



OPEN ACCESS

EDITED BY

Massimiliano Bonamente,
University of Alabama in Huntsville, United States

REVIEWED BY

Takuya Yamano,
Kanagawa University, Japan
Mohd Azmi Haron,
University of Malaya, Malaysia

*CORRESPONDENCE

Shamsul Rijal Muhammad Sabri*
✉ rjal@usm.my

RECEIVED 07 July 2025

ACCEPTED 16 October 2025

PUBLISHED 06 November 2025

CITATION

Sabri SRM and AL Hourani MAM (2025)
Inference on the scale-inflated gamma
distribution applied to Malaysian household
income. *Front. Appl. Math. Stat.* 11:1660916.
doi: 10.3389/fams.2025.1660916

COPYRIGHT

© 2025 Sabri and AL Hourani. 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.

Inference on the scale-inflated gamma distribution applied to Malaysian household income

Shamsul Rijal Muhammad Sabri* and
Mallak Ahmad Mohammad AL Hourani

School of Mathematical Sciences, Universiti Sains Malaysia, USM, Pulau Pinang, Malaysia

Modeling income distributions is crucial for understanding inequality and providing evidence-based policy support. A key challenge, however, lies in evaluating the extent to which household income inflates over time. While income is inherently random, it exhibits a persistent upward trend, and fitting income distributions using conventional models often leads to inconsistent parameter estimates. This highlights the necessity of explicitly incorporating inflation-adjusted scaling to preserve proper statistical properties. To address this gap, we introduce the Scale-Inflated Gamma (SIG) distribution, which extends the standard Gamma distribution by including an inflation-adjusted scale parameter (δ), thereby providing greater flexibility in capturing heterogeneous income dynamics. Standard models such as the Lognormal, Pareto, or Generalized Beta of the Second Kind (GB2) systematically underestimate upper-tail incomes and fail to capture inflation-adjusted heterogeneity across subgroups (B40, M40, T20). The SIG model, in contrast, strikes a balance between parsimony and flexibility by directly adjusting for inflationary scale shifts. For instance, while the Gamma distribution underestimates the 95th percentile by 10%–12% in 2019, the SIG model reduces this bias to approximately 3%, accurately reflecting income dynamics across B40, M40, and T20 groups. We develop the theoretical foundations of the SIG distribution by deriving its probability density function (PDF), cumulative distribution function (CDF), and moments. Parameters are initially estimated using the method of moments and then refined through maximum likelihood estimation (MLE). To assess estimator precision, we derive the Fisher information matrix, using the inverse Hessian to approximate the variance–covariance matrix, thus ensuring reliable inference. A Monte Carlo simulation study is conducted to evaluate the consistency and efficiency of the estimators under various sample sizes. The SIG model is subsequently applied to Malaysian Household Income Survey (HIS) data spanning the period from 2007 to 2022. Results demonstrate that the SIG distribution offers a superior fit for modeling income inequality and upper-tail behavior compared to conventional models. Overall, the study establishes the SIG distribution as a theoretically robust and policy-relevant framework for analyzing income patterns in inflation-sensitive and structurally diverse economies.

KEYWORDS

Scale-Inflated Gamma (SIG) distribution, income distribution, upper-tail modeling, maximum likelihood estimation, fisher information, Malaysian household income survey

1 Introduction

Rising income inequality has become a critical concern in Malaysia over the past two decades, particularly as disparities between the B40, M40, and T20 groups have widened [1, 4]. These disparities stem not only from structural economic changes but also from persistent inflation, which erodes purchasing power unevenly across income segments [5, 6]. Such dynamics underscore the necessity for statistical tools capable of accurately representing income distributions under inflationary pressures, especially to inform targeted policy interventions [3].

Classical income distribution models, such as the Lognormal and Gamma distributions, have long been favored for their tractability in estimation and interpretation [7, 8]. However, they systematically underestimate upper-tail incomes and fail to accommodate inflation-adjusted shifts in the distribution. More flexible models like GB2 address some of these limitations, yet they often introduce complexity that impedes interpretability in applied policy settings [9]. In light of these challenges, scale-augmented approaches that adjust for inflation and heterogeneity have garnered increasing attention [10, 11]. A critical gap remains in evaluating how much household income increases over time. Although income is inherently random, it exhibits a persistent upward trend, and fitting income using conventional distributions without explicitly incorporating inflationary effects often leads to inconsistent parameter estimates and biased statistical properties [1, 2].

This gap is especially evident in upper-tail modeling: while conventional models consistently underestimate high-income observations, particularly in the upper 95th percentile, the proposed SIG model captures these dynamics with substantially greater accuracy, thereby offering a more faithful representation of income inequality. To fill this gap, we propose the Scale-Inflated Gamma (SIG) distribution. By introducing an inflation-adjusted scale parameter (δ), SIG flexibly models upper-tail behavior while preserving the familiar shape properties of the Gamma distribution. Visual evidence, via boxplots and kernel density estimates, reveals progressively heavier upper tails in Malaysian household income distributions from 2007 to 2022 (Figures 1, 2), illustrating the inadequacy of conventional models under inflationary dynamics.

This study applies the Scale-Inflated Gamma (SIG) distribution to Malaysian Household Income Survey (HIS) data from 2007 to 2022 to assess its performance relative to conventional alternatives. We estimate parameters via maximum likelihood estimation (MLE) and assess model fit using information criteria such as AIC and BIC. Additionally, we utilize visual tools (e.g., empirical cumulative distribution function plots, boxplots) and statistical diagnostics to validate the tail behavior modeling. The results position the SIG distribution as a theoretically grounded and practically relevant framework for analyzing income distributions under inflationary conditions, with implications for measuring inequality, monitoring subgroup dynamics, and informing policy design in Malaysia and similar economies.

In addition to the theoretical formulation, we conduct formal model comparisons using likelihood-based hypothesis testing and model selection metrics such as AIC and BIC. While our empirical focus centers on comparing the SIG model with the conventional Gamma distribution, we also situate our findings

within the broader literature. Specifically, prior studies have shown that the Lognormal distribution tends to underestimate tail inequality [6], and the Pareto distribution is suited only for modeling the upper tail. It lacks inflation adjustment [4], and the GB2 distribution, though flexible, suffers from interpretability challenges in policy contexts [9]. Moreover, Yu et al. [12] highlight the advantages of using information criteria, such as AIC and BIC, for distributional comparison, which supports our evaluation framework. Collectively, our findings confirm that the SIG model offers a statistically robust and practically relevant alternative for modeling inflation-adjusted income dynamics and inequality.

2 General properties of the scale-inflated gamma distribution

This section formally introduces the Scale-Inflated Gamma (SIG) distribution by extending the conventional Gamma model to account for inflation-adjusted scaling across time. The SIG framework is governed by three primary parameters: the shape parameter $\alpha > 0$, the baseline scale parameter $\theta > 0$, and the scale-inflation parameter $\delta > -1$. The inflation-adjusted scale parameter is denoted by $\beta = (1 + \delta)^k \theta$, where $k \in \mathbb{N}$ represents the number of time periods since the baseline. This formulation enables the model to dynamically adjust the scale of the income distribution over time, effectively capturing the erosion of purchasing power due to inflation. Each parameter plays a distinct role: α controls the skewness and dispersion of the distribution, θ anchors the scale in the base period, and δ modulates how the scale evolves. These structural components jointly enable the SIG distribution to model income dynamics under inflationary regimes with flexibility. The following subsections develop the full probabilistic specification of the model, including its probability density function (PDF), log-likelihood function, gradient vector, and Fisher Information Matrix.

Let $X \sim F_X(\alpha, \beta)$ be a random variable with mean $\mathbb{E}(X | \beta)$ and variance $\text{Var}(X | \beta)$, where $\alpha > 0$ and $\beta > 0$ denote the shape and scale parameters, respectively. These parameters can be written in vector form as $\varepsilon = (\alpha, \beta)'$, where the prime symbol ($'$) denotes the transpose of a vector. Inflationary scaling is introduced through the transformation

$$Y = (1 + \delta)^k X, \quad \delta > -1, k \in \mathbb{N},$$

where δ is the scale-inflation parameter and k is the number of periods. The extended parameter set is expressed as $\nu = (\theta, \delta)'$, so that Y follows the scale-inflated distribution

$$Y \sim F_Y(\alpha, \theta, \delta) \equiv F_X[\alpha, \beta(\nu)],$$

with $\beta = (1 + \delta)^k \theta$, and the complete parameter vector is denoted as $\lambda = (\alpha, \theta, \delta)'$, where the prime symbol ($'$) indicates vector transposition.

Given the distribution functions of X , with cumulative distribution function (CDF) $F_X(x; \varepsilon)$ and probability density function (PDF) $f_X(x; \varepsilon)$, the corresponding functions for Y are written as

$$F_Y(y; \alpha, \theta, \delta) \equiv F_Y[y; \alpha, \beta(\nu)], \quad f_Y(y; \alpha, \theta, \delta) \equiv f_Y[y; \alpha, \beta(\nu)],$$

for $X > 0$, $Y > 0$.

For the log-PDF of X , denoted $\ln f_X(x; \varepsilon)$, the score functions (first derivatives) are defined as

$$g_{\varepsilon_i}(x) = \frac{\partial}{\partial \varepsilon_i} \ln f_X(x; \varepsilon). \quad (2.1)$$

For a complete random sample, the unbiasedness condition holds: $\mathbb{E}[g_{\varepsilon_i}(X)] = 0$, which follows from standard likelihood theory [13].

The second derivatives (Hessian components) of the log-PDF of X are given by

$$h_{\varepsilon_i \varepsilon_j}(x) = \frac{\partial^2}{\partial \varepsilon_i \partial \varepsilon_j} \ln f_X(x; \varepsilon). \quad (2.2)$$

For the inflated variable Y , the log-PDF has derivatives

$$g_{\varepsilon_i}(y) = \frac{\partial}{\partial \varepsilon_i} \ln f_Y(y; \varepsilon), \quad (2.3)$$

and, defining $g_{\beta}(y) = \frac{\partial}{\partial \beta} \ln f_Y(y; \varepsilon)$, the score functions with respect to $v = (\theta, \delta)'$ are

$$g_{v_i}(y) = g_{\beta}(y) \cdot \frac{\partial \beta}{\partial v_i}. \quad (2.4)$$

The corresponding Hessian terms are obtained as

$$h_{\varepsilon_i \varepsilon_j}(y) = \frac{\partial}{\partial \varepsilon_i} g_{\varepsilon_j}(y), \quad (2.5)$$

and for $v = (\theta, \delta)'$, where the prime symbol ($'$) denotes the transpose of a vector,

$$h_{v_i v_j}(y) = h_{\beta \beta}(y) \frac{\partial \beta}{\partial v_i} \frac{\partial \beta}{\partial v_j} + h_{\beta}(y) \frac{\partial^2 \beta}{\partial v_i \partial v_j}. \quad (2.6)$$

Finally, since the SIG distribution involves both the shape parameter α and the inflation parameters $v = (\theta, \delta)'$, mixed derivatives are given by

$$h_{\alpha v_i}(y) = h_{\alpha \beta}(y) \frac{\partial \beta}{\partial v_i} + h_{\beta}(y) \frac{\partial^2 \beta}{\partial \alpha \partial v_i}. \quad (2.7)$$

Throughout this section, the symbols $g_{\alpha}, g_{\theta}, g_{\delta}$ are used to denote the score functions, i.e., the first-order derivatives of the log-likelihood, while $h_{\alpha \theta}, h_{\theta \delta}$ represent the corresponding Hessian components, i.e., the second-order derivatives. The sample sizes are denoted by n_X for the baseline distribution and n_Y for the inflated distribution. Equations 2.1–2.7 establish the score functions and curvature structure of the Scale-Inflated Gamma (SIG) distribution, providing the analytical foundation for maximum likelihood estimation (MLE) and asymptotic inference. These expressions enable the systematic derivation of the gradient and Hessian matrices, which in turn allow the construction of the Fisher Information Matrix (FIM). The latter forms the basis for parameter estimation, variance–covariance analysis, and hypothesis testing within the SIG framework, and will be developed explicitly in the subsequent subsection.

