



On the capacity of the Gini index to represent income distributions

Yang Liu¹ · Joseph L. Gastwirth¹ 

Received: 23 May 2019 / Accepted: 14 February 2020 / Published online: 24 February 2020
© Sapienza Università di Roma 2020

Abstract

Almost all governmental and international agencies use the Gini index to summarize income inequality in a nation or the world. The index has been criticized because it can have the same value for two different distributions. It will be seen that other commonly used summary measures of inequality are subject to the same criticism. The Gini index has the advantage that it is able to distinguish between two distributions that have identical integer valued generalized entropy measures. Because no single measure can fully summarize a distribution, researchers should consider combining the Gini index with another measure appropriate for the topic being studied.

Keywords Generalized entropy measures · Gini index · Lorenz curve · Measures of income inequality · Pareto distribution · Moment problem

1 Introduction

Although the Gini index is the most commonly used measure of income inequality and has many desirable properties (see [12,16,21]), it has often been criticized (see [3,4,13,20]), because its values calculated from two different distributions can be equal. While this is true, the other major summary indices of inequality, e.g., the generalized entropy family including the Theil's index, the mean log deviation, one half of the squared coefficient of variation, as well as the Palma index (the ratio of the shares of income received by the top 10% to the bottom 40%) and other ratios of the top $100q\%$ to the bottom $100p\%$ shares e.g., Dorling [6] $p = q = 0.2$, Jasso [15] $p = 1 - q$, also can have the same value for different distributions. From a statistical viewpoint, this is not surprising, as two parameters, the mean and the standard deviation are needed to determine a normal distribution. Thus, it is unreasonable to expect any single summary measure to uniquely determine the underlying distribution and

✉ Joseph L. Gastwirth
jlgast@gwu.edu

Yang Liu
louisliu@gwu.edu

¹ Department of Statistics, George Washington University, Washington, DC, USA

researchers should consider combining the Gini index with another measure focusing on the portion of the income distribution most relevant to the topic under investigation.

This paper presents several examples clarifying the relationship between income inequality measures and the underlying income distribution. In Sect. 2, examples of two distributions with the same value of one of the three commonly used inequality measures (e.g., the Gini index), but different values of the other two (e.g., the mean log deviation and the Theil index) are presented. A classical example in the theory of moments is used in Sect. 3 to demonstrate that the values of an infinite family of the generalized entropy indices can be identical for two different distributions. In particular, this implies that two distributions can have identical values of one half of the squared coefficient of variation, the mean log deviation and the Theil index but their Gini indices are not equal. In Sect. 4, several recently proposed measures based on the ratio of the share of income received by the top $100q\%$ to the share of income received by the bottom $100p\%$ are examined. Given the Lorenz curve of an income distribution, the share of the top $100q_0\%$ and the share of the bottom $100p_0\%$, using the approach described in [9, 11], one can construct two Lorenz curves having the same values of $L(p_0)$ and $L(1 - q_0)$ as the given Lorenz curve, which bound the original Lorenz curve from above and below. Consequently, the given Lorenz curve and the two bounding Lorenz curves have the same ratio of the incomes received by the top $100q_0\%$ and bottom $100p_0\%$. The Gini indices of the three distributions, however, are different and the Gini index of the given income distribution lies in between the Gini indices of the bounding Lorenz curves. Furthermore, an example of a class of discrete income distributions with the same ratio of the top to bottom shares but very different Gini indices is presented in Sect. 5. A summary and discussion of the implications of these results are of given in Sect. 6.

2 Examples of distributions with the same value of one of three commonly used inequality but not the others

Consider three commonly used inequality indices: the mean log deviation (MLD), the Theil index (TI) and the Gini index (G), which are defined in [5] as

$$MLD_Y = -E \left[\log \frac{Y}{\mu_Y} \right] \quad TI_Y = E \left[\frac{Y}{\mu_Y} \log \frac{Y}{\mu_Y} \right] \quad \text{and} \quad G_Y = \frac{E |Y_1 - Y_2|}{2\mu_Y}, \quad (1)$$

where Y is the random variable (r.v.) generating the income data, $\mu_Y = E[Y]$ and Y_1 and Y_2 are i.i.d. as Y . The construction of the examples uses the fact that for any r.v. Y and a specific inequality index (e.g., MLD, the Theil index, or the Gini index) with value γ , there exists a discrete r.v. X_k of the form

$$X_k = \begin{cases} 1 & \text{with probability } p \\ k & \text{with probability } 1 - p, \end{cases} \quad (2)$$

with the same value (γ) of that inequality measure. For any given value of k , one solves for p after setting the value of the specified inequality index of X_k to γ . The following example demonstrates that while X_k and Y have the same value of the specified index, their values of the other inequality indices can be noticeably different.

To illustrate the construction of the r.v.'s X_k , let Y follow a Pareto distribution with (α, θ) . From the formulas given in [1,5], one obtains

$$MLD_Y = -\frac{1}{\alpha} + \log \frac{\alpha}{\alpha - 1} \quad TI_Y = \frac{1}{\alpha - 1} + \log \frac{\alpha - 1}{\alpha} \quad \text{and} \quad G_Y = \frac{1}{2\alpha - 1}. \quad (3)$$

Table 1 MLD, Theil Index, Gini Index for Y and different settings of X_3

	MLD	Gini index	Theil index
$Y: Pareto(\alpha = 3, \theta = 1)$	0.072	0.200	0.095
$X_3: p = 0.187$ (same MLD)	0.072	0.116	0.055
$X_3: p = 0.334$ (same TI)	0.115	0.191	0.095
$X_3: p = 0.896$ (same TI)	0.074	0.154	0.095
$X_3: p = 0.355$ (same Gini)	0.120	0.200	0.100
$X_3: p = 0.845$ (same Gini)	0.100	0.200	0.120

Table 2 MLD, Theil Index, Gini Index for Y and different settings of X_{10}

	MLD	Gini index	Theil index
$Y: Pareto(\alpha = 3, \theta = 1)$	0.072	0.200	0.095
$X_{10}: p = 0.052$ (same MLD)	0.072	0.047	0.036
$X_{10}: p = 0.136$ (same TI)	0.182	0.120	0.095
$X_{10}: p = 0.993$ (same TI)	0.047	0.061	0.095
$X_{10}: p = 0.229$ (same Gini)	0.296	0.200	0.164
$X_{10}: p = 0.972$ (same Gini)	0.164	0.200	0.296

From (2), the values of the three indices for a discrete r.v. of the form X_k are

$$MLD_{X_k} = p \log k - \log \frac{k}{p + (1 - p)k}, \tag{4}$$

$$TI_{X_k} = \frac{k(1 - p)}{p + k(1 - p)} \log k - \log(p + k(1 - p)) \quad \text{and} \tag{5}$$

$$G_{X_k} = \frac{p(1 - p)(k - 1)}{p + k(1 - p)}. \tag{6}$$

When $Y \sim Pareto(\alpha = 3, \theta = 1)$, by (3), the mean log deviation for Y is 0.072. Consider a r.v. of the form X_3 , i.e. $k = 3$. From the condition $MLD_Y = MLD_{X_3}$ and (4), one obtains that $p = 0.187$. For this value of p , the values of the other two indices (the Gini index and the Theil index) for X_3 are determined by (5) and (6), which are 0.116 and 0.055, respectively. Similarly, for the Gini index or the Theil index, one can obtain the discrete r.v. X_3 by solving p from the condition that the value of the specific index for X_3 equals the corresponding value for Y . Once the value of p is calculated, the values of the other two indices for X_3 are determined.

When $k = 3$, Table 1 reports the values of p determined from Eqs. (4), (5) and (6) so that one of the values of the mean log deviation, the Gini index, and the Theil index for X_3 matches the corresponding value for $Y \sim Pareto(3, 1)$. The values of the other two indices for that distribution are also reported. Table 2 presents the similar results when $k = 10$.

Tables 1 and 2 show that one can have different members of the X_k family with the same value of any specific inequality index and noticeably different values of the others. For example, consider the second row of both tables referring to the r.v.'s X_3 and X_{10} with the same value of the mean log deviation. Notice that the Gini index of X_{10} is less than one-half of the Gini index of X_3 . The Theil indices for X_3 and X_{10} are also different. The results in Tables 1 and 2 also illustrate that for any X_k , there can be two solutions for the parameter p ,

so that two different X_k s has the same value of the specific index, e.g., consider the third and fourth row in Table 1, X_3 with $p = 0.334$ and $p = 0.896$ and the same value of the Theil index. Therefore, their corresponding X_k would have the same value of the specific index but different values of the other inequality indices.

3 An example of two distributions with the same integer-valued generalized entropy measure

The generalized entropy family of indices are defined as

$$GE_Y(a) = \frac{1}{a(a-1)} \cdot \left[\frac{E[Y^a]}{E[Y]^a} - 1 \right], \tag{7}$$

where a gives more weight to the upper portion of the distribution as it increases. $GE_Y(0)$ and $GE_Y(1)$ are defined as $GE_Y(0) = \lim_{a \rightarrow 0} GE_Y(a)$ and $GE_Y(1) = \lim_{\alpha \rightarrow 1} GE_Y(a)$ respectively. Recall that $GE_Y(0)$ is the mean log deviation, $GE_Y(1)$ is the Theil index, and $GE_Y(2)$ is one half of the squared coefficient of variation. It follows from (7) that if two underlying income distributions have the same a th moments for $a = 1, 2, \dots$, the corresponding sequence of the generalized entropy measures for these two underlying income distributions will be the same.

It is well known that the lognormal distribution does not satisfy the conditions [2,19] required for a distribution to be uniquely determined by its moments. In particular, Heyde [14] presented a family of distributions indexed by a parameter ϵ , having the same moments as the lognormal. Hence, their values of $GE(a)$ are the same for each integer value of a . The Heyde’s family of r.v.’s is given as the following.

Lemma 1 (Heyde’s family of r.v.’s) *Let Y be the lognormal random variable such that $\log Y \sim N(\mu, \sigma)$ and let Y_ϵ be a random variable with the pdf*

$$f_{Y_\epsilon}(y) = f_Y(y) \cdot [1 + \epsilon \sin(2\pi(\log y - \mu))] \text{ for } y > 0 \text{ and } -1 \leq \epsilon \leq 1, \tag{8}$$

then we have

$$E[Y_\epsilon^a] = E[Y^a] + I_a, \tag{9}$$

where $I_a = \epsilon \int_{-\infty}^{\infty} y^a f_Y(y) \sin(2\pi(\log y - \mu)) dy = 0$ for all integer values of a . Hence, for all integer values of a ,

$$GE_{Y_\epsilon}(a) = GE_Y(a). \tag{10}$$

The density function of the Heyde’s family of r.v.’s differs from the log normal random variable by a function that oscillates with a magnitude of ϵ , but integrates to zero. Lemma 1 implies that for any lognormal distribution, there is an infinite family of distributions with the same general entropy index for all integer values of a . On the other hand, the values of the corresponding Gini indices differ. The Gini indices for several members of Heyde’s family of distributions having the same moments as the lognormal distribution with $\mu = 0, \sigma = 1$ are given in Table 3.

Table 3 shows that when ϵ takes different values, the Gini indices for Y_ϵ are different. Because the distributions of Y_ϵ and Y have the same moment sequence, their distributions should be very close to one another, which is why the values of the Gini index for different ϵ are similar. In [19], one can find other examples of families of distributions with the same moment sequence as well as the same values of $GE(a)$ for $a = 1, 2, \dots$

Table 3 The Gini index of Y_ϵ with $\log Y \sim N(0, 1)$

ϵ	-0.9	-0.5	0	0.5	0.9
G_{Y_ϵ}	0.5182	0.5198	0.5205	0.5198	0.5182

4 Example of distributions with the same ratio of the top to bottom shares

The inequality measure $r(p, q)$ having been used by Palma [18] ($p = 0.4, q = 0.1$) and Dorling [6] ($p = 0.2, q = 0.2$) is the ratio of the share of the income received by the top 100 $q\%$ to the share of the income received by the bottom 100 $p\%$. This family of inequality measures focuses on the difference of the upper 100 $q\%$ and the lower 100 $p\%$ of the distribution, but ignores the middle portion of the distribution. It is obvious that there are many Lorenz curves having the same values of $L(p)$ and $L(1 - q)$ but differ noticeably over the range of $(p, 1 - q)$.

The ratio of the top to bottom shares for an income Y is defined as

$$r_Y(p, q) = \frac{1 - L_Y(1 - q)}{L_Y(p)}, \tag{11}$$

where $L_Y(p) = \mu_Y^{-1} \int_{-\infty}^p F^{-1}(t)dt = \mu_Y^{-1} \int_{-\infty}^{y_p} t \cdot f_Y(t)dt$ is the Lorenz curve [8] of Y , and $y_p = F^{-1}(p) = \inf\{y : F_Y(y) \geq p\}$ is the p th quantile of Y .

For any pair of values (p_0, q_0) and an income distribution (Y) with its Lorenz curve $L_Y(p)$, following the approach in [9,11], one can obtain two piecewise linear Lorenz curves that bound $L_Y(p)$ from above (by $L_A(p)$) and below (by $L_B(p)$) with the same values of $L_Y(p_0)$ and $L_Y(1 - q_0)$, i.e., $L_A(p_0) = L_B(p_0) = L_Y(p_0)$ and $L_A(1 - q_0) = L_B(1 - q_0) = L_Y(1 - q_0)$. By construction, the ratio $r(p_0, q_0)$ of $L_A(p)$, $L_B(p)$ and $L_Y(p)$ are equal.

The formulas for $L_A(p)$ and $L_B(p)$ will now be described. Let $L_A(p)$ be the piecewise linear Lorenz curve bounding $L_Y(p)$ from above. $L_A(p)$ consists of line segments $T_{A_1}(p)$, $T_{A_2}(p)$ and $T_{A_3}(p)$, where $T_{A_1}(p)$ is the line segment connecting $(0, 0)$ and $(p_0, L_Y(p_0))$; $T_{A_2}(p)$ is the line segment connecting $(p_0, L_Y(p_0))$ and $(1 - q_0, L_Y(1 - q_0))$; and $T_{A_3}(p)$ is the line segment connecting $(1 - q_0, L_Y(1 - q_0))$ and $(1, 1)$. Therefore, it follows that

$$L_A(p) : \begin{cases} T_{A_1}(p) = \frac{L_Y(p_0)}{p_0} \cdot p & 0 \leq p \leq p_0 \\ T_{A_2}(p) = \frac{L_Y(1-q_0) - L_Y(p_0)}{1-q_0-p_0} \cdot (p - p_0) + L_Y(p_0) & p_0 < p \leq 1 - q_0 \\ T_{A_3}(p) = \frac{1 - L_Y(1-q_0)}{q_0} \cdot (p - 1) + 1 & 1 - q_0 < p < 1 \end{cases} \tag{12}$$

$L_B(p)$ consists four linear segments $T_{B_1}(p)$, $T_{B_2}(p)$, $T_{B_3}(p)$, and $T_{B_4}(p)$ that circumscribes $L_Y(p)$ at $(p_0, L_Y(p_0))$ and $(1 - q_0, L_Y(1 - q_0))$. Here $T_{B_2}(p)$ and $T_{B_3}(p)$ are the two tangent lines segments circumscribing $L_Y(p)$ at $(p_0, L_Y(p_0))$ and $(1 - q_0, L_Y(1 - q_0))$ respectively. They intersect at (x_0, y_0) . In addition, $T_{B_2}(p)$ intersects $L(p) = 0$ at $(x_1, 0)$ and $T_{B_1}(p)$ is the line segment connecting $(0, 0)$ and $(x_1, 0)$. $T_{B_3}(p)$ intersects $p = 1$ at $(1, T_{B_3}(1))$, and $T_{B_4}(p)$ connects $(1, T_{B_3}(1))$ to $(1, 1)$. Therefore, we have

$$L_B(p) : \begin{cases} T_{B_1}(p) = 0 & 0 \leq p \leq x_1 \\ T_{B_2}(p) = L'_Y(p_0)(p - p_0) + L_Y(p_0) & p_0 < p \leq x_0 \\ T_{B_3}(p) = L'_Y(1 - q_0)(p - (1 - q_0)) + L_Y(1 - q_0) & x_0 < p < 1 \\ T_{B_4}(p) = [q_0 \cdot L'_Y(1 - q_0) + L_Y(1 - q_0), 1] & p = 1 \end{cases} \tag{13}$$

where $L'_Y(p) = F_Y^{-1}(p)/\mu$ is the derivative of $L_Y(p)$.

By the construction, the area between the line of equality and $L_A(p)$ will be smaller than the area between the line of equality and $L_Y(p)$, and the area between the line of equality and $L_B(p)$ will be the largest. This implies that $G_B > G_Y > G_A$. In addition, $L_A(p_0) = L_Y(p_0) = L_B(p_0)$ and $L_A(1 - q_0) = L_Y(1 - q_0) = L_B(1 - q_0)$, yielding $r_A(p_0, q_0) = r_B(p_0, q_0) = r_Y(p_0, q_0)$. Hence, the three Lorenz curves have the same value of $r(p_0, q_0)$ but different values of the Gini index.

To illustrate an example of how the curves L_A and L_B bounding a Lorenz curve of an income distribution, consider $Y \sim \text{Pareto}(\alpha, \theta)$. From [1], we have

$$G_Y = \frac{1}{2\alpha - 1}, \tag{14}$$

$$L_Y(p) = 1 - (1 - p)^{1 - \frac{1}{\alpha}}, \text{ and} \tag{15}$$

$$L'_Y(p) = \frac{\alpha - 1}{\alpha} (1 - p)^{-\frac{1}{\alpha}}. \tag{16}$$

When $p_0 = 0.4, q_0 = 0.1$ and $\alpha = 1.125$, from (12),(13), (15), and (16), $L_A(p)$ and $L_B(p)$ are

$$L_A(p) = \begin{cases} 0.138 \cdot p & 0 \leq p \leq 0.4 \\ 0.341 \cdot (p - 0.4) + 0.055 & 0.4 < p \leq 0.9 \\ 7.743 \cdot (p - 1) + 1 & 0.9 < p < 1 \end{cases} \text{ and} \tag{17}$$

$$L_B(p) = \begin{cases} 0 & 0 \leq p \leq 0.121 \\ 0.175(p - 0.4) + 0.055 & 0.121 < p \leq 0.779 \\ 0.860(p - 0.9) + 0.226 & 0.779 < p < 1 \\ [0.312, 1] & p = 1 \end{cases}. \tag{18}$$

Fig. 1 presents $L_Y(p)$ (solid curve), $L_A(p)$ (dotted curve) and $L_B(p)$ (dash curve) for $Y \sim \text{Pareto}(1, 1.125)$, $p_0 = 0.4$ and $q_0 = 0.1$. From the figure, it can be seen that the $L_Y(p)$ is between $L_A(p)$ and $L_B(p)$. The three Lorenz curves have the same $r(0.4, 0.1)$ as $L_A(p)$ and $L_B(p)$, because their values are the same when $p = 0.4$ and 0.9 . Calculating the areas between each of the Lorenz curve and the line of equality, it follows that

$$G_A = 0.715 < G_Y = \frac{1}{2(1.125) - 1} = 0.800 < G_B = 0.820. \tag{19}$$

In this example, the difference between the Gini indices of the two bounding Lorenz curves is about 0.1. The reason G_A is less than G_Y is because its distribution assumes that all individuals in the region between the 40th and 90th percentiles have the same income.

5 A family of discrete distributions with the same value of the ratio of top to bottom shares

This section describes a family of discrete distributions having the same pre-specified ratio of the top to bottom shares $r(p_0, q_0)$, with noticeably different values of the Gini indices. This means that the middle portion of these distributions varies substantially while the shares of income received by the top $100q_0\%$ and the bottom $100p_0\%$ remain unchanged.

Suppose a population only has three classes of households, lower, middle and upper, and every person within each class receives the same income. For a specified value of the ratio of the top to bottom shares $r(p_0, q_0)$, by varying the share received by the households in the middle class, one can generate many populations with the same value of the ratio of the top

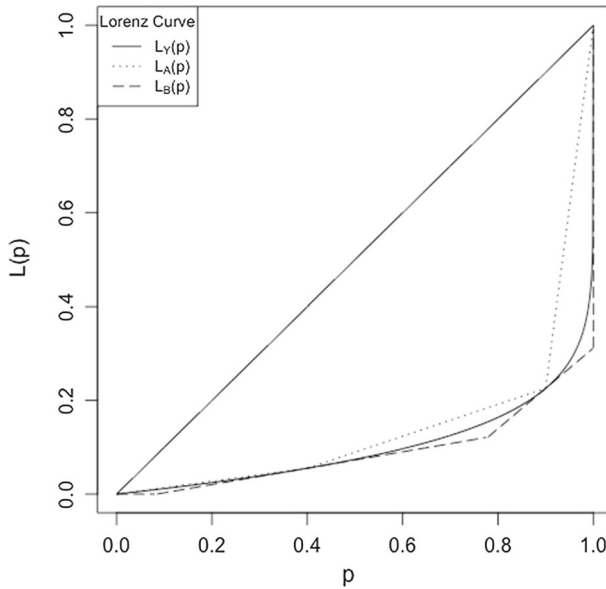


Fig. 1 The Lorenz curves of $L_Y(p)$, $L_A(p)$ and $L_B(p)$ with $Y \sim \text{Pareto}(1.125, 1)$ and the same value of $r(0.4, 0.1)$

to bottom shares, but different values of the Gini index. For example, let $p_0\%$ ($q_0\%$) be the proportion of the households belonging to the lower (upper) class, so that $1 - p_0 - q_0\%$ is proportion of the households in the middle class. Consider the family of income distributions Y_x with the same value $r = r(p_0, q_0)$ defined by

$$Y_x = \begin{cases} 1 & \text{with probability } p_0 \\ x & \text{with probability } 1 - p_0 - q_0, \\ rp_0/q_0 & \text{with probability } q_0 \end{cases} \tag{20}$$

where

$$r_{Y_x}(p_0, q_0) = \frac{1 - L_{Y_x}(1 - q_0)}{L_{Y_x}(p_0)} = \frac{q_0(rp_0/q_0)}{1 \cdot p_0} = r, \tag{21}$$

for any $1 \leq x \leq rp_0/q_0$ and $rp_0/q_0 > 1$. While r is determined by $L(p_0)$ and $L(1 - q_0)$, and does not vary for any distributions of the form Y_x , the share of the income received by the middle class depends on x will increase as x goes from 1 to rp_0/q_0 , i.e., as the income recieved by the middle class approaches that of the upper class.

For any value of x , the Gini index of Y_x is

$$G_{Y_x} = \frac{\Delta_{Y_x}}{2\mu_{Y_x}} = p_0 + \frac{(rp_0 - xq_0)(1 - p_0 - q_0) - p_0(1 - p_0)}{p_0 + x(1 - p_0 - q_0) + rp_0} \tag{22}$$

$$= p_0 - q_0 + \frac{p_0(1 - p_0)}{p_0 + x(1 - p_0 - q_0) + rp_0} \cdot \left(r - \frac{1 - q_0}{1 - p_0} \right). \tag{23}$$

Table 4 The Gini index of Y_x for different values of x , where $r = 10$, $p_0 = 0.4$ and $q_0 = 0.1$

x	1	8	16	24	32	40
G_{Y_x}	0.716	0.543	0.465	0.424	0.400	0.384

When $r > (1 - q_0)/(1 - p_0)$, G_{Y_x} is monotone decreasing in x , and when $r < (1 - q_0)/(1 - p_0)$, G_{Y_x} is monotone increasing in x . In either case, G_{Y_x} is in between

$$G_{Y_1} = 1 - q_0 - \frac{1 - q_0}{1 - q_0 + rp_0} \text{ and} \tag{24}$$

$$G_{Y_r} = p_0 - \frac{p_0q_0(1 - p_0)}{p_0q_0 + rp_0(1 - p_0)}. \tag{25}$$

Therefore, as the income x , that all members of the middle class receive, varies, the Gini index of Y_x also varies.

To illustrate the wide range of the values that the Gini index can have for different distributions as the Y_x family, consider the case that $r = 10$, $p_0 = 0.4$, and $q_0 = 0.1$. Table 4 presents the values of the Gini index for several values of x .

From Table 4, one sees that the Gini indices of Y_1 and Y_{40} are 0.716 and 0.384 respectively. These values reflect that as x increases from 1^+ to 40^- , the income of a middle class household goes from being lower class to being near upper class and the Gini index decreases.

6 Conclusion

This article demonstrates that the major indices of income inequality can take the same value for different distributions so that criticizing the Gini index or any other single measure of income inequality due to this reason is inappropriate. An example of two different distributions, arising in the theory of moments [2], having identical moments of all orders, indicates that a single index will not be able to distinguish between all pairs of distributions. The Gini index, however, is able to distinguish between members of a family of distributions with the same values for the generalized entropy measures with an integer index.

The results in Sect. 3 demonstrate that it is unreasonable to expect a single measure to completely describe the entire income distribution. Hence combining several measures may provide more information about the income distribution. For example, Oancea and Pirjol [17] show that for a heavily skewed distribution, the Theil index can increase to infinity, but the Gini index is bounded by one. However, the modified Gini index [10] which replaces the mean by the median, i.e., multiplies the Gini index by the ratio of the mean to the median, a measure of skewness, resolves this problem. Foster and Wolfson [7] combine the Gini index with the relative median deviation, i.e., $(\mu_U - \mu_L)/\mu$ where μ_U is the mean of those above the median and μ_L is the mean of those below the median, to measure the polarization of an income distribution.

Because no single measure can summarize the entire income distribution, researchers may benefit from combining the Gini index with another measure which emphasizes the portion of the underlying distribution most relevant to the research problem.

Acknowledgements It is a pleasure to thank two anonymous reviewers and the editor for many helpful suggestions.

References

1. Arnold, B.: Pareto Distributions, 2nd edn. CRC, Boca Raton (2015)
2. Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York (2008)
3. Ceriani, L., Verme, P.: The origins of the Gini index: extracts from *variabilità e mutabilità* (1912) by Corrado Gini. *J. Econ. Inequal.* **10**(3), 421–443 (2012)
4. Charles-Coll, J.A.: Understanding income inequality: concept, causes and measurement. *Int. J. Econ. Manag. Sci.* **1**(3), 17–28 (2011)
5. Cowell, F.A.: Measuring Inequality, 3rd edn. Oxford University Press, Oxford (2011)
6. Dorling, D.: Inequality and the 1%, 1st edn. Verso, Brooklyn (2014)
7. Foster, J.E., Wolfson, M.C.: Polarization and the decline of the middle class: Canada and the US. *J. Econ. Inequal.* **8**(2), 247–273 (2010)
8. Gastwirth, J.L.: A general definition of the Lorenz curve. *Econometrics* **39**(6), 1037–1039 (1971)
9. Gastwirth, J.L.: The estimation of the Lorenz curve and Gini index. *Rev. Econ. Stat.* **54**(3), 306–316 (1972)
10. Gastwirth, J.L.: A robust Gini type-index better detects shifts in the income distribution: a reanalysis of income inequality in the United States from 1967–2011. Available at SSRN 2164745 (2012)
11. Gastwirth, J.L.: Is the Gini index of inequality overly sensitive to changes in the middle of the income distribution? *Stat. Public Policy* **4**(1), 1–11 (2017)
12. Giorgi, G.M., Gigliarano, C.: The Gini concentration index: a review of the inference literature. *J. Econ. Surv.* **31**(4), 1130–1148 (2017)
13. Hesse, J.O.: Fact or fiction? complexities of economic inequality in twenties century Germany. In: Hudson, P., Tribe K. (eds.) *The Contradictions of Capital in the Twenty-first Century*, chap. II., pp. 87–107. Agenda, Newcastle upon Tyne, UK (2016)
14. Heyde, C.C.: On a property of the lognormal distribution. *J. R. Stat. Soc. Ser. B (Methodological)* **25**(2), 392–393 (1963)
15. Jasso, G.: Anything Lorenz curves can do, top shares can do: assessing the topbot family of inequality measures. *Sociol. Methods Res.* (2018). <https://doi.org/10.1177/0049124118769106>
16. Lambert, P.J., Decoster, A.: The Gini coefficient reveals more. *Metron* pp. 373–400
17. Oancea, B., Pirjol, D.: Extremal properties of the Theil and Gini measures of inequality. *Qual. Quant.* **53**(2), 859–869 (2019)
18. Palma, G.: Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-u': It's all about the share of the rich. *Dev. Change* **42**(1), 87153 (2011)
19. Stoyanov, J.M.: Counterexamples in Probability. Dover Publications, Newburyport (2013)
20. Wikipedia: Gini coefficient. https://en.wikipedia.org/wiki/Gini_coefficient
21. Yitzhaki, S., Schechtman, E.: *The Gini Methodology: A Primer on a Statistical Methodology*. Springer Series in Statistics, vol. 272. Springer, New York (2013)

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