SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104

The New Measures of Lorenz Curve Asymmetry: Formulation and Hypothesis Testing

Muhammad Fajar[1,2], Setiawan Setiawan[1,*], Nur Iriawan[1]

[1] Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] Badan Pusat Statistik-Statistics Indonesia, Jakarta, Indonesia

ARTICLE INFO

Article history:
Received 24 August 2023
Received in revised form 23 October 2023
Accepted 27 October 2023
Available online 14 November 2023

Keywords: Measure; Lorenz curve; Income inequality; Asymmetry; Hypothesis; Bootstrap.

ABSTRACT

The existence of an asymmetric empirical Lorenz curve requires a measure of asymmetry that directly involves the geometry of the Lorenz curve as a component of its formulation. Therefore, establishing hypothesis testing for Lorenz curve asymmetry is necessary to conclude whether the Lorenz curve exhibits symmetry in actual data. Consequently, this study aims to construct a measure of Lorenz curve asymmetry that utilizes the area and perimeter elements of the inequality subzones as its components and establish a procedure for hypothesis testing the symmetry of the Lorenz curve. This study proposes two types of asymmetry measures, $\mathcal{R}_A$ and $\mathcal{R}_P$, constructed based on the ratio of area and perimeter obtained from the inequality subzone. These measures effectively capture the asymmetric phenomenon of the Lorenz curve and provide an economic interpretation of the values of $\mathcal{R}_A$ and $\mathcal{R}_P$. The Lorenz curve symmetry hypothesis testing, based on $\mathcal{R}_A$ and $\mathcal{R}_P$ through a nonparametric bootstrap, yields reliable results when applied to actual data.

1. Introduction

The Lorenz Curve is a graphical representation that describes income distribution inequality. It was developed by Lorenz [1] to depict wealth distribution, evaluate income disparities among various income groups in society, and analyze income inequality. Figure 1 presents the Lorenz curve ($\overline{\mathrm{OEC}}$), showing the relationship between the cumulative proportion of income ($q;\ 0 \leq q \leq 1$) and households ($p;\ 0 \leq p \leq 1$). The mathematical formulation of Lorenz curve with the parameter vector $\boldsymbol{\theta}$ is denoted as $L(p;\boldsymbol{\theta})$ [2–13].

Various household classes with differing socioeconomic characteristics across regions drive income distribution, as evident in the Lorenz curve. The assertion is that the shape may adopt asymmetry due to the unequal contributions of the two major income groups to income inequality. This notion establishes the groundwork for the economic interpretation of Lorenz curve's asymmetry measure developed in this study. The Lorenz curve asymmetry is categorized into two conditions, namely Asymmetry Conditions 1 and 2.

* *Corresponding author.*
E-mail address: setiawan@statistika.its.ac.id

https://doi.org/10.31181/dmame712024875

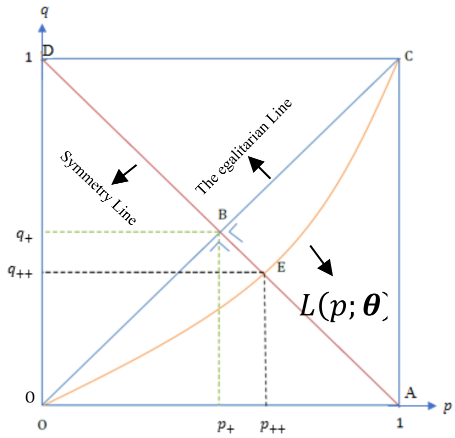
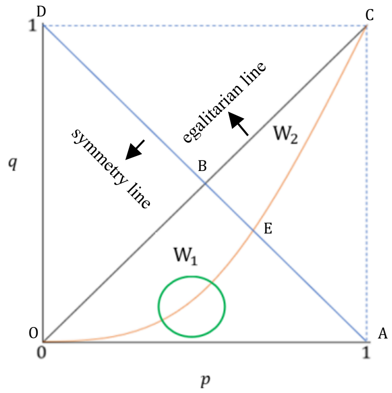
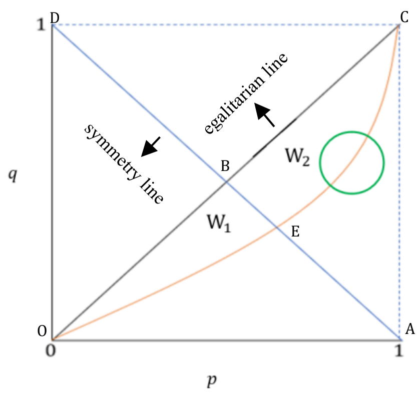


Fig. 1. The Lorenz curve and inequality zone $\overline{\mathrm{OBCE}}$

Asymmetry Condition 1 of the Lorenz curve showcases asymmetry shape, thereby presenting a bulge in the lower left section, indicated by the green circle in Figure 2. This condition shows inequality in income distribution in the population, in which the group with a cumulative proportion of income $q \leq q_+$ contributes more to inequality than those with $q > q_+$. These disparities in inequality sharing lead to larger area and perimeter of the inequality subzone W_1, created by the group with a cumulative proportion of income $q \leq q_+$, compared to the area and perimeter W_2 formed by $q > q_+$.





Note: the orange line is the Lorenz curve, and the blue line is the symmetry line. The inequality subzones, W_1 and W_2, formed from the inequality zone bounded by the egalitarian line and the asymmetric Lorenz curve Condition 1 divided by the line $q = 1 - p$ (as the symmetry line). The inequality subzones, W_1 and W_2, formed from the inequality zone bounded by the egalitarian line and the asymmetric Lorenz curve Condition 2 divided by the line $q = 1 - p$.

Fig. 2. The asymmetry Lorenz curve Condition 1

Fig. 3. The asymmetry Lorenz curve Condition 2

The Lorenz curve of Asymmetry Condition 2 shows an asymmetry shape with a noticeable bulge in the upper right portion, denoted by the green circle in Figure 3. This asymmetry signifies income inequality in the population, where the high-income group surpasses the low-income group in their contributions. Consequently, the inequality subzone W_2, shaped by the high-income group, comprises a larger area and perimeter compared to W_1 attributed to the low-income group.

To address this issue, the study introduces the novel measure of the Lorenz curve asymmetry, consisting of two components, including (1) asymmetry measure denoted as $\mathcal{R}_A$, based on the ratio of the area $\overline{\mathrm{OBE}}$ to the area $\overline{\mathrm{BEC}}$, and (2) asymmetry measure designated as $\mathcal{R}_P$, based on the ratio of the perimeter $\overline{\text{OBE}}$ to the perimeter $\overline{\text{BEC}}$. It is important to be aware that the two measures are inherently positive. Further interpretation of these two measures measure take into account the cumulative proportions of low-income and high-income groups, segmented by point q_+, which represents the intersection of the symmetry line $\overline{\text{AD}}$ and the egalitarian line $\overline{\text{OC}}$.

In the first scenario, Asymmetry Condition 1 occurs when both $\mathcal{R}_A$ and $\mathcal{R}_p$ fall within the interval $[1, \infty)$. In this case, subzone $\overline{\text{OBE}}$ (W_1) has a larger area and perimeter than subzone $\overline{\text{BEC}}$ (W_2). This suggests that the group with a cumulative proportion of income, $q \leq q_+$, significantly contributes to income inequality compared to $q > q_+$.

In the second scenario, Asymmetry Condition 2 occurs when both $\mathcal{R}_A$ and $\mathcal{R}_P$ values lie in the interval $(0, 1]$. This condition shows that subzone $\overline{\text{BEC}}$ (W_2) possesses a larger area and perimeter than the subzone $\overline{\text{OBE}}$ (W_1). This implies that the group with a cumulative proportion of income, $q > q_+$, plays a more substantial role in generating income inequality than $q > q_+$.

In the third scenario presents the symmetrical Lorenz curve when $\mathcal{R}_A = 1$ or $\mathcal{R}_P = 1$. This implies that subzone $\overline{\text{OBE}}$ (W_1) and subzone $\overline{\text{BEC}}$ (W_2) share equivalent areas and perimeters. In this case, income inequality is equally contributed to by both the groups with a cumulative proportion of income, $q \leq q_+$, and $q > q_+$.

The study necessitates hypothesis testing to determine whether the Lorenz curve is symmetrical or asymmetrical, specifically when the values of $\mathcal{R}_A$ or $\mathcal{R}_A$ closely method 1. As a result, confirming the symmetry or asymmetry of curve is essential. This determination relies on specific value criteria and also on a probabilistic framework, ensuring that the conclusions drawn are objective and robust.

The proposed measures, $\mathcal{R}_A$ and $\mathcal{R}_P$, offer a precise analysis of income inequality. These two measures help identify the income class that significantly contributes to income inequality. The information enables analysts and economists to recommend more targeted and effective measures to reduce inequality based on the income class that plays a substantial role in inequality. For example, when $\mathcal{R}_A$ and $\mathcal{R}_P$ both equal 1.2 (greater than 1), it suggests that the lower middle-income class is a significant contributor to income inequality. This implies that authorities can consider policies to increase the minimum wage, provide regular cash assistance to those in poverty, and create labor-intensive jobs. On the other hand, if $\mathcal{R}_A$ and $\mathcal{R}_P$ are both 0.8 (less than 1), it shows that the middle-income class and above are significant contributors to income inequality. In this case, authorities can explore measure such as increasing tax rates for individuals with income growth exceeding 50% from the previous year, optimizing and expanding Corporate Social Responsibility, and distributing zakat funds more strategically to reduce income inequality.

The two measures introduced in this study differ from another measure of asymmetry, the Zanardi index [14,15]. The Zanardi index is based on the ratio of the difference in the Gini index of subzones $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$ to the Gini index of the inequality zone $(\overline{\text{OBCE}})$. This can result in the Zanardi index value that is positive or negative. However, the Zanardi index lacks a clear economic interpretation and does not directly incorporate the geometry of the inequality zone $(\overline{\text{OBCE}})$, including its area and perimeter. Apart from this scenario, the development of Lorenz curve asymmetry measure, $\mathcal{R}_A$ and $\mathcal{R}_P$, directly incorporates the geometry of the inequality zone $\overline{(\text{OBCE})}$ and is more straightforward. As a result, $\mathcal{R}_A$ and $\mathcal{R}_P$ represent the state of income inequality in society, provide a clear interpretation of income inequality, and have formulation related to the Lorenz function specification, $L(p; \boldsymbol{\theta})$.

2. Basic Theory

2.1 Property of the Lorenz Function

Assuming the defined and continuous function $L(p; \boldsymbol{\theta})$ has a parameter vector $\boldsymbol{\theta}$ in the interval [0,1]. A function $L(p; \boldsymbol{\theta})$ is the Lorenz function if it satisfies the following criteria:

$L(p=0;\boldsymbol{\theta})=0; L(p=1;\boldsymbol{\theta})=1; L'(p;\boldsymbol{\theta})\geq 0;\ L''(p;\boldsymbol{\theta})\geq 0,$

where $L'(p;\boldsymbol{\theta})$ and $L''(p;\boldsymbol{\theta})$ are the first and second derivatives of $L(p;\boldsymbol{\theta})$ with respect to p, respectively. The Lorenz function $L(p;\boldsymbol{\theta})$ is a function with one predictor variable p, so the criteria $L'(p;\boldsymbol{\theta})\geq 0$ and $L''(p;\boldsymbol{\theta})\geq 0$ guarantee that the Lorenz function $L(p;\boldsymbol{\theta})$ is convex in the interval $0\leq p\leq 1$. This convexity is appropriate for describing income inequality concerning the egalitarian line.

2.2 Area

In Figure 4, there is a zone ב bounded by the interval $[\mathrm{u},\mathrm{v}]$, $\mathscr{k}(x)$, and $\mathscr{m}(x)$. The area of this zone can be formulated as shown in reference [16]:

$$\text{area ב} = \int_{\mathrm{u}}^{\mathrm{v}} \left(\mathscr{k}(x)-\mathscr{m}(x)\right) dx. \tag{1}$$

The concept in Eq. (2) is used to construct asymmetry measure $\mathcal{R}_A$, which uses the components in the inequality zone $\overline{\mathrm{OBCE}}$.

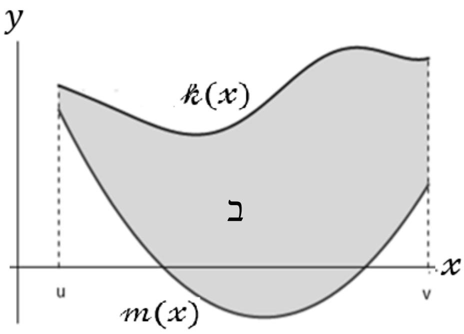


Fig. 4. zone ב bounded by functions $\mathscr{k}(x)$ and $\mathscr{m}(x)$.

2.3 Curve Length

According to Figure 4, curve length formulation of a function is [16]:

$$\text{curve length } \mathscr{k}(x) \text{ on the interval } [\mathrm{u},\mathrm{v}] = \int_{\mathrm{u}}^{\mathrm{v}} \sqrt{1+\left(\frac{d\mathscr{k}(x)}{dx}\right)^2}\, dx. \tag{2}$$

In Eq. (2), when the function $\mathscr{k}(x)$ in the context of income distribution is Lorenz function $L(p;\boldsymbol{\theta})$ bounded by the interval [0,1], then the length of the Lorenz curve represents the Amato index. A longer Lorenz curve implies that the inequality will also have a larger area. The concept presented in Eq. (2) is used for constructing asymmetry measure $\mathcal{R}_P$, drawing from the components in the inequality zone $\overline{\mathrm{OBCE}}$.

2.4 Monte Carlo Integration

If solving the integral cannot be done analytically, then an alternative solution is to use the Monte Carlo approach. The Monte Carlo approach's integral solution is as follows [17,18].

a. Suppose a function $\mathscr{k}(x)$ with upper bound u and lower bound v:

$$\mathcal{J} = \int_{\mathrm{u}}^{\mathrm{v}} \mathscr{k}(x)\, dx\ .$$

b. Assume that $\mathcal{X} \sim \mathrm{Uniform(u, v)}$, then:

$$E\big(\mathcal{k}(\mathcal{X})\big) = \int_{\mathrm{u}}^{\mathrm{v}} \mathcal{k}(x) \frac{1}{\mathrm{v}-\mathrm{u}}\, dx$$

$$\mathcal{J} = (\mathrm{v}-\mathrm{u})\, E\big(\mathcal{k}(\mathcal{X})\big) \approx (\mathrm{v}-\mathrm{u}) \frac{1}{M} \sum_{\hbar=1}^{M} \mathcal{k}(x_{\hbar}), \quad (3)$$

where $x_{\hbar}$ is generated from the Uniform $(\mathrm{u,v})$ distribution. There are M units of $x_{\hbar}$ generated. $\mathcal{J}$ is the integral solution of $\int_{\mathrm{u}}^{\mathrm{v}} \mathcal{k}(x)\, dx$. Completing the integral using the Monte Carlo approach anticipates the formulation of a measure of Lorenz curve asymmetry based on specification of Lorenz functions, which cannot be solved analytically.

2.5 Hypothesis Testing Using Nonparametric Bootstrap

Hypothesis testing through nonparametric bootstrap methods is designed to assess statistical measure when the probability distribution of the population is unknown. In this method, resampling from the empirical distribution of bootstrap sample approximates the population probability distribution. The following outlines the procedure for hypothesis testing using nonparametric bootstrap for the univariate case [19–21].

