frontiers | Frontiers in Sustainable Food Systems

TYPE Original Research
PUBLISHED 12 January 2026
DOI 10.3389/fsufs.2025.1622916

OPEN ACCESS

EDITED BY
Samina Mehnaz,
Forman Christian College, Pakistan

REVIEWED BY
Abbas Adnan,
Anhui Polytechnic University, China
Tulus Joseph Marpaung,
Universitas Sumatera Utara, Indonesia

*CORRESPONDENCE
Yipeng Zhuang
✉ yipengzh@hku.hk
Yuxiang Bai
✉ cotsbyx@163.com
Ge Wang
✉ wangge302@126.com
Dingchun Zi
✉ suzdc2018@163.com

RECEIVED 06 May 2025
REVISED 16 September 2025
ACCEPTED 30 November 2025
PUBLISHED 12 January 2026

CITATION
Song J, Wu J, Li X, Ning Q, Liu W, Liu Z, Yuan G, Huang H, Guo Y, Zhuang Y, Bai Y, Wang G and Zi D (2026) Tobacco cultivation in Yunnan: an integrated analysis of economic trends and diversification.
Front. Sustain. Food Syst. 9:1622916.
doi: 10.3389/fsufs.2025.1622916

Tobacco cultivation in Yunnan: an integrated analysis of economic trends and diversification

Jie Song[1], Jiang Wu[1], Xiangyang Li[1], Qingong Ning[1], Wei Liu[1], Zefan Liu[2], Guoyi Yuan[3], Haichang Huang[4], Yifeng Guo[5], Yipeng Zhuang[5,6]*, Yuxiang Bai[3]*, Ge Wang[3]* and Dingchun Zi[3]*

[1]China Tobacco Yunnan Import and Export Co., Ltd., Kunming, China, [2]Yunnan Tobacco Company Qujing City Company Luliang Branch, Luliang, China, [3]College of Tobacco Science, Yunnan Agricultural University, Kunming, China, [4]Tsinghua Shenzhen International Graduate School, Shenzhen, China, [5]Demetech Limited, Hong Kong, China, [6]LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, China

This study conducts an integrated analysis of the impacts of tobacco cultivation on farmers in Yunnan Province using machine learning and econometric methods, drawing on detailed micro-level survey data collected from 253 households. Our findings highlight the significant role of tobacco income within household income structures, demonstrating its competitiveness compared to other crops. Through interpretable machine learning techniques and advanced econometric methods, the analysis reveals that expanding tobacco cultivation is associated with improvements in certain indicators of farmers' quality of life, although the benefits vary with the scale of production. Additionally, the promotion of agricultural diversification by the Tobacco Bureau has shown positive effects, despite challenges such as localized disadvantages, such as poor soil quality, limited access to irrigation, and remoteness from markets. The results emphasize the need for optimizing cropping structures by adjusting crop combinations, planting schedules, and resource allocation to improve yield and sustainability and improving management practices to maximize the benefits of tobacco cultivation while supporting sustainable agricultural development.

KEYWORDS

tobacco cultivation, household income, scale efficiency, agricultural diversity, quality of life, interpretable machine learning

1 Introduction

As global agricultural practices evolve, the need for sustainable and diversified farming systems has become increasingly important. Tobacco cultivation continues to attract both farmers and governments due to its considerable economic profitability and revenue-generating potential. In many regions, where tobacco cultivation has historically been a significant economic activity, understanding the dynamics of this crop within the broader agricultural landscape is essential. Tobacco farming, while profitable, poses various challenges related to income stability, environmental sustainability, and social welfare (Rahman et al., 2020; Wan et al., 2022). Recent research further highlights the importance of integrated policy and resource management, such as Song et al. (2025) on collaborative governance models for optimizing resource allocation in waste management, and Abbas et al. (2022) on energy input–output efficiency and greenhouse gas emissions in crop production. As farmers face pressures from market fluctuations and changing regulations, there is a pressing need to evaluate the

role of tobacco cultivation in their income structures and consider the potential benefits of agricultural diversification.

Controversies surrounding tobacco cultivation become increasingly apparent. On the one hand, the resource-intensive nature of tobacco cultivation, characterized by the extensive use of pesticides and fertilizers, causes irreversible damage to the land and other crops, resulting in long-term negative impacts on agriculture (Bondurant et al., 2001). On the other hand, large areas of fertile land are dedicated to tobacco cultivation, reducing the land available for food crops and exacerbating food security challenges (Achterbosch et al., 2014). Furthermore, the pollution generated during the tobacco curing process and the well-documented health risks associated with smoking have become more widely recognized by the public (World Health Organization, 2017). Besides, with the decline of the agricultural population in China and the growing issue of aging farmers, exploring sustainable agricultural development has become more critical. In this context, tobacco farming, as one of Yunnan's key economic crops, plays a significant role in achieving sustainable development by analyzing their income and expenditure structures.

This study investigates the economic reliance of farmers on tobacco cultivation in Yunnan Province, China, along with the associated challenges faced by farming communities. Yunnan possesses favorable geographical conditions for tobacco farming, making it one of the largest tobacco-producing regions in the country. Under the framework of national and industry-specific monopoly regulations, the tobacco industry has become one of the province's top five pillar industries, significantly contributing to industrial development, increasing farmers' incomes, promoting plateau-specific agriculture, and alleviating rural poverty (Tai and Yang, 2024). The economic significance of tobacco cultivation has shaped both the livelihoods of farmers and the broader socioeconomic development of the region, making it a vital component of Yunnan's agricultural economy.

The socioeconomic context in Yunnan presents both opportunities and challenges for local farmers. While tobacco cultivation offers farmers a relatively stable and profitable income source, it also exposes them to long-term risks. The reliance on a single cash crop makes farmers vulnerable to economic shocks, particularly in the face of fluctuating market conditions and changing agricultural policies. This dependency limits farmers' ability to diversify their income sources, posing risks to their long-term economic resilience. Therefore, it is essential to examine how tobacco cultivation influences farmers' income structures and assess the potential benefits of agricultural diversification in mitigating these risks and improving livelihoods. Moreover, this analysis can reveal the factors that influence income levels and the overall quality of life for farmers, providing insights into their economic resilience.

