

10-2021

Beware the Gini Index! A New Inequality Measure

Sabiou M. Inoua

Chapman University, inoua@chapman.edu

Follow this and additional works at: https://digitalcommons.chapman.edu/esi_working_papers



Part of the [Econometrics Commons](#), [Economic Theory Commons](#), and the [Other Economics Commons](#)

Recommended Citation

Inoua, S. (2021). Beware the Gini index! A new inequality measure. *ESI Working Paper 21-18*.
https://digitalcommons.chapman.edu/esi_working_papers/355/

This Article is brought to you for free and open access by the Economic Science Institute at Chapman University Digital Commons. It has been accepted for inclusion in ESI Working Papers by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

Beware the Gini Index! A New Inequality Measure

Comments

ESI Working Paper 21-18

Beware the Gini Index! A New Inequality Measure

Sabiou Inoua

Chapman University

inoua@chapman.edu

Abstract. The Gini index underestimates inequality for heavy-tailed distributions: for example, a Pareto distribution with exponent 1.5 (which has infinite variance) has the same Gini index as any exponential distribution (a mere 0.5). This is because the Gini index is relatively robust to extreme observations; while a statistic's robustness to extremes is desirable for data potentially distorted by outliers, it is misleading for heavy-tailed distributions, which inherently exhibit extremes. We propose an alternative inequality index: the variance normalized by the second moment. This ratio is more stable (hence more reliable) for large samples from an infinite-variance distribution than the Gini index paradoxically. Moreover, the new index satisfies the normative axioms of inequality measurement; in particular, it is decomposable into inequality within and between subgroups, unlike the Gini index.

Keywords: inequality, Gini index, heavy tail, power law, infinite variance, generalized central limit theorem, robustness

JEL Codes: C10, D63

1 Overview

The Gini index, proposed by C. Gini [1], is the most popular inequality measure. Iconic by its geometric interpretation in terms of the Lorenz curve [2], the Gini index is also fascinating by its rich mathematical properties and alternative formulations [3]. Yet the Gini index has some limitations; for example, it does not apply to a zero-mean distribution (such as the normal distribution) and it may behave poorly, falling outside the interval $[0,1]$, for variables assuming negative values. This problem is easily fixed, however; for one can adjust the definition of the index for negative values [4]. In fact, the Gini ratio's denominator is more rigorously defined for a signed variable to be the variable's mean absolute value (rather than the raw mean): then the ratio is well-defined and well-behaved for any nonzero variable (Section 4).

More importantly, the Gini index underestimates inequality for heavy-tailed distributions.¹ Thus a Pareto distribution with an exponent of 1.5 (an infinite-variance variable) has the same Gini index as any exponential distribution (a mere 0.5), because the Gini index is relatively insensitive to a distribution's tail, namely to extreme realizations of a variable (Section 2). While a statistic's robustness to extremes is desirable for data potentially distorted by outliers, it is misleading for heavy-tailed distributions, which inherently exhibit extremes. We propose an alternative inequality index (Section 3) that may seem paradoxical at first sight but that is well-behaved upon scrutiny: the variance normalized by the second moment. This ratio is stable for large samples even when the theoretical (population) variance is infinite: it is then more stable statistically than the Gini index, surprisingly (Section 4). Moreover, the new index satisfies the normative axioms of inequality

¹ Seemingly related limitations of the Gini index are suggested in the literature in terms of the Gini index's over-sensitive to changes in the middle of a distribution compared to changes in the distribution's tails [5-8], a claim that does not stand closer scrutiny, however [9]. This paper focuses not so much on the Gini's different treatments of portions of a distribution than the index's mild treatment of heavy tails. Other reported flaws of the index include a small-sample bias and a bias due to grouping [5, 6, 8, and refs. therein].

measurement: notably it is decomposable into inequality within and between subgroups, unlike the Gini index.

2 The Gini Index and Heavy Tails

The Gini index of a non-negative variable X with positive mean is by definition:²

$$G(X) = \frac{\frac{1}{2} \mathbb{E} |X - X'|}{\mathbb{E}(X)}, \quad (1)$$

where X' is an independent copy of X (that is, X and X' are independent but identically distributed) and \mathbb{E} is the expectation operator. For a continuous variable X , the Gini index is more easily computed from the distribution function $F(x) = \text{prob}\{X \leq x\}$ through the formula:

$$G(X) = \frac{\int_0^\infty F(x)[1 - F(x)]dx}{\mathbb{E}(X)}. \quad (2)$$

For an empirical variable (or sample) $X = [x_1, \dots, x_N]$, the Gini index becomes

$$G(X) = \frac{\frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N |x_i - x_j|}{\text{mean}(X)}, \quad (3)$$

where $\text{mean}(X)$ stands for the arithmetic average of X . Formula (3) becomes simpler if one arranges X in ascending order: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(N)}$. Then

$$G(X) = \frac{2}{N} \frac{\sum_{i=1}^N i x_{(i)}}{\sum_{i=1}^N x_i} - \frac{N+1}{N}. \quad (4)$$

These are standard facts about the Gini index that we remind for the sequel.³

The following result illustrates the above-mentioned inadequacy of the Gini index for heavy-tailed distributions:

² Throughout this paper, the term “variable” (used interchangeably with “distribution”) is used preferably to “random variable”.

³ For a review and proof of these facts and many others, see e.g. [3].

Proposition 1. *If Z is an exponential distribution with mean $2/3$, then Z and $\exp(Z)$ have the same Gini index $G = 0.5$. Moreover, $G(Z) \geq G[\exp(Z)]$ if $\text{mean}(Z) \geq 2/3$.*

Proof. It is easy to show that a variable Z has an exponential distribution with mean $1/\alpha$ if and only if $\exp(Z)$ has a Pareto distribution with exponent α (see below). The Gini index of any exponential distribution is 0.5 and the Gini index of a Pareto distribution with exponent $\alpha > 1$ is $G = G(\alpha) = 1/(2\alpha - 1)$. [These are known facts that one can derive from formula (2), for example.] Thus $G(\alpha) \leq 0.5$ if $\alpha \geq 3/2$, with equality if $\alpha = 3/2$. ■

That Z and $\exp(Z)$ could have the same Gini index is not a minor anomaly, for Z and $\exp(Z)$ are not different just quantitatively (though this is important enough in itself): $\exp(Z)$ is an infinite-variance heavy-tailed variable (Figure 1).

