011 |
| International relations[a] | B-7 | 6,567 | 32 | 271 | 0.930 | 0.852 | 0.044 | 0.003 |
| Social sciences-other topics[a] | B-10 | 4,282 | 27 | 546 | 0.818 | 0.958 | 0.026 | 0.017 |
| Sociology[a] | B-14 | 2,589 | 15 | 273 | 0.895 | 0.885 | 0.017 | 0.006 |
| Social issues[a] | B-18 | 743 | 11 | 56 | 0.971 | 0.742 | 0.003 | 0.001 |
| Area studies[a] | B-23 | 181 | 4 | 16 | 1.000 | 0.697 | 0.001 | 0.000 |
| Environmental sciences and ecology | C-1 | 80,444 | 500 | 9,497 | 0.783 | 1.000 | 0.598 | 0.023 |
| Business and economics[a] | C-2 | 72,624 | 535 | 10,737 | 0.783 | 1.000 | 0.585 | 0.023 |
| Energy and fuels | C-3 | 47,617 | 255 | 6,096 | 0.783 | 1.000 | 0.481 | 0.023 |
| Forestry | C-13 | 2,704 | 21 | 228 | 0.930 | 0.852 | 0.027 | 0.003 |
| Development studies[a] | D-8 | 5,104 | 29 | 383 | 0.818 | 0.958 | 0.033 | 0.017 |
| Transportation | D-12 | 3,081 | 18 | 251 | 0.967 | 0.767 | 0.026 | 0.001 |
| Agriculture | D-16 | 1,143 | 5 | 76 | 0.930 | 0.852 | 0.010 | 0.003 |
| Geography[a] | E-5 | 12,052 | 69 | 1,479 | 0.783 | 1.000 | 0.096 | 0.023 |
| Psychology[a] | E-24 | 27 | 3 | 16 | 1.000 | 0.590 | 0.000 | 0.000 |
| Physical geography | F-15 | 1,444 | 16 | 628 | 0.963 | 0.793 | 0.008 | 0.002 |
| Geology | F-17 | 744 | 9 | 349 | 1.000 | 0.767 | 0.003 | 0.000 |

[a]*Disciplines ("Research Areas") as assigned by Clarivate Analytics' WoS.*

failures and suggestions for policymakers to measure the contribution to global decarbonization more accurately.

A key "social sciences-other topics" work consists of an article on Norwegian attitudes towards the impact of air travel on climate change. The article concludes that Norway is sensitive to climate issues and thus a vanguard European tourism market (Higham and Cohen, 2011). Additionally, on the topic of tourism Gössling (2009) provides a critical review of carbon neutrality within the context of the UNWTO's Davos Declaration, which ascribed the responsibility to reduce greenhouse gas emissions to the main tourism industrial actors (Gössling, 2009).

One sociology paper by Scott et al. (2019) also contributes to the understanding of the consequences facing the tourism sector through the systematic construction of the Climate Change Vulnerability Index for Tourism (CVIT) (Scott et al., 2019). This research found that Africa, the Middle East, South Asia, and Small Island Developing States are highly vulnerable and that climate change will hinder the contributions of tourism to the Sustainable Development Goals. From the perspective of everyday life and devices of accounting, another sociology work carried out by Marres (2011) examined the role of technology in public participation in carbon accounting, suggesting the capacity to "co-articulate" participation (Marres, 2011).

From the discipline of social issues, the exemplar work by Pidgeon and Demski (2012) on the public's attitudes towards energy transformation highlights the importance of public consent and acceptability in the adoption of emerging renewable energy projects (Pidgeon and Demski, 2012). The work by Boehmer-Christiansen (2003) on the need for global treaty obligations to advance the transition to green fuels and technologies asks several questions regarding equity and political instability during the decarbonization of global energy supplies (Boehmer-Christiansen, 2003).

From area studies, the work by Benjaminsen (2017) examined why the global policy framework known as Reducing Emissions from Deforestation and Forest Degradation (REDD+) in Zanzibar is at risk of early enthusiasm and then ensuing abandonment due to its lack of legitimacy in terms of durable community forest management (Benjaminsen, 2017). Another area studies work by Gilley (2017) compared the local success in reducing greenhouse gas emissions in China (Guangzhou) and India (Gujarat), revealing similar patterns of integrated governance solutions but different institutional linkages: mainly intra-governmental linkages for China and mainly state-society linkages for India. The research also showed the insignificant to marginal effects of international negotiations and national frameworks, with implications for central–local relations regarding environmental issues and for international aid and assistance (Gilley, 2017).

## Other Peripheral Disciplinary Work

Other social science disciplines consist of Development Studies, Geography, Psychology, Communication, Linguistics, Social Work, Archaeology, and Cultural Studies, each with a few publications (some four and often two). Nonetheless, it is helpful to examine how each social science discipline has contributed to this topic so far.

One major development studies work by Zuo et al. (2012) was previously discussed in the urban studies discipline, not only because it was published in the cross-disciplinary journal Habitat International but also because it discusses the need for cultural change in the construction industry towards sustainability (Green, 2013). Another research article by Andrews (2008) investigated how land use and greenhouse gas emissions are related using exploratory case studies to illustrate the varying per capita carbon dioxide emissions. This study found that buildings typically contribute more emissions than personal transportation and argued that holistic solutions to this problem must incorporate technical fixes (especially green buildings) and land-use pattern changes (considering population densities and transportation).

One important work in the field of geography is an article published in Economic Geography by Bumpus and Liverman (2008) on the governance of carbon offsets. Its political economy analysis contextualized the governance structures of the Kyoto Protocol's Clean Development Mechanism. The analysis illustrates how carbon offsets represent the strategies of capital accumulation. Another geography research article tested six decarbonization pathways for the United Kingdom, with specific discussions of the role of water in electricity generation and cooling water use (Byers et al., 2014).

Following the traditions of pro-environmental behavior research, one essential psychology work by Quimby and Angelique (2011) analyzed the perceived barriers and catalysts for encouraging pro-environmental behavior within the environmental movement. Based on the perceived barriers they found (such as hopelessness, low efficacy, and time and money resources) and the catalysts identified (such as changing social norms through institutional and educational transformation), the article argues that knowledge of community psychology can contribute to the fight for a carbon-neutral future (Quimby and Angelique, 2011). Based on the fieldwork examining cobalt mining in the Democratic Republic of the Congo (DRC), another important psychology work by Payne (2020) highlighted the injustice of the "decarbonization divide", connecting decarbonization with issues of environmental destruction, child labor, energy dispossession, and gender inequality, with certain ethnic groups being rendered more psychologically vulnerable (Payne, 2020).

Coming from the discipline of communication research, Hopke and Hestres (2018) visualized the media and climate stakeholder relationship based on Twitter activity relating to the Paris Climate Talks (Hopke and Hestres, 2018). The findings showed that the Twitter accounts of fossil fuel and trade association, as major outliers, had put more emphasis on former US President Barack Obama's climate policy instead of the more common climate change visual framing used by activists, movement organizations, multinational representatives, and scientific experts. These accounts' messages aimed to create a visual narrative of the lack of domestic support for his climate policies in the global arena. Additionally, focusing on the discourse surrounding carbon, the media discourse analysis by Koteyko (2012) examined the "market-driven sustainability" in the British media, specifically United Kingdom national newspapers, between 1990 and 2009 (Koteyko, 2012). The findings reveal that finance-related carbon compound phrases (1990–2005) occur before other compound phrases, such as "low-carbon", "zero-carbon", and "carbon-neutral" (popular after 2005). This study argues that the climate change debate has been narrowed because carbon compound phrases suggest that carbon emissions management starts from calculation and monetization.

Linguistics works include the communications research work carried out by Koteyko (2012), as discussed earlier, as it was published in a cross-disciplinary journal called Language and Communication. Additionally related to communication research is an article by Dury (2008) from the linguistics journal Terminology, based on a bilingual (French and English) comparable corpus, which shows that the English term "carbon neutral" has followed a particular development pattern: it becomes more widespread in the media over time (i.e., determinologized) and then moves into the ecology lexicon (i.e., terminologized) (Dury, 2008).

From the discipline of social work, the previously discussed work by Quimby and Angelique (2011) is included, since it was published in the cross-disciplinary journal *American Journal of Community Psychology* (Quimby and Angelique, 2011). The social work article by Chapman and Boston (2007) examines the social implications of decarbonization in the New Zealand economy, outlining the likely economic and distributional impacts of related policies on public attitudes towards and behaviors concerning carbon emission reduction (Chapman and Boston, 2007). The archaeology work by Pigliautile et al. (2019) examines historical buildings to explore an innovative approach to microclimate enhancement using thin-envelope materials (Pigliautile et al., 2019). The cultural studies work by Pallesen (2016) examines the cultural and framing politics of pricing by examining the adoption of tariffs and wind power pricing in France during the decarbonization of the electricity sector, raising the issues of the social and governance dynamics of value attribution and negotiation (Pallesen, 2016).

## DISCUSSION

Overall, our findings confirm our assumption that the issues of carbon neutrality and decarbonization are multi- and cross-disciplinary; these detailed and systematic findings provide specific empirical evidence to quantify and qualify the existing knowledge in this area. As carbon-neutral efforts are shown to be made in response to various issues of climate change, collective action requires not only technical innovations for transforming various modern industrial sectors, but also social science understanding of the ways in which carbon information can be instrumental in reshaping our social, economic, and political interactions. Researchers and policy-makers have so far contributed to the discussions on the role of technologies, policies, and markets in achieving carbon neutrality. Facing

the climate change mitigation opportunities and challenges, issues such as measuring and accounting decarbonization efforts, sustainability transition, and just transition have been discussed in specific sectors such as energy, transport, forestry, agriculture, forestry, and tourism.

Our descriptive summary of 907 articles revealed the fast-growing publications patterns, especially after the Paris Agreement in 2015, with temporal evolution in intellectual and conceptual structures of the knowledge.

In terms of intellectual structure, the established state-of-the-art knowledge has been built based on the intellectual origin of 1) systems and technology cluster, 2) accounting and accountability cluster, and 3) transition and scenario clusters. Both historiography and author co-citation analysis have revealed the intellectual origins and authoritative sources that have founded and shaped the body of knowledge. The largely macro-, regional, sectorial and country-level of analysis aim to address the fundamental concerns of human technologies' impact on the environment, with the initial system modelling work growing into technology-, policy- and organization-level analysis towards more sustainable scenarios and pathways.

In terms of conceptual structure, the last few years (2019–2022) have seen the rise of keywords such as decarbonization, energy transition, the Paris Agreement, just transition, European Union, and net-zero. Such a trend indicates the impacts of political actions and policies on carbon neutrality research for just and equal transitions towards sustainability. Undoubtedly, what decarbonization opportunities and challenges may occur for the industry and environment has become the basis for the most relevant and cutting-edge research topics across various social science disciplines.

In terms of social structure of knowledge publication, the bibliographic coupling analysis of journals shows that the two main core journal clusters are multidisciplinary while the other three smaller clusters focus on cross-disciplinary topics such as forestry, transportation, and sustainable tourism. Throughout the years, main journals such as *Energy Policy*, *Climate Policy*, *Energy Economics*, *Ecological Economics*, *Global Environmental Change-Human and Policy Dimensions*, and *Technological Forecasting and Social Change* have accumulated a substantial number of studies, connecting social science disciplines such as Business and Economics, Public Administration, and Geography with other disciplines such as Energy and Fuels and Environmental Sciences and Ecology. Publishing works on carbon neutrality that connect social science and non-social science categories; the journals have been identified in their contribution to have bridged the related disciplines. The empirical evidence on top-performing countries shows the relatively low presence of Asia and Africa, indicating a gap in the knowledge production and collaboration needed to tackle the issues of decarbonization as part of the global goals for sustainable and just transitions.

In terms of social discipline perspective, the scientometric approach of science mapping has revealed the otherwise partial and incomplete impressions of the social science literature in a systematic whole. The overall picture is much more diverse than the aforementioned intellectual structural base. Specific connections have been explored in discussing the specific main research works in **section 3.4**, revealing a vibrant picture of policy, professional, educational, and academic work that contributes to human knowledge on carbon neutrality and decarbonization. While the backbone of cross-disciplinary carbon neutrality research remains at the intersection between the "Business and Economics" and "Life Sciences and Biomedicine" disciplines, an emerging critical body of work has begun to question the limitations of "market-driven sustainability" and the social and emotional costs and implications of decarbonization, which coincides with the emerging notion of just transitions.

## CONCLUSION

Significant progress has been made, especially since the Paris Agreement in 2015, on the issues of decarbonization and carbon neutrality, as part of the arguably the largest human efforts in transitioning energy, transport, industries, and cities towards low-carbon and net-zero sustainability. It is significant because human-caused climate change needs human efforts and knowledge to change the socio-technical and socio-ecological systems we have constructed. The emphasis on the "socio" requires a comprehensive overview of the state of art in social science knowledge on the topic.

This scientometric and systematic review provides a comprehensive and disciplinary examination of the body of social science knowledge produced so far on the topic of carbon neutrality. However, it is acknowledged that the scope of this review does not include conference proceedings and other works that are not indexed by the Web of Science database. It has limitations in depth because more specific carbon-neutrality keywords such as forest carbon sinks, carbon capture, utilization and storage (CCUS), and so on were not included in the search query design. It is also limited to the trade-off between comprehensiveness and depth that any scientometric and systematic review must decide and justify. More in-depth analysis in the future should include more specific keywords by snowballing specific set of keywords, and the conceptual structure findings here can be useful references. Future work must determine such trade-off in using scientometric and systematic reviews to understand the multi-disciplinary topic of decarbonization.

Nonetheless, the findings presented here amount to the first scientometric analysis of the social science knowledge of carbon neutrality. This study provides an updated understanding and highlights the established and emerging relationship between policies and technologies, especially those relating to energy, transport, construction, education, and other sectors. The main body of work have been policy-oriented to examine the industrial and economic activities, especially those surrounding energy, transport, forestry, and cities-related technologies, with the practical purpose to outline transitioning pathways and scenarios. The emerging conceptual structure findings reveal the substantial impact of just transition, European Union, net-

zero, cites, and energy transition on the most recent literature. Such a phenomena can be explained by the social structure of the current knowledge production, where European and North American regions remain dominant.

The main controversy appears to be about justice in the decarbonization processes, as reflected in the recent literature and European Union policy on "just transition." Research gaps in the case studies beyond European and North American regions, cities, and people must be filled, especially the impact of climate change is expected to be more severe in developing regions such as Asia, Africa, and South America.

In terms of the potential development, we believe that, based on the findings presented here, the existing backbone of cross-disciplinary carbon neutrality research, which remains at the intersection between the "Business and Economics" and "Life Sciences and Biomedicine" disciplines, must engage the emerging critical literature to reflect on the limitations of the mainstream market-driven, business-driven, and data-driven approaches of decarbonization efforts aiming at carbon neutrality. Given the increasing knowledge demand for the development of policies and technologies relating to carbon neutrality across various industrial sectors and different academic disciplines, such a systematic science-mapping work provides significant results that will help to foster new conversations and enrich old ones. The descriptions on the most-cited work in each social science discipline in **section 3.4** should begin to provide a more balanced perspective in terms of disciplines. Such knowledge, for instance, can be more personal and emotional at the micro-levels, beyond the initial knowledge base on country- and organization-level scenarios and pathways. Future work is needed, for instance, to deliver in-depth systematic reviews for each social science discipline.

These science-mapping outcomes lead to the final research question we will raise upfront: How can researchers and policy-makers understand this overview of social science literature relating to the topic of carbon neutrality, with a conceptual or taxonomic framework, to address the issues in achieving decarbonization and carbon neutrality, especially in terms of the design and application of relevant policies and technologies? While the findings described here may not be sufficient to generate a definitive conceptual framework, a set of common taxonomic features across disciplines and sectors can be built that includes decarbonization, carbon management, carbon accounting, sustainability, sustainable development, and just transition. Future work can advance more in-depth analysis within a specific social science discipline or a specific cross-disciplinary mix. As carbon-neutral efforts are shown to be carried out in response to various issues of climate change, a collective action that requires not only technical innovation for transforming various modern industrial sectors but also a social science understanding of how carbon information can be instrumental in reshaping our social, economic, and political interactions, especially for coming up with new strategy for green and digital transitions, is needed (European Commission, 2020).

Several directions for future research relating to social science knowledge about carbon neutrality are discussed here. First, since carbon information, including GHG emission information, is key for constructing a sustainable future, it is important to understand the overall and specific datafication processes of selecting, recording, monitoring, and acting upon any specific piece and set of carbon information concerning human activities and the environment, as a necessary and evidence-based grounding. The datafication of carbon information requires better knowledge in data science and social data science that can identify relevant and critical carbon information that contributes to the socio-technical transition to a sustainable future, especially the unjust cost of decarbonization processes. Second, as carbon information-based products, systems, and stories have increasingly involved several important industrial bodies and various human daily activities, it is vital to understand the conflicts, struggles, and negotiations that institutions, organizations, and individuals are likely to encounter. Decisions, interactions, and participation processes relating to carbon information, ranging from carbon trading to sustainable consumption, demands adequate empirical and theoretical knowledge that can help in integrating relevant carbon information into the socio-technical systems of everyday life. Third, collaboration surrounding carbon information, carbon accounting, and carbon neutrality will and should have become the focus of research and development for stakeholders across geographic boundaries, industrial sectors, and academic disciplines. Digital collaborations, especially those leveraging digital technologies such as AI, Blockchains, cloud computing, and data analytics, should track the social and environmental impacts of the human footprint on Earth, and act upon such impacts. Acting upon the understanding of human impacts, it is then possible to create a future that is not only carbon neutral but also sustainable for the planet and its inhabitants. In conclusion, as many aspects of the planet and humans have become datafied, digitized, and networked, carbon neutrality must take on a central role in guiding our conscious green digital transformation of many political, economic, social and psychological aspects of our societies according to the existing and emerging social science knowledge.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

Conceptualization, YZ, H-TL, and C-LP; methodology, H-TL and C-LP; software, H-TL and C-LP; data curation, YZ, H-TL, and C-LP; writing—original draft preparation, H-TL, C-LP, and YZ; writing—review and editing, YZ, H-TL, and C-LP; visualization, H-TL, YZ, and C-LP; funding acquisition, YZ. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This research was funded mainly by the Department of Education of Guangdong Province for the project "Enterprise energy privilege trading scheme in Guangdong-Hong Kong-Macao Greater Bay Area" (Grant Number 2020WTSCX138), partly by the Department of Education of Guangdong Province in the project called "Smart App Design Innovation Research in the Age of New Business, Arts and Engineering Disciplines" (Grant Number 2019GXJK186), and partly by Guangzhou Nanfang College for the curriculum project of "Information Visualization Design" (Grant Number NFU 02-40250).

## ACKNOWLEDGMENTS

The authors wish to express their gratitude to all colleagues who directly or indirectly contributed to the making of the VOSviewer, Bibliometrix, and Citespace software.

## REFERENCES

Altieri, K. E., Trollip, H., Caetano, T., Hughes, A., Merven, B., and Winkler, H. (2016). Achieving Development and Mitigation Objectives through a Decarbonization Development Pathway in South Africa. *Clim. Pol.* 16, S78–S91. doi:10.1080/14693062.2016.1150250

Andrews, C. J. (2008). Greenhouse Gas Emissions along the Rural-Urban Gradient. *J. Environ. Plann. Manage.* 51, 847–870. doi:10.1080/09640560802423780

Ang, B. W. (2004). Decomposition Analysis for Policymaking in Energy:. *Energy Policy* 32, 1131–1139. doi:10.1016/S0301-4215(03)00076-4

Aria, M., and Cuccurullo, C. (2017). Bibliometrix : An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetrics* 11, 959–975. doi:10.1016/j.joi.2017.08.007

Ascui, F., and Lovell, H. (2011). As Frames Collide: Making Sense of Carbon Accounting. *Acc. Auditing Acc. J* 24, 978–999. doi:10.1108/09513571111184724

Bataille, C., Waisman, H., Colombier, M., Segafredo, L., and Williams, J. (2016a). The Deep Decarbonization Pathways Project (DDPP): Insights and Emerging Issues. *Clim. Pol.* 16, S1–S6. doi:10.1080/14693062.2016.1179620

Bataille, C., Waisman, H., Colombier, M., Segafredo, L., Williams, J., and Jotzo, F. (2016b). The Need for National Deep Decarbonization Pathways for Effective Climate Policy. *Clim. Pol.* 16, S7–S26. doi:10.1080/14693062.2016.1173005

Benjaminsen, G. (2017). The Bricolage of REDD+ in Zanzibar: from Global Environmental Policy Framework to Community forest Management. *J. East. Afr. Stud.* 11, 506–525. doi:10.1080/17531055.2017.1357103

Böhringer, C., Müller, A., and Schneider, J. (2015). Carbon Tariffs Revisited. *J. Assoc. Environ. Res. Econom.* 2, 629–672. doi:10.1086/683607

Boehmer-Christiansen, S. (2003). Science, Equity, and the War against Carbon. *Sci. Technol. Hum. Values* 28, 69–92. doi:10.1177/0162243902238496

Bowen, F., and Wittneben, B. (2011). Carbon Accounting. *Acc. Auditing Acc. J* 24, 1022–1036. doi:10.1108/09513571111184742

Boyack, K. W., and Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling, and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? *J. Am. Soc. Inf. Sci.* 61, 2389–2404. doi:10.1002/asi.21419

Bressler, R. D. (2021). The Mortality Cost of Carbon. *Nat. Commun.* 12, 4467. doi:10.1038/s41467-021-24487-w

Bulkeley, H., and Betsill, M. M. (2003). *Cities and Climate Change: Urban Sustainability and Global Environmental Governance*. London, New York: Routledge.

Bumpus, A. G., and Liverman, D. M. (2008). Accumulation by Decarbonization and the Governance of Carbon Offsets. *Econ. Geogr.* 84, 127–155. doi:10.1111/j.1944-8287.2008.tb00401.x

Byers, E. A., Hall, J. W., and Amezaga, J. M. (2014). Electricity Generation and Cooling Water Use: UK Pathways to 2050. *Glob. Environ. Change* 25, 16–30. doi:10.1016/j.gloenvcha.2014.01.005

Cai, Y., Ramis Ferrer, B., and Luis Martinez Lastra, J. (2019). Building University-Industry Co-innovation Networks in Transnational Innovation Ecosystems: Towards a Transdisciplinary Approach of Integrating Social Sciences and Artificial Intelligence. *Sustainability* 11, 4633. doi:10.3390/su11174633

Chapman, R., and Boston, J. (2007). The Social Implications of Decarbonising the New Zealand Economy. *Soc. Pol. J. N. Z.*, 104–136.

Chen, K.-S., and Liao, H.-T. (2020). "Mapping the Field of Educational Assessment and Professional Accreditation in Higher Education: Bibliographic Coupling and Co-Citation Analyses for Green and Digital Transformation," in 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID, 18-20 Dec. 2020, Guangzhou, China), New York, NY, USA (IEEE), 281–284. doi:10.1109/MSIEID52046.2020.00058

Chen, C. (2004). Searching for Intellectual Turning Points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci.* 101, 5303–5310. doi:10.1073/pnas.0307513100

Chen, C. (2016). *CiteSpace: A Practical Guide for Mapping Scientific Literature*. New York: Nova Science Publishers, Inc.

Clarivate Analytics (2020). Research Areas (Categories/Classification). Web of Science Core Collection Help. Available at: http://images.webofknowledge.com//WOKRS535R111/help/WOS/hp_research_areas_easca.html (Accessed February 1, 2021).

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., and Herrera, F. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. *J. Am. Soc. Inf. Sci.* 62, 1382–1402. doi:10.1002/asi.21525

Cui, T., and Zhang, J. (2018). Bibliometric and Review of the Research on Circular Economy through the Evolution of Chinese Public Policy. *Scientometrics* 116, 1013–1037. doi:10.1007/s11192-018-2782-y

Diniz, E. H., Yamaguchi, J. A., Rachael dos Santos, T., Pereira de Carvalho, A., Alégo, A. S., and Carvalho, M. (2021). Greening Inventories: Blockchain to Improve the GHG Protocol Program in Scope 2. *J. Clean. Prod.* 291, 125900. doi:10.1016/j.jclepro.2021.125900

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., and Lim, W. M. (2021). How to Conduct a Bibliometric Analysis: An Overview and Guidelines. *J. Bus. Res.* 133, 285–296. doi:10.1016/j.jbusres.2021.04.070

Downie, J., and Stubbs, W. (2013). Evaluation of Australian Companies' Scope 3 Greenhouse Gas Emissions Assessments. *J. Clean. Prod.* 56, 156–163. doi:10.1016/j.jclepro.2011.09.010

Dury, P. (2008). The Rise of Carbon Neutral and Compensation Carbone. *Term* 14, 230–248. doi:10.1075/term.14.2.06dur

Elisha, I. L., and Viljoen, A. (2021). Trends in Rooibos Tea (Aspalathus Linearis) Research (1994-2018): A Scientometric Assessment. *South Afr. J. Bot.* 137, 159–170. doi:10.1016/j.sajb.2020.10.004

European Commission (2020). A New Industrial Strategy for a Green and Digital Europe. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_416 (Accessed June 28, 2021).

Farnworth, E. (2018). *We CAN Build a Carbon-Neutral World by 2050. Here's How*. World Economic Forum. Available at: https://www.weforum.org/agenda/2018/06/the-world-will-be-carbon-neutral-by-2050-but-at-what-cost/.

Fragkos, P., Tasios, N., Paroussos, L., Capros, P., and Tsani, S. (2017). Energy System Impacts and Policy Implications of the European Intended Nationally Determined Contribution and Low-Carbon Pathway to 2050. *Energy Policy* 100, 216–226. doi:10.1016/j.enpol.2016.10.023

Freeman, L. C. (1978). Centrality in Social Networks Conceptual Clarification. *Soc. Networks* 1, 215–239. doi:10.1016/0378-8733(78)90021-7

Fridell, M., Edwin, S., von Schreeb, J., and Saulnier, D. D. (2019). Health System Resilience: What Are We Talking about? A Scoping Review Mapping Characteristics and Keywords. *Int. J. Health Pol. Manag.* 9, 6–16. doi:10.15171/ijhpm.2019.71

Garfield, E. (1994). Research Fronts. Current Comments. Available at: https://clarivate.com/webofsciencegroup/essays/research-fronts/ (Accessed January 27, 2020)..

Garfield, E. (2004). Historiographic Mapping of Knowledge Domains Literature. *J. Inf. Sci.* 30, 119–145. doi:10.1177/0165551504042802

Geels, F. W., Sovacool, B. K., Schwanen, T., and Sorrell, S. (2017). Sociotechnical Transitions for Deep Decarbonization. *Science* 357, 1242–1244. doi:10.1126/science.aao3760

Geels, F. W. (2005). The Dynamics of Transitions in Socio-Technical Systems: A Multi-Level Analysis of the Transition Pathway from Horse-Drawn Carriages to Automobiles (1860-1930). *Technol. Anal. Strateg. Manage.* 17, 445–476. doi:10.1080/09537320500357319

Geels, F. W. (2010). Ontologies, Socio-Technical Transitions (To Sustainability), and the Multi-Level Perspective. *Res. Pol.* 39, 495–510. doi:10.1016/j.respol.2010.01.022

Geels, F. W. (2018). Socio-Technical Transitions to Sustainability. *Oxford Res. Encycl. Environ. Sci.* doi:10.1093/acrefore/9780199389414.013.587

Geels, F. W. (2019). Socio-technical Transitions to Sustainability: a Review of Criticisms and Elaborations of the Multi-Level Perspective. *Curr. Opin. Environ. Sustainability* 39, 187–201. doi:10.1016/j.cosust.2019.06.009

Gibassier, D., and Schaltegger, S. (2015). Carbon Management Accounting and Reporting in Practice. *Sustainability* 6, 340–365. doi:10.1108/SAMPJ-02-2015-0014

Gilley, B. (2017). Local Governance Pathways to Decarbonization in China and India. *China Q.* 231, 728–748. doi:10.1017/S0305741017000893

Gössling, S., Bredberg, M., Randow, A., Sandström, E., and Svensson, P. (2006). Tourist Perceptions of Climate Change: A Study of International Tourists in Zanzibar. *Curr. Issues Tourism* 9, 419–435. doi:10.2167/cit265.0

Gössling, S. (2009). Carbon Neutral Destinations: a Conceptual Analysis. *J. Sustain. Tourism* 17, 17–37. doi:10.1080/09669580802276018

Goldemberg, J., Coelho, S. T., and Guardabassi, P. (2008). The Sustainability of Ethanol Production from Sugarcane. *Energy Policy* 36, 2086–2097. doi:10.1016/j.enpol.2008.02.028

Green, J. F. (2013). Order Out of Chaos: Public and Private Rules for Managing Carbon. *Glob. Environ. Polit.* 13, 1, 25. doi:10.1162/GLEP_a_00164

Grübler, A., Nakićenović, N., and Victor, D. G. (1999). Dynamics of Energy Technologies and Global Change. *Energy Policy* 27, 247–280. doi:10.1016/S0301-4215(98)00067-6

Hansen, D. L., Schneiderman, B., and Smith, M. A. (2020). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Second edition. Amsterdam: Morgan Kaufmann.

Harvey, F. (2020). *China Pledges to Become Carbon Neutral before 2060*. The Guardian. Available at: https://www.theguardian.com/environment/2020/sep/22/china-pledges-to-reach-carbon-neutrality-before-2060.

Hawken, P. (2017). *Drawdown: The Most Comprehensive Plan Ever Proposed to Reverse Global Warming*. New York, New York: Penguin Books.

Helm, D. (2010). Government Failure, Rent-Seeking, and Capture: the Design of Climate Change Policy. *Oxford Rev. Econ. Pol.* 26, 182–196. doi:10.1093/oxrep/grq006

Higham, J. E. S., and Cohen, S. A. (2011). Canary in the Coalmine: Norwegian Attitudes towards Climate Change and Extreme Long-Haul Air Travel to Aotearoa/New Zealand. *Tourism Manage.* 32, 98–105. doi:10.1016/j.tourman.2010.04.005

Höhne, N., Kuramochi, T., Warnecke, C., Röser, F., Fekete, H., Hagemann, M., et al. (2017). The Paris Agreement: Resolving the Inconsistency between Global Goals and National Contributions. *Clim. Pol.* 17, 16–32. doi:10.1080/14693062.2016.1218320

Hopke, J. E., and Hestres, L. E. (2018). Visualizing the Paris Climate Talks on Twitter: Media and Climate Stakeholder Visual Social Media during COP21. *Soc. Media + Soc.* 4, 205630511878268. doi:10.1177/2056305118782687

Hübler, M., and Löschel, A. (2013). The EU Decarbonisation Roadmap 2050-What Way to Walk? *Energy Policy* 55, 190–207. doi:10.1016/j.enpol.2012.11.054

IPCC (2018). "Global Warming of 1.5 °C, an IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways," in *The Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*. Editors V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, et al. (Switzerland: IPCC). Available at: http://ipcc.ch/report/sr15/.

Jarneving, B. (2007). Bibliographic Coupling and its Application to Research-Front and Other Core Documents. *J. Informetrics* 1, 287–307. doi:10.1016/j.joi.2007.07.004

Kibwami, N., and Tutesigensi, A. (2016). Enhancing Sustainable Construction in the Building Sector in Uganda. *Habitat Int.* 57, 64–73. doi:10.1016/j.habitatint.2016.06.011

Köhler, J., Geels, F. W., Kern, F., Markard, J., Onsongo, E., Wieczorek, A., et al. (2019). An Agenda for Sustainability Transitions Research: State of the Art and Future Directions. *Environ. Innovation Soc. Trans.* 31, 1–32. doi:10.1016/j.eist.2019.01.004

Kolk, A., Levy, D., and Pinkse, J. (2008). Corporate Responses in an Emerging Climate Regime: The Institutionalization and Commensuration of Carbon Disclosure. *Eur. Account. Rev.* 17, 719–745. doi:10.1080/09638180802489121

Koteyko, N. (2012). Managing Carbon Emissions: A Discursive Presentation of 'market-Driven Sustainability' in the British media. *Lang. Commun.* 32, 24–35. doi:10.1016/j.langcom.2011.11.001

Lilliestam, J., Bielicki, J. M., and Patt, A. G. (2012). Comparing Carbon Capture and Storage (CCS) with Concentrating Solar Power (CSP): Potentials, Costs, Risks, and Barriers. *Energy Policy* 47, 447–455. doi:10.1016/j.enpol.2012.05.020

Lohmann, L. (2009). Toward a Different Debate in Environmental Accounting: The Cases of Carbon and Cost-Benefit. *Account. Organ. Soc.* 34, 499–534. doi:10.1016/j.aos.2008.03.002

Loulou, R., Goldstein, G., and Noble, K. (2004). Documentation for the MARKAL Family of Models. *Energy Technol. Syst. Anal. Program.*, 65–73.

MacDonald, K. I., and Dressler, V. (2018). Using Citation Analysis to Identify Research Fronts: A Case Study with the Internet of Things. *Sci. Technol. Libraries* 37, 171–186. doi:10.1080/0194262X.2017.1415183

Mallapaty, S. (2020). How China Could Be Carbon Neutral by Mid-Century. *Nature* 586, 482–483. doi:10.1038/d41586-020-02927-9

Marcucci, A., and Fragkos, P. (2015). Drivers of Regional Decarbonization through 2100: A Multi-Model Decomposition Analysis. *Energ. Econ.* 51, 111–124. doi:10.1016/j.eneco.2015.06.009

Marres, N. (2011). The Costs of Public Involvement: Everyday Devices of Carbon Accounting and the Materialization of Participation. *Econ. Soc.* 40, 510–533. doi:10.1080/03085147.2011.602294

Mathy, S., Criqui, P., Knoop, K., Fischedick, M., and Samadi, S. (2016). Uncertainty Management and the Dynamic Adjustment of Deep Decarbonization Pathways. *Clim. Pol.* 16, S47–S62. doi:10.1080/14693062.2016.1179618

McGlade, C., and Ekins, P. (2014). Un-burnable Oil: An Examination of Oil Resource Utilisation in a Decarbonised Energy System. *Energy Policy* 64, 102–112. doi:10.1016/j.enpol.2013.09.042

Meseguer-Sánchez, V., Gálvez-Sánchez, F. J., Molina-Moreno, V., and Wandosell-Fernández-de-Bobadilla, G. (2021). The Main Research Characteristics of the Development of the Concept of the Circular Economy Concept: A Global Analysis and the Future Agenda. *Front. Environ. Sci.* 9, 304. doi:10.3389/fenvs.2021.704387

METI (2020). *"Green Growth Strategy through Achieving Carbon Neutrality in 2050" Formulated*. Japan: Ministry of Economy, Trade and Industry (METI). Available at: https://www.meti.go.jp/english/press/2020/1225_001.html.

Milne, M. J., and Grubnic, S. (2011). Climate Change Accounting Research: Keeping it Interesting and Different. *Acc. Auditing Account. J* 24, 948–977. doi:10.1108/09513571111184715

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G.The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *Plos Med.* 6, e1000097. doi:10.1371/journal.pmed.1000097

N. Stern (Editor) (2007). *The Economics of Climate Change: The Stern Review* (Cambridge, UK New York: Cambridge University Press).

Nakićenović, N. (1996). Decarbonization: Doing More with Less. *Technol. Forecast. Soc. Change* 51, 1–17. doi:10.1016/0040-1625(95)00167-010.1016/s0040-1625(96)90029-3

Onnela, J.-P., Saramäki, J., Kertész, J., and Kaski, K. (2005). Intensity and Coherence of Motifs in Weighted Complex Networks. *Phys. Rev. E* 71, 065103. doi:10.1103/PhysRevE.71.065103

Oshiro, K., Kainuma, M., and Masui, T. (2016). Assessing Decarbonization Pathways and Their Implications for Energy Security Policies in Japan. *Clim. Pol.* 16, S63–S77. doi:10.1080/14693062.2016.1155042

Pallesen, T. (2016). Valuation Struggles over Pricing - Determining the worth of Wind Power. *J. Cult. Econ.* 9, 527–540. doi:10.1080/17530350.2016.1212084

Payne, C. (2020). The Decarbonisation Divide. *Nat. Hum. Behav.* 4, 236. doi:10.1038/s41562-020-0843-8

Persson, O. (1994). The Intellectual Base and Research Fronts ofJASIS 1986-1990. *J. Am. Soc. Inf. Sci.* 45, 31–38. doi:10.1002/(sici)1097-4571(199401)45:1<31:aid-asi4>3.0.co;2-g

Peters, G. P., Andrew, R. M., Solomon, S., and Friedlingstein, P. (2015). Measuring a Fair and Ambitious Climate Agreement Using Cumulative Emissions. *Environ. Res. Lett.* 10, 105004. doi:10.1088/1748-9326/10/10/105004

Pidgeon, N., and Demski, C. C. (2012). From Nuclear to Renewable: Energy System Transformation and Public Attitudes. *Bull. At. Sci.* 68, 41–51. doi:10.1177/0096340212451592

Pigliautile, I., Castaldo, V. L., Makaremi, N., Pisello, A. L., Cabeza, L. F., and Cotana, F. (2019). On an Innovative Approach for Microclimate Enhancement and Retrofit of Historic Buildings and Artworks Preservation by Means of Innovative Thin Envelope Materials. *J. Cult. Heritage* 36, 222–231. doi:10.1016/j.culher.2018.04.017

Pihl, E., Alfredsson, E., Bengtsson, M., Bowen, K. J., Cástan Broto, V., Chou, K. T., et al. (2021). Ten New Insights in Climate Science 2020 - a Horizon Scan. *Glob. Sustain.* 4, e5. doi:10.1017/sus.2021.2

Pye, S., and Bataille, C. (2016). Improving Deep Decarbonization Modelling Capacity for Developed and Developing Country Contexts. *Clim. Pol.* 16, S27–S46. doi:10.1080/14693062.2016.1173004

Pye, S., McGlade, C., Bataille, C., Anandarajah, G., Denis-Ryan, A., and Potashnikov, V. (2016). Exploring National Decarbonization Pathways and Global Energy Trade Flows: a Multi-Scale Analysis. *Clim. Pol.* 16, S92–S109. doi:10.1080/14693062.2016.1179619

Quimby, C. C., and Angelique, H. (2011). Identifying Barriers and Catalysts to Fostering Pro-environmental Behavior: Opportunities and Challenges for Community Psychology. *Am. J. Commun. Psychol.* 47, 388–396. doi:10.1007/s10464-010-9389-7

Rogelj, J., den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., et al. (2016). Paris Agreement Climate Proposals Need a Boost to Keep Warming Well below 2 °C. *Nature* 534, 631–639. doi:10.1038/nature18307

Schwartz, J. (2021). A Carbon Calculation: How Many Deaths Do Emissions Cause? *the New York Times*. Available at: https://www.nytimes.com/2021/07/29/climate/carbon-emissions-death.html (Accessed August 1, 2021).

Scott, D., Hall, C. M., and Gössling, S. (2019). Global Tourism Vulnerability to Climate Change. *Ann. Tourism Res.* 77, 49–61. doi:10.1016/j.annals.2019.05.007

Small, H. (1999). Visualizing Science by Citation Mapping. *J. Am. Soc. Inf. Sci.* 50, 799–813. doi:10.1002/(sici)1097-4571(1999)50:9<799:aid-asi9>3.0.co;2-g

Smith, C. B., and Fletcher, W. D. (2020). *Reaching Net Zero: What it Takes to Solve the Global Climate Crisis*. 1st ed. Cambridge: Elsevier.

Sovacool, B. K. (2017). Contestation, Contingency, and justice in the Nordic Low-Carbon Energy Transition. *Energy Policy* 102, 569–582. doi:10.1016/j.enpol.2016.12.045

Stadel, A., Eboli, J., Ryberg, A., Mitchell, J., and Spatari, S. (2011). Intelligent Sustainable Design: Integration of Carbon Accounting and Building Information Modeling. *J. Prof. Issues Eng. Educ. Pract.* 137, 51–54. doi:10.1061/(ASCE)EI.1943-5541.0000053

Steckel, J. C., Jakob, M., Marschinski, R., and Luderer, G. (2011). From Carbonization to Decarbonization?-Past Trends and Future Scenarios for China's CO2 Emissions. *Energy Policy* 39, 3443–3455. doi:10.1016/j.enpol.2011.03.042

Strachan, N., and Kannan, R. (2008). Hybrid Modelling of Long-Term Carbon Reduction Scenarios for the UK. *Energ. Econ.* 30, 2947–2963. doi:10.1016/j.eneco.2008.04.009

Strachan, N., Pye, S., and Kannan, R. (2009). The Iterative Contribution and Relevance of Modelling to UK Energy Policy. *Energy Policy* 37, 850–860. doi:10.1016/j.enpol.2008.09.096

Tapio, P., Banister, D., Luukkanen, J., Vehmas, J., and Willamo, R. (2007). Energy and Transport in Comparison: Immaterialisation, Dematerialisation and Decarbonisation in the EU15 between 1970 and 2000. *Energy Policy* 35, 433–451. doi:10.1016/j.enpol.2005.11.031

Twidale, S. (2021). Global Initiative Aims to Bring Transparency to Voluntary Carbon Markets. Reuters. Available at: https://www.reuters.com/business/environment/global-initiative-aims-bring-transparency-voluntary-carbon-markets-2021-07-29/(Accessed August 1, 2021).

Unruh, G. C. (2000). Understanding Carbon Lock-In. *Energy Policy* 28, 817–830. doi:10.1016/S0301-4215(00)00070-7

van der Ven, H., Bernstein, S., and Hoffmann, M. (2017). Valuing the Contributions of Nonstate and Subnational Actors to Climate Governance. *Glob. Environ. Polit.* 17, 1–20. doi:10.1162/GLEP_a_00387

van Eck, N. J., and Waltman, L. (2014). "Visualizing Bibliometric Networks," in *In* Measuring Scholarly Impact. Editors Y. Ding, R. Rousseau, and D. Wolfram (Cham: Springer International Publishing), 285–320. doi:10.1007/978-3-319-10377-8_13

Vandenbergh, M., and Steinemann, A. (2007). The Carbon-Neutral Individual. *N. Y. Univ. L. Rev.* 82, 1673–1745.

Vesty, G. M., Telgenkamp, A., and Roscoe, P. J. (2015). Creating Numbers: Carbon and Capital Investment. *Account. Auditing Account. J.* 28, 302–324. doi:10.1108/AAAJ-10-2013-1507

Waheed, B., Khan, F., Veitch, B., and Hawboldt, K. (2011). An Integrated Decision-Making Framework for Sustainability Assessment: A Case Study of Memorial University. *High Educ. Pol.* 24, 481–498. doi:10.1057/hep.2011.17

Wang, Z., Liao, H.-T., Zhou, Y., and Pang, S. (2020). "Toward the Design of Socio-Technical Systems for Sustainability: A Scientometric Analysis," in Proceedings of the 6th International Conference on Humanities and Social Science Research (ICHSSR2020), Hangzhou, China, 2020.04.11, Dordrecht (Atlantis Press). doi:10.2991/assehr.k.200428.095

Wang, R.-Y., Lin, P.-a., Chu, J.-Y., Tao, Y.-H., and Ling, H.-C. (2021). A Decision Support System for Taiwan's forest Resource Management Using Remote Sensing Big Data. *Enterprise Inf. Syst.*, 1–22. doi:10.1080/17517575.2021.1883123

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *J. Stat. Soft.* 40, 1–29. doi:10.18637/jss.v040.i01

Yigitcanlar, T., and Lee, S. H. (2014). Korean Ubiquitous-Eco-City: A Smart-Sustainable Urban Form or a Branding Hoax? *Technol. Forecast. Soc. Change* 89, 100–114. doi:10.1016/j.techfore.2013.08.034

Yu, Y., Li, Y., Zhang, Z., Gu, Z., Zhong, H., Zha, Q., et al. (2020). A Bibliometric Analysis Using VOSviewer of Publications on COVID-19. *Ann. Transl Med.* 8, 816. doi:10.21037/atm-20-4235

Zuo, J., Read, B., Pullen, S., and Shi, Q. (2012). Achieving Carbon Neutrality in Commercial Building Developments - Perceptions of the Construction Industry. *Habitat Int.* 36, 278–286. doi:10.1016/j.habitatint.2011.10.010

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Zhang, Pan and Liao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 21 October 2021
doi: 10.3389/fenvs.2021.724095

# A Counterfactual Baseline for Assessing Future Environmental Impact: A Case Study of the Belt and Road Initiative

*Jing Victor Li*[1], *Tsun Se Cheong*[2] *and Xunpeng Shi*[3]*

[1]Department of Geography and Resource Management and Institute of Future Cities, The Chinese University of Hong Kong, Shatin, China, [2]Department of Economics and Finance, Hang Seng University of Hong Kong, Hong Kong, China, [3]Australia-China Relations Institute, University of Technology Sydney, Sydney, NSW, Australia

The impact of the Belt and Road Initiative (BRI) on the environment is an important but controversial topic. But assessing it faces a significant challenge for separating its policy impact from the overall effect of economic development that will also have environmental impacts. This paper attempts to provide an evolutionary and counterfactual baseline to evaluate the environmental impact of BRI, based on the distribution dynamics approach and the mobility probability plots (MPPs). Our estimation results suggest that while the majority of BRI economies will lower their emission levels compared with the world average, there are significant differences between BRI and non-BRI countrie's emission patterns and dynamics. The majority of non-BRI economies will have lower emission levels than their BRI counterparts in the absence of BRI policy, indicating that the difference in future emissions between BRI and non-BRI countries should not be completely attributed to the environmental impact of BRI. Instead, BRI should take measures to prevent certain countries from moving upwards energy intensity paths through policy intervention, international cooperation, and an inclusive project assessment process.

**Keywords: belt and road initiative, emission convergence, distribution dynamics, mobility probability plots, counterfactual baseline**

OPEN ACCESS

***Edited by:***
*Cosimo Magazzino,*
*Roma Tre University, Italy*

***Reviewed by:***
*Yaoping Wang,*
*The Ohio State University, United States*
*Abdul Rauf,*
*Nanjing University of Information Science and Technology, China*
*W Selmier,*
*Indiana University, United States*

***Correspondence:**
*Xunpeng Shi*
*xunpeng.shi@gmail.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *11 June 2021*
***Accepted:*** *08 October 2021*
***Published:*** *21 October 2021*

***Citation:***
*Li JV, Cheong TS and Shi X (2021) A Counterfactual Baseline for Assessing Future Environmental Impact: A Case Study of the Belt and Road Initiative. Front. Environ. Sci. 9:724095. doi: 10.3389/fenvs.2021.724095*

## INTRODUCTION

The impact of BRI on the environment has been a controversial topic. The international community is concerned that the BRI may worsen the environmental performance of participating countries due to the relocation of China's "outdated industries" to these countries (Han et al., 2018), the ecological impact of infrastructure building, emissions from productive activities, and lock into fossil-fuel dependency (Ascensão et al., 2018; Magazzino and Mele, 2020). China's experience in environmental degradation seems to justify concerns (Yang et al., 2021). Such controversy may discourage the engagement of certain BRI countries which may otherwise benefit from the initiative. Nevertheless, the negative outlook of BRI could be turned into opportunities (Ascensão et al., 2018), if its environmental impact is adequately and fairly assessed.

One challenge to evaluate the environmental impact of the BRI is to separate its impact from the overall impact of development. Economic development will naturally lead to economic growth, energy consumption, foreign direct investment (FDI), financial development and urbanization,

which will pollute the environment (Baloch et al., 2019; Udemba et al., 2020). BRI should not be blamed for the negative impact, which is a natural consequence of economic development. However, the estimation of the environmental effect of BRI is challenged by two factors. For one thing, the BRI is a new initiative and thus empirical data are not available. For another, even if BRI has empirical data, it is difficult to establish a counterfactual baseline without BRI against which to estimate the impact of BRI.

The prevailing criticism of BRI based on industry relocation highlights the importance of a counterfactual baseline in the debates on BRI's environmental consequences. The international community often worry that China's relocation of its "excess capacity," considered "outdated" in China and usually energy-intensive, would pollute the environment in the receiving countries (Han et al., 2018; Dong, 2017). The relocation of "outdated" industries, however, is neither new nor unique to China's BRI. One perspective suggests that the East Asian Miracle was created due to the relocation of Japan's "outdated" to Korea, Taiwan, Hong Kong and Singapore, and later, further to China (Lloyd, 1996). The relocation creates win-win outcomes for both the outsourcing countries and the receiving countries as an "outdated" industry in Japan was still very advanced in China by that time (Han et al., 2018). However, such views lack an evolutionary dynamics perception: what was advanced for China in the 1980s and into the 1990s was not so by the 2000s. And what underpins the East Asian Miracle is more profound at the institutional levels (Haggard, 1999). Without access to the relocated industries and associated technologies, China will have to rely on less advanced technologies to engine its economic development.

The different perceptions of industry relocation depend on the baseline case. If the baseline is the pre-development stage, the relocation will generate more pollution. However, if compared with realistic alternatives during the development stage, the relocation can be desirable. Since the alternatives in the development stage are seldom projected, it is very difficult to validate the argument. Validation of the controversies related to the environmental impact of BRI will help both China and the BRI participating countries to understand the potential adverse impacts to a wider scale and prepare measures to prevent or mitigate the adverse impacts. For instance, debates are arising about whether BRI indicates China's ambition to reassess the nature of power and influence the contemporary international system (Beeson, 2018), such as discourses stemming from the global debt incurred through BRI (Hurley et al., 2019) and the heterogeneous and contrasting effects of BRI on participating countries in the global value chains (Wu et al., 2020).

The existing literature on the environmental impact of BRI mainly focuses on the hypothetical problems that have not been demonstrated by empirical studies due to the nascent status of BRI. The impact of ecological systems is justified based on factors such as infrastructure construction, fossil fuel development, energy consumption, and material use (Ascensão et al., 2018), as well as water consumption and embedded water trade (Zhang et al., 2018). The quantitative studies of the potential environmental impact of BRI are based on reasonable assumptions and hypotheses (Han et al., 2018; Qiu et al., 2018; Dong et al., 2015). Although efforts have been made to prepare a baseline for evaluating the impact of BRI in the future (Liu et al., 2018), it is essential to establish a dynamic baseline to assess BRI's environmental impact, given the importance of understanding the environmental impact of BRI.

This paper attempts to provide a counterfactual baseline to evaluate the environmental impact of BRI. We use $CO_2$ emissions as the proxy of environmental performance. We then adopt the distribution dynamics approach to examine the trend of carbon emission globally using their contour maps of transition probability kernels, and their ergodic distributions and mobility probability plots (MPP). By separating the sample countries into BRI and non-BRI countries, we provide detailed information on their carbon emission's current pattern and future development.

Our contributions to the debates on BRI literature are two folds. First, this paper is the first quantitative study that can address the future impact of BRI on the environment. The observed systematic difference between BRI and non-BRI countries provides useful information for future policy decisions and a baseline for evaluating the future impact of BRI. Second, the findings of this paper will help developing countries involved in BRI call for development power and space while setting more objective and realistic goals of emission reduction so that they can achieve adequate growth opportunities at lower environmental costs.

The remainder of this paper is structured as follows. *Literature Review* reviews the literature on the environmental impact of BRI. *Data and Methodology* describes the data source and research methodology. *Results and Discussion* reports and discusses the results with various alternative GDP measurements. *Conclusion* concludes the paper with policy implications.

## LITERATURE REVIEW

The Belt and Road Initiative aims to "reconfigure China's external sector in order to continue its strong growth" (Huang, 2016). While there is a close and multi-facet relationship between the SDGs and the five priority areas of the Belt and Road Initiative (Hong, 2016; Yin, 2019), questions about the potential impacts of BRI on accelerating energy consumption and increasing the global carbon emissions are arising (Zhang et al., 2017; Huang, 2019). For instance, the carbon emissions of BRI countries significantly increased together with their GDP growth (Fan et al., 2019; Han et al., 2020). Li et al. (2018) maintain that BRI countries lack an effective driving mechanism to promote green energy and low-carbon development in relation to this grand initiative. Huang and Li (2020) demonstrate that while BRI has spatial spillover effects on the green development of participating countries, it cannot break the spatial club imbalance.

However, some others argue that BRI can contribute to carbon emission abatement through green technology (Chen et al., 2020). Moreover, Liu et al. (2020) conclude that Chinese OFDIs tend to invest in green projects where "the host countries had better

political environments, natural resource endowments and higher energy efficiencies." These OFDIs contributed to increased green total factor productivity of BRI countries (Wu et al., 2020). Coenen et al. (2021) further find that China is actively and rapidly developing an institutional architecture of green BRI. However, it remains unclear if the claim of green BRI will eventually lead to green development or greenwashing (Harlan, 2021). Zhao et al. (2021) find that the productivity growth effect does not contribute significantly to green growth in BRI countries, but the potential output change effect does. Rauf et al. (2018a) demonstrate the existence of EKC approximately in every region.

Some other studies focus on the environmental impact of infrastructure building and argue that BRI could adversely affect the environment through habitat loss, the overexploitation of resources and the degradation of the surrounding landscape (Ascensão et al., 2018; Li and Shvarts, 2017) and the destruction of the traditional way of life, and the fragility and vulnerability of many ecosystems (Dong et al., 2017). A few studies suggest that increasing productive activities due to BRI could cause environmental degradation. For example, the production of cement that is used for road construction is expected to add emission to China's one-third of the GHG emissions (Ascensão et al., 2018). Intensive human activities due to BRI is also considered to aggravate the water crisis in Central Asia, deteriorate the vulnerable environment, and accelerate energy consumption in this area (Li et al., 2015). Rauf et al. (2018b; 2020) find that trade, investment and economic growth have different impacts on BRI countrie's environmental performance. In contrast, one study finds that after the BRI was implemented, China's net export growth concentrated in the pollution-intensive and resource-intensive industries, indicating that China has not shifted environmental polluting industries to the BRI countries (Tian et al., 2019).

The environmental impact of infrastructure building due to BRI is a hot topic. Hughes (2019) predicted biodiversity hotspots in BRI's proposed road and rail routes on key biodiversity areas and suggested that BRI may pose a significant risk due to construction and development along the route and mining. Teo et al. (2019) shows how the infrastructure's environmental impacts interact and aggregate across multiple spatiotemporal scales and describes how those environmental impacts are influenced by economic and socio-political drivers. Yao et al. (2019) discuss how BRI can bring investment and technology to assist electricity market integration in the Southeast Asian region. Shi and Yao (2019) investigates the problems of China's investment into the ASEAN energy sector and suggests that building a sense of ownership within host countries is essential for the success of China's outward energy infrastructure investment. Selmier (2018) provides interesting ties into infrastructure, energy and BRI which could be used to develop environmental themes and are generally supportive of BRI.

The limited empirical studies cannot test the BRI impact due to its nascent status. Empirical studies which use the year 2013 and the starting point of BRI, i.e., (Tian et al., 2019), are not convincing as the BRI action plan was only unveiled in 2015. While BRI-like development projects may be traced back to the early 1990s (Selmier, 2018), such project cooperation would not impact the government-to-government cooperation in BRI. It is difficult to argue that China's unilateral policy declaration in 2015 could have a significant impact on other countries in a few years. Therefore, most quantitative studies of the BRI's environmental impacts are based on reasonable assumptions and other hypothetical scenarios. Han et al. (2018) suggest that BRI can promote energy efficiency convergence by inferring from the positive roles of regional cooperation and trade integration promoted by BRI. While tourism is not energy-intensive, a study of the environmental impact of tourism in China's BRI provinces shows that tourism could also lead to environmental degradation (Ahmad et al., 2018). Assuming effective liner shipping vessel sharing is critical to building efficient maritime transport networks under BRI (2018), BRI can boost liner shipping companie's profits and reduce $CO_2$ emissions dramatically. Dong et al. (2015) propose main modes and paths to achieve sustainable economic development of BRI.

Various studies have shown positive views on avoiding the negative environmental impact of BRI. The BRI Ecological and Environmental Cooperation Plans aspire to achieve sustainable development. Dong et al. (2018) argue that the BRI is consistent with other major sustainable development initiatives and strengthening the linkage between BRI and 2030 sustainable development goals (SDGs) could enhance international cooperation and make up other global environmental governance deficits.

More recent studies find that the environmental impact of BRI is heterogeneous across various indicators. Using the panel dataset covering 178 countries for the period of 2002–2018, Y. Wu et al. (2021) find that the impacts of BRI on energy efficiency depend on energy efficiency quantiles, resource endowments, and income levels. Wu et al. (2021) find that more imports from China will reduce the per capita CO2 emissions of the BRI countries, but more exports to China will have the opposite impact.

Some scholars suggest using environment safeguard measures, such as Strategic Environmental and Social Assessments (SESAs); or other instruments, such as green finance and stakeholder consultation, to prevent damage to the natural environment (Ascensão et al., 2018). International and grassroots environmental pressures are also found to be effective in informing broader policy approaches to BRI implementation (Klinger, 2019). Others generally call for advancing scientific research, reinforcing international collaboration and enhancing education to build a sustainable Silk Road economic belt (Li et al., 2015). In particular, data collection for monitoring and evaluation is usually advocated. Although China has experienced environmental degradation during its development stage, she has successfully initiated many measures to tackle the environmental problems, such as improving environmental regulations and adopting new green technologies and higher environmental standards (Ascensão et al., 2018).

In the case of methodologies, recently, there have been increasing trends of revisiting traditional issues with big data methods, such as LASSO, machine learning and Artificial Neural

Networks (ANNs). Issues examined include economic growth and emissions (Magazzino et al., 2021b; Mele and Magazzino, 2020), economic convergence (Magazzino et al., 2021c), information technology and pollution (Magazzino et al., 2021a) and drivers of emissions (Hsu, 2015; Shi et al., 2020). The use of newspaper coverage frequency also generates new variables, such as China's provincial economic policy uncertainty index (Yu et al., 2021b).

There is a lack of a useful baseline to evaluate the environmental impact of BRI. The limited effort to prepare a baseline for future comparison is not useful at this moment. For example, Liu et al. (2018) monitored and analyzed the ecological environment status of the BRI in 2015 using remote Sensing technology as the basis for future policy decisions in eco-environmental evaluation and protection.

To better reveal the difference between BRI and non-BRI countries in the future environmental dynamics, we will adopt the distribution dynamics approach and Ergodic distributions and mobility probability plots. The distribution dynamics analysis was first proposed by (Quah, 1993a; 1993b) and updated continuously (Quah, 1996, 2006). The convergence analysis, mainly used in the economic growth and household consumption literature (Yu et al., 2021a), was recently extended to the energy and environmental literature. Examples include energy intensity convergence (Ezcurra, 2007; Le Pen and Sévi, 2010), carbon emission convergence (Apergis and Payne, 2017; Wu et al., 2016), and energy (electricity) consumption convergence (Fallahi and Voia, 2015; Payne et al., 2017; Cheong et al., 2019; Shi et al., 2021). There are also interactions between economics and energy issues, such as a recent study on the living standards and energy consumption convergence (Shi et al., 2021).

Distribution dynamics is a useful tool to analyze the evolutionary pattern of different sectors through building counterfactual baselines, such as examining the housing price of China (Li et al., 2017; Cheong et al., 2021) and the housing affordability of United States, Canada and Australia (Cheong and Li, 2018). As far as we know, there is not yet a particular study on building counterfactual baselines for carbon emissions in non-BRI countries. This gap makes our research more worthwhile.

We will further use the mobility probability plots (MPPs), a methodology developed by Cheong and Wu (2013) to improve the display tools of the stochastic kernel approach. (Cheong and Wu (2018) improved MPP for analyzing the mobility probability of the entities within the distribution. The MPP has been employed to analyze transitional dynamics in various research areas, such as industrial output (Cheong and Wu, 2018), energy markets (Wu et al., 2018; Cheong et al., 2019), carbon dioxide emissions (Cheong et al., 2016; Wu et al., 2016), and foreign direct investment (Cheong et al., 2019).

## DATA AND METHODOLOGY

The data of this study are compiled from the World Bank's *World Development Indicators* (WDI) database which provides global data on carbon dioxide emissions. In this research, three different measurements of carbon dioxide emissions are used so as to provide a comprehensive analysis on the underlying trend and future development of carbon emissions in the world, namely, kg per 2010 US$ of GDP, metric tons per capita, and kg per 2011 purchasing power parity (PPP)$ of GDP. The data cover the period from 1987 to 2013 for a total of 27 years, and all the countries in the WDI database have been included (except a few countries whose data are not available from the World Bank). Given that the dataset covers a very long period and a large number of countries, thereby making the distribution dynamics analysis fairly robust in this study.

Our data cover 178 countries and regions, including 71 BRI countries. The scope of BRI countries was defined as those that have signed BRI cooperation by September 2018. A full list is presented in the **Appendix**.

For each year, the world average of the three measurements is also obtained from the WDI database. Carbon emissions for each country are then divided by this year's world average to calculate the relative emissions. Therefore, all the values are transformed into a ratio, of which a value greater than one means that the emission is greater than the world average, while a value smaller than one suggests that the emission of this country is lower than the world average. However, it is worth noting that the average world data on kg per 2011 PPP$ of GDP are only available from 1990, so that this data series is a bit shorter than the other series.

The data of relative emission are then used for the distribution dynamics analysis. Distribution dynamics analysis is an important tool for studying the changes in distributions across time. Distribution dynamics analysis may prove better than traditional time series econometrics as the advantage of the former lies in providing information on the whole shape of the distribution instead of discrete forecasted values. By employing this technique, not only can we fill a gap in the literature but also complement existing findings derived from econometrics only.

Distribution dynamics analysis can be broadly divided into two different categories. One is the traditional Markov transition matrix analysis and the other is the stochastic kernel approach. Given that there is an issue of demarcation of the state which is associated with the selection of grid values for the former approach; and thus the analytical results are dependent on the selection of grid lines. As a result, it has received criticism. On the contrary, the latter is deemed better as it can circumvent the issue of demarcation. Hence the stochastic kernel approach is used in this study.

The bivariate kernel estimator used in this study is defined as:

$$\hat{f}(x,y) = \frac{1}{nh_1h_2}\sum_{i=1}^{n} K\left(\frac{x - X_{i,t}}{h_1}, \frac{y - X_{i,t+1}}{h_2}\right) \quad (1)$$

where $K$ is the normal density function, $n$ is the number of observations, $x$ is a variable representing the relative emission value of one of the three measurements of a country at time t, $y$ is a variable representing the relative emission value of one of the three measurements of that country at time t+1, $X_{i,t}$ is an observed value of relative emission value at time t, $X_{i,t+1}$ is the observed value of relative emission at time t+1, and $h_1$ and $h_2$ are the bandwidths which are calculated based on the procedure proposed by Silverman (2018).

An adaptive kernel with flexible bandwidth is employed (Silverman, 2018) to consider the sparseness of the data. Assuming that the process is first-order and time-invariant, and the distribution at time $t+\tau$ depends on $t$ only and not on any previous distributions, then the relationship between the distributions at time $t$ and time $t+\tau$ is shown as:

$$f_{t+\tau}(z) = \int_0^{\infty} g_\tau(z|x) f_t(x) dx \quad (2)$$

where $f_{t+\tau}(z)$ is the $\tau$-period-ahead density function of $z$ conditional on $x$, $g_\tau(z|x)$ is the transition probability kernel which maps the distribution from time $t$ to $t+\tau$, whilst $f_t(x)$ is the kernel density function of the distribution of the relative emission at time $t$.

The ergodic density function can then be calculated by:

$$f_\infty(z) = \int_0^{\infty} g_\tau(z|x) f_\infty(x) \mathrm{d}x \quad (3)$$

where $f_\infty(z)$ is the ergodic density function when $\tau$ is infinite. This is the final distribution in the long run. It is worth noting that distribution dynamics analysis assumes that the dynamics remain unchanged when one prepares the forecast by computing ergodic distribution. This practice is similar to those used in econometrics when one calculates the beta value based on information derived from previous data, and these computed beta values may change in the future. So similarly, as the ergodic distribution is also based on information derived from previous data, intervention from the government can alter the distribution and change the final shape of the distribution by implementing relevant policies in mitigating carbon emissions.

The shape of the ergodic distribution is dependent on the distribution dynamics computed from the dataset, therefore, one can gain comprehensive knowledge on the underlying trend and evolution by examining the ergodic distribution. Moreover, the ergodic distribution also provides a forecast into the future and one can know what would happen to the shape of the distribution if the effects are persistent in nature.

The mobility probability plot (MPP) is further used to improve the traditional display tools of the stochastic kernel approach which are mainly based on the three-dimensional plots and the contour maps. The MPP can be constructed by computing $p(x)$ which is the net upward mobility probability. It can be represented as:

$$p(x) = \int_x^{\infty} g_\tau(z|x) \mathrm{d}z - \int_0^x g_\tau(z|x) \mathrm{d}z \quad (4)$$

The MPP shows the net upward mobility probability against the relative emission of the three measurements. It is worth noting that a positive value implies that the country will have a net probability of moving upwards in the future; whereas a negative value of net upward mobility probability implies that the country has a net probability of moving downwards within the distribution. Interested readers can refer to Cheong and Wu (2018), Shi et al. (2021) for more technical details.

The MPP shows the probability of future carbon emissions of each country within the distribution, therefore, one can know more about the future trajectory and development path of each country in carbon emissions. Moreover, a list of priority can be set up so that attention can be paid to those countries with a high probability of increasing emissions in the future.

# RESULTS AND DISCUSSION

We adopt the distribution dynamics approach to examine the trend of carbon emission globally. An overview of the carbon emission patterns is presented first, using their transitional probability kernels. It is followed by a comparison between BRI and non-BRI economies, by means of their contour maps of transition probability kernels. Ergodic distributions and mobility probability plots are then constructed for BRI and non-BRI economies to provide detailed information on the current pattern and future development of their carbon emissions. Three measures of carbon emissions are used, namely per GDP, per capita, and per purchasing power parity (PPP). Emission per GDP is used as the baseline model to capture the evolution trend of BRI and non-BRI economies; emissions per capita and per purchase power parity (PPP) are used as alternative models to check the robustness of our findings. For the cross-country comparison, emissions per capita is a preferred metric as it indicates equality across countries. The other two metrics, however, provide possible alternatives as a robustness check.