In light of the challenges faced by tobacco farmers, this paper seeks to explore the intricate income structures, quality of life, and agricultural diversification strategies among these farmers in Yunnan. By employing interpretable machine learning and advanced econometric methods, this study aims to provide a comprehensive analysis of income portfolios and identify the key determinants influencing farmers' economic wellbeing. The insights gained from this research will contribute to actionable recommendations, empowering farmers to transition toward more sustainable agricultural practices while enhancing their economic resilience. Understanding the economic and social complexities of tobacco cultivation is crucial, as it enables a deeper appreciation of its role in farmers' livelihoods and highlights opportunities for improvement. Although this study focuses on Yunnan Province, the analytical framework and findings, especially regarding income composition, diversification strategies, and policy interventions, offer valuable insights for other tobacco-growing regions. The methodological approach and policy implications may be adapted to similar agricultural contexts in other provinces, and even international settings where tobacco remains a key economic crop. This study aspires to foster a more sustainable future for tobacco farmers, equipping them with the knowledge to navigate the evolving agricultural landscape effectively.

The rest of this paper is organized as follows. Section 2 presents our research design, including data acquisition details and the analysis methodologies employed. In Section 3, we report the analysis results along with a thorough discussion of their implications. Finally, we present our conclusions and outline the implications of our findings for agricultural practices among tobacco farmers in Section 4.

2 Materials and methods

2.1 Tobacco cultivation in Yunnan

Before the mid-20th century, Yunnan Province's rugged terrain and limited transportation infrastructure constrained agricultural production, resulting in low overall income levels for farmers, particularly in remote mountainous areas. Many farmers relied primarily on traditional crop cultivation, along with limited forestry and animal husbandry, leading to relatively single-source incomes.

The shift toward tobacco cultivation began in the 1940s, when Yunnan's comparative natural advantages for tobacco farming were recognized. Under government guidance, China's tobacco industry gradually relocated from northern provinces such as Shandong and Henan to Yunnan. By 1990, Yunnan surpassed Henan in tobacco cultivation area, securing its position as the country's top tobacco-producing province. Government policies played a pivotal role in this transformation, offering planting subsidies, technical training, and financing assistance to farmers. Additionally, the State Tobacco Monopoly Administration ensured stable farmer incomes through protective purchase prices for tobacco leaves, thereby reducing market risks and incentivizing tobacco farming.

Since then, tobacco cultivation has become a crucial component of farmers' incomes in Yunnan, especially in areas suitable for growing tobacco, where the high added value of tobacco has significantly enhanced farmers' economic returns (Peng et al., 2000; Feng and Ji, 2010; Luo et al., 2015). Figure 1 illustrates the scale of tobacco cultivation in Yunnan from 2001 to 2022 and its proportion of the national total.

However, tobacco cultivation in Yunnan has faced mounting challenges in recent years. With growing public awareness of the health risks associated with smoking, tobacco farming has come under increasing criticism for encroaching on land that could be used for food crops and causing long-term soil degradation. In response, the provincial government has introduced environmentally friendly farming practices, promoted agricultural restructuring, and encouraged the cultivation of alternative cash crops such as vegetables, fruits, and medicinal plants. Despite these

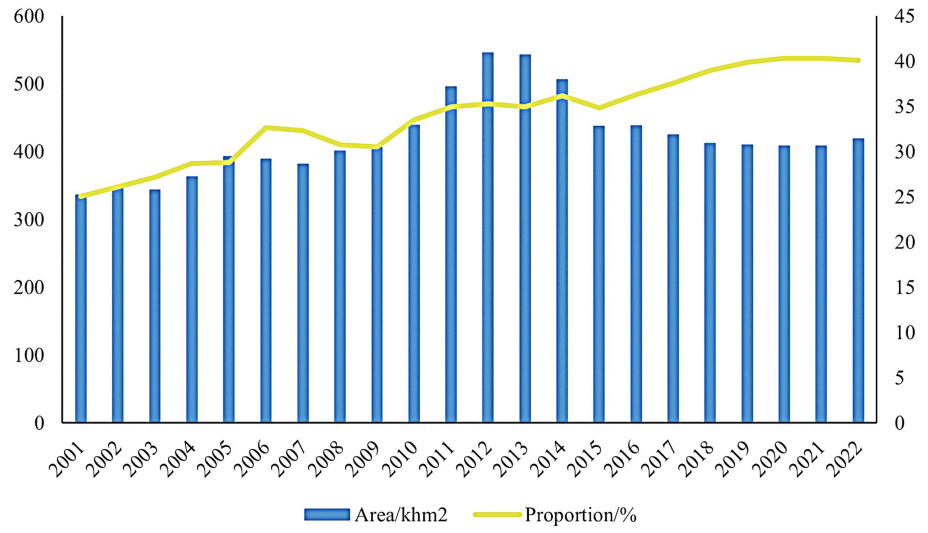

FIGURE 1
The scale of tobacco cultivation in Yunnan from 2001 to 2022 and its proportion of the national total. The data is sourced from the National Bureau of Statistics of China.

efforts, Yunnan's unique natural advantages continue to make tobacco one of the most competitive agricultural options. Compared to staple crops like corn and potatoes, tobacco offers higher economic returns, reinforcing its role as a pillar of the province's agricultural economy. Going forward, balancing economic returns with sustainable farming practices will be essential to ensure the long-term viability of Yunnan's agricultural sector.

2.2 Farm selection and sampling approach

This study utilized a structured questionnaire survey combined with quantitative analysis to investigate the economic and social conditions of tobacco farmers in Yunnan Province. To achieve representative sampling, we selected 10 township-level regions across three major tobacco-producing areas: Dali, Honghe, and Qujing Prefecture. These districts were chosen based on their significance in tobacco production and their diverse agricultural and socioeconomic characteristics.

The selection process was guided by preliminary consultations with key stakeholders, including tobacco industry employees, township government officials, village leaders, and tobacco farmers. These discussions provided essential insights into the income composition, production practices, and living conditions of local farmers, ensuring that the survey design captured relevant and context-specific variables.

The questionnaire covered a broad range of topics, including household income sources, agricultural diversification, farming practices, and perceptions of economic resilience. The survey aimed to capture both quantitative data on income structures and qualitative insights into farmers' challenges and adaptation strategies. This mixed-methods approach enhances the reliability of the findings by providing a comprehensive understanding of the socioeconomic dynamics within tobacco-farming communities.

