AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 6(4) (2025) 327-340
https://doi.org/10.22060/AJMC.2024.22945.1205

Original Article

The chi-square statistic as an income inequality index

Shahryar Mirzaei*[a], Sayed Mahdi Amir Jahanshahi[b]

[a]*Department of Statistics, Payame Noor University, Tehran, Iran*
[b]*Department of Statistics, University of Sistan and Baluchestan, Zahedan, Iran*

ABSTRACT: This article presents a novel concept known as the chi-square inequality index, developed through the utilization of the chi-square distance function. The study delves into the essential characteristics necessary for an effective inequality index. Additionally, a detailed formulation for the chi-square inequality curve is provided within key inequality models. A comparative analysis between the chi-square curve and the conventional Lorenz curve is conducted. Furthermore, a stochastic order based on the chi-square inequality curve is introduced. The research includes a simulation analysis to explore the statistical properties of the proposed sampling estimator. To conclude, the article highlights the effectiveness of this index through an application to real-world data.

Review History:

Received:27 January 2024
Revised:08 May 2024
Accepted:20 May 2024
Available Online:01 October 2025

Keywords:

Chi-square index
Income inequalty index
Lorenz curve

MSC (2020):

62P05; 91B82

1. Introduction

In mathematics, the concept of distance is considered essential and foundational. It is defined as the measurement of the separation between two objects. Today, distance is a key element in various fields such as probability theory and applied statistics within science and technology. There are a substantial number of distance measures in many different fields of sciences. Cha (2007) [4] collected a set of these metrics in an article.

Distance measures are essential to solving many pattern recognition problems such as income inequality problems. The topic of inequality indicators is naturally used to measure inequality in the population. One of the important issues within the talk of income inequality is finding a suitable numerical metric to measure disparity in population. In the discussion of income inequality, the level of inequality is measured by the distance between the Lorenz curve and the line of perfect equality. In fact, the Lorenz curve measures how income is distributed among individuals in society. This curve shows the cumulative share of income from different sections of the population. The line of perfect equality, on the other hand, is also a Lorenz curve. In fact, the equality line assumes that all incomes are equal. Therefore, finding the appropriate distance between the Lorenz curve and the perfect equality line to explain inequality in society has always been of interest to researchers. Pietra index (see Sarabia and jordia

**Corresponding author.*
E-mail addresses: sh_mirzaee@pnu.ac.ir, mjahan@math.usb.ac.ir

(2014) [13]) for measure inequality was defined based on Chebyshev distance. In mathematics, Chebyshev distance or maximum metric is a metric defined on a vector space, where the distance between two vectors is the greatest of their differences along any coordinate dimension. Gini (1912) considered the Manhattan distance or absolute value distance to introduce a new index (see Giorgi and Gigliarano (2017) [6]). Gini index examines the absolute differences between coordinate of a pair of the Lorenz curve and the perfect line of equality. Also Mirzaei et al. (2017) [9] introduced the Canberra inequality index based on the Canberra distance function.

There are metrics that, although they have the usual properties for the income inequality index, have not received much attention. Such a metric is the chi-square measure. The chi-square measure is often used to compare two observed distributions to determine either the independence between two criteria or homogeneity in a contingency table. Also, The chi-square test is referred to as the goodness-of-fit criterion (see Andersen (1973) [1]).

The use of the chi-square metric in econometrics has been new and can be increased in its applications widely. Although this distance measure has not been introduced for exactly the same purpose, it has some features that required to an inequality index. In this document, while introducing the chi-square index, we study some properties of this criterion and its corresponding curve.

The structure of this article is as follows. Section 2 includes some basic concepts in the topic of income inequality that are useful for the next sections. In Section 3, we present a new definition of the chi-square index. Some properties of an optimal inequality measure are investigated. One of the most useful uses of the chi-squared curve is to define the ordering of the curve. This is dealt with in Section 5. In Section 6, we provide explicit expressions for chi-squared index and the corresponding curve with graph representation in some significant inequality distributions. In section 7, we suggest the chi-squared sampling estimator. In order to investigate some statistical properties of the discussed estimator, a simulation study has been conducted in Section 8. Next, using a real data set, the superiority and sensitivity of the new inequality curve compared to the Lorenz curve has been shown in Section 9. At the end, results are plased in the last section.

2. Preliminaries

The Lorenz curve is like a graph that shows how wealth or income is spread out among people in a group. It was created by a guy named Max O. Lorenz to help us see how unequal things are. The Lorenz curve is mainly represented by a function $L(p)$, where p, the cumulative portion of the population, is depicted on the horizontal axis, and $L(p)$, the cumulative portion of the total wealth or income, is depicted on the vertical axis.

Let $\mathcal{L}$ denote the set of continuous random variables that are non-negative and have a positive finite expectation ($\mu = E(X) > 0$). For the random variable X in $\mathcal{L}$ with cumulative distribution function (cdf) F, the quantile function is defined by $F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad u \in [0, 1]$.
Pasquazzi and Zenga (2018) [10] believe, the Lorenz function of X is denoted by

$$L(p) = \frac{1}{\mu}\int_0^p F^{-1}(u)du, \qquad p \in [0, 1]. \tag{1}$$

The Lorenz function depicts the cumulative percentage of the total and final income of a cumulative proportion p of the population. To visualize ratios (1), like Figure 1, points are shown $(p, L(p))$.

As a result, we have the L curve. We can call this curve the Lorenz curve. The L curve is defined on the whole interval $(0, 1)$. The values of $L(0) = 0$ and $L(1) = 1$ are defined. Notice that the L is always under the diagonal

$$I(p) = p, \qquad p \in [0, 1].$$

The straight line I, is also a Lorenz curve. In fact, supposing that all the incomes are equal. So the interpretation of I (the straight line) represents perfect equality and any deviation from this bisector line indicates inequality.

According to the literature on Lorenz curve and equality line, it is natural to measure the economic inequality by applying some distance between the egalitarian line and the Lorenz curve, then this distance can be considered as a measure for economic inequality in the population.

Lorenz (1905) used a curve to show the inequality of income distribution and evaluate it. This curve showed a clear interpretation of inequality. Since then, many indices have been presented based on the Lorenz function. Pietra inequality index was defined based on the greatest distance between the Lorenz curve and the perfect equality line. Also, the Gini coefficient was defined based on the area between the Lorenz curve and the perfect equality line (see Sarabia (2008) [12]).

The most important discussion in the interpretation of inequality indices is to determine the characteristics that each inequality index should have so that it can be used to measure income inequality. These special characteristics are called the principles governing income inequality indicators. Comparatively, the more characteristics the index has, the more suitable the index is. Some of these features are listed in the table below.

