How to (not) estimate Gini indices for fat tailed variables

Nassim Nicholas Taleb*
*Tandon School of Engineering, NYU, and Real World Risk Institute, LLC.

***Abstract*—Direct measurements of Gini index by conventional arithmetic calculations are poor estimators, even if paradoxically, they include the entire population, as because of super-additivity they cannot lend themselves to comparisons between units of different size; further, intertemporal analyses are vitiated by the population changes. The Gini of aggregated units A and B will be higher than those of A and B computed separately. This effect becomes more acute with fatness of tails. When the sample size is smaller than entire population, the estimation error is extremely high.**

The subadditivity has been proved by Zagier in 1983; we found a new proof through a more general lemma for ordered subsums applicable for ranked variables. We also show the effect of kurtosis on the subadditivity and the consequences for statistical estimation.

We compare the standard methodologies to the indirect methods via maximum likelihood estimation of tail exponent.

The conventional literature on Gini index cannot be trusted and comparing countries of different sizes makes no sense; nor does it make sense to make claims of "changes in inequality" based on such a measure.

We suggest a simple but efficient methodology to calculate the Gini index.

I. Introduction/Summary

Consider 10 separate countries, cities, or other units of equal size, with population 10^3. Assume the wealth in each unit follows a power law distribution, say a Pareto-Lomax, all with the exact same parameters. Assume a tail exponent of 1.1. The average Gini index as obtained by direct measurement will be $\approx .71$ per country. Now aggregate them into a single country. The composite Gini —as traditionally and currently measured — will be $\approx .75$, that is 6% higher –for the *same* sample. This inconsistency implies not only that the Gini cannot lend itself to comparisons between units of different size but that intertemporal assessments are vitiated by the population changes.

Further, the sampling error remains high throughout.

The effect is similar to the one about percentile in [1].

This note shows that

- as a consequence of the inequality, the Gini index obtained by conventional "direct" measurement as estimator is not consistent, downward biased and lends itself to illusions
- the superadditivity increases with the variance and fat-tailedness
- maximum likelihood (ML) parametrization of tail exponent is more efficient, unbiased, and economical of data: its error rate can be more than one order of magnitude smaller than the "direct" Gini measurement.

Further, wee get explicit distributions for the maximum likelihood estimator.

Table I shows biases and errors in the computation the Gini index via different methods, which presents our story and its conclusion. It compares the Gini index obtained by conventional arithmetic calculations to the Maximum Likelihood estimation via tail exponent by varying the population size n. These calculations are for the same data, generated by Monte Carlo simulations (10^8) for a Pareto distribution with exponent $\alpha = 1.1$, meaning finite mean and infinite variance (close to the "Pareto 80/20" in popular and managerial discussions). For the first category, "direct", we estimated the Gini using conventional methods of summing individual units (say wealth or income per person, or another unit in the physical domain). For the second, corresponding to Maximum Likelihood methods (ML), we estimated the tail exponent from the data using ML estimation methods and expressed the corresponding Gini.

TABLE I
Comparison of direct Gini to ML estimator, assuming tail $\alpha = 1.1$

n (popul or sample)	Direct Mean	Direct Bias	Direct STD	ML Mean	ML STD	Error ratio
10^3	0.711	-0.122	0.0648	.8333	0.0476	1.4
10^4	0.750	-0.083	0.0435	.8333	0.015	3
10^5	0.775	-0.058	0.0318	.8333	0.0048	6.6
10^6	0.790	-0.043	0.0235	.8333	0.0015	156
10^7	0.802	-0.033	0.0196	.8333	≈ 0	$> 10^5$

We note that there are many wealth distributions, and we took the one with the thickest tails, in the class of powerlaws. Thinner tail distributions do not generate significant bias.

II. The Gini estimated and its Superadditivity

A. "Direct" Estimators

Where g is the Gini index and X and X' are independent (etc., etc.) with mean μ:

$$g = \frac{1}{2}\frac{\mathbb{E}\left(|X - X'|\right)}{\mu}. \tag{1}$$

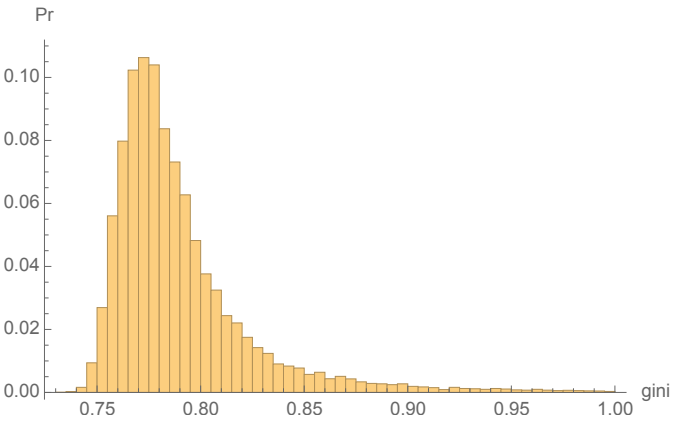


Fig. 1. Histogram of the distribution of direct estimation, population $= 10^6$. We notice a long right tail bounded at 1.