To ensure data accuracy, the survey was conducted through in-person interviews with farmers, allowing for clarification of questions and reducing potential biases. The collected data was subsequently subjected to rigorous statistical analysis to identify patterns and key determinants affecting farmers' income and livelihood quality.

2.3 Data acquisition and description

During the 2022 field survey, 310 questionnaires were distributed across selected regions in Yunnan Province. After removing incomplete or invalid responses—such as those with short completion times or repetitive answers—253 valid questionnaires were retained, yielding an 75.8% response rate. The sample size of 253 households was determined based on regional representation and statistical reliability. Yunnan is the largest tobacco-growing region globally, and we selected two of its largest city-level tobacco zones for randomized surveying. This ensures that the income sources and structures of tobacco farmers across different terrains and administrative divisions are adequately represented. Moreover, this sample size provides sufficient data volume for robust econometric and machine learning analysis, enhancing the credibility of the findings.

The surveyed households generally consisted of 4 to 6 members, a typical family size in the region. Most farmers reported cultivating tobacco on land plots ranging from 0.26 to 0.50 hectares, which aligns with the average tobacco farm size in Yunnan. Additionally, the geographic distribution of participants was relatively balanced, with approximately equal representation from mountainous and basin terrains. This balance ensures that the dataset captures diverse agricultural and socioeconomic conditions across different terrains.

Table 1 presents the descriptive statistics of the key variables used in this study. The table highlights significant variation in farmers' agricultural practices, including differences in land use, income sources, and crop diversification strategies. These variations provide

valuable insights into the heterogeneity of tobacco farming in Yunnan, allowing for a nuanced analysis of the factors influencing farmers' economic outcomes. For further details on the questionnaire design and content, refer to the Appendix.

2.4 Methodology

2.4.1 Analysis of income structure

To investigate the role of tobacco cultivation income as a revenue source for farmers in Yunnan, micro-level data were collected and used to calculate and summarize household income. The analysis focused on the profitability of tobacco in comparison to other crops, specifically among tobacco-growing households that also cultivate corn, potatoes, wheat, fruits, and medicinal plants. Paddy production was excluded from this comparison, as it is typically consumed directly by the households and does not generate explicit cash income.

Data on growth cycles were gathered through field surveys and professional references. The adjusted per-mu income was calculated using the following formula:

$$Adjusted\ Per-Mu\ Income = \frac{Crop\ Income\left(yuan\right)}{Planting\ Area\left(mu\right)\times Growth\ Cycle\left(day\right)}.$$

Given the data did not meet the assumptions of normality or homogeneity of variance, the Kruskal-Wallis test was applied using Stata 17.0 to compare the profitability of the six crops. This non-parametric method compares the medians of three or more independent groups. Subsequently, post-hoc tests were conducted to perform pairwise comparisons.

TABLE 1 Descriptive analysis results.

Variables		Frequency	Relative frequency (%)
Tobacco planting area (hm^2)	≦0.25	36	14.23
	0.26–0.50	104	41.11
	0.51–0.75	37	14.62
	0.76–1.00	31	12.25
	>1.01	45	17.79
Whether to grow other crops	Yes	248	98.02
	No	5	1.98
Whether to breed	Yes	226	89.33
	No	27	10.67
Prefecture of investigation	Dali	77	30.43
	Qujing	84	33.20
	Honghe	92	36.36
Number of family members	≦3	45	17.79
	4–6	193	76.28
	≧7	15	5.93
Mountainous terrains or dam terrains	Mountainous	139	54.94
	Dam	114	45.06

2.4.2 Analysis of factors affecting income and quality of life

To explore the determinants of household income, an Extreme Gradient Boosting (XGBoost) regressor was trained, followed by SHapley Additive exPlanations (SHAP) analysis to interpret the model's outputs (Lundberg et al., 2020). XGBoost is a decision-tree-based ensemble algorithm that optimizes performance through gradient descent and regularization. It is particularly effective for regression tasks involving complex, non-linear relationships. The SHAP framework allows for a detailed breakdown of feature importance by quantifying the marginal contribution of each predictor to the model's output, thereby improving interpretability and ensuring transparency in identifying income determinants (Dwivedi et al., 2023).

The explanatory variable set (X) was constructed from questionnaire responses and comprised 46 variables, including tobacco farm size, household population, livestock ownership, and other socioeconomic factors. To ensure consistency in model training, categorical variables were encoded as dummy variables. The response variable (Y) was defined as the average income per household member. Mathematically, the XGBoost model minimizes the following objective function to predict the household income:

$$L_{XGBoost} = \sum l\left(y_i, \widehat{y_i}\right) + \sum \Omega\left(f_k\right),$$

where l is the loss function (squared error was used here), and Ω is the regularization term.

To interpret the model, SHAP values are computed. These values represent the marginal contribution of each feature to the prediction, based on cooperative game theory:

$$\phi_i = \sum_{S\subseteq F\{i\}} \frac{|S|!\left(|F|-|S|-1\right)!}{|F|!}\left(val\left(S\cup\{i\}\right)-val\left(S\right)\right),$$

where S is the set of all features, $val\left(S\right)$ is the prediction for feature values.

It provided a granular understanding of feature impacts across different households. The resulting SHAP values ranked the variables based on their influence on income predictions, highlighting key factors driving income variation. For example, variables such as tobacco farm size and household composition were found to be among the most significant predictors. Visual interpretations of SHAP values further elucidated how these variables contribute positively or negatively to income levels, offering a nuanced perspective on the socioeconomic dynamics of tobacco-farming households.

Furthermore, we explore the concept of scale efficiency in relation to tobacco cultivation. This methodology involves visualizing the relationship between cultivation area, per capita income, and its growth rate across different scales. By understanding how these variables interact, we can identify optimal cultivation practices that maximize income.

Following this, we utilized Stata to assess the benefits of expanding tobacco cultivation areas from multiple perspectives. In this analysis, tobacco cultivation area serves as the explanatory variable, while three key factors are selected as dependent variables to construct three distinct linear regression models:

A. Household education expenditure plays a crucial role in enhancing human capital, fostering income growth potential, improving

social status, and promoting sustainable economic development (Chapman, 2016). Model 1 (Equation 1) is formulated with the natural logarithm of "household education expenditure" (Edu_i) as the dependent variable, "tobacco cultivation area" (A_i), "Family Members" (FM_i), "Terrain" (T_i, coded as 1 for mountainous terrains and 0 for plain terrains), and P_{ik} ($k = 1,2,3$) represents the different prefecture as the explanatory variable.

$$\ln\left(Edu_i\right) = \beta_0 + \beta_1 A_i + \beta_2 FM_i + \beta_3 T_i + \sum_{k=1}^{3} \gamma_k P_{ik} + \varepsilon_i, \quad (1)$$

where ε_i is the error term.