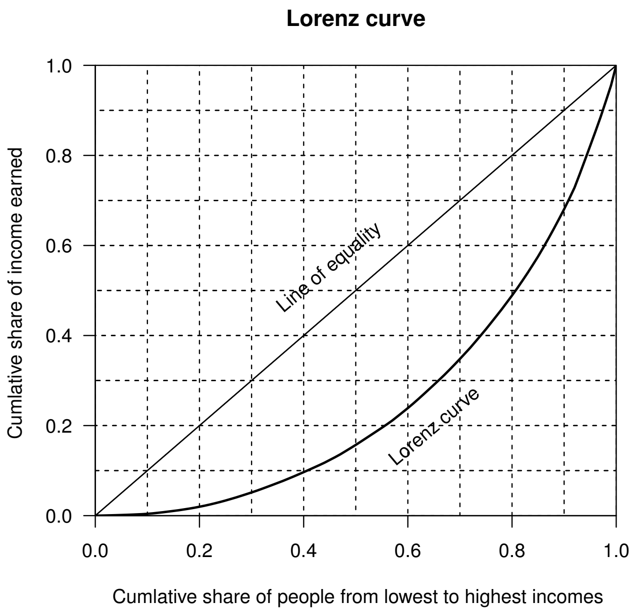

Figure 1: An example of Lorenz curve

Table 1: The principles governing income inequality indices

Normalization	The range of the inequality index must be between zero and one.
Replication reliability	If incomes are repeated equally, income inequality will not change.
Symmetry	By changing the income status of people with different characteristics, there should be no change in the size of the income inequality index.
Sensitivity to location reliability	If a certain amount is added or subtracted from the income of all members of the society, the index of income inequality should decrease or increase respectively.
Insensitivity to scale reliability	If the incomes of people in the society change in the same proportion, the inequality index does not change.
Decomposability	If the society is divided into subgroups of people, then the inequality index of the society is equal to the sum of the inequality indices of the subgroups.
Sensitivity to income transfer	If the income of a rich person is transferred to a poor person, provided that the amount of transfer is such that the relationship between the rich and the poor person is not reversed, then the inequality index will decrease.

One of the other important features of the income inequality index is the ease of its calculation and estimation, as well as the intuitive interpretation of this index and its corresponding curve. Researchers are always looking for indicators to measure income inequality that follow the principles governing the income inequality index.

3. The main chi-square index

In this segment, we present the chi-square measure and its associated curve derived from the chi-square distance function. Furthermore, we explore several essential properties of this index in detail.

Definition 3.1. *Let X be a random variable belonging to $\mathcal{L}$ class. The chi-square index (Ψ) based on chi-square distance is defined as*

$$\Psi = d(p, L(p)) = \int_0^1 \Psi(p)dp, \tag{2}$$

where the chi-square curve is defined as:

$$\Psi(p) = \frac{[p - L(p)]^2}{p}, \qquad p \in (0, 1]. \tag{3}$$

When all individuals in the population possess the same value, the chi-square curve aligns with the perfect equality line connecting the points $(0, 1)$, $(0, 0)$, $(1, 0)$. Conversely, the line representing maximum inequality connects the points $(0, 1)$, $(1, 1)$, $(1, 0)$. As the random variable X approaches minimum inequality scenario, the curve $\Psi(p)$ converges to 0 for all $p \in (0, 1)$. On the other hand, as X tends towards maximum inequality, the $\Psi(p)$ curve approaches 1 for all $p \in (0, 1)$. Similar to the Lorenz curve, the chi-square curve also exhibits fixed behavior at the start and end of the interval. So we always have:

$$\lim_{p \to 0^+} \Psi(p) = 0, \quad \lim_{p \to 1^-} \Psi(p) = 0.$$

It is worth mentioning that this curve is always concave. This can be expressed in the following lemma.

Lemma 3.2. *Suppose $\Psi(p)$ is a chi-square curve defined and continuous on (0,1] with second derivative $\Psi^{''}(p)$. Then $\Psi(p)$ is concave curve.*

Proof. Let X be a non-negative random variable of income with the Lorenz function $(L(p))$. Since the first and second derivatives of the Lorenz function are positive ($L^{'}(p) > 0$ and $L^{''}(p) > 0$, see Kleiber and Kotz (2003) [7]) and also $L(p) < p$, taking the second derivative of the chi-square function gives the desired result, i.e. the convexity of the chi-square function. □

Like other measures of income inequality , the chi-square index is calculated based on the mean value of the $\Psi(p)$ curve, but it specifically represents the area under the curve. In Figure 2 the chi-square index can be visualized as the area between the chi-square curve and the line of perfect equality. A value of 0 indicates perfect equality, while a value of 1 indicates complete inequality.

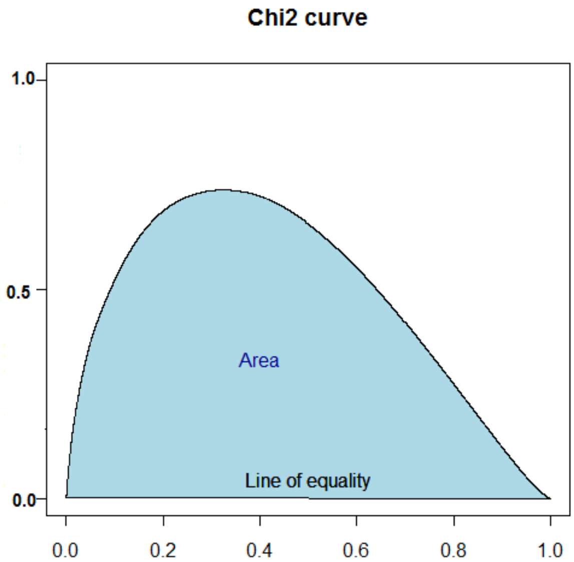

Figure 2: Example of the chi-square curve

The new inequality index fulfills all the main features of a composite inequality measure. It is symmetrical in the sense that its invariance lies in the permutation of incomes. The symmetry property follows from the fact that we defined Ψ directly in the ordered distribution. This fulfills the Pigue-Dalton condition, the postulate that, other things being equal, including their relative position in the distribution, income transfers from the rich to the poor reduce the extent of income inequality. Here, some main properties of chi-square index have been investigated.

4. Some main properties of the new index

Here we examine some important principles on income inequality indicators. Among these features are the property of normalization, scale reliability and sensitivity to location reliability or sensitivity to income transfer, which we refer to in the following propositions.

