Sensitivity of Inequality Measures
to Extreme Values

by

Frank A. Cowell
STICERD, London School of Economics and Political Science

and

Emmanuel Flachaire
EUREQUA, University of Paris I Panthéon-Sorbonne

Discussion Paper
No. DARP 60
March 2002

Distributional Analysis Research Programme
The Toyota Centre
Suntory and Toyota International
Centres for Economics and
Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Distributional Analysis Research Programme

The Distributional Analysis Research Programme was established in 1993 with funding from the Economic and Social Research Council. It is located within the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics and Political Science. The programme is directed by Frank Cowell. The Discussion Paper series is available free of charge and most papers are downloadable from the website. To subscribe to the DARP paper series, or for further information on the work of the Programme, please contact our Research Secretary, Sue Coles on:

Telephone:	UK+20 7955 6678
Fax:	UK+20 7955 6951
Email:	s.coles@lse.ac.uk
Web site:	http://sticerd.lse.ac.uk/research/darp.asp

Abstract

We examine the sensitivity of estimates and inequality indices to extreme values, in the sense of their robustness properties and of their statistical performance. We establish that these measures are very sensitive to the properties of the income distribution. Estimation and inference can be dramatically affected, especially when the tail of the income distribution is heavy.

Keywords: inequality measures, statistical performance, robustness
JEL classification: C1, D63

Acknowledgements

This paper is part of the research programme of the TMR network 'Living Standards, Inequality and Taxation" [Contract No. ERBFMRXCT 980248] of the European Communities, whose financial support is gratefully received.

Correspondence to: F.A. Cowell, STICERD, LSE, Houghton Street, London WC2A 2AE. (f.cowell@lse.ac.uk) and to: Emmanuel Flachaire, Universite Paris I Pantheon-Sorbonne, EUREQua, Maison des Sciences Economiques, 106-112 bd de l'Hopital, 75013 Paris - France

1 Introduction

There is a folk wisdom about inequality measures concerning their empirical performance. Some indices are commonly supposed to be particularly sensitive to specific types of change in the income distribution and may be rejected *a priori* in favour of others that are presumed to be "safer". This folk wisdom is only partially supported by formal analysis and it is appropriate to examine the issue by considering the behaviour of inequality measures with respect to extreme values. An extreme value is a observation that is highly influential on the estimate of an inequality measure. It is clear that an extreme value is not necessarily an error or some form of contamination. It could in fact be an informative observation belonging to the true distribution – *a high leverage observation*. In this paper, we study sensitivity of different inequality measures to extreme values, in both cases of contamination and of high leverage observations.

What is a "sensitive" inequality measure? This issue has been addressed in ad hoc discussion of individual measures in terms of their empirical performance on actual data (Braulke 1983). Some of the welfare-theoretical literature focuses on transfer sensitivity (Shorrocks and Foster 1987) [1] and related concepts. But it is clear that informal discussion is not a satisfactory approach for characterising; furthermore the welfare properties of inequality measures in terms of the relative impact of a transfer at different income levels will not provide a reliable guide to the way in which the measures may respond to extreme values. We need a general and empirically applicable tool.

The principal tool that we use for evaluating the influence of observations on estimates is the influence function (*IF*), taken from the theory of robust estimation. If the *IF* is unbounded for some values of z it means that the index estimate may be catastrophically affected by an extreme value close to z. Cowell and Victoria-Feser (1996) show that "if the mean has to be estimated from the sample then all scale independent or translation independent and decomposable measures have an unbounded *IF*" and that inequality measures are typically not robust to data contamination. However we will show that the *IF* has a role to play beyond the consideration of contamination on empirical estimates of inequality.

In section 2, we present some key inequality measures and their influence functions. In section 3, we study sensitivity of these inequality measures to contamination in the data, both in high and small incomes: we calculate the rates of increase to infinity of their influence functions. In section 4, we study the sensitivity of inequality measures to high leverage observations. We investigate Monte Carlo simulations to study the error in the probability rejection of a test in finite samples. Section 5 concludes.

[1] Consider a small transfer from a person with income y to one with income $y' < y$: if the inequality index I satisfies the principle of transfers then this change in the distribution reduces inequality, by an amount ΔI; if the same small income transfer were to take place from a person with income $y + k$ to one with income $y' + k$ $(k > 0)$ then the inequality-reduction ΔI would be smaller if I satisfies the principle of *transfer-sensitivity*.

2 Inequality Measure and Influence Function

Let y be a random variable – for example income – and F its probability distribution. We define the two moments

$$\mu_F = \int y \, dF(y) \qquad \text{and} \qquad \nu_F = \int \phi(y) \, dF(y) \tag{1}$$

An inequality measure fulfilling the property of decomposability can be written as

$$I(F) = \int f(y, \mu_F) \, dF(y)$$

where f is a function $\mathbb{R}^2 \rightarrow \mathbb{R}$ which is monotonic increasing in its first argument in order to respect the principle of transfers – see Cowell and Victoria-Feser (1996). We can also express most of the commonly-used indices as a function of the two moments μ_F and ν_F,

$$I(F) = \psi \left(\nu_F \, ; \, \mu_F \right) \tag{2}$$

where ϕ and ψ are functions $\mathbb{R}^2 \rightarrow \mathbb{R}$ and ψ is monotonic increasing in its first argument. For example this is true for the Generalised Entropy, Theil, Mean Logarithmic Deviation and Atkinson measures. However, the Gini index does not belong to the class of decomposable measure and cannot be reduced to the form (2): this important index will be treated separately below.

The influence function[2] is defined as the effect of an infinitesimal proportion of "bad" observations on the value of the estimator. Let us define the following *mixture distribution*

$$G_\epsilon = (1 - \epsilon) F + \epsilon H \tag{3}$$

where $0 < \epsilon < 1$ and H is some perturbation distribution; we consider that H is the cumulative distribution function which puts a point mass 1 at an arbitary income level z:

$$H(y) = \iota(y \geq z) \tag{4}$$

where $\iota(.)$ is a Boolean indicator – it takes the value 1 if the argument is true and 0 if it is false. The influence of an infinitesimal model deviation on the estimate is given by $\lim_{\epsilon \rightarrow 0} [(I(G_\epsilon) - I(F))/\epsilon]$, or $\partial I(G_\epsilon)/\partial \epsilon|_{\epsilon=0}$ when the derivative exists. Then, the influence function for an inequality measure defined as (2), is given by

$$IF \; (z \, ; I, F) \; = \; \frac{\partial \psi}{\partial \nu_{G_\epsilon}} \frac{\partial \nu_{G_\epsilon}}{\partial \epsilon} \bigg|_{\epsilon=0} + \frac{\partial \psi}{\partial \mu_{G_\epsilon}} \frac{\partial \mu_{G_\epsilon}}{\partial \epsilon} \bigg|_{\epsilon=0}$$

[2]The use of the IF to assess the robustness properties of any estimator originated in the work of Hampel (1968, 1974) and further developed in Hampel et al. (1986). Cowell and Victoria-Feser (1996) used it to study robustness properties of inequality measures.