B. Engel's coefficient is the proportion of total food expenditure in total household consumption expenditure. In the 19th century, German statistician Engel derived a rule based on statistical data regarding changes in consumption structure: the lower the household income, the larger the proportion of income (or total expenditure) spent on food. As household income increases, the proportion of income (or total expenditure) spent on food decreases. Therefore, Engel's coefficient is often used to reflect the living standards of a household or a country's residents. A lower Engel's coefficient indicates that the proportion of expenditure on food is smaller, reflecting a higher standard of living. The model 2 is constructed with "Engel's coefficient" as the dependent variable and "tobacco cultivation area" as the explanatory variable (Equation 2).

$$Engel_i = \beta_0 + \beta_1 A_i + \beta_2 FM_i + \beta_3 T_i + \sum_{k=1}^{3} \gamma_k P_{ik} + \varepsilon_i. \quad (2)$$

Engel's coefficient has certain limitations in measuring living standards, especially since its theoretical foundation does not align with the traditional Chinese view of "food as the most important thing for the people." An increasing number of scholars believe that Engel's coefficient cannot accurately reflect the actual living standards of Chinese residents. A quality of life index for tobacco-growing households was proposed. Theoretical foundations are based on the OECD's introduction of the Better Life Index (BLI) in 2011, which measures the quality of life of residents in countries around the world. This index incorporates multidimensional factors such as income, employment, housing, health, and education, making it highly valuable as a reference for constructing a comprehensive "quality of life" evaluation indicator (Resce and Maynard, 2018).

C. Based on the expenditure structure of tobacco-growing households—comprising transportation and communication expenses, education expenses, entertainment expenses, housing expenses, household goods and services expenses, food expenses, tobacco, tea, and sugar expenses, clothing expenses, and healthcare expenses—the relative importance of the indicators was initially determined through literature analysis and thematic discussions (Koronakos et al., 2020; Aydan et al., 2022; De Moura Pereira et al., 2022). Subsequently, the Analytic Hierarchy Process (AHP) was applied to finalize the weights of each indicator. The detailed construction of this comprehensive evaluation index can be found in the Appendix. The model 3 can be expressed as shown in Equation 3.

$$Q_i = \beta_0 + \beta_1 A_i + \beta_2 FM_i + \beta_3 T_i + \sum_{k=1}^{2} \gamma_k P_{ik} + \varepsilon_i. \quad (3)$$

The integration of interpretable machine learning (e.g., SHAP analysis) and econometric regression models enables a dual-layered analytical approach. Machine learning is used to uncover non-linear relationships and rank feature importance across diverse household profiles. Econometric models then quantify these relationships, allowing for statistical inference and control of confounding factors.

2.4.3 Analysis of the effect of agricultural diversification

To analyze the agricultural diversity, the Agricultural Diversity (Diversity) was introduced, which measures the extent of diversification in farmers' planting patterns. The formula is:

$$Diversity_i = \sum_{m=1}^{M} \alpha_{im},$$

where α_{im} is a dummy variable that equals 1 if farmer i grows or raises crop or livestock m; otherwise, it is 0. This variable provides valuable insights into the sustainability and economic stability of farming households. A higher diversity score indicates a more varied and potentially more stable agricultural system, which can be less susceptible to market fluctuations and environmental stresses.

3 Results and discussion

The results presented in this section follow a structured analytical pipeline:

Step 1: Data preprocessing and feature engineering were conducted to clean and encode survey responses.

Step 2: Descriptive statistics were computed (Tables 1, 2) to understand income distribution and crop profitability (Figure 2).

Step 3: XGBoost model was trained to predict income, and SHAP values were extracted to rank feature importance (Figure 3).

Step 4: Scale efficiency was analyzed by plotting income growth across cultivation sizes (Figure 4).

Step 5: Regression models were estimated to assess the impact of tobacco area on education, Engel's coefficient, and QoL (Table 3).

Step 6: Agricultural diversity scores were calculated and visualized (Table 4 and Figure 5) to evaluate diversification benefits.

3.1 The position of tobacco cultivation income in the income structure

As it stands, the position of tobacco cultivation income as the primary source of revenue for farmers in Yunnan remains difficult to challenge. This perspective is supported by detailed micro-level data as Table 2.

The average total income of households reached 96,900 Yuan, with a maximum of 405,000 Yuan. For tobacco-growing households, the proportion of income derived from tobacco cultivation averaged 43%, with a maximum of 96%. This highlights that tobacco income remains a significant pillar of household revenue.

Then we focus on tobacco-growing households to compare the planting profitability of tobacco with other local crops and discuss the economic competitiveness of tobacco. The survey revealed that, in

TABLE 2 Descriptive statistics of household income.

Variables	Obs.	Mean	SD	Min	Median	Max
Tobacco cultivation income (10,000 of Yuan/CNY)	253	4.95	3.896	1.5	3.78	18.91
Total income (10,000 of Yuan/CNY)	253	11.13	5.815	3	9.69	40.50
Tobacco cultivation income / Total income	253	0.43	0.225	0.12	0.43	0.96

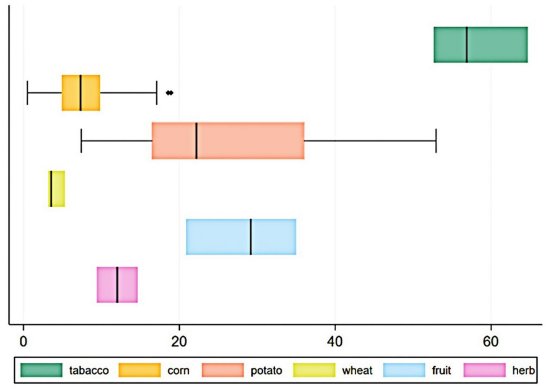

FIGURE 2
The box plot of adjusted per-mu income (CNY) for the growth cycle above was created to visualize the test results.

addition to tobacco cultivation, these households also grow crops such as corn, potatoes, wheat, fruits, and medicinal plants.