Suppose a study conducts a hypothesis test as follows:

$$\mathrm{H}_0: \Psi = \Psi_0 \text{ versus } \mathrm{H}_1: \Psi \neq \Psi_0$$

a. Result an estimator Ψ calculated from the original sample data, and the test statistic τ as follows:

$$\tau = \mathcal{N}\big(\widehat{\Psi} - \Psi_0\big)^2, \quad (4)$$

where $\widehat{\Psi}$ is the estimator for Ψ, and τ is a Wald-type test statistic, with a weight of 1 [22,23].

b. Resampling the observational data of size $\mathcal{N}$, for $\mathbb{B}$ times (for $n = 1,2,\dots,\mathbb{B}$), with each resampling having a sample size of $\mathcal{N}$size [20,21]. The statistics for each resampling is then calculated as:

$$\tau_n = \mathcal{N}\big(\widehat{\Psi}^{[n]} - \widehat{\Psi}\big)^2, \quad (5)$$

where $\widehat{\Psi}^{[n]}$ is the estimator of Ψ, calculated from the n-th bootstrap sample.

c. Calculating the p-value $\rho(\tau)$ for a two-way hypothesis test [20,21] as follows:

$$\rho(\tau) = 2 \min\left(\frac{1}{\mathbb{B}} \sum_{n=1}^{\mathbb{B}} I(\tau_n \leq \tau), \frac{1}{\mathbb{B}} \sum_{n=1}^{\mathbb{B}} I(\tau_n > \tau)\right), \quad (6)$$

where $I(.)$ is the indicator function.

d. Based on Step (c), if $\rho(\tau) < \alpha$, then H_0 is rejected. Also, if $\rho(\tau) \geq \alpha$, then H_0 fails to be rejected, where α is the significance level, $0 < \alpha < 1$.

2.6 Transformation of proportion cumulative

Constructing p_i and q_i derived from the observation order index data and household income $(x_j, j = 1, \dots, \mathcal{N},$ where $\mathcal{N}$ represents the number of sample households) is an initial and critical step in estimating the Lorenz function parameters. The procedure for constructing p_i and q_i are presented as follows.

a. Arranging the dataset as shown in Table 1.

Table 1
Layout dataset

Sample Household (j)	Household income
1st household	x_1
2nd household	x_2
⋮	⋮
$\mathcal{N}$-th household	$x_{\mathcal{N}}$

b. Sorting x_j variables in ascending order, signifying the arrangement of households based on their income ownership, as seen in Table 2.

Table 2
Sorting results based on the dataset in Table 1

Ordered sample household (j)	Ordered household income $x_{(j)}$
(1)-st household	$x_{(1)}$
(2)-nd household	$x_{(2)}$
⋮	⋮
($\mathcal{N}$)-th household	$x_{(\mathcal{N})}$

c. Computing p_i and q_i using the cumulative proportion obtained from Table 2:

Table 3
p_i and q_i

p_i	q_i
$p_0 = 0$	$q_0 = (x_{(0)})/\sum_{i=0}^{\mathcal{N}} x_{(i)} = 0, x_{(0)} = 0$
$p_1 = (0+1)/\sum_{i=0}^{\mathcal{N}} i$	$q_1 = (x_{(0)} + x_{(1)})/\sum_{i=0}^{\mathcal{N}} x_{(i)}$
$p_2 = (0+1+2)/\sum_{i=0}^{\mathcal{N}} i$	$q_2 = (x_{(0)} + x_{(1)} + x_{(2)})/\sum_{i=0}^{\mathcal{N}} x_{(i)}$
⋮	⋮
$p_i = (0+1+2+\cdots+i)/\sum_{i=0}^{\mathcal{N}} i$	$q_i = (x_{(0)} + x_{(1)} + x_{(2)} + \cdots + x_{(i)})/\sum_{i=0}^{\mathcal{N}} x_{(i)}$
⋮	⋮
$p_{\mathcal{N}} = \frac{\sum_{i=0}^{\mathcal{N}} i}{\sum_{i=0}^{\mathcal{N}} i} = 1$	$q_{\mathcal{N}} = \frac{\sum_{i=0}^{\mathcal{N}} x_{(i)}}{\sum_{i=0}^{\mathcal{N}} x_{(i)}} = 1$

2.7 Nonlinear Least Squares Using Levenberg-Marquardt Algorithm

The nonlinear least squares estimation method is used in this study to estimate the parameters of Lorenz function $L(p;\boldsymbol{\theta})$. Estimation of parameters in $L(p;\boldsymbol{\theta})$ to obtain an estimator $\widehat{\boldsymbol{\theta}}$ by minimizing $S^*(\boldsymbol{\theta})$:

$$\begin{aligned}\min_{\boldsymbol{\theta}} S^*(\boldsymbol{\theta}) &= \min_{\boldsymbol{\theta}} \frac{1}{2}\big(\boldsymbol{q} - L(\boldsymbol{p};\boldsymbol{\theta})\big)^{\mathrm{T}}\big(\boldsymbol{q} - L(\boldsymbol{p};\boldsymbol{\theta})\big) \\ &= \min_{\boldsymbol{\theta}} \frac{1}{2}\boldsymbol{f}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{\theta}) \\ &= \min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{i=0}^{\mathcal{N}}\big(f_i(\boldsymbol{\theta})\big)^2.\end{aligned} \tag{7}$$

$$\boldsymbol{q} = \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_{\mathcal{N}} \end{bmatrix}, \boldsymbol{p} = \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_{\mathcal{N}} \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_j \\ \vdots \\ \theta_d \end{bmatrix}, \text{and } d\text{: number of parameters.}$$

In Eq. (7), the concept of distance between the empirical and theoretical distribution functions, which is employed for the Kolmogorov-Smirnov test, is used indirectly. It denotes a distance between $\boldsymbol{q}$, representing the experimentally determined cumulative proportion of income, and $L(\boldsymbol{p};\boldsymbol{\theta})$. Because the Lorenz function $L(\boldsymbol{p};\boldsymbol{\theta})$ is also a nonlinear function, it complicates the objective function $S^*(\boldsymbol{\theta})$, which is dependent on the specification of the functional form of $L(\boldsymbol{p};\boldsymbol{\theta})$, so it is difficult to solve analytically for the minimization process of the objective function $S^*(\boldsymbol{\theta})$.

$$S^*(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=0}^{\mathcal{N}}\big(q_i - L(p_i;\boldsymbol{\theta})\big)^2.$$

The first derivative $S^*(\boldsymbol{\theta})$ with respect to p_i, is:

$$\frac{dS^*(\boldsymbol{\theta})}{dp_i} = -\sum_{i=0}^{\mathcal{N}}\big(q_i - L(p_i;\boldsymbol{\theta})\big)L'(p_i;\boldsymbol{\theta})\ .$$

The second derivative $S^*(\boldsymbol{\theta})$ with respect to p_i is:

$$\frac{d^2S^*(\boldsymbol{\theta})}{dp_i^2} = \sum_{i=0}^{\mathcal{N}}\Big(\big(L'(p_i;\boldsymbol{\theta})\big)^2 - \big(q_i - L(p_i;\boldsymbol{\theta})\big)L''(p_i;\boldsymbol{\theta})\Big).$$

The objective function $S^*(\boldsymbol{\theta})$ is a convex function if $\frac{dS^*(\boldsymbol{\theta})}{dp_i} \geq 0$ and $\frac{d^2S^*(\boldsymbol{\theta})}{dp_i^2} \geq 0$, however this depends on how the functions $L'(p;\boldsymbol{\theta})$ and $L''(p;\boldsymbol{\theta})$ are specified. This rule means the objective function $S^*(\boldsymbol{\theta})$ is not always convex. Hence, the estimation procedure of $\boldsymbol{\theta}$ involves iteration to produce a convergent estimated value of $\boldsymbol{\theta}$ and includes supplying a local minimum on $S^*(\boldsymbol{\theta})$. The estimator $\widehat{\boldsymbol{\theta}}$ is a local minimizer [24–26]:

$S^*(\widehat{\boldsymbol{\theta}}) \leq S^*(\boldsymbol{\theta})$ for $\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\| \leq \epsilon$, where ϵ is positive and very small.

In each iteration of $L(\boldsymbol{p};\boldsymbol{\theta})$, the parameter vector $\boldsymbol{\theta}$ will be updated with $\boldsymbol{\theta}+\boldsymbol{\delta}$. To determine $\boldsymbol{\delta}$, the function $L(\boldsymbol{p},\boldsymbol{\theta}+\boldsymbol{\delta})$ is approximated by linearization:

$$\boldsymbol{L}(\boldsymbol{p};\boldsymbol{\theta}+\boldsymbol{\delta}) \approx \boldsymbol{L}(\boldsymbol{p};\boldsymbol{\theta}) + \boldsymbol{J}^*\boldsymbol{\delta}, \qquad \boldsymbol{J}^* = \frac{\partial \boldsymbol{L}(\boldsymbol{p};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

So,

$$\boldsymbol{S}^*(\boldsymbol{\theta}+\boldsymbol{\delta}) \approx \frac{1}{2}(\boldsymbol{q}^* - \boldsymbol{L}(\boldsymbol{p};\boldsymbol{\theta}) - \boldsymbol{J}^*\boldsymbol{\delta})^{\mathrm{T}}(\boldsymbol{q}^* - \boldsymbol{L}(\boldsymbol{p};\boldsymbol{\theta}) - \boldsymbol{J}^*\boldsymbol{\delta})$$

$$\boldsymbol{S}^*(\boldsymbol{\theta}+\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{f}(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{\delta}$$

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\delta}) = \boldsymbol{S}^*(\boldsymbol{\theta}) + \boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{\delta}, \qquad (8)$$

where $-\boldsymbol{J}^* = \boldsymbol{J}$, $\quad \boldsymbol{J} = \dfrac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \boldsymbol{J} \in \mathbb{R}^{\mathcal{N} \times d}$.

In Eq. (8), the first derivative of $\boldsymbol{\delta}$ with respect to $\boldsymbol{\theta}$ is performed, so it becomes:

$$\mathcal{S}'(\boldsymbol{\delta}) = S^{*\prime}(\boldsymbol{\theta} + \boldsymbol{\delta}) = \boldsymbol{J}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{J}^{\mathrm{T}} \boldsymbol{J} \boldsymbol{\delta}. \tag{9}$$

The Gauss-Newton step $\boldsymbol{\delta}_{\mathrm{gn}}$ obtained through the first order condition in Eq. (9) becomes:

$$\boldsymbol{\delta}_{\mathrm{gn}} = -(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{J})^{-\mathbf{1}} \boldsymbol{J}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{\theta}), \tag{10}$$

$\boldsymbol{J}$ is a full rank matrix. In Eq. (10), a modification is made to the component $(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{J})$ by adding the damping parameter λ in the component so that it becomes:

$$\boldsymbol{\delta}_{\mathrm{lm}} = -(\boldsymbol{J}^{\mathrm{T}} \boldsymbol{J} + \lambda \mathbf{I}_d)^{-1} \boldsymbol{J}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{\theta}), \lambda \geq 0, \tag{11}$$

The Levenberg-Marquardt step is denoted by $\boldsymbol{\delta}_{\mathrm{lm}}$, while the identity matrix of dimension $d \times d$ is denoted by $\mathbf{I}_d$. Eq. (11) becomes a Gauss-Newton step if $\lambda = 0$, and it tends to be a Gauss-Newton step if λ is close to zero. When λ is sufficiently large,

$$\boldsymbol{\delta}_{\mathrm{sd}} \cong -\frac{1}{\lambda} \boldsymbol{J}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{\theta}). \tag{12}$$

Eq. (12) represents the steepest-descent step. Because the damping parameter λ greatly influences the direction and size of the step, an iteration process is used to obtain the estimator $\widehat{\boldsymbol{\theta}}$. During the iteration process, λ can be updated, which is controlled by the gain ratio [26]:

$$\Phi = \frac{S^*(\boldsymbol{\theta}) - S^*(\boldsymbol{\theta} + \boldsymbol{\delta}_{\mathrm{lm}})}{\frac{1}{2} \boldsymbol{\delta}_{\mathrm{lm}}^{\mathrm{T}} \left(\lambda \boldsymbol{\delta}_{\mathrm{lm}} - \boldsymbol{J}^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{\theta})\right)}. \tag{13}$$

If Φ in Eq. (13) is large, it indicates that $\mathcal{S}(\boldsymbol{\delta}_{\mathrm{lm}})$ is a good approximation for $S^*(\boldsymbol{\theta} + \boldsymbol{\delta}_{\mathrm{lm}})$, and the value of λ can be lowered, which makes the Levenberg-Marquardt step close to the Gauss-Newton step. If Φ is small, it indicates that $\mathcal{S}(\boldsymbol{\delta}_{\mathrm{lm}})$ is a poor approximation for $S^*(\boldsymbol{\theta} + \boldsymbol{\delta}_{\mathrm{lm}})$, so the value of λ should be increased.

The following Levenberg-Marquardt algorithm to minimize $S^*(\boldsymbol{\theta})$ is [24–26] :

1. Set the initial value $\widehat{\boldsymbol{\theta}}_{(0)}$ and $\hat{\lambda}_{(0)} = \lambda^* \max\left(\mathrm{diag}\left(\hat{\boldsymbol{J}}_{(0)}^{\mathrm{T}} \hat{\boldsymbol{J}}_{(0)}\right)\right)$, λ^* is a very small positive number.
2. For $r = 0,1, \ldots.$ calculate:

$$\boldsymbol{\delta}_{\mathrm{lm(r)}} = -\left(\hat{\boldsymbol{J}}_{(r)}^{\mathrm{T}} \hat{\boldsymbol{J}}_{(r)} + \lambda_{(r)} \mathbf{I}_d\right)^{-1} \hat{\boldsymbol{J}}_{(r)}^{\mathrm{T}} \boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}_{(r)}\right)$$

$$\Phi_{(r)} = \frac{S^*\left(\widehat{\boldsymbol{\theta}}_{(r)}\right) - S^*\left(\widehat{\boldsymbol{\theta}}_{(r)} + \boldsymbol{\delta}_{\mathrm{lm(r)}}\right)}{\frac{1}{2} \boldsymbol{\delta}_{\mathrm{lm}(r)}^{\mathrm{T}} \left(\lambda_{(r)} \boldsymbol{\delta}_{\mathrm{lm}(r)} - \hat{\boldsymbol{J}}_{(r)}^{\mathrm{T}} \boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}_{(r)}\right)\right)}$$

   a. If $\Phi_{(r)} > 0$, then:

$$\widehat{\boldsymbol{\theta}}_{(r+1)} = \widehat{\boldsymbol{\theta}}_{(r)} + \boldsymbol{\delta}_{\text{lm}(r)};\ \lambda_{(r+1)} = \lambda_{(r)} \max\left(\frac{1}{3}, 1 - \left(2\Phi_{(r)} - 1\right)^3\right);\ \ell_{(r)} = 2$$

b. If $\Phi_{(r)} < 0$, then:

$$\widehat{\boldsymbol{\theta}}_{(r+1)} = \widehat{\boldsymbol{\theta}}_{(r)} + \boldsymbol{\delta}_{\text{lm}(r)};\ \lambda_{(r+1)} = \lambda_{(r)}\ell_{(r)};\ \ell_{(r+1)} = 2\ell_{(r)}$$

3. The r-th iteration stops if $\left\|\widehat{\boldsymbol{\theta}}_{(r+1)} - \widehat{\boldsymbol{\theta}}_{(r)}\right\| \le \epsilon$, where ϵ is a very small positive number and close to zero.
4. From Step 3, the algorithm produces the estimator $\widehat{\boldsymbol{\theta}}$.