From (1) and (3), we have

$$\nu_{G_\epsilon} = (1-\epsilon)\int \phi(y)\, dF(y) + \epsilon\phi(z) \qquad \text{and} \qquad \mu_{G_\epsilon} = (1-\epsilon)\int y\, dF(y) + \epsilon z \tag{5}$$

Then, using (5) in (2) leads to

$$IF\ (z\,; I, F) = \frac{\partial\psi}{\partial\nu_F}\cdot\left[\phi(z) - \nu_F\right] + \frac{\partial\psi}{\partial\mu_F}\cdot\left[z - \mu_F\right] \tag{6}$$

If IF is unbounded for some value of z it means that the estimate of the index may be catastrophically affected by data-contamination at income values close to z. In standard asymptotic theory (see econometric manuals), the dominant power of the sample size n of an asymptotic expansion is commonly used as an indicator of the rate of convergence of an estimator. Similarly, as an indicator of the rate of increase of IF to infinity, we can use the dominant power of z in (6).

In the light of this consider the impact of an extreme value on inequality. To assess this we require the influence function for an observation at arbitrary point z and the rate of increase to infinity of IF for specific inequality measures.

- **Generalised Entropy class** $(\alpha \neq 0, 1)$

$$I_E^\alpha = \int \frac{1}{\alpha(\alpha-1)}\left[\left(\frac{y}{\mu}\right)^\alpha - 1\right] dF(y) = \frac{1}{\alpha(\alpha-1)}\left(\frac{\nu}{\mu^\alpha} - 1\right) \tag{7}$$

where $\nu = \int y^\alpha\, dF(y)$. From (6) we can derive its influence function,

$$IF\ (I_E^\alpha) = \left[z^\alpha - \nu\right] - \frac{\nu}{(\alpha-1)\mu^{\alpha+1}}\left[z - \mu\right] \tag{8}$$

It is clear that for any fixed value of α the influence function is unbounded and we have the following situations. If $\alpha > 1$: IF tends to infinity when $z \to \infty$ at the rate of z^α. If $0 < \alpha < 1$: IF tends to infinity when $z \to \infty$ at the rate of z. If $\alpha < 0$: IF tends to infinity when $z \to \infty$ at the rate of z, and when $z \to 0$ at the rate of z^α.

- **Theil index**: this is the special case of the Generalised Entropy class where $\alpha = 1$,

$$I_E^1 = \int \frac{y}{\mu}\log\left(\frac{y}{\mu}\right) dF(y) = \frac{\nu}{\mu} - \log\mu \tag{9}$$

where $\nu = \int y \log y\, dF(y)$. From (6) we can derive its influence function,

$$IF\ (I_E^1) = \frac{1}{\mu}\left[z\log z\ - \nu\right] - \frac{\nu+\mu}{\mu^2}\left[z - \mu\right] \tag{10}$$

This influence function tends to infinity at the rate of z when $z \to \infty$.

- **Mean Logarithmic Deviation** : this is the special case of the Generalised Entropy class where $\alpha = 0$,

$$I_E^0 = -\int \log\left(\frac{y}{\mu}\right) dF(y) = \log\mu - \nu \tag{11}$$

where $\nu = \int \log y \, dF(y)$. From (6) we can derive its influence function,

$$IF\ (I_E^0) = -\Big[\log z - \nu\Big] + \frac{1}{\mu}\Big[z - \mu\Big] \tag{12}$$

This influence function tends to infinity, at the IF rate of z when $z \to \infty$, and at the rate of $\log z$ when $z \to 0$.

• **Atkinson class of measures** ($\varepsilon > 0$)

$$I_A^\varepsilon = 1 - \left[\int \left(\frac{y}{\mu}\right)^{1-\varepsilon} dF(y)\right]^{\frac{1}{1-\varepsilon}} = 1 - \frac{\nu^{\frac{1}{1-\varepsilon}}}{\mu} \qquad \varepsilon \neq 1 \tag{13}$$

where $\nu = \int y^{1-\varepsilon} \, dF(y)$. From (6) we can derive its influence function,

$$IF\ (I_A^\varepsilon) = \frac{\nu^{\frac{\varepsilon}{1-\varepsilon}}}{(\varepsilon - 1)\mu}\Big[z^{1-\varepsilon} - \nu\Big] + \frac{\nu^{\frac{1}{1-\varepsilon}}}{\mu^2}\Big[z - \mu\Big] \tag{14}$$

This influence function is unbounded and we have the following situations. If $0 < \varepsilon < 1$: IF tends to infinity when $z \to \infty$ at the rate of z. If $\varepsilon > 1$: IF tends to infinity when $z \to \infty$ at the rate of z, and when $z \to 0$ at the rate of $z^{1-\varepsilon}$.[3]

For $\varepsilon = 1$, the Atkinson index is equal to

$$I_A^1 = 1 - e^{-I_E^0} = 1 - \frac{e^\nu}{\mu} \qquad \text{where} \qquad \nu = \int \log y \, dF(y) \tag{15}$$

and then, its influence function is

$$IF\ (I_A^1) = -\frac{e^\nu}{\mu}\Big[\log z - \nu\Big] + \frac{e^\nu}{\mu^2}\Big[z - \mu\Big] \tag{16}$$

which tends to infinity, when $z \to \infty$ at the rate of z, and when $z \to 0$ at the rate of $\log z$.

• **Logarithmic Variance**

$$I_{LV} = \int \left[\log\left(\frac{y}{\mu}\right)\right]^2 dF(y) = \nu_1 - 2\,\nu_2 \log\mu + (\log\mu)^2 \tag{17}$$

[3]Note that the Atkinson index I_A^ε is ordinally equivalent to the Generalised Entropy index I_E^α for $\varepsilon = 1 - \alpha > 0$, and is a non-linear transformation of the latter:

$$I_A^\varepsilon = 1 - [(\alpha^2 - \alpha)I_E^\alpha + 1]^{\frac{1}{\alpha}}$$

where $\nu_1 = \int (\log y)^2 \, dF(y)$ and $\nu_2 = \int \log y \, dF(y)$. By extension of (6) to three parameters, we derive its influence function as

$$IF\ (I_{LV}) = \left[(\log z)^2 - \nu_1\right] - 2\log\mu\left[\log z - \nu_2\right] - \frac{2}{\mu}(\nu_2 - \log\mu)\left[z - \mu\right] \tag{18}$$

which tends to infinity at the rate of z when $z \to \infty$, and when $z \to 0$ at the rate of $(\log z)^2$.

• **Gini index**: there are several equivalent forms of this index, the most useful here is

$$I_{Gini} = 1 - 2R(F) \qquad \text{with} \qquad R(F) = \frac{1}{\mu}\int_0^1 C(F;q)\, dq \tag{19}$$

where, for all $0 \leq q \leq 1$, the Quantile function $Q(F;q)$ and the Cumulative income function $C(F;q)$ are respectively defined by

$$Q(F;q) = \inf\left\{y | F(y) \geq q\right\} \qquad \text{and} \qquad C(F;q) = \int_0^{Q(F;q)} y\, dF(y) \tag{20}$$

The IF of I_{Gini} is given by (see e.g. Monti 1991)

$$IF\ (I_{Gini}) = 2\left[R(F) - C\big(F; F(z)\big) + \frac{z}{\mu}\Big(R(F) - \big(1 - F(z)\big)\Big)\right] \tag{21}$$

which tends to infinity at the rate of z when $z \to \infty$.

