The Econometrics Journal

Econometrics Journal (2012), volume **15**, pp. 125–153.
doi: 10.1111/j.1368-423X.2011.00356.x

On the problem of inference for inequality measures for heavy-tailed distributions

CHRISTIAN SCHLUTER[†,‡]

[†]*DEFI, Université de la Méditerranée (Aix-Marseille II), 14 Avenue Jules Ferry, 13621 Aix-en-Provence, France.*

[‡]*University of Southampton, Southampton SO17 1BJ, UK.*
E-mail: C.Schluter@soton.ac.uk

First version received: February 2011; final version accepted: August 2011

Summary We consider the class of heavy-tailed income distributions and show that the shape of the income distribution has a strong effect on inference for inequality measures. In particular, we demonstrate how the severity of the inference problem responds to the exact nature of the right tail of the income distribution. It is shown that the density of the studentized inequality measure is heavily skewed to the left, and that the excessive coverage failures of the usual confidence intervals are associated with excessively low estimates of both the point measure and the variance. For further diagnostics, the coefficients of bias, skewness and kurtosis are derived and examined for both studentized and standardized inequality measures. These coefficients are also used to correct the size of confidence intervals. Exploiting the uncovered systematic relationship between the inequality estimate and its estimated variance, variance stabilizing transforms are proposed and shown to improve inference significantly.

Keywords: *Asymptotic expansions*, *Inequality measures*, *Inference*, *Statistical performance*, *Variance stabilization*.

1. INTRODUCTION

While first-order asymptotics for estimators of measures of inequality, such as Generalized Entropy (GE) indices, are well known, it is now also well known that this theory is a poor guide to actual behaviour in samples of even moderate size when the population (income) distribution exhibits a right tail that decays sufficiently slowly. Such distributions not only include the class of heavy-tailed distributions, whose tail decays like a power function, but also, for instance, the lognormal distribution, whose tail decays exponentially fast, provided the shape parameter is sufficiently large. For instance, Schluter and van Garderen (2009) have shown that the actual (finite sample) densities of the estimators are substantially skewed and far from normal. Standard one-sided and equi-tailed two-sided confidence intervals are too short, exhibiting coverage errors significantly larger than their nominal rates thus rendering inference unreliable. Davidson and Flachaire (2007) have shown that these problems persist for standard bootstrap inference.

Following the contributions of Schluter and Trede (2002), several authors have focused on the tail behaviour of the population income distribution. If the distribution is heavy-tailed,

samples are likely to contain 'extremes' or 'outliers', i.e. income realizations from the tail of the distribution which are substantially larger than income realizations associated with the main body of the distribution. An intuition, discussed in e.g. Cowell and Flachaire (2007) and Davidson and Flachaire (2007, p. 142), is to surmise that these extremes are the root cause of the inference problem since most inequality measures are not robust to such extremes (Cowell and Victoria-Feser, 1997). Alternatively, it might be that the sample drawn from a heavy-tailed distribution does not contain enough extreme drawings. We show that the coverage failures of standard confidence intervals are associated with estimates of the inequality measure and estimates of its variance which are both *too low* compared to their population values. This also holds for income distributions whose right tail decays faster than a power function, such as the lognormal provided its shape parameter is sufficiently large.

A principal contribution of the paper is the diagnosis of the underlying problem for inference, and we carefully show how the severity of the inference problem responds to the exact nature of the right tail of the income distribution. Denoting $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$ the standard estimators of the inequality measure and its variance, the problem is made visible via simulations in plots of realizations of $\widehat{\mathrm{var}}(\widehat{I})$ against $\widehat{I}$ and identifying those $(\widehat{I}, \widehat{\mathrm{var}}(\widehat{I}))$ pairs which are associated with a coverage failure of standard two-sided confidence intervals. Since the actual density of the studentized measure is shown to have a substantial *left* tail, this implies that the usual right confident limit is too often too small. Almost all coverage failures are on this side (despite the fact that the standard confidence intervals are two-sided and symmetric), and these wrong confidence limits, it turns out, are associated with particularly *low* realizations of both $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$. Exploiting the systematic relation between $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$, we propose variance stabilizing transforms. This constitutes the second principal contribution of the paper. We show that these succeed, in conjunction with a bootstrap, in reducing the inference problems significantly.