3. Method and Data

3.1 Construction Method of $\mathcal{R}_A$ and $\mathcal{R}_P$ (The Proposed Measure)

The development of the asymmetry measure and testing the symmetry of the Lorenz curve to achieve the research objectives follow the following steps:

(1) Identify the equations that form the inequality subzones of $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$ as presented in Figure 1.
(2) Identify the intersection points of the inequality subzones of $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$ based on Figure 1.
(3) Step (2) results are the upper and lower bounds used to calculate the area and perimeter of the area bounded by the egalitarian line, the Lorenz curve, and the line $\overline{\text{AD}}$.
(4) Construct a general formulation of the area for the inequality subzones of $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$ bounded by the interval $[\text{u}, \text{v}]$ using Eq. (1). This asymmetry measure is formed based on the ratio of the area $\overline{\text{OBE}}$ to the area $\overline{\text{BEC}}$. This asymmetry measure is denoted as $\mathcal{R}_A$.
(5) Construct a general formulation of the perimeter for the inequality subzones of $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$ bounded by the interval [u, v] using Eq. (2). This second measure of asymmetry is formed based on the ratio of the perimeter $\overline{\text{OBE}}$ to the perimeter $\overline{\text{BEC}}$, which is denoted as $\mathcal{R}_P$.
(6) Formulate the asymmetry measures of the Lorenz function $L(p; \boldsymbol{\theta})$, $\mathcal{R}_A$ and $\mathcal{R}_P$, based on Steps (4) and (5). $\mathcal{R}_A$ and $\mathcal{R}_P$ can be calculated when the estimated value of the vector parameter $\boldsymbol{\theta}$ is known. The Levenberg-Marquardt algorithm is used to estimate the Lorenz function parameters as described in Section 2.7.
(7) Suppose the integral solution in Steps (4), (5), and (6) cannot be solved analytically so that no close-form solution is obtained. The integral solution is approached using the Monte Carlo through Eq. (3) in Section 2.4.
(8) Construct a formulation that establishes a relationship between $\mathcal{R}_A$ and the Gini index. This formulation complements the information provided by $\mathcal{R}_A$ since it is a relative measure that does not reflect the magnitude of inequality. Therefore, it is essential to formulate the relationship between $\mathcal{R}_A$ and the Gini index.

3.2 Data

The data utilized in this analysis came from three sources:

a. Household consumption expenditure data for one month. This information is derived from SUSENAS March 2020, conducted by BPS-Statistics of Banten Province. The object of study is a sample of Banten Province households totaling 6,964 households. Regardless of the origin of the items, household consumption is distinguished from food and non-food consumption. It is restricted to expenditure for home needs, omitting business expenditure, or expenditure supplied to third parties. Food expenditures are estimated for the previous week, whereas non-food expenditures are calculated for the previous month and 12 months. The average monthly spend for food and non-food consumption is calculated. So, consumption expenditure is a proxy of income.

b. The total income employment data per family sourced from the Ghana Statistical Service's Ghana Living Standards Survey IV was completed in 1998. Out of the targeted 6,000 households, 5,999 are successfully enumerated. However, one household is excluded from the dataset due to incomplete data, resulting in a total of 5,998 households in the dataset.
c. Household income data obtained from Statistics South Africa's Living Conditions Survey 2014/2015, performed from October 13, 2014, to October 25, 2015. The dataset included 23,380 households. Household income includes all receipts in the form of money and goods received by all household members in exchange for employment or capital investment, as well as receipts collected from other sources such as social grants, pensions, Etc.

4. Results and Discussion

4.1 The Measure of Lorenz Curve Asymmetry Based on Area Ratio: $\mathcal{R}_A$

The asymmetry measure of the Lorenz curve is constructed based on the ratio of the areas and perimeters of the subzones $\overline{\mathrm{OBE}}$ and $\overline{\mathrm{BEC}}$, respectively. This measure aims to identify the dominant concentration of income inequality between two income groups: the group with cumulative proportion of income $q \le q_+$ and the group with cumulative proportion of income $q > q_+$, as shown in Figure 1. It illustrates three component lines used in constructing the asymmetry measure: the egalitarian line $\overline{\mathrm{OC}}$, the line $\overline{\mathrm{AD}}$ serving as the symmetry line, and the Lorenz curve $\overline{\mathrm{OEC}}$. The egalitarian line is formed by points O, B, and C, with the angle between this line and the horizontal axis p being 45° (or the angle between the egalitarian line and the vertical axis q being 45°). The Lorenz curve is the curve that connects points O, E, and C, lying below the egalitarian line. The area between the egalitarian line and the Lorenz curve represents the magnitude of income inequality. When the value of income inequality is zero, the Lorenz curve coincides with the egalitarian line, indicating equal income distribution in society. The egalitarian line is mathematically represented by the equation $q = p$ (the line connecting points O, B, and C in Figure 1).

Various measures of income inequality have emerged from this concept of inequality areas, including the Gini index (G). Based on the visualization in Figure 1, the Gini index is given by the following formula [27]:

$$G = \frac{\text{Area }\ \overline{\mathrm{OBE}} + \text{Area }\overline{\mathrm{OEC}}}{\text{Area }\overline{\mathrm{OAC}}}\ , 0 \le G \le 1. \tag{14}$$

Point B represents the intersection between the egalitarian line and the line $\overline{\mathrm{AD}}$. Therefore, point B has coordinates (p_+, q_+), serving as the midpoint on the egalitarian line and the center point of the zone $\overline{\mathrm{OACD}}$. Since p and q denote the cumulative proportion of income and the cumulative proportion of households, respectively, they fall within the interval $[0,1]$. As a result, point B, the center point of the zone $\overline{\mathrm{OACD}}$, has coordinates $(0.5, 0.5)$, with $p_+ = 0.5$ and $q_+ = 0.5$.

The equation of the egalitarian line is:

$$q = p\ . \tag{15}$$

The equation of the line $\overline{\mathrm{AD}}$ is:

$$p + q = 1, \tag{16}$$

input Eq. (15) into (16) thus:

$$p + p = 1$$
$$p = 0.5. \tag{17}$$

Then, input Eq. (17) into (16), thus obtained $q = 0.5$. So, point B has coordinates (0.5, 0.5), so $p_+ = 0.5$ and $q_+ = 0.5$.

Point q_+ represents the midpoint on the vertical axis q, dividing the cumulative proportion of income into two levels: the cumulative proportion of income 50% and below ($q \leq 0.5$), and the cumulative proportion of income more than 50% ($q > 0.5$). Point E represents the intersection between the Lorenz curve and the line $\overline{\text{AD}}$, with coordinates (p_{++}, q_{++}). The line $\overline{\text{BE}}$, which is a segment of the line $\overline{\text{AD}}$, divides the inequality zone into two subzones: $\overline{\text{OBE}}$ and $\overline{\text{BEC}}$. The asymmetry measure constructed in this study is based on the ratio between the area $\overline{\text{OBE}}$ to the area $\overline{\text{BEC}}$, which is formulated as follows:

$$\mathcal{R}_A = \frac{\text{Area } \overline{\text{OBE}}}{\text{Area } \overline{\text{BEC}}}. \tag{18}$$

$\mathcal{R}_A$ measures the Lorenz curve asymmetry based on the area ratio between the two inequality subzones. $\mathcal{R}_A$ can take several possible values, namely:

a. $\mathcal{R}_A = 1$ occurs when the area $\overline{\text{OBE}}$ equals the area $\overline{\text{BEC}}$. It indicates that both groups equally contribute to income inequality in the group with cumulative proportion of income $q \leq 0.5$ (the cumulative proportion of income 50% and below) and the group with cumulative proportion of income $q > 0.5$ (the cumulative proportion of income more than 50%).
b. $\mathcal{R}_A > 1$ occurs when the area $\overline{\text{OBE}}$ is greater than the area $\overline{\text{BEC}}$. This condition indicates that the group with cumulative proportion of income $q \leq 0.5$ contributes more to the creation of income inequality (income gap level) compared to the group with cumulative proportion of income $q > 0.5$.
c. $\mathcal{R}_A < 1$ occurs when the area $\overline{\text{OBE}}$ is less than the area $\overline{\text{BEC}}$. This condition indicates that the group with the cumulative proportion of income $q > 0.5$ has a greater contribution to the creation of income inequality (income gap level) than the group with cumulative proportion of income $q \leq 0.5$.

Proposition 1

Based on Eq. (18), the measure of Lorenz curve asymmetry based on the ratio of area subzone inequality is:

$$\mathcal{R}_A = \frac{-\frac{1}{4} - \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_0^{p_{++}} L(p;\boldsymbol{\theta})\,dp}{\frac{3}{4} + \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_{p_{++}}^1 L(p;\boldsymbol{\theta})\,dp}, \tag{19}$$

where p_{++} is the Kolkata index derived from the Lorenz function $L(p;\boldsymbol{\theta})$.

Proof:

a. Area $\overline{\text{OBE}}$

$\overline{\text{OBE}}$ is formed from the intersection of the egalitarian line ($q = p$), the Lorenz curve, and the line $\overline{\text{AD}}$ ($q = 1 - p$).

$$\text{Area } \overline{\text{OBE}} = \int_0^{p_+} p - L(p;\boldsymbol{\theta})dp + \int_{p_+}^{p_{++}} \big((1-p) - L(p;\boldsymbol{\theta})\big)dp. \tag{20}$$

Because $p_+ = 0.5$, Eq. (20) becomes:

$$\begin{aligned}\text{Area } \overline{\text{OBE}} &= \int_0^{0.5} \big(p - L(p;\boldsymbol{\theta})\big)\, dp + \int_{0.5}^{p_{++}} \big((1-p) - L(p;\boldsymbol{\theta})\big)\, dp \\ &= -\frac{1}{4} - \left(\frac{1}{2}p_{2++}^2 - p_{++}\right) - \int_0^{p_{++}} L(p;\boldsymbol{\theta})\, dp\ .\end{aligned} \tag{21}$$

b. Area $\overline{\text{BEC}}$
$\overline{\text{BEC}}$ is formed by the intersection of the egalitarian line $(q = p)$, the Lorenz curve, and the line $\overline{\text{AD}}$ $(q = 1 - p)$.

$$\text{Area } \overline{\text{BEC}} = \int_{p_+}^{p_{++}} \big(p - (1-p)\big)\, dp + \int_{p_{++}}^{1} \big(p - L(p;\boldsymbol{\theta})\big)\, dp \tag{22}$$

Since $p_+ = 0.5$, Eq. (22) becomes:

$$\begin{aligned}\text{Area } \overline{\text{BEC}} &= \int_{0,5}^{p_{++}} \big(p - (1-p)\big)\, dp + \int_{p_{++}}^{1} \big(p - L(p;\boldsymbol{\theta})\big)\, dp \\ &= \frac{3}{4} + \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_{p_{++}}^{1} L(p;\boldsymbol{\theta})\, dp.\end{aligned} \tag{23}$$

Eqs. (21) and (23) are inserted into (18) to obtain the following:

$$\mathcal{R}_A = \frac{-\frac{1}{4} - \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_0^{p_{++}} L(p;\boldsymbol{\theta})\, dp}{\frac{3}{4} + \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_{p_{++}}^{1} L(p;\boldsymbol{\theta})\, dp}\ \text{, Eq (19) is proved.}$$

In Eq. (19), we have p_{++}, which is part of the coordinates of point E, denoted as (p_{++}, q_{++}). Point E represents the intersection of the Lorenz curve and the line $q = 1 - p$. Geometrically, p_{++} is known as the Kolkata index [28–30], and it satisfies the following equations:

$$L(p_{++}) + p_{++} = 1 \tag{24}$$
$$q_{++} + p_{++} = 1\ . \tag{25}$$

The Kolkata index's determination depends on the Lorenz function specification used in Eq. (24). The solution of Eq. (24) can be obtained analytically or numerically, depending on the specification of $L(p;\boldsymbol{\theta})$. The Monte Carlo approach is employed to approximate the measure of Lorenz curve asymmetry $\mathcal{R}_A$ (21), which is given by:

$$\mathcal{R}_A \approx \frac{-\frac{1}{4} - \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \left(\frac{p_{++}}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} L(\tilde{z}_\hbar;\boldsymbol{\theta})\right)}{\frac{3}{4} + \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \left(\frac{(1-p_{++})}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} L(\widetilde{w}_\hbar;\boldsymbol{\theta})\right)}, \hbar = 1, \dots, \mathcal{M}, \tag{26}$$

where $\tilde{z}_{\hbar}$ is generated from the Uniform distribution $(0, p_{++})$. The number of $\tilde{z}_{\hbar}$ units generated is $\mathcal{M}$ units. $\widetilde{w}_{\hbar}$ is generated from the Uniform distribution $(p_{++}, 1)$. There are $\mathcal{M}$ units of $\widetilde{w}_{\hbar}$ generated. $\mathcal{M}$ is a large number.

Since the Lorenz function $L(p;\boldsymbol{\theta})$ contains a vector of parameters to be estimated using the Levenberg-Marquardt algorithm as described in Section 2.7. So, Eqs. (19) and (26) become Eqs. (27) and (28), respectively.

$$\hat{\mathcal{R}}_A = \frac{-\frac{1}{4} - \left(\frac{1}{2}(\widehat{p_{++}})^2 - \widehat{p_{++}}\right) - \int_0^{\widehat{p_{++}}} L\left(p;\widehat{\boldsymbol{\theta}}\right) dp}{\frac{3}{4} + \left(\frac{1}{2}(\widehat{p_{++}})^2 - \widehat{p_{++}}\right) - \int_{\widehat{p_{++}}}^1 L\left(p;\widehat{\boldsymbol{\theta}}\right) dp}, \tag{27}$$

$$\approx \frac{-\frac{1}{4} - \left(\frac{1}{2}(\widehat{p_{++}})^2 - \widehat{p_{++}}\right) - \left(\frac{\widehat{p_{++}}}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} L\left(\tilde{z}_{\hbar};\widehat{\boldsymbol{\theta}}\right)\right)}{\frac{3}{4} + \left(\frac{1}{2}(\widehat{p_{++}})^2 - \widehat{p_{++}}\right) - \left(\frac{(1-\widehat{p_{++}})}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} L\left(\widetilde{w}_{\hbar};\widehat{\boldsymbol{\theta}}\right)\right)}, \hbar = 1, \dots, \mathcal{M}\ . \tag{28}$$

$L\left(p;\widehat{\boldsymbol{\theta}}\right)$ is the Lorenz Function based on sample data of size $\mathcal{N}$, where the specification contains the estimator $\widehat{\boldsymbol{\theta}}$.