## Baseline Model (Emission per GDP)

Stochastic kernel analyses are first performed to calculate the transitional dynamics of carbon emission per GDP for each economy. **Figure 1A** shows the three-dimensional plots of transitional probability kernels for carbon emission per GDP in economies which joined the BRI. **Figure1B** shows the three-dimensional plots of transitional probability kernels for carbon emission per GDP in economies that have not joined the BRI. In the 3D plot, X-axis presents the carbon emission per GDP at period t. Y-axis presents the carbon emission per GDP at period t+1. The width of the transition probability kernel for BRI and non-BRI economies are dispersed with the density mass concentrated along the 45-degree diagonal line, indicating that future carbon emission can vary significantly. In other words, there is inconsiderable persistence in the changes in the distribution for future carbon emission. There is one peak for BRI economies located near 0.8, indicating that their carbon emissions per GDP tend to cluster. BRI economies used to have high carbon emissions per GDP before joining BRI. In comparison, there are two peaks for non-BRI economies: one is located around 0.3; the other is located around 0.6. The implication is that non-BRI economies used to have lower carbon emission per GDP even before the introduction of BRI.

To capture more features of the transitional dynamics, **Figures 1C,D** displays the associated overhead view of their contour maps in **Figures 1A,B**. **Figure 1C** shows the clustering effect of BRI economie's emission at one heat point, while **Figure 1D** shows the clustering effect of BRI economie's emission at two heat points. Whether the heat points of BRI and non-BRI economies

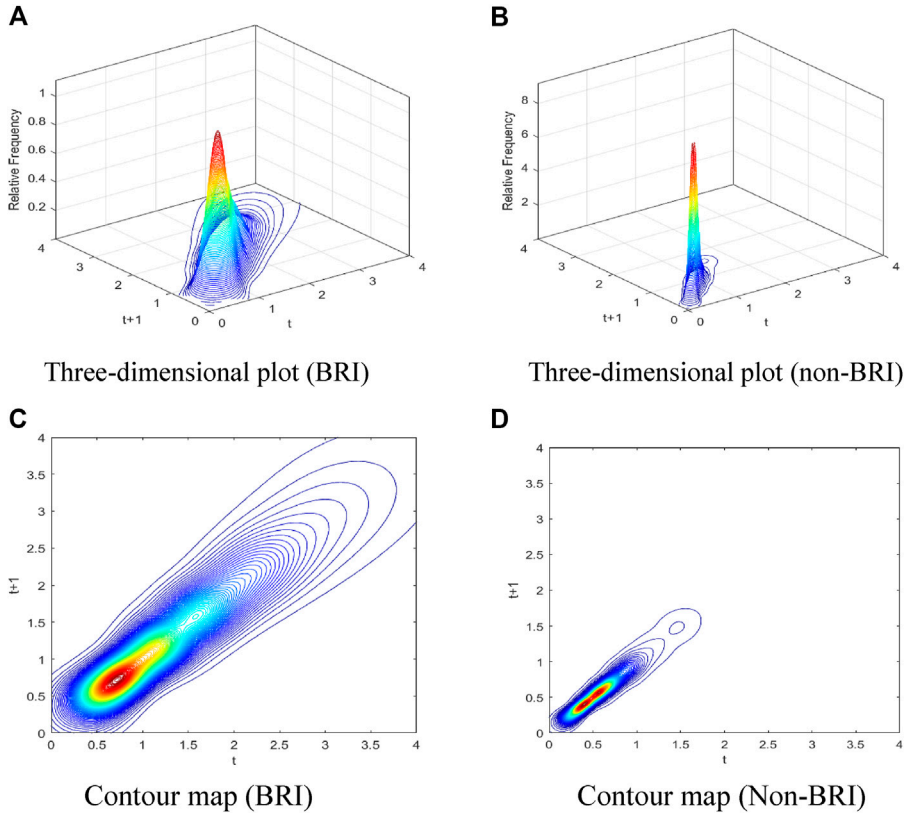



**FIGURE 1 |** Three-dimensional plot and Contour map of transitional probability kernel for carbon emission per GDP with yearly transitions, from 1987 to 2014. **(A)** Three-dimensional plot (BRI). **(B)** Three-dimensional plot (non-BRI). **(C)** Contour map (BRI). **(D)** Contour map (Non-BRI).

also represent their convergence clubs of carbon emission trends, in the long run, will require further investigation into their ergodic distribution.

As such, **Figure 2A** shows the ergodic distribution of the carbon emission per GDP of BRI economies, which is its long-run steady-state. The convergence club of 0.8 for BRI economies is more clearly observed in the ergodic distribution. While the ergodic distribution in **Figure 2A** has clarified the existence of one convergence club for BRI economies, the mobility of their future carbon emission cannot be identified. The new framework of MPP can effectively tackle the problem and offer a direct interpretation of the probability mass. In **Figure 2B**, the MPP on distributions of the probability mass is above zero when the carbon emission per GDP is no more than 0.8. However, this result alone does not suffice to say that BRI economie's carbon emission per GDP will always decline, as at round 1.0 and 1.5 the probability masses are still above zero, indicating that a few BRI economies may have higher carbon emission probabilities in the future.

In comparison, **Figure 2C** displays the ergodic distribution of the carbon emission per GDP of non-BRI economies. Unlike BRI economies, the ergodic distribution shows that the convergence clubs of non-BRI economies are primarily around 0.3 and 0.6. **Figure 2D** further reveals that the MPP on distributions of the probability mass is complicated in non-BRI economies: MPP of non-BRI economies is above zero for those with carbon emission levels lower than 0.59. Yet for carbon emission levels from 0.60 to 0.84, the MPP lies below zero which shows a net probability of moving downward. It displays a net tendency to move downward again in the coming periods for carbon emission levels ranged from 0.87 to 1.29. The MPP turns positive for carbon emission levels from 1.30 to 1.48. Afterwards, the MPP stays below the horizontal axis, except for carbon emission levels from 3.35 to 3.97. The results suggest that in general non-BRI emissions are lower and tend to decrease in the future despite that a few economies may have higher emission levels (those not clustered in the convergence clubs of 0.3 or 0.6).

To summarize, the baseline model of carbon emission per GDP has shown distinctive patterns among BRI and non-BRI economies before BRI was introduced: The average carbon emission level was higher in BRI economies (convergence club of 0.8) than in non-BRI economies (convergence clubs of 0.3 and 0.6). Such gap may be narrowed over time as the MPP of BRI economies almost always stays below zero when the carbon emission level is larger than 0.8, while the MPP of non-BRI economies has several intersects with the horizontal axis when the carbon emission level is larger than 1.0.

Our finding complimentary the empirical literature that studies the environmental impacts of BRI. Our counterfactual baseline framework can avoid the endogeneity issues as identified

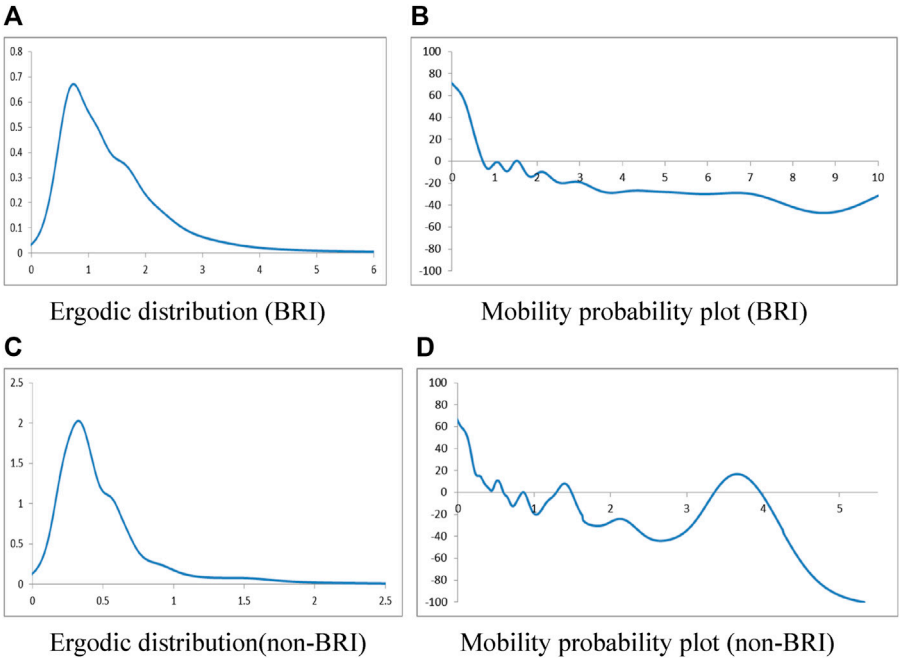



**FIGURE 2 |** Ergodic distribution and Mobility probability plot of carbon emission per GDP. **(A)** Ergodic distribution (BRI). **(B)** Mobility probability plot (BRI). **(C)** Ergodic distribution (non-BRI). **(D)** Mobility probability plot (non-BRI).

by Liu et al. (2020), and Wu et al. (2020) who find that the BRI may preferred curtained countries. This counterfactual baseline may lead to slight different conclusions of the emission impact of BRI, such as Zhang et al. (2017), Li et al. (2018), Huang (2019; and Huang and Li (2020), who find that BRI may have negative environmental impacts. Our aggregated findings provide additional support to the findings of Rauf et al. (2018a), Chen et al. (2020) and Coenen et al. (2021) who find that various factors have heterogeneous impacts. However, whether such finding is significant requires further robustness tests as follows.

## Alternative Model 1 (Emission per Capita)

We first perform stochastic kernel analyses to compute the transitional dynamics for carbon emission per capita for all economies. **Figure 3A** shows the three-dimensional plots of carbon emission per capita for BRI economies. Unlike emission per GDP, the peak for emission per capita of BRI economies is located around 0.15. The peak for emission per capita of non-BRI economies is located around 0.10 only (see **Figure 3B**). While both BRI and non-BRI economies have clustered at lower peaks when measured by the emission per capita, the peak for BRI economies remains higher than the peak for non-BRI economies. Similar to **Figure 1**, the width of the transition probability kernel for BRI and non-BRI economies are dispersed with the density mass concentrated along the 45-degree diagonal line.

**Figures 3C,D** displays the associated overhead view of their contour maps in **Figures 3A,B**, with the stochastic transition probability kernel of emission per capita across the range 0.10–0.15. The implication is that carbon emissions per capita of most economies are below the world average (among the clustered group), while some economies (both BRI and non-BRI) have extremely high levels of carbon emission per capita. It is also interesting to note that in the contour map of **Figure 3A** there is another concentrated point around 0.9 for BRI economies. That is to say, both peaks of BRI economies are higher than the peak of non-BRI economies, consistent with the findings for emission per GDP that BRI economies had higher average carbon emission levels.

**Figures 4A,B** further explores the evolution trend of carbon emission per capita for BRI economies. **Figure 4A** shows two convergence clubs, one at 0.14 and another at 0.85. BRI economies have both lower and higher convergence levels of carbon emission per capita in the long run. In particular, when looking into the MPP in **Figure 4B**, it is obvious that MPP is below zero for economies with emission per capita ranged from 0.14 to 0.55. Yet for emission per capita from 0.56 to 0.85, the MPP lies above zero which shows a net probability of moving upward. Afterwards, it displays a net tendency to move downward in the coming periods again for emission per capita greater than 0.86. The distribution suggests that economies with emission per capita will have a net probability of downward emission trends if current emission levels are between 0.14 and 0.55, or above 0.86. Thus the two convergence clubs of 0.14 and 0.85 are well-identified for BRI economies.

In contrast, **Figure 4C** displays one convergence club (0.11) for non-BRI economie's emission per capita only, indicating that the average emission levels will remain low in the future. The MPP for non-BRI economies in **Figure 4D** has a straightforward

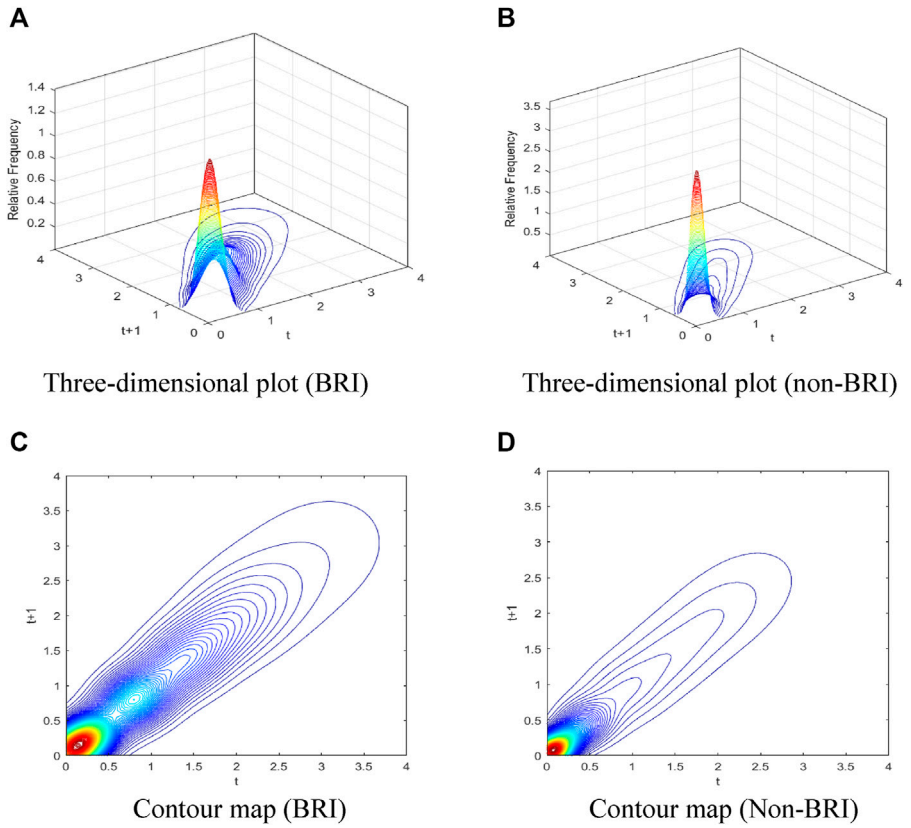



**FIGURE 3 |** Three-dimensional plot and Contour map of transitional probability kernel for carbon emission per capita with yearly transitions, from 1987 to 2014. **(A)** Three-dimensional plot (BRI). **(B)** Three-dimensional plot (non-BRI). **(C)** Contour map (BRI). **(D)** Contour map (Non-BRI).

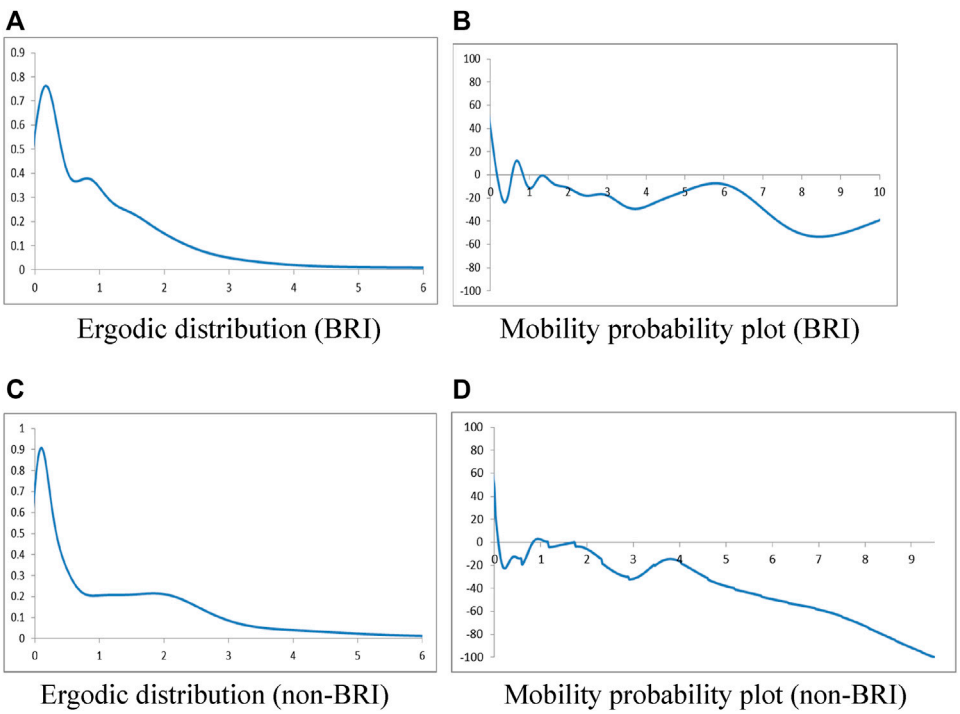



**FIGURE 4 |** Ergodic distribution and Mobility probability plot of carbon emission per capita. **(A)** Ergodic distribution (BRI). **(B)** Mobility probability plot (BRI). **(C)** Ergodic distribution (non-BRI). **(D)** Mobility probability plot (non-BRI).

pattern as well: The probability mass is above zero only when the carbon emission level is no more than 0.09, or between 0.83 and 1.13. In other ranges, the MPP has a net tendency of moving downwards.

In brief, while changing the measurement of emission from per GDP to per capita will have an impact on the number of convergence clubs for BRI and non-BRI economies, it does not affect the result that the average emission level is and will be higher in BRI economies based on the baseline model and alternative model 1.

## Alternative Model 2 (Emission per PPP)

Our baseline model (per GDP) and alternative model 1 (per capita) have indicated a contrast of carbon emission patterns and trends between BRI economies and non-BRI economies. However, purchasing power parity can be more suitable than GDP or capita to calculate the emission levels of the individual economy because PPP is closely related with the international exchange rate parities (Taylor and Taylor, 2004). Therefore, it is essential to explore whether the measurement of carbon emission per PPP has similar or different evolution pathways with the measurements of emission per GDP or per capita.

**Figures 5A,B** provides the three-dimensional plot of transitional probability kernels by BRI and non-BRI economies, measured by emission per PPP. It is worth noting that the distribution of carbon emission is bimodal in both **Figure 5A** and **Figure 5B**, justifying the importance of conducting distributional analysis to avoid deceiving results (Quah, 1997). **Figures 5C,D** displays their relevant contour maps. The width of the transition probability kernel for all economies are dispersed with the density mass concentrated along the 45° diagonal line, implying scattered future distribution of carbon emission per PPP. Specifically, the two peaks for BRI economies are located at around 0.6 and 0.9 respectively, while the two peaks for non-BRI economies are located at around 0.3 and 1.1 accordingly. Unlike emission per GDP or emission per capita, the peaks of BRI economies' emission per PPP are in-between the higher and lower bands of peaks for non-BRI economies.

**Figure 6A** shows the long run ergodic distribution of carbon emission per PPP in BRI economies. Convergence clubs can be observed at 0.59 and 0.86. Combined with the MPP in **Figure 6B**, it is evident that for BRI economies emitting up to 63%; or 80–89% of the world average, they tend to further increase their carbon emission per PPP in the future. For BRI economies

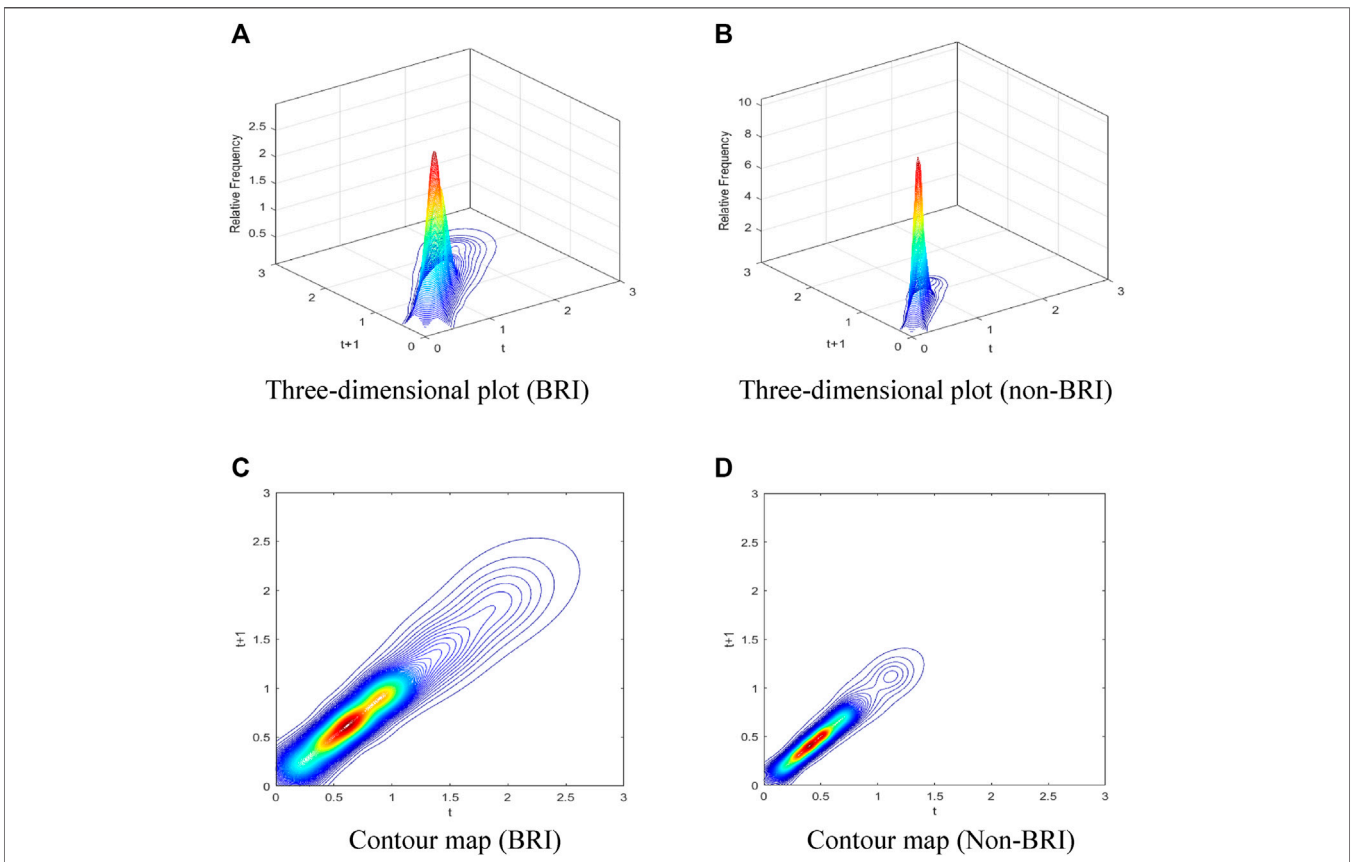



**FIGURE 5 |** Three-dimensional plot and Contour map of transitional probability kernel for carbon emission per PPP with yearly transitions from 1990 to 2014. **(A)** Three-dimensional plot (BRI). **(B)** Three-dimensional plot (non-BRI). **(A)** Three-dimensional plot (BRI) **(B)** Three-dimensional plot (non-BRI). **(C)** Contour map (BRI) **(D)** Contour map (Non-BRI).

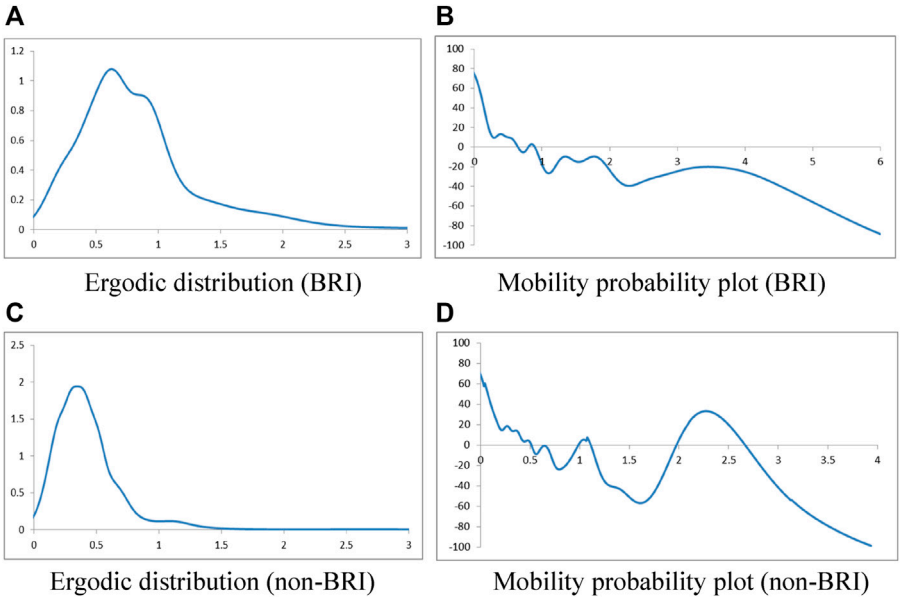



**FIGURE 6 |** Ergodic distribution and Mobility probability plot of carbon emission per PPP. **(A)** Ergodic distribution (BRI) **(B)** Mobility probability plot (BRI). **(C)** Ergodic distribution (non-BRI) **(D)** Mobility probability plot (non-BRI).

emitting 64–79% of the world average, they tend to reduce their future emission per PPP. Similarly, if their current emission per PPP is higher than 90% of the world average, BRI economies tend to reduce their carbon emission per PPP in the future. Hence, in the long run, the emission per PPP of BRI economies will tend to converge to lower than the world average levels.

**Figures 6C,D** provides a different picture for non-BRI economie's carbon emission per PPP. Convergence clubs can be observed at 0.34 and 1.12 respectively in **Figure 6C**. The future evolution trend is more complicated: MPP in **Figure 6D** shows that for non-BRI economies emitting 52–98% of the world average, they tend to reduce their carbon emission per PPP to the level of 34% of the world average. For non-BRI economies emitting 99% to 1.11 times of the world average, they tend to increase their future emission to 1.12 times of the world average. If their current emission level is 1.12–1.97 times the world average, non-BRI economies tend to reduce their levels of emission to 1.12 times the world average. Yet if non-BRI economies emit 1.98 times to 2.68 times the world average, they tend to increase their future carbon emission despite there is no convergence club. To conclude, although a majority of economies will reduce their carbon emissions, several economies will increase their emission and some of them will even reach an alarming level of two times or above (those outside the cluster, which is not fully shown in **Figure 5**). The overall results of alternative model two concur with the baseline model in that BRI economies are more likely to lower their emission levels in the future than non-BRI economies.

It is not surprising that the number of convergence clubs is different across the three different metrics because the three variables are different. However, the key hypothesis of this paper—the emissions dynamics are different between the BRI and non-BRI countries—is consistent across all three cases. This suggests that our finding is robust.

## CONCLUSION

The impact of the Belt and Road Initiative (BRI) on the environment is controversial. A major challenge to evaluate the environmental impact of the BRI is to separate its impact from the overall impact of development. Validation of the controversy about the environmental impact of BRI will help both China and the BRI participating countries to better understand the potential adverse impacts and prepare measures to prevent or mitigate the potential adverse effects. Despite many studies on the BRI's environmental impacts, there is an urgent need to create further baseline information to facilitate comparison and benchmark future evolution of emission patterns.

This paper attempts to provide an evolutionary and counterfactual baseline to evaluate the environmental impact of BRI, based on the distribution dynamics approach and the mobility probability plots (MPPs). Our estimation results suggest that the majority of BRI economies will lower their emission levels compared with the world average, despite that a substantial proportion of them are developing or emerging economies. This suggests that most BRI countries will not follow carbon-intensive development paths as other newly industries economies have done.

Furthermore, our estimations find that there is a systematic difference between BRI countries and non-BRI countries. Most non-BRI economies will have lower emission levels measured by per GDP or per capita than BRI countries, despite that a few

developed economies will have higher emission levels in the future. Specifically, in terms of emission per GDP the convergence clubs show that the majority of BRI economies tend to have emission levels at 80%, while non-BRI economies tend to have emission levels at 30% or 60%, of the global average emission level, in the long run. In terms of emission per capita, BRI economies tend to have emission levels at 14% or 85%, while non-BRI economies tend to have an emission level at 10% of the global average as their long-term steady states. In terms of emission per capita of PPP price, the convergence clubs of emission levels for BRI economies are 59% or 86% of the global average. In contrast, the convergence clubs of emission levels for non-BRI economies become 34% or 112% of the global average. Both BRI and non-BRI economies have outliers above the global average emission level, though.

Our findings have rich implications for understanding and managing the environmental impact of the Belt and Road Initiative: First, a counterfactual baseline for BRI countries is needed to assess the environmental impact of BRI. This study finds that non-BRI countries will have worsened environmental performances than their BRI counterparts in the absence of BRI. Therefore, the BRI's environmental impact should not be assessed against the current situation of BRI countries or the performance of non-BRI countries. Second, further policy intentions, such as environmental regulations and public pressure, are needed to promote green growth in BRI countries to low their future emission intensity. Third, international cooperation in regulations, data collection and technologies is necessary to mitigate the negative environmental impact across countries. Effective operation and management of the initiative require collaboration among a broad range of disciplines comprising civil engineering, material science, government operations, economics and finance, social and political science, and environmental science. Last but not least, an inclusive project assessment process that involves all stakeholders and the scientific community is useful to help relevant stakeholders address large-scale and complex challenges in the lifecycle management of these projects.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://databank.worldbank.org/source/world-development-indicators.

## AUTHOR CONTRIBUTIONS

JL: Formal analysis; Investigation; Validation; Writing - original draft; Writing - review and editing. XS: Conceptualization; Formal analysis; Investigation; Validation; Writing - original draft; Writing - review and editing. TC: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Visualization; Writing - review and editing.

## FUNDING

This study is supported by the National Science Foundation of China (No. 72174056, 71828401) and the CUHK Direct Grant for Research (Ref. 4052249).

## REFERENCES

Ahmad, F., Draz, M., Su, L., Ozturk, I., and Rauf, A. (2018). Tourism and Environmental Pollution: Evidence from the One Belt One Road Provinces of Western China. *Sustainability* 10, 3520. doi:10.3390/su10103520

Apergis, N., and Payne, J. E. (2017). Per Capita carbon Dioxide Emissions across U.S. States by Sector and Fossil Fuel Source: Evidence from Club Convergence Tests. *Energ. Econ.* 63, 365–372. doi:10.1016/j.eneco.2016.11.027

Ascensão, F., Fahrig, L., Clevenger, A. P., Corlett, R. T., Jaeger, J. A. G., Laurance, W. F., et al. (2018). Environmental Challenges for the Belt and Road Initiative. *Nat. Sustain.* 1, 206–209. doi:10.1038/s41893-018-0059-3

Baloch, M. A., Zhang, J., Iqbal, K., and Iqbal, Z. (2019). The Effect of Financial Development on Ecological Footprint in BRI Countries: Evidence from Panel Data Estimation. *Environ. Sci. Pollut. Res.* 26, 6199–6208. doi:10.1007/s11356-018-3992-9

Chen, Y., Liu, S., Wu, H., Zhang, X., and Zhou, Q. (2020). How can Belt and Road Countries Contribute to Glocal Low-Carbon Development?. *J. Clean. Prod.* 256, 120717.

Cheong, T. S., Li, V. J., and Shi, X. (2019). Regional Disparity and Convergence of Electricity Consumption in China: A Distribution Dynamics Approach. *China Econ. Rev.* 58, 101154. doi:10.1016/j.chieco.2018.02.003

Cheong, T. S., and Wu, Y. (2018). Convergence and Transitional Dynamics of China's Industrial Output: A County-Level Study Using a New Framework of Distribution Dynamics Analysis. *China Econ. Rev.* 48, 125–138. doi:10.1016/j.chieco.2015.11.012

Cheong, T. S., and Wu, Y. (2013). Regional Disparity, Transitional Dynamics and Convergence in China. *J. Asian Econ.* 29, 1–14. doi:10.1016/j.asieco.2013.08.007

Cheong, T. S., Wu, Y., and Wu, J. (2016). Evolution of Carbon Dioxide Emissions in Chinese Cities: Trends and Transitional Dynamics. *J. Asia Pac. Economy* 21, 357–377. doi:10.1080/13547860.2016.1176642

Dong, L., Yang, X., and Li, H. (2018). The Belt and Road Initiative and the 2030 Agenda for Sustainable Development: Seeking Linkages for Global Environmental Governance. *Chin. J. Popul. Resour. Environ.* 16, 203–210. doi:10.1080/10042857.2018.1487745

Ezcurra, R. (2007). Distribution Dynamics of Energy Intensities: A Cross-Country Analysis. *Energy Policy* 35, 5254–5259. doi:10.1016/j.enpol.2007.05.006

Fallahi, F., and Voia, M.-C. (2015). Convergence and Persistence in Per Capita Energy Use Among OECD Countries: Revisited Using Confidence Intervals. *Energ. Econ.* 52, 246–253. doi:10.1016/j.eneco.2015.10.004

Fan, J. L., Da, Y. B., Wan, S. L., Zhang, M., Cao, Z., and Wang, Y. (2019). Determinants of Carbon Emissions in 'Belt and Road initiative' countries: A production technology perspective. *Applied Energy* 239, 268–279.

Han, L., Han, B., Shi, X., Su, B., Lv, X., and Lei, X. (2018). Energy Efficiency Convergence across Countries in the Context of China's Belt and Road Initiative. *Appl. Energ.* 213, 112–122. doi:10.1016/j.apenergy.2018.01.030

Han, M., Lao, J., Yao, Q., Zhang, B., and Meng, J. (2020). Carbon inequality and economic development across the Belt and Road regions. *J. Environ. Manage.* 262, 110250.

Harlan, T. (2021). Green Development or Greenwashing? A Political Ecology Perspective on China's Green Belt and Road. *Eurasian Geography and Economics* 62 (2), 202–226.

Hong, P. (2016). Jointly Building the 'Belt and Road' Towards the Sustainable Development Goals. *Inclusive and Sustainable City Exhibition and Dialogue* (United Nations Industrial Development Organization.

Hsu, D. (2015). Identifying Key Variables and Interactions in Statistical Models of Building Energy Consumption Using Regularization. *Energy* 83, 144–155. doi:10.1016/j.energy.2015.02.008

Huang, Y. (2016). Understanding China's Belt & Road initiative: motivation, framework and assessment. *China Econ. Rev.* 40, 314–321.

Huang, Y. (2019). Environmental risks and opportunities for countries along the Belt and Road: Location choice of China's investment. *J. Clean. Prod.* 211, 14–26.

Huang, M. X., and Li, S. Y. (2020). The analysis of the impact of the Belt and Road initiative on the green development of participating countries. *Sci. Total Environ.* 722, 137869.

Hughes, A. C. (2019). Understanding and Minimizing Environmental Impacts of the Belt and Road Initiative. *Conservation Biol.* 33, 883–894. doi:10.1111/cobi.13317

Klinger, J. M. (2019). Environment, Development, and Security Politics in the Production of Belt and Road Spaces. *Territory, Polit. Governance* 8, 657–675. doi:10.1080/21622671.2019.1582358

Le Pen, Y., and Sévi, B. (2010). On the Non-convergence of Energy Intensities: Evidence from a Pair-wise Econometric Approach. *Ecol. Econ.* 69, 641–650. doi:10.1016/j.ecolecon.2009.10.001

Li, H., Li, F., and Yu, X. (2018). China's Contributions to Global Green Energy and Low-Carbon Development: Empirical Evidence Under the Belt and Road Framework. *Energies* 11 (6), 1527.

Li, N., and Shvarts, E. (2017). *The Belt and Road Initiative: WWF Recommendations and Spatial Analysis*. China: WWF.

Li, P., Qian, H., Howard, K. W. F., and Wu, J. (2015). Building a New and Sustainable "Silk Road Economic belt". *Environ. Earth Sci.* 74, 7267–7270. doi:10.1007/s12665-015-4739-2

Liu, Q., Wu, J., Li, L., Yu, L., Li, J., Xin, X., et al. (2018). Ecological Environment Monitoring for Sustainable Development Goals in the Belt and Road Region. *Yaogan Xuebao/Journal Remote Sens* 22, 686–708. doi:10.11834/jrs.20187264

Lloyd, P. J. (1996). The Role of Foreign Investment in the success of Asian Industrialization. *J. Asian Econ.* 7, 407–433. doi:10.1016/S1049-0078(96)90019-6

Magazzino, C., Mele, M., Morelli, G., and Schneider, N. (2021a). The Nexus between Information Technology and Environmental Pollution: Application of a New Machine Learning Algorithm to OECD Countries. *Utilities Policy* 72, 101256. doi:10.1016/j.jup.2021.101256

Magazzino, C., and Mele, M. (2020). On the Relationship between Transportation Infrastructure and Economic Development in China. *Res. Transportation Econ.*, 100947. doi:10.1016/j.retrec.2020.100947

Magazzino, C., Mele, M., and Schneider, N. (2021b). A Machine Learning Approach on the Relationship Among Solar and Wind Energy Production, Coal Consumption, GDP, and CO2 Emissions. *Renew. Energ.* 167, 99–115. doi:10.1016/j.renene.2020.11.050

Magazzino, C., Mele, M., and Schneider, N. (2021c). Testing the Convergence and the Divergence in Five Asian Countries: from a GMM Model to a New Machine Learning Algorithm. *J. Econ. Stud.*. doi:10.1108/JES-01-2021-0027

Mele, M., and Magazzino, C. (2020). A Machine Learning Analysis of the Relationship Among Iron and Steel Industries, Air Pollution, and Economic Growth in China. *J. Clean. Prod.* 277, 123293. doi:10.1016/j.jclepro.2020.123293

Payne, J. E., Vizek, M., and Lee, J. (2017). Stochastic Convergence in Per Capita Fossil Fuel Consumption in U.S. States. *Energ. Econ.* 62, 382–395. doi:10.1016/j.eneco.2016.03.023

Qiu, X., Wong, E. Y. C., and Lam, J. S. L. (2018). Evaluating Economic and Environmental Value of Liner Vessel Sharing along the Maritime Silk Road. *Maritime Pol. Manage.* 45, 336–350. doi:10.1080/03088839.2018.1437285

Quah, D. (1993b). Empirical Cross-Section Dynamics in Economic Growth. *Eur. Econ. Rev.* 37, 426–434. doi:10.1016/0014-2921(93)90031-5

Quah, D. (1993a). Galton's Fallacy and Tests of the Convergence Hypothesis. *Scand. J. Econ.* 95, 427. doi:10.2307/3440905

Quah, D. (2006). Galton's Fallacy and Tests of the Convergence Hypothesis. *Scand. J. Econ.* 95, 427. doi:10.2307/3440905

Quah, D. T. (1996). Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Econ. J.* 106, 1045. doi:10.2307/2235377

Rauf, A., Liu, X., and Amin, W. (2018a). Testing EKC Hypothesis with Energy and Sustainable Development Challenges: A Fresh Evidence From Belt and Road Initiative Economies. *Environ Sci Pollut. Res.* 25, 32066–32080.

Shi, X., Cheong, T. S., Yu, J., and Liu, X. (2021). Quality of Life and Relative Household Energy Consumption in China. *China & World Economy* 29, 127–147. doi:10.1111/cwe.12390

Shi, X., Wang, K., Cheong, T. S., and Zhang, H. (2020). Prioritizing Driving Factors of Household Carbon Emissions: An Application of the LASSO Model with Survey Data. *Energ. Econ.* 92, 104942. doi:10.1016/j.eneco.2020.104942

Shi, X., and Yao, L. (2019). Prospect of China's Energy Investment in Southeast Asia under the Belt and Road Initiative: A Sense of Ownership Perspective. *Energ. Strategy Rev.* 25, 56–64. doi:10.1016/j.esr.2019.100365

Silverman, B. W. (2018). *Density Estimation for Statistics and Data Analysis*. Boca Raton, Florida: CRC Press. doi:10.1201/9781315140919

Suocheng, D., Kolosov, V., Yu, L., Zehong, L., Fujia, L., Minyan, Z., et al. (2017). Green Development Modes of the belt and Road. *Geogr. Environ. Sustain.* 10, 53–69. doi:10.24057/2071-9388-2017-10-1-53-69

Suocheng, D., Zehong, L., Yu, L., Guangyi, S., Huilu, Y., Juanle, W., et al. (2015). Resources, Environment and Economic Patterns and Sustainable Development Modes of the Silk Road Economic Belt. *J. Resour. Ecol.* 6, 65–72. doi:10.5814/j.issn.1674-764x.2015.02.001

Taylor, A. M., and Taylor, M. P. (2004). The Purchasing Power Parity Debate. *J. Econ. Perspect.* 18, 135–158. doi:10.1257/0895330042632744

Teo, H. C., Lechner, A. M., Walton, G. W., Chan, F. K. S., Cheshmehzangi, A., Tan-Mullins, M., et al. (2019). Environmental Impacts of Infrastructure Development under the belt and Road Initiative. *Environments* 6, 72. doi:10.3390/environments6060072

Tian, X., Hu, Y., Yin, H., Geng, Y., and Bleischwitz, R. (2019). Trade Impacts of China's Belt and Road Initiative: From Resource and Environmental Perspectives. *Resour. Conservation Recycling* 150, 104430. doi:10.1016/j.resconrec.2019.104430

Udemba, E. N., Magazzino, C., and Bekun, F. V. (2020). Modeling the Nexus between Pollutant Emission, Energy Consumption, Foreign Direct Investment, and Economic Growth: New Insights from China. *Environ. Sci. Pollut. Res.* 27, 17831–17842. doi:10.1007/s11356-020-08180-x

Wu, J., Wu, Y., Guo, X., and Cheong, T. S. (2016). Convergence of Carbon Dioxide Emissions in Chinese Cities: A Continuous Dynamic Distribution Approach. *Energy Policy* 91, 207–219. doi:10.1016/j.enpol.2015.12.028

Wu, J., Wu, Y., Se Cheong, T., and Yu, Y. (2018). Distribution Dynamics of Energy Intensity in Chinese Cities. *Appl. Energ.* 211, 875–889. doi:10.1016/j.apenergy.2017.10.097

Wu, Y., Hu, C., and Shi, X. (2021). Heterogeneous Effects of the Belt and Road Initiative on Energy Efficiency in Participating Countries. *Energies* 14, 5594. doi:10.3390/en14185594

Wu, Y., Shi, X., and Hu, C. (2021). Per Capita CO2 Emissions Divergence Influenced by Bilateral Trade with china under the belt and Road Initiative. *Sustainable Prod. Consumption* 27, 1589–1601. doi:10.1016/j.spc.2021.04.005

Yao, L., Andrews-Speed, P., and Shi, X. (2019). Asean Electricity Market Integration: How Can Belt and Road Initiative Bring New Life to it. *Singapore Econ. Rev.* 66, 85–103. doi:10.1142/s0217590819500413

Yin, W. (2019). Integrating Sustainable Development Goals into the Belt and Road Initiative: Would It Be a New Model for Green and Sustainable Investment?. *Sustainability* 11 (24), 6991

Yu, J., Shi, X., and Cheong, T. S. (2021a). Distribution Dynamics of China's Household Consumption Upgrading. *Struct. Change Econ. Dyn.* 58, 193–203. doi:10.1016/j.strueco.2021.05.012

Yu, J., Shi, X., Guo, D., and Yang, L. (2021b). Economic Policy Uncertainty (EPU) and Firm Carbon Emissions: Evidence Using a China Provincial EPU index. *Energ. Econ.* 94, 105071. doi:10.1016/j.eneco.2020.105071

Zhang, Y., Zhang, J.-H., Tian, Q., Liu, Z.-H., and Zhang, H.-L. (2018). Virtual Water Trade of Agricultural Products: A New Perspective to Explore the Belt and Road. *Sci. Total Environ.* 622-623, 988–996. doi:10.1016/j.scitotenv.2017.11.351

Zhang, N., Liu, Z., Zheng, X., and Xue, J. (2017). Carbon footprint of China's belt and road. *Science* 357 (6356), 1107.

Zhao, Z., Shi, X., Zhao, L., and Zhang, J. (2021). Extending Production-Theoretical Decomposition Analysis to Environmentally Sensitive Growth: Case Study of

Belt and Road Initiative Countries. *Technol. Forecast. Soc. Change* 161, 120289. doi:10.1016/j.techfore.2020.120289

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Li, Cheong and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

# APPENDIX

**Table A1 |** Scope of BRI and non-BRI countries

| **BRI countries** |
| --- |
| Afghanistan, Albania, Armenia, Azerbaijan, Bahrain, Bangladesh, Belarus, Bhutan, Bosnia and Herzegovina, Brunei Darussalam, Bulgaria, Cambodia, China, Croatia, Czech Republic, Egypt, Arab Rep., Estonia, Ethiopia, Georgia, Hungary, India, Indonesia, Islamic Rep. of Iran, Iraq, Israel, Jordan, Kazakhstan, Rep. of Korea, Kuwait, Kyrgyz Republic, Lao PDR, Latvia, Lebanon, Lithuania, Macedonia, Madagascar, Malaysia, Maldives, Moldova, Mongolia, Montenegro, Morocco, Myanmar, Nepal, New Zealand, Oman, Pakistan, Panama, Philippines, Poland, Qatar, Romania, Russian Federation, Saudi Arabia, Serbia, Singapore, Slovak Republic, Slovenia, South Africa, Sri Lanka, Syrian Arab Republic, Tajikistan, Thailand, Timor-Leste, Turkey, Turkmenistan, Ukraine, United Arab Emirates, Uzbekistan, Vietnam, Yemen |
| **Non-BRI countries** |
| Algeria, American Samoa, Andorra, Angola, Antigua and Barbuda, Argentina, Aruba, Australia, Austria, Bahamas, Barbados, Belgium, Belize, Benin, Bermuda, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cabo Verde, Cameroon, Canada, Cayman Islands, Central African Republic, Chad, Channel Islands, Chile, Colombia, Comoros, Congo, Dem. Rep., Congo, Rep., Costa Rica, Cote d'Ivoire, Cuba, Curacao, Cyprus, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Eritrea, Faroe Islands, Fiji, Finland, France, French Polynesia, Gabon, Gambia, Germany, Ghana, Greece, Greenland, Grenada, Guam, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong SAR, Iceland, Ireland, Isle of Man, Italy, Jamaica, Japan, Kenya, Kiribati, Dem. People's Rep. of Korea, Kosovo, Lesotho, Liberia, Libya, Liechtenstein, Luxembourg, Macao SAR, Malawi, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mexico, Fed. Sts. of Micronesia, Monaco, Mozambique, Namibia, Netherlands, New Caledonia, Nicaragua, Niger, Nigeria, Northern Mariana Islands, Norway, Palau, Papua New Guinea, Paraguay, Peru, Portugal, Puerto Rico, Rwanda, Samoa, San Marino, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Sint Maarten (Dutch part), Solomon Islands, Somalia, South Sudan, Spain, St. Kitts and Nevis, St. Lucia, St. Martin (French part), St. Vincent and the Grenadines, Sudan, Suriname, Sweden, Switzerland, Tanzania, Togo, Tonga, Trinidad and Tobago, Tunisia, Turks and Caicos Islands, Tuvalu, Uganda, United Kingdom, United States, Uruguay, Vanuatu, Venezuela, Virgin Islands, West Bank and Gaza, Zambia, Zimbabwe |

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fenvs.2021.779358

# Exposing the Effects of Environmental Regulations on China's Green Total Factor Productivity: Results From Econometrics Analysis and Machine Learning Methods

*Junyi Feng, Jianjun Yan* and Xia Tao*

*Hunan University of Technology and Business, Changsha, China*

OPEN ACCESS

**Edited by:**
*Xunpeng Shi,*
*University of Technology Sydney, Australia*

**Reviewed by:**
*Hao-Kuang Li,*
*Shanghai Academy of Social Sciences, China*
*Solomon Prince Nathaniel,*
*University of Lagos, Nigeria*
*Yongran Lin,*
*Henan Administration Institute, China*

***Correspondence:**
*Jianjun Yan*
*yan751225@163.com*

**Specialty section:**
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

**Received:** *18 September 2021*
**Accepted:** *28 October 2021*
**Published:** *18 November 2021*

**Citation:**
*Feng J, Yan J and Tao X (2021) Exposing the Effects of Environmental Regulations on China's Green Total Factor Productivity: Results From Econometrics Analysis and Machine Learning Methods. Front. Environ. Sci. 9:779358. doi: 10.3389/fenvs.2021.779358*

With the increasingly obvious restriction of the ecological environment on economic development, environmental regulations are widely used to achieve "green production," that is, to improve green total factor productivity (GTFP). First, through the econometric model, it can be concluded that command-based environmental regulations could improve GTFP, while market-based environmental regulations have no significant impact on GTFP. Unlike traditional econometric models, machine learning has no specific data requirements and research assumptions. We use Lasso regression to verify the above results by obtaining the optimal tuning parameter. Furthermore, considering that the leap of China's economy is inseparable from foreign direct investment (FDI), we use FDI as a threshold variable. The threshold model results showe that when the intensity of FDI in China ranges between 1.2492 and 1.588, both types of environmental regulations can significantly promote GTFP. These conclusions passed the robustness test. Given the differences in economy and resource endowment among different regions in China, a regional heterogeneity test is conducted. The results show that the current environmental regulations in eastern and central China have no significant impact on GTFP. However, when the intensity of FDI in central China is greater than 3.6868, environmental regulations have a significant promoting effect on GTFP. In western China, when FDI intensity ranges between 1.3950 and 1.5880, market-based environmental regulations can significantly promote GTFP. Further, the path test of the mediation effect model reveals that command-based environmental regulations reduce GTFP by reducing FDI. The above conclusions provide empirical data for the intensity of FDI in different regions of China to improve GTFP.

**Keywords: environmental regulations, green total factor productivity, LASSO model, threshold model, mediation effect model**

# INTRODUCTION

Since the reform and opening up in 1978, China has adopted a sloppy production strategy in the industrial sector. While this approach and inappropriate economic growth strategies (Nathaniel et al., 2021a) have greatly contributed to the economic growth of China, China's environmental problems have become increasingly serious, for example, the haze in the Beijing-Tianjin-Hebei region in 2015 and dust storms in Inner Mongolia in 2017. Thus, achieving green development is urgently required in China. The essence of green development is to reduce resource consumption, reduce environmental pollution, and achieve comprehensive, coordinated, and sustainable development of the economic, social, and ecological environment (Wu et al., 2021a). This proposes increased requirements for national sustainability practices (Nathaniel et al., 2021b).

The Chinese government has made substantial efforts to achieve sustainable development by adopting various measures, such as establishing carbon markets and implementing environmental taxes/charges (Liao and Shi, 2018) and green energy infrastructure (Cheng et al., 2021). Li et al. (2021), and has focused on the application of hydrogen energy and fuel cell electric vehicles in ASEAN countries. Its economic competitiveness and environmental implications may inspire future developments (Li and Kimura, 2021).

Our study focuses on the effects of environmental regulations on green development. Given that the proportion of public participation in environmental protection in China is still low, the main forms of environmental regulation are command-based environmental regulations and market-based environmental regulations. The effectiveness of different environmental regulations also depends on the degree of marketization and government decision-making behavior in each region, as well as FDI (Jiang et al., 2020). Whether environmental regulations can improve the GTFP and whether this effect varies regionally in China are two issues that need to be addressed.

To quantify "green production," we used the Malmquist-Luenberger index (ML index) under the Slacks-Based Measure (SBM) model and max-Data Envelopment Analysis (DEA) to measure green total factor productivity (GTFP) in each Chinese province (except Tibet). Simultaneously, we also used China's provincial macro data to measure environmental regulation. We not only studied the impact of environmental regulations on GTFP, but also considered the impacts of FDI intensity. Considering the regional differences in China (Wu et al., 2021b), the impact of environmental regulations in various regions of China on GTFP will be studied separately. Finally, through the intermediary effect model, we explored the mechanism path of the effects of environmental regulations on GTFP.

In view of the research on environmental regulations and GTFP, few studies have introduced machine learning into this field. Moreover, traditional econometric models often have certain data requirements and research assumptions, but in reality, we cannot make assumptions about real data. Machine learning has no such limitations. Therefore, we used machine learning to introduce Lasso regression in this research field to verify the results of the traditional econometric model to achieve breakthrough research on the effect of the traditional single econometric model on environmental regulations. We hope to provide some advice for different regions of China to improve GTFP. To the best of our knowledge, few papers have focused on the effects of environmental regulations by combining econometrics and machine learning.

This paper is structured as follows: *Literature Review* section reviews the literature, *Data and Methodology* section presents the data and methods used, *Results and Discussion* section presents the results, and *Conclusion* section presents the conclusions.

# LITERATURE REVIEW

## Review of Theoretical Research

### Measurement of GTFP

Since the last century, some scholars have attempted to obtain adjusted GDP values by deducting the market value of environmental externalities from normal GDP values (Nordhaus and Tobin, 1972), namely green GDP. However, green GDP cannot easily quantify environmental and natural resource damage. This has prompted academics to develop GTFP by considering environmental performance in conjunction with environmental variables based on traditional total factor productivity (TFP) indicators.

In the past decades, the GTFP has become an important indicator for examining the quality of China's economic development. Consequently, a large body of relevant literature has emerged (Long et al., 2016; Wang et al., 2018; Ai et al., 2020). Various studies on GTFP were conducted considering environmental performance. The main productivity measures are the traditional Malmquist index (Yue et al., 2019) and Fischer index. However, these indexes do not consider the non-desired output in the production process during measurement, consequently, leading to errors.

Therefore, researchers have introduced the ML index (Wang et al., 2020), which considers a non-desired output. ML indices have also been used in combination with directional distance functions (Chung et al., 1997), which relaxed the assumption of equal proportionality between the expected output increase, energy saving, and emission reduction (Zhou et al., 2012). Zhang and Choi (2013) considered relaxation errors and infeasible solution problems to construct a total factor $CO_2$ emission indicator based on the non-radial directional distance function. The combination of ML and distance function considers the non-desired output while calculating GTFP, and comprehensively measures the quality of economic growth.

To summarize, because traditional methods, such as Malmquist index and Fischer index, do not account for undesirable output, we preferred to use the ML index method under the SBM model and Max DEA to calculate GTFP for all Chinese provinces (except Tibet) from 2000 to 2017.

### Environmental Regulations and GTFP

Several scholars have studied the relationship between environmental regulations and TFP (Ambec et al., 2013; Zhao et al., 2015; Dong et al., 2015). Most of the earlier studies concluded that environmental regulations increased the

production costs of firms, leading to varying degrees of productivity loss (Barbera and Mcconnell, 1990; Wayne and Ronald, 1998; Gray and Shadbegian, 2003; Greenstone, et al., 2012; Rexhäuser and Rammer, 2014). Moreover, manufacturing firms prefer to use cheap and polluting fossil fuels to address the rising economic policy uncertainty (Yu et al., 2021), which is not conducive to GTFP. However, some studies have come to a more optimistic conclusion that environmental regulations may improve productivity by forcing firms to carry out technological innovation and reduce compliance costs (Porter and Van, 1995). The above two completely opposite opinions are because researches are based on two different theories, the "compliance cost theory" and the "innovation compensation theory."

Based on the classification of environmental regulations, we analyzed the effects of command-based environmental regulations and market-based environmental regulations on GTFP separately, as shown in **Figure 1**. The use of legislation by the government to enact a series of legal policies that can directly regulate emitters is known as command-based environmental regulations. Emission standards, manufacturing process standards, and energy or waste reduction standards are examples of specific initiatives. It is generally believed that command-based environmental regulations do not play a strong role in promoting GTFP because of its administrative nature.

The use of market mechanisms to incentivize polluters to reduce pollution is referred to as market-based environmental regulations. Emissions levies, tradable permit systems, and emission reduction subsidies are some of the specific approaches. By taking German enterprises as the research sample, Bitat (2018) found that only long-term objectives and market incentives are positively associated with eco-innovation. Greater the government's intervention in markets, greater is the incentive for firms to "rent-seek." The "rent-seeking" behavior populates most corporate innovations (Li et al., 2020) and weakens the incentive for corporates' green technology innovation. This is not conducive to GTFP.

Thus, the impact of environmental regulations on GTFP is still uncertain. Environmental regulations may increase costs, thus, crowding out the amount of firms' innovation expenditure, consequently, resulting in low GTFP. However, environmental regulations may also force firms to conduct green technology innovation to promote GTFP. In addition, the impact of environmental regulations on GTFP may change with the intensity of environmental regulations or other variables. On this basis, we further considered the nonlinear relationships between environmental regulations and GTFP.

## Environmental Regulations, FDI, and GTFP

In the past years, China has actively promoted FDI in the hope that it will bring advanced thinking technology, management

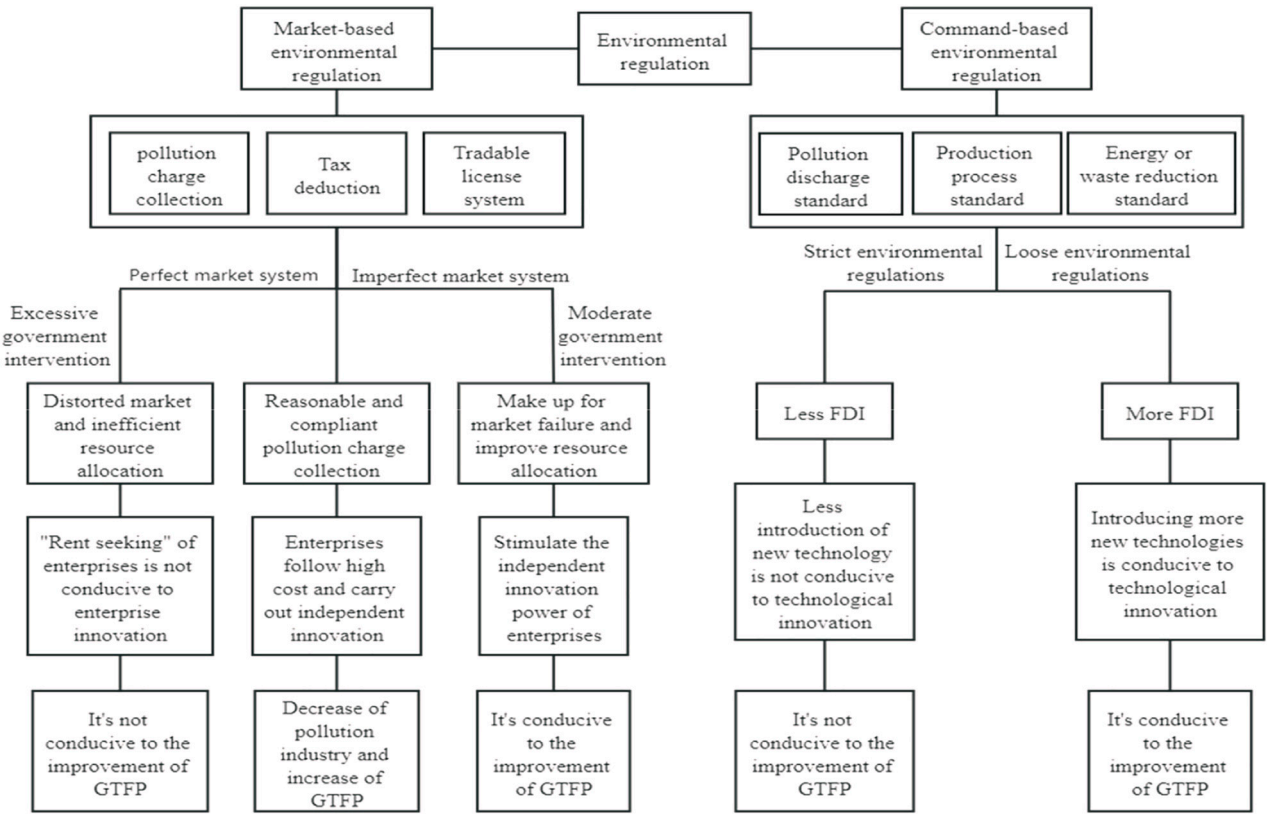



**FIGURE 1 |** The policy effect of two types of environmental regulations.

experience, and green technology. However, national environmental regulation policies determine whether or not FDI has a spillover effect in China. Qiu et al. (2021) suggested that the impact of environmental regulations on FDI inflows was negative. Under the constraint of environmental regulations, enterprises must increase investment in pollution control, which directly increases the cost of enterprises and reduces FDI.

Murshed et al. (2021) confirmed the validity of "pollution heaven hypothesis" in South Asia. FDI inflows are detrimental to the prospects of achieving environmental sustainability within the concerned South Asian countries. Further, FDI brings forward advanced technology while causing environmental pollution in developing countries and regions (Dou and Han, 2019). Qiu et al. (2021) conducted an empirical analysis and confirmed that the "pollution heaven hypothesis" existed in the eastern and central regions of China. In addition, some scholars have pointed out that the FDI spillover effect hinders the progress of industrial technology in China. The solution is to strengthen environmental regulations (Yu and Li, 2020). However, some scholars disagreed and argued that the "pollution heaven hypothesis" was not valid in China. Some scholars have suggested that the positive interaction between FDI and environmental regulations has indirectly led China to strengthen its environmental regulations. In turn, strict environmental regulations could effectively raise the environmental threshold for foreign investment entry and played a "screening" role for FDI (Qiu et al., 2021).

In fact, there is still no unified conclusion on whether the "pollution heaven hypothesis" has been established in China. However, we observed the impact of FDI on the environmental regulations' effect on GTFP. The existing studies have focused more on the relationship between FDI and environmental regulations or GTFP, but few studies have considered FDI as a threshold variable to study the impact of its intensity change on the effects of environmental regulations. Therefore, in this study, we focused on the effect of changes in FDI intensity on the effect of environmental regulations on GTFP. We used FDI as a threshold variable to consider the effect of FDI intensity change on the effect of environmental regulations on GTFP.

## Review of Research Methods

The effect of environmental regulations on GTFP is studied using the linear probability and the proportional hazards models (Ai et al., 2020), systematic generalized estimation method (GMM), dynamic count data model (Bitat, 2018), difference-in-difference (DID) model (Zhang et al., 2021), bootstrap panel Granger causality test (Liu et al., 2021), spatial model (Peng, 2020) and evolutionary game theory (Ulph, 2000). It is concluded that environmental regulations have an "innovation compensation effect," thus, increasing productivity. The above econometric models are not only the mainstream research approaches in the study of environmental regulations and GTFP, but also one of the main methods used in this study.

In addition, in the field of environment, several researchers have used machine learning to study air pollution, $CO_2$ emissions, and energy consumption. Some researchers have used machine learning to identify causal relationships (Payne, 2012). Mele and Magazzino (2020) used a causal direction from depth (D2C) algorithm to derive higher pollution caused by economic growth concentrations, which may contribute to neocrown pneumonia by making the respiratory system more susceptible to infection. Similarly, Cosimo et al. (2021) used the D2C algorithm to identify and predict the causality of coal consumption, solar wind production, economic growth, and $CO_2$ emissions in China, India, and the United States.

Among the existing studies, most of the research on environmental regulations and GTFP has been based on econometric models. In their linear regression studies, none of the scholars have introduced machine learning into the field. Therefore, we combined econometric models with machine learning and used the least absolute shrinkage and selection operator (Lasso) regression to test the results of the econometric empirical. Subsequently, the regional heterogeneity of the impact of environmental regulations on GTFP was analyzed.

# DATA AND METHODOLOGY

## Data

### Core Variables

The explained variable included green total factor productivity (GTFP), which is the inclusion of environmental pollution caused by production as a non-desired output in the calculation of TFP. We used the ML index method under the SBM model to measure the GTFP of Chinese provinces from 2000 to 2017. According to the Malmquist- Luenberger index (ML index)-based approach proposed by Chung et al. (1997):

$$
\begin{aligned}
ML_{t}^{t+1} &= \left\{\frac{\left[1+D_{i}^{t}\left(x^{t},y^{t},b^{t};g^{t}\right)\right]}{\left[1+D_{i}^{t}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)\right]}\times\frac{\left[1+D_{i}^{t+1}\left(x^{t},y^{t},b^{t};g^{t}\right)\right]}{\left[1+D_{i}^{t+1}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)\right]}\right\}^{\frac{1}{2}}\\
&= \left\{\frac{\left[1+D_{i}^{t+1}\left(x^{t},y^{t},b^{t};g^{t}\right)\right]}{\left[1+D_{i}^{t}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)\right]}\times\frac{\left[1+D_{i}^{t+1}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)\right]}{1+D_{i}^{t}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)}\right\}^{\frac{1}{2}}\\
&\quad\times\left\{\frac{\left[1+D_{i}^{t}\left(x^{t},y^{t},b^{t};g^{t}\right)\right]}{\left[1+D_{i}^{t}\left(x^{t+1},y^{t+1},b^{t+1};g^{t+1}\right)\right]}\right\}\\
&= TECH_{t}^{t+1}\times EFFCH_{t}^{t+1}
\end{aligned}
\tag{1}
$$

The ML index can be decomposed into two components, the efficiency improvement index (EFFCH) and the technological progress index (TECH). An ML > 1 indicates that TFP grows. An EFFCH > 1 indicates an improvement in efficiency. A TECH > 1 indicates technological progress. ML is the growth rate of the GTFP, and t represents time. D is the production unit, and x, y, and b are the input factors, desired output, and non-desired output, respectively. The input elements include 1) the capital stock of each province: $K_{it} = I_{it}+(1-\delta_{it})K_{it}-1$, where $I_{it}$ is the total fixed capital formation. δ is the depreciation rate of fixed assets, 2) the number of employed persons in each province at the end of each year, and 3) resources (standard coal) consumed by each province annually. The expected output is the total GDP of each province. The non-desired output is the industrial solid waste

**TABLE 1 |** Control variables.

| Variable | Name | Meanings |
|---|---|---|
| PGDP | Economic development level | Per capita GDP |
| INS | Industrial structure | Percentage of secondary industry in total output |
| INN | Intensity of research and development (R&D) | Percentage of internal expenditure on R&D funds to the province's GDP |
| FDI | Foreign investment | Percentage of foreign direct investment in each province to the province's fixed assets |
| GOV | Government intervention | The percentage of environmental protection expenditure in the general budget expenditure of each province's government |
| POL | Pollution control investment | Investment in emissions control by province as a percentage of the province's GDP government intervention |

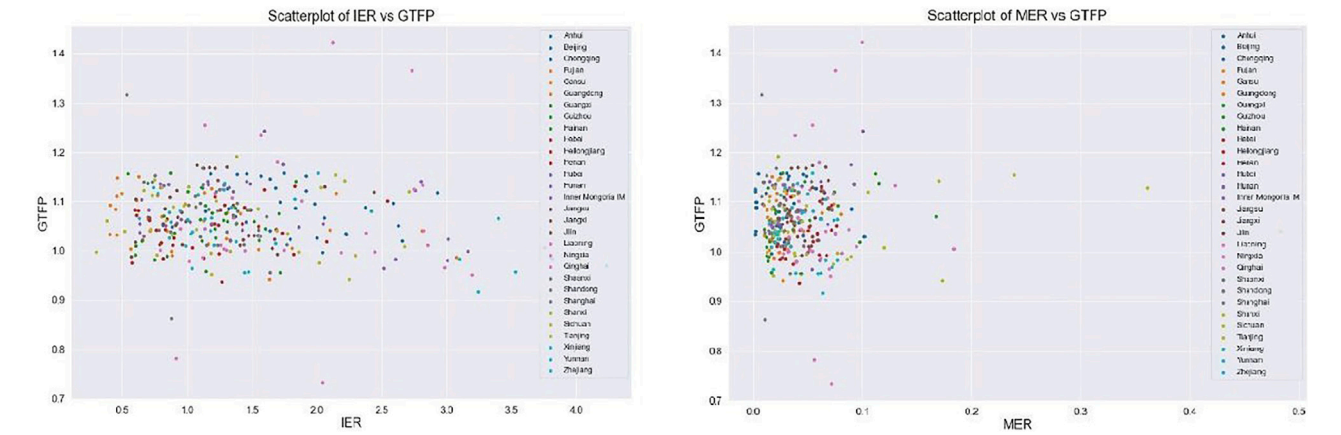



**FIGURE 2 |** Visualization of the correlation between environmental regulations and GTFP.

emissions of each province. By applying Max DEA software, the GTFP of 30 Chinese provinces from 2000 to 2017 was measured.

