



Estimating Income Inequality Using Single-Parameter Lorenz Curves: A New Proposal

José María Sarabia¹ · Vanesa Jordá¹ · Mercedes Tejería¹

Received: 18 October 2024 / Accepted: 24 April 2025 / Published online: 22 May 2025
© The Author(s) 2025

Abstract

In a recent paper, Paul and Shankar (2020) introduced a single-parameter Lorenz curve that provides an improved fit compared to many existing uniparametric models. This paper explores new properties of their model, offering a refined representation in terms of convex linear combinations of Lorenz curves. We also derive closed-form expressions for several inequality measures and examine the Lorenz ordering. However, we identify a key limitation: The Gini index for this curve is lower bounded at 0.418, making the model unsuitable for income distributions with lower inequality. To address this issue, we propose an alternative model that extends the range of the Gini index, allowing for greater flexibility in representing income distributions across a wider range of inequality levels. Our results suggest that the Lorenz curve proposed in this paper surpasses the proposal by Paul and Shankar, even in countries with high inequality, where the constraint imposed by the Gini index is not binding.

Keywords Gini index · functional form · Lorenz ordering · inequality

Mathematics Subject Classification C80 · D31

1 Introduction

Parametric models are valuable tools for the empirical analysis of income and wealth distributions. Previous research suggests that parametric functional forms provide more accurate estimates of inequality measures than nonparametric methods, especially when data are limited to a few income shares (Dhongde and Minoiu 2013; Jorda et al. 2021). This superiority of parametric models appears to hold even for simple

✉ José María Sarabia
sarabiaj@uncan.es

Vanesa Jordá
vanesa.jorda@uncan.es

Mercedes Tejería
mercedes.tejeria@uncan.es

¹ Department of Economics, University of Cantabria, Santander, Spain

functional forms with only two parameters. As a result, parametric models become essential when individual records are unavailable, and only aggregated income distribution data are accessible.

The unavailability of individual data was particularly common in the 1980s and 1990s, and it remains a significant challenge today. For instance, when analyzing the distribution of income in Bhutan, one must rely on grouped data, as individual records are not accessible to the general public. In fact, grouped data are often more readily available than individual records. Several comprehensive secondary databases have been developed to provide access to aggregated income distribution data, including the World Income Inequality Database (WIID) hosted by the World Institute for Development Economics Research, the Poverty and Inequality Platform (PIP) managed by the World Bank, and the World Inequality Database (WID) operated by the Paris School of Economics. While these resources are invaluable, they typically rely on grouped data, making parametric models essential for an accurate analysis of income distributions.

The challenge of identifying an adequate functional form for income distributions remains unresolved, with several alternative specifications proposed.¹ Among these, the log-normal distribution is likely the most widely used model. However, despite its popularity, the log-normal model has notable limitations and is often outperformed by alternative functional forms, including those with just two parameters (Bandourian et al. 2002; Jorda et al. 2021). Consequently, scholars have devoted significant attention to finding a functional form that better fits observed distributions. Among the range of alternatives, the generalized beta family of distributions stands out as the most promising option (Kleiber and Kotz 2003).²

Another branch of the literature on statistical distributions has focused on developing *ad hoc* Lorenz curves. Since the pioneering work of Kakwani and Podder (1973), directly modeling the Lorenz curve has become a common practice, as it provides an excellent fit to observed income shares while maintaining a straightforward estimation process. The World Bank has a long-standing tradition of using such models to estimate income distributions from grouped data. Notably, Kakwani's (1980) beta Lorenz curve and Villaseñor and Arnold's (1989) quadratic Lorenz curve are both integrated into the PIP methodology to derive poverty estimates. Alternative models with fewer parameters, such as those proposed by Aggarwal (1984), Chotikapanich (1993), Kakwani and Podder (1973), Gómez-Déniz (2016), and the Lamé Lorenz curves (Henle et al. 2008; Sarabia et al. 2017), also provide good approximations of the Lorenz curve. Developing simpler models remains crucial for ensuring ease of implementation while retaining accuracy, particularly when working with grouped data.

Paul and Shankar (2020) introduced a one-parameter model that appears to provide an excellent fit for Australian data (hereafter referred to as the PS Lorenz curve). In this paper, we explore new properties of their model, offering a new representation that enables us to define the Lorenz ordering and derive closed-form expressions for various inequality measures. Additionally, we demonstrate that, despite the model's strong apparent performance, it is only suitable for representing income distributions

¹ See Ryu and Slotte (1996) and Sarabia Sarabia (2008) for a detailed discussion on selecting a parametric model.

² For an extensive review of statistical models developed since the late nineteenth century, see Kleiber and Kotz (2003).

with high levels of inequality, as the Gini index is bounded below by 0.418. To address this limitation, we propose a new model that, while still bounded, has a Gini index range starting at 0.164, allowing for a more flexible representation of income distributions across a broader range of inequality levels.³ We also compare the performance of our model with the PS Lorenz curve across various countries exhibiting different levels of inequality. Our proposed model demonstrates superior flexibility and accuracy, even for countries with high inequality, where the constraint imposed by the Gini index of the PS model is not binding.

The contents of this paper are structured as follows. In Section 2, we explore a new representation of the PS Lorenz curve and introduce closed-form expressions for several inequality measures. Section 3 introduces an alternative Lorenz curve model, which offers a much wider range for the Gini index. Additionally, this new model is simpler and provides closed-form expressions for both the quantile function and the cumulative distribution function (CDF), making it easier to compute several economic quantities and improving the accuracy of inequality estimates, particularly in cases where the PS Lorenz curve is limited by its bounded Gini index. In Section 4, we analyze the performance of our new model compared to the PS Lorenz curve across various countries. The paper concludes by discussing the practical and theoretical implications of our findings.

2 New properties of the PS Lorenz curve

In this section, we introduce a new representation of the PS Lorenz curve, which simplifies the definition of Lorenz ordering. We also present closed-form expressions for several inequality indices, including the Gini index and the Donaldson, Weymark, and Kakwani (DWK) index.

2.1 A new representation of the PS Lorenz curve

The PS Lorenz curve is defined as (Paul and Shankar 2020):

$$L(p; \gamma) = p \left[\frac{e^{-\gamma(1-e^p)} - 1}{e^{-\gamma(1-e)} - 1} \right], \quad 0 \leq p \leq 1, \quad (1)$$

where $\gamma > 0$. The limit case of (1) when γ goes to zero is,

$$\lim_{\gamma \rightarrow 0} L(p; \gamma) = p \frac{e^p - 1}{e - 1},$$

which is also a Lorenz curve, as it is the product of the egalitarian Lorenz curve and the model proposed by Chotikapanich (1993) when the parameter value is equal to 1.

³ Only 18 out of 24,366 datasets available in the WIID report Gini indices lower than 0.164. Therefore, despite the lower bound of the Gini index, the new curve remains an excellent candidate for representing Lorenz curves in the vast majority of countries.

Table 1 Weight values for different γ parameters

$\gamma = 0.228$		$\gamma = 0.187$		$\gamma = 0.107$	
w_k	cumulative	w_k	cumulative	w_k	cumulative
0.81687	0.81687	0.84793	0.84793	0.91089	0.91089
0.16001	0.97689	0.13623	0.98416	0.08374	0.99462
0.02090	0.99778	0.01459	0.99875	0.00513	0.99975
0.00205	0.99983	0.00117	0.99992	0.00024	0.99999
0.00016	0.99999	0.00008	1.00000	0.00001	1.00000

We present the following representation of the PS Lorenz curve, which simplifies the derivation of analytical expressions for the Gini index and other relevant inequality measures, as well as the Lorenz ordering.

Theorem 1 *The PS Lorenz curve defined in (1) can be written as an (infinite) convex linear combination of LCs of the form,*

$$L(p; \gamma) = \sum_{k=1}^{\infty} w_k L_k(p; \gamma), \quad (2)$$

where

$$L_k(p) = p \left(\frac{e^p - 1}{e - 1} \right)^k, \quad k = 1, 2, \dots \quad (3)$$

are genuine LC, with weights

$$w_k = \frac{\gamma^k (e - 1)^k}{(e^{\gamma(e-1)} - 1)k!}, \quad k = 1, 2, \dots \quad (4)$$

Note that $0 < w_k < 1$, $k = 1, 2, \dots$ and $\sum_{k=1}^{\infty} w_k = 1$. The proof is provided in Appendix.

This representation as a convex linear combination allows us to make several important observations. The weights w_k decrease as k increases, and the components of the linear combination $L_k(p)$ in equation (3) represent a special case of the class of Lorenz curves defined by Sarabia et al. (2001). Table 1 presents the values of these weights for three selected values of the γ parameter, based on the results of Paul and Shankar (2020). A notable finding is that a significant portion of the overall Lorenz curve might be captured with just two of these components.

2.2 Lorenz ordering

The Lorenz ordering is a relevant aspect in the analysis of income and wealth distributions. Let $\mathcal{L}$ be the class of all non-negative random variables with positive and finite

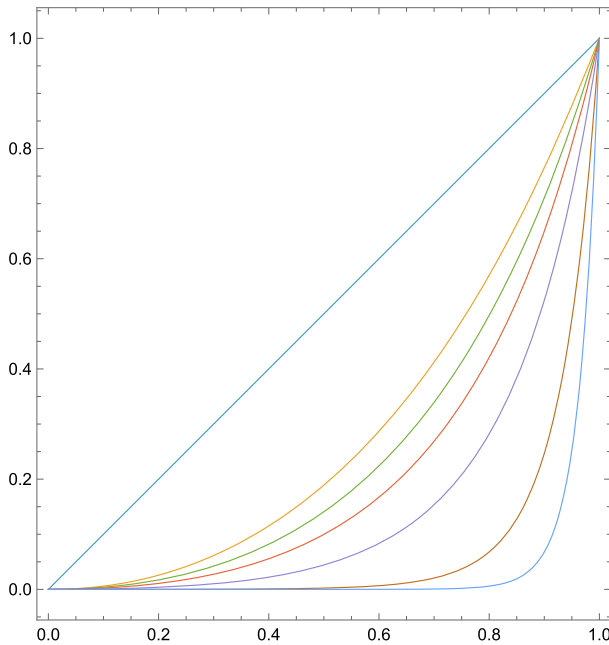


Fig. 1 PS Lorenz curve (1) for $\gamma=0, 0.5, 1, 2, 5$, and 10 (left to right).

mathematical expectations. The Lorenz partial ordering $\leq_L$ on the class $\mathcal{L}$ is defined as:

$$X \leq_L Y \iff L_X(p) \geq L_Y(p), \quad \forall p \in [0, 1],$$

where X and Y are random variables in $\mathcal{L}$. If $X \leq_L Y$, then X exhibits less inequality than Y in the Lorenz sense. We will now show that the family of Lorenz curves in (1) is ordered with respect to the γ parameter.

Theorem 2 *If $L(p; \gamma)$ is defined in equation (1) and $\gamma_1 \leq \gamma_2$, then $L(p; \gamma_1) \geq L(p; \gamma_2)$, for $0 \leq p \leq 1$.*

The proof of Theorem 2 is presented in Appendix. The Lorenz ordering is illustrated in Figure 1, which shows the PS Lorenz curves for $\gamma = 0, 0.5, 1, 2, 5$, and 10 . The Lorenz curves do not intersect, indicating that the models are ordered according to the γ parameter.

2.3 Inequality measures

In this section, we obtain analytical expressions for relevant inequality measures. For the Gini index, we have the next result

Lemma 1 *The Gini index of the LC (1) is given by*

$$G_L = 1 - 2 \sum_{k=1}^{\infty} \frac{w_k}{(e-1)^k} A_k, \quad (5)$$

where $w_k, k = 1, 2, \dots$ are given in (4) and A_k are defined as

$$A_k = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{1 + e^{k-j}(k-j-1)}{(k-j)^2}. \quad (6)$$

Proof See Appendix. □

As a consequence of Theorem 2, we demonstrate that the Gini index of the curve given in (1) is bounded from below, as stated in the following lemma.

Lemma 2 *The Gini index of (1) is lower bounded. Specifically, if $\gamma > 0$, then $G_\gamma > \frac{e-2}{e-1} = 0.418$, where G_γ denotes the Gini index of the family (1). The lower bound of the Gini index is reached in the limit case $\gamma = 0$.*

The proof of Lemma 2 is presented in Appendix. An important consequence of this lemma is that, since the Gini index for the PS Lorenz curve is always greater than 0.418, this model can only be applied to countries whose empirical Gini index exceeds this threshold. This limitation must be taken into account when selecting the model for cross-country comparisons or when applying it to countries with less pronounced income inequality.

An important generalization of the Gini index was proposed by Donaldson and Weymark (1980) and Yitzhaki (1983) and was studied by Kakwani (1980a). This generalization is given by

$$G_X(\nu) = 1 - \nu(\nu+1) \int_0^1 (1-p)^{\nu-1} L_X(p) dp, \quad (7)$$

where $\nu \geq 1$ and $L_X(p)$ is the LC. If we set $\nu = 1$ in equation (7), we obtain the usual Gini index. As ν increases, more weight is attached to the lower tail of the distribution. In the limit when ν goes to infinity, the index only considers the minimum income, which is congruent with the Rawlsian criterion, and the judgment that social welfare depends only on the poorest society member. The next result provides the DKW index for the LC (1).

Lemma 3 *The DKW index for the LC (1) is given by*

$$G_X(\nu) = 1 - \nu(\nu+1) \sum_{k=1}^{\infty} \frac{w_k}{(e-1)^k} A_k(\nu), \quad (8)$$

where $w_k, k = 1, 2, \dots$ are given in (4) and $A_k(v)$ are defined by

$$A_k(v) = \sum_{j=0}^k \frac{(-1)^j}{(k-j)^2} \binom{k}{j} \left\{ (k-j)^{v+1} + e^{k-j} (k-j-v) [\Gamma(v+1) - \Gamma(v+1, k-j)] \right\}, \quad (9)$$

where $\Gamma(a)$ is the usual gamma function and $\Gamma(a, b)$ is the uncomplete gamma function.

Proof See Appendix. $\square$

3 An alternative model for the Lorenz curve

The PS proposal has limitations which deem it unsuitable for distributions with low inequality levels. In this section, we propose a simpler functional form than (1) and with more satisfactory properties. The new Lorenz curve is defined as follows:

$$\tilde{L}(p; \gamma) = \frac{e^{\gamma(e^p-1)} - 1}{e^{\gamma(e-1)} - 1}, \quad 0 \leq p \leq 1, \quad (10)$$

with $\gamma > 0$ (hereafter referred to as the SJT Lorenz curve). Relevant properties of this new family are obtained in the following sections.

3.1 Lorenz ordering and the Gini index

As a first property, the curve (10) is also ordered in the Lorenz sense based on the γ parameter.

Theorem 3 If $\tilde{L}(p; \gamma)$ is defined in equation (10) and $0 < \gamma_1 \leq \gamma_2$, then $\tilde{L}(p; \gamma_1) \geq \tilde{L}(p; \gamma_2)$, for $0 \leq p \leq 1$.

The proof of the result is direct as will be omitted. The Gini index of (10) is given by

$$G_{\tilde{L}}(\gamma) = 1 - 2 \frac{(\text{Ei}(e\gamma) - \text{Ei}(\gamma))e^{-\gamma} - 1}{\exp(\gamma(e-1)) - 1}, \quad (11)$$

where $\gamma > 0$ and $\text{Ei}(z)$ denotes the exponential integral, defined by,

$$\text{Ei}(z) = - \int_{-z}^{\infty} \frac{e^{-t}}{t} dt,$$

with $z > 0$. It can also be proved that

$$0.164 = \frac{e-3}{e-1} \leq G_{\tilde{L}}(\gamma) \leq 1.$$

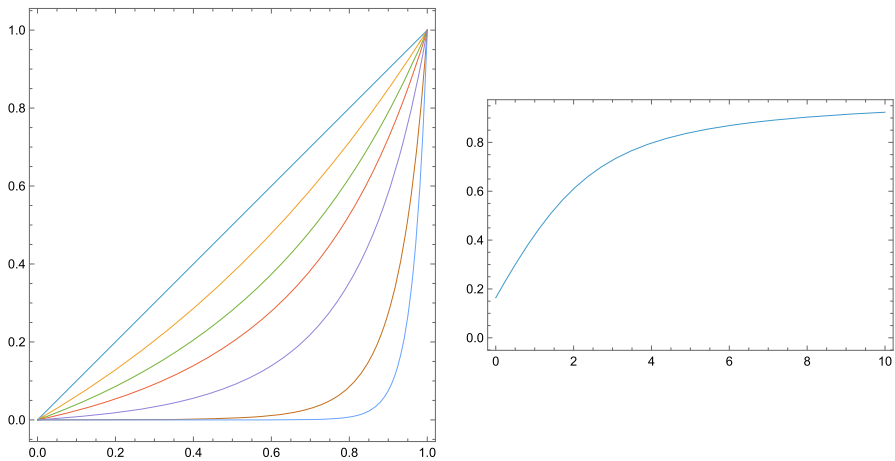


Fig. 2 Lorenz curve of the new proposal (10) for $\gamma=0, 0.5, 1, 2, 5$, and 10 (left) and the corresponding Gini index (right) as a function of the parameter γ .

In consequence, the new proposal sets a lower bound for the Gini index at 0.164. Figure 2 presents the Lorenz curve of (10) in the left panel and the Gini index as a function of γ in the right panel. When $\gamma = 0$, the curve is closer to the egalitarian line, indicating greater flexibility compared to the PS Lorenz curve.

3.2 Quantile and cumulative distribution functions

A potential limitation of using *ad hoc* Lorenz curves to model income distributions is that the underlying CDF is often unavailable, complicating the analysis of specific features such as poverty and absolute inequality. While the PS proposal lacks a closed-form expression for the CDF, the alternative Lorenz curve introduced in this paper provides explicit expressions for both the CDF and the quantile function. Using the formulas in Sarabia (2008) (see also Arnold and Sarabia (2018), chapters 3 and 6), it is straightforward to prove that the quantile function of the new Lorenz curve (10) is:

$$Q(p; \gamma, \mu) = \frac{\mu\gamma \exp[\gamma(e^p - 1) + p]}{\exp[\gamma(e - 1) + 1]}, \quad 0 \leq p \leq 1, \quad (12)$$

where $\gamma > 0$ and $\mu > 0$ is the mean of the random variable.

On the other hand, the CDF associated with the LC (10) is given by (see Appendix):

$$F(x; \gamma, \mu) = \gamma + \log \left[\frac{x}{\mu\gamma} (e^{e\gamma - \gamma} - 1) \right] - W \left[\frac{x}{\mu} (e^{e\gamma} - e^\gamma) \right], \\ a(\gamma, \mu) \leq x \leq b(\gamma, \mu), \quad (13)$$

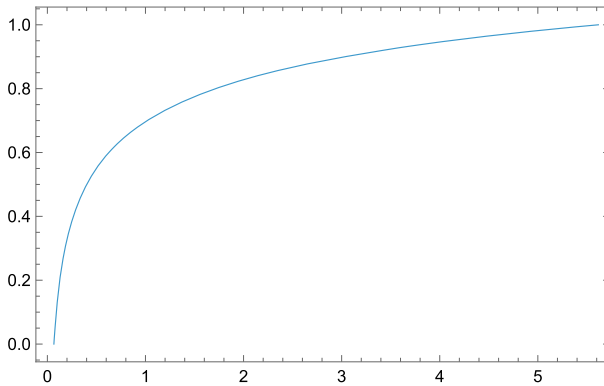


Fig. 3 Cumulative distribution function (13) corresponding to the new Lorenz curve for $\mu = 1$ and $\gamma = 2$.

and $F(x; \gamma, \mu) = 0$ if $x \leq a(\gamma, \mu)$, and $F(x; \gamma, \mu) = 1$ if $x \geq b(\gamma, \mu)$, where

$$a(\gamma, \mu) = \frac{\mu\gamma}{e^{\gamma(e-1)} - 1},$$

and

$$b(\gamma, \mu) = \frac{\mu\gamma e^{\gamma(e-1)+1}}{e^{\gamma(e-1)} - 1},$$

being $W(z)$ the Lambert function, or product logarithm function, which gives the principal solution for w in the equation $z = we^w$.

Figure 3 shows the cumulative distribution function (13) for $\mu = 1$ and $\gamma = 2$.

4 Performance comparison of Lorenz curves

In this section, we compare the performance of the PS Lorenz curve and our proposed model in estimating income inequality across various countries using goodness-of-fit (GOF) measures. This analysis is based on grouped data in the form of income shares from the WIID, which includes 24,366 datasets from 201 countries, spanning from the beginning of the nineteenth century. To ensure a globally representative sample of income distributions, we selected 10 countries from different world regions. The income shares and inequality measures for these countries are presented in Appendix (Table 3). The selected countries exhibit a wide range of inequality levels, with Gini indices between 0.2846 and 0.4880. Consequently, we believe our results can be generalized to most countries worldwide.

We turn now our attention to the estimation of the curves given in (1) and (10). Let $\mathbf{x}$ be a random sample of size N of a continuous random variable X . Assume that the sample $(\mathbf{x})$ is divided into J mutually exclusive intervals $H_j = [h_{j-1}, h_j)$, $j = 1, \dots, J$. Let $\mathbf{1}_{[h_{j-1}, h_j)}(x_i)$ be an indicator function, taking the value 1 if $x_i \in [h_{j-1}, h_j)$ and zero otherwise. Denote $c_j = \sum_{i=1}^N \mathbf{1}_{[h_{j-1}, h_j)}(x_i)x_i / \sum_{i=1}^N x_i$, $j = 1, \dots, J$ as the proportion of total income held by individuals in the j^{th} interval and the cumulative

proportion by $s_j = \sum_{k=1}^j c_k$. Let $u_j = \sum_{i=1}^N \mathbf{1}_{[h_{j-1}, h_j)}(x_i)/N$, $j = 1, \dots, J$ denote the frequency of the sample $\mathbf{x}$ in the j^{th} interval and $p_j = \sum_{k=1}^j u_k$ the cumulative frequency.

Therefore, the proportion of observations in each group is determined before sampling, so the population proportions (p_j) are fixed, while the income shares (s_j) are treated as random variables. Estimating parametric distributions by maximum likelihood would be misspecified in this case, due to the non-stochastic nature of the group frequencies (Hajargasht and Griffiths 2020). To avoid this potential issue, nonlinear least squares are conventionally used to estimate the vector of parameters of interest by minimizing the distance between the observed income shares and the functional form of the Lorenz curve. In this context, nonlinear least squares can be referred to as the equally weighted minimum distance (EWMD) estimator, which takes the following form:

$$\hat{\theta} = \underset{\gamma}{\operatorname{argmin}} \mathbf{M}(\gamma)' \mathbf{M}(\gamma), \quad (14)$$

where $\mathbf{M}(\theta)' = [m_1(\gamma), \dots, m_{J-1}(\gamma)]$ is the vector of moment conditions, given by

$$\mathbf{M}(\theta) = L(\mathbf{p}; \gamma) - \mathbf{s}, \quad (15)$$

where $\mathbf{s}' = (s_1, \dots, s_{J-1})$ is a vector of cumulative income shares associated with the population proportions $\mathbf{p}' = (p_1, \dots, p_{J-1})$. Closed expressions for $L(\mathbf{p}; \gamma)$ are given in (1) and (10).

The EWMD estimator given in (14) overlooks that, by definition, the sum of the income shares is equal to one, introducing dependence between the income shares. As a result, EWMD provides consistent but inefficient estimates of parameters, including those that depend on the γ parameter, such as inequality measures. To obtain more efficient estimates, optimally weighted minimum distance (OMD) estimators could be used. However, Jorda et al. (2021) find that, in most cases, EWMD yields more accurate estimates of inequality measures than OMD. Therefore, since our primary goal is to obtain consistent estimates and examine the practical implications of using these two Lorenz curves to estimate inequality measures, we opt to use the EWMD estimator.

Table 2 presents the estimates of the gamma parameter for the PS Lorenz curve and the new proposal (SJT curve), along with the bootstrap standard errors of these estimates, the mean squared errors (MSE) for each model, and the estimated Gini index. As expected, the gamma parameter for the PS curve approaches zero in countries with Gini indices below 0.418, reflecting the model's limitation in capturing lower inequality levels. This limitation is further evidenced by the higher MSE for these countries, indicating a poor fit for the PS curve. Our model also shows better performance in countries with high inequality levels, where the Gini index constraint of the PS Lorenz curve is not binding. Notably, the MSE for the PS Lorenz curve is three to five times higher in Honduras, the USA, and Brazil, where it reaches 0.0009, compared to the SJT model's MSE, which ranges between 0.0002 and 0.0003.

Table 2 Parameter estimates, MSE, and estimated Gini indices of the PS and SJT curves

Country	$\gamma_{PS}^{(a)}$	$\gamma_{SJT}^{(a)}$	MSE _{PS}	MSE _{SJT}	Gini _{PS}	Gini _{SJT}	Diff. _{PS} ^(b)	Diff. _{SJT} ^(b)
Bhutan 2022	$2.3340 \cdot 10^{-6}$ (0.0112)	0.4224 (0.0023)	0.0069	0.0001	0.4180	0.2787	0.1334	0.0059
Guinea 2019	$1.6587 \cdot 10^{-6}$ (0.0038)	0.4688 (0.0022)	0.0058	0.0001	0.4180	0.2908	0.1221	0.0051
Canada 2019	$3.8660 \cdot 10^{-6}$ (0.0038)	0.4753 (0.0020)	0.0055	0.0001	0.4180	0.2925	0.1179	0.0076
Israel 2018	$1.2570 \cdot 10^{-6}$ (0.0205)	0.6100 (0.0021)	0.0032	0.0001	0.4180	0.3272	0.0837	0.0071
China 2018	$1.9749 \cdot 10^{-6}$ (0.0550)	0.7559 (0.0024)	0.0020	0.0002	0.4180	0.3636	0.0489	0.0055
Belgium 2022	$7.2599 \cdot 10^{-6}$ (0.0684)	0.8392 (0.0023)	0.0008	0.0000	0.4180	0.3838	0.0284	0.0058
Australia 2018	$2.4280 \cdot 10^{-6}$ (0.0878)	0.8233 (0.0021)	0.0010	0.0001	0.4180	0.3800	0.0275	0.0105
Honduras 2018	0.3950 (0.2517)	1.1848 (0.0028)	0.0009	0.0002	0.4783	0.4621	0.0082	0.0080
United States 2018	0.3827 (0.2422)	1.1773 (0.0024)	0.0009	0.0002	0.4764	0.4605	0.0055	0.0104
Brazil 2021	0.4836 (0.2518)	1.2501 (0.0025)	0.0009	0.0003	0.4917	0.4758	0.0037	0.0122

(a) Bootstrapped standard errors for γ_{PS} and γ_{SJT} are presented in parentheses. These estimates are based on samples of size 1000 and 200 repetitions. (b) Diff. refers to the absolute difference between the estimated and the empirical Gini indices. Source: authors' compilation

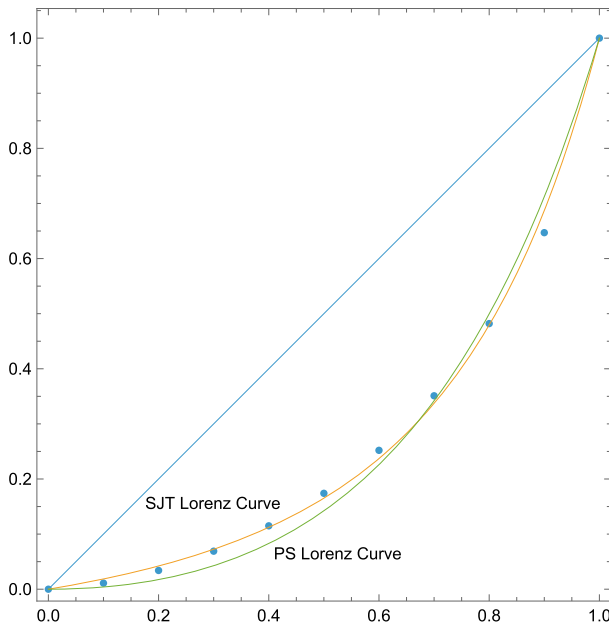


Fig. 4 Lorenz curves for Brazil (2021): PS model (green) and SJT model (orange).

In terms of the Gini coefficient, the SJT curve provides a much closer approximation to empirical inequality estimates in countries with low inequality levels. However, in Brazil, Honduras, and the USA, which exhibit empirical Gini indices between 0.47 and 0.49, the PS model seems to provide more accurate estimates of this inequality measure, although the differences with the SJT curve are not important.

To investigate this surprising result, Figure 4 presents the observed income shares in Brazil in 2021 (black points) alongside the estimated PS and SJT models. The graph shows that the SJT model provides a highly accurate fit and clearly outperforms the PS model, consistent with the MSE values. However, the gap between the observed and estimated Gini index is three times smaller for the PS Lorenz curve (see Table 2). This occurs because the Gini index is proportional to the area between the Lorenz curve and the egalitarian line. As a result, underestimations in some income shares can be offset by overestimations in others. Thus, the seemingly better performance of the PS model is a statistical artifact arising from the way the Gini coefficient is calculated.

5 Conclusions and final remarks

In this paper, we have derived new and important economic properties of the PS Lorenz curve, focusing on its enhanced representation through convex linear combinations of Lorenz curves. We have provided explicit formulas for the Gini index, as well as the DWK index, and established the Lorenz ordering of the curve. Additionally, we

demonstrated that the Gini index for this model is constrained by a lower bound of 0.418, which significantly limits its applicability to countries with low inequality.

To address this limitation, we proposed a new Lorenz curve with a broader range for the Gini index, offering greater flexibility in measuring inequality in income distributions. Our proposal also presents several additional advantages. First, it offers a simpler functional form compared to the PS curve, making it more practical to use. Second, the Gini index derived from our model has a more straightforward expression than that of the PS curve. Finally, both the quantile function and the CDF are available in closed form (see equations 12 and 13).

Our results suggest that the SJT curve consistently outperforms the PS curve, yielding lower MSE statistics in all cases. For countries with low inequality, the Gini index of the PS model tends to its lower bound, limiting its accuracy. In contrast, for countries with higher inequality, both models provide reasonably accurate estimates of inequality measures. However, the apparent good performance of the PS model seems to be influenced by the way the Gini coefficient is defined, as underestimations in some income shares can be offset by overestimations elsewhere. Overall, the Lorenz curve proposed in this paper surpasses the PS model by offering a wider range of Gini estimates and consistently delivering a better fit, making it a more robust tool for measuring inequality across diverse income distributions.

Appendix

Proof of Theorem 1 If $c^{-1} = e^{\gamma(e-1)} - 1$, the representation (2) is obtained by the following expansions,

$$\begin{aligned} L(p) &= cp \left[e^{\gamma(e^p-1)} - 1 \right] \\ &= cp \sum_{k=1}^{\infty} \frac{\gamma^k (e^p - 1)^k}{k!} \\ &= \sum_{k=1}^{\infty} \frac{\gamma^k (e - 1)^k}{(e^{\gamma(e-1)} - 1)k!} p \left(\frac{e^p - 1}{e - 1} \right)^k, \end{aligned}$$

which corresponds to Equation (2). $\square$

Proof of Theorem 2 If $\gamma > 0$, the function $L(p; \gamma)$ is differentiable with respect to γ . If $v(p) = 1 - e^p$, we can write $L(p; \gamma) = \frac{e^{-\gamma v(p)} - 1}{e^{-\gamma v(1)} - 1}$. Then,

$$\frac{\partial L(p; \gamma)}{\partial \gamma} = \frac{e^{-\gamma v(p) + \gamma v(1)}}{(e^{\gamma v(1)} - 1)} \cdot g(p),$$

where

$$g(p) = v(1)(1 - e^{\gamma v(p)}) - v(p)(1 - e^{\gamma v(1)}).$$

It can be shown that $g(p) \leq 0$, and then, $L(p; \gamma)$ is decreasing in γ and we have the result. $\square$

Table 3 Income shares by decile groups and empirical Gini index for ten selected countries

Country	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	Gini index
Bhutan (2022)	3.63	5.18	6.21	7.25	8.28	9.35	10.6	12.16	14.63	22.71	0.2846
Guinea (2019)	3.48	5.02	6.03	7.02	8.11	9.23	10.59	12.35	15.09	23.09	0.2959
Canada (2019)	2.96	4.88	6.05	7.18	8.29	9.48	10.81	12.51	15.06	22.78	0.3001
Israel (2018)	2.66	4.32	5.55	6.70	7.83	9.12	10.69	12.66	15.53	24.93	0.3343
China (2018)	2.70	3.91	5.03	6.14	7.30	8.60	10.16	12.26	15.77	28.13	0.3691
Belgium (2022)	2.26	3.32	4.30	5.59	6.96	8.80	11.09	13.76	17.21	26.71	0.3896
Australia (2018)	1.85	3.33	4.63	5.96	7.41	8.99	10.83	13.12	16.17	27.72	0.3905
Honduras (2018)	1.24	2.41	3.59	4.86	6.29	7.85	9.81	12.78	17.51	33.66	0.4701
United States (2018)	1.05	2.43	3.65	4.91	6.3	8.02	10.04	12.78	17.11	33.70	0.4709
Brazil (2021)	1.10	2.30	3.50	4.60	5.90	7.80	9.90	13.10	16.50	35.40	0.4880

Source: UNU-WIDER (2024)—version 28 November 2023

Proof of Lemma 1 According to the representation (2), the Gini index can be written as

$$G_L = 1 - 2 \int_0^1 L(p, \gamma) dp = 1 - 2 \sum_{k=1}^{\infty} w_k \int_0^1 L_k(p) dp,$$

where w_k and $L_k(p)$ are defined in (4) and (3), respectively. Then,

$$\begin{aligned} \int_0^1 L_k(p) dp &= \frac{1}{(e-1)^k} \int_0^1 p(e^p - 1)^k dp \\ &= \frac{1}{(e-1)^k} \sum_{j=0}^k (-1)^j \binom{k}{j} \int_0^1 p e^{(k-j)p} dp \\ &= \frac{1}{(e-1)^k} \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{1 + e^{k-j}(k-j-1)}{(k-j)^2}, \end{aligned}$$

and we get the formula. $\square$

Proof of Lemma 2 Using Theorem 2, since $\gamma \geq 0$, then $G_\gamma \geq G_0$. Now, the Gini index of $L(p; 0)$ is given by

$$\begin{aligned} G_0 &= 1 - 2 \int_0^1 L(p, 0) dp \\ &= 1 - 2 \int_0^1 p \frac{e^p - 1}{e - 1} dp \\ &= \frac{e - 2}{e - 1}, \end{aligned}$$

which is equal to 0.418, and we have the result. $\square$

Proof of Lemma 3 For computing the DWK index, we need to calculate the integral,

$$\int_0^1 (1-p)^{v-1} L_X(p) dp = \sum_{k=1}^{\infty} w_k \int_0^1 (1-p)^{v-1} L_k(p) dp,$$

where $L_k(p)$ is defined in (3) and w_k in (4). Now, we have

$$\int_0^1 (1-p)^{v-1} L_k(p) dp = \frac{1}{(e-1)^k} \sum_{j=0}^k \binom{k}{j} \int_0^1 p(1-p)^v e^{(k-j)p} dp,$$

where this last integral can be computed in terms of the Gamma and the incomplete Gamma functions, and then, we get (8). $\square$

Proof of Equation (13) Here we briefly describe how to obtain the cumulative distribution function of (10) given in Equation (13). The Lorenz curve is defined as

$$L(p) = \frac{1}{\mu} \int_0^p F^{-1}(u) du, \quad 0 \leq p \leq 1,$$

where $F^{-1}(u)$ is the inverse function of the CDF. Computing the first derivative in the above formula we have,

$$F^{-1}(p) = \mu L'(p),$$

where $p \in (\mu L'(0), \mu L'(1))$ (see also Arnold and Sarabia (2018), chapter 3). In our case,

$$\tilde{L}'(p) = \frac{\gamma \exp[\gamma(e^p - 1) + p]}{\exp[\gamma(e - 1)] - 1},$$

and solving for p the equation $\mu \tilde{L}'(p) = x$, we get (13), using the Lambert function.

Acknowledgements The authors thank the anonymous reviewers and the Associate Editor for their valuable comments and suggestions.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aggarwal V (1984) On optimum aggregation of income distribution data. *Sankhyā: The Indian Journal of Statistics, Series B*, 46(3):343–355
- Arnold BC, Sarabia JM (2018) Majorization and the Lorenz order with applications in applied mathematics and economics. Springer
- Bandourian R, McDonald J, Turley RS (2002) A comparison of parametric models of income distribution across countries and over time. *Luxembourg income study working paper*
- Chotikapanich D (1993) A comparison of alternative functional forms for the Lorenz curve. *Econ Lett* 41(2):129–138
- Dhongde S, Minoiu C (2013) Global poverty estimates: a sensitivity analysis. *World Dev* 44(1):1–13
- Donaldson D, Weymark J (1980) A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory* 22(1):67–86
- Gómez-Déniz E (2016) A family of arctan Lorenz curves. *Empirical Economics* 51:1215–1233
- Hajargasht G, Griffiths WE (2020) Minimum distance estimation of parametric Lorenz curves based on grouped data. *Economet Rev* 39(4):344–361

- Henle J, Horton N, Jakus S (2008) Modelling inequality with a single parameter. In: Chotikapanich D (ed) Modeling income distributions and Lorenz curves. Springer, pp 255–269
- Jorda V, Sarabia JM, Jäntti M (2021) Inequality measurement with grouped data: parametric and non-parametric methods. *J R Stat Soc Ser A Stat Soc* 184(3):964–984
- Kakwani N (1980a) *Income inequality and poverty*. World Bank New York
- Kakwani N (1980) On a class of poverty measures. *Econometrica* 48(2):437–446
- Kakwani NC, Podder N (1973) On the estimation of lorenz curves from grouped observations. *Int Econ Rev* 14(2):278–292
- Kleiber C, Kotz S (2003) Statistical size distributions in economics and actuarial sciences, vol 470. John Wiley & Sons, New Jersey
- Paul S, Shankar S (2020) An alternative single parameter functional form for lorenz curve. *Empirical Economics* 59(3):1393–1402
- Ryu HK, Slottje DJ (1996) Two flexible functional form approaches for approximating the lorenz curve. *Journal of Econometrics* 72(1–2):251–274
- Sarabia JM (2008) Parametric lorenz curves: models and applications. In: Chotikapanich D (ed) Modeling income distributions and Lorenz curves. Springer, pp 167–190
- Sarabia JM, Castillo E, Slottje DJ (2001) An exponential family of lorenz curves. *South Econ J* 67(3):748–756
- Sarabia JM, Jordá V, Trueba C (2017) The lamé class of lorenz curves. *Communications in Statistics-Theory and Methods* 46(11):5311–5326
- UNU-WIDER (2024) World Income Inequality Database. Accessed: October 14, 2024
- Villaseñor J, Arnold BC (1989) Elliptical lorenz curves. *Journal of Econometrics* 40(2):327–338
- Yitzhaki S (1983) On an extension of the gini inequality index. *Int Econ Rev* 24(3):617–628

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.