Geometrically, the Gini index is the ratio of the inequality area to the triangle $\overline{\mathrm{OAC}}$ in Figure1 presented in Eq. (14). Based on Figure 1, the information that the inequality area is the sum of areas $\overline{\mathrm{OBE}}$ and $\overline{\mathrm{BEC}}$, and the area of the triangle $\overline{\mathrm{OAC}}$ is 0.5, so Eq. (14) becomes:

$$G = \frac{\text{Area } \overline{\mathrm{OBE}} + \text{Area } \overline{\mathrm{BEC}}}{\text{Area } \overline{\mathrm{OAC}}}$$

$$= 2 \times (\text{Area } \overline{\mathrm{OBE}} + \text{Area } \overline{\mathrm{BEC}})\ . \tag{29}$$

Proposition 2

The relation between the asymmetry measures $\mathcal{R}_A$, and the Gini index is:

$$\mathcal{R}_A = \frac{G}{\frac{3}{2} + (p_{++}^2 - 2p_{++}) - 2\int_{p_{++}}^1 L(p;\boldsymbol{\theta})dp} - 1\ . \tag{30}$$

Proof:

Relate Eq. (29) and $\mathcal{R}_A$ so that it becomes:

$$G = 2((\mathcal{R}_A \times \text{Area } \overline{\mathrm{BEC}}) + \text{Area } \overline{\mathrm{BEC}})$$

$$\mathcal{R}_A = \frac{G}{2 \times \text{Area } \overline{\mathrm{BEC}}} - 1\ . \tag{31}$$

Insert Eq. (23) into (31), and it becomes:

$$\mathcal{R}_A = \frac{G}{2\left(\frac{3}{4} + \left(\frac{1}{2}p_{++}^2 - p_{++}\right) - \int_{p_{++}}^1 L(p;\boldsymbol{\theta})dp\right)} - 1$$

$$= \frac{G}{\frac{3}{2} + (p_{++}^2 - 2p_{++}) - 2\int_{p_{++}}^{1} L(p;\boldsymbol{\theta})dp} - 1\ .\ \text{So, Eq. (30) is proved.}$$

4.2 The Measure of Lorenz Curve Asymmetry Based on Perimeter Ratio: $\mathcal{R}_P$

The asymmetry measure is constructed based on the perimeter ratio between the two subzones of inequality. Therefore, this study also proposes a Lorenz curve asymmetry measure based on the ratio between the perimeters $\overline{\mathrm{OBE}}$ and $\overline{\mathrm{BEC}}$, as illustrated in Figure 1. The measure of Lorenz curve asymmetry using the perimeter approach is formulated as follows:

$$\mathcal{R}_P = \frac{\text{perimeter } \overline{\mathrm{OBE}}}{\text{perimeter } \overline{\mathrm{BEC}}} = \frac{\text{length } \overline{\mathrm{OB}} + \text{length } \overline{\mathrm{BE}} + \text{length } \overline{\mathrm{OE}}}{\text{length } \overline{\mathrm{BC}} + \text{length } \overline{\mathrm{BE}} + \text{length } \overline{\mathrm{EC}}}. \tag{32}$$

$\mathcal{R}_P$ represents the Lorenz curve asymmetry measure based on the perimeter ratio of subzone inequality. It can take on several possible values, including:

a. $\mathcal{R}_P = 1$ occurs when the perimeter $\overline{\mathrm{OBE}}$ equals the perimeter $\overline{\mathrm{BEC}}$. This value indicates that both groups contribute equally to income inequality (income gap level) in the group with cumulative proportion of income $q \le 0.5$ and the group with the cumulative proportion of income $q > 0.5$.
b. $\mathcal{R}_P > 1$ occurs when the perimeter $\overline{\mathrm{OBE}}$ is greater than the perimeter $\overline{\mathrm{BEC}}$. This value indicates that the group with cumulative proportion of income $q \le 0.5$ contributes more to creating income inequality (income gap level) than the group with cumulative proportion of income $q > 0.5$.
c. $\mathcal{R}_P < 1$ occurs when the perimeter $\overline{\mathrm{OBE}}$ is less than the perimeter $\overline{\mathrm{BEC}}$. This value indicates that the group with cumulative proportion of income $q > 0.5$ has a greater contribution to the creation of income inequality (income gap level) compared to the group with cumulative proportion of income $q \le 0.5$.

Based on Figure 1 and Section 4.1, the following information can be obtained:

(1) Length $\overline{\mathrm{OA}} = \overline{\mathrm{OD}} = \overline{\mathrm{AC}} = \overline{\mathrm{CD}} = 1$.
(2) Length $\overline{\mathrm{OC}} = \overline{\mathrm{AD}}$.
Length $\overline{\mathrm{OC}} = \sqrt{(\overline{\mathrm{OA}})^2 + (\overline{\mathrm{AC}})^2} = \sqrt{1^2 + 1^2} = \sqrt{2}$.
(3) Length $\overline{\mathrm{OB}} = \overline{\mathrm{BC}} = \overline{\mathrm{AB}} = \overline{\mathrm{BD}} = \overline{\mathrm{OC}}/2 = \overline{\mathrm{AD}}/2$, thus:
Length $\overline{\mathrm{OB}} = \overline{\mathrm{BC}} = \overline{\mathrm{AB}} = \overline{\mathrm{BD}} = \frac{\sqrt{2}}{2}$.

The length $\overline{\mathrm{OB}}$ can also be obtained from the coordinates of point B, which are $p_+ = 0.5, q_+ = 0.5$, and the triangle $\overline{\mathrm{O}p_+\mathrm{B}}$ is a right triangle, thus:

$$\text{Length } \overline{\mathrm{OB}} = \sqrt{p_+^2 + q_+^2} = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}\ . \tag{33}$$

Proposition 3

The length $\overline{\mathrm{BE}}$ is a hypotenuse of a right triangle (see Figure 1):

$$\text{Length } \overline{\mathrm{BE}} = \left(p_{++} - \frac{1}{2}\right)\sqrt{2}\,. \tag{34}$$

Proof:
Since the coordinates of point B, i.e. $p_{+} = 0.5, q_{+} = 0.5$ and $\overline{\mathrm{BE}}$ is the hypotenuse, it follows:

$$\text{Length } \overline{\mathrm{BE}} = \sqrt{\left(p_{++} - \frac{1}{2}\right)^2 + \left(\frac{1}{2} - q_{++}\right)^2}\,. \tag{35}$$

Based on Eq. (25), Eq. (35) becomes (34).

Proposition 4
The measure of Lorenz curve asymmetry based on the ratio of the perimeter of $\overline{\mathrm{OBE}}$ to $\overline{\mathrm{BEC}}$ is:

$$\mathcal{R}_P = \frac{p_{++}\sqrt{2} + \int_0^{p_{++}} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp}{p_{++}\sqrt{2} + \int_{p_{++}}^{1} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp}\,. \tag{36}$$

Proof:
Length $\overline{\mathrm{OE}}$ and $\overline{\mathrm{EC}}$ are, respectively:

$$\text{Length } \overline{\mathrm{OE}} = \int_0^{p_{++}} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp\,. \tag{37}$$

$$\text{Length } \overline{\mathrm{EC}} = \int_{p_{++}}^{1} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp\,. \tag{38}$$

Eqs. (33), (34), (37), and (38) are inserted into Eq. (32), thus:

$$\mathcal{R}_P = \frac{p_{++}\sqrt{2} + \int_0^{p_{++}} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp}{p_{++}\sqrt{2} + \int_{p_{++}}^{1} \sqrt{1 + \left(L'(p;\boldsymbol{\theta})\right)^2}\, dp}\,.$$

So, Eq. (36) is proved. It comprises integral components in both the numerator and denominator. If the integral component cannot be solved analytically, the Monte Carlo approach can obtain the solution using Eq. (3). Therefore, the measure of asymmetry $\mathcal{R}_P$, using the Monte Carlo approach, is given by:

$$\mathcal{R}_P \approx \frac{p_{++}\sqrt{2} + \frac{p_{++}}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(L'(\tilde{z}_{\hbar};\boldsymbol{\theta})\right)^2}}{p_{++}\sqrt{2} + \left(\frac{(1 - p_{++})}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(L'(\widetilde{w}_{\hbar};\boldsymbol{\theta})\right)^2}\right)}\,, \tag{39}$$

where $\tilde{z}_{\hbar}$ is generated from the Uniform distribution $(0, p_{++})$, the number of $\tilde{z}_{\hbar}$ units generated is $\mathcal{M}$ units, and $\widetilde{w}_{\hbar}$ is generated from the Uniform distribution $(p_{++}, 1)$. There are $\mathcal{M}$ units of $\widetilde{w}_{\hbar}$ generated.

Since the Lorenz function $L(p;\boldsymbol{\theta})$ contains a vector of parameters to be estimated using the Levenberg-Marquardt algorithm as described in Section 2.7, the formulations (36) and (39) become Eqs. (40) and (41), respectively:

$$\hat{\mathcal{R}}_P = \frac{\widehat{p_{++}}\sqrt{2} + \int_0^{\widehat{p_{++}}} \sqrt{1 + \left(L'(p;\widehat{\boldsymbol{\theta}})\right)^2}\, dp}{\widehat{p_{++}}\sqrt{2} + \int_{\widehat{p_{++}}}^1 \sqrt{1 + \left(L'(p;\widehat{\boldsymbol{\theta}})\right)^2}\, dp} \tag{40}$$

$$\approx \frac{\widehat{p_{++}}\sqrt{2} + \frac{\widehat{p_{++}}}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(L'(\tilde{z}_{\hbar};\widehat{\boldsymbol{\theta}})\right)^2}}{\widehat{p_{++}}\sqrt{2} + \left(\frac{(1-\widehat{p_{++}})}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(L'(\widetilde{w}_{\hbar};\widehat{\boldsymbol{\theta}})\right)^2}\right)}\,. \tag{41}$$

4.3 Formulations of $\mathcal{R}_A$ and $\mathcal{R}_P$ Based on Four Types of Lorenz functions

Several well-known Lorenz function specifications, namely derive the $\mathcal{R}_A$ and $\mathcal{R}_P$ formulations:

a. Lorenz-Sarabia Function

The Lorenz-Sarabia function has a specification as presented in Eq. (42) [11]:

$$L_{SB}(p;\gamma) = p^{\gamma}, \gamma \geq 1;\ 0 \leq p \leq 1\,. \tag{42}$$

The $\mathcal{R}_A$ formulation derived from Eq. (42) is

$$\mathcal{R}_{A_{SB}} = \frac{-\frac{1}{4} - \left(\frac{1}{2}\kappa_{SB}^2 - \kappa_{SB}\right) - \frac{(\kappa_{SB})^{\gamma+1}}{\gamma+1}}{\frac{3}{4} + \left(\frac{1}{2}\kappa_{SB}^2 - \kappa_{SB}\right) - \frac{1}{\gamma+1}\left(1 - (\kappa_{SB})^{\gamma+1}\right)}, \tag{43}$$

where κ_{SB} is the Kolkata index derived from Eq. (42).

The $\mathcal{R}_P$ formulation derived from Eq. (42) is

$$\mathcal{R}_{P_{SB}} = \frac{\kappa_{SB}\sqrt{2} + \mathcal{V}}{\kappa_{SB}\sqrt{2} + \left({}_2\mathrm{F}_1\left(\frac{1}{2(\gamma-1)}, -\frac{1}{2}, 1 + \frac{1}{2(\gamma-1)}, -\gamma^2\right) - \mathcal{V}\right)}, \tag{44}$$

where:

$$\mathcal{V} = \kappa_{SB}\ {}_2\mathrm{F}_1\left(\frac{1}{2(\gamma-1)}, -\frac{1}{2}; 1 + \frac{1}{2(\gamma-1)}; -\left(\gamma(\kappa_{SB})^{\gamma-1}\right)^2\right)$$

${}_2\mathrm{F}_1(\mathcal{A},\mathcal{B};\mathcal{C};\mathcal{D})$ is the hypergeometric function.

b. Lorenz-Hossain & Saeki (Lorenz-HS) Function

The Lorenz-HS function has a specification as presented in Eq. (45) [4]:

$$L_{HS}(p;\sigma_1,\sigma_2,\sigma_3) = p^{\sigma_1} e^{\sigma_2(p-1)}\left(1 - (1-p)^{\sigma_3}\right), \sigma_1 \geq 0; \sigma_2 \geq 0; 0 < \sigma_3 \leq 1; 0 \leq p \leq 1. \tag{45}$$

The $\mathcal{R}_A$ and $\mathcal{R}_P$ formulations derived based on Eq. (45) are not close-form, so the formulations use a Monte Carlo approach. The $\mathcal{R}_A$ formulation derived from Eq. (45) using the Monte Carlo approach is

$$\mathcal{R}_{A_{HS}} \approx \frac{-\frac{1}{4}-\left(\frac{1}{2}\kappa_{HS}^2-\kappa_{HS}\right)-\left(\frac{\kappa_{HS}}{\mathcal{M}}\sum_{\hbar=1}^{\mathcal{M}}\left(\widetilde{\tilde{z}_J}\right)^{\sigma_1}e^{\sigma_2\left(\widetilde{\tilde{z}_\hbar}-1\right)}\left(1-\left(1-\widetilde{\tilde{z}_\hbar}\right)^{\sigma_3}\right)\right)}{\frac{3}{4}+\left(\frac{1}{2}\kappa_{HS}^2-\kappa_{HS}\right)-\left(\frac{(1-\kappa_{HS})}{\mathcal{M}}\sum_{\hbar=1}^{\mathcal{M}}\left(\widetilde{\tilde{w}_J}\right)^{\sigma_1}e^{\sigma_2\left(\widetilde{\tilde{w}_\hbar}-1\right)}\left(1-\left(1-\widetilde{\tilde{w}_\hbar}\right)^{\sigma_3}\right)\right)}, \quad (46)$$

where κ_{HS} is the Kolkata index derived from $L_{HS}(p)$, $\widetilde{\tilde{z}_\hbar}$ is generated from the Uniform distribution $(0, \kappa_{HS})$. The number of $\widetilde{\tilde{z}_\hbar}$ units generated is M units. $\widetilde{\tilde{w}_\hbar}$ is generated from the Uniform distribution $(\kappa_{HS}, 1)$. The number of units $\widetilde{\tilde{w}_\hbar}$ generated is $\mathcal{M}$ units. $\mathcal{M}$ is a large number. The $\mathcal{R}_P$ formulation derived from Eq. (45) using the Monte Carlo approach is

$$\mathcal{R}_{P_{HS}} \approx \frac{\kappa_{HS}\sqrt{2}+\frac{\kappa_{HS}}{\mathcal{M}}\sum_{\hbar=1}^{\mathcal{M}}\sqrt{1+\left(L'_{HS}\left(\widetilde{\tilde{z}_\hbar};\sigma_1,\sigma_2,\sigma_3\right)\right)^2}}{\kappa_{HS}\sqrt{2}+\left(\frac{(1-\kappa_{HS})}{\mathcal{M}}\sum_{\hbar=1}^{\mathcal{M}}\sqrt{1+\left(L'_{HS}\left(\tilde{w}_\hbar;\widetilde{\sigma_1,}\sigma_2,\sigma_3\right)\right)^2}\right)}, \quad (47)$$

$L'_{HS}(.)$ is the first derivative of $L_{HS}(p;\sigma_1,\sigma_2,\sigma_3)$ with respect to p.

c. Lorenz-Rohde Function

The Lorenz-Rohde function has a specification as presented in Eq. (48) [7]:

$$L_R(p;a_R) = \frac{(a_R-1)p}{a_R-p}, a_R > 1, 0 \le p \le 1 \quad (48)$$

The $\mathcal{R}_A$ formulation derived from Eq. (48) is

$$\mathcal{R}_{A_R} = \frac{-\frac{1}{4}-\left(\frac{1}{2}\kappa_R^2-\kappa_R\right)-(1-a_R)\left(\kappa_R+a_R\log\left(\left|1-\frac{\kappa_R}{a_R}\right|\right)\right)}{\frac{3}{4}+\left(\frac{1}{2}\kappa_R^2-\kappa_R\right)-(1-a_R)\left(1-\kappa_R+a_R\log\left(\left|\frac{1-a_R}{\kappa_R-a_R}\right|\right)\right)}, \quad (49)$$

$\kappa_R = a_R - \sqrt{a_R^2 - a_R}$, and κ_R is the Kolkata index derived from Eq. (48).