3 Data contamination

We can briefly summarise the different rates of increase to infinity of the influence function for the different class of measures in the following table[4]:

Measure	Generalised Entropy, I_E^α $\alpha>1$	$0<\alpha\leq 1$	$\alpha=0$	$\alpha<0$	Atkinson, I_A^ε $0<\varepsilon<1$	$\varepsilon=1$	$\varepsilon>1$	LogVar	Gini
$z \to \infty$	z^α	z	z	z	z	z	z	z	z
$z \to 0$	-	-	$\log z$	z^α	-	$\log z$	$z^{1-\varepsilon}$	$(\log z)^2$	-

Table 1: Rates of increase to infinity of the influence function

First of all, we can see that all the inequality measures discussed here have an influence function which tends to infinity when z tends to infinity. Clearly, the measures are not robust to data contamination in high incomes – see Cowell and Victoria-Feser (1996). However, we can go further and see that the rate of increase of IF to infinity, when z tends to infinity, is faster for the Generalised Entropy measures with $\alpha > 1$. In other words:

[4] A "-" in the table corresponds to the case where IF converges to a constant coefficient.

Remark 1: *Generalised Entropy measures with $\alpha > 1$ are very sensitive to high incomes in the data.*

Furthermore, we can see that, for some measures, IF tends to infinity when z tends to zero: the rate of increase is faster for the Generalised Entropy measures with $\alpha < 0$ and for the Atkinson measures with $\varepsilon > 1$. This results suggests that,

Remark 2: *Generalised Entropy measures with $\alpha < 0$, and Atkinson measures with $\varepsilon > 1$ are very sensitive to small incomes in the data.*

Based on the rates of convergence from table 1, we cannot compare the speed of increase of IF for different values of $0 < \alpha < 1$ and $0 < \varepsilon < 1$. If we do not know the income distribution, we cannot compare influence functions of different class of measures, because the IFs are functions of moments. However, we can choose an income distribution and then plot and compare their influence functions for this special case. Let us take the distribution proposed by Singh and Maddala (1976), which is a member of the Burr family (type 12), and can quite successfully mimic observed income distributions in various countries, as shown by Brachman, Stich, and Trede (1996). The cumulative distribution function, or CDF, of the Singh-Maddala distribution is

$$F(y) = 1 - \frac{1}{(1 + ay^b)^c} \tag{22}$$

We use the parameter values $a = 100$, $b = 2.8$, $c = 1.7$, which closely mirrors the net income distribution of German households, apart from a scale factor. It can be shown that the moments of the distribution with CDF (22) are

$$E(y^\alpha) = \int_0^\infty y^\alpha dF(y) = a^{-\alpha/b} \frac{\Gamma(1 + \alpha b^{-1})\,\Gamma(c - \alpha b^{-1})}{\Gamma(c)} \tag{23}$$

where $\Gamma(z)$ is the gamma function (see e.g. McDonald 1984). Using (23) in (7) gives us true values of Generalised Entropy measures for the Singh-Maddala distribution, from which we can derive true values of Theil (9) and Mean Logarithmic Deviation (11) measures by l'Hopital's rule, we obtain

$$E(y\log y) = \mu_F\, b^{-1}(\psi(b^{-1} + 1) - \psi(c - b^{-1}) - \log a)$$
$$E(\log y) = b^{-1}(\psi(1) - \psi(c) - \log a)$$

where $\psi(z) \equiv \Gamma'(z)/\Gamma(z)$ is the digamma function. With these moments, we can compute true values of Generalized Entropy measures. For our choice of parameter values, we have $I_E^2 = 0.162037$, $I_E^1 = 0.140115$, $I_E^{0.5} = 0.139728$, $I_E^0 = 0.146011$, $I_E^{-1} = 0.189812$, $I_E^{-2} = 0.386647$. In addition, we approximate true values of Logarithmic Variance and Gini measures from a sample of 10 millions observations drawn from the Singh-Maddala distribution: $I_{LV} = 0.332128$ and $I_{Gini} = 0.288714$.

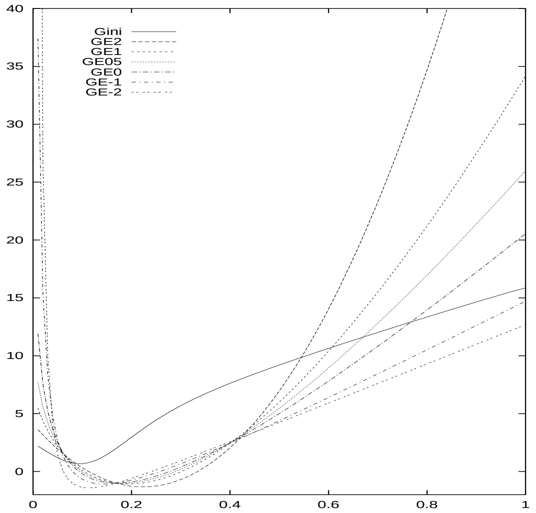

Figure 1: *IF*s of Generalised Entropy I_E^α

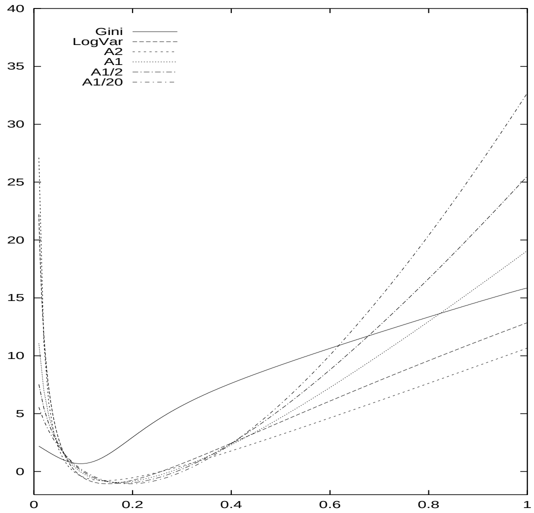

Figure 2: *IF*s of Atkinson I_A^ε

In order to compare different *IF*s we need to normalize them: we divide the influence function by its index: *IF* $(I(\hat{F}))/I(\hat{F})$, that we call *relative influence function.* In figure 1, we plot the relative influence functions for different Generalised Entropy measures, with $\alpha = 2, 1, 0.5, 0, -1, -2$, and for Gini index, as functions of z in the x-axis. For Generalised Entropy measures, we can see that, when z increases, *IF* increases faster with high values of α ; when z tends to 0, *IF* increases faster with small values of α. *IF* of Gini index increases slower than others but is larger for moderate values of z. In figure 2, we plot relative influence functions for different Atkinson measures, with $\varepsilon = 2, 1, 0.5, 0.05$, for the Logarithmic Variance and for the Gini index. For Atkinson measures, we can see an opposite relation: when z increases, *IF* increases as quickly as ε is positive and small ; when z tends to 0 *IF* increases as quickly as ε is large. *IF* of Logarithmic Variance increases slowly as z tends to infinity, but quickly as z tends to zero. Once again, the *IF* of Gini index increases slower and is larger for moderate values of z.