2.1 Curvature and information structure

By construction, the expectation of the score functions vanishes, $\mathbb{E}[g(Y)] = 0$. Hence, the expected second derivatives are summarized as

$$\mathbb{E}[h_{\varepsilon_i \varepsilon_j}(Y)] = \mathbb{E}\left[\frac{\partial}{\partial \varepsilon_i} g_{\varepsilon_j}(Y)\right], \quad (2.8)$$

$$\mathbb{E}[h_{v_i v_j}(Y)] = \mathbb{E}[h_{\beta \beta}(Y)] \frac{\partial \beta}{\partial v_i} \frac{\partial \beta}{\partial v_j}, \quad (2.9)$$

$$\mathbb{E}[h_{\alpha v_i}(Y)] = \mathbb{E}[h_{\alpha \beta}(Y)] \frac{\partial \beta}{\partial v_i}. \quad (2.10)$$

These expectations capture the curvature of the log-likelihood under the Scale-Inflated Gamma (SIG) distribution, and serve as the analytical foundation for subsequent likelihood-based inference. The use of digamma and trigamma functions follows the classical treatment of special functions provided in Abramowitz and Stegun [14].

2.2 Log-likelihood and gradient structure

Let $X \sim F(\alpha, \theta)$ and $Y \sim F[\alpha, \beta(v)]$, then for a sample size n_X from X and n_Y from Y , the log-likelihood function is:

$$\ell_{XY}(\alpha, \theta, \delta) = \sum_{i=1}^{n_X} \ln f_X(x_i; \alpha, \theta) + \sum_{i=1}^{n_Y} \ln f_Y(y_i; \alpha, \beta(v)). \quad (2.11)$$

The total gradient (a 3×1 vector) from X and Y is:

$$G = G_X + G_Y = \begin{bmatrix} \sum_{i=1}^{n_X} g_{\alpha}(x_i) \\ \sum_{i=1}^{n_X} g_{\theta}(x_i) \\ 0 \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{n_Y} g_{\alpha}(y_i) \\ \sum_{i=1}^{n_Y} g_{\theta}(y_i) \\ \sum_{i=1}^{n_Y} g_{\delta}(y_i) \end{bmatrix}. \quad (2.12)$$

These gradient expressions highlight how information is accumulated from both baseline and inflated samples, allowing consistent parameter estimation through maximum likelihood.

2.3 Hessian and fisher information

The total Hessian matrix (3×3) is:

$$H = H_X + H_Y = \begin{bmatrix} \sum h_{\alpha \alpha}(x_i) & \sum h_{\alpha \theta}(x_i) & 0 \\ \sum h_{\alpha \theta}(x_i) & \sum h_{\theta \theta}(x_i) & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \sum h_{\alpha \alpha}(y_i) & \sum h_{\alpha \theta}(y_i) & \sum h_{\alpha \delta}(y_i) \\ \sum h_{\alpha \theta}(y_i) & \sum h_{\theta \theta}(y_i) & \sum h_{\theta \delta}(y_i) \\ \sum h_{\alpha \delta}(y_i) & \sum h_{\theta \delta}(y_i) & \sum h_{\delta \delta}(y_i) \end{bmatrix}. \quad (2.13)$$

The Fisher Information Matrix (FIM) plays a pivotal role in formulating the SIG distribution because it determines the variances and covariances of parameter estimates. Following Fisher

[15], the FIM is obtained using the negative expectation of the Hessian:

$$I_{XY}(\alpha, \theta, \delta) = \mathbb{E}(-H)$$

$$= n_X \begin{bmatrix} \mathbb{E}[-h_{\alpha\alpha}(X)] & \mathbb{E}[-h_{\alpha\theta}(X)] & 0 \\ \mathbb{E}[-h_{\alpha\theta}(X)] & \mathbb{E}[-h_{\theta\theta}(X)] & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$+ n_Y \begin{bmatrix} \mathbb{E}[-h_{\alpha\alpha}(Y)] & \mathbb{E}[-h_{\alpha\beta}(Y)] \frac{\partial \beta}{\partial \theta} & \mathbb{E}[-h_{\alpha\beta}(Y)] \frac{\partial \beta}{\partial \delta} \\ \mathbb{E}[-h_{\alpha\beta}(Y)] \frac{\partial \beta}{\partial \theta} & \mathbb{E}[-h_{\beta\beta}(Y)] \left(\frac{\partial \beta}{\partial \theta} \right)^2 & \mathbb{E}[-h_{\beta\beta}(Y)] \frac{\partial \beta}{\partial \theta} \frac{\partial \beta}{\partial \delta} \\ \mathbb{E}[-h_{\alpha\beta}(Y)] \frac{\partial \beta}{\partial \delta} & \mathbb{E}[-h_{\beta\beta}(Y)] \frac{\partial \beta}{\partial \theta} \frac{\partial \beta}{\partial \delta} & \mathbb{E}[-h_{\beta\beta}(Y)] \left(\frac{\partial \beta}{\partial \delta} \right)^2 \end{bmatrix}. \quad (2.14)$$

This decomposition clarifies how baseline and inflated observations contribute separately to the information structure. Similar approaches have been used in the analysis of gamma-type models [9].

2.4 Extension to multiple periods

For a system consisting of the base year dataset X and m datasets $Y_1, \dots, Y_m$, the log-likelihood becomes:

$$\ell_{XY_1 \dots Y_m}(\alpha, \theta, \delta) = \ell_X(\alpha, \theta) + \sum_{j=1}^m \ell_{Y_j}(\alpha, \theta, \delta)$$

$$= \sum_{i=1}^{n_X} \ell_X(x_i; \alpha, \theta) + \sum_{j=1}^m \sum_{i=1}^{n_Y} \ell_Y(y_{ij}; \alpha, \theta, \delta). \quad (2.15)$$

This extension generalizes the likelihood across multiple inflation-adjusted periods, which is crucial for empirical applications such as household income dynamics under inflationary regimes.

$$I_{XY_1 \dots Y_m}(\alpha, \theta, \delta) = I_X + \sum_{j=1}^m I_{Y_j}$$

$$= n_X \begin{bmatrix} \mathbb{E}[-h_{\alpha\alpha}(X)] & \mathbb{E}[-h_{\alpha\theta}(X)] & 0 \\ \mathbb{E}[-h_{\alpha\theta}(X)] & \mathbb{E}[-h_{\theta\theta}(X)] & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$+ \sum_{j=1}^m n_{Y_j} \begin{bmatrix} \mathbb{E}[-h_{\alpha\alpha}(Y_j)] & \mathbb{E}[-h_{\alpha\beta}(Y_j)] \frac{\partial \beta}{\partial \theta} & \mathbb{E}[-h_{\alpha\beta}(Y_j)] \frac{\partial \beta}{\partial \delta} \\ \mathbb{E}[-h_{\alpha\beta}(Y_j)] \frac{\partial \beta}{\partial \theta} & \mathbb{E}[-h_{\beta\beta}(Y_j)] \left(\frac{\partial \beta}{\partial \theta} \right)^2 & \mathbb{E}[-h_{\beta\beta}(Y_j)] \frac{\partial \beta}{\partial \theta} \frac{\partial \beta}{\partial \delta} \\ \mathbb{E}[-h_{\alpha\beta}(Y_j)] \frac{\partial \beta}{\partial \delta} & \mathbb{E}[-h_{\beta\beta}(Y_j)] \frac{\partial \beta}{\partial \theta} \frac{\partial \beta}{\partial \delta} & \mathbb{E}[-h_{\beta\beta}(Y_j)] \left(\frac{\partial \beta}{\partial \delta} \right)^2 \end{bmatrix}. \quad (2.16)$$

Finally, the variance-covariance matrix of the estimators α , θ , and δ is obtained from the inverse of the Fisher Information Matrix, ensuring asymptotic normality of the maximum likelihood estimators.

3 Gamma distribution: preliminaries and scale-inflated extension

Let X be Gamma distributed with shape parameter α and scale parameter θ , then X has CDF denoted as

$$F_X(x; \alpha, \theta) = \Gamma(x; \alpha, \theta).$$

The PDF of X is defined as

$$f_X(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)} \left(\frac{1}{\theta} \right)^\alpha x^{\alpha-1} e^{-x/\theta}, \quad (3.1)$$

where $x > 0$. The above distribution has the mean $\alpha\theta$ and the variance $\alpha\theta^2$. Halliwell [16] documented the properties of the Log-Gamma distribution that can lead to the derivation of the Fisher Information Matrix of this distribution.

By defining the log function of X , $U = \ln(X)$, it follows that U is well-defined only for $X > 0$, which is consistent with the support of the Gamma distribution. As documented by Halliwell [16], we obtain the mean of U as:

$$\mathbb{E}[\ln X] \equiv \mathbb{E}[U] = \ln \theta + \psi(\alpha). \quad (3.2)$$

where $\psi(\alpha)$ denotes the Digamma function.

Furthermore, the expected function of $X \ln X$ is equivalent to the term Ue^U , and thus we derive:

$$\mathbb{E}[X \ln X] \equiv \mathbb{E}[Ue^U] = \alpha\theta[\ln \theta + \psi(\alpha + 1)]. \quad (3.3)$$

The log PDF of X is:

$$\ln f(x; \alpha, \theta) = -\ln \Gamma(\alpha) - \alpha \ln \theta + (\alpha - 1) \ln x - \frac{x}{\theta}. \quad (3.4)$$

The first derivatives of Equation 3.4 are:

$$g_\alpha(x) = -\psi(\alpha) - \ln \theta + \ln x, \quad (3.5)$$

$$g_\theta(x) = -\frac{\alpha}{\theta} + \frac{x}{\theta^2}. \quad (3.6)$$

From Equation 3.2, it follows that $\mathbb{E}[g_\alpha(X)] = 0$. Furthermore, since X has mean $\alpha\theta$, it also holds that $\mathbb{E}[g_\theta(X)] = 0$.

We denote the Digamma function of α as $\psi(\alpha) = \frac{\partial}{\partial \alpha} \ln \Gamma(\alpha)$. From Equation 3.2, it directly follows that $\mathbb{E}[l_\alpha(X)] = 0$. Moreover, since X has mean $\alpha\theta$, we also obtain $\mathbb{E}[l_\theta(X)] = 0$ as established in Casella and Berger [13].

We denote the score functions with respect to the parameters α and θ as $l_\alpha(X) = \frac{\partial \ln f_X(X; \alpha, \theta)}{\partial \alpha}$ and $l_\theta(X) = \frac{\partial \ln f_X(X; \alpha, \theta)}{\partial \theta}$, respectively. The second derivatives of the log-PDF are given by

$$h_{\alpha\alpha}(x) = -\psi'(\alpha), \quad (3.7)$$

$$h_{\theta\theta}(x) = \frac{\alpha}{\theta^2} - \frac{2x}{\theta^3}, \quad (3.8)$$

$$h_{\alpha\theta}(x) = h_{\theta\alpha}(x) = -\frac{1}{\theta}. \quad (3.9)$$

Here, $\psi'(\alpha) = \frac{\partial^2}{\partial \alpha^2} \ln \Gamma(\alpha)$ denotes the Trigamma function.