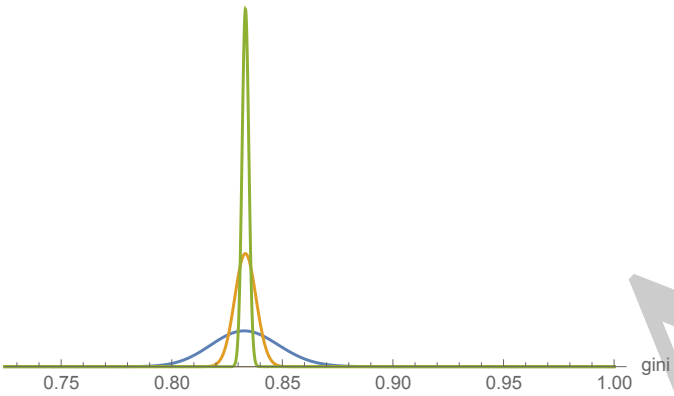


Fig. 2. Distribution of indirect estimator via exponent $n = 10^4, 10^5, 10^6$.

In other words the Gini is the mean expected deviation between any two random variables ("mean difference") scaled by the mean.

The "direct estimator" of the Gini of a sample becomes half the relative mean difference, where sample $Y = (Y_i)_{1\leq i\leq n}$ as follows:

$$\widehat{G}_d(Y) = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n}|Y_i - Y_j|}{2(n-1)\sum_{i=1}^{n} Y_i} \tag{2}$$

which can be further simplified

$$\widehat{G}_d(Y) = \frac{2\sum_{i=1}^{n} iY_{(i)}}{n\sum_{i=1}^{n} Y_i} - \frac{1}{n} + 1 \tag{3}$$

where $Y_{(1)}, Y_{(2)}, ..., Y_{(n)}$ are the ordered statistics of $Y_1, ..., Y_n$ such that: $Y_{(1)} < Y_{(2)} < ... < Y_{(n)}$.

B. Superadditivity and Results around the Zagier inequality

It turned out after we devised the proof that Zagier proved the inequality in [2] and [3]. The paper appears to be unnoticed by the economics community concerned with measurement of inequality. As we used another route, we proved a lemma that is useful for sums and sub-sums of ordered variables, and facilitates results about inequalities in other contexts.

We can show that $\widehat{G}$ is a slowly converging estimator, downward biased, inconsistent under aggregation. We start with an inequality:

Theorem 1 (The Zagier inequality)**.** *Partition the n data into p sub-samples $N = N_1 \cup \ldots \cup N_p$ of respective sizes $n_1, \ldots, n_p$, with $\sum_{i=1}^{m} n_i = n$, and let $S_1, \ldots, S_p$ be the sum of variables over each sub-sample, and $S = \sum_{i=1}^{p} S_i$ be that over the whole sample. Then we have:*

$$\widehat{G}(N) \geq \sum_{i=1}^{p} \frac{n_i}{n} \widehat{G}(N_i)$$

Proof. We start with $p = 2$ and use elementary recursion to generalize.

$$\widehat{G}(Y^\frown Z) \geq \frac{n_Y}{n_Y + n_Z}\widehat{G}(Y) + \frac{n_X}{n_Y + n_Z}\widehat{G}(Z). \tag{4}$$

Proving the Gini estimators can be done by proving the following lemma about ordered sums:

Lemma 1. *Let n_Y and n_Z be the relative sample sizes of vectors Y and Z respectively, with $Y = (Y_1, \ldots, Y_{n_y})$, $n_y \geq 2$, and $Z = (Z_1, \ldots, Z_{n_z})$, $n_z \geq 2$, having $Y_{(1)}, Y_{(2)}, ..., Y_{(n_y)}$ the ordered values of Y such that: $Y_{(1)} < Y_{(2)} < ... < Y_{(n)}$ (and the same for $Z_{(i)}$), assuming only positive values for Y and Z, and setting the merged vector (concatenated) $(Y^\frown Z) = (Y_1, \ldots, Y_{n_y}, Z_1, \ldots, Z_{n_z})$.*

$$\frac{\sum_{i=1}^{n_y+n_z} i(Y^\frown Z)_{(i)}}{\sum_{i=1}^{n_y} Y_i + \sum_{i=1}^{n_z} Z_i} - \frac{\sum_{i=1}^{n_y} iY_{(i)}}{\sum_{i=1}^{n_y} Y_i} - \frac{\sum_{i=1}^{n_z} iZ_{(i)}}{\sum_{i=1}^{n_z} Z_i} + \frac{1}{2} \geq 0 \tag{5}$$

Define the multiset of cardinality n, $\chi(\mathrm{n}) \triangleq \{\sum_{j=1}^{i} \delta_j : \delta_j \geq 0, 2 \leq i \leq n\}$.