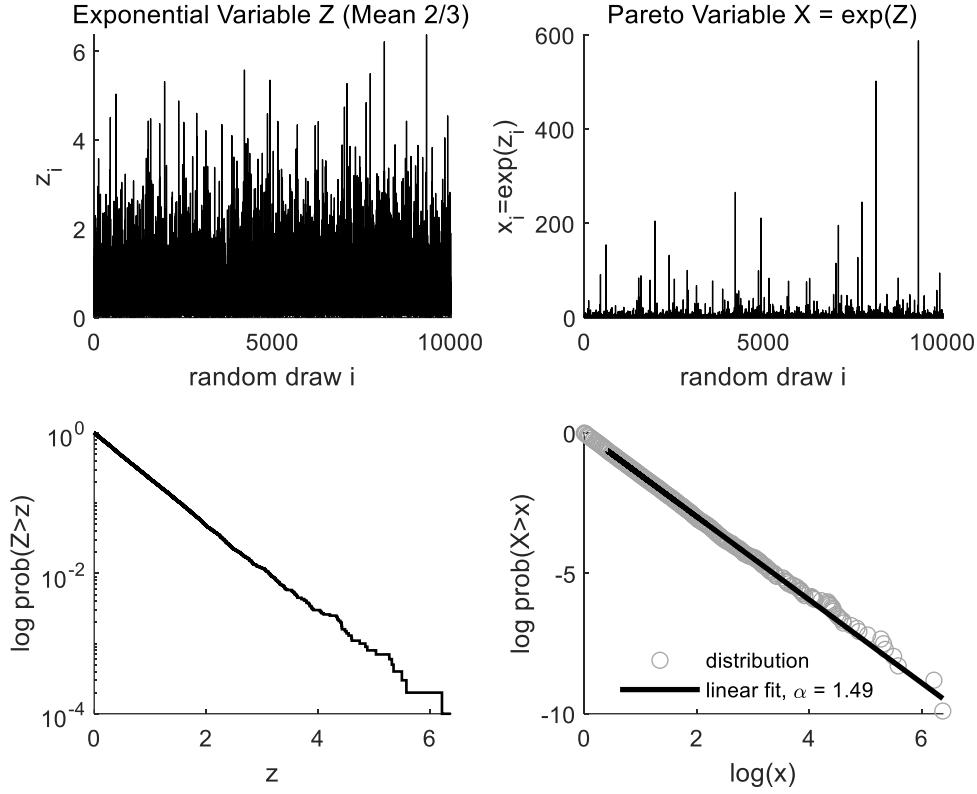


Figure 1. Exponential versus Pareto distribution. The two distributions, Z and $X = \exp(Z)$, have the same Gini index $G=0.5$!

A heavy-tailed distribution is precisely one that is more prone to exhibit extremes than any exponential distribution, by having a tail that decays more slowly than any exponential tail. Formally, a variable X is (right) heavy-tailed if⁴

$$\limsup_{x \rightarrow \infty} \frac{\text{prob}\{X \geq x\}}{\exp(-\lambda x)} = \infty \text{ for all } \lambda > 0. \quad (5)$$

A Pareto distribution is a simple such distributions. A Pareto distribution, recall, is a continuous variable X with $\text{prob}\{X \geq x\} = (x/x_{\min})^{-\alpha}$, $x \geq x_{\min} > 0$, $\alpha > 0$, where α is called the tail exponent.⁵ It corresponds to a line in a log-log plot (Figure 2: 4th Subplot) and can be characterized as the exponential of an exponential distribution:

⁴ Alternatively, X is heavy tailed if $\mathbb{E}[\exp(\lambda X)] = \infty$ for all $\lambda > 0$. The two definitions are equivalent [10, theorem 2.6]. We prefer definition (5) to the just-mentioned and more common one, because the former makes it clear that heavy-tailed-ness is relatively to exponential distributions.

⁵ For a more detailed introduction to the theory and empirics of Pareto distributions, see, e.g., [11].

if $\text{prob}\{Z \geq z\} = \exp[-\alpha(z - z_{\min})]$, for $z \geq z_{\min}$, then $\text{prob}\{\exp(Z) \geq x\} = \text{prob}\{Z \geq \log x\} = \exp[-\alpha(\log x - z_{\min})] = (x / x_{\min})^{-\alpha}$, where $x_{\min} = \exp(z_{\min})$; conversely, $\log X$ is exponential if X is Pareto. A power law X is a distribution that is Pareto asymptotically, in the sense that there is $\alpha > 0$ such that $x^\alpha \text{prob}\{|X| \geq x\} \rightarrow C > 0$ as $x \rightarrow \infty$. Power laws themselves belong to the class of regularly varying distributions, defined by $\text{prob}\{|X| \geq x\} = L(x)x^{-\alpha}$, where L is a slowly varying function, that is, a function that behaves asymptotically like constant, in the sense that $L(tx)/L(x) \rightarrow 1$ as $x \rightarrow \infty$ for any $t > 0$ (two common examples being a constant and a logarithmic function).⁶ Regularly varying functions are special distributions because they need not obey the standard laws of large numbers due to their high variability (the lower α , the more extreme the variability); and when they do, their sum converges more slowly to a Gaussian distribution compared to more common distributions. One can show that for any regularly varying variable X , we have $\mathbb{E}(|X|^\mu) < \infty$ if $0 < \mu < \alpha$ and $\mathbb{E}(|X|^\mu) = \infty$ if $\mu > \alpha$, which follow by simple integration for power laws, for which in addition $\mathbb{E}(|X|^\mu) = \infty$ if $\mu = \alpha$, and for a Pareto, $\mathbb{E}(X^\mu) = \alpha x_{\min}^\mu / (\alpha - \mu)$ if $0 < \mu < \alpha$.

Thanks to seminal work by P. Levy [12] and the contributions of other influential probability theorists of the past century, the central limit theorem has been extended to any variable (with light or heavy tails) into what we may call the “stable limit theorem”, a complete characterization of all the possible limit laws of the (properly normalized) sum of independent copies of a variable. (See this paper’s Appendix for a specialization of this theorem invoked throughout this paper, notably in Proposition 4, Section 4.) The only possible limit laws are called alpha-stable or Levy variables, in reference to their stability by convolution (that is, up to standardization, a stable law is preserved by addition of independent copies of it) and their characteristic parameter $\alpha \in (0, 2]$: the Gaussian is the only finite-variance stable law

⁶ A constant function C satisfies $C(xt) = C(x)$.

and it corresponds to $\alpha=2$; the other stable variables are a special subclass of infinite-variance power laws whose exponents $\alpha \in (0,2)$.⁷ The limit of the sum of independent copies of any variable X , if this limit exists, is either an infinite-variance variable, namely a Levy with $\alpha \in (0,2)$, or a Gaussian when the tail of X is not heavy enough, in the sense that the function $x \mapsto \mathbb{E}(X^2 1_{|X| \leq x})$ is slowly varying: this last condition is the general condition for convergence to a Gaussian (generalizing the finite second moment condition).⁸ A necessary condition for the sum of independent copies of a variable X to converge to a stable law is that X have a regularly varying distribution. We focus in this paper mostly on power laws, for the economic data of interest here (notably income or wealth distributions) are indeed power laws, as is known since V. Pareto's discovery of this class of distributions [18].⁹

Proposition 1 suggests that the typically moderate income or wealth Gini index (below 0.5 in most countries: Figure 2) might be an understatement.

⁷ For a synthetic textbook exposition of the general central limit theorem see, e.g., [13, sec. 2.2, th. 2.2.15] and [14, sec. 3.8, th. 3.8.2]. For intuitive derivations in a physics context, see e.g. [15, 16]. For an economic application, see e.g. [17].

⁸ Throughout, 1_A stands for the indicator function of A : that is, $1_A = 1$ if A is true, $1_A = 0$ if A is false.

⁹ For a review of power laws in natural, social, and economic data see [11, 19].