Since the data do not satisfy the assumptions of normality or homogeneity of variance, the Kruskal-Wallis test was applied. The test produced a statistic $\chi^2 = 265.03$ ($p < 0.01$). Thus, the null hypothesis was rejected, indicating significant differences in the medians across the groups.

A post-hoc test was further conducted using pairwise comparisons to determine whether the differences between groups were significant. The results showed that the profitability of tobacco was significantly different from the other five crops, with all comparisons yielding p-values less than 0.05 ($p < 0.05$).

In summary, the per-mu income data for tobacco demonstrates a concentrated distribution that is significantly higher than the profitability of other crops (Figure 2). Notably, while corn does not offer a distinct profitability advantage as a standalone crop, it is frequently intercropped with tobacco, wheat, and other crops, making it a common choice among tobacco farmers. In fact, all farmers in the sample reported growing corn alongside their tobacco crops. Besides, fruits and medicinal herbs showed higher unadjusted per-mu income figures. However, it is essential to consider that these crops typically have longer growth cycles compared to tobacco and other seasonal crops. Therefore, the comparative results presented in the figure are reasonable and provide valuable insight into the income dynamics of the farming community.

It is clear that tobacco is the most profitable crop for farmers in the region, representing a crucial component of their overall income. This underscores the importance of tobacco cultivation in the agricultural landscape, as it not only contributes significantly to farmers' revenue but also plays a vital role in their economic stability and livelihood.

3.2 Factors affecting income and quality of life

3.2.1 SHAP analysis for tobacco income

Figure 3 presents a violin summary plot of the top 10 variables influencing average household income per person (Y), based on SHAP values from the XGBoost regressor. This plot provides a comprehensive visualization of each variable's contribution to income predictions, showing both the magnitude and distribution of their impacts across households.

The analysis highlights 'Farm Size (Mu)' as the most influential variable, with a substantial impact on income predictions in both positive and negative directions. The SHAP values reveal that larger farm sizes generally contribute positively to income, while smaller farm sizes (represented by blue values) are associated with negative income predictions. This aligns with expected trends, as larger cultivation areas typically generate higher revenue.

Another significant factor is 'Family Population', which shows a negative association with income levels. The SHAP plot indicates that larger household sizes may exert financial pressure due to a higher proportion of non-working members, reducing per capita income. Additionally, geographic location plays a key role, with households in mountainous terrains experiencing a negative income impact compared to those in flatland terrains.

The plot also reveals that variables such as 'Growing Size (Corn) (Mu)' display a more balanced SHAP distribution, suggesting that their impact on income is modulated by interactions with other variables. This indicates that these factors are context-dependent, contributing to income predictions in specific scenarios.

Overall, the SHAP analysis provides valuable insights into the complex and non-linear relationships between household characteristics and income levels. It highlights both direct effects and interaction effects, offering a nuanced understanding of the socioeconomic factors shaping income disparities in tobacco-farming communities.

3.2.2 Scale economies analysis of tobacco cultivation

The absolute economic competitiveness of tobacco has been confirmed, serving as an encouragement for expanding tobacco cultivation areas. To ensure rigor, this section aims to explore whether scale efficiency exists in expanding tobacco cultivation areas and to

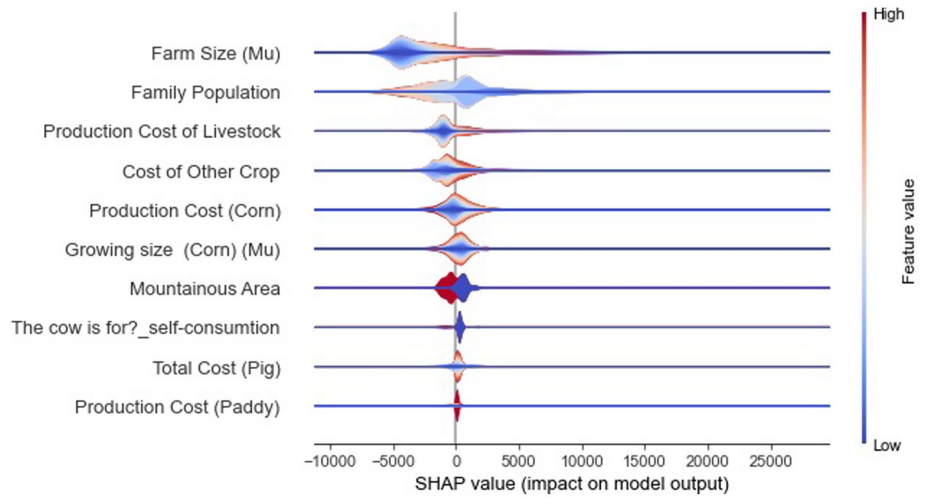

FIGURE 3
Layered violin plot for SHAP value in XGBoost regressor predicting the income.

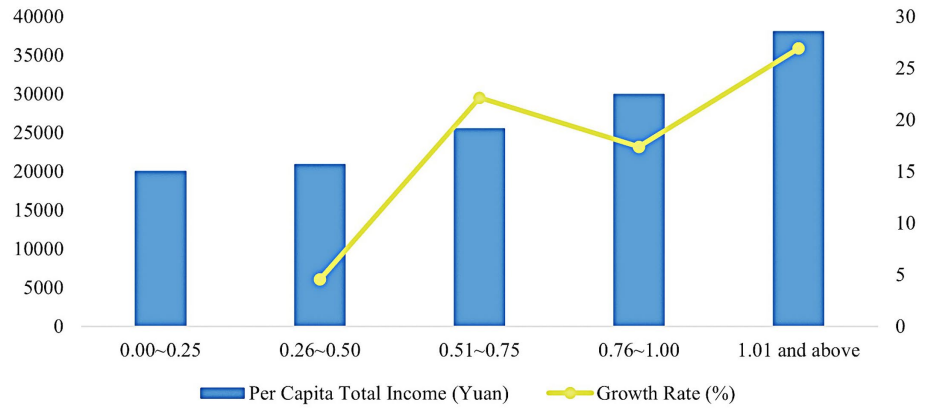

FIGURE 4
Per capita income and growth rate of tobacco-growing households at different planting scales.

what extent. Figure 4 visually presents the per capita income and its growth rate at different cultivation scales.

