Modeling urban household income in Malaysia using parametric approach

Muhammad Aslam Mohd Safari¹, Nurulkamal Masseran² and Saiful Izzuan Hussain³

¹Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
²,³Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

E-mail: aslam.safari@upm.edu.my

Abstract. The parametric approach is commonly used to estimate the actual distribution and thus measure income inequalities while analyzing the income distribution. The parametric estimate of the income distribution is useful as it helps in inferences about the estimated inequalities to be made more easily. Centered on household income survey data from 2007, 2009, 2012, and 2014, a comparative assessment of six parametric distribution candidates that represent urban household income in Malaysia was conducted in this research. The gamma, lognormal, Weibull, Dagum, Singh–Maddala, and generalized beta of the second kind distributions were some of the two, three, and four parameters considered in this study. Based on our analysis, it was found that the generalized beta of second kind distribution was the most adequate model for explaining the urban household incomes in Malaysia. Then, on the basis of the generalized beta of second kind distribution, income inequality analysis of the urban households in Malaysia was assessed using Lorenz curve and Gini coefficient.

1. Introduction

Human life and survival depend on income [1]. Simple services and needs, such as food, lodging, schooling, and clothes, can be obtained with the money earned [1]. Commonly, nonetheless, an individual's wealth varies, and the resulting income disparity in society will contribute to income inequality. The consequences of income inequality have been extensively explored in the literature. A high level of income inequality, which is caused by broad income distribution, leads to a country's sluggish economic growth [2,3]. Furthermore, research on the relationship between income inequality and crime has discovered that income inequality is positively correlated with crime, including violent crimes like burglary and robbery [4,5]. As a result, attempts to minimize wealth inequality become critical to a country's overall well-being and economic growth.

The Department of Statistics, Malaysia (DOSM) has carried out income distribution analysis through an official survey known as the Household Income Surveys (HIS) [1]. The HIS' key goal is to gather data on household income and social background to assess the Malaysian population's economic well-being. The data and statistics obtained are then used to establish policies and economic plans for the world, especially for eradicating poverty and developing strategies for equal distribution of income. To calculate income inequality among Malaysian households, the DOSM usually employs an analytical
approach based on the Lorenz curve (LC) and the Gini index. The DOSM's solution, however, ignores the heavy-tailed properties that occur in the upper part of the income distribution [6]. For modeling income data, the parametric approach is commonly used to overcome this issue. The use of a parametric distribution to model income data has many benefits. Instead of describing an entire curve, a parametric distribution can be defined using only a few parameters [7]. Furthermore, such a distribution will capture the heavy-tailed property in the upper part of the income distribution, making further research on income inequalities easier [8–12]. Kleiber and Kotz [7] provided a thorough overview of parametric income distribution models including the Pareto, gamma, lognormal, Weibull, Dagum, Singh–Maddala, and generalized beta of the second kind (GB2) distributions. Some scholars have considered the semi-parametric method, which combines the empirical distribution and the Pareto model to explain both the lower and upper portions of the income distribution [13,14]. The parametric approach based on the Pareto and reverse Pareto distributions are used to model upper- and lower-tailed data for Malaysian household income data, respectively, whereas the semi-parametric approach is used to describe the whole income distribution [6,15–20].

The strata can be divided into urban and rural areas using the HIS. A gazetted area with adjoining built-up areas with a combined population of 10,000 or more is known as an urban area [1]. Rural areas, on the other hand, are defined as gazetted areas with a population of fewer than 10,000 people and non-gazetted areas [1]. The objective of this research is to use a parametric method to explain the income distribution of urban households using HIS data from 2007, 2009, 2012, and 2014. The gamma, lognormal, Weibull, Dagum, Singh–Maddala, and GB2 distributions consisting of two, three, and four parameters were used in this research. Furthermore, this study utilized the LC and Gini coefficient to measure income inequality among urban households based on the best parametric model.