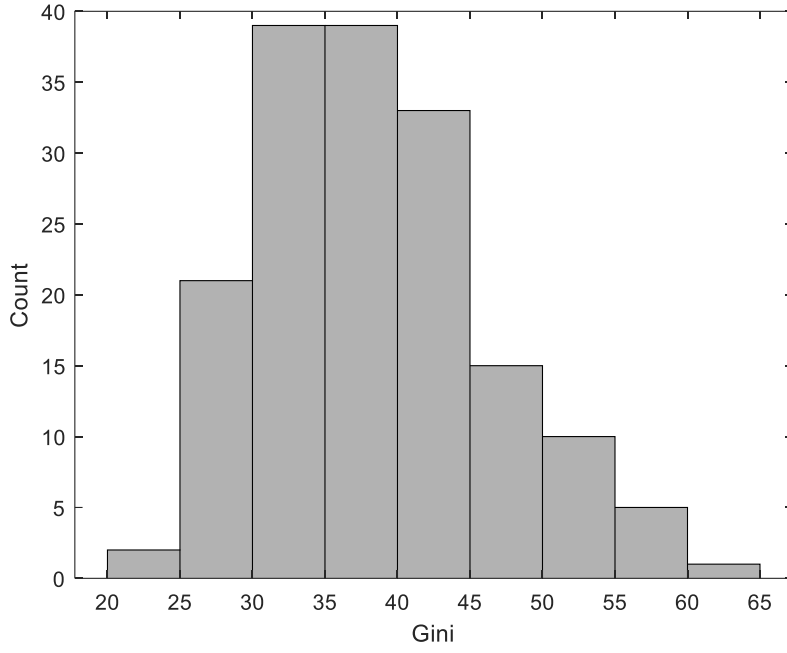


Figure 2. Histogram of Countries' Income Gini Indices (Percent).¹⁰

The cause of this drawback of the Gini index is due to the nature of its dispersion concept, its numerator also known as the Gini mean difference (GMD), whose empirical (or sample) version is:¹¹

$$GMD(X) = \frac{\sum_{i=1}^N \sum_{j=1}^N |x_i - x_j|}{2N^2}. \quad (6)$$

The GMD being the numerator in formula (4), it also reduces to

$$GMD(X) = \frac{2}{N^2} \sum_{i=1}^N ix_{(i)} - \frac{N+1}{N} \text{mean}(X). \quad (7)$$

The sensitivity of the GMD to a rank-preserving variation of a particular observation (holding the others constant) follows from (7) by simple differentiation:

$$\frac{\partial GMD(X)}{\partial x_{(k)}} = \frac{2}{N^2} \left[k - \frac{N+1}{2} \right]. \quad (8)$$

¹⁰ Data source: World Bank https://data.worldbank.org/indicator/SI.POV.GINI?most_recent_value_desc=true (the years vary across countries: but we took the most recent values for each country).

¹¹ Some authors define the GMD as twice the numerator of the Gini ratio. The properties of the GMD are reviewed e.g. in [20].

Thus, the sensitivity of the Gini index to an observation depends merely on the observation's rank, and a similar conclusion holds for the Gini ratio.¹² Compare the sensitivity of the variance, which depends on the size of the observation:

$$\frac{\partial \text{var}(X)}{\partial x_i} = \frac{2}{N}[x_i - \text{mean}(X)]. \quad (9)$$

Because empirical data are possibly contaminated by a few outliers, one usually assumes (for good reasons) that the lower a statistic's sensitivity to an individual observation, the better: hence a certain advocacy for robust dispersion measures like the mean absolute deviation (MAD), or the GMD, compared to the variance.¹³ Yet it is a mistake to transpose the outlier argument to a heavy-tailed distribution, which intrinsically exhibits extremes that, moreover, represent a sizable portion of the distribution. It is in fact a defining property of heavy-tailed variables, or at least the most common ones in applications, the so-called sub-exponential distributions (which includes regularly varying distributions, hence power laws), to exhibit the "one-big jump phenomenon", whereby a distribution is dominated by one observation (the maximum one), in the sense that for independent copies $\{X_1, \dots, X_N\}$ of a sub-exponential X we have (by definition):¹⁴

$$\frac{\text{prob}\{X_1 + \dots + X_N > x\}}{\text{prob}\{\max(X_1, \dots, X_N) > x\}} \rightarrow 1 \quad (x \rightarrow \infty, N \geq 2). \quad (10)$$

For power laws (or regularly varying distributions more generally), the one-big jump phenomenon presents itself in a specific manner that is the more striking, the smaller the power law's tail exponent α : in particular, when $\alpha < 1$ one can show that the maximum observation in a large random sample from a (positive) power

¹² For a more detailed such sensitivity calculation, see [9].

¹³ The MAD is even less sensitive to extremes in the sense that its sensitivity to an observation depends merely on the sign of this observation's distance from the median.

¹⁴ For a review of subexponential variables, see [21-23].

law is on average of the same order of magnitude as the sum of all observations [24, p. 465]:¹⁵

$$\mathbb{E}\left(\frac{X_1 + \dots + X_N}{\max\{X_1, \dots, X_N\}}\right) \rightarrow \frac{1}{1-\alpha} \quad (0 < \alpha < 1, N \rightarrow \infty). \quad (11)$$

If $\alpha = 0.5$, for example, one single observation in a sufficiently large sample will represent on average 50% of the sum of all observations (Figure 3).

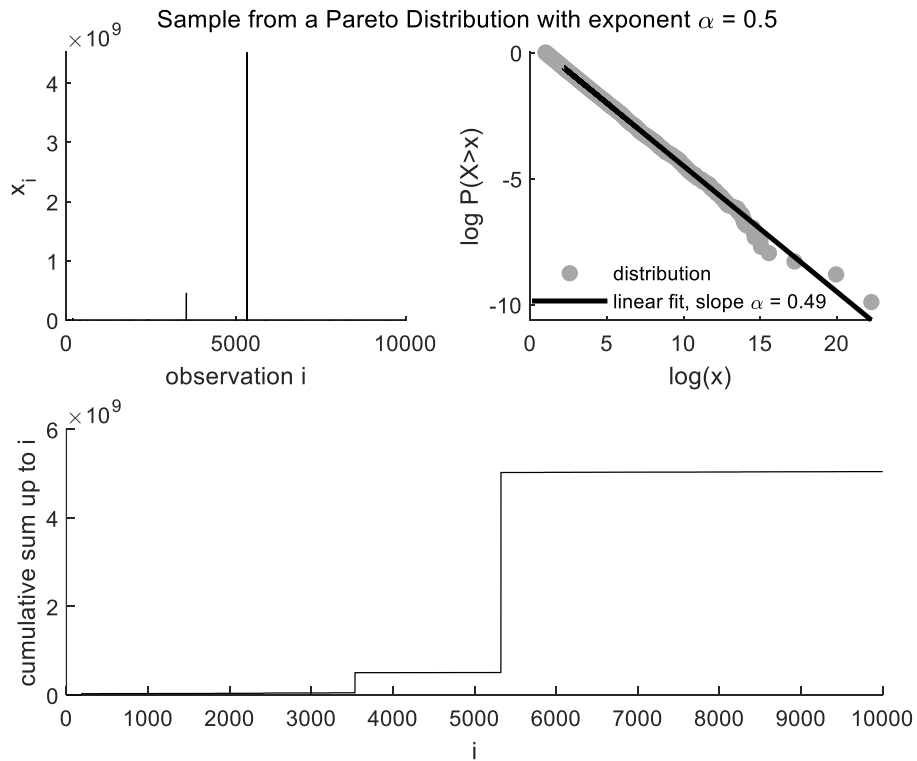


Figure 3. The one-big jump phenomenon (or dominance of a few observations) is an intrinsic feature of heavy-tailed distribution: here, a Pareto distribution with exponent 0.5; one observation (the 5324th one) accounts for much of the sum of the distribution.

Thus, any robust summary of a heavy-tailed variable is a poor summary of that distribution.

By its greater sensitivity to extremes, the variance is a more faithful measure of dispersion for power laws than the GMD. That the variance is infinite for a broader

¹⁵ For an intuitive derivation of this formula, and the corresponding ones for $\alpha > 1$, see [16, p. 106].

class of distributions can itself be counted as a virtue in this respect.¹⁶ The concept of infinite variance should not deter us, for an infinite theoretical (population) variance empirically simply means that the sample variance cannot reasonably be reduced to any point estimate, but rather it is a variable that fluctuates wildly from sample to sample.¹⁷ But if the variance is properly normalized into a ratio, through a denominator whose sample fluctuation is of the same order of magnitude as that of the variance, then it yields a statistically stable measure of variability. The right denominator to that effect is none other than the second moment: as it turns out, the variance normalized by the second moment is a natural inequality measure.