The $\mathcal{R}_P$ formulation derived from Eq. (48) is:

$$\mathcal{R}_{P_R} = \frac{\kappa_R\sqrt{2}+(\kappa_R-a_R)\ {}_2F_1\left(-\frac{1}{2},-\frac{1}{4};\frac{3}{4};-\frac{a_R^2(a_R-1)^2}{(a_R-\kappa_R)^4}\right)+a_R\ {}_2F_1\left(-\frac{1}{2},-\frac{1}{4};\frac{3}{4};-\frac{(a_R-1)^2}{a_R^2}\right)}{\kappa_R\sqrt{2}+(1-a_R)\ {}_2F_1\left(-\frac{1}{2},-\frac{1}{4};\frac{3}{4};-\frac{a_R^2}{(a_R-1)^2}\right)-(\kappa_R-a_R)\ {}_2F_1\left(-\frac{1}{2},-\frac{1}{4};\frac{3}{4};-\frac{a_R^2(a_R-1)^2}{(a_R-\kappa_R)^4}\right)}. \quad (50)$$

d. Lorenz-Chotikapanich Functions

The Lorenz-Chotikapanich function is formulated as follows [9]:

$$L_C(p;\aleph) = \frac{e^{\aleph p}-1}{e^{\aleph}-1}, \aleph > 1, 0 \le p \le 1. \quad (51)$$

The $\mathcal{R}_A$ formulation derived from the Lorenz-Chotikapanich function (51) using the Monte Carlo approach is as follows:

$$\mathcal{R}_{A_C} \approx \frac{-\frac{1}{4} - \left(\frac{1}{2}\kappa_C^2 - \kappa_C\right) - \left(\frac{\kappa_C}{\mathcal{M}} \sum_{j=1}^{\mathcal{M}} \frac{e^{\aleph \tilde{z}_{\hbar}^{***}} - 1}{e^{\aleph} - 1}\right)}{\frac{3}{4} + \left(\frac{1}{2}\kappa_C^2 - \kappa_C\right) - \left(\frac{(1-\kappa_C)}{\mathcal{M}} \sum_{j=1}^{\mathcal{M}} \frac{e^{\aleph \widetilde{w}_{\hbar}^{***}} - 1}{e^{\aleph} - 1}\right)}, \tag{52}$$

where κ_C is the Kolkata index derived from the Lorenz-Chotikapanich function (51), $\tilde{z}_{\hbar}^{***}$ is generated from the Uniform distribution $(0, \kappa_C)$ with the number of $\tilde{z}_{\hbar}^{***}$ units generated being $\mathcal{M}$ units. $\widetilde{w}_{\hbar}^{***}$ is generated from the Uniform distribution $(\kappa_C, 1)$ with the number of $\widetilde{w}_{\hbar}^{***}$ units generated being $\mathcal{M}$ units.

The $\mathcal{R}_P$ formulation derived from the Lorenz-Chotikapanich function (51) using the Monte Carlo approach is as follows:

$$\mathcal{R}_{P_C} \approx \frac{\kappa_C\sqrt{2} + \frac{\kappa_C}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(\frac{\aleph e^{\aleph \tilde{z}_{\hbar}^{***}}}{e^{\aleph} - 1}\right)^2}}{\kappa_C\sqrt{2} + \left(\frac{(1-\kappa_C)}{\mathcal{M}} \sum_{\hbar=1}^{\mathcal{M}} \sqrt{1 + \left(\frac{\aleph e^{\aleph \widetilde{w}_{\hbar}^{***}}}{e^{\aleph} - 1}\right)^2}\right)}. \tag{53}$$

4.3 Hypothesis Testing of Lorenz Curve Symmetry Using $\mathcal{R}_A$ and $\mathcal{R}_P$

The Lorenz curve symmetry hypothesis testing is constructed using the nonparametric bootstrap, as explained in Section 2.5. The flowchart of the Lorenz curve symmetry hypothesis testing based on $\mathcal{R}_A$ and $\mathcal{R}_P$ is presented in Figure 5. The following is an explanation of the steps to test the Lorenz curve symmetry hypothesis:

(1) Hypothesis statement
$\mathrm{H}_0: \mathcal{R}_A = 1$ (Symmetrical Lorenz curve) versus $\mathrm{H}_1: \mathcal{R}_A \neq 1$ (Asymmetrical Lorenz curve)
Or
$\mathrm{H}_0: \mathcal{R}_P = 1$ (Symmetrical Lorenz curve) versus $\mathrm{H}_1: \mathcal{R}_P \neq 1$ (Asymmetrical Lorenz curve)

(2) From the observation data $x_j, j = 1, \dots, \mathcal{N}$ is cumulatively proportioned to obtain data p_i and $q_i, i = 0,1, \dots, \mathcal{N}; p_0 = 0, p_{\mathcal{N}} = 1, q_0 = 1, q_{\mathcal{N}} = 1,$ where the cumulative proportion process is explained in Section 2.6. Furthermore, these p_i and q_i data are input materials for Lorenz curve fitting. Determine the best fitting Lorenz function to the empirical Lorenz curve from the candidates of the Lorenz function based on the minimum mean squared error (MSE) and/or mean absolute error (MAE).

(3) From Step (2):

i. Calculate $\hat{\mathcal{R}}_A$ using Eq. (27) or (28), and calculate $\tau(\hat{\mathcal{R}}_A)$ based on Eq. (4).

$$\tau(\hat{\mathcal{R}}_A) = (\mathcal{N} + 1)(\hat{\mathcal{R}}_A - 1)^2. \tag{54}$$

Suppose the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions fit the empirical Lorenz curve. In this case, $\hat{\mathcal{R}}_A$ can be replaced by estimation of $\mathcal{R}_{A_{SB}}$ (43), $\mathcal{R}_{A_{HS}}$ (46), $\mathcal{R}_{A_R}$ (49), or $\mathcal{R}_{A_C}$ (52), respectively.

ii. Calculate $\hat{\mathcal{R}}_P$ using Eq. (40) or (41), and calculate $\tau(\hat{\mathcal{R}}_P)$ based on Eq. (4).

$$\tau(\hat{\mathcal{R}}_P) = (\mathcal{N} + 1)(\hat{\mathcal{R}}_P - 1)^2. \tag{55}$$

Suppose the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions fit the empirical Lorenz curve. In that case, $\hat{\mathcal{R}}_A$ can be replaced by estimation of $\mathcal{R}_{P_{SB}}$ (44), $\mathcal{R}_{P_{HS}}$ (47), $\mathcal{R}_{P_R}$ (50), or $\mathcal{R}_{P_C}$ (53), respectively.

(4) Perform resampling $\mathbb{B}$ times ($n = 1, \dots, \mathbb{B}$) on observation data x_j with a bootstrap sample size of $\mathcal{N}$ units. From the resampling result data:

i. Based on the n-th bootstrap sample, $x_j^{[n]}; n = 1, \dots, \mathbb{B}$, of size N units, process the cumulative proportions to obtain $p_i^{[n]}$ and $q_i^{[n]}$ as described in Section 2.6.

ii. Use the $p_i^{[n]}$ and $q_i^{[n]}$ from Step (4).i for Lorenz curve fitting using the Lorenz function defined.

iii. From the Step (4).ii:

a. Calculate $\hat{\mathcal{R}}_A^{[n]}$ using Eq. (27) or (28). Suppose the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions fit the empirical Lorenz curve. In this case, $\hat{\mathcal{R}}_A$ can be replaced by estimation of $\mathcal{R}_{A_{SB}}$ (43), $\mathcal{R}_{A_{HS}}$ (46), $\mathcal{R}_{A_R}$ (49), or $\mathcal{R}_{A_C}$ (52), respectively.

b. Calculate $\hat{\mathcal{R}}_P^{[n]}$ using Eq. (40) or (41). Suppose the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions fit the empirical Lorenz curve. In this case, $\hat{\mathcal{R}}_P$ can be replaced by estimation of $\mathcal{R}_{P_{SB}}$ (44), $\mathcal{R}_{P_{HS}}$ (47), $\mathcal{R}_{P_R}$ (50), or $\mathcal{R}_{P_C}$ (53), respectively.

iv. From Step (4).iii:

a. Calculate the statistics $\tau_n(\hat{\mathcal{R}}_A)$ based on Eq. (5).

$$\tau_n(\hat{\mathcal{R}}_A) = (\mathcal{N} + 1)\left(\hat{\mathcal{R}}_A^{[n]} - \hat{\mathcal{R}}_A\right)^2. \tag{56}$$

b. Calculate the statistics $\tau_n(\hat{\mathcal{R}}_P)$ based on Eq. (5).

$$\tau_n(\hat{\mathcal{R}}_P) = (\mathcal{N} + 1)\left(\hat{\mathcal{R}}_P^{[n]} - \hat{\mathcal{R}}_P\right)^2. \tag{57}$$

(5) Based on Step (4), construct the p-value $\rho\left(\tau(\hat{\mathcal{R}}_A)\right)$ and $\rho\left(\tau(\hat{\mathcal{R}}_P)\right)$ based on Eq. (6).

$$\rho\left(\tau(\hat{\mathcal{R}}_A)\right) = 2\min\left(\frac{1}{\mathbb{B}}\sum_{n=1}^{\mathbb{B}} I\left(\tau_n(\hat{\mathcal{R}}_A) \le \tau(\hat{\mathcal{R}}_A)\right), \frac{1}{\mathbb{B}}\sum_{n=1}^{\mathbb{B}} I\left(\tau_n(\hat{\mathcal{R}}_A) > \tau(\hat{\mathcal{R}}_A)\right)\right). \tag{58}$$

$$\rho\left(\tau(\hat{\mathcal{R}}_P)\right) = 2\min\left(\frac{1}{\mathbb{B}}\sum_{n=1}^{\mathbb{B}} I\left(\tau_n(\hat{\mathcal{R}}_P) \le \tau(\hat{\mathcal{R}}_P)\right), \frac{1}{\mathbb{B}}\sum_{n=1}^{\mathbb{B}} I\left(\tau_n(\hat{\mathcal{R}}_P) > \tau(\hat{\mathcal{R}}_P)\right)\right). \tag{59}$$

(6) Based on Step (5):

i. The hypothesis testing decision using $\rho\left(\tau(\hat{\mathcal{R}}_A)\right)$ is:

a. If $\rho\left(\tau(\hat{\mathcal{R}}_A)\right) < \alpha$, then H_0 is rejected, and the conclusion is that the Lorenz curve is asymmetric.

b. If $\rho\left(\tau(\hat{\mathcal{R}}_A)\right) \ge \alpha$, then H_0 failed to be rejected, and the conclusion is that the Lorenz curve is symmetric.

x

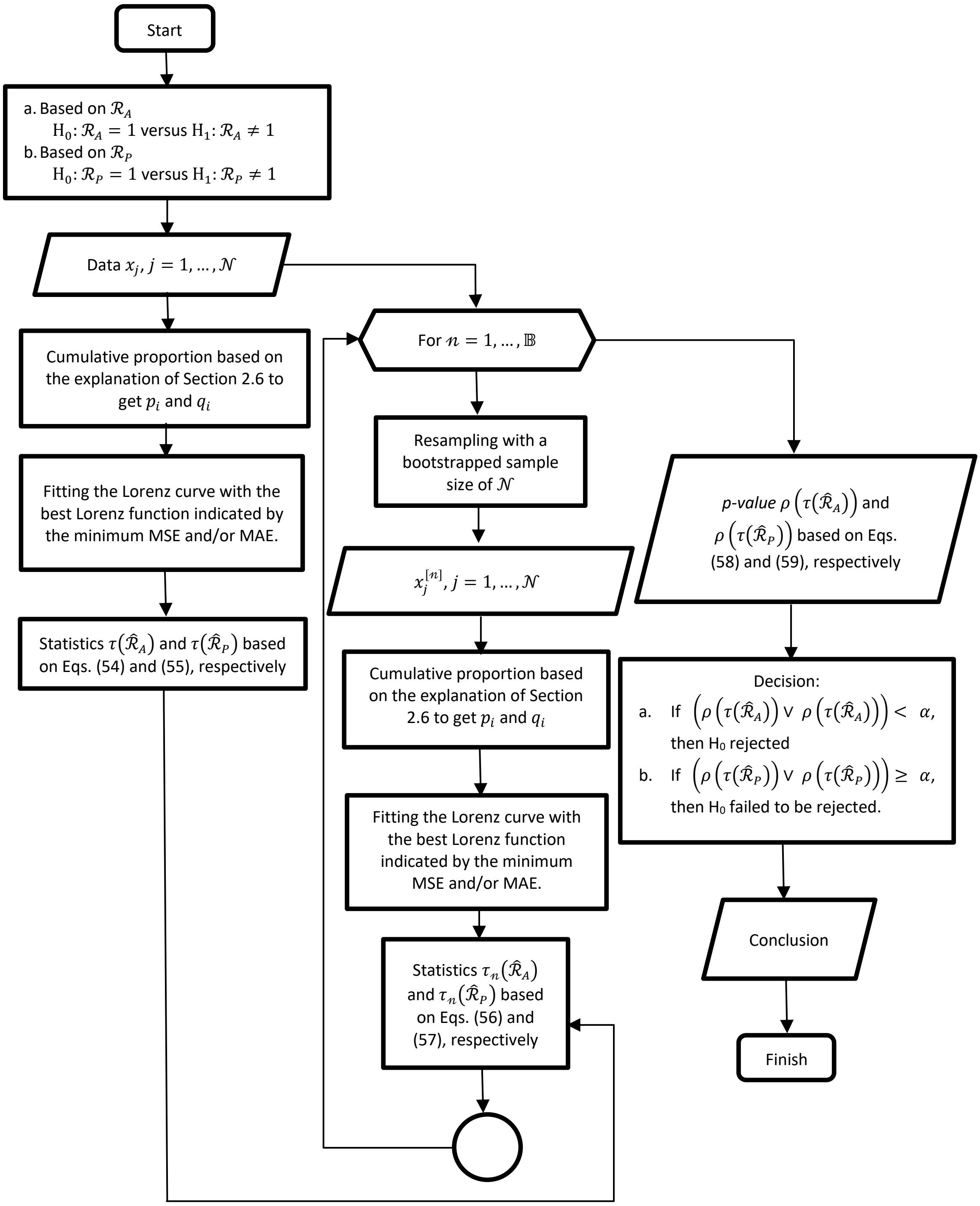


Fig. 5. Steps for hypothesis testing of Lorenz curve symmetry using nonparametric bootstrap

ii. The hypothesis testing decision using $\rho\left(\tau(\hat{\mathcal{R}}_P)\right)$ is:

a. If $\rho\left(\tau(\hat{\mathcal{R}}_P)\right) < \alpha$, then H_0 is rejected, and the conclusion is that the Lorenz curve is asymmetric.

b. If $\rho\left(\tau(\hat{\mathcal{R}}_P)\right) \geq \alpha$, then H_0 failed to be rejected, and the conclusion is that the Lorenz curve is symmetric.