The horizontal axis of the following chart represents the tobacco cultivation area, the left vertical axis represents per capita total income, and the right vertical axis shows the year-on-year growth rate of per capita income at different cultivation scales. When the tobacco cultivation area increases from [0.26, 0.50] to [0.51, 0.75], the income growth rate is higher, indicating that a moderate increase in the planting area can quickly raise per capita income in the early stages. As the tobacco cultivation area grows from [0.51, 0.75] to [0.76, 1.00], the income growth rate slows down, which may be due to increased costs, resource limitations, or declining efficiency, leading to diminishing marginal returns. However, when the tobacco cultivation area increases from [0.76, 1.00] to 1.01 and beyond, the income growth rate rebounds. This could be attributed to the fact that, after reaching a certain scale, farmers may have achieved economies of scale and optimized resource allocation, thus driving a leap in income.

Given the objective resource constraints, it is recommended to maintain a moderate scale of [0.51–0.75] to enable as many households as possible to achieve favorable income levels. At the same time, it is essential to thoroughly analyze the advantages effectively leveraged and the disadvantages successfully avoided within this scale. Efforts should focus on improving technology and management practices, developing transferable experiences, and seeking higher income growth opportunities.

3.2.3 Regression analysis of benefits of tobacco cultivation

This section provides a broader analysis of the overall impact of expanding tobacco cultivation areas on farmers' livelihoods through the development of linear regression models.

The analysis indicates that increasing the area of tobacco cultivation is significantly associated with higher household education expenditure, showing a 3.8% rise for every unit increase in cultivated area. It also correlates with a reduction in Engel's coefficient by 0.124%, and an improvement in the Quality of Life Index by 0.043. This suggests that expanding tobacco cultivation may enhance household welfare, not just by boosting income, but also by shifting

TABLE 3 Regression results.

Variable	Model 1	Model 2	Model 3
	Ln (Edu)	Engel	Q
Cultivated Area	0.038***	−0.124**	0.043***
(A_i)	(2.91)	(−2.34)	(5.54)
Family Members	−0.061	0.237	0.080**
(FM_i)	(−0.87)	(0.94)	(2.18)
Terrain	0.594**	1.427	0.429***
(T_i)	(2.37)	(1.48)	(3.06)
Dali Prefecture	0.666**	3.111***	0.594***
(P_{i1})	(2.55)	(3.02)	(3.96)
Qujing Prefecture	−0.231	2.659***	−0.295**
(P_{i2})	(−1.02)	(3.26)	(−2.48)
Constant	8.338***	3.578**	6.879***
	(19.04)	(2.31)	(30.44)
Observations	152	229	229
R-squared	0.080	0.112	0.207

t-statistics in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 4 Frequency distribution of agricultural diversity among tobacco-growing households.

Diversity	Frequency	Relative frequency (%)
2	13	5.68
3	75	32.75
4	55	24.02
5	59	25.76
6	20	8.73
7	7	3.06
Total	229	100.00

family expenditure toward education and reducing the proportion of income spent on food. This reallocation of resources potentially offers more stable living conditions and long-term economic prospects. Furthermore, such cultivation might influence other economic sectors and community health, necessitating a balance between economic gains and social wellbeing.

On the other hand, the number of family members shows a notable positive correlation with the Quality of Life Index, yet a negative relationship with education expenditure. This contrast reflects how larger family sizes in Yunnan may exert financial pressure, potentially limiting investment in education. Culturally, larger families might prioritize immediate survival needs over educational investments, which could have long-term implications for future generations' economic mobility and development.

The results also reveal that households in mountainous terrains experience a significant increase in their education expenditure and Quality of Life Index compared to those in plains. This indicates that the challenging terrain, with its less fertile soil and unfavorable climate, hinders agricultural productivity, impacting the overall living standards of these households. Geographical challenges like poor soil quality, steep slopes, and adverse weather conditions make agricultural practices difficult and less profitable. Comparing the economic opportunities and living conditions of households in different terrains can highlight the need for targeted interventions to support those in less favorable environments.

Lastly, this geographical inequality is also manifested in administrative divisions. The result highlights prefectural variations in education expenditure, Engel's coefficient, and Quality of Life Index, attributed to differing administrative policies. These policies evidently play a crucial role in shaping the economic conditions of tobacco-growing households, influencing their living standards and investment in education. Regions with more supportive policies tend to show better outcomes in household welfare. Examining successful policy implementations can provide insights into best practices that can be replicated in other regions to promote balanced and sustainable development.

Overall, while the expansion of tobacco cultivation presents economic benefits, addressing regional and demographic challenges remains vital for achieving balanced and sustainable development. Understanding the interplay between cultivated area, family size, terrain, and regional policies can help in formulating comprehensive strategies that enhance household welfare and promote long-term economic stability.

3.3 Agricultural diversification status of tobacco households

The previous discussion centered on the scale of planting areas, thoroughly addressing the benefits of tobacco cultivation and its role in improving farmers' livelihoods. However, field observations reveal that the Yunnan Tobacco Monopoly Bureau (Company) actively explores intercropping and crop rotation with other grains, rather than being confined to simply expanding the scale of tobacco cultivation. Therefore, incorporating agricultural diversity and analyzing farmers' production and living conditions through the lens of planting structures offers a valuable extension to the study. Specifically, the Bureau has implemented intercropping programs, crop rotation strategies, and financial incentives for cultivating alternative crops such as legumes and medicinal plants. These initiatives aim to reduce dependency on tobacco and improve soil health and income stability.

For example, intercropping tobacco with legumes can improve soil fertility and reduce pests, while crop rotation with grains can enhance soil structure and break pest cycles. These practices not only benefit the environment but also provide farmers with multiple income streams, reducing the risks associated with dependence on a single crop. The "tobacco + N" industrial complex model integrates tobacco cultivation with other agricultural activities and processing industries, creating a more resilient agricultural economy. This model includes livestock rearing, fruit and vegetable production, and the establishment of agro-processing units, all of which contribute to increased household income and employment opportunities.

Table 4 illustrates the frequency distribution of the agricultural diversity among tobacco-growing households.

