Article

Not peer-reviewed version

Some Notes on the Gini Index and New Inequality Measures: The nth Gini Index

José Manuel Gavilan Ruiz , África Ruiz-Gándara * , Francisco Javier Ortega Irizo , Luis Gonzalez Abril

Posted Date: 2 October 2024

doi: 10.20944/preprints202410.0088.v1

Keywords: inequality measures; mean difference; order statistics; rank-dependent inequality measurement

 doi:10.20944/preprints202410.0088.v1

Article

Some Notes on the Gini Index and New Inequality Measures: The nth Gini Index

José Manuel Gavilan-Ruiz, África Ruiz-Gándara *, Francisco Javier Ortega-Irizo and Luis Gonzalez-Abril

Dept. Applied Economic I, Faculty of Sciences of Economics and Business, University of Seville, Avda. Ramón y Cajal, 1, Sevilla E-41018, Spain

* Correspondence: africaruiz@us.es

Abstract: A new family of inequality indices based on the deviation between the expected maximum and the expected minimum of random samples, called the nth Gini index is presented. These indices generalize the Gini index. At the same time, this family of indices and the S-Gini index are generalised by proposing the $uv-$Gini index, which turns out to be a convex combination of the S-Gini index and the Lorenz family of inequality measures. This family of Gini indices provides a methodology for achieving perfect equality in a given distribution of incomes. This is achieved through a series of successive and equal increases in the incomes of each individual.

Keywords: inequality measures; mean difference; order statistics; rank-dependent inequality measurement

1. Introduction

The Gini index [1,2] is one of the principal inequality measures. It could be argued that it is the most well-known and widely accepted quantitative gauge in the field of economics and social sciences, used with the aim of quantifying income and wealth inequality [3–6]. A significant form of injustice is the unequal access to economic resources [7]. In this context, Gini coefficients are used as a primary summary measure of inequality by many government agencies and policy makers [8]. More generally, the Gini index is a quantitative measure of statistical evenness in the contex of non-negative datasets with positive means [9–13]. The Gini coefficient has been employed in a multitude of disciplines beyond its traditional applications in socioeconomics, where it is used to quantify inequality in wealth distributions. Its use has expanded to encompass a diverse range of scientific fields, including gender parity, access to education and health services, and environmental regulations [14]. The extant literature illustrates the extensive range of applications, with recent examples including: agriculture [15], anthropology [16], astrophysics [17], biomedical engineering [18], computational chemistry [19], criminology [20], ecology [21,22], econophysics [23], environmental sciences [24], epidemiology and public health [25,26],finance [27,28], geosciences and remote sensing [29], materials science and surface engineering [30], medical chemistry [31], molecular biology and genetics [32], population biology [33], sustainability science [34], and transportation [35].

Eliazar and Sokolov [36] presents an impressive toolbox of quantitative measures of societal egalitarianism. There is a plethora of inequality measures [37], and Charles *et al.* [38] points out that only two, the Gini index and Theil [39], satisfy the five most sought-after characteristics or properties, including the well-known transfer principle [40], the Gini index being more sensitive to income transfers towards the middle of the distribution [41]. In addition, the Gini index satisfies other essential conditions often imposed on any good poverty index [42].

The attractiveness of the Gini index often relies on the fact that it has an intuitive geometric interpretation, that is, it can be defined geometrically as the area between the line of perfect equality (the 45-degree line in the unit box) and the Lorenz curve multiplied by two. The Gini index is also interpreted in the context of interpersonal comparisons as the average gain to be expected, expressed as a proportion of the average level of income, if each member of the population is allowed to have the best income of either their own income or the income of another member of the population drawn

 doi:10.20944/preprints202410.0088.v1

at random [43]. Furthermore, this index is an important component of the Sen index of poverty intensity [44].

Regarding the computability of the Gini index, it has received some criticism: in Osberg [45] it is determined mathematically as the average of the absolute value of the relative mean difference in income between all possible pairs of individuals, in Liao [46] it is obtained directly from the Lorenz curve, Furman *et al.* [47] propose an alternative expression and interpretation of the Gini index based on the concept of a size-biased distribution, in [48] eight possible formulations of the Gini index are presented.

A profound connection between the Gini index and extreme value statistics is presented in [49]. The recent review by Eliazar [50] of the Gini coefficient as an elegant mathematical object sheds light on the Gini coefficient from a multitude of perspectives, culminating in a profound comprehension of the Gini coefficient that may potentially give rise to innovative applications in the realms of science and engineering. Multivariate extensions of the Gini index can be found in Lunetta1972 [51] and Taguchi [52], who proposed a two-dimensional version, while Koshevoy and Mosler [53], Arnold [54] and Grothe *et al.* [55] extended it to more dimensions.

There are generally two different approaches to the analysis of the theoretical results of the Gini index: one is based on discrete distributions; the other on continuous distributions. These two approaches can be unified [56,57], but for certain purposes the continuous formulation, which is the one that will be followed in this paper, is more convenient, since it yields insights that are not easily accessible when the considered random variable is discrete [58].

One of the major drawbacks of using the Gini index is that two distributions that behave differently in terms of their concentration (for example, through their Lorenz curves) can have the same value of this index. An illustrative example can be seen in Figure 1, which shows the Lorenz curves of two distributions, which will be revisited in Examples 3.5 and 3.6, that have the same Gini index. In these types of situations it is not possible to declare which distribution has a greater concentration using the Gini index, neither is it possible using the Lorenz curve, since neither dominates the other, but they do intersect. This problem has already been addressed in the literature on stochastic dominance [59] and inverse stochastic dominance [60]. Furthermore, the use of the square of the coefficient of variation is proposed in Gonzalez-Abril *et al.* [61], since it can be considered as an inequality measure similar to the Gini index. Liu and Gastwirth [8] mention that it is unrealistic to expect a single measure to describe the behaviour of an entire distribution with respect to inequality.