For instance we have $\chi(2) = \{\delta_1, \delta_1 + \delta_2\}$, $\chi(3) = \{\delta_1, \delta_1 + \delta_2, \delta_1 + \delta_2 + \delta_3\}$, etc.

Now, with $m < n$, $m \in \mathbb{N}^+$ we select $p =_n P_m$ permutations indexed by $k = 1, 2, \ldots, p$ of subsets of size m in $\chi(n)$.

More precisely, we define the subset $\chi_m(n) = \chi_m(n, l), \ldots \chi_m(n, p)$ of cardinality m in $\chi(n)$ and $\chi_m^c(n)$ its complement such that (indexing by l), $\forall l \leq p$,

$$\chi(n) = \chi_m(n, l) \cup \chi_m^c(n, l).$$

Let $\sigma \in \mathcal{S}_n$ be any permutation $\sigma_1, \ldots, \sigma_n$ of $\{1, 2, \ldots, n\}$.

It is a standard result (from the rearrangement inequality) that

$$\sum_{i=1}^{n} i\chi(n)_{(i)} \geq \sum_{i=1}^{n} \sigma_i \chi(n)_{\sigma_i}.$$

Hence, applying to our case we can derive further inequalities across subsets such as:

$$\begin{aligned}\sum_{k=1}^{n} k(\chi(n))_{(k)} &\geq \sum_{j=i}^{m} j\chi_m(n, l)_{(j)} + \sum_{i=1}^{m-n} (i+m)\chi_m^c(n, l)_{(i)} \\ &\geq \sum_{j=i}^{m} j\chi_m(n, l)_{(j)} + \sum_{i=1}^{m-n} (i)\chi_m^c(n, l)_{(i)} \\ &\quad + m\sum_{i=1}^{m-n} \chi_m^c(n, l)_i,\end{aligned} \tag{6}$$

the latter is a non-ordered sum.

We calculate the LHS of the inequality in Equation 5:

$$I_l(m,n) = \frac{\sum_{i=1}^{n} i\chi(n,l)_{(i)}}{\sum_{i=1}^{m}\chi_m(n,l)_i + \sum_{i=1}^{n-m}\chi_m^c(n,l)_i} - \frac{\sum_{i=1}^{m} i\chi_m(n,l)_{(i)}}{\sum_{i=1}^{m}\chi_m(n,l)_i} - \frac{\sum_{i=1}^{n-m} i\chi_m^c(n,l)_{(i)}}{\sum_{i=1}^{n-m}\chi_m^c(n,l)_i} + \frac{1}{2}. \tag{7}$$

For $n = 2$: ${}_1P_2 = 2$, $\chi(2) = \{\delta_1, \delta_1 + \delta_2\}$ and the various subsets of cardinality 1 and their complements are:

$$\left\{ \begin{array}{ccc} \{\delta_1\} & , & \{\delta_1+\delta_2\} \\ \{\delta_1+\delta_2\} & , & \{\delta_1\} \end{array} \right\}$$

hence

$$I_1(1,2) = I_2(1,2) = \frac{\delta_2}{4\delta_1 + 2\delta_2} \geq 0$$

[1] Let us see how when it holds for n it necessarily holds for $n+1$. I continue with the following simplification of notation:

$$I = \frac{S_s}{S_1+S_2} - \frac{Ss_1}{S_1} - \frac{S_{s2}}{S_2} + \frac{1}{2}$$

for $n+1$

$$I' = \frac{S'_s}{S_1+S'_2} - \frac{Ss_1}{S_1} - \frac{S_{s2'}}{S'_2} + \frac{1}{2}$$

$Ss' = Ss + n(S_s + \delta_{n+1})$
$S'_2 = 2S_2 + \delta_{n+1}$
$S_{s2} = S_{s2} + (n-m)(S_{s2} + \delta_{n+1})$
Reexpressing $I^{'}$:

$$I' = \frac{(n+1)\delta_{n+1} + (n+2)S_s}{S_1 + 2S_2 + \delta_{n+1}} - \frac{S_{s1}}{S_1} - \frac{(n-m+1)(\delta_{n+1} + S_{s2}) + S_{s2}}{\delta_{n+1} + 2S_2} + \frac{1}{2} \tag{8}$$

From 6, $S_s \geq S_1 + S_2 + mS_2$ and $\delta_{n+1} \geq 0$, as well as the integers $m > 1$ and $n > m$. Normalizing with $S_2 = 1$ and from these inequalities: $I' \geq I''$, with

$$I'' = I'/.\ \{S_2 \to 1, S_{s2} \to 1, S_{s1} \to S_1, S_s \to m + S_1 + 1\}$$

Taking the numerator I''' of I'' after expressing it with a single denominator, given that the denominator $2(\delta_{n+1}+2)(\delta_{n+1}+S_1+2) \geq 0$:

$$I''' = (10m\delta_{n+1} - 4\delta_{n+1}) + \left(2m\delta_{n+1}^2 - \delta_{n+1}^2\right) + (4mn-4) + \delta_{n+1}(2mn + (2m+1)S_1) + 2S_1(m+n+1) + 12m \tag{9}$$