2. Source of data and descriptive statistics

The monthly gross income of urban households in 2007, 2009, 2012, and 2014 was collected from the DOSM and used in this research. The DOSM collects the data through the HIS, which takes place twice every five years. The mean, median, deviation, maximum and minimum skewness, and kurtosis are all included in table 1's descriptive statistics for urban household income. Furthermore, table 1 indicates a growing pattern demonstrated by the mean and median, showing that urban household income rose overall from 2007 to 2014. For all years, the mean and median values are ranged between 59.17–375.00 and 70400.00–186892.00, respectively. These figures show that the distribution of urban household income in different years is incredibly broad, suggesting that the data is highly dispersed. Per year, the variances in urban household income are also high, suggesting that the data is dispersed widely across the mean. All of the skewness coefficients appear positive, meaning that the urban household income distribution does not obey a normal distribution and skews to the right. All of the distributions of urban household income exhibit a heavy-tailed property, as shown by the high value of kurtosis coefficients.

| Year | Mean  | Median | Min   | Max       | Variance       | Skewness | Kurtosis |
|------|-------|--------|-------|-----------|----------------|----------|----------|
| 2007 | 3991.58 | 2833.00 | 59.17 | 109036.00 | 19117840      | 6.4741   | 89.2593  |
| 2009 | 4404.16 | 3245.54 | 100.00| 70400.00  | 16931676      | 3.9953   | 34.2811  |
| 2012 | 5554.47 | 4095.42 | 228.00| 105958.30 | 32686684      | 5.7150   | 63.5546  |
| 2014 | 6659.99 | 4985.33 | 375.00| 186892.00 | 42782507      | 6.5998   | 104.3536 |

3. Methodology

3.1. Parametric distributions and parameter estimation

The income distribution of urban households was modeled using six parametric distributions in this research. To estimate the parameters of all the distributions considered in this research, the maximum
likelihood estimator (MLE) was used. All of the models are defined below, including the probability density function (PDF), cumulative distribution function (CDF), and estimation of parameters based on MLE.

3.1.1. Gamma distribution. The gamma distribution's PDF and CDF can be described as following [21]:

\[
f(x) = \frac{x^{a-1}}{\Gamma(a)b^a} e^{-\frac{x}{b}}, \quad x > 0, a > 0, b > 0,
\]

\[
F(x) = \frac{\gamma(a, x/b)}{\Gamma(a)}, \quad x > 0, a > 0, b > 0,
\]

where the shape and scale parameters are presented by \(a\) and \(b\), respectively. Furthermore, \(\Gamma(\cdot)\) denotes the gamma function while \(\gamma(\cdot, \cdot)\) denotes the partial gamma function. The gamma distribution's log likelihood function was defined as follows:

\[
\ell(a, b; x_1, \ldots, x_n) = (a-1)\sum_{i=1}^{n} \log x_i - n \log \Gamma(a) - na \log b - \frac{1}{b} \sum_{i=1}^{n} x_i.
\]

The MLE for \(b\) is given by

\[
\hat{b} = \frac{1}{na} \sum_{i=1}^{n} x_i.
\]

By substituting \(\hat{b}\) into the log likelihood function in equation (3), the MLE for parameter \(a\) was derived. After that, the log likelihood function was then refined in R software utilizing the optimization function.

3.1.2. Lognormal distribution. The lognormal distribution's PDF and CDF are demonstrated as following [21]:

\[
f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\log x - \mu\right)^2}{2\sigma^2}\right], \quad x > 0, -\infty < \mu < \infty, \sigma > 0,
\]

\[
F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right), \quad x > 0, -\infty < \mu < \infty, \sigma > 0,
\]

where the scale parameter is \(\mu\), the shape parameter is \(\sigma\), and the CDF of the ordinary normal distribution is \(\Phi(\cdot)\). For parameters \(\mu\) and \(\sigma\), the MLE is as follows:

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \log x_i,
\]

\[
\hat{\sigma} = \left(\frac{1}{n} \sum_{i=1}^{n} (\log x_i - \hat{\mu})^2\right)^{1/2}.
\]

3.1.3. Weibull distribution. The Weibull distribution's corresponding PDF and CDF are as following [21]:

\[
f(x) = \frac{\lambda x^{\lambda-1}}{\beta^\lambda} e^{-(x/\beta)\lambda}, \quad x > 0, \lambda > 0, \beta > 0,
\]
\[ F(x) = 1 - e^{-x^{(\lambda/\beta)}} , \quad x > 0, \lambda > 0, \beta > 0, \quad (10) \]

where \( \beta \) denotes the scale parameter and \( \lambda \) represents the shape parameter. The Weibull distribution is also described as the stretched exponential distribution in the literature [22]. The Weibull distribution's log likelihood function is as follows:
\[
\ell(\lambda, \beta; x_1, \ldots, x_n) = n \log \lambda - n \lambda \log \beta + (\lambda - 1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left( \frac{x_i}{\beta} \right)^\lambda .
\quad (11)
\]

The log likelihood function in equation (11) was optimized by disregarding it and calculating its minimum to obtain the MLE for parameters \( \lambda \) and \( \beta \). The optimization function in R software was used to perform this operation iteratively.