A whole family of inequality indices, called the nth Gini index, is proposed in this paper to overcome this situation, as opposed to the approach of combining the Gini index with another measure, as for example in Foster and Wolfson [62].

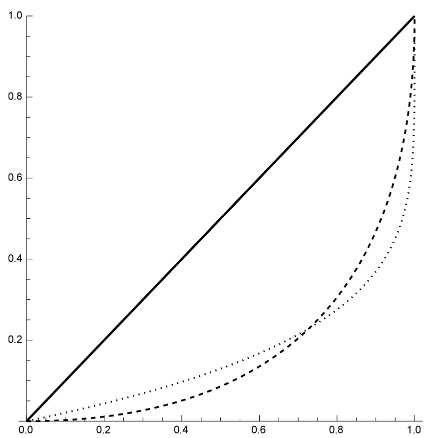

Figure 1. Lorenz curves of two distributions that have the same Gini index. The dashed line corresponds to a Singh-Maddala model and the dotted line to a Pareto model.

 doi:10.20944/preprints202410.0088.v1

The rest of the paper is organized as follows. The Gini index is studied in Section 2. The generalizations of the Gini index are given in Section 3. Conclusions are drawn in the final section.

2. The Gini Index

Let $X \geq 0$ be a non-negative random variable with cumulative distribution function (cdf) $F(x)$ and probability density function $f(x)$, where $\mu = E[X] > 0$ exists.

One of the most widely used definitions of the Gini index of the random variable X is [63]

$$IG(X) = \frac{E[|X_1 - X_2|]}{2\mu}$$

where X_1, X_2 are independent and identically distributed random variables with the same distribution as X. $E[|X_1 - X_2|]$ is the mean difference and is denoted as Δ (that is, the Gini index is the relative mean difference). Nevertheless, $|X_1 - X_2| = \max\{X_1, X_2\} - \min\{X_1, X_2\}$ and hence

$$IG(X) = \frac{E[\max\{X_1, X_2\} - \min\{X_1, X_2\}]}{2\mu} = \frac{1}{\mu}\int_0^\infty F(x)(1 - F(x))dx \tag{1}$$

This result, which is also obtained in Section 3, was first given in Yitzhaki [64] when $support(X) \triangleq \{x \,/\, f(x) > 0\} = (a, b)$ with $-\infty < a < b < +\infty$.

This formulation is useful whenever the interpretation of the Gini index is related to extreme value theory, since it is established in terms of the difference between the maximum and the minimum (the most extreme values) of a random sample (see Yitzhaki and Schechtman [65] for more details).

The Gini index admits many other formulations and interpretations. By way of example, it is expressed below as an expected value of a transformation of X and also as the covariance between two transformations of X.

Given a non-negative random variable X, the transformation $Q^*(X) = \frac{1}{\mu}\int_X^\infty (1 - F(t))dt$ is considered, and its expected value is given by

$$\begin{aligned} E[Q^*(X)] &= \frac{1}{\mu}\int_0^\infty \left(\int_x^\infty (1 - F(t))dt\right) f(x)dx \\ &= \left(\text{by integration by parts, } u = \int_x^\infty (1 - F(t))dt \text{ and } dv = f(x)dx\right) \\ &= \frac{1}{\mu}\left(\left[F(x)\int_x^\infty (1 - F(t))dt\right]_0^\infty + \int_0^\infty F(x)(1 - F(x))dx\right) \\ &= IG(X). \end{aligned}$$

Note that $E[\int_X^\infty (1 - F(t))dt] = \int_0^\infty F(x)(1 - F(x))dx$. A better-known equality is $IG(X) = E[1 - 2Q(X)]$, where $Q(x) = \int_0^x tf(t)dt$.

Let us examine another formulation of the Gini index. By integration by parts [66], with $u = F(x)(1 - F(x))$ and $dv = dx$,

$$\int_0^\infty F(x)(1 - F(x))dx = [xF(x)(1 - F(x))]_0^\infty + 2\int_0^\infty x(F(x) - \frac{1}{2})f(x)dx$$

and since $\lim_{x\to\infty} x(1 - F(x)) = 0$ and $\int_0^\infty x(F(x) - \frac{1}{2})f(x)dx = cov(X, F(X))$, due to $\mu < \infty$ and $\mu_{F(X)} = \frac{1}{2}$, then

$$IG(X) = \frac{2}{\mu}cov(X, F(X)) = cov\left(\frac{X}{\mu_X}, \frac{F(X)}{\mu_{F(X)}}\right).$$

Therefore, the Gini index is the covariance between $\frac{X}{\mu_X}$ and $\frac{F(X)}{\mu_{F(X)}}$. Furthermore, this result can be used to obtain a well-known inequality between the Gini index and Pearson's variation coefficient, denoted

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2024 doi:10.20944/preprints202410.0088.v1

by $CV(X)$. From the equality $var(F(X)) = \frac{1}{12}$ it follows that $cov^2(X, F(X)) \leq var(X)\, var(F(X)) = \frac{1}{12} var(X)$ and, therefore

$$IG(X) \leq \sqrt{\frac{4}{12}\frac{var(X)}{\mu^2}} = \frac{1}{\sqrt{3}} CV(X).$$

On the other hand, if the linear regression between the variables X and $F(X)$ is considered, that is $X = a + bF(X)$, then the regression coefficient b is $b = 6\mu IG(X)$. This expression allows the Gini index to be easily obtained by employing any software to carry out the linear regression between the variables X and $F(X)$.

3. The nth Gini Index

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (iid) random variables with the same distribution as X. If the transformations given by $U_n = \max\{X_1, X_2, \cdots, X_n\}$ and $V_n = \min\{X_1, X_2, \cdots, X_n\}$ are considered, then their cdf's are: $F_{U_n}(u) = F(u)^n$ and $F_{V_n}(v) = 1 - (1 - F(v))^n$, and $0 \leq V_n \leq U_n$ for each integer n. Hence,

$$0 \leq E[V_n] \leq E[U_n] = \int_0^\infty x\, nF(x)^{n-1} f(x)dx \leq n \int_0^\infty x f(x)dx = n\mu$$

that is, $E[U_n]$ and $E[V_n]$ exist, and therefore $E[U_n - V_n]$ exists for any n.