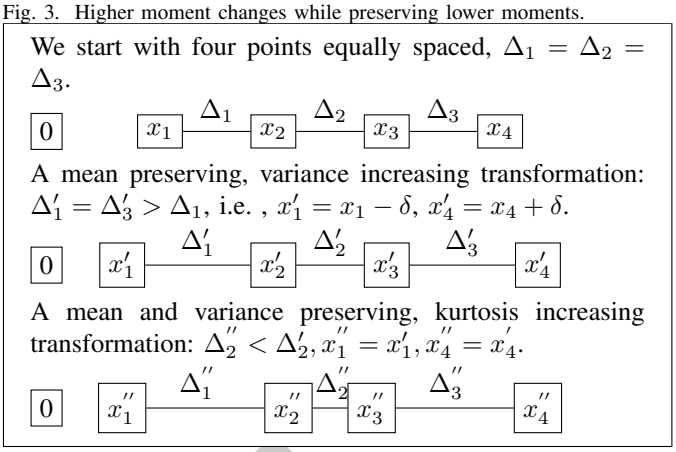
we have $I''' \geq 0$, which proves the lemma. To prove the rest of the theorem, it suffices to see that the lemma holds for all integers $0 < m < n$.

□

This inequality is similar to the inequality in theorem 1 in [1].

The theorem holds for, of course, any one-tailed distribution. But our focus is on fat tails. .

Fig. 3. Higher moment changes while preserving lower moments.



C. Extracting the Effect of Fattailedness in the inequality and sampling problem

Proposition 1. *The inequality becomes equality when all subsets are identical.*

Proof. Consider a random variable $X > 0$ that has four (positive) realizations $x_1 < x_2 < x_3 < x_4$. We construct three units, $A = (x_1, x_2, x_3)$, $B = (x_1, x_2, x_4)$ and the merged unit $C = (x_1, x_1, x_2, x_2, x_3, x_4)$.

The "direct" estimators can be calculated from Equation 3 as follows: $\widehat{G}(A) = \frac{2(x_3-x_1)}{3(x_1+x_2+x_3)}$, $\widehat{G}(B) = \frac{2(x_4-x_1)}{3(x_1+x_2+x_4)}$, and $\widehat{G}(C) = \frac{-8x_1+3x_3+5x_4}{6(2x_1+2x_2+x_3+x_4)}$.

Replacing x_4 with $x_3 + \delta$, with $\delta \geq 0$, we get

$$\widehat{G}(C) - \frac{1}{2}\left(\widehat{G}(A) + \widehat{G}(B)\right) = \frac{\delta\left(\frac{2(2x_1+x_2)\delta}{(x_1+x_2+x_3)(x_1+x_2+x_3+\delta)} + 1\right)}{6\left(2x_1+2x_2+2x_3+\delta\right)} \tag{10}$$

which can be shown to be strictly positive for all values of $\delta > 0$ and to increase with δ. Further it will be 0 for identical samples A and B.

By introducing m additional intermediate values between any two units, say $x_{1.1}, x_{1.2}, \ldots, x_{1.m}$ we can show the property holds for all values of N that produce equal size subsets (that is for 2 subsets, $N = 8, 10, \ldots$ □

Proposition 2. *The inequality increases with a mean-preserving difference between the variances of the subsets.*

Consider a mean-preserving symmetric transformation: two samples A and B have a different variance. $A = (x_1, x_2, x_3)$, $B = (x_1 - \delta, x_2, x_3 + \delta)$ and the merged unit $C = (x_1 - \delta, x_1, x_2, x_2, x_3, x_3 + \delta)$. With $0 \geq \delta \geq x_1$, we have:

$$\Delta(\delta) \triangleq \widehat{G}(C) - \frac{1}{2}\left(\widehat{G}(A) + \widehat{G}(B)\right) = \frac{\delta}{6\left(x_1+x_2+x_3\right)} \tag{11}$$

and we can tighten the inequality for that specific case:

$$0 \leq \Delta(\delta) \leq \frac{x_1}{6(x_1+x_2+x_3)} \tag{12}$$

The difference between the Gini estimator for the aggregate C and the average of the Gini estimates for the components A and B will be a convex function in δ, hence increases with the fatness of tails (how far the highest realization is away from the rest), as we can see next.