3.1.4. Dagum distribution. Camilo Dagum suggested the Dagum distribution for modeling income data [23]. The Dagum distribution's accompanying PDF and CDF are as following:
\[
f(x) = \frac{apx^{wp-1}}{b^w \left[ 1 + (x/b)^w \right]^{wp+1}} , \quad x > 0, \quad a > 0, \quad b > 0, \quad p > 0,
\quad (12)
\]
\[
F(x) = \left[ 1 + \left( \frac{x}{b} \right)^a \right]^{-p} , \quad x > 0, \quad a > 0, \quad b > 0, \quad p > 0,
\quad (13)
\]

The shape parameters are represented by \( a \) and \( p \), while the scale parameter is denoted by \( b \). For \( a, b, \) and \( p \), there is no closed form of MLE expression. The log likelihood function of the Dagum distribution was optimized using the optimization function in R software to produce the MLE for \( a, b, \) and \( p \). The Dagum distribution's log likelihood function is as follows:
\[
\ell(a, b, p; x_1, \ldots, x_n) = n \log a + n \log p + (ap - 1) \sum_{i=1}^{n} \log x_i - nap \log b - (p + 1)
\times \sum_{j=1}^{n} \log \left[ 1 + \left( \frac{x_i}{b} \right)^a \right].
\quad (14)
\]

3.1.5. Singh–Maddala distribution. Singh and Maddala [24] proposed the Singh–Maddala distribution, which has gotten a lot of coverage in the literature of income distributions and is also regarded as the Burr distribution [23]. The Singh–Maddala distribution's PDF and CDF are as described in the following:
\[
f(x) = \frac{aqx^{q-1}}{b^q \left[ 1 + (x/b)^q \right]^{q+1}} , \quad x > 0, \quad a > 0, \quad b > 0, \quad q > 0,
\quad (15)
\]
\[
F(x) = 1 - \left[ 1 + \left( \frac{x}{b} \right)^a \right]^{-q} , \quad x > 0, \quad a > 0, \quad b > 0, \quad q > 0,
\quad (16)
\]

where the shape parameters are presented as \( a \) and \( q \), and the scale parameter is denoted by \( b \). The MLE for parameters \( a, b, \) and \( q \) was calculated by optimizing the log likelihood function in R software with the optimization function. The Singh–Maddala distribution's log likelihood function is given as follows:
\begin{equation}
\ell(a, b, q; x_1, \ldots, x_n) = n \log q + n \log a + (a - 1) \sum_{i=1}^{n} \log x_i - na \log b
- (1 + q) \sum_{i=1}^{n} \log \left[ 1 + \left( \frac{x_i}{b} \right)^{\frac{a}{p}} \right].
\end{equation}

\subsection*{3.1.6. GB2 distribution.}
McDonald [25] introduced the GB2 distribution, which is generally regarded as an excellent definition of income distribution. The following are the PDF and CDF for the GB2 distribution:

\begin{equation}
f(x) = \frac{a x^{a-1}}{b^a B(p, q) \left[ 1 + (x/b)^a \right]^{p+q}}, \quad x > 0, a > 0, b > 0, p > 0, q > 0,
\end{equation}

\begin{equation}
F(x) = \frac{B(w, p, q)}{B(p, q)} = I(w, p, q), \quad x > 0, a > 0, b > 0, p > 0, q > 0,
\end{equation}

where \(a, p,\) and \(q\) are the shape parameters, whereas \(b\) is the scale parameter. Moreover, \(B(\cdot, \cdot)\) is the beta function, \(B(\cdot, \cdot, \cdot)\) represents the incomplete beta function, \(I(\cdot, \cdot, \cdot)\) denotes the regularized beta function, and \(w = (x/b)^a / [1 + (x/b)^a]\). The Dagum and Singh–Maddala distributions are special cases of the GB2 distribution, with \(q = 1\) and \(p = 1\), respectively. The log likelihood function of the GB2 distribution is optimized using the optimization function in R software to obtain the MLE for parameters \(a, b, p,\) and \(q\). The log likelihood function of the GB2 distribution is given by the following:

\begin{equation}
\ell(a, b, p, q; x_1, \ldots, x_n) = n \log \Gamma(p + q) + n \log a + (ap - 1) \sum_{i=1}^{n} \log x_i - nap \log b
- n \log \Gamma(p) - n \log \Gamma(q) - (p + q) \sum_{i=1}^{n} \log \left[ 1 + \left( \frac{x_i}{b} \right)^{\frac{a}{p}} \right].
\end{equation}

Note that \(B(p, q) = \Gamma(p)\Gamma(q) / \Gamma(p + q)\).

\subsection*{3.2. Model selection criteria}
The coefficient of determination \(R^2\) and Akaike information criterion (AIC) was used to find the best formula for representing urban household income. Under an expected distribution, \(R^2\) measures the degree of association between observed and theoretical probabilities [19]. An \(R^2\) value close to 1 means that the assumed theoretical model suits the data well. The \(R^2\) can be calculated using the following formula [19]:

\begin{equation}
R^2 = \frac{\sum_{i=1}^{n} \left[ \hat{F}(x_i) - \bar{F}(x) \right]^2}{\sum_{i=1}^{n} \left[ \hat{F}(x_i) - \bar{F}(x) \right]^2 + \sum_{i=1}^{n} \left[ F(x_i) - \hat{F}(x_i) \right]^2},
\end{equation}

where \(F(x_i)\) is the empirical cumulative probability for the \(i\)-th household income data, \(\hat{F}(x_i)\) is the estimated cumulative probability for the \(i\)-th household income data under an assumed theoretical model, and \(\bar{F}(x)\) is the average of \(\hat{F}(x)\).

The AIC calculates how much information a model loses in comparison to other models. The lower the volumes of information lost by a model, the higher the model's accuracy. Simply put, the one with the lowest AIC value would be chosen. The AIC can be determined using the following formula [26]:

\begin{equation}
AIC = -2 \log(L) + 2k,
\end{equation}
where \( L \) is the likelihood and \( k \) is the number of parameters of the fitted model.

### 3.3. Parametric Lorenz Curve and Gini index

The LC is a graphical instrument that can be used to measure income inequalities in a target group [27]. The more the LC deviates from the 45° line of equality, the more unequal the income distribution [27]. The Gini index is the ratio of the area between the 45° line of equality and LC to the area of the triangle below the 45° line of equality. The Gini index varies from 0 to 1, with 0 indicating perfect income equality and 1 indicating perfect income inequality [28].

The following are the definitions of the LC and Gini index [27,28]:

\[
L(y) = \frac{1}{E(x)} \int_0^y F^{-1}(t) dt = \frac{1}{E(x)} \int_0^{F^{-1}(y)} t f(t) dt, \quad 0 < y < 1, \quad (23)
\]

\[
G = 1 - 2 \int_0^1 L(y) dy, \quad (24)
\]

where \( E(x) \) represents a specific distribution's mean, \( F^{-1} \) denotes a specific distribution's quantile function, and \( f(x) \) represents a specific distribution's PDF. Based on the parametric distributions investigated in this research, table 2 shows the parametric form of the LC and Gini index.