In the sample, 82.53% of tobacco farmers have an agricultural diversity score ranging from 3 to 5. Specifically, 32.75% of the farmers

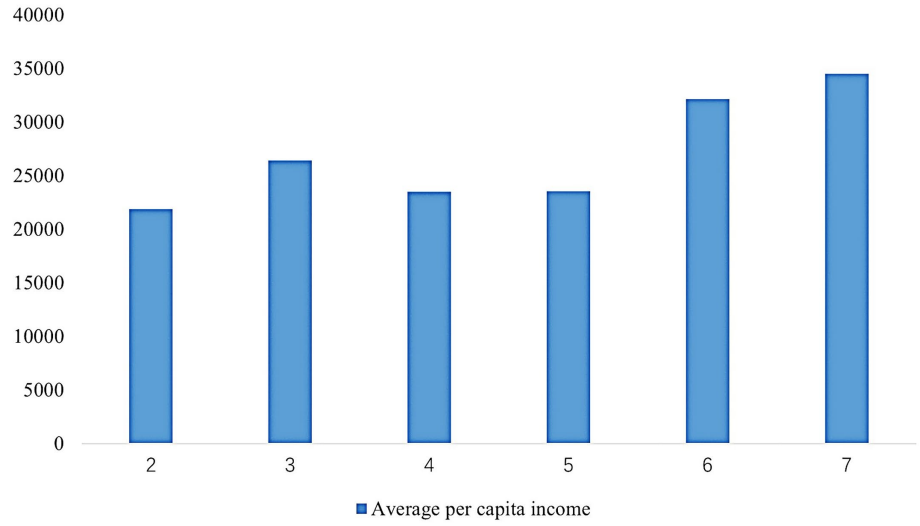

FIGURE 5
Average per capita income of tobacco-growing households under different levels of agricultural diversity.

exhibit an agricultural diversity score of 3, which typically involves cultivating tobacco and corn while raising pigs, accounting for 70.31% of the sample. This particular agricultural structure yields the highest per capita income among those with a diversity score of 5 or less. As agricultural diversity increases to 6 or above, there is a noticeable leap in per capita income. However, only a small number of farmers achieve this level of diversity, possibly due to the substantial financial and labor resources required. See Figure 5 for details.

It is advisable for farmers to consider adjusting their agricultural structures to promote greater diversity in their farming practices. Specifically, farmers are encouraged to aim for an agricultural diversity score of 6, which would involve integrating six or more types of crops and livestock into their operations. This diversified approach can significantly enhance income stability and bolster overall economic resilience (Altieri, 1999; Lin, 2011).

However, for households with fewer members or limited resources, achieving such a high level of diversity may be challenging. In these cases, maintaining a minimum diversity score of 3 is recommended. For instance, farmers could focus on cultivating a combination of tobacco and corn, which are known to complement each other by improving soil nutrition. Additionally, incorporating livestock, such as cows, can further benefit soil health through natural fertilization and provide additional sources of income.

In addition to economic considerations, tobacco cultivation poses significant externalities. The intensive use of pesticides and fertilizers contributes to soil degradation and water pollution. The curing process generates air pollutants, and the end product—tobacco—has well-documented health risks. Furthermore, market volatility and regulatory shifts can destabilize farmer incomes, underscoring the need for resilient and diversified agricultural systems. By fostering agricultural diversity, farmers can create a more sustainable and resilient farming system that not only supports their livelihoods but also contributes to the health of their ecosystems.

4 Conclusion

This study provides a comprehensive analysis of the role of tobacco cultivation in the income structure of farming households in Yunnan Province, based on survey data from 253 households. The findings underscore tobacco's significant contribution to household income, reaffirming its economic competitiveness compared to other crops. Descriptive statistics demonstrate that tobacco income constitutes a substantial portion of farmers' earnings, particularly in regions where natural conditions favor its cultivation.

Further analysis using both general linear regression and interpretable machine learning models identifies key factors influencing household income and quality of life. The results indicate that larger areas dedicated to tobacco farming generally lead to higher incomes, though the marginal benefits vary across households depending on farm size, household composition, and geographic factors. The study highlights the broader socioeconomic impacts of tobacco cultivation, including its positive effects on living standards and rural economic stability.

The research also examines the role of agricultural diversification in enhancing household resilience. Initiatives by the Tobacco Bureau to encourage crop diversification show potential in improving overall farm productivity and reducing income risks associated with reliance on a single cash crop. However, the study identifies persistent challenges, such as limited access to resources and localized geographic disadvantages, that may constrain farmers' diversification efforts. Addressing these barriers is crucial to optimizing the economic and environmental benefits of diversification.

Overall, this study highlights that tobacco cultivation is an essential economic activity for farmers in Yunnan Province. However, attaining long-term sustainability necessitates a strategic focus on agricultural diversification. By optimizing planting structures, tobacco farmers can continually increase their incomes. Policymakers and agricultural stakeholders must assist farmers in adapting their cropping systems and enhancing farm management practices to strike a balance between income generation and resilience. By

promoting diversification and sustainable farming methods, rural communities in Yunnan can more effectively address the challenges posed by an evolving agricultural landscape while protecting their livelihoods, ultimately achieving sustainable development in tobacco agriculture.

Data availability statement

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

Ethics statement

This study has been approved by the Ethics Committee of China Yunnan Tobacco Import and Export Co., Ltd. The tobacco farmers who participated in the survey have signed the informed consent forms.

Author contributions

JS: Resources, Project administration, Writing – review & editing, Writing – original draft. JW: Writing – review & editing, Resources. XL: Resources, Writing – review & editing. QN: Writing – review & editing, Resources. WL: Writing – review & editing, Resources. ZL: Resources, Writing – review & editing. GY: Writing – review & editing, Visualization. HH: Writing – review & editing, Formal analysis. YG: Formal analysis, Writing – review & editing. YZ: Writing – original draft, Formal analysis, Methodology, Validation. YB: Conceptualization, Project administration, Supervision, Writing – review & editing, Funding acquisition. GW: Funding acquisition, Project administration, Supervision, Writing – review & editing, Conceptualization. DZ: Writing – original draft, Validation, Methodology.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by the General Scientific and Technological Projects of China Tobacco Yunnan Import and Export Co., Ltd. [Grant no. KJXM/YNYC(JCK)2023001].

Acknowledgments

The authors would like to express thanks to the reviewers and editors for their invaluable contributions to the improvement of this paper.