As shown in Kleiber and Kotz [9], the Fisher information of the Gamma distribution can be derived

by evaluating the expectations of the second derivatives of the log-PDF.

$$\begin{aligned}\mathbb{E}[h_{\alpha\alpha}(X)] &= -\psi'(\alpha), \quad \mathbb{E}[h_{\theta\theta}(X)] = -\frac{\alpha}{\theta^2}, \quad \mathbb{E}[h_{\alpha\theta}(X)] \\ &= \mathbb{E}[h_{\theta\alpha}(X)] = -\frac{1}{\theta}.\end{aligned}$$

Now, let the scale parameter β be defined to incorporate inflationary adjustment through $\delta > -1$ across $k \in \mathbb{N}$ periods:

$$\beta = (1 + \delta)^k \theta.$$

This function is differentiable with respect to both θ and δ , with first derivatives

$$\frac{\partial \beta}{\partial \theta} = (1 + \delta)^k, \quad \frac{\partial \beta}{\partial \delta} = k(1 + \delta)^{k-1} \theta.$$

Then, as defined earlier, if $Y = (1 + \delta)^k X$, then $Y \sim \text{Gamma}(\alpha, (1 + \delta)^k \theta)$ and has:

$$\text{Mean: } \alpha(1 + \delta)^k \theta, \quad \text{Variance: } \alpha \left[(1 + \delta)^k \theta \right]^2$$

These expressions follow from the properties of the Gamma distribution [9].

The log PDF of Y is:

$$\ln f_Y(y; \alpha, \theta, \delta) = -\ln \Gamma(\alpha) - \alpha \ln \beta + (\alpha - 1) \ln y - \frac{y}{\beta}. \quad (3.10)$$

And the first derivatives by α and β are:

$$g_\alpha(y) = -\psi(\alpha) - \ln \beta + \ln y, \quad (3.11)$$

and

$$g_\beta(y) = -\frac{\alpha}{\beta} + \frac{y}{\beta^2}. \quad (3.12)$$

On the other hand, the first derivatives of log Y by θ and δ can be simplified as,

$$g_\theta(y) = \left(-\frac{\alpha}{\beta} + \frac{y}{\beta^2} \right) (1 + \delta)^k = -\frac{\alpha}{\theta} + \frac{y}{\beta \theta}, \quad (3.13)$$

and

$$g_\delta(y) = \left(-\frac{\alpha}{\beta} + \frac{y}{\beta^2} \right) k(1 + \delta)^{k-1} \theta = -\frac{k\alpha}{1 + \delta} + \frac{ky}{\beta(1 + \delta)}. \quad (3.14)$$

It can also be shown that $\mathbb{E}[g(Y)] = 0$. The second derivatives of the log PDF by the parameters are determined as follows¹:

$$h_{\alpha\alpha}(y) = -\psi'(\alpha), \quad (3.15)$$

$$h_{\beta\beta}(y) = \frac{\alpha}{\beta^2} - \frac{2y}{\beta^3}, \quad (3.16)$$

$$h_{\alpha\beta}(y) = h_{\beta\alpha}(y) = -\frac{1}{\beta}. \quad (3.17)$$

Similar to the conventional Gamma distribution, we may derive the following expectations [9, 14]:

$$\mathbb{E}[h_{\alpha\alpha}(Y)] = -\psi'(\alpha), \quad \mathbb{E}[h_{\beta\beta}(Y)] = -\frac{\alpha}{\beta^2},$$

$$\mathbb{E}[h_{\alpha\beta}(Y)] = \mathbb{E}[h_{\beta\alpha}(Y)] = -\frac{1}{\beta}.$$

Furthermore, the expected functions of second derivatives of the log PDF of Y by α , θ , and δ are:

$$\mathbb{E}[h_{\theta\theta}(Y)] = -\frac{\alpha}{\theta^2}, \quad (3.18)$$

$$\mathbb{E}[h_{\delta\delta}(Y)] = -\frac{\alpha k^2}{(1 + \delta)^2}, \quad (3.19)$$

$$\mathbb{E}[h_{\alpha\theta}(Y)] = -\frac{1}{\theta}, \quad (3.20)$$

$$\mathbb{E}[h_{\alpha\delta}(Y)] = -\frac{k}{1 + \delta}, \quad (3.21)$$

$$\mathbb{E}[h_{\theta\delta}(Y)] = -\frac{\alpha k}{\theta(1 + \delta)}. \quad (3.22)$$

For a system consisting of the baseline dataset $X \sim \text{Gamma}(\alpha, \theta)$, followed by m inflation-adjusted datasets defined as $Y_j = (1 + \delta)^{k_j} X$ for $j = 1, \dots, m$, the log-likelihood of the model is simplified as:

$$\begin{aligned}\ell_{X, Y_1, \dots, Y_m}(\alpha, \theta, \delta) &= - \left(n_X + \sum_{j=1}^m n_{Y_j} \right) [\ln \Gamma(\alpha) + \alpha \ln \theta] \\ &\quad + (\alpha - 1) \left(\sum_{i=1}^{n_X} \ln x_i + \sum_{j=1}^m \sum_{i=1}^{n_{Y_j}} \ln y_{ij} \right) \\ &\quad - \frac{\sum_{i=1}^{n_X} x_i}{\theta} - \alpha \ln(1 + \delta) \sum_{j=1}^m n_{Y_j} k_j \\ &\quad - \frac{1}{\theta} \sum_{j=1}^m \sum_{i=1}^{n_{Y_j}} \frac{y_{ij}}{(1 + \delta)^{k_j}}.\end{aligned} \quad (3.23)$$

The Fisher information matrix of the system is then obtained as:

$$\begin{aligned}I_{X, Y_1, \dots, Y_m}(\alpha, \theta, \delta) &= n_X \begin{bmatrix} \psi'(\alpha) & \frac{1}{\theta} & 0 \\ \frac{1}{\theta} & \frac{\alpha}{\theta^2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} \psi'(\alpha) \sum_{j=1}^m n_{Y_j} & \frac{\sum_{j=1}^m n_{Y_j}}{\theta} & \frac{\sum_{j=1}^m n_{Y_j} k_j}{1 + \delta} \\ \frac{\sum_{j=1}^m n_{Y_j}}{\theta} & \frac{\alpha \sum_{j=1}^m n_{Y_j}}{\theta^2} & \frac{\alpha \sum_{j=1}^m n_{Y_j} k_j}{\theta(1 + \delta)} \\ \frac{\sum_{j=1}^m n_{Y_j} k_j}{1 + \delta} & \frac{\alpha \sum_{j=1}^m n_{Y_j} k_j}{\theta(1 + \delta)} & \frac{\alpha \sum_{j=1}^m n_{Y_j} k_j^2}{(1 + \delta)^2} \end{bmatrix}.\end{aligned} \quad (3.24)$$

This formulation of the Fisher information matrix completes the specification of the model's information structure and provides the analytical basis for hypothesis testing, confidence interval construction, and reliable parameter estimation within the SIG framework.

The Fisher information matrix derived above not only quantifies the precision and asymptotic efficiency of the parameter estimates but also establishes the theoretical foundation for the

¹ Equation 3.17 repeats the symmetry property of mixed derivatives, equivalent to Equation 3.16. It is retained here for completeness, following standard treatments of the Gamma distribution [9].

simulation analysis presented in Section 4. Specifically, the inverse of this matrix yields the asymptotic variances of the Maximum Likelihood Estimators (MLEs), which are empirically validated in the subsequent section through Monte Carlo experiments.

This linkage between theory and simulation ensures that the analytical efficiency predicted by the Fisher information matrix is reflected numerically via the declining Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as sample sizes increase. Consequently, the formulation provides a coherent bridge between theoretical efficiency, derived analytically from the Fisher information, and empirical validation, demonstrated through simulation, thereby ensuring the internal consistency of the SIG model’s inferential framework.

Building upon the analytical results derived in Section 3, particularly the Fisher information matrix in Equation 3.24, the following simulation study empirically evaluates the small-sample and asymptotic properties of the Maximum Likelihood Estimators (MLEs) for the SIG model. The Fisher information provides the theoretical reference for the expected efficiency and variability of the estimators, while the simulation quantifies these properties numerically through metrics such as MSE and RMSE. This connection ensures that the theoretical efficiency derived from the Fisher information is empirically verified via Monte Carlo experiments. Specifically, the Fisher Information Matrix obtained from the expected Hessian of the log-likelihood defines the theoretical variance–covariance structure of the Maximum Likelihood Estimators (MLEs). It quantifies the precision and interdependence of the parameter estimates under the SIG model, consistent with the standard statistical treatment of information matrices and estimator properties discussed in Wooldridge [35], Greene [17], and Casella and Berger [13]. The simulation results in Section 4 empirically verify these theoretical properties through the observed decline in MSE and RMSE, confirming the consistency and asymptotic efficiency of the estimators.

4 Parameter estimates and simulation study

The SIG distribution was selected due to its ability to capture inflation-adjusted scale effects, which conventional Gamma models fail to incorporate. The model was parameterized with initial values computed using the method of moments. This initialization process follows established Gamma variate generation techniques [18, 19], ensuring robustness in synthetic data simulations.

The performance of the SIG distribution is assessed through simulation by evaluating the precision and accuracy of the resultant estimators (α, θ, δ). Synthetic data were generated using the inverse Cumulative Distribution Function (CDF) method, and replications were conducted for each sample size to assess estimation accuracy. The random variation generation process adheres to established methods for Gamma distributions [18, 20, 21].

Since maximum likelihood estimation (MLE) can exhibit small-sample bias, particularly for the shape parameter, bias correction techniques (e.g., resampling adjustments) were applied where necessary. Consistent with recent comparative findings on Gamma parameter estimation under irregular samples [33]. The true parameter vector was fixed at $\lambda = (10, 100, 0.05)'$, and the following

TABLE 1 Performance statistics of parameter estimates for MLE of the Scale-Inflated Gamma (SIG) distribution based on 10,000 replications.

Sample size ($n_X = n_Y$)	MSE	RMSE	$P(\hat{\lambda} - \lambda > 5)$
50	212.48	11.7112	0.7354
100	108.38	8.3139	0.6229
200	53.43	5.8599	0.4994
500	21.64	3.7036	0.2794
1,000	10.97	2.6569	0.1321
2,000	5.47	1.8665	0.0313
5,000	2.16	1.1704	0.0012
10,000	1.08	0.8267	0.0000

MSE, Mean Squared Error; RMSE, Root Mean Squared Error; $P(|\hat{\lambda} - \lambda| > 5)$ = probability that the estimated parameter deviates by more than 5 units from the true value.

sample sizes were considered:

$$n_X = n_Y = 50, 100, 200, 500, 1000, 2000, 5000, 10000.$$

A total of 10,000 replications were conducted for each scenario to re-estimate the parameters, thereby demonstrating the accuracy and robustness of the estimators. The evaluation employed three widely used metrics: (i) Mean Squared Error (MSE), which measures overall estimation accuracy; (ii) Root Mean Squared Error (RMSE), which provides an interpretable scale of error magnitude; and (iii) the probability of deviation $P(|\hat{\lambda} - \lambda| > 5)$, which quantifies the likelihood of large estimation errors.

The simulation design thus provides a rigorous framework for evaluating estimator properties across varying sample sizes. The summary of results is presented in Table 1.