Proposition 3. *The inequality increases with total kurtosis of the population.*

Proof.

$$A = (x_1, x_2, x_3, x_4)\,,$$

$$B = \left(x_1 - \delta, x_2 + \frac{\epsilon}{2}, x_3 - \frac{\epsilon}{2}, \delta + x_4\right)$$

and

$$C = \left(x_1 - \delta, x_1, x_2, x_2 + \frac{\epsilon}{2}, x_3 - \frac{\epsilon}{2}, x_3, x_4, \delta + x_4\right).$$

Setting

$$\epsilon = -\sqrt{4\delta x_1 - 4\delta\left(\delta + x_4\right) + \left(x_2 - x_3\right)^2} - x_2 + x_3,$$

As outlined in the methodology in Figure 3 we can isolate a parameter δ that controls kurtosis without affecting lower moments; we need to have as bounds to preserve the variance:

$$0 \leq \delta \leq \frac{1}{2}\left(\sqrt{\left(x_1 - x_4\right)^2 - 3\left(x_2 - x_3\right)^2} + x_1 - x_4\right.$$

$$\begin{aligned}\Delta(\delta) &\triangleq \widehat{G}(C) - \frac{1}{2}\left(\widehat{G}(A) + \widehat{G}(B)\right)\\ &= -\frac{-2\delta + \sqrt{4\delta x_1 - 4\delta\left(\delta + x_4\right) + \left(x_2 - x_3\right)^2} + x_2 - x_3}{16\left(x_1 + x_2 + x_3 + x_4\right)}\\ &\geq 0\end{aligned} \tag{13}$$

□

1) Known distribution: If we know the distribution of X, then Equation 1 is straightforward. In the event of known cumulative distribution function Φ, consider that $|X - X'| = X + X' - 2\min(X, X')$. Hence the expectation becomes:

$$\mathbb{E}\left(|X - X'|\right) = 2\left(\mu - \mathbb{E}(X, X')^-\right)$$

We have the joint cumulative

$$F\left((x, x')^-\right) = 1 - \mathbb{P}(X > x)\mathbb{P}(X' > x)$$

hence, with $X \in [L, \infty)$:

$$G = 1 - \frac{1}{\mu}\int_L^\infty \left(1 - \Phi(x)\right)^2 \mathrm{d}x \tag{14}$$

Rewriting $\Phi = \Phi(x, \lambda)$, where λ is the parameter of the distribution that is stochasticized. With $\lambda > \Delta \geq 0$:

$$G(\lambda) = 1 - \frac{1}{\mu}\int_L^\infty \left(1 - \int_0^\infty \Phi(x, \lambda) d\lambda\right)^2 \mathrm{d}x \tag{15}$$

This is in case the distribution depends on the scale. Heterogeneous distributions:

Remark 1. *(i) Let $\Phi_i(x)$, be a distribution function associated with random variable X_i , and let $s_i(x)$ be a square integrable function $s_i : [L, \infty) \to (-1, 1)$, with $s_i(\infty) = s_i(L) = 0$ and $\forall i\,, -\Phi(x) \leq s_i(x) \leq 1 - \Phi(x)$. We can express*

$$\Phi_i(x) = \Phi(x) + s_i(x)$$

.

(ii)Assume the random variables X_i are scaled by μ ($\mathbb{E}(X_i) = \mu$). Consider the normalized weights α_i, $\sum_{i\leq n}\alpha_i = 1$ and $0 \leq \alpha_i \leq 1$. Then Δ_n the lower bound of the subadditivity for n heterogeneous subsamples, that is, $\Delta_n = G\left(\sum_{i\leq n}\alpha_i\Phi_i(x)\right) - \sum_{i\leq n} G(\alpha_i\Phi_i(x)$ is expressed as:

$$\Delta_n = \frac{1}{\mu}\int_L^\infty \sum_{i=1}^N \alpha_i s_i(x)^2 - \left(\sum_{i=1}^N \alpha_i s_i(x)\right)^2 \mathrm{d}x$$

We further assume that $s_1(x) = 0$ that is, $i = 1$ is the baseline. For $n = 2$,

$$\Delta_2 = -\frac{\left(\alpha_1 - 1\right)\alpha_1}{\mu}\int_L^\infty s_2(x)^2\, dx$$

Further, we can use a lemma.

Lemma 2. *To be mean preserving, the square-integrable function $s(x)$ modifying a cumulative distribution $\Phi(x)$ requires $\int s(x)\, dx = 0$. And to be mean and variance preserving requires $\int xs(x)dx = 0$.*

Proof. Integrating by parts for positive r.v., $\mu = K + \int x\frac{\partial\Phi'(x)}{\partial x}\, dx$, we get $\int_L^\infty \left(1 - \Phi(x)\right)\, dx = \int_L^\infty \left(1 - \Phi(x) - s(x)\right)\, dx$. A similar reasoning for the variance. □

We can also extract the effect of the Kurtosis on the Gini as follows:

$$\begin{aligned}G\left(\sum_{i\leq n}\alpha_i\Phi_(x)\right) - G(\Phi) = \frac{1}{\mu}\Bigg(&\int_L^\infty s(x)\, dx\\ &- \int_L^\infty s(x)\Phi(x)\, dx\\ &- \frac{1}{4}\int_L^\infty s(x)^2\, dx\Bigg)\end{aligned} \tag{16}$$

Additional constraints. We have $2s(x)\Phi(x) + s(x)^2 + \Phi(x)^2 \leq 1$ and by Cauchy-Schwartz $\int s(x)\Phi(x) \leq \int s(x)^2 \times \int \Phi(x)^2$ and $\int xs(x) \leq \int xdx$ Since $s(x)^2 \leq \left(1 - \Phi(x)\right)^2$...