Core explanatory variables are market-based environmental regulations (MER) and command-based environmental regulations (IER). We used the amount of sewage charges collected by each Chinese province as a percentage of the province's GDP during 2007–2017 as market-based environmental regulations. We chose the percentage of actual pollution control investment in the GDP of each Chinese province from 2007 to 2017 as the indicator of command-based environmental regulations. The actual pollution input refers to industrial pollution control investment, construction projects' "three simultaneous" investment in environmental protection, waste gas, and wastewater pollution control facilities operating costs. Specific control variables can be seen in **Table 1**.

### Sources of Data and Descriptive Analysis

We selected the panel data of 30 provinces, autonomous regions, and municipalities directly under the central government in China from 2007 to 2017 as sample data. All data were obtained from the China Statistical Yearbook, China Statistical Yearbook of Industrial Economy, China Statistical Yearbook of Energy, China Environmental Yearbook, and the statistical yearbooks of each province in previous years. Descriptive statistics of the variables are shown in **Supplementary Table S1**. The correlations between the core explanatory variables and the explained variables are visualized and analyzed in **Figure 2**.

From **Figure 2**, GTFP is concentrated between 1.0 and 1.2. The values of IER are concentrated between 0.5 and 1.5. The values of MER are all concentrated between 0 and 0.1. IER was negatively correlated with GTFP, but MER was not correlated with GTFP. This was further validated using econometric models and machine learning.

## Methodology

### Orthogonal Least Square Regression

To explore the relationship between the two variables, we used the regression method of econometrics. The regression model equation can be expressed as follows:

$$y_i = \alpha + \beta x_i + \varepsilon_i \, (i = 1.....n) \tag{2}$$

where $y_i$ is the dependent variable, $x_i$ is the explanatory variable, $\alpha$ is the intercept term, and $\varepsilon_i$ is the residual term.

For the above regression models, OLS is the most commonly used method for estimating model coefficients (Helwig, 2017). The OLS method estimates the parameters in the regression model by minimizing the error between the predicted and observed values of the outcome variable, which can provide the most accurate linear unbiased estimates for the current sample (Chartterjee and Hadi, 2006).

In the standard OLS regression, the parameter estimates of the regression model can be obtained by the following optimization problem:

$$\hat{\beta}_{ols} = arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2\right\} \quad (3)$$

However, the OLS approach focuses on unbiased estimation of the current dataset, which can easily lead to overfitting of the model (Yarkoni and Westfall, 2017).

## Lasso Regression

With the rise and development of machine learning, more statistical tools have emerged. These tools continue to compensate for the limitations of the traditional methods. Among them, the regularization method represented by the least absolute shrinkage and selection operator (Lasso) method (Tibshirani, 1996) can effectively optimize OLS estimation and deal with the overfitting problem (Candes and Tao, 2007). By adding a penalty term to the model estimation, the regularization method can compress regression coefficients that are too small or zero at the cost of some estimation bias to obtain higher model prediction accuracy and model generalization capability.

Compared to the OLS estimation mentioned above, the Lasso method adds $\lambda(\sum_{j=1}^{p}|\beta_j|)$ as a penalty term (also called the regularization function):

$$\hat{\beta} = arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 + \lambda\left(\sum_{j=1}^{p}\left|\beta_j\right|\right)\right\} \quad (4)$$

where $\sum_{j=1}^{p}|\beta_j|$ is the $L_1$ norm of $\beta$ ($L_q$ norm is $\sum_{j=1}^{p}|\beta_j|^q$). $\lambda(\sum_{j=1}^{p}|\beta_j|)$ denotes the penalty function and $\lambda$ is the tuning parameter used to control the severity of punishment. Cross-validation (CV), adaptive, and plugin are usually used to determine the value of $\lambda$, that is, to determine the $\lambda$ that minimizes the mean square error (MSE) of the sample. When $\lambda = 0$, the loss function does not penalize the model. $\hat{\beta}$ is the loss function of the OLS model.

In addition, there is another use of Lasso, called postselection. In other words, Lasso regression was used to screen independent variables, and then the model without penalty was used for regression. The results of this method can be used as a reference for comparing the results of various models and can also be used for variable screening. However, because too many variables were eliminated when Lasso screened variables, Zou and Hastie (2005) proposed elastic net regression, which can be represented as the following **Eq. 5**:

$$\hat{\beta} = arg\min_{\beta}\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - x_i'\beta\right)^2 + \lambda\left[\left(\alpha\sum_{j=1}^{p}\left|\beta_j\right| + \frac{(1-\alpha)}{2}\right)\times\sum_{j=1}^{p}\beta_j^2\right]\right\} \quad (5)$$

where $\alpha$ is the penalty coefficient of the elastic net. $\alpha = 0$ is ridge regression, and $\alpha = 1$ is Lasso regression.

## Threshold Model

The threshold effect refers to the phenomenon in which when one parameter reaches a certain value or range, its effect on another parameter is reversed. This threshold value is the threshold value referred to in the threshold model. Compared with the linear model, the threshold model can explore the relationship between explanatory variables and explained variables more accurately. Threshold regression eliminates the interference of subjective factors of data, and it can also complete a significance test of the threshold value when calculating the threshold value of the sample data. Its model form is as follows:

$$\begin{cases} y_{it} = u_i + \lambda' x_{it}' + \beta_1' x_{it} + \varepsilon_{it}, \text{if} q_{it} \le \gamma \\ y_{it} = u_i + \lambda' x_{it}' + \beta_2' x_{it} + \varepsilon_{it}, \text{if} q_{it} > \gamma \end{cases} \quad (6)$$

where $y_{it}$ is the dependent variable, $x_{it}$ is an independent variable, $q_{it}$ is a threshold variable, $\gamma$ is the threshold value, and $\varepsilon_{it}$ is the interference term. The individual intercept term $u_i$ represents the fixed-effect mode. Using the indicative function $I(\cdot)$, the model can be simplified as follows:

$$y_{it} = u_i + \lambda' x_{it}' + \beta_1' x_{it}\cdot I(q_{it} \le \gamma) + \beta_2' x_{it}\cdot I(q_{it} > \gamma) + \varepsilon_{it} \quad (7)$$

where $x_{it}'$ indicates other dependent variables. $\beta_1'$ and $\beta_2'$ represent the influence of $x_{it}$ on $y_{it}$ of the threshold variables under $q_{it} \le \gamma$ and $q_{it} > \gamma$, respectively.

Double threshold panel regression model:

$$y_{it} = u_i + \lambda' x_{it}' + \alpha_1' x_{it}\cdot I(q_{it} \le \gamma_1) + \alpha_2' x_{it}\cdot I(\gamma_1 < q_{it} \le \gamma_2) + \alpha_3' x_{it}\cdot I(q_{it} > \gamma_2) + \varepsilon_{it} \quad (8)$$

## Mediation Model

The mediating effect is an important statistical method. The influence of independent variable $X$ on dependent variable $Y$ indirectly acts on the dependent variable $Y$ through intermediary variable $M$. This type of effect is an intermediary effect. According to Baron and Kenny (1986), the following model can be constructed to describe the relationship between the variables:

$$Y = i_1 + c'X + e_2 \quad (9)$$

$$M = i_2 + aX + e_1 \quad (10)$$

$$Y = i_3 + cX + bM + e_3 \quad (11)$$

where $M$ is an intermediate variable, $Y$ is the dependent variable, and $X$ is an independent variable. Baron and Kenny (1986) recommend testing using **Eqs 9–11**. Further, the Sobel Z-test used to test the indirect path of $a \times b$ can be represented as:

$$z = \frac{a \times b}{\sqrt{b^2 s_a^2 + a^2 s_b^2}} \quad (12)$$

where $s_a^2$ and $s_b^2$ are the squared standard errors of $a$ and $b$, respectively.

However, Zhao et al. (2010) criticized the above test method and suggested that Baron and Kenny's "three tests and Sobel" could be replaced with one test: a Bootstrap test of indirect effect

**TABLE 2 |** Regression results of the effect of environmental regulations on GTFP.

| Variables | GTFP | GTFP |
|---|---|---|
| IER | –0.018[b] | |
| | (–2.20) | |
| MER | | –0.012 |
| | | (–0.08) |
| PGDP | –0.002 | –0.003 |
| | (–0.43) | (–0.76) |
| INS | 0.005[c] | 0.004[c] |
| | (4.29) | (3.96) |
| INN | –0.007 | –0.007 |
| | (–0.70) | (–0.64) |
| FDI | 0.003 | 0.003 |
| | (1.22) | (1.19) |
| GOV | 0.013[c] | 0.013[c] |
| | (2.66) | (2.61) |
| POL | –0.039 | –0.049 |
| | (–0.71) | (–0.84) |
| _cons | 0.835[c] | 0.831[c] |
| | (12.98) | (12.80) |
| $R^2$ | 0.174 | 0.160 |
| N | 330 | 330 |

[a]*Note:Indicates significance at the 10% level.*
[b]*Indicates significance at 5% level.*
[c]*Indicates significance at 1% level. The t-values are in parentheses.*

$a \times b$. Thus, we adopted the Bootstrap test to test the mediating effect of environmental regulations on GTFP.

# RESULTS AND DISCUSSION

## Empirical Results

To verify the effects of different types of environmental regulations on GTFP, we examined the effects of command-based environmental regulations and market-based environmental regulations on GTFP using data from 30 provinces in China as a sample. The following panel linear regression model was constructed.

$$GTFP_{\mathrm{it}} = \alpha_0 + \alpha_1 ER_{\mathrm{it}} + \alpha_2 \mathrm{Control}_{\mathrm{it}} + \xi_{it} \quad (13)$$

where $GTFP_{it}$ denotes the GTFP of province $i$ in year $t$. Further, $ER_{it}$ denotes the environmental regulations of province $i$ in year $t$, including $IER_{it}$ and $MER_{it}$, which denote the command-based environmental regulations and market-based environmental regulations of province $i$ in year $t$, respectively. Control variables include economic development level (PGDP), industrial structure (INS), intensity of research and development (INN), foreign direct investment (FDI), government intervention (GOV), and pollution control investment (POL). $\xi_{it}$ is the error term. We used a panel model to regress **Eq. 13**, and the results are listed in **Table 2**.

After the Hausman test, we chose a fixed-effects panel model. The empirical evidence shows that command-based environmental regulations have a negative effect on GTFP, which is significant at the 5% significance level. We believe that command-based environmental regulations increase firm costs, which is counterproductive to improving GTFP. Market-based environmental regulations have no significant impact on GTFP. Most studies show that market-based environmental regulation provides enterprises with increased autonomy and is more conducive to stimulate the green technology innovation of enterprises (Peng et al., 2021). However, there is a time lag in the response of enterprises to market-based environmental regulation (Liu et al., 2021). In addition, the basis for the market-based environmental regulation is a sound market economy system (Zhang, 2021). China's incomplete and inadequate market economy system results in no significant effect of market-based environmental regulations on GTFP. Among the control variables, industrial structure and government intervention can significantly contribute to the development of GTFP, while other control variables have no significant effect on GTFP.

## Machine Learning

Econometric models have concluded that command-based environmental regulations are negatively associated with GTFP, while market-based environmental regulations have no significant effect on GTFP. Furthermore, the above empirical results were validated by Lasso regression in machine learning.

### Selection of the Optimal $\lambda$

Lasso regression is usually used to determine the value of the tuning parameter $\lambda$, that is, to find the $\lambda$ that minimizes the out-of-sample MSE. We selected the optimal $\lambda$ based on the above four methods of CV, adaptive, plugin, and elastic network regression. It is worth noting that the default of elastic network regression is still a CV test. Initially, we compared the MSEs of the above four methods to select the optimal model.

As seen in **Table 3**, when the core variable was IER, the elastic network regression had the smallest MSE; therefore, the optimal model should be the elastic network model. When the core variable was MER, the adaptive method of Lasso regression had the smallest MSE. Therefore, the optimal model should be the adaptive method of Lasso regression. Based on the elastic network model and adaptive method of Lasso regression, we selected the optimal $\lambda$. The penalty coefficients $\alpha$ of the elastic net were 0.25, 0.5, and 0.75.

According to **Table 4**, the core variable was IER, and when the penalty coefficients of elastic net were $\alpha = 0.75$ and $\alpha = 0.5$, seven variables were included. There were two final values of $\lambda$, but neither was optimal. When $\alpha = 0.25$, the optimal $\lambda$ value was $\lambda = 0.0005494$. The CV mean prediction error was 0.004464, which was the minimum value. When the core variable was MER, Lasso regression selected the optimal $\lambda = 0.0012456$. At this time, the number of variables was 7, implying that some variables were removed. Moreover, the CV mean prediction error was 0.0044455.

**Figures 3**, **4** indicate that $\lambda_{cv}$ is the minimum $\lambda$ value under the elastic network regression cross test and Lasso regression. The optimal $\lambda$ values were 0.0005494 and 0.0012456. The process of $\lambda$ selection in the elastic network model and Lasso regression is shown in **Table 5**. As the optimal $\lambda$ occurred when $\alpha = 0.25$, we only showed the results of $\alpha = 0.25$ in the elastic network regression.

As the optimal $\lambda$ value had been obtained previously, the selection process for different $\lambda$ values is further examined comprehensively in **Table 5**. When the core variable was IER, $\alpha = 0.25$, ID = 276, and $\lambda =$

**TABLE 3 |** MSE of different methods.

| Type | Name | MSE | | R-squared | | obs | |
|---|---|---|---|---|---|---|---|
| | | IER | MER | IER | MER | IER | MER |
| OLS | ols | 0.004873 | 0.0050495 | 0.0557 | 0.0215 | 82 | 82 |
| Lasso | CV | 0.0048688 | 0.0050087 | 0.0565 | 0.0294 | 82 | 82 |
| | adaptive | 0.004877 | 0.0049676 | 0.0549 | 0.0374 | 82 | 82 |
| | plugin | 0.0051958 | 0.0051958 | −0.0069 | −0.0069 | 82 | 82 |
| Elastic network regression | enet | 0.0048664 | 0.0050024 | 0.0570 | 0.0306 | 82 | 82 |

**TABLE 4 |** Optimal $\lambda$ value selection of elastic net regression.

| Core variable | | | | IER | | |
|---|---|---|---|---|---|---|
| Alpha | ID | Description | Lambda | No. of nonzero coef. | Out-of-sample R-squared | CV mean prediction error |
| 0.750 | 1 | First lambda | 0.0469402 | 0 | 0.0144 | 0.0049279 |
| | 113 | Last lambda | 1.56e-06 | 7 | 0.0811 | 0.0044643 |
| 0.500 | 114 | First lambda | 0.0469402 | 0 | 0.0144 | 0.0049279 |
| | 226 | Last lambda | 1.56e-06 | 7 | 0.0811 | 0.0044643 |
| 0.250 | 227 | First lambda | 0.0469402 | 0 | 0.0164 | 0.0049376 |
| | 275 | Lambda before | 0.000603 | 7 | 0.0811 | 0.004464 |
| | *276 | Selected lambda | 0.0005494 | 7 | 0.0811 | 0.004464 |
| | 277 | Lambda after | 0.0005006 | 7 | 0.0811 | 0.004464 |
| | 339 | Last lambda | 1.56e-06 | 7 | 0.0811 | 0.0044643 |

| Core variable | | | MER | | |
|---|---|---|---|---|---|
| ID | Description | Lambda | No. of nonzero coef. | Out-of-sample R-squared | CV mean prediction error |
| 61 | First lambda | 0.0294522 | 0 | 0.0143 | 0.0049275 |
| 94 | Lambda before | 0.0013671 | 6 | 0.0849 | 0.0044457 |
| *95 | Selected lambda | 0.0012456 | 6 | 0.0849 | 0.0044455 |
| 96 | Lambda after | 0.0011349 | 6 | 0.0849 | 0.0044457 |
| 131 | Last lambda | 0.0000437 | 7 | 0.0831 | 0.0044544 |

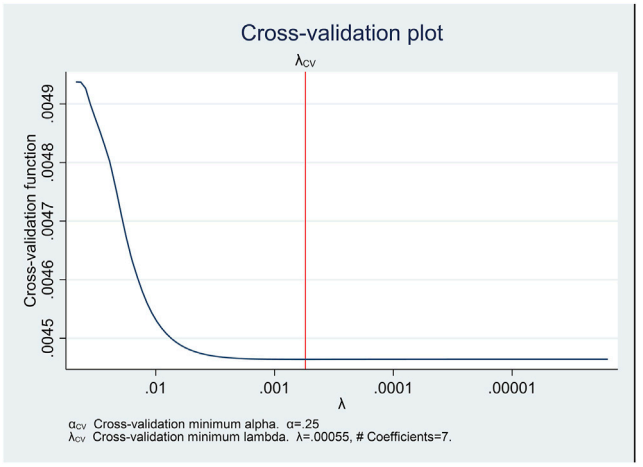



**FIGURE 3 |** Cross-validation plot (elastic net regression).

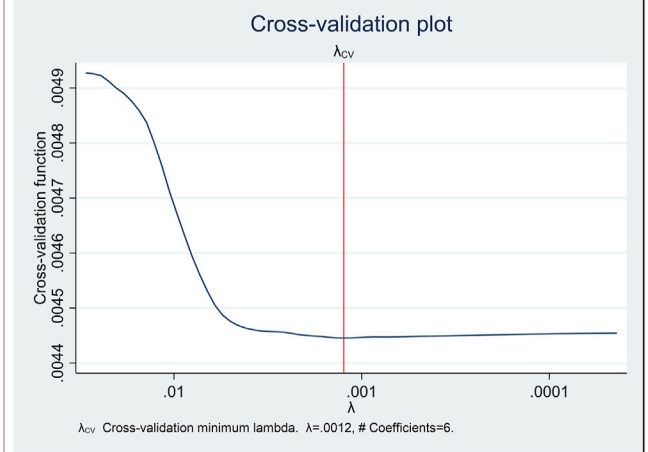



**FIGURE 4 |** Cross-validation plot (adaptive method of Lasso regression).

0.0005494, all seven variables were entered. At this time, the minimum CV mean prediction error was 0.004464, and $\lambda$ = 0.0005494 was also the optimal $\lambda$. When the core variable was MER, ID = 95, $\lambda$ = 0.0012456 and the number of variables was 6. At this time, CV mean prediction error reached the minimum of 0.0044455, and $\lambda$ is also the optimal $\lambda$. However, when ID = 98, the last variable MER entered, but $\lambda$ was not the optimal value, and CV mean prediction error was bigger than that of ID = 95. This indicates

**TABLE 5 |** Process of $\lambda$ value selection.

| Core variable | | | IER | | |
|---|---|---|---|---|---|
| $\alpha$ | ID | $\lambda$ | No. of nonzero coef. | CV mean prediction error | Variables (A)dded, (R)emoved, or left (U)nchanged |
| 0.250 | 228 | 0.0427701 | 2 | 0.0049372 | A INS FDI |
| | 229 | 0.0389705 | 3 | 0.0049267 | A POL |
| | 232 | 0.0294798 | 4 | 0.004853 | A PGDP |
| | 233 | 0.0268609 | 5 | 0.0048284 | A GOV |
| | 234 | 0.0244746 | 6 | 0.0048023 | A INN |
| | 235 | 0.0234701 | 7 | 0.0047866 | A IER |
| | *276 | 0.0005494 | 7 | 0.004464 | U |
| | 339 | 1.56e-06 | 7 | 0.0044643 | U |
| **Core variable** | | | **MER** | | |
| **ID** | | $\lambda$ | **No. of nonzero coef.** | **CV mean prediction error** | **Variables (A)dded, (R)emoved, or left (U)nchanged** |
| 62 | | 0.0268358 | 1 | 0.0049258 | A INS |
| 66 | | 0.0184969 | 2 | 0.0048902 | A FDI |
| 69 | | 0.0139922 | 3 | 0.0048375 | A INN |
| 71 | | 0.0116166 | 4 | 0.0047593 | A PGDP |
| 74 | | 0.0087875 | 5 | 0.0046323 | A POL |
| 83 | | 0.0038039 | 6 | 0.0044607 | A GOV |
| *95 | | 0.0012456 | 6 | 0.0044455 | U |
| 98 | | 0.0009423 | 7 | 0.0044471 | A MER |
| 131 | | 0.0000437 | 7 | 0.0044544 | U |

**TABLE 6 |** Results of Lasso regression.

| | GTFP | GTFP |
|---|---|---|
| IER | –0.011[a] | |
| | (–1.77) | |
| MER | | –0.021 |
| | | (–0.20) |
| PGDP | –0.004[b] | –0.004[b] |
| | (–2.41) | (–2.46) |
| INS | 0.002[c] | 0.002[c] |
| | (4.36) | (4.16) |
| INN | 0.010[c] | 0.009[c] |
| | (3.46) | (3.33) |
| FDI | 0.003[c] | 0.003[c] |
| | (3.50) | (4.11) |
| GOV | 0.008[b] | 0.008[a] |
| | (2.01) | (1.85) |
| POL | –0.032 | –0.061 |
| | (–0.71) | (–1.45) |
| N | 330 | 330 |

[a]*Note:Indicates significance at the 10% level.*
[b]*Indicates significance at 5% level.*
[c]*Indicates significance at 1% level. The t-values are in parentheses.*

that when $\lambda$ reaches the optimal value, variable MER is not included. We further performed sensitivity analysis.

Finally, a sensitivity analysis was conducted according to the results in **Table 5**. The tuning parameter $\lambda'$ (ID = 277/ID = 94) near the optimal $\lambda$ value was manually selected, and the result was denoted as "Hand." The results are shown in **Supplementary Table S2** that indicated that the MSEs of the Hand was larger. Hence, $\lambda = 0.0005494$ and $\lambda = 0.0012456$ were the optimal $\lambda$.

### Causal Inference

It has been concluded that the optimal $\lambda = 0.0005494$ when the core explanatory variable includes command-based environmental regulations, and the optimal $\lambda = 0.0012456$ when the core explanatory variable includes market-based environmental regulations. In this section, we explore the causal relationship between the core explanatory variables and GTFP. In this study, cross-fit partialing out Lasso linear regression was used to conduct a Lasso regression for causal inference.

Lasso regression results are shown in **Table 6** and indicate that the regression coefficient of command-based environmental regulations on GTFP is negative and significant; that is, command-based environmental regulations cannot promote GTFP. The regression coefficient of market-based environmental regulations on GTFP is negative but not significant, that is, market-based environmental regulations have no impact on GTFP. These results are consistent with those of the econometric model.

## Threshold Effect Model

Most of the previous studies have used linear models to study the effects of environmental regulations on GTFP. However, if the relationship between environmental regulations and GTFP is nonlinear, the former results are biased. We further analyzed the nonlinear relationships between environmental regulations and GTFP using a threshold panel model with FDI as the threshold variable.

The threshold panel model was set as given below. It is assumed that for a specific threshold $\xi$, the effect of environmental regulations on GTFP differs significantly when FDI $\leq \xi$ and FDI $> \xi$.

**TABLE 7 |** Threshold estimation results of environmental regulations on GTFP.

| Variables | GTFP | GTFP |
|---|---|---|
| PGDP | −0.006[a] | −0.006[a] |
| | (−1.89) | (−1.85) |
| INS | 0.004[b] | 0.004[b] |
| | (4.02) | (3.84) |
| INN | 0.001 | 0.001 |
| | (0.17) | (0.12) |
| GOV | 0.008[a] | 0.008[a] |
| | (1.88) | (1.72) |
| POL | 0.007 | −0.025 |
| | (0.14) | (−0.45) |
| IER_1 (FDI ≤ 1.2492) | −0.219[c] | |
| | (−2.54) | |
| IER_2 (1.2492 < FDI < 1.588) | 0.043[b] | |
| | (3.17) | |
| IER_3 (FDI ≥ 1.588) | −0.009 | |
| | (−1.03) | |
| MER_1 (FDI ≤ 1.2492) | | −0.248 |
| | | (−1.00) |
| MER_2 (1.2492 < FDI < 1.588) | | 1.756[b] |
| | | (4.60) |
| MER_3 (FDI ≥ 1.588) | | 0.141 |
| | | (0.92) |
| _cons | 0.881[b] | 0.877[b] |
| | (15.83) | (15.80) |
| $R^2$ | 0.253 | 0.082 |
| N | 330 | 330 |

[a]*Note:Indicates significance at the 10% level.*
[b]*Indicates significance at 1% level. The t-values are in parentheses.*
[c]*Indicates significance at 5% level.*

**TABLE 8 |** Empirical results of heterogeneity test in eastern China.

| Variables | GTFP | GTFP | GTFP |
|---|---|---|---|
| IER | −0.017 | | |
| | (−1.10) | | |
| MER | | 0.284 | |
| | | (0.37) | |
| PGDP | 0.013[a] | 0.013[a] | 0.007 |
| | (1.89) | (1.78) | (0.99) |
| INS | 0.010[b] | 0.009[b] | 0.006[c] |
| | (3.79) | (3.63) | (2.46) |
| INN | 0.003 | 0.002 | 0.013 |
| | (0.21) | (0.12) | (0.89) |
| FDI | 0.004 | 0.004 | |
| | (1.44) | (1.27) | |
| GOV | 0.000 | −0.003 | −0.008 |
| | (0.03) | (−0.31) | (−0.83) |
| POL | −0.101 | −0.089 | −0.118 |
| | (−0.68) | (−0.59) | (−0.83) |
| MER_1 (FDI ≤ 11.1268) | | | −0.078 |
| | | | (−0.10) |
| MER_2 (11.1268 < FDI < 11.2256) | | | −4.003[b] |
| | | | (−2.78) |
| MER_3 (FDI ≥ 11.2256) | | | 0.736 |
| | | | (0.96) |
| _cons | 0.531[b] | 0.559[b] | 0.764[b] |
| | (3.51) | (3.71) | (5.24) |
| $R^2$ | 0.178 | 0.170 | 0.128 |
| N | 121 | 121 | 121 |

[a]*Note:Indicates significance at the 10% level.*
[b]*Indicates significance at 1% level. The t-values are in parentheses.*
[c]*Indicates significance at 5% level.*

$$GTFP_{it} = \alpha_1 + \beta_{11} ER_{it} \times I(FDI_{it} \le \zeta) + \beta_{12} ER_{it} \times I(FDI_{it} > \zeta) + \gamma_2 Control_{it} + \xi_{it} \quad (14)$$

where $I(\cdot)$ are indicative functions. FDI is a threshold variable. $\beta_{11}$ and $\beta_{12}$ denote the coefficients of the effect of environmental regulations on GTFP for FDI ≤ $\xi$ and FDI > $\xi$, respectively.

First, we conducted a self-sampling test on the threshold effect of environmental regulations on the GTFP. The corresponding results are shown in **Supplementary Table S3** that revealed that two types of environmental regulations had a double threshold effect on GTFP, with FDI as the threshold variable. The first and second threshold values were 1.2492 and 1.5880, respectively.

The regression results of the panel threshold model are presented in **Table 7**. When FDI is less than 1.2492, command-based environmental regulations are significantly unfavorable to the increase in GTFP. When FDI is between 1.2492 and 1.588, it promotes GTFP. When FDI is greater than 1.588, there is no significant influence on GTFP. Similarly, when FDI intensity is between 1.2492 and 1.588, market-based environmental regulations significantly promote GTFP. Further, at other FDI intensities, market-based environmental regulations have no significant impact on GTFP. Therefore, to enhance the promoting effect of environmental regulations on GTFP, FDI intensity must be controlled between 1.2492 and 1.588.

## Robustness Test

First, we added the control variable human capital level (the natural logarithm of the average number of institutions of higher learning per 100,000 population in each province). Second, to overcome the influence of the outliers and non-randomness of GTFP on the model estimation results, we eliminated the 1% maximum and minimum values of GTFP. The results are shown in **Supplementary Table S4**.

The robustness results showed that command-based environmental regulations were still significantly unfavorable to the improvement of GTFP after the addition of control variable education, while market-based environmental regulations had no significant impact on GTFP. In addition, the impact of environmental regulations on GTFP had a double threshold effect, and the optimal FDI intensity should be between 1.2492 and 1.5880, so that the two types of environmental regulations could promote GTFP. After screening the outliers of GTFP, we can obtain the same results. The robustness of the results is demonstrated in **Supplementary Table S4**.

## Regional Heterogeneity Analysis

Environmental regulations as a policy will have a time lag. In addition, traditional panel models may suffer from problems, such as endogeneity of variables, different initial conditions, and the development of each province. Therefore, China's provinces were divided into eastern, central, and western regions for

**TABLE 9 |** Empirical results of heterogeneity test in central China.

| | GTFP | GTFP | GTFP | GTFP |
|---|---|---|---|---|
| IER | 0.006 | | | |
| | (0.43) | | | |
| MER | | −0.017 | | |
| | | (−0.10) | | |
| PGDP | 0.008 | 0.008 | 0.004 | 0.013 |
| | (0.77) | (0.72) | (0.54) | (1.37) |
| INS | 0.002[a] | 0.002[b] | 0.002[a] | 0.003[a] |
| | (2.01) | (1.94) | (2.13) | (2.57) |
| INN | −0.005 | −0.010 | 0.057 | 0.063 |
| | (−0.05) | (−0.10) | (0.61) | (0.65) |
| FDI | 0.017[a] | 0.017[a] | | |
| | (2.24) | (2.19) | | |
| GOV | 0.025[c] | 0.025[a] | 0.023[a] | 0.030[c] |
| | (2.65) | (2.60) | (2.63) | (3.18) |
| POL | −0.072 | −0.067 | −0.044 | −0.094 |
| | (−0.73) | (−0.52) | (−0.46) | (−0.75) |
| IER_1 (FDI ≤ 3.6868) | | | −0.010 | |
| | | | (−0.78) | |
| IER_2 (FDI > 3.6868) | | | 0.029[b] | |
| | | | (1.82) | |
| MER_1 (FDI ≤ 3.6868) | | | | 0.035 |
| | | | | (0.21) |
| MER_2 (FDI > 3.6868) | | | | 1.432[c] |
| | | | | (3.14) |
| _cons | 0.782[c] | 0.794[c] | 0.844[c] | 0.739[c] |
| | (7.89) | (8.31) | (11.19) | (8.16) |
| $R^2$ | 0.301 | 0.299 | 0.051 | 0.124 |
| N | 88 | 88 | 88 | 88 |

[a]*Note:Indicates significance at 5% level.*
[b]*Indicates significance at the 10% level.*
[c]*Indicates significance at 1% level. The t-values are in parentheses.*

**TABLE 10 |** Empirical results of heterogeneity test in western China.

| | GTFP | GTFP | GTFP | GTFP |
|---|---|---|---|---|
| IER | −0.023[b] | | | |
| | (−2.01) | | | |
| MER | | 0.186 | | |
| | | (0.58) | | |
| PGDP | −0.011[b] | −0.015[b] | −0.011 | −0.016 |
| | (−1.97) | (−2.52) | (−1.01) | (−1.35) |
| INS | 0.004[c] | 0.004[c] | 0.005[b] | 0.005[c] |
| | (2.62) | (2.70) | (2.16) | (2.68) |
| INN | −0.015 | −0.014 | −0.002 | −0.005 |
| | (−1.13) | (−1.05) | (−0.20) | (−0.51) |
| FDI | 0.003 | 0.004 | | |
| | (0.86) | (1.10) | | |
| GOV | 0.005 | 0.002 | 0.001 | −0.008 |
| | (0.91) | (0.34) | (0.14) | (−0.81) |
| POL | 0.010 | −0.071 | −0.066 | −0.030 |
| | (0.14) | (−1.03) | (−0.81) | (−0.44) |
| IER_1 (FDI ≤ 1.2098) | | | −0.023 | |
| | | | (−1.56) | |
| IER_2 (1.2098 < FDI < 1.2404) | | | −0.067[b] | |
| | | | (−2.43) | |
| IER_3 (FDI ≥ 1.2404) | | | 0.003 | |
| | | | (0.15) | |
| MER_1 (FDI ≤ 1.3950) | | | | 0.094 |
| | | | | (0.21) |
| MER_2 (1.3950 < FDI < 1.5880) | | | | 2.853[c] |
| | | | | (4.23) |
| MER_3 (FDI ≥ 1.5880) | | | | 0.459 |
| | | | | (0.89) |
| _cons | 0.919[c] | 0.886[c] | 0.892[c] | 0.886[c] |
| | (15.02) | (13.04) | (6.83) | (6.84) |
| $R^2$ | 0.385 | 0.183 | 0.186 | 0.367 |
| N | 121 | 121 | 121 | 112 |

[a]*Note:Indicates significance at the 10% level.*
[b]*Indicates significance at 5% level.*
[c]*Indicates significance at 1% level. The t-values are in parentheses.*

analysis. The linear and nonlinear effects of environmental regulations in different regions on GTFP were tested using the econometric model, as shown in **Tables 8–11**.

**Table 8** shows that environmental regulations in eastern China had no significant impact on GTFP. Moreover, there was no threshold effect of command-based environmental regulations on GTFP in eastern China, but there was a threshold effect of market-based environmental regulations on GTFP with FDI as the threshold variable. When FDI intensity was between 11.1268 and 11.2256, market-based environmental regulations had a significant negative effect on GTFP.

As shown in **Table 9**, the environmental regulations in central China had no significant impact on GTFP. However, there is a threshold effect of FDI as the threshold variable. When the FDI intensity was greater than 3.6868, both types of environmental regulation had a significant promoting effect on GTFP. Therefore, it is necessary to increase the intensity of FDI in the central region.

**Table 10** showed that command-based environmental regulations in western China were significantly detrimental to the growth of GTFP, while market-based environmental regulations had no significant impact on GTFP. In addition, through the threshold test, it was found that when FDI intensity was between 1.2098 and 1.2404, command-based environmental regulations had a significant hindrance effect on GTFP. When FDI intensity was between 1.3950 and 1.5880, market-based environmental regulations could significantly improve GTFP. Therefore, the optimal FDI intensity in western China should be between 1.3950 and 1.5880.

## Intermediation Effect Test

In the previous econometric model and machine learning validation, command-based environmental regulations had a significant effect on GTFP. Furthermore, we analyzed the mechanism of command-based environmental regulations on GTFP. The following mediation model was developed:

$$GTFP_{it} = i_1 + c'IER_{it} + d_1 Control_{it} + \mu_{it} \quad (15)$$

$$FDI_{it} = i_2 + aIER_{it} + d_2 Control_{it} + \mu_{it} \quad (16)$$

$$GTFP_{it} = i_3 + cIER_{it} + bFDI_{it} + d_3 Control_{it} + \mu_{it} \quad (17)$$

where $t$ represents time, $i$ represents the province, and $\mu_{it}$ is a random disturbance term. The intermediate effect model estimation results are presented in **Table 11**.

Since the Sobel test is prone to Type I errors in the hypothesis test (Zhao et al., 2010), the Bootstrap method was used for the mediation test. The test results are presented in **Table 11**. The result of the bootstrap test was negative and significant at the 1% significance level, implying

**TABLE 11 |** Test of the mediating effect of command-based environmental regulations on GTFP.

| Variables | GTFP | FDI | GTFP |
|---|---|---|---|
| IER | –0.018[a] | | –0.011[b] |
| | (–2.18) | | (–1.82) |
| FDI | | | 0.003[c] |
| | | | (3.12) |
| PGDP | –0.005 | 0.332[c] | –0.004[a] |
| | (–1.64) | (3.09) | (–2.27) |
| INS | 0.004[c] | 0.079[a] | 0.002[c] |
| | (4.24) | (2.37) | (3.75) |
| GOV | 0.011[a] | –0.683[c] | 0.008[c] |
| | (2.39) | (–3.84) | (2.82) |
| POL | –0.033 | –3.486 | –0.032 |
| | (–0.61) | (–1.23) | (–0.69) |
| INN | 0.000 | 1.072[c] | 0.010[c] |
| | (0.06) | (5.68) | (3.01) |
| _cons | 0.869[c] | 5.095[c] | 0.953[c] |
| | (14.94) | (2.93) | (33.14) |
| Bootstrap test | | –0.005[c] (–2.89) | |
| $R^2$ | 0.169 | 0.316 | 0.139 |
| N | 330 | 330 | 330 |

[a]*Note:Indicates significance at 5% level.*
[b]*Indicates significance at the 10% level.*
[c]*Indicates significance at 1% level. The t-values are in parentheses.*

that a negative mediating effect existed. The results indicated that command-based environmental regulations were significantly detrimental to the increase in FDI, which in turn, increased GTFP. Therefore, command-based environmental regulations are detrimental to the increase in GTFP by reducing FDI.

## Discussion

First, we studied the effects of environmental regulations on GTFP using econometric analysis and machine learning methods, such as panel models, machine learning Lasso regression. The results confirmed that the current command-based environmental regulations are significantly detrimental to GTFP, and market-based environmental regulations have no significant effect on GTFP. However, these results were not consistent with those by Zhang (2021), who argued that both types of environmental regulations promote green productivity. However, command-based environmental regulations can be effective but are often less efficient (Peng et al., 2021). Further, there is a time lag in the response of enterprises to market-based environmental regulation (Liu et al., 2021) and China's market system does not function well, which will affect the impact of market-based environmental regulations. We then used machine learning Lasso regression to confirm the previous findings and derived the optimal $\lambda$ = 0.0005494 and $\lambda$ = 0.0012456, which indicate the breakthrough of machine learning applications in estimating the effect of environmental regulation policy.

Second, we observed that when the intensity of FDI ranged between 1.2492 and 1.5880, both types of environmental regulations significantly affected GTFP. According to an analysis of regional heterogeneity, environmental regulations in eastern China have no significant impact on GTFP. Central China's environmental regulations have a significant promoting effect on GTFP when the FDI intensity is greater than 3.6868. In western China, when the FDI intensity is between 1.2098 and 1.2404, command-based environmental regulations significantly hinder GTFP. Further, when the FDI intensity ranges between 1.3950 and 1.5880, market-based environmental regulations can significantly improve GTFP. Therefore, the optimal FDI intensity in western China should range between 1.3950 and 1.5880. Moreover, command-based environmental regulations are detrimental to the increase in GTFP by reducing FDI. This is consistent with the study of Qiu et al. (2021), who reported that FDI reduces GTFP. This provides a theoretical basis for formulating environmental regulation policies according to local conditions in China.

However, our research can be improved further. First, while selecting samples, we can consider the data of other Chinese cities. However, corresponding macro data are lacking in China. Second, we only considered the impacts of FDI on the effects of environmental regulations on GTFP. Other influencing factors can be further considered in future studies.

## CONCLUSION

As environmental regulations have become the main factors of environmental governance in China, whether environmental regulation can achieve a "win-win" between environmental conservation and economic development has become the focus of attention. Further, whether environmental regulations can promote GTFP is a main indicator to judge the effectiveness of environmental regulations. This is also a significant factor for China's sustainable economic development and high-quality development.

In addition to the traditional econometric models, our study also used Lasso model of machine learning, which is a breakthrough in the research of environmental regulations. Machine learning does not work on the data and model assumptions of traditional econometric models, and can better explore the causal relationship of the acquired data. We observed that command-based environmental regulations are significantly detrimental to GTFP, and market-based environmental regulations have no significant effect on GTFP. Additionally, we found that when the FDI intensity increases between 1.2492 and 1.5880, both types of environmental regulations significantly impact GTFP. There are also regional differences in the effects of FDI on environmental regulations on GTFP. Further, command-based environmental regulations are detrimental to the increase in GTFP by reducing FDI.

The following policy recommendations are proposed: First, the intensity of FDI should be rationally controlled. When the FDI-to-fixed-assets ratio is between 1.2492 and 1.5880, command-based environmental regulations can significantly boost China's GDP. Second, reasonable environmental

regulations should be selected according to regional differences because the impact of environmental regulations on GTFP varies regionally. Moreover, the introduction of FDI should also be accompanied by the establishment of a corresponding system for foreign investment, which should encourage foreign investment with advanced technologies to enter China and prevent polluting enterprises from moving to China and consider China as their "pollution heaven." Finally, the share of environmental protection expenditure in the general government budget should be further increased. Government action is a powerful signal. Increased government spending on environmental protection can release market and policy signals for green production and green consumption patterns, which influence business behavior and consumer behavior, thus, further affecting GTFP.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

JF proposed the research idea, developed the model and wrote the first manuscript draft. JY analyzed and visualized the data. XT wrote, reviewed, and edited the manuscript.

## FUNDING

We are grateful to the National Natural Science Foundation of China (72091511), the Natural Science Foundation of Hunan Province (2020JJ4259), the Scientific research project of Hunan Education Department (20A121) and Hunan Provincial Innovation Foundation For Postgraduate (CX20211117) for providing financial assistance for this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenvs.2021.779358/full#supplementary-material

## REFERENCES

Ai, H., Hu, S., Li, K., and Shao, S. (2020). Environmental Regulation, Total Factor Productivity, and enterprise Duration: Evidence from China. *Bus Strat Env* 29, 2284–2296. doi:10.1002/bse.2502

Barbera, A. J., and Mcconnell, V. D. (1990). The Impact of Environmental Regulations on Industry Productivity: Direct and Indirect Effects. *J. Environ. Econ. Manag.* 18, 50–65. doi:10.1016/0095-0696(90)90051-Y

Baron, R. M., and Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *J. Personal. Soc. Psychol.* 51, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Bitat, A. (2018). Environmental Regulation and Eco-Innovation: the Porter Hypothesis Refined. *Eurasian Bus. Rev.* 8, 299–321. doi:10.1007/s40821-017-0084-6

Candes, E., and Tao, T. (2007). The Dantzig Selector: Statistical Estimation when P Is Much Larger Than N. *Ann. Statist.* 35, 2313–2351. doi:10.1214/009053606000001523

Chatterjee, S., and Hadi, A. S. (2006). *Regression by Example*. 4th Edition. Hoboken: John Wiley & Sons. doi:10.1002/0470055464

Cheng, D., Shi, X., and Yu, J. (2021). The Impact of green Energy Infrastructure on Firm Productivity: Evidence from the Three Gorges Project in China. *Int. Rev. Econ. Finance* 71, 385–406. doi:10.1016/j.iref.2020.09.022

Chung, Y. H., Färe, R., and Grosskopf, S. (1997). Productivity and Undesirable Outputs: A Directional Distance Function Approach. *J. Environ. Manage.* 51, 229–240. doi:10.1006/jema.1997.0146

Dong, J., Xue, G., Dong, M., and Xu, X. (2015). Energy-saving Power Generation Dispatching in China: Regulations, Pilot Projects and Policy Recommendations-A Review. *Renew. Sust. Energ. Rev.* 43, 1285–1300. doi:10.1016/j.rser.2014.11.037

Dou, J., and Han, X. (2019). How Does the Industry Mobility Affect Pollution Industry Transfer in China: Empirical Test on Pollution Haven Hypothesis and Porter Hypothesis. *J. Clean. Prod.* 217, 105–115. doi:10.1016/j.jclepro.2019.01.147

Gray, W. B., and Shadbegian, R. J. (2003). Plant Vintage, Technology, and Environmental Regulation. *J. Environ. Econ. Manag.* 46, 384–402. doi:10.1016/S0095-0696(03)00031-7

Greenstone, M., List, J., and Syverson, C. (2012). The Effects of Environmental Regulation on the Competitiveness of US Manufacturing. Cambridge, United States: NBER working paper No.18392. doi:10.3386/w18392

Helwig, N. E. (2017). Adding Bias to Reduce Variance in Psychological Results: A Tutorial on Penalized Regression. *Tqmp* 13, 1–19. doi:10.20982/tqmp.13.1.p001

Jiang, X., Fu, W., and Li, G. (2020). Can the Improvement of Living Environment Stimulate Urban Innovation?--Analysis of High-Quality Innovative Talents and Foreign Direct Investment Spillover Effect Mechanism. *J. Clean. Prod.* 255, 120212. doi:10.1016/j.jclepro.2020.120212

Lanoie, P., Ambec, S., Cohen, M. A., and Elgie, S. (2013). The porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness? *SSRN J.* 7, 2–22. doi:10.2139/ssrn.1682001

Li, J., Shan, Y., Tian, G., and Hao, X. (2020). Labor Cost, Government Intervention, and Corporate Innovation: Evidence from China. *J. Corporate Finance* 64, 101668. doi:10.1016/j.jcorpfin.2020.101668

Li, Y. F., Jiang, H., Farhad, T. H., and Wei, W. (2021). Review on the Economics of Hydrogen Production Technology from Renewable Energy and its Application Evaluation in ASEAN. *J. Glob. Energ. Interconnection* 4, 292–300. doi:10.19705/j.cnki.issn2096-5125.2021.03.009

Li, Y., and Kimura, S. (2021). Economic Competitiveness and Environmental Implications of Hydrogen Energy and Fuel Cell Electric Vehicles in ASEAN Countries: The Current and Future Scenarios. *Energy Policy* 148, 111980. doi:10.1016/j.enpol.2020.111980

Liao, X., and Shi, X. (2018). Public Appeal, Environmental Regulation and green Investment: Evidence from China. *Energy Policy* 119, 554–562. doi:10.1016/j.enpol.2018.05.020

Liu, L., Jiang, J., Bian, J., Liu, Y., Lin, G., and Yin, Y. (2021). Are Environmental Regulations Holding Back Industrial Growth? Evidence from China. *J. Clean. Prod.* 306, 127007. doi:10.1016/j.jclepro.2021.127007

Long, R., Shao, T., and Chen, H. (2016). Spatial Econometric Analysis of China's Province-Level Industrial Carbon Productivity and its Influencing Factors. *Appl. Energ.* 166, 210–219. doi:10.1016/j.apenergy.2015.09.100

Magazzino, C., Mele, M., and Schneider, N. (2021). A Machine Learning Approach on the Relationship Among Solar and Wind Energy Production, Coal Consumption, GDP, and CO2 Emissions. *Renew. Energ.* 167, 99–115. doi:10.1016/j.renene.2020.11.050

Mele, M., and Magazzino, C. (2020). Pollution, Economic Growth, and COVID-19 Deaths in India: a Machine Learning Evidence. *Environ. Sci. Pollut. Res.* 28, 2669–2677. doi:10.1007/s11356-020-10689-0

Murshed, M., Rahman, M. A., Alam, M. S., Ahmad, P., and Dagar, V. (2021). The Nexus between Environmental Regulations, Economic Growth, and Environmental Sustainability: Linking Environmental Patents to Ecological

Footprint Reduction in South Asia. *Environ. Sci. Pollut. Res.* 28, 49967–49988. doi:10.1007/s11356-021-13381-z

Nathaniel, S. P., Murshed, M., and Bassim, M. (2021a). The Nexus between Economic Growth, Energy Use, International Trade and Ecological Footprints: the Role of Environmental Regulations in N11 Countries. *Energ. Ecol. Environ.* 6, 496–512. doi:10.1007/s40974-020-00205-y

Nathaniel, S. P., Yalçiner, K., and Bekun, F. V. (2021b). Assessing the Environmental Sustainability Corridor: Linking Natural Resources, Renewable Energy, Human Capital, and Ecological Footprint in BRICS. *Resour. Pol.* 70, 101924. doi:10.1016/j.resourpol.2020.101924

Nordhaus, W. D., and Tobin, J. (1972). "Is Growth Obsolete?," in *Economic Research: Retrospect and Prospect*. Editors W. D. Nordhaus and J. Tobin (Cambridge: National Bureau of Economic Research), 1–80.

Payne, J. E. (2012). The Causal Dynamics between US Renewable Energy Consumption, Output, Emissions, and Oil Prices. *Energ. Sourc. B: Econ. Plann. Pol.* 7, 323–330. doi:10.1080/15567249.2011.595248

Peng, J., Xie, R., Ma, C., and Fu, Y. (2021). Market-based Environmental Regulation and Total Factor Productivity: Evidence from Chinese Enterprises. *Econ. Model.* 95, 394–407. doi:10.1016/j.econmod.2020.03.006

Peng, X. (2020). Strategic Interaction of Environmental Regulation and green Productivity Growth in China: Green Innovation or Pollution Refuge? *Sci. Total Environ.* 732, 139200. doi:10.1016/j.scitotenv.2020.139200

Porter, M. E., and Van, D. L. C. (1995). Green and Competitive: Ending the Stalemate. *Harv. Bus. Rev.* 28, 128–129.

Qiu, S., Wang, Z., and Geng, S. (2021). How Do Environmental Regulation and Foreign Investment Behavior Affect green Productivity Growth in the Industrial Sector? an Empirical Test Based on Chinese Provincial Panel Data. *J. Environ. Manage.* 287, 112282. doi:10.1016/j.jenvman.2021.112282

Rexhäuser, S., and Rammer, C. (2014). Environmental Innovations and Firm Profitability: Unmasking the Porter Hypothesis. *Environ. Resource Econ.* 57, 145–167. doi:10.1007/s10640-013-9671-x

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

Ulph, A. (2000). Harmonization and Optimal Environmental Policy in a Federal System with Asymmetric Information. *J. Environ. Econ. Manag.* 39, 224–241. doi:10.1006/jeem.1999.1098

Wang, K.-L., Pang, S.-Q., Ding, L.-L., and Miao, Z. (2020). Combining the Biennial Malmquist-Luenberger index and Panel Quantile Regression to Analyze the green Total Factor Productivity of the Industrial Sector in China. *Sci. Total Environ.* 739 (15), 140280. doi:10.1016/j.scitotenv.2020.140280

Wang, K., Wei, Y.-M., and Huang, Z. (2018). Environmental Efficiency and Abatement Efficiency Measurements of China's thermal Power Industry: A Data Envelopment Analysis Based Materials Balance Approach. *Eur. J. Oper. Res.* 269, 35–50. doi:10.1016/j.ejor.2017.04.053

Wayne, B. G., and Ronald, J. S. (1998). Environmental Regulation, Investment Timing, and Technology Choice. *J. Ind. Econ.* 46, 235–256. doi:10.1111/1467-6451.00070

Wu, W.-p., Zeng, W.-k., Gong, S.-w., and Chen, Z.-g. (2021b). Does Energy Poverty Reduce Rural Labor Wages? Evidence from China's Rural Household Survey. *Front. Energ. Res.* 9, 1–8. doi:10.3389/fenrg.2021.670026

Wu, W.-p., Zhu, Y.-f., Zeng, W.-k., Wang, M., Yang, D.-x., and Chen, W.-f. (2021a). Green Efficiency of Water Resources in Northwest China: Spatial-Temporal Heterogeneity and Convergence Trends. *J. Clean. Prod.* 320, 128651. doi:10.1016/j.jclepro.2021.128651

Yarkoni, T., and Westfall, J. (2017). Choosing Prediction over Explanation in Psychology: Lessons from Machine Learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi:10.1177/1745691617693393

Yu, J., Shi, X., Guo, D., and Yang, L. (2021). Economic Policy Uncertainty (EPU) and Firm Carbon Emissions: Evidence Using a China Provincial EPU index. *Energ. Econ.* 94, 105071. doi:10.1016/j.eneco.2020.105071

Yu, X., and Li, Y. (2020). Effect of Environmental Regulation Policy Tools on the Quality of Foreign Direct Investment: an Empirical Study of China. *J. Clean. Prod.* 270, 122346. doi:10.1016/j.jclepro.2020.122346

Yue, S., Shen, Y., and Yuan, J. (2019). Sustainable Total Factor Productivity Growth for 55 States: An Application of the New Malmquist index Considering Ecological Footprint and Human Development index. *Resour. Conservation Recycling* 146, 475–483. doi:10.1016/j.resconrec.2019.03.035

Zhang, D. (2021). Marketization, Environmental Regulation, and Eco-Friendly Productivity: A Malmquist-Luenberger index for Pollution Emissions of Large Chinese Firms. *J. Asian Econ.* 76, 101342. doi:10.1016/j.asieco.2021.101342

Zhang, N., and Choi, Y. (2013). Total-factor Carbon Emission Performance of Fossil Fuel Power Plants in China: A Metafrontier Non-radial Malmquist index Analysis. *Energ. Econ.* 40, 549–559. doi:10.1016/j.eneco.2013.08.012

Zhang, Q., Xu, L., Wang, K., and Shi, X. (2021). What Effect Did the Green Credit Policy Have on China's Energy or Emission Intensive Firms? *Ijoem* ahead-of-print, 1. doi:10.1108/IJOEM-04-2021-0555

Zhao, X., Lynch, J. G., and Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *J. Consum. Res.* 37, 197–206. doi:10.1086/651257

Zhao, X., Zhao, Y., Zeng, S., and Zhang, S. (2015). Corporate Behavior and Competitiveness: Impact of Environmental Regulation on Chinese Firms. *J. Clean. Prod.* 86, 311–322. doi:10.1016/j.jclepro.2014.08.074

Zhou, P., Ang, B. W., and Wang, H. (2012). Energy and CO2 Emission Performance in Electricity Generation: A Non-radial Directional Distance Function Approach. *Eur. J. Oper. Res.* 221, 625–635. doi:10.1016/j.ejor.2012.04.022

Zou, H., and Hastie, T. (2005). Addendum: Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc B* 67, 768. doi:10.1111/j.1467-9868.2005.00527.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Feng, Yan and Tao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fenvs.2021.774343

# Assessing the Environmental Efficiency of Grain Production and Their Spatial Effects: Case Study of Major Grain Production Areas in China

*Hanxiao Xu, Bei Ma and Qiang Gao**

*College of Management, Ocean University of China, Qingdao, China*

Environmental factors have a profound impact on the evaluation of grain production efficiency. Taking environmental factors into account can more accurately measure grain production efficiency and identify the development stage of grain production. Based on the global super-efficiency SBM model, environmental factors are included in the grain production efficiency measurement system, while the temporal and spatial evolution characteristics of grain production environmental efficiency and the sources of inefficiency are analyzed. In addition, the Spatial Durbin Model is used to investigate the influencing factors and spatial spillover effects of China's grain production environmental efficiency. The results show that: environmental factors have a significant impact on the efficiency of grain production in China; the environmental efficiency of grain production in China fluctuates with a downward trend; the environmental efficiency of grain production in different provinces show obvious spatial differentiation and geographical agglomeration; the main influencing factors of the environmental efficiency loss of grain production in major grain production areas are carbon emissions, non-point source pollution of grain production and labor input; the level of education, the level of technological development, and labor transfer have a significant positive spillover effect on the environmental efficiency of China's grain production, yet disaster rate has a significantly negative one. Thus, it is important to pay attention to resource conservation and pollution control, strengthen cooperation and exchanges between provinces, and take the path of coordinated development between the environment and grain production.

**Keywords: environmental factors, carbon sinks, grain production, spatial spillover effect, the global super-efficiency SBM model**

OPEN ACCESS

***Edited by:***
*Xunpeng Shi,*
*University of Technology Sydney, Australia*

***Reviewed by:***
*Huaping Sun,*
*Jiangsu University, China*
*Xu Zhao,*
*Shandong University, China*

****Correspondence:***
*Qiang Gao*
*gao12211@126.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *11 September 2021*
***Accepted:*** *25 October 2021*
***Published:*** *23 November 2021*

***Citation:***
*Xu H, Ma B and Gao Q (2021) Assessing the Environmental Efficiency of Grain Production and Their Spatial Effects: Case Study of Major Grain Production Areas in China. Front. Environ. Sci. 9:774343. doi: 10.3389/fenvs.2021.774343*

## 1 INTRODUCTION

Food security has always been the foundation of national security, social stability and economic development, and a major strategic issue that concerns the overall situation. In recent years, China's overall grain production capacity has steadily increased, but many problems are worsening, such as waste of resources, water and soil pollution, and ecological damage in the process of grain production. The study pointed out that from 2010 to 2050, due to expected changes in population and income levels, without technological changes and special mitigation measures,

the environmental impact of the grain system may increase by 50–90% (Springmann et al., 2018). In the face of challenges, in 2015, the Ministry of Agriculture of China launched a zero-growth campaign to contain the use of chemical fertilizers and pesticides, vigorously promoting the reduction of chemical fertilizers and efficiency, and the reduction of pesticides to control the harm. The No. One Central Document of 2021 further proposes to "continue to promote the reduction of fertilizers and pesticides and increase efficiency, and to promote green products and technologies for the prevention and control of crop diseases and insect pests." In general, the scope of environmental governance has been expanded and the governance has been continuously strengthened (Liao et al., 2018; Shi et al., 2020). Correspondingly, the grain production environment gets improved. However, there is still a gap in the sustainable development capacity of grain production between China and agricultural developed countries. In order to better carry out the construction of ecological civilization in the new era and implement the development concept of harmonious coexistence between man and nature, China's grain production in the future needs to consider the rigid constraints of resources and the environment and take the road of coordinated development of environmental protection and grain production in order to achieve the high-quality development goal of China's grain industry.

Based on the above-mentioned practical problems, in order to fill the gap of existing research, this paper incorporates environmental indicators such as grain carbon sinks, grain production carbon emissions, and non-point source pollution into the grain production efficiency evaluation system. The global super-efficiency SBM model is adopted to analyze the spatial and temporal evolution of the environmental efficiency of grain production and the source of inefficiency in the main grain production area from 2004 to 2019. The Moran's I index is also used to investigate the spatial agglomeration and diffusion effects of the environmental efficiency of grain production in the main grain production area. On this basis, a spatial Durbin model is constructed to explore the spatial spillover effects of China's grain production environmental efficiency. It aims to provide some reference for the ecological transformation of China's grain production and promote the high-quality development of grain production. In comparison with the existing studies, the possible marginal contributions of this paper are as follows:

1) It has incorporated environmental factors (especially grain carbon sinks) into the grain production efficiency analysis framework and conducted effective calculations, which has a certain corrective effect on grain production efficiency and helps to understand the true performance of grain production efficiency in various provinces.
2) It has constructed the global super-efficiency SBM model that takes undesired output into consideration. Super-efficiency SBM model is superior to other models as it can ① effectively solve the radial and perspective problems; ② fully consider the problem of undesired output; and ③ solve the ranking problem when the efficiency value of the decision unit is 1. Additionally, given that the factor endowments of the main grain production areas are basically stable, and the input-output factors and development objectives are similar between periods, it is possible to compare the efficiency of the same sample across periods. Therefore, the study has constructed the global super-efficiency SBM model based on the Global Malmquist-Luenberger proposed by Pastor and Lovell (2005). The SBM model uses the input and output data of all decision units in the whole sample period to construct the best production Frontier as a unified benchmark, and each decision unit is compared with the best production Frontier to obtain efficiency values, thus solving the problem of inter-period comparability of DEA models and the problem of linearity without feasible solutions due to too many input and output variables.
3) It has examined and identified the source of inefficiency in the environmental efficiency of grain production. At the national and inter provincial levels, it finds a realistic basis for improving the potential sources, enriches the relevant research on China's grain production efficiency, and provides a decision-making reference for deeply implementing the concept of green development and promoting the high-quality development of grain production.
4) It has analyzed the spillover effects of external environmental influences such as the level of education, the level of technological development, the labor transfer and the disaster rate on the environmental efficiency of grain production in China to further explore the improvement path of grain production environmental efficiency in China.

## 2 LITERATURE REVIEW

At present, scholars have conducted abundant studies on grain production efficiency, the perspectives of which mainly focused on the measurement of grain production efficiency, influencing factors and spatial effect studies.

### 2.1 Calculation of Grain Production Efficiency

The measurement of grain production efficiency mainly includes two aspects: first, the research on the measurement method of grain production efficiency; and second, the selection of indicators for grain production efficiency evaluation. In terms of measurement methods, there are mainly ratio method, life cycle method, ecological footprint method, energy analysis method, stochastic Frontier method and data envelopment analysis method (Wilson et al., 1998; Sachdeva et al., 2011). Among them, the parametric method represented by the stochastic Frontier method and the non-parametric method represented by the data envelopment analysis method are most widely used. Compared with SFA, the DEA method ignores the influence of random errors, leaving out the need to set the functional form of the production boundary in advance (Battese and Coelli, 1995). With the advantages of simultaneous treatment of multiple input-output elements and non-parametric processing of effective boundaries, the DEA has become the most

used method for efficiency evaluation and is widely used in energy, transportation, finance and other fields (Xian et al., 2018; Wang et al., 2020; Sun et al., 2021a). Some scholars have also applied the method to the measurement of grain production efficiency, for example, Dolgikh (2019) and Pishgar-Komleh et al. (2020) measured the grain production efficiency in Poland and Ukraine respectively by using the DEA method. In the selection of evaluation indicators, some early scholars took traditional production factors such as land, intermediate factors and labor as input factors (Lin, 1992; Wen, 1993), and most scholars took production factors such as land, water resources, labor, pesticides and chemical fertilizers as input factors (Benedetti et al., 2019; Zhang et al., 2021). In the selection of output factor indicators, scholars generally choose the total grain output as the output index (Yao and Xuan, 2019), while some use the total agricultural output value per unit of cultivated land (Armagan et al., 2010; Kumbhakar et al., 2014) or agricultural added value (Gebregziabher et al., 2012).

## 2.2 Influencing Factors of Grain Production Efficiency

As a key guarantee of national food security and socio-economic sustainability, food grain production efficiency is influenced by many factors (Maxime et al., 2006). Existing studies tend to incorporate external environmental factors that can impact grain production efficiency into the research framework and explore effective ways to improve grain production efficiency based on varying research scales, research perspectives and focuses (Picazo-Tadeo et al., 2011). Among these external factors, the urbanization level, the level of openness to the outside world, the rate of disasters, and the non-rural employment opportunities all have a significant bearing on grain production (Liu et al., 2020; Zou et al., 2020).

## 2.3 Space Effects of Grain Production Efficiency

Early literature mainly described or compared the results of grain production efficiency in different regions (Vlontzos et al., 2014). With the deepening of the research, various spatial analysis methods such as Gini coefficient and Theil index were widely used to investigate the temporal and spatial pattern, regional gap, distribution dynamic evolution trend of efficiency (Pang et al., 2016). In recent years, with the development of spatial econometrics and geosciences, more and more scholars begin to use spatial econometric methods to analyze production efficiency by combining data points with their corresponding spatial locations (Wang et al., 2019; Sun et al., 2021b). This method has extended the traditional regression model from the spatial dimension and compensates for the limitations of instability and spatial correlation of traditional spatial data. For example, Cao et al. (2019) used Exploratory Spatial Data Analysis (ESDA) to analyze the evolutionary characteristics of land use efficiency in China. Yuan et al. (2019) adopted ESDA to analyze the spatial correlation of green total factor productivity in the urban cluster on the Shandong Peninsula.

Through the review of the previous literature, it can be found that the existing literature has conducted detailed research on grain production efficiency, but there is still room for further improvement.

1) Previous research has important theoretical value for guiding the development of China's grain production, but most scholars have not considered the impact of environmental factors on the efficiency of grain production. In recent years, although a few scholars have taken environmental factors into consideration when studying the efficiency of grain production, they only considered undesired output (such as chemical oxygen demand, total nitrogen and total phosphorus) without attaching importance to carbon sink output resulted from the positive externalities of grain production (Bai et al., 2019). However, ignoring the environmental factors of grain production (especially grain carbon sinks) will be detrimental to the accuracy of grain production efficiency assessment.
2) Most of the existing literature adopts the traditional DEA methods to measure the environmental efficiency of grain production, which cannot take account of the effects of "slack variables" on the efficiency value, nor does it consider the change in technical efficiency when the desired output increases and the undesired output decreases, and thus could lead to biased efficiency value. In recent years, although a few scholars have considered using the SBM model to measure grain production efficiency, most of them have not solved the ranking problem when the efficiency value of the decision unit is one in the SBM model. In addition, the DEA method measures relative efficiency, which usually treats the frontiers constructed in each period as independent of each other. As a result, the efficiency values obtained in different periods are not comparable because the frontiers are not at the same level. Beyond that, the construction of frontiers with current decision-making units (DMUs) is likely to violate the rule of thumb due to the large number of input-output variables, resulting in overestimation of efficiency.
3) Most scholars have not included the breaking down of inefficient sources of grain production efficiency and the direction of improvement into the research content, ignored the inbuilt driving factors of grain production efficiency improvement, which is not conducive to the research on the potential of China's grain production efficiency improvement.
4) Although existing studies have provided an important reference value for the spatial perspective of grain production efficiency, most of them have only focused on the analysis of the spatial and temporal divergence of grain production efficiency and the verification of the existence of spatial spillover effects. There are fewer analyses of the spatial spillover effects of various influencing factors on grain production efficiency.