Proposition 4.1. *The chi-square inequality index has the property of normalization.*

Proof. According to the structure of the definition of chi-square index,

$$(p - L(p))^2 \le p, \qquad \forall p \in (0, 1]. \tag{4}$$

It can be seen that this criterion is always between zero and one. This indicates that the new index has the important property of the normalization principle. □

Proposition 4.2. *The chi-square curve and corresponding index are scale invariant.*

Proof. Let $Y = aX$ for any $a > 0$, then

$$F_Y^{-1}(p) = aF_X^{-1}(p), \qquad p \in (0, 1), \tag{5}$$

where F_X^{-1} and F_Y^{-1} are the inverse function of X, Y respectively. Also, we have

$$E(Y) = aE(X). \tag{6}$$

Since $L_X(p) = \frac{1}{E(X)} \int_0^p F_X^{-1}(t)dt$, is scale invariant, from (5) and (6), then the chi-square curve as a function of Lorenz curve, also is scale invariant. So, we have

$$\Psi_Y(p) = \frac{[p - L_Y(p)]^2}{p} = \frac{[p - L_X(p)]^2}{p} = \Psi_X(p).$$

Consequently, the scale invariant of the chi-square index is also obtained. □

Proposition 4.3. *The chi-square inequality index is sensitivity to location reliability.*

Proof. Let $Y = X + b$ for any $b > 0$, $\Psi_X(p)$ and $\Psi_Y(p)$ be point measures of the chi-square index of X and Y, respectively. So, we have

$$\begin{aligned}\Psi_Y(p) &= \frac{[\mu_Y - \mu_Y(p)]^2 \times p}{\mu_Y^2} \\ &= \frac{[(\mu_X + b) - (\mu_X(p) + b)]^2 \times p}{(\mu_X + b)^2} \\ &= \Psi_X(p) \times (\frac{\mu_X}{\mu_X + b})^2.\end{aligned}$$

For every fixed value $b > 0$, the ratio $(\frac{\mu_X}{\mu_X+b})^2$ is the values in $(0, 1)$.Therefore

$$\Psi_Y(p) < \Psi_X(p), \qquad \forall p \in (0, 1). \tag{7}$$

Consequently, the consistency with transiation of the chi-square index follows from (7). □

Remark 4.4. *An similar consideration holds for any $b < 0$, in this case $\Psi_Y > \Psi_X$.*

5. Stochastic orders based on the $\Psi(p)$ curve

Inequality curves can be used to establish certain rankings, enabling the comparison of distributions in terms of inequality. This type of comparison within a single model helps us understand how model parameters impact inequality. The section discusses a partial order derived from the Lorenz curve and another from the chi-square curve, showing that both curves yield equivalent partial orders.

Definition 5.1. *Let X_1 and X_2 be random variables belonging to $\mathcal{L}$ class. The Lorenz order $\le_L$ on $\mathcal{L}$ is defined by,*

$$X_1 \le_L X_2 \Leftrightarrow L_{X_1}(p) \ge L_{X_2}(p), \qquad p \in [0, 1].$$

If $X_1 \le_L X_2$, then X_1 expresses less inequality than X_2 in the Lorenz's sense. From a graphic perspective, the random variable X_1 is smaller than X_2 in this order, if its Lorenz curve lies above the Lorenz curve of X_2 for all $p \in (0, 1)$. Similarly, to the ordering based on the Lorenz curve, other stochastic orders have been used in inequality analysis. The links between some of the different stochastic orders and their relationships with inequality have been extensively studied (see Polisicchio and Porro (2009) [11] and Arnold (2015) [3]). Compared with the Lorenz stochastic order, we can define stochastic order curves for the chi-square inequality curve as follows.

Definition 5.2. *Let X_1 and X_2 be random variables belonging to $\mathcal{L}$ class. The chi-square order $\leq_\Psi$ on $\mathcal{L}$ is defined by*

$$X_1 \leq_\Psi X_2 \Leftrightarrow \Psi_{X_1}(p) \leq \Psi_{X_2}(p), \qquad p \in (0,1].$$

If $X_1 \leq_\Psi X_2$, then X_1 shows less inequality than X_2 in chi-square sense.

Lemma 5.3. *Let $X_1, X_2 \in \mathcal{L}$. Then*

$$X_1 \leq_L X_2 \Leftrightarrow X_1 \leq_\Psi X_2. \tag{8}$$

Proof. According to the definition, $X_1 \leq_L X_2$, means $L_{X_1}(p) \geq L_{X_2}(p)$ for any fixed $p \in (0,1]$, also, since the chi-square curve can be written in the form

$$\Psi(p) = \frac{[p - L(p)]^2}{p},$$

and conversely

$$L(p) = p[1 - \sqrt{\frac{\Psi(p)}{p}}].$$

It is obvious that

$$L_{X_1}(p) \geq L_{X_2}(p) \Leftrightarrow \Psi_{X_1}(p) \leq \Psi_{X_2}(p), \qquad \forall p \in (0,1).$$

In other words, the result can be rewritten as:

$$X_1 \leq_L X_2 \Leftrightarrow X_1 \leq_\Psi X_2.$$

□

The stochastic order determined by the Bonferroni and Zenga curves appears to be essentially the same as the stochastic order determined by the Lorenz curve. This equivalence is discussed in studies by Arcagni and Porro (2014) [2], as well as Arnold (2015) [3]. As chi-square and Lorenz stochastic orders are equivalent in (8), the relationship based on curve order leads to Proposition 5.4.

Proposition 5.4. *Let $X_1, X_2 \in \mathcal{L}$. Then the following statements are equivalent:*

$$X_1 \leq_\Psi X_2 \Leftrightarrow X_1 \leq_L X_2 \Leftrightarrow X_1 \leq_B X_2 \Leftrightarrow X_1 \leq_Z X_2,$$

where $\leq_B$ and $\leq_Z$ denote the Bonferroni order and the Zenga order, respectively.

Arcagni and Porro (2014) [2] and Arnold (2015) [3] provided information on the definitions and characteristics of Bonferroni and Zenga stochastic order.

6. The Behavior of the chi-square curve in some income models

A comparison of the Lorenz and chi-square curves leads to a comparison of the curves of some income models. Two curves are studied in this section. The Lorenz curve is currently the oldest but also the most used curve with forced behavior at the beginning and end of the interval. The chi-square curve is the most recent, although it is related to the Lorenz curve. This curve, like the Lorenz curve, has forced behavior at the beginning and end of its range. The chi-square curve has different shapes that allow us to distinguish between different situations in terms of inequality. Several densities have been predicted in the literature to model the income distribution. However all these densities are arranged in positive support. The simplest and thus more widely used distributions are the exponential, uniform and Pareto distributions. The distribution models just studied have one or two parameters, so the curves behave differently. A three-parameter distribution is presented to better fit the curved tails. We study the Burr XII distribution as a three-parameter model.

6.1. Exponential model

Let X be a random variable with exponential distribution, $F_X(x) = 1 - e^{-\alpha x} I_{(x>0)}$, where $\alpha > 0$ is its parameter distribution, then the $L(p)$ curve is

$$L(p) = p + (1-p)\log(1-p), \qquad p \in [0,1],$$

and from that, the chi-square curve is obtained as

$$\Psi(p) = \frac{[(1-p)\log(1-p)]^2}{p}, \qquad p \in (0,1].$$

Here it is important to pay attention to the fact that the scale parameter α does not affect the inequality, i.e. the chi-square curve, the Lorenz curve and the resulting measures of inequality do not depend on α. (see Figure 3).