In order to understand better the separate and joint contributions of $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$ to the inference problem we also develop asymptotic expansions for both studentized (using the estimated variance) and standardized (using the theoretical variance) inequality measures. Building on the second-order expansions of Schluter and van Garderen (2009), we now derive third-order expansions. In particular, we derive the bias, skewness and kurtosis coefficients. These coefficients are used as diagnostic tools and also to correct the sizes of one-sided confidence intervals. Used as diagnostics, the cumulant coefficients enable us to quantify the departure from normality of the finite sample distributions, and to quantify the distortions caused by the variance estimate $\widehat{\mathrm{var}}(\widehat{I})$ by comparing the studentized and the standardized inequality measures. In our settings, the cumulant coefficients for the former are found to be substantially larger in magnitude than for the latter. Hence, while the density of the standardized inequality measure is close to normal and skewness is only modest, the studentized density exhibits significant skewness and a 'fat' left tail. This good performance of the standardized inequality measure contrasts starkly with the poor performance of the studentized measure. This confirms that the poor performance arises from the need to estimate the variance of the inequality measure, and it is the correlation of this variance estimator with the inequality estimator that plays an important role. Exploiting this relationship, we show that our variance stabilizing transform exhibits cumulant coefficients that are smaller in magnitude than those of the studentized inequality measure.

The plan of paper is as follows. The classes of inequality measures and income distributions considered in this paper are defined in Section 2. Section 3 presents the simulation evidence which shows that it is particularly low realizations of both $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$ which are associated with excessive coverage errors of the usual two-sided confidence intervals. We propose asymptotic

expansions for the cumulants of both standardized and studentized inequality measures as diagnostic tools to better understand the inference problem. These are considered abstractly and numerically in Section 4. In order to maintain readability, the precise statements and the derivation of the cumulant coefficients are collected in the Web Appendix to this paper, which also contains the remaining proofs. In Section 4.1, the cumulant coefficients are examined quantitatively. In Section 4.2, we use them to correct the size of one-sided confidence intervals. The availability of the cumulant coefficients also enables us to examine the performance of alternative distributional approximations. In particular, in Section 4.3 we consider saddlepoint approximations. The specific uncovered relationship between $\widehat{I}$ and the estimated variance $\widehat{\text{var}}(\widehat{I})$ suggests the application of variance stabilizing transforms. This is done in Section 5, and we present performance evidence that shows that these succeed in considerably lessening the inference problem.

2. GENERALIZED ENTROPY INDICES OF INEQUALITY

We consider the popular and leading class of inequality indices, the GE indices. These are of particular interest because it is the only class of inequality measures that simultaneously satisfies the key properties of anonymity and scale independence, the principles of transfer and decomposability, and the population principle. For an extensive discussion of the properties of the GE index see Cowell (2000). The class of indices is defined for any real α by

$$I(\alpha; F) = \begin{cases} \dfrac{1}{\alpha^2 - \alpha}\left[\dfrac{\mu_\alpha(F)}{\mu_1(F)^\alpha} - 1\right] & \text{for } \alpha \neq \{0, 1\} \\ -\displaystyle\int log\left(\frac{x}{\mu_1(F)}\right) dF(x) & \text{for } \alpha = 0 \\ \displaystyle\int \frac{x}{\mu_1(F)} log\left(\frac{x}{\mu_1(F)}\right) dF(x) & \text{for } \alpha = 1 \end{cases}, \tag{2.1}$$

where α is a sensitivity parameter, F is the income distribution and $\mu_\alpha(F) = \int x^\alpha dF(x)$ is the moment functional, and we will assume incomes to be positive. The index is continuous in α. The larger the parameter α, the larger is the sensitivity of the inequality index to the uppertail of the income distribution. It is not monotonic in α, however. Although the index is defined for any real value of α, in practice only values between 0 and 2 are used and we confine our examination to this range. The limiting cases 0 and 1 are treated implicitly below since all key quantities are continuous in α.