3 An Alternative Inequality Measure

We propose, therefore, for any (nonzero) empirical variable $X=[x_1, \dots, x_N]$, the inequality measure:

$$I(X) = \frac{\text{variance}(X)}{\text{mean}(X^2)}. \quad (12)$$

Clearly, $0 \leq I \leq 1$. Besides the obvious interpretation as a normalized variance, the ratio I has various other interpretations that illustrate its potential versatility as a general concept (beyond the specific use as an inequality measure). Suffice it here to mention a few known results across the sciences that implicitly involve the inequality index I and on which therefore new light may be shed if this connection is made explicit. Consider, to begin with, the following simple but powerful result in probability theory (usually known as the “second moment method”):

Second moment method. For any finite-variance $X \geq 0$ we have $\text{prob}\{X = 0\} \leq I(X)$.

¹⁶ The GMD of a distribution is finite whenever the mean of the distribution is finite; thus, the GMD of a power law with exponent $\alpha > 1$ is finite. More on this in Section 4.

¹⁷ The second moment of a large sample from an infinite-variance power law is an infinite variance Levy variable by the general central limit theorem (see Appendix): more precisely, if $\{X_1, \dots, X_N\}$ are independent copies of an (infinite-variance) power law with $\alpha \in]0, 2[$, then $N^{-2/\alpha} \sum_{i=1}^N x_i^2$ converges in distribution to Levy variable with index $\alpha/2$.

This result is a consequence of Cauchy-Schwarz inequality, it is known:

$$\mathbb{E}(X) = \mathbb{E}(X1_{\{X>0\}}) \leq [\mathbb{E}(X^2)]^{1/2} [\mathbb{E}(1_{\{X>0\}}^2)]^{1/2} = [\mathbb{E}(X^2)]^{1/2} [\text{prob}\{X > 0\}]^{1/2},$$

from which (13) follows, if one solves for $\text{prob}\{X > 0\}$.

The I index has also a series of potential applications, notably in economics, inherited from its simple relationship with the Hirschman-Herfindahl index [25, 26]:

$$H(X) = \frac{\sum_{i=1}^N x_i^2}{(\sum_{i=1}^N x_i)^2}. \quad (14)$$

Clearly, we have¹⁸

$$I(X) = 1 - \frac{1}{NH(X)}. \quad (15)$$

Thus, any interpretation of the H index can be rephrased in terms of the I index. The H index is of course primarily used as a concentration measure: H is then interpreted as measuring an inverse effective number of observations in a distribution (the number of dominant elements in it so to speak). That is, a variable X is mostly concentrated on $1/H(X)$ number of its elements (e.g. $1/H(X)=2.05$ for $X=[1,2,1,90,110]$); in particular, perfect equality means $1/H(X)=N$, and perfect inequality is approached by a large distribution concentrated on one individual, for which $1/H(X)=1-1/N$. It makes sense therefore to measure the degree of equality in a distribution X by the fraction of its dominant elements, namely by the ratio $[1/H(X)]/N = 1 - I(X)$, or the effective number of elements in X relatively to that of a perfectly equal distribution of the same size.

A similar interpretation of I follows from the following probabilistic interpretation of H . To any nonnegative variable $X = [x_1, \dots, x_N]$ we associate the two probability systems $P = [p_i]$ and $Q = [q_i]$, where $p_i = 1/N$ and $q_i = x_i / \sum_k x_k$ ($i = 1, \dots, N$), and

¹⁸ An author, G. Prathap, brought to my attention an unpublished draft of his (https://www.academia.edu/2562606/A_Tale_of_Four_Indices_-) in which he mentions this formula as an inequality measure.

denote by \mathbb{E}_p and \mathbb{E}_Q the corresponding expectation operators. The probability systems can be called the uniform versus “rich-get-richer” probability measures (as would suggest the following experiment: pick at random a value from X , where the probability of picking x_i is p_i , and increase the picked value by some amount; consider the same experiment, but with probability q_i instead.) We have $\mathbb{E}_p(P) = 1/N$ and $\mathbb{E}_Q(Q) = H$. Thus, $1 - I(X)$ measures the degree of equality in X by comparing the two associated probability measures through the ratio $\mathbb{E}_p(P)/\mathbb{E}_Q(Q)$.

The index I also has an informational interpretation: since the inverse ratio $1/H$ measures the effective number of elements in a distribution, it follows that H itself measures the frequency of each element in the distribution adjusted for the relative size of each element: that is, H is effectively a probability. But since to each probability, one can associate an information measure by taking the log-probability, we have an information interpretation of H in terms of

$$R_2(Q(X)) = -\log H(X), \quad (16)$$

namely the second order Rényi entropy of the “rich get richer” probability system associated to X . The Rényi entropy [27] of a probability system $P = [p_1, \dots, p_n]$ is more generally defined as

$$R_\lambda(P) = \frac{1}{1-\lambda} \log \left(\sum_{i=1}^N p_i^\lambda \right) \quad (\lambda \geq 0, \lambda \neq 1). \quad (17)$$

The I index is also a normalized version of the second order generalized inequality entropy measure, the general class of which plays a central role in the axiomatic approach to inequality measurement discussed shortly:

$$E_2 = \frac{1}{2} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i}{\text{mean}(X)} \right)^2 - 1 \right] = \frac{1}{2} [NH - 1]. \quad (18)$$

In other words

$$I = \frac{2E_2}{2E_2 + 1}. \quad (19)$$

Finally, the new index has this other general interpretation also inherited from the H index: usually this latter emerges as a measure of how much of the variability of an aggregate variable (e.g. GDP [17]) is due to the typical fluctuation of the individual components (firms' sales), provided these latter are statistically independent from one another. That is, the following interpretation holds in many contexts:

$$H = \frac{\text{aggregate volatility}}{\text{individual volatility}}. \quad (20)$$

In terms of the new inequality index, we have

$$\frac{\text{aggregate volatility}}{\text{individual volatility}} = \frac{1}{N} \frac{1}{1-I}. \quad (21)$$

Thus, a large aggregate volatility can emerge even in a very large system of independent components if these latter are highly unequal: $1/N$ is the damping factor due to the aggregation of independent homogenous components; but this damping factor is offset if the individual components are highly unequal. The argument in terms of the H index is known [17].¹⁹ But the reformulation in terms of the new index uncovers an

$$\text{Inequality Multiplier} = \frac{1}{1-I} = \sum_{k=0}^{\infty} I^k. \quad (22)$$

Any situation where the H index plays a role, so could also do the inequality multiplier; and this is potentially the case of any sum of individual variables. Consider, for example, the return of an index or portfolio of assets, which is the average return of the individual assets, say $R = \sum_{i=1}^N x_i R_i$, where $[x_i] = X$ are the asset weights. In the simple case of independent, identically distributed, zero-mean individual asset returns, the portfolio's volatility is $\text{var}(R|X) = H(X)\text{var}(R_1)$.²⁰

Relaxing the assumption of independent individual components, one can also define a linkage multiplier, the part of aggregate fluctuation due to the network

¹⁹ According to [17], the micro shocks of 100 largest firms in the US could account for one-third of aggregate fluctuations of the country's GDP.