Conflict of interest

JS, JW, XL, QN, and WL were employed by the China Tobacco Yunnan Import and Export Co., Ltd. ZL was employed by the Yunnan Tobacco Company Qujing City Company Luliang Branch. YG and YZ were employed by the Demetech Limited.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from China Tobacco Yunnan Import and Export Co., Ltd. The funder was involved in the study by providing insights into the study design and making the decision to submit it for publication.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Abbas, A., Zhao, C., Waseem, M., Ahmed Khan, K., and Ahmad, R. (2022). Analysis of energy input–output of farms and assessment of greenhouse gas emissions: a case study of cotton growers. *Front. Environ. Sci.* 9:826838. doi: 10.3389/fenvs.2021.826838

Achterbosch, T., Van Berkum, S., and Meijerink, G. (2014). Cash crops and food security: Contributions to income, livelihood risk and agricultural innovation. LEI Report, 15, 57.

Altieri, M. A. (1999). "The ecological role of biodiversity in agroecosystems" in Invertebrate biodiversity as bioindicators of sustainable landscapes. ed. M. G. Paoletti (Amsterdam: Elsevier), 19–31.

Aydan, S., Donar, G. B., and Arikan, C. (2022). Impacts of economic freedom, health, and social expenditures on well-being measured by the better life index in OECD countries. *Soc. Work Public Health* 37, 435–447. doi: 10.1080/19371918.2021.2018083

Bondurant, S., Wallace, R., Shetty, P., and Stratton, K. (2001). Clearing the smoke: assessing the science base for tobacco harm reduction. Available online at: https://books.google.com/books?hl=en&lr=&id=pT6tE0GfgbQC&oi=fnd&pg=PT20&dq=the+pollution+generated+during+the+tobacco+curing+process+and+the+well-documented+health+risks+associated+with+smoking+have+become+more+widely+recognized+by+the+public&ots=_bV_WSPWhJ&sig=GOFtO28EwZgY_D2QVXF7JFhqCoY (accessed June 21, 2001).

Chapman, B. (2016). The knowledge capital of nations: education and the economics of growth. *Econ. Rec.* 92, 667–669. doi: 10.1111/1475-4932.12298

De Moura Pereira, D. A., Dos Santos, M., De Araujo Costa, I. P., Lellis Moreira, M. A., Terra, A. V., Rocha Junior, C. D. S., et al. (2022). Multicriteria and statistical approach to support the outranking analysis of the OECD countries. *IEEE Access* 10, 69714–69726. doi: 10.1109/ACCESS.2022.3187001

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., et al. (2023). Explainable AI (XAI): core ideas, techniques, and solutions. *ACM Comput. Surv.* 55, 1–33. doi: 10.1145/3561048

Feng, Y., and Ji, Q. (2010). Macroeconomic development trends of tobacco during the "Twelfth Five-Year Plan" and strategies for Yunnan tobacco. Exploration of *Economic Issues*, (in Chinese) 151–156.

Koronakos, G., Smirlis, Y., Sotiros, D., and Despotis, D. K. (2020). Assessment of OECD better life index by incorporating public opinion. *Socio Econ. Plan. Sci.* 70:100699. doi: 10.1016/j.seps.2019.03.005

Lin, B. B. (2011). Resilience in agriculture through crop diversification: adaptive management for environmental change. *Bioscience* 61, 183–193. doi: 10.1525/bio.2011.61.3.4

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Int.* 2, 56–67. doi: 10.1038/s42256-019-0138-9

Luo, L., Luo, C., Qiao, Z., and Yi, M. (2015). Analysis of modern tobacco agriculture development in Yunnan. *Ecological Economy, (in Chinese)* 31, 130–134.

Peng, J., Ma, M., Bao, X., and Liu, J. (2000). Problems and countermeasures facing the development of Yunnan tobacco industry in the 21st century. Exploration of *Economic Issues*, (in Chinese) 125–128.

Rahman, M. S., Ahmed, N. A. M. F., Ali, M., Abedin, M. M., and Islam, M. S. (2020). Determinants of tobacco cultivation in Bangladesh. *Tob. Control.* 29, 692–694. doi: 10.1136/tobaccocontrol-2019-055167

Resce, G., and Maynard, D. (2018). What matters most to people around the world? Retrieving better life index priorities on twitter. *Technol. Forecast. Soc. Change* 137, 61–75. doi: 10.1016/j.techfore.2018.06.044

Song, W., Elahi, E., Hou, G., and Wang, P. (2025). Collaborative governance for urban waste management: a case study using evolutionary game theory. *Susadd then. Cities Soc.* 126:106380. doi: 10.1016/j.scs.2025.106380

Tai, L., and Yang, H. (2024). Evolution characteristics and influencing factors of the human-earth system in minority areas of Yunnan. *China. J. Resour. Ecol*, 15, 372–384.

Wan, X., Jin, J., Ran, S., Guan, T., and Qiu, X. (2022). Are farmers willing to substitute tobacco cultivation? Evidence from Lichuan City, China. *Nicotine Tob. Res.* 24, 897–903. doi: 10.1093/ntr/ntab264

World Health Organization (2017). Tobacco and its environmental impact: An overview. Geneva: World Health Organization.

Appendix

The detailed account of the construction process for the Quality of Life Index Q are provided. Table A1 presents the indicator weights determined through the Analytic Hierarchy Process (AHP), based on literature analysis and thematic discussions.

TABLE A1 Indicator weight vector

Indicator	Weight
Transportation and communication expenses	0.06
Education expenses	0.28
Entertainment expenses	0.22
Housing expenses	0.16
Household goods and services expenses	0.11
Food expenses	0.07
Tobacco, tea, and sugar expenses	0.05
Clothing expenses	0.03
Healthcare expenses	0.02
Total expenses	1.00

The consistency ratio (CR) is calculated to be approximately 0.049, which is less than 0.1, indicating that the consistency of the judgment matrix is good.

To construct a comprehensive evaluation index Q to measure the quality of life level of every tobacco-growing household in this study, the formula for the indicator construction is as follows:

$$Q = \ln\left(\sum_{i=1}^{n} w_i x_i\right),$$

where i is indicator. The w_i represents the weight of indicator, satisfying the condition $\sum_{i=1}^{n} w_i = 1$. The x_i represents the quantity. Given that the marginal benefit decreases with an increase in individual expenditure, the logarithm is applied to adjust for this.