GE indices constitute a large class which nests some popular inequality measures as special cases. If $\alpha = 2$ the index is also known as the (Hirschman-)Herfindahl index and equals half the coefficient of variation squared. Herfindahl's index plays an important role as a measure of concentration in industrial organization and merger decisions. In empirical work on income distributions this value of α is considered large. Two other popular inequality measures are the so-called Theil indices, which are the limiting cases $\alpha = 0$ and $\alpha = 1$. Finally, the Atkinson index is ordinally equivalent to the GE index.

We follow the literature cited above and assume that incomes X are independent and identically distributed according to income distribution F, and that we have samples of size n at our disposal. I is usually estimated by $\widehat{I} = I(\widehat{F})$ where $\widehat{F}$ is the empirical distribution function,

hence the estimator replaces the population moments in (2.1) by the sample moments. We denote the sample analogue of $\mu_\alpha(F)$ by $m_\alpha = \mu_\alpha(\widehat{F})$. For a sample of size n define the studentized index

$$S_n = \sqrt{n}\left(\frac{\widehat{I} - I}{\widehat{\sigma}}\right), \tag{2.2}$$

where $\widehat{\sigma}$ is an estimate of the asymptotic standard deviation of $\sqrt{n}(\widehat{I} - I)$, derived by the delta method and given by $\sigma = [(\alpha^2 - \alpha)\mu_1^{\alpha+1}]^{-1}B_0{}^{1/2}$ with $B_0 = \alpha^2\mu_\alpha^2\mu_2 - 2\alpha\mu_1\mu_\alpha\mu_{\alpha+1} + \mu_1^2\mu_{2\alpha} - (1-\alpha)^2\mu_1^2\mu_\alpha^2$ for $\alpha \notin \{0, 1\}$. $\widehat{B}_0$ and thus $\widehat{\sigma}$ is obtained by replacing population moments with sample moments. In order to examine the role played by the estimated variance $\widehat{\sigma}^2$ we also consider the standardized inequality measure

$$\widetilde{S}_n = \sqrt{n}\left(\frac{\widehat{I} - I}{\sigma}\right). \tag{2.3}$$

We will distinguish standardized quantities from their studentized counterparts throughout by tildes. Simplifying a little we have thus $S_n = \sqrt{n}\widehat{B}_0^{-1/2}[m_\alpha m_1 - \mu_1^{-\alpha}\mu_\alpha m_1^{\alpha+1}]$ and $\widetilde{S}_n = \sqrt{n}B_0^{-1/2}[\mu_1^{\alpha+1}m_\alpha m_1^{-\alpha} - \mu_1\mu_\alpha]$.

By standard central limit arguments, S_n has a distribution that converges asymptotically to the standard Normal distribution (see e.g. Cowell, 1989), and by the arguments of Section 4 later

$$\Pr(S_n \le x) = \Phi(x) + O(n^{-1/2})$$

where Φ denotes the Gaussian distribution.

2.1. Heavy-tailed income distributions

We follow the previous literature cited above and consider the leading parametric income distributions which are regularly used to fit real-world income data. Specifically, we consider first the heavy-tailed Singh-Maddala distribution $SM(a, b, c)$ with density $f(x; a, b, c) = abcx^{b-1}/(1 + ax^b)^{c+1}$. Schluter and Trede (2002) have shown that the right tail of SM is of the form $L_0(x)x^{-bc}$ where $L_0(x)$ is a slowly varying function. Hence the tail decays like a power function so the distribution is heavy-tailed, and the right tail index equals bc. Since this right tail is close to Paretian, it is of interest to consider also directly the heavy-tailed Pareto distribution with parameter and tail index equal to λ. In empirical settings, the Pareto distribution is often used to fit wealth distributions. Finally, we consider the lognormal $LN(m, sd)$ distribution. Although the tail of LN decays exponentially fast, we will show that for large sd the inequality measure suffers the same problems as in the heavy-tailed cases.

