Gini estimation under non-finite variance Paretan tails

Andrea Fontanari - TU Delft and CWI
Nassim Nicholas Taleb - Tandon School of Engineering, NYU
Pasquale Cirillo - TU Delft

***Abstract*—We study different approaches to the estimation of the Gini index in presence of a heavy tailed data generating process, that is, one with Paretan tails and/or in the stable distribution class with non-finite variance, that is with tail index $1 < \alpha < 2$.**

In particular we show how the "non-parametric" estimator of the Gini index undergoes a phase transition in the symmetry structure of its asymptotic distribution as the data distribution shifts from the domain of attraction of a light tail distribution to the domain of attraction of a fat tailed, infinite variance one. This explains the downward bias often encountered in applications. We then show how the maximum likelihood estimator outperforms the "non-parametric" requiring a much smaller sample size to reach efficiency.

Finally we provide a simple correction mechanism to the small sample bias of the "non-paramteric" estimator based on the distance between the mode and the mean of its asymptotic distribution for the case of heavy tailed data generating process.

Examples and simulation experiments are also provided to support our results.

I. Introduction

Wealth inequality studies represent a field of economics and statistics exposed to fat-tailed data generating process, often with infinite variance. This is not surprising especially if we think that the prototype of fat tailed distribution, the Pareto distribution, has been derived for the first time to model data of households income [4].

However, fat tails can be problematic in the context of statistical estimations since most of the basic statistic results of efficiency and consistency do not hold anymore.

The scope of this work is, hence, to show how fat tails affect the estimation of one of the most celebrated measure of income inequality: the Gini index.

The literature about Gini index estimator is wide and comprehensive, however, strangely enough almost no attention has been posed on its behavior in presence of fat tails. This paper aims to close this gap deriving the limiting distribution of the so called "non-parametric" Gini estimator when the data are fat tailed and to analyze possible strategies to reduce fat tails impact on the estimation.

In particular, we will show how a maximum likelihood approach, despite the risk of misspecify the model for the data, needs much less observation to reach efficiency compared to a "non-parametric" one.

By heavy tails we mean data coming from a power law distribution which is the generalization of the type I Pareto distribution with density:

$$f(x) = \rho x_m^{\rho} x^{-\rho-1}, x > x_m \quad (1)$$

Note that in this paper we will assume $x_m = 1$ without loss of generality.

In particular we restrict our focus on distributions with finite mean and infinite variance and therefore we can limit the class of power laws only to those with tail exponent $\rho \in (1, 2)$.

Table I and Figure 1 present numerically and graphically our story and suggest its conclusion.

Table I compares the Gini index obtained by the "non-parametric" estimator and the one obtained via Maximum Likelihood Estimation (MLE) of the tail exponent.

As the first column shows, the convergence of the "non-parametric" estimator to the true value $g = 0.8\overline{33}$ is extremely slow and monotonic increasing when the data distribution is infinite variance. This suggests an issue not only in the tail structure of the distribution of the "non-parametric" estimator but also in its symmetry.

As we shall see, in order to avoid the poor quality in the convergence of the "non-parametric" estimator to the true value of the Gini index, in presence of infinite variance data generating process, we suggest a MLE approach to recover normality of the asymptotic distribution as well as to improve the speed of convergence. We believe that in such a heavy tail framework the issues of a parametric approach are offset by the gains in the quality of the approximation.

TABLE I: Comparison of "non-parametric" Gini to ML estimator, assuming tail $\alpha = 1.1$

n *(number of observations)*	Non-Parametric		MLE		Error Ratio
	Mean	*Bias*	*Mean*	*Bias*	
10^3	*0.711*	*-0.122*	*0.8333*	*0*	1.4
10^4	*0.750*	*-0.083*	*0.8333*	*0*	3
10^5	*0.775*	*-0.058*	*0.8333*	*0*	6.6
10^6	0.790	-0.043	0.8333	0	156
10^7	0.802	-0.033	0.8333	0	$> 10^5$

Figure 1 provides additional evidence that the limiting distribution of the "non-parametric" Gini index loses its properties of normality and symmetry shifting to a fatter tailed and skewed limit as the distribution of the data enters an infinite variance domain. As we will prove in Section II this is exactly what happens. When the data generating process is in the domain of attraction of a fat tail distribution, the asymptotic distribution of the Gini index moves away from gaussianity towards a totally skewed to the right α-stable limit.

This change of behavior is the responsible for the main problem of the "non-parametric" estimators for the Gini index: a downwards bias, for almost every sample size, when data are heavy tailed.

This result, suggests another possible solution to improve the quality of the "non-parametric" estimator in case a maximum likelihood approach is not preferred.

The idea is to correct for the skewness of the distribution of the "non-parametric" estimator in order to place its mode on the true value of the Gini. This correction, we will show, improves the consistency and the bias of the estimator, while still allowing the use of a non-parametric approach. The correction reduces the risk of underestimating the Gini index.

The remainder of the paper is organized as follows.

In Section II we derive the asymptotic distribution of the sample Gini index when the data distribution is fat tailed and infinite variance. We then, provide an example with Pareto distributed data and we compare the quality of the limit distribution between maximum likelihood estimator and the "non-parametric" one. Finally in Section III we propose a simple correction mechanism based on the mode-mean distance of the asymptotic distribution of the "non-parametric" estimator to adjust the bias in small samples.

II. Asymptotic and pre-asymptotic results on "Non-parametric" Estimator

In this Section, we derive the asymptotic distribution for the "non-parametric" estimator of the Gini index when the data generating process is fat tailed with finite mean but infinite variance.

In this work we choose the, so called, stochastic representation of the Gini, denoted by g,

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

where g is the Gini index and X' and X'' are independent identical copies of a random variable X with c.d.f. $F(x) \in [c, \infty), c > 0$ and finite mean $\mathbb{E}[X] = \mu$.

The quantity $\mathbb{E}\left(|X' - X''|\right)$ is known as the "Gini mean difference" (GMD).

The Gini index is then the mean expected deviation between any two independent draws from a random variable scaled by twice its mean [9], [1].