Comparison of the Gini index with the Logarithmic Variance, Generalised Entropy or Atkinson influence functions does not lead to clear conclusions. We used a simulation study to evaluate the impact of a contamination in large and small observations for different measures of inequality. We simulated 100 samples of 200 observations from the Singh-Maddala distribution. Then, we contaminated one observation chosen randomly by multiplying it by 10 in the case of contamination in high values, and by dividing it by 10 in the case of contamination in small values. Let Y_i, $i = 1, \ldots, N$, be an IID sample from the distribution F, $\hat{F}$ is the empirical distribution function of this sample and $\hat{F}^\star$ the empirical distribution function of the contaminated sample. Then, for each sample, we compute the quantity

$$RC(I) = \frac{I(\hat{F}) - I(\hat{F}^\star)}{I(\hat{F})} \tag{24}$$

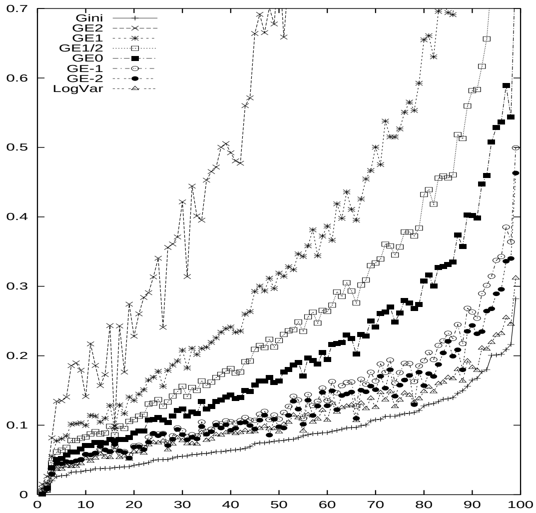

Figure 3: Contamination in high values

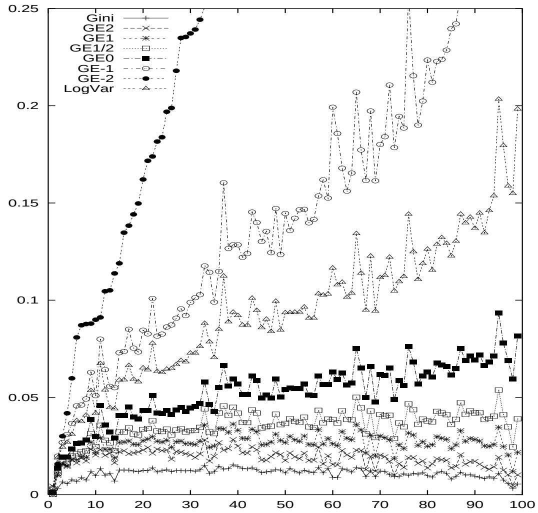

Figure 4: Contamination in small values

which evaluates the relative impact of a contamination on the index $I(\hat{F})$. We can then plot and compare realizations of $RC(I)$ for different measures. In order to have a plot that is easy to interpret, we sorted the samples such that realizations of $RC(I)$ for one chosen measure are increasing.

Figure 3 plots realizations of $RC(I)$ for the Gini index, the Logarithmic Variance index and the Generalised Entropy measures with $\alpha = 2, 1, 0.5, 0, -1, 2$, when contamination is in high values. The y-axis is $RC(I)$ and the x-axis is the 100 different samples, sorted such that Gini realizations are increasing. Figure 4 plots realizations of $RC(I)$ for the same measures when contamination is in small values. We can see from figures 3 and 4 that Gini index is less affected by contamination than Generalised Entropy measures. However, the impact of the contamination on Logarithmic Variance and Generalised Entropy measures with $0 \leq \alpha \leq 1$ is relatively small compared to measures with $\alpha < 0$ or $\alpha > 1$. In addition, the Generalised Entropy measures with $0 \leq \alpha \leq 1$ is less sensitive to contamination in high values if α is small. Finally, the Logarithmic Variance index is slightly less sensitive to high incomes than the Gini, but nearly as sensitive to small incomes than the Generalised Entropy with $\alpha = -1$. These results suggest that,

Remark 3: *The Gini index is less sensitive to contamination in high incomes than the Generalised Entropy class of measures, which is less sensitive as α is small.*

Similarly to the Generalised Entropy class of measures, we find the same results for the Atkinson class of measures with an opposite relation to its parameter ($\varepsilon = 1 - \alpha$).

4 High leverage observations

An extreme value is not necessarily an error or some sort of contamination: it could be an observation belonging to the true distribution and that conveys important information. This observation is extreme in the sense that its influence on the inequality measure estimate is very important. Such observations can have important consequences on the statistical performance of the measure. Davidson and Flachaire (2001) showed that, even in very large samples, the error in the rejection probability, or ERP, of an asymptotic or bootstrap test based on the Theil index, can be significant and that tests are therefore not reliable. They investigated three main causes of these bad performances. First, almost all indices are nonlinear functions of sample moments, thereby inducing biases and nonnormality in estimates of these indices. Second, estimates of the covariances of the sample moments used to construct indices are often very noisy. Third, the indices are often extremely sensitive to the exact nature of the tails of the distribution, with the result that a bootstrap sample in which nothing is resampled from the tail can have properties very different from those of the population. Their simulation experiments show that the third cause is often quantitatively the most important. Our remark 3 suggests that statistical performance should be better with the Mean Logarithmic Deviation index and Generalised Entropy measures with $0 < \alpha < 1$, than with Theil index. In this section, we study statistical performance of different measures based on Monte Carlo simulation, we study ERPs of asymptotic and bootstrap tests for different measures of inequality.

If Y_i, $i = 1, \dots, N$, is an IID sample from the distribution F, then the empirical distribution function of this sample is

$$\hat{F}(y) = \frac{1}{N} \sum_{i=1}^{N} \iota(Y_i \leq y), \tag{25}$$

A decomposable inequality measure can be estimated by using the sample moments

$$\mu_{\hat{F}} = \frac{1}{N} \sum_{i=1}^{N} Y_i, \qquad \text{and} \qquad \nu_{\hat{F}} = \frac{1}{N} \sum_{i=1}^{N} \phi(Y_i) \tag{26}$$

which are consistent and asymptotically normal, and

$$I(\hat{F}) \equiv \psi\left(\nu_{\hat{F}}\,;\,\mu_{\hat{F}}\right) \tag{27}$$

This estimate is also consistent and asymptotically normal, with asymptotic variance that can be calculated by the delta method. Specifically, if $\hat{\Sigma}$ is the estimate of the covariance matrix of $\nu_{\hat{F}}$ and $\mu_{\hat{F}}$[5], the variance estimate for $I(\hat{F})$ is

$$\hat{V}\big(I(\hat{F})\big) = \quad \big[\partial\psi/\partial\nu_{\hat{F}}\,;\, \partial\psi/\partial\mu_{\hat{F}}\big] \quad \hat{\Sigma} \quad \begin{bmatrix} \partial\psi/\partial\nu_{\hat{F}} \\ \partial\psi/\partial\mu_{\hat{F}} \end{bmatrix} \tag{28}$$

Using the estimate (27) and the estimate (28) of its variance, it is possible to test hypotheses about $I(F)$ and to construct confidence intervals for it. The obvious way to proceed is to base inference on asymptotic t statistics computed using (27) and (28). Consider a test of the hypothesis that $I(F) = I_0$, for some given value I_0. The asymptotic t statistic for this hypothesis, based on $\hat{I} \equiv I(\hat{F})$, is

$$W = (\hat{I} - I_0)/(\hat{V}(\hat{I}))^{1/2}, \tag{29}$$

where by $\hat{V}(\hat{I})$ we denote the variance estimate (28). We compute an asymptotic P value based on the standard normal distribution or on the Student distribution with N degree of freedom. For the particular case of the Gini index, its estimate is computed by

$$\hat{I}_{Gini} = \frac{1}{\mu_{\hat{F}}} \sum_{i=1}^{N} \sum_{j=1}^{N} |Y_i - Y_j| \tag{30}$$

and its standard deviation is computed as defined by Cowell (1989). Then, an asymptotic P value is calculated from (29).