2) Special case of the Lognormal: In the case of the lognormal $L(\mu, \sigma)$, while σ is not strictly the scale, it can be used to simulate stochastic volatility, and $\frac{1}{2}\left(\mu(\sigma + \Delta) + \mu(\sigma - \Delta)\right) = \frac{1}{2}e^\mu\left(e^{\frac{1}{2}(\Delta - \sigma)^2} + e^{\frac{1}{2}(\Delta + \sigma)^2}\right)$. Hence

$$\begin{aligned}G(\sigma, \Delta) = 1 - \frac{2}{\mu(\sigma + \Delta) + \mu(\sigma - \Delta)}\int_L^\infty \Bigg(&1\\ &- \Phi(x, \sigma) - \frac{1}{2}\Delta^2\Phi^{(0,2)}(x, \sigma)\Bigg)^2 \mathrm{d}x\end{aligned} \tag{17}$$

There is a term missing we can ignore for now or add later.

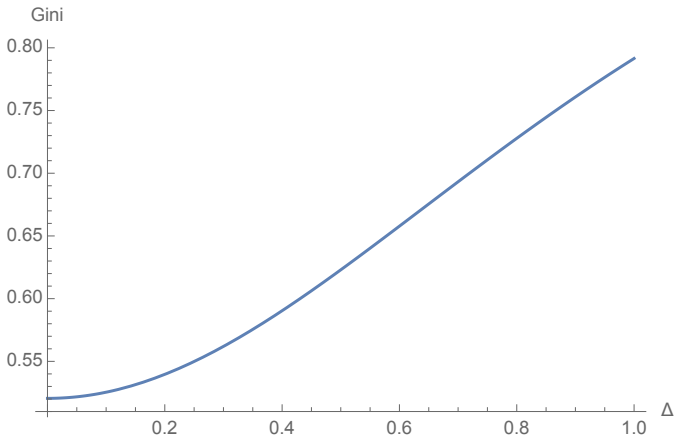


Fig. 4. Fatness of Tails expressed as Δ in the case of the Lognormal

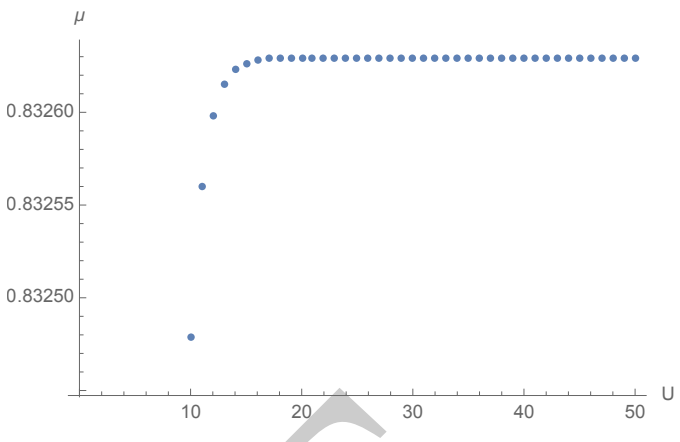


Fig. 5. convergence of the index with number of summands U

Another Route

$$G(\sigma) = 2\Psi\left(\frac{\sigma}{2}\right) - 1 \tag{18}$$

For variance preserving perturbation: $\sigma^+ = \sqrt{\log\left(1 - e^{(\Delta-\sigma)^2}\right)}$, $\sigma^- = \sigma + \Delta$ Which is the equivalent of having part of the sample with variance σ^{+^2} and the other part with σ^{+^2}, each with half the ratio. We already know that

$$G(\sigma) \geq \frac{1}{2}\left(G(\sigma^+) + G(\sigma^-)\right) \tag{19}$$

and more generally for any linear combination of variables from separate distributions.

We can show that for the stochastic volatility case

$$G(\sigma, \Delta) \geq G(\sigma, 0) \tag{20}$$

In other words, an increase in the kurtosis of the distribution translates into an increase in the theoretical Gini.

This section is to be completed

III. Estimation from tail α via Maximum Likelihood

So, next we show that for the case power law, an "indirect" estimation via the Hill estimator of the tail exponent is a more efficient way to estimate the Gini index.

A. Distribution of the exponent

Next we calculate the distribution of the tail exponent of a power law. We start with the standard Pareto distribution for random variable X with pdf:

$$\phi_X(x) = \alpha L^\alpha x^{-\alpha-1}, x > L \tag{21}$$

Assume $L = 1$ by scaling.