# 3 MODEL CONSTRUCTION AND DATA DESCRIPTION

## 3.1 Research Model and Estimation Method

### 3.1.1 The Global Super-Efficiency SBM Model

Considering the current situation of the grain production process, this paper constructs a global super-efficiency SBM model that

takes the undesired output into account. The grain eco-economic system has DMUs (this paper takes the provinces of the main grain production areas as decision-making units), of which there are m kinds of inputs, $s_1$ kinds of expected outputs, and $s_2$ kinds of undesired outputs. For the $k$th DMUk, its input, expected output, and undesired output vectors are denoted as $x_{jk}^{T}$, $y_{jk}^{g,T}$ and $y_{jk}^{b,T}$ respectively, and they meet the following conditions: $x_j \in R_+^m$, $y_j^{g,t} \in R_+^{s_1,t}$ and $y_j^{b,t} \in R_+^{s_2,t}$. Under the assumption of constant return to scale (CRS), the constructed production possibility set (PPS) is shown in **Eq. 1**; the global super-efficiency SBM model considering undesired output is shown in **Eq. 2**.

$$PPS = \left\{ \left( \bar{x}\ \bar{y}^{-g-b} \right) | \bar{x}^{-T} \geq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{L} \lambda_j^t x_j^t; \bar{y}^{-g,T} \leq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{L} \lambda_j^t y_j^{g,t}; \bar{y}^{-b,T} \geq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{L} \lambda_j^t y_j^{b,t}, \lambda \geq 0 \right\} \tag{1}$$

$$\min \rho = \frac{\frac{1}{m} \sum_{i=1}^{m} x_i^{-T} / x_{ik}}{\frac{1}{s_1 + s_2} \left( \sum_{r=1}^{s_1} y_r^{-g,T} / y_{rk}^{g,T} + \sum_{q=1}^{s_2} y_q^{-b,T} / y_{qk}^{b,T} \right)}$$

$$s.t. \ x_i^{-T} \geq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{N} \lambda_j^t x_j^t; y_r^{-g,T} \leq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{N} \lambda_j^t y_j^{g,t}; y_q^{-b,T} \geq \sum_{t=1}^{T} \sum_{j=1,j\neq k}^{N} \lambda_j^t y_j^{b,t}$$

$$\bar{x}^{T} \geq x_k^T, \ \bar{y}^{-g,T} \leq y_k^{g,T}, \ \bar{y}^{-b,T} \geq y_k^{b,T}, \ \bar{y}^{-g,T} \geq 0, \ \bar{y}^{-b,T} \geq 0, \lambda_j^t \geq 0 \tag{2}$$

where $\lambda_j^t$ is the weight vector. $x_i^{-T}$, $y_r^{-g,T}$ and $y_q^{-b,T}$ are the optimal solutions of the evaluated DMUk in the global super-efficiency SBM model.

In order to incorporate slack variables into **Eq. 2**, $x_i^{-T}$, $y_r^{-g,T}$ and $y_q^{-b,T}$ can be further expressed by **Eq. 3**.

$$x_i^{-T} = x_{ik}^T + s_i^-; y_r^{-g.T} = y_{rk}^{g,T} - s_r^+; y_q^{-b,T} = y_{qk}^{b,T} + s_q^- \tag{3}$$

where $s_i^-$,$s_r^+$, and $s_q^-$ are the slack variables of input, expected output and undesired output respectively.

Further, the efficiency calculation formula of each input-output element can be decomposed. Based on (Cooper et al., 2007) the inefficiency decomposition idea, the inefficiency of each input and output element is completely decomposed. The efficiency decomposition formula and inefficiency decomposition formula of each input-output element are expressed by **Eqs 4, 5** respectively.

$$e_i = \frac{x_{ik}^T - s_i^-}{x_{ik}^T}, e_r = \frac{y_{rk}^{g,T}}{y_{rk}^{g,T} + s_r^+}, e_q = \frac{y_{qk}^{b,T} - s_q^-}{y_{qk}^{b,T}} \tag{4}$$

$$ie_i = \frac{1}{2m} \frac{s_i^-}{x_{ik}^T}, ie_r = \frac{1}{2(s_1 + s_2)} \frac{s_r^+}{y_{rk}^{g,T}}, ie_q = \frac{1}{2(s_1 + s_2)} \frac{s_q^-}{y_{qk}^{b,T}} \tag{5}$$

The three equations from left to right in **Eqs 4, 5** represent the efficiency decomposition formula and the inefficiency decomposition formula of input elements, expected output elements, and undesired output elements. The efficiency decomposition formula obtains the efficiency level of each element, and the inefficiency decomposition formula obtains the decomposition of the total inefficiency, reflecting the potential of each element to improve the efficiency of grain production.

### 3.1.2 Spatial Durbin Model

Space panel model can be divided into space lag model (SAR), space error model (SEM) and spatial durbin model (SDM). Among them, the spatial Dubin model is a general form of spatial lag model and spatial error model, thus it possesses both advantages of spatial autocorrelation and spatial interaction effects. This article examines the effect of each explanatory variable on the environmental efficiency of grain production and the spatial spillover. The basic model is as follows:

$$Y_{it} = \rho \sum_{j=1}^{N} W_{ij} Y_{jt} + \beta X_{it} + \sum_{j=1}^{N} \theta W_{ij} X_{jt} + \mu_i + \phi_t + \varepsilon_{it} \tag{6}$$

where $Y_{it}$ denotes the environmental efficiency of grain production in region $i$ in year $t$; $W$ is a spatial weight matrix of order $N \times N$; $W_{ij}Y_{jt}$ denotes the interaction effect of $Y_{jt}$ on $Y_{it}$ in regions adjacent to $i$; $X_{it}$ denotes each explanatory variable, i.e. factors affecting the environmental efficiency of grain production, in region $i$ in year $t$; $W_{ij}X_{jt}$ denotes the interaction effect of $X_{jt}$ on $X_{it}$ in regions adjacent to $i$; $\beta$ is the explanatory elasticity coefficient of the variable; $\theta$ is the elasticity coefficient of the spatially lagged term of the explanatory variable; $\mu_i$ denotes individual fixed effects; $\phi_t$ denotes time fixed effects; and $\varepsilon_{it}$ is a random error term. To ensure the robustness of the research conclusions, this paper also includes the spacial contiguity weights matrix ($W^1$), distance-based spatial weight matrix ($W^g$) and economic distance spatial weights matrix ($W^e$) in the research framework.

## 3.2 Index Selection

### 3.2.1 Definition of Input Variables

In this paper, land, irrigation, fertiliser, pesticide, agricultural film and agricultural machinery inputs are selected as input variables for grain production. Among them, the land input is expressed by the area planted with grain[1]; the irrigation input is expressed by the effective irrigated area of each province; the labor input is expressed by the number of grain laborers in each province; the fertilizer input is expressed by the annual amount of chemical fertilizer used by each province (including nitrogen, phosphate, potash, and compound fertilizers); pesticide input is expressed by the annual total pesticide input of each province; agricultural film input is expressed by the total agricultural film input of each province which is obtained by adding plastic film and film; and the input of agricultural machinery is expressed by the total power of agricultural machinery in each province.

### 3.2.2 Definition of Output Variables

1) The desirable output of grain production mainly includes grain production output and grain ecological output, with the

[1]Due to data availability limitations in the food carbon sequestration rate measurement section, the food sown area in this paper is mainly obtained by summing the area of four crops: rice, wheat, maize, beans and potatoes.

grain production output being expressed by the total grain output of each province and the grain ecological output being expressed by the grain carbon sinks in each province, that is, the amount of grain carbon absorbed by photosynthesis of grain crops[2]. Its calculation formula can be expressed as:

$$C = \sum_{i}^{k} C_i = \sum_{i}^{k} c_i \times Y_i \times (1 - r)/HI_i \quad (7)$$

where $C$ is the carbon sum of grain crops; $C_i$ is the carbon sink of a certain grain crop; $K$ is the number of grain crops; $c_i$ is the amount of carbon that can be absorbed by photosynthesis; $Y_i$ is the yield; $r$ is the water content, and $HI_i$ is the economic coefficient of grain crops.

2) Undesired output of grain production includes carbon emissions and non-point source pollution.

① Carbon emissions from grain production. According to the actual production of grain crops, this paper mainly calculates the carbon emissions from grain production in term of both agricultural materials and soil: the former is the carbon emissions from agricultural materials input. The carbon emissions of agricultural materials are mainly from directly or indirectly the input of agricultural materials such as fertilizers, pesticides, and agricultural films, and the electricity consumption of irrigation activities. The latter is $N_2O$ emissions from the soil. In the process of planting grain crops, the damage of $N_2O$ to the soil surface can easily lead to the emission of large amounts of greenhouse gases. It should be noted that, in order to facilitate analysis, when carbon emissions are being calculated, the standard carbon equivalent is replaced uniformly[3]. The formula for calculating the total carbon emissions from grain production is as follows:

$$E = \sum E_i = \sum (T_i \times \delta_i) \quad (8)$$

where $E$ is the total carbon emission from grain production; $E_i$ is the carbon emission of various carbon sources; $T_i$ is the amount of each carbon emission source; and $\delta_i$ is the carbon emission coefficient of each carbon emission source.

② Grain non-point source pollution. Grain non-point source pollution is represented by a comprehensive index of the loss of fertilizer nitrogen, the loss of fertilizer phosphorus, the amount of ineffective use of pesticides, and the residual amount of agricultural film. Among them, chemical fertilizer nitrogen loss is estimated by multiplying nitrogen loss coefficient by the sum of compound fertilizer nitrogen content and nitrogen fertilizer usage; chemical fertilizer phosphorus loss is estimated by multiplying the sum of compound fertilizer phosphorus content and phosphate fertilizer usage by phosphorus loss coefficient; the amount of ineffective use of pesticides is reversely estimated by subtracting the amount of effective use of pesticides; the residual amount of agricultural film is estimated by multiplying the amount of agricultural film used by the residue coefficient.

In addition, in order to further explore the spatial spillover effect of the environmental efficiency of grain production, this paper selects the environmental efficiency of grain production as the core explanatory variable, and the level of education, the level of financial support for agriculture, the loss of agricultural labor, the governance of the ecological environment, the level of urbanization, and the rate of disaster as the explanatory variables. Among them, the level of education (*EDU*) is expressed by the average years of education received by rural residents; the level of financial support for agriculture (*FS*) is expressed by the proportion of local financial expenditure on agriculture, forestry and water affairs in the overall budget of local finance; the loss of agricultural labor force (*ALT*) is expressed by the proportion of the labor force engaged in the primary sector in the total labor force engaged in the three major industries; the ecological environment control (*EPC*) is expressed by the proportion of environmental pollution control investment in GDP; the urbanization level (*UR*) is expressed by the ratio of the urban population to the permanent population; the disaster rate (*DR*) is expressed by the proportion of the area affected by the crops in the total sown area of crops. The descriptive statistics of each variable are shown in **Table 1**.

**TABLE 1 |** Descriptive analysis results.

| Variable | Obs | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|---|
| EDU | 208 | 7.532 | 0.475 | 6.197 | 8.35 |
| FS | 208 | 0.096 | 0.027 | 0.047 | 0.19 |
| ALT | 208 | 0.416 | 0.093 | 0.177 | 0.631 |
| EPC | 208 | 1.188 | 0.520 | 0.400 | 3.16 |
| UR | 208 | 0.474 | 0.097 | 0.246 | 0.677 |
| DR | 208 | 0.245 | 0.148 | 0.016 | 0.689 |

## 3.3 Data Sources

The input-output indicator data of this study is mainly derived from the "China Statistical Yearbook", "China Grain Yearbook", "China Rural Statistical Yearbook", and have been compiled to compose panel data of 13 provinces in the major grain production areas from 2004 to 2019. Considering the limitations of the statistical data in the statistical yearbooks, this study constructs a weight coefficient to separate the agricultural input elements to obtain more accurate input data for grain production (Coelli et al., 2003). In addition, to control the quantity of input-output indicators, this study uses the entropy method to integrate the four types of indicators of grain non-point source pollution accounting into one pollution index.

[2]Due to the availability of data, the calculation of grain carbon sinks in the current study only considers the carbon uptake during the entire life cycle of major grain crops.

[3]According to the IPCC Fourth Assessment Report (2007), 1 ton of $CO_2$ contains 0.272 tons of *C*; the greenhouse effect caused by 1 ton of $CH_4$ is equivalent to the greenhouse effect of 25 tons of $CO_2$ (about 6.818 tons *C*); The greenhouse effect caused by tons of $N_2O$ is equivalent to the greenhouse effect of 298 tons of $CO_2$ (about 81.272 tons of *C*).

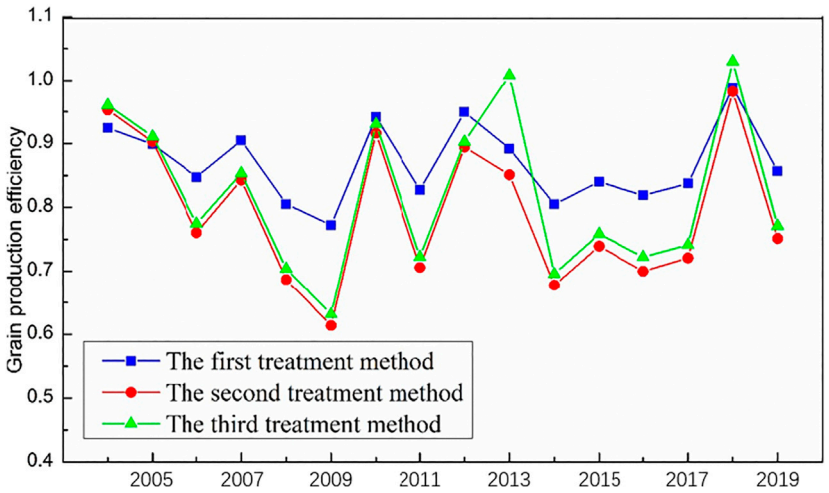



**FIGURE 1 |** Comparative analysis and trend of grain production efficiency in major grain production areas from 2004 to 2019.

# 4 EMPIRICAL RESULTS AND ANALYSIS

## 4.1 Measures of Environmental Efficiency in Grain Production and Sources of Inefficiency

### 4.1.1 Comparative Analysis of Efficiency of Grain Production With Different Treatment Methods

To reveal the influence of environmental factors on the efficiency of grain production, and comprehensively reflect the actual performance of the grain production environment efficiency in the major grain production areas, three different treatments was adopted to measure the efficiency value.

1) The first treatment method is the grain production efficiency taking no account of environmental factors;
2) The second is the grain production efficiency taking account of the only pollution factors of grain production environment;
3) The third is to simultaneously consider take account of both the grain production efficiency of grain ecological output and environmental pollution factors of grain production (hereinafter the environmental efficiency of grain production).

The three treatment methods are mainly used for the comparative analysis of the environmental efficiency of grain production. This study analyzes the environmental production efficiency of grain production in major grain production areas while considering the pollution factors of grain production environment and the ecological output of grain production. The MAXDEA5.0 software was used to measure the grain production efficiency of major grain production areas from 2004 to 2019 adopting the three treatment methods respectively, and to draw its annual trend chart (**Figure 1**).

As shown in **Figure 1**, the trends of the development stages are similar while adopting three different treatments. Except for 2004 and 2013, the grain production efficiency values taking no account of environmental factors were significantly higher than that in the other two treatments. The result indicated that the environmental factors impact the efficiency of grain production. This difference reflects both the degree of environmental pollution caused by the current mode of grain production, and to some extent reflects the magnitude of the environmental reaction to grain production and the cost of environmental governance. The greater the difference between them, the weaker the sustainability of grain production. Therefore, it is easy to misjudge the actual development capacity of grain production by measuring the efficiency of grain production without considering environmental pollution factors. This is likely to lead to misjudgment by policy makers. Grain production efficiency taking account of only environmental pollution factors is slightly lower than that of grain production environment. This may be due to the fact that the environmental efficiency of grain production considering ecological output is more approximate to the realistic situation of grain production. However, the positive externality of grain production is neglected when only considering environmental pollution is being considered. This will underestimate China's grain production efficiency. Therefore, the introduction of grain ecological output factors has a certain corrective effect on grain production efficiency, which helps us to understand the actual performance of grain production efficiency in each province.

### 4.1.2 Analysis of Time Series Evolution Characteristics

From 2004 to 2019, the average environmental efficiency of grain production was 0.820 (**Figure 1**). This result indicates that if the current input and output levels are maintained while eliminating efficiency losses, the efficiency value can be increased by 18%. It also shows that resource consumption and environmental pollution have caused greater efficiency losses in China's grain production process. China's grain production, which has not achieved coordinated development with the resources and the environment. At the same time, it shows that there is still much room for improvement in terms of China's grain production

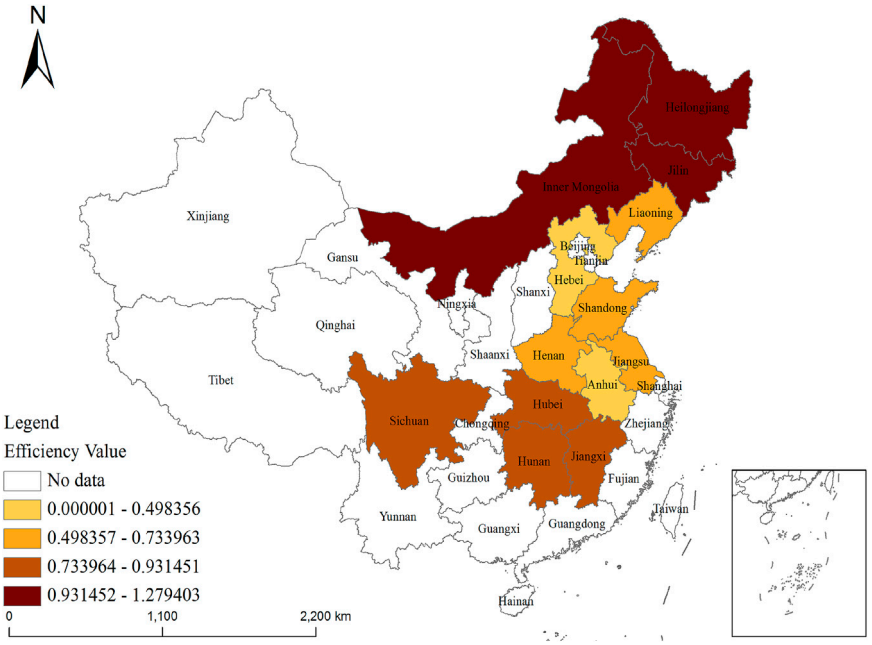



**FIGURE 2 |** Environmental efficiency of grain production in 13 provinces in China's major grain production areas.

methods. To achieve sustainable development of the grain industry, more efforts are needed.

Based on the trend of the mean change in the efficiency of grain production environment, the overall environmental efficiency of China's grain production fluctuated with a downward trend, from 0.961 in 2004 to 0.771 in 2019, and experienced three stages. Efficiency value witnessed a volatile period of decline from 2004 to 2009, with the environmental efficiency of grain production decreasing from 0.961 to 0.631 and hitting the bottom in 2009. Meanwhile, environmental efficiency increased in 2007, which might be due to the positive impact of the grain production subsidy policy and the agricultural tax pilot reform that was launched at that time in China. This policy has greatly stimulated the enthusiasm of farmers for production. However, there was a downward trend, indicating that the improvement of grain production efficiency in 2007 was only temporary, and the agricultural tax pilot reform had only temporarily promoted the improvement of production efficiency. There was a M-shaped turn, during which time, the efficiency value increased from 0.631 to 0.695 yet the average efficiency hitting a bottom in 2011 and 2014 due to a lack of production factors. The environmental efficiency in grain production peaked in 2010 and 2013, with efficiency values reaching 0.931 and 1.008, respectively. This was due to the promulgation of the agricultural tax exemption policy, which had increased the enthusiasm of farmers for growing grain. Therefore, the environmental efficiency value has been greatly improved in 2010 and natural disasters may be the main cause of the decline in environmental efficiency. In 2014, with the improvement of the technology and the expansion of the planting scale, the environmental efficiency value increased and reached a peak. From 2015 to 2019, another period of fluctuation, it increased from 0.758 to 0.771, and peaked in 2015. However, as shown in **Figure 1**, the efficiency value dropped sharply in 2019. The possible reason is that the floods and locust outbreaks in 2019 caused drop in grain production. At the same time, the structural reform of the agricultural supply side has changed the model of grain production from being previously "production-oriented" to "quality and benefit-oriented." This might have resulted in a reduction in the area of grain in the year and a significant decline in grain production in 2019 compared to 2018.

### 4.1.3 Spatial Distribution Characteristics

According to the results of the environmental efficiency of grain production in China's major grain production areas from 2004 to 2019, the natural breaks were used to divide the average efficiency into four grades (**Figure 2**). For the spatial distribution characteristics of the environmental efficiency, the deeper the color, the higher the environmental efficiency.

As shown in **Figure 2**, the environmental efficiency of grain production in provinces of Jilin, Inner Mongolia, and Heilongjiang is highest. The abundant land resources and large-scale operation are important reasons for the high efficiency of the grain production environment in this region. The environmental efficiency values of provinces of Sichuan, Jiangxi, Hubei, Hunan are merely above the average. These provinces have good ecological resources and grain production activities there are supported by more reliable resource elements. However, carbon emissions in the grain production process are high, and reducing carbon emissions is the focus of these provinces to improve the efficiency of the grain production

**TABLE 2 |** Input redundancy rates for environmental efficiency in grain production.

| Provience | Land | Labour | Fertilizer | Pesticides | Film | Machinery | Irrigation |
|---|---|---|---|---|---|---|---|
| Anhui | −22.12% | −57.45% | −31.60% | −48.01% | −23.54% | −47.65% | −38.14% |
| Hebei | −21.25% | −55.40% | −28.81% | −37.07% | −43.70% | −67.73% | −50.11% |
| Henan | −16.92% | −64.68% | −33.74% | −21.55% | −16.30% | −51.23% | −23.94% |
| Hubei | −16.31% | −52.57% | −29.66% | −54.96% | −4.17% | −17.35% | −18.41% |
| Hunan | −5.06% | −44.53% | −8.56% | −37.65% | −16.27% | −29.39% | −8.89% |
| Jiangsu | −8.42% | −30.72% | −26.73% | −33.03% | −25.44% | −27.32% | −41.55% |
| Jiangxi | −9.81% | −26.72% | −1.19% | −54.79% | −14.37% | −23.16% | −16.56% |
| Liaoning | −7.63% | −37.55% | −7.25% | −42.69% | −64.80% | −31.99% | −21.87% |
| Shandong | −9.50% | −58.79% | −28.99% | −53.35% | −69.30% | −64.19% | −34.82% |
| Sichuan | −26.46% | −60.90% | −5.28% | −8.87% | −45.93% | −7.22% | −6.90% |
| National | −14.35% | −48.93% | −20.18% | −39.20% | −32.38% | −36.72% | −26.12% |

**TABLE 3 |** Output redundancy rates for environmental efficiency in grain production.

| Provience | Total production (%) | Carbon sequestration (%) | Carbon emissions | Surface pollution |
|---|---|---|---|---|
| Anhui | 0.00 | 26.86 | −72.95% | −37.67% |
| Hebei | 0.00 | 14.91 | −28.06% | −57.38% |
| Henan | 0.00 | 16.85 | −24.90% | −65.67% |
| Hubei | 0.00 | 33.32 | −83.48% | −78.14% |
| Hunan | 0.00 | 29.83 | −78.44% | −44.00% |
| Jiangsu | 0.00 | 30.76 | −82.76% | −72.44% |
| Jiangxi | 0.00 | 32.06 | −55.53% | −27.75% |
| Liaoning | 0.00 | 17.60 | −28.18% | −52.61% |
| Shandong | 0.00 | 16.18 | −23.74% | −68.91% |
| Sichuan | 0.00 | 49.82 | −82.48% | −47.95% |
| National | 0.00 | 26.82 | −56.05% | −55.25% |

environment. The environmental efficiency levels in provinces of Liaoning, Shandong, Jiangsu, Henan, Hebei, and Anhui are relatively low. Among them, Hebei and Anhui have the lowest environmental efficiency. Most of these provinces are areas where agricultural machinery and fertilizers, pesticides, and agricultural film inputs are high. These regions have achieved higher grain output depending thanks to the high input of production factors, but they cannot achieve the highest levels of environmental efficiency given the environmental pollution, which makes it difficult for grain production to develop sustainably.

### 4.1.4 Breakdown of Sources of Inefficiency and Directions for Improvement

**Table 2** and **Table 3** report the decomposition and improvement directions of inefficient sources of grain production environment in each province of the main grain production areas. It shows that the environmental efficiency values of the three provinces of Inner Mongolia, Jilin and Heilongjiang are greater than 1, and the DEA is effective without efficiency improvement.

From the perspective of the production process, the total grain output of each province, has zero redundancy, yet the input factors and other output variables have certain redundancy. This indicates that China's environmental efficiency loss is not resulted from insufficient grain production, but mainly the following two aspects: production factor input and undesired output. Excessive consumption of input elements and excessive emission of environmental pollutants are the main reasons for the low environmental efficiency of China's grain production at this stage.

From a national perspective, the main influential factors of the environmental efficiency loss of grain production in major grain production areas are carbon emissions, non-point source pollution of grain production, and labor input in order. Among them, carbon emissions account for the highest proportion of the environmental efficiency loss of China's grain production. This shows the excessive carbon emissions in China's grain production process are excessive. This may be due to the input of large amount of high-carbon production materials such as pesticides, fertilizers, and agricultural machinery in China's grain production process currently. It leads to a continuous increase in carbon emissions from grain production and energy. Thus, China's grain production is facing tremendous pressure to reduce carbon emissions. Non-point source pollution in the process of grain production is the second most influential factor in the loss of environmental efficiency of grain production. This shows that China's grain production has serious non-point source pollution. The consumption of pesticides throughout the country is increasing at a rate of 10% each year. Problems such as excessive pesticide residues and water and soil pollution are very prominent. As China has a large amount of surplus labor in rural areas, labor input is the third most influential factor in the loss of environmental efficiency of grain production. However, it is worth noting that although China's main grain production areas have a relatively high redundancy rate of labor for growing grain, the human capital stock is low, and the aging problem is

**TABLE 4 |** The Moran's I of China's grain production environmental efficiency between 2004 and 2019.

| Year | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---|---|---|---|---|---|---|---|---|
| Moarn's I | –0.101 | 0.160 | 0.575 | 0.400 | 0.435 | 0.550 | 0.114 | 0.442 |
| Z-value | –0.043 | 1.353 | 3.292 | 2.427 | 2.663 | 3.523 | 1.060 | 2.676 |
| *p*-value | 0.452 | 0.095 | 0.004 | 0.016 | 0.016 | 0.003 | 0.139 | 0.013 |
| Year | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| Moarn's I | 0.057 | –0.124 | 0.516 | 0.356 | 0.345 | 0.276 | 0.136 | 0.321 |
| Z-value | 0.712 | –0.248 | 3.080 | 2.211 | 2.214 | 1.834 | 1.309 | 2.047 |
| *p*-value | 0.229 | 0.467 | 0.006 | 0.25 | 0.027 | 0.055 | 0.099 | 0.29 |

prominent. The aging rate has been higher than the national average in the past 11 years, and the gap is widening.

From an inter-provincial perspective, the reasons for the loss of environmental efficiency of grain production vary from province to province. The main influential factor of the environmental efficiency loss of grain production in Anhui, Sichuan, Hubei, Hunan, Jiangsu, Jiangxi and other provinces is carbon emission, which shows that these provinces have not effectively controlled carbon emissions and there is huge room for improvement. The main influential factor of the environmental efficiency loss of grain production in provinces of Liaoning and Shandong is the input of agricultural film, which shows that there is still a large redundancy in the use of agricultural film in the two provinces. The input of machinery in Hebei has a greater impact on the environmental efficiency loss of grain production in the region. This may be due to the implementation of agricultural machinery subsidy policies such as agricultural machinery purchase subsidies. The province's machinery ownership has increased year by year, resulting in machinery redundancy and declining efficiency. The main influential factor for the loss of environmental efficiency in Henan province is non-point source pollution, which might be resulted from lack of scientific protective development, and necessary agricultural technology expertise among farmers, and could lead to improper fertilization and drug use, and even serious non-point source pollution.

## 4.2 Analysis of Spatial Spillovers and Impact Factors for Environmental Efficiency of Grain Production

### 4.2.1 Spatial Correlation Analysis

#### *(1) Global Spatial Autocorrelation Analysis*

Using Geoda software based on Queen's first-order spatial weight matrix, Monte Carlo simulation 999 tests were used to calculate the global Moran's I, Z-statistic test value, and significance level of grain production environmental efficiency in China's major grain production areas. The test results are reported in **Table 4**.

From **Table 4**, we can observe that the environmental efficiency of grain production in China's major grain production areas in 2004 and 2013 was negative and did not pass the significance test. This indicates that the correlation between environmental efficiency level and adjacent areas is low. It also presents a random distribution state. Moran's I for 2005, 2010, 2012, and 2018 were positive and of small magnitude, and none of them passed the significance test, indicating that environmental efficiency was spatially randomized over the 4 years. Moran's I was positive in 2006–2009, 2014–2017, 2011 and 2019, and showed significant positive spatial correlation according to the 5% significance test, meaning that there was spatial dependence. This demonstrates that the spatial distribution of grain production environmental efficiency levels shows obvious cluster characteristics.

#### *(2) Local Spatial Autocorrelation Analysis*

Local spatial autocorrelation analysis can make up for the shortcomings of global spatial autocorrelation that cannot clearly give spatially clustered locations with similar attribute regions. Moran's I scatter plot and a LISA spatial agglomeration map can be used to evaluate the heterogeneity of environmental efficiency in different regions in the spatial range. To reflect the spatial differentiation characteristics of 13 provinces in China's major grain production areas, Moran's I scatter plots for 2004, 2009, 2014 and 2019 were drawn using GeoDa software, based on the environmental efficiency values obtained above. Moran's I scatter plot distributes the environmental efficiency values of the aforementioned provinces in four quadrants, reflecting the spatial correlation between provinces and their neighboring provinces. The first quadrant is high-high value cluster (H-H), the second low-high value cluster (L-H), the third low-low value cluster (L-L), and the fourth high-low value cluster (H-L). **Figure 3** shows Moran's I scatter plots for grain production environmental efficiency in 2004, 2009, 2014, and 2019. The number of provinces in H-H clusters and L-L clusters were 5, 13, 12, and eight in 2004, 2009, 2014, and 2019, respectively. Numbers increased rapidly and then gradually decreased, indicating that the environmental production efficiency of grain production H-H clusters and L-L clusters increased first and then went down. The concentration of L-H clusters fluctuated, and H-L clusters fell first and then gradually increased. By comparison, we found:

① H-H clusters were identified in Heilongjiang, Jilin, and Inner Mongolia. The reason is that all three provinces are in the northeast of China. These provinces have good resource endowment, few natural disasters such as floods and pests, and are sparsely populated, which is conducive to the large-scale mechanization of grain production. Therefore, the environmental efficiency of these areas is relatively high, which has a diffusion effect on the surrounding areas.

② L-H clusters were identified in Anhui, Liaoning, Hebei, Hubei, and other areas adjacent to the H-H clusters. The possible reason is that Liaoning Province is adjacent to Inner Mongolia

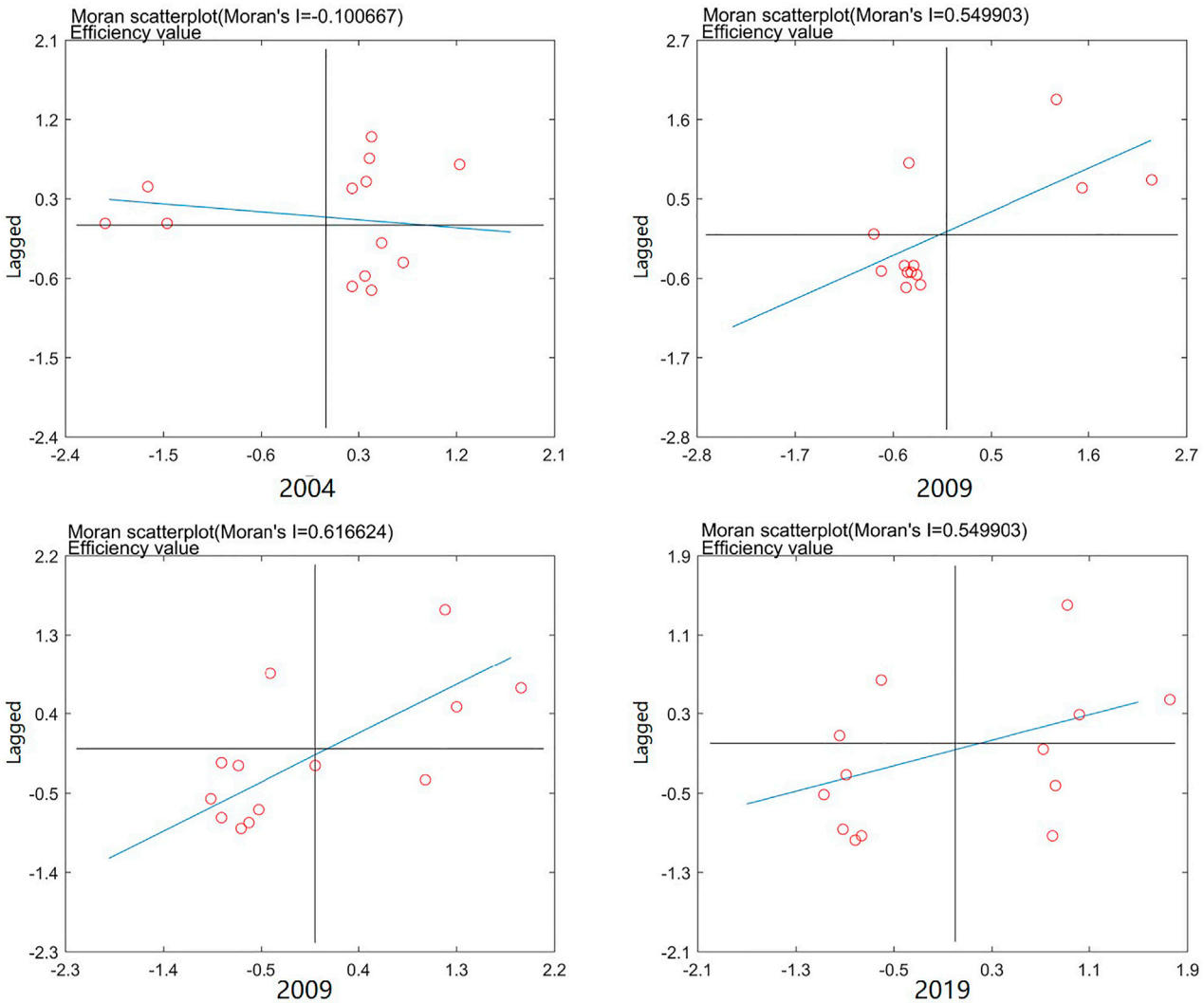



**FIGURE 3 |** Moran's I scatter plot of environmental efficiency of grain production in 2004, 2009, 2014 and 2019.

and Jilin, which has a positive impact on production technology, environmental protection investment, and environmental protection policies. Anhui and Hebei are in L-L clusters and L-H clusters, respectively. Their efficiency value fluctuate significantly, which may be resulted from the fact that the selection of grain production modes in the two areas is influenced by their neighboring provinces.

③ L-L clusters account for a high proportion in the major grain production areas, and Anhui, Shandong, Henan, and Jiangsu are in this cluster area in 2009, 2014, and 2019. The provinces of the cluster area have a certain downward trend from 2014 to 2019, which may be due to the positive radiation effects of the provinces, especially from those with high efficiency in grain production.

④ The distribution of H-L clusters is relatively scattered, covering seven regions in the initial stage of research to four regions at the end of the study. Among them, Hunan, Jiangxi, and Sichuan changed from L-L clusters in 2009 and 2014 to H-L clusters in 2019. This may be because the three provinces increased their investment in grain production technology and environmental management throughout the study period, reducing carbon emissions and non-point source pollution in grain production, thereby improving the environmental efficiency of grain production. However, the surrounding environment with relatively low environmental efficiency has a good geographical advantage. In the future, due to the spillover of policies, technologies, and factors in high-efficiency areas, the efficiency of the grain environment will be greatly improved.

The Moran's I scatter plot reflects the spatial dependence and heterogeneity of the provinces in the major grain production areas. However, the salient status of each province is not given. Therefore, this study draws a LISA spatial cluster map (**Figure 4**) to reflect the degree of correlation between the provinces and the environmental efficiency of their neighboring provinces.

According to the analysis results of the LISA aggregation map, 13 provinces in the major grain-production areas did

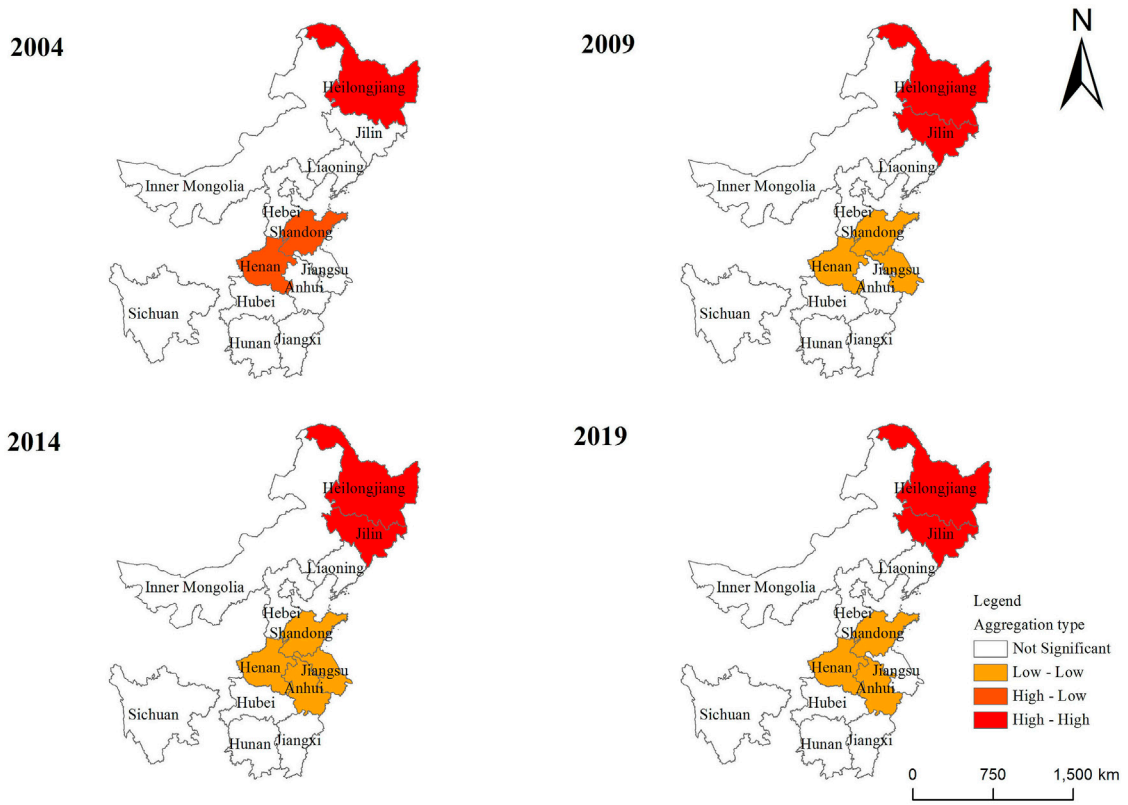



**FIGURE 4 |** LISA clusters map of China's grain production environmental efficiency in 2004, 2009, 2014 and 2019.

not pass the significance test in 4 years. However, the spatial agglomeration of grain production environmental efficiency in various provinces in the region is becoming increasingly obvious, and more provinces are showing significant relationships. The transformation of spatial cluster type in some provinces may be directly related to the environmental efficiency of grain production in each province in the same year. In this study, Heilongjiang Province was always in the H-H clusters, while Jilin Province, which was not insignificant in 2004, was in the H-H clusters in other years. It indicates that the two provinces have good resource endowment, advanced technology, proper environmental protection measures, high environmental efficiency and environmental efficiency in surrounding provinces, and high concentration of resources and technology. Moreover, it has played a leading role in the surrounding areas and has a strong positive radiation effect. Shandong and Henan provinces were in H-L clusters in 2004, but converted to L-L clusters in 2009, 2014, and 2019. This indicates that the two provinces had higher levels of efficiency during the initial study period, but were later assimilated by the lower environmental efficiency levels of neighboring provinces and were in L-L clusters. Jiangsu and Anhui provinces have been in L-L clusters during the study period. This indicates that the environmental efficiency of the two provinces and the environmental efficiency of the surrounding areas were very low, and there was no timely response to the ecological development of China's grain production.

### 4.2.2 Regression Analysis of Spatial Durbin Model

Through spatial autocorrelation analysis, it is found that there is a certain spatial correlation among China's grain production environment in terms of the efficiency. In order to further explore the spatial effect and influential factors for that case, the study uses MATLAB R2015b to estimate the spatial econometric model from the perspective of robustness. The study has also used three different spatial weight matrices to estimate the model. The regression results are shown in **Table 5**. After the comparison of the estimation results under the application of three spatial weight matrices, it is found that the signs and significance test results of each variable are consistent, indicating that the regression results are robust enough. The R2 of the model under spatial weight matrices are 0.782, 0.714, and 0.770, respectively, indicating that the model has a convincing explanatory. In the specific coefficient analysis below, this article will analyze the regression results of spatial contiguity weights matrix case that has the highest degree of fit.

Before estimating the space spillover effect (robust) LM test should be used to select the models between space lag or space error. The results are shown in **Table 5**. The test value of LM-err is 6.724, which is significant at the 1% level; the R-LM-err value is

**TABLE 5 |** Estimation results of SDM.

| Variable | $SDM(W^l)$ | | $SDM(W^g)$ | | $SDM(W^e)$ | |
|---|---|---|---|---|---|---|
| | Coefficient | z-Statistic | Coefficient | z-Statistic | Coefficient | z-Statistic |
| EDU | 0.018** | 2.092 | 0.014 | 1.503 | 0.015* | 1.661 |
| FS | 0.200 | 1.474 | 0.177 | 1.305 | 0.166 | 1.278 |
| ALT | -0.553 | −1.167 | −0.820* | -1.803 | -1.023** | −2.326 |
| EPC | 0.061 | 1.591 | 0.069 | 1.588 | 0.077* | 1.858 |
| UR | 0.157 | 0.213 | 1.506 | 1.564 | 1.589*** | 1.652 |
| DR | −0.295** | −2.198 | −0.278** | −2.137 | −0.277** | −2.170 |
| W×EDU | 0.031** | 2.012 | 0.068* | 1.829 | 0.080** | 2.007 |
| W×FS | 1.197** | 2.48 | 2.519** | 2.033 | 2.903** | 2.372 |
| W×ALT | -2.116*** | −3.772 | -4.568*** | −2.705 | -4.402*** | −2.846 |
| W×EPC | 0.120 | 1.450 | 0.412* | 1.755 | 0.524** | 2.381 |
| W×UR | 2.779** | 2.064 | 1.666 | 0.570 | 1.737 | 0.627 |
| W×DR | −0.059 | −0.315 | −0.509 | −1.242 | −0.653 | −1.581 |
| $\lambda\backslash\rho$ | 0.122*** | 3.537 | 0.603*** | 3.388 | 0.581*** | 3.292 |
| $R^2$ | 0.782 | | 0.714 | | 0.771 | |
| log-likelihood | 111.370 | | 77.304 | | 111.591 | |
| LM-lag test | 19.431** | | 21.308** | | 25.921** | |
| R-LM-lag test | 49.163** | | 9.787** | | 10.577* | |
| LM-err-test | 6.724*** | | 15.989*** | | 19.763*** | |
| R-LM-err-test | 36.457*** | | 4.467*** | | 4.419*** | |
| Wald-lag test | 21.112*** | | 12.685*** | | 3.460** | |
| LR-lag test | 21.492*** | | 12.196*** | | 12.122*** | |
| Wald-err test | 21.060*** | | 12.379*** | | 8.233** | |
| LR-err test | 19.947*** | | 11.785*** | | 8.297** | |
| Hausman test | 42.221*** | | 49.147*** | | 46.555*** | |

*Note:* ***p < 0.01, **p < 0.05, *p < 0.1.

**TABLE 6 |** Spatial spillover effect of environmental efficiency of Grain Production.

| Variable | $SDM(W^l)$ | | | $SDM(W^g)$ | | | $SDM(W^e)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | Direct Effect | Indirect Effect | Total Effect | Direct Effect | Indirect Effect | Total Effect | Direct Effect | Indirect Effect | Total Effect |
| EDU | 0.019** | 0.036* | 0.055** | 0.005* | 0.039 | 0.044 | 0.005* | 0.036 | 0.041 |
| | (2.192) | (2.073) | (2.278) | (1.953) | (0.752) | (0.757) | (1.854) | (0.696) | (0.711) |
| FS | 0.267* | 3.104* | 3.371** | 0.346* | 2.555* | 2.901* | 0.156* | 1.888* | 2.045* |
| | (1.937) | (2.094) | (2.363) | (1.945) | (2.089) | (2.041) | (1.918) | (1.975) | (1.874) |
| ALT | −0.470* | 2.281*** | 1.811* | −1.073** | 2.188*** | 1.116*** | −1.140** | 1.854 | 0.714 |
| | (−1.957) | (−3.618) | (−2.059) | (−2.187) | (−3.856) | (−3.402) | (−2.337) | (−0.759) | (−0.271) |
| EPC | 0.067 | 0.143 | 0.210 | −0.025 | −0.280 | −0.305 | −0.023 | −0.217 | −0.239 |
| | (−1.669) | (−1.476) | (−1.708) | (−0.638) | (−1.305) | (−1.309) | (−0.582) | (−1.084) | (−1.093) |
| UR | 0.299** | −0.100 | 0.399** | 0.337** | −0.368* | 0.031 | 0.338** | −0.373 | 0.035 |
| | (−2.239) | (−0.518) | (−2.263) | (−2.595) | (−1.969) | (−0.059) | (−2.674) | (−0.695) | (−0.067) |
| DR | −0.251* | −1.348** | −1.599** | −0.007* | −0.910** | −0.903** | −0.012* | −0.904* | −0.892** |
| | (−1.829) | (−2.449) | (−2.753) | (−2.054) | (−2.325) | (−2.256) | (−1.807) | (−1.932) | (−2.310) |

*Note: The values in brackets below the coefficients are the corresponding t-values;* ***p < 0.01, **p < 0.05, *p < 0.1.

36.457, which is significant at the 1% level; the LM-lag value is 19.431, which indicates that the 5% significance test is passed; the value of R-LM-lag is 49.163, which also satisfies the 5% significance level test. Therefore, it can be concluded that the spatial error model is more applicable than the spatial lag model. The Hausman test results show that the statistic passes the chi-square test at a 1% significance level, so a fixed-effect model is more reasonable to apply. The statistics of Wald test and LR test are both significant at the level of 1%, indicating that the spatial Durbin model cannot be simplified to the SAR model or the SEM model. The $\rho$ value is 0.122, and it is significant at the level of 1%, which once again proves that there is a significant spatial correlation of the environmental efficiency of grain production. This correlation effect is more obvious under the geographical weight, which means that the geographical distribution of environmental efficiency of grain production exerts an aggregation effect.

This paper breaks down the effects of various influential factors on the environmental efficiency of grain production to calculate the direct and indirect effects. The results are shown in **Table 6**. Among them, the direct effect, indirect effect and total effect of education level all show significant positive effects. It indicates that the improvement of farmers' education level can not only improve the environmental efficiency of grain

production in the region, but also increase the environmental efficiency of grain production in other regions through spatial spillover effects. This may be because the improvement of education level means the improvement of the quality of farmers. In the process of grain production, they tend to choose more ecologically and environmentally friendly production methods, reduce grain production pollution, and improve grain production environmental efficiency. The direct effect, indirect effect and total effect of the level of fiscal support to agriculture all show significant positive effects. This shows that financial support for agriculture can not only improve the environmental efficiency of grain production in the region, but also effectively promote the environmental efficiency of grain production in the neighboring areas. The direct effect of labor transfer is significantly negative, but the indirect effect presents a more significant positive effect. It indicates that the transfer of rural labor is not conducive to the improvement of the environmental efficiency of grain production in the region, but it is conducive to improving the environmental efficiency of grain production in the neighboring areas. The indirect spillover effect is greater than the direct spillover effect in the region.

The direct effect, indirect effect and total effect of ecological environment governance are negative but not significant, indicating that the current ecological environment governance has an insignificant effect on improving the environmental efficiency of grain production. This may be because the current investment in environmental governance is mainly concentrated in industry, while the investment in agricultural production environmental governance is relatively low. Therefore, the effect of improving the environmental efficiency of grain production is not obvious. The direct and total effects of the urbanization level are both significantly positive, and the indirect effects are negative but not significant. This shows that the increase in the urbanization rate has effectively promoted the growth of the local grain production environment efficiency, but the spillover effect on the surrounding areas is not obvious. This may be because the development of urbanization has created many favorable conditions for improving the environmental efficiency of grain production by promoting the advancement of agricultural science and technology, accelerating the development of modern agriculture, and attracting and promoting capital investment in agriculture. The direct effect, indirect effect and total effect of the disaster rate all show significant negative effects. It means that an increase in the disaster rate will reduce grain production, which is not conducive to the improvement of the environmental efficiency of grain production. At the same time, the farmland of neighboring provinces is often connected. When the grain production in one region suffers from natural disasters, grain production in neighboring regions will be affected to a certain extent.

## 5 CONCLUSION

This paper incorporates the environmental factors in the grain production process into the grain production efficiency evaluation system and employs the undesired global super-efficiency SBM model to calculate and analyze the temporal and spatial characteristics of grain production environmental efficiency, sources of inefficiency, and spatial spillover in major grain production areas from 2004 to 2019. The main conclusions are as follows:

1) Except for 2004 and 2013, the grain production efficiency values that take no account of environmental factors (environmental pollution and grain carbon sinks) were significantly higher than those that take environmental factors into consideration. There was a certain corrective effect on the efficiency of grain production in China.
2) From 2004 to 2019, the average environmental efficiency of grain production in China's main grain production areas was 0.820, showing generally fluctuations with a downward trend, and the level of development was generally low. The main influential factors of the environmental efficiency loss of grain production in major grain production areas were carbon emissions, non-point source pollution of grain production, and labor input. In provinces with inefficient grain production environments, there were widespread problems of redundant element inputs and serious environmental pollution in grain production.
3) The environmental efficiency of grain production in various provinces shows more and more obvious spatial differentiation characteristics in spatial distribution. Liaoning, Shandong, Jiangsu, Henan, Hebei, Anhui and other provinces have low levels of grain environmental efficiency and suffer serious environmental efficiency losses. In addition, the environmental efficiency level of grain production shows obvious cluster characteristics. Among them, the Northeast region and the Huang-Huai-Hai region shows high-high aggregation and low-low aggregation respectively.
4) The environmental efficiency of grain production has a significant positive spillover effect. In terms of direct effects, the level of education, the level of financial support for agriculture, and the rate of urbanization all have positive effects on the environmental efficiency of grain production in the region. On the contrary, the transfer of labor and the rate of disasters have a significant negative impact. From the perspective of spillover effects, the level of education, the level of financial support for agriculture, and the transfer of labor are all conducive to the improvement of the environmental efficiency of grain production in the neighboring areas, while the disaster rate has a negative impact on the environmental efficiency of grain production in the neighboring areas.

Based on the above conclusions, the following suggestions for improving the environmental efficiency of grain production are proposed:

1) In view of the overall low environmental efficiency of grain production in China, there is still large room for improvement in the prevention and control of pollution and the discovery of ecological value in the process of grain production. In order to

maintain the sustainable development of grain production, it is necessary to actively promote the reduction of chemical fertilizers and pesticides, control the carbon emissions of grain planting, highlight the ecological function of grain planting, and take the path of coordinated development of resources, environment and grain production while ensuring grain production.

2) In the process of grain production, the natural and resource endowment of the local area must be fully considered. The emphasis should be laid on supporting provinces where the environmental efficiency of grain production is lower than that of the surrounding areas. The decline in the sustainability of grain production caused by low efficiency and serious ecological damage should be addressed.
3) The scope and intensity of the positive spillover including education level, financial support for agriculture, labor transfer should be expanded and deepened, while the negative spillover effects of grain disasters deserve equal attention. The relevant policies, measures and technologies of green grain production in neighboring areas should also be highlighted. A multilateral or bilateral regular exchange mechanism can be established to strengthen the exchange of technology and policy experience among the provinces of the main grain production areas and guide the proper flow, transfer and communication of policy experience, technology, talent and other elements.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by HX, BM, and QG. The first draft of the manuscript was written by HX and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported in part by grants from the Ministry of education of Humanities and Social Science project (Grant No. 21YJC790149), Qingdao social science planning and research project in 2021 (QDSKL2101031), Social Science Planning Research Project of Shandong "Research on the Stagnation and Breakthrough Path of Relative Poverty Management in Shandong Province after 2020" (Grant No. 20CDCJ21), "The key project of plan in research and development of Shandong "Research on the protection and utilization of marine fishery germplasm resources in Shandong Province" (Grant No. 2020RZE29007), The Natural Science Foundation of Shandong "Research on the Long-term Path of Poverty Alleviation in Agricultural Industry from the Perspective of High-Quality De velopment" (Grant No. ZR2020QG045), The Fundamental Research Funds for the Central Universities (Grant No.202013011), The China Postdoctoral Science Foundation (Grant No. 2019M652486).

## REFERENCES

Armagan, G., Ozden, A., and Bekcioglu, S. (2010). Efficiency and Total Factor Productivity of Crop Production at Nuts1 Level in turkey: Malmquist index Approach. *Qual. Quant* 44, 573–581. doi:10.1007/s11135-008-9216-5

Bai, X., Salim, R., and Bloch, H. (2019). Environmental Efficiency of Apple Production in china: A Translog Stochastic Frontier Analysis. *Agric. Resour. Econom. Rev.* 48, 199–220. doi:10.1017/age.2018.25

Battese, G. E., and Coelli, T. J. (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Econ.* 20, 325–332. doi:10.1007/bf01205442

Benedetti, I., Branca, G., and Zucaro, R. (2019). Evaluating Input Use Efficiency in Agriculture through a Stochastic Frontier Production: An Application on a Case Study in Apulia (italy). *J. Clean. Prod.* 236, 117609. doi:10.1016/j.jclepro.2019.117609

Cao, X., Liu, Y., Li, T., and Liao, W. (2019). Analysis of Spatial Pattern Evolution and Influencing Factors of Regional Land Use Efficiency in china Based on Esda-Gwr. *Scientific Rep.* 9, 1–11. doi:10.1038/s41598-018-36368-2

Coelli, T., Rahman, S., and Thirtle, C. (2003). A Stochastic Frontier Approach to Total Factor Productivity Measurement in Bangladesh Crop Agriculture, 1961-92. *J. Int. Dev.* 15, 321–333. doi:10.1002/jid.975

Cooper, W. W., Seiford, L. M., and Tone, K. (2007). A Comprehensive Text with Models, Applications, References and Dea Solver Software. *Data Envelopment Anal.*

Dolgikh, Y. (2019). Evaluation and Analysis of Dynamics of Change of Efficiency of Grain Production in ukraine by Dea Method. *Agric. Res. Econ.* 5 (3), 47–62.

Gebregziabher, G., Namara, R. E., and Holden, S. (2012). Technical Efficiency of Irrigated and Rain-Fed Smallholder Agriculture in Tigray, ethiopia: A Comparative Stochastic Frontier Production Function Analysis. *Q. J. Int. Agric.* 51, 203–226.

Kumbhakar, S. C., Lien, G., and Hardaker, J. B. (2014). Technical Efficiency in Competing Panel Data Models: a Study of Norwegian Grain Farming. *J. Prod. Anal.* 41, 321–337. doi:10.1007/s11123-012-0303-1

Liao, X., and Shi, X. (2018). Public Appeal, Environmental Regulation and green Investment: Evidence from china. *Energy Policy* 119, 554–562. doi:10.1016/j.enpol.2018.05.020

Lin, J. Y. (1992). Rural Reforms and Agricultural Growth in china. *Am. Econ. Rev.*, 34–51.

Liu, Y., Zou, L., and Wang, Y. (2020). Spatial-temporal Characteristics and Influencing Factors of Agricultural Eco-Efficiency in china in Recent 40 Years. *Land Use Policy* 97, 104794. doi:10.1016/j.landusepol.2020.104794

Maxime, D., Marcotte, M., and Arcand, Y. (2006). Development of Eco-Efficiency Indicators for the canadian Food and Beverage Industry. *J. Clean. Prod.* 14, 636–648. doi:10.1016/j.jclepro.2005.07.015

Pang, J., Chen, X., Zhang, Z., and Li, H. (2016). Measuring Eco-Efficiency of Agriculture in china. *Sustainability* 8, 398. doi:10.3390/su8040398

Pastor, J. T., and Lovell, C. A. K. (2005). A Global Malmquist Productivity index. *Econ. Lett.* 88, 266–271. doi:10.1016/j.econlet.2005.02.013

Picazo-Tadeo, A. J., Gómez-Limón, J. A., and Reig-Martínez, E. (2011). Assessing Farming Eco-Efficiency: a Data Envelopment Analysis Approach. *J. Environ. Manag.* 92, 1154–1164. doi:10.1016/j.jenvman.2010.11.025

Pishgar-Komleh, S. H., Zylowski, T., Rozakis, S., and Kozyra, J. (2020). Efficiency under Different Methods for Incorporating Undesirable Outputs in an LCA+DEA Framework: A Case Study of winter Wheat Production in Poland. *J. Environ. Manag.* 260, 110138. doi:10.1016/j.jenvman.2020.110138

Sachdeva, J., Sharma, J., and Chahal, S. (2011). Technical Efficiency in Crop Production and Dairy Farming in punjab: A Zone-wise Analysis1. *Productivity* 52, 177. doi:10.3109/10428194.2011.574756

Shi, X., Yu, J., and Cheong, T. S. (2020). Convergence and Distribution Dynamics of Energy Consumption Among China's Households. *Energy Policy* 142, 111496. doi:10.1016/j.enpol.2020.111496

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B. L., Lassaletta, L., et al. (2018). Options for Keeping the Food System within Environmental Limits. *Nature* 562, 519–525. doi:10.1038/s41586-018-0594-0

Sun, H., Edziah, B. K., Kporsu, A. K., Sarkodie, S. A., and Taghizadeh-Hesary, F. (2021a). Energy Efficiency: The Role of Technological Innovation and Knowledge Spillover. *Technol. Forecast. Soc. Change* 167, 120659. doi:10.1016/j.techfore.2021.120659

Sun, H., Edziah, B. K., Sun, C., and Kporsu, A. K. (2021b). Institutional Quality and its Spatial Spillover Effects on Energy Efficiency. *Socio-Economic Plann. Sci.*, 101023. doi:10.1016/j.seps.2021.101023

Vlontzos, G., Niavis, S., and Manos, B. (2014). A Dea Approach for Estimating the Agricultural Energy and Environmental Efficiency of Eu Countries. *Renew. Sustain. Energ. Rev.* 40, 91–96. doi:10.1016/j.rser.2014.07.153

Wang, J., Wang, K., Shi, X., and Wei, Y.-M. (2019). Spatial Heterogeneity and Driving Forces of Environmental Productivity Growth in china: Would it Help to Switch Pollutant Discharge Fees to Environmental Taxes. *J. Clean. Prod.* 223, 36–44. doi:10.1016/j.jclepro.2019.03.045

Wang, K., Xian, Y., Yang, K., Shi, X., Wei, Y.-M., and Huang, Z. (2020). The Marginal Abatement Cost Curve and Optimized Abatement Trajectory of Co 2 Emissions from china's Petroleum Industry. *Reg. Environ. Change* 20, 1–13. doi:10.1007/s10113-020-01709-3

Wen, G. J. (1993). Total Factor Productivity Change in China's Farming Sector: 1952-1989. *Econ. Dev. Cult. Change* 42, 1–41. doi:10.1086/452063

Wilson, P., Hadley, D., Ramsden, S., and Kaltsas, I. (1998). Measuring and Explaining Technical Efficiency in uk Potato Production. *J. Agric. Econ.* 49, 294–305. doi:10.1111/j.1477-9552.1998.tb01273.x

Xian, Y., Wang, K., Shi, X., Zhang, C., Wei, Y.-M., and Huang, Z. (2018). Carbon Emissions Intensity Reduction Target for China's Power Industry: An Efficiency and Productivity Perspective. *J. Clean. Prod.* 197, 1022–1034. doi:10.1016/j.jclepro.2018.06.272

Yao, Z., and Xuan, P. (2019). "The Influence Degree of Minimum purchase price Policy on Grain Production Efficiency," in IOP Conference Series: Earth and Environmental Science (Xian: IOP Publishing), 032106. doi:10.1088/1755-1315/252/3/032106*IOP Conf. Ser. Earth Environ. Sci.*252.

Yuan, W., Li, J., Meng, L., Qin, X., and Qi, X. (2019). Measuring the Area green Efficiency and the Influencing Factors in Urban Agglomeration. *J. Clean. Prod.* 241, 118092. doi:10.1016/j.jclepro.2019.118092

Zhang, Q., Zhang, F., Wu, G., and Mai, Q. (2021). Spatial Spillover Effects of Grain Production Efficiency in china: Measurement and Scope. *J. Clean. Prod.* 278, 121062. doi:10.1016/j.jclepro.2020.121062

Zou, L., Liu, Y., Wang, Y., and Hu, X. (2020). Assessment and Analysis of Agricultural Non-point Source Pollution Loads in China: 1978-2017. *J. Environ. Manag.* 263, 110400. doi:10.1016/j.jenvman.2020.110400

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Xu, Ma and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 21 December 2021
doi: 10.3389/fenvs.2021.800179

# Tracking Cloud Forests With Cloud Technology and Random Forests

*Pasky Pascual*[1]* *and Cam Pascual*[2]

[1]United States Environmental Protection Agency (Retired), Washington, DC, United States, [2]World Wildlife Fund, Washington, DC, United States

Hotspots of endemic biodiversity, tropical cloud forests teem with ecosystem services such as drinking water, food, building materials, and carbon sequestration. Unfortunately, already threatened by climate change, the cloud forests in our study area are being further endangered during the Covid pandemic. These forests in northern Ecuador are being razed by city dwellers building country homes to escape the Covid virus, as well as by illegal miners desperate for money. Between August 2019 and July 2021, our study area of 52 square kilometers lost 1.17% of its tree cover. We base this estimate on simulations from the predictive model we built using Artificial Intelligence, satellite images, and cloud technology. When simulating tree cover, this model achieved an accuracy between 96 and 100 percent. To train the model, we developed a visual and interactive application to rapidly annotate satellite image pixels with land use and land cover classes. We codified our algorithms in an R package—loRax—that researchers, environmental organizations, and governmental agencies can readily deploy to monitor forest loss all over the world.

**Keywords: cloud forest, random forest, artificial intelligence, machine learning, cloud technology, Sentinel-2, raster index, biodiversity 1**

OPEN ACCESS

***Edited by:***
*Xunpeng (Roc) Shi,*
*University of Technology Sydney, Australia*

***Reviewed by:***
*Jizhong Wan,*
*Qinghai University, China*
*Miguel Alfonso Ortega-Huerta,*
*National Autonomous University of Mexico, Mexico*

***Correspondence:**
*Pasky Pascual*
*pasky012@gmail.com*

***Specialty section:***
*This article was submitted to Environmental Informatics and Remote Sensing, a section of the journal Frontiers in Environmental Science*

***Received:*** *22 October 2021*
***Accepted:*** *01 December 2021*
***Published:*** *21 December 2021*

***Citation:***
*Pascual P and Pascual C (2021) Tracking Cloud Forests With Cloud Technology and Random Forests. Front. Environ. Sci. 9:800179. doi: 10.3389/fenvs.2021.800179*

## INTRODUCTION

CLOUD FORESTS (CF) are mountain forests persistently shrouded by clouds and mist. They have been likened to island archipelagos because, like an island, each CF constitutes a unique, isolated habitat where heterogeneous, endemic species proliferate (Foster, 2001).

Although they cover less than one per cent of the planet's land surface, CFs disproportionately host its most biodiverse ecoregions, which are regions of natural communities where species share environmental conditions and interact in ways sustaining their collective existence (Bruijnzeel et al., 2010). Nearly 90% of CFs across the globe are on the World Wildlife Fund's list of the 200 priority ecoregions—a list based on species richness, number of endemic species, rarity of habitats, and other features of biodiversity (Olson and Dinerstein, 2002).

Twenty-five percent of the earth's CFs can be found in the Americas. In particular, the tropical Andes mountains are hotspots of biodiversity, including the CFs of Ecuador, site of this paper's study area (**Figure 1**). These forests comprise 10% of the country's territory, but they contain half of the country's species, of which 39% are endemic. In the previous 5 years alone, new species of frogs (Guayasamin et al., 2019), bamboo (Clark and Mason, 2019), lizards (Reyes-Puig et al., 2020), hummingbirds (Sornoza-Molina et al., 2018), and orchids (Romero et al., 2017) have been discovered in Ecuador's CFs.

Huston's Dynamic Equilibrium Hypothesis helps explain the explosion of endemic biodiversity in tropical CFs (Huston, 1979). The hypothesis proposes optimal diversity perches between environmental productivity and disturbance, between sustenance and stress.

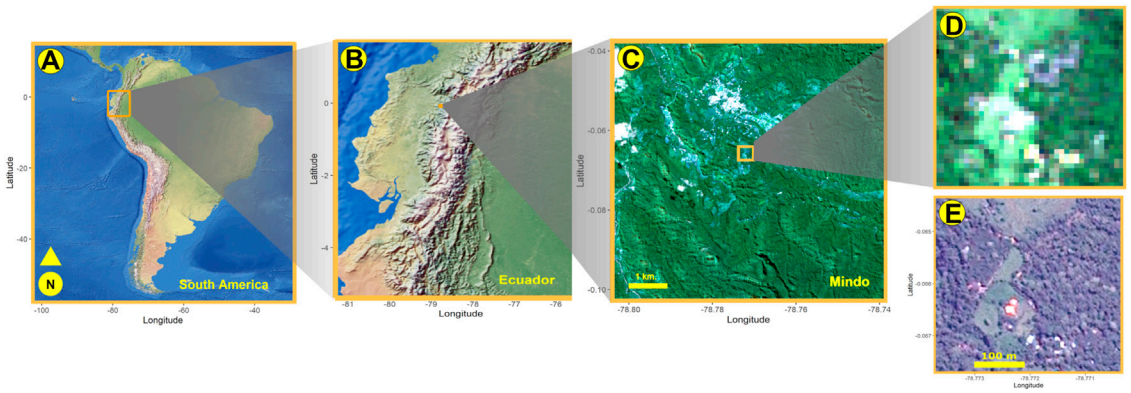



**FIGURE 1 | (A)** Running north and south across South America, the Andes supports 15 and 12 percent of globally known plant and vertebrate species, respectively. The study area is in northern Ecuador **(B)**, centered around a pueblo called Mindo, for which Sentinel-2 satellite rasters were cropped **(C)**. From these rasters, we extracted tiles of 38 × 38 pixel resolution **(D)**, with each pixel having a spatial resolution of 10 $m^2$ To manually classify individual pixels, we compared these Sentinel-derived tiles to higher resolution Airbus satellite images **(E)**.