GE indices are scale invariant, and thus independent of the parameters m and a for the LN and SM distributions, respectively. For notational convenience, we suppress these irrelevant parameters later. Since I is scale invariant, so is σ and thus S_n. The population values are in the lognormal case $I(\alpha; sd) = (\alpha^2 - \alpha)^{-1} \times [\exp(0.5(\alpha^2 - \alpha)(sd)^2) - 1]$, in the Singh-Maddala case, defined only for $bc > \alpha$, $I(\alpha; b, c) = (\alpha^2 - \alpha)^{-1}c^{-(\alpha-1)}B(1 + \alpha/b, c - \alpha/b)/[B(1 + 1/b, c - 1/b)^\alpha - 1]$ where $B(\cdot, \cdot)$ denotes the Beta function, and in the Pareto case $I(\alpha; \lambda) = (\alpha^2 - \alpha)^{-1}[(\lambda - 1)^\alpha(\lambda - \alpha)^{-1}\lambda^{-\lambda+1} - 1]$. The asymptotic variance σ^2 of the inequality measure is always finite in the LN case, but in the SM and the Pareto case we require that bc and λ exceed max $\{2, 1 + \alpha, 2\alpha\}$. Note that, for given α, I increases as the tail of the income distribution becomes heavier (as sd increases in the LM case, or λ decreases in the Pareto case, or b decreases for fixed

c in the SM case). We therefore depict many results below as functions of I, in order to facilitate comparisons across income distributions, and to show how the severity of the inference problem responds to the exact nature of the right tail of the income distribution.

3. SIMULATION EVIDENCE: THE ROLE OF $\widehat{I}$, $\widehat{\mathrm{VAR}}(\widehat{I})$, AND THE TAIL BEHAVIOUR OF F

In order to fix ideas and illustrate the main insights about the inferential problem, we consider the Theil indices and $I(2)$, and samples of size $n = 500$. We parameterize the lognormal case with $sd \in \{0.3, 0.7\}$ and the SM case with $b = 2.8$ and $c = 1.7$.[1,2] Extensive results of several experiments (varying sample size n, α, and the extent of tail heaviness) are reported in Section 5.2 later. The simulation exercises of this section, based on 10,000 repetitions, are meant to be illustrative, and not exhaustive. Complementary simulation evidence is provided in Section 5 later, and in Davidson and Flachaire (2007) and Schluter and van Garderen (2009). Our qualitative conclusions also hold for these other settings.[3]

The first set of experiments simply consists in estimating, using standard kernel density estimators, the actual densities of the studentized inequality measure S_{500} and of the standardized inequality measure $\widetilde{S}_{500}$, focusing on the skewness of the densities. The juxtaposition of S_{500} and $\widetilde{S}_{500}$ is a first illustration of the distributional impact of having to estimate σ^2.

Figure 1 depicts the results ($\alpha = 2$ (solid line), $\alpha = 1.05$ (broken line), $\alpha = 0.05$ (dotted line)) The kernel density estimates for S_{500} in the *SM* and the Pareto case clearly reveal the substantial skewness the density of the studentized measure suffers when incomes are generated by a heavy-tailed distribution. The problem increases as the sensitivity parameter α of the inequality measure increases. The problem is not, however, exclusively associated with tails which decay like power function. While the density estimates for S_{500} in the lognormal case look fairly standard Normal when the shape parameter is 0.3, increasing the shape parameter to 0.7 induces again substantial skewness. As a shorthand, we will refer to these two cases as income distributions which exhibit 'sufficiently slow tail decay'.

By contrast, the density estimates for the standardized inequality measure $\widetilde{S}_{500}$ do appear very symmetric. However, the densities also exhibit a greater concentration around 0 than the standard Normal density when the tails of the income distribution decay sufficiently slowly and the sensitivity parameter α equals 2. For the lower values of α the densities appear close to the standard Normal.