²⁰ The H index appears in portfolio theory as an effective portfolio diversification measure [28, 29].

amplification or propagation of interdependent individual shocks [30], or even consider a multiplier due to the inequality of network degrees (since some individuals are more heavily connected than others), usually measured by the coefficient of variation (CV) of the degrees [30]. The link between CV and I is a direct one:²¹

$$I = \frac{CV^2}{CV^2 + 1}. \quad (23)$$

To summarize: the new index has potentially a range of applications due to its simple connections with a few core concepts across the sciences, connections given by the conversion formulas (already proven or easy to establish):

$$NH = CV^2 + 1 = (1 - I)^{-1} = \sum_{k=0}^{\infty} I^k, \quad (24)$$

$$CV^2 = I(1 - I)^{-1} = \sum_{k=1}^{\infty} I^k. \quad (25)$$

$$R_2 = \log[N(1 - I)]. \quad (26)$$

$$E_2 = \frac{1}{2} \frac{I}{1 - I}. \quad (27)$$

As an inequality measure, more specifically, the I index enjoys the axioms of the normative approach to inequality measurement, which requires that an inequality measure $J = J(X)$ satisfy a list of properties, notably:

1. Normalization: $0 \leq J(X) \leq 1$.
2. Scale invariance: $J(\lambda X) = J(X)$ for $\lambda > 0$.

²¹ In this form (23), the index I appeared (p. 183) as the dual of CV in a recent book [31], as was pointed in a comment to an earlier draft of this paper by one of the authors of that book: the dual $J^*(X)$ of an inequality measure $J(X)$, as I understand it from the comment, is defined in that book as follows: to a variable X , associate the hypothetical transform variable $X^* = [a, \dots, a, 0, \dots, 0]$, $a > 0$, such that $J(X) = J(X^*)$; then the dual inequality index $J^*(X)$ is the proportion of zero elements in X^* .

3. Anonymity: $J(X)$ should be invariant to the way the elements of X are indexed (invariance to permutations of X).
4. Transfer principle: a rank-preserving transfer from a greater to lower element of X (from a richer to a poorer person) should reduce inequality.
5. Population principle: $J(X, X) = J(X)$ (invariance to pooled replications of X).
6. Decomposability: if X comes in subgroups, then J should be decomposable into inequality within subgroups and inequality between subgroups.

We know of no inequality measure in the literature that satisfy all the 6 axioms. The Gini index satisfies Axioms 1-5, but not necessarily Axiom 6 (decomposability), except in the case of nonoverlapping subgroups [32, Appendix A]. By an important theorem [33-35], an inequality measure $J(X)$, continuous with respect to each element of X , satisfies the principle of transfers, scale independence, and decomposability if and only if it is some increasing function of a generalized entropy measure (where θ can be any real number including $\theta = 1$ if interpreted as $\theta \rightarrow 1$)²²

$$E_\theta = \frac{1}{\theta(\theta-1)} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i}{\text{mean}(X)} \right)^\theta - 1 \right]. \quad (28)$$

Proposition 2. *The index I satisfies the Axioms 1-6 listed above.*

Proof. The index I obeys the population principle and anonymity in a straightforward way, and it is normalized. Since $I = 2E_2/(2E_2 + 1)$, a strictly increasing function of E_2 , it satisfies the principle of transfers, scale independence, and decomposability by virtue of the above-mentioned theorem. ■

One can also derive the 6 properties for I by direct calculation; for example, the transfer principle: a rank-preserving transfer $t > 0$ from richer to poorer, or $(x_i, x_j) \rightarrow (x_i - t, x_j + t)$, is by definition one such that $x_i - t > x_j + t$, hence it is one such that $2t(x_j - x_i) < -4t^2$. The transfer preserves $\text{mean}(X)$ but decreases the second moment:

²² This is a central theorem of the axiomatic approach [36: see Theorem 4]. See also [37].

$\text{mean}(X^2) \rightarrow \text{mean}(X^2) + 2t^2 + 2t(x_j - x_i)$. Thus, the transfer lowers I , namely $1 - (\text{mean}(X))^2 / \text{mean}(X^2)$. The decomposition of I by subgroup components can be established from the corresponding decomposition formula of the generalized entropy E_2 through a standard (if tedious) procedure [32, Appendix A].

4 Comparing the Two Inequality Measures

At least two reasons are in favor of the new index over the Gini (that I reflects better extreme variability and that it is always decomposable). A systematic empirical investigation of the two inequality measures using income or wealth distributions (in view notably of the first argued limitation of the Gini index) is beyond this paper's scope: for such study to be conclusive one would need to work with complete income or wealth distributions, rather than the relatively aggregated and often partial data, coming, say, by income classes, such as the US census data used in Figure 4. (We should not expect a major difference between the two indices using aggregated data that even out extremes, although we note a turning point around 1995 when G becomes greater than I .)²³ Thus we will be content here with theoretical comparisons of the two indices illustrated, when needed, with simulated heavy-tailed data.

²³ Source: Historical Income Tables, census.gov. These are income data (in dollars) grouped in classes: [under 15000], [25,000 to 34,999], ..., [200000 and over]. We take the lowest income to be zero and roughly estimate the highest income in such a way as to recover the summary statistics (such as the mean income) reported in the data. An overview of the limitations of common inequality data can be found e.g. in the recent review paper [38].

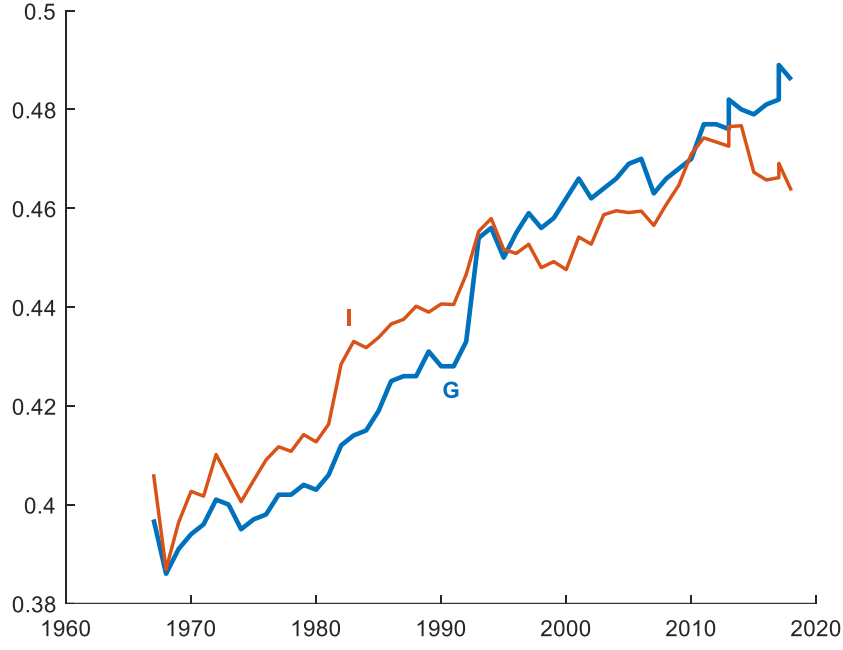


Figure 4. The inequality measures G and I for the USA census income data.

The two measures I and G share a few similarities. They coincide for at least two distributions: the exponential one, for which $I=G=0.5$, and the Bernoulli one, or the indicator function of an event E , for which $I = G = 1 - \text{prob}(E)$. For a Pareto with exponent α , we have $G = G(\alpha) = (2\alpha - 1)^{-1}$ if $\alpha > 1$, as already mentioned earlier; in contrast $I = I(\alpha) = (\alpha - 1)^{-2}$ for $\alpha > 2$, as one can establish using the moment formula for the Pareto mentioned in the brief reminder on power laws (Section 2). Simple calculations yield $I(\alpha) \geq G(\alpha)$ if $2 < \alpha < 2 + \sqrt{2}$ and $I(\alpha) \leq G(\alpha)$ if $\alpha \geq 2 + \sqrt{2}$.