**FIGURE 2 |** Mountainous, cloud forests **(A)** completely surround our study area in Mindo, Ecuador. Ecosystem services from these forests include **(B)** food, such as plantains; **(C)** building materials and carbon sequestration by *Guada angustifolia*, the native bamboo; **(D)** water; and biodiversity of epiphytes, such as orchids, and vertebrate species, such as hummingbirds **(E)**.

In tropical CFs, sustenance and stress are caused by these forests' low latitude and high altitude. At low latitudes near the equator, sustenance is spurred by year-long growth, uninterrupted by winter, and by relatively constant thermal conditions. Sustenance is further enhanced in the high altitudes of mountains by ubiquitous clouds and fog, which provide moisture for epiphytes—such as moss, lichens, bromeliads, and orchids—that have evolved the ability to draw water from the atmosphere (Gradstein et al., 2008).

But without stress in an ecosystem, only few species would dominate by outcompeting all others for resources. In tropical climates, stress is generated when high precipitation leaches minerals from the soil, lowering its fertility. Competition for nutrients prevents dominance by the few, thereby creating opportunities for the many to survive and for biodiversity to thrive.

Biodiversity is boosted further by the topographic structure of CFs. Mountains occupy a lower surface area than lowlands; they are physically separated from other mountains; and each mountain is distinct from other mountains in terms of slope, height, and lighting conditions. These topographic features create plant and animal communities that are small and isolated, where populations can diverge genetically to form endemic taxa (Myster, 2020).

Beyond biodiversity, CFs provide other ecosystem services (**Figure 2**). They control soil erosion and regulate water supply by

intercepting, retaining, and filtering rainfall. CFs surrounding Quito, Ecuador's capital of 1.85 million people, provide the majority of the city's water supply (Bubb et al., 2004).

Researchers concluded that contributions by CFs to carbon sequestration have been underestimated because their vegetative cover lies vertically on slopes, rather than the planar, horizontal orientation traditionally used to estimate coverage of carbon sinks (Spracklen and Righelato, 2016). In particular, *Guada angustifolia*, the bamboo indigenous to South America, thrives in CFs and is notably effective at converting carbon dioxide to plant matter (Muñoz-López et al., 2021). Because of its tensile strength, *Guada* is also used as a primary source of building materials.

Given the value of their ecosystem services, it is indeed tragic that CFs are being threatened by climate change and habitat destruction. With temperatures predicted to rise by four degrees centigrade, the cloud uplifts to which life in CFs have evolved are dissipating (Helmer et al., 2019).

Beyond climate change, other human activities are having a more direct and immediate influence on CF destruction. From 2016 to 2017, Ecuador's Ministry of Mining increased mining concessions from 3 to 13 percent of the country's continental land area, with the majority of this territory occurring within CFs (Roy et al., 2018).

The 2019 outbreak of the Corona virus has made this bad situation worse. The world's unstable economy has driven up gold prices to record highs. Without jobs and money, desperate illegal miners in South America have taken advantage of pandemic lockdowns and their government's focus on Covid-related activities to expand their activities in cloud forests (Brancalion et al., 2020).

A literature review noted recently that researchers increasingly recognize the global threats to forests, as well as the important role these forests play in creating a sustainable world. To investigate forest loss and deforestation, these researchers have been relying heavily on digital technologies—such as Remote Sensing and Artificial Intelligence (Nitoslawski et al., 2021). In this project, we based much of our investigations on these technologies.

Broadly defined, REMOTE SENSING refers to any technique obtaining information about an object without physical contact (Aggarwal, 2004). In the case of satellites, information comes in the form of fluxes in electromagnetic radiation from the remotely-sensed object to the satellite sensor.

The technology is based on two capabilities of remote-sensing satellites (Boyd, 2005). They emit electro-magnetic energy of various wavelengths; and they gather information about the distinctive manner in which an individual object reflects back the energy from each wavelength, so that its singular reflectance curves can be used to identify the object's nature.

These capabilities have extended the scientific community's ability to observe the world. For example, remote sensing is now being put to field-scale and site-specific agricultural use (Huang et al., 2018); to identify and classify marine debris on beaches (Acuña-Ruz et al., 2018); to assess the adequacy of environmental conditions for freshwater fisheries (Dauwalter et al., 2017); to map areas of severe soil erosion (Sepuru and Dube, 2018); and to track oil spills in the ocean (Fingas and Brown, 2018).

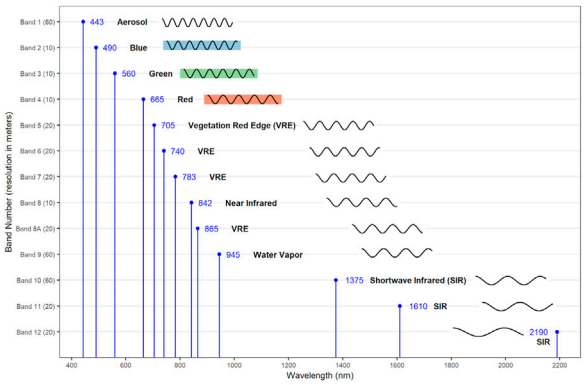



**FIGURE 3 |** European Space Agency's Sentinel-2 series collects information across 13 bands of the electromagnetic spectrum, sampled at four different spatial resolutions. These bands and resolution are shown on the *y*-axis. The *x*-axis shows the central wavelength value for each individual band. Bands 2–4, respectively, cover the blue, green, and red bands of light visible to humans. Bands 1, 9, and 10 are used to correct for atmospheric distortion. The remaining bands are designed for specific uses, such as detecting vegetation and moisture.

In 2015, the European Space Agency (ESA) launched the first of the series of Sentinel-2 satellites, which contain sensors for 13 bands of spectral wavelengths (Kaplan and Avdan 2017). **Figure 3** summarizes the portion of the electromagnetic spectrum for which each band was designed, as well as the spatial resolution at which their data are collected. These bands can be classified according to their respective foci. Bands two through four extend across the red, blue, and green portions of the electromagnetic spectrum corresponding to colors perceived by humans. Bands 1, 9, and 10 have been designed to detect information for correcting distortions from atmospheric and cloud conditions. (Note that as of 2019, ESA adjusted its methods, so it no longer provides data on Band 10.) The remaining bands have been designed to collect data for specific purposes, such as detecting vegetation and moisture.

Once obtained, the values of these bands can be manipulated to compute indicators—or indices—of land use and land cover (LULC) classes. For example, one such index is the Normalized Difference Vegetation Index (NDVI) for delineating vegetative cover (**Figure 4**). NDVI is based on plant biology (Segarra et al., 2020). To carry out photosynthesis, green plant cells absorb solar energy in the visible light region of 400–700 μm wavelength, known as the region of photosynthetically active radiation (PAR). To avoid cellular damage, they reflect near-infrared light (NIR) wavelengths of 700–1,100 μm because NIR energy is too low to synthesize organic molecules; its absorption would merely overheat and possibly destroy plant tissue. NDVI capitalizes on this evolutionary artifact. It is the ratio of the difference between and the sum of the band values at the PAR and NIR regions. The NDVI enables researchers to translate remotely sensed band values into indicators of plant cover.

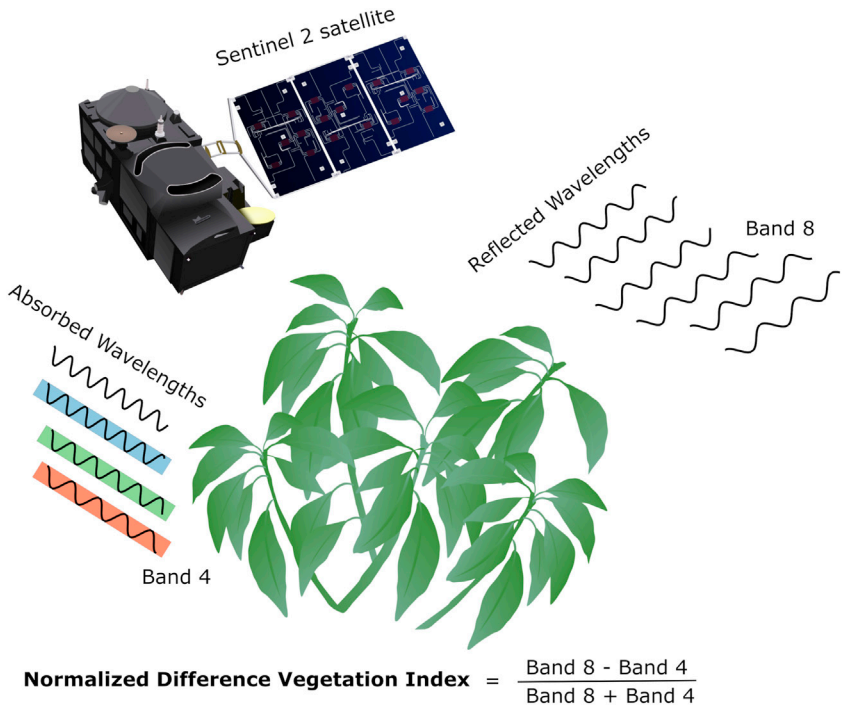



**FIGURE 4 |** Raster indices capitalize on the fact that different objects absorb and reflect different regions of the electromagnetic spectrum. For example, green plants absorb energy in the 400–700 μm wavelengths to photosynthesize, while reflecting wavelengths of 700–1,100 μm. To serve as an indicator for vegetation, the Normalized Difference Vegetation Index calculates the ratio of the difference between and sums of Sentinel's Bands 4 and 8.

**TABLE 1 |** These Raster Indices derive from calculations based on Sentinel-2 band values. Using the formulas in column 4, we estimated the indices in Column 1 to serve as predictors for our Random Forest models. In our project, we categorize them according to the Land Use and Land Cover type for which they were developed (column 3).

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| **Index** | **Acronym** | **Type** | **Formula** | **References** |
| Normalized Difference Vegetation Index | NDVI | vegetation | (b8 – b4)/(b8 + b4) | Rokni and Musa (2019) |
| Normalized Pigment Chlorophyll Ratio Index | NPCRI | vegetation | (b4 – b2)/(b4 + b2) | Peñuelas et al. (1994) |
| Shadow Index | SI | vegetation | sqrt((256 – b2) * (256 – b3)) | Ono et al. (2010) |
| Normalized Difference Snow Index | NDSI | water | (b4 – b6)/(b4 + b6) | Gascoin et al. (2020) |
| Normalized Difference Water Index | NDWI | water | (b3 – b8)/(b3 + b8) | Du et al. (2016) |
| Modified Normalized Difference Water Index | MNDWI | water | (b3 – b11)/(b3 + b11) | Xu (2006) |
| Bare Soil Index | BSI | soil | ((b11 + b4)–(b8 + b2))/((b11 + b4)+(b8 + b2)) | Nguyen et al. (2021) |
| Normalized Built-up Area Index | NBAI | built area | ((b12 – b8)/b2)/((b12 + b8)/b2) | Bouhennache et al. (2019) |
| Normalized Difference Built-up Index | NDBI | built area | (b11 – b8)/(b11 + b8) | Zha et al. (2003) |

**Table 1** summarizes the raster-based indices we used in this project, along with their formulas and pertinent references. It will be useful, for later analyses, to categorize these indices by type, according to the purposes for which they were developed, such as to identify vegetation or moisture. These types are listed along **Table 1**, Column 3.

Many researchers have used these raster indices to classify LULC classes. Ma et al., for example, used three indices—the BSI, the NDVI, and the MNDWI—to model urbanization across several cities in China (Ma et al., 2019). Wasniewski et al. used the NDVI and digital elevation maps to map forest cover and forest type in Western Africa (Waśniewski et al., 2020). To develop the predictive model for their project, these latter researchers relied on Sentinel-2 images, together with the Random Forest classifier, a machine learning algorithm.

RANDOM FORESTS are called forests because they aggregate the results of many individual decision trees (Schonlau and Zou, 2020). To construct a decision tree, the analyst assumes a predictive function relates the response variable Y to the vector of predictor variables X. The decision tree algorithm maximizes the joint probability P(X,Y) by iteratively minimizing an associated loss function. In a later section, in relation to **Figure 13**, we discuss the graphical depiction of the algorithm's logic.

Because the response variable Y is defined, training a decision tree risks over-fitting, a problem that sometimes besets empirical models (Pothuganti, 2018). Over-fitting occurs when the model is so dependent on the data used for training that it generates poor results when applied to other, new data. This happens because the model estimates not only the systemic relationship between predictors and the response variable, but also the random noise inherent in real-world data.

To constrain over-fitting, Breiman proposed two additional algorithms to supplement decision trees at each node (Breiman, 2001). The first randomly selects rows across the set of observations. The second randomly selects a subset of predictors in each randomly selected row.

In the Random Forest algorithm, any individual decision tree might be biased towards the particular dataset upon which it has been trained. But if analysts create a forest of thousands of such trees—with each tree modeling P(X,Y) based on randomly chosen observations—they could then pool the results from each tree in the forest, thereby reducing the influence of noisy information. To put it differently, if analysts consistently get the same model results from iteratively and randomly sampled data, then they can more confidently conclude the results are based on signals rather than noise. The algorithm's randomizing features increase their predictive capabilities. We use this principle of randomization later, when we run simulations from an ensemble of models.

Previous studies have usefully applied Random Forests to Sentinel-2 images. For example, Ghorbanian et al. achieved accuracy rates of 85% in identifying pixels of mangrove plants (Ghorbanian et al., 2021). They report that the lowest rates of accuracy occurred in areas with mixed or boundary conditions, such as areas where mangroves tended to mix with mudflats or shallow waters. They therefore speculated that accuracy might be improved with satellite images of finer spatial scale than the 10 $m^2$ resolution of Sentinel images. In another study, the investigators applied Random Forest to Sentinel-2 images to classify LULC classes (such as cropland, water, forests, and bare soil) in Croatia, achieving an overall accuracy of 89% (Dobrinić et al., 2021). The investigators noted the usefulness of predictors such as raster indices and landscape textural features.

One other feature of Random Forest enhances its utility in this project. Because the algorithm randomly selects subsets of predictor variables at each individual tree, they are not heavily influenced by correlated predictors. That is, multicollinearity does not affect the predictive capacity of models produced by random forests, thereby enabling us to use, as predictor variables, both the computed raster indicators listed in **Table 1**, along with the original band values upon which they are based.

The models produced by the random forest algorithm can then be run against the predictors to generate simulations. Running random forests and generating simulations across voluminous rows and columns of data can be time-consuming. In our project, we ultimately had to analyze 1.56 $\times$ $10^6$ rows of observations with 58 columns of predictors. CLOUD TECHNOLOGY provided us with the tools we needed to analyze these data.

We received a grant from Microsoft's Artificial Intelligence for Earth program, which provided us with a cloud platform enabling us to parallelize many of our calculations in a virtual machine with 14 CPU cores and 64 gigabytes of memory. The grant also provided us access to time-stamped and geocoded Airbus satellite rasters at 1.5 $m^2$ resolution that we used as reference images against which to compare our Sentinel images.

Cloud technology further enabled us to integrate computational efficiency with data integrity. One of our project's objectives is to provide a low-cost method for researchers and environmental groups to apply our algorithms to monitor other vulnerable forests in the world. For this reason, we developed an R package codifying our methods. Our package, loRax, is available on the web at http://pax.green/lorax/.

We hope other researchers will use loRax to share data and metadata from their own projects. To preserve data integrity, we developed an algorithm to encode our data and metadata within blockchains. The technical foundation for bitcoins and other digital currency, the blockchain algorithm provides a way to maintain data provenance while also minimizing data tampering. Data are linked, block by block in a chain, so that each block connects to the previous one with a hash value, i.e., a randomly generated key that maps to the data in the previous block (Beck et al., 2017). The blockchain algorithm validates the series of hash values along this chain of blocks and detects any attempt to corrupt or replace the information within any block. Through this algorithm, data providers who use our R package can maintain proprietorship and provenance over their data.

# MATERIALS AND METHODS

## Study Area

The Andes cordilleras run down Ecuador's central spine from north to south, a narrow (150–180 km wide) stretch of mountains about 600 km long. The study area lies on the northern end of this length, on the western foothills of the volcano, Mount Pichincha. The area centers around a pueblo called Mindo, and it lies within one of the world's five "megadiversity hotspots," a term referring to areas containing 5 $\times$ $10^3$ vascular plants species per $10^4$ $km^2$.

The study area, which hereinafter we refer to as Mindo, covers 52.13 $km^2$. It has an average elevation of 1.49 kms and is bounded by the following longitudes and latitudes in the NW and SE, respectively: (−78.80, −0.04) and (−78.74, −0.1).

## Data and Tools

**Figure 5** provides an overview of our methods and our workflow for this project. We derived our data from three sources of satellite information: the European State Agency's series of Sentinel-2 satellite images; the United States National Aeronautics and Space's Administration (NASA's) elevation data, estimated by its Global Digital Elevation Model (GDEM); and Airbus, a private aerospace company that provided us with satellite images with pixel resolution of 1.5 $m^2$. At 10 $m^2$ resolution, images from Sentinel rasters can be difficult to interpret. We therefore used the higher resolution of Airbus images to compare with the Sentinel rasters (see **Figures 1D,E**).

Our primary analytical tool was R, the public-domain statistical package (R Core Team, 2020). We also used SNAP,

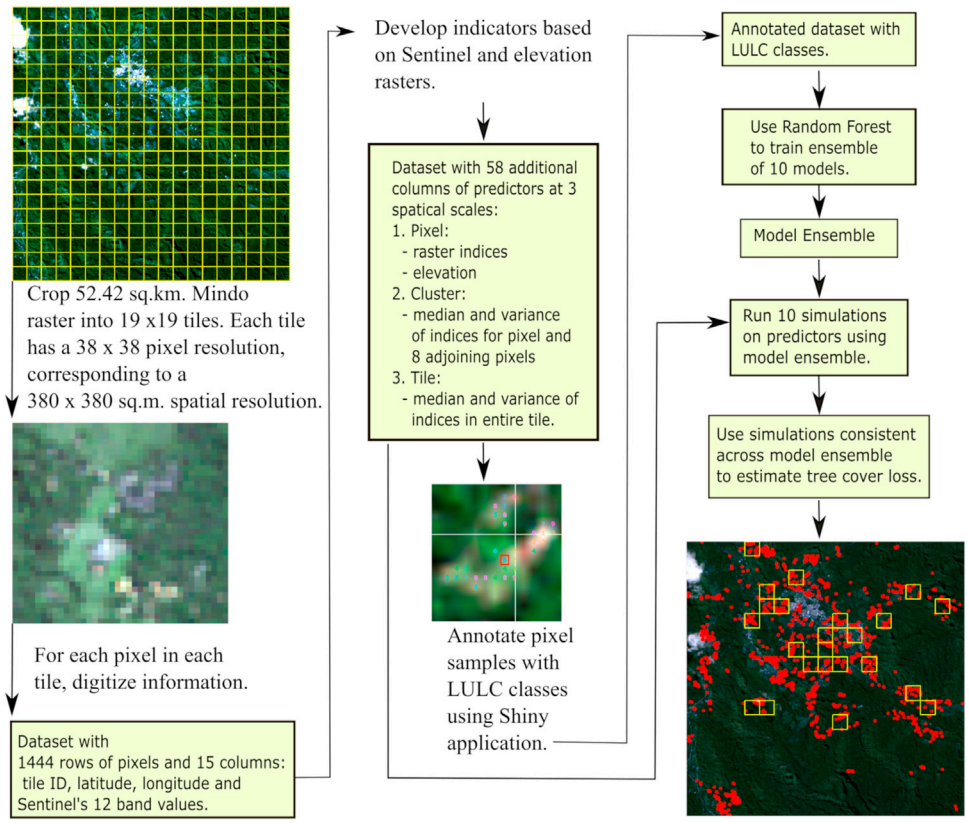



**FIGURE 5 |** This project takes Sentinel-2 rasters as its input. From these rasters, we estimate predictors based on Sentinel's 12 band values, as well as raster indices and elevation. To derive spatial information, we estimate the median and variance of the raster indices at three spatial scales: pixel-, cluster-, and tile-levels. We use annotated images to train an ensemble of 10 models using the Random Forest algorithm. With this model ensemble, we select pixels for which consistent simulations have been made across all 10 models. We then use these simulations to map areas of tree cover loss.

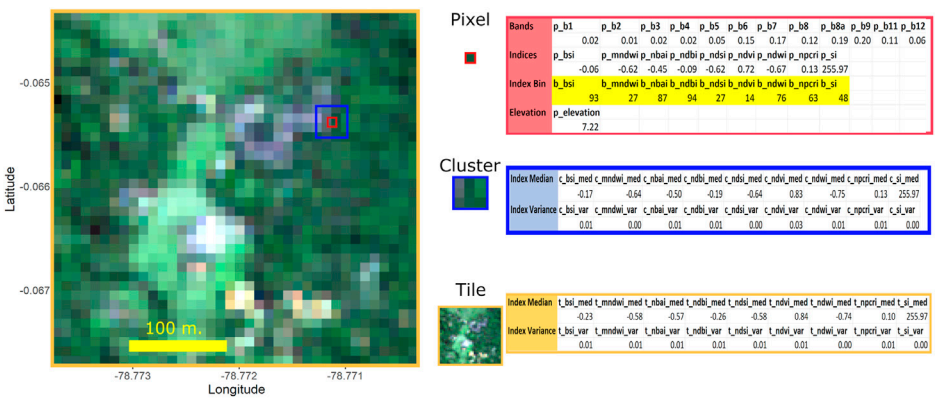



**FIGURE 6 |** The raster above is the same one featured in **Figure 1D**. For each pixel (such as the one above enclosed within a red square), we developed predictors at three spatial scales: the individual pixel; the cluster of nine adjoining pixels; and all pixels within the tile. At the pixel-level, predictors consist of the 12 normalized band values; 9 raster indices; and the logarithm of the elevation level. At the cluster-level, we took the median and variance for each raster index across the nine pixels within the cluster. At the tile-level, we took the median and variance of each index across all pixels within the tile.

**TABLE 2 |** We developed a Model Ensemble based on 120 iterations of the Random Forest algorithm on our annotated dataset. We sampled and manually classified 2.65 × $10^5$ pixels. For each iteration, pixel numbers for training, testing, and oversampling are listed in columns 3 through 5. To evaluate the accuracy of our model ensemble, we developed Confusion Matrices for two types of test data. In column 6, we combined all test results across all iterations. In column 7, we selected only those pixels that were classified consistently across at least 10 iterations.

| For each individual iteration | | | | | Across all iterations | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Land use and land cover class** | **No. of annotated pixels** | **No. of pixels for training** | **No. of pixels for testing** | **No. of pixels for oversampling** | **No. of pixels across all iterations** | **No. of pixels restricted by consistent classification across at least 10 iterations** |
| water | 6,132 | 4,292 | 1839 | | 220,680 | 217,388 |
| built | 7,221 | 4,292 | 1839 | 1,090 | 220,680 | 218,803 |
| cloud | 12,365 | 4,292 | 1839 | | 220,680 | 219,523 |
| soil | 17,539 | 4,292 | 1839 | 5,000 | 220,680 | 101,136 |
| grass | 58,268 | 4,292 | 1839 | 5,000 | 220,680 | 10,136 |
| tree | 164,041 | 4,292 | 1839 | 5,000 | 220,680 | 146 |
| TOTAL | 265,566 | | | | | |

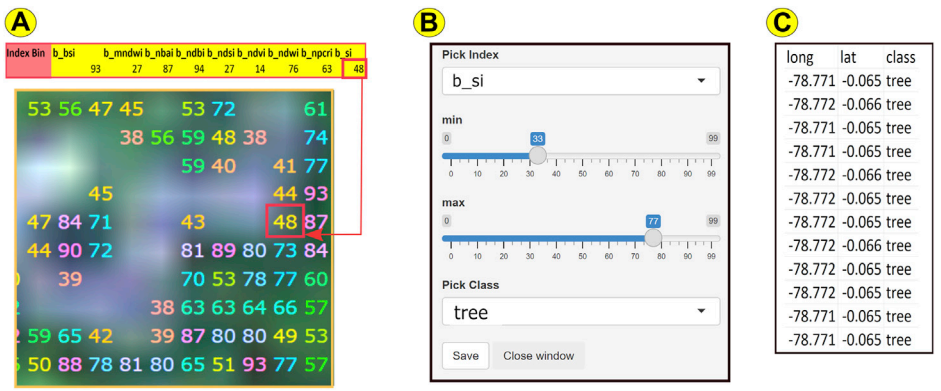



**FIGURE 7 |** The raster above **(A)** is the upper-right cross-section of the raster featured in **Figure 6**, with a red square around the same pixel from **Figure 6**. Within each tile, for each raster index for each pixel, we took the 99% percentile value of the index, which we refer to as the Index Bin. For example, the selected pixel had an SI (shadow index, used to monitor tree canopy) value of 255.97 (see **Figure 6**). The pixel has an SI Index Bin (b_si) value of 48. To annotate or tag a pixel, we developed a Shiny, interactive application **(B)** for our R package, *loRax*. In the example above, we selected "b_si" from the drop-down box and manipulated the scroll bars to select minimum/maximum values of 33/77. Colored numbers corresponding to the Index Bin values of each pixel within this range appeared on the raster image **(A)**. After comparing the resulting image with the higher resolution Airbus image (see **Figures 1D,E**), we selected the class "tree" from another drop-down list **(B)** to save the class and location of the selected pixels to file **(C)**.

a software application developed by ESA to process, analyze, and visualize the Sentinel rasters. We codified our data, algorithms, and workflow in an R package, loRax, information about which is available at: http://pax.green/lorax/.

## Select and Process Sentinel Satellite Images

We searched ESA's data hub for satellite rasters for Mindo having the least amount of cloud cover. To minimize seasonal differences across the three time periods we downloaded, we tried choosing rasters for months that were as close to each other as possible, but this desideratum was constrained by the availability of relatively cloud-free rasters for Mindo. Listed below are the monitoring dates and filenames of the Sentinel raster we downloaded. Ultimately, to maximize our number of cloud-free pixels, we decided to compare only two time periods, 2019 and 2021. Rasters can be accessed by filename from the European Space Agency at: https://scihub.copernicus.eu/dhus/#/home.

- 2019-08-30 (S2B_MSIL2A_20190830T153619_N0213_R068_T17MQV_20190830T205757.SAFE)
- 2020-08-24 (S2B_MSIL2A_20200824T153619_N0214_R068_T17MQV_20200824T193939.SAFE)
- 2021-07-05 (S2A_MSIL2A_20210705T153621_N0301_R068_T17NQA_20210705T213848.SAFE)

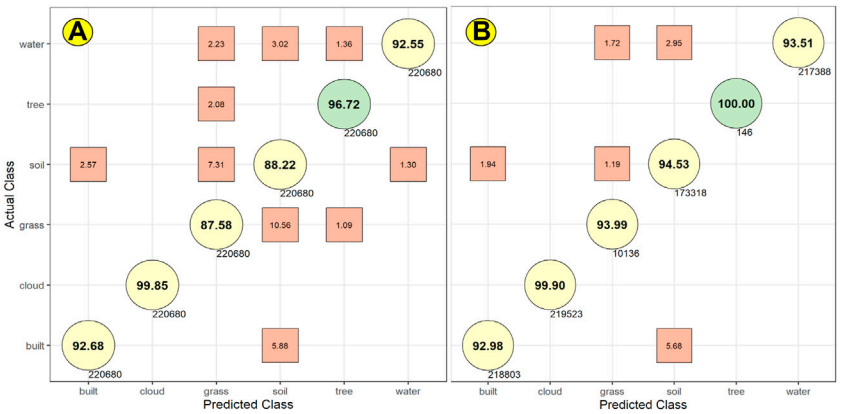



**FIGURE 8 |** These Confusion Matrices are based on results from Random Forest tests for two different types of data (see **Table 2**). In **(A)**, we estimated accuracy scores based on all the test data derived from the 120 iterations we ran of Random Forest. In **(B)**, we selected pixels that were classified consistently across at least 10 iterations.

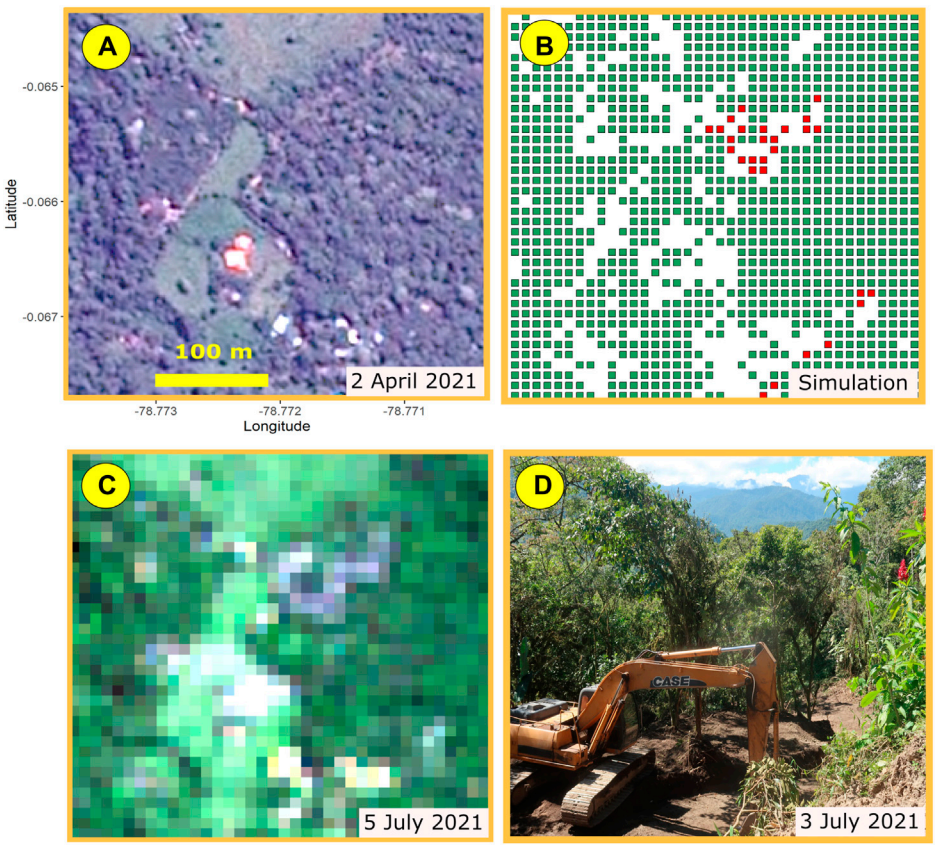



**FIGURE 9 | (A)** Airbus mid-resolution image of a Mindo tile for which **(B)** the model simulation showed tree cover loss in the tile. The Airbus image (taken on April 2021) does not show tree cover loss. But the model was based on a Sentinel raster **(C)** taken on July 2021. A site visit clarified the discrepancy. Construction of a new subdivision broke ground **(D)** after the Airbus image was taken in April. Sentinel's July image and the Random Forest model detected the more recent pattern of tree cover loss.

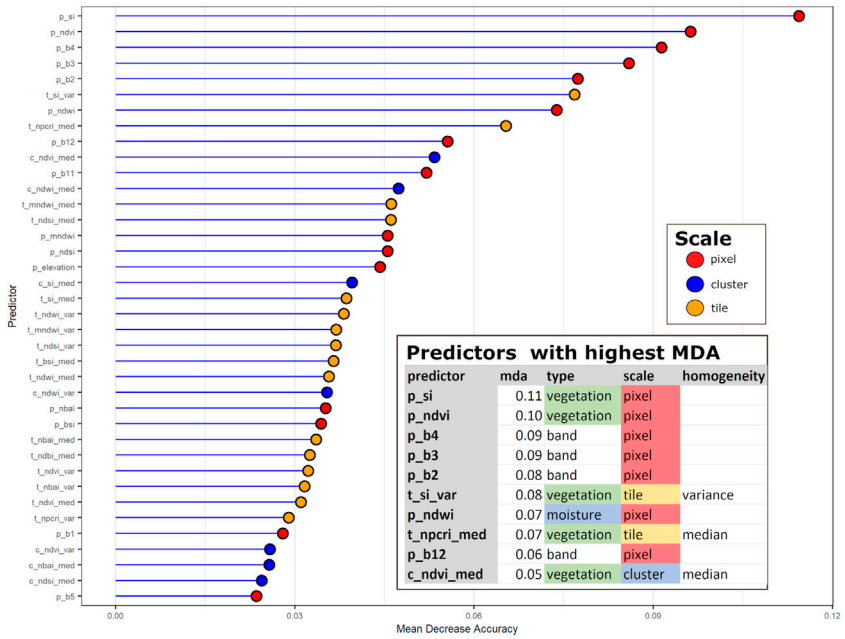



**FIGURE 10 |** Mean Decrease Accuracy measures each predictor's influence on the LULC classes simulated by the Random Forest Model. Each predictor has three attributes: type, scale, and homogeneity. Type refers to how they are used (e.g., to detect vegetation, water, etc.). Scale is the spatial level at which each was measured. Homogeneity is the median or variance of each raster index at the cluster or tile levels. The most influential predictors are shown inset. They are a mix of predictors varying by type, scale and homogeneity.

To prepare these rasters for analysis, we used ESA's SNAP software to resample and reproject the rasters to a 10 $m^2$ pixel resolution with the World Geodetic System (WGS84) coordinate system. We then exported the results to GeoTIFF format for further analysis with R.

## Crop Raster Into Tiles

Each raster listed in above covers an area in northern Ecuador with spatial resolution of 110 × 110 kms. From each of these, we cropped rasters for our study area in Mindo. Each Mindo raster had a spatial resolution of 7.22 × 7.22 kms and pixel resolution of 722 × 722 pixels. To create even smaller rasters that were more computationally manageable, we further cropped each Mindo raster into tiles (see **Figures 1C,D**). Each tile has a pixel resolution of 38 × 38 pixels, corresponding to a spatial resolution of 380 × 380 m. Each Mindo raster supplies us with 361 tiles, and each tile provides us with 1,444 (38 × 38) rows of data. Ultimately, our dataset consisted of $1.56 \times 10^6$ rows, i.e. (1444 pixels per tile) x (361 tiles per Mindo raster) x (3 time periods per raster). Ultimately however, to maximize the number of cloud-free pixels we could track over time, we decided to use only the rasters for 2019 and 2021.

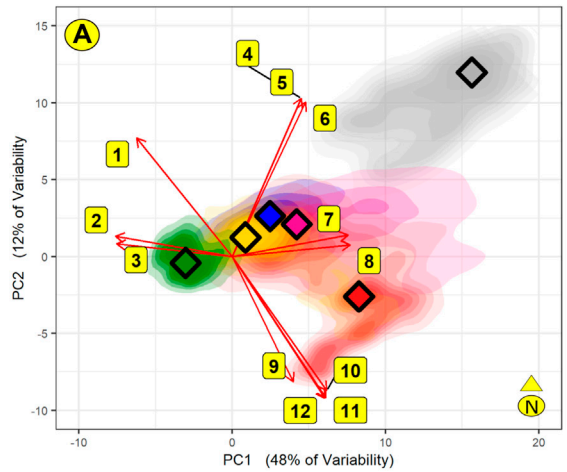
## Transform the Information in Each Tile Into a Set of Predictors

To prepare for our Random Forest model, we converted the information in the GeoTIFF of each tile into a set of predictor variables. **Figure 6** summarizes our approach (The tile in **Figure 6** is the same one shown in **Figure 1D**). All in all, we developed 58 predictors representing information for each pixel at three different spatial scales.

At the scale of individual pixels, we extracted the numeric values of each of the 12 bands recorded by the Sentinel raster and calculated the normalized values of each band. From the original, non-normalized band values, we calculated the various raster indices shown in **Table 1**. Lastly, we calculated the logarithm of the elevation, in meters, provided by NASA's GDEM. To implement the annotation method we describe in the next section, we also needed to calculate the 99-percentile value within *each* tile for each raster index for each pixel. In the example in **Figure 6**, we highlight these values in yellow and refer to each as an "Index Bin."

We hypothesized we could achieve a higher prediction accuracy if we captured spatial information about our pixel-level predictors. We therefore calculated indicators of raster index homogeneity at two other spatial scales. At a cluster level—for each pixel and its eight adjoining pixels—we calculated the median and variance of each raster index. At the tile level, we performed the same calculations for median and variance of each raster index across all pixels in the tile.

In **Figure 6** and in the rest of the paper, we use the following the notation to refer to predictors. We use the prefix "*p*", "*c*", and "*t*" to refer to the scale (pixel, cluster, and tile) of the predictor. This is followed by the acronym for each index (e.g., NDVI, BSI, SI, etc.) as denoted in **Table 1**. For the cluster- and tile-level predictors, we use the suffix "*med*" or "*var*" to refer to whether the predictor represents the median or variance of the group of pixels.

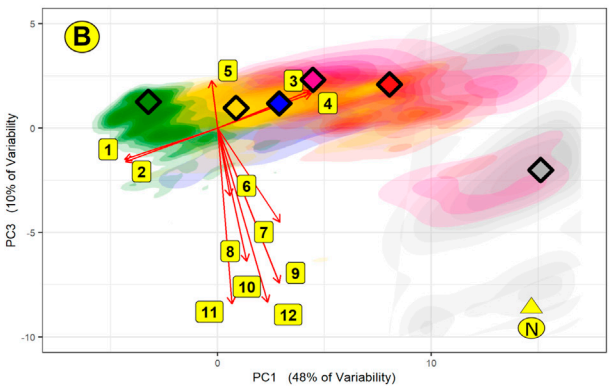



**Predictors**

| label | quadrant | predictor | type | scale | homogeneity |
|---|---|---|---|---|---|
| 1 | W | t_ndvi_med | vegetation | tile | median |
| 2 | W | c_ndvi_med | vegetation | cluster | median |
| 3 | W | p_ndvi | vegetation | pixel | |
| 4 | NE | p_ndsi | moisture | pixel | |
| 5 | NE | p_mndwi | moisture | pixel | |
| 6 | NE | c_ndsi_med | moisture | cluster | median |
| 7 | E | p_ndwi | moisture | pixel | |
| 8 | E | c_ndwi_med | moisture | cluster | median |
| 9 | SE | t_npcri_med | vegetation | tile | median |
| 10 | SE | t_nbai_med | built | tile | median |
| 11 | SE | t_bsi_med | soil | tile | median |
| 12 | SE | t_ndbi_med | built | tile | median |

**Class**

tree grass water
built soil cloud




**Predictors**

| label | quadrant | predictor | type | scale | homogeneity |
|---|---|---|---|---|---|
| 1 | W | c_ndvi_med | vegetation | cluster | median |
| 2 | W | p_ndvi | vegetation | pixel | |
| 3 | E | p_ndwi | moisture | pixel | |
| 4 | E | c_ndwi_med | moisture | cluster | median |
| 5 | N | p_elevation | elevation | pixel | |
| 6 | S | t_npcri_var | vegetation | tile | variance |
| 7 | SE | t_nbai_var | built | tile | variance |
| 8 | S | t_si_var | vegetation | tile | variance |
| 9 | SE | t_ndvi_var | vegetation | tile | variance |
| 10 | S | t_ndsi_var | moisture | tile | variance |
| 11 | S | t_mndwi_var | moisture | tile | variance |
| 12 | SE | t_ndwi_var | moisture | tile | variance |

**FIGURE 11 |** We conducted a PCA on our annotated dataset. The first three dimensions (PC1, PC2, and PC3) explain 70% of the data variability. Individual PCA scores are shown above as differently colored density clouds, with the highest density points for each class shown as diamonds. The Random Forest's high accuracy scores for cloud and tree conform to the observation that their PCA scores cluster distinctly at the east and west quadrants of the PCA. Contrariwise, the individual scores for the other classes merge closely together near the point of origin. PC1's west and east quadrants are dominated by loadings related to vegetation and moisture, respectively. Both PC2's and PC3's southern quadrant is dominated by tile-level measures of homogeneity—median and variance for PC2 and PC3, respectively.

So, for example, the notation "*t_ndvi_med*" refers to the NDVI's median value across all pixels in a tile.

## Annotate Tile Samples With Land Use and Land Cover Classes

Before training our model, we first needed to annotate samples from our data. To perform our annotations, we developed an interactive application using R's Shiny package. An interactive demonstration of the Shiny app is available at http://pax.green/lorax/. The application is summarized in **Figure 7** (The raster in **Figure 7A** is the upper-right cross-section of the raster from **Figure 6**. The pixel within the red square is the same pixel featured in **Figure 6**).

As we stated earlier, for each raster index for each tile, we categorized each pixel value into 0-99 percentiles, which we refer to as an Index Bin. For example, in **Figure 7A**, the selected pixel's value for the Shadow Index (si)—a raster index used to estimate tree canopy cover—falls into the 48th percentile, or Index Bin.

Using loRax's Shiny application, we selected "b_si" (for binned Shadow Index) from the drop-down list that allows one to choose from any of the indices listed in **Table 1** (see **Figure 7B**). We then manipulated scroll bars to choose 33 and 77 as the minimum and maximum values, respectively, for the selected tile's *Index Bins* for SI. On the tile's satellite image (**Figure 7A**), pixels with *Index Bin* values between and including these min/max numbers interactively appeared as colored numbers (their *Index Bin* values). We then compared this image to the higher resolution image from Airbus (*see* **Figures 1D,E**), as well as **Figures 9A,C**). Once we were satisfied that the colored pixels corresponded to the corresponding LULC class (in this example, tree cover), loRax enabled us to select the LULC class from another drop-down list (**Figure 7B**) and to save the results—*i.e.* the pixels' LULC class, along with their location (**Figure 7C**)—to a file for inclusion in the dataset for annotated LULC classes.

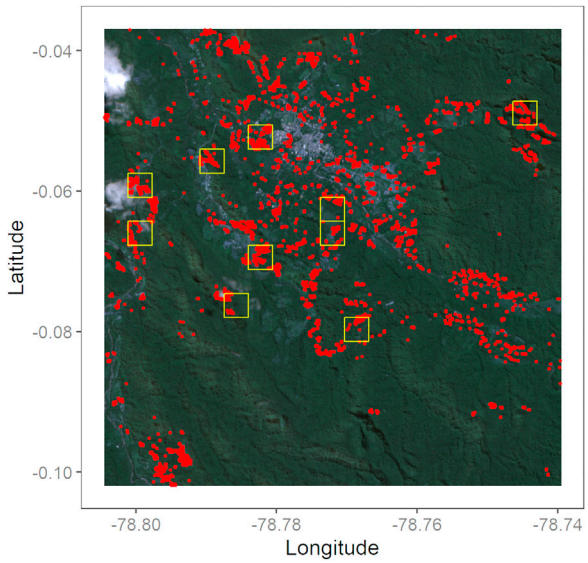



**FIGURE 12 |** Individual red dots show areas of tree cover loss in Mindo. Yellow boxes indicate tiles where tree cover loss equaled or exceeded 5% of the tile area. Based on our model results, the total area of tree cover loss between August 2019 and July 2021 is 0.61 $km^2$—1.17% of the study area.

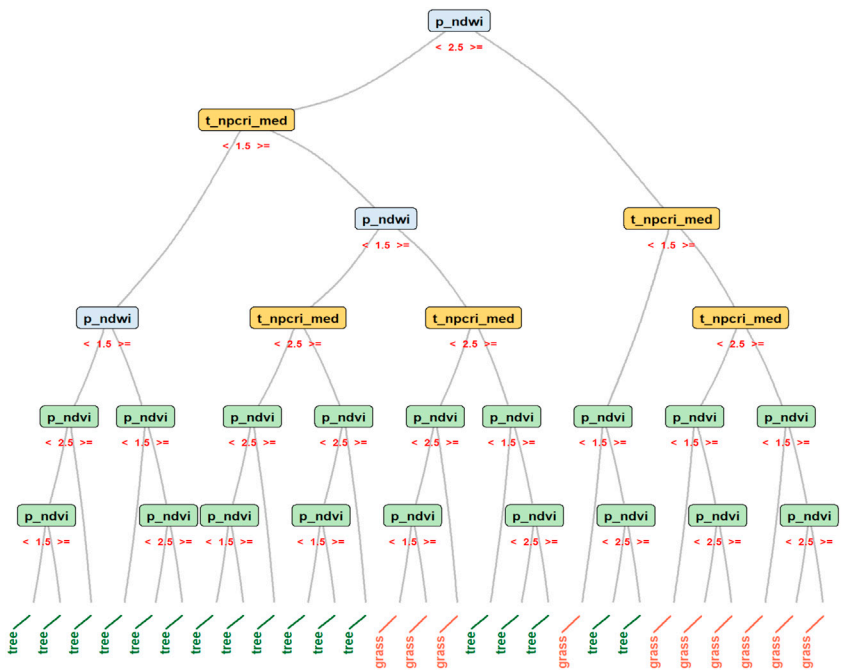



**FIGURE 13 |** This graphic visualizes the Random Forest Algorithm's logic for a single Decision Tree for a simplified annotated dataset. The original dataset was filtered for data with the LULC classes of "tree" and "grass;" only three predictors were selected; and values of the predictors were grouped into three percentile bins. In this example, rules govern whether a branch in the Decision Tree ultimately leads to a pixel being classified as "grass." Generally, **(A)** *p_ndwi* must be≥2.5; and **(B)** *t_npcri_med* must be≥1.5. If **(A)** is true and **(B)** is false, then *p_ndvi* must be <1.5. If **(A)** and **(B)** are false, then *p_ndwi* must be >1.5 and *p_ndvi* must be <2.5. Simulations from our ensemble of Random Forest models are the net, probabilistic result from thousands of such trees for all predictors and response variables in our annotated dataset.

## Train the Random Forest Models and Run Simulations

These annotated LULC classes served as the response variables for the Random Forest algorithm, and the 58 variables shown in **Figure 6** served as predictors. Before running the algorithm, we split the data into two bins—one for training and another for testing the model. **Table 2** summarizes the number of annotated pixels within each class. In total, we had 265,566 pixels in our annotated dataset, with six classes: tree, grass, water, soil, built, and cloud. Water had the fewest annotations, with 6,132 pixels. For each iteration of training our Random Forest model, we took 70 and 30 percent of 6,132 as the number of pixels within each class to respectively train and test the model. Because we had sufficient annotations for the other classes, we over-sampled from these classes, as detailed in **Table 2**.

To create an ensemble of models, we ran Random Forest 120 times, with each iteration's model based on randomly sampled data. The testing data thus produced were of two types. For the first type, we merged the test data from across the 120 iterations, generating 220,680 (120 iterations x 1839 rows of test data per iteration) as shown in **Table 2**'s Column 6. For the second type, we created a second test dataset, one that was restricted by two criteria: 1) the pixel needed to have been simulated by at least 10 iterations; and 2) the classification of the pixel must have been consistent across all 10 iterations (**Table 2**'s Column 7).

For this second, "restricted" dataset, trees had fewer data (146) because it was a class with many annotated pixels (164,041). Each pixel therefore had a low probability of being randomly selected over 120 iterations, making it difficult for this class to fulfill the first criteria of needing to have been simulated for at least 10 iterations.

R's Random Forest package computes the Mean Decrease Accuracy (MDA), a measure of each predictor's influence on model accuracy. After an initial run of the Random Forest using all predictors, we selected those predictors with an MDA value of at least 0.02. Our objective was to choose predictors that were influential, but not to have so many predictors that the models would run the risk of over-fitting.

After running the Random Forest, we applied Principal Component Analysis to our annotated dataset to better understand the influence of each predictor on each LULC class (Jolliffe and Cadima, 2016).

# RESULTS

## Assess Simulation Accuracy With Confusion Matrix

To evaluate accuracy across the simulations from our ensemble of 10 models, we developed the Confusion Matrices shown in **Figure 8**. **Figure 8A**'s Confusion Matrix is based on test data of the first type, which merged all test data across the 120 iterations. In this matrix, the models simulated tree cover correctly 96.72% of the time; they misidentified tree cover as grass 2.08% of the time (for clarity, only error rates greater than or equal to 1 percent are shown).

The highest rates of error of the Random Forest simulations occurred in the models' simulation of grass cover. These grass simulations generated a true positive accuracy rate of 87.58%. For grass, the models generated false positives 11.62% of the time, classifying as grass those pixels that were actually soil (7.31%), tree cover (2.08%), and water (2.23%). The models also generated false negatives for grass, misclassifying as soil (10.56%) and tree (1.09%) pixels that should have been classified as grass.

## Assess Simulation Accuracy With Satellite Images and Field Visits

We tried evaluating simulation accuracy by comparing our simulations against the higher resolution Airbus satellite images. But because of the temporal mismatch between the Sentinel and Airbus rasters, this approach can lead to misleading results. **Figure 9** provides an example of the problems of temporal mismatch.

To calculate loss of tree cover, we filtered our data in two steps. First, we selected only those pixels in Mindo that were cloud-free across two time periods (2019 and 2021). Second, we used only the "restricted" test data. For these data, we defined a pixel as having lost tree cover if the pixel was classified as tree cover in 2019 but was classified as some other class in 2021.

We also identified tiles having a relatively higher proportion of tree cover loss—at least 5%. We refer to these tiles as "tree loss hotspots." To avoid the bias of small samples, we selected only those tiles that retained at least 1,000 pixels after filtering for cloud cover. We then calculated the proportion within these tiles of pixels identified as having lost tree cover.

Given the Covid lock-down, we could only visit "tree loss hotspots" within walking distance of Mindo's central plaza. Fortunately, within one such area, we obtained results enabling us to verify our simulations.

In **Figure 9B**, the model simulation for 2021 indicates loss of tree cover (shown as red pixels) that do not appear in the corresponding Airbus image (**Figure 9A**). A visit to the site clarified the reasons for this apparent discrepancy. The Airbus image was taken in April, while the simulation was derived from the Sentinel raster for July (**Figure 9C**). A site visit enabled us to confirm that new construction had broken ground between April and July (**Figure 9D**).

## Evaluate Influence of Predictor Variables With Mean Decrease in Accuracy

**Figure 10** shows the relative importance of the various predictors, based on the mean decrease in accuracy (MDA). This metric estimates relative importance by calculating how removing a predictor affects prediction accuracy. As discussed in **Figure 4** and its accompanying text, each predictor can be characterized by type (i.e., by the purpose for which each index was developed, such as to identify vegetation or moisture); scale (pixel-, cluster-, or tile-levels); and, at the cluster- and tile-level, by measures of homogeneity (median and variance). **Figure 10**'s coloration of points characterizes each predictor by scale. As shown in the figure's inset, the most influential predictors, as measured by

MDA, are a mix of predictors varying by type, scale, and measure of homogeneity.

## Evaluate Influence of Predictor Variables With Principal Component Analysis

Principal Component Analysis (PCA) simplifies high-dimensional, complex data by extracting features of the data and then projecting these features on to lower dimensions (called principal components). These features are extracted and projected one dimension at a time; each dimension cumulatively explains the variability of data features, and each succeeding dimension is estimated independently from preceding dimensions.

We conducted a PCA on the annotated dataset used to train and test our Random Forest model. The first three dimensions, results for which are summarized in **Figure 11**'s biplots, explain 70% of the variability across our dataset. The density of points for each class across the PCA quadrants is shown as color gradients on the biplots, with diamond points indicating each class's center of highest density.

The influence of a predictor along one dimension is measured by its so-called loading, represented on the biplot as the length of each arrow. The angles between arrows show the correlation between predictors. For example, in **Figure 11A**, along the PCA's first dimension, points 2 and 3 (*c_ndvi_med* and *p_ndvi*, the cluster-level median and pixel-level values of NDVI) are closely related. Points 7 and 8 (*p_ndwi* and *c_ndwi_med*, the pixel-level and cluster-level median values for NDWI) are closely related; however, compared to points 2 and 3, they exert an opposite influence on the first dimension. On the other hand, along the second dimension, point 12 lies on the PCA's southern quadrant. This point, representing *t_ndbi_med*, exerts an influence that is orthogonal to any of the points along the first dimension.

## Map of Tree Loss

One of our primary objectives was to produce a map of individual pixels of tree loss and of tiles that are "tree loss hotspots." We have shared this map, shown in **Figure 12** with environmentally-motivated stakeholders in Mindo to help them focus their efforts on these areas of greatest concern.

# DISCUSSION

The convergence of remote sensing, Artificial Intelligence, and cloud computing supplied us with massive amounts of data and with the computing power to process them. We set out to develop a workflow for monitoring deforestation with satellite data and to build a corresponding R package that is scalable and transferrable for use by others to protect forests across the globe. For our tool, loRax, to be credibly implemented, it must deliver results that are accurate, in a way that is relatively transparent.

Artificial Intelligence has been criticized as being a "black box" through which one feeds copious data to be digested by obtuse algorithms, and from which predictive simulations are regurgitated (Castelvecchi, 2016). Although the mathematical computations of the Random Forest can be obtuse, we have tried in this paper to provide an intuitive explanation of its theoretical underpinnings, which we believe can be communicated more easily than those of, say, a Convolutional Neural Network (CNN) (LeCun et al., 2015). Additionally, most pre-trained CNNs have been built on RGB images (Senecal et al., 2019). In the future, researchers may yet develop CNNs that can analyze more than three bands and that would not require the computational resources typically demanded by deep CNNs.

For the present, we hypothesized we could use Random Forest to obtain useful information from the 12 bands of Sentinel rasters. We further hypothesized that we could improve the accuracy of our simulations:

- by developing predictors to capture spatial information; and
- by generating these simulations with a model ensemble.

**Figure 8B**'s Confusion Matrix suggests there is much analytical value to be gained from these two innovations.

## Influence of Individual Predictors on Simulated Classes

As an additional step towards model transparency, we also investigated how the predictors might be influencing model simulations. **Figure 10**'s MDA indicates that the predictors wielding the most influence are a combination of normalized band values; raster indices for vegetation and moisture; and measures of homogeneity of these indices across spatial scales. Out of the 38 predictors used for the Random Forest, 15 were at the pixel level and another 16 at the tile level—9 for the median and seven for the variance. This result confirms the importance of information provided by our predictors of spatial homogeneity. Out of all predictors, 13 were for vegetation-related indices and 12 for moisture-related indices.

**Figure 11**'s PCAs provide additional insight about the predictors. While Random Forests maximize the joint probability between predictor and response variables, PCAs aim to group predictors into fewer dimensions by eliminating information redundancies in the data. In doing so, the PCA provides insight on how individual predictors might be influencing the Random Forest simulations.

The PCAs indicate 70% of variability in the dataset could be accounted for in the PCA's first three dimensions: respectively 48, 12, and 10 per cent in the first, second, and third dimensions. The first dimension is defined by raster indices for vegetation and moisture lying west and east, respectively, of the point of origin (POI). In the second and third dimensions, predictors south of the POI relate to spatial information at the tile level—the tile median in the second dimension and the tile variance in the third. These results conform to **Figure 10**'s MDA indicating the importance of spatial information to complement water- and vegetation-related raster indices at the individual pixel level.

Looking beyond the PCA loadings to the scores for individual observations, we observe that these PCA scores are consistent with **Figure 8**'s Confusion Matrices. **Figure 8** shows simulations for tree and cloud cover to be highly accurate (more than 99%). In

**Figure 11**'s PCA, individual scores for tree and clouds form distinct clusters lying west and east on the PCA quadrants. Contrariwise, data for grass, soil, water, and built classes are widely dispersed around the POI. Although the highest density points for these classes are distinct, their individual points merge with one another across the four quadrants of the PCA. This might account for their relatively lower accuracies in **Figure 8**'s Confusion Matrices. We hypothesize that the electromagnetic, spectral signals sent by objects in the grass, soil, water, and built LULC classes are less distinctive than those sent by the tree- and cloud-cover classes.

At this juncture, one point merits mention. This paper's senior author has co-written several policy- and law-oriented papers on the importance of understanding a model's epistemic framework, i.e., its assumptions and underlying theoretical foundations (Fisher et al., 2014). PCA and Random Forest are built upon different epistemic foundations. That each corroborates the other's results should provide one with a measure of confidence in the models' simulations.

## Visualizing the Random Forest

With a better understanding of the influence of individual predictors on the modeling process, with a mind to enhance the transparency of our modeling process, we now visualize the logic of the Random Forest algorithm. To do so, we trained a model using a highly simplified annotated dataset which we created by modifying the original annotated dataset as follows:

- We subset the response variables to only two classes—tree and grass.
- We subset the predictors to only three—*p_ndwi*, *p_ndwi*, and *t_mndwi_var*.
- We binned the raw predictor values into three percentile groups.

We ran the Random Forest on this simplified annotated dataset to produce a model distinguishing between grass and tree cover. **Figure 13** graphically depicts the logic through which the algorithm produced a single decision tree. The decision tree goes through a series of Yes/No choices based on the value of each predictor, with each tree branch leading to either "grass" or "tree." During the training period with our full, annotated dataset, Random Forest aggregates the results of thousands of such trees to generate the set of rules the models use to classify the different LULC classes.

We glean a plausible narrative from **Figure 13**'s decision tree and **Figure 11**'s PCAs. Although NDVI and NPCRI have both been developed to detect vegetation, **Figure 13** suggests higher NPCRI values are associated with grass cover. **Figure 11B**'s PCA suggests that the tile-level variance of NPCRI exerts an influence on the third dimension that is opposite that of elevation. These results suggest a spatial structure consistent with Mindo's settlement pattern. Tree cover tends to prevail at higher elevations. At lower elevations, in Mindo's valleys, there is greater variability in the distribution of trees and pastureland.

## Addressing Model Uncertainties

Uncertainties are inherent in the modeling process (Wagner et al., 2010). As highlighted in **Figure 9**, a significant source of uncertainty stems from the temporal mismatch between the Sentinel rasters and whichever high-resolution image one uses to verify the simulations. To put it another way, "model errors" may have as much to do with the manual annotations of the training data as it does with the algorithms of Artificial Intelligence.

To constrain the latter type of error, we attempted to deal with its uncertainties in three ways. First, for those LULC classes for which we had sufficient data—soil, tree, and grass—we oversampled for the training set for the Random Forest.

Second, we ran simulations based on an ensemble of models. Each iteration of the Random Forest algorithm was based on randomly sampled data from the annotated dataset. In our earlier discussion on the theory of Random Forests, we noted that if analysts consistently get the same model results from iteratively and randomly sampled data, they can more confidently conclude the results are based on signals rather than noise. We extended this principle to ensemble modeling, and **Figure 8B**'s Confusion Matrix shows promising results.

Third, to identify those areas with tree loss between 2019 and 2021, we selected only those pixels which were classified consistently across all simulations from an ensemble of at least 10 model iterations.

Given these precautions and given that we filtered out pixels that were cloud-covered, we suggest that if anything, our results likely underestimate forest loss in Mindo. We were hampered by our inability to find more cloud-free satellite rasters for Mindo than we would have wanted for our analyses. We suspect this would *not* have been the case for areas thought to offer more commercial opportunities than Mindo, where ecosystem services redounding to global sustainability go largely unpriced by the market.

## Topics for Future Research

There is considerable future work to be done. First, it is generally recognized that data-intensive, place-based models may have limited application beyond the region from which data were collected. **Figure 11**'s PCAs are consistent with Mindo's spatial structure. In Mindo, people have tended to settle in valleys at lower elevations, where the landscape is crisscrossed by rivers, pastures, and homes. This would account for the importance of spatial measures of homogeneity along dimensions 2 and 3, particularly in dimension 3's measures of variance, which have an inverse relationship with elevation. It would be fruitful to investigate how our models would compare to models run in areas with vastly different settlement patterns.

Second, at the code level, we have tried to speed our computations by using parallel-processing whenever possible. We also have tried to use transparent data structures by organizing our data hierarchically and grouping pixels within tiles. We hope that other data scientists can improve the efficiency of our workflow while maintaining clarity and transparency in their algorithms and data structures.

Third, we cropped the Mindo raster into 380 × 380 m tiles because this resolution roughly corresponded to the spatial resolution of Google's satellite images at Google Map's API's 16x resolution. (We did this before being notified we received a grant to purchase Airbus images.) It would be worth exploring whether a different spatial resolution for tiles would generate more accurate results.

Fourth, to generate spatial information, our current algorithms estimate tile- and cluster-level median and variance. We can improve upon these algorithms by clustering pixels within each tile into segments where pixels share similar attributes. The shape and area of the segments might serve as predictors that would provide greater accuracy to our predictions.

Fifth, we are presently trying to classify LULC classes and predictors across the Mindo tiles into some typology of tiles. We then intend to integrate this typology of tiles with bird observations. In Mindo, in a roughly 100-hectare area of forest reserve, 356 species of birds have been identified (Stevens et al., 2021). Our objective is to build a hierarchical, Bayesian model relating Mindo's land attributes with bird counts obtained from Cornell's ebird database (Wood et al., 2011).

Finally, a fruitful ground for social science research is how to use computational models for forest governance. Elinor Ostrom's work on how communities collectively manage shared resources emphasizes the importance of performance measures (Anderies et al., 2013). It would be extremely useful to explore how deforestation models based on public data and Artificial Intelligence can be communicated more clearly and credibly to community stakeholders. As a corollary, it would be worth investigating whether model results can be used as performance measures to encourage community-based tree conservation activities.

## CONCLUSION

To summarize our project in concrete terms, we are accessing data from observations made from a distance of 786 km. (Sentinel's orbiting altitude) to make predictions about whether or not an area half the size of a singles tennis court (roughly the spatial resolution of an individual pixel from Sentinel) is covered by trees. Despite this task's challenges, **Figure 8**'s Confusion Matrices indicate a probability of more than 96% that our simulations of tree cover loss would be correct. To summarize **Figure 12**'s results in concrete terms, in our study area of 52 $km^2$ (about a third the size of Washington, DC), tree cover loss over the past 2 years added up to an area of 0.61 $km^2$, about the size of 3,106 tennis courts.

We had set out to extract the signals in our satellite images and to transform them into actionable information based on evidence. For this reason, we exercise conservatism in our modeling approach, as described in the previous section. For the same reason, we developed our R package, loRax, and based it primarily on open-access data. It is the reason our package uses blockchain technology to preserve the integrity and provenance of the annotated dataset.

Briefly put, we hope both our algorithms and data will be used by researchers, environmental groups, and governmental agencies to protect forests across the globe. We hope the results generated by anyone using loRax can serve as presumptive evidence that an area is losing forest cover, particularly in those countries that are resource-poor and that need to prioritize areas for conservation or for other forms of intervention. We hope that others who use the R package will block-chain and share their annotated data (or improve upon our algorithms) so that collectively, we can increase the accuracy, and therefore the defensibility, of its simulations of tree cover loss, thereby boosting our ability to conserve this precious resource.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://pax.green/lorax/.

## AUTHOR CONTRIBUTIONS

PP conceived of the project, wrote the code, developed the methodology, obtained the data, and conducted the literature review. CP annotated several of the tiles used to train the Random Forest model. They also clarified the over-all concept and methodology with the senior author.

## FUNDING

Microsoft's AI for Earth program provided free access to their Azure cloud platform. The program also arranged for free access to Airbus satellite rasters. This project was supported by the community of Mindo. The pueblo's people sustained the work by offering hospitality and by providing us with the oral history of the area. This helped inform our understanding of Mindo's spatial structure.

## REFERENCES

Acuña-Ruz, T., Uribe, D., Taylor, R., Amézquita, L., Guzmán, M. C., Merrill, J., et al. (2018). Anthropogenic marine Debris over Beaches: Spectral Characterization for Remote Sensing Applications. *J. Remote Sensing Environ.* 217, 309–322. doi:10.1016/j.rse.2018.08.008

Aggarwal, S. (2004). *Principles of Remote Sensing*. Geneva, Switzerland: World Meteorological Oganisation.

Anderies, J. M., Folke, C., Walker, B., and Ostrom, E. (2013). Aligning Key Concepts for Global Change Policy: Robustness, Resilience, and Sustainability. *J. Ecol. Soc.* 18 (2), 8. doi:10.5751/es-05178-180208

Beck, R., Avital, M., Rossi, M., and Thatcher, J. B. (2017). *Blockchain Technology in Business and Information Systems Research*. Wiesbaden, Germany: Springer.

Bouhennache, R., Bouden, T., Taleb-Ahmed, A., and Cheddad, A. (2019). A New Spectral index for the Extraction of Built-Up Land Features from Landsat 8 Satellite Imagery. *Geocarto Int.* 34 (14), 1531–1551. doi:10.1080/10106049.2018.1497094

Boyd, D. (2005). Physical Basis of Remote Sensing. *J. J. Geoinformatics* 1, 67–81.

Brancalion, P. H. S., Broadbent, E. N., de-Miguel, S., Cardil, A., Rosa, M. R., Almeida, C. T., et al. (2020). Emerging Threats Linking Tropical Deforestation and the COVID-19 Pandemic. *J. Perspect. Ecol. conservation Biol.* 18 (4), 243–246. doi:10.1016/j.pecon.2020.09.006

Breiman, L. (2001). Random Forests. *J. Machine Learning* 45 (1), 5–32. doi:10.1023/a:1010933404324

Bruijnzeel, L., Kappelle, M., Mulligan, M., and Scatena, F. (2010). *Tropical Montane Cloud Forests: State of Knowledge and Sustainability Perspectives in a Changing World*. Cambridge, United Kingdom: Cambridge University Press.

Bubb, P., May, I. A., Miles, L., and Sayer, J. (2004). *Cloud forest Agenda*. Cambridge, United Kingdom: UNEP-WCMC biodiversity series.

Castelvecchi, D. (2016). Can We Open the Black Box of AI? *Nature* 538 (7623), 20–23. doi:10.1038/538020a

Clark, L. G., and Mason, J. J. (2019). Redescription of Chusquea Perligulata (Poaceae: Bambusoideae: Bambuseae: Chusqueinae) and Description of a Similar but New Species of Chusquea from Ecuador. *Phytotaxa* 400 (4), 227–236. doi:10.11646/phytotaxa.400.4.2

Dauwalter, D. C., Fesenmyer, K. A., Bjork, R., Leasure, D. R., and Wenger, S. J. (2017). Satellite and Airborne Remote Sensing Applications for Freshwater Fisheries. *Fisheries* 42 (10), 526–537. doi:10.1080/03632415.2017.1357911

Dobrinić, D., Gašparović, M., and Medak, D. (2021). Sentinel-1 and 2 Time-Series for Vegetation Mapping Using Random Forest Classification: A Case Study of Northern Croatia. *Remote Sensing* 13 (12), 2321. doi:10.3390/rs13122321

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., and Li, X. (2016). Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. *Remote Sensing* 8 (4), 354. doi:10.3390/rs8040354

Fingas, M., and Brown, C. E. (2018). A Review of Oil Spill Remote Sensing. *Sensors* 18 (1), 91. doi:10.3390/s18010091

Fisher, E., Pascual, P., and Wagner, W. (2014). Rethinking Judicial Review of Expert Agencies. *Tex. L. Rev.* 93, 1681.