Testing the Lorenz curve symmetry hypothesis using both $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ provides a consistent decision because the Monte Carlo process generates sequences $\tau_n(\hat{\mathcal{R}}_A)$ and $\tau_n(\hat{\mathcal{R}}_P)$ that have varying values so that the determination of the symmetrical or asymmetrical decision of the Lorenz curve is based on the proportion which is the p-value presented in Eqs. (58) and (59). Then, due to the nature of the Lorenz curve, which is essentially a visualization of a convex function bounded by the interval [0,1], the change between the perimeter and the area of the inequality zone is consistent and coherent. Hence, this condition also makes the decision results from testing the Lorenz curve symmetry hypothesis using either $\hat{\mathcal{R}}_A$ or $\hat{\mathcal{R}}_P$ consistent and coherent.

4.5 Bootstrap Confidence Interval of $\mathcal{R}_A$ and $\mathcal{R}_P$

To corroborate the results of the Lorenz curve symmetry test through bootstrapping, we constructed bootstrap confidence intervals of $\mathcal{R}_A$ and $\mathcal{R}_P$ by adopting the Chen-Shao algorithm [31], which was originally used to construct the highest probability density (HPD) interval from Markov chain Monte Carlo (MCMC) samples. We have argued for using the Chen-Sao algorithm to anticipate the asymmetry of the sampling distribution of statistic/estimator calculated from bootstrap samples. The following are the steps to form confidence intervals based on bootstrap samples:

(1) Determine the bootstrap samples of $\mathcal{R}_A$ and $\mathcal{R}_P$, which are $\hat{\mathcal{R}}_A^{[n]}$ and $\hat{\mathcal{R}}_P^{[n]}$, respectively, $n = 1, \dots, \mathbb{B}$.
(2) Based on the result of Step (1), sort $\hat{\mathcal{R}}_A^{[n]}$ and $\hat{\mathcal{R}}_P^{[n]}$ from the lowest to the largest value so that the set of $\mathcal{R}_A$ and $\mathcal{R}_P$ containing their respective member elements is obtained: $\hat{\mathcal{R}}_A^{(1)} \le \hat{\mathcal{R}}_A^{(2)} \le \dots \le \hat{\mathcal{R}}_A^{(\mathbb{B})}$ and $\hat{\mathcal{R}}_P^{(1)} \le \hat{\mathcal{R}}_P^{(2)} \le \dots \le \hat{\mathcal{R}}_P^{(\mathbb{B})}$.
(3) Based on the results from Step (2), calculate the $100(1-\alpha)\%$ interval for each of $\mathcal{R}_A$ and $\mathcal{R}_P$ is:
    a. $\mathfrak{R}_{\mathcal{R}_A}^k = \left(\hat{\mathcal{R}}_A^{(k)}, \hat{\mathcal{R}}_A^{(k+(1-\alpha)\mathbb{B})}\right), k = 1,2,\dots,\mathbb{B} - \left((1-\alpha)\mathbb{B}\right)$.
    b. $\mathfrak{R}_{\mathcal{R}_P}^k = \left(\hat{\mathcal{R}}_P^{(k)}, \hat{\mathcal{R}}_P^{(k+(1-\alpha)\mathbb{B})}\right), k = 1,2,\dots,\mathbb{B} - \left((1-\alpha)\mathbb{B}\right)$,
(4) $100(1-\alpha)\%$ HPD interval, only one unit is symbolized by $\mathfrak{R}^{k^*}$. It is the bootstrap confidence interval with the shortest width among all confidence intervals based on Step (3).

4.6 Empirical Study

In this section's empirical study, the Lorenz curve in three regions, namely Banten Province in Indonesia, Ghana, and South Africa, is modeled using four specifications of the Lorenz function.

4.6.1. The Case of Banten Province in Indonesia

Figure 6 illustrates the visual fitting of the Lorenz curve using the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions. The Lorenz-Rohde and Lorenz-HS functions closely align with the empirical Lorenz curve. Judging whether the Lorenz curve is symmetric or asymmetric and assessing the significance of any asymmetry in the empirical Lorenz curve presents a challenge. Consequently, determining the Lorenz function that best fits the empirical data is essential in acquiring a credible measure of asymmetry for the empirical Lorenz curve. Among the four candidate Lorenz functions (Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich), the selected Lorenz function yields the lowest MSE or MAE. This selection determines the value of the asymmetry measure corresponding to the condition of the empirical Lorenz curve.

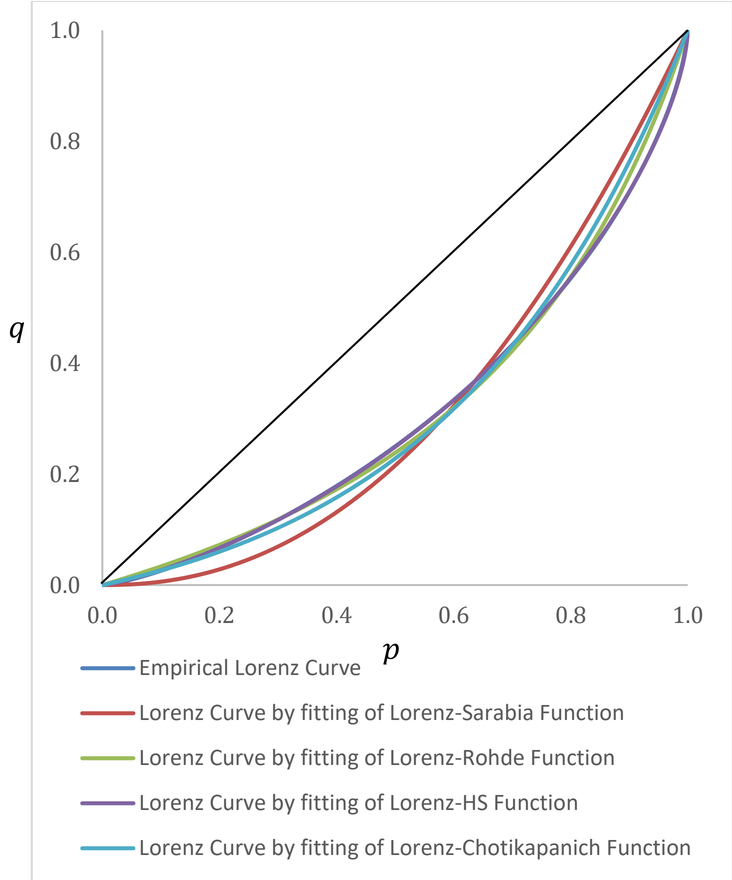


Fig. 6. Empirical Lorenz curve Fitting by Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich functions, Banten Province 2020

The findings in Table 4 demonstrate that the Lorenz-HS function exhibits the lowest MSE and MAE values compared to the Lorenz-Sarabia, Lorenz-Rohde, and Lorenz-Chotikapanich functions. This information indicates that the Lorenz-HS function is the most precise in fitting the empirical Lorenz curve in Banten Province for 2020. Consequently, the asymmetry measures $\mathcal{R}_A$ and $\mathcal{R}_P$ calculated by the Lorenz-HS function accurately reflect the true condition of the empirical Lorenz curve.

Table 4
Parameter estimation results, MSE, and MAE of Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich Functions, Banten Province in 2020

Lorenz Function	Estimator	Value estimate	MSE	MAE
Lorenz-Sarabia	$\hat{\gamma}$	2.2158	2.0105×10^{-3}	3.8762×10^{-2}
Lorenz-HS	$\hat{\sigma}_1$	0.2530	**1.0262×10^{-6}**	**7.8351×10^{-4}**
	$\hat{\sigma}_2$	0.2693		
	$\hat{\sigma}_3$	0.5984		
Lorenz-Rohde	$\hat{a}_R$	1.4553	2.3323×10^{-4}	1.1231×10^{-2}
Lorenz-Chotikapanich	$\hat{\aleph}$	2.4404	6.2743×10^{-4}	1.8799×10^{-2}

Table 5 displays the estimated values of the Kolkata index and asymmetry measures ($\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$) of the Lorenz curves, which were computed using the Lorenz-HS function. The Kolkata index is instrumental in constructing the asymmetry measures in this study. Its interval is [0.5,1], signifying that a value such as 0.6345 indicates a lower degree of income distribution inequality. Utilizing the Lorenz-HS function, we obtained estimated values of $\mathcal{R}_A$ and $\mathcal{R}_P$ as 0.9446 and 0.9974, respectively. If we consider only mathematical criteria without any probabilistic mechanism, the Lorenz curve, represented by the Lorenz-HS function, would be classified as asymmetric under Condition 2. However, the values of $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ produced by the Lorenz-HS function are proximate to 1. Consequently, it is imperative to test the hypothesis of Lorenz curve symmetry to determine whether the empirical Lorenz curve is symmetric or asymmetric

Table 5
Value estimate of asymmetry measures of Kolkata Index, $\hat{\mathcal{R}}_A$, and $\hat{\mathcal{R}}_P$ Based on the Lorenz-HS Function, Banten Province in 2020

Lorenz Function	Kolkata Index ($\widehat{p_{++}}$)	$\hat{\mathcal{R}}_A$	$\hat{\mathcal{R}}_P$
Lorenz-HS	0.6345	0.9446	0.9974

Table 6
Results of Lorenz curve symmetry test and bootstrap confidence interval of $\mathcal{R}_A$ and $\mathcal{R}_P$, Banten Province in 2020

Statistical Hypothesis	Test Statistic	p-value	Decision
H$_0$: $\mathcal{R}_A = 1$ H$_1$: $\mathcal{R}_A \neq 1$	$\tau(\hat{\mathcal{R}}_A) = 21.3566$	0.0000	H$_0$ is rejected at 5% significance level ($\alpha = 0.05$).
H$_0$: $\mathcal{R}_P = 1$ H$_1$: $\mathcal{R}_P \neq 1$	$\tau(\hat{\mathcal{R}}_P) = 0.0477$	0.0000	H$_0$ is rejected at 5% significance level.
The bootstrap confidence interval of $\mathcal{R}_A(\mathfrak{R}^{k^*}_{\mathcal{R}_A})$: $0.9297 \leq \mathcal{R}_A \leq 0.9599$			
The bootstrap confidence interval of $\mathcal{R}_P(\mathfrak{R}^{k^*}_{\mathcal{R}_P})$: $0.9966 \leq \mathcal{R}_P \leq 0.9982$			

Note: In testing the hypothesis through a nonparametric bootstrap, replication was conducted in 4,000 replications ($\mathbb{B}$=4,000).

The results of the Lorenz curve symmetry test, as presented in Table 6, corroborate the information previously discussed. As can be seen in Table 6, both p-values of $(\hat{\mathcal{R}}_A)$ and $\tau(\hat{\mathcal{R}}_P)$ are 0.0000, which is below the 5% significance level. This finding is further substantiated by the bootstrap confidence intervals of $\mathcal{R}_A$ and $\mathcal{R}_P$, which do not encompass the value 1. Both the upper and lower bounds of these intervals are less than 1. Consequently, this observation suggests the existence of an upper right bulge, indicative of the Lorenz curve's asymmetry. Furthermore, it reveals that the group with a cumulative proportion of income greater than 0.5 ($q > 0.5$) contributes more to inequality generation than the group with a cumulative proportion of income less than or equal to 0.5 (the Lorenz curve of asymmetric Condition 2).

The hypothesis testing of Lorenz curve symmetry, based on $\mathcal{R}_A$ and $\mathcal{R}_P$ in Banten Province in 2020, concluded that the most significant contribution to income inequality originated from the upper middle-income group (the group with cumulative proportion of income greater than 0.5 $q > 0.5$). Therefore, the local government can implement policies prioritizing the upper middle-income group to reduce income inequality and foster conditions for income redistribution to the lower middle-income group. These policies include:

a. Increasing taxes for individuals earning more than IDR 120 million per year. The values of $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ indicate that the group primarily contributes income inequality with a cumulative proportion of income more than 0.5 (upper middle-income class). According to Deloitte Southeast Asia [32], consumer income is classified into four income classes:
   - Upper income class (More than IDR 120 million per year)
   - Upper middle-income class (IDR 60 million - IDR 120 million per year)

- Lower middle-income class (IDR 36 million - IDR 60 million per year)
- Lower-income class (Less than IDR 36 million per year).

The tax increase for households earning more than IDR 120 million per year should correspond to the increase in income reported on the tax return. This tax increase aims to enhance the social security rate, such as Program Keluarga Harapan (the Family Hope Program).

a. Increasing social security transfers, such as the Family Hope Program and direct cash transfers, but with improved targeting accuracy of beneficiaries.
b. Implementing a policy of transparency and bureaucratic efficiency for labor-intensive investment licensing to stimulate job growth and absorb labor from the constructed industrial region.
c. Providing tax incentives for businesses or industries affected by the COVID-19 pandemic to mitigate massive layoffs.
d. Optimizing the distribution of community zakat funds collected by the National Zakat Agency (BAZNAS). Zakat collected by BAZNAS serves as an alternative financing source for social security for individuals without social security.
e. Allocating an additional social security budget in village funds. Village funds should not only focus on village infrastructure but also need an additional social security post for individuals experiencing poverty who have yet to receive any social security program.

4.6.2. The Case of Ghana

Table 7 displays the results of parameter estimation, MSE, and MAE obtained from fitting the empirical Lorenz curve of Ghana in 1998 using the Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, and Lorenz-Chotikapanich functions. Among these functions, the Lorenz-HS function exhibits the lowest MSE and MAE values compared to the other three functions (The Lorenz-Sarabia, Lorenz-Rohde, and Lorenz-Chotikapanich). This information suggests that the Lorenz-HS function is the most appropriate choice for fitting the empirical Lorenz curve of Ghana in 1998. The Lorenz-HS function effectively encapsulates the natural occurrence of an asymmetric Lorenz curve, precisely Condition 1, characterized by a lower left bulge. Figure 7 visually substantiates this finding by depicting the fitting of the four Lorenz functions to the empirical Lorenz curve of Ghana in 1998. It is evident that the Lorenz-HS function provides a superior fit to the empirical Lorenz curve of Ghana compared to the other three functions.