Conceptually, moreover, both indices are just normalized dispersion measures. In fact, the two indices belong to a general class of inequality measures:

$$I_p(X) = \frac{\frac{1}{2} \mathbb{E} |X - X'|^p}{\mathbb{E} |X|^p}, \quad (29)$$

where, again, X' is an independent copy of X , and $p \geq 1$.²⁴ The general index I_p is well-defined for any nonzero variable X with $\mathbb{E}|X|^p < \infty$, and it is in the range $[0,1]$. This is true by the triangle inequality:²⁵

$$(\mathbb{E}|X - X'|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|X'|^p)^{1/p} = 2(\mathbb{E}|X|^p)^{1/p}. \quad (30)$$

For $p=2$, the numerator in (29) is $\mathbb{E}(X - X')^2/2 = \mathbb{E}(X^2 - 2X'X + X'^2)/2 = \text{var}(X)$, so $I_2 = I$. For $p=1$ we recover G if $X \geq 0$ and $\mathbb{E}(X) > 0$. In fact we should define the Gini index more generally as $G(X) = I_1(X)$, namely

$$G(X) = \frac{\frac{1}{2} \mathbb{E}|X - X'|}{\mathbb{E}|X|}. \quad (31)$$

As announced earlier (Section 2), this general definition circumvents the undesirable behavior of the common definition of G vis-à-vis signed variables. In particular, if X is a symmetric variable (that is, X and $-X$ have the same distribution), then $G(X) = G(-X)$ by the extended definition (31). However, one can show that $G = 1/\sqrt{2} \approx 0.7$ for the Gaussian, suggesting this latter is more unequal distribution than an exponential variable, for example; more generally, a significant presence of negatives in a distribution tends to produce relatively high values for both indices, which is perhaps intuitively reasonable: a society with symmetric income distribution would be one in which only half of the population have positive incomes and the other half are indebted to them, an extreme form of inequality! This is even better reflected in the new index, since $I(X) = 1$ if $\text{mean}(X) = 0$. But then both indices (especially the new one) are poor measures of heavy-tailed-ness for signed variables (by ranking higher a Gaussian compared to say a Pareto with index 1.5, for example: the next

²⁴ If we define dispersion more strictly in terms of the distance function $(X, Y) \mapsto \|X - Y\|_p$ induced by the norm $X \mapsto \|X\|_p = (\mathbb{E}|X|^p)^{1/p}$, then we would consider instead $I^{1/p}$ as inequality measure. But statisticians use more flexible “distance” concepts (such as a square distance).

²⁵ The triangle inequality is applied, that is, to the distance function referred to in the preceding note.

paragraphs define I for infinite-variance variables). This is easily fixed however: the heavy-tailed-ness of X can be measured by $I(|X|)$.

We close the comparison of the two indices by a discussion of their respective statistical stability. One might be inclined to prejudge that I is a more volatile statistic compared to G , and to assume that I is the more volatile, the heavier the tail of X . But this is only true of the ratios' numerators. In fact, I is more stable statistically than G for infinite-variance variables, paradoxically: moreover, we have the curious pattern that $I(X)$ is the more stable compared to $G(X)$, the more volatile is X ; and the reverse pattern for G (Figure 5).

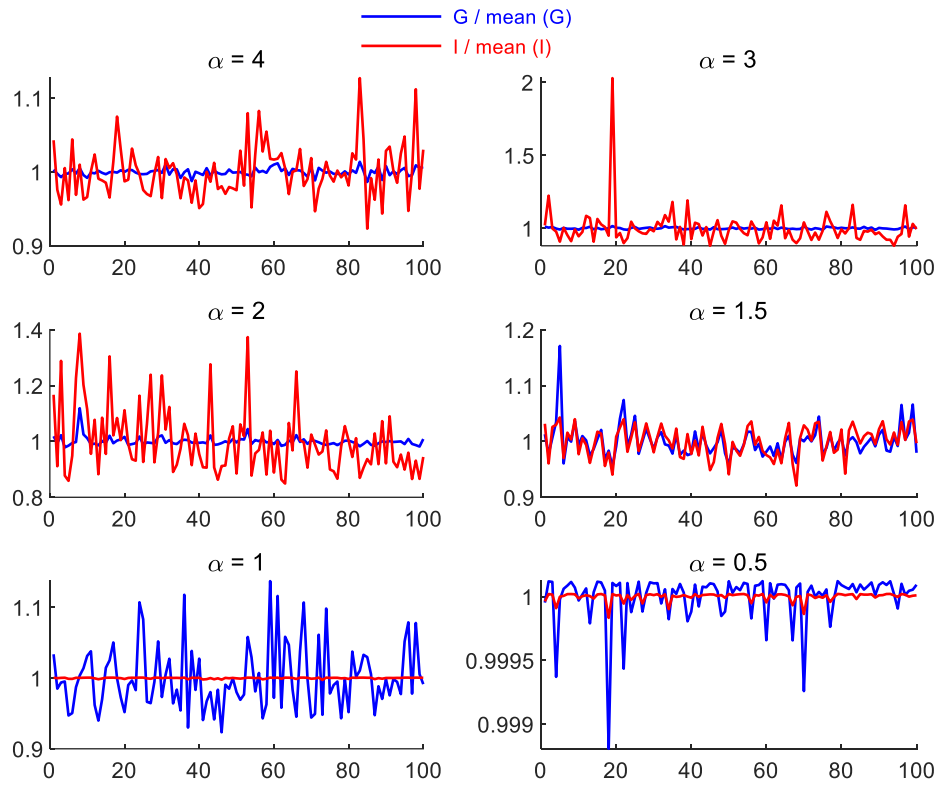


Figure 5. Statistical Stability of G versus I for Pareto variables (100 large samples of 100000 random draws). Notice the curious pattern: the lower is alpha (hence the heavier-tailed the Pareto), the more stable is I (red) compared to G (blue), and vice versa.

The slow convergence pattern of G is not surprising per se and it has been documented [8]. What is surprising is the inverse pattern of I compared to that of G .

This suggests a scrutiny of the limit behavior of the I index. Formally, we model the possible sample realization of $I(X)$ by the following estimator, based on N independent copies $\{X_1, \dots, X_N\}$ of X , to investigate its large-sample behavior:

$$I_N(X) = 1 - \frac{\sum_{i=1}^N X_i^2}{N(\sum_{i=1}^N X_i)^2}. \quad (32)$$

Because $I_N(X)$ involves sums of independent copies of X and X^2 , its limit behavior is regulated by the general central limit theorem (in the same way as for the H index, whose limit behavior for a Pareto with $\alpha \geq 1$ is known [17, prop. 2]).

Proposition 3. *If X is a positive power law with exponent $\alpha \in (0, 2)$, then $I_N(X) \rightarrow 1$ as $N \rightarrow \infty$. Moreover, $1 - I_N(X) \rightarrow 0$ like N^{-1} if $\alpha \in (0, 1)$, $N^{-1+2/\alpha}$ if $\alpha \in (1, 2)$, and $N^{-1}(\log N)^2$ if $\alpha = 1$ and X is Pareto.²⁶*

This proposition, proven in the Appendix, explains the above-mentioned seemingly paradoxical statistical stability of the index I (Figure 5): the heavier the tail of the power law, the more stable the index I , with the possible exception of $\alpha = 1$ (which more generally is a singular parameter in the theory of the limit sum of power laws).

²⁶ We assume $X \geq 0$ merely to avoid worrying about an additional condition: see Appendix.