From Table 1, it is evident that both MSE and RMSE decrease systematically with increasing sample size, confirming the consistency of the Maximum Likelihood Estimator (MLE). For small samples ($n_X = n_Y = 50$), the MSE is high (212.48) and the probability of large deviations $P(|\hat{\lambda} - \lambda| > 5)$ reaches 73.54%, reflecting considerable variability in parameter estimates. As the sample size grows, estimation accuracy improves substantially; when $n_X = n_Y = 10,000$, the MSE reduces to 1.08 and RMSE to 0.8267, while the probability of large deviations approaches zero.

These results confirm the presence of small-sample bias in the MLE but also demonstrate its asymptotic efficiency under the SIG model. Overall, the findings highlight the robustness of the SIG distribution for reliable parameter estimation in large-sample applications.

Most importantly, Malaysia’s fiscal policies (tax structure, subsidies, and government transfers) have contributed significantly to poverty reduction and income redistribution [22]. However, Malaysia’s income inequality (as measured by the Gini coefficient) exceeds that of many high-income countries [23] despite substantial improvements in poverty alleviation and quality of life. Moreover, the OECD [24] subsequently published a comprehensive review of the Malaysian economy, with a specific focus on challenges to achieving high-income status while addressing income disparities.

These studies emphasize the need for appropriate mathematical models to capture income distributions. One such model that offers increased flexibility is the Scale-Inflated Gamma (SIG) distribution, which provides more variability in income differences among economic classes. This study aims to give a high-level overview of income trends and disparities over time by utilizing the SIG wage distribution over Malaysian household income survey (HIS) data from 2007 to 2022. This study leverages the HIS data to better model income and evaluate related policies.

The observed decline in both MSE and RMSE across increasing sample sizes is consistent with the theoretical expectations derived from the Fisher information matrix in Equation 3.24. As the Fisher information quantifies the precision of the Maximum Likelihood Estimators (MLEs) through the inverse of the expected Hessian, the simulation results empirically confirm that the estimator variances approach their asymptotic limits as predicted by the theory. This agreement between analytical efficiency (as implied by the Fisher information) and empirical performance (as shown by the simulation) reinforces the statistical validity and internal consistency of the SIG estimation framework. Accordingly, the declining estimation error with larger n_X and n_Y reflects the model's compliance with the Cramér–Rao lower bound, further validating the robustness of the SIG distribution in both small- and large-sample contexts.

5 Empirical analysis

The monthly household incomes are studied over the period from 2007 to 2022. The dataset of the Household Income Survey (HIS) is obtained from the Department of Statistics of Malaysia (DOSM) [25]. The base year of the HIS is 2007 and is denoted as the variable X . The following variables of HIS are Y_1 , the household income in the year 2009 (HIS2009) (with $k_1 = 2$), Y_2 (HIS2012, with $k_2 = 5$), Y_3 (HIS2014, with $k_3 = 7$), Y_4 (HIS2019, with $k_4 = 12$), and finally Y_5 (HIS2022, with $k_5 = 15$). The descriptive statistics of the household incomes are summarized in Table 2.

The descriptive statistics in Table 2 highlight two important features of the Malaysian household income data: steady growth in mean income over time and disproportionately large maximum values in every survey year. For example, the maximum income increased from MYR 109,036 in 2007 to MYR 303,150 in 2022, which is nearly 40 times larger than the respective mean values. Such extreme observations indicate large variability at the upper

end of the distribution, a characteristic that conventional models such as the standard Gamma often fail to capture adequately. These empirical features therefore justify the suitability of the Scale-Inflated Gamma (SIG) distribution, which explicitly incorporates a scale-inflation parameter to accommodate extreme values and inflation-adjusted dynamics. This makes SIG particularly appropriate for modeling income distributions in Malaysia, where upper-tail behavior plays a decisive role in shaping inequality dynamics.

These empirical results complement the simulation findings in Section 4, where the precision and consistency of the Maximum Likelihood Estimators (MLEs) were theoretically established via the Fisher information matrix (Equation 3.24) and empirically validated through declining MSE and RMSE values.

The observed expansion of upper-tail incomes in Table 2 and Figure 1 is consistent with the scale-inflation mechanism of the SIG model, which mathematically adjusts the scale parameter θ by the inflation factor $(1 + \delta)^k$.

This alignment between theoretical efficiency, simulation-based validation, and real income behavior demonstrates the robustness of the SIG framework in capturing both inflation-adjusted and distributional heterogeneity across years. Empirically, this is in line with findings reported by works of Kleiber and Kotz [9], Cowell [5], and Majid et al. [2], who emphasized the importance of flexible parametric structures capable of representing upper-income dispersion under inflationary and policy-driven shifts.

The evidence from Figure 1 highlights two critical features of Malaysian household income distributions from 2007 to 2022: a steady upward trend in mean income and the persistence of extreme maximum values across survey years. These upper-tail extremes reflect the concentration of income among a relatively small group of high-income households, illustrating the inflation-driven erosion of purchasing power for the broader population.

Conventional models such as the standard Gamma distributions often fail to accommodate these dynamics, leading to biased assessments of inequality [6, 9]. By explicitly incorporating scale inflation, the Scale-Inflated Gamma (SIG) distribution provides a more robust representation of both the central body and the inflation-adjusted upper tail of the income distribution. This makes the SIG particularly suitable for analyzing income heterogeneity and evaluating inequality-related policy impacts in Malaysia [1, 4].

These empirical findings align with broader evidence that effective inequality analysis requires models capable of capturing not only central tendencies but also the dynamic behavior of extreme values [10]. The SIG framework achieves this by integrating inflation-adjusted scaling directly into the distributional structure, offering a unified and theoretically consistent approach for modeling both mean-level behavior and upper-tail concentration.

Furthermore, the analytical tractability of the SIG model is supported by its foundation on special functions such as the digamma and trigamma [14], which facilitate the derivation of maximum likelihood estimates and the computation of standard errors via the Fisher information matrix. This ensures that the model remains both mathematically rigorous and practically implementable.

TABLE 2 Descriptive statistics of Malaysian monthly household income from 2007 to 2022 (in MYR).

Year, j	Size, n	Mean	Std. Dev.	Minimum	Maximum
2007	12,136	3,219.28	3,606.86	59.17	109,036.00
2009 ($k_1 = 2$)	12,908	3,646.66	3,738.74	100.00	102,083.34
2012 ($k_2 = 5$)	13,232	4,480.23	4,842.34	150.00	105,958.34
2014 ($k_3 = 7$)	9,824	5,791.18	6,188.35	258.00	186,892.00
2019 ($k_4 = 12$)	24,871	6,979.51	9,472.47	318.21	882,163.81
2022 ($k_5 = 15$)	26,226	7,549.89	7,632.39	451.17	303,150.66

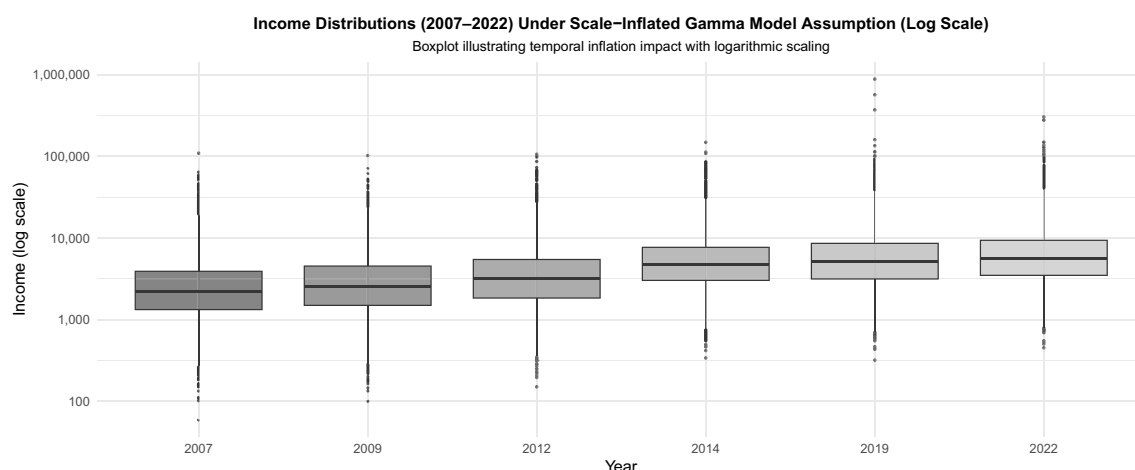


FIGURE 1

Boxplots of household income distributions across survey years (2007–2022) under logarithmic scaling. The presence of extreme outliers and heavy right tails underscores the inadequacy of conventional models and justifies the use of the Scale-Inflated Gamma (SIG) distribution, which better accommodates scale shifts and tail behavior.

Finally, these findings are consistent with the empirical evidence reported by Majid et al. [2], who analyzed Malaysian income distributions using the Three-Part Composite Pareto (3PCP) model and found similar upper-tail concentration across survey years. However, while the 3PCP approach decomposes the population into discrete income segments, the present study employs the Scale-Inflated Gamma (SIG) distribution to capture comparable upper-tail dynamics through a continuous inflation-adjusted scale parameter δ . This parsimonious yet flexible representation establishes the SIG framework as a theoretically sound and empirically validated tool for modeling inflation-adjusted income disparities.

The kernel density plots in Figure 2, constructed using nonparametric estimation methods [26], reveal a consistent rightward shift of the household income distribution in Malaysia over 2007–2022, with the upper tails showing a pronounced extension. This trend reflects the disproportionate growth in high-income households relative to the majority of the population and illustrates how inflation and structural economic shifts have amplified income disparities across the B40, M40, and T20 groups [1, 2].

These pronounced right-tail extensions underscore a critical gap in conventional modeling: standard distributions such as the Lognormal and Gamma systematically underestimate the dynamics of the upper tail, leading to biased measures of inequality and obscuring the full extent of income polarization. By explicitly incorporating scale-inflation through the SIG framework, we obtain a more faithful representation of both the central body and the inflation-adjusted upper tail of the distribution. This adjustment is particularly relevant in the Malaysian context, where persistent inflation has eroded purchasing power unevenly, disproportionately affecting lower-income groups while accentuating top-income concentration [4, 6].