Table 7
Parameter estimation results of Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich Functions of empirical Lorenz curve based on total income employment per household, Ghana in 1998

Lorenz Function	Estimator	Value estimate	MSE	MAE
Lorenz-Sarabia	$\hat{\gamma}$	15.1426	1.3787×10^{-4}	5.5773×10^{-3}
Lorenz-HS	$\hat{\sigma}_1$	12.6137	$\mathbf{1.2365 \times 10^{-4}}$	$\mathbf{5.6894 \times 10^{-3}}$
	$\hat{\sigma}_2$	0.0000		
	$\hat{\sigma}_3$	0.6600		
Lorenz-Rohde	$\hat{a}_R$	1.0303	2.3389×10^{-3}	3.5982×10^{-2}
Lorenz-Chotikapanich	$\hat{\aleph}$	15.9529	1.6829×10^{-4}	6.7903×10^{-3}

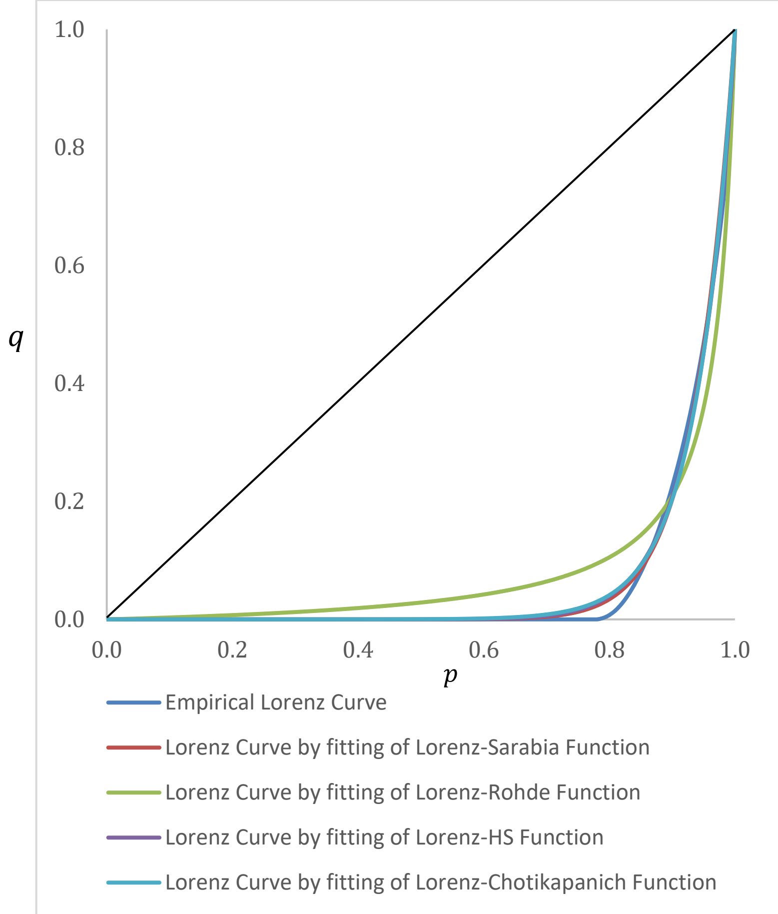


Fig. 7. Empirical Lorenz curve fitting by Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich functions based on total employment income per household, Ghana in 1998

The Lorenz-HS function yields an estimated value of 0.8707 for the Kolkata index, signifying a substantial income distribution disparity (as shown in Table 8). The inequality zone of Ghana is larger than that of the Banten Province case. Moreover, the estimated values of $\mathcal{R}_A$ and $\mathcal{R}_P$ generated by the Lorenz-HS function for Ghana are 1.1492 and 1.0184, respectively. These values suggest that the empirical Lorenz curve for Ghana adheres to an asymmetric Lorenz curve of Condition 1.

Table 8
Value estimate of asymmetry measures of Kolkata Index, $\hat{\mathcal{R}}_A$, and $\hat{\mathcal{R}}_P$ Based on the Lorenz-HS Function, Ghana in 1998

Lorenz Function	Kolkata Index ($\widehat{p_{++}}$)	$\hat{\mathcal{R}}_A$	$\hat{\mathcal{R}}_P$
Lorenz-HS	0.8707	1.1492	1.0184

This result is corroborated by the results of the hypothesis testing for Lorenz curve symmetry, as presented in Table 9. The test reveals that the p-values of $\tau(\hat{\mathcal{R}}_A)$ and $\tau(\hat{\mathcal{R}}_P)$ are 0.0000, which is below the 5% significance level. This finding is further substantiated by the bootstrap confidence intervals of $\mathcal{R}_A$ and $\mathcal{R}_P$, which do not encompass the value 1. Both the upper and lower bounds of these intervals exceed 1. Consequently, this observation confirms the asymmetry of the Lorenz curve, as indicated by the presence of a lower left bulge in the curve. Therefore, it can be concluded that the group with a cumulative proportion of income less than or equal to 0.5 contributes significantly more to income inequality among workers than the group with cumulative proportion of income greater than 0.5 (the Lorenz curve of asymmetric Condition 1).

Table 9
Results of Lorenz curve symmetry test and bootstrap confidence interval of $\mathcal{R}_A$ and $\mathcal{R}_P$, Ghana in 1998

Statistical Hypothesis	Test Statistic	p-value	Decision
H_0: $\mathcal{R}_A = 1$ H_1: $\mathcal{R}_A \neq 1$	$\tau(\hat{\mathcal{R}}_A) = 133.5311$	0.0000	H_0 is rejected at 5% significance level ($\alpha = 0.05$).
H_0: $\mathcal{R}_P = 1$ H_1: $\mathcal{R}_P \neq 1$	$\tau(\hat{\mathcal{R}}_P) = 2.0244$	0.0000	H_0 is rejected at 5% significance level.
The bootstrap confidence interval of $\mathcal{R}_A(\mathfrak{R}^{k^*}_{\hat{\mathcal{R}}_A})$: $1.1400 \leq \mathcal{R}_A \leq 1.1598$			
The bootstrap confidence interval of $\mathcal{R}_P(\mathfrak{R}^{k^*}_{\hat{\mathcal{R}}_P})$: $1.0178 \leq \mathcal{R}_P \leq 1.0193$			

Note: In testing the hypothesis through a nonparametric bootstrap, replication was conducted in 4,000 replications ($\mathbb{B}$ =4,000).

The hypothesis testing of Lorenz curve symmetry based on $\mathcal{R}_A$ and $\mathcal{R}_P$ in Ghana concludes that the most significant contribution to income inequality originates from the lower middle-income group (the group with a cumulative proportion of income less than or equal to 0.5, $q \leq 0.5$). Therefore, the local government can implement policies prioritizing this group to reduce income inequality and improve economic welfare. These policies include:

a. Increasing the minimum wage. This policy can reduce income inequality by ensuring that low-income workers earn wages sufficient for their livelihoods [33]. It can also help alleviate poverty and enhance economic mobility.
b. Expanding the scope of the Earned Income Tax Credit (EITC). Intended for low- to moderate-income workers, this policy can help reduce poverty and promote upward economic mobility [33].
c. Encouraging asset accumulation for working households. Policy measures that facilitate asset accumulation, such as savings accounts or homeownership opportunities, can play a role in reducing income inequality by encouraging economic mobility among working families [33].
d. Ensuring equal opportunities: Policies that promote equality of opportunity, including removing discriminatory regulations and practices and investing in skills development, can reduce income inequality by ensuring equal opportunities for all individuals to succeed.

4.6.3. The Case of South Africa

The Lorenz-HS function demonstrates superior performance over the Lorenz-Sarabia, Lorenz-Rohde, and Lorenz-Chotikapanich functions in generating the lowest MSE and MAE values, as evidenced in Table 10. This finding underscores the preeminence of the Lorenz-HS function in modeling the empirical Lorenz curve for South Africa. Figure 8 provides a visual representation of the congruence between the empirical Lorenz curve and the fits of the four Lorenz functions. However, ascertaining whether the Lorenz curve is symmetric or asymmetric based solely on visual inspection remains a task. Consequently, the asymmetry measure developed in this study emerges as an alternative indicator derived from the Lorenz curve, demonstrating effective applicability to actual data.

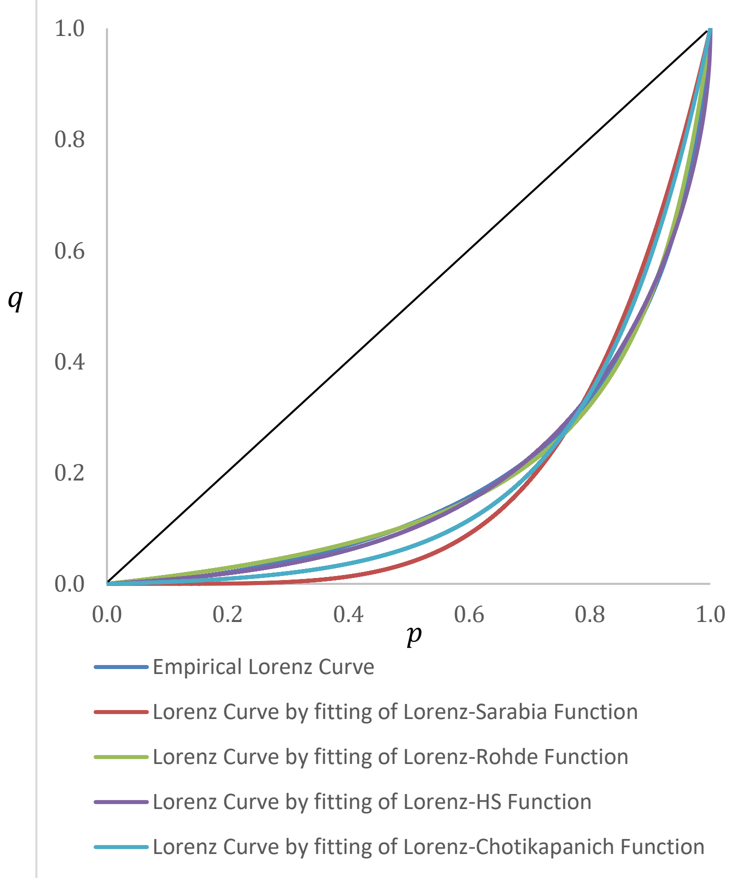


Fig. 8. Empirical Lorenz curve fitting by Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich functions based on household income, South Africa in 2015

Table 10
Parameter estimation results of Lorenz-Sarabia, Lorenz-HS, Lorenz-Rohde, Lorenz-Chotikapanich Functions of empirical Lorenz curve based on household income, South Africa in 2015

Lorenz Function	Estimator	Value estimate	MSE	MAE
Lorenz-Sarabia	$\hat{\gamma}$	4.7203	3.0075×10^{-3}	4.5448×10^{-2}
Lorenz-HS	$\hat{\sigma}_1$	0.0000	**6.6606×10^{-5}**	**6.2529×10^{-3}**
	$\hat{\sigma}_2$	1.9891		
	$\hat{\sigma}_3$	0.4403		
Lorenz-Rohde	$\hat{a}_R$	1.1343	1.0512×10^{-4}	7.5726×10^{-3}
Lorenz-Chotikapanich	$\hat{\aleph}$	5.3218	1.5100×10^{-3}	3.0411×10^{-2}

As indicated in Table 11, the Lorenz-HS function yields an estimated Kolkata index value of 0.7367 for South Africa. This value is lower than the Kolkata index observed in the case of Ghana but higher than the Kolkata index of Banten Province, suggesting that income distribution inequality in South Africa is not as severe as in Ghana but is still more pronounced than in Banten Province. The Lorenz-HS function produces values of $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ as 1.0236 and 1.0031, respectively, for South Africa. Given that the values of $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ for the empirical Lorenz curve are greater than 1 (Asymmetry Condition 1), this indicates an asymmetric Lorenz curve characterized by a lower left bulge. However, it would be premature to conclude that the curve is asymmetric based solely on these values since $\hat{\mathcal{R}}_A$ and $\hat{\mathcal{R}}_P$ are proximate to 1. Consequently, it is imperative to conduct a test for the asymmetry of the Lorenz curve to draw definitive conclusions.

Table 11
Value estimate of asymmetry measures of Kolkata Index, $\hat{\mathcal{R}}_A$, and $\hat{\mathcal{R}}_P$ Based on the Lorenz-HS Function, South Africa in 2015

Lorenz Function	Kolkata Index ($\widehat{p_{++}}$)	$\hat{\mathcal{R}}_A$	$\hat{\mathcal{R}}_P$
Lorenz-HS	0.7367	1.0236	1.0031

Therefore, testing the Lorenz curve symmetry hypothesis in South Africa is paramount. As evidenced in Table 12, the test result unequivocally indicates that the empirical Lorenz curve in South Africa for 2015 exhibits asymmetry. This finding is substantiated by the bootstrap confidence intervals of $\mathcal{R}_A$ and $\mathcal{R}_P$, which do not encompass the value 1. Furthermore, both the upper and lower bounds of these intervals exceed 1. Consequently, this compelling evidence strongly suggests that the group with a cumulative proportion of income less than or equal to 0.5 ($q \leq 0.5$) contributes significantly more to income inequality among households than the group with a cumulative proportion of income greater than 0.5 (resulting in Lorenz curve of asymmetry Condition 1).

Table 12
Results of Lorenz curve symmetry test and bootstrap confidence interval of $\mathcal{R}_A$ and $\mathcal{R}_P$, South Africa in 2015

Statistical Hypothesis	Test Statistic	p-value	Decision
H_0: $\mathcal{R}_A = 1$ H_1: $\mathcal{R}_A \neq 1$	$\tau(\hat{\mathcal{R}}_A) = 13.0415$	0.0000	H_0 failed to be rejected at 5% significance level ($\alpha = 0.05$)
H_0: $\mathcal{R}_P = 1$ H_1: $\mathcal{R}_P \neq 1$	$\tau(\hat{\mathcal{R}}_P) = 0.2219$	0.0000	H_0 failed to be rejected at 5% significance level
The bootstrap confidence interval of $\mathcal{R}_A(\mathfrak{R}^{k^*}_{\hat{\mathcal{R}}_A})$: $1.0158 \leq \mathcal{R}_A \leq 1.0305$			
The bootstrap confidence interval of $\mathcal{R}_P(\mathfrak{R}^{k^*}_{\hat{\mathcal{R}}_P})$: $1.0026 \leq \mathcal{R}_P \leq 1.0036$			

Note: In testing the hypothesis through a nonparametric bootstrap, replication was conducted in 4,000 replications ($\mathbb{B}$ =4,000).

The hypothesis testing of Lorenz curve symmetry based on $\mathcal{R}_A$ and $\mathcal{R}_P$ for South Africa concludes that the most significant contribution to income inequality originates from the lower-middle income group (the group with a cumulative proportion of income less than or equal to 0.5 ($q \leq 0.5$)). Therefore, local authorities can implement several policies to reduce income inequality, including:

a. Progressive Fiscal Redistribution: The Government of South Africa has instituted progressive fiscal redistribution as a fundamental policy approach for mitigating income inequality [34,35]. This multifaceted strategy encompasses the amplification of social spending, the targeted disbursement of government transfers, and the implementation of affirmative action initiatives aimed at diversifying wealth ownership and fostering entrepreneurial endeavors among previously marginalized demographic segments.
b. Expansion of Social Grants: The expansion of social grants has a significant equalizing effect in South Africa [36,37]. Social grants are crucial in helping people out of poverty and effectively reducing income inequality.
c. Creation of Skill-Oriented Employment Opportunities for lower-income class: The Government of South Africa duly recognizes the imperative of establishing skill-oriented employment prospects for low-income people as a strategic measure to reduce inequality [38]. Drawing upon the counsel of the World Bank, a suite of policy interventions, including the creation of skilled employment opportunities for those experiencing poverty and the facilitation of economic growth through heightened competition, policy stability, and the promotion of skilled migration,

are considered instrumental in the potential reduction of South Africa's poverty rate by half by the year 2030.

d. Enhancement of the Efficiency and Effectiveness of Social Spending: In concordance with recommendations from the World Bank, a policy avenue of paramount significance involves augmenting the efficiency and effectiveness of social spending initiatives among member countries of the Southern African Customs Union (SACU) [39].This policy framework entails refining quality, targeting precision, and overall efficiency of social spending programs, amplifying their contributory role in diminishing income inequality.