Foster, P. (2001). The Potential Negative Impacts of Global Climate Change on Tropical Montane Cloud Forests. *Earth-Science Rev.* 55 (1-2), 73–106. doi:10.1016/s0012-8252(01)00056-3

Gascoin, S., Barrou Dumont, Z., Deschamps-Berger, C., Marti, F., Salgues, G., López-Moreno, J. I., et al. (2020). Estimating Fractional Snow Cover in Open Terrain from sentinel-2 Using the Normalized Difference Snow index. *Remote Sensing* 12 (18), 2904. doi:10.3390/rs12182904

Ghorbanian, A., Zaghian, S., Asiyabi, R. M., Amani, M., Mohammadzadeh, A., and Jamali, S. (2021). Mangrove Ecosystem Mapping Using Sentinel-1 and Sentinel-2 Satellite Images and Random forest Algorithm in Google Earth Engine. *Remote Sensing* 13 (13), 2565. doi:10.3390/rs13132565

Gradstein, S. R., Homeier, J., and Gansert, D. (2008). *The Tropical Mountain forest; Patterns and Processes in a Biodiversity Hotspot*. Göttingen, Germany: Universitätsverlag Göttingen.

Guayasamin, J. M., Cisneros-Heredia, D. F., Vieira, J., Kohn, S., Gavilanes, G., Lynch, R. L., et al. (2019). A New Glassfrog (Centrolenidae) from the Chocó-Andean Río Manduriacu Reserve, Ecuador, Endangered by Mining. *PeerJ* 7, e6400. doi:10.7717/peerj.6400

Helmer, E. H., Gerson, E. A., Baggett, L. S., Bird, B. J., Ruzycki, T. S., and Voggesser, S. M. (2019). Neotropical Cloud Forests and Páramo to Contract and Dry from Declines in Cloud Immersion and Frost. *PloS one* 14 (4), e0213155. doi:10.1371/journal.pone.0213155

Huang, Y., Chen, Z.-x., Yu, T., Huang, X.-z., and Gu, X.-f. (2018). Agricultural Remote Sensing Big Data: Management and Applications. *J. Integr. Agric.* 17 (9), 1915–1931. doi:10.1016/s2095-3119(17)61859-8

Huston, M. (1979). A General Hypothesis of Species Diversity. *The Am. Naturalist* 113 (1), 81–101. doi:10.1086/283366

Jolliffe, I. T., and Cadima, J. (2016). Principal Component Analysis: a Review and Recent Developments. *Phil. Trans. R. Soc. A.* 374 (2065), 20150202. doi:10.1098/rsta.2015.0202

Kaplan, G., and Avdan, U. (2017). Mapping and Monitoring Wetlands Using Sentinel-2 Satellite Imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* IV-4/W4, 271–277. doi:10.5194/isprs-annals-iv-4-w4-271-2017

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Ma, X., Li, C., Tong, X., and Liu, S. (2019). A New Fusion Approach for Extracting Urban Built-Up Areas from Multisource Remotely Sensed Data. *Remote Sensing* 11 (21), 2516. doi:10.3390/rs11212516

Muñoz-López, J., Camargo-García, J. C., and Romero-Ladino, C. (2021). Valuation of Ecosystem Services of Guadua Bamboo (Guadua Angustifolia) forest in the Southwestern of Pereira, Colombia. *Caldasia* 43 (1), 186–196. doi:10.15446/caldasia.v43n1.63297

Myster, R. W. (2020). Disturbance and Response in the Andean Cloud forest: A Conceptual Review. *Bot. Rev.* 86 (2), 119–135. doi:10.1007/s12229-020-09219-x

Nguyen, C. T., Chidthaisong, A., Kieu Diem, P., and Huo, L.-Z. (2021). A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. *Land* 10 (3), 231. doi:10.3390/land10030231

Nitoslawski, S. A., Wong-Stevens, K., Steenberg, J. W. N., Witherspoon, K., Nesbitt, L., and Konijnendijk van den Bosch, C. C. (2021). The Digital Forest: Mapping a Decade of Knowledge on Technological Applications for forest Ecosystems. *Earth's Future* 9 (8), e2021EF002123. doi:10.1029/2021ef002123

Olson, D. M., and Dinerstein, E. (2002). The Global 200: Priority Ecoregions for Global Conservation. *Ann. Mo. Bot. garden* 89, 199–224. doi:10.2307/3298564

Ono, A., Kajiwara, K., and Honda, Y. (2010). Development of New Vegetation Indexes, Shadow index (SI) and Water Stress Trend (WST). *Intern. Arch. Photogrammetry, Remote Sensing Spat. Inf. Sci.* 38, 710–714.

Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., and Field, C. B. (1994). Reflectance Indices Associated with Physiological Changes in Nitrogen-And Water-Limited sunflower Leaves. *Remote sensing Environ.* 48 (2), 135–146.

Pothuganti, S. (2018). Review on Over-fitting and Under-fitting Problems in Machine Learning and Solutions. *Int. J. Adv. Res. Electr. Electron. Instrumentation Eng.* 7 (9), 3692–3695. doi:10.15662/IJAREEIE.2018.070901

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reyes-Puig, C., Wake, D. B., Kotharambath, R., Streicher, J. W., Koch, C., Cisneros-Heredia, D. F., et al. (2020). Two Extremely Rare New Species of Fossorial Salamanders of the Genus Oedipina (Plethodontidae) from Northwestern Ecuador. *PeerJ* 8, e9934. doi:10.7717/peerj.9934

Rokni, K., and Musa, T. A. (2019). Normalized Difference Vegetation Change index: A Technique for Detecting Vegetation Changes Using Landsat Imagery. *Catena* 178, 59–63. doi:10.1016/j.catena.2019.03.007

Romero, B. J. Z., Solano-Gomez, R., and Wilson, M. (2017). A New Species of Pleurothallis (Orchidaceae: Pleurothallidinae) from Southwestern ecuador: Pleurothallis Marioi. *Phytotaxa* 308 (1), 80–088. doi:10.11646/phytotaxa.308.1.6

Roy, B. A., Zorrilla, M., Endara, L., Thomas, D. C., Vandegrift, R., Rubenstein, J. M., et al. (2018). New Mining Concessions Could Severely Decrease Biodiversity and Ecosystem Services in Ecuador. *Trop. Conservation Sci.* 11, 1940082918780427. doi:10.1177/1940082918780427

Schonlau, M., and Zou, R. Y. (2020). The Random forest Algorithm for Statistical Learning. *Stata J.* 20 (1), 3–29. doi:10.1177/1536867x20909688

Segarra, J., Buchaillot, M. L., Araus, J. L., and Kefauver, S. C. (2020). Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy* 10 (5), 641. doi:10.3390/agronomy10050641

Senecal, J. J., Sheppard, J. W., and Shaw, J. A. (2019). "Efficient Convolutional Neural Networks for Multi-Spectral Image Classification," in 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, June 14, 2019 (IEEE).

Sepuru, T. K., and Dube, T. (2018). An Appraisal on the Progress of Remote Sensing Applications in Soil Erosion Mapping and Monitoring. *Remote Sensing Appl. Soc. Environ.* 9, 1–9. doi:10.1016/j.rsase.2017.10.005

Sornoza-Molina, F., Freile, J. F., Nilsson, J., Krabbe, N., and Bonaccorso, E. (2018). A Striking, Critically Endangered, New Species of Hillstar (Trochilidae: Oreotrochilus) from the Southwestern Andes of Ecuador. *The Auk* 135 (4), 1146–1171. doi:10.1642/auk-18-58.1

Spracklen, D. V., and Righelato, R. (2016). Carbon Storage and Sequestration of Re-Growing Montane Forests in Southern Ecuador. *For. Ecol. Manage.* 364, 139–144. doi:10.1016/j.foreco.2016.01.001

Stevens, H. C., Re, B., and Becker, C. D. (2021). *Avian Species Inventory and Conservation Potential of Reserva Las Tangaras, Ecuador*. Neotropical Bird Clud.

Wagner, W., Fisher, E., and Pascual, P. (2010). Misunderstanding Models in Environmental and Public Health Regulation. *NYU Envtl. LJ* 18, 293.

Waśniewski, A., Hościło, A., Zagajewski, B., and Moukétou-Tarazewicz, D. (2020). Assessment of Sentinel-2 Satellite Images and Random Forest Classifier for Rainforest Mapping in Gabon. *Forests* 11 (9), 941. doi:10.3390/f11090941

Wood, C., Sullivan, B., Iliff, M., Fink, D., and Kelling, S. (2011). eBird: Engaging Birders in Science and Conservation. *Plos Biol.* 9 (12), e1001220. doi:10.1371/journal.pbio.1001220

Xu, H. (2006). Modification of Normalised Difference Water index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. *Int. J. remote sensing* 27 (14), 3025–3033. doi:10.1080/01431160600589179

Zha, Y., Gao, J., and Ni, S. (2003). Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *Int. J. remote sensing* 24 (3), 583–594. doi:10.1080/01431160304987

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 Pascual and Pascual. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Energy Research

ORIGINAL RESEARCH
published: 24 December 2021
doi: 10.3389/fenrg.2021.795413

# Electricity-Water Consumption and Metropolitan Economic Growth: An Empirical Dual Sectors Dynamic Equilibrium Model

*Yiming He*[1]* *and Shaoui Gao*[2]*

[1]School of National Agricultural Institution and Development, South China Agricultural University, Guangzhou, China, [2]College of Economics and Management, South China Agricultural University, Guangzhou, China

This paper develops a dual sectors dynamic equilibrium model and introduces electricity consumption and water consumption in a growth model that tested by using a time series data set from 1950 to 2014 in Guangzhou, China. It presents a theoretical prediction on the interactions between electricity consumption, water consumption, and the metropolitan economic growth. Consistent with this prediction, electricity consumption and water consumption by themselves appear to have significant effects on metropolitan economic performance. The cointegration techniques show that electricity consumption, water consumption, and the metropolitan economic performance have long-run equilibrium relationship. The results of kernel-based regularized least squares reveal that metropolitan economic growth is positively correlated with electricity consumption. Also consistent with the theory, water consumption is positively associated with metropolitan economic performance. These results are generally stable and hold with alternative measures of unit roots, with alternative estimation strategies, and with or without controlling for trends, intercepts, and break points.

**Keywords: electricity consumption, water consumption, metropolitan development, cointegration, kernel-based regularized least squares, dynamic general equilibrium, Guangzhou**

OPEN ACCESS

***Edited by:***
*Xunpeng (Roc) Shi, University of Technology Sydney, Australia*

***Reviewed by:***
*Festus Victor Bekun, Gelişim Üniversitesi, Turkey*
*Abdul Rehman, Henan Agricultural University, China*

***Correspondence:**
*Yiming He*
*hym0509@scau.edu.cn*
*Shaoui Gao*
*gsh0728@126.com*

***Specialty section:***
*This article was submitted to Sustainable Energy Systems and Policies, a section of the journal Frontiers in Energy Research*

***Received:*** *15 October 2021*
***Accepted:*** *08 November 2021*
***Published:*** *24 December 2021*

***Citation:***
*He Y and Gao S (2021) Electricity-Water Consumption and Metropolitan Economic Growth: An Empirical Dual Sectors Dynamic Equilibrium Model. Front. Energy Res. 9:795413. doi: 10.3389/fenrg.2021.795413*

## 1 INTRODUCTION

Casual empiricism suggests the presence of significant differences in metropolitan economic growth in China. For example, China's electricity consumption, heavy reliance on fossil fuel, and watering down environmental regulations have significantly affected the economic development of the country. Nevertheless, there has not been a systematic analysis of long-run differences in metropolitan economic growth with both electricity consumption and water consumption. Our primary motivation in this paper is to make a first attempt at such a systematic analysis and to investigate the relationship between electricity consumption, water consumption, and metropolitan economic growth in Guangzhou, China.

Electricity is one of the most studied energy classes—and for good reason. Growing demands for electricity and water resources and increasingly serious environmental challenges, particularly in metropolitan areas of developing countries like China, have elevated the urgency of studying the water-energy-growth nexus (Lai et al., 2011). China is the largest emitter of greenhouse gases (GHGs) in the world. In 2012, China was the largest contributor to carbon emissions, and with 8.50 Gt in $CO_2$ emissions from fossil burning and cement production in 2012, China was responsible for 25% of global carbon emissions, which was equivalent to the emissions of the U.S. and the E.U. combined

(Liu, 2015). Between 2000 and 2010, the electricity production in China had increased threefold and accounted for 50% of domestic and 12% of global $CO_2$ emissions in 2010. For these reasons alone, researchers, policy-makers, consumers, and power plants all have a significant interest in accurately valuing the nexus between electricity consumption and economic development.

In addition to emitting large amounts of GHGs, electricity production also requires significant amounts of water, so water use is an important factor for electricity sector and for sustainable development in China. Continuous economic growth in China can have significant implications for water resources and environmental quality. Therefore, it is important to understand the interdependencies between electricity consumption, water use, and economic growth particularly when it comes to formulating policies targeted at environmental quality protection and sustainability.

Since the founding of China in 1949, electricity and water consumption has become an important issue for the economic development during the process of industrialization and urbanization (He et al., 2017). After 1978 China enforced the openness and reform policy, electricity consumption, water consumption, and economic performance increase dramatically (Kahrl and Roland-Holst, 2008; Jiang et al., 2014). Continuous economic growth of China will have significant implications for electricity and water. Therefore, understanding of the implications of the electricity and water use policies is an important factor for policy formulation targeted at driving economic growth. Moreover, the electricity production is related with the water use. Hence, water withdrawal is also an important factor for economy sustainable growth in China (Feng et al., 2014).

Electric grid and water infrastructure have been critical factors for the economic growth, industrialization, and urbanization of China since founding in 1949. In particular, energy consumption and water use increased dramatically after 1978 when China enacted the openness and reform policies that stimulated economic growth (Kahrl and Roland-Holst, 2008; Jiang et al., 2014). Over time, maintaining balance between electricity supply and demand in China required increase in power generation, which has been primarily based on thermal power generation from coal until recent years (Ohdoko et al., 2013). In 2001, thermal power generation accounted for 81.2% of total power generation, and 95% of thermal power was produced in coal power plants (Zhu et al., 2005). However, concerns have been mounting pertaining to emissions from coal power plant operations including carbon dioxide, nitrogen oxides, sulfur oxides, and particulate matter, including PM2.5 and PM10.

Water use in China rose by 12.47% from 2000 to 2013. On average, 64.06% of water use is attributed to agricultural industry, 22.66% to industry, 11.91% to household use, and 1.36% to biology use from 2000 to 2013 (Zhang et al., 2016). In Guangzhou, 18.43% of water use came from industry, 24.25% of water use came from public services, 55.19% of water use came from household, and others are occupied at 2.13% in 2014.

Guangzhou is the third largest metropolitan area in China, after Beijing and Shanghai, and the largest city in south central China. Moreover, the Guangzhou Statistical Division provides the most complete and longest duration time series dataset (from 1950 to 2014) among the metropolitans in China. It helps us to observe and estimate the electricity consumption function with institutional reform. Throughout its 2100-year history, Guangzhou, the capital of Guangdong province, has been a major commercial center in south China. From 1949 to 1977, collectivist policies abolished much of the trade and commercial activities long associated with Guangzhou. The major breakthrough for repositioning of Guangzhou as a regional commercial hub came with implementation of the open-door policy in 1978, which allowed for international trade and investment introduced. In 1950, the GDP of Guangzhou reached $43 million and per capita GDP was $16. However, in 1978, the GDP reached $0.72 billion and per capita was $149. After economic reform, in 2014, the GDP reached $278,448 million and per capita GDP was $21419 (GSD, 2015). The 1950 census found the population of Guangzhou to be 2.5 million. As of 2014, it was estimated at 13 million. The industry sector accounted for 33.46% of its GDP in 2014, agriculture accounted for 1.32%, and the service sector represented 65.22%. Major industries in Guangzhou include automobiles, electronics, and petrochemicals.

Electricity in Guangzhou comes from the power stations in Guangdong province. There are 33 coal-, gas-, and fuel oil–based power stations, five nuclear power stations, three hydro power stations, three wind power stations, and four pumped-storage power stations in Guangdong province. Except for Zhujiang, Huangpu, Guangzhou (coal-, gas-, and fuel oil–based), and Guangzhou Pumped Storage Power Station, all other power stations are outside Guangzhou. The grid that supplies electricity to the city is Guangzhou Power Supply Co., Ltd. (GZPS). GZPS is one of the enterprises owned by China Southern Power Grid Co., Ltd. (CSG), which is state owned. In 1996, the first "People's Republic of China Electricity Law" was enforced and followed by a series of supporting laws, marking the electricity industry of Guangzhou into legal constraint. This "Electricity Law" makes it clear on electricity infrastructure construction, electricity production and management, electricity supply and use, supervision, and legal responsibility. On the other hand, in 2005, the Chinese Council promulgated the "Electricity Control System", which marked the electricity industry of Guangzhou has established a clear monitoring framework. Meanwhile, the "Electricity Business License", "Electricity Markets Operation Basic Management", and "Electricity Markets Supervision Measures" are the three regulations that came into effect. As a result, the electricity consumption in Guangzhou grows following the economic growth.

The water in Guangzhou comes from Nanzhou Water Supply Plant (NWSP) in Guangzhou that puts into production at full-scale with the capacity of one million cubic meters per day and from Guangzhou Water Supply Co. (GWSC) that is a large state-owned water supply enterprise and provides all water supply services, including water treatment and and diversified business development. The electricity of Guangzhou comes from the power stations in Guangdong province. There are 33 coal-,

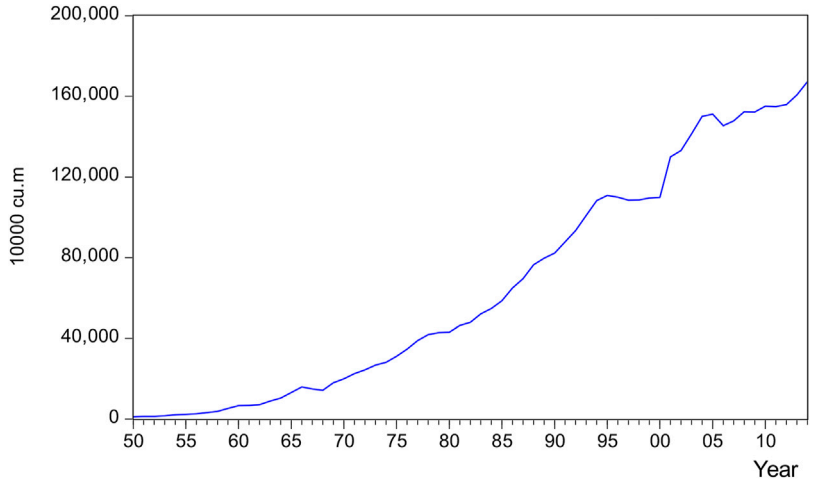



**FIGURE 1 |** Water consumption in Guangzhou, 1950–2014.

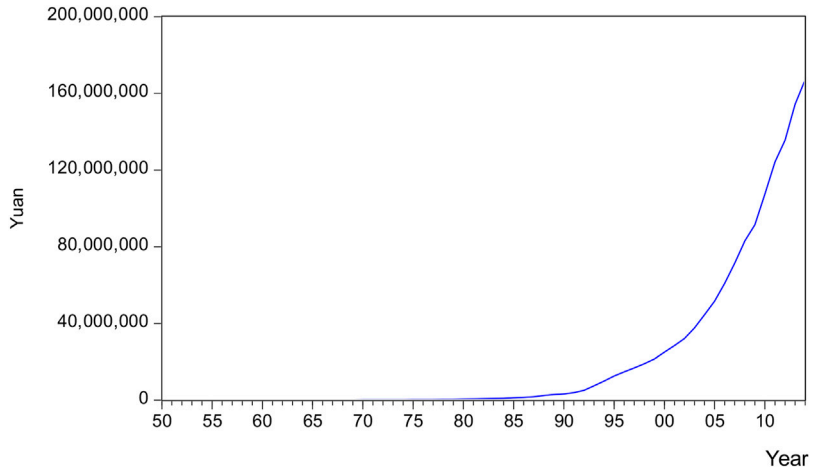



**FIGURE 2 |** GDP in Guangzhou, 1950–2014.

gas, and fuel oil–based power stations, five nuclear power stations, three hydro power stations, three wind power stations, and four pumped-storage power stations in Guangdong province. Except for Zhujiang Power Station, Guangzhou Huarun Thermal Power Station, and Guangzhou Pumped Storage Power Station, other power stations are outside Guangzhou that is the capital of Guangdong province. The water in Guangzhou comes from Nanzhou Water Supply Plant (NWSP) in Guangzhou that puts into production at full-scale with the capacity of one million cubic meters per day and from Guangzhou Water Supply Co. (GWSC) that is a large state-owned water supply enterprise and provides all water supply services, including water treatment and diversified business development. **Figure 1** shows the increase in water consumption for Guangzhou, China, from 1950 to 2014. Before 2004, the water in Guangzhou just came from state-owned NWSP and GWSC. NWSP was built in 2004 and put into production at full-scale with the capacity of one million cubic meters per day. It is the largest advanced water treatment plant in the nation. GWSC is a large state-owned water supply enterprise and provides all water supply services, including water treatment.

From **Figure 2**, it is found that the metropolitan economy grow after 1980s. Actually, China is a marketization transition country, and the Chinese government enforced the Open and Reform Policy at 1978. Since then, the economic institution in China gradually has been transferred from planning system into market system (Xu, 2011). In particular, being the capital of Guangdong province and the third largest Chinese city, Guangzhou was the earliest metropolitan in China to enforce that marketization policy. Furthermore, 1994 was the other important year for Chinese metropolitan economic development. Because the Chinese tax institution was

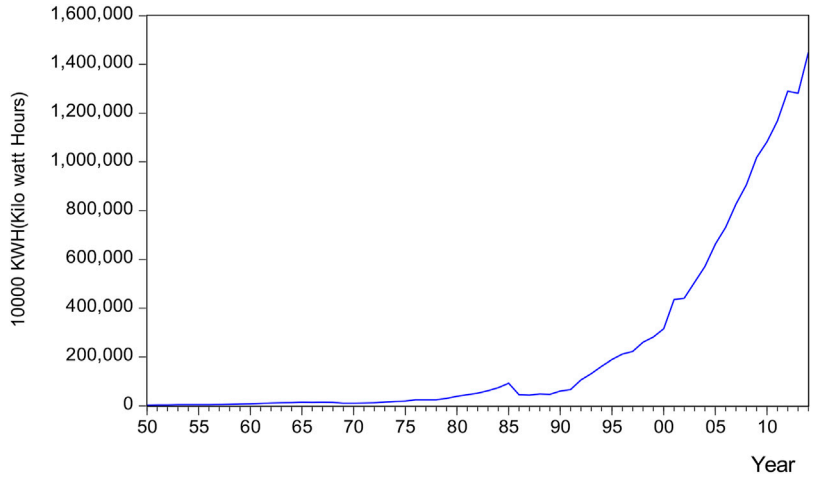



**FIGURE 3 |** Electricity consumption in Guangzhou, 1950–2014.

transferred from the tax contracting system into the tax sharing system at 1994 (Li and Kung, 2015), the metropolitan economy grows dramatically after 1994.

GDP in Guangzhou is divided into three industrial sectors product values: primary industry (agriculture), secondary industry (mainly manufacture and construction), and tertiary industry (service, business, tourism, etc.). In 2014, the ratios of agriculture, manufacture, and service respect to GDP are 1.31%, 33.47%, and 65.22%, respectively. The total electricity consumption of agriculture, manufacture, and service are 575,520,000 (kWh), 39,527,160,000 (kWh), and 20,421,400,000 (kWh) for Guangzhou in 2014, respectively. The water for production use, public services, household use, and others are 296,263,400 (cu.m), 389,842,300 (cu.m), 8,87,217,300 (cu.m), and 34,117,600 (cu.m) for Guangzhou in 2014, respectively. The electricity of Guangzhou comes from the power stations in Guangdong province. There are 33 coal-, gas, and fuel oil–based power stations, five nuclear power stations, three hydro power stations, three wind power stations, and four pumped-storage power stations in Guangdong province. Except for Zhujiang Power Station, Guangzhou Huarun Thermal Power Station, and Guangzhou Pumped Storage Power Station, other power stations are outside Guangzhou that is the capital of Guangdong province.

Therefore, the time series data with 65 observations from Guangzhou to investigate the interactions between electricity consumption, water consumption, and metropolitan economic growth were carried out. This dataset from Guangzhou can provide us a long-run data sample, and the data are all historic and realistic, which is the only metropolitan level statistical dataset including the data from 1950 to 2014 in China.

The rest of the paper is organized as follows. **Section 2** reviews related literature. **Section 3** introduces the theoretical framework, characterizes the equilibrium, and derives the major hypotheses and results. **Section 4** provides the data description and econometric strategy. **Section 5** shows empirical results. **Section 6** concludes.

## 2 LITERATURE REVIEW

Actually, limited attention has been devoted in academic literature to the nexus between water consumption and economic growth, and studies have provided evidence that national per capita water consumptive use seem to follow an inverted U-shaped path, with respect to per capita income, consistent with Kuznets type of relation. The research finds some support for the existence of an Environmental Kuznets Curve, but results are highly dependent on choice of datasets and statistical technique. For example, Katz (2015) presented the results of the analysis of the relationship between national per capita water use and per capita income using international cross-sectional data and panel data for OECD nations and U.S. states. He found that the water consumption initially rise and then decline with respect to income.

Barbier (2004) provided strong support for the inverted U relationship between economic growth and the rate of water use across countries. Barbier and Chaudhry (2014) showed that, in urban counties in the U.S., higher water use and population growth are associated with increase in income per capita. Cazcarro et al. (2013) found that a significant growth of per capita income has been the main factor driving the increase in water consumption growth. Ngoran et al. (2016) suggested that economic growth in 38 Sub-Saharan African countries for the period 1980–2011 is driven mainly by water and labor. Hence, their study supports the growth hypothesis for water consumption.

However, Gleick (2003) even found no relationship between per capita national water consumption and income, which is consistent with neutrality hypothesis. Obviously, previous empirical studies above have come under scrutiny in distinct literatures, and the literatures remain disjointed. Although some empirical researches have calculated empirically the links between water consumption GDP, they do not examine the causality among GDP and water consumption, comparing empirical results among different metropolitans. According the

literature above, the nexus between electricity consumption and economic output has been extensively studied, but the evidence so far is contradictory and inconclusive (Stern et al., 2018).

On the other hand, Magazzino (2014) applied time series methodologies to examine the causal relationship among electricity demand, real per capita GDP, and total labor force for Italy from 1970 to 2009. The results of estimation indicate that one cointegrating relationship exists among these variables. This equilibrium relation implies that, in the long-run, GDP and labor force are correlated negatively, as well as GDP and electricity. Moreover, there is a bidirectional Granger causality flow between real per capita GDP and electricity demand; whereas labor force does not Granger cause neither real per capita GDP nor electricity demand. This implies that electricity demand and economic growth are jointly determined at the same time for the Italian case. The forecast error variance decomposition shows that forecast errors in real per capita GDP are mainly caused by the uncertainty in GDP itself, whereas forecast errors in labor force are mainly resulted from the labor force itself, although aggregate income and electricity are important, too.

Magazzino (2017) investigated the stationary properties of electric power consumption in 18 countries in the Middle East and North Africa by using yearly data over the period 1971–2013. After having controlled for the presence of cross-sectional dependence, the "second-generation" panel unit root tests reveal mixed results.

However, concerning the academic literature, there are limited studies about the nexus among economic growth, water consumption, and electricity consumption. For example, Wang et al. (2016) only discussed the electricity consumption in Beijing. Because it is the third largest urban economy in China, that is a surprising omission from the resource economics and China metropolitan economy issue literature (Mele and Magazzino, 2020). Udemba et al. (2020) conducted an empirical research to analyze the relationship between pollutant emission, energy consumption, foreign direct investment, and economic growth.

Indeed, there is a large literature on related problems, including Rehman and Deyuan, (2018), Rehman, (2020), Rehman et al. (2020), Rehman et al. (2021a), Rehman et al. (2021b), Rehman et al. (2021c), Rehman et al. (2021d), Muhammad Awais Baloch et al. (2021), Andrew Adewale Alola et al. (2021), Bright Akwasi Gyamfi et al., 2021, Festus Victor Bekun et al. (2021a), Festus Victor Bekun et al. (2021b), and Festus Victor Bekun et al. (2021a). For example, He (2020) identified the causal effects of urbanization and metropolitan economic performance on electricity consumption by a dynamic general equilibrium theoretical framework and time series econometric models using data from 1949 to 2016 in Guangzhou, China. Both autoregressive distributed lag (ARDL) and Johansen cointegration techniques show that electricity consumption, urbanization, and metropolitan economic performance are cointegrated. The results of two-stage least squires demonstrate that the marginal propensity to electricity consumption in Guangzhou is about 1.28.

Meanwhile, He and Fullerton (2020) examined the nexus between water consumption and economic growth. Water consumption function is derived using an optimal dynamic equilibrium model. Two instrument variable models are proposed with real per capita economic output specified as a function of institutional reform and urbanization, which are used to examine the nexus among water consumption, reform, urbanization, and economic growth in Guangzhou, China.

On the basis of the literature above, this study builds a dual sectors dynamic equilibrium model where the economic performance is expressed as a linear combination of electricity consumption and water consumption. The objectives and contributions of this paper are both theoretical and empirical. The theory described in this paper 1) applied dynamic general equilibrium models to metropolitan economic growth using a social optimal method and 2) describes an estimation procedure that connects electricity consumption and water consumption and also yields interpretable implications. Empirically, this study finds the following: 1) electricity consumption, water consumption, and the metropolitan economic performance have long-run equilibrium relationship; 2) the kernel-based regularized least squares (KRLS) approach reveals that metropolitan economic growth is positively correlated with electricity consumption and water consumption is positively associated with metropolitan economic performance; and 3) results of unit roots are generally robust with alternative estimation strategies and with or without controlling for trends, intercepts, and break points.

# 3 THEORETIC MODEL

## 3.1 Electricity Sector

In the model of this study, metropolitan economy is composed of only electricity and water sectors. There are many companies (utilities) in both sectors. Consider that the problem of a representative company i in electricity sector cares about the net benefit and production function per capita is supposed as the Cobb-Douglas form: $y_i^e(k_i^e, e_i^e) = \mathrm{A}_i^e (k_i^e)^a (e_i)^b \, (0 < a < 1,\ 0 < b < 1)$. The maximization problem can be written as follows:

$$\max_{k_i^e, e_i^e} NB_i^e = A_i^e (k_i^e)^a (e_i)^b - \left(pe_i + \theta k_i^e\right), \tag{1}$$

where $\mathrm{A}_i^e$ denotes the technology level in the $i$th electricity company, $k_i^e$ is capital per capita in the $i$th electricity company, $e_i$ is electricity per capita in the $i$th electricity company, p is price of electricity, and $\theta$ is interest rate. The first-order conditions of **Equation 1** are as follows:

$$\frac{\partial NB_i^e}{\partial k_i^e} = a\mathrm{A}_i^e \left(k_i^e\right)^{a-1} (e_i)^b - \theta = 0, \tag{2}$$

$$\frac{\partial NB_i^e}{\partial e_i^e} = b\mathrm{A}_i^e \left(k_i^e\right)^a (e_i)^{b-1} - \mathrm{p} = 0. \tag{3}$$

Combining **Equations 2** and **3**, the following was obtained:

$$k_i^e = \frac{a}{b}\frac{p}{\theta} e_i. \tag{4}$$

Suppose there are n homogeneous companies in electricity sector, so the electricity consumption per capita in the whole sector $E = ne_i$ and capital per capita in the whole sector

$K^e = nk_i^e = n\frac{a}{b}\frac{p}{\theta}e_i = \frac{a}{b}\frac{p}{\theta}ne_i = \frac{a}{b}\frac{p}{\theta}E$. Hence, given technology level in the whole sector is identical $\mathrm{A}_i^e = A^e$, the output per capita of electricity sector is as follows:

$$\begin{aligned} Y^e = ny_i^e &= n\mathrm{A}_i^e\left(k_i^e\right)^a\left(e_i\right)^b = nA^e\left(\frac{K^e}{n}\right)^a\left(\frac{E}{n}\right)^b, \\ &= n^{1-a-b}A^e\left(K^e\right)^a(\mathrm{E})^b = n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\mathrm{E}\right)^a(\mathrm{E})^b, \\ &= n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\right)^a\left(E^e\right)^{a+b}. \end{aligned} \tag{5}$$

## 3.2 Water Sector

Furthermore, the economy activities in water sector are considered. Suppose the production function per capita in the *j*th water company also takes C-D form with the price of water (r):

$$y_j^w\left(k_j^w, w_j\right) = \mathrm{A}_j^w\left(k_j^w\right)^a\left(w_j\right)^b. \tag{6}$$

The maximization problem can be written as follows:

$$\max_{k_j^w, e_j^w} NB_j^w = \mathrm{A}_j^w\left(k_j^w\right)^a\left(w_j\right)^b - \left(rw_j + \theta k_j^w\right). \tag{7}$$

The first-order conditions of **Equation 7** are as follows:

$$\frac{\partial NB_i^w}{\partial k_j^w} = a\mathrm{A}_j^w\left(k_j^w\right)^{a-1}\left(w_j\right)^b - \theta = 0, \tag{8}$$

$$\frac{\partial NB_i^w}{\partial k_j^w} = aA_j^w\left(k_j^w\right)^{a-1}\left(w_j\right)^b - \theta = 0. \tag{9}$$

Combining **Equations 8** and **9**, the following was obtained:

$$k_i^w = \frac{a}{b}\frac{r}{\theta}w_j. \tag{10}$$

Suppose there are m homogeneous water companies in this sector, water consumption per capita in the whole sector $W = mw_j$ and capital per capita in the whole sector.

$$K^w = mk_j^w = m\frac{a}{b}\frac{r}{\theta}w_j = \frac{a}{b}\frac{r}{\theta}mw_j = \frac{a}{b}\frac{r}{\theta}\mathrm{W}. \tag{11}$$

Given technology level in water sector is identical $\mathrm{A}_j^w = A^w$, the output per capita in water sector is as follows:

$$\begin{aligned} Y^w = my_j^w &= mA_j^w\left(k_j^w\right)^a\left(w_j\right)^b = mA^w\left(\frac{K^w}{m}\right)^a\left(\frac{W}{m}\right)^b, \\ &= m^{1-a-b}A^w\left(K^w\right)^a(W)^b = m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}W\right)^a(W)^b, \\ &= m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}\right)^a(\mathrm{W})^{a+b}. \end{aligned} \tag{12}$$

## 3.3 Dynamic General Equilibrium

Now, the optimal electricity and water consumption function to find the dynamic paths for electricity sector and water sector over time are derived. Hence, the problem that this study aims to solve can be expressed as follows:

$$W_m = \int_0^T U_m\left(c_{m,t}\right)e^{-\theta t}\mathrm{d}t, \tag{13}$$

where $c_m$ denotes private *numeraire* consumption per capita in metropolitan so as to maximize the social welfare ($W_m$) in metropolitan. It is assumed that the metropolitan utility function is an iso-elastic utility function with the coefficient of relative risk aversion $\sigma$:

$$U_m\left(c_m\right) = \frac{c_m^{1-\sigma}}{1-\sigma}. \tag{14}$$

Hence, the social welfare maximization problem with utility discount rate $\theta$ (assume being identical to interest rate) in the metropolitan is expressed as follows:

$$\max_{c_m} \int_0^T \frac{c_m^{1-\sigma}}{1-\sigma}e^{-\theta t}dt. \tag{15}$$

There are two constraints that must be satisfied by optimal solution.

First, the average stock of electricity s is to be used. Electricity consumption per capita and water consumption per capita are $E^u$ and $E^r$. Since electricity and water in Metropolitan of China with limited observation period both are considered as renewable resources. Assume the net generating growth of them is the function of their capacity $\mathrm{s_c}$, $\mathrm{G} = \mathrm{G}\left(\mathrm{s_c}\right)$. Hence, the average renewable resources stock constraint is expressed as follows:

$$\dot{s} = \mathrm{G}\left(\mathrm{s_c}\right) - E - \mathrm{W}. \tag{16}$$

A second constraint on the change of capital stock per capita derives from the accounting identity relating private *numeraire* consumption per capita, output per capita in electricity sector and water sector:

$$\dot{\mathrm{K}} = Y^e + Y^w - c_m = Y - c_m. \tag{17}$$

According to **Equations 5**, **12**, and **17**, the following was obtained:

$$\dot{\mathrm{K}} = n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\right)^a(\mathrm{E})^{a+b} + m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}\right)^a(\mathrm{W})^{a+b} - \mathrm{c}_m. \tag{18}$$

Then, the optimal social welfare problem is as follows:

$$\begin{aligned} &\max_{c_m} \int_0^T \frac{c_m^{1-\sigma}}{1-\sigma}\mathrm{e}^{-\theta t}\mathrm{d}t \\ \text{s.t.}\dot{\mathrm{K}} &= n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\right)^a(\mathrm{E})^{a+b} + m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}\right)^a(\mathrm{W})^{a+b} - \mathrm{c}_m \\ &\dot{s} = \mathrm{G}\left(\mathrm{s_c}\right) - E - \mathrm{W}. \\ &E + W \geq 0 \text{ and } \int_0^T (E + \mathrm{W})\,dt \leq \mathrm{s} \end{aligned}$$

Set up the current Hamiltonian function:

$$\begin{aligned} \mathcal{H}_C = \frac{c_m^{1-\sigma}}{1-\sigma} + \varphi\Big[&n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\right)^a(\mathrm{E})^{a+b} \\ &+ m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}\right)^a(\mathrm{W})^{a+b} - c_m\Big] + \tau\left[\mathrm{G}\left(\mathrm{s_c}\right) - E - \mathrm{W}\right]. \end{aligned} \tag{19}$$

**TABLE 1 |** Variable definitions.

| Variable | Notation | Definition | Unit |
|---|---|---|---|
| Metropolitan economic performance | Y | Annual GDP in Guangzhou | 10,000 Yuan RMB |
| Electricity consumption | E | Total annual consumption of electricity in Guangzhou | 10,000 kWh (kilowatt hours) |
| Water consumption | W | Total annual consumption of water in Guangzhou | 10,000 cu.m |

The maximum principle conditions are as follows:

$$\frac{\partial \mathcal{H}_C}{\partial c_m} = c_m^{-\sigma} - \varphi = 0, \tag{20}$$

$$\frac{\partial \mathcal{H}_C}{\partial \mathrm{E}} = \varphi\left[ (a+b)n^{1-a-b}A^e\left(\frac{a}{b}\frac{p}{\theta}\right)^a (\mathrm{E})^{a+b-1}\right] - \tau = 0, \tag{21}$$

$$\frac{\partial \mathcal{H}_C}{\partial W} = \varphi\left[ (a+b)m^{1-a-b}A^w\left(\frac{a}{b}\frac{r}{\theta}\right)^a (\mathrm{W})^{a+b-1}\right] - \tau = 0, \tag{22}$$

$$\frac{\partial \mathcal{H}_C}{\partial s} = -(\dot{\tau} - \theta\tau) = 0, \tag{23}$$

$$\frac{\partial \mathcal{H}_C}{\partial K} = -\left(\dot{\phi} - \theta\varphi\right) = 0. \tag{24}$$

The solution of the optimal control model above is as follows:

$$\varphi^*(t) = \varphi(0)e^{\theta t}, \tag{25}$$

$$\tau^*(t) = \tau(0)e^{\theta t} \tag{26}$$

$$E = \frac{a+b}{\frac{\tau^*(t)}{\varphi^*(t)}}Y^e = \frac{a+b}{\frac{\tau(0)}{\varphi(0)}}Y^e \tag{27}$$

$$W = \frac{a+b}{\frac{\tau^*(t)}{\varphi^*(t)}}Y^w = \frac{a+b}{\frac{\tau(0)}{\varphi(0)}}Y^w. \tag{28}$$

Actually, let $\beta_e = \frac{a+b}{\frac{\tau(0)}{\varphi(0)}}$, and then, the electricity consumption function is as follows:

$$E_t = \beta_e Y_t^e. \tag{29}$$

Similarly, the water consumption function is as follows:

$$W_t = \beta_w Y_t^w. \tag{30}$$

Finally, presume that $Y_t = Y_t^e + Y_t^w + \nu_t$, where $\nu_t$ represents other factors. Combining **Equations 29** and **30**, $Y_t = \frac{\beta_w}{\beta_e\beta_w}E_t + \frac{\beta_w}{\beta_e\beta_w}W_t + \nu_t$ is obtained, which is the expression about the nexus among water consumption, electricity consumption, and economic output for long run.

# 4 DATA AND ECONOMETRIC METHODOLOGY

## 4.1 Data and Statistical Description

Annual data from 1950 to 2014 was obtained from the "Guangzhou Statistical Yearbook", within the Guangzhou Municipal Statistics Bureau. The empirical research by Stata 15 was conducted.

**Table 1** lists all variables and their definitions used in the empirical analysis. **Table 2** lists the summary statistics for the sample. The sample data exhibit good variability.

**TABLE 2 |** Descriptive statistics.

| Variable | E | Y | W |
|---|---|---|---|
| Mean | 1,346,930 | 20,713,531 | 65,740.22 |
| Median | 322,323 | 721,533.0 | 48,022.00 |
| Max | 7,658,542 | 1.67E+08 | 167,314.6 |
| Min | 4,872 | 25,620.00 | 1,140 |
| Std. Deviation | 2,094,516 | 40,189,753 | 55,943.97 |
| Skewness | 1.77870 | 2.249,855 | 0.418,239 |
| Kurtosis | 4.90564 | 7.171,236 | 1.698,786 |
| Observation | 65 | 65 | 65 |

## 4.2 Econometric Methodologies

### 4.2.1 Stationarity Tests

Standard Granger causality tests have to be conducted on stationary time series. Following this line, the unit roots of Xt to confirm the stationary properties of each variable were first tested. This is achieved by using the augmented Dickey–Fuller test. The augmented Dickey–Fuller test (Granger, 1969) can be revised and derived from the Perron (1989) test:

$$\Delta Z_t = \hat{\mu}^1 + \hat{\theta}^1 DU_t + \hat{\beta}^1 t + \hat{d}^1 D(T_B)_t + \hat{\alpha}^1 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^1 \Delta Z_{t-i} + \hat{\varepsilon}_t, \tag{31}$$

$$\Delta Z_t = \hat{\mu}^2 + \hat{\gamma}^2 DT_t^* + \hat{\beta}^2 t + \hat{\alpha}^2 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^2 \Delta Z_{t-i} + \hat{\varepsilon}_t, \tag{32}$$

$$\Delta Z_t = \hat{\mu}^3 + \hat{\theta}^3 DU_t + \hat{\beta}^3 t + \hat{\gamma}^3 DT_t^* + \hat{d}^3 D(T_B)_t + \hat{\alpha}^3 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^3 \Delta Z_{t-i} + \hat{\varepsilon}_t, \tag{33}$$

where $Z_t = (Y_t, E_t, W_t)^T$.

Thus, the choice of the break point is correlated with the data in hand and the choice of break point cannot be considered as independent of the data. Zivot and Andrews (1992) addressed this issue by estimating the structural break data endogenously instead of considering an exogenous break date. The following equations for the Zivot and Andrews (ZA) test are estimated, when $Z_t = (Y_t, E_t, W_t)^T$:

$$\Delta Z_t = \hat{\mu}^4 + \hat{\theta}^4 DU_t(\hat{\lambda}) + \hat{\beta}^4 t + \hat{\alpha}^4 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^4 \Delta Z_{t-i} + \hat{\varepsilon}_t, \tag{34}$$

$$\Delta Z_t = \hat{\mu}^5 + \hat{\gamma}^5 DT_t^*(\hat{\lambda}) + \hat{\beta}^5 t + \hat{\alpha}^5 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^5 \Delta Z_{t-i} + \hat{\varepsilon}_t, \tag{35}$$

$$\Delta Z_t = \hat{\mu}^6 + \hat{\theta}^6 DU_t(\hat{\lambda}) + \hat{\beta}^6 t + \hat{\gamma}^6 DT_t^*(\hat{\lambda}) + \hat{\alpha}^6 Z_{t-1} + \sum_{i=1}^{k} \hat{c}_i^6 \Delta Z_{t-i} + \hat{\varepsilon}_t. \tag{36}$$

#### 4.2.2 Cointegration Analyses

Following the application of cointegration approach in electricity consumption and GDP by Shiu and Lam (2004), and Lai et al. (2011), the Johansen cointegration test (Johansen, 1991) was used:

$$\Delta Z_t = \mu + \Pi Z_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Z_{t-i} + \varepsilon_t. \tag{37}$$

In addition, Tang et al. (2013) also used the ARDL model to test the cointegration relationship. Hence, the ARDL model in this study can be expressed as follows:

$$D(Y_t) = a_{01} + b_{11}Y_{t-1} + b_{21}E_{t-1} + b_{31}W_{t-1} + \sum_{i=1}^{p} a_{1i}^{1} D(Y_{t-i}) + \sum_{i=1}^{q} a_{2i}^{1} D(E_{t-i}) + \sum_{i=1}^{q} a_{3i}^{1} D(W_{t-i}) + \varepsilon_{1t}, \tag{38}$$

$$D(E_t) = a_{02} + b_{12}Y_{t-1} + b_{22}E_{t-1} + b_{32}W_{t-1} + \sum_{i=1}^{p} a_{1i}^{2} D(E_{t-i}) + \sum_{i=1}^{q} a_{2i}^{2} D(Y_{t-i}) + \sum_{i=1}^{q} a_{3i}^{2} D(W_{t-i}) + \varepsilon_{2t}, \tag{39}$$

$$D(W_t) = a_{3j} + b_{13}Y_{t-1} + b_{23}E_{t-1} + b_{33}W_{t-1} + \sum_{i=1}^{p} a_{1i}^{3} D(W_{t-i}) + \sum_{i=1}^{q} a_{2i}^{3} D(Y_{t-i}) + \sum_{i=1}^{q} a_{3i}^{3} D(E_{t-i}) + \varepsilon_{3t}, \tag{40}$$

where D () represents the change of $Z_t$, $Z_t = (Y_t, E_t, W_t, K_t, L_t)^T$.

#### 4.2.3 Kernel-Based Regularized Least Squares

After cointegration, the KRLS method to estimate the cointegration relationship among metropolitan economic growth, electricity consumption, and water consumption was used:

$$k(X, X_t) = e^{-\frac{||X - X_t||^2}{\sigma^2}}, \tag{41}$$

where $X_t = W_t$ , *or* $E_t$. The target function Y = f(X) is as follows:

$$Y = f(X) = \sum_{t=1}^{T} c_i k(X, X_t) = \begin{bmatrix} k(X_1, X_1) & \cdots & k(X_1, X_T) \\ \vdots & \ddots & \vdots \\ k(X_T, X_1) & \cdots & k(X_T, X_T) \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_T \end{bmatrix} = KC, \tag{42}$$

where $c_i$ is a weight for each covariate vector. Hence, the solution is follows: $C^* = argmin \sum_t V(Y_t, f(X_t)) + \pi R(f(X_t)) = argmin(Y - KC)^T (Y - KC) + \pi C^T KC$.

## 5 EMPIRICAL RESULTS

The data of electricity consumption, water consumption, and GDP may have structural change during 1950–2014. Thus, the stability with break point to complement to the ADF approach that does not consider the data structural change has to be examined. **Table 3** shows the unit root test results from the ADF test. **Table 4** shows the unit root test results from the Perron test, and **Table 5** shows the unit root test results from the Zivot–Andrews test. All of them confirm that Y, E, and W are integrated at I (1).

In terms of cointegration tests, the Jonansen test results in **Table 6** and the ARDL bounds test results in **Table 7** both show that the economic performance, water, and electricity consumption have long-run equilibrium relationship.

## 5.1 Unit Root Tests

### 5.1.1 ADF Test and PP Test

The ADF test is applied to detect the possible presence of unit roots in $Y_t$, $W_t$, and $E_t$. The null hypothesis of unit root can be rejected in favor of the alternative hypothesis of no unit root when the *p*-value large. **Table 4** represents that no variable is stationary in their levels. On the other hand, $Y_t$, $W_t$, and $E_t$ are the stationary process in their first.

### 5.1.2 Perron Modified ADF Test With Exogenous Break Point

The framework that follows the work of Perron (1989), Perron and Vogelsang (1992), Vogelsang and Perron (1998), and Banerjee et al. (1992) supports the computation of the modified Dickey–Fuller tests that allow for levels and trends that differ across a single break date. The results of the Perron's modified ADF test in **Table 5** and the Zivot–Andrews test are detailed in **Table 6**, which show that non-stationary process is found in all series at level with intercept and trend, but variables are found to be stationary at first difference. This confirms that $Y_t$, $W_t$, and $E_t$ are integrated at I (1).

## 5.2 Cointegration Tests

According to the unit root test results, the integration of the variables is of the same order, and testing whether these variables are cointegrated over the sample period was continued.

### 5.2.1 Johansen Cointergration Test

**Table 7** shows the results of the Johansen test. Because the trace statistic of non-cointegrating equation and at most one cointegrating equation are greater than the 5% critical values, respectively, the test rejects the hypothesis of no cointegration and indicates that there is one cointegrating equation at the 5% significance level, so there is a long-run relationship between $Y_t$, $W_t$, and $E_t$ for Guangzhou.

### 5.2.2 ARDL Bounds Test Approach to Cointegration

Pesaran et al. (2001) critical values are based on the assumption that the variables are integrated of order I (0) or I (1). Unit root tests insure that none of the series is integrated of I (2) or higher. Armed with information about stationarity, the ARDL bounds testing approach to cointegration was applied. The results of the bound test are given in **Table 8**. From these results, it is clear that there is a long-run relationship between $Y_t$, $W_t$, and $E_t$, because their F-statistic are higher than the upper-bound critical value at the 1% level. This implies that the null

**TABLE 3 |** ADF unit root test results.

| Levels | | | | First differences | | | |
|---|---|---|---|---|---|---|---|
| **Variables** | **ADF test statistics** | **Lag length** | **Critical values** | **Variables** | **ADF test statistics** | **Lag length** | **Critical values** |
| Y | −1.613,673 | 10 | −2.916,566* | ΔY | 1.721,083 | 9 | −1.612,934* |
| E | 2.853,394 | 1 | −3.482,763*** | ΔE | 2.822,923 6 | 15 | −2.614,029*** |
| W | 1.891,674 | 0 | −2.907,660*** | ΔW | −6.004387 | 2 | −3.538,362*** |

*Note: * shows significance at 10% level; ** shows significance at 5% level; and *** shows significance at 1% level.*

**TABLE 4 |** Perron's modified ADF unit root test results.

| Break Type: Innovative outlier | | | | Break Type: Innovative outlier | | | |
|---|---|---|---|---|---|---|---|
| **Variables** | **T-statistic** | **Break data** | **5% critical values** | **Variables** | **T-statistic** | **Break data** | **5% critical values** |
| Y | −4.283,895 | 2006 | −4.443,649 | ΔY | −5.743,232** | 2003 | −4.443,649 |
| E | −0.960,870 | 1984 | −4.443,649 | ΔE | −7.42739 *** | 2001 | −4.949,133 |
| W | −0.960,871 | 1984 | −4.443,649 | ΔW | −7.427,391** | 2001 | −4.193,627 |

*Note: 1) ** indicates significance at 5% level; 2) innovative outlier affecting every member of a set of autoregressive time series at the same time point are represented as independent random effects.*

**TABLE 5 |** Zivot–Andrews structural break trended unit root test results.

| Variables | At level | | | Variables | At first difference | | |
|---|---|---|---|---|---|---|---|
| | **T-statistic** | **Break data** | **5% critical values** | | **T-statistic** | **Break data** | **5% critical values** |
| Y | −2.293,951 | 1999 | −4.93 | ΔY | −5.264,743** | 1998 | −5.08 |
| E | −0.944,716 | 2003 | −4.93 | ΔE | −4.352,876* | 1988 | −4.11 |
| W | −3.251,608 | 2001 | −4.93 | ΔW | −5.805,366** | 1995 | −4.93 |

*Note: **indicates significance at 5% level.*

**TABLE 6 |** Johansen cointegration test results.

| Hypothesized number of cointegrating equation | Maximum eigenvalues statistics | Trace statistic | 5% Critical value |
|---|---|---|---|
| None** | 100.3432 | 151.1318 | 69.81889 |
| At most 1** | 78.4543 | 77.37808 | 47.85613 |
| At most 2 | 34.1234 | 35.69042 | 29.79707 |
| At most 3 | 14.8879 | 15.78851 | 15.49471 |
| At most 4 | 1.2333 | 1.516,098 | 3.841,466 |

*Note: **indicates significance at 5% level.*

**TABLE 7 |** Bounds test results.

| Bounds testing to cointegration | | |
|---|---|---|
| **Estimated model** | **Lag length** | **F-statistic** |
| f (Y/E,W) | (4,4,3,4,3) | 9.976,565** |
| f (E/W,Y) | (4,3,4,4,4) | 11.85424** |
| f (W/E,Y) | (2,1,3,2,3) | 8.490,854** |
| 5% critical values | I (0) | I (1) |
| | 3.29 | 4.37 |

*Note: ***indicates significance at 5% level. The estimated models just show which is the dependent variable.*

**TABLE 8 |** Results of KRLS (dependent variable = $Y_t$).

| | |
|---|---|
| $W_t$ | 130.899 ***(33.0933) |
| $E_t$ | 10.1467***(0.801,614) |
| R-squared | 0.9335 |
| Number of observations | 67 |
| Lambda | 0.07082 |
| Tolerance | 0.067 |
| Sigma | 2 |
| Looloss | 52,180 |

*Notes: 1)*** indicates that the variable is significant at the 1% level, ** indicates significant at the 5% level, and * indicates significant at the 10% level; 2) robust standard errors in parentheses.*

hypothesis of no cointegration between $Y_t$, $W_t$, and $E_t$ is rejected.

### 5.3 Further Discussion: Results From Kernel-Based Regularized Least Squares

At the end, to estimate the relation in (2.31) above, KRLS to tackle regression problem without a specification search was utilized. The results from **Table 8** suggest a statistically significant relationship between Y, E, and W. The KRLS results also suggest that the $R^2$ from KRLS is high enough, which reveals that this regression model fit results. The empirical results indicate that the metropolitan economic growth is positively correlated with the electricity consumption and the water consumption is positively associated with the metropolitan economic performance.

The empirical results are consistent with other studies in developing countries (Ghosh, 2002; Jumbe, 2004; Mehrara, 2007; Narayan and Prasad, 2008; Jamil and Ahmad, 2010; Shahbaz et al., 2011). However, the empirical results in this study contradict with the empirical results of He and Huang (2020). This result is also helpful for metropolitan policy-making in balancing the relationship between electricity use, water consumption, and economic growth. In particular, the experience of economic growth in Guangzhou provides evidence that the resources consumption generates more GDP.

Indeed, both ARDL and Johansen cointegration techniques show that electricity consumption, urbanization, and metropolitan economic performance are cointegrated (He, 2020). Hence, the empirical results regarding to the relationship between electricity consumption and metropolitan economic growth are consistent with his findings, but our work considers the interaction effect of water consumption. Actually, the water consumption is determined by the intersection of endogenous growth function and water consumption function, neither function can be consistently identified by comparing average quantities of water consumed at different values of observed real per capita output (He and Fullerton, 2020). Obviously, our results confirm their arguments and provide new evidence for the long-run relationship between water consumption, electricity consumption, and endogenous economic growth.

## 6 CONCLUSION AND POLICY RECOMMENDATION

The electricity consumption and water consumption are vitally important for an economic developing region. This study builds a dynamic general equilibrium model that connects electricity consumption and water consumption with metropolitan economic performance. It also conducts econometric research on the interconnection among them for the third largest metropolitan in China (Guangzhou) from 1950 to 2014. The cointegration techniques show that electricity consumption, water consumption, and the metropolitan economic performance have long-run equilibrium relationship. The results of KRLS reveal that the metropolitan economic growth is positively correlated with electricity consumption. Also consistent with the theory, the water consumption is positively associated with the metropolitan economic performance. These results are generally stable and hold with alternative measures of unit roots, with alternative estimation strategies, and with or without controlling for trends, intercepts, and break points.

Thus, the policy implication of that is, if to improve the electricity and water consumption efficiency, then the resources consumption pattern should be changed into a pattern of low resource intensity. The policy decision-makers should 1) set up R&D funds to drive up distributed energy resources technology, 2) implement subsidy policies to develop resource-saving business models, 3) design an efficient system of prices and regulate charges for electricity services, and 4) enlarge the scale of investment on water utilities and development of water plants. Furthermore, on the basis of the analytical findings of this study, it is proposed that policy-makers and officials continue to enhance their interventions aimed at promoting successful economic development, water use, and increasing electricity consumption. In China, energy, water, and economic developments have had a severe effect. China has faced both the beneficial and detrimental impacts of economic development, as have many other developed countries.

However, this study has limitations that include the following: 1) the theoretical model assumes utility optimization and dynamic general equilibrium; 2) the specification of econometric model does not consider macroeconomic factors such as interest rates; and 3) important uncertainties referring to the export and import between districts, electricity market structures, and decarbonizing technology are not included in the empirical models. It is supported that a more systematic investigation and measurement of joint policy of energy and environment and how they interact with economic growth factors are promising areas for future research (Henryk and Łukasz, 2012; Han and Kung, 2015; Alola et al., 2021; Gyamfi et al., 2021; Dagar, 2021; Gyamfi et al., 2021).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

YH collected and analyzed data, and built theoretical models. SG wrote and revised the manuscript.

## FUNDING

This work is financially supported by the National Ten Thousand Outstanding Young Scholar Program (W02070352) and the National Social Science Foundation in China (19FJYB050).

## REFERENCES


Alola, A. A., Ozturk, I., and Bekun, F. V., (2021). Is Clean Energy prosperity and Technological Innovation Rapidly Mitigating Sustainable Energy-Development Deficit in Selected Sub-saharan Africa A Myth or Reality. *Energy Policy* 158, 112520. doi:10.1016/j.enpol.2021.112520

Baloch, M. A., Ozturk, I., Bekun, F. V., and Khan, F. V. (2021). Modeling the Dynamic Linkage between Financial Development, Energy Innovation, and Environmental Quality: Does Globalization Matter. *Bus Strat Env.* 30 (1), 176–184. doi:10.1002/bse.2615

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992). Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence. *J. Business Econ. Stat.* 10 (3), 271–287. doi:10.2307/1391542

Barbier, E. B., and Chaudhry., A. M. (2014). Urban Growth and Water. *Water Resour. Econ.* 6, 1–17. doi:10.1016/j.wre.2014.05.005

Barbier, E. B. (2004). Water and Economic Growth. *Econ. Rec.* 80 (248), 1–16. doi:10.1111/j.1475-4932.2004.00121.x

Bekun, F. V., Alola, A., Gyamfi, B. A., and Ampomah, A. B. (2021a). The Environmental Aspects of Conventional and Clean Energy Policy in Sub-saharan Africa: Is N-Shaped Hypothesis Valid. *Environ. Sci. Pollut. Res. Int.*, 1, 1–14. doi:10.1007/s11356-021-14758-w

Bekun, F. V., Gyamfi, B. A., Onifade, S. T., and Agboola, M. O. (2021b). Beyond the Environmental Kuznets Curve in E7 Economies: Accounting for the Combined Impacts of Institutional Quality and Renewables. *J. Clean. Prod.*, 314. 127924. doi:10.1016/j.jclepro.2021.127924

Cazcarro, I., Duarte, R., and Sánchez-Chóliz, J. (2013). Economic Growth and the Evolution of Water Consumption in Spain: A Structural Decomposition Analysis. *Ecol. Econ.* 96 (June), 51–61. doi:10.1016/j.ecolecon.2013.09.010

Dagar, V. (2021). Reinvigorating the Role of Clean Energy Transition for Achieving a Low-Carbon Economy: Evidence From Bangladesh. *Environ. Sci. Pollut. Res.*.

Feng, K., Hubacek, K., Siu, Y. L., and Li, X. (2014). The Energy and Water Nexus in Chinese Electricity Production: A Hybrid Life Cycle Analysis. *Renew. Sustainable Energ. Rev.* 39, 342–355. doi:10.1016/j.rser.2014.07.080

Ghosh, S. (2002). Electricity Consumption and Economic Growth in India. *Energy Policy* 30, 125–129. doi:10.1016/s0301-4215(01)00078-7

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37 (3), 424–438. doi:10.2307/1912791

Gleick, P. H. (2003). Water Use. *Annu. Rev. Environ. Resour.* 28 (1), 275–314. doi:10.1146/annurev.energy.28.040202.122849

GSD (2015). *Guangzhou, CN: Guangzhou Municipal Bureau*. Guangzhou Statistical Yearbook. Available at: http://Data.Gzstats.Gov.Cn/GzStat1/Chaxun/Njsj.Jsp.

Gyamfi, B. A., Bein, M. A., Ozturk, I., and Bekun, F. V. (2021). The Moderating Role of Employment in an Environmental Kuznets Curve Framework Revisited in G7 Countries. *Indonesian J. Sustainability Account. Management* 4 (2), 241–248.

Han, L., and Kung, J. K.-S., (2015). Fiscal Incentives and Policy Choices of Local Governments: Evidence from China. *J. Development Econ.* 116, 89–104. doi:10.1016/j.jdeveco.2015.04.003

He, Y., and Huang, H. (2020). Energy Intensity in Guangdong of China (2006–2015): A Spatial Dynamic General Equilibrium Econometric Model. *Nat. Resource Model.* 33, e12247. doi:10.1111/nrm.12247

He, Y. (2020). Agricultural Population Urbanization, Long-Run Economic Growth, and Metropolitan Electricity Consumption: An Empirical Dynamic General Equilibrium Model. *Energ. Strategy Rev.* 30, 100498. doi:10.1016/j.esr.2020.100498

He, Y., Fullerton, T. M., and Walke, A. G. (2017). Electricity Consumption and Metropolitan Economic Performance in Guangzhou: 1950-2013. *Energ. Econ.* 63, 154–160. doi:10.1016/j.eneco.2017.02.002

He, Y., and Fullerton, T. (2020). The Economic Analysis of Instrument Variables Estimation in Dynamic Optimal Models with an Application to the Water Consumption. *Agric. Econ. - Czech* 66, 413–423. doi:10.17221/270/2020-agricecon

Henryk, G., and Łukasz, L. (2012). The Electricity Consumption versus Economic Growth of the Polish Economy. *Energ. Econ.* 34, 500–510.

Jamil, F., and Ahmad, E. (2010). The Relationship between Electricity Consumption, Electricity Prices and GDP in Pakistan. *Energy Policy* 38 (10), 6016–6025. doi:10.1016/j.enpol.2010.05.057

Jiang, L., Folmer, H., and Ji, M. (2014). The Drivers of Energy Intensity in China: A Spatial Panel Data Approach. *China Econ. Rev.* 31, 351–360. doi:10.1016/j.chieco.2014.10.003

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59 (6), 1551–1580. doi:10.2307/2938278

Jumbe, C. B. L. (2004). Cointegration and Causality between Electricity Consumption and GDP: Empirical Evidence from Malawi. *Energ. Econ.* 26 (1), 61–68. doi:10.1016/s0140-9883(03)00058-6

Kahrl, F., and Roland-Holst, D. (2008). China's Water-Energy Nexus. *Water Policy* 10 (1), 51–65. doi:10.2166/wp.2008.052

Katz, D. (2015). Water Use and Economic Growth: Reconsidering the Environmental Kuznets Curve Relationship. *J. Clean. Prod.* 88, 205–213. doi:10.1016/j.jclepro.2014.08.017

Lai, T. M., To, W. M., Lo, W. C., Choy, Y. S., and Lam, K. H. (2011). The Causal Relationship between Electricity Consumption and Economic Growth in a Gaming and Tourism Center: The Case of Macao SAR, the People's Republic of China. *Energy* 36, 1134–1142. doi:10.1016/j.energy.2010.11.036

Liu, Z., Guan, D., Wei, W., Davis, S. J., Ciais, P., Bai, J., et al. (2015). Reduced Carbon Emission Estimates From Fossil Fuel Combustion and Cement Production in China. *Nature.* 524, 335–338. doi:10.1038/nature14677

Magazzino, C. (2014). Electricity Demand, GDP and Employment: Evidence from Italy. *Front. Energ.* 8 (1), 31–40. doi:10.1007/s11708-014-0296-8

Magazzino, C. (2017). Stationarity of Electricity Series in MENA Countries. *Electricity J.* 30 (10), 16–22. doi:10.1016/j.tej.2017.11.014

Mehrara, M. (2007). Energy Consumption and Economic Growth: The Case of Oil Exporting Countries. *Energy Policy* 35 (5), 2939–2945. doi:10.1016/j.enpol.2006.10.018

Mele, M., and Magazzino, C. (2020). Machine Learning Analysis of the Relationship Among Iron and Steel Industries, Air Pollution, and Economic Growth in China. *J. Clean. Prod.* 277. 123293. doi:10.1016/j.jclepro.2020.123293

Narayan, P. K., and Prasad, A. (2008). Electricity Consumption-Real GDP Causality Nexus: Evidence from a Bootstrapped Causality Test for 30 OECD Countries. *Energy Policy* 36 (2), 910–918. doi:10.1016/j.enpol.2007.10.017

Ngoran, S. D., Xue, X. Z., and Wesseh, P. K. (2016). Signatures of Water Resources Consumption on Sustainable Economic Growth in Sub-saharan African Countries. *Int. J. Sustainable Built Environ.* 5 (1), 114–122. doi:10.1016/j.ijsbe.2016.04.002

Ohdoko, T., Komatsu, S., and Kaneko, S. (2013). Residential Preferences for Stable Electricity Supply and a Reduction in Air Pollution Risk: a Benefit Transfer Study Using Choice Modeling in China. *Environ. Econ. Pol. Stud.* 15, 309–328. doi:10.1007/s10018-013-0061-y

Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica* 57 (6), 1361–1401. doi:10.2307/1913712

Perron, P., and Vogelsang, T. J. (1992). Testing for a Unit Root in a Time Series With a Changing Mean: Corrections and Extensions. *J. Business Econ. Stat.* 10 (4), 467–470. doi:10.2307/1391823

Pesaran, M. H., Shin, Y., and Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *J. Appl. Econ.* 16, 289–3326. doi:10.1002/jae.616

Rehman, A., and Deyuan, Z. (2018). Pakistan's Energy Scenario: a Forecast of Commercial Energy Consumption and Supply from Different Sources through 2030. *Energ Sustain. Soc.* 8, 26. doi:10.1186/s13705-018-0167-y

Rehman, A., Irfan, M., Hena, S., and Chandio, A. A. (2020). By Applying an ARDL Bounds Testing Approach and Causality Test to Investigate the Electricity Consumption and Production with Economic Growth. *Wjstsd* 17 (2), 182–199. doi:10.1108/wjstsd-08-2019-0054

Rehman, A., Ma, H., Ozturk, I., Ahmad, M., and Irfan, M. (2021d). Another Outlook to Sector-Level Energy Consumption in Pakistan from Dominant Energy Sources and Correlation with Economic Growth. *Environ. Sci. Pollut. Res.* 28, 33735–33750. doi:10.1007/s11356-020-09245-7

Rehman, A., Ma, H., and Ozturk, I. (2021b). Do industrialization, Energy Importations, and Economic Progress Influence Carbon Emission in Pakistan. *Environ. Sci. Pollut. Res.* 28, 45840–45852. doi:10.1007/s11356-021-13916-4

Rehman, A., Ma, H., Radulescu, M., Sinisi, C. I., and Yousaf, Z. (2021a). Energy Crisis in Pakistan and Economic Progress: Decoupling the Impact of Coal Energy Consumption in Power and Brick Kilns. *Mathematics* 9, 2083. doi:10.3390/math9172083

Rehman, A., Radulescu, M., Ma, H., Dagar, V., Hussain, I., and Khan, M. K. (2021c). The Impact of Globalization, Energy Use, and Trade on Ecological Footprint in Pakistan: Does Environmental Sustainability Exist. *Energies* 14, 5234. doi:10.3390/en14175234

Rehman, A. (2020). "The Nexus of Electricity Access, Population Growth, Economic Growth in Pakistan and Projection through 2040: An ARDL to Co-integration Approach", *Int. J. Energ. Sector Management*, 13 (3), 747–763.