These estimated densities have several important implications for inference when incomes are drawn from distributions with sufficiently slow tail decay. The non-Gaussian shape of the density of S_{500} suggests that standard inference is likely to be unreliable. The substantial left tail of the densities indicates that there are too many realizations $\widehat{I}$ which are *too small*. In conjunction with the steep increase of the densities at the depicted right tail, coverage errors of standard symmetric two-sided confidence intervals are likely to be one-sided. A comparison of the densities of S_{500} and $\widetilde{S}_{500}$ suggests that the distributional problem arises from the need to estimate σ^2. It is not the non-linearity of the inequality measure I which induces the non-Gaussian shape of the density of S_{500}, but the systematic relation between $\widehat{I}$ and $\widehat{\mathrm{var}}(\widehat{I})$ on which we focus on next.

[1]To be precise, we set the sensitivity parameter α of the inequality index equal to 0.05 and 1.05, respectively.
[2]$SM(., 2.8, 1.7)$ and $LN(., 0.3)$ are good fitting parameterizations of the income distribution in Germany.
[3]Results for different parameterizations, sample sizes and different αs are available on request.

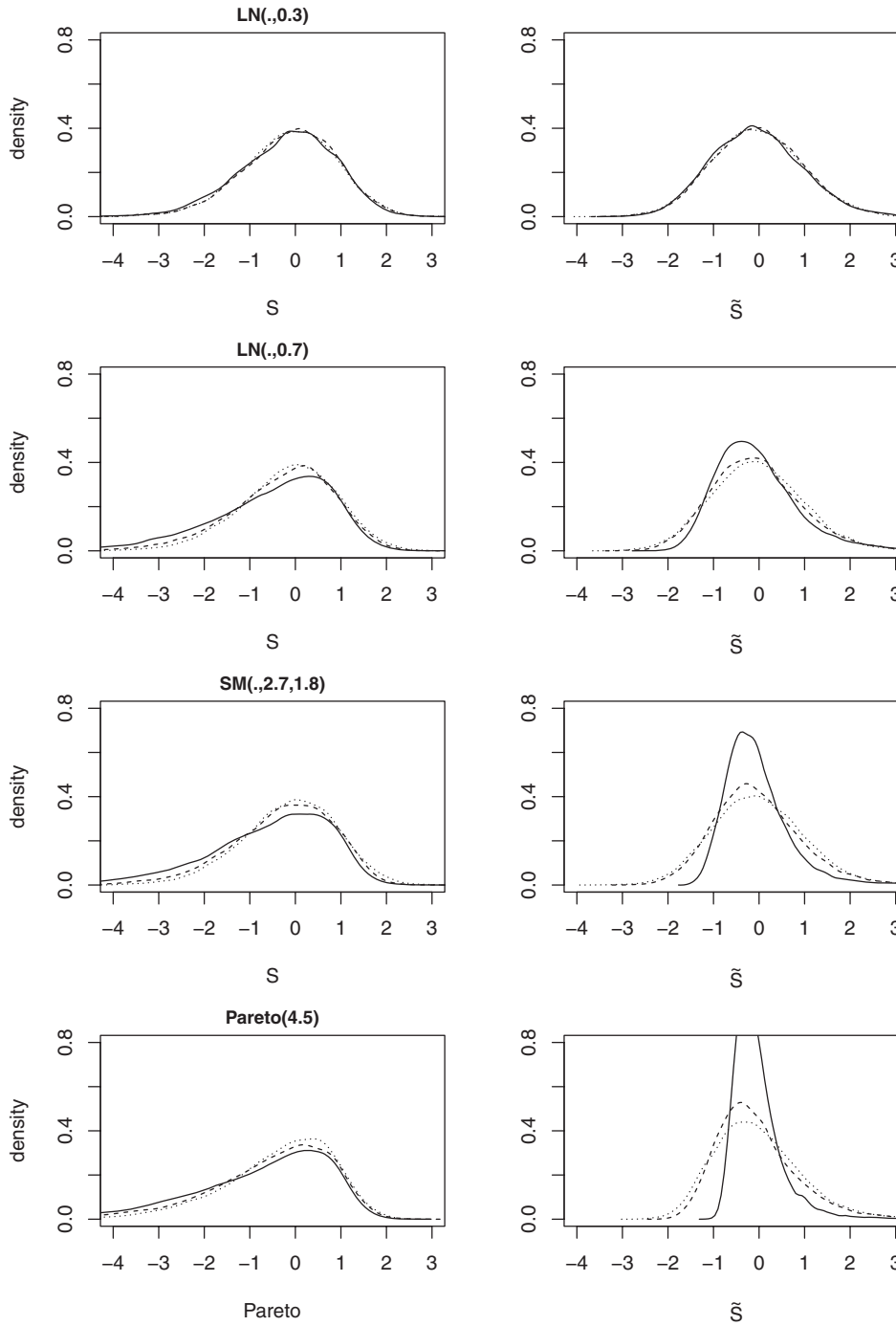

Figure 1. Density estimates for S_{500} and $\widetilde{S}_{500}$.

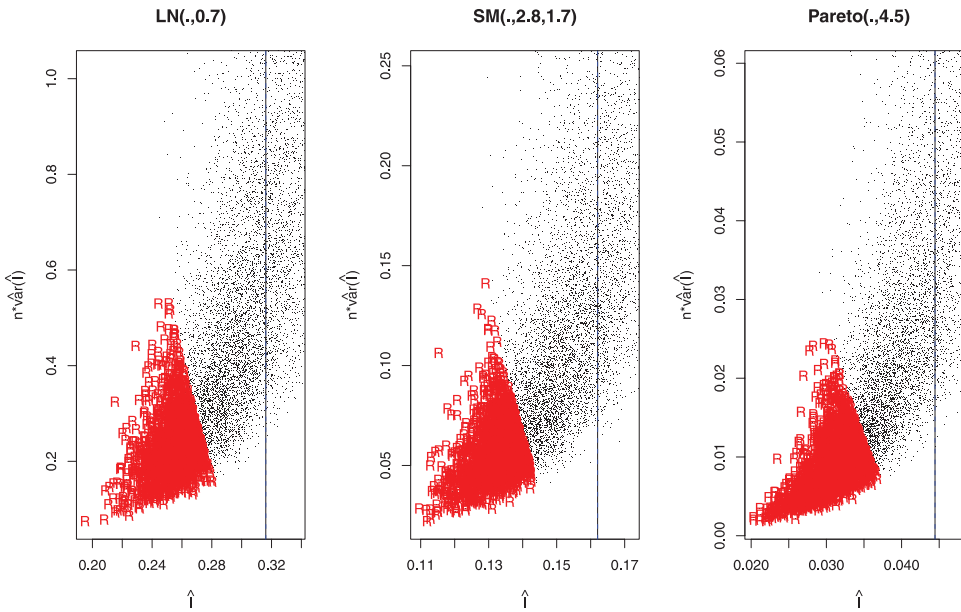

Figure 2. Coverage errors in $\widehat{I}$ versus $n \times \widehat{\text{var}}(\widehat{I})$ plots.