The likelihood function is $\mathcal{L} = \prod_{i=1}^n \alpha x_i^{-\alpha-1}$. Maximizing the Log of the likelihood function (assuming we set the minimum value) $\log(\mathcal{L}) = n(\log(\alpha) + \alpha\log(L)) - (\alpha + 1)\sum_{i=1}^n \log(x_i)$ yields: $\hat{\alpha} = \frac{n}{\sum_{i=1}^n \log(x_i)}$. Now consider $l = -\frac{\sum_{i=1}^n \log X_i}{n}$. Using the characteristic function to get the distribution of the average logarithm yield:

$$\psi(t)^n = \left(\int_1^\infty f(x)\exp\left(\frac{it\log(x)}{n}\right)dx\right)^n = \left(\frac{\alpha n}{\alpha n - it}\right)^n$$

which is the characteristic function of the gamma distribution $(n, \frac{1}{\alpha n})$. A standard result is that $\hat{\alpha}' \triangleq \frac{1}{l}$ will follow the inverse gamma distribution with density:

$$\phi_{\hat{\alpha}}(a) = \frac{e^{-\frac{\alpha n}{\hat{\alpha}}}\left(\frac{\alpha n}{\hat{\alpha}}\right)^n}{\hat{\alpha}\Gamma(n)},\ a > 0$$

.

1) Debiasing: Since $\mathbb{E}(\hat{\alpha}) = \frac{n}{n-1}\alpha$ we elect another – unbiased– random variable $\hat{\alpha'} = \frac{n-1}{n}\hat{\alpha}$ which, after scaling, will have for distribution $\phi_{\hat{\alpha'}}(a) = \frac{e^{\frac{\alpha-\alpha n}{a}}\left(\frac{\alpha(n-1)}{a}\right)^{n+1}}{\alpha\Gamma(n+1)}$.

2) Truncating for $\alpha > 1$: Given that values of $\alpha \leq 1$ lead to infinite mean (hence no Gini) we restrict the distribution to values greater than $1+\epsilon$, $\epsilon > 0$. Our sampling now applies to lower-truncated values of the estimator, those strictly greater than 1, with a cut point $\epsilon > 0$, that is, $\sum \frac{n-1}{\log(x_i)} > 1 + \epsilon$, or $\mathbb{E}(\hat{\alpha}|_{\hat{\alpha}>1+\epsilon})$: $\phi_{\hat{\alpha''}}(a) = \frac{\phi_{\hat{\alpha'}}(a)}{\int_{1+\epsilon}^\infty \phi_{\hat{\alpha'}}(a)\,da}$, hence the distribution of the values of the exponent conditional of it being greater than 1 becomes:

$$\phi_{\hat{\alpha''}}(a) = \frac{e^{\frac{\alpha n^2}{a-an}}\left(\frac{\alpha n^2}{a(n-1)}\right)^n}{a\left(\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)\right)},\ a \geq 1+\epsilon \tag{22}$$

B. The distribution of the α-derived Gini

Now define the "derived Gini" from estimated α, $G \triangleq \frac{1}{2\hat{\alpha''}-1}$. After some manipulation, we have $\phi_G(g)$ the distribution of the derived Gini:

$$\phi_G(g) = \frac{2^n e^{-\frac{2\alpha g n^2}{(g+1)(n-1)}}\left(\frac{\alpha g n^2}{(g+1)(n-1)}\right)^n}{g(g+1)\left(\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)\right)},$$
$$g \in (0, \frac{1}{2\epsilon+1}) \tag{23}$$

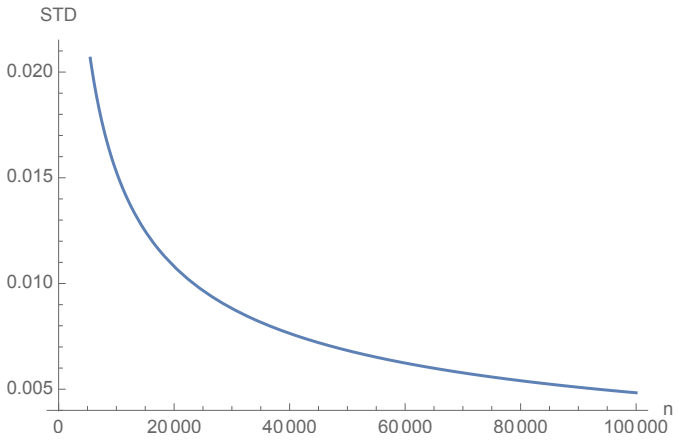


Fig. 6. Standard deviation of the ML estimator with an increase of population

C. Moments of the estimated Gini index

We are looking for moment of order m, that is $\mu(m)$ as $\int_0^{\frac{1}{2\epsilon+1}} g^m \phi_G(g) dg$. By substitution, with $u = \frac{g}{g+1}$,

$$\mu(m) = \int_0^{\frac{1}{2\epsilon+2}} \frac{2^n \left(\frac{1}{1-u}\right)^m u^{m-1} e^{-\frac{2\alpha n^2 u}{n-1}} \left(\frac{\alpha n^2 u}{n-1}\right)^n}{\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)} du$$

using the property that $\sum_{i=0}^{\infty} u^i \binom{i+m-1}{i} = (1-u)^{-m}$ and that

$$\int_0^{\frac{1}{2\epsilon+2}} \frac{2^n u^i u^{m-1} e^{-\frac{2\alpha n^2 u}{n-1}} \left(\frac{\alpha n^2 u}{n-1}\right)^n}{\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)} du = \left(\frac{1}{2\epsilon+2}\right)^{i+m} \frac{\left(\frac{\alpha n^2}{(n-1)(\epsilon+1)}\right)^{-i-m} \left(\Gamma(i+m+n) - \Gamma\left(i+m+n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)\right)}{\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)} \tag{24}$$

we finally have, with U a natural number:

$$\mu(m) = \lim_{U \to +\infty} \sum_{i=0}^{U} \frac{\binom{i+m-1}{i} \left(\left(\frac{1}{2\epsilon+2}\right)^{i+m} \left(\frac{\alpha n^2}{(n-1)(\epsilon+1)}\right)^{-i-m}\right)}{\Gamma(n) - \Gamma\left(n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)} \left(\Gamma(i+m+n) - \Gamma\left(i+m+n, \frac{n^2\alpha}{(n-1)(\epsilon+1)}\right)\right) \tag{25}$$

which, in practice, with values of $U \approx 7$ produces appropriate approximations, see Figure 5. We get explicit (rather, semi-explicit) expressions of the standard deviations and show their decline in Figure 6.

D. Some comments

For recent wealth data restating Pareto and Mandelbrot's point [4], see [5]. Some authors missed the point: see [6], [7], [8]. In some cases, researchers get it backwards, getting α from G [9].

ACKNOWLEDGMENT

Raphael Douady, Pasquale Cirillo, Mike Lawler.

NOTES

[1]We don't need for this proof to examine $n = 3$, but let's do it for verification. For $n = 3$, ${}_2P_3 = 3$, $\chi(3) = \{\delta_1, \delta_1+\delta_2, \delta_1+\delta_2+\delta_3\}$ and the various subsets where cardinality 1 and their complements (or, alternatively, the subsets of cardinality 2 and their complements) are

$$\left\{ \begin{array}{ccc} \{\delta_1, \delta_1+\delta_2\} & , & \{\delta_1+\delta_2+\delta_3\} \\ \{\delta_1, \delta_1+\delta_2+\delta_3\} & , & \{\delta_1+\delta_2\} \\ \{\delta_1+\delta_2, \delta_1+\delta_2+\delta_3\} & , & \{\delta_1\} \end{array} \right\}$$

hence

$$\begin{aligned} I_1(2,3) = I_3(1,3) &= \tfrac{\delta_2\delta_3+\delta_1(\delta_2+4\delta_3)}{2(2\delta_1+\delta_2)(3\delta_1+2\delta_2+\delta_3)} \geq 0 \\ I_2(2,3) = I_2(1,3) &= \tfrac{(\delta_1+\delta_3)(\delta_2+\delta_3)}{2(2\delta_1+\delta_2+\delta_3)(3\delta_1+2\delta_2+\delta_3)} \geq 0 \\ I_3(2,3) = I_1(1,3) &= \tfrac{(2\delta_2+\delta_3)^2+\delta_1(4\delta_2+\delta_3)}{2(2\delta_1+2\delta_2+\delta_3)(3\delta_1+2\delta_2+\delta_3)} \geq 0 \end{aligned}$$

REFERENCES

[1] N. N. Taleb and R. Douady, "On the super-additivity and estimation biases of quantile contributions," *Physica A: Statistical Mechanics and its Applications*, vol. 429, pp. 252–260, 2015.
[2] D. B. Zagier, *On the decomposability of the Gini coefficient and other indices of inequality*. Inst. für Ges.-und Wirtschaftswiss., Wirtschafts-theoretische Abt., Univ., 1983.
[3] D. Zagier, "Inequalities for the gini coefficient of composite populations," *Journal of Mathematical Economics*, vol. 12, no. 2, pp. 103–118, 1983.
[4] B. Mandelbrot, "The pareto-levy law and the distribution of income," *International Economic Review*, vol. 1, no. 2, pp. 79–106, 1960.
[5] J. K. Dagsvik, Z. Jia, B. H. Vatne, and W. Zhu, "Is the pareto–lévy law a good representation of income distributions?" *Empirical Economics*, vol. 44, no. 2, pp. 719–737, 2013.
[6] R. I. Lerman and S. Yitzhaki, "Improving the accuracy of estimates of gini coefficients," *Journal of econometrics*, vol. 42, no. 1, pp. 43–47, 1989.
[7] J. L. Gastwirth and M. Glauberman, "The interpolation of the lorenz curve and gini index from grouped data," *Econometrica: Journal of the Econometric Society*, pp. 479–483, 1976.
[8] F. Alvaredo, "A note on the relationship between top income shares and the gini coefficient," *Economics Letters*, vol. 110, no. 3, pp. 274–277, 2011.
[9] J. Wildman, H. Gravelle, and M. Sutton, "Health and income inequality: attempting to avoid the aggregation problem," *Applied Economics*, vol. 35, no. 9, pp. 999–1004, 2003.