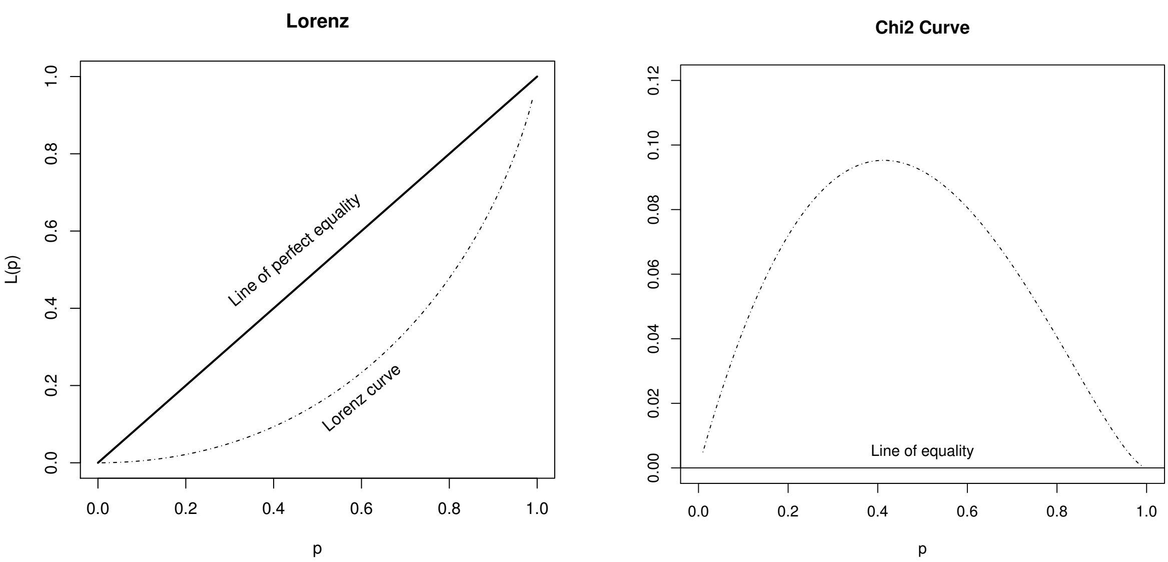

Figure 3: Lorenz and chi-square curves for exponential model

6.2. Pareto distribution

The random variable X has a Pareto model if its distribution function has the following form:

$$F_X(x) = 1 - (\frac{x}{x_0})^{-\theta} I_{(x>x_0)},$$

where $x_0 > 0$ and $\theta > 1$. In this case,

$$L(p) = 1 - (1-p)^{(1-\frac{1}{\theta})}, \qquad p \in [0,1],$$

and from that, the chi-square curve can be obtained as

$$\Psi(p) = \frac{[p - (1-(1-p)^{(1-\frac{1}{\theta})})]^2}{p}, \quad p \in (0,1].$$

It can be seen that the changes of the two curves are not affected by the changes of the scale parameter x_0. In another sense, inequality indices and curves are invariant relative to the scale parameter.

Some $L(p)$ and $\Psi(p)$ curves for the Pareto model are drawn in Figures 4. Each curve corresponds to a different choice of the distribution parameter θ.

As θ increases, the Lorenz curve also increases (while inequality decreases) for any fixed $p \in (0,1)$. The $\Psi(p)$ curve decreases as θ increases, indicating a decrease in inequality. Thus, the distribution parameter θ serves as an inverse indicator of inequality for both curves.

6.3. The uniform distribution

The random variable X has a classical uniform model with $F_X(x) = \frac{x-\alpha}{\beta-\alpha} I_{(\alpha<x<\beta)}$, where $0 < \alpha < \beta$ are parameters model, then the $L(p)$ curve is

$$L(p) = \frac{2\alpha p + (\beta - \alpha)p^2}{\alpha + \beta}, \qquad p \in [0,1],$$

and therefore the $\Psi(p)$ curve can be shown as

$$\Psi(p) = p[(\frac{\beta - \alpha}{\beta + \alpha})(1-p)]^2, \qquad p \in (0,1].$$

Here, the role of these two parameters α and β in the uniform model with respect to the Lorenz and chi-square curves is examined. Figure 5 shows some $L(p)$ and $\Psi(p)$ curves with different values of α and $\beta = 10$. In the Figure 6, the distribution parameter α is kept constant at 2, and the value of β changes. It is clear that if the value of the distribution parameter β is fixed, then α is an inverse inequality index, and if α is held, then β is a direct index.

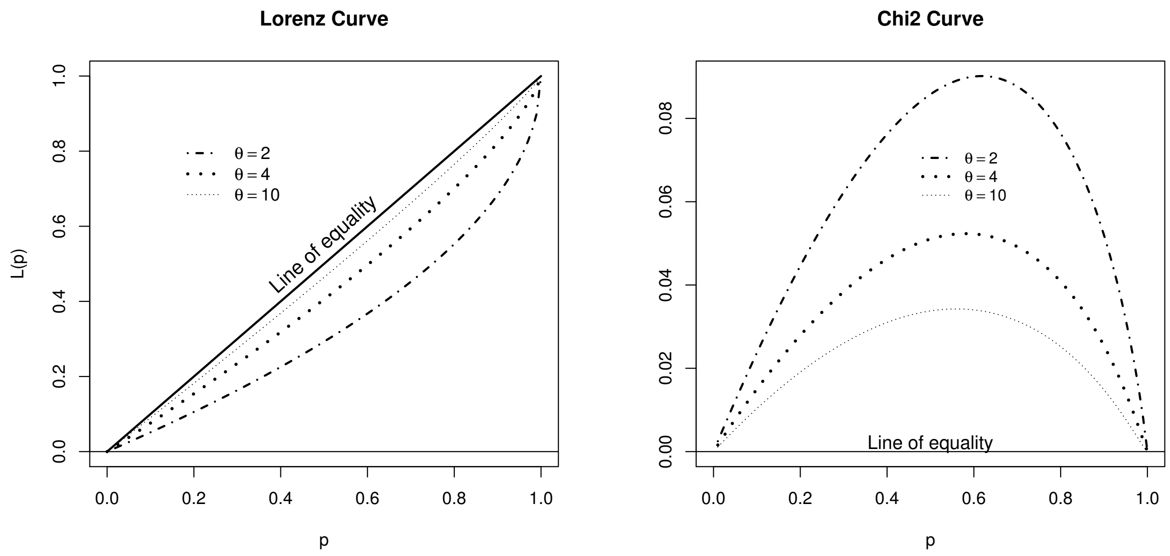

Figure 4: Lorenz and chi-square curves for Pareto model

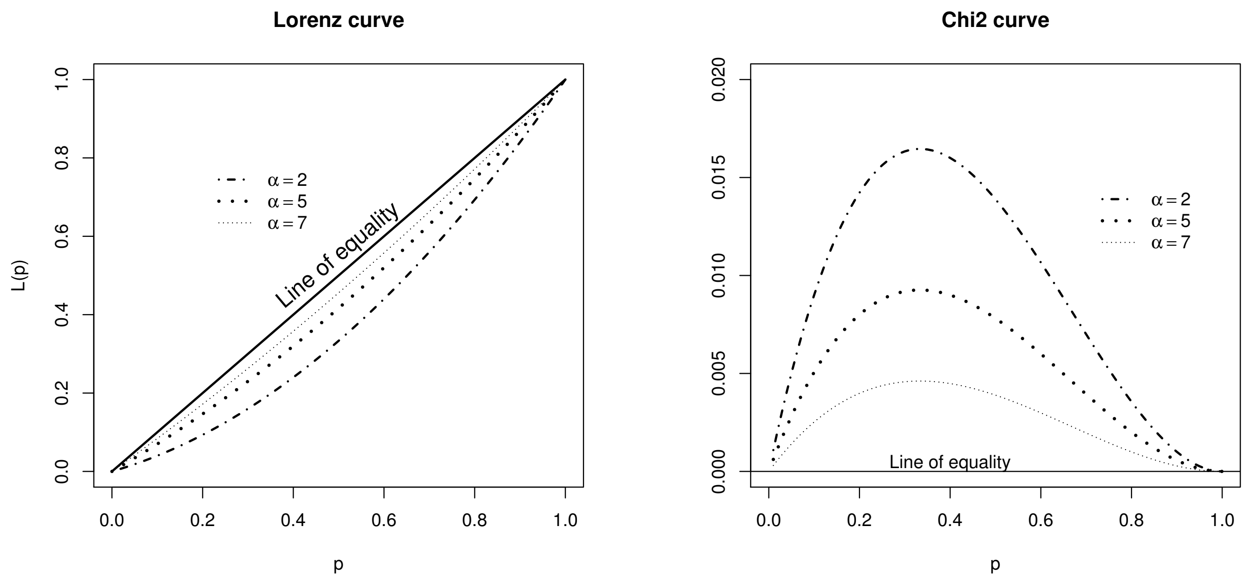

Figure 5: Lorenz and chi-square curves for the uniform distribution with $\beta = 10$ and different values of α.

6.4. The Burr XII distribution

Suppose X be a non-negative random income variable with Burr XII model and corresponding cdf,

$$F(x) = 1 - \frac{1}{[1+(\frac{x}{b})^{a}]^{q}},$$

where $a, b, q > 0$. If $q > \frac{1}{a}$, then using (1), the Lorenz curve can be obtained as:

$$\begin{aligned} L(p) &= \frac{1}{\mu}\int_{0}^{p} b[(1-y)^{-\frac{1}{q}}-1]^{\frac{1}{a}}dy \\ &= \frac{bq}{\mu}\int_{0}^{z} t^{\frac{1}{a}}(1-t)^{q-\frac{1}{a}-1}dt \\ &= I_z(1+\frac{1}{a}, q-\frac{1}{a}), \qquad p \in [0,1], \end{aligned}$$

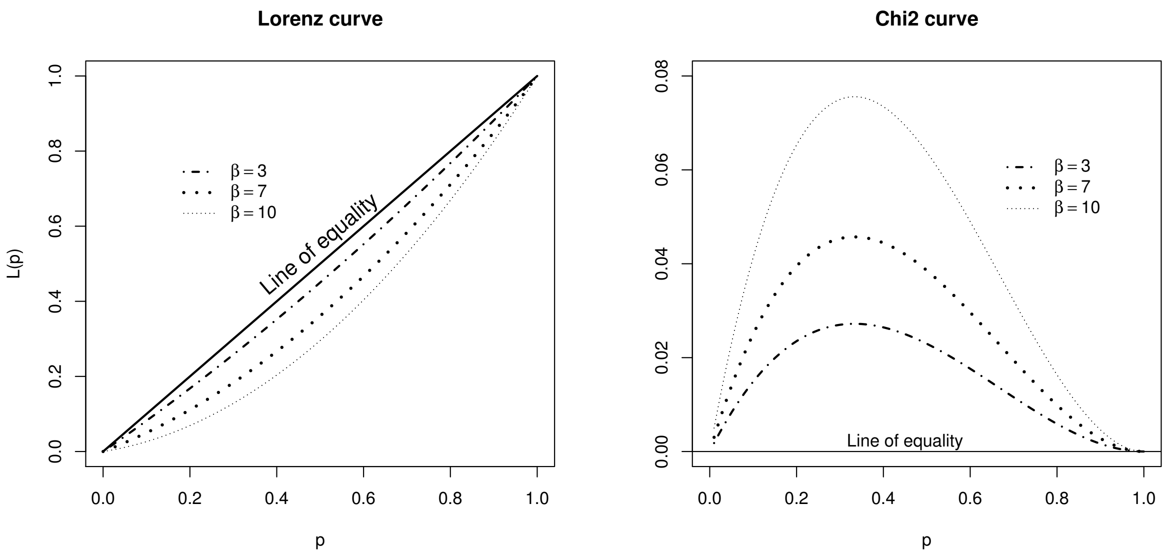

Figure 6: Lorenz and chi-square curves for the uniform model with $\alpha = 2$ and different values of β.

where $z = 1 - (1-p)^{\frac{1}{q}}$ and $I_x(a, b)$ shows the incomplete beta function ratio defined as

$$I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}.$$

Therefore, $\Psi(p)$ curve can be written as:

$$\Psi(p) = \frac{[p - I_z(1+\frac{1}{a}, q-\frac{1}{a})]^2}{p}, \qquad p \in (0,1].$$

In this study, the impact of the parameter q on the Burr XII model was examined in comparison to Lorenz and chi-square curves. It was observed that as b serves as the scale parameter in the Burr XII model, the inequality indices and curves remain unchanged regardless of its value, ensuring that our findings are not influenced by this choice.

Figure 7 displays various $L(p)$ and $\Psi(p)$ curves for different values of q with a fixed value of $a = 2$. It is clear that when the distribution parameter a is constant, the value of q serves as an indicator of inverse inequality. Similarly, when q is constant, a also acts as an indicator of inverse inequality.

7. Sampling estimator of chi-square index

Suppose $X_1, \dots, X_n$ be a random sample of size n from cdf F and $X_{1:n} \leq \dots \leq X_{j:n} \leq \dots \leq X_{n:n}$ be the corresponding order statistics. The chi-square estimator can be defined by combining empirical cdf of F ($\hat{F}_n$) instead of F in (2) and (3) as

$$\hat{\Psi}_n = \frac{1}{n}\sum_{j=1}^{n}\frac{j}{n}(1 - \frac{\bar{X}_{j:n}}{\bar{X}})^2, \qquad j = 1, 2, \dots, n,$$

where $\bar{X}_{j:n}$ is the partial ordered mean and $\bar{X}$ denoting the sample mean of $X_1, \dots, X_n$.

One of the advantages of using the chi-square estimator is that it assumes normal distribution for large samples. It can be mentioned that $\Psi(p)$ curve is a function of F based on (3) and (1). Thus, for more emphasize, we will show $\Psi(F)$ instead of $\Psi(p)$ in the following theorem.

Theorem 7.1. *Let X be a random variable belonging to $\mathcal{L}$ class with $E(|X|^{2+\alpha}) < \infty$ for $\alpha \geq 1$, the estimator of chi-square index is asymptotically normal.*

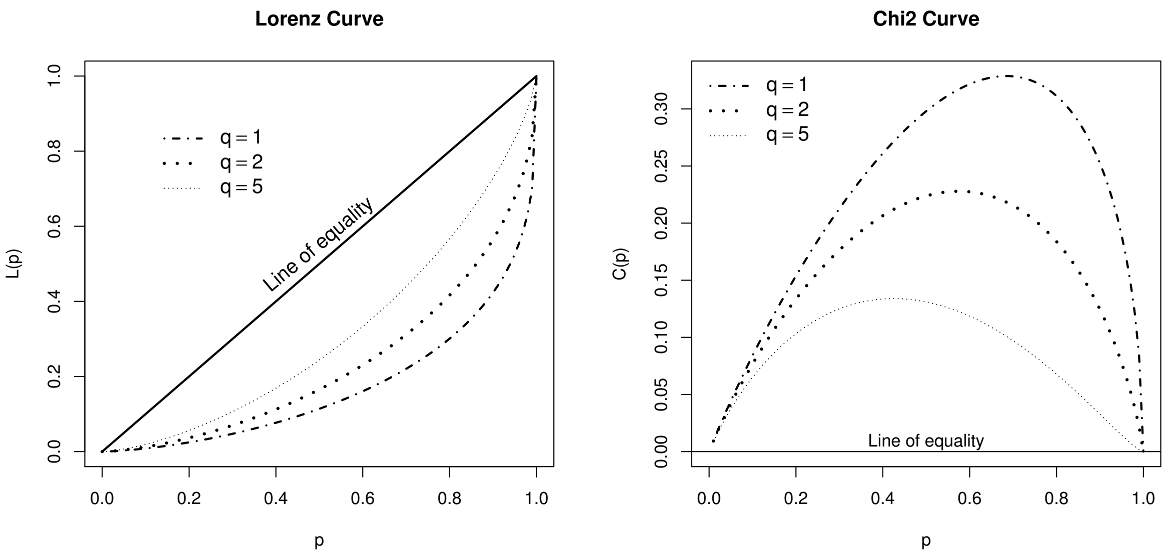

Figure 7: Lorenz and chi-square curves for the Burr XII model with $a = 2$ and different values of q.

Proof. Suppose that the random variable X has a continuous and differentiable distribution function, the chi-square estimator may be represented as (see Ruper, 2012 [8])

$$\hat{\Psi}_n = \Psi + \frac{1}{n}\sum_{j=1}^{n} \mathrm{K}_\Psi(X_j) + o(n^{\frac{-1}{2}}),$$

where $\mathrm{K}_\Psi(X_j)$ denotes the influence function evaluated at the point X_j, i.e.

$$\mathrm{K}_\Psi(X_j) = \lim_{\gamma\to 0} \frac{\Psi(F + \gamma(\tau_{X_j} - F)) - \Psi(F)}{\gamma},$$

and τ_X shows the distribution with unit mass at X. It results that the chi-square inequality estimator has normal distribution asymptotically. Then

$$\sqrt{n}(\hat{\Psi}_n - \Psi) \xrightarrow{d} N(0, \sigma^2_\Psi),$$

where $\xrightarrow{d}$ gives the meaning convergence in distribution and $\sigma^2_\Psi = var(\mathrm{K}_\Psi(X))$. □

8. Data analysis and simulation study

The analysis of the information relies on the total income of American households from the year 2019, as reported by the US Census Bureau's Current Population Survey (CPS). The income figures provided can be found in Table 2, displaying the relative frequencies of different income groups out of a total of $N = 128,451$ individuals. These data sets are available for review on the Census Bureau's website at www.census.gov/hhes/www/income/data/incpovhlth. It is important to mention that due to government regulations and privacy statutes, income data is typically shared in the form of frequency distribution tables.

Table 2: 2019 United States Household Income Relative Frequency Table

Total	[0,15)	[15,25)	[25,35)	[35,50)	[50,75)	[75,100)	[100,150)	[150,200)	[200,∞)
100	9.1	8	8.3	11.7	16.5	12.3	15.5	8.3	10.3

The data correspond to a sample of 128,451 households total money income families combined into 9 groups. Since the data are reported in a grouped format, maximum likelihood estimates are obtained by maximizing the multinomial log-likelihood function (see Cowell, 2000 [5]). The results of estimation and goodness of fit show that the GB2 distribution with probability density function

$$f_{GB2}(x; a, b, p, q) = \frac{a x^{a\,p-1}}{b^{a\,p} B(p, q)\left[1 + (\frac{x}{b})^a\right]^{p+q}},$$

with $x > 0$, $a, b > 0$ and where $B(p, q)$ is the Beta function given by

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^\infty \frac{t^{p-1}}{(1+t)^{p+q}} dt, \qquad p, q > 0.$$

has optimal performance based on goodness of fit criteria. The GB2 model has the advantage that many of the densities are special cases of this distribution and thus provides a good framework for discussion.

Here, we take Monte Carlo samples from a GB2 model with shape parameters equal to $a = 1.544, p = 0.851, q = 2.314$ and scale parameter equal to 158.461. We note that b as a scale parameter of the GB2 distribution does not affect the inequality. The parameters considered are the maximum likelihood estimates of the GB2 model based on 2019 households total money income of the US family from the US Census Bureau, Current Population Survey (CPS).

8.1. Asymptotic normality

Here, we examine the extent to which the chi-square estimator proposed here provides valid inferences using simulation. First, to see if the assumption of asymptotic normality provides a good approximation, the simulations are based on samples from the appropriate distribution of the real data. The true value of the chi-square index of this distribution is indicated by the symbol (Ψ_0) . Figure 8 shows graphs are of the empirical distribution function of 10,000 substantiations of the statistic $\tau_\Psi = \frac{\hat{\Psi}-\Psi_0}{\hat{\sigma}_{\hat{\Psi}}}$. It should be noted that the estimation of the standard error was done by the bootstrapping method. For sample sizes $n = 50$ and 100 the graph of the standard normal cdf is also shown as a template in Figure 8.

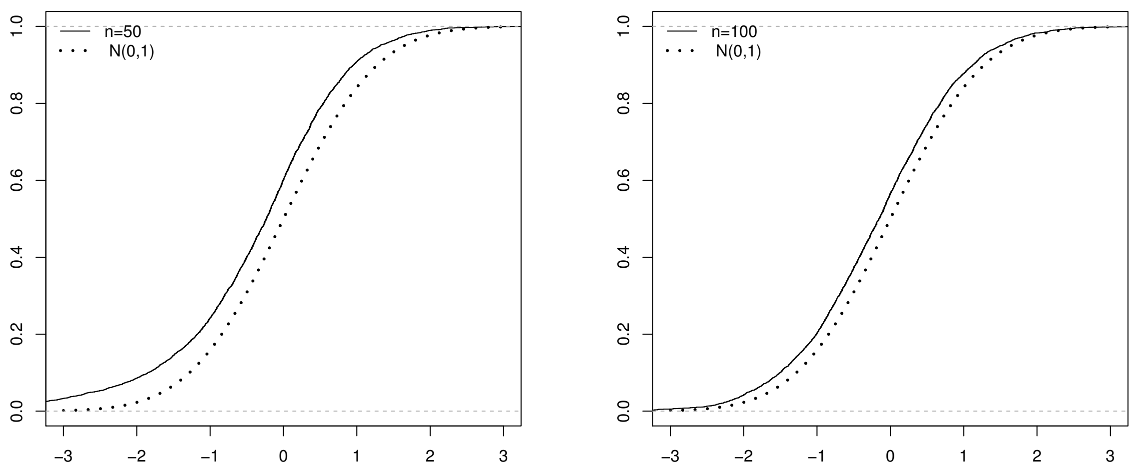

Figure 8: Distribution of chi-square standardized statistic as a function of n

It is observed that the chi-square estimator is consistent and it has a good asymptotic standard normal.

For more information, here we compare the new index with the usual Gini inequality index (see Giorgi and Gigliarano (2017) [6]). To do this, it simulates, 10000 simulated samples of size $n = 100$ from the fitted distribution to income data and evaluates both indices in each sample. The histograms in Fig 9 show the corresponding sampling distributions for this set of simulations.

The figures show that the sample distributions of the chi-square index and the Gini coefficient are both approximately normal. Moreover, the sampling distributions of both estimators are similar.

8.2. Skewness and kurtosis

In this study, we simulate to evaluate how well the new chi-square inequality index performs compared to the conventional Gini coefficient. We base our comparison on skewness and kurtosis under a distribution fitted to actual data using specified parameters. By generating 10,000 estimates for both indices across sample sizes of 10, 20, 30, 50, 100, and 500, we present the findings in Table 3.

The results of Table 3 show that the sample distributions of Gini and chi-square indices do not differ much in terms of skewness and kurtosis. Also, as the sample size increases, the distribution of estimators tends to approach a normal distribution due to the Central Limit Theorem.

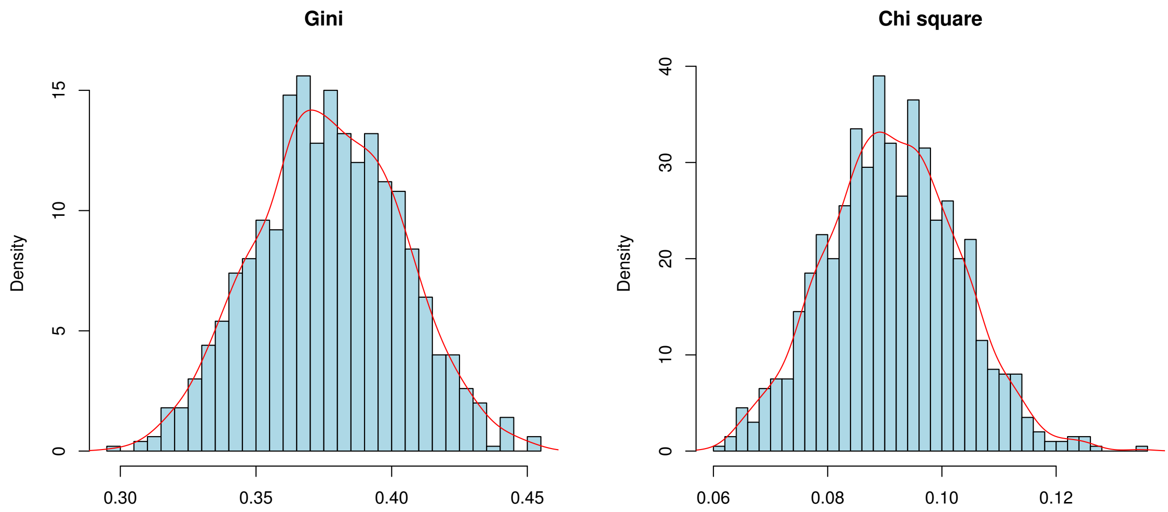

Figure 9: Histogram of the distributions of Gini and chi-square estimates computed on 10000 samples of size $n = 100$

Table 3: Comparision the skewness and kurtosis for Gini and chi-square measures

	Skewness		Kurtosiss	
n	Gini	chi-square	Gini	chi-square
10	0.311	0.260	0.359	0.301
20	0.220	0.196	0.235	0.210
30	0.178	0.163	0.185	0.171
50	0.137	0.129	0.141	0.132
100	0.097	0.093	0.099	0.094
500	0.069	0.066	0.069	0.067

8.3. Coverage probability

Here, we compare the performance of asymptotic confidence intervals for a rival inequality measure of chi-square with the traditional Gini measure. Normal, t-bootstrap and BCa confidence intervals are considered for the fitted distribution of the real income data. Table 4 presents coverage probabilities for 10,000 confidence intervals (relative frequencies of confidence intervals that contain the true value of inequality) for each of the three types: normal, t-bootstrap and BCa with a nominal confidence level of 0.95.

Table 4: Coverage probability of Gini and chi-square measures

	Normal		t-bootstrap		BCa	
n	Gini	chi-square	Gini	chi-square	Gini	chi-square
10	0.828	0.885	0.866	0.930	0.840	0.845
20	0.899	0.925	0.911	0.932	0.886	0.885
30	0.901	0.926	0.916	0.935	0.892	0.893
50	0.918	0.929	0.925	0.941	0.905	0.905
100	0.935	0.940	0.939	0.942	0.921	0.925
500	0.938	0.946	0.940	0.948	0.923	0.926

The results show that for a large sample size, the confidence interval coverage accuracy of the resampling methods is quite close to the nominal confidence level. As can be seen, there is no significant difference in coverage probability of the two inequality indices. However, we can see from Table 4 that the chi-square estimator performs well in terms of confidence interval coverage probabilities compared to the Gini estimator.

9. Testing with real data

This section displays Lorenz and chi-square curves visualizing US household income distribution for 2000, 2006, and 2013, based on US Census data. The income data used, subject to privacy regulations, comes from the US Census Bureau and is accessible via frequency tables on their website. These specific data sets are available for review on the Census Bureau's website[1].