Shahbaz, M., Tang, C. F., and Shabbir, M. S. (2011). Electricity Consumption and Economic Growth Nexus in Portugal Using Cointegration and Causality Approaches. *Energy Policy* 39 (6), 3529–3536. doi:10.1016/j.enpol.2011.03.052

Shiu, A., and Lam, P.-L. (2004). Electricity Consumption and Economic Growth in China. *Energy Policy* 32 (1), 47–54. doi:10.1016/s0301-4215(02)00250-1

Stern, S., Cimarelli, C., Gaudin, D., Scheu, B., and Dingwell, D. B. (2019). Electrification of Experimental Volcanic Jets With Varying Water Content and Temperature. *Geophys. Res. Lett.* 46, 11136–11145. Received 24 JUL 2019 Accepted 23 SEP 2019 Accepted article online 10 OCT 2019 STERN ET AL. doi:10.1029/2019GL084678

Tang, C., Muhammad, F. S., and Mohamed, A. (2013). Re-investigating the Electricity Consumption and Economic Growth Nexus in Portugal. *Energy Policy* 36 (8), 3077–3085. doi:10.1016/j.enpol.2013.07.073

Udemba, E. N., Magazzino, C., and Bekun, F. V. (2020). Modeling the Nexus between Pollutant Emission, Energy Consumption, Foreign Direct Investment, and Economic Growth: New Insights from China. *Environ. Sci. Pollut. Res. Int.* 27, 17831–17842. doi:10.1007/s11356-020-08180-x

Vogelsang, T. J., and Perron, P. (1998). Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time. *Int. Econ. Rev.* 39, 1073–1100. doi:10.2307/2527353

Wang, Z., Han, B., and Lu, M. (2016). Measurement of Energy Rebound Effect in Households: Evidence from Residential Electricity Consumption in Beijing, China. *Renew. Sustainable Energ. Rev.* 58, 852–861. doi:10.1016/j.rser.2015.12.179

Xu, C. (2011). The Fundamental Institutions of China's Reforms and Development. *J. Econ. Lit.* 49 (4), 1076–1151. doi:10.1257/jel.49.4.1076

Zhang, C., Zhong, L., Fu, X., Wang, J., and Wu, Z. (2016). Revealing Water Stress by the Thermal Power Industry in China Based on a High Spatial Resolution Water Withdrawal and Consumption Inventory. *Environ. Sci. Technol.* 50 (4), 1642–1652. doi:10.1021/acs.est.5b05374

Zhu, Q., Sarkis, J., and Geng, Y. (2005). Green Supply Chain Management in China: Pressures, Practices and Performance. *Int. Jrnl Op Prod. Mnagemnt.* 25 (5), 449–468. doi:10.1108/01443570510593148

Zivot, E., and Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *J. Business Econ. Stat.* 10 (3), 251–270. doi:10.1080/07350015.1992.10509904

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2021 He and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fenvs.2021.793221

# Financial Inclusion and Carbon Reduction: Evidence From Chinese Counties

*Zhenkai Yang[1], Lu Yu[2], Yinwei Liu[2], Zhichao Yin[2] and Zumian Xiao[2]**

[1]School of Economics and Management, Wuhan University, Wuhan, China, [2]School of Finance, Shandong University of Finance and Economics, Jinan, China

With the improvement of inclusive financial system, China's economy has made significant development and growth. It worth in-depth investigation on environmental impact of financial inclusion, since growing GDP usually accompanied by more intensive carbon emission. This paper aims to reveal whether financial inclusion contributes to the carbon reduction in China using county-level dataset. A fixed-effect panel regression approach is adopted to examine the impact of financial inclusion on county-level regional carbon emissions. The estimation results imply that financial inclusion plays an important role in reducing carbon emissions. The mediation effect analysis reveals two channels through which financial inclusion imposes negative impact on the level of regional carbon emissions. One is to elevate the carbon sequestration capacity by increasing vegetation coverage, and the other is to improve the industrial structure through enhanced financial support. In addition to being a bridge between economic opportunity and output, financial inclusion can also act as an effective measure for addressing climate change.

**Keywords: financial inclusion, carbon reduction, industrial structure, carbon emission, carbon sequestration**

OPEN ACCESS

***Edited by:***
*Yanfei Li,*
*Hunan University of Technology and Business, China*

***Reviewed by:***
*Luigi Aldieri,*
*University of Salerno, Italy*
*Sevda Kuşkaya,*
*Erciyes University, Turkey*

****Correspondence:***
*Zumian Xiao*
*xiaozumian@sdufe.edu.cn*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *11 October 2021*
***Accepted:*** *24 December 2021*
***Published:*** *12 January 2022*

***Citation:***
*Yang Z, Yu L, Liu Y, Yin Z and Xiao Z (2022) Financial Inclusion and Carbon Reduction: Evidence From Chinese Counties. Front. Environ. Sci. 9:793221. doi: 10.3389/fenvs.2021.793221*

## INTRODUCTION

Global consensus has been reached that the emission of carbon dioxide and other greenhouse gases is the main cause of climate deterioration and its social and economic consequences. In addition, it can also cause significant harm to human health (Dong et al., 2021). In December 2015, 196 parties around the world entered into the "Paris Climate Change Agreement". The goal of this agreement is to control the increase in global average temperature below 2°C. To achieve this goal, more effective policies are needed to control the growing carbon emissions (Cosmas et al., 2019).

This study examines the impact of financial inclusion on carbon emissions. We focus on issues in China as its carbon dioxide emissions account for a large proportion of the world. Financial inclusion has played an important role in China's economic growth. Because of the fact that economic expansions usually come along with higher level of carbon emission, it is of great significance to study the impact of financial inclusion on carbon emissions. In addition, our research can shed light on similar situation that other developing countries may encounter with.

The Paris Climate Change Agreement emphasizes the important role of financial support for global climate governance. Recent studies support this notion. It is shown that the development of financial inclusion is an important factor affecting carbon emission of a country or region (Le et al., 2020). In theory, the development of financial inclusion may affect carbon emissions in opposite ways. On one hand, financial inclusion can lead to investment in technological innovation by

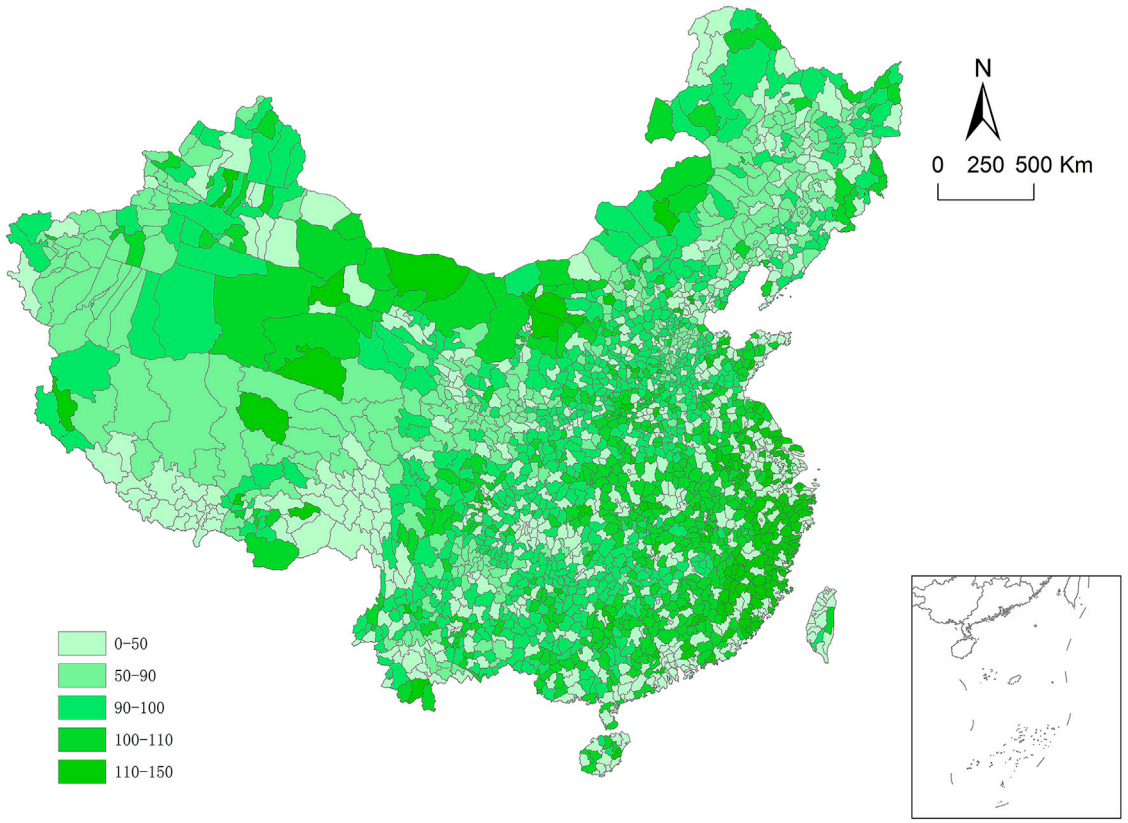



**FIGURE 1 |** Financial inclusion index of Chinese counties in 2016. Data source: IDF, Available at https://en.idf.pku.edu.cn/index.htm.

broadening financing channels and reducing financing costs, which resulting in more energy-saving and environment friendly production (Tamazian et al., 2009). On the other hand, financial inclusion promotes economic growth, which further increases the demand for energy. An increase in energy demand usually means an increase in carbon emission (Sadorsky, 2010). The existing literatures have not yet reached a consistent conclusion about the impact of financial inclusion on carbon emissions, which is important and requires further research. At the same time, a considerate financial policy needs to be formulated by a country to achieve climate goals.

Comparing to alternative measures, improving the capacity of ecosystems to absorb carbon seems to be an appealing solution. However, either restoring land or afforestation require notable investments in labor, capital and other resources. The reality is that it is notoriously difficult to make the necessary investments in rural China. Another feasible approach, low carbon transition of industrial production, also requires considerable investment. Under such a background, financial inclusion can be adopted as an additional financing channel aiming at incentivizing low-carbon activities. Therefore, in combination with China's institutional environment, it is of great significance to thoroughly examine the impact of financial inclusion on carbon emissions.

**Figures 1**, **2** depict the development of financial inclusion and carbon emissions of Chinese counties in 2016, respectively. The two indicators shown are highly heterogeneous in geographical distribution. No obvious correlation or causality relationships are observed. The investigation of the relationship between financial inclusion and carbon emissions relies on in-depth empirical analysis.

In recent years, China has been vigorously promoting the implementation of inclusive finance, which has greatly promoted economic development, especially in remote districts and counties. The continued involvement of businesses will be vital for unlocking opportunities to expand financial inclusion (World Bank, 2018). However, we should also pay attention to environmental quality while pursuing economic development. Dasgupta et al. (2001) noted that the capital market in developing countries may provide appropriate financial and reputation incentives for corporate pollution control, which may be an effective way of environmental governance. Thus, the research aims to explore the relationship between inclusive financial development and carbon emissions.

This paper evaluates the impact of the development of financial inclusion on carbon emissions based on panel data at county level. By providing more convenient and affordable access to finance, financial inclusion is able to elevate the efficiency and breadth of financial services. We adopt panel regression and

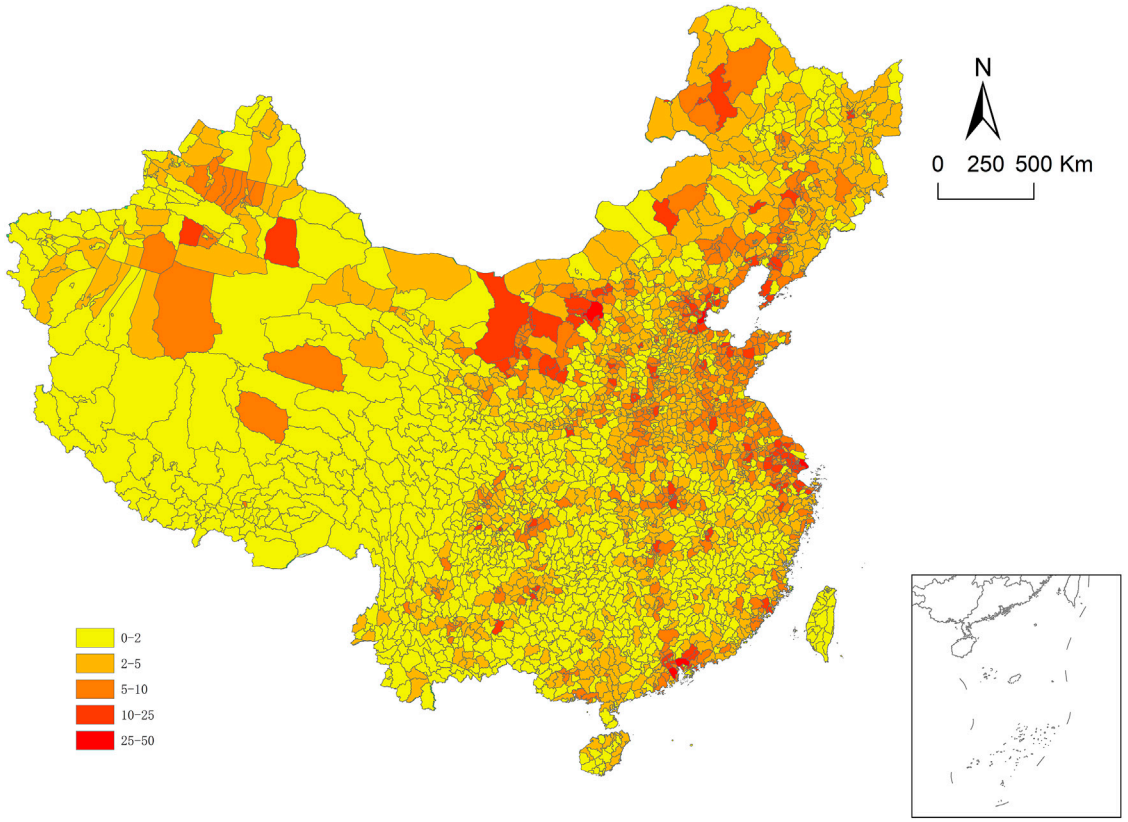



**FIGURE 2 |** Carbon emissions of Chinese counties in 2016 (Million tons). Data source: Chen et al. (2020).

mediation effect model and discover the carbon reduction effect of financial inclusion. Our findings are contrary to the relevant studies, such as Le et al. (2020) and Zaidi et al. (2021). They provide evidence that the development of financial inclusion is positively correlated with carbon emissions based on panel data from multiple countries. Compared with these studies, this paper suffers less from aggregation bias because of the county-level dataset, which is gathered from the basic administrative unit in China.

The marginal contributions of this paper to the literature are as follows: First, to the best of our knowledge, this paper is the first attempt to study the carbon reduction effect of financial inclusion in China at the county level. We use a dataset of greater data granularity compared to existing researches, which makes it possible to excavate deeper into the relationship between financial inclusion and carbon emission. Second, mediation effect model is introduced to unravel the mechanisms that how financial inclusion contributes to regional carbon reduction. The results help us to better understand why and how financial inclusion changes emission. Third, we decompose the financial inclusion index into three different dimensions, including the breadth of coverage, depth of use, and degree of digitization. By doing so, different aspects of financial inclusion development are taken into account. Thus, the impact of financial inclusion on carbon emission can be discussed more thoroughly. The empirical results shed light on the formation of China's climate change policies.

The rest of this paper proceeds as follows. Second 2 is literature review. Second 3 presents the baseline model setting, methodologies, and variable selection. Second 4 discusses the empirical results and the mechanisms. Second 5 draws the conclusion and puts forward policy implications.

## LITERATURE REVIEW

Since greenhouse gas emissions have triggered global climate problems, the academic community has conducted a great deal of research on the factors that influence carbon emissions. Existing literature shows that plenty factors significantly affect carbon emissions, such as urbanization, population size, income, energy consumption, GDP, financial development, and so on.

Wang W.-Z. et al. (2021) believe urbanization will affect carbon emissions through three paths that affect economic growth, energy efficiency, and consumption structure. Urbanization in OECD countries is conducive to reducing carbon dioxide emissions. Shafiei and Salim (2014) document that the impact of urbanization on carbon dioxide emissions is non-linear. When the level of urbanization in a country is low, it is usually accompanied by more serious environmental pollution; when the level of urbanization reaches a certain level, the degree

of environmental pollution will decrease. Farhani (2015) examined the relationship between renewable energy consumption, economic growth and carbon dioxide emissions in 12 countries in Africa from 1975 to 2008. They find that renewable energy consumption played an important role in reducing carbon dioxide emissions. Li and Yuan (2014) argue that population size, income level and technological progress are the essential causes of the increase in global carbon emissions and climate warming. Saidi and Mbarek (2017) use a dataset from emerging market economies and find that financial development, income levels, trade openness, and urbanization are all driving factors of carbon dioxide emissions. Research from Malik et al. (2020) implies negative correlation between oil prices and carbon emissions. Rising oil prices reduce carbon emissions, and rising oil prices aggravate carbon emissions.

Except for the aforementioned factors, scholars also find that structural and technological factors could impose significant impact on carbon emission. Wang B. et al. (2021) point out that the allocation of industrial resources is the key factor affecting regional carbon emissions. Improving the efficiency of industrial resource allocation can significantly reduce carbon emissions. Khan and Yahong. (2021) believe that income inequality and population density play important role in changing the level of carbon emissions. Erdoğan et al. (2020) have studied the impact of innovation on sectoral carbon emissions based on data from 14 states of the G20 from 1991 to 2017. They find that innovation reduces carbon emissions in the industrial sector. However, innovation has no significant impact on carbon emissions in transportation, energy and other sectors.

Regarding the impact of financial development on carbon emissions, academic researchers have different views. On one hand, scholars have documented empirical evidence implying financial development leads to more carbon emissions. Sadorsky (2010) notes that financial development can make it easier for consumers to obtain credit, thereby stimulating consumers to purchase energy-consuming products such as cars, and increasing carbon emissions. Zhang (2011) adopts co-integration examination, Granger causality test, and variance decomposition methods to explore the impact of financial development on carbon emissions. The study suggests that China's financial development is an important driving force for the increase in carbon emissions. And the size of financial intermediaries has the greatest impact on carbon emissions among other factors. Mahalik et al. (2016) believes that financial development makes it easier for companies to finance, and the expansion of production scale promotes carbon emissions.

On the other hand, there are also scholars argue that financial development can reduce carbon emissions. Tamazian et al. (2009) believe that financial development will increase R&D investment and improve resource utilization efficiency, resulting in reduced carbon emissions. Tamazian and Rao (2010) point out that there are certain prerequisites for financial development to improve environmental quality. Only under a well-functioning institutional framework can financial development help reducing carbon dioxide emissions, otherwise, it will lead to environmental degradation. Jalil and Feridun. (2011) use the autoregressive distributed lag (ARDL) method to examine the relationship between financial development and environmental pollution, and found that the coefficient of financial development was significantly negative, indicating that financial development reduced carbon emissions. Brunnschweiler (2010) uses panel data from 119 non-economic cooperation organizations from 1980 to 2006 and found that financial intermediaries represented by commercial banks can effectively increase the production of renewable energy, thereby reducing carbon emissions from traditional energy consumption.

The reason for the diametrically opposite conclusions, Nasreen and Anwar (2015) speculate, is that countries or regions are in different stages of economic development. In high-income countries or regions, financial development can help reducing carbon emissions, and in low- and middle-income countries or regions, financial development increases carbon emissions.

Previous literature has explored the impact of financial development on carbon emission (e.g., Brunnschweiler,2010). Our contribution to this growing literature is to focus on data with finer granularity and more dimentions, and to draw more explicit policy recommendations.

For the most relevant studies, Le et al. (2020) and Zaidi et al. (2021), positive correlation between the development of financial inclusion and carbon emissions has been observed. Using cross country panel data, these two studies argue that the improvement in financial inclusion leads to more intensive carbon emission. Compared with these studies, we construct a county-level dataset trying to minify the aggregation bias comes from the country-level macro data. And, the county-level dataset enables us to identify the mechanism and to perform in-depth analysis. In addition, we deeply analyze the impact of financial inclusion on carbon emission by decomposing the financial inclusion index into three different dimensions, including the breadth of coverage, depth of use, and degree of digitization.

# METHODOLOGY AND DATA

## Panel Regression Model

In terms of research methods, panel regression with time and individual fixed effect is introduced to estimate the impact of financial inclusion on carbon emission, which is consistent with the previous literatures. And the finer data granularity makes our estimates suffer much less from the aggregation bias. The baseline econometric model is as follow:

$$E\text{mission}_{it} = \alpha + \beta Fin - inclusion_{it} + \boldsymbol{\gamma}\mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (1)$$

where, i and t denote county and year respectively. $Emission_{it}$ represents the carbon dioxide emissions of county i in year t $Fin - inclusion_{it}$ is the index of financial inclusion development for a specific county in a given year. $\mathbf{X}_{it}$ is a vector of control variables, including the total output of firms (Output), the population of given county (Population), the area of county-level administrative unit (Area) and fixed asset investment (Invest). $\mu_t$ and $\eta_i$ represent time and individual fixed effect. $\varepsilon_{it}$ is the error term.

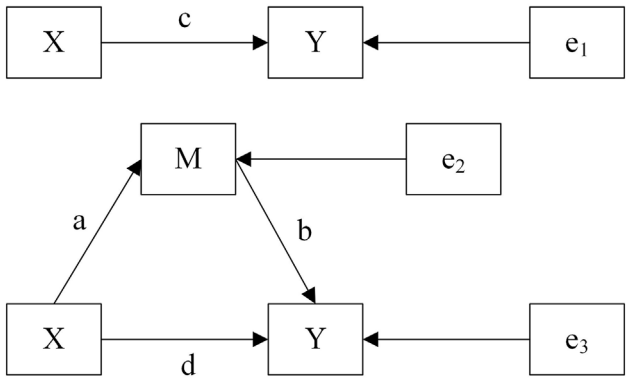



**FIGURE 3 |** Illustration of mediation effect model.

## Mediation Effect Model

This study evaluates the direct and indirect effects of financial inclusion on carbon emissions by estimating the mediation effect model. **Figure 3** illustrates that the independent variable X can influence the dependent variable Y indirectly, through a mediator M. When there is no mediating effect, based on the coefficient estimation of independent variable X we can interpret the direct effect (c) of X on Y. And this is the only effect exists. When there is a mediator, the indirect effect (a-b) of X on Y through M (the mediator) should be considered, along with the direct effect (d). The mediation effect model is estimated using two-stage regression method.

Both carbon absorption and carbon dioxide production as channels through which financial inclusion could impose influence on carbon emission. Carbon sequestration of vegetation cover is the most important means of carbon absorption. And the industrial structure, industrial production value as a percent of gross regional product, determines the carbon intensity of regional production activities and decides carbon dioxide production. These two variables are good proxies for carbon absorbing and releasing. So, they are introduced as the mediators in our mechanism analysis.

1) In order to test the mediating effect of carbon sequestration, we adopt the following regression model:

$$\text{Sequestration}_{it} = \xi + \zeta Fin - inclusion_{it} + \varsigma \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (2)$$

$$\text{Emission}_{it} = \alpha + \beta Fin - inclusion_{it} + \varphi \text{Sequestration}_{it} + \boldsymbol{\gamma} \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (3)$$

2) In order to test the mediating effect of industrial structure, we adopt the following regression model:

$$Industry_{it} = \xi + \zeta Fin - inclusion_{it} + \varsigma \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (4)$$

$$\text{Emission}_{it} = \alpha + \beta Fin - inclusion_{it} + \theta Industry_{it} + \boldsymbol{\gamma} \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (5)$$

By constructing these models, this study first examines the direct impact of financial inclusion on carbon reduction, and then we focus on the dual-channel chain carbon emission reduction effect of inclusive finance. The research framework of this study is shown in **Figure 4**.

## Data Description

We chose the largest possible time span. The financial inclusion index data is available from 2014 to 2020, and the county-level carbon emissions data is available from 1997 to 2016. Therefore, after merging the two databases, this paper uses environmental and economic data for the counties in China over the period of 2014–2016 for the empirical analysis. The data sources involved in the study are as follows: 1) Financial Inclusion Index is publicly available from the Institute of Digital Finance, Peking University (IDF)[1]. 2) Carbon emission and sequestration data for Chinese counties comes from Chen et al. (2020). 3) Control variables are gathered from "China Statistical Yearbook (County-Level)" published by National Bureau of Statistics of China. **Table 1** reports the variable description and our data sources.

# EMPIRICAL RESULTS

## Descriptive Statistics

After converting the dataset into a balanced panel, we have obtained 5,105 county-year observations. Descriptive statistics are reported in **Table 2**. The mean value of the financial inclusion index is 0.684, and the standard deviation is 0.212, indicating that there are significant differences in the development level of financial inclusion. The maximum of carbon emission is 25.71 and the average is 3.039, implying significant heterogeneity among regions.

## Variable Correlation

Pearson test is performed to examine the correlation between variables. The results are shown in **Table 3**. The correlation coefficients are presented along with significance levels. If the absolute value of correlation above 0.8, it implies possible colinearity between the two variables. The correlation analysis of variables shows that the variables involved in the table do not have the problem of colinearity, indicating that the selection of variables is feasible.

## Hausman Test

This study uses Hausman test to determine whether a fixed-effects model or a random-effects model should be adopted. The null hypothesis is that individual effects are not related to other independent variables. If the estimator obtained by OLS regression for the fixed-effect model is a consistent and efficient estimator of the true parameter, there will be no systematic difference between this estimator and a comparison estimator. The comparison estimator here refers to the estimator obtained by GLS regression for a random-effect model. If the null

[1]The official website for IDF is: https://en.idf.pku.edu.cn/index.htm.

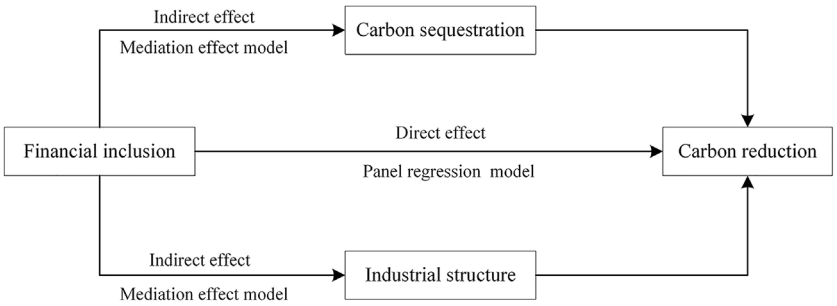



**FIGURE 4 |** Financial inclusion and carbon reduction: analytical framework.

**TABLE 1 |** Variable description and data source.

| Variable name | Variable description | Data source |
|---|---|---|
| Emission | Carbon dioxide emissions (Million tons) | Chen et al. (2020) |
| Fin-inclusion | Digital financial inclusion index | IDF |
| Sequestration | Carbon sequestration due to vegetation cover (Million tons) | Chen et al. (2020) |
| Industry | Industrial production value as a percent of gross regional product (%) | National Bureau of Statistics of China |
| Output | Total output of industrial enterprises above designated size (ten thousand yuan) | National Bureau of Statistics of China |
| Population | The total number of households at the end of the year | National Bureau of Statistics of China |
| Area | The area of county-level administrative unit (100,000 square kilometers) | National Bureau of Statistics of China |
| Invest | Fixed asset investment (billion yuan) | National Bureau of Statistics of China |

**TABLE 2 |** Descriptive statistics.

| Variable | N | Mean | Sd | Min | p50 | Max |
|---|---|---|---|---|---|---|
| Emission | 5105 | 3.039 | 2.831 | 0.0628 | 2.255 | 25.71 |
| Fin-inclusion | 5105 | 0.684 | 0.212 | 0.102 | 0.709 | 1.318 |
| Industry | 5105 | 0.437 | 0.142 | 0.017 | 0.443 | 0.887 |
| Sequestration | 5105 | 4.901 | 3.561 | 0.795 | 3.925 | 14.00 |
| Output | 5105 | 13.92 | 1.602 | 2.944 | 14.12 | 18.24 |
| Population | 5105 | 3.693 | 0.756 | 0.000 | 3.738 | 5.505 |
| Area | 5105 | 0.036 | 0.089 | 0.001 | 0.020 | 2.023 |
| Invest | 5105 | 15.16 | 14.23 | 0.194 | 11.32 | 133.4 |

hypothesis is not true, the parameter estimates of the fixed-effect model are still consistent, while that of the random-effect model are not. Hausman test provides a statistical test quantity based on the difference of the parameter estimates of the two. **Table 4** reports the Hausman test results. The *p*-value for the Chi-Sq statistic is 0.00. So, the null hypothesis should be rejected, implying a fixed-effect model should be used.

## Panel Regression Results

In order to investigate whether financial inclusion imposes significant impact on carbon emission, fixed-effect panel regression approach is adopted based on empirical model (1). **Table 5** shows the estimates and statistics. The control variables are gradually added in columns (1) to (5). The coefficients of financial inclusion are all significantly negative and significant at the level of 1%, which indicates that the development of digital financial inclusion can effectively reduce carbon dioxide emissions. Aluko and Obalade (2020) have already pointed out that due to the low level of financial development of sub-Saharan Africa, the carbon dioxide emissions keep rising along with the booming energy consumption. Our findings are consistent with evidence provide by this study. It can be seen that the

**TABLE 3 |** Correlation coefficients.

| | Emission | Fin-inclusion | Sequestration | Industry | Output | Population | Area | Invest |
|---|---|---|---|---|---|---|---|---|
| Emission | 1 | — | — | — | — | — | — | — |
| Fin-inclusion | 0.226*** | 1 | — | — | — | — | — | — |
| Sequestration | −0.174*** | −0.106*** | 1 | — | — | — | — | — |
| Industry | 0.310*** | 0.121*** | −0.230*** | — | — | — | — | — |
| Output | 0.544*** | 0.267*** | −0.302*** | 0.606*** | 1 | — | — | — |
| Population | 0.346*** | 0.0100 | −0.178*** | 0.090*** | 0.608*** | 1 | — | — |
| Area | 0.043*** | −0.00700 | 0.427*** | −0.046*** | −0.163*** | −0.270*** | 1 | — |
| Invest | 0.669*** | 0.350*** | −0.199*** | 0.369*** | 0.687*** | 0.505*** | −0.087*** | 1 |

*Note: *, **, *** indicate significance at 10, 5, and 1% level, and the t-values are in parentheses.*

**TABLE 4 |** Hausman test results.

| Variables | (1) | (2) |
|---|---|---|
| | Random Effect | Fixed Effect |
| | Emission | Emission |
| Fin-inclusion | −0.2839*** | −0.2556*** |
| | (−19.29) | (−19.67) |
| Output | 0.0527*** | 0.0164** |
| | (6.13) | (2.13) |
| Population | 0.6522*** | −0.0092 |
| | (12.52) | (−0.14) |
| Area | 2.7634*** | 3.1567* |
| | (4.62) | (1.87) |
| Invest | 0.0090*** | 0.0039*** |
| | (10.13) | (4.87) |
| Constant | −0.0510 | 2.8476*** |
| | (−0.23) | (10.86) |
| Observations | 5,105 | 5,105 |
| R-squared | — | 0.111 |
| No. of Counties | 1,844 | 1,844 |
| Statistic for Hausman test | Chi-Sq | *p*-value |
| Value | 153.95 | 0.00 |

*Note: ***p < 0.01, **p < 0.05, *p < 0.1 and t-statistics in parentheses. Column (1) and (2) report estimations from random-effect and fixed-effect panel regression respectively.*

**TABLE 6 |** Mediation effect estimation results: Carbon sequestration as mediator.

| Variables | (1) | (2) |
|---|---|---|
| | Sequestration | Emission |
| Fin-inclusion | 0.0967*** | −0.2536*** |
| | (3.51) | (−17.30) |
| Sequestration | — | −0.0213** |
| | — | (−2.53) |
| Output | −0.0177 | 0.0161* |
| | (−1.04) | (1.82) |
| Population | −0.1068 | −0.0115 |
| | (−1.23) | (−0.23) |
| Area | −6.8111 | 3.0116* |
| | (−1.14) | (1.67) |
| Invest | −0.0052*** | 0.0038*** |
| | (−3.38) | (2.71) |
| Constant | 5.7988*** | 2.9711*** |
| | (14.04) | (12.77) |
| Observations | 5,105 | 5,105 |
| R-squared | 0.009 | 0.112 |
| No. of Counties | 1,844 | 1,844 |

*Note: *, **, *** indicate significant at the levels of 10, 5, and 1%, and the t-values are in parentheses. The dependent variables for regressions (1) and (2) are carbon sequestration and carbon emission, respectively.*

**TABLE 5 |** Fixed-effect panel regression results.

| Variables | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Emission | Emission | Emission | Emission | Emission |
| Fin-inclusion | −0.2323*** | −0.2336*** | −0.2336*** | −0.2332*** | −0.2556*** |
| | (−19.37) | (−19.44) | (−19.42) | (−19.42) | (−17.38) |
| Output | — | 0.0257** | 0.0257** | 0.0248** | 0.0164* |
| | — | (2.56) | (2.56) | (2.51) | (1.85) |
| Population | — | — | 0.0013 | −0.0021 | −0.0092 |
| | — | — | (0.03) | (−0.04) | (−0.18) |
| Area | — | — | — | 2.9651* | 3.1567* |
| | — | — | — | (1.69) | (1.77) |
| Invest | — | — | — | — | 0.0039*** |
| | — | — | — | — | (2.77) |
| Constant | 3.1981*** | 2.8407*** | 2.8358*** | 2.7548*** | 2.8476*** |
| | (389.97) | (20.37) | (12.76) | (11.82) | (12.42) |
| Observations | 5,105 | 5,105 | 5,105 | 5,105 | 5,105 |
| R-squared | 0.100 | 0.103 | 0.103 | 0.104 | 0.111 |
| No. of Counties | 1,844 | 1,844 | 1,844 | 1,844 | 1,844 |

*Note: *, **, *** indicate significant at the levels of 10, 5, and 1%, and the t-values are in parentheses. The dependent variable for regressions (1)–(5) is carbon emission at county-level.*

development of financial inclusion plays an important role in the process of carbon emission reduction.

The signs of the coefficients estimation of control variables are in line with expectations. In column (5), the coefficient of the county-level output value of industrial enterprises (Output) above designated size is significantly positive at the 10% level, indicating that the increase in regional output brings more carbon emissions. The coefficient of population is negative but not significant, which implies the scale of population is not necessarily a driving force of carbon reduction. Nevertheless, the area of counties (Area) is positively correlated with carbon emission, with a significance level of 10%. The coefficient of investment in fixed assets (Invest) is also significantly positive at the 1% level. Investment in fixed assets leads to a considerable increase in carbon emissions.

# Mechanisms

## Carbon Sequestration

We examine the mediation effects using a two-stage regression approach according to model (2)–(3). Estimation results are presented in **Table 6**. The estimated coefficient of digital financial inclusion in column (1) is 0.9675, which is significant at the 1% level, suggesting that the development of financial inclusion has significantly increased the amount of carbon sequestration from vegetation cover. The estimated coefficient of carbon sequestration in column (2) is significant negative,

**TABLE 7 |** Mediation effect estimation results: Industrial structure as mediator.

| Variables | (1) | (2) |
|---|---|---|
| | Industry | Emission |
| Fin-inclusion | –0.0866*** | –0.2009*** |
| | (0.034) | (0.167) |
| Industry | — | 0.6326*** |
| | — | (0.090) |
| Output | 0.0257*** | 0.0002 |
| | (0.007) | (0.007) |
| Population | 0.0017 | –0.0103 |
| | (0.014) | (0.053) |
| Area | 0.9312 | 2.5676 |
| | (0.880) | (1.976) |
| Invest | 0.0010*** | 0.0032** |
| | (0.000) | (0.001) |
| Constant | 0.0834 | 2.7949*** |
| | (0.111) | (0.229) |
| Observations | 5,105 | 5,105 |
| R-squared | 0.335 | 0.123 |
| No. of Counties | 1,844 | 1,844 |

*Note: *, **, *** indicate significant at the levels of 10, 5, and 1%, and the t-values are in parentheses. The dependent variables for regressions (1) and (2) are industry and carbon emission, respectively.*

which indicates that financial inclusion can reduce carbon dioxide emissions by elevating the carbon sequestration from vegetation cover.

Financing constraints (such as limited financing channels and insufficient bank credit) have been identified as one of the biggest obstacles to the development of clean energy (Baulch et al., 2018). At the corporate level, the development of financial inclusion helps mitigating the financing constraints. Through such a channel, financial inclusion provides the companies opportunities to invest more in the research and development activities for energy-saving and environmentally friendly production technologies. Inclusive finance also makes clean energy more affordable for the consumers, which could arouse the motivation of households to choose more sustainable energy sources other than fossil fuels (IPA, 2017; Le et al., 2020). Substantial progress has been made in this regard in China's land degraded areas, especially desert areas (Zhang et al., 2021). As an important terrestrial ecosystem, forests play a significant role in absorbing carbon emission and mitigating climate change. The development of financial inclusion can help increase vegetation coverage and improving carbon sequestration capacity.

### Industrial Structure

In China, the concentration of carbon intensity industries is high. Compared to agriculture and service sector, the industrial production sector contributes much larger proportion of total carbon emission. In the wave of financial liberalization, digital financial inclusion, by providing diversified and professional wealth management, has effectively supplemented the insufficient and traditional financial system. It also improves the allocation efficiency of capital and labor. That's because inclusive finance can create new investment opportunities and jobs by loosing the financial constraints of startups and small firms.

We use the variable of industrial structure as a mediator for another mediation effect model estimation. The model (4)–(5) is estimated, and the regression results are reported in **Table 7**. From column (1) of **Table 7**, we find that financial inclusion reduces the proportion of industrial output, which is positively correlated with carbon emissions. And column (2) shows that the financial inclusion impacts the level of emission through the channel of industrial structure variation. The literature documents that changes in the industrial structure can be driven by changes in consumer demand (Kongsamut et al., 2001) or the unbalanced development of technological progress in different sectors on the supply side (Acemoglu and Guerrieri, 2008). As a matter of fact, financial inclusion is able to affect both demand and supply side of the economy. That's the reason why industrial structure can act as an effective mediator. Our results consistent with Mao. (2013), who argues that China's development of a low-carbon economy relies heavily on financial technology innovation and industrial structure optimization.

## Heterogeneity

### Geographic Heterogeneity

Significant heterogeneity exists in Chinese economy. According to the level of economic development in different regions, we divide the sample data into three groups, including eastern, central, and western region. **Table 8** reports the impact of financial inclusion on carbon emissions in these three regions. Columns (1) to (3) contain coefficients estimates for the three regions respectively. The coefficients of financial inclusion in all columns imply that financial inclusion imposes significant inhibitory effect on carbon emissions across three regions. By comparing the absolute value of coefficients, we find that the carbon reduction effect of financial inclusion is strongest in the eastern region and weakest in the western region. The possible explanation is that the eastern region possesses the most

**TABLE 8 |** Regional effects of financial inclusion on carbon emissions.

| Variables | (1) | (2) | (3) |
|---|---|---|---|
| | Eastern region | Central region | Western region |
| | Emission | Emission | Emission |
| Fin-inclusion | –0.2783*** | –0.2465*** | –0.2201*** |
| | (0.285) | (0.195) | (0.413) |
| Output | 0.0078 | 0.0311 | 0.0032 |
| | (0.013) | (0.021) | (0.011) |
| Population | 0.0785* | –0.2691 | –0.2648 |
| | (0.047) | (0.176) | (0.202) |
| Area | 27.8657 | 2.5721 | 3.5261 |
| | (29.373) | (2.309) | (2.658) |
| Invest | 0.0069*** | –0.0027 | 0.0057 |
| | (0.001) | (0.002) | (0.007) |
| Constant | 3.5385*** | 3.5099*** | 2.8348*** |
| | (0.548) | (0.707) | (0.697) |
| Observations | 1,423 | 2,157 | 1,525 |
| R-squared | 0.113 | 0.151 | 0.095 |
| No. of Counties | 521 | 755 | 568 |

*Note: *, **, *** indicate significant at the levels of 10, 5, and 1%, and the t-values are in parentheses. The dependent variables for regressions (1)–(3) are carbon emission in eastern, central and western region of China, respectively.*

**TABLE 9 |** Carbon reduction effect of sub-indicators on carbon emissions.

| Variables | (1) | (2) | (3) |
|---|---|---|---|
| | Emission | Emission | Emission |
| Coverage_breadth | −0.2146*** | — | — |
| | (0.013) | — | — |
| Usage_depth | — | −0.2724*** | — |
| | — | (0.015) | — |
| Digitization_level | — | — | −0.1562*** |
| | — | — | (0.013) |
| Output | 0.0172* | 0.0171* | 0.0192** |
| | (0.009) | (0.009) | (0.008) |
| Population | −0.0100 | −0.0108 | −0.0342 |
| | (0.050) | (0.052) | (0.061) |
| Area | 3.2009* | 2.6785 | 3.8145** |
| | (1.816) | (1.818) | (1.709) |
| Invest | 0.0029** | 0.0040*** | 0.0023 |
| | (0.001) | (0.001) | (0.001) |
| Constant | 2.8281*** | 2.8820*** | 2.8163*** |
| | (0.228) | (0.236) | (0.260) |
| Observations | 5,105 | 5,105 | 5,105 |
| R-squared | 0.102 | 0.108 | 0.059 |
| No. of Counties | 1,844 | 1,844 | 1,844 |

*Note: *, **, *** indicate significant at the levels of 10, 5, and 1%, and the t-values are in parentheses. The independent variables for regressions (1)–(3) are three sub-indicators of financial inclusion index.*

developed financial system, which provides ideal infrastructure for the implementation of financial inclusion.

### Decomposition of Financial Inclusion Index

Financial inclusion index is a composition of three sub-indicators: breadth of coverage, depth of use, and degree of digitization. The breadth of coverage reflects the coverage of inclusive finance services. The depth of use reflects the penetration of inclusive finance services in the financial field. And the degree of digitization reflects the degree that inclusive finance services are digitized. This study uses sub-indicators of the financial inclusion index to further examine the impact of various dimensions of financial inclusion on carbon emissions. **Table 9** reports the regression results. Columns (1) to (3) investigate how the coverage, depth of use, and degree of digitization of financial inclusion would change regional carbon emission. The results show highly significant carbon reduction effect of all three sub-indicators, suggesting improvement in any of these three aspects could help reducing regional carbon emission.

## CONCLUSIONS AND POLICY IMPLICATIONS

This paper uses balanced panel data from 1,844 counties in China from 2014 to 2016 to investigate the relationship between the development of digital financial inclusion and carbon emissions. The study suggests that the development of digital financial inclusion can help reduce carbon dioxide emissions. The analysis of the mediation effect shows that digital financial inclusion mainly reduces carbon emissions by increasing the amount of carbon sequestration by vegetation cover and optimizing the industrial structure.

In China, the rural communities are commonly financially constrained. A large portion of them rely on agriculture. Afforestation and restoration activities may hamper their income growth and the way of living. As Wild et al. (2021) once pointed out that it is feasible to incentivize ecosystem management or restoration activities by introducing inclusive finance. Financial inclusion could provide the rural communities alternative working opportunities and financial support. It enables rural households to address the problems caused by restoration and afforestation, and even helps them to achieve more prosperous economic outcome.

The development of financial inclusion could improve industrial structure, which leads to lower level of carbon emission. Government should pay serious attention to the implementation of inclusive finance under the strategic goal of carbon neutrality. On one hand, it is necessary to strengthen the support of application financial inclusion for low-carbon industries and encourage the continuous flow of capital to finance the green industries. On the other hand, financial inclusion could smooth the adjustment path of economic industrial structure. Financial inclusion creates jobs and investment opportunities in agriculture and service industry, which help absorbing the unemployment caused by the exit of firms with high emission intensity.

Given the conclusions drawn above, we argue that the policy authorities and local institutions should fully promote the implementation of inclusive financial policy and give priority to providing financial support for low-carbon development. Further, we also stress that the construction of inclusive financial system in western regions should be strengthened. These revised policies may further assist country to improve environmental quality and meet its climate change targets. Limited by the availability of data, this paper only studies the impact of financial inclusion on carbon emissions. The future studies may aim to focus on the analysis of inclusive finance and environmental protection investment, energy consumption and green innovation to draw more explicit policy recommendations.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

ZYa: software, validation, resources, data curation, writing original draft. LY: writing—review and editing, and visualization. YL: conceptualization, methodology, formal analysis, investigation. ZYi: data curation, software, funding acquisition. ZX: writing—review and editing, supervision, project administration, and funding acquisition. All authors: contributed to the article and approved the submitted version.

## FUNDING

We acknowledge the financial support from Shandong Provincial Natural Science Foundation (Grant Number: ZR2020QG032), Shandong Provincial Social Science Planning Office (Grant Numbers: 21DGLJ12; 21DJJJ02), Taishan Scholars Program of Shandong Province, China (Grant Numbers: ts201712059; tsqn201909135) and Youth Innovative Talent Technology Program of Shandong Province, China (Grant Number: 2019RWE004).

## REFERENCES

Acemoglu, D., and Guerrieri, V. (2008). Capital Deepening and Nonbalanced Economic Growth. *J. Polit. Economy* 116 (3), 467–498. doi:10.1086/589523

Aluko, O. A., and Obalade, A. A. (2020). Financial Development and Environmental Quality in Sub-saharan Africa: Is There a Technology Effect? *Sci. Total Environ.* 747, 141515. doi:10.1016/j.scitotenv.2020.141515

Baulch, B., Duong Do, T., and Le, T.-H. (2018). Constraints to the Uptake of Solar home Systems in Ho Chi Minh City and Some Proposals for Improvement. *Renew. Energ.* 118, 245–256. doi:10.1016/j.renene.2017.10.106

Brunnschweiler, C. N. (2010). Finance for Renewable Energy: An Empirical Analysis of Developing and Transition Economies. *Envir. Dev. Econ.* 15 (3), 241–274. doi:10.1017/S1355770X1000001X

Chen, J., Gao, M., Cheng, S., Hou, W., and Shan, Y. (2020). County-level co2 emissions and sequestration in china during 1997-2017. *Scientific Data* 7 (1). doi:10.1038/s41597-020-00736-3

Cosmas, N. C., Chitedze, I., and Mourad, K. A. (2019). An Econometric Analysis of the Macroeconomic Determinants of Carbon Dioxide Emissions in Nigeria. *Sci. Total Environ.* 675, 313–324. doi:10.1016/j.scitotenv.2019.04.188

Dasgupta, S., Laplante, B., and Mamingi, N. (2001). Pollution and Capital Markets in Developing Countries. *J. Environ. Econ. Manage.* 42 (3), 310–335. doi:10.1006/jeem.2000.1161

Dong, H., Xue, M., Xiao, Y., and Liu, Y. (2021). Do carbon Emissions Impact the Health of Residents? Considering China's Industrialization and Urbanization. *Sci. Total Environ.* 758 (10120), 143688. doi:10.1016/j.scitotenv.2020.143688

Erdoğan, S., Yıldırım, S., Yıldırım, D. Ç., and Gedikli, A. (2020). The Effects of Innovation on Sectoral Carbon Emissions: Evidence from G20 Countries. *J. Environ. Manage.* 267, 110637. doi:10.1016/j.jenvman.2020.110637

Farhani, S. (2015). Renewable Energy Consumption, Economic Growth and CO2 Emissions: Evidence from Selected MENA Countries. Department of Research, Ipag Business School. Working Papers 2015-612.

IPA (2017). "Innovation for Poverty Action," in *Climate Change and Financial Inclusion* (New Haven, Connecticut: Innovation for poverty action).

Jalil, A., and Feridun, M. (2011). The Impact of Growth, Energy and Financial Development on the Environment in China: A Cointegration Analysis. *Energ. Econ.* 33 (2), 284–291. doi:10.1016/j.eneco.2010.10.003

Khan, S., and Yahong, W. (2021). Symmetric and Asymmetric Impact of Poverty, Income Inequality, and Population on Carbon Emission in Pakistan: New Evidence from ARDL and NARDL Co-integration. *Front. Environ. Sci.* 9, 666362. doi:10.3389/fenvs.2021.666362

Kongsamut, P., Rebelo, S., and Xie, D. (2001). Beyond Balanced Growth. *Rev. Econ. Stud.* 68 (4), 869–882. doi:10.1111/1467-937X.00193

Le, T.-H., Le, H.-C., and Taghizadeh-Hesary, F. (2020). Does Financial Inclusion Impact CO2 Emissions? Evidence from Asia. *Finance Res. Lett.* 34, 101451. doi:10.1016/j.frl.2020.101451

Li, G., and Yuan, Y. (2014). Impact of Regional Development on Carbon Emission: Empirical Evidence across Countries. *Chin. Geogr. Sci.* 24 (5), 499–510. doi:10.1007/s11769-014-0710-5

Mahalik, M. K., Babu, M. S., Loganathan, N., and Shahbaz, M. (2017). Does Financial Development Intensify Energy Consumption in Saudi Arabia? *Renew. Sustain. Energ. Rev.* 75, 1022–1034. doi:10.1016/j.rser.2016.11.081

Malik, M. Y., Latif, K., Khan, Z., Butt, H. D., Hussain, M., and Nadeem, M. A. (2020). Symmetric and Asymmetric Impact of Oil price, FDI and Economic Growth on Carbon Emission in Pakistan: Evidence from ARDL and Non-linear ARDL Approach. *Sci. Total Environ.* 726, 138421. doi:10.1016/j.scitotenv.2020.138421

Mao, G., Dai, X., Wang, Y., Guo, J., Cheng, X., Fang, D., et al. (2013). Reducing Carbon Emissions in China: Industrial Structural Upgrade Based on System Dynamics. *Energ. Strategy Rev.* 2 (2), 199–204. doi:10.1016/j.esr.2013.07.004

Nasreen, S., and Anwar, S. (2015). The Impact of Economic and Financial Development on Environmental Degradation: An Empirical Assessment of EKC Hypothesis. *Stud. Econ. Finance* 32 (4), 485–502. doi:10.1108/sef-07-2013-0105

Sadorsky, P. (2010). The Impact of Financial Development on Energy Consumption in Emerging Economies. *Energy Policy* 38 (5), 2528–2535. doi:10.1016/j.enpol.2009.12.048

Saidi, K., and Mbarek, M. B. (2017). The Impact of Income, Trade, Urbanization, and Financial Development on CO2 Emissions in 19 Emerging Economies. *Environ. Sci. Pollut. Res.* 24 (14), 12748–12757. doi:10.1007/s11356-016-6303-3

Shafiei, S., and Salim, R. A. (2014). Non-renewable and Renewable Energy Consumption and CO2 Emissions in OECD Countries: A Comparative Analysis. *Energy Policy* 66, 547–556. doi:10.1016/j.enpol.2013.10.064

Tamazian, A., and Bhaskara Rao, B. (2010). Do economic, Financial and Institutional Developments Matter for Environmental Degradation? Evidence from Transitional Economies. *Energ. Econ.* 32 (1), 137–145. doi:10.1016/j.eneco.2009.04.004

Tamazian, A., Chousa, J. P., and Vadlamannati, K. C. (2009). Does Higher Economic and Financial Development lead to Environmental Degradation: Evidence from BRIC Countries. *Energy Policy* 37 (1), 246–253. doi:10.1016/j.enpol.2008.08.025

Wang, B., Yu, M., Zhu, Y., and Bao, P. (2021a). Unveiling the Driving Factors of Carbon Emissions from Industrial Resource Allocation in China: A Spatial Econometric Perspective. *Energy Policy* 158, 112557. doi:10.1016/j.enpol.2021.112557

Wang, W.-Z., Liu, L.-C., Liao, H., and Wei, Y.-M. (2021b). Impacts of Urbanization on Carbon Emissions: An Empirical Analysis from OECD Countries. *Energy Policy* 151, 112171. doi:10.1016/j.enpol.2021.112171

Wild, R., Egaru, M., Ellis-Jones, M., Bugembe, B., Mohamed, A., Ngigi, O., et al. (2021). *Using Inclusive Finance to Significantly Scale Climate Change Adaptation*. Berlin: Springer International Publishing, 2565–2590. doi:10.1007/978-3-030-45106-6_127

World Bank (2018). The Little Data Book on Financial Inclusion. Available at http://hdl.handle.net/10986/29654.

Zaidi, S. A. H., Hussain, M., and Uz Zaman, Q. (2021). Dynamic Linkages between Financial Inclusion and Carbon Emissions: Evidence from Selected OECD Countries. *Resour. Environ. Sustainability* 4, 100022. doi:10.1016/j.resenv.2021.100022

Zhang, Y.-J. (2011). The Impact of Financial Development on Carbon Emissions: An Empirical Analysis in China. *Energy Policy* 39 (4), 2197–2203. doi:10.1016/j.enpol.2011.02.026

Zhang, Y., ChenHan, J. Y., Han, Y., Qian, M., Guo, X., Chen, R., et al. (2021). The Contribution of Fintech to Sustainable Development in the Digital Age: Ant forest and Land Restoration in China. *Land Use Policy* 103, 105306. doi:10.1016/j.landusepol.2021.105306

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Yang, Yu, Liu, Yin and Xiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fenvs.2021.810305

# Does Green Credit Policy Move the Industrial Firms Toward a Greener Future? Evidence From a Quasi-Natural Experiment in China

*Zumian Xiao, Lu Yu, Yinwei Liu, Xiaoning Bu* and Zhichao Yin*

*School of Finance, Shandong University of Finance and Economics, Jinan, China*

How to utilize financial instrument to deal with environmental issues has been a focal topic. Taking the introduction of green credit program as a "quasi-natural experiment," the propensity score matching and difference-in-difference approach (PSM-DID) are used to investigate the impact of the green credit policy implemented by Chinese government on firm-level industrial pollutant emissions. The estimation results indicate that the green credit policy significantly reduces corporate sulfur dioxide emissions. Heterogeneity analysis shows this impact is more pronounced for large-scale enterprises and enterprises located in the eastern region. The estimated mediation models reveal that after the implementation of the green credit policy, reduction in sulfur dioxide emissions can be attribute to the increased environmental investment and improved energy consumption intensity. Moreover, the green credit policy is also significantly effective in mitigating the discharge of other common industrial pollutants. Our findings highlight the importance of green credit policies in achieving greener industrial production and more sustainable economic development.

**Keywords: green credit policy, $SO_2$ emissions, environmental investment, energy consumption intensity, PSM-DID**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong,*
*Hong Kong SAR, China*

***Reviewed by:***
*Shujahat Haider Hashmi,*
*Bahria University, Pakistan*
*Dervis Kirikkaleli,*
*European University of Lefka, Turkey*

****Correspondence:***
*Xiaoning Bu*
*buxiaoning@sdufe.edu.cn*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *06 November 2021*
***Accepted:*** *20 December 2021*
***Published:*** *12 January 2022*

***Citation:***
*Xiao Z, Yu L, Liu Y, Bu X and Yin Z (2022) Does Green Credit Policy Move the Industrial Firms Toward a Greener Future? Evidence From a Quasi-Natural Experiment in China. Front. Environ. Sci. 9:810305. doi: 10.3389/fenvs.2021.810305*

## 1 INTRODUCTION

As the largest developing country, China's booming economy development accounts for a large proportion of the world's energy consumption and pollutant emissions (Yan and Toshihiko, 2009). It is well known that air pollutant emissions impose a negative impact on both human health and environmental quality (Hollingsworth et al., 2021) and a strong spatial spillover effect, which leads to a wider range of harm to the society (Liu et al., 2017). Therefore, air pollution has caused major concerns (Wang L. et al., 2021). Achieving environmental sustainability is urgent issue especially for emerging and developing economies (Nawaz et al., 2020). Level of air pollutant emissions depend on many factors, including energy consumption, financial development, transportation infrastructure, etc. (Nawaz et al., 2020; Umar et al., 2020). Shen et al. (2020) noted that China's air pollution mainly comes from its large consumption of fossil energy, such as coal and fuel oil. Previous studies also find that both international export and interprovincial trade exacerbate the health burdens of air pollution in China's less developed interior provinces (Wang et al., 2017). In addition, population density is also an important factor affecting air pollution in China (Chen et al., 2020). Therefore, it can be inferred that air pollution is caused by a combination of many factors.

It is the common responsibility of all countries in the world to promote the green and sustainable development of the global economy. Ma and He (2016) have found that a license plate-based traffic

control measure implemented in Beijing in 2008 has effectively improved air quality, though the long-term impacts of this policy remain controversial. Davis and Lucas (2017) have evaluated the effects of traffic control in Mexico City, note that there is little evidence supports that the program has improved air quality. On the contrary, air pollution becomes more serious on weekends. Kumar and Foster (2015) have studied the impact of New Delhi's air quality control on the distribution of air pollutants. They find that the air quality in New Delhi has been improved, while is deteriorated in the surrounding areas. The reason for this deterioration is that environmental regulations have caused the high-polluting industries in New Delhi to move to the surrounding areas. The air quality in those areas has thus become worse. These studies have explored the relationship between command-control environmental regulation and air quality, but ignored the discussion of market-oriented environmental regulation tools. Wu et al. (2017) also have found that the strict environmental policy accelerates the transfer of pollution, which may cause worse damage to the more fragile ecosystem in the western provinces. In initial studies, the focus was on using $SO_2$ emissions and energy consumption to measure environmental impact (Kihombo et al., 2021), relevant research has been gradually enriched in recent years.

The green credit policy is an important part of China's national environmental governance system. It supplements the traditional environmental regulation toolbox (Zhang et al., 2021). This policy was formally implemented in 2007. The literature has studied the impacts of this policy on economic development (Wang Y. et al., 2021) and corporate investment (Liu et al., 2015). And, several studies have found that the green credit policy imposes positive impact on environmental quality (Sun et al., 2019). Nevertheless, prior researches are insufficient in investigating the mechanisms that how the green credit policy works. In fact, the fossil fuel energy consumption and green investment have significant impact on air quality (Rafindadi et al., 2014). Our study attempts to fill this research gap and make an in-depth analysis on the mechanisms that how green credit policy affect pollutant emissions. Our findings are designated to provide reference for developing countries to reduce industrial pollutant emissions.

We chose sulfur dioxide ($SO_2$) emissions as our explained variable for the following reasons: first, sulfur dioxide is a typical air pollutant, and regulation for $SO_2$ emission has become a key policy agenda in China (Kim and Lee, 2015; Chen and Yan, 2020). Second, $SO_2$ imposes a negative impact on human health and environmental quality (Yan and Wu, 2017). Third, compared with other air pollutants, the dataset of $SO_2$ emission contains much less missing values. $SO_2$ emission has been chosen as explained variable to ensure robust empirical results. In addition to our main empirical analysis, the effects of green credit policy on other industrial pollutants are also examined.

Our study is motivated by the implementation of green credit policy in China. We aim to explore whether the green credit policy imposes $SO_2$ emission reduction effect on industrial firms. To identify the relationship between the implementation of green credit policy and level of $SO_2$ emissions, we apply a PSM-DID approach to a dataset contains 121,163 Chinese industrial firms during the period of 1999–2013. We have found that the green credit policy can significantly reduce $SO_2$ emissions in high-polluting industries, implying this policy is highly effective in improving air quality. The results also suggest that green credit policy affects $SO_2$ by curbing energy consumption, such as reducing the use of fuel and coal. At the same time, it also enhances the motivation of enterprises to take environmental responsibility and promotes the upgrading of production technology.

The contribution of this research lies in the following three aspects: first, we construct a unique dataset by matching survey data of polluting firms with the annual survey of industrial firms. The dataset includes more than 120,000 industrial enterprises in China. Therefore, it is possible to excavate deeper into the relationship between the green credit policy and $SO_2$ emissions. More rigorous conclusions can be drawn from our empirical analysis. Second, green credit policy is market-oriented, which is different from traditional environmental regulations. Our study enriches the literature on the evaluation of environmental protection policies. Finally, based on China's institutional background, we thoroughly analyze the two channels through which the green credit policy takes effect. Our mechanism analysis provides better understanding of the channels through which green credit policy change the level of firm $SO_2$ emissions.

The remainder of this paper is organized as follows. **Section 2** introduces background and develops our hypotheses. **Section 3** discusses our research design and provides the descriptive sample statistics. **Section 4** reports the empirical results. Finally, **Section 5** concludes our paper.

## 2 LITERATURE REVIEW AND HYPOTHESES

### 2.1 Green Credit Policy and $SO_2$ Emission

In 2007, the State Environmental Protection Administration and the China Banking Regulatory Commission adopted green credit to curb the expansion of high-pollution and energy-intensive industries and promulgated the "Opinions on Implementing Environmental Protection Policies and Regulations to Prevent Credit Risks." Green credit policy integrates economic decisions with environmental decisions to arrive at optimally beneficial outcomes (Wang and Zhi, 2016). Wang Y. et al. (2021) note that the implementation of green credit policies can improve economic performance and reduce pollutant emission, which is an effective way to achieve sustainable economic development. The green credit policy incorporates environmental risks into the credit management framework, and more credit funds are invested in low-carbon green industries (Cao et al., 2021).

Green credit is an important supplement to the traditional administrative environmental regulations. It affects sulfur dioxide emissions in two ways. On one hand, the green credit policy has a punitive effect on enterprises sulfur dioxide emissions. Enterprises with high energy consumption intensity and level of pollution usually face tight financing constraints. The green credit policy curbs investment in industries that consume large amount of energy (Liu et al., 2015). These industries are also

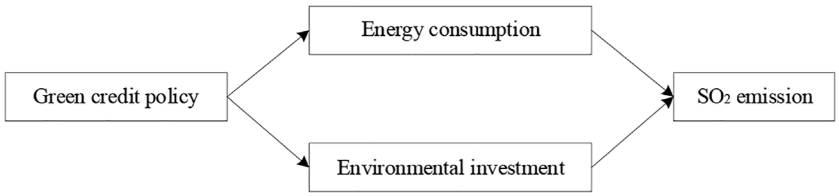



**FIGURE 1 |** Mechanisms that how green credit policy affect $SO_2$ emission.

**TABLE 1 |** Variable description and data sources.

| Variable name | Variable description | Data source |
|---|---|---|
| $SO_2$ | Natural logarithm of sulfur dioxide emission in million tons | Annual Environmental Survey of Polluting Firms |
| Lev | Total liabilities divided by total assets (ten thousand yuan) | National Bureau of Statistics of China |
| Age | Natural logarithm of a company's age | National Bureau of Statistics of China |
| Roa | Net income divided total assets | National Bureau of Statistics of China |
| Far | Fixed asset investment (billion yuan) | National Bureau of Statistics of China |
| Admin | Management expenses divided by gross revenue | National Bureau of Statistics of China |
| State | Takes the value of 1 if a firm is a state-owned enterprise, and 0 otherwise | National Bureau of Statistics of China |

under strict environmental regulations. On the other hand, the green credit policy creates incentives for enterprises to realize energy conservation and emission reduction. Sun et al. (2019) note that to obtain credit, companies are motivated to adopt green and low-carbon production techonologies. Our main research hypothesis is as follows:

**H1:** Green credit policies can significantly reduce the $SO_2$ emissions of enterprises.

## 2.2 The Mediation Effect of Energy Consumption Intensity

In recent years, with its rapid economic development, China's energy consumption has been rising quickly. As a result, air pollution has become a serious problem (Shen et al., 2020). Jiang et al. (2019) note that coal use, population, area, and thermal power can significantly increase national air pollution. Among these, coal consumption has had the biggest impact on national $SO_2$ concentration. It is possible that replacing traditional energy with renewable energy is an effective measure for controlling environmental pollution (Su et al., 2021a). Moreover, the transition to renewable energy is beneficial to the energy supply and reduce geopolitical risks (Su et al., 2021b). The green credit policy mainly adjusts the energy consumption structure of enterprises through resource credit rationing. If companies do not restrict the consumption of non-renewable energy, such as coal and fuel oil, they will face higher financing constraints. After the implementation of the green credit policy, companies are more inclined to reduce fossil energy consumption. Therefore, the green credit policy will impose strict credit constraints on high-polluting companies, forcing them to reduce energy consumption and investment and convert to energy-saving and environmentally friendly practices. Based on the discussion above, we hypothesize the following:

**H2:** Energy consumption intensity mediates the reduction effect of green credit policy on $SO_2$ emissions.

## 2.3 The Mediation Effect of Corporate Environmental Investment

Studies have documented that environmental policies affect corporate behavior (Albrizio et al., 2017; Ji et al., 2017). Green credit policy influence corporate environmental investment through the channels of credit supply and financial constraint. Green credit policy puts enterprises under greater financial pressure regarding environmental governance. Polluting companies will face tight financing constraints, which can only be eased for the companies to increase green investment. The government subsidies and credit supply create incentives for companies' environmental investment. As a result, technological innovation can be carried out, which will impose a positive impact on environment (Ran et al., 2021). A large number of literatures emphasize the important role of technological innovation in improving the environment (such as Samargandi, 2017; Yu and Du, 2019). However, it is ignored that the environmental investment is necessary condition for technological upgrades. Although corporate environmental protection investment has increased costs in the short-term, this environmental governance behavior can generate greater social and environmental benefits and has won the support of stakeholders including capital market investors. Managers anticipate investors' reaction and thus overwhelmingly disclose their investment and tend to focus their disclosure on the societal benefits of the investment rather than on the cost to the company (Martin and Moser, 2016). Sun et al. (2019) note that the green credit policy is an important tool to drive companies to assume environmental responsibility. Therefore, we have developed the third hypothesis as follows:

**TABLE 2 |** Descriptive statistics.

| Variable | N | Mean | Sd | Min | p50 | Max |
|---|---|---|---|---|---|---|
| $SO_2$ | 499,070 | 2.923 | 2.261 | −6.908 | 2.975 | 14.59 |
| Lev | 637,091 | 0.606 | 0.294 | 0.023 | 0.610 | 1.551 |
| Age | 402,529 | 2.475 | 0.957 | 0.000 | 2.485 | 7.606 |
| Roa | 637,062 | 0.078 | 0.177 | −0.225 | 0.026 | 0.942 |
| Far | 632,987 | 0.396 | 0.218 | 0.015 | 0.375 | 0.914 |
| Admin | 636,655 | 0.074 | 0.082 | 0.002 | 0.050 | 0.529 |
| State | 694,835 | 0.181 | 0.385 | 0.000 | 0.000 | 1.000 |

**H3:** Corporate environmental investment mediates the reduction effect of green credit policy on $SO_2$ emissions.

Based on the above discussion, we hypothesize that green credit policy affects $SO_2$ emission through energy consumption intensity and environmental investment. The specific impact paths are depicted in **Figure 1**.