By using $E[X] = \int_0^\infty (1 - F(x))dx$, then $E[U_n] = \int_0^\infty (1 - F(x)^n)dx$ and $E[V_n] = \int_0^\infty (1 - F(x))^n dx$. Thus, for any integer:

$$E[U_n - V_n] = \int_0^\infty (1 - F(x)^n - (1 - F(x))^n)dx.$$

Let us denote $a_n = E[U_n - V_n]$ and study the sequence $\{a_n\}$.

Clearly, this sequence is non-negative and non-decreasing since $E[U_{n+1}] \geq E[U_n]$ and $E[V_{n+1}] \leq E[V_n]$. Furthermore, $E[U_n - V_n] \leq \frac{n}{2}\Delta = \frac{n}{2} E[U_2 - V_2]$ as can be seen in the following result.

Proposition 3.1. *Let $X_1, X_2, \cdots, X_n$ be iid with cdf $F(x)$, then*

$$E[U_n - V_n] \leq n \int_0^\infty F(x)(1 - F(x))dx, \text{ for } n \geq 2 \tag{2}$$

and if $n = 2$ or 3, then the equality holds.

Proof: Let us find a bound of a_n by induction:
For $n = 2$: $a_2 = \int_0^\infty (1 - F(x)^2 - (1 - F(x))^2)dx = 2\int_0^\infty F(x)(1 - F(x))dx$
For $n = 3$: $a_3 = \int_0^\infty (1 - F(x)^3 - (1 - F(x))^3)dx = 3\int_0^\infty F(x)(1 - F(x))dx$
For $n + 1$ $(n \geq 3)$:

$$\begin{aligned}
a_{n+1} - a_n &= \int_0^\infty (F(x)^n - F(x)^{n+1} + (1 - F(x))^n - (1 - F(x))^{n+1})dx \\
&= \int_0^\infty (F(x)^n(1 - F(x)) + (1 - F(x))^n(1 - (1 - F(x)))dx \\
&= \int_0^\infty F(x)(1 - F(x))(F(x)^{n-1} + (1 - F(x))^{n-1})dx \\
&\leq \int_0^\infty F(x)(1 - F(x))dx
\end{aligned}$$

since $0 \leq (x^n + (1 - x)^n) \leq (x + (1 - x))^n = 1$ when $0 \leq x \leq 1$ for each integer n. Therefore $a_{n+1} \leq a_n + \int_0^\infty F(x)(1 - F(x))dx$, and by induction $a_{n+1} \leq a_2 + (n - 1)\int_0^\infty F(x)(1 - F(x))dx$, and the proof is complete. ∎

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2024 doi:10.20944/preprints202410.0088.v1

Note that the inequality (2) is also true if $n = 1$ since $U_1 = V_1$ and therefore

$$0 = a_1 \leq \int_0^\infty F(x)(1-F(x))dx < \int_0^\infty (1-F(x))dx = \mu \tag{3}$$

Another important result which must also be borne in mind is presented.

Proposition 3.2. *Let X be a non-negative random variable with $\mu > 0$. The sequence $\left\{\frac{1}{n}E[U_n - V_n]\right\}_{n=3}^\infty$ is non-increasing and tends towards zero.*

Proof: From Proposition 3.1:

$$a_{n+1} - a_n = \int_0^\infty (F(x)^n(1-F(x)) + (1-F(x))^n(1-(1-F(x)))dx$$

On the other hand:

$$1 = (x + (1-x))^{n+1} = x^{n+1} + (n+1)x^n(1-x) + R(x) + (n+1)x(1-x)^n + (1-x)^{n+1}$$

where $R(x) \geq 0$ if $0 \leq x \leq 1$ and $n \geq 3$, and hence

$$x^n(1-x) + x(1-x)^n \leq \frac{1}{n+1}\left(1 - x^{n+1} - (1-x)^{n+1}\right)$$

Therefore:

$$a_{n+1} - a_n \leq \frac{1}{n+1}\int_0^\infty \left(1 - F(x)^{n+1} - (1-F(x))^{n+1}\right)dx = \frac{1}{n+1}a_{n+1}$$

which implies that $\frac{1}{n+1}a_{n+1} \leq \frac{1}{n}a_n$ and therefore $\left\{\frac{a_n}{n}\right\}$ is non-increasing. In fact, if $V_{n+1} = X_{(1)} \leq X_{(2)} \cdots \leq X_{(n)} \leq X_{(n+1)} = U_{n+1}$ are the order statistics, then it is easy to prove that $\frac{E[U_{n+1}]}{n+1} = \frac{E[U_n]}{n} - \frac{1}{n(n+1)}E[X_{(n)}]$, $\frac{E[V_{n+1}]}{n+1} = \frac{E[V_n]}{n} - \frac{1}{n(n+1)}E[X_{(2)}]$, and hence $\frac{a_{n+1}}{n+1} - \frac{a_n}{n} = \frac{-1}{n(n+1)}\left(E[X_{(n)}] - E[X_{(2)}]\right) \leq 0$.

On the other hand, as $\left\{\frac{a_n}{n}\right\}$ is non-increasing and $0 \leq a_n$ for all n, then $g = \lim_{n\to\infty}\frac{a_n}{n}$ exists.

Let us see that $g = 0$:

If $support(X) \triangleq \{x \,/\, f(x) > 0\} \subset [0, b]$ where b is a constant then $a_n \leq b$, and therefore $0 \leq \frac{a_n}{n} \leq \frac{b}{n} \to 0$ if $n \to \infty$.