In order to construct a bootstrap test, we resample from the original data. Since the test statistic we have looked at so far is asymptotically pivotal, bootstrap inference should be superior to asymptotic inference because of Beran's (1988) result on prepivoting. After computing W from the observed sample, one draws B bootstrap samples, each of the same size N as the observed sample, by making N draws with replacement from the N observed incomes Y_i, $i = 1, \ldots, N$, where each Y_i has probability $1/N$ of being selected on each draw. Then, for bootstrap sample j, $j = 1, \ldots, B$, a bootstrap statistic $W_j^\star$ is computed in exactly the same way as W from the original data, except that I_0 in the numerator (29) is replaced by the index $\hat{I}$ estimated from the original data. This replacement is necessary in order that the hypothesis that is tested by the bootstrap statistics should actually be true for the population from which the bootstrap samples are drawn, that is, the original sample. Details of the theoretical reasons for this replacement can be found in many standard references, such as Hall (1992). This method is known as the *percentile-t* or *bootstrap-t* method. The bootstrap P value is just the proportion of the bootstrap samples for which the bootstrap statistic is more extreme than the statistic computed from the original data.

[5] Then, $\hat{\Sigma}$ is a symmetric 4×4 matrix with arguments calculated as: $\hat{\Sigma}_{11} = N^{-1} \sum_{i=1}^{N} (\phi(Y_i) - \nu_{\hat{F}})^2$, $\hat{\Sigma}_{22} = N^{-1} \sum_{i=1}^{N} (Y_i - \mu_{\hat{F}})^2$ and $\hat{\Sigma}_{12} = \hat{\Sigma}_{21} = N^{-1} \sum_{i=1}^{N} (Y_i - \mu_{\hat{F}})(\phi(Y_i) - \nu_{\hat{F}})$.

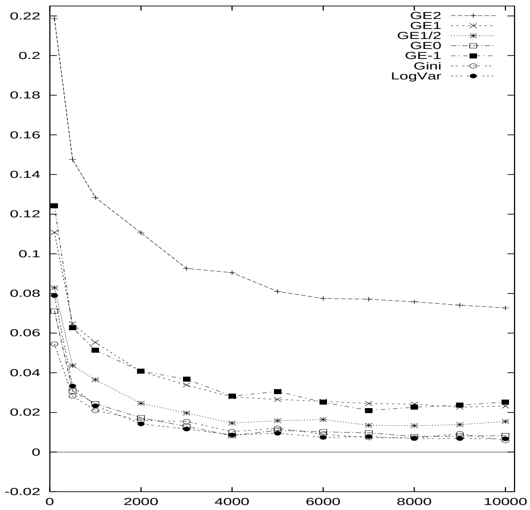

Figure 5: ERP of asymptotic tests

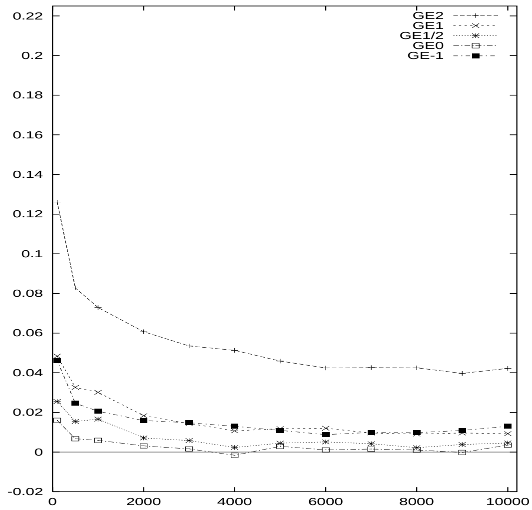

Figure 6: ERP of bootstrap tests

Thus, for the usual two-tailed test, the bootstrap P value, $P^\star$, is

$$P^\star = \frac{1}{B}\sum_{j=1}^{B} \iota(|W_j^\star| > |W|), \tag{31}$$

where $\iota(.)$ is the indicator function. For the investigation carried out here, it is more revealing to consider one-tailed tests, for which rejection occurs when the statistic is too negative.

Figure 5 shows ERPs of asymptotic tests at the nominal level 0.05, that is, the difference between the actual and nominal probabilities of rejection, for different Generalised Entropy measures, for Logarithmic Variance index and for the Gini index, when the sample size increases[6]. The plot for a statistic that yields tests with no size distortion coincides with the horizontal axis. Let us take an example from figure 5: for $N = 2{,}000$ observations, the ERP of the Generalised Entropy measure with $\alpha = 2$ (GE2) is approximately equal to 0.11: it means that asymptotic test over-rejects the null hypothesis and that the actual level is 16%, when the nominal level is 5%. In our simulations, the number of replications is 10,000. From Figure 5, it is clear that the error in the rejection probabilities, or ERP, of asymptotic tests is very large in moderate samples and decreases very slowly as the sample size increases; the distortion is still significant in very large samples. We can see that the Gini index, Logarithmic Variance index and Generalised Entropy measure with $\alpha = 0$ perform similarly. We investigate additional experiments with two extremely large samples, 50,000 and 100,000 observations. The results are shown in table 2 from which we can see that the actual level is still nearly twice the nominal level for Generalised Entropy

[6] $N = 100; 500; 1{,}000; 2{,}000; 3{,}000; 4{,}000; 5{,}000; 6{,}000; 7{,}000; 8{,}000; 9{,}000; 10{,}000$

Measure	I_E^2	I_E^1	$I_E^{0.5}$	I_E^0	I_E^{-1}
N=50,000	0.0492	0.0096	0.0054	0.0024	0.0113
N=100,000	0.0415	0.0096	0.0052	0.0043	0.0125

Table 2: ERP of asymptotic tests at nominal level 5% in huge sample.

measures with $\alpha = 2$. The distortion is very small in this case only for $\alpha = 0, 0.5$.