**Table 2.** Parametric form of LC and Gini coefficient based on six considered distributions.

| Distributions | LC | Gini |
|---------------|----|------|
| Gamma         | \( \frac{\gamma(a+1,F^{-1}(y)/b)}{\Gamma(a+1)} \) | \( 1 - \frac{\Gamma(2a+1)}{a(a+1)\Gamma^2(a)} \cdot \frac{1}{2} F_1(a+1,2a+1;a+2;-1) \) |
| Lognormal     | \( \Phi \left[ \frac{\log F^{-1}(y) - \mu - \frac{\sigma^2}{2}}{\sigma} \right] \) | \( 1 - \frac{2}{\sqrt{2\pi}\sigma^2} \int \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \frac{1}{\sigma} \Phi \left( \frac{x - \mu - \frac{\sigma^2}{2}}{\sigma} \right) dx \) |
| Weibull       | \( \frac{\gamma(1+1/\lambda,(F^{-1}(y)/\beta)^\lambda)}{\Gamma(1+1/\lambda)} \) | \( 1 - 2^{-1/\lambda} \) |
| Dagum         | I \((v\mid p+1/a,1-1/a)\), where \( v = y^{1/p} \) | \( \frac{\Gamma(p)\Gamma(2p+1/a)}{\Gamma(2p)\Gamma(p+1/a)} - 1 \) |
| Singh-Maddala | I \((z;1+1/a,q-1/a)\), where \( z = 1-(1-y)^{1/q} \) | \( \frac{\Gamma(q)\Gamma(2q-1/a)}{\Gamma(q-1/a)\Gamma(2q)} - 1 \) |
| GB2           | I \((F^{-1}(y); p, q); p+1/a, q-1/a) \) | \( \frac{2B(2p+1/a,2q-1/a)}{pB(p,q)B(p+1/a,q-1/a)} \times \left[ \frac{1}{p} \cdot \left( F_1(1, p+q, 2p+1/a; p+1, 2(p+q); 1) \right)^2 - \frac{1}{p+1/a} \cdot \left( F_2(1, p+q, 2p+1/a; p+1/a+1, 2(p+q); 1) \right) \right] \) |

In table 2, \( I^{(*)(*)} \) is the inverse of the regularized beta function, \( \Phi \) denotes the Gaussian hypergeometric function and \( \Phi \) denotes the generalized hypergeometric function.
4. Results and discussion

For all of the models considered, Table 3 displays the parameter estimates, $R^2$, and AIC values. The fitting of all the considered distributions to the urban household income distribution can be seen in Figure 1. In Table 3, the subscript symbols “a” and “b” denote the highest and lowest $R^2$ and AIC values, respectively. In comparison to the other versions, the GB2 distribution has the highest $R^2$ and the lowest AIC, as seen in Table 3. The GB2 distribution was observed as the most appropriate model for representing urban household wealth, according to these findings. Furthermore, the $R^2$ values for all years using the GB2 distribution were greater than 0.99, implying that the GB2 model explained more than 99% of the variance in the results, while the residual (less than 1%) variation was due to errors and cannot be explained by the model.

![Histogram of urban household income](image)

**Figure 1.** Fitting of all the considered distributions to the urban household income distribution for the years of (a) 2007, (b) 2009, (c) 2012, and (d) 2014.
Table 3. Results of fitting of six parametric distributions to urban household income for each year.