If $support(X) = (0, \infty)$ then $0 \leq F(x) < 1$ for all x. Hence:

$$0 \leq \frac{a_n}{n} \leq \frac{E[U_n]}{n} = \int_0^\infty x\,F(x)^{n-1}\,f(x)\,dx$$

and by taking the limit when n tends towards ∞, and since the integral is positive then:

$$\lim_{n\to\infty}\frac{a_n}{n} \leq \lim_{n\to\infty}\int_0^\infty x\,F(x)^{n-1}\,f(x)\,dx = \int_0^\infty x\,f(x)\left(\lim_{n\to\infty}F(x)^{n-1}\right)dx = 0$$

and the proof is complete. ■

A family of indices of X is defined from the properties of $E[U_n - V_n]$ as follows:

Definition 3.3. *Let X be a non-negative random variable with $\mu > 0$. For each integer n, the nth Gini index of X is defined as follows:*

$$IG_n(X) = \frac{E[U_n - V_n]}{n\mu}, \qquad \textit{for } n \geq 2. \tag{4}$$

Let us enumerate some properties of the nth Gini index:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2024 doi:10.20944/preprints202410.0088.v1

1. The nth Gini index exists for any non-negative random variable with $\mu > 0$. This property is of major significance because if X is an income distribution with $\mu > 0$, then the nth Gini index can always be calculated even if some of its conventional moments do not exist.
2. $IG(X) = IG_2(X) = IG_3(X)$.
3. $0 \leq IG_n(X) < 1$ for all $n > 3$ (from Proposition 3.1 and inequality (3)).
4. $IG_n(bX) = IG_n(X)$ for any $b > 0$, that is, the nth Gini index is not affected by ratio-scale changes of the X variable.
5. $IG_n(a+X) = \frac{\mu}{a+\mu} IG_n(X)$ if $Y = a + X$ is a non-negative random variable (translation-scale changes). Therefore, transforming X into $a + X$ with $a > 0$ diminishes the nth Gini index.
6. $IG_n(X) = \frac{1}{\mu} Cov(X, F_{U_n}(X) - F_{V_n}(X))$, that is, the nth Gini index is the covariance between X and a transformation of X.
7. The sequence $\{IG_n(X)\}_{n=3}^{\infty}$ is non-increasing and $\lim_{n\to\infty} IG_n(X) = 0$. (straightforward from Proposition 3.2).

Proposition 3.4. *Let X be a non-negative random variable with $\mu > 0$. For each integer n, a value r_n exists such that $IG_n(X) = IG(r_n + X)$.*

Proof: Let $g = IG_n(X)$, and by taking into consideration that $0 < g \leq IG(X)$ (if $g = 0$ then $IG(X) = 0$) then the following equation in terms of r is considered: $IG_n(X) = \frac{\mu}{r+\mu} IG(X)$, whose solution is given by $r_n = \mu\left(\frac{IG(X)}{g} - 1\right) \geq 0$. Therefore, the Gini index of the random variable $Y = r_n + X$ is equal to $IG_n(X)$. ■

The most important consequence of Proposition 3.4 is that the nth Gini index is a Gini index and can be interpreted in this way.

One of major drawbacks when the Gini index is utilized, is the fact that there are non-negative random variables X and Y such that $IG(X) = IG(Y)$ and, therefore, it is not possible using the Gini index to quantify which distribution is more unequal. To overcome this situation, the nth Gini index can be calculated for several values of n. Let us consider an example based on two significant parametric models for describing incomes, whose Lorenz curves can be seen in Figure 1.

Example 3.5. *Let $X \sim SM(q = 2, a = 1, b = 1)$ be a Singh-Maddala distribution with shape parameters q and a, and scale parameter b. By using any calculus program, it can be calculated that*

$$\{IG_n(X)\}_{n=2}^{10} = \left\{\frac{2}{3}, \frac{2}{3}, \frac{22}{35}, \frac{62}{105}, \frac{386}{693}, \frac{122}{231}, \frac{3238}{6435}, \frac{3098}{6435}, \frac{106762}{230945}\right\} \tag{5}$$

$$= \left\{0.\widehat{6}, 0.\widehat{6}, 0.6286, 0.5905, 0.5570, 0.5281, 0.5032, 0.4814, 0.4622\right\} \tag{6}$$

On the other hand, let $Y \sim Pa(k = \frac{1}{5}, \alpha = \frac{5}{4})$ be a Pareto distribution with minimum value parameter k and shape parameter α. It can be obtained that

$$\{IG_n(Y)\}_{n=2}^{10} = \left\{\frac{2}{3}, \frac{2}{3}, \frac{57}{88}, \frac{83}{132}, \frac{14701}{24024}, \frac{2391}{4004}, \frac{372893}{638352}, \frac{45691}{79794}, \frac{9402119}{16721276}\right\} \tag{7}$$

$$= \left\{0.\widehat{6}, 0.\widehat{6}, 0.6477, 0.6288, 0.6119, 0.5972, 0.5841, 0.5726, 0.5623\right\} \tag{8}$$

Note that $IG(X) = IG(Y)$ and therefore these two distributions cannot be compared in relation to their concentration using the Gini index, neither can their Lorenz curves be used for that purpose because, as shown in Figure 1, there is no dominance relation, but they do intersect. Since $IG_n(X) < IG_n(Y)$ for $n \geq 4$, then the $Pa(k = \frac{1}{5}, \alpha = \frac{5}{4})$ distribution can be considered to be more concentrated than the $SM(q = 2, a = 1, b = 1)$ distribution, despite having the same value for the Gini index.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2024 doi:10.20944/preprints202410.0088.v1

It is easy to make two variables take the same Gini index by means of a positive shift of the variable that has the greatest Gini index. Let X and Y be non-negative random variables with $IG(X) > IG(Y)$. As in Proposition 3.4, it is possible to find a value r^* such that $IG(r^* + X) = IG(Y)$. Therefore, the two random variables $r^* + X$ and Y are similar from the point of view of their concentration. Note that no particular supposition about the distribution has been made.