We turn to the induced inferential problems by considering the actual coverage errors of standard 95% two-sided symmetric confidence intervals for $\alpha = 2$. Extensive results of several experiments (varying sample size n, α and the extent of tail heaviness) are reported in Section 5.2 later. Compared to a nominal coverage error rate of 5%, the actual total coverage error rate in the lognormal case $LN(., 0.7)$ is 14.3%, but almost all rejections (13.8 percentage points) are rejections on the right (i.e. the population value I exceeds the right confidence limit). In the *SM* case the total rate is 15.5%, and 15.2 percentage points are rejections on the right. In the Pareto case, the total and right rejection rates are 21.1%. This is the flip-side of the substantial *left* tail and the heavy skewness of the density of S_{500}. The importance of the distortion caused by $\widehat{\text{var}}(\widehat{I})$ can also be assessed in terms of the actual coverage errors for $\widetilde{S}_{500}$ compared to those for S_{500}. For the particular LN, SM and Pareto distributions we have actual total coverage errors of 3.37%, 1.97% and 1.47%, respectively. Hence the impact of the distortion is substantial, as these are far closer to the nominal rate. These lower than nominal coverage error rates are consistent with the observed greater concentration, relative the Gaussian density, of the densities of $\widetilde{S}_{500}$ for $\alpha = 2$ depicted in Figure 1.