# 3 METHODOLOGY AND DATA

## 3.1 Model Specification

### 3.1.1 Propensity Score Matching

We choose the one-to-one nearest neighbor approach to match the sample firms according to the propensity score value. Moreover, we selected the financial leverage, company age, return on assets, fixed assets, management expense as the matching characteristic variables. A logit regression has been constructed to perform the propensity score matching as follows:

$$Logit\left(\mathrm{du}_{it}=\lambda\right)=\alpha+\beta\mathbf{X}_{it}+\varepsilon_{it} \quad (1)$$

In model (1), $\mathrm{du}_{it}$ is the dummy variable of the planning. When $\lambda = 1$, the firm is located in the city with the green credit policy implemented, and 0 otherwise. The covariates $\mathbf{X}_{it}$ were the characteristic variables mentioned above. Individuals in the same value range were matched after using the one-to-one nearest neighbor approach.

### 3.1.2 Difference-in-Difference Estimation

To investigate the effects of the green credit policy on $SO_2$ emission, we estimate the following Difference-in-Difference model:

$$SO_2_emission_{it}=\alpha+\beta Policy\times Time+\gamma\mathbf{X}_{it}+\mu_t+\eta_i+\varepsilon_{it} \quad (2)$$

The dependent variable is $SO_2_emission_{it}$, the subscript i and t denote firm and year, respectively. Policy is an indicator variable equal to 1 if the firm is located in the city implementing the green credit policy, and 0 otherwise. In 2007, the State Environmental Protection Administration and the China Banking Regulatory Commission used green credit to curb the expansion of high-pollution and energy-intensive industries and promulgated the "Opinions on Implementing Environmental Protection Policies and Regulations to Prevent Credit Risks."[1] The green credit policy mainly targets six high-pollution and high-energy-consuming industries,[2] and has no or minimal impact on clean industries. Therefore, we divide the samples into experimental groups (heavily polluting industries) and control groups (non-heavily polluting industries). Time is a dummy variable equal to 1 after 2007, the time of implementation of the green credit policy, and equal to 0 otherwise. Before we conduct DID estimation according to model (1), a one-on-one propensity score matching procedure is performed to disentangle the treatment effect and selection effect of the green credit policy on $SO_2$ emissions based on observable characteristics. The coefficient $\beta$ captures the effect of the green credit policy on $SO_2$ emissions with the implement of the green credit policy. The variable $X_{it}$ is a vector of control variables, including corporate financial leverage (Lev), return on assets (Roa), the proportion of fixed assets in total assets (Far), the proportion of management expenses to operating income (Admin), and enterprise nature (State). The variable $\mu_t$ and $\eta_i$ represent time and individual fixed effect. $\varepsilon_{it}$ is the error term.

## 3.2 Variable Definition

### 3.2.1 Dependent Variable

The dependent variable in this research is sulfur dioxide emissions, which are calculated by the natural logarithm of sulfur dioxide emissions from industrial enterprises. The data comes from the Annual Environmental Survey of Polluting Firms (AESPF) of China.

### 3.2.2 Independent Variable

The independent variable is the green credit policy, which is an indicator to reveal whether the firm's industry has implemented the green credit policy. Policy = 1 if the firm's industry has implemented the green credit policy, and Policy = 0 otherwise.

### 3.2.3 Control Variables

According to Xing et al. (2020), the control variables in this paper include corporate financial leverage, return on assets, corporate age, the proportion of fixed assets in sales revenue, the proportion of management expenses, and the company's properties. These data are obtained from the National Bureau of Statistics. **Table 1** reports the variable description and our data sources.

## 3.3 Data Sources

### 3.3.1 Firm Level Pollution Data

The data on firms' pollution emissions comes from Annual Environmental Survey of Polluting Firms (AESPF).[3] Established by the Ministry of Ecology and Environment (formerly known as the Ministry of Environmental Protection) in the 1980s in a bid to document the state of environmental pollution and abatement in China, AESPF provides information on firms' environmental performance, including the emission of main pollutants, pollution treatment equipment and information of energy consumptions.

[1]Source: http://jjckb.xinhuanet.com/wzpd/2007-07/30/content_60171.htm.

[2]The six major industries include: petrochemical, chemical, building materials, steel, non-ferrous and power industries.

[3]Lacking of an official English name, this dataset was also called as China Environmental Statistics dataset (CESD) (Zhang et al., 2018) or Environmental Statistics Data (ESD) (Wu et al., 2017).

**TABLE 3 |** Baseline results: the green credit policy and $SO_2$ emissions.

| Variables | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | $SO_2$ | $SO_2$ | $SO_2$ | $SO_2$ | $SO_2$ | $SO_2$ |
| Policy×Time | −0.1843*** | −0.1921*** | −0.1955*** | −0.1914*** | −0.2074*** | −0.2055*** |
| | (−16.16) | (−15.78) | (−16.06) | (−15.71) | (−17.05) | (−17.02) |
| Lev | | 0.2585*** | 0.1881*** | 0.2657*** | 0.2809*** | 0.2014*** |
| | | (25.68) | (17.92) | (24.96) | (26.40) | (18.92) |
| Roa | | | −0.4367*** | −0.4398*** | −0.2695*** | −0.1368*** |
| | | | (−23.64) | (−23.73) | (−14.23) | (-7.24) |
| Far | | | | 0.8297*** | 0.8035*** | 0.7074*** |
| | | | | (57.76) | (55.96) | (49.68) |
| Admin | | | | | −1.7447*** | −1.8223*** |
| | | | | | (−46.31) | (-48.31) |
| State | | | | | | 0.7352*** |
| | | | | | | (81.18) |
| Constant | 2.4138*** | 2.2230*** | 2.2633*** | 1.8193*** | 2.0076*** | 1.5910*** |
| | (10.75) | (9.92) | (10.07) | (8.07) | (8.95) | (7.39) |
| Year | YES | YES | YES | YES | YES | YES |
| Ind | YES | YES | YES | YES | YES | YES |
| Regional | YES | YES | YES | YES | YES | YES |
| Observations | 499,070 | 459,742 | 459,490 | 456,558 | 455,902 | 455,902 |
| R-squared | 0.261 | 0.264 | 0.265 | 0.271 | 0.271 | 0.284 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

### 3.3.2 Other Firm-Level Data

We match AESPF dataset with the Annual Survey of Industrial Firms (ASIF), which is one of the most comprehensive and widely used Chinese firm-level dataset maintained by the National Bureau of Statistics of China (NBSC). ASIF dataset surveys industrial firms with annual sales above 5 million RMB. It contains detailed financial and characteristical information. After eliminating outliers and excluding records which violate the accounting standards, we obtain a panel of 455,902 observations for 121,163 unique firms.[4]

## 3.4 Descriptive Statistics

The descriptive statistics of variables are reported in **Table 2**. The mean value of the $SO_2$ emissions is 2.923, and the standard deviation is 2.261, indicating that there are significant differences among firms.

# 4 EMPIRICAL RESULTS

## 4.1 Main Results

**Table 3** reports the results of the PSM-DID regression of the main analysis. Column (1) shows that the coefficients on the interaction between *Policy* and *Time* (*Policy*Time*) are significantly negative at a 1% level without adding any control variables. After gradually adding variables that may affect $SO_2$ emissions, and controlling time, industry, and regional effects, estimation results showthat the coefficients of *Policy*Time* are significantly negative in all settings. The implementation of the green credit policy reduced $SO_2$ emissions from industrial firms in China. Firms in areas where the green credit policy has been implemented show better environmental performance. Our results consistent withUmar et al. (2020). They argue that financial development can reduce carbon dioxide emissions. It is noted that financial resources should be allocated to environmentally-friendly sectors of the economy. Green technologies are long-term projects that require financing, especially the support from green finance.

## 4.2 Test for Assumption of Parallel Trend

The necessary precondition for the validity of the PSM-DID model is that the treatment group and the control group are assumed to have a parallel trend before treatment effect takes place. Instead of one discrete policy dummy, we instantiate a series of pre- and post-treatment dummies. The model is set as follows:

$$SO_2_emission_{it} = \mu_t + \eta_i + \delta_{i,t-3}\mathrm{d}_{t-3} + \delta_{i,t-2}\mathrm{d}_{t-2} + \delta_{i,t-1}\mathrm{d}_{t-1} + \delta_{i,t}\mathrm{d}_t + \delta_{i,t+1}\mathrm{d}_{t+1} + \delta_{i,t+2}\mathrm{d}_{t+2} + \delta_{i,t+3}\mathrm{d}_{t+3} + \boldsymbol{\gamma}\mathbf{X}_{it} + \varepsilon_{it} \quad (3)$$

where $d_t$ is the immediate effect of the exposure for all firms to the treatment. To be clear, the immediate or instantaneous effect is equal to 1 for a treated firm in the initial adoption year (i.e., the year of change). The control firms, on the other hand, remain consistently equal to 0 in all periods. The "lags" investigate how effects have evolved since the initial adoption year (e.g., $d_{t-1}$, $d_{t-2}$, and $d_{t-3}$). It should be noted that each policy dummy is the product of a treatment indicator with a series of pre- and post-exposure year dummies. Again, the interaction is implicit in the coding of each policy variable; thus, no explicit multiplicative term is required in the model formula. The constituent elements of the interaction term are not required either as the relevant information is already captured by the fixed effects.

[4]We exclude firms with missing values in $SO_2$ emissions.

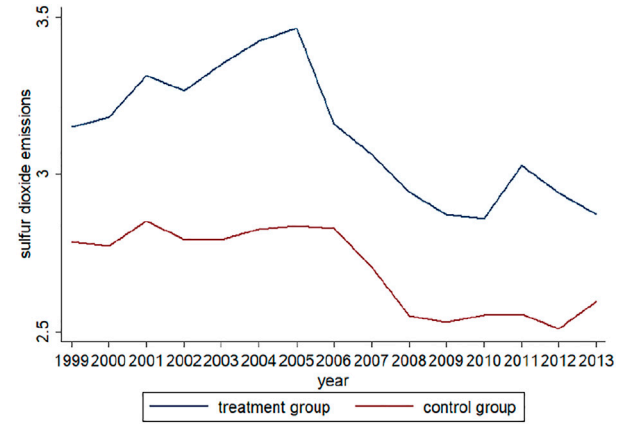



**FIGURE 2 |** Trends of sulfur dioxide emissions during 1999–2013 (Million tons).

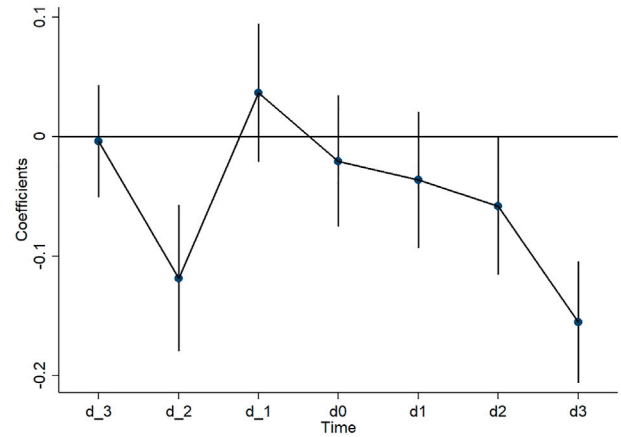



**FIGURE 3 |** Test results for the assumption of parallel trend. Notes: d0 represents the year in which green credit policy implemented. d_3, d_2, d_1 represent three, two and 1 year before the policy is implemented, respectively. d3, d2, and d1 represent one, two and 3 year after the policy is implemented, respectively.

The trends of sulfur dioxide emissions for the treatment group and the control group are shown in **Figure 2**. Indeed, it is not feasible to identify a common trend before the implementation of the green credit policy from the visual representations of the data. We estimate model (3) to test against the hypothesis of existence of parallel trend. The estimation results are presented in **Figure 3**.

As shown in **Figure 3**, the dot represents the value of the estimated coefficient, while the vertical line represents the confidence interval. It is documented that the curves of the treatment group and the control group of the full sample were almost parallel before policy implementation. However, the $SO_2$ emissions of the treatment group were lower than that of the control group, and the gap between the two trends increased noticeably. This result proves that the PSM-DID model is an appropriate method to test the impact of the green credit policy on $SO_2$ emissions.

**TABLE 4 |** The green credit policy and $SO_2$ emissions: regional heterogeneity.

| Variables | (1) | (2) | (3) |
|---|---|---|---|
| | Eastern region | Central region | Western region |
| | $SO_2$ | $SO_2$ | $SO_2$ |
| Policy×Time | −0.2959*** | −0.0924*** | −0.1493*** |
| | (−18.75) | (−4.06) | (−4.40) |
| Lev | 0.2545*** | 0.1322*** | 0.1739*** |
| | (17.42) | (7.05) | (6.51) |
| Roa | 0.0301 | −0.3729*** | −0.1894*** |
| | (1.17) | (−11.79) | (−3.47) |
| Far | 0.7485*** | 0.6373*** | 0.5225*** |
| | (38.82) | (24.88) | (14.31) |
| Admin | −2.1573*** | −1.3151*** | −1.8691*** |
| | (−42.60) | (−19.76) | (−18.31) |
| State | 0.6859*** | 0.7967*** | 0.8101*** |
| | (51.20) | (52.59) | (37.96) |
| Constant | 0.8027*** | 1.4129*** | 2.5326*** |
| | (11.65) | (6.46) | (34.85) |
| Year | YES | YES | YES |
| Ind | YES | YES | YES |
| Regional | YES | YES | YES |
| Observations | 246,198 | 130,465 | 79,239 |
| R-squared | 0.297 | 0.283 | 0.265 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

**TABLE 5 |** The green credit policy and $SO_2$ emissions: corporate size.

| Variables | (1) | (2) |
|---|---|---|
| | SEMs | Large enterprises |
| | $SO_2$ | $SO_2$ |
| Policy×Time | −0.0821*** | −0.3586*** |
| | (−5.01) | (−16.55) |
| Lev | 0.2066*** | 0.3149*** |
| | (16.56) | (16.94) |
| Roa | 0.0148 | −0.2672*** |
| | (0.56) | (−10.16) |
| Far | 0.6185*** | 0.7605*** |
| | (35.97) | (32.34) |
| Admin | −1.4543*** | −1.7084*** |
| | (−30.78) | (−28.97) |
| State | 0.3893*** | 0.8794*** |
| | (38.46) | (50.77) |
| Constant | 1.4996*** | 1.5388*** |
| | (6.56) | (3.24) |
| Year | YES | YES |
| Ind | YES | YES |
| Regional | YES | YES |
| Observations | 276,921 | 178,981 |
| R-squared | 0.298 | 0.342 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

## 4.3 Heterogeneity

### 4.3.1 Regional Heterogeneity

In China, there are obvious differences in the natural environment, social development, and economic level in different regions. Therefore, this research further analyzes

**TABLE 6 |** Mediation effect estimation results: energy consumption as mediator.

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Econsumption | Econsumption | $SO_2$_emission | $SO_2$_emission |
| Policy×Time | −0.0154*** | −0.0141*** | −0.0825*** | −0.0554*** |
| | (−4.93) | (−3.60) | (−5.90) | (−3.49) |
| Econsumption | | | 0.8493*** | 0.7660*** |
| | | | (4.55) | (4.27) |
| Lev | | 0.0290*** | | 0.0864*** |
| | | (10.30) | | (6.18) |
| Roa | | −0.0396*** | | −0.0720*** |
| | | (−8.63) | | (−2.75) |
| Far | | 0.1208*** | | 0.5803*** |
| | | (31.68) | | (21.15) |
| Admin | | −0.0240** | | −1.6346*** |
| | | (−2.57) | | (-35.18) |
| State | | −0.0042* | | 0.7971*** |
| | | (−1.89) | | (69.28) |
| Constant | 0.8813*** | 0.8133*** | 1.5136*** | 0.9854*** |
| | (42.68) | (38.51) | (5.58) | (3.91) |
| Year | YES | YES | YES | YES |
| Ind | YES | YES | YES | YES |
| Regional | YES | YES | YES | YES |
| Observations | 268,141 | 235,411 | 266,675 | 234,027 |
| R-squared | 0.130 | 0.129 | 0.294 | 0.323 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

**TABLE 7 |** Mediation effect estimation results: environmental investment as mediator.

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Ginvest1 | Ginvest2 | $SO_2$_emission | $SO_2$_emission |
| Policy×Time | 0.0305*** | 0.0413*** | −0.1740*** | −0.1154*** |
| | (12.90) | (7.39) | (−10.09) | (−5.88) |
| Ginvest1 | | | −0.0698*** | |
| | | | (-4.57) | |
| Ginvest2 | | | | −0.0171* |
| | | | | (−1.80) |
| Lev | −0.0308*** | −0.0170*** | 0.1793*** | 0.1033*** |
| | (−13.62) | (−4.05) | (12.46) | (7.16) |
| Roa | 0.0053 | −0.0709*** | −0.0585** | 0.0832*** |
| | (1.45) | (−8.38) | (−2.07) | (2.62) |
| Far | 0.1070*** | 0.2636*** | 0.7554*** | 0.7606*** |
| | (35.66) | (44.62) | (38.89) | (38.49) |
| Admin | −0.2337*** | −0.5974*** | −1.9917*** | −1.9566*** |
| | (−31.46) | (−38.86) | (−36.08) | (−34.39) |
| State | 0.1497*** | 0.2724*** | 0.7375*** | 0.7090*** |
| | (70.00) | (77.23) | (65.12) | (63.22) |
| Constant | 0.4025*** | 0.3783*** | 1.9452*** | 2.2251*** |
| | (9.52) | (4.66) | (7.00) | (8.81) |
| Year | YES | YES | YES | YES |
| Ind | YES | YES | YES | YES |
| Regional | YES | YES | YES | YES |
| Observations | 493,698 | 373,401 | 256,046 | 224,405 |
| R-squared | 0.175 | 0.213 | 0.297 | 0.296 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

and compares the regional differences in the impact of green credit policies on $SO_2$ emissions. **Table 4** reports the impact of green credit policies on $SO_2$ emissions in these three regions. Columns (1) to (3) contain coefficients estimates for the three regions respectively. The coefficients of $SO_2$ emissions in all columns imply that the green credit policy imposes significant inhibitory effects on $SO_2$ emissions across the three regions. By comparing the absolute value of the coefficients, we find that the $SO_2$ emission reduction effect the green credit policy is strongest in the eastern region and weakest in the central region. One possible reason for this difference is that the financial infrastructure in the eastern region is ideal, which has created a strong foundation for the implementation of the green credit policy. However, the economic growth pressure in the central region is high, so green credit has less effect in restraining the flow of funds to high-polluting enterprises. Thus, the effect of green credit on reducing industrial environmental pollution is small.

### 4.3.2 Corporate Size Heterogeneity

Companies of different sizes have large differences in operating efficiency, internal control, and financing capabilities, which in turn affects the impact of green credit policies on $SO_2$ emissions. This study divides the full sample into two sub-sample groups of small and medium-sized enterprises and large enterprises. **Table 5** reports the regression results. The results show that the green credit policy has a significant inhibitory effect on sulfur dioxide emissions in the two sub-sample groups, and the impact on large enterprises is more significant. This difference may be attributed to the fact that the business strategies and management of large enterprises are often more susceptible to the overall interests of the government, and they can actively cooperate with the implementation of government policies and better meet the environmental requirements than smaller

**TABLE 8 |** The green credit policy and other pollutant emission.

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | smoke_emission | cod_discharge | NH3_N_discharge | dust_discharge |
| Policy×Time | −0.0898*** | −0.0974*** | −0.3397*** | −0.3677*** |
| | (−7.32) | (−7.01) | (−16.21) | (−6.77) |
| Lev | 0.2123*** | 0.0826*** | 0.1793*** | 0.2034*** |
| | (18.96) | (6.41) | (9.50) | (7.51) |
| Roa | −0.0573*** | −0.0062 | −0.0692** | −0.0685 |
| | (−2.94) | (−0.26) | (−2.18) | (-1.14) |
| Far | 0.6768*** | 0.8101*** | 0.9645*** | 0.6690*** |
| | (45.48) | (46.50) | (38.39) | (17.57) |
| Admin | −1.6376*** | −1.7052*** | −1.9017*** | −1.2878*** |
| | (−42.03) | (−37.98) | (−30.27) | (−11.69) |
| State | 0.6188*** | 0.9253*** | 1.0445*** | 0.3714*** |
| | (67.81) | (88.50) | (59.59) | (16.65) |
| Constant | 2.0833*** | 2.1599*** | −1.0231** | 5.3445*** |
| | (7.97) | (9.06) | (−2.14) | (11.02) |
| Year | YES | YES | YES | YES |
| Ind | YES | YES | YES | YES |
| Regional | YES | YES | YES | YES |
| Observations | 441,615 | 413,016 | 204,099 | 84,697 |
| R-squared | 0.286 | 0.242 | 0.228 | 0.445 |

*Note: *, **, and *** indicate significance at 10, 5 and 1% level, respectively.*

enterprises. This corroborates with the work of Woo et al. (2014) who argued that large firms implement environmental activities more than small ones.

## 4.4 Mechanisms

The benchmark regression shows that the green credit policy leads to lower $SO_2$ emissions. In this section, we further explore the possible mechanism behind this relationship. According to the above theoretical analysis, enterprises mainly achieve reductions in $SO_2$ emissions by reducing energy consumption intensity and increasing environmental investment.

According to Fan et al. (2021), the first stage is to verify the impact of the green credit policy on the two major effects. A mediation model is constructed to test the effects of policy variables on mediating variables, see model (4). $Econsumption_{it}$ and $Ginvest_{it}$ represent energy consumption intensity and environmental investment respectively. If $\beta_1$ is significant, it means the policy impacts the mediating variables, and the second stage will be entered. First stage regression performed as follows:

$$Econsumption_{it}\,(Ginvest_{it}) = \alpha + \beta_1 Policy \times Time + \gamma \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \tag{4}$$

The second stage is to verify the effect of two major effect of the green credit policy on $SO_2$ emissions by setting up a comprehensive model (5) based on the mediating model (4). It takes a certain amount of time for environmental protection investment to turn into real capital, which contributes to firm emission reduction. So, the mediator Ginvest enters the regression with one period lag. If $\beta_3$ is significant, there exists mediating effect, regardless of whether $\beta_2$ is significant. If $\beta_2$ is insignificant, it indicates that the mediating variable is the only transmission path of the policy effect to $SO_2$ emissions. Otherwise, other transmission paths exist. Second stage regression performed as follows:

$$SO_2_emission_{it} = \alpha + \beta_2 Policy \times Time + \beta_3 Econsumption_{it}\,(Ginvest_{i,t-1}) + \gamma \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \tag{5}$$

### 4.4.1 Energy Consumption Intensity

Using Korean data, Park and Lee (2011) find that energy consumption seems to be the most significant variable in explaining air pollution. Indeed, industrialized production and corresponding energy consumption in China are also the main sources of sulfur dioxide emission (Jiang et al., 2019). We use coal consumption intensity (total coal consumption/total industrial output value) as proxy indicators of energy consumption intensity to test the intermediary effect of energy consumption. **Table 6** reports the regression results. Columns (1) and (2) show that the green credit policy can significantly decrease $SO_2$ emissions whether control variables are added or not. In columns (3) and (4), The coefficient of Policy×Time is significantly negative while the coefficient of $Econsumption_{it}$ is significantly positive, suggesting that the green credit policy has significantly reduced the consumption of coal, thereby achieving the effect of reducing sulfur dioxide emissions. Our results are

consistent with Wang K.-H. et al. (2021), who believes that reduce oil dependence and renewable energy development are very important to alleviate environmental pressures in China. In addition, Shen et al. (2020) also point out that reducing the proportion of traditional fossil energy consumption and increasing investment in renewable energy are important for air quality improvement.

#### 4.4.2 Corporate Environmental Investment

Research by Yu and Du (2019) showed that technical equipment upgrade is closely related to the reduction of carbon emissions. Similarly, Samargandi (2017) pointed technological innovations in production actives has made more significant contributions to the reduction of pollutant emissions. We use log difference of the number of wastewater treatment facilities and the number of waste gas facilities as proxy variables for corporate environmental investment. The regression results are reported in **Table 7**. The positively significant coefficients of treatment variable in columns (1) and (2) imply that the green credit policy raises the corporate environmental investment. Columns (3) and (4) test the effect of environment investment on $SO_2$ emissions. The coefficients of $Ginvest_{it}$ are significantly negative, suggesting that the green credit policy will encourage companies to strengthen pollution control.

### 4.5 The Green Credit Policy and Other Pollutant Discharges

This research mainly focuses on the impact of green credit policy on sulfur dioxide emission. However, we are also interested in the question that whether green credit policy also imposes mitigating effect on other industrial pollutants. In line with Huang (2012), this paper further investigates the impact of green credit policies on other pollutant discharges, including industrial smoke, industrial dust, ammonia nitrogen compounds, and chemical oxygen demand (COD). COD is a measure for the amount of organic pollutants in wastewater, while other pollutant discharges are sources of air pollution. The empirical model is as follow:

$$Pollution_{it} = \alpha + \beta Policy \times Time + \gamma \mathbf{X}_{it} + \mu_t + \eta_i + \varepsilon_{it} \quad (6)$$

Estimation results are presented in **Table 8**. The estimated coefficients for all pollutants mentioned above are negative with a significant level at the 1%. Results imply that the green credit policy also imposes a significant mitigating effect on other pollution discharges.

## 5 CONCLUSION AND IMPLICATIONS

This paper investigates whether China's green credit policy contributes to the reduction of firms' $SO_2$ emissions. We hypothesize that the implementation of green credit policies can negatively affect $SO_2$ emissions. To establish causality, we employ aPSM-DID method to explore the impact of the green credit policy on $SO_2$ emissions based on the data of Chinese industrial enterprises. Our estimation indicates the implementation of the green credit policy has successfully reduced firm level $SO_2$ emissions. The effects of the green credit policy on $SO_2$ emissions are more pronounced for large enterprises and enterprises in the eastern region. We also find that the green credit policy reduces $SO_2$ emissions by reducing energy consumption and encouraging environmental investment.

Our results provide several policy implications. First, faced with the challenge of reducing polluting gas emissions in absolute terms, the Chinese government should formulate green financial policies based on local conditions and provide credit resources to favor environmentally-friendly companies. With the deepening of China's market-oriented reforms, green credit policies will play a greater role in improving environmental quality. For financial credit institution, they should strictly implement green credit standards and modify financial products and services. Traditional environmental regulations may not achieve the expected environmental protection and emission reduction effects. Future environmental policies should focus on market-oriented emission reduction methods, and use market tools to restrict pollution emissions.

Secondly, it is very important to strengthen the coordinated economic development of different regions, the construction of financial infrastructure in the central and western regions needs to be improved, to maximize the energy saving and emission reduction of the green credit policy. At the same time, the local authorities should provide a fair competitive external environment for small and medium-sized enterprises. Specifically, policy makers should design differentiated green credit policies for companies of different sizes, which will help them actively assume environmental responsibility and eliminate outdated production capacity.

Finally, energy structure and energy consumption intensity directly affect sulfur dioxide emissions of industrial enterprises, so it should vigorously promote the development of clean energy industries and reduce the use of fossil fuels. An effective measure is to allocate more financial resources to the environmental-friendly industry through the green credit policy. When companies receive external financial support, they are more motivated to carry out green innovations and improve pollution treatment capabilities.

Although this paper has discovered and discussed the negative relationship between green credit policy and industrial pollutant emissions, further investigation is necessary. Future research may aim to answer the question that how financial institutions can help facilitating green credit and promoting the green development of economy.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

ZX: conceptualization, validation, resources, data curation, funding acquisition and writing original draft. LY: methodology, software, writing—review and editing, and visualization. YL: formal analysis, data curation, and investigation. XB: writing—review and editing, supervision, project administration, and funding acquisition. ZY: software, visualization, investigation and resources. All authors: contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We acknowledge the financial support from Shandong Provincial Natural Science Foundation (Grant Number: ZR2020QG032), Shandong Provincial Social Science Planning Office (Grant Numbers: 21DGLJ12; 21DJJJ02), Taishan Scholars Program of Shandong Province, China (Grant Numbers: ts201712059; tsqn201909135) and Youth Innovative Talent Technology Program of Shandong Province, China (Grant Number: 2019RWE004). All errors remain our own.

## REFERENCES

Albrizio, S., Kozluk, T., and Zipperer, V. (2017). Environmental Policies and Productivity Growth: Evidence across Industries and Firms. *J. Environ. Econ. Management* 81, 209–226. doi:10.1016/j.jeem.2016.06.002

Cao, Y., Zhang, Y., Yang, L., Li, R. Y. M., and Crabbe, M. J. C. (2021). Green Credit Policy and Maturity Mismatch Risk in Polluting and Non-polluting Companies. *Sustainability* 13, 3615. doi:10.3390/su13073615

Chen, J., Wang, B., Huang, S., and Song, M. (2020). The Influence of Increased Population Density in China on Air Pollution. *Sci. Total Environ.* 735, 139456. doi:10.1016/j.scitotenv.2020.139456

Chen, W., and Yan, W. (2020). Impact of Internet Electronic Commerce on SO2 Pollution: Evidence from China. *Environ. Sci. Pollut. Res.* 27 (1), 25801–25812. doi:10.1007/s11356-020-09027-1

Davis, L. W., and Lucas, W. (2017). Saturday Driving Restrictions Fail to Improve Air Quality in Mexico City. *Sci. Rep.* 7, 41652. doi:10.1038/srep41652

Fan, H., Tao, S., and Hashmi, S. H. (2021). Does the Construction of a Water Ecological Civilization City Improve Green Total Factor Productivity? Evidence from a Quasi-Natural Experiment in China. *Ijerph* 18 (22), 11829. doi:10.3390/ijerph182211829

Hollingsworth, A., Konisky, D., and Zirogiannis, N. (2021). The Health Consequences of Excess Emissions: Evidence from Texas. *OSF Preprints*. doi:10.31219/osf.io/gc73x

Huang, C. (2012). The Impact of Local Environmental Quality on International Tourism Demand: The Case of China. Master's thesis. San Francisco: University of San Francisco. Available at: https://repository.usfca.edu/thes/21.

Ji, X., Li, G., and Wang, Z. (2017). Impact of Emission Regulation Policies on Chinese Power Firms' Reusable Environmental Investments and Sustainable Operations. *Energy Policy* 108, 163–177. doi:10.1016/j.enpol.2017.05.035

Jiang, J., Zha, Y., and Li, L. (2019). Simulation Analysis of Atmospheric SO2 Contributions from Different Regions in China. *Atmos. Pollut. Res.* 10, 913–920. doi:10.1016/j.apr.2018.12.019

Kihombo, S., Ahmed, Z., Chen, S., Adebayo, T. S., and Kirikkaleli, D. (2021). Linking Financial Development, Economic Growth, and Ecological Footprint: what Is the Role of Technological Innovation? *Environ. Sci. Pollut. Res.* 28, 61235–61245. doi:10.1007/s11356-021-14993-1

Kim, K. Y., and Lee, J. S. (2015). An Analysis of the Impact of China's Pollution Reduction Policy on Sulfur Dioxide Emissions. *ksccr* 6 (4), 367–377. doi:10.15531/KSCCR.2015.6.4.367

Kumar, N., and Foster, A. D. (2009). Air Quality Interventions and Spatial Dynamics of Air Pollution in Delhi and its Surroundings. *Ijewm* 4 (1-2), 85–111. doi:10.1504/IJEWM.2009.026886

Liu, H., Fang, C., Zhang, X., Wang, Z., Bao, C., and Li, F. (2017). The Effect of Natural and Anthropogenic Factors on Haze Pollution in Chinese Cities: A Spatial Econometrics Approach. *J. Clean. Prod.* 165, 323–333. doi:10.1016/j.jclepro.2017.07.127

Liu, J.-Y., Xia, Y., Fan, Y., Lin, S.-M., and Wu, J. (2017). Assessment of a green Credit Policy Aimed at Energy-Intensive Industries in China Based on a Financial CGE Model. *J. Clean. Prod.* 163, 293–302. doi:10.1016/j.jclepro.2015.10.111

Ma, H., and He, G. (2016). Effects of the Post-Olympics Driving Restrictions on Air Quality in Beijing. *Sustainability* 8 (9), 902. doi:10.3390/su8090902

Martin, P. R., and Moser, D. V. (2016). Managers' green Investment Disclosures and Investors' Reaction. *J. Account. Econ.* 61 (1), 239–254. doi:10.1016/j.jacceco.2015.08.004

Nawaz, M. A., Seshadri, U., Kumar, P., Aqdas, R., Patwary, A. K., and Riaz, M. (2020). Nexus between green Finance and Climate Change Mitigation in N-11 and BRICS Countries: Empirical Estimation through Difference in Differences (DID) Approach. *Environ. Sci. Pollut. Res.* 28, 6504–6519. doi:10.1007/s11356-020-10920-y

Park, S., and Lee, Y. (2011). Regional Model of EKC for Air Pollution: Evidence from the republic of Korea. *Energy Policy* 39 (10), 5840–5849. doi:10.1016/j.enpol.2011.06.028

Rafindadi, A. A., Yusof, Z., Zaman, K., Kyophilavong, P., and Akhmat, G. (2014). The Relationship between Air Pollution, Fossil Fuel Energy Consumption, and Water Resources in the Panel of Selected Asia-Pacific Countries. *Environ. Sci. Pollut. Res.* 21 (19), 11395–11400. doi:10.1007/s11356-014-3095-1

Samargandi, N. (2017). Sector Value Addition, Technology and CO2 Emissions in Saudi Arabia. *Renew. Sustainable Energ. Rev.* 78, 868–877. doi:10.1016/j.rser.2017.04.056

Shen, N., Wang, Y., Peng., H., and Hou, Z. (2020). Renewable Energy green Innovation, Fossil Energy Consumption, and Air Pollution-Spatial Empirical Analysis Based on China. *Sustainability* 12 (16), 6397. doi:10.3390/su12166397

Su, C.-W., Khan, K., Umar, M., and Zhang, W. (2021b). Does Renewable Energy Redefine Geopolitical Risks? *Energy Policy* 158 (6), 112566. doi:10.1016/j.enpol.2021.112566

Su, C.-W., Yuan, X., Tao, R., and Umar, M. (2021a). Can New Energy Vehicles Help to Achieve Carbon Neutrality Targets? *J. Environ. Manage.* 297, 113348. doi:10.1016/j.jenvman.2021.113348

Sun, J., Wang, F., Yin, H., and Zhang, B. (2019). Money Talks: the Environmental Impact of China's green Credit Policy. *J. Pol. Anal. Manage.* 38, 653–680. doi:10.1002/pam.22137

Tao, R., Umar, M., Naseer, A., and Razi, U. (2021). The Dynamic Effect of Eco-Innovation and Environmental Taxes on Carbon Neutrality Target in Emerging Seven (E7) Economies. *J. Environ. Manage.* 299, 113525. doi:10.1016/j.jenvman.2021.113525

Umar, M., Ji, X., Kirikkaleli, D., and Xu, Q. (2020). COP21 Roadmap: Do Innovation, Financial Development, and Transportation Infrastructure Matter for Environmental Sustainability in China? *J. Environ. Manage.* 271, 111026. doi:10.1016/j.jenvman.2020.111026

Wang, H., Zhang, Y., Zhao, H., Lu, X., Zhang, Y., Zhu, W., et al. (2017). Trade-driven Relocation of Air Pollution and Health Impacts in China. *Nat. Commun.* 8 (1), 738. doi:10.1038/s41467-017-00918-5

Wang, K.-H., Su, C.-W., Lobonţ, O.-R., and Umar, M. (2021). Whether Crude Oil Dependence and Co2 Emissions Influence Military Expenditure in Net Oil Importing Countries? *Energy Policy* 153, 112281. doi:10.1016/j.enpol.2021.112281

Wang, L., Wang, H., Dong, Z., Wang, S., and Cao, Z. (2021). The Air Pollution Effect of Government Economic Growth Expectations: Evidence from China's Cities Based on green Technology. *Environ. Sci. Pollut. Res.* 28, 27639–27654. doi:10.1007/s11356-021-12536-2

Wang, Y., Lei, X., Zhao, D., Long, R., and Wu, M. (2021). The Dual Impacts of green Credit on Economy and Environment: Evidence from China. *Sustainability* 13 (8), 4574. doi:10.3390/su13084574

Wang, Y., and Zhi, Q. (2016). The Role of green Finance in Environmental protection: Two Aspects of Market Mechanism and Policies. *Energ. Proced.* 104, 311–316. doi:10.1016/j.egypro.2016.12.053

Woo, C., Chung, Y., Chun, D., Han, S., and Lee, D. (2014). Impact of green Innovation on Labor Productivity and its Determinants: an Analysis of the Korean Manufacturing Industry. *Bus. Strat. Env.* 23 (8), 567–576. doi:10.1002/bse.1807

Wu, H., Guo, H., Zhang, B., and Bu, M. (2017). Westward Movement of New Polluting Firms in China: Pollution Reduction Mandates and Location Choice. *J. Comp. Econ.* 45 (1), 119–138. doi:10.1016/j.jce.2016.01.001

Xing, C., Zhang, Y., and Wang, Y. (2020). Do banks Value green Management in China? the Perspective of the green Credit Policy. *Finance Res. Lett.* 35, 101601. doi:10.1016/j.frl.2020.101601

Xu, Y., and Masui, T. (2009). Local Air Pollutant Emission Reduction and Ancillary Carbon Benefits of SO2 Control Policies: Application of AIM/CGE Model to China. *Eur. J. Oper. Res.* 198 (1), 315–325. doi:10.1016/j.ejor.2008.07.048

Yan, S., and Wu, G. (2017). SO2 Emissions in China - Their Network and Hierarchical Structures. *Sci. Rep.* 7, 46216. doi:10.1038/srep46216

Yu, Y., and Du, Y. (2019). Impact of Technological Innovation on CO2 Emissions and Emissions Trend Prediction on 'New Normal' Economy in China. *Atmos. Pollut. Res.* 10, 152–161. doi:10.1016/j.apr.2018.07.005

Zhang, B., Chen, X., and Guo, H. (2018). Does central Supervision Enhance Local Environmental Enforcement? Quasi-Experimental Evidence from China. *J. Public Econ.* 164, 70–90. doi:10.1016/j.jpubeco.2018.05.009

Zhang, K., Li, Y., Qi, Y., and Shao, S. (2021). Can green Credit Policy Improve Environmental Quality? Evidence from China. *J. Environ. Manage.* 298, 113445. doi:10.1016/j.jenvman.2021.113445

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Xiao, Yu, Liu, Bu and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Environmental Science

POLICY BRIEF
published: 01 February 2022
doi: 10.3389/fenvs.2022.807216

# Dynamic Supervision and Control of VOCs Emission From China's Furniture Manufacturing Based on Big Data and IoT

*Zi-Gui Chen[1], Wei-Ping Wu[2]*, Jing Li[3]* and Yu-Heng Zeng[4]*

*[1]Design and Art Institute, Hunan University of Technology and Business, Changsha, China, [2]School of Economy and Trade, Hunan University of Technology and Business, Changsha, China, [3]Business College, Hong Kong Baptist University, Hong Kong, China, [4]Law School, Hunan University of Technology and Business, Changsha, China*

How to effectively monitor and prevent volatile organic compounds (VOCs) in furniture manufacturing has become the key to solving the dilemma between furniture industry development and public health. Based on the combination of VOCs emission standards and prevention policies of China's furniture manufacturing, this article explores the path of dynamic monitoring, intelligent prevention, and control of VOCs embedded in furniture manufacturing with big data and IoT and puts forward feasible VOCs control technologies in the whole process of China's furniture manufacturing, including source, process, and end. Herein, dynamic monitoring, intelligent prevention, and control of VOCs in furniture manufacturing have important reference significance for the prevention and control of VOCs in other industries, such as coal chemical industry, rubber products, engineering machinery, and coiled materials.

**Keywords: VOCs emission, furniture manufacturing, big data, IoT, dynamic supervision**

OPEN ACCESS

***Edited by:***
*Tsun Se Cheong,*
*Hang Seng University of Hong Kong, China*

***Reviewed by:***
*Vincent Chen,*
*The University of Sheffield, United Kingdom*
*Wen Zezhou,*
*Huazhong University of Science and Technology, China*

****Correspondence:***
*Wei-Ping Wu*
*wuweiping.2007@163.com*
*Jing Li*
*694280519@qq.com*

***Specialty section:***
*This article was submitted to Environmental Economics and Management, a section of the journal Frontiers in Environmental Science*

***Received:*** *01 November 2021*
***Accepted:*** *03 January 2022*
***Published:*** *01 February 2022*

***Citation:***
*Chen Z-G, Wu W-P, Li J and Zeng Y-H (2022) Dynamic Supervision and Control of VOCs Emission From China's Furniture Manufacturing Based on Big Data and IoT. Front. Environ. Sci. 10:807216. doi: 10.3389/fenvs.2022.807216*

## INTRODUCTION

Human beings spend about 87% of their time indoors. Therefore, indoor air quality directly affects human health (Wang et al., 2021). With the continuous advancement of economic and social development, people pay more and more attention to the aesthetic and visual enjoyment of furniture, which leads to the diversified and personalized demand for furniture products, making furniture manufacturing enterprises constantly seek new materials, new technologies, and new processes to innovate furniture manufacturing. However, because furniture production is inseparable from adhesive and coating, a large number of irritating or toxic VOC gases will be generated in the production process, forming precursors of PM2.5 and ozone, which will aggravate atmospheric pollution such as haze and photochemical smog. Therefore, how to effectively monitor and prevent VOCs in furniture production has become an important means to solve the dilemma between the development of the furniture manufacturing industry and public health (Menghi et al., 2018). With the continuous maturity of new generation information technologies, such as big data and Internet of things (IoT), people began to explore the application of big data and IoT to VOCs dynamic monitoring and intelligent prevention in furniture production. They achieved certain results (Kaginalkarab et al., 2021). Since China overtook the United States to become the world's largest furniture producer in 2010, it has always ranked first in the world. Therefore, taking China as an example, this article summarizes the evolution of VOCs monitoring and prevention policies and the application of big data and IoT technology in China's furniture manufacturing from

**TABLE 1 |** VOCs emission standards and policies of China's furniture manufacturing.

| National policies | | Provincial policies | |
|---|---|---|---|
| **Policies** | **Content** | **Policies** | **Content** |
| Technical specification for dust prevention and gas prevention in furniture manufacturing industry (2010) | • Technical requirements and management measures for dust prevention<br>• Isolation and remote control shall be adopted for the process and equipment | Emission standard of VOCs for furniture manufacturing industry in Guangdong province (2010) | • VOCs emission concentration limit, concentration limit, and monitoring requirements<br>• Production process and management requirements for controlling VOCs emission in furniture manufacturing industry<br>• VOCs emission monitoring technology shall be implemented in accordance with HJ/T 55 |
| Technical policy for prevention and control of VOCs pollution (2013) | • Strategies of pollution prevention in production, storage, transportation, and consumption of products containing VOCs<br>• Encourage enterprises to carry out VOCs monitoring<br>• Research and development of new technologies and equipment for VOCs | Emission standard of air pollutants from furniture manufacturing industry in Shanghai (2017) | • Concentration limits of VOCs and other air pollutants at monitoring points<br>• The automatic monitoring equipment shall be implemented in accordance with the management measures for automatic monitoring of pollution sources and HJ/T 75 |
| Emission standard of pollutants in synthetic resin industry (2015) | • Emission limit management of VOCs materials in synthetic resin industrial enterprises<br>• Install automatic monitoring equipment for on-site monitoring of pollution sources<br>• HJ/T 76 and other standards can be adopted for monitoring | Emission standard of VOCs in industrial coating process of Fujian Province (2018) | • VOCs emission and monitoring standard for industrial coating process<br>• VOC monitoring equipment and system shall be implemented according to HJ 75, HJ 76, and other relevant methods<br>• VOCs data acquisition and control, data processing, and remote communication system |
| Control standard for unorganized emission of VOCs (2019) | • Supervision of VOCs material storage, transfer, and transportation<br>• VOCs unorganized exhaust gas collection and treatment system<br>• VOCs monitoring process and online analysis system | Emission standard of VOCs in Jiangxi province part 6: furniture manufacturing industry (2019) | • VOCs emission limit, production process, monitoring, and supervision requirements for furniture manufacturing enterprises<br>• VOC monitoring equipment and system shall be implemented according to HJ 75, HJ 76, and other relevant methods |
| Technical guide for pollution prevention and control in furniture manufacturing industry (2021) | • Feasible technologies for VOCs prevention and control in furniture manufacturing industry<br>• Automatic equipment such as manipulator, robot, and reciprocating automatic spraying box is used to reduce paint consumption and VOCs | Comprehensive treatment plan of VOCs in Zhejiang province during the 14th five-year plan (2021) | • Technical requirements of key industries producing VOCs such as furniture manufacturing<br>• Install VOCs automatic monitoring facilities and comprehensively use navigation monitoring technologies |

historical and practical logic, so as to provide scientific basis and technical support for the country to promote VOCs pollution prevention and control in furniture manufacturing industry.

## THE CARDING OF VOCS EMISSION STANDARDS AND POLICIES OF CHINA'S FURNITURE MANUFACTURING

Chinese governments at all levels have issued a series of policies on the standards, monitoring, and prevention of VOCs emission from the furniture manufacturing industry (**Table 1**). From the national level, in September 2010, the State Administration of Work Safety distributed "Technical Specification for Dust Prevention and Gas Prevention in Furniture Manufacturing Industry," which put forward normative requirements for pollution prevention and control in the furniture manufacturing industry for the first time. However, the specification does not put forward hard constraint requirements in technology but recommends that relevant enterprises adopt isolation and remote control operations. In 2013, the Ministry of Natural Resources of the People's Republic of China issued "Technical Policy for Prevention and Control of VOCs Pollution" to encourage relevant enterprises to carry out VOCs monitoring, prevention, and control by themselves and develop new technologies and equipment in the form of suggestions. In 2015, the Ministry of Ecology and Environment of the People's Republic of China issued the

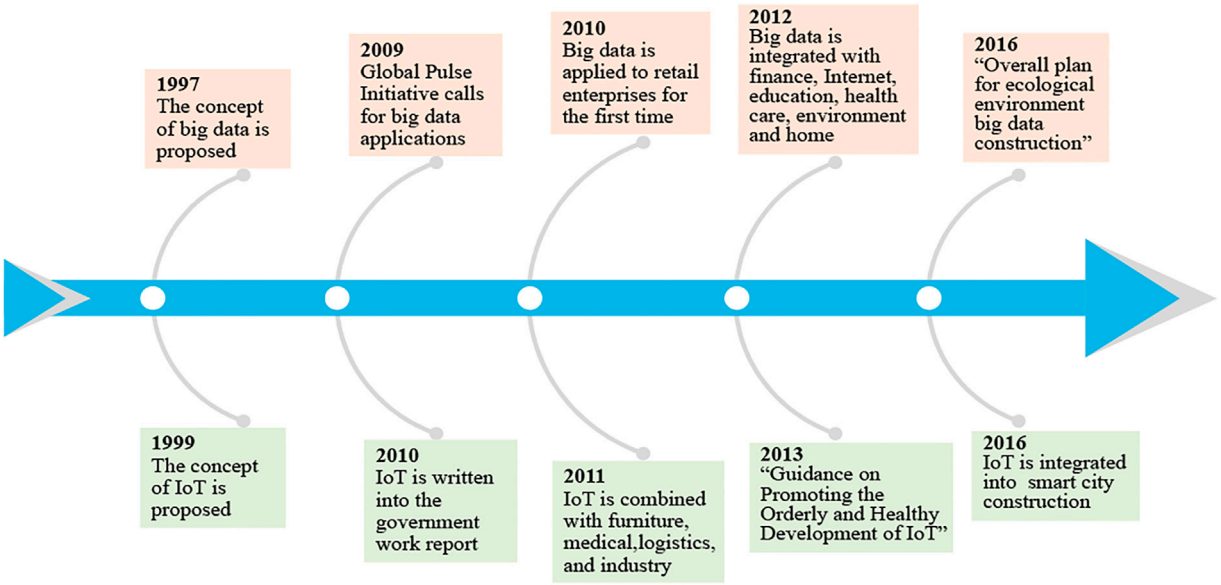



**FIGURE 1 |** Evolution of big data and IoT in China.

"Emission Standard of Pollutants in Synthetic Resin Industry," requiring enterprises associated with furniture manufacturing to install corresponding automatic monitoring equipment for various pollution sources, including VOCs materials. On May 24, 2019, the Ministry of Ecology and Environment of the People's Republic of China and the State Administration for Market Regulation jointly issued "Control Standard for Unorganized Emission of VOCs" and proposed an online sampling and analysis system for VOCs control. On May 12, 2021, the Ministry of Ecology and Environment of the People's Republic of China issued the latest "Technical Guide for Pollution Prevention and Control in Furniture Manufacturing Industry," which put forward guiding suggestions on the use of automatic and intelligent equipment and technologies such as manipulator, robot, and Internet of things for production in terms of equipment and process innovation technology for the first time.

At the provincial level, Guangdong, Shanghai, Zhejiang, Jiangxi, Fujian, and other major furniture manufacturing provinces have issued corresponding policies for VOCs emission monitoring and prevention. For example, on October 22, 2010, Guangdong issued the "Emission Standard of VOCs for Furniture Manufacturing Industry," which put forward specific requirements for instrument use and technological innovation of VOCs emission monitoring in furniture production. In June 2017, Shanghai issued the "Emission Standard of Air Pollutants from Furniture Manufacturing Industry," which required that the automatic monitoring equipment shall be implemented in accordance with the management measures for automatic monitoring of pollution sources and HJ/T 75. Moreover, the "Emission Standard of VOCs in Industrial Coating Process" issued by Fujian province in 2018 clearly stipulates the VOCs emission limit of furniture manufacturing industry and requires manufacturers to establish a variety of monitoring units such as gaseous pollutants, as well as data acquisition and control system, data processing and remote communication system, fixed source monitoring system, and pollution control facility operation monitoring system. On July 17, 2019, Jiangxi province had formulated technical requirements consistent with those of Shanghai and Fujian for VOCs pollution control policies for the furniture manufacturing industry, such as "Emission Standard of VOCs in Part 6: Furniture Manufacturing Industry." In 2021, Zhejiang issued the "Comprehensive Treatment Plan of VOCs in Zhejiang Province during the 14th Five-Year Plan," which prospectively proposed that furniture manufacturing needs to install VOCs automatic monitoring facilities and strengthen the application of intelligent technologies such as navigation monitoring.

In conclusion, it can be seen that a series of policies have been issued for VOCs emission, monitoring, and prevention in furniture manufacturing at both the national and provincial levels. Moreover, with the development and deepening of the new generation of information technology, government departments began to encourage furniture manufacturing enterprises to seek emerging technical means such as IoT and big data to improve the efficiency of VOCs monitoring and prevention.

## APPLICATION SCENARIOS OF BIG DATA AND IOT IN VOCS EMISSION FROM CHINA'S FURNITURE MANUFACTURING

### Evolution of Big Data and IoT

At present, the new generation of information technology represented by big data and IoT is developing rapidly, gradually accelerating its penetration into new materials, intelligent medicine, intelligent manufacturing, and other fields (**Figure 1**). Looking

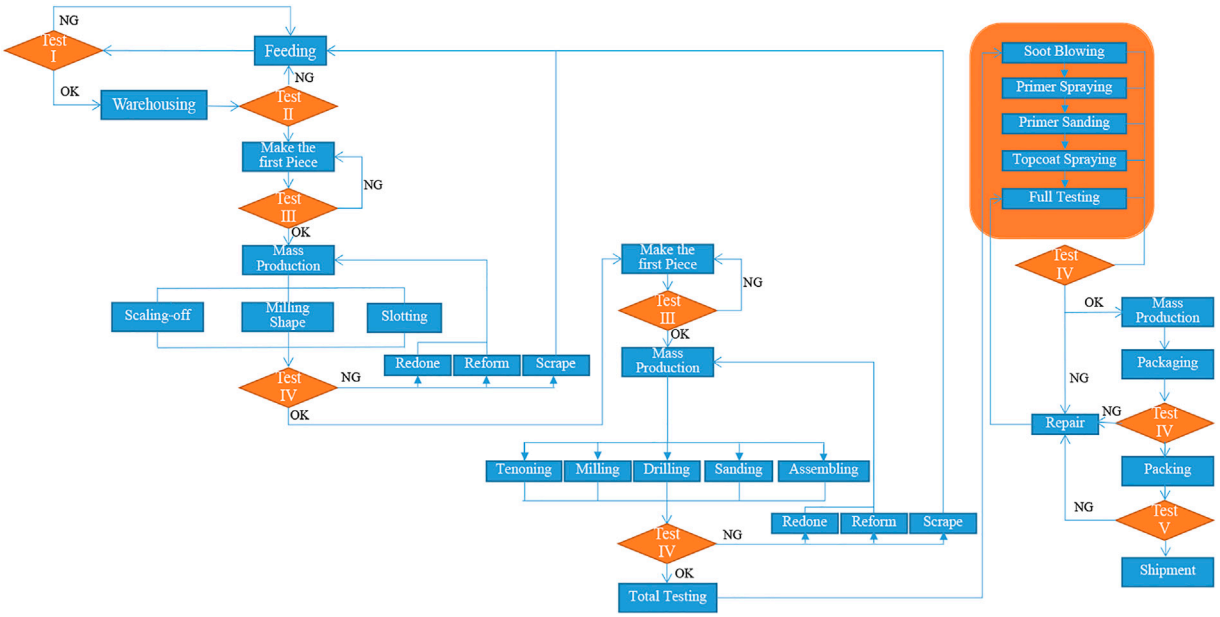



**FIGURE 2 |** Big data and IoT in furniture manufacturing. Note: the IoT intelligent perception system and big data technology are embedded in the module with orange shadow.

back on history, both big data and IoT were initiated in the late 20th century and flourished in the early 21st century. Among them, the concept of big data was first proposed by Michael and David Ellsworth in 1997. The concept of IoT was first proposed by the Auto-ID laboratory in 1999 and formally proposed at the World Summit on the Information Society (WSIS) in 2005. As big data and IoT become gradually known and recognized, they begin to enter people's vision of life and production, which has changed the life and production mode and its inherent thinking to a certain extent. In 2009, the United Nations issued the "Global Pulse Plan," calling for big data and IoT as the driving force to promote the economic development of a country or region. Therefore, big data and IoT began to be rapidly promoted and applied in China, India, and other countries. Meanwhile, Wen Jiabao, then Chinese Premier of the State Council, inspected and delivered an important speech at Wuxi Micro Nano Sensor Network Engineering Technology R&D Center, saying that China should seize the opportunity and vigorously develop Internet of things technology.

At the national two sessions held in March 2010, it was clearly pointed out that the use of big data and IoT to promote the transformation of economic development mode, and the IoT was established as one of the strategic pillars of national economic and technological development. In addition, in the "12th Five Year Plan for the development of the IoT" issued by the Ministry of Industry and Information of the People's republic of China in 2011, the Ministry further defined the direction and focus of the development of the IoT, and promoted the interactive development of IoT technology and medical, logistics, home, industry, agriculture and other fields. In 2013, the State Council of China issued the guiding opinions on promoting the orderly and healthy development of the IoT, defined the guiding ideology, development objectives and main tasks of the development of the Internet of things, and proposed to promote the construction of a new smart city based on the classification of IoT technologies. In 2016, the Ministry of Ecology and Environment of the People's Republic of China also issued the overall plan for the construction of ecological and environmental big data, which clearly pointed out the use of big data and IoT to monitor air pollution, environmental governance and life health, marking the formal integration of big data and Internet of things technology into all fields of life and production.

## Application Scenarios and Platform Architecture for Big Data and IoT

1) Big data and IoT implanted furniture manufacturing link

The development of IoT is based on the extension and expansion of the Internet. Its ultimate development goal is to realize comprehensive perception, reliable transmission, and intelligent processing with the help of big data technology. With the continuous integration and application of IoT and big data technology. They are more and more respected in VOCs monitoring and control of furniture manufacturing. As we all know, the constituent materials of furniture mainly fall into two categories: one is dry materials, which mainly refers to all kinds of wood panels (solid wood panels, man-made panels, PVC panels, *etc.*), and soft materials (mattresses, sofa mattresses, *etc.*). One is wet materials (Wu et al., 2021), which mainly refer to paints, wood colorants, adhesives, and so on. As the base material of the furniture, the wood board itself will release a variety of aldehydes and pinenes. In the process of furniture manufacturing, the use of adhesives, paints, and other materials will introduce new organics, such as esters, ketones, and aromatic hydrocarbons.

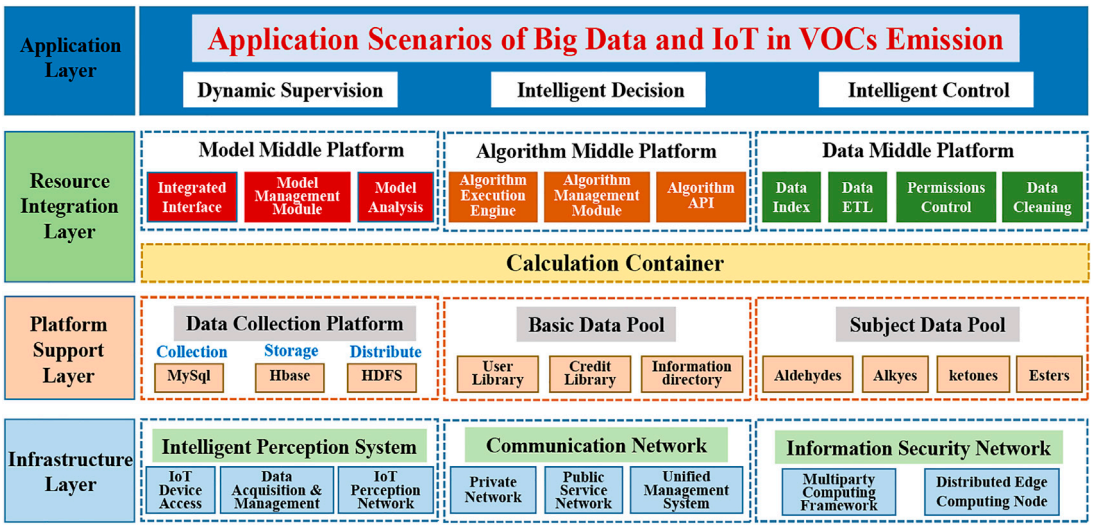



**FIGURE 3 |** Platform architecture for big data and IoT in VOCs emission.

Therefore, the intelligent monitoring system based on IoT and big data technology is embedded into the important links and modules of furniture manufacturing, which helps collect VOCs monitoring data from time to time, and supports the intelligent and green selection of materials and production process of furniture manufacturing.

**Figure 2** reports the production process of furniture manufacturing and the implantation of big data and IoT. See the orange shaded module in the figure for details. It can be seen that the IoT intelligent perception system and big data technology are mainly applied to VOCs detection and collection before material warehousing (Test I); VOCs detection and collection when material warehousing (Test II); VOCs detection and collection of the first product produced (Test III); VOCs detection and collection after various production links in mass production (Test IV); VOCs detection and collection after packaging of furniture parts (Test V); VOCs detection and collection in paint production; and so on. Among them, the paint manufacturing process includes soot blowing, primer spraying, primer sanding, topcoat spraying, paint full inspection, and other modules, which are the main links of VOCs generation and emission.

2) Platform architecture for big data and IoT in furniture manufacturing

The design of infrastructure platform architecture is the key to solving the intelligent monitoring system of big data and IoT. From **Figure 3**, the basic platform architecture for VOCs dynamic monitoring, intelligent decision-making, and control in furniture manufacturing is mainly composed of four layers, namely, application layer, resource integration layer, platform support layer, and infrastructure layer. Among them, the application layer mainly solves the dynamic monitoring, intelligent decision-making, and control of VOCs emission data. The resource integration layer includes modules such as model platform, data platform, algorithm platform, and computing container. The platform support layer includes three modules: data collection platform (collection, storage, distribution), basic data pool (user library, credit library, information directory), and subject data pool (such as aldehydes, alkies, ketones, and esters). The infrastructure layer consists of three modules: intelligent sensing system, communication integrated network, and information security network.

## CONCLUDING REMARKS

The furniture manufacturing industry is a typical industry with high pollution and low added value. The VOCs produced in furniture manufacturing seriously affect people's health. Based on big data and IoT and combining the VOCs emission standards and prevention policies of China's furniture manufacturing, we explore the path and mode of dynamic monitoring and intelligent prevention and control of VOCs embedded in furniture manufacturing by big data and IoT and put forward the feasible VOCs control technology of the whole process of China's furniture manufacturing industry, including source, process, and end. In short, the prevention and control of VOCs in furniture manufacturing are not achieved overnight. In addition to the implementation of key technology and production process innovation in important links such as raw material selection, production adhesive, and coating, it is also necessary to combine emerging technologies such as big data and IoT to improve the efficiency of VOCs monitoring, prevention, and control in the whole process of furniture production. In addition, in reality, there are many industries involving VOCs emission, including coal chemical industry, rubber products, engineering machinery, steel structure, and coiled materials. These industries all need to carry out more refined monitoring, prevention, and control research. In this article, the dynamic monitoring, intelligent prevention, and control of VOCs in furniture manufacturing have important reference

significance for the prevention and control of VOCs in other industries mentioned above.

## AUTHOR CONTRIBUTIONS

Z-GC: conceptualization, writing—original draft, and supervision. W-PW: writing—original draft and methodology. JL: software, writing—review and editing. Y-HZ: Methodology, writing—review and editing.

## REFERENCES

Kaginalkar, A., Kumar, S., Gargava, P., and Niyogi, D. (2021). Review of Urban Computing in Air Quality Management as Smart City Service: An Integrated IoT, AI, and Cloud Technology Perspective. *Urban Clim.* 39, 100972. doi:10.1016/j.uclim.2021.100972

Menghi, R., Ceccacci, S., Papetti, A., Marconi, M., and Germani, M. (2018). A Method to Estimate the Total VOC Emission of Furniture Products. *Proced. Manuf.* 21, 486–493. doi:10.1016/j.promfg.2018.02.148

Wang, Y., Wang, H., Tan, Y., Liu, J., Wang, K., Ji, W., et al. (2021). Measurement of the Key Parameters of VOC Emissions from Wooden Furniture, and the Impact of Temperature. *Atmos. Environ.* 259 (7), 118510. doi:10.1016/j.atmosenv.2021.118510

Wu, W. p., Zhu, Y. f., Zeng, W. k., Wang, M., Yang, D. x., and Chen, W. f. (2021). Green Efficiency of Water Resources in Northwest China: Spatial-Temporal Heterogeneity and Convergence Trends. *J. Clean. Prod.* 320, 128651. doi:10.1016/j.jclepro.2021.128651

## FUNDING

This research was supported by the National Natural Science Foundation of China (no. 71903056), National Science Foundation for Post-Doctoral Scientists of China (no. 2020M672464), Natural Science Foundation of Hunan Province in China (nos. 2020JJ5104 and 2020JJ4259), Project of Hunan Social Science Achievement Appraisal Committee (no. XSP21YBZ094), Scientific Research Project of Hunan Education Department in China (nos. 19C1038 and 21B0573).

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Chen, Wu, Li and Zeng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

# Advantages of publishing in Frontiers

**OPEN ACCESS**
Articles are free to read for greatest visibility and readership

**FAST PUBLICATION**
Around 90 days from submission to decision

**HIGH QUALITY PEER-REVIEW**
Rigorous, collaborative, and constructive peer-review

**TRANSPARENT PEER-REVIEW**
Editors and reviewers acknowledged by name on published articles

**Frontiers**
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

**Visit us:** www.frontiersin.org
**Contact us:** frontiersin.org/about/contact

**REPRODUCIBILITY OF RESEARCH**
Support open data and methods to enhance research reproducibility

**DIGITAL PUBLISHING**
Articles designed for optimal readership across devices

**FOLLOW US**
@frontiersin

**IMPACT METRICS**
Advanced article metrics track visibility across digital media

**EXTENSIVE PROMOTION**
Marketing and promotion of impactful research

**LOOP RESEARCH NETWORK**
Our network increases your article's readership