In Forcina and Giorgi [67] is pointed out that the political and economic debate on how to achieve a more equal distribution of income and wealth was particularly lively at the beginning of the last century. Proposition 3.2 provides us with a statistical way to attain perfect equality, that is, if no economic nor any other kind of consideration is made, then the perfect equality of a distribution X is attained by considering the shifted variables $X_n = r_n + X$ such that $IG(X_n) = IG_n(X)$.

Example 3.6. *Let X be a Singh-Maddala distribution $SM(q = 2, a = 1, b = 1)$. The first values of $IG_n(X)$ are given in Example 3.5. Thus, the values r_n such that $IG(r_n + X) = IG_n(X)$ are $r_n = \left\{\frac{2}{33}, \frac{4}{31}, \frac{38}{193}, \frac{16}{61}, \frac{526}{1619}, \frac{596}{1549}, \frac{70802}{160143}\right\}$ for $n = 4, \cdots, 10$. Therefore, perfect equality is achieved with the sequence of random variables $X_n = X_{n-1} + b_n$, where $b_n = \left\{\frac{2}{33}, \frac{70}{1023}, \frac{406}{5983}, \frac{770}{11773}, \frac{6182}{98759}, \frac{150150}{2507831}, \frac{14227070}{248061507}\right\}$ for $n = 4, \cdots, 10$. Hence, the initial concentration measured through the Gini index, which has a value equal to $IG_3(X) = 2/3 \simeq 0.67$, is reduced to $IG_{10}(X) \simeq 0.46$ through the seven successive increases given by b_n, which means a decrease of 31%.*

3.1. The Extended Gini Index, the Lorenz Family and the uv-Gini Index

A similar generalization of the Gini index is given in Yitzhaki [68] and its development is given in Yitzhaki and Schechtman [58] which is called extended Gini or S-Gini index in the mathematically oriented literature. This family of indices, denoted by $EG(X, v)$, is defined as

$$EG(X, v) = \frac{1}{\mu} \int_0^\infty (1 - F(x) - (1 - F(x))^v) dx \tag{9}$$

where[1] $v \geq 1$. Several authors suggest using the extended Gini index to characterize the distribution [69].

Hence, $EG(X, v) = \frac{1}{\mu}(E[X] - E[\min\{X_1, X_2, \cdots, X_v\}])$ if v is an integer. Note that $E[U_2 - V_2] = \int_0^\infty (1 - F(x)^2 - (1 - F(x))^2) dx = 2\int_0^\infty (1 - F(x) - (1 - F(x))^2) dx$, and therefore, $IG(X) = EG(X, 2) = IG_2(X)$.

By following a similar expression of (9), a generalization of the nth Gini index is proposed by considering two positive real numbers instead of an integer as in (4):

Definition 3.7. *Let X be a non-negative random variable with $\mu > 0$. For two positive real numbers $u, v \geq 1$, the uv-Gini index of X is defined as follows:*

$$IG(X, u, v) = \frac{1}{u\mu} \int_0^\infty (1 - F(x)^u - (1 - F(x))^v) dx, \qquad \text{for } u, v \geq 1 \tag{10}$$

This index is more general than the extended Gini since $EG(X, v) = IG(X, 1, v)$.

It is well-known from the economic literature that the basic condition for a measure of inequality is that it should satisfy the Pigou-Dalton principle of transfers or eventually a similar normative principle. We now show that the *uv*-Gini index satisfies this principle from rank-dependent inequality measurements.

It is interesting to observe that the *uv*-Gini index is a linear combination of indices that belong to two alternative families of rank-dependent measures of inequality. One of these families is the

[1] The inequality index is negative if $0 < v < 1$ is considered.

generalized Gini family introduced by Kakwani [70], Donaldson and Weymark [71], and later by Yitzhaki [68]. This family is defined by

$$G(X,v) = 1 - \frac{1}{\mu}\int_0^\infty (1-F(x))^v dx, \qquad \text{for } v \geq 1, \tag{11}$$

and it is identical to expression (9).

The second family, the Lorenz family of inequality measures, was introduced in Aaberge [72] and is defined by

$$D(X,u) = \frac{1}{u\mu}\int_0^\infty F(x)(1-F(x)^u)dx, \qquad \text{for } u \geq 1. \tag{12}$$

Hence, by inserting (11) and (12) into $IG(X,u,v)$ (defined by (10)), it follows that

$$u\,IG(X,u,v) = (u-1)D(X,u-1) + G(X,v), \qquad \text{for } u,v \geq 1$$

and, therefore,

$$IG(X,u,v) = (1-\frac{1}{u})D(X,u-1) + \frac{1}{u}G(X,v), \qquad \text{for } u,v \geq 1. \tag{13}$$

It follows directly from expression (13) that the uv-Gini index is a convex combination of $G(x,v)$ and $D(X,u)$. Thus, since $G(x,v)$ and $D(X,u)$ satisfy the Pigou-Dalton principle of transfers, it follows from (13) that $IG(X,u,v)$ also satisfies the Pigou-Dalton principle of transfers. Furthermore, by observing that $G(x,v)$ and $D(X,u)$ are (0,1)-normalized measures of inequality, it holds that $IG(X,u,v)$ is also a (0,1)-normalized measure, and takes its maximum value when one unit receives the total income and its minimum value when all the units receive the same income.

3.2. The nth Gini Index in Terms of the Lorenz Curve

Taking into consideration that $IG_n(X) = IG(X,n,n)$, it follows that the nth Gini index verifies the Pigou-Dalton principle of transfers and that this index assumes the maximum value in the case of maximum concentration.

It is straightforward to write $IG(X,n,n)$ in terms of the Lorenz curve from (13) and the Lorenz curve expression of $G(X,n)$ and $D(X,n)$.