| Year | Distributions     | Estimated parameters | $R^2$   | AIC      |
|------|-------------------|----------------------|---------|----------|
| 2007 | Gamma             | $\hat{a} = 1.6864$  | $\hat{b} = 2366.8790$ | 0.9749  | 120344.10 |
|      | Lognormal         | $\hat{\mu} = 7.9670$ | $\hat{\sigma} = 0.7821$ | 0.9992  | 119285.50 |
|      | Weibull           | $\hat{\lambda} = 1.1978$ | $\hat{\beta} = 4285.1747$ | 0.9613  | 120856.60 |
|      | Dagum             | $\hat{a} = 2.0370$  | $\hat{b} = 2261.4626$ | 0.9985  | 119303.80 |
|      | Singh–Maddala     | $\hat{a} = 2.3452$  | $\hat{b} = 2647.2641$ | 0.9981  | 119335.40 |
|      | GB2               | $\hat{a} = 0.9615$  | $\hat{b} = 1820.2006$ | 0.9996a | 119247.50 |
|      |                   | $\hat{p} = 4.9965$  | $\hat{q} = 3.3800$   |          |           |
| 2009 | Gamma             | $\hat{a} = 1.8561$  | $\hat{b} = 2372.8170$ | 0.9819  | 134860.10 |
|      | Lognormal         | $\hat{\mu} = 8.0973$ | $\hat{\sigma} = 0.7542$ | 0.9996  | 133975.40 |
|      | Weibull           | $\hat{\lambda} = 1.2860$ | $\hat{\beta} = 4806.5631$ | 0.9705  | 135435.70 |
|      | Dagum             | $\hat{a} = 2.1061$  | $\hat{b} = 2636.1315$ | 0.9981  | 134078.60 |
|      | Singh–Maddala     | $\hat{a} = 2.3418$  | $\hat{b} = 3195.6874$ | 0.9979  | 134111.30 |
|      | GB2               | $\hat{a} = 0.5483$  | $\hat{b} = 1255.0405$ | 0.9997a | 133965.00 |
|      |                   | $\hat{p} = 16.2453$ | $\hat{q} = 9.7858$   |          |           |
| 2012 | Gamma             | $\hat{a} = 1.7733$  | $\hat{b} = 3132.3110$ | 0.9824  | 150210.90 |
|      | Lognormal         | $\hat{\mu} = 8.3146$ | $\hat{\sigma} = 0.7722$ | 0.9993  | 149199.70 |
|      | Weibull           | $\hat{\lambda} = 1.2377$ | $\hat{\beta} = 6011.4612$ | 0.9672  | 150852.30 |
|      | Dagum             | $\hat{a} = 2.2725$  | $\hat{b} = 4034.8008$ | 0.9993  | 149242.00 |
|      | Singh–Maddala     | $\hat{a} = 2.2069$  | $\hat{b} = 4411.5972$ | 0.9993  | 149236.00 |
|      | GB2               | $\hat{a} = 1.0387$  | $\hat{b} = 4025.2289$ | 0.9995a | 149178.80 |
|      |                   | $\hat{p} = 3.6050$  | $\hat{q} = 3.5588$   |          |           |
| 2014 | Gamma             | $\hat{a} = 2.0319$  | $\hat{b} = 3277.6900$ | 0.9722  | 327629.50 |
|      | Lognormal         | $\hat{\mu} = 8.5380$ | $\hat{\sigma} = 0.7029$ | 0.9979  | 324964.50 |
|      | Weibull           | $\hat{\lambda} = 1.3003$ | $\hat{\beta} = 7297.1171$ | 0.9459  | 329543.60 |
|      | Dagum             | $\hat{a} = 2.3480$  | $\hat{b} = 4328.7295$ | 0.9997  | 324770.20 |
|      | Singh–Maddala     | $\hat{a} = 2.6989$  | $\hat{b} = 4601.5463$ | 0.9995  | 324804.30 |
|      | GB2               | $\hat{a} = 1.6867$  | $\hat{b} = 4116.2238$ | 0.9998a | 324741.90 |
|      |                   | $\hat{p} = 2.1741$  | $\hat{q} = 1.6475$   |          |           |

[a] Highest value of $R^2$ of each year.
[b] Lowest value of AIC of each year.
Figure 2 shows the LC of urban households on the basis of the GB2 distribution for all of the years considered. Table 4 summarizes the share of overall income shared by three groups: the bottom 40% (B40), the middle 40% (M40), and the top 20% (T20). Figure 2 and table 4 indicate modest but steady growth in the proportion of income gained by the B40 community from 14.85% to 16.89% over the period of study. The percentage of income gained by the M40 party fluctuated ranging from 35.84%–36.99%, showing no consistent trend. In 2007, the T20 party obtained 49.31% of total income, compared to 46.40% in 2014, signaling a small decline in the percentage of total income gained. Thus, it can be concluded that the income distribution of urban households changed during the course of study based on the rise and decline in income shares of the B40 and T20 classes, respectively. However, there was still a significant gap in the proportion of income earned by the T20 and B40 categories, reflecting Malaysia's income disparities among urban households.

![Lorenz curve of urban households](image)

**Figure 2.** Lorenz curve of urban households in 2007, 2009, 2012, and 2014 based on the GB2 model.

**Table 4.** Income shares of different income groups and Gini coefficient of urban households.

| Year | Percentage of income (%) | Cumulative percentage of income (%) | Gini   |
|------|--------------------------|-----------------------------------|--------|
|      | B40                  | M40                  | T20    | B40                  | M40                  | T20    |        |
| 2007 | 14.85                 | 35.84                 | 49.31  | 14.85                 | 50.69                 | 100.00 | 0.4343 |
| 2009 | 15.55                 | 36.99                 | 47.46  | 15.55                 | 52.54                 | 100.00 | 0.4143 |
| 2012 | 15.35                 | 36.99                 | 47.66  | 15.35                 | 52.34                 | 100.00 | 0.4181 |
| 2014 | 16.89                 | 36.71                 | 46.40  | 16.89                 | 53.60                 | 100.00 | 0.3966 |
Table 4 illustrates the estimated Gini coefficients of urban households based on the GB2 model. From 0.4343 in 2007 to 0.3966 in 2014, the Gini coefficients displayed a downward trend. According to these figures, only 56.57% of urban households shared gross household income in 2007, while the remaining 43.43% received none. In the meantime, in 2014, the index of income segregation fell, with 60.43% of urban families sharing gross household income while the remaining 39.66% receiving none. These results support the notion that urban household income distribution has changed over time.