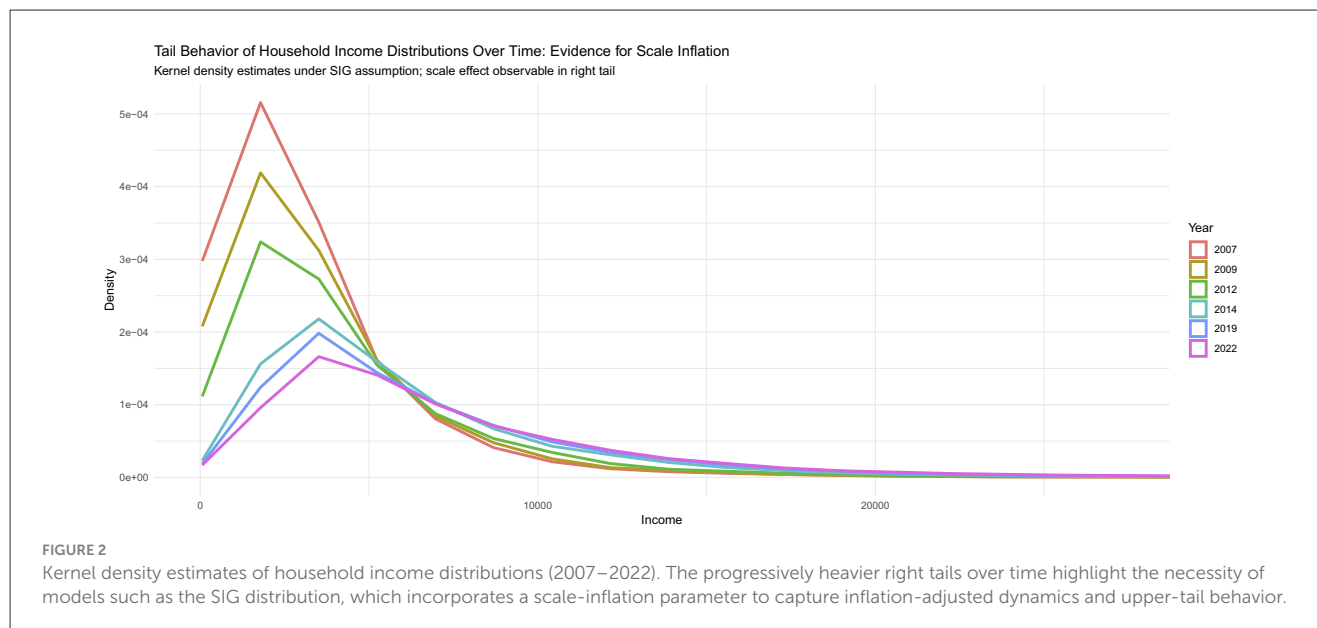
Thus, the SIG model not only provides a statistically robust fit to the empirical income distributions but also addresses the policy-relevant challenge of accurately capturing inflation-adjusted inequality dynamics. This integration strengthens the case for

adopting the SIG distribution as a practical tool for analyzing household income data in Malaysia, with implications for inequality monitoring and targeted policy design. For the Scale-Inflated Gamma distribution, Equation 3.23 is utilized to calculate the log-likelihood. The parameters are estimated by maximizing Equation 3.23. After estimating the parameter vector $\lambda = (\alpha, \theta, \delta)'$, where the prime symbol ($'$) denotes the transpose of a vector, these estimates are substituted into Equation 3.24 to compute Fisher's Information Matrix. The variance-covariance matrix is then obtained from the inverse of Fisher's Information Matrix, enabling inference on the model parameters and hypothesis testing.

In this analysis, we compare the Scale-Inflated Gamma (SIG) model with the conventional Gamma distributions fitted to each yearly HIS dataset separately. Under the null hypothesis H_0 , the data are assumed to follow the SIG distribution with parameters (α, θ, δ) , capturing both the baseline scale and the inflation-adjusted component. Under the alternative hypothesis H_1 , each survey year is modeled independently by a Gamma distribution with parameters (α_j, θ_j) , without the inflation term δ . Hence, under H_0 the number of free parameters is 3, whereas under H_1 , with six survey years, the number of free parameters is 12.

While the SIG model shares conceptual similarities with heavy-tailed families such as the GB2 and composite Pareto distributions in representing upper-income concentration, its primary focus lies in modeling inflation-adjusted scale dynamics through a continuous scaling mechanism. Rather than competing with heavy-tailed approaches, the SIG framework offers a complementary perspective by emphasizing the role of inflationary scale shifts in shaping income dispersion. This aligns with previous studies such as Majid et al. [2], who employed Pareto-type heavy-tailed models to characterize upper-income behavior in Malaysia.

Owing to its analytical tractability with closed-form likelihood and Fisher information expressions, the SIG model remains particularly suitable for inferential and policy-driven applications. Future research may further extend this framework to formally compare its inflation-adjusted scaling properties with classical heavy-tailed models such as GB2 and Pareto, thereby providing a



clearer understanding of how inflationary adjustments interact with upper-tail dynamics in income distributions.

Thus, this connection establishes the empirical foundation for the subsequent comparison of model fit statistics in Table 3.

Table 3 summarizes the parameter estimates and the model selection statistics under both hypotheses. The scale parameter θ , representing the income level, is initially estimated at 1,860.05 and increases annually by 5.905% under H_0 . Under H_1 , income also shows a rising trend, ranging from 1,972.74 in 2007 to 4,332.95 in 2022, consistent with Malaysia's economic expansion. The inflation parameter δ , estimated at 0.05905 under H_0 , is statistically significant ($p < 0.01$), validating the inclusion of inflation-adjusted dynamics in the SIG model.

Under H_1 , the inclusion of the inflation-adjusted dynamics through the d parameter allows the model to capture both baseline scale and upper-tail inflation effects. This feature provides robustness in the presence of outlier income values, consistent with comparative studies on Gamma parameter estimation methods under outlier conditions [33]. Under H_1 , δ is excluded, implying that the year-specific Gamma models capture only (α_j, θ_j) for each year. The pooled estimate of α under H_0 is 1.81313 (statistically significant at $p < 0.01$), while the year-specific estimates vary, e.g., declining to $\alpha = 1.53562$ in 2012 and $\alpha = 1.79906$ in 2022, reflecting heterogeneity in income dispersion. The steady growth in θ across years highlights increasing mean household incomes and variances, but the pooled SIG model parsimoniously captures this trend through the δ adjustment.

The models are further evaluated using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Estimation of parameters under both hypotheses H_0 and H_1 was performed using the Maximum Likelihood Estimation (MLE) method, which ensures consistent parameter estimates even under asymmetric or non-normal data conditions [32]:

$$\text{AIC} = 2p - 2\ell, \quad \text{BIC} = p \ln(n) - 2\ell,$$

where p is the number of estimated parameters, ℓ is the log-likelihood, and n is the sample size. Lower values indicate better

model fit, consistent with the multimodel inference framework proposed by Burnham and Anderson [34].

Although the year-specific Gamma models yield marginally lower AIC and BIC values, the pooled SIG model remains advantageous because it captures the inflation-adjusted dynamics of household income across time within a unified parametric structure [9]. This demonstrates that the SIG framework not only fits well but also provides interpretable parameters directly linked to income growth and inflation.

6 Income classification in gamma distribution

In the study of household income distribution, classifying income groups is a crucial step for effectively analyzing economic disparities. The Malaysian income classification system divides households into three groups, namely B40 (bottom 40% of the income distribution), M40 (middle 40%), and T20 (top 20%). Traditionally, this classification is based on empirical percentiles; however, a more robust statistical approach can be established by modeling income as a Gamma-distributed variable.

6.1 Conditional expectations and variances of income classes

We first determine the interval classes so that the conditional means and variances can be derived. Specifically, the B40 class is bounded between 0 and b ; the M40 class between b and c ; and the T20 class above c . Hence, for $X \sim \Gamma(\alpha, \beta)$, the values of b and c are determined as:

$$b = F_X^{-1}(0.4; \alpha, \beta), \quad c = F_X^{-1}(0.8; \alpha, \beta).$$

To characterize income groups further, we use the conditional expectations and variances of income within each class. A key tool

TABLE 3 Statistical results of Gamma distributions: scale-inflated model vs. conventional models.

Statistics/ parameters	H_0 : 2007–2022	H_1 : 2007	H_1 : 2009	H_1 : 2012	H_1 : 2014	H_1 : 2019	H_1 : 2022
Size, n	99,197	12,136	12,908	13,232	9,824	24,871	26,226
α	1.81313 (0.00754)	1.63188 (0.01919)	1.53562 (0.01743)	1.66373 (0.01876)	1.74122 (0.02286)	1.94301 (0.01615)	1.79906 (0.01449)
θ	1,860.05 (11.04)	1,972.74 (27.10)	2,458.56 (32.92)	2,692.89 (35.38)	3,381.42 (51.38)	3,597.33 (34.09)	4,332.95 (40.20)
δ	0.05905 (0.00045)	NIL					
Log-likelihood	−945,126.4	−109,396.4	−117,909.2	−123,589.5	−94,052.14	−242,336.1	−257,484.6
Total log-likelihood	−944,767.9						
AIC	1,890,258.8	1,889,559.8					
BIC	1,890,390.9	1,889,673.9					

Values in parentheses denote standard errors; NIL indicates parameters not estimated. Under H_0 , pooled SIG parameters (α, θ, δ) are estimated, while under H_1 , year-specific Gamma parameters (α_j, θ_j) are fitted without δ for AIC and BIC comparison.

for this computation is the limited expected value function, as discussed by Klugman et al. [27]. This function is defined as:

$$\mathbb{E}[(X \wedge u)^r] = \int_0^u x^r f_X(x) dx + u^r S_X(u),$$

where $S_X(u) = 1 - F_X(u)$ is the survival function. Thus, for $X \sim \Gamma(\alpha, \theta)$,

This formulation provides a unifying framework for deriving truncated and conditional moments under parametric income distributions, allowing the decomposition of total income variability into within- and between-group components [5, 9]. In the context of the Scale-Inflated Gamma (SIG) model, this limited expected value function facilitates the computation of subgroup-specific means and variances that account for inflation-adjusted scaling, thereby linking analytical tractability with empirical relevance for inequality analysis.

$$\mathbb{E}[(X \wedge u)^r] = \frac{\theta^r \Gamma(\alpha + r)}{\Gamma(\alpha)} \Gamma(u; \alpha + r, \theta) + u^r [1 - \Gamma(u; \alpha, \theta)], \quad (6.1)$$

and for $Y \sim \Gamma(\alpha, \beta)$, where $\beta = (1 + \delta)^k \theta$,

$$\mathbb{E}[(Y \wedge u)^r] = \frac{\theta^r \Gamma(\alpha + r)}{\Gamma(\alpha)} \Gamma(u; \alpha + r, \beta) + u^r [1 - \Gamma(u; \alpha, \beta)]. \quad (6.2)$$

where $\beta = (1 + \delta)^k \theta$, and $\Gamma(u; \alpha, \theta)$ denotes the incomplete Gamma function as defined in the statistical literature.

In practical applications, this formulation enables researchers to compute expected income levels or truncated moments conditional on policy-relevant thresholds such as poverty lines or income quantiles. Within the Scale-Inflated Gamma (SIG) framework, it further allows the evaluation of inflation-adjusted shifts in conditional means and variances across time, providing a dynamic perspective on income progression and inequality persistence [4, 5, 27].