What we call the "non-parametric" estimator of the Gini index of a sample $(X_i)_{1\leq i\leq n}$ is the following:

$$G^{NP}(X_n) = \frac{\sum_{1\leq i<j\leq n} |X_i - X_j|}{(n-1)\sum_{i=1}^{n} X_i} \tag{3}$$

which can also be expressed as:

$$G^{NP}(X_n) = \frac{\sum_{i=1}^{n}(2(\frac{i-1}{n-1} - 1)X_{(i)}}{\sum_{i=1}^{n} X_{(i)}} = \frac{\frac{1}{n}\sum_{i=1}^{n} Z_{(i)}}{\frac{1}{n}\sum_{i=1}^{n} X_i} \tag{4}$$

where $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are the ordered statistics of $X_1, ..., X_n$ such that: $X_{(1)} < X_{(2)} < ... < X_{(n)}$.

The asymptotic normality of (4) under finite variance assumption has been shown already by different authors. The result follows from the properties of the U-statistics and the L-estimators involed in foulation 4.

For more details see for example [4].

For later reference we introduce the following notation for what concerns α-stable distributions.

A random variable X is distributed accordingly to an α-stable distribution if:

$$X \sim S(\alpha, \beta, \gamma, \delta)$$

where $\alpha \in (0, 2)$ is the tail parameter, $\beta \in [-1, 1]$ is the skewness, $\gamma \in \mathbb{R}^+$ is the scale and $\delta \in \mathbb{R}$ is the location.

For future reference we also define the standardized α-stable random variable as:

$$Z_{\alpha,\beta} \sim S(\alpha, \beta, 1, 0) \tag{5}$$

α-stable distributions, are a subclass of infinitely divisible distributions, in particular, thanks to their stability under summation property they can be used to describe the entire class of distributions arising from a Central Limit Theorem type of argument. In particular for $\alpha = 2$ we obtain the normal distribution, the limit distribution for the most classical central limit theorem, for different values of α we obtain other type of limiting distributions. In particular we use the following notation to describe a random variable in the domain of attraction of an α-stable distribution: $X \in DA(S_\alpha)$ [3].

Given the possible confusion that can arise in the parametrization of α-stable distribution we shall clarify which definition we are using. We use the parametrization presented in [7] also called $S1$ parametrization [6], therefore the characteristic function defining the α-stable is given by:

$$E(e^{itX}) = e^{-\gamma^\alpha |t|^\alpha (1 - i\beta\,\mathrm{sign}(t))\tan(\frac{\pi\alpha}{2}) + i\delta t}, \alpha \neq 1$$

$$E(e^{itX}) = e^{-\gamma |t| (1 + i\beta\frac{2}{\pi}\,\mathrm{sign}(t))\ln|t| + i\delta t}, \alpha = 1$$

Since we are dealing with finite mean distributions we restrict ourselves to the case of $\alpha \in (1, 2)$. Therefore our parametrization define a location-scale family for every choice of $\alpha \in (1, 2)$ and there is not need to use the $S0$ parametrization introduced by J. Nolan in [6]. Additionally in the case of $\alpha \in (1, 2)$ the expected value

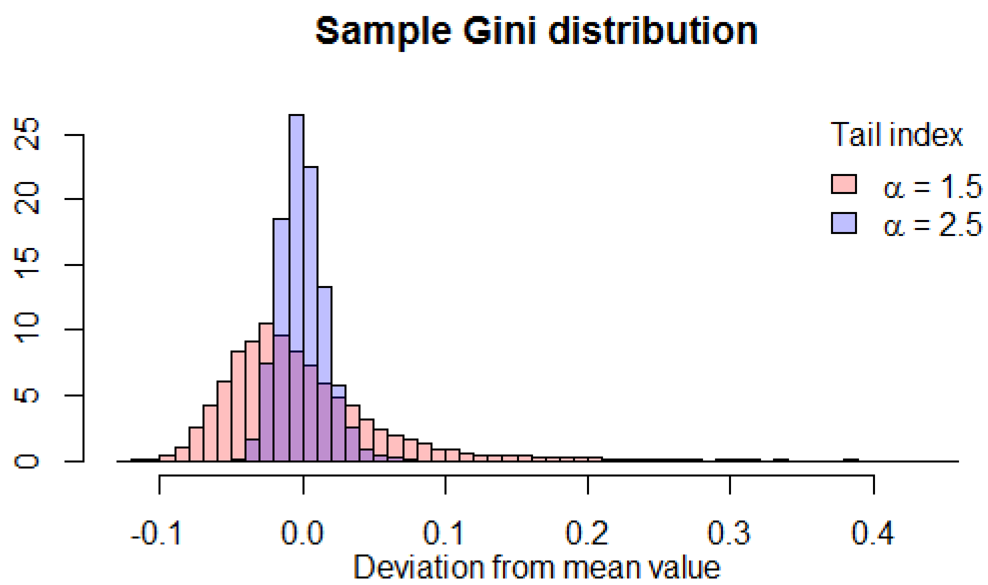


Fig. 1: Empirical distribution for Gini "non-parametric" estimator for Type I Pareto ditribution with different tail index (results has been centered to ease comparison)

of an α-stable random variable coincide with the location parameter: $\mathbb{E}(X) = \delta$

For more reference the interested reader can consult [7], [6].

A. The α-Stable asymptotic limit for the Gini index

We are now ready to present our main result: the asymptotic distribution of the Gini index estimator, as presented in (4), when the data generating process is fat tailed and more specifically with infinite variance.

More formally we assume that our observations $(X_i)_{1\leq i\leq n}$ are i.i.d. generated by a random variable X in the maximum domain of attraction of a Frechet distribution [2], $X \in MDA(\Phi(\rho))$ with $\rho \in (1,2)$ such that

$$P(\max(X_1, ..., X_n) \leq x) \xrightarrow{d} e^{-x^{-\rho}}$$

The result is divided into two theorems, Theorem II.1 takes care of the limiting distribution of the "Gini Mean Difference" (GMD) (numerator in Equation 3), while Theorem II.2 completes the proof for the whole Gini index.

Theorem II.1. *Consider a sequence $(X_i)_{1\leq i\leq n}$ of i.i.d random variables from a distribution X on $[c, +\infty)$ $c > 0$, such that F is in the maximum domain of attraction of an Frechet random variable: $X \in MDA(\Phi(\rho))$ where we restrict to the case in which $\rho \in (1,2)$. Then the sample Gini mean deviation (GMD) $\frac{\sum_{i=1}^{n} Z_{(i)}}{n} = GMD_n$ satisfies the following limit in distribution.*

$$\frac{n^{-\frac{1}{\rho}}\left(\sum_{i=1}^{n} Z_{(i)} - n\theta\right)}{L_0(n)} \xrightarrow{d} Z_{\rho,1} \quad (6)$$

where $\theta = E[Z_{(i)}]$, $L_0(n)$ is a slowly varying function such that Equation 7 holds and $Z_{\rho,1}$ is a standardized α-stable random variable defined as 5.

Proof. Theorem 3.1 (ii) in [5] proves the existence of a weak limit for the GMD prior to the existence of the right scaling sequences to apply to the sequence of i.i.d. random variables $Z_i = (2F(X_i) - 1)X_i$ where $F(X)$ is the integral probability transform of $X \sim F$. Therefore what is left to prove is the characterization of scaling sequences and the limiting distribution.

To start recall that by assumption $X \in MDA(\Phi(\rho))$. A standard result in extreme value theory , [2], characterizes the tail of the distribution in the Frechet domain as: $P(|X| > x) \sim L(x)x^{-\rho}$, where $L(x)$ is a slowly varying function. Recall [3] that this is a characterization for distributions in the domain of attraction of α-stable distribution as well.

Therefore $X \in MDA(\Phi(\rho))$ is equivalent to $X \in DA(S_\rho)$ with $\rho \in (1,2)$. This result enables us to use a Central Limit Theorem type argument for the convergence of the sum in the estimator.

However, we first need to prove that the r.v. $Z \in DA(S_\alpha)$ as well. i.e. $P(|Z| > z) \sim L(z)z^{-\rho}$ with $\rho \in (1,2)$ and $L(z)$ slowly varying.

Note that:

$$P(|\tilde{Z}| > z) \leq P(|Z| > z) \leq P(2X > z)$$

where $\tilde{Z} = (2U-1)X$ with $U \sim unif[0,1]$ and $U \perp X$.

The first bound hold because of positive dependence between X and $F(X)$ and can be proven rigorously by noting that $2UX \leq 2F(X)X$ by re-arrangement inequity. The upper bound is trivial.

By assumption $P(2X > z) \sim 2^{\rho} L(z) z^{-\rho}$.[1] In order to show that also $\tilde{Z} \in DA(S_\alpha)$ we exploit the Breiman's Theorem [8] which ensure the stability of the Fréchet class under the product as long as the second random variable is not too heavy tailed. In order to apply the theorem we re-write $P(|\tilde{Z}| > z)$ as:

$$P(|\tilde{Z}| > z) = P(\tilde{Z} > z) + P(-\tilde{Z} > z) = P(\tilde{U}X > z) + P(-\tilde{U}X > z)$$

where $\tilde{U} \sim unif[-1,1]$ and $\tilde{U} \perp X$.
We focus on $P(\tilde{U}X > z)$ since for $P(-\tilde{U}X > z)$ the procedure is the same.

We have:

$$P(\tilde{U}X > z) = P(\tilde{U}X > z|\tilde{U} > 0)P(\tilde{U} > 0) + P(\tilde{U}X > z|\tilde{U} \leq 0)P(\tilde{U} \leq 0)$$

for $z \to +\infty$, $P(\tilde{U}X > z|\tilde{U} \leq 0) \to 0$ while applying Breiman's Theorem, $P(\tilde{U}X > z|\tilde{U} > 0)$ becomes:

$$P(\tilde{U}X > z|\tilde{U} > 0) \to E((\tilde{U})^{\rho}|U > 0)P(X > z)P(U > 0)$$

Therefore:

$$P(|\tilde{Z}| > z) \to \frac{1}{2}E((\tilde{U})^{\rho}|U > 0)P(X > z) + \frac{1}{2}E((-\tilde{U})^{\rho}|U \leq 0)P(X > z)$$

From this:

$$P(|\tilde{Z}| > z) \to \frac{1}{2}P(X > z)[E((\tilde{U})^{\rho}|U > 0) + E((-\tilde{U})^{\rho}|U \leq 0)] = \frac{2^{\rho}}{1-\rho}P(X > z) \sim \frac{2^{\rho}}{1-\rho}L(z)z^{-\rho}$$

We conclude then that by squeezing theorem:

$$P(|Z| > z) \sim L(z)z^{-\rho}$$

as $z \to \infty$ and therefore $Z \in DA(S_\alpha)$ with $\alpha = \rho$
We are now ready to invoke the generalized Central Limit Theorem [2] for the sequence of Z_i.

$$\frac{\sum_{i=1}^{n} Z_i - n\theta}{L_0(n) n^{\frac{1}{\rho}}} \xrightarrow{d} Z_{\rho,\beta}$$

Where $\theta = E(Z_i)$, $L_0(n) = c_n n^{-\frac{1}{\rho}}$ where c_n is a sequence which must satisfy the following relation:

[1]Note that $L(\frac{z}{2}) \sim L(z)$ by definition of slowly varying function

$$\lim_{n\to\infty} \frac{nL(c_n)}{c_n^{\rho}} = \frac{1-\rho}{\Gamma(2-\rho)\cos(\frac{\pi\rho}{2})} = C_\rho \quad (7)$$

and $Z_{\rho,\beta}$ is a standardized α-stable r.v.
The skewness parameter β must respect the following equation:

$$\frac{P(Z > z)}{P(|Z| > z)} \to \frac{1+\beta}{2}$$

Recalling that by construction $Z \in [-c, +\infty)$ the above expression reduces to,

$$\frac{P(Z > z)}{P(Z > z) + P(-Z > z)} \to \frac{P(Z > z)}{P(Z > z)} = 1 \to \frac{1+\beta}{2} \quad (8)$$

and therefore $\beta = 1$.
Hence by applying result (ii) of Theorem 3.1 in [5] we conclude the existence of the same limiting α-stable distribution also for the ordered version $Z_{(i)}$.

□

Theorem II.2. *Given the same assumptions of Theorem II.1, the estimated Gini index $G^{NP}(X_n)$ satisfies the following limit in distribution:*

$$n^{\frac{\rho-1}{\rho}}\left(\frac{G^{NP}(X_n) - g}{L_0(n)}\right) \xrightarrow{d} Q \quad (9)$$

where $g = \mathbb{E}(G^{NP}(X_n))$ and Q is an α-stable random variable $S(\rho, 1, \frac{1}{\mu}, 0)$

Proof. In Theorem II.1 we proved that $\sum Z_{(i)} \xrightarrow{d} Z_{\rho,\beta}$ if $\sum Z_i \xrightarrow{d} Z_{\rho,\beta}$. Recall that by (4) the Gini index is given by $\frac{\sum Z_{(i)}}{\sum X_i}$. Therefore it is sufficient to prove that $\frac{\sum Z_i}{\sum X_i} \xrightarrow{d} \Lambda$ to prove that $\frac{\sum Z_{(i)}}{\sum X_i} \xrightarrow{d} \Lambda$. We achieve this through a Slutsky type argument.
Call Y_n the sequence $\frac{\sum_{i=1}^{n} Z_{(i)} - n\theta}{n^{\frac{1}{\rho}} L_0(n)}$. By Theorem II.1 we have that $Y_n \xrightarrow{d} Z_{\rho,1}$. By Weak Law of Large Numbers we also have that $m_n = \frac{\sum X_i}{n} \xrightarrow{p} \mu$. By Slutsky Theorem: $\frac{Y_n}{m_n} \xrightarrow{d} \frac{1}{\mu} Z_{\rho,1}$.
What is left to prove is that also (9) converges in distribution to the same limit.
A well known theorem in probability theory [3] states that if a sequence $W_n \xrightarrow{d} \Lambda$ and $W_n - V_n = o_p(1)$ with V_n another sequence, then $V_n \xrightarrow{d} \Lambda$.
Take $W_n = \frac{Y_n}{m_n}$ and V_n the sequence defined by $V_n = n^{\frac{\rho-1}{\rho}}\left(\frac{G^{NP}(X_n) - g}{L_0(n)}\right)$, we prove that

$$\frac{Y_n}{m_n} - V_n \xrightarrow{p} 0 \quad (10)$$

Which reduces to showing that

$$\frac{n^{\frac{\rho-1}{\rho}}\theta}{L_0(n)}\left(\frac{n}{\sum X_i} - \frac{1}{\mu}\right) \xrightarrow{p} 0 \quad (11)$$

Thanks to the continuous mapping theorem $\frac{n}{\sum X_i} \xrightarrow{p} \frac{1}{\mu}$ in particular $\frac{n}{\sum X_i} - \frac{1}{\mu} \to o_p(n^{-1})$
Therefore Equation (11) goes to zero as $n \to \infty$ being $\rho \in (1,2)$ by assumption.
We conclude the proof by noting that an α-stable random variable is closed under scaling by constant [7]. In particular this parameters changes as follows: the tail parameter is unchanged and equal to ρ, the skewed parameter $\beta = 1$ and the scale parameter $\gamma = \frac{1}{\mu}$.

□

The next remark highlights some insight on the behavior of the asymptotic distribution of the "non-parametric" estimator provided by Theorem II.2.

Remark 1. In view of equation (8), in case of fat tails, the asymptotic distribution of the Gini index estimator, is always totally right skewed regardless the distribution of the data generating process. Comparing this result with the case of a finite variance data generating process (leading to a Gaussin limit distribution), we can see how the limiting distribution of the estimator undergoes a phase transition in its skewness when variance becomes infinite. Shifting from a symmetric Gaussian to a totally skewed α-stable. Thereofre a fat tailed data distribution not only induces a fatter tail limit but it also changes its shape. Therefore the estimator, whose asymptotic consistency is still ensured even in the fat tail case, [5], will approach its true value slower and from below. Evidence for these behaviors are given in in Table I. A perhaps less risky outcome would have been if the limiting distributions would have exhibit only fat tails but still symmetric behavior with a $\beta = 0$. Unfortunately this is not the case.

B. The Maximum Likelihood estimator

Theorem II.2 shows that the "non-parametric" estimator for Gini index is not the best option when dealing with infinite variance parent distributions due to the skewness ad "fatness" of its limiting distribution. A way out could be to seek estimators that still preserve asymptotic normality under fat tails. In general this is not possible in view of the α-stable Central Limit Theorem to which any "non-parametric" estimator will eventually fall into. However, a possible solution is to adopt parametric techniques. Theorem II.3 shows how, once a parametric family for the data generating process has been identified, it is possible to estimate the Gini index via Maximum Likelihood (ML) over the parameters of the chosen model for the parent distributions. The so obtained estimator will not only be asymptotically normal but it will be asymptotically efficient.

Theorem II.3. *Let $X \sim F_\theta$ such that F_θ is a parametric family belonging to the exponential family. Then the Gini index obtained by plug-in of maximum likelihood estimator of θ, $G^{ML}(X_n)_\theta$, is asymptotically normal and efficient. Namely:*

$$\sqrt{n}(G^{ML}(X_n)_\theta - g_\theta) \xrightarrow{d} N(0, g_\theta'^2 I^{-1}(\theta)) \tag{12}$$

With $g'_\theta = \frac{dg_\theta}{d\theta}$ and $I(\theta)$ being the Fisher information.

Proof. The result follows easily from the asymptotic efficiency of MLE estimators of exponential family and the invariance principle of MLE estimators. In particular the validity of the inviariance principle for the Gini index is granted by continuity and monotonicity of g_θ with respect to θ.
The asymptotic variance is obtained by application of the delta-method.

□

C. A Paretian illustration

We provide an illustration of the above results using a Pareto type I distribution as distribution for the data. More formally: consider the standard Pareto distribution for a random variable X with density given in Equation 1.

Corollary 1. *Let (X_i) be a sequence of i.i.d. observations with distribution type I Pareto with tail index $\rho \in (1,2)$.
Then the "non-parametric" Gini estimator has the following limit:*

$$G^{NP}(X_n) - g \sim S\left(\rho, \frac{d}{n^{\frac{\rho-1}{\rho}}}\frac{(\rho-1)}{\rho}, 0, 1\right) \tag{13}$$

Proof. The results is a mere application of Theorem II.2, recalling that a Pareto distribution satisfy the domain of attraction of α-stable random variables with slowly varying function $L(x) = 1$. Therefore the sequence c_n which satisfies Equation 7 becomes: $c_n = n^{\frac{1}{\rho}} C_\rho^{-1}$, and therefore $L_0(n) = C_\rho^{-1}$ independent on n, for convenience we call $C_\rho^{-1} = d$. Additionally the mean of the distribution is also a function of ρ: $\mu = \frac{\rho}{\rho-1}$.

□

Corollary 2. *Let the sample (X_i) be distributed as in Corollary 1, let G_θ^{ML} be the MLE for the Gini index as defined in Theorem II.3. In particular $G_\rho^{ML} = \frac{1}{2\rho^{ML}-1}$. Then the asymptotic distribution of G_ρ^{ML} is:*

$$G_\rho^{ML}(X_n) - g \sim N\left(0, \frac{4\rho^2}{n(2\rho-1)^4}\right) \tag{14}$$

Proof. The result follows from the fact that the Pareto distribution belongs to the exponential family and therefore satisfies the "regularity" conditions necessary for asymptotic normality and efficiency of maximum likelihood estimator.
Recall also that the Fisher information for a Pareto distribution is $\frac{1}{\rho^2}$. □

Now that we have worked out both the asymptotic distributions we can show by how much the quality of the convergence in the MLE case compared to the "non-parametric" is better.
In particular we approximate the distribution of the

deviations from the true value of the Gini index for finite sample size by Equation 13 and Equation 14.

To start we visualize in Figure 2 how the noise around the mean of the two different type of estimators is distributed and how these distributions change as the number of observations increase. In particular, to ease the comparison between MLE and "non-parametric" estimator we have fixed the number of observation in the MLE case and vary them in the "non-parametric" one. We perform this study for different type of tail index showing how big impact it has on the consistency of the estimator. It worth to point out that as the tail index goes towards 1, the threshold for infinite mean, the mode of the distribution of the "non-parametric" estimator shifts farther and farther away from the mean of the distribution (centered on 0 by definition). This effect is responsible for the small sample bias observed in applications. This phenomena is not present in the MLE case because of the normality of the limit for every value of the tail parameter.

We now make more rigorous our argument by assessing the number of observations needed for the "non-parametric" estimator to be as good as the MLE one, where the following concentration measure is taken as quantitative measure for the expression "as good as".

$$\int \mathbb{1}_{|x|>c} dP_i = P_i(|X| > c) \tag{15}$$

With P_i, $i \in S, N$ being the distribution of each estimator as in Equation 14 and Equation 13. $\mathbb{1}_A$ is the indicator function.
More precisely, we wish to compare for a fixed number of observation in 14 how many observation are required to reach the same value of the concentration measure in 13.
The problem can be rephrased in the following way.

Consider the function:

$$r(c,n) = \frac{P_S^n(|X| > c)}{P_N(|X| > c)}$$

we wish to find $\tilde{n}$ such that $r(c,\tilde{n}) = 1$ for fixed c.

Table II displays the results for some thresholds and some tail parameters.
In particular, we can see how for just a sample size of $n = 100$ the MLE estimator outperform the "non-parametric" one. In fact, a much bigger amount of observations is needed to obtain similar tail probabilities for the "non-parametric" estimator.
One thing to notice is that the number of observations needed to match the tail probability does not vary uniformly with the threshold. However this is correct since as the threshold goes to infinity: $\lim_{c\to\infty} P(|X| > c)$ and to zero: $\lim_{c\to 0} P(|X| > c)$ the tail probabilities are the same for every number of n. Therefore given the unimodality of the distributions we expect that there will be a threshold which maximize the number of observations needed to match the tail probabilities, while for all the others threshold level the number of observations will be smaller.

Figure 3 additionally shows some examples on how the parity of the tail probabilities is reached for different thresholds c and different tails index.
We conclude that when in presence of fat tailed wealth distribution with infinite variance a plug-in MLE based estimator should be preferred with respect to the "non-parametric" one due to its normality and more efficient use of the observations.

TABLE II: Inside the Table, the optimal number of observations $\tilde{n}$ needed in order to match tail probably of asymptotic MLE distribution with fixed $n = 100$.

	Treshold c:			
α	0.005	0.01	0.015	0.02
1.8	321	1242	2827	2244
1.5	844	836	2925	23036
1.2	402200	194372	111809	73888

III. Small samples correction

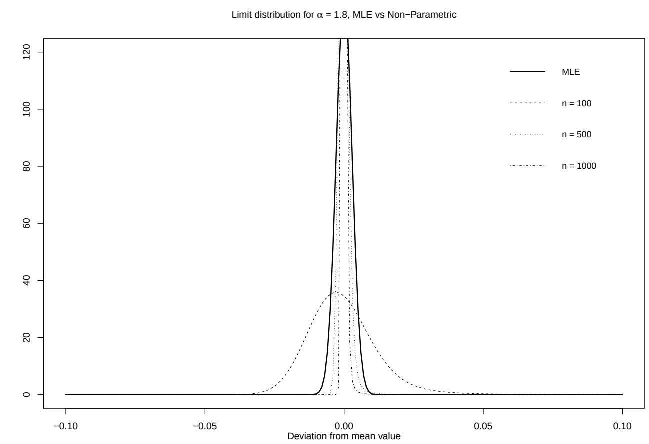
An alternative approach, instead of assuming a parametric data distribution and compute the Gini through MLE, could be to capitalize the information given by Theorem II.2 to try to build a correction mechanism for the bias in the "non-parametric" estimator which arises especially for small samples size.

The key idea is to recognize that for unimodal distributions most of the observations comes from around the mode of the distribution, in symmetric distributions the mode and the mean coincide and therefore we expect that most of the observations will be close to the mean value as well. For skewed distributions this is not the case. In particular for right skewed continuous unimodal distributions the mode is lower than the mean. Therefore, given that the distribution of the "non-parametric" Gini index is right skewed we expect that the realized (i.e. observed) value of the Gini index will be usually lower than the true value of the Gini placed at the mean level. In particular we can quantify this difference (i.e. the bias) by looking at the distance between the mode and the mean of its distribution and once this distance is known we can adjust our estimate of the Gini index by adding it.
More formally we would like to derive a "corrected non-parametric" estimator $G^C(X_n)$ such that:

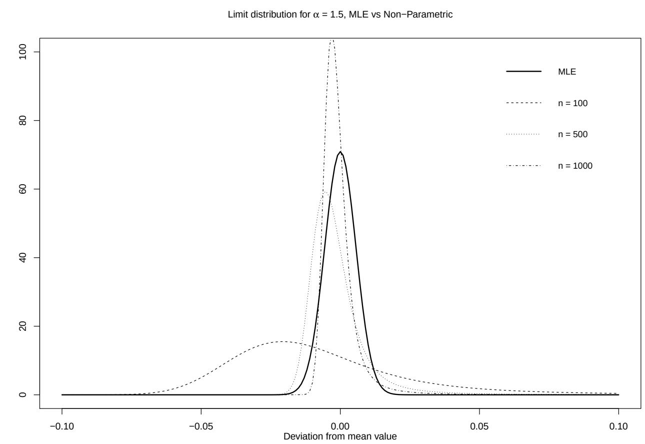
$$G^C(X_n) = G^{NP}(X_n) + ||m(G^{NP}(X_n)) - E(G^{NP}(X_n))|| \tag{16}$$

where $||m(G^{NP}(X_n)) - E(G^{NP}(X_n))||$ is the distance between the mode and the mean of the distribution of the Gini "non-parametric" estimator $G^{NP}(X_n)$ and according to our reasoning above is what we will call it the correction term.

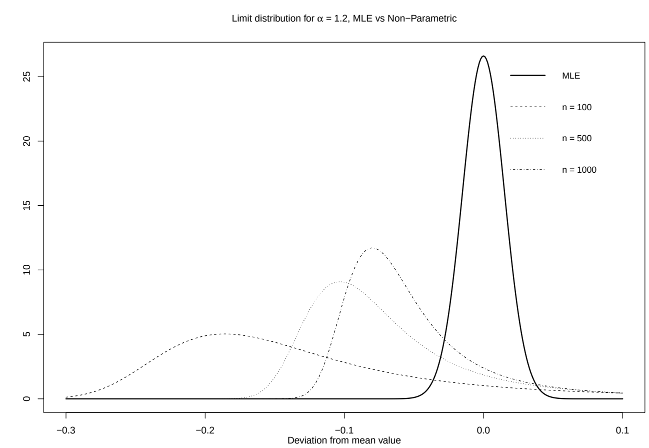
In particular, performing the type of correction described in the Equation 16 is equivalent to shift the



(a) $\alpha = 1.8$



(b) $\alpha = 1.5$



(c) $\alpha = 1.2$

Fig. 2: Comparison between MLE and "non-parametric" asymptotic distribution for different values of tail index α. Number of observation for MLE is fixed to $n = 100$. Note that despite all the distributions has mean zero the mode of the "non-parametric" one is different form it.

distribution of $G^{NP}(X_n)$ in order to place its mode on the true value of the Gini index in order to increase the probability of observing values produced by the estimator closer to it.

Ideally, we would like to measure this mode-mean distance on the exact distribution of the Gini index to get the most accurate correction. However, the finite distribution is not always easily derivable and it requires assumptions on the parametric structure of the data generating process.

Therefore, we propose to use the limiting distribution for the "non-parametric" Gini obtained in Section II to approximate the finite sample distribution and to estimate on it the mode-mean distance to use in the correction term. This procedure allows for more freedom in the modeling assumptions and potentially decrease the number of parameters to estimate since the limiting distribution only depends on the tail index of the data ρ and possibly the mean μ, which however can be assumed to be a function of the tail index itself as in the Pareto case i.e. $\mu = \frac{\rho}{\rho-1}$.

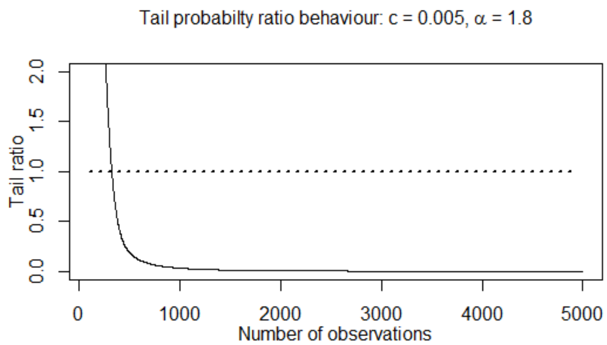
In particular by exploiting the location-scale property of α-stable distributions and Equation 9 we approximate the distribution of $G^{NP}(X_n)$ for finite sample by:

$$G^{NP}(X_n) \sim S\left(\rho, 1, \gamma(n), g\right) \tag{17}$$

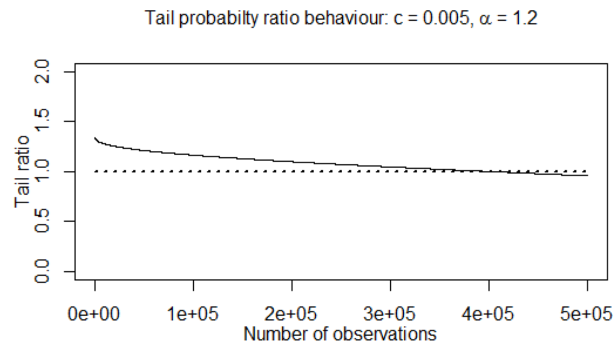
where $\gamma(n) = \frac{1}{n^{\frac{\rho-1}{\rho}}} \frac{L_0(n)}{\mu}$ is the scale parameter of the limiting distribution.

Recalling once again the location-scale property of α-stable distributions we can reduce the approximated version of the correction term in Equation 16 in the following way:

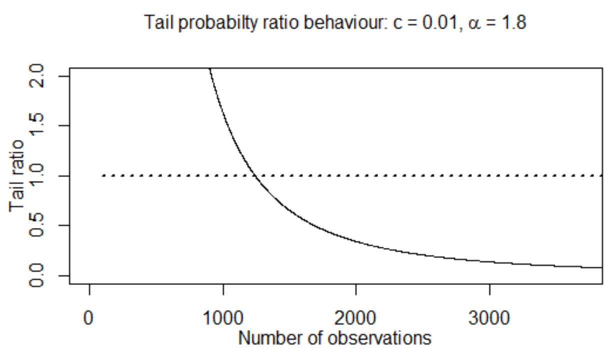
$$\begin{aligned} ||m(G^{NP}(X_n)) - E(G^{NP}(X_n))|| &\approx ||m(\rho, \gamma(n)) + g - g|| \\ &= ||m(\rho, \gamma(n))|| \end{aligned}$$



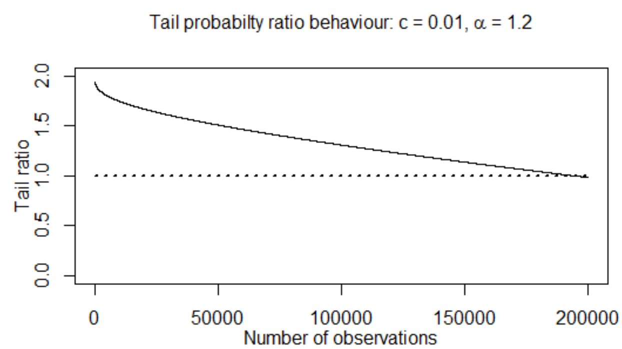
(a)



(b)



(c)



(d)

Fig. 3: Speed of convergence of probability ratio $r(c,n) = \frac{P_S^n(|X|>c)}{P_N(|X|>c)}$ as n grows, note that MLE observations are fixed to $n = 100$

where $m(\rho, \gamma(n))$ is the mode function of the α-stable distribution in Equation 17.
This means that in order to obtain the correction term the knowledge of the true Gini index is not necessary since $m(\rho, \gamma(n))$ does not depend on g.

We then proceed in computing the correction term which will be so obtained:

$$\zeta(\rho, \gamma(n)) = \arg\max_x f(x) \tag{18}$$

where $f(x)$ is the numerical density[2] of the associated α-stable distribution in Equation 17 but centered in 0. Recalling that α-stable distributions are unimodal continuous distributions we conclude that: $\zeta(\rho, \gamma(n)) = \arg\max_x f(x) = m(\rho, \gamma(n))$.
Hence our "corrected non-paramteric" estimator will have the following form:

$$G^C(X_n) = G^{NP}(X_n) + \zeta(\rho, \gamma(n)) \tag{19}$$

and asymptotic distribution:

$$G^C(X_n) \sim S\left(\rho, 1, \gamma(n), g + \zeta(\rho, \gamma(n))\right) \tag{20}$$

where the mode (i.e. the correction term $\zeta(\rho, \gamma(n))$) cannot be expressed in closed form.
Note that the correction term $\zeta(\rho, \gamma(n))$ is a function of the tail index of the data ρ and is connected to the sample size n by the scale parameter $\gamma(n)$ of the associated limiting distribution . In particular it is important to note that it is decreasing in n, and that $\lim_{n\to\infty} \zeta(\rho, \gamma(n)) \to 0$. This happens because as n increases the distribution described in 17 becomes more and more centered around its mean value pushing to zero the distance between the mode and the mean. This ensure the asymptotic equivalence of the "corrected" estimator and the "non-parametric" one. Indeed note that:

$$\begin{aligned}
&\lim_{n\to\infty} |\bar{G}(X_n)^C - G^{NP}(X_n)| \\
&\quad = \lim_{n\to\infty} |G^{NP}(X_n) + \zeta(\rho, \gamma(n)) - G^{NP}(X_n)| \\
&\quad\quad = \lim_{n\to\infty} |\zeta(\rho, \gamma(n))| \to 0
\end{aligned}$$

[2]Note also that for α-stable distributions the mode is not available in closed form, however it can be computed numerically by optimizing the numerical density [6].

However, because of the correction, $G^C(X_n)$ will behave better in small samples.
Note also that, from 20 the distribution of the corrected estimator have now mean $g + \zeta(\rho, \gamma(n))$ which converges to the true Gini g as $n \to \infty$ but its mode will be not placed in g. Therefore the estimator is placing most of its probability mass to values closed to the true Gini value.
In general, the quality of this correction depends on the distance between the exact distribution of $G^{NP}(X_n)$ and its α-stable limit, more the two are close to each other better the approximation of the mode-mean distance of the finite sample Gini distribution with its asymptotic counterpart $\zeta(\rho, \gamma(n))$ will be.

Additionally, as it is clear from the notation that the correction term depends on the tail index of the data and possibly also on their mean. These parameters, if not assumed to be known a priori must be estimated as well either from the same dataset or through other calibration procedure. Therefore the additional uncertainty due to the estimation will reflect as well on the quality of the correction.

We conclude the Section by showing the effect of this correction procedure through the following experiment. We assume that the data were coming from a Pareto distribution where the tail index ρ is know, this assumption can be weaken without leading to a big difference in the results. We simulate 1000 samples of growing size from $n = 10$ to $n = 2000$ and for each of the sample size we compute both the original "non-parametric" estimator $G^{NP}(X_n)$ and the corrected one $G^C(X_n)$. We repeat the experiment for different tail index ρ.
Figure 4 presents our results. In particular it is clear that in all the examples the corrected estimators performs equally or better than the original one. In particular for small sample size $n \leq 500$ the gain is quite remarkable. As expected the difference between the estimators decreases as a function of the sample size and the tail index reflect the fact that the correction term is decreasing both in n and the tail index ρ. This effect is due to the fact that as the tail index of an α-stable approaches the value of 2 the skewness parameter β loses his influence on the distribution, reducing the skewness and so the difference between the mode and the mean. Ultimately when the value of the tail index is equal to 2 we obtain the symmetric Gaussian distribution and the two estimators will coincide.

IV. Conclusions

In this paper we addressed the issue of asymptotic behavior of the Gini index estimator in presence of infinite variance data distribution. The issue has been unexpectedly ignored by literature so far.
In particular we derived the asymptotic distribution for the Gini index estimator when the parent distribution is in the domain of attraction of an α-stable random variable with $\alpha \in (1, 2)$. We showed that the limiting distribution is a totally skewed to the right α-stable with same tail index of the parent distribution. This result matches the empirical observations and underlines how dangerous could be using a "non-parametric" estimator of the Gini index if infinite variance is present in the data, despite asymptotic consistency of the statistics is still granted.
We then show that a parametric approach provides better asymptotic results thanks to maximum likelihood asymptotic properties.
In view of the above results we strongly suggest that, if the collected data are thought to be fat tailed distributed parametric methods should be preferred to strict "non-parametric" estimation.
However, if a parametric approach cannot be performed we proposed a simple correction mechanism for the "non-parametric" estimator based on the distance between the mode and the mean of its asymptotic distribution. We then show though an experiment with Pareto data how this correction improves the quality of the estimations. However we suggest caution in its use because of possible additional uncertainty deriving from the estimation of the correction term.

References

[1] I. Eliazar, I.M. Sokolov, *Measuring statistical evenness: A panoramic overview,* Physica A 391 (2012) 1323-1353.

[2] P. Embrechts, C. Kluppelberg, T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, (2003).

[3] Feller, Willliam. *An introduction to probability theory and its applications*. Vol. 2. John Wiley and Sons, 2008.

[4] C. Kleiber, S.Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley (2003).

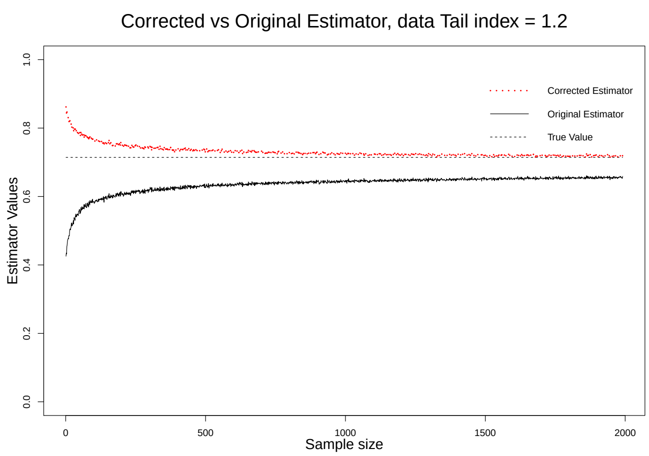
[5] Li, Deli and Rao, M Bhaskara and Tomkins, RJ, *The law of the iterated logarithm and central limit theorem for L-statistics* Journal of multivariate analysis 78 (2001) 2.

[6] Nolan, John P. *Parameterizations and modes of stable distributions.* Statistics and Probability Letters 38.2 (1998): 187-195.

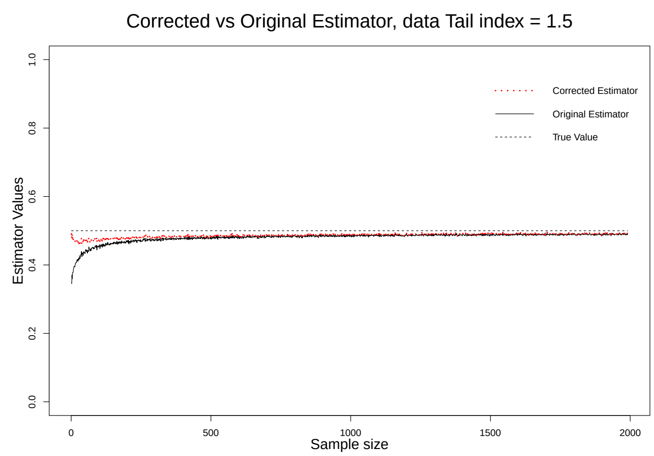
[7] Samorodnitsky, Gennady, and Murad S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance.* Vol. 1. CRC press, (1994).

[8] Yingying Yang, Shuhe Hu, Tao Wu *The tail probability of the product of depended random variables from max-domains of attraction* Statistics and Probability Letters 81 (2011) 1876-1882

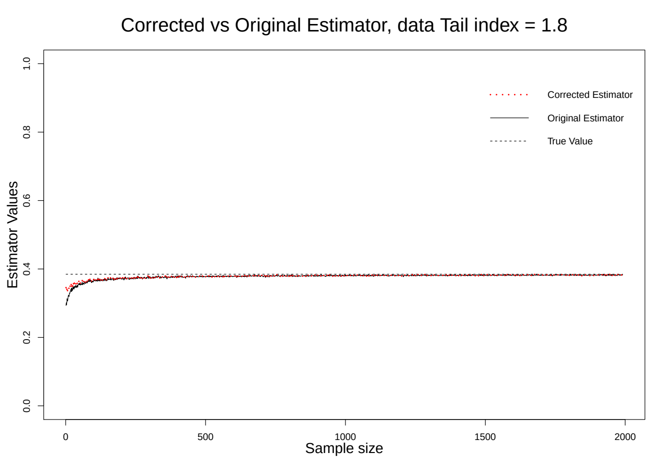
[9] Yitzhaki, Shlomo and Schechtman, Edna, *The Gini Methodology: A primer on a statistical methodology* Springer Science & Business Media (2012).



(a) $\alpha = 1.2$



(b) $\alpha = 1.5$



(c) $\alpha = 1.8$

Fig. 4: Comparison between corrected "non-parametric" estimator (in red) and usual "non-parametric" estimator (black), it is clear how especially from small samples size the corrected one improves the quality of the estimation.