The left side of Figure 10 displays the Lorenz curves for the household distributions being studied. These curves indicate varying levels of income inequality, but it is not easy to discern any significant changes in inequality over the years mentioned. Obtaining more specific information about how inequality has evolved dynamically is challenging. The right side of Figure 10 shows the chi-square curves corresponding to these distributions. The chi-square curve provides a more detailed and dynamic view of income inequality over time, allowing for a better understanding of how income distribution has changed. The increase in inequality for higher incomes from 2000 to 2013 is clearly visible in the chi-square curve, highlighting the growing gap between the rich and the poor. In contrast, the Lorenz curve may not show these changes as clearly, as it only provides a snapshot of income distribution at a specific point in time. By using both curves together, researchers can gain a more comprehensive understanding of income inequality trends and patterns.

Overall, the chi-square curve offers a valuable alternative to the Lorenz curve for analyzing income inequality, providing insights that may not be apparent from traditional measures. Its ability to capture changes over time makes it a useful tool for policymakers and researchers seeking to address issues of economic inequality.

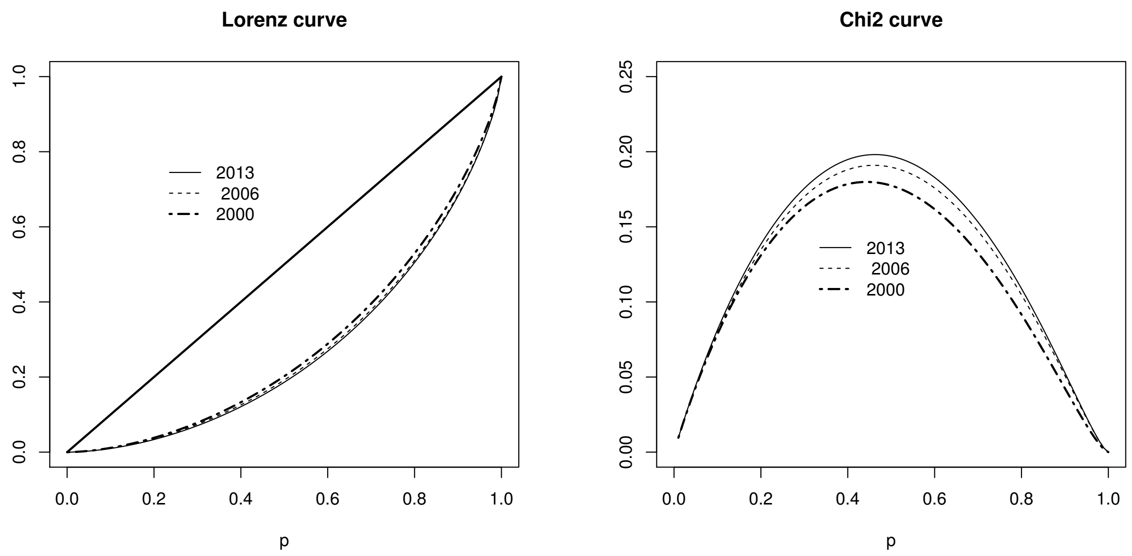

Figure 10: The Lorenz curves (left) and the chi-square curves (right) of household income in different years

10. Results

For the first time, this article introduced a new concept called the chi-square inequality index, which was created using the chi-square distance function. The study examined the important features needed for a reliable inequality index and provided a detailed explanation of the chi-square inequality curve in various inequality models. A comparison between the chi-square curve and the traditional Lorenz curve was also carried out. Overall, the chi-square curve and its related index appear to be a more robust and reliable alternative to the Lorenz curve for analyzing income distributions and measuring inequality. The chi-square curve is more explanatory and flexible, and all direct inequality indicators for the Lorenz curve are also applicable to the chi-square curve. Additionally, a stochastic order based on the chi-square inequality curve was introduced. Furthermore, while introducing a suitable estimator for this inequality index, we showed that this estimator has a normal asymptotic distribution. Also, the simulation results showed that the estimation of the new index has a more favorable coverage probability than the estimation of the common Gini index.

[1] www.census.gov/hhes/www/income/data/incpovhlth

These findings suggest that researchers and policymakers may benefit from using the chi-square curve and its associated index when analyzing income inequality. The chi-square curve provides a comprehensive and accurate representation of income distribution, making it a valuable tool for studying inequality dynamics and designing effective policy interventions.

Acknowledgement

The authors are grateful to the anonymous referees for their careful reading and useful comments to improve the quality of the paper.

References

[1] E. B. Andersen, *A goodness of fit test for the rasch model*, Psychometrika, 38 (1973), pp. 123–140.

[2] A. Arcagni and F. Porro, *The Graphical Representation of Inequality*, Revista Colombiana de Estadística, 37 (2014), pp. 419–437.

[3] B. C. Arnold, *On zenga and bonferroni curves*, METRON, 73 (2015), pp. 25–30.

[4] S.-H. Cha, *Comprehensive survey on distance/similarity measures between probability density functions*, Int. J. Math. Model. Meth. Appl. Sci., 1 (2007), pp. 300–307.

[5] F. A. Cowell, *Measurement of inequality*, in Handbook of Income Distribution, vol. 1, Elsevier, 2000, ch. 2, pp. 87–166.

[6] G. M. Giorgi and C. Gigliarano, *The Gini concentration index: A review of the inference literature*, Journal of Economic Surveys, 31 (2017), pp. 1130–1148.

[7] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc, 2003.

[8] R. G. Miller, *Simultaneous Statistical Inference*, Springer New York, NY, 2nd ed., 2012.

[9] S. Mirzaei, G. R. Mohtashami Borzadaran, M. Amini, and H. Jabbari, *Some properties of the Canberra inequality index*, Hacet. J. Math. Stat., 46 (2017), pp. 1159–1174.

[10] L. Pasquazzi and M. Zenga, *Components of Gini, Bonferroni, and Zenga inequality indexes for EU income data*, Journal of Official Statistics, 34 (2018), pp. 149–180.

[11] M. Polisicchio and F. Porro, *A comparison between Lorenz $L(p)$ curve and Zenga $I(p)$ curve*, Statistica Applicata, 21 (2011), pp. 289–301.

[12] J. M. Sarabia, *Parametric lorenz curves: Models and applications*, in Modeling Income Distributions and Lorenz Curves, vol. 5, Springer New York, 2008, ch. 9, pp. 167–190.

[13] J. M. Sarabia and V. ía, *Explicit expressions of the Pietra index for the generalized function for the size distribution of income*, Physica A: Statistical Mechanics and its Applications, 416 (2014), pp. 582–595.

Please cite this article using:

Shahryar Mirzaei, Sayed Mahdi Amir Jahanshahi, The chi-square statistic as an income inequality index, AUT J. Math. Comput., 6(4) (2025) 327-340
https://doi.org/10.22060/AJMC.2024.22945.1205