Next, the interplay between $\widehat{I}$, $\widehat{\text{var}}(\widehat{I})$, and the coverage errors is examined by simply plotting in Figure 2 (The vertical line corresponds to the population value of I, pairs labelled R correspond to coverage errors on the right of standard 95% two-sided symmetric confidence intervals) the $(\widehat{I}, \widehat{\text{var}}(\widehat{I}))$ pairs, and by identifying those pairs which are associated with a coverage error. Given that almost all coverage errors are right rejections, we restrict the depicted range of $\widehat{I}$, and re-label those $(\widehat{I}, n \times \widehat{\text{var}}(\widehat{I}))$ pairs associated with such a coverage error to the right by R. The population value of I is indicated by the vertical line, the population value of $n \times \text{var}(\widehat{I})$ exceeds the depicted range. It is evident that the wrong confidence limits are associated with particularly *low* realizations of both $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$. The intuition underlying these results is that

as tail heaviness increases, the population moments increase and eventually cease to exist for heavy-tailed distributions, whilst the (finite) sample moments tend to underestimate these. In their investigation Davidson and Flachaire (2007) conclude that the persistent inference problem is due to the poor tail estimation of the underlying income distribution.

The juxtaposition of the densities of studentized and standardized inequality measure suggests that the problem is the non-linearity of S_n, and in particular the systematic relationship between $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$. This relationship is exploited in Section 5, where variance stabilizing transforms are proposed and performance evidence for these is examined. Before turning to these, however, we proceed to examine the issues of bias, skewness and kurtosis formally using asymptotic expansions.

4. ASYMPTOTIC EXPANSIONS

Asymptotic expansions of the cumulants of S_n provide measures for the departures of the distribution of S_n from the Gaussian limit. These will be used later as diagnostic tools for the anatomy of the above inference problems, to correct the size of confidence intervals, and to assess possible remedies.

Expanding the first four cumulants of S_n in powers of $n^{-1/2}$ yields

$$\begin{aligned} K_1 &= n^{-1/2}k_{1,2} + O(n^{-3/2}) \\ K_2 &= 1 + n^{-1}k_{2,2} + O(n^{-2}) \\ K_3 &= n^{-1/2}k_{3,1} + O(n^{-3/2}) \\ K_4 &= n^{-1}k_{4,1} + O(n^{-2}). \end{aligned} \tag{4.1}$$

Since S_n is studentized, the coefficient $k_{1,2}$ is the bias coefficient, $k_{3,1}$ is the coefficient of skewness, and $k_{4,1}$ is the kurtosis coefficient.[4,5] In terms of the cumulant generating function of S_n, given by $\exp(K_{S_n}(s)) = E\{\exp(sS_n)\}$), the cumulant coefficients define the second- and third-order term in the approximation to K_{S_n}, i.e. we have

$$K_{S_n}(s) = \frac{1}{2}s^2 + n^{-1/2}\left(sk_{1,2} + \frac{1}{6}s^3k_{3,1}\right) + n^{-1}\left(\frac{1}{2}s^2k_{2,2} + \frac{1}{24}s^4k_{4,1}\right) + O(n^{-3/2}). \tag{4.2}$$

In the exact Gaussian case, all these coefficients are zero.

It is an important contribution of this paper to derive explicitly these cumulant coefficients for both studentized and standardized inequality measures. In order to maintain readability of the exposition, these cumulant coefficients are stated explicitly in Web Appendix for this paper, since the resulting expressions are lengthy and involve expectations of products of certain mean-zero random variables.

[4]The cumulant of order r exists if the all moments of S_n up to order r exist.