The Lorenz curve $L(\cdot)$ for X is defined by $L(u) = \frac{1}{\mu}\int_0^u x\,dF(x)$ for any $0 \leq u \leq 1$, and it can be observed in Yitzhaki [68] and Aaberge [72] that

$$G(X,n) = n(n-1)\int_0^1 (1-u)^{n-2}(u-L(u))du, \quad D(X,n) = (n+1)\int_0^1 u^{n-1}(u-L(u))du,$$

therefore

$$IG_n(X) = (n-1)\int_0^1 \left(u^{n-2} + (1-u)^{n-2}\right)(u-L(u))du.$$

Hence, in order to be more operative since $IG_2(X) = IG_3(X) = IG(X)$, it can be defined

$$IG_n^*(X) = (n+1)\int_0^1 (u^n + (1-u)^n)(u-L(u))du, \quad n = 1,2,3,\cdots \tag{14}$$

It can be seen that the criterion of Lorenz-dominance, which recognizes the highest of the Lorenz curves as preferable [73], is verified for the family $IG_n^*(X)$ from (14). Nevertheless, due to fact that the Lorenz curves may intersect, the criterion of Lorenz-dominance does not apply in many practical situations. No single measure is able to retain all the features concerning inequality exhibited by the Lorenz curve. For this reason, it is proposed that the first few nth indices $IG_n^*(X)$ are obtained together with other possible measures for the purpose of application.

4. Conclusions

In this paper, a new family of inequality indices, the nth Gini index, is proposed and studied. These indices are based on the deviation between the expected maximum and the expected minimum of independent and identically distributed random samples. This family generalizes the Gini index. At the same time, these indices and the S-Gini index are generalized by proposing the $uv-$Gini index, which turns out to be a convex combination of the S-Gini index and the Lorenz family of inequality measures, thereby verifying the principle of transfers. These new formulations can be useful whenever the preferred interpretation is related to extreme value theory and the use of a few measures of this family of Gini indices is proposed to obtain a summary of the basic information provided by the Lorenz curve.

This family of Gini indices enables a path to be found to the perfect equality for a given distribution of incomes through successive and equal increases of the incomes of each individual. It is also specially useful to compare the concentration of two distributions that have the same Gini index and intersecting Lorenz curves.

Author Contributions: Conceptualization, J.M.G., A.R.G., F.J.O. and L.G.; methodology, J.M.G., A.R.G., F.J.O. and L.G.; formal analysis, J.M.G., A.R.G., F.J.O. and L.G.; investigation, J.M.G., A.R.G., F.J.O. and L.G.; writing—original draft preparation, J.M.G., and A.R.G.; writing—review and editing, F.J.O., and L.G.; funding acquisition, J.M.G., A.R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially supported by the "ARTIFACTS: Generation of Reliable Synthetic Health Data for Federated Learning in Secure Data Spaces" Research Project (PID2022-1410450B-C42 (AEI/FEDER, UE)) funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe" by the "European Union".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Gini, C. Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti* **1914**, *73*, 1203–1248. Reprinted in: C. Gini, (2005). On the measurement of concentration and variability of characters. Metron - International Journal of Statistics 63, 3-38.
2. Gini, C. Measurement of inequality of incomes. *The Economic Journal* **1921**, *31*, 124–126.
3. Karsu, Ö.; Morton, A. Inequity averse optimization in operational research. *European Journal of Operational Research* **2015**, *173*, 343–359. doi:10.1016/j.ejor.2015.02.035.
4. Giorgi, G.; Gigliarano, C. The Gini concentration index: a review of the inference literature. *Journal of Economic Surveys* **2016**, *31*, 1130–1148. doi:10.1111/joes.12185.
5. Wu, W.C.; Chang, Y.T. Income inequality, distributive unfairness, and support for democracy: evidence from East Asia and Latin America. *Democratization* **2019**, *26*, 1475–1492. doi:10.1080/13510347.2019.1656198.
6. Giorgi, G., Gini Coefficient. In *In: P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, and R.A. Williams*; SAGE Research Methods Foundations, 2020. doi:10.4135/9781526421036883328.
7. Atkinson, A. *Inequality*; Harvard University Press, 2015.
8. Liu, Y.; Gastwirth, J. On the capacity of the Gini index to represent income distributions. *METRON* **2020**, *78*, 61–69. doi:10.1007/s40300-020-00164-8.
9. Xu, K. How has the Literature on Gini's Index Evolved in the Past 80 Years? Economics working paper, Dalhousie University, Canada, 2004.
10. Farris, F. The Gini Index and Measures of Inequality. *American Mathematical Monthly* **2010**, *117*, 851–864. doi:10.4169/000298910X523344.
11. Eliazar, I. Harnessing inequality. *Physics Reports* **2016**, *649*, 1–29. doi:10.1016/j.physrep.2016.07.005.
12. Eliazar, I. A tour of inequality. *Annals of Physics* **2018**, *389*, 306–332. doi:10.1016/j.aop.2017.12.010.