Figure 6 shows the ERPs of bootstrap tests. Firstly, we can see that distortions are reduced for all measures when we use the bootstrap. However, ERP of the Generalised Entropy measure with $\alpha = 2$ is still very large even in large samples, ERPs of Generalised Entropy measure with $\alpha = 1, 0.5, -1$ is small only for large samples. The Generalised Entropy measure with $\alpha = 0$, or Mean Logarithmic Deviation index, performs better than others and its ERP is quite small for 500 or more observations. These results suggest that,

Remark 4: *The rate of convergence to zero of the error in the rejection probability of asymptotic and bootstrap tests is very slow. Tests based on Generalised Entropy measures can be unreliable even in large samples.*

Computations for the Gini index are very time-intensive and we computed ERPs of this index only for $N = 100; 500; 1{,}000$ observations: we found similar ERPs with Generalised Entropy index with $\alpha = 0$. Experiments on Logarithmic Variance index show that it performs similarly to the Gini index and the Generalised Entropy measure with $\alpha = 0$. Moreover, additional experiments have been done for the Atkinson class of measures, we find similar results with an opposite relation to ε. These results lead us to conclude that,

Remark 5: *Generalised Entropy with $\alpha = 0$, Atkinson with $\varepsilon = 1$, Logarithmic Variance and Gini indexes perform similarly in finite samples.*

In practice, we can detect extreme values by considering the sensitivity of the index estimate to influential observations, in the sense that deleting them would change the estimate substantially. The effect of a single observation on $\hat{I}$ can be seen by comparing $\hat{I}$ with $\hat{I}^{(i)}$, the estimate of $I(F)$ that would be obtained if we used a sample from which the i^{th} observation was omitted. Let us define $\widehat{IF}_i$ as a measure of the influence of observation i, as follows:

$$\widehat{IF}_i = (\hat{I} - \hat{I}^{(i)})/\hat{I} \tag{32}$$

Figure 7 plots the values of $\widehat{IF}_i$ for different inequality measures and for the 10 highest, the 10 in the middle and the 10 smallest observations of a sorted sample of $N = 5,000$ observations drawn from the Singh-Maddala distribution. We can see that observations in the middle of the sorted sample don't affect estimates compared to smallest or highest

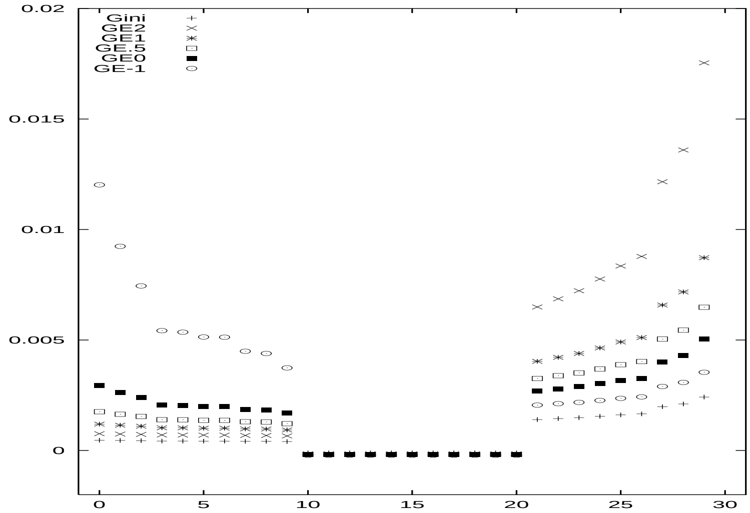

Figure 7: Influential observations

observations. Moreover, we can see that highest values are more influential than smallest values. Furthermore, we can see that the highest value is very influential for the Generalised Entropy measure with $\alpha = 2$, its estimate should be modified by nearly 0.018 if we remove it, respectively more influential than for $\alpha = 1$, $\alpha = 0.5$, $\alpha = 0$, $\alpha = -1$ and Gini index. On the other hand, we can see that the Generalised Entropy index with $\alpha = -1$ is highly influenced by the smallest observation. Finally, this plot can be extremely useful in practice to identify extreme values which may affect estimates substantially, and to select an inequality measure more or less sensitive to these outliers relatively to our preceding study and remarks.

5 Income distribution

Results of preceding sections are based on a specific choice of the income distribution: Singh-Maddala distribution with special choice of parameters. In this section, we extend our preceding results to other choice of parameters and distributions. In addition to the Singh-Maddala, we use the Pareto and the Lognormal distributions with different choices of parameters.

The CDF of the **Singh-Maddala** distribution is defined in (22). In our simulation, we use $a = 100$, $b = 2.8$ and $c = 0.7, 1.2, 1.7$. Singh-Maddala distributions for these different values of c are plotted in figure 8, we can see that the upper tail is thicker and longer as c decreases.

The CDF of the **Pareto** distribution (type I) is defined by

$$\Pi(y\,;y_l,\alpha) = 1 - \left(\frac{y_l}{y}\right)^{\alpha} \tag{33}$$

where $y_l > 0$ is a scale parameter and $\alpha > 0$. The formulas for Theil and Mean Logarithmic deviation measures, given that the underlying distribution is Pareto, are respectively

$$I_E^1(\Pi) = \frac{1}{\alpha-1} + \log\frac{\alpha-1}{\alpha} \qquad \text{and} \qquad I_E^0(\Pi) = -\frac{1}{\alpha} - \log\frac{\alpha-1}{\alpha} \tag{34}$$

see Cowell (1977). In our simulation, we use $y_l = 0.1$ and $\alpha = 1.5, 2, 2.5$. Pareto distributions for these different values of α are plotted in figure 10, we can see that the upper tail is thicker and longer as α decreases.

The CDF of the **Lognormal** distribution is defined by

$$\Lambda(y\,;\mu,\sigma^2) = \int_0^y \frac{1}{\sqrt{2\pi}\sigma x}\, exp\Big(-\frac{1}{2\sigma^2}[\log x - \mu]^2\Big)\, dx \tag{35}$$

The formulas for Theil and Mean Logarithmic deviation measures, given that the underlying distribution is Lognormal, are both equal to

$$I_E^1(\Lambda) = I_E^0(\Lambda) = \sigma^2/2 \tag{36}$$

see Cowell (1977). In our simulation, we use $\mu = -2$ and $\sigma = 1, 0.7, 0.5$. Lognormal distributions for these different values of σ are plotted in figure 12, we can see that the upper tail is thicker and longer as σ increases.

As described in section 3 at figure 3, we use a similar simulation study to evaluate the impact of a **contamination** in large observations for the Mean Logarithmic Deviation index and for the different underlying distributions described above. Respectively, figures 9, 11 and 13 plot results of our three different choices of parameters for Singh-Maddala, Pareto and Lognormal distributions. In all cases, we can see that the Mean Logarithmic Deviation index is more sensitive to contamination when the upper tail is heavy, it is to say thick and long ($c = 0.7$, $\alpha = 1.5$, $\sigma = 1$), and less sensitive when the upper tail is thin and short ($c = 1.7$, $\alpha = 2.5$, $\sigma = 0.5$)

Remark 6: *Mean Logarithmic Deviation index is more sensitive to contamination in high incomes when the underlying distribution upper tail is heavy.*

As described in section 4, we study statistical performance of Theil and Mean Logarithmic Deviation measures for the different underlying distributions described above. Table 3 gives result of **ERPs** of asymptotic test for Mean Logarithmic Deviation index. The column 4, Singh-Maddala with $c = 1.7$ is similar to GE_0 curve in figure 5. First, we can see that in many cases ERP is quite large and decreases slowly as the number of observations increases. In addition, if we compare columns, we can see that ERPs are more significant with heavy upper tails ($c = 0.7$, $\alpha = 1.5$, $\sigma = 1$) than with thin and shorter upper tails ($c = 1.7$,

$\alpha = 2.5$, $\sigma = 0.5$). Finally, ERPs are smaller with Lognormal than with Singh-Maddala distributions, which ERPs are smaller than with Pareto distributions. Lognormal upper tail are known to decrease faster, as exponential function, than Singh-Maddala and Pareto distributions, which decrease as power function. Table 4 presents ERPs of bootstrap tests for Mean Logarithmic Deviation index.