[5]Expanding cumulant i, which is of order $n^{-(i-2)/2}$, as a power series in n^{-1} yields $K_i = n^{-(i-2)/2}(k_{i,1} + n^{-1}k_{i,2} + \cdots)$ with $k_{1,1} = 0$ and $k_{2,1} = 1$ because of centering and the studentization. Hence $k_{i,j}$ refers to the coefficient of $n^{-(j-1)}$ in this power series in the expansion for the i's cumulant.

These coefficients are also the key quantities in the Edgeworth expansion of the CDF of S_n given by

$$\Pr\{S_n \leq x\} = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + O(n^{-3/2}) \tag{4.3}$$

with

$$p_1(x) = -\left(k_{1,2} + \frac{1}{6}k_{3,1}(x^2 - 1)\right)$$
$$p_2(x) = -x\left(\frac{1}{2}\left(k_{2,2} + k_{1,2}^2\right) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2 - 3) + \frac{1}{72}k_{3,1}^2(x^4 - 10x^2 + 15)\right).$$

The right-hand side of equation (4.3) is to be interpreted as an asymptotic expansion since it does not necessarily converge as an infinite series. See e.g. Hall (1992) for an extensive discussion of Edgeworth expansions, and in his Section 2.4 a statement of the regularity conditions for the validity of the expansion; chapter 3 in Hall (1992) justifies the bootstrap given the Edgeworth expansions.[6] The GE index is a smooth function of the moments with continuous third derivatives and $\mu_1 > 0$ since we assume incomes be positive. This implies that Hall's Theorem 2.2 applies and we require that (a) X has a proper density function (implying that Cramér's condition is satisfied), and with $\alpha^* = \max\{2, \alpha + 1, 2\alpha\}$ that (b) $\mu_{3\alpha^*} < \infty$ for the first-order expansion which includes the $O(n^{-1/2})$ term and $\mu_{4\alpha^*} < \infty$ for the second-order expansion which includes the $O(n^{-1})$ term. If $\mu_{4\alpha^*} < \infty$ then the regularity condition of footnote 4 with $r = 4$ is satisfied. These moment conditions are satisfied in the case of the lognormal distribution, in case of the *SM* and the Pareto distribution bc and λ must exceed $(j + 2)\alpha^*$ for the Edgeworth expansion of order j. For the standardized inequality measure $\widetilde{S}_n$, α^* appearing in the regularity conditions is replaced by $\max\{1, \alpha\}$.

4.1. Diagnostics

We proceed to use the cumulant coefficients to examine how increased tail-heaviness of the income distribution induces more severe departures from normality.

Consider first $\alpha = 2$, the studentized measured S_n, and the lognormal case as the shape parameter sd of the income distribution, and thus $I(2)$, increases. Table 1 reports the results, and Figure 3 ($\alpha = 2$ (solid line), $\alpha = 1.05$ (broken line) and $\alpha = 0.05$ (dotted line)) depicts the cumulant coefficients as functions of I. It is evident that the magnitudes of the cumulant coefficients not only increase sharply, but also that these become large relative to their $n^{-1/2}$ and n^{-1} coefficients (for instance, for $n = 500$, $n^{1/2} = 22.3$ and $k_{3,1} = -62$ for $\alpha = 2$ and $sd = 0.6$; note too that σ is substantially larger than I). These problems are less pronounced, but still not negligible, for the smaller values of α.

In the case of the heavy-tailed SM and Pareto distributions, the explosions of the cumulant coefficients are even more pronounced, as depicted in Figure 4 (Values same as Figure 3) as functions of I.[7] The tails of these income distributions become more heavy as the tail indices (bc and λ) decrease; fixing c, I decreases in b and also in λ. Recall that for the Edgeworth expansion of order j for the distribution of S_n to exist that $\mu_{(j+2)\alpha^*}$ be finite. Hence, for sufficiently heavy tails, the cumulants will cease to exist.

[6]Biewen (2002) has explicitly justified the bootstrap for inequality measures.
[7]The figure for $k_{4,1}$ is less insightful because the coefficient exhibits non-monotonicity, and is therefore not depicted.

Table 1. Cumulant coefficients when $X \sim LