Heliyon 10 (2024) e26482

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

New approach to measuring income inequality

Youngsoon Kim [a], Joongyang Park [a,*], Ae-Jin Ju [b]

[a] Department of Information and Statistics, Gyeongsang National University, 501 Jinju-daero, Jinju-si, Gyeongsangnam-do, 52828, Republic of Korea

[b] Department of Information and Statistics, Graduate School, Gyeongsang National University, 501 Jinju-daero, Jinju-si, Gyeongsangnam-do, 52828, Republic of Korea

ARTICLE INFO

JEL classification:
C43
D31
D63

Keywords:
Cumulative distribution function
Income inequality
Norm
Progressive transfer
Quantile function
Unequally distributed income

ABSTRACT

We show that the conventional income inequality indexes assess income inequality incorrectly because of three problems. The unequally distributed (UD) income-based approach solves the problems, decomposes income inequality into two kinds of departure from equality, and provides two indexes. The comprehensive assessment of income inequality requires the integration of two kinds of departure. This paper proposes the relative UD (RUD) income-based approach. The RUD income-based approach combines the cumulative distribution function and quantile function of the RUD income and produces a new index integrating two kinds of departure. We investigate the properties of the new index and demonstrate its applicability through example income distributions.

1. Introduction

The measurement of income inequality has been an important topic in economics. Since the introduction of the Lorenz curve by Lorenz [1], many indexes have been developed to measure the degree of income inequality. We refer the readers to Hao and Naiman [2], Jenkins and Kerm [3], and Cowell [4] for a general overview of income inequality measurement. Though Hao and Naiman [2, p. 42] and Cowell [4, p. 155] provide lists of income inequality indexes, the list is still expanding. The Palma ratio [5,6] was added by Cobham and Sumner [7]. Gallegati et al. [8] and Clementi et al. [9] added the Zanardi index developed by Zanardi [10] to measure the asymmetry of the Lorenz curve. Henceforth, we refer to these indexes as conventional indexes. We will look into expressions for the conventional indexes in Section 2.

This paper will show that the conventional income inequality indexes assess income inequality incorrectly because of three problems, propose a new approach to measuring income inequality, and develop a new index. This paper is organized as follows. Section 2 presents three problems that make the conventional income inequality indexes incorrect. Section 3 reviews the unequally distributed (UD) income-based approach proposed by Park et al. [11,12]. The UD income-based approach solves the problems and provides two indexes for two kinds of departure from equality. Section 4 discusses the insufficiency of the UD income-based approach and proposes the relative UD (RUD) income-based approach. The RUD income-based approach provides a new index by evaluating the discrepancy between equality and the combination of the cumulative distribution function (CDF) and quantile function (QF) of

* Corresponding author.
E-mail address: joongyang.park@gmail.com (J. Park).

https://doi.org/10.1016/j.heliyon.2024.e26482
Received 11 December 2022; Received in revised form 3 February 2024; Accepted 14 February 2024

Available online 21 February 2024

the RUD income. We investigate the properties of the new index in Section 5 and demonstrate the applicability of the new index through example income distributions in Section 6. Section 7 presents concluding remarks.

2. Problems of conventional indexes

Suppose that $y_1 \le y_2 \le \cdots \le y_n$ are the incomes of n individuals in a population. The income distribution of the population is written as $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. The total income and mean income of the income distribution $\mathbf{y}$ are denoted by $S_y = \sum_{i=1}^{n} y_i$ and $\mu_y = S_y/n$, respectively. As Cowell (2011, p. 1) defined, inequality is a departure from equality [4]. Equality in income inequality is an income distribution in which all individuals in a population have the same income. Such an income distribution is referred to as perfect equality and is denoted by $\mathbf{y}_{pe} = (\mu_y, \ldots, \mu_y)$. Therefore, the income inequality of $\mathbf{y}$ is the departure of $\mathbf{y}$ from $\mathbf{y}_{pe}$.

In this section, we present three problems of the conventional indexes.

2.1. Mixture of information about equality and inequality

All the conventional indexes do not consider that an income distribution includes information about equality and inequality. To assess income inequality, we need to extract information about inequality from the income distribution. Consider, for example, income distribution $\mathbf{y}_{ex} = (1, 2, 3, 4, 5)$, where the total income is 15 and the mean income is 3. Each individual's income is at least 1. That is, 5, one-third of the total income, is equally distributed over five individuals. We can represent this information about equality as $(1, 1, 1, 1, 1)$. Therefore, $\mathbf{y}_{ex}$ decomposes into two distributions $(1, 1, 1, 1, 1)$ and $(0, 1, 2, 3, 4)$. The distribution $(1, 1, 1, 1, 1)$ carries information about equality, while the distribution $(0, 1, 2, 3, 4)$ carries information about inequality. Similarly, perfect equality corresponding to $\mathbf{y}_{ex}$, $(3, 3, 3, 3, 3)$, decomposes into $(3, 3, 3, 3, 3)$ and $(0, 0, 0, 0, 0)$. We should measure the income inequality of $\mathbf{y}_{ex}$ by comparing $(0, 1, 2, 3, 4)$ and $(0, 0, 0, 0, 0)$ along with $\mathbf{y}_{ex}$.

The direct comparison between $\mathbf{y}$ and $\mathbf{y}_{pe}$ without information separation will lead us to the measurement of some mixture of equality and inequality. All the conventional indexes have the information mixture problem. They do not extract information about inequality from the income distribution. Park et al. [11,12] raised the information mixture problem.

2.2. Variation within distribution

According to the definition mentioned above, income inequality is about the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$. However, most conventional indexes measure the variation within $\mathbf{y}$. Such indexes involve the comparison of $\mathbf{y}$ with μ_y such as $(y_i - \mu_y)$, $(y_i - y_j) = (y_i - \mu_y) - (y_j - \mu_y)$, and y_i/μ_y (equivalently y_i/S_y). For example, the most popular Gini coefficient

$$G_y = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| y_i - y_j \right|}{2n^2 \mu_y}$$

involves $(y_i - y_j)$. The coefficient of variation (CV)

$$CV_y = \frac{\sqrt{\sum_{i=1}^{n} (y_i - \mu_y)^2}}{\sqrt{n}\mu_y}$$

and the Pietra index, known as Hoover index, the Robin Hood index, and the Ricci-Schutz index,

$$\frac{\sum_{i=1}^{n} \left| y_i - \mu_y \right|}{n\mu_y}$$

involve $(y_i - \mu_y)$ [13]. The Atkinson index

$$A_\epsilon = 1 - \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i}{\mu_y} \right)^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}},$$

the Theil index

$$T = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\mu_y} \log \left(\frac{y_i}{\mu_y} \right),$$

the generalized entropy index

$$E_\theta = \frac{1}{\theta^2 - \theta} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i}{\mu_y} \right)^{\theta} - 1 \right],$$

Herfindahl index

$$H=\sum_{i=1}^{n}\left(\frac{y_i}{n\mu_y}\right)^2,$$

and the mean log deviation

$$MLD=\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\mu_y}{y_i}\right)$$

involve y_i/μ_y [14–18].

Though μ_y is a representative value of $\mathbf{y}_{pe}$, the representation of $\mathbf{y}_{pe}$ as μ_y accompanies dimension reduction. Due to the dimension reduction, the comparison between $\mathbf{y}$ and μ_y reflects the comparison between $\mathbf{y}$ and $\mathbf{y}_{pe}$ incompletely. Moreover, since μ_y is the mean of $\mathbf{y}$, the comparison of $\mathbf{y}$ with μ_y results in measuring the dispersion of $y_i, i=1,2,\dots,n$. We refer to the dispersion of $y_i, i=1,2,\dots,n$ as the variation within distribution $\mathbf{y}$. The conventional indexes measure the variation within distribution $\mathbf{y}$. Cowell (2011, p. 7) defined an income inequality index as a numerical representation of the interpersonal differences in income within a given population [4]. The conventional indexes are in line with this definition.

We can not measure a departure from equality without equality. We can not measure a departure of $\mathbf{y}$ from $\mathbf{y}_{pe}$ without $\mathbf{y}_{pe}$. We can not measure the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$ without $\mathbf{y}_{pe}$. Inequality of $\mathbf{y}$ is a relationship between $\mathbf{y}$ and $\mathbf{y}_{pe}$. The variation within $\mathbf{y}$ neither require $\mathbf{y}_{pe}$ nor describe a relationship with $\mathbf{y}_{pe}$. The variation within $\mathbf{y}$ has nothing to do with $\mathbf{y}_{pe}$. Therefore, the conventional indexes are not inequality measures. This variation within distribution problem has never been considered before in other literature.

The Gini coefficient intends to measure the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$. The Gini coefficient compares the Lorenz curves for $\mathbf{y}$ and $\mathbf{y}_{pe}$. The Gini coefficient measures the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$ by the area enclosed by the Lorenz curves for $\mathbf{y}$ and $\mathbf{y}_{pe}$. Therefore, the Gini coefficient does not have the variation within distribution problem. However, the Gini coefficient has the information mixture problem and results in the variation within $\mathbf{y}$.

2.3. Negative incomes

One fundamental assumption, which all the conventional indexes rely on, is that income is non-negative. Perfect inequality refers to an income distribution in which one individual takes all the income and each of the rest takes zero income. Perfect inequality depends on the non-negative income assumption. Perfect inequality is used for computing the upper bound of an income inequality index. For example, the Gini coefficient takes a value between zero and one. The upper bound is the Gini coefficient for perfect inequality. The list of conventional indexes in Cowell (2011, p. 155) shows the upper bounds [4].

However, we frequently encounter negative incomes in reality. Park et al. [12,19] analyzed the LIS income datasets of forty-two countries. Negative incomes were observed in twenty-seven countries. Negative incomes are collected when the expense of the self-employed exceeds the revenue, and when the debt repayment of an employee is more than his earnings. Negative incomes can result from accounting conventions, tax laws, and data collection procedures that differ from country to country. A negative income in one country can be positive in another country. Conversely, a positive income in one country can be negative in another country. Therefore, negative incomes are valid values and should be dealt with as they are.

Negative incomes incur problems in computing the conventional indexes. The indexes based on information theory and income shares are neither computable nor interpretable [14–16]. The popular Gini coefficient requires the normalization proposed by Chen et al. [20] and Raffinetti [21]. Usually, the negative or non-positive incomes are adjusted to cope with the problems. Typical adjustments are the deletion of non-positive incomes [2,4] and the replacement of negative incomes with either zero incomes [22] or arbitrarily small positive incomes [23].

The non-negative income assumption does not represent reality. The indexes developed under unrealistic assumptions can not assess income inequality correctly. Equally problematic is that the inconsistency between reality and the assumption is resolved by adjusting the data. Data adjustment is equivalent to fitting the data into a model. We should fit a model to the data.

3. UD income-based approach

Park et al. [11] introduced the UD income to solve the information mixture problem. Park et al. [12] proposed a UD income-based approach to income inequality measurement. The UD income-based approach allows negative income values and assumes that the total income (equivalently, the mean income) is positive. The approach solves the variation within distribution problem by assessing the discrepancy between either the CDFs, the QFs, or the unscaled Lorenz curves for the UD income distribution and perfect equality.

The UD income-based approach begins with the expression $y_i=y_1+\left(y_i-y_1\right),\ i=1,2,\dots n$. We can derive the following from this expression.

(i) ny_1 of S_y is evenly distributed among the n individuals.
(ii) $\left(S_y-ny_1\right)$ is unequally distributed among the n individuals as $x_i=\left(y_i-y_1\right),\ i=1,2,\dots,n$.
(iii) The unequally distributed portions of the incomes, $x_i,\ i=1,2,\dots,n$, contains information about inequality.
(iv) $x_i,\ i=1,2,\dots,n$, are non-negative, and x_1 is zero.

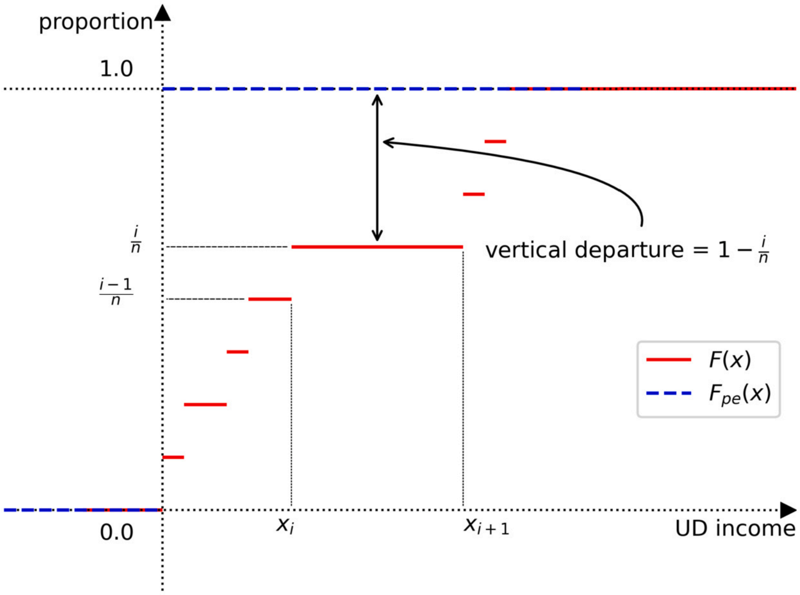

Fig. 1. CDFs of UD income distributions $\mathbf{x}$ and $\mathbf{x}_{pe}$.

x_i, $i = 1, 2, \ldots, n$ are called the UD incomes. We denote the UD income distribution by $\mathbf{x} = \left(x_1, x_2, \ldots, x_n\right)$. The total and mean of the UD incomes are

$$S_x = \left(S_y - n y_1\right) = n\left(\mu_y - y_1\right) \text{ and } \mu_x = \left(\mu_y - y_1\right).$$

Similarly, the UD income distribution of $\mathbf{y}_{pe}$ is obtained as $\mathbf{x}_{pe} = (0, 0, \cdots, 0)$.

We can derive $\mathbf{x}$ and S_y from $\mathbf{y}$. We can restore $\mathbf{y}$ from $\mathbf{x}$ and S_y. Therefore, $\mathbf{y}$ is equivalent to $\mathbf{x}$ and S_y. Similarly, $\mathbf{y}_{pe}$ is equivalent to $\mathbf{x}_{pe}$ and S_y. The UD income-based approach focuses on $\mathbf{x}$, $\mathbf{x}_{pe}$, and S_y (equivalently, μ_y) instead of $\mathbf{y}$ and $\mathbf{y}_{pe}$. The UD income-based approach assesses the discrepancy between $\mathbf{x}$ and $\mathbf{x}_{pe}$ in three ways. The first is to evaluate the discrepancy between the CDFs. The CDFs for $\mathbf{x}$ and $\mathbf{x}_{pe}$ are

$$F(x) = \begin{cases} 0 & \text{for } x < x_1, \\ \frac{i}{n} & \text{for } x_i \le x < x_{i+1}, i = 1, 2, \cdots, n-1, \\ 1 & \text{for } x \ge x_n, \end{cases} \tag{1}$$

and

$$F_{pe}(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \ge 0 \end{cases}, \tag{2}$$

respectively. Mathematically, the CDFs are step functions depicted in Fig. 1. The departure of $F(x)$ from $F_{pe}(x)$ is called the vertical departure. We can measure the magnitude of the vertical departure by ℓ_1 and ℓ_2 norms. Dorfman [24] and Yitzhaki [25] showed that

$$\int_0^{x_n} [1 - F(x)]\, dx = \mu_x \text{ and } \int_0^{x_n} [1 - F(x)]^2\, dx = \mu_x\left(1 - G_x\right),$$

where G_x is the Gini coefficient of $\mathbf{x}$. Therefore, ℓ_1 and ℓ_2 norms of the vertical departure are

$$\ell_1^v = \int_0^{x_n} [1 - F(x)]\, dx = \mu_x \text{ and } \ell_2^v = \left[\int_0^{x_n} [1 - F(x)]^2\, dx\right]^{\frac{1}{2}} = \left[\mu_x\left(1 - G_x\right)\right]^{\frac{1}{2}}.$$

To make ℓ_1^v and ℓ_2^v unitless, we normalize ℓ_1^v and ℓ_2^v by μ_y and $\sqrt{\mu_y}$, respectively. The normalized ℓ_1^v and ℓ_2^v, denoted by $\widetilde{\ell_1^v}$ and $\widetilde{\ell_2^v}$, are

$$\widetilde{\ell_1^v} = \frac{\mu_x}{\mu_y} \text{ and } \widetilde{\ell_2^v} = \left[\frac{\mu_x}{\mu_y}\left(1 - G_x\right)\right]^{\frac{1}{2}} = \left(\frac{\mu_x}{\mu_y} - G_y\right)^{\frac{1}{2}}. \tag{3}$$

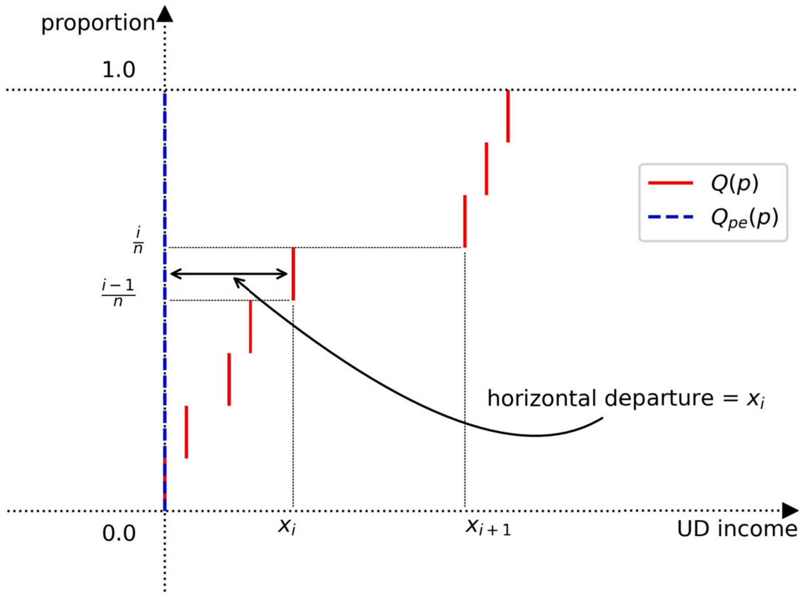

Fig. 2. QFs of UD income distributions $\mathbf{x}$ and $\mathbf{x}_{pe}$.

The second is to evaluate the discrepancy between the QFs. The QFs for $\mathbf{x}$ and $\mathbf{x}_{pe}$ are

$$Q(p) = \inf\{x : F(x) \geq p\} = \begin{cases} 0 & \text{for } p = 0 \\ x_i & \text{for } \frac{i-1}{n} < p \leq \frac{i}{n}, i = 1, 2, \dots, n, \end{cases} \tag{4}$$

and

$$Q_{pe}(p) = 0 \text{ for } 0 \leq p \leq 1. \tag{5}$$

The QFs are also step functions depicted in Fig. 2. The departure of $Q(p)$ from $Q_{pe}(p)$ is called the horizontal departure. Similarly, we can measure the magnitude of the horizontal departure by normalized ℓ_1 and ℓ_2 norms. ℓ_1 and ℓ_2 norms for the horizontal departure are

$$\ell_1^h = \int_0^1 Q(p)dp = \frac{\sum_{i=1}^n x_i}{n} = \mu_x \text{ and } \ell_2^h = \left[\int_0^1 [Q(p)]^2 dp\right]^{\frac{1}{2}} = \left[\frac{\sum_{i=1}^n x_i^2}{n}\right]^{\frac{1}{2}} = \left(\mu_x^2 + V_x\right)^{\frac{1}{2}},$$

where $V_x = \sum_{i=1}^n \left(x_i - \mu_x\right)^2 / n$ is the variance of the UD income. Normalization of ℓ_1^h and ℓ_2^h by μ_y results in

$$\widetilde{\ell_1^h} = \widetilde{\ell_1^v} \text{ and } \widetilde{\ell_2^h} = \frac{\mu_x}{\mu_y}\left(1 + CV_x^2\right)^{\frac{1}{2}} = \left[\left(\frac{\mu_x}{\mu_y}\right)^2 + CV_y^2\right]^{\frac{1}{2}}. \tag{6}$$

The third is to evaluate the discrepancy between the unscaled Lorenz curves. Park et al. [12] showed that the normalized ℓ_1 norm of the discrepancy between the unscaled Lorenz curves of $\mathbf{x}$ and $\mathbf{x}_{pe}$ is equivalent to $\widetilde{\ell_2^h}$. It is because the Lorenz curve derives from the QF. $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ were proposed as income inequality index.

The noteworthy points of the UD income-based approach are as follows:

(i) Income inequality breaks down into two kinds of departure from equality. The kind of departure depends on how to represent a distribution.
(ii) $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ consist of two components. One is μ_x/μ_y associated with the locations of $\mathbf{y}$ and $\mathbf{x}$. The other is either G_y or CV_y associated the variation within $\mathbf{y}$. The variation within $\mathbf{y}$ is a part of income inequality.
(iii) It is well-known that the progressive transfer preserving μ_y reduces both G_y and CV_y. Thus, such a transfer has been considered an effective method to improve income inequality. The progressive transfer preserving μ_x is such a transfer. If the poorest is not the beneficiary of the transfer, μ_x does not change. Equations (3) and (6) show that the progressive transfer preserving μ_x reduces the horizontal departure, but increases the vertical departure. Consequently, the progressive transfer preserving μ_y does not guarantee the improvement of income inequality.

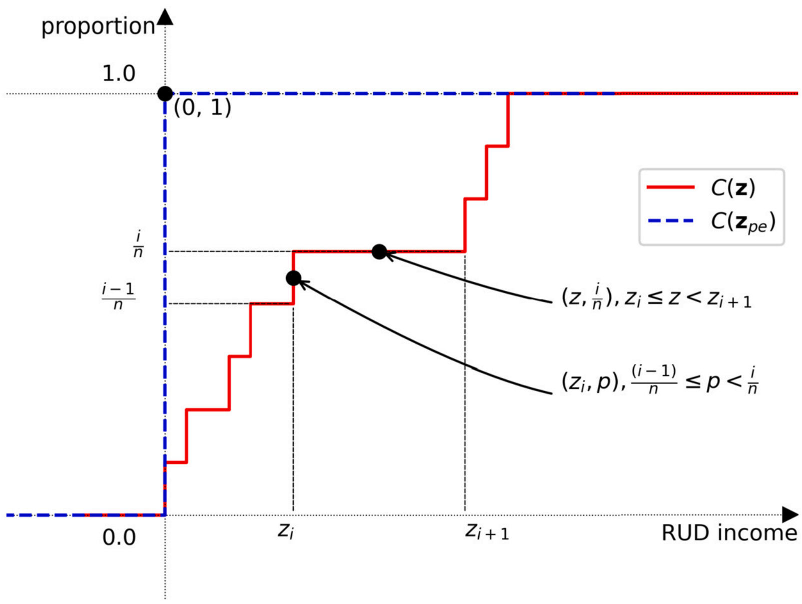

Fig. 3. Combinations of CDFs and QFs of RUD income distributions $\mathbf{z}$ and $\mathbf{z}_{pe}$.

We will discuss the effect of progressive transfers further in Section 5.

4. RUD income-based approach

The UD income-based approach reviewed in the previous section solves the problems mentioned in Section 2. However, it provides two indexes for two kinds of departure from perfect equality. It is difficult to evaluate income inequality comprehensively. We need to integrate two kinds of departure. The vertical departure is unitless, while the horizontal departure is a monetary unit. Therefore, the integration of the two kinds of departure is not straightforward. We thus introduce the RUD income which is the UD income relative to the mean income and defined as $z_i = x_i/\mu_y$. The RUD income distribution for $\mathbf{y}$ and the corresponding perfect equality are denoted by $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ and $\mathbf{z}_{pe} = (0, 0, \ldots, 0)$. The CDFs and QFs for $\mathbf{z}$ and $\mathbf{z}_{pe}$ are defined as Equations (1), (2), (4), and (5) with x replaced with z. If we replace x by z in Figs. 1 and 2, we have the graphs for the CDFs and QFs for $\mathbf{z}$ and $\mathbf{z}_{pe}$.

To take both kinds of departure into income inequality measurement, we combine the CDF and QF. The combination of the CDF and QF is the superimposition of the CDF and the QF. We denote the combinations for $\mathbf{z}$ and $\mathbf{z}_{pe}$ by $C(\mathbf{z})$ and $C(\mathbf{z}_{pe})$, which are depicted in Fig. 3.

The coordinates of an arbitrary point on $C(\mathbf{z})$ are (z, p). Since $C(\mathbf{z}_{pe})$ provides the vertical and horizontal baselines, we can assess the vertical and horizontal departures of (z, p). The horizontal and vertical departures of (z, p) are z and $(1-p)$, respectively. Note that $(0, 1)$ on $C(\mathbf{z}_{pe})$ corresponds to perfect equality. The horizontal and vertical departures of (z, p) are integrated into $(-z, 1-p)$ that is the discrepancy of (z, p) from $(0, 1)$. If (z, p) is on the step height of $C(\mathbf{z})$, ℓ_1 and squared ℓ_2 norms of $(-z, 1-p)$ are $[Q(p) + (1-p)]$ and $\left[Q(p)^2 + (1-p)^2\right]$. If (z, p) is on the step width of $C(\mathbf{z})$, ℓ_1 and squared ℓ_2 norms of $(-z, 1-p)$ are $[z + (1 - F(z))]$ and $\left[z^2 + (1 - F(z))^2\right]$. Therefore, ℓ_1 and ℓ_2 norms of the discrepancy between $C(\mathbf{z})$ and $(0, 1)$ are obtained as

$$\begin{aligned}
\ell_1(\mathbf{z}) &= \int_0^{z_n} [z + (1 - F(z))]\, dz + \int_0^1 [Q(p) + (1-p)]\, dp \\
&= \int_0^{z_n} z\, dz + \int_0^{z_n} [1 - F(z)]\, dz + \int_0^1 Q(p)\, dp + \int_0^1 (1-p)\, dp \\
&= \frac{1}{2} z_n^2 + \mu_z + \mu_z + \frac{1}{2} \\
&= 2\mu_z + \frac{1}{2}\left(1 + z_n^2\right)
\end{aligned} \tag{7}$$

and

$$\ell_2(\mathbf{z}) = \left[\int_0^{z_n} \left[z^2 + (1 - F(z))^2\right] dz + \int_0^1 \left[Q(p)^2 + (1-p)^2\right] dp\right]^{\frac{1}{2}}$$

$$
\begin{aligned}
&= \left[\int_0^{z_n} z^2 dz + \int_0^{z_n} [1 - F(z)]^2 dz + \int_0^1 Q(p)^2 dp + \int_0^1 (1-p)^2 dp \right]^{\frac{1}{2}} \\
&= \left[\frac{1}{3} z_n^3 + \mu_z \left(1 - G_z\right) + \mu_z^2 \left(1 + CV_z^2\right) + \frac{1}{3} \right]^{\frac{1}{2}} \\
&= \left[\mu_z \left(1 - G_z\right) + \mu_z^2 \left(1 + CV_z^2\right) + \frac{1}{3} \left(1 + z_n^3\right) \right]^{\frac{1}{2}},
\end{aligned}
\tag{8}
$$

where μ_z, G_z, and CV_z are the mean, Gini coefficient, and CV of $\mathbf{z}$, respectively. Next we compute ℓ_1 and ℓ_2 norms of the discrepancy between $C\left(\mathbf{z}_{pe}\right)$ and $(0,1)$. A point on the height of $C\left(\mathbf{z}_{pe}\right)$ is $(0,p)$ for $0 \le p \le 1$, while a point on the width of $C\left(\mathbf{z}_{pe}\right)$ is $(z,1)$ for $0 \le z \le z_n$. Their discrepancies from $(0,1)$ are $(0,1-p)$ and $(z,0)$, respectively. ℓ_1 and squared ℓ_2 norms of $(0,1-p)$ are $(1-p)$ and $(1-p)^2$. ℓ_1 and squared ℓ_2 norms of $(z,0)$ are z and z^2. Therefore, ℓ_1 and ℓ_2 norms of the discrepancy between $C\left(\mathbf{z}_{pe}\right)$ and $(0,1)$ are obtained as

$$
\ell_1\left(\mathbf{z}_{pe}\right) = \int_0^{z_n} z dz + \int_0^1 (1-p) dp = \frac{1}{2}\left(1 + z_n^2\right) \tag{9}
$$

and

$$
\ell_2\left(\mathbf{z}_{pe}\right) = \left[\int_0^{z_n} z^2 dz + \int_0^1 (1-p)^2 dp \right]^{\frac{1}{2}} = \left[\frac{1}{3}\left(1 + z_n^3\right) \right]^{\frac{1}{2}}. \tag{10}
$$

We derive from Equations (7)-(10) two indexes, L_1 and L_2, which are

$$
L_1 = \ell_1(\mathbf{z}) - \ell_1\left(\mathbf{z}_{pe}\right) = 2\mu_z = \widetilde{\ell_1^v} + \widetilde{\ell_1^h}
$$

and

$$
L_2 = \ell_2(\mathbf{z}) - \ell_2\left(\mathbf{z}_{pe}\right) = \left[\left(\widetilde{\ell_2^v}\right)^2 + \left(\widetilde{\ell_2^h}\right)^2 + \left(\ell_2\left(\mathbf{z}_{pe}\right)\right)^2 \right]^{\frac{1}{2}} - \ell_2\left(\mathbf{z}_{pe}\right).
$$

Since $\mu_z G_z = G_y$ and $\mu_z CV_z = CV_y$, we can write L_2 as

$$
\begin{aligned}
L_2 = &\left[\left(\frac{\mu_y - y_1}{\mu_y} \right) - G_y + \left(\frac{\mu_y - y_1}{\mu_y} \right)^2 + CV_y^2 + \frac{1}{3}\left(1 + \left(\frac{y_n - y_1}{\mu_y} \right)^3 \right) \right]^{\frac{1}{2}} \\
&- \left[\frac{1}{3}\left(1 + \left(\frac{y_n - y_1}{\mu_y} \right)^3 \right) \right]^{\frac{1}{2}}.
\end{aligned}
$$

L_1 and L_2 integrate two kinds of departure. L_1 and L_2 are greater than or equal to zero. L_1 and L_2 are zero for perfect equality. We can think of more unequal income distributions for any income distribution because negative income values are allowed. Therefore, the L_1 and L_2 indexes are not bounded above. μ_z is the total UD income ratio relative to the total income. We can interpret L_1 as the average distance between $\mathbf{z}$ and $\mathbf{z}_{pe}$. The dispersion measures such as G_z, CV_z, and z_n assess the variation within $\mathbf{z}$, that is, how unevenly the total RUD income is distributed. L_2 integrates the average distance between $\mathbf{z}$ and $\mathbf{z}_{pe}$ and the variation within $\mathbf{z}$. We propose L_2 as an income inequality index.

5. Properties of L_2 index

We described in the previous section the basic properties and interpretation of the L_2 index. Many studies commonly mention that scale invariance, replication invariance, the anonymity axiom, and the principle of progressive transfers are the desirable properties of an income inequality index [2–4,26–28]. As explained in Subsection 2.2, such studies consider the variation within $\mathbf{y}$. Ebert [27, p. 366] mentions that these properties are about inequality within a population. In general, scale invariance, replication invariance, and the anonymity axiom are also desirable properties for the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$. However, we should investigate whether the principle of progressive transfers is a desirable property for an income inequality index. Because a decrease of the variation within $\mathbf{y}$ by progressive transfers does not mean a decrease of the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$. In this section, we examine the L_2 index for these properties.

Multiplication of $\mathbf{y}$ by some $\alpha > 0$ results in a new income distribution $\mathbf{y}_m = \alpha\mathbf{y}$. Then $\mathbf{x}_m = \alpha\mathbf{x}$ and $S_{y_m} = \alpha S_y$, where S_{y_m} is the total income for $\mathbf{y}_m$. Therefore, $\mathbf{z}_m = \mathbf{x}_m / S_{y_m} = \mathbf{z}$, $C\left(\mathbf{z}_m\right)$ equals $C(\mathbf{z})$, and L_2s for $\mathbf{y}_m$ and $\mathbf{y}$ are the same. The L_2 index is scale-invariant.

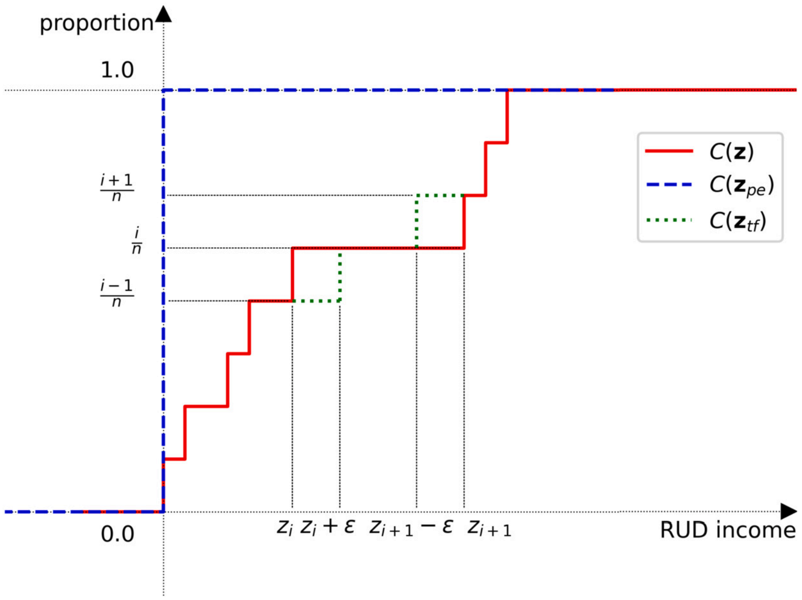

Fig. 4. Difference between $C(\mathbf{z})$ and $C(\mathbf{z}_{tf})$.

Let $\mathbf{y}_r = (y_1, y_1, y_2, y_2, ..., y_n, y_n)$ be an income distribution induced by replication of $\mathbf{y}$. Then $\mathbf{y}$ and $\mathbf{y}_r$ have the same mean and minimum. The RUD income distribution $\mathbf{z}_r$ for $\mathbf{y}_r$ is simply replication of $\mathbf{z}$. Therefore, $C(\mathbf{z}_r)$ equals $C(\mathbf{z})$, and L_2s for $\mathbf{y}_r$ and $\mathbf{y}$ are the same. The L_2 index is replication-invariant.

Suppose that $\mathbf{y}_p$ is an income distribution obtained by permutating y_is in $\mathbf{y}$. Permutation does not change the mean and minimum. Therefore, $\mathbf{z}_p = \mathbf{z}$, and L_2s for $\mathbf{y}_p$ and $\mathbf{y}$ are the same. The L_2 index satisfies the anonymity axiom.

The principle of progressive transfers is that some income transfer from a high-income person to a low-income person reduces inequality, provided that the total income and the post-transfer ranking remain the same. Consider an income distribution $\mathbf{y}$ with $y_1 = 0$. Then $\mathbf{y} = \mathbf{x}$ and $\mathbf{z} = \mathbf{y}/\mu_y$. Suppose that $\mu_y\epsilon$ is transferred from y_{i+1} to y_i and $(y_i + \mu_y\epsilon) < (y_{i+1} - \mu_y\epsilon)$ for some $\epsilon > 0$. Denote the resulting income distribution by $\mathbf{y}_{tf} = (y_1, y_2, ..., y_i + \mu_y\epsilon, y_{i+1} - \mu_y\epsilon, ..., y_n)$. This progressive transfer preserves the mean, minimum, and ranking. The corresponding RUD income distribution for $\mathbf{y}_{tf}$ is $\mathbf{z}_{tf} = \mathbf{y}_{tf}/\mu_y = (z_1, z_2, ..., z_i + \epsilon, z_{i+1} - \epsilon, ..., z_n)$. Then, $C(\mathbf{z})$ and $C(\mathbf{z}_{tf})$ are different only for $(i-1)/n \le p \le (i+1)/n$ and $z_i \le z \le z_{i+1}$ as shown in Fig. 4. We can derive the following from Fig. 4.

(i) The horizontal departure increases from z_i to $(z_i + \epsilon)$ for $(i-1)/n \le p \le i/n$, while it decreases from z_{i+1} to $(z_{i+1} - \epsilon)$ for $i/n \le p \le (i+1)/n$. The amount increased, ϵ, equals the amount decreased. Intervals, $(i-1)/n \le p \le i/n$ and $i/n \le p \le (i+1)/n$, have the same length.
(ii) The vertical departure increases from $(1 - i/n)$ to $(1 - (i-1)/n)$ for $z_i \le z \le (z_i + \epsilon)$, while it decreases from $(1 - i/n)$ to $(1 - (i+1)/n)$ for $(z_{i+1} - \epsilon) \le z \le z_{i+1}$. The amount increased, $1/n$, equals the amount decreased. Intervals, $z_i \le z \le (z_i + \epsilon)$ and $(z_{i+1} - \epsilon) \le z \le z_{i+1}$, have the same length.

The departure increase for an interval accompanies the departure decrease for another interval of the same length. The amount increased equals the amount decreased. Therefore, it is not evident that progressive transfers reduce the departure from perfect equality. The effect of progressive transfers depends on y_i, y_{i+1}, ϵ, and how to integrate the vertical and horizontal departures. If we integrate by the ℓ_1 norm, $C(\mathbf{z})$ and $C(\mathbf{z}_{tf})$ have the same departure from $C(\mathbf{z}_{pe})$. It is because $\mathbf{z}$ and $\mathbf{z}_{tf}$ have the same mean. Consequently, $\mathbf{y}$ and $\mathbf{y}_{tf}$ have the same income inequality. If we integrate by the ℓ_2 norm, the departure of $C(\mathbf{z}_{tf})$ from $C(\mathbf{z}_{pe})$ can be larger or smaller than the departure of $C(\mathbf{z})$ from $C(\mathbf{z}_{pe})$. We will present in the next section several progressive transfers showing different effects on the departure of $C(\mathbf{z})$ from $C(\mathbf{z}_{pe})$. Progressive transfers can fail to reduce the discrepancy between $\mathbf{y}$ and $\mathbf{y}_{pe}$. The principle of progressive transfers does not apply to the RUD income-based approach.

6. Application of L_2 index

This section applies the proposed L_2 index to the example income distributions listed in Table 1. We present G_y, CV_y, L_2, and its components μ_z, z_n, $\widetilde{\ell_2^v}$, and $\widetilde{\ell_2^h}$ in Table 1. [12] used the first eight income distributions $\mathbf{y}_i$, $i = 1, 2, \ldots, 8$ to demonstrate the deficiencies of G_y and CV_y and the applicability of $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$. We show the insufficiency of $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ and the applicability of L_2. We include an additional income distribution $\mathbf{y}_9$. All the income distributions except $\mathbf{y}_8$ have $n = 5$, $S_y = 15$, and $\mu_y = 3$. Each of the

Table 1
Example income distributions.

$\mathbf{y}$	G_y	CV_y	μ_z	z_n	$\widetilde{\ell_2^v}$	$\widetilde{\ell_2^h}$	L_2
$\mathbf{y}_1 = (1,2,2,5,5)$	0.2933	0.5578	0.6667	1.3333	0.6110	0.8692	0.4408
$\mathbf{y}_2 = (0,3,3,4,5)$	0.2933	0.5578	1.0000	1.6667	0.8406	1.1450	0.6035
$\mathbf{y}_3 = (2,2,2,4,5)$	0.2133	0.4216	0.3333	1.0000	0.3464	0.5375	0.2206
$\mathbf{y}_4 = (1,3,3,3,5)$	0.2133	0.4216	0.6667	1.3333	0.6733	0.7888	0.4230
$\mathbf{y}_5 = (1,3,3,4,4)$	0.1867	0.3651	0.6667	1.0000	0.6928	0.7601	0.4967
$\mathbf{y}_6 = (2,3,3,3,4)$	0.1067	0.2108	0.3333	0.6667	0.4761	0.3944	0.2451
$\mathbf{y}_7 = (-1,4,4,4,4)$	0.2667	0.6667	1.3333	1.6667	1.0328	1.4907	0.9029
$\mathbf{y}_8 = (0,4,4,4,4)$	0.2000	0.5000	1.0000	1.2500	0.8944	1.1180	0.7498
$\mathbf{y}_9 = (1,2,3,4,5)$	0.2667	0.4714	0.6667	1.3333	0.6325	0.8165	0.4200

distributions derives from others by a series of transfers. Using these income distributions, we investigate the effect of progressive transfer. $\mathbf{y}_7$ has a negative income value. Replacing the negative income with zero, we obtain $\mathbf{y}_8$ from $\mathbf{y}_7$.

According to G_y and CV_y, the income inequality of $\mathbf{y}_1$ is the same as $\mathbf{y}_2$. G_y and CV_y also evaluate the income inequality of $\mathbf{y}_3$ and $\mathbf{y}_4$ the same. G_y and CV_y fail to differentiate these two sets of income distributions concerning income inequality. G_y also fails to differentiate $\mathbf{y}_7$ and $\mathbf{y}_9$. Besides the three problems in Section 2, G_y and CV_y are not sensitive to the change in distribution. By contrast, the L_2 index is sensitive to the change in distribution enough to successfully differentiate these income distributions. G_y and CV_y measure the variation within income distribution. Two income distributions with the same within variation can have different distances from equality, consequently different income inequalities. Since $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ for $\mathbf{y}_1$ are smaller than $\mathbf{y}_2$, we can say that $\mathbf{y}_1$ is less unequal than $\mathbf{y}_2$. Considering $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$, we can differentiate between $\mathbf{y}_1$ and $\mathbf{y}_2$, between $\mathbf{y}_3$ and $\mathbf{y}_4$, and between $\mathbf{y}_7$ and $\mathbf{y}_9$. Note that L_2 gives the same comparison result with the simultaneous use of $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$.

Next, we make an inequality comparison between $\mathbf{y}_1$ and $\mathbf{y}_4$. $\widetilde{\ell_2^v}$ says that $\mathbf{y}_1$ is less unequal than $\mathbf{y}_4$, while $\widetilde{\ell_2^h}$ says that $\mathbf{y}_1$ is more unequal than $\mathbf{y}_4$. The comparison by $\widetilde{\ell_2^v}$ conflicts with the comparison by $\widetilde{\ell_2^h}$. We observe similar conflicting results when comparing $\mathbf{y}_3$ with $\mathbf{y}_6$, and $\mathbf{y}_5$ with $\mathbf{y}_9$. Conflicting results can happen because $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ measure different kinds of departure from equality. $\widetilde{\ell_2^v}$ and $\widetilde{\ell_2^h}$ are not sufficient for the comprehensive income inequality comparison. According to the L_2 index, $\mathbf{y}_4$ ($\mathbf{y}_3$, $\mathbf{y}_9$) is less unequal than $\mathbf{y}_1$ ($\mathbf{y}_6$, $\mathbf{y}_5$).

We can compute G_y and CV_y when there are negative income values. However, negative income values are usually adjusted to meet the non-negative income assumption. If we delete a negative value in $\mathbf{y}_7$, $\mathbf{y}_7$ becomes perfect equality. L_2 says that $\mathbf{y}_7$ is the most unequal among nine distributions in Table 1. If we replace a negative value in $\mathbf{y}_7$ with zero, we have $\mathbf{y}_8$. Comparing L_2 and its components for $\mathbf{y}_7$ and $\mathbf{y}_8$, we see that such adjustment incurs the underestimation of income inequality.

Finally, we investigate the effect of progressive transfer in the sense of L_2. $\mathbf{y}_5$ derives from $\mathbf{y}_4$ by the transfer from the richest to the second richest. This transfer decreases $\widetilde{\ell_2^h}$ and z_n, increases $\widetilde{\ell_2^v}$, and does not change μ_z. L_2 says that the overall income inequality increases. It is noteworthy that the transfer between rich individuals can worsen income inequality. $\mathbf{y}_9$ derives from $\mathbf{y}_1$ by the transfer between the middle class. This transfer does not change μ_z and z_n, increases $\widetilde{\ell_2^v}$, decreases $\widetilde{\ell_2^h}$, and decreases L_2. $\mathbf{y}_4$ derives from $\mathbf{y}_9$ by the transfer between the middle class. These transfers do not change μ_z and z_n, increases $\widetilde{\ell_2^v}$, decreases $\widetilde{\ell_2^h}$, and increases L_2. $\mathbf{y}_6$ derives from $\mathbf{y}_4$ by the transfer from the richest to the poorest. $\mathbf{y}_3$ derives from $\mathbf{y}_9$ by the transfer from the middle class to the poorest. The transfer involving the poorest decreases L_2 and all of its components. In summary, progressive transfers can fail to improve income inequality. However, we should note that the progressive transfer is essential for improving income inequality. We need to examine whether the progressive transfer improves income inequality before application.

7. Conclusions

An important topic in economics is the measurement of income inequality. We showed that the conventional income inequality indexes assessed income inequality incorrectly because of three problems presented in Section 2. We raised the variation within distribution problem for the first time. The conventional indexes measure the variation within the given income distribution. By contrast, the UD income-based approaches intend to measure the discrepancy between two distributions. The UD income-based approach first extracts information about inequality by deriving UD income distributions for the given income distribution and perfect equality. Then it focuses on the discrepancy between two UD income distributions. The UD income-based approach compares the CDFs and QFs of the UD income distributions. It breaks down income inequality into two kinds of departure from equality and provides two indexes. It is unsuccessful in integrating two kinds of departure and assessing income inequality comprehensively. We proposed the RUD income-based approach, developed the L_2 index, and examined the properties of the L_2 index. We demonstrated the applicability of the L_2 index and the failure of progressive transfers to improve income inequality.

To apply the L_2 index in practice, we need an estimator of the L_2 index. This study does not provide an estimator of the L_2 index. Some components of the L_2 index such as the minimum and the range are not easy to estimate. The development of a good estimator can be a challenging task. In addition, we need to identify the characteristics of the progressive transfers improving income inequality. The characteristics will help formulate policies for improving inequality.

CRediT authorship contribution statement

Youngsoon Kim: Writing – original draft, Visualization, Methodology. **Joongyang Park:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Ae-Jin Ju:** Visualization, Software.

Declaration of competing 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.

Data availability

No datasets were created or analyzed in this study.

References

[1] O. Lorenz, Methods of measuring the concentration of wealth, Publ. Am. Stat. Assoc. 9 (1905) 209–2019.
[2] L. Hao, D.Q. Naiman, Assessing Inequality, Sage Publications, Inc., 2010.
[3] S.P. Jenkins, P.V. Kerm, The measurement of economic inequality, in: B. Nolan, W. Salverda, T.M. Smeeding (Eds.), The Oxford Handbook of Economic Inequality, Oxford University Press, New York, 2011, pp. 40–67.
[4] F.A. Cowell, Measuring Inequality, 3rd ed., Oxford University Press, New York, 2011.
[5] J.G. Palma, Globalizing Inequality: Centrifugal and centripetal forces at work, DESA Working Paper 35, UN Department of Economic and Social Affairs, New York, 2006.
[6] J.G. Palma, Homogeneous middles vs. heterogeneous tails, and the end of the inverted-U: the share of the rich is what it's all about, Cambridge Working Papers in Economics 1111, University of Cambridge Department of Economics, Cambridge, 2011, later published in Dev. Change 42 (1) (2011) 87–153.
[7] A. Cobham, A. Sumner, Is it all about the tails? The Palma measure of income inequality, Working Paper 343, Center for Global Development (CGD), Washington DC, 2013.
[8] M. Gallegati, S. Landini, J.E. Stiglitz, The inequality multiplier, Research paper 16–29, Columbia Business School, New York, 2016.
[9] F. Clementi, M. Gallegati, L. Gianmoena, S. Landini, J.E. Stiglitz, Mis-measurement of inequality: a critical reflection and new insights, J. Econ. Interact. Coord. 14 (2019) 891–921.
[10] G. Zanardi, Della asimmetria condizionata delle curve di concentrazione. Lo scentramento, Riv. Ital. Econ. Demogr. Stat. 18 (1964) 431–466.
[11] J. Park, Y. Kim, S. Heo, Dual-index measurement of income inequality, Bull. Econ. Res. 70 (3) (2018) 277–284.
[12] J. Park, Y. Kim, A. Ju, Measuring income inequality based on unequally distributed income, J. Econ. Interact. Coord. 16 (2021) 309–322.
[13] R.R. Schutz, On the measurement of income inequality, Am. Econ. Rev. 41 (1951) 107–122.
[14] A.B. Atkinson, On the measurement of inequality, J. Econ. Theory 2 (1970) 244–263.
[15] H. Theil, Economics and Information Theory, North-Holland, Amsterdam, 1967.
[16] A.F. Shorrocks, The class of additively decomposable inequality measures, Econometrica 48 (1980) 613–625.
[17] O. Herfindahl, Concentration in the steel industry, Dissertation, Columbia University, New York, 1950.
[18] J. Haughton, S.R. Kandker, The Handbook on Poverty and Inequality, The World Bank, Washington, DC, 2009.
[19] Luxembourg Income Study (LIS) Database, LIS, http://www.lisdatacenter.org.
[20] C.N. Chen, T.W. Tsau, T.S. Rhab, The Gini coefficient and negative income, Oxf. Econ. Pap. 34 (1982) 473–478.
[21] E. Raffinetti, E. Siletti, A. Vernizzi, On the Gini coefficient normalization when incomes with negative values are considered, Stat. Methods Appl. 24 (3) (2015) 507–521.
[22] OECD, OECD framework for statistics on the distribution of household income, consumption and wealth, available at https://doi.org/10.1787/9789264194830-en, 2013.
[23] L.G. Bellù, P. Liberati, Describing income inequality: Theil index and entropy class indices, available at http://www.fao.org/docs/up/easypol/445/theil_index_051en.pdf, 2006.
[24] R. Dorfman, A formula for Gini coefficient, Rev. Econ. Stat. 61 (1979) 146–149.
[25] S. Yitzhaki, More than a dozen alternative ways of spelling Gini, Res. Econ. Inequal. 8 (1998) 13–30.
[26] J.E. Foster, An axiomatic characterization of the Theil measure of income inequality, J. Econ. Theory 31 (1983) 14–16.
[27] U. Ebert, A family of aggregative compromise inequality measures, Int. Econ. Rev. 29 (2) (1988) 363–376.
[28] L.A. Litchfield, Inequality: Methods and Tools, World Bank, Washington DC, 1999, available at http://siteresources.worldbank.org/INTPGI/Resources/Inequality/litchfie.pdf.