5. Conclusion

This research has compared six parametric distribution options for representing urban household income in Malaysia, including the gamma, lognormal, Weibull, Dagum, Singh–Maddala, and GB2 distributions. The monthly gross income of urban households in 2007, 2009, 2012, and 2014 comprised the dataset used in this report. Meanwhile, the GB2 distribution was considered the most reliable model for representing urban household income, according to this research. Further, this study has used the LC and Gini coefficient to measure income inequality among urban households focusing on the GB2 distribution. According to the equipped LCs, it was discovered that during the course of the study, from 2007 to 2014, the B40 group's share of total urban household income increased marginally, while the T20 group's share of total urban household income decreased slightly. Nevertheless, there was no discernible difference in the M40 group's share of overall urban household revenue. Finally, it was estimated that urban households in Malaysia observed a declining trend in income disparity based on the calculated Gini coefficients.

Acknowledgments

The authors truly acknowledge the help by the Department of Statistics Malaysia and Bank Data of UKM for the HIS data supplied. This work is supported by the Universiti Kebangsaan Malaysia [grant no. DIP-2018-038].

References

[1] Department of Statistics Malaysia 2017 Household Income and Basic Amenities Survey Report 2016 (Putrajaya: DOSM)
[2] Persson T and Tabellini G 1994 Amer. Econ. Rev. 84 600–621
[3] Amarante V 2014 Growth Change 45 571–589
[4] Patterson E B 1991 Criminology 29 755–776
[5] Kelly M 2000 Rev. Econ. Stat. 82 530–539
[6] Safari M A M, Masseran N and Ibrahim K 2018 Physica A 512 1–13
[7] Kleiber C and Kotz S 2003 Statistical Size Distributions in Economics and Actuarial Sciences (Hoboken, NJ: John Wiley & Sons)
[8] Alfons A, Templ M and Filzmoser P 2013 J. Royal Stat. Soc. Ser. C Appl. Stat. 62 271–286
[9] Atkinson A B 2017 Economica 84 129–156
[10] Jenkins S P 2017 Economica 84 261–289
[11] Jenkins S P 2009 Rev. Income Wealth 55 392–398
[12] Chotikapanich D, Griffiths W E, Hajargasht G, Karunarathne W and Rao D S 2018 Econometrics 6 21
[13] Cowell F A and Victoria-Feser M P 2007 J. Econ. Inequal. 5 21–37
[14] Hlasny V and Verme P 2018 World Bank Econ. Rev. 32 428–455
[15] Safari M A M, Masseran N and Ibrahim K 2018 Physica A 509 169–180
[16] Safari M A M, Masseran N and Ibrahim K 2019 J. Appl. Stat. 46 1886–1902
[17] Safari M A M, Masseran N, Ibrahim K and Hussain S I 2019 Physica A 517 431–439
[18] Masseran N, Yee L H, Safari M A M and Ibrahim K 2019 Mathematics and Statistics 7 70–77
[19] Safari M A M, Masseran N, Ibrahim K and AL-Dhurafi N A 2020 Physica A 557 124893
[20] Safari M A M, Masseran N, Ibrahim K and Hussain S I 2021 Physica A 562 125359
[21] Giorgi G M and Nadarajah S 2010 Metron 68 23–46.
[22] Brzezinski M 2015 *Scientometrics* **103** 213–228
[23] Dagum C 1977 *Economie Appliquée* **30** 413–437
[24] Singh S K and Maddala G S 1976 *Econometrica* **44** 963–970
[25] McDonald J B 1984 *Econometrica* **52** 647–663
[26] Masseran N 2018 *Energy Convers. Manag.* **173** 56–64.
[27] Lorenz M O 1905 *J. Amer. Statist. Assoc.* **9** 209–219
[28] Gini C 1921 *Econ. J.* **31** 124–126