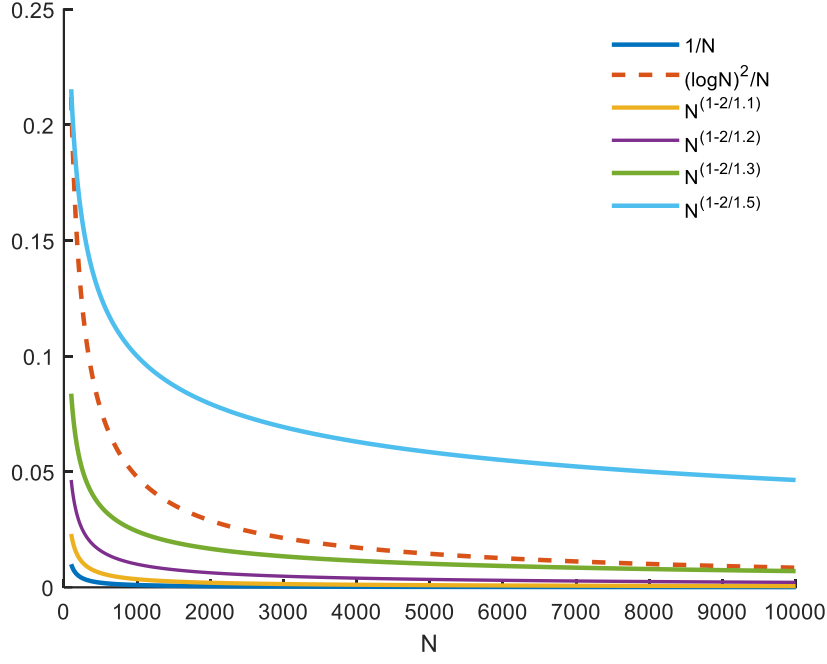


Figure 6. Convergence rate of the empirical (or sample) inequality index $I_N(X)$ for infinite-variance Pareto variable X . The lower is the tail exponent α , the more rapid the convergence rate (for $\alpha \neq 1$).

Proposition 3 suggests that we extend the definition of the I index to a theoretical variable X as the large sample limit of its empirical version, namely as $I(X) = \lim_{N \rightarrow \infty} I_N(X)$, which for a finite-variance distribution reduces of course to $\text{var}(X)/\mathbb{E}(X^2)$, by the law of large numbers, and for an (infinite-variance) power law X with $\alpha \in (0, 2)$ is $I(X) = 1$ (by Proposition 3). An apparent limitation of this extension to infinite-variance theoretical distributions is that it does not reflect the gradation of inequality for power laws of exponent in the range $\alpha \in (0, 2)$, a scale of inequality that we know is inversely related to α . But this is hardly a real problem, for the gradation of inequality for power laws with $\alpha \in (0, 2)$ is recovered empirically in the speed of the convergence $I_N(X) \rightarrow 1$, which is the more rapid, the lower is α , so that for a comparably large sample size N , the index $I_N(X)$ tends to be higher, the lower is α .

The limit behavior of the Gini index, on the other hand, is that of the estimator

$$G_N(X) = 2 \frac{\sum_i^N iX_{(i)}}{N \sum_i^N X_i} - 1. \quad (33)$$

The study of $G_N(X)$ involves more care and technique because $G_N(X)$ involves a sum of ranked (hence dependent) variables in the numerator. For a power law X , a characterization of the limit behavior of $G_N(X)$ has already been done for the case $1 < \alpha < 2$ [8], which show the slow convergence of $G_N(X)$ for infinite-variance power laws: the complement case $0 < \alpha < 1$ would involve essentially the same technicalities, which we do not repeat here. We mention however:

Proposition 4. *If X is a power law with exponent $\alpha < 1$, then $G_N(X) \rightarrow 1$ as $N \rightarrow \infty$.*

In lieu of a formal but long proof, we just emphasize here a simple intuitive argument as to why this result is to be expected (referring the reader to [8] for the preliminary technique needed before one can apply the general central limit theorem for a formal proof). The intuitive argument is based on Atkinson's formula [39], according to which "when a very top group of the income distribution, infinitesimal in numbers, owns a finite share S of total income, the Gini coefficient G can be approximated by $G^*(1-S) + S$, where G^* is the Gini coefficient for the rest of the population" [40]. In a large sample from a power law with $\alpha < 1$, the highest "income" on average has the share $S = 1 - \alpha$ according to equation (11), and we also have $G = G^*$ approximately (because G is only mildly affected if one observation, here the biggest one, is removed). So $G = \alpha G + (1 - \alpha)$, hence $G \equiv \lim_{N \rightarrow \infty} G_N = 1$.

5 Summary

Beyond the specific discussion on inequality measurement, this paper revisits received doctrine on robustness, whose desirability might be questionable for heavy tailed variables. The Gini index remains a powerful inequality measure for light-

tailed distributions potentially distorted by outliers. But the variance normalized by the second moment better suits heavy-tailed variables, and it satisfies the normative axioms of inequality measurement, including decomposability. As a concept, the new index is potentially useful beyond the specific purpose of inequality measurement due to its simple relationships with fundamental concepts across the sciences, such as: variance, Herfindahl-Hirschman index, and Rényi entropy.

Appendix

For two functions F and G of x , the notation $F \sim G$ means $F(x)/G(x) \rightarrow 1$ as $x \rightarrow \infty$. For variables $\{X_N\}$ and Y , $X_N \sim a_N + b_N Y$ means the convergence in distribution $(X_N - a_N)/b_N \rightarrow Y$ holds. Many claims in the main text rely on the following [13: sec. 2.2, th. 2.2.15, 14: sec. 3.8, th. 3.8.2]:

Limit Theorem for Power Laws. Let $\{X_1, \dots, X_N\}$ be independent copies of a variable X with $\text{prob}\{|X| \geq x\} \sim Cx^{-\alpha}$, $C > 0$, $\alpha \in (0, 2)$, and $\lim_{x \rightarrow \infty} [\text{prob}\{|X| \geq x\} / \text{prob}\{X \geq x\}] = \theta \in [0, 1]$. Then $N^{-1/\alpha}(\sum_{i=1}^N X_i - b_N) \rightarrow \Lambda(\alpha)$ as $N \rightarrow \infty$, where $\Lambda(\alpha)$ is a Levy variable with exponent α , and $b_N = 0$ if $\alpha \in (0, 1)$, $N\mathbb{E}(X)$ if $\alpha \in (1, 2)$, and $N\mathbb{E}(X1_{|X| \leq N})$ if $\alpha = 1$.

For a Pareto $X \geq 1$, the theorem says $N^{-1/\alpha}(\sum_{i=1}^N X_i - b_N) \rightarrow \Lambda(\alpha)$ as $N \rightarrow \infty$, where

$$b_N = \begin{cases} 0, & \alpha < 1 \\ N \log N, & \alpha = 1 \\ E(X)N, & \alpha > 1. \end{cases} \quad (34)$$

Proof of Proposition 3. Without loss of generality, assume $X \geq 0$ (otherwise, simply add the condition $\theta \in [0, 1]$). We investigate the limit behavior of

$$I_N(X) = 1 - \frac{(\sum_{i=1}^N X_i)^2}{N \sum_{i=1}^N X_i^2}.$$

By the limit theorem, $\sum_{i=1}^N x_i \sim b_N(\alpha) + N^{1/\alpha} \Lambda(\alpha)$ and $\sum_{i=1}^N x_i^2 \sim N^{2/\alpha} \Lambda(\alpha/2)$.²⁷ Thus

$$I_N(X) \sim 1 - \frac{[b_N(\alpha) + N^{1/\alpha} \Lambda(\alpha)]^2}{N^{1+2/\alpha} \Lambda(\alpha/2)}. \quad (35)$$

Hence

²⁷ The second sum is just an application of the theorem to X^2 which is a Pareto with exponent $\alpha/2$, since $\text{prob}\{X^2 \geq z\} = \text{prob}\{X \geq z^{1/2}\} = (z^{1/2})^{-\alpha}$.

