

Income Inequality Measures

Johan Fellman

Hanken School of Economics, Helsinki, Finland

Email: fellman@hanken.fi

How to cite this paper: Fellman, J. (2018) Income Inequality Measures. *Theoretical Economics Letters*, 8, 557-574.
<https://doi.org/10.4236/tel.2018.83039>

Received: January 19, 2018

Accepted: February 11, 2018

Published: February 14, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Income distributions are commonly unimodal and skew with a heavy right tail. Different skew models, such as the lognormal and the Pareto, have been proposed as suitable descriptions of income distribution and applied in specific empirical situations. More wide-ranging tools have been introduced as measures for general comparisons. In this study, we review the income analysis methods and apply them to specific Lorenz models.

Keywords

Lorenz Curve, Gini Index, Pietra Index, Pareto Distribution, Simplified Rao-Tam Distribution, Chotikapanich Distribution

1. Introduction

Income distributions are commonly unimodal and skew with a heavy right tail. Therefore, different skew models, such as the lognormal and the Pareto, have been proposed as suitable descriptions of income distribution, but they are usually applied in specific empirical situations [1]. For general studies, more wide-ranging tools have been considered. The target for them is to introduce measures that are useable for comparisons of different distributions. Primary income data yield the most exact estimates of income inequality coefficients such as Gini and Pietra. Earlier studies have shown that no method is always optimal. Therefore, different attempts are still worth studies. In this study, we review income analysis methods based on Lorenz curves. The theory is applied to specific models.

Fellman [2] analyzed different methods for numerical estimation of Gini coefficients. As an application of these methods, he considered Pareto distributions. Using Lorenz curves, various numerical integration attempts were made to obtain accurate estimates. Mettle *et al.* [3] considered Lorenz curves and estimated the Gini coefficient of income by Newton-Cotes methods, and compared the

accuracy of these estimates for some (Ghanaian) data.

2. Methods

The Lorenz curve. The most commonly used theory is based on the Lorenz curve. Lorenz [4] developed it in order to analyze the distribution of income and wealth within populations. He described the Lorenz curve, $L(p)$, for wealth within populations in the following way:

“Plot along one axis accumulated per cents of the population from poorest to richest, and along the other, wealth held by these per cents of the population”.

Consequently, $L(p)$ is an accumulated amount of income (wealth) defined as a function of the proportion p of the population. It satisfies the condition $L(p) \leq p$ because the income share of the poor is less than their proportion of the population. The increase ΔL caused by a fixed increase Δp of the population is a growing function of p , and accordingly, the derivative $L'(p)$ is an increasing function of p and $L(p)$ is a convex function [5].

Consider the income distribution $F_X(x)$ of a non-negative variable X . Let $f_X(x)$ be the corresponding frequency distribution and let the mean of X be $\mu_X = \int_0^\infty x f_X(x) dx$. Then, the Lorenz curve $L_X(p)$ is

$$L_X(p) = \frac{1}{\mu_X} \int_0^{x_p} x f_X(x) dx, \tag{1}$$

where x_p is the p quantile, that is $F_X(x_p) = p$. The Lorenz curve is not defined if the mean is zero or infinite. A Lorenz curve always starts at $(0,0)$ and ends at $(1,1)$. The higher the Lorenz curve, the lesser is the inequality of the income distribution. The diagonal $L(p) = p$ is commonly interpreted as the Lorenz curve for complete equality between income receivers, but according to [6] it is not perfectly associated with the Lorenz curve. As everyone has the same income level, strictly speaking, no one can be said to be at the lowest or highest level of the population. The associated Lorenz curve then exists only at the origin and the termination point by the definition of the curve. To overcome this problem, they adopted the convention of allocating any fraction $0 < x < 1$ of the population to be the lowest/highest x percent. This convention then allowed the 45 degree line through the origin to be associated with complete equality, as usually loosely taken to be so. This permitted Wang & Smyth [6] to use $L(p) = p$ as a useful component in the creation of Lorenz curves.

On the other hand, increasing inequality lowers the Lorenz curve, and theoretically, it can converge towards the lower right corner of the square. A sketch of a Lorenz curve is given in **Figure 1**.

Variable transformations. Consider a transformed variable $Y = g(X)$, where $g(\cdot)$ is positive and monotone increasing. Then, the distribution of $F_Y(y)$ is

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq g(x)) = P(X \leq x) = F_X(x). \tag{2}$$

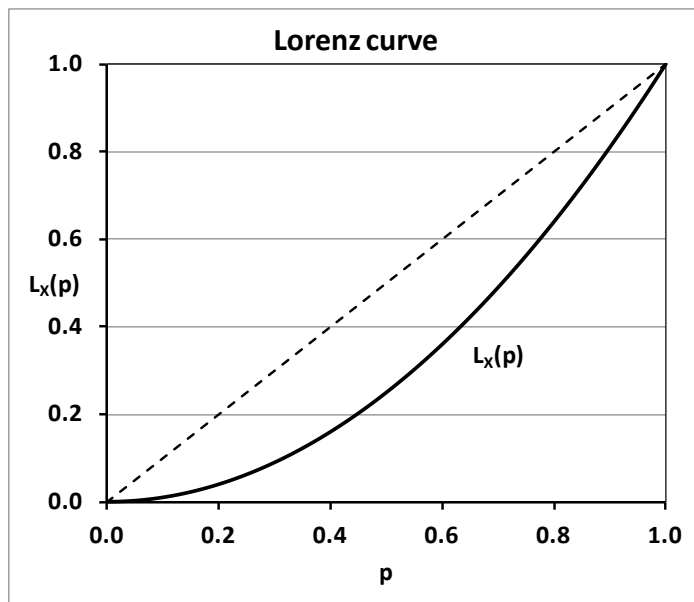


Figure 1. A sketch of a Lorenz curve $L_X(p)$. The diagonal interpreted as complete equality is included in the figure.

For the transformed variable Y , the p quantile y_p is $F_Y(y_p) = p$, that is, $y_p = g(x_p)$.

Now

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \frac{dx}{dy} = f_X(x) \frac{dx}{dy}. \tag{3}$$

Hence,

$$\mu_Y = \int_0^\infty y f_Y(y) dy = \int_0^\infty g(x) f_X(x) \frac{dx}{dy} dy = \int_0^\infty g(x) f_X(x) dx \tag{4}$$

and

$$L_Y(p) = \frac{1}{\mu_Y} \int_0^{x_p} g(x) f_X(x) dx. \tag{5}$$

If the transformation is linear $g(x) = \theta x$, then $Y = \theta X$, $\mu_Y = \theta \mu_X$,

$$L_Y(p) = \frac{1}{\theta \mu_X} \int_0^{x_p} \theta x f_X(x) dx = L_X(p), \tag{6}$$

and consequently, the Lorenz curve is invariant under linear transformations. A simple example of this property is that the Lorenz curve of income distributions is independent of the currency used. Another not so obvious result is that proportional income increase and flat tax policies are linear transformations and do not influence the Lorenz curve. Consequently, the Lorenz curve satisfies the general rules [1]:

To every distribution $F(x)$ with finite mean corresponds a unique Lorenz curve, $L_X(p)$. The contrary does not hold because every curve $L_X(p)$ is a common Lorenz curve for a whole class of distributions $F(\theta x)$, where θ is

an arbitrary positive constant.

Consider two variables X and Y , their distributions $F_X(x)$ and $F_Y(y)$, and their Lorenz curves $L_X(p)$ and $L_Y(p)$. If $L_X(p) \geq L_Y(p)$ for all p , then measured by the Lorenz curves, the distribution $F_X(x)$ has lower inequality than the distribution $F_Y(y)$ and $F_X(x)$ is said to *Lorenz dominate* $F_Y(y)$. We denote this relation $F_X(x) \succ_L F_Y(y)$ [1]. An example of Lorenz dominance is given in **Figure 2**.

Income inequalities can be of different type and the corresponding Lorenz curves may intersect and for these no Lorenz ordering can be identified (cf. **Figure 3**). The Lorenz curve $L_1(p)$ in **Figure 3** corresponds to a population with

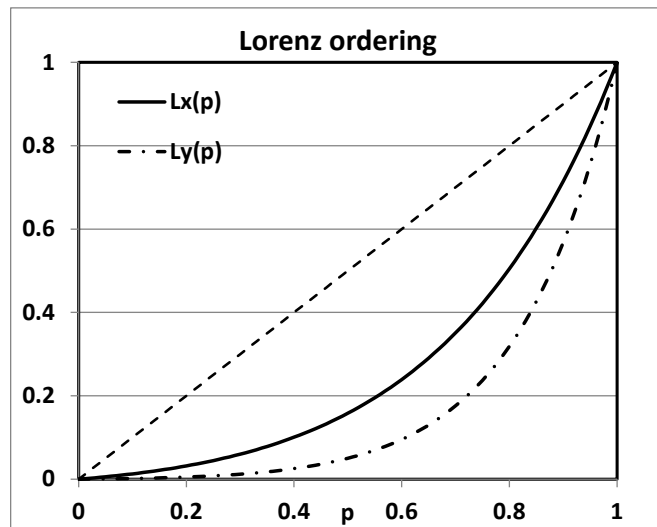


Figure 2. Lorenz curves with Lorenz ordering, that is, $L_X(p) \succ_L L_Y(p)$.

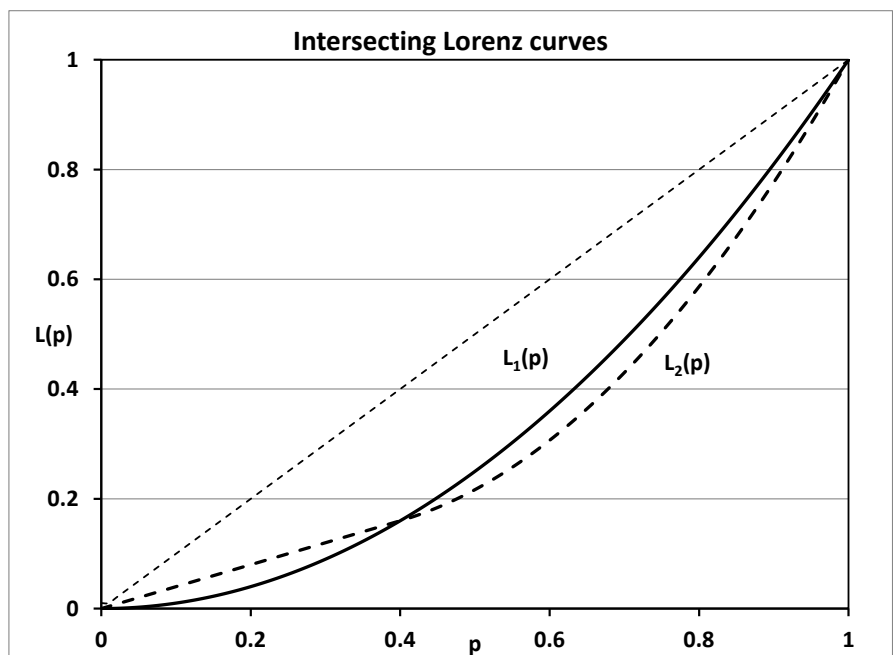


Figure 3. Two intersecting Lorenz curves.

very poor among the poor and rich who are not so rich. On the other hand, Lorenz curve $L_2(p)$ corresponds to a population where the poor are relatively not so poor and the rich are relatively rich. For intersecting Lorenz curves, alternative inequality measures have to be defined.

Properties of Lorenz curves. The Lorenz curve has the following general properties:

- i) $L(p)$ is monotone increasing,
- ii) $L(p) \leq p$,
- iii) $L(p)$ is convex,
- iv) $L(0) = 0$ and $L(1) = 1$.

If the Lorenz curve is differentiable, the derivative has the following properties.

Let $L_X(p) = \frac{1}{\mu_X} \int_0^{x_p} x f_X(x) dx$, $F_X(x_p) = p$ and the density function $f_X(x)$.

When we differentiate the equation $F_X(x_p) = p$, we obtain

$$\frac{dF_X(x_p)}{dp} = \frac{dF_X(x_p)}{dx_p} \frac{dx_p}{dp} = 1,$$

$$f_X(x_p) \frac{dx_p}{dp} = 1 \tag{7}$$

and

$$\frac{dx_p}{dp} = \frac{1}{f_X(x_p)}. \tag{8}$$

The differentiation of $L_X(p) = \frac{1}{\mu_X} \int_0^{x_p} x f_X(x) dx$ yields

$$\frac{dL_X(p)}{dp} = \frac{1}{\mu_X} \frac{d \int_0^{x_p} x f_X(x) dx}{dx_p} \frac{dx_p}{dp} = \frac{1}{\mu_X} x_p f_X(x_p) \frac{dx_p}{dp} = \frac{x_p}{\mu_X}, \tag{9}$$

and consequently,

$$\frac{dL_X(p)}{dp} = \frac{x_p}{\mu_X}. \tag{10}$$

If the Lorenz curve is differentiable twice, then the second derivative is

$$\frac{d^2L_X(p)}{dp^2} = \frac{1}{\mu_X} \frac{dx_p}{dp} = \frac{1}{\mu_X} \frac{1}{f_X(x_p)}.$$

Hence,

$$\frac{d^2L(p)}{dp^2} = \frac{1}{\mu_X f_X(x_p)}. \tag{11}$$

If $\lim_{p \uparrow 1}$ denotes the limit from the left, we can prove the following theorem [1]:

Theorem 1. *If μ_X exists, then $\lim_{p \uparrow 1} L'(p)(1-p) = 0$.*

Proof. Consider the integral $\int_x^\infty t f_X(t) dt$. If μ_X exists, then $\int_0^\infty t f_X(t) dt = \mu_X$ and for every $\varepsilon > 0$ there exists an x' such that $\int_x^\infty t f_X(t) dt < \varepsilon$ if $x > x'$. Choose p so that $x_p > x'$, then

$$\varepsilon > \int_{x_p}^\infty t f_X(t) dt \geq x_p \int_{x_p}^\infty f_X(t) dt = x_p(1-p) \tag{12}$$

and $\lim_{p \uparrow 1} x_p(1-p) = 0$.

As a consequence of (12),

$$\lim_{p \uparrow 1} L'_X(p)(1-p) = \lim_{p \uparrow 1} \frac{x_p}{\mu_X}(1-p) = \frac{1}{\mu_X} \lim_{p \uparrow 1} x_p(1-p) = 0.$$

Consider a one-parametric class of cumulative distribution functions $F(x, \theta)$, defined on the positive x-axis. If we assume that $F(x, \theta) = F(\theta x)$, i.e. it depends only on the product θx , then the following theorem holds [1]:

Theorem 2. Let $F(x, \theta)$ be an one-parametric class of distributions with the properties

- i) $F(x, \theta) = F(\theta x)$,
- ii) $F(\theta x)$ is defined on the positive x-axis,
- iii) $F(\theta x)$ and its derivative are continuous,
- iv) $\mu_x = E(X)$ exists.

Let $T = \theta X$, then

$$x_p(\theta) = \frac{t_p}{\theta} \tag{13}$$

and

$$\mu_x(\theta) = \frac{c}{\theta}, \tag{14}$$

where t_p and c are independent of θ .

Proof. Let θ be an arbitrary, positive parameter. Then the quantile $x_p(\theta)$ is defined by the equation $F(\theta x_p) = p$. If we define t_p by the equation $F(t_p) = p$, then t_p does not depend on θ and $\theta x_p(\theta) = t_p$, and (13) is proved. The formula (14) and the statement that $L(p) = \frac{1}{\mu(\theta)} \int_0^{x_p(\theta)} x dF(\theta x)$ is independent of θ is proved by using the substitution $t = \theta x$ in the integrals $E(X) = \int_0^\infty x dF(\theta x)$ and $L(p) = \frac{1}{\mu(\theta)} \int_0^{x_p(\theta)} x dF(\theta x)$.

Furthermore, we can prove the following [1]:

Theorem 3. Consider a function $L(p)$ defined on the interval $[0,1]$ with the properties

- 1) $L(p)$ is monotone increasing and convex to the p-axis,
- 2) $L(0) = 0$ and $L(1) = 1$,
- 3) $L(p)$ is differentiable,

iv) $\lim_{p \uparrow 1} L'(p)(1-p) = 0$,
 then $L(p)$ is a Lorenz curve of a distribution with finite mean.

Proof. If we denote the unknown distribution $F(x)$ and its derivative $f(x)$, then necessarily $L'(p) = \frac{x_p}{\mu}$. The derivative $L'(p)$ is a monotone-increasing function. If its inverse is denoted $M(p)$, we get the necessary relation

$$F(x_p) = p = M\left(\frac{x_p}{\mu}\right).$$

If $\theta = \frac{1}{\mu}$, then $F(x) = M(\theta x)$. Now we shall prove the sufficiency, that is, that $M(\theta x)$ is a distribution whose mean is $\mu = \frac{1}{\theta}$ and whose Lorenz curve is $L(p)$. We denote $M(\theta x) = F(x)$, then $f(x) = F'(x) = \theta M'(\theta x)$. After observing that the property (iv) indicates that $L'(p)$ is integrable from 0 to 1, we introduce the variable transformation

$$\begin{aligned} y &= M(\theta x) \\ dy &= \theta M'(\theta x) dx \\ x &= \frac{1}{\theta} L'(y). \end{aligned}$$

We obtain

$$\mu = \lim_{t \rightarrow \infty} \int_0^t x \theta M'(\theta x) dx = \lim_{p \uparrow 1} \int_0^p \frac{1}{\theta} L'(y) dy = \frac{1}{\theta} \lim_{p \uparrow 1} \int_0^p L'(y) dy = \frac{1}{\theta}.$$

The given function $L(p)$ has a monotone-increasing inverse function whose mean is μ .

Using the same transformation, we obtain that the Lorenz curve $\tilde{L}(p)$ of $F(x) = M(\theta x)$ is

$$\tilde{L}(p) = \theta \int_0^{x_p} x \theta M'(\theta x) dx = \int_0^p L'(y) dy = \int_0^p L'(y) dy,$$

and the theorem is proved.

These results have been collected in the following theorem [7] [8]:

Theorem 4. Consider a given function $L(p)$ with the properties

(i) $L(p)$ is monotone increasing and convex to the p -axis,

(ii) $L(0) = 0$ and $L(1) = 1$,

(iii) $L(p)$ is differentiable,

(i) $\lim_{p \uparrow 1} L'(p)(1-p) = 0$,

then $\tilde{L}(p)$ is the Lorenz curve of a whole class of distribution functions $F(\theta x)$, where θ is an arbitrary positive constant and the function $F(\cdot)$ is the inverse function to $L'(p)$.

Fellman [7] presented this result and later Fellman [8] presented the following theorem:

Theorem 5. A class of continuous distributions $F(x, \theta)$ with finite mean

has a common Lorenz curve if and only if $F(x, \theta) = F(\theta x)$.

Additional properties of the Lorenz curves. Consider the vertical difference D , between the diagonal and the Lorenz curve

$$D = p - L_X(p)$$

$$\frac{dD}{dp} = 1 - L'_X(p) = 1 - \frac{x_p}{\mu_X}$$

$$\frac{d^2D}{dp^2} = -L''_X(p) = -\frac{1}{\mu_X} \frac{dx_p}{dp} = -\frac{1}{\mu_X f_X(x)} < 0.$$

The maximum of D implies $1 - \frac{x_p}{\mu_X} = 0$, that is, $x_p = \mu$.

For $x_p = \mu_X$, $L'_X(p) = \frac{\mu_X}{\mu_X} = 1$ and at the point $p_\mu = F_X(\mu_X)$ the tangent

is parallel to the line of perfect equality. This is also the point at which the vertical distance between the Lorenz curve and the egalitarian line attains its maximum $D_{\max} = F(\mu_X) - L_X(F(\mu_X))$. This maximum is defined as the Pietra index, in this study denoted P , and discussed below [9].

Kleiber and Kotz have outlined a progressive development of how the income distributions can be characterized by their Lorenz curves [10] [11].

Income inequality indices. When Lorenz curves intersect, the corresponding distributions cannot be compared by the Lorenz curves. Consequently, the distributions have to be compared by numerical indices mainly based on the Lorenz curves.

Gini index. The most frequently used index is the Gini coefficient, G [12]. Using the Lorenz curves, this coefficient is the ratio of the area between the diagonal and the Lorenz curve and the whole area under the diagonal (cf. Figure 1). The formula is

$$G = 1 - 2 \int_0^1 L(p) dp. \tag{15}$$

This definition yields Gini coefficients satisfying the inequalities $0 < G < 1$. The higher the G value, the lower the Lorenz curve and the stronger the inequality. If $G_X < G_Y$, then the distribution $F_X(x)$, measured by the Gini coefficient, has lower inequality than the distribution $F_Y(y)$ and we say that $F_X(x)$ *Gini dominates* $F_Y(y)$, and denote this relation $F_X(x) \succ_G F_Y(y)$ [13]. Finally, one observes the obvious result $F_X(x) \succ_L F_Y(y) \Rightarrow F_X(x) \succ_G F_Y(y)$, that is, Lorenz dominance implies Gini dominance.

The coefficient allows direct comparison of the income of two populations' distributions, regardless of their sizes. The Gini's main limitation is that it is not easily decomposable or additive. Also, it does not respond in the same way to income transfers between people in opposite tails of the income distribution as it does to transfers in the middle of the distribution. The reason for the popularity of the Gini coefficient is that it is easy to compute being a ratio of two areas in

Lorenz curve diagrams. As a disadvantage, the Gini index only maps a number to the properties of a diagram, but the diagram itself is not based on any model of a distribution process. The “meaning” of the Gini index can only be understood empirically. Hence, the Gini does not capture where in the distribution the inequality occurs. As an additional result, two very different distributions of income having different Lorenz curves can have the same Gini index.

Using the Gini coefficient presented in the text, one can compare the Gini coefficients for $L_1(p)$ and $L_2(p)$ in **Figure 3**; $L_1(p)$ has less inequality ($G_1 = 0.333$) than $L_2(p)$ ($G_2 = 0.360$).

There are other inequality measures defined by the Gini coefficient. Yitzhaki [14] proposed a generalized Gini coefficient

$$G(\nu) = 1 - \nu(1 - \nu) \int_0^1 (1-p)^{\nu-2} L(p) dp, \quad (16)$$

where $\nu > 1$. Different ν 's are used in order to identify different inequality properties. For low ν 's greater weights are associated with the rich and for high ν 's greater weights are associated with the poor.

Using the mean income (μ) and the Gini coefficient (G), Sen [15] proposed a welfare index

$$W = \mu(1 - G). \quad (17)$$

Pietra index. The Pietra index P is defined as the maximum $D_{\max} = F(\mu_x) - L_x(F(\mu_x))$, presented above. It can be graphically represented as the longest vertical distance between the diagonal and the Lorenz curve, or the cumulative portion of the total income held below a certain income percentile, with the 45 degree line representing perfect equality. The definitions yield Pietra coefficients satisfying the inequality $0 \leq P < 1$. The lower bound of P is obtained when there is total income equality, that is, the Lorenz curve coincides with the diagonal. The upper bound can be obtained when the Lorenz curve converges towards the lower right corner. The limits in the inequalities can be obtained, and this is outlined in **Figure 4**. The Pietra index can be interpreted as the income of the rich that should be redistributed to the poor in order to obtain total income equality. In other words, the value of the index approximates the share of total income that must be transferred from households above the mean to those below the mean to achieve equality in the distribution of incomes. Higher values of P indicate more inequality, and more redistribution is needed to achieve income equality. Therefore, the index is sometimes named the Robin Hood index. The Pietra index is also known as the Hoover index and it is still better known as the Schutz index [16] [17] [18].

If $P_x < P_y$, then the distribution $F_x(x)$ measured by the Pietra index has lower inequality than the distribution $F_y(y)$, and we say that $F_x(x)$ *Pietra dominates* $F_y(y)$. We denote this relation $F_x(x) \succ_P F_y(y)$. For the Lorenz curves in **Figure 3**, $L_1(p)$ is more equal than $L_2(p)$. In general, the Pietra and the Gini orderings are not identical [1]. However, one observes the similar

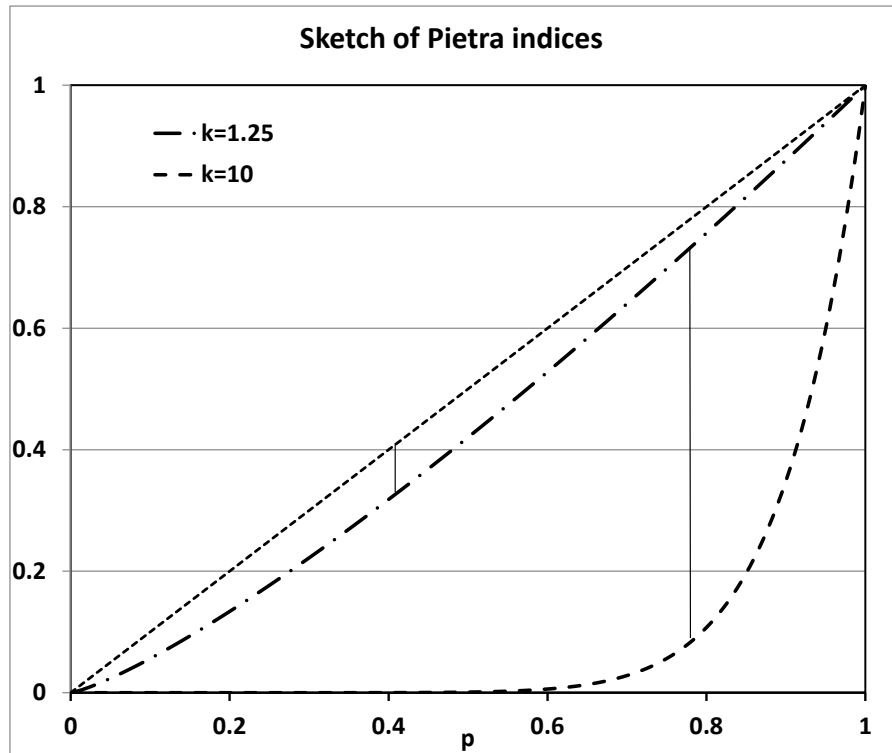


Figure 4. Sketches of two extreme cases of simplified Rao-Tam Lorenz curves with corresponding P indices. For the Lorenz curve $k = 1.25$, the Pietra index is 0.0819 and for the Lorenz curve $k = 10$ the index is 0.697 [1].

obvious result $F_x(x) \succ_L F_y(y) \Rightarrow F_x(x) \succ_P F_y(y)$, that is, Lorenz dominance implies Pietra dominance.

An alternative definition of the Pietra index has also been given. It can be defined as twice the area of the largest triangle inscribed in the area between the Lorenz curve and the diagonal line [9]. In **Figure 5**, one observes that the triangle obtains its maximum when the corner lies on the Lorenz curve where the tangent is parallel to the diagonal. The height of the triangle is $h = \frac{P}{\sqrt{2}}$, and the base is the diagonal $b = \sqrt{2}$. The double of the area is $2\text{area} = 2 \cdot \frac{1}{2} \frac{P}{\sqrt{2}} \sqrt{2} = P$.

In comparison, the Gini index, G , is twice the area between the Lorenz curve and the diagonal, and the Pietra index is twice the area of the triangle inscribed in this area. Hence, the inequality $G \geq P$ holds generally [1].

3. Applications

In this section, we collect some examples in order to elucidate the theory. The models Pareto [19], the simplified Rao-Tam [20] and the Chotikapanich [21] contain only one parameter. Therefore, they can easily be analyzed. Rohde [22] and Fellman [2] [13] paid these models special attention and examined them in more detail. However, they are so simple that it is impossible to distinguish between the estimated length of the range for the income distribution function and

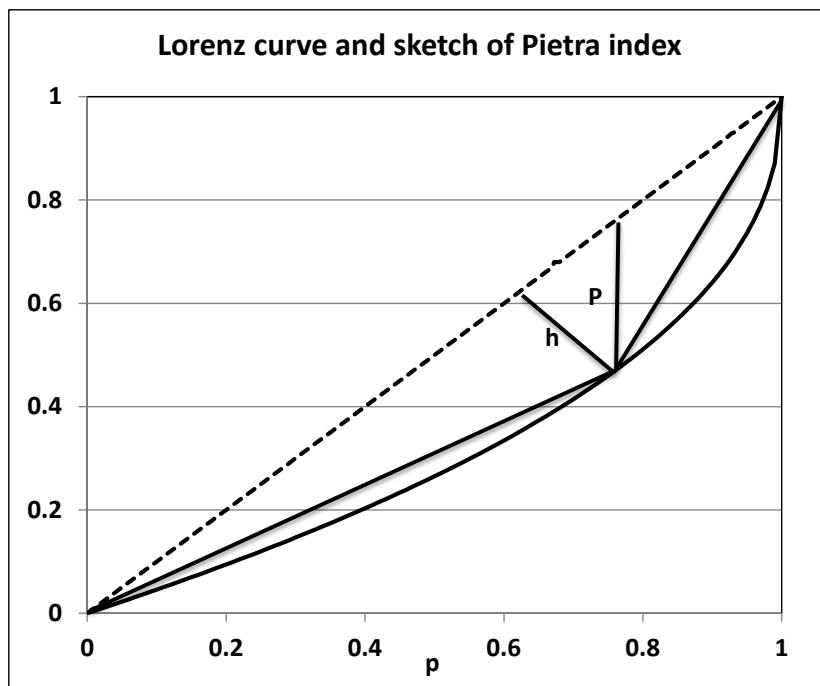


Figure 5. The Lorenz curve and the geometric interpretations of the Pietra index.

the Gini coefficient. If one of these properties is estimated, the other is fixed. We consider these three theoretical Lorenz curve models. We present how the Lorenz curves and the Gini and the Pietra indices depend on the model parameters. In addition, we compare the Lorenz curves of the models when their Gini indices are equal.

Pareto model. We define the Pareto distribution as $F(x) = 1 - x^{-\alpha}$, where $x \geq 1$ and $\alpha > 1$.

The frequency function is $f(x) = \alpha x^{-\alpha-1}$, the mean is $\mu = \frac{\alpha}{\alpha-1}$, the quantiles are $x_p = \left(\frac{1}{1-p}\right)^{\frac{1}{\alpha}}$, the Lorenz curve $L(p) = 1 - (1-p)^{\frac{\alpha-1}{\alpha}}$ and the Gini coefficient $G = \frac{1}{2\alpha-1}$. If $\alpha \rightarrow 1$, $G \rightarrow 1$ and if $\alpha \rightarrow \infty$, $G \rightarrow 0$. In Figure 6, we present the Pareto distribution as a function of the parameter α .

Finally, the Pietra index is $P = \frac{1}{\alpha} \left(\frac{\alpha-1}{\alpha}\right)^{\alpha-1}$. According to the general theory, the inequality $G \geq P$ holds for all parameter values, and consequently, $P \rightarrow 0$ when $\alpha \rightarrow \infty$. Let $\beta = \alpha - 1$, then $P = \frac{1}{\beta+1} \left(\frac{\beta}{\beta+1}\right)^\beta$. When $\beta \rightarrow 0$, ($\alpha \rightarrow 1$), then, $P \rightarrow 1$. In Figure 7, we compare the Gini and Pietra indices as functions of the model parameter α . One observes that the inequality $G \geq P$ holds.

Simplified Rao-Tam model. Consider the simplified Rao-Tam model whose Lorenz curve is $L(p) = p^\alpha$ ($\alpha > 1$) [20]. When $\alpha \rightarrow 1$, then $L(p) \rightarrow p$ and

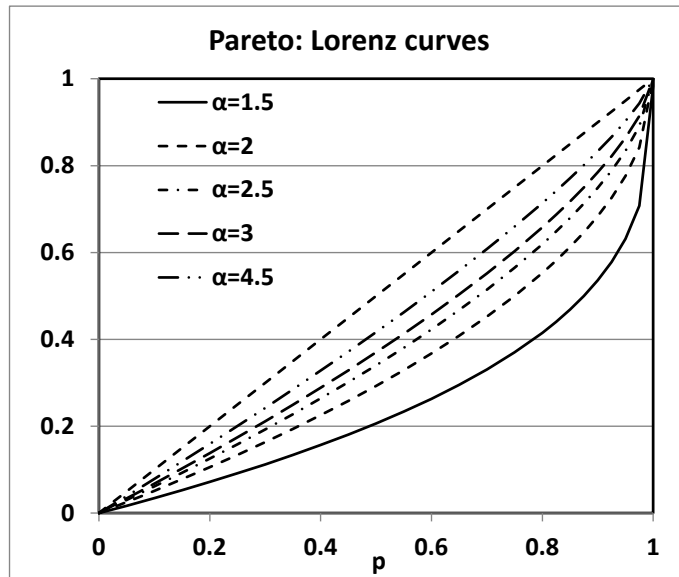


Figure 6. Lorenz curves for Pareto distributions with different parameter values.

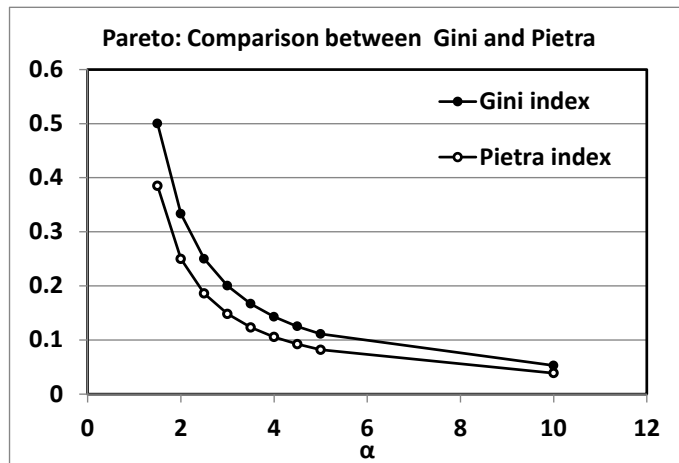


Figure 7. Gini and Pietra indices for Pareto distributions with different parameter values.

when $\alpha \rightarrow \infty$ $L(p) \rightarrow 0$ for all $0 \leq p < 1$ and the Lorenz curve converges, towards the lower right corner of the square. In **Figure 8**, we present the Lorenz curve for a set of α values.

The Gini coefficient is $G = \frac{\alpha - 1}{\alpha + 1}$. When $\alpha \rightarrow 1$, then $G \rightarrow 0$ and when

$\alpha \rightarrow \infty$ then $G \rightarrow 1$. The Pietra index is $P = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}$. Using the ver-

tical difference $D = p - p^\alpha$, the index inequalities $D \leq P < G$ hold, and for $\alpha \rightarrow 1$ $G \rightarrow 0$, and consequently, $P \rightarrow 0$. For increasing α values, the supremum of $D = p - p^\alpha$ is one. This must also hold for the supremum of

$P = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}}$. Consequently, the interval $0 < P < 1$ cannot be short-

ened. In **Figure 9**, we present G and P for different α .

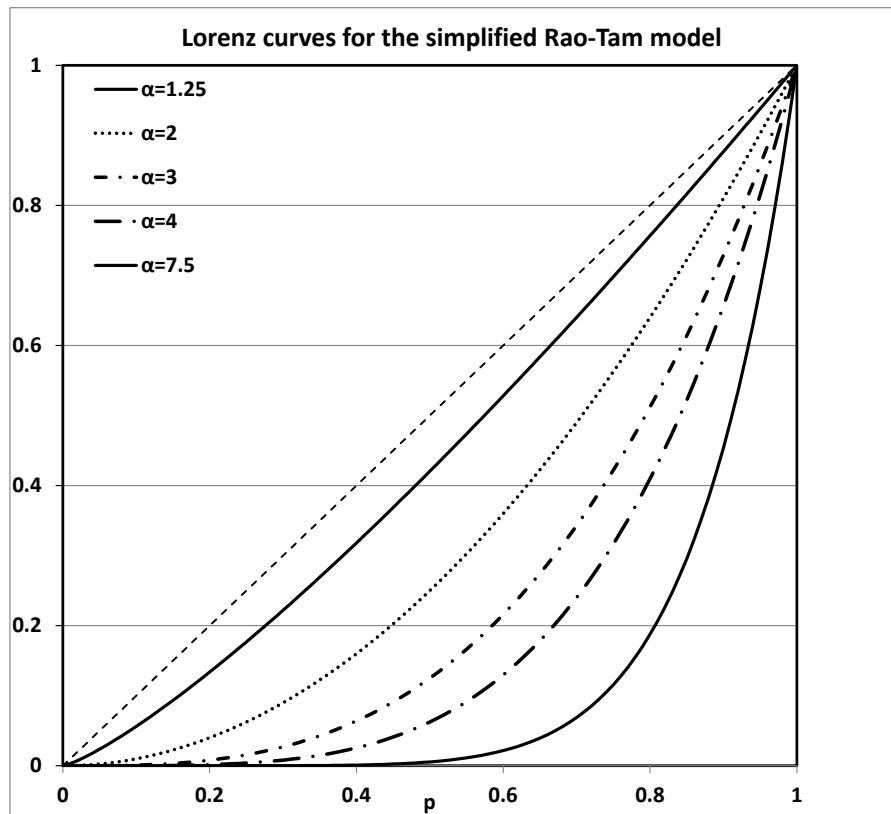


Figure 8. Lorenz curves for simplified Rao-Tam distributions with different parameter values.

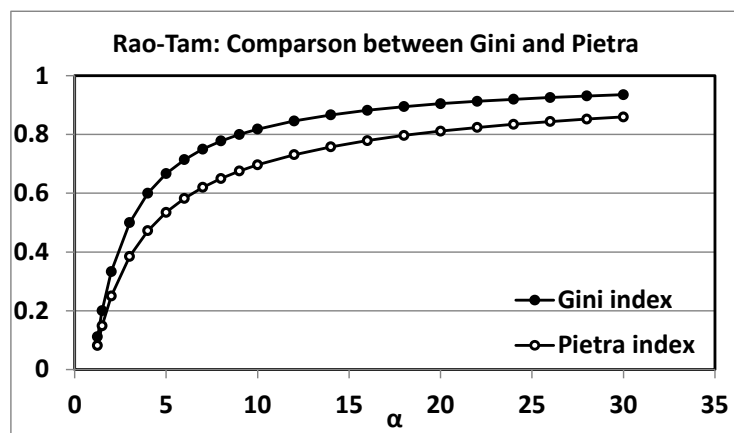


Figure 9. Gini and Pietra indices for simplified Rao-Tam distributions with different parameter values.

Chotikapanich. Chotikapanich [21] defined the Lorenz curve

$$L_C(p) = \frac{e^{kp} - 1}{e^k - 1} \text{ for } k > 0.$$

The limits of the fractions studied below result in indefinite forms $\frac{0}{0}$ and

l'Hospital's rule $\lim_{k \rightarrow a} \frac{f(x)}{g(x)} = \lim_{k \rightarrow a} \frac{f'(x)}{g'(x)}$ has to be applied several times. For

$k \rightarrow 0$ and arbitrary $0 < p < 1$, we obtain $\lim_{k \rightarrow 0} \frac{e^{kp} - 1}{e^k - 1} = \lim_{k \rightarrow 0} \frac{e^{kp} p}{e^k} = p$. This means that the Lorenz curve converges towards the diagonal. For $k \rightarrow \infty$, one obtains that for all $0 < p < 1$

$$\lim_{k \rightarrow \infty} \frac{e^{kp} - 1}{e^k - 1} = \lim_{k \rightarrow \infty} \frac{e^{kp} p}{e^k} = \lim_{k \rightarrow \infty} \frac{e^{kp} p}{e^k} = \lim_{k \rightarrow \infty} p e^{-k(1-p)} = 0.$$

This means that the Lorenz curve converges towards the lower right corner of the square.

The extreme Lorenz curves can be obtained by the limit studies $k \rightarrow 0$ and $k \rightarrow \infty$, and the Lorenz curves as functions of the parameter k are presented in **Figure 10**.

The Gini index is for the Chotikapanich model

$$G = 1 - 2 \left(\frac{\frac{1}{k} e^k - 1}{e^k - 1} - \frac{1}{k(e^k - 1)} \right) = 1 - 2 \left(\frac{e^k - k - 1}{k(e^k - 1)} \right).$$

When $k \rightarrow 0$, one obtains

$$\begin{aligned} \lim_{k \rightarrow 0} G &= \lim_{k \rightarrow 0} \left(1 - 2 \left(\frac{e^k - k - 1}{k(e^k - 1)} \right) \right) = 1 - 2 \lim_{k \rightarrow 0} \left(\frac{e^k - 1}{(e^k - 1) + k e^k} \right) \\ &= 1 - 2 \lim_{k \rightarrow 0} \left(\frac{e^k}{e^k + k e^k + e^k} \right) = 1 - 2 \frac{1}{2} = 0 \end{aligned}$$

When $k \rightarrow \infty$, one obtains

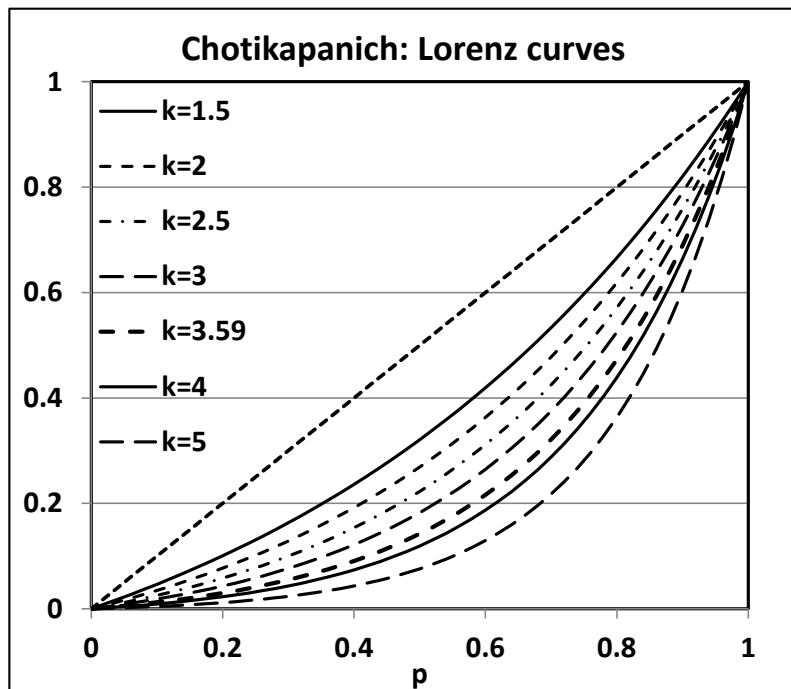


Figure 10. Lorenz curves for Chotikapanich distributions with different parameter values.

$$\begin{aligned} \lim_{k \rightarrow \infty} G &= \lim_{k \rightarrow \infty} \left(1 - 2 \left(\frac{e^k - k - 1}{k(e^k - 1)} \right) \right) = 1 - 2 \lim_{k \rightarrow \infty} \left(\frac{e^k - 1}{(e^k - 1) + ke^k} \right) \\ &= 1 - 2 \lim_{k \rightarrow \infty} \left(\frac{e^k}{e^k + ke^k + e^k} \right) = 1 - 2 \lim_{k \rightarrow \infty} \left(\frac{1}{2 + k} \right) = 1 - 0 = 1 \end{aligned}$$

Consequently, for G , the inequalities $0 < G < 1$ hold and the range cannot be shortened.

The Pietra index is

$$\begin{aligned} P &= \frac{1}{k} \ln \left(\frac{e^k - 1}{k} \right) - \frac{1}{k} + \frac{1}{e^k - 1} = \frac{1}{k} \ln \left(\frac{e^k - 1}{k} \right) - \frac{e^k - 1 - k}{k(e^k - 1)} \\ &= \frac{1}{k} k + \frac{1}{k} \ln \frac{(1 - e^{-k})}{k} - \frac{e^k - 1 - k}{k(e^k - 1)} = 1 + \frac{1}{k} \ln \frac{(1 - e^{-k})}{k} - \frac{e^k - 1 - k}{k(e^k - 1)} \end{aligned}$$

In general, $P < G$, and consequently, $P \rightarrow 0$ when $k \rightarrow 0$. When $k \rightarrow \infty$, one obtains

$$\begin{aligned} &\lim_{k \rightarrow \infty} \left(1 + \frac{1}{k} \ln \frac{(1 - e^{-k})}{k} - \frac{e^k - 1 - k}{k(e^k - 1)} \right) \\ &= \lim_{k \rightarrow \infty} \left(1 + \frac{1}{k} \ln \frac{(1 - e^{-k})}{k} \right) - \lim_{k \rightarrow \infty} \left(\frac{e^k - 1}{(e^k - 1) + ke^k} \right) \\ &= \lim_{k \rightarrow \infty} \left(1 - \frac{\ln(k)}{k} + \frac{1}{k} \ln(1 - e^{-k}) \right) - 0 = 1 \end{aligned}$$

Hence, $\lim_{k \rightarrow \infty} P = 1$ and the inequalities $0 < P < 1$ hold and cannot be shortened.

The G and P as functions of the parameter k indices are presented in **Figure 11**.

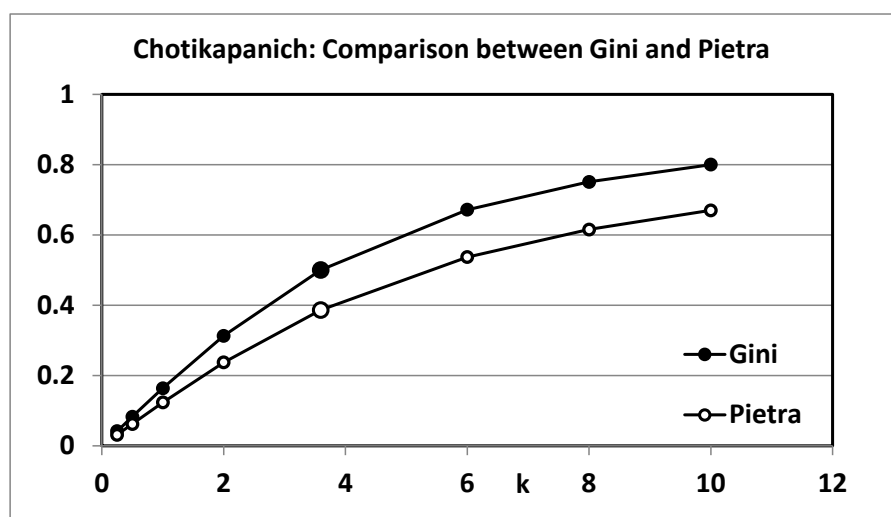


Figure 11. Gini and Pietra indices for Chotikapanich distributions with different parameter values.

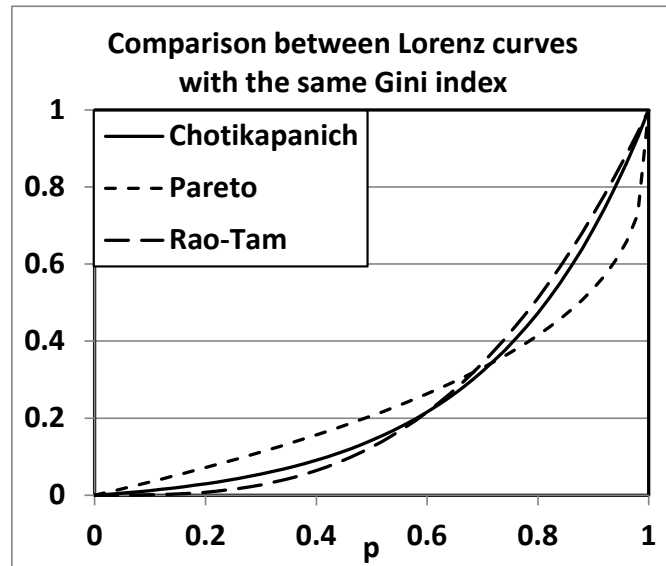


Figure 12. Lorenz curves for distributions with the same Gini index ($G = 0.5$). Note that the Lorenz curves for the simplified Rao-Tam and Chotikapanich models are rather similar, but the Lorenz curve for the Pareto model is markedly different.

Above, we made the general remark that different distributions can result in the same Gini index. In **Figure 12**, we present a simple example of this finding. We compare a Chotikapanich model with the Gini index 0.500 ($k = 3.593525$) with a Pareto model (with $\alpha = 1.5$) and a simplified Rao-Tam distribution (with $\alpha = 3.0$), all having the same Gini index. The Lorenz curves for the simplified Rao-Tam and Chotikapanich models are rather similar, but the Lorenz curve for the Pareto model is markedly different. This is caused by the fact that the Pareto model is introduced as a model for distributions with high income levels.

4. Discussion

In general, the step from the Lorenz curve to the income distribution starts from the formula

$$L'(p) = \frac{x_p}{\mu}, \tag{18}$$

where x_p is the p -percentile and μ is the mean of the corresponding distribution $F(x)$. We define $M(\cdot)$ as the inverse function of the derivative $L'(\cdot)$. From (18), we obtain

$$p = M\left(\frac{x_p}{\mu}\right). \tag{19}$$

Equation (19) indicates that $M(\cdot)$ is the income distribution function corresponding to the given Lorenz curve, that is, $F(x) = M\left(\frac{x}{\mu}\right)$. This connection between the Lorenz curve and the distribution function is easily defined, but for

most of the exact Lorenz curves, it is difficult or even impossible to mathematically obtain the distribution.

Primary income data yield the most exact estimates of the income inequality coefficients, such as Gini and Pietra. Fellman [2] analyzed different methods for numerical estimation of Gini coefficients based on Lorenz curves. As an application of these methods, he considered Pareto distributions. Using Lorenz curves, various numerical integration attempts were applied to obtain accurate estimates. The trapezium rule is simple, but yields a positive bias for the area under the Lorenz curve, and consequently, a negative bias for the Gini coefficient. Simpson's rule is better fitted to the Lorenz curve, but this rule demands an even number of subintervals of the same length. Lagrange polynomials of second degree can be considered as a generalization of Simpson's rule. Fellman [2] compared different methods and he also gave references concerning numerical integration. To include Simpson's rule in his study, he considered Lorenz curves with deciles. Compared with Simpson's rule, he used the trapezium rule, Lagrange polynomials and generalizations of Golden's method [23]. No method was uniformly optimal, but the trapezium rule was almost always inferior and Simpson's rule was superior. Golden's method is usually of medium quality. Mettle *et al.* [3] estimated the Lorenz curve by Newton-Cotes methods, namely, the trapezium rule, Simpson's 1/3 rule and Simpson's 3/8 rule. Using these, they estimated the Gini coefficients of income and compared the accuracy of these estimates on some (Ghanaian) data. The curves in **Figure 12** in this study indicate weaknesses in the inequality indices. Lorenz curves with the same Gini index may show marked geometrical differences and consequently different income distributions.

Acknowledgements

This work was supported in part by a grant from the Magnus Ehrnrooth Foundation.

References

- [1] Fellman, J. (2015) *Mathematical Analysis of Distribution and Redistribution of Income*. Science Publishing Group, 166 p. <http://www.sciencepublishinggroup.com/book/B-978-1-940366-25-8.aspx>
- [2] Fellman, J. (2012) Estimation of Gini Coefficients Using Lorenz Curves. *Journal of Statistical and Econometric Methods*, **1**, 31-38.
- [3] Mettle, F.O., Darkwah, K.A., Nortey, E.N. and Lotsi, C.A. (2016) An Estimation of the Gini Coefficient of Income Using Newton-Cotes Methods. Conference Paper February 2016.
- [4] Lorenz, M.O. (1905) Methods for Measuring Concentration of Wealth. *Publ. of the Amer. Statist. Assoc.*, **9**, 209-219.
- [5] Fellman, J. (2011) Lorenz Curve. In: Lovric, M., Ed., *International Encyclopedia of Statistical Science*, Part 12, 760-762. https://doi.org/10.1007/978-3-642-04898-2_345
- [6] Wang, Z.X., Ng, Y.-K. and Smyth, R. (2011) A General Method for Creating Lorenz

- Curves. *Review of Income and Wealth*, **57**, 561-582.
<https://doi.org/10.1111/j.1475-4991.2010.00425.x>
- [7] Fellman, J. (1976) The Effect of Transformations on Lorenz Curves. *Econometrica*, **44**, 823-824. <https://doi.org/10.2307/1913450>
- [8] Fellman, J. (1980) Transformations and Lorenz Curves. Swedish School of Economics and Business Administration Working Papers, 48, 18 p.
- [9] Lee, W.-C. (1999) Probabilistic Analysis of Global Performances of Diagnostic Tests: Interpreting the Lorenz Curve Based Summary Measures. *Statistics in Medicine*, **18**, 455-471.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19990228\)18:4<455::AID-SIM44>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0258(19990228)18:4<455::AID-SIM44>3.0.CO;2-A)
- [10] Kleiber, Ch. and Kotz, S. (2001) Characterizations of Income Distributions and the Moment Problem of Order Statistics. *The 53rd Session of the International Statistical Institute*, Seoul, Korea, 22-29 August 2001.
<http://isi.cbs.nl/iamamember/CD2/pdf/462.pdf>
- [11] Kleiber, Ch. and Kotz, S. (2002) A Characterization of Income Distributions in Terms of Generalized Gini Coefficients. *Social Choice and Welfare*, **19**, 789-794.
<https://doi.org/10.1007/s003550200154>
- [12] Gini, C. (1914) Sulla misura della concentrazione e della variabilita dei caratteri. *Premiate officine grafiche C. Ferrari*, 73c, 1203-1248.
- [13] Fellman, J. (2012) Modelling Lorenz Curve. *Journal of Statistical and Econometric Methods*, **1**, 53-62.
- [14] Yitzhaki, S. (1983) On an Extension of the Gini Index. *International Economic Review*, **24**, 617-628. <https://doi.org/10.2307/2648789>
- [15] Sen, A. (1973) *On Economic Inequality*. Clarendon Press, Oxford.
<https://doi.org/10.1093/0198281935.001.0001>
- [16] Hoover, E. (1936) The Measurement of Industrial Localization. *The Review of Economics and Statistics*, **18**, 162-171. <https://doi.org/10.2307/1927875>
- [17] Schutz, R.R. (1951) On the Measurement of Income Inequality. *American Economic Review*, **41**, 107-122.
- [18] Atkinson, A.B. (1970) On the Measurement of Inequality. *Journal of Economic Theory*, **2**, 244-263. [https://doi.org/10.1016/0022-0531\(70\)90039-6](https://doi.org/10.1016/0022-0531(70)90039-6)
- [19] Rasche, R.H., Gaffney, J., Koo, A.Y.C. and Obst, N. (1980) Functional Forms for Estimating the Lorenz Curve. *Econometrica*, **48**, 1061-1062.
<https://doi.org/10.2307/1912948>
- [20] Rao, U.L.G. and Tam, A.Y.-P. (1987) An Empirical Study of Selection and Estimation of Alternative Models of the Lorenz Curve. *Journal of Applied Statistics*, **14**, 275-280. <https://doi.org/10.1080/02664768700000032>
- [21] 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](https://doi.org/10.1016/0165-1765(93)90186-G)
- [22] Rohde, N. (2009) An Alternative Functional Form for Estimating the Lorenz Curve. *Economics Letters*, **105**, 61-63. <https://doi.org/10.1016/j.econlet.2009.05.015>
- [23] Golden, J. (2008) A Simple Geometric Approach to Approximating the Gini Coefficient. *The Journal of Economic Education*, **39**, 68-77.
<https://doi.org/10.3200/JECE.39.1.68-77>