4.7.4. Discussion

The asymmetry measures $\mathcal{R}_A$ and $\mathcal{R}_P$ in this study still hinge on selecting Lorenz functions that align with the data conditions. The criteria for selecting the Lorenz function are based on the minimum MSE or MAE generated from several candidate functions. However, these criteria's procedure is not yet under a probabilistic framework, such as the Kolmogorov-Smirnov test and its development [40], the Anderson-Darling test and its development [41,42], the Chi-square test and its extensions [43], and Bayesian model selection using the Bayes factor [44]. The procedure for selecting the Lorenz function could be constructed by adopting one of the ideas from the aforementioned studies.

In assessing the asymmetry of the empirical Lorenz curve, $\mathcal{R}_A$ and $\mathcal{R}_P$ serve as reliable indicators of asymmetrical measure. However, to ensure the accuracy and reliability of these measures, it is crucial to supplement them with additional supporting information. This may include visual representations of the empirical Lorenz curve and other inequality measures such as the Kolkata, Gini, Amato, and Pietra indices. The accuracy of the Lorenz function used to model the empirical Lorenz curve is pivotal in ensuring the credibility of $\mathcal{R}_A$ and $\mathcal{R}_P$.

Supporting information is crucial because $\mathcal{R}_A$ and $\mathcal{R}_P$ are ratio measures that do not consider the size of the inequality zone. For instance, if two subzones of the empirical Lorenz curve generate identical $\mathcal{R}_A$ values, it is impossible for an analyst to determine whether the $\mathcal{R}_A$ value corresponds to an inequality zone with a large or small size. It is important to note that a larger inequality zone indicates a more unequal income distribution. Therefore, the Kolkata index, which is based on Eq. (24), can be used as an additional indicator for the asymmetry analysis of the Lorenz curve, serving as a proxy for the size of the Lorenz curve. For example, the value of the Kolkata index increases as the size of the inequality zone expands.

The hypothesis testing for Lorenz curve symmetry in this study employs a nonparametric bootstrap method, which requires the transformation of cumulative proportions to calculate $\mathcal{R}_A$ and $\mathcal{R}_P$ measures, utilizing p and q from the original data (x). Nevertheless, the omission of the cumulative proportion process from hypothesis testing might be feasible if it is possible to represent $\mathcal{R}_A$ and $\mathcal{R}_P$ formulations within the context of the original data (x).

Regarding policies that can be implemented based on the results of income inequality analysis through Lorenz curve visualization and asymmetry measures ($\mathcal{R}_A$ and $\mathcal{R}_P$), priority should be given to policies that target the reduction of income inequality caused by the main contributing groups before addressing other groups. Essentially, policies intended to reduce income inequality should be based on a commitment to improving welfare while sustaining long-term economic growth by considering the Government Budget [45–47].

5. Conclusions

Based on the empirical results and discussions, it can be concluded that $\mathcal{R}_A$ and $\mathcal{R}_P$ effectively encapsulate the inherent asymmetric phenomenon of the empirical Lorenz curve. However, the credibility of both measures is contingent upon the utilization of a Lorenz function that accurately fits the empirical Lorenz curve. The inclusion of supporting indicators, such as the Kolkata index, is pivotal for a comprehensive analysis of the symmetry of the Lorenz curve and for obtaining more precise information. The hypothesis testing for Lorenz curve symmetry, developed in this study and employing a nonparametric bootstrap, provides a probabilistic framework for assessing the symmetry or asymmetry of a Lorenz curve rather than merely comparing a single value to a specific threshold. Based on the three cases examined in this study, this hypothesis test for the symmetry of the Lorenz curve proves effective in drawing definitive conclusions about the asymmetric condition of the empirical Lorenz curve. Consequently, the government can implement policies to reduce income inequality while prioritizing the target beneficiaries of these executed policies.

Author Contributions

Conceptualization, M.F., S. and N.I.; methodology, M.F.; software, M.F.; validation, M.F., S. and N.I.; formal analysis, M.F., S. and N.I.; investigation, M.F., S. and N.I.; resources, M.F.; data curation, M.F.; writing—original draft preparation, M.F., S. and N.I.; writing—review and editing, M.F., S. and N.I.; visualization, M.F.; supervision, S. and N.I.; project administration, M.F., S. and N.I.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding

This research was financially supported by Badan Pusat Statistik-Statistics Indonesia.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

The authors are grateful for the support of Badan Pusat Statistik-Statistics Indonesia and Institut Teknologi Sepuluh Nopember.

References

[1] Lorenz, M. O. (1905). Methods of measuring of wealth. *American Statistical Association*, *9*(70), 209–219. https://doi.org/10.2307/2276207
[2] Rasche, R. H., Gaffney, J., Koo, A. Y. C., & Obst, N. (1980). Functional Forms for Estimating the Lorenz Curve. *Econometrica*, *48*(4), 1061–1062. https://doi.org/10.2307/1912948
[3] Gupta, M. R. (1984). Functional Form for estimating the Lorenz curve. *Econometrica*, *52*(5), 1313–1314. https://doi.org/10.2307/1911001
[4] Hossain, S., & Saeki, C. (2003). A New functional form for estimating Lorenz curves. *Journal of Business and Economics Research*, *1*(5), 81–94. https://doi.org/https://doi.org/10.19030/jber.v1i5.3014
[5] Ortega, P., Martin, G., Fernandez, A., Ladoux, M., & Garcia, A. (1991). A new functional form for estimating Lorenz curves. *Review of Income and Wealth Series*, *37*(4), 447–452. https://doi.org/10.1111/j.1475-4991.1991.tb00383.x
[6] Basmann, R. L., Hayes, K. J., Slottje, D. J., & Johnson, J. D. (1990). A general functional form for approximating the Lorenz curve. *Journal of Econometrics*, *43*(1–2), 77–90. https://doi.org/10.1016/0304-4076(90)90108-6
[7] Rohde, N. (2009). An alternative functional form for estimating the Lorenz curve. *Economics Letters*, *105*(1), 61–63. https://doi.org/10.1016/j.econlet.2009.05.015
[8] Paul, S., & Shankar, S. (2020). An alternative single parameter functional form for Lorenz curve. *Empirical Economics*, *59*(3), 1393–1402. https://doi.org/10.1007/s00181-019-01715-3
[9] Chotikapanich, D. (1993). A comparison of alternative functional forms for the Lorenz curve. *Economics Letters*, *41*, 129–138. https://doi.org/10.1016/0165-1765(93)90186-g
[10] Schader, M., & Schmid, F. (1994). Fitting Parametric Lorenz Curves to Grouped Income Distributions - A Critical Note. *Empirical Economics*, *19*(3), 361–370. https://doi.org/10.1007/bf01205943

[11] Sarabia, J. M. (1997). A hierarchy of Lorenz curves based on the Generalized Tukey's Lambda distribution. *Econometric Reviews*, *16*(3), 305–320. https://doi.org/10.1080/07474939708800389
[12] Aggarwal, V. (1984). On optimum aggregation of income distribution data. *Sankhyā: The Indian Journal of Statistics, Series B*, *46*(3), 343–355. https://doi.org/10.2307/25052360
[13] Qingtao, Y., Tong, Y., Wenchao, W., & Shenghua, F. (2014). A new Lorenz curve model based on income distribution research. *Proceedings - 2014 International Conference on Management of e-Commerce and e-Government, ICMeCG 2014*, 295–300. https://doi.org/10.1109/ICMeCG.2014.67
[14] Gallegati, M., Landini, S., & Stiglitz, J. E. (2016). The Inequality Multiplier. *Columbia Business School Research Paper*, *16–29*. http://ssrn.com/abstract=2766301Electroniccopyavailableat:http://ssrn.com/abstract=2766301
[15] Clementi, F., Gallegati, M., Gianmoena, L., Landini, S., & Stiglitz, J. E. (2019). Mis-measurement of inequality: a critical reflection and new insights. *Journal of Economic Interaction and Coordination*, *14*(4), 891–921. https://doi.org/10.1007/s11403-019-00257-2
[16] Mendelson, E. (2022). *Schaum's outlines Calculus: 7th Edition* (7th ed.). McGraw-Hill.
[17] Givens, G. H., & Hoeting, J. A. (2013). *Computational Statistics, 2nd Edition*. John Wiley & Sons, Inc.
[18] Mishra, D. R., Dumka, P., Mishra, P., & Tewari, P. (2023). Integration based on Monte Carlo Simulation. *I. J. Mathematical Sciences and Computing*, *3*, 58–65. https://doi.org/10.5815/ijmsc.2023.03.05
[19] Smaga, Ł. (2017). Bootstrap methods for multivariate hypothesis testing. *Communications in Statistics - Simulation and Computation*, *46*(10), 7654–7667. https://www.tandfonline.com/doi/abs/10.1080/03610918.2016.1248573?journalCode=lssp20
[20] MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record*, *82*(SPEC. ISS. 1), S2–S18. https://doi.org/10.1111/j.1475-4932.2006.00328.x
[21] MacKinnon, J. G. (2007). Bootstrap Hypothesis Testing. In D. A. Belsley & E. J. Kontoghiorghes (Eds.), *Handbook of Computational Econometrics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470748916.ch6
[22] Duchesne, P., & Francq, C. (2015). Multivariate hypothesis testing using generalized and {2}-inverses – with applications. *Statistics*, *49*(3), 475–496. https://doi.org/10.1080/02331888.2014.896917
[23] Duchesne, P., & Lafaye De Micheaux, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis*, *54*(4), 858–862. https://doi.org/10.1016/j.csda.2009.11.025
[24] Marumo, N., Okuno, T., & Takeda, A. (2023). Majorization-minimization-based Levenberg–Marquardt method for constrained nonlinear least squares. *Computational Optimization and Applications*, *84*(3), 833–874. https://doi.org/10.1007/s10589-022-00447-y
[25] Umar, A. O., Sulaiman, I. M., Mamat, M., Waziri, M. Y., & Zamri, N. (2021). On damping parameters of Levenberg-Marquardt algorithm for nonlinear least square problems. *Journal of Physics: Conference Series*, *1734*(1). https://doi.org/10.1088/1742-6596/1734/1/012018
[26] Madsen, K., Nielsen, H. B., & Tingleff, O. (2004). *Methods for Non-linear Least Squares Problems* (2nd ed.). Informatics and Mathematical Modelling, Technical University of Denmark. http://www2.imm.dtu.dk/pubdb/edoc/imm3215.pdf
[27] Sitthiyot, T., & Holasut, K. (2020). A simple method for measuring inequality. *Palgrave Communications*, *6*(1). https://doi.org/10.1057/s41599-020-0484-6
[28] Banerjee, S., Chakrabarti, B. K., Mitra, M., & Mutuswami, S. (2019). On the Kolkata index as a measure of income inequality. *Physica A: Statistical Mechanics and Its Applications*, *545*. https://doi.org/10.1016/j.physa.2019.123178
[29] Banerjee, S., Chakrabarti, B. K., Mitra, M., & Mutuswami, S. (2020). Inequality Measures: The Kolkata Index in Comparison With Other Measures. *Frontier in Physics*, *8*(562182). https://doi.org/10.3389/fphy.2020.562182
[30] Chatterjee, A., Ghosh, A., & Chakrabarti, B. K. (2017). Socio-economic inequality: Relationship between Gini and Kolkata indices. *Physica A: Statistical Mechanics and Its Applications*, *466*, 583–595. https://doi.org/10.1016/j.physa.2016.09.027
[31] Le, H., Pham, U., Nguyen, P., & Pham, T. B. (2020). Improvement on Monte Carlo estimation of HPD intervals. *Communications in Statistics: Simulation and Computation*, *49*(8), 2164–2180. https://doi.org/10.1080/03610918.2018.1513141
[32] Deloitte Southeast Asia. (2015). *Deloitte Consumer Insights Capturing Indonesia's latent markets*.
[33] Powell, J. A. (2014, October). *Six policies to reduce economic inequality*. https://belonging.berkeley.edu/six-policies-reduce-economic-inequality
[34] IMF. (2020, January). *Six Charts Explain South Africa's Inequality* . https://www.imf.org/en/News/Articles/2020/01/29/na012820six-charts-on-south-africas-persistent-and-multi-faceted-inequality
[35] IMF. (2020, February). *South Africa: Bridging the Income Divide*. https://www.imf.org/en/Blogs/Articles/2020/02/07/south-africa-bridging-the-income-divide

[36] Leibbrandt, M. V., Wegner, Eva., & Finn, Arden. (2011). *The policies for reducing income inequality and poverty in South Africa* (64). SALDRU. http://hdl.handle.net/11090/79

[37] UNU-WIDER. (2019, February). *Reducing inequalities in South Africa*. Policy Brief. https://www.wider.unu.edu/publication/reducing-inequalities-south-africa

[38] World Bank. (2018, April). *Policy Interventions, Skilled Jobs Can Reduce Inequality in South Africa*. https://www.worldbank.org/en/country/southafrica/publication/south-africa-economic-update-policy-interventions-skilled-jobs-can-reduce-inequality-in-south-africa

[39] World Bank. (2022, March 14). *In Southern Africa, Leveling the Playing Field at Birth Critical to Reducing Inequality, Intergenerational Poverty*. https://www.worldbank.org/en/region/afr/publication/in-southern-africa-leveling-the-playing-field-at-birth-critical-to-reducing-inequality-intergenerational-poverty

[40] Finner, H., & Gontscharuk, V. (2018). Two-sample kolmogorov-smirnov-type tests revisited: Old and new tests in terms of local levels. *Annals of Statistics*, *46*(6A), 3014–3037. https://doi.org/10.1214/17-AOS1647

[41] Coronel-Brizio, H. F., & Hernández-Montoya, A. R. (2010). The Anderson-Darling test of fit for the power-law distribution from left-censored samples. *Physica A: Statistical Mechanics and Its Applications*, *389*(17), 3508–3515. https://doi.org/10.1016/j.physa.2010.03.041

[42] Li, Y., Wei, Y., Li, B., & Alterovitz, G. (2014). Modified anderson-darling test-based target detector in non-homogenous environments. *Sensors (Switzerland)*, *14*(9), 16046–16061. https://doi.org/10.3390/s140916046

[43] Rolke, W., & Gongora, C. G. (2021). A chi-square goodness-of-fit test for continuous distributions against a known alternative. *Computational Statistics*, *36*(3), 1885–1900. https://doi.org/10.1007/s00180-020-00997-x

[44] Van Ravenzwaaij, D., Monden, R., Tendeiro, J. N., & Ioannidis, J. P. A. (2019). Bayes factors for superiority, non-inferiority, and equivalence designs. *BMC Medical Research Methodology*, *19*(1). https://doi.org/10.1186/s12874-019-0699-7

[45] Cingano, F. (2014). Trends in Income Inequality and its Impact on Economic Growth. *OECD Social, Employment and Migration Working Papers*, *163*. https://doi.org/10.1787/5jxrjncwxv6j-en

[46] Business and Economic Research Ltd. (2020, December 15). *Inequality and New Zealand*. https://berl.co.nz/our-pro-bono/inequality-and-new-zealand

[47] Green, D. A., Riddell, W. C., & St-Hilaire, F. (2017, February 23). *Income Inequality in Canada: Driving Forces, Outcomes and Policy*. https://irpp.org/research-studies/income-inequality-in-canada/