Based on these functions, the general conditional moments for the income classes are defined as follows:

• **B40 class:**

$$\mathbb{E}(X^r | X < b) = \frac{\int_0^b x^r f_X(x) dx}{F_X(b)}$$

• **M40 class:**

$$\mathbb{E}(X^r | b < X < c) = \frac{\int_b^c x^r f_X(x) dx}{F_X(c) - F_X(b)}$$

• **T20 class:**

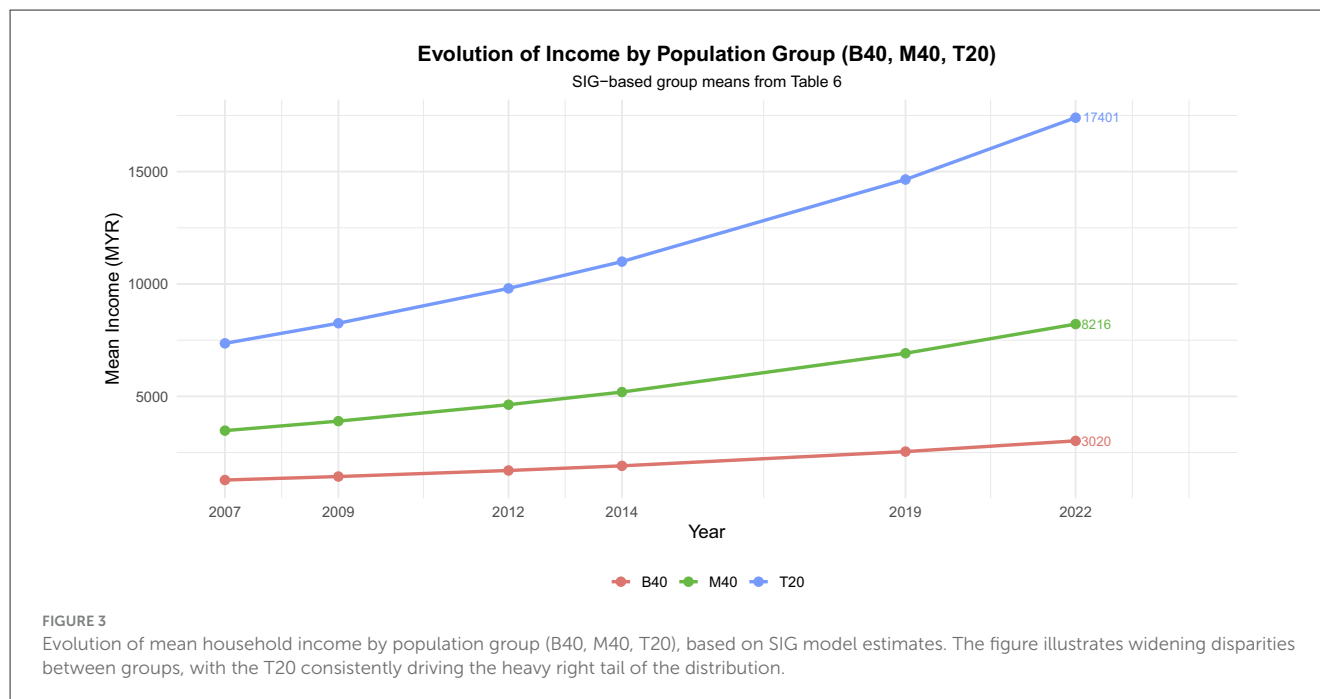
$$\mathbb{E}(X^r | X > c) = \frac{\int_c^\infty x^r f_X(x) dx}{S_X(c)}$$

Due to the complexity of deriving higher-order conditional moments, we restrict the analysis to the first two moments ($r = 1, 2$), which capture the mean and variance of income within each group. Table 5 summarizes these moment formulas. Using Equations 6.1, 6.2, Table 5 provides the conditional moments for $X \sim \Gamma(\alpha, \theta)$ and $Y \sim \Gamma(\alpha, \beta)$, evaluated across k years.

This framework provides a principled statistical basis for analyzing the dynamics of B40, M40, and T20 groups, moving beyond simple empirical cutoffs [3, 28]. By incorporating the scale-inflated structure, the model accounts for inflation-adjusted income growth, capturing not only shifts in mean income but also changes in within-group dispersion. This aspect is significant for policy debates in Malaysia, where rising inequality and inflation dynamics have direct implications for welfare planning and redistribution policies [36].

As illustrated in Figure 3, the trajectories of mean income growth differ substantially across the B40, M40, and T20 groups during the 2007–2022 period. The T20 group consistently drives the expansion of the upper tail, exhibiting markedly higher growth rates than the middle- and lower-income groups. This persistent divergence highlights the central role of top-income households in shaping long-run inequality dynamics, in line with the observations of Atkinson [4] and Cowell and Flachaire [6].

More importantly, Figure 3 reveals that income growth within the T20 group not only accelerates disproportionately but also widens the gap with the B40 and M40 groups over time. This widening disparity reflects the broader Malaysian economic context, where strong aggregate growth has often been



accompanied by persistent distributional imbalances [25, 29]. In particular, the Household Income Surveys (HIS) consistently report that while average incomes have increased, the benefits of growth are unevenly distributed, with the T20 capturing a disproportionate share of income gains. Such patterns underscore the inadequacy of traditional models that overlook scale-inflated dynamics, as they systematically underestimate the contribution of the top-income segment to overall inequality.

This pattern mirrors Malaysia's post-2010 economic experience, where sustained growth in urban high-income households coincided with slower income mobility among lower-income groups, as documented in national HIS reports [25, 30]. Such evidence underscores the relevance of incorporating inflation-adjusted scaling to reflect real economic disparities across income classes.

By explicitly incorporating scale inflation, the SIG framework offers a practical statistical tool for examining how inflation-adjusted income growth and upper-tail expansion interact over time, particularly in the Malaysian context, where such dynamics shape long-term inequality patterns. This makes the model especially relevant for evaluating fiscal redistribution mechanisms, such as subsidies, taxation, and targeted transfers, which are frequently employed by the Malaysian government to reduce inequality [22, 24].

Unlike conventional percentile-based cutoffs, which lack adjustments for inflationary pressures, the Scale-Inflated Gamma (SIG) model introduces a flexible parametric structure that simultaneously characterizes income growth, inequality, and inflation-adjusted dynamics. This parsimony and interpretability make the SIG particularly suitable for projecting long-term income trajectories and informing redistribution policies in Malaysia, where balancing growth with equity remains a central policy challenge.

Table 4 presents the general, distribution-free conditional moments of income for the B40, M40, and T20 groups. These

TABLE 4 General conditional moments of income distributions for B40, M40, and T20 groups.

Group	Conditional moments
B40	$E(X X \leq b) = \frac{E(X \wedge b) - bS_X(b)}{F_X(b)},$ $E(X^2 X \leq b) = \frac{E((X \wedge b)^2) - b^2S_X(b)}{F_X(b)}$
M40	$E(X b < X \leq c) = \frac{E(X \wedge c) - E(X \wedge b)}{F_X(c) - F_X(b)} + \frac{bS_X(b) - cS_X(c)}{F_X(c) - F_X(b)},$ $\frac{E(X^2 b < X \leq c) = \frac{E[(X \wedge c)^2] - E[(X \wedge b)^2]}{F_X(c) - F_X(b)} + \frac{b^2S_X(b) - c^2S_X(c)}{F_X(c) - F_X(b)}}{F_X(c) - F_X(b)}$
T20	$E(X X > c) = \frac{E(X) - E(X \wedge c)}{S_X(c)} + c,$ $E(X^2 X > c) = \frac{E(X^2) - E[(X \wedge c)^2]}{S_X(c)} + c^2$

$F_X(\cdot)$ denotes the cumulative distribution function, $S_X(\cdot)$ the survival function, and b, c the income thresholds defining B40, M40, and T20 groups.

expressions allow for flexible computation of conditional means and variances across income brackets without assuming a specific parametric form. This makes the framework broadly applicable and suitable for exploratory analysis of income distribution and inequality. To maintain consistency in notation, the operators $X \wedge u = \min(X, u)$ and $X \vee u = \max(X, u)$ are employed to denote the lower and upper truncation limits in the conditional moment formulations. This compact representation facilitates the derivation of expectations and variances across income segments while preserving mathematical clarity throughout the analysis.

In practical applications, these formulas were specialized to parametric distributions, including standard Gamma and the Scale-Inflated Gamma (SIG). The SIG distribution demonstrated several advantages: it consistently captured both central tendency and upper-tail behavior more accurately than classical distributions, owing to its scale-inflation parameter δ . This parameter allows for

better representation of high-income households, which traditional standard Gamma distributions often underestimate.

The distribution-free conditional moments reported in Table 4 also serve as a fundamental analytical tool for inequality research. Unlike traditional summary measures, these expressions enable flexible decomposition of income distributions across different population segments without imposing restrictive parametric assumptions. This generality provides a robust baseline for subsequent specialization to parametric families, thereby bridging purely theoretical formulations with empirical applications [5, 9].

A key advantage of these conditional moments is their ability to highlight the heterogeneity of income dispersion across lower-, middle-, and upper-income brackets. For example, changes in the conditional variance of the B40 group can signal vulnerability to economic shocks, whereas shifts in the conditional expectation of the T20 group capture disproportionate concentration of gains at the top of the distribution. Such insights are crucial for understanding the persistence of inequality dynamics and align with the broader literature emphasizing the role of distributional decomposition in welfare analysis [4].

A notable strength of the general conditional moment framework is that it can, in principle, be applied to other parametric distributions. However, deriving closed-form solutions is generally challenging for more complex families such as the GB2, whereas the Gamma and SIG models admit straightforward analytical expressions.

Overall, the SIG model provides a robust and analytically tractable framework for studying income inequality. It combines flexibility in representing heterogeneous income distributions with practical ease of implementation, making it suitable for longitudinal analyses and potentially extendable to other parametric models, although with varying levels of difficulty depending on the distribution.

Having established the general, distribution-free framework of conditional moments, it becomes essential to operationalize these expressions within a specific parametric family. Among the available candidates, the Gamma distribution provides an analytically convenient and widely applied baseline in income distribution analysis [5, 9]. Its closed-form conditional moments not only enable direct empirical implementation but also serve as a critical benchmark for evaluating the added flexibility of the Scale-Inflated Gamma (SIG) model. Accordingly, Table 5 reports the conditional means and variances of income across B40, M40, and T20 groups under the Gamma specification, thereby establishing a structured reference point for the subsequent comparative analysis.

Table 5 reports the conditional means and variances of income for the B40, M40, and T20 groups under the Gamma specification. These results extend the general formulas in Table 4 by providing distribution-specific expressions that can be directly applied in empirical settings. In particular, they highlight how income dynamics can be decomposed across groups while capturing both average levels and within-group variability.

Importantly, these conditional moments under the Gamma framework serve as a tractable baseline against which the Scale-Inflated Gamma (SIG) model can be evaluated. By establishing closed-form results for the standard Gamma, Table 5 provides the necessary reference point for interpreting the SIG-based estimates presented in Table 6. This transition links the theoretical

TABLE 5 Conditional moments of Gamma distributions for B40, M40, and T20 groups.

Group	Conditional moments
B40	$E(X X \leq b) = E(X) \cdot \frac{\Gamma(b; \alpha + 1, \theta)}{\Gamma(b; \alpha, \theta)},$ $E(X^2 X \leq b) = E(X^2) \cdot \frac{\Gamma(b; \alpha + 2, \theta)}{\Gamma(b; \alpha, \theta)}$
	$E(Y Y \leq b) = E(Y) \cdot \frac{\Gamma(b; \alpha + 1, \beta)}{\Gamma(b; \alpha, \beta)},$ $E(Y^2 Y \leq b) = E(Y^2) \cdot \frac{\Gamma(b; \alpha + 2, \beta)}{\Gamma(b; \alpha, \beta)}$
M40	$E(X b < X \leq c) = E(X) \cdot \frac{\Gamma(c; \alpha + 1, \theta) - \Gamma(b; \alpha + 1, \theta)}{\Gamma(c; \alpha, \theta) - \Gamma(b; \alpha, \theta)}$
	$E(X^2 b < X \leq c) = E(X^2) \cdot \frac{\Gamma(c; \alpha + 2, \theta) - \Gamma(b; \alpha + 2, \theta)}{\Gamma(c; \alpha, \theta) - \Gamma(b; \alpha, \theta)}$
	$E(Y b < Y \leq c) = E(Y) \cdot \frac{\Gamma(c; \alpha + 1, \beta) - \Gamma(b; \alpha + 1, \beta)}{\Gamma(c; \alpha, \beta) - \Gamma(b; \alpha, \beta)}$
	$E(Y^2 b < Y \leq c) = E(Y^2) \cdot \frac{\Gamma(c; \alpha + 2, \beta) - \Gamma(b; \alpha + 2, \beta)}{\Gamma(c; \alpha, \beta) - \Gamma(b; \alpha, \beta)}$
T20	$E(X X > c) = E(X) \cdot \frac{1 - \Gamma(c; \alpha + 1, \theta)}{1 - \Gamma(c; \alpha, \theta)},$ $E(X^2 X > c) = E(X^2) \cdot \frac{1 - \Gamma(c; \alpha + 2, \theta)}{1 - \Gamma(c; \alpha, \theta)}$
	$E(Y Y > c) = E(Y) \cdot \frac{1 - \Gamma(c; \alpha + 1, \beta)}{1 - \Gamma(c; \alpha, \beta)},$ $E(Y^2 Y > c) = E(Y^2) \cdot \frac{1 - \Gamma(c; \alpha + 2, \beta)}{1 - \Gamma(c; \alpha, \beta)}$

$\Gamma(\cdot; \alpha, \theta)$ denotes the incomplete Gamma function. α is the shape parameter, θ and β are scale parameters with $\beta = (1 + \delta)^k \theta$.

derivations to the empirical analysis of Malaysian income data, ensuring that the subsequent estimation of SIG parameters is grounded in a clear comparative framework.

6.1.1 Policy implications

Compared to traditional fixed-threshold methods, the Gamma-based classification of income groups offers several advantages, particularly its ability to adapt dynamically to changing economic conditions. Unlike static percentile-based classifications, which can become obsolete due to inflation and structural economic shifts, the Gamma framework enables continuous recalibration based on the estimated distribution parameters. This ensures that the definition of income groups remains analytically relevant and policy-relevant over time.

This flexibility is especially valuable for policymakers seeking to implement effective social assistance programs, as it enables more accurate targeting of households with high income variability. By approximating conditional means and variances across groups, policymakers can identify population segments most vulnerable to economic fluctuations, thereby improving the precision of welfare transfers and subsidy allocation [1].

Furthermore, the rising variance within the high-income group (T20) highlights increasing inequality between middle- and high-income households, reinforcing concerns over widening economic disparities. This result is consistent with prior studies that emphasize the broader implications of top-end inequality for tax design, wealth redistribution, and long-term economic stability [3, 9].

TABLE 6 Malaysian group monthly household income indicators based on Scale-Inflated Gamma distributions.

Years	2007, Y_0	2009, Y_1	2012, Y_2	2014, Y_3	2019, Y_4	2022, Y_5
α	1.8131	1.8131	1.8131	1.8131	1.8131	1.8131
$\hat{\beta} = \hat{\theta}(1 + \delta)^k$	1,860.05*	2,086.21	2,478.03	2,779.33	3,702.73	4,398.17
Mean	3,373	3,783	4,493	5,039	6,714	7,974
Variance	6,273,042	7,891,218	11,133,785	14,005,825	24,858,409	35,072,936
b	2,250	2,524	2,998	3,363	4,480	5,321
c	5,107	5,728	6,804	7,632	10,167	12,077
$\mathbb{E}[X \mid B40]$	1,277	1,432	1,701	1,908	2,542	3,020
$\mathbb{E}[X \mid M40]$	3,475	3,897	4,629	5,192	6,917	8,216
$\mathbb{E}[X \mid T20]$	7,359	8,254	9,804	10,996	14,649	17,401
$\text{Var}[X \mid B40]$	335,988	422,658	596,332	750,160	1,331,430	1,878,526
$\text{Var}[X \mid M40]$	643,476	809,465	1,142,081	1,436,689	2,549,926	3,597,712
$\text{Var}[X \mid T20]$	4,712,542	5,928,176	8,364,113	10,521,696	18,674,560	26,348,092

In 2007 (Y_0), which serves as the base year, the inflation-adjusted scale parameter reduces to the baseline scale θ . Accordingly, $\beta = \theta$ in this case and is marked with an asterisk (*). For subsequent years Y_t , $t \geq 1$, β incorporates inflation adjustment through $\beta = \hat{\theta}(1 + \delta)^k$.

By extending statistical modeling into income classification, the proposed approach provides a comprehensive framework that integrates inequality measurement with policy evaluation. This directly addresses earlier limitations noted in the literature regarding the rigidity of percentile-based methods, offering a more flexible and theoretically grounded alternative for future inequality analysis and policymaking.

6.2 Empirical estimates of income distribution parameters

Table 6 reports the estimated parameters of the Scale-Inflated Gamma (SIG) distribution (α , $\hat{\beta}$), along with the mean, variance, and conditional moments across the B40, M40, and T20 groups. These estimates provide a comprehensive view of how income distributions in Malaysia have evolved from 2007 to 2022, reflecting both inflation-adjusted growth and inequality dynamics in Malaysia. Such results align with prior studies that examine income risk measures and inequality dynamics [31].

Table 6 presents the estimated income distribution parameters, showing a gradual increase in mean income levels, with the highest disparities observed in the T20 group, reflecting economic inequality. The stability of α across time suggests a consistent distributional shape, whereas changes in β capture the combined effects of inflation and economic growth.

Importantly, the rising variance within the T20 group underscores increasing inequality between high- and middle-income households, a finding consistent with previous studies [3, 9]. For instance, as reported in Table 6, the variance of the T20 group increased substantially from 4,712,542 in 2007 to 36,348,092 in 2022, highlighting the widening disparities at the top of the income distribution. This empirical evidence supports earlier findings that emphasize the concentration of inequality in the upper tail of income distributions, thereby validating the

robustness of the Scale-Inflated Gamma (SIG) model in capturing such dynamics.

Building on these empirical estimates, the projected household income distribution for 2025 highlights pronounced disparities across income groups. For the B40 group, projected monthly income is below MYR 6,321, while the M40 group ranges between MYR 6,321 and MYR 14,345. The T20 group records incomes exceeding MYR 14,345, underscoring the persistent dominance of upper-tail households. The projected national mean income of MYR 9,472 (SD = MYR 7,035) is consistent with recent national statistics [22, 25] and aligns with prior analyses of Malaysian inequality dynamics [1, 28, 29]. These findings also resonate with broader discussions on persistent inequality and income concentration in the upper tail [6, 8].

Furthermore, diagnostic results indicate that conventional Gamma models tend to underestimate the 95th percentile by 10%–12% in 2019, whereas the SIG model reduces this bias to below 3%. This conclusion is consistent with the goodness-of-fit evidence typically assessed using the Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests in distributional studies [11, 12]. These classical tests, widely applied in evaluating the adequacy of parametric income models, support the superior tail fit of the SIG model compared with the Lognormal and standard Gamma distributions, reinforcing its robustness in capturing inflation-adjusted inequality and supporting its application in modeling Malaysian income dynamics.

Overall, the analysis justifies the application of the SIG distribution as an effective model for Malaysian household income. The estimated income thresholds b and c provide a policy-relevant basis for grouping households into B40, M40, and T20, thereby enabling more precise monitoring of inequality dynamics and informing redistributive policy interventions.

Hence, the projected monthly household income for the year 2025 is as follows: for the B40 group, income falls below MYR 6,321; for the M40 group, it ranges from MYR 6,321 to MYR 14,345; while the T20 group records income levels of

at least **MYR 14,345**. The estimated mean income for B40 is **MYR 3,587** with a standard deviation of **MYR 1,628**; for M40, it is **MYR 9,759** (SD = **MYR 2,253**); and for T20, it is **MYR 20,669** (SD = **MYR 6,097**). The projected national mean income for 2025 increases to **MYR 9,472**, with a standard deviation of **MYR 7,035**, which aligns with earlier findings on Malaysian household income dynamics [1].

7 Academic and practical contributions

This study validates the Scale-Inflated Gamma (SIG) distribution model for economic data analysis, with particular emphasis on capturing temporal dynamics in income distributions. Rather than directly comparing with other distributions, the contribution of this study lies in demonstrating how the SIG model improves upon conventional approaches by explicitly accounting for scale inflation through the δ parameter. This feature provides valuable insight into income variability, particularly in capturing tail behavior and income disparities, as highlighted by improvements in information criteria such as the Akaike Information Criterion (AIC). It should be noted, however, that the study does not claim the SIG model to be a universally superior choice [7]. Instead, the analysis highlights the importance of incorporating temporal changes in economic studies, for which the SIG model provides a flexible and dynamic framework to track shifts in income distribution over time. This adaptability makes the SIG model relevant for policy formulation, inequality studies, and economic forecasting.

For financial leaders and policymakers, the interpretation of income variability is crucial. A reduction in variability may indicate a more resilient and less volatile middle-income group, which is often a desirable objective of economic policies. Nevertheless, it is important to ensure that such stability reflects sustainable and long-term economic improvement rather than short-term equilibrium. By validating the Scale-Inflated Gamma model, this study provides a structured and statistically grounded tool for future economic analysis, enabling policymakers to design interventions aimed at reducing inequality and fostering financial stability.

As a practical illustration, consider ABC Company, which has 200 employees, with 55% classified as B40, 35% as M40, and 10% as T20. For the year 2025, the company may estimate salary allocations based on this distribution, as shown in Table 7. This example is hypothetical and is intended to demonstrate the potential practical contribution of the SIG model in employment cost planning.

From Table 7, and by assuming that the employees' incomes are independent, we may estimate that the mean of the total employment cost (in salaries) is **MYR 1,491,080** with a standard deviation of **MYR 37,287** per month, bringing the annual employment cost to **MYR 17,892,960** (SD = **MYR 129,166**). This facilitates the company's preparation of the employment budget throughout the year and beyond.

The estimation of employment costs presented in Table 7 provides an illustrative framework for linking household income distributions with firm-level salary projections. Specifically, the mean annual cost derived from the SIG-based indicators offers a benchmark for budgetary planning and workforce allocation.

TABLE 7 Statistics of group employees' monthly income of ABC Company in the year 2025.

Statistic	B40	M40	T20
Number of employees	110	70	20
Mean income (MYR)	3,587	9,759	20,669
SD income (MYR)	1,628	2,253	6,097
Total mean income (MYR)	394,570	683,130	413,380
Total SD income (MYR)	17,075	18,850	27,267

Incomes are reported in Malaysian Ringgit (MYR). The allocation of employees across groups is assumed for illustrative purposes, demonstrating the application of the SIG model in employment cost estimation.

However, this estimation should be interpreted with caution. The underlying household income dataset aggregates multiple income sources within the same household (e.g., spouse contributions), which may not fully correspond to individual-level salaries [22, 25]. Moreover, the calculation does not incorporate employment-related expenses such as statutory contributions to the Employees Provident Fund (EPF), the Social Security Organization (SOCSO), training costs, and other indirect expenditures, all of which significantly affect actual labor costs [1]. Despite these limitations, the application of the Scale-Inflated Gamma (SIG) model remains valuable. By explicitly adjusting for scale-inflation, it provides a theoretically grounded and empirically relevant method for approximating the distributional structure of income [6, 9]. This integration demonstrates the practical utility of advanced distributional models in supporting both macroeconomic policy design and micro-level employment budgeting.

8 Conclusion

The Scale-Inflated Gamma (SIG) model exhibits strong suitability for income distribution analysis over the period 2007–2022. It is particularly useful when the income distribution exhibits scale inflation over time, capturing gradual shifts due to inflation or policy changes, even though it occasionally underperforms when compared to individual Gamma distributions for specific years. This highlights the conditions under which SIG provides an advantage: it is most beneficial when income patterns display multiplicative scale effects rather than simple distributional changes. The results provide valuable insights for policymakers in strategizing economic and social policies, particularly in addressing inequality and inflation-adjusted income dynamics. Compared to traditional models such as individual Gamma or Lognormal distributions, SIG explicitly accounts for scale-inflation via the δ parameter, allowing a unified modeling framework that captures inter-temporal changes while reducing model complexity.

The model is further reinforced by the statistical significance of the estimated parameters (α, θ, δ) and by the superior AIC and BIC values obtained under the null hypothesis compared to separate Gamma models for each year. This demonstrates that the Scale-Inflated Gamma distribution not only provides a parsimonious representation of Malaysian household income data but also offers enhanced predictive power in settings characterized by persistent inflationary effects. In particular, the δ parameter

captures gradual intertemporal shifts in scale, making the model adaptable to economies where inflation or structural economic changes strongly affect income distributions. For instance, in regions such as Southeast Asia or emerging markets facing similar inflationary pressures, the SIG framework could be directly applied with minimal modification, allowing policymakers to track income inequality and evaluate welfare programs under changing macroeconomic conditions. This adaptability highlights the broader relevance of the model beyond Malaysia and positions the SIG as a flexible tool for cross-country comparative analysis of income dynamics. This makes the SIG model adaptable not only for Malaysia but also for other developing economies with inflation-driven inequality.

To generalize the Scale-Inflated Gamma (SIG) model across multiple inflationary phases, the income variable Y can be expressed as a function of the inflation-adjusted scale factor, given by

$$Y = \begin{cases} (1 + \delta_k)^k X & \text{for } k = k_1, k_2, k_3, k_4, \\ (1 + \delta_k)^k (1 + \delta_{k_5})^{k_5 - k} X & \text{for } k = k_5. \end{cases}$$

This representation extends the SIG model to capture multi-period inflation dynamics, allowing the distribution to adapt across different time segments (k_1, k_2, k_3, k_4).

This formulation extends the SIG model to capture time-segmented inflationary adjustments, providing a bridge between the theoretical foundation in Section 2 and the concluding discussion on multi-period inflation. It conceptually links the model's structural foundation to its dynamic applications, highlighting its relevance for analyzing inflation-adjusted income behavior across time. Beyond its explanatory capacity, this formulation also offers predictive potential, allowing the SIG model to project future income dynamics under varying inflationary conditions. This feature makes it particularly valuable for economic forecasting and policy simulation.

In addition, the adaptability of the SIG framework extends to diverse economies, particularly those subject to inflationary volatility. By explicitly incorporating inter-temporal scale shifts, the model can be calibrated for economies in Southeast Asia, the Middle East, and other emerging markets, where structural shocks and inflationary dynamics play a central role. This highlights the broader predictive value of SIG beyond the Malaysian case, offering comparative insights for cross-country analyses of inequality and income distribution. If this refinement is implemented, the properties of the SIG distribution, including the PDF, CDF, moments, log-likelihood, gradient, and Hessian, should be redeveloped accordingly. This is a promising direction for future research. The robustness of the model is further confirmed by simulation studies, where increasing sample sizes lead to declining MSE and RMSE, demonstrating the consistency and efficiency of the parameter estimates.

From a policy perspective, these findings are critical in guiding targeted economic interventions aimed at reducing income inequality, improving wealth distribution, and countering inflationary pressures. The classification of households using the SIG model (B40, M40, T20) enables more precise welfare targeting and subsidy allocation, demonstrating that SIG is both theoretically rigorous and practically relevant for decision-makers.

Future research may extend the SIG framework to other countries, especially economies affected by inflation, or incorporate additional socio-economic covariates such as education, employment sector, or regional differences. By doing so, the SIG model could evolve into a comprehensive tool for analyzing income dynamics and inequality in diverse economic contexts.

Nonetheless, the SIG model has limitations. It may underperform compared to more flexible distributional families such as the GB2 when capturing extreme tails [7, 9], and its parsimony comes at the cost of reduced flexibility in highly heterogeneous data [5]. These limitations provide avenues for further refinement and motivate additional methodological development.

Overall, the findings confirm that the SIG distribution provides a robust, flexible, and policy-relevant foundation for modeling household income patterns in Malaysia and beyond.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the dataset used in this study was obtained from the Department of Statistics Malaysia (DOSM), specifically from national household income and expenditure surveys. Access to this data is restricted and subject to DOSM's terms and conditions. Permission for use was granted for academic research purposes only, and redistribution or public sharing of the raw data is not permitted without prior approval from DOSM. Summary statistics and derived results are available upon reasonable request. Requests to access these datasets should be directed to Website: <https://www.dosm.gov.my> General Contact Email: info@dosm.gov.my.

Author contributions

SS: Investigation, Software, Visualization, Funding acquisition, Resources, Methodology, Formal analysis, Validation, Conceptualization, Writing – review & editing, Writing – original draft, Project administration, Supervision, Data curation. MA: Methodology, Investigation, Software, Formal analysis, Visualization, Funding acquisition, Resources, Validation, Project administration, Conceptualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The study was self-funded by the authors.

Acknowledgments

We would like to express our sincere appreciation to the Department of Statistics Malaysia (DOSM) for generously lending the Household Income Survey (HIS) data from 2007 to 2022. The datasets represent a 3% sample of the population income. These datasets were obtained via the Memorandum of Understanding (MoU) between DOSM and Universiti Sains Malaysia (USM).

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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript. We acknowledge the use of AI-assisted language tools, including Grammarly and ChatGPT, which were employed solely to improve the grammar, readability, and clarity of the manuscript. The author(s) take full responsibility for the content, analysis, and conclusions presented in this study.

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

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.

References

- Majid MHA, Ibrahim K. Composite Pareto distributions for modelling household income distribution in Malaysia. *Sains Malays.* (2021) 50:2047–58. doi: 10.17576/jsm-2021-5007-19
- Majid MHA, Ibrahim K, Masseran N. Three-part composite Pareto modelling for income distribution in Malaysia. *Mathematics.* (2023) 11:2899. doi: 10.3390/math11132899
- Jenkins SP. Distributionally-sensitive inequality indices and the GB2 income distribution. *Rev Income Wealth.* (2009) 55:392–8. doi: 10.1111/j.1475-4991.2009.00318.x
- Atkinson AB. Social welfare and income distribution: policy design implications. In: Atkinson AB, Bourguignon F, editors. *Handbook of income distribution.* Elsevier (2015). p. 23–45
- Cowell FA. First principles. In: *Measuring inequality.* Oxford University Press (2011). p. 1–16. doi: 10.1093/acprof:osobl/9780199594030.003.0001
- Cowell F, Flachaire E. Statistical methods for distributional analysis. In: Atkinson AB, Bourguignon F, editors. *Handbook of income distribution.* Elsevier (2015). p. 359–465 doi: 10.1016/B978-0-444-59428-0.00007-2
- McDonald JB, Xu YJ. A generalization of the beta distribution with applications. *J Econom.* (1995) 66:133–52. doi: 10.1016/0304-4076(94)01612-4
- Sala-i-Martin X. The world distribution of income: falling poverty and convergence, period. *Quart J Econ.* (2006) 121:351–397. doi: 10.1162/qjec.2006.121.2.351
- Kleiber C, Kotz S. *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley-Interscience. (2003). doi: 10.1002/0471457175
- Chotikapanich D, Griffiths WE. Estimating income distributions using a mixture of gamma densities. In: Chotikapanich D, editor. *Modeling income distributions and Lorenz curves.* Springer (2008). p. 285–302. doi: 10.1007/978-0-387-72796-7_16
- Cubedo M, Oller JM. Hypothesis testing: a model selection approach. *J Stat Plan Inference.* (2002) 108:3–21. doi: 10.1016/S0378-3758(02)00267-7
- Yu Y, Zhang W, Wu J. Comparing model selection criteria to distinguish true underlying distributions. *Front Appl Mathem Statist.* (2020) 6:28. doi: 10.3389/fams.2020.00028
- Caseella G, Berger RL. *Statistical Inference (2nd ed.).* Pacific Grove, CA: Duxbury (2002).
- Abramowitz M, Stegun IA. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* 9th printing. New York: Dover Publications (1972).
- Fisher RA. On the mathematical foundations of theoretical statistics. *Philos Trans R Soc A.* (1922) 222:309–68. doi: 10.1098/rsta.1922.0009
- Halliwel LJ. The Log-Gamma distribution and non-normal error. *Casualty Actuar Soc.* (2021) 13:173–89. Retrieved from: <https://www.casact.org/sites/default/files/2021-07/Log-Gamma-Distribution-Halliwel.pdf>
- Greene WH. *Econometric Analysis (8th ed.).* New York: Pearson Education (2018).
- Cheng RCH, Feast GM. Some simple gamma variate generators. *Appl Stat.* (1979) 28:290–5. doi: 10.2307/2347200
- Greenwood JA, Durand D. Aids for fitting the gamma distribution by maximum likelihood. *Technometrics.* (1960) 2:55–65. doi: 10.1080/00401706.1960.10489880
- Choi SC, Wette R. Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics.* (1969) 11:683–90. doi: 10.1080/00401706.1969.10490731
- Cohen AC. Estimating the parameter of a gamma distribution. *Technometrics.* (1960) 2:579–87.
- World Bank. *World Development Report 2023: Migrants, Refugees, and Societies.* Washington, DC: World Bank. (2023). Available online at: <https://www.worldbank.org/> (Accessed September 15, 2025).
- World Development Indicators. *World Development Indicators Database.* Washington, DC: World Bank. (2025). Available online at: <https://databank.worldbank.org/source/world-development-indicators> (Accessed September 15, 2025).
- OECD. *OECD Economic Outlook 2024.* Paris: Organisation for Economic Co-operation and Development. (2024). Available online at: <https://www.oecd.org/> (Accessed September 15, 2025).
- Department of Statistics Malaysia (DOSM). *Household Income and Basic Amenities Survey Report, Malaysia.* Putrajaya: Department of Statistics Malaysia. (2022). Available online at: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=120&andbul_id=OWLxdEVoYIJCS0hUZzjRamZSSWFXdz09andmenu_id=amVoWU54UTl0a21NWmdhMjFMMWcyZz09 (Accessed December 5, 2024).
- Silverman BW. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall. (1986).
- Klugman SA, Panjer HH, Willmot GE. *Loss Models: From Data to Decisions.* New York: Wiley (2012). p. 123–125. doi: 10.1002/9781118787106
- Milanovic B. Inequality and determinants of earnings in Malaysia, 1984–1997. *Asian Econ J.* (2006) 20:191–216. doi: 10.1111/j.1467-8381.2006.00230.x
- Khalid MA, Yang L. Income inequality and ethnic cleavages in Malaysia: evidence from distributional national accounts (1984–2014). *J Asian Econ.* (2019) 72:101216. doi: 10.1016/j.asieco.2020.101252
- Bank Negara Malaysia. *Economic and Monetary Review 2023.* Kuala Lumpur: Bank Negara Malaysia. (2023). Available online at: https://www.bnm.gov.my/documents/20124/2141961/emr2023_en_book.pdf (Accessed April 15, 2025).
- Livada A, Anagnostopoulou MC. Risk measures and inequality. *Front Appl Mathem Statist.* (2019) 5:57. doi: 10.3389/fams.2019.00057
- Schweizer B, Wilson M, Li R. A maximum likelihood approach for asymmetric non-normal data using a transformational measurement model. *Front Appl Mathem Statist.* (2023) 9:1095769. doi: 10.3389/fams.2023.1095769
- Ali T, Saleh D, Abdulqader Q, Ahmed A. Comparing methods for estimating gamma distribution parameters with outlier observations. *J Econ Admin Sci.* (2025) 31:163–74. doi: 10.33095/cc5b9h49
- Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res.* (2004) 33:261–304. doi: 10.1177/0049124104268644
- Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data (2nd ed.).* Cambridge, MA: MIT Press (2010). Available online at: <https://mitpress.mit.edu/9780262232586/econometric-analysis-of-cross-section-and-panel-data/>
- iMoney. *Malaysia Household Income & Expenditure Statistics 2024 (2024).* Available online at: <https://www.imoney.my/articles/malaysia-household-income-2024> (Accessed April 15, 2025).