Tables 5 and 6 present respectively ERPs of asymptotic and bootstrap tests for Theil index. We can see the same results from all those tables,

> **Remark 7:** *Error in the rejection probability, or ERP, of an asymptotic and bootstrap test based on the Mean Logarithmic Deviation or Theil index, is more significant when the underlying distribution upper tail is heavy.*

In addition, we can see that bootstrap tests largely improve numerical performance.

6 Conclusion

Extreme values are important for the empirical analysis of income distribution. This is already known in the case of data-contamination, but it is also true in cases where the extreme values "really belong" to the data. Very large incomes matter both in principle and practice when it comes to inequality judgments.

A careful analysis of the behaviour of inequality indices using standard statistical techniques yields a number of insights into their performance with respect to extreme values. We can summarise some of the main points that emerge from our approach by focusing on two issues that often arise in the context of applied work on income distribution

1. Why use the Gini coefficient? Apart from its intrinsic attractions (for example its undeniable intuitive appeal) there is sometimes a supposition that it is going to be less prone to the influence of outliers than some of the alternative candidate inequality indices. As might be expected the Gini coefficient is indeed less sensitive than Generalised entropy indices to contamination in high incomes. However, in terms of performance in finite samples there is little to choose between the Gini coefficient and I_E^0 the GE index with $\alpha = 0$ (or equivalently the Atkinson index with $\varepsilon = 1$). There is also little to choose between the Gini and the logarithmic variance.

2. There is piece of folk wisdom which suggests that "the bootstrap will get you out of trouble." Our results make clear that the bootstrap performs better than asymptotic methods, but does it perform well enough? From Figure 6 we see that, in terms of the ERP, the bootstrap does well only for the Gini, I_E^0 and the logarithmic variance. Furthermore if we use a distribution with a heavy upper tail (Singh-Maddala with a low value of c and Pareto with low value of the Pareto coefficient) Table 4 shows that the bootstrap performs poorly in the case of I_E^0, even in large samples.

References

Beran, R. (1988). "Prepivoting test statistics: a bootstrap view of asymptotic refinements". *Journal of the American Statistical Association 83*(403), 687–697.

Brachman, K., A. Stich, and M. Trede (1996). "Evaluating parametric income distribution models". *Allgemeines Statistisches Archiv 80*, 285–298.

Braulke, M. (1983). An approximation to the Gini coefficient for population based on sparse information for sub-groups. *Journal of Development Economics 2*(12), 75–81.

Cowell, F. A. (1977). *Measuring Inequality.* Philip Allan Publishers Limited, Oxford.

Cowell, F. A. (1989). "Sampling variance and decomposable inequality measures". *Journal of Econometrics 42*, 27–42.

Cowell, F. A. and M.-P. Victoria-Feser (1996). "Robustness properties of inequality measures". *Econometrica 64*, 77–101.

Davidson, R. and E. Flachaire (2001). "Asymptotic and bootstrap inference for inequality and poverty measures". unpublished.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer Series in Statistics. New York: Springer Verlag.

Hampel, F. R. (1968). *Contribution to the Theory of Robust Estimation.* Ph. D. thesis, University of California, Berkeley.

Hampel, F. R. (1974). "The influence curve and its role in robust estimation". *Journal of the American Statistical Association 69*, 383–393.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions.* New York: John Wiley.

McDonald, J. B. (1984). "Some generalized functions for the size distribution income". *Econometrica 52*, 647–663.

Monti, A. C. (1991). "The study of the gini concentration ratio by means of the influence function". *Statistica 51*, 561–577.

Shorrocks, A. F. and J. E. Foster (1987). Transfer-sensitive inequality measures. *Review of Economic Studies 54*, 485–498.

Singh, S. K. and G. S. Maddala (1976). "A function for size distribution of incomes". *Econometrica 44*, 963–970.