13. Eliazar, I.; Giorgi, G. From Gini to Bonferroni to Tsallis: an inequality-indices trek. *METRON* **2020**, *78*, 119–153. doi:10.1007/s40300-020-00171-9.
14. Mukhopadhyay, N.; Sengupta, P.E. *Gini Inequality Index: Methods and Applications*; Chapman and Hall/CRC, 2021. doi:10.1201/9781003143642.
15. Sadras, V.; Bongiovanni, R. Use of Lorenz curves and Gini coefficients to assess yield inequality within paddocks. *Field Crops Research* **2004**, *90*, 303–310. doi:10.1016/j.fcr.2004.04.003.
16. Hammel, E. Demographic dynamics and kinship in anthropological populations. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 2248–53. doi:10.1073/pnas.0409762102.
17. Abraham, R.; Bergh, S.; Nair, P. A New Approach to Galaxy Morphology: I. Analysis of the Sloan Digital Sky Survey Early Data Release. *The Astrophysical Journal* **2003**, *588*. doi:10.1086/373919.
18. Karmakar, A.; Banerjee, P.; De, D.; Bandyopadhyay, S.; Ghosh, P. MedGini: Gini index based sustainable health monitoring system using dew computing. *Medicine in Novel Technology and Devices* **2022**, *16*, 100145. doi:10.1016/j.medntd.2022.100145.
19. Pernot, P.; Savin, A. Using the Gini coefficient to characterize the shape of computational chemistry error distributions. *Theoretical Chemistry Accounts* **2021**, *140*. doi:10.1007/s00214-021-02725-0.
20. Hasisi, B.; Perry, S.; Ilan, Y.; Wolfowicz, M. Concentrated and Close to Home: The Spatial Clustering and Distance Decay of Lone Terrorist Vehicular Attacks. *Journal of Quantitative Criminology* **2020**, *36*. doi:10.1007/s10940-019-09414-z.
21. Wittebolle, L.; Marzorati, M.; Clement, L.; Balloi, A.; Daffonchio, D.; Heylen, K.; De Vos, P.; Verstraete, W.; Boon, N. Initial community evenness favours functionality under selective stress. *Nature* **2009**, *458*, 623–626. doi:10.1038/nature07840/.
22. Naeem, S. Ecology: Gini in the bottle. *Nature* **2009**, *458*, 579–80. doi:10.1038/458579a.
23. Ho, K.; Chow, F.; Chau, H. Study of the Wealth Inequality in the Minority Game. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2004**, *70*, 066110. doi:10.1103/PhysRevE.70.066110.
24. Beaugrand, G.; Edwards, M.; Legendre, L. Marine biodiversity, ecosystem functioning, and carbon cycles. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 10120–4. doi:10.1073/pnas.0913855107.
25. Arbel, Y.; Fialkoff, C.; Kerner, A.; Kerner, M. Do Population Density, Socio-Economic Ranking and Gini Index of Cities Influence Infection Rates from Coronavirus? Israel as a case Study. *The Annals of regional science* **2020**, *68*, 181–206. doi:10.21203/rs.3.rs-40387/v1.
26. Cima, E.; Uribe Opazo, M.; Bombacini, M.; Rocha, W.; Pagliosa Carvalho Guedes, L. Spatial Analysis: A Socioeconomic View on the Incidence of the New Coronavirus in Paraná-Brazil. *Stats* **2022**, *5*, 1029–1043. doi:10.3390/stats5040061.
27. Sazuka, N.; ichi Inoue, J. Fluctuations in time intervals of financial data from the view point of the Gini index. *Physica A: Statistical Mechanics and its Applications* **2007**, *383*, 49–53. Econophysics Colloquium 2006 and Third Bonzenfreies Colloquium, doi:https://doi.org/10.1016/j.physa.2007.04.088.
28. Sazuka, N.; Inoue, J.i.; Scalas, E. The distribution of first-passage times and durations in FOREX and future markets. *Physica A: Statistical Mechanics and its Applications* **2009**, *388*, 2839–2853. doi:10.1016/j.physa.2009.03.027.
29. Tu, J.; Sui, H.; Feng, W.; Sun, K.; Xu, C.; Han, Q. Detecting building façade damage from oblique aerial images using local symmetry feature and the Gini Index. *Remote Sensing Letters* **2017**, *8*, 676–685. doi:10.1080/2150704X.2017.1312027.
30. Lechthaler, B.; Pauly, C.; Mücklich, F. Objective homogeneity quantification of a periodic surface using the Gini coefficient. *Scientific Reports* **2020**, *10*, 14516. doi:10.1038/s41598-020-70758-9.
31. Graczyk, P. Gini Coefficient: A New Way To Express Selectivity of Kinase Inhibitors against a Family of Kinases †. *Journal of medicinal chemistry* **2007**, *50*, 5773–9. doi:10.1021/jm070562u.
32. O'Hagan, S.; Wright Muelas, M.; Day, P.; Lundberg, E.; Kell, D. GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles. *Cell Systems* **2018**, *6*. doi:10.1016/j.cels.2018.01.003.
33. Woolhouse, M.; Dye, C.; Etard, J.F.; Smith, T.; Charlwood, J.; Garnett, G.; Hagan, P.; Hii, J.; Ndhlovu, P.; Quinnell, R.; Watts, C.; Chandiwana, S.; Anderson, R. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 338–42. doi:10.1073/pnas.94.1.338.

34. Rindfuss, R.; Walsh, S.; Turner, B.; Fox, J.; Mishra, V. Developing a Science of Land Change: Challenges and Methodological Issues. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 13976–81. doi:10.1073/pnas.0401545101.
35. Hörcher, D.; Graham, D. The Gini index of demand imbalances in public transport. *Transportation* **2021**, *48*. doi:10.1007/s11116-020-10138-4.
36. Eliazar, I.; Sokolov, I. Measuring statistical evenness: A panoramic overview. *Physica A: Statistical Mechanics and its Applications* **2012**, *391*, 1323–1353. doi:10.1016/j.physa.2011.09.007.
37. Kokko, H.; Mackenzie, A.; Reynolds, J.; Lindström, J.; Sutherland, W. Measures of Inequality Are Not Equal. *The American naturalist* **1999**, *154*, 358–382. doi:10.1086/303235.
38. Charles, V.; Gherman, T.; Paliza, J., The Gini Index: A Modern Measure of Inequality; Palgrave Macmillan, Cham., 2022; pp. 55–84. doi:10.1007/978-3-030-84535-3_3.
39. Theil, H. *Economics and Information Theory*; North-Holland Publishing Company, 1967.
40. Dalton, H. Measurement of the Inequality of Income. *Economic Journal* **1920**, *30*, 348–361.
41. Allison, P.D. Measures of Inequality. *American Sociological Review* **1978**, *43*, 865–880.
42. Stefanescu, S. About the Accuracy of Gini Index for Measuring the Poverty. *Romanian Journal of Economic Forecasting* **2011**, *43*, 255–266.
43. Pyatt, G. On the intepretation and disaggragation of Gini coefficient. *Economic Journal* **1976**, *86*, 243–255.
44. Xu, K.; Osberg, L. The social welfare implications, decomposability, and geometry of the Sen family of poverty indices. *Canadian Journal of Economics* **2002**, *35*, 138–152.
45. Osberg, L. On the Limitations of Some Current Usages of the Gini Index. *Review of Income and Wealth* **2017**, *63*, 574–584. doi:10.1111/roiw.12256.
46. Liao, T. Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-Based Clustering. *Sociological Methodology* **2006**, *36*, 201 – 224. doi:10.1111/j.1467-9531.2006.00179.x.
47. Furman, E.; Kye, Y.; Su, J. Computing the Gini index: A note. *Economics Letters* **2019**, *185*, 108753. doi:10.1016/j.econlet.2019.108753.
48. Ceriani, L.; Verme, P. Individual Diversity and the Gini Decomposition. *Social Indicators Research* **2015**, *121*, 637–646. doi:10.1007/s11205-014-0656-0.
49. Eliazar, I.I.; Sokolov, I.M. Gini characterization of extreme-value statistics. *Physica A: Statistical Mechanics and its Applications* **2010**, *389*, 4462–4472. doi:10.1016/j.physa.2010.07.005.
50. Eliazar, I. Beautiful Gini. *METRON* **2024**. doi:10.1007/s40300-024-00271-w.
51. Lunetta1972. Sulla concentrazione delle distribuzioni doppie. *In Atti della XXVII Riunione Scientifica della Società Italiana di Statistica, vol. II, Palermo* **1972**, pp. 127–150.
52. Taguchi, T. On the two-dimensional concentration surface and extensions of concentration coefficient and pareto distribution to the two dimensional. *Annals of the Institute of Statistical Mathematics* **1972-1973**, *part I (1972), vol. 24 355-381; part II (1972), vol. 24, 599-619; part III (1973), vol. 25, 215-237*.
53. Koshevoy, G.; Mosler, K. Multivariate Gini Indices. *Journal of Multivariate Analysis* **1997**, *60*, 252–276. doi:10.1006/jmva.1996.1655.
54. Arnold, B.C. Inequality measures for multivariate distributions. *Metron - International Journal of Statistics* **2005**, *LXIII*, 317–327.
55. Grothe, O.; Kächele, F.; Schmid, F. A multivariate extension of the Lorenz curve based on copulas and a related multivariate Gini coefficient. *The Journal of Economic Inequality* **2022**, *20*, 727–748. doi:10.1007/s10888-022-09533-x.
56. Pietra, G. Delle relazioni tra gli indici di variabilità (Nota I). *Atti del Reale Istituto Veneto di Scienze, Lettere e Arti* **1915**, *LXXIV, Part I*, 775–792. Reprinted in: G. Pietra, (2014). On the relationships between variability indices (Note I). Metron - International Journal of Statistics 72, 5-16.
57. Dorfman, R. A Formula for the Gini coefficient. *Review of Economics and Statistics* **1979**, *61*, 146–149.
58. Yitzhaki, S.; Schechtman, E. The Properties of the Extended Gini Measures of Variability and Inequality. Technical report, Social Science Research Network, 2005. Available at SSRN: http://ssrn.com/abstract=815564.
59. Fishburn, P. Stochastic dominance and moments of distributions. *Mathematics of Operations Research* **1980**, *5*, 94–100.
60. Muliere, P.; Scarsini, M. A note on stochastic dominance and inequality measures. *Journal of Economic Theory* **1989**, *49*, 314–323.

61. Gonzalez-Abril, L.; Velasco, F.; Gavilan, J.M.; Sanchez-Reyes, L.M. The Similarity between the Square of the Coeficient of Variation and the Gini Index of a General Random Variable. *Journal of Quantitative Methods for Economics and Business Administration* **2010**, *10*, 5–18.
62. Foster, J.; Wolfson, M. Polarization and the Decline of the Middle Class in Canada and the U.S. *The Journal of Economic Inequality* **2010**, *8*, 247–273. doi:10.1007/s10888-009-9122-7.
63. Kendall, M.; Stuart, A. *The Advanced Theory of Statistics, Vol. 1, Distribution Theory*; Hafner Publishing Company, 1958.
64. Yitzhaki, S. Stochastic Dominance, Mean Variance and Gini's Mean Difference. *American Economic Review* **1982**, *72*, 178–185.
65. Yitzhaki, S.; Schechtman, E. *The Gini methodology: A primer on a statistical methodology*; Springer, New York, 2013.
66. Lerman, R.; Yitzhaki, S. A Note on the Calculation and Interpretation of the Gini Index. *Economics Letters* **1984**, *15*, 363–368.
67. Forcina, A.; Giorgi, G. Early Gini's Contributions to Inequality Measurement and Statistical Inference. *Electronic Journal for History of Probability and Statistics* **2005**, *1*, 1–15.
68. Yitzhaki, S. On a Extension of Gini inequality index. *International Economic Review* **1983**, *24*, 617–628.
69. Kotz, S.; Kleiber, C. A characterization of income distributions in terms of generalized Gini coefficients. *Social Choice and Welfare* **2002**, *19*, 789–794.
70. Kakwani, N. On a Class Poverty Measures. *Econometrica* **1980**, *48*, 437–446.
71. Donaldson, D.; Weymark, J. A Single Parameter Generalization of the Gini Indices of Inequality. *Journal of Economic Theory* **1980**, *22*, 67–86.
72. Aaberge, R. Characterizations of Lorenz Curves and Income Distributions. *Social Choice and Welfare* **2000**, *17*, 639–653.
73. Atkinson, A. On the measurement of inequality. *Journal of Economic Theory* **1970**, *2*, 244–263.