$$1 - I_N(X) \sim \begin{cases} \frac{[\Lambda(\alpha)]^2}{\Lambda(\alpha/2)} \frac{1}{N}, & \alpha \in (0,1) \\ \frac{[\mathbb{E}(X)]^2}{\Lambda(\alpha/2)} N^{1-2/\alpha}, & \alpha \in (1,2). \end{cases} \quad (36)$$

In particular, $I_N(X) \rightarrow 1$ as $N \rightarrow \infty$ in either case. The case $\alpha = 1$ requires special care; for a Pareto, we have $1 - I_N(\alpha) \sim C_1 N^{-1} [\log N]^2 / \Lambda(1/2) \rightarrow 0$. More generally, $b_N(1) = N \mathbb{E}(X 1_{X \leq N}) = N \int_0^N L(x) x^{-1} dx$, where by assumption $L(x) \rightarrow C$ as $x \rightarrow \infty$, a slowly-varying function, which therefore obeys $[L(N)]^{-1} \int_0^N L(x) x^{-1} dx \rightarrow \infty$ as $N \rightarrow \infty$ [23, Remark 1.2.7], hence $\int_0^N L(x) x^{-1} dx \rightarrow \infty$ [since $L(N) \rightarrow C$], that is, $b_N(1)/N \rightarrow \infty$.

This suggests writing (35) for $\alpha = 1$ as follows:

$$[1 - I_N(X)]^{1/2} \sim \frac{[b_N(1) + N\Lambda(1)]}{N^{3/2} [\Lambda(1/2)]^{1/2}} = \frac{1 + [N/b_N(1)]\Lambda(1)}{[b_N(1)/N] N^{5/2} [\Lambda(1/2)]^{1/2}} \rightarrow 0. \quad (37)$$

All in all, $I_N(X) \rightarrow 1$ for every $\alpha \in (0,2)$. ■

References

- [1] C. Gini, Variabilita e mutabilita: contributo allo studio delle relazioni statistiche, Studi Economico-giuridici. Facolta di Giurisprudenza della R. Universita di Cagliari, 3 (1912).
- [2] M.O. Lorenz, Methods of measuring the concentration of wealth, Publications of the American statistical association, 9 (1905) 209-219.
- [3] S. Yitzhaki, More than a dozen alternative ways of spelling Gini, Research on Economic Inequality, 8 (1998) 13-30.
- [4] C.-N. Chen, T.-W. Tsaur, T.-S. Rhai, The Gini coefficient and negative income, Oxford Economic Papers, 34 (1982) 473-478.
- [5] G. Deltas, The small-sample bias of the Gini coefficient: results and implications for empirical research, Review of Economics and Statistics, 85 (2003) 226-234.
- [6] T. Van Ourti, P. Clarke, A simple correction to remove the bias of the Gini coefficient due to grouping, Review of Economics and Statistics, 93 (2011) 982-994.
- [7] F. Cowell, E. Flachaire, Inequality Measures and the Median: Why inequality increased more than we thought, in: Seventh Meeting of the Society for the Study of Economic Inequality (ECINEQ), New York City, 2017.
- [8] A. Fontanari, N.N. Taleb, P. Cirillo, Gini estimation under infinite variance, Physica A, 502 (2018) 256-269.
- [9] J.L. Gastwirth, Is the Gini index of inequality overly sensitive to changes in the middle of the income distribution?, Statistics and Public Policy, 4 (2017) 1-11.
- [10] S. Foss, D. Korshunov, S. Zachary, An introduction to heavy-tailed and subexponential distributions, Springer-Verlag, New York, 2013.
- [11] M.E. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics, 46 (2005) 323-351.
- [12] P. Lévy, Théorie des erreurs. La loi de Gauss et les lois exceptionnelles, Bulletin de la Société mathématique de France, 52 (1924) 49-85.
- [13] P. Embrechts, C. Kluppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer Science & Business Media, Berlin, 1997.
- [14] R. Durrett, Probability: Theory and Examples, 5 ed., Cambridge University Press, Cambridge, 2019.
- [15] J.-P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications, Physics Reports, 195 (1990) 127-293.
- [16] D. Sornette, Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools, 2 ed., Springer, Berlin, 2006.
- [17] X. Gabaix, The granular origins of aggregate fluctuations, Econometrica, 79 (2011) 733-772.
- [18] V. Pareto, La courbe de la répartition de la richesse, Viret-Genton, Lausanne, 1896.

- [19] X. Gabaix, Power laws in economics: An introduction, *The Journal of Economic Perspectives*, 30 (2016) 185-205.
- [20] S. Yitzhaki, Gini's mean difference: A superior measure of variability for non-normal distributions, *Metron*, 61 (2003) 285-316.
- [21] J.L. Teugels, The class of subexponential distributions, *The Annals of Probability*, 3 (1975) 1000-1011.
- [22] C.M. Goldie, C. Klüppelberg, Subexponential distributions, in: R.E.F. Robert J. Adler, Murad S. Taqqu (Ed.) *A practical guide to heavy tails: statistical techniques and applications*, Birkhauser Boston, Cambridge, MA, 1998, pp. 435-459.
- [23] T. Mikosch, Regular variation, subexponentiality and their applications in probability theory, in: *Eindhoven University of Technology Eindhoven*, 1999.
- [24] W. Feller, *An Introduction to Probability Theory and its Applications*, 2 ed., John Wiley & Sons, New York, 1966 [1971].
- [25] A. Hirschman, *National Power and the Structure of Foreign Trade*, University of California Press, Berkeley, 1945.
- [26] O.C. Herfindahl, *Concentration in the Steel Industry*, in: *Columbia University*, 1950.
- [27] A. Rényi, On measures of entropy and information, in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, University of California Press, 1961, pp. 547-561.
- [28] W. Woerheide, D. Persson, An index of portfolio diversification, *Financial Services Review*, 2 (1992) 73-85.
- [29] J.-P. Bouchaud, M. Potters, J.-P. Aguilar, Missing information and asset allocation, *arXiv preprint cond-mat/9707042*, (1997).
- [30] D. Acemoglu, V.M. Carvalho, A. Ozdaglar, A. Tahbaz-Salehi, The network origins of aggregate fluctuations, *Econometrica*, 80 (2012) 1977-2016.
- [31] R. Hoffmann, D. Botassio, J. Jesus, *Distribuição de renda: medidas de desigualdade, pobreza, concentração, segregação e polarização*, Editora da Universidade de São Paulo, São Paulo 2019.
- [32] F. Cowell, *Measuring Inequality*, Oxford University Press, Oxford, 2011.
- [33] F. Bourguignon, Decomposable income inequality measures, *Econometrica*, 47 (1979) 901-920.
- [34] F.A. Cowell, On the structure of additive inequality measures, *The Review of Economic Studies*, 47 (1980) 521-531.
- [35] A.F. Shorrocks, The class of additively decomposable inequality measures, *Econometrica*, 48 (1980) 613-625.
- [36] F.A. Cowell, Measurement of Inequality, in: *Handbook of Income Distribution*, 2000, pp. 87-166.

- [37] F.A. Cowell, E. Flachaire, Inequality with ordinal data, *Economica*, 84 (2017) 290-321.
- [38] T. McGregor, B. Smith, S. Wills, Measuring inequality, *Oxford Review of Economic Policy*, 35 (2019) 368-395.
- [39] A.B. Atkinson, T. Piketty, Top incomes over the twentieth century: a contrast between continental european and english-speaking countries, in: A.B. Atkinson, T. Piketty (Eds.), *Oxford University Press*, Oxford, 2007.
- [40] F. Alvaredo, A note on the relationship between top income shares and the Gini coefficient, *Economics Letters*, 110 (2011) 274-277.