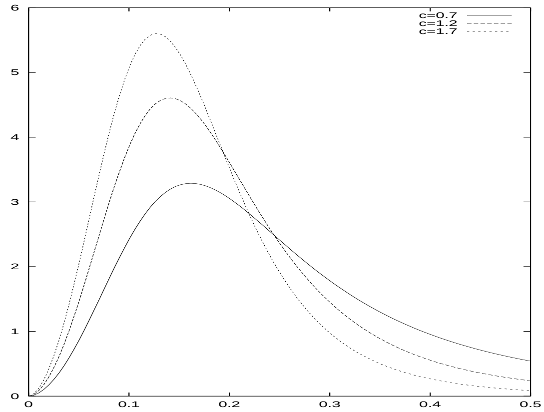

Figure 8: Singh-Maddala distributions

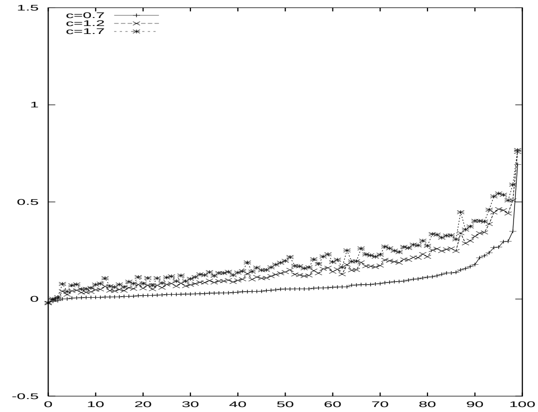

Figure 9: Contamination, Singh-Maddala

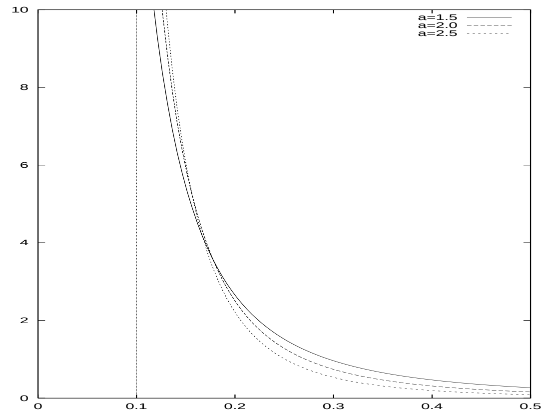

Figure 10: Pareto distributions

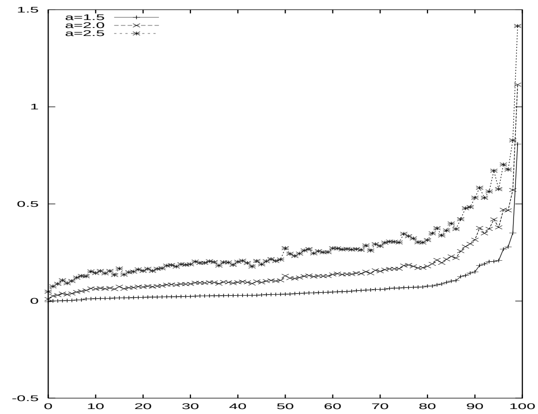

Figure 11: Contamination, Pareto

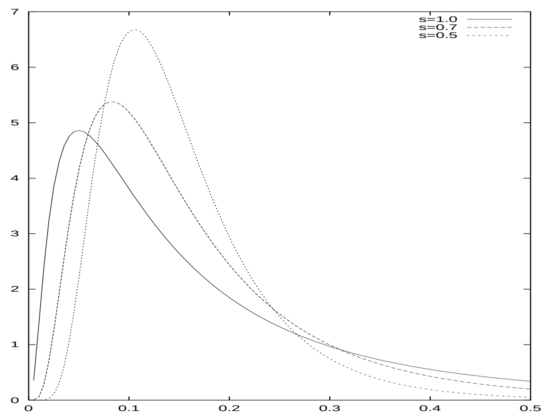

Figure 12: LogNormal distribution

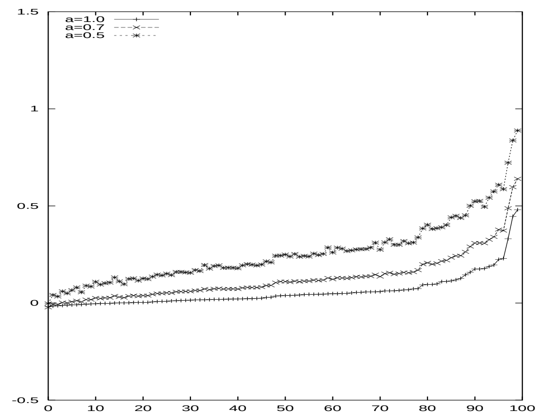

Figure 13: Contamination, Lognormal

Nobs	Singh-Maddala $c = 0.7$	Singh-Maddala $c = 1.2$	Singh-Maddala $c = 1.7$	Pareto $\alpha = 1.5$	Pareto $\alpha = 2$	Pareto $\alpha = 2.5$	Lognormal $\sigma = 1$	Lognormal $\sigma = 0.7$	Lognormal $\sigma = 0.5$
100	0.2504	0.1030	0.0711	0.4078	0.2876	0.2392	0.0947	0.0712	0.0581
500	0.1749	0.0578	0.0308	0.3226	0.1894	0.1427	0.0418	0.0238	0.0186
1000	0.1553	0.0503	0.0244	0.2933	0.1647	0.1188	0.0390	0.0248	0.0207
2000	0.1345	0.0350	0.0173	0.2753	0.1392	0.0958	0.0225	0.0134	0.0088
3000	0.1163	0.0285	0.0129	0.2625	0.1243	0.0785	0.0208	0.0125	0.0094
4000	0.1135	0.0236	0.0083	0.2607	0.1168	0.0741	0.0140	0.0071	0.0048
5000	0.1033	0.0222	0.0110	0.2509	0.1080	0.0655	0.0134	0.0042	0.0028

Table 3: ERP of asymptotic tests at nominal level 5%, MLD measure (I_E^0).

Nobs	Singh-Maddala $c = 0.7$	Singh-Maddala $c = 1.2$	Singh-Maddala $c = 1.7$	Pareto $\alpha = 1.5$	Pareto $\alpha = 2$	Pareto $\alpha = 2.5$	Lognormal $\sigma = 1$	Lognormal $\sigma = 0.7$	Lognormal $\sigma = 0.5$
100	0.1378	0.0370	0.0160	0.2155	0.1399	0.1097	0.0398	0.0197	0.0118
500	0.0957	0.0251	0.0067	0.1859	0.0940	0.0654	0.0136	0.0041	-0.0020
1000	0.0886	0.0236	0.0059	0.1742	0.0862	0.0566	0.0164	0.0070	0.0051
2000	0.0767	0.0127	0.0031	0.1655	0.0755	0.0455	0.0073	0.0007	-0.0024
3000	0.0668	0.0116	0.0016	0.1644	0.0637	0.0380	0.0030	-0.0001	-0.0009
4000	0.0677	0.0073	-0.0016	0.1639	0.0652	0.0374	-0.0001	-0.0024	-0.0027
5000	0.0629	0.0090	0.0029	0.1566	0.0623	0.0334	-0.0013	-0.0034	-0.0045

Table 4: ERP of bootstrap tests at nominal level 5%, MLD measure (I_E^0).

Nobs	Singh-Maddala $c = 0.7$	Singh-Maddala $c = 1.2$	Singh-Maddala $c = 1.7$	Pareto $\alpha = 1.5$	Pareto $\alpha = 2$	Pareto $\alpha = 2.5$	Lognormal $\sigma = 1$	Lognormal $\sigma = 0.7$	Lognormal $\sigma = 0.5$
100	0.4527	0.2008	0.1108	0.6567	0.4478	0.3549	0.1990	0.1234	0.0880
500	0.3476	0.1315	0.0647	0.5497	0.3413	0.2442	0.1109	0.0588	0.0357
1000	0.3099	0.1121	0.0553	0.5265	0.3011	0.2086	0.0907	0.0518	0.0325
2000	0.2823	0.0938	0.0406	0.4982	0.2722	0.1792	0.0620	0.0326	0.0184
3000	0.2661	0.0773	0.0338	0.4830	0.2544	0.1620	0.0584	0.0277	0.0175
4000	0.2585	0.0738	0.0278	0.4741	0.2490	0.1548	0.0455	0.0210	0.0106
5000	0.2450	0.0646	0.0265	0.4662	0.2362	0.1424	0.0419	0.0174	0.0087

Table 5: ERP of asymptotic tests at nominal level 5%, Theil measure (I_E^1).

Nobs	Singh-Maddala $c = 0.7$	Singh-Maddala $c = 1.2$	Singh-Maddala $c = 1.7$	Pareto $\alpha = 1.5$	Pareto $\alpha = 2$	Pareto $\alpha = 2.5$	Lognormal $\sigma = 1$	Lognormal $\sigma = 0.7$	Lognormal $\sigma = 0.5$
100	0.2717	0.1156	0.0484	0.3933	0.2421	0.1789	0.1040	0.0578	0.0343
500	0.2049	0.0718	0.0326	0.3440	0.1917	0.1268	0.0516	0.0239	0.0095
1000	0.1832	0.0610	0.0301	0.3279	0.1730	0.1128	0.0462	0.0233	0.0129
2000	0.1669	0.0488	0.0183	0.3139	0.1556	0.0993	0.0296	0.0130	0.0042
3000	0.1594	0.0402	0.0144	0.3108	0.1505	0.0886	0.0266	0.0087	0.0013
4000	0.1537	0.0383	0.0107	0.3070	0.1447	0.0847	0.0215	0.0023	-0.0012
5000	0.1478	0.0357	0.0118	0.3022	0.1378	0.0788	0.0166	0.0022	-0.0037

Table 6: ERP of bootstrap tests at nominal level 5%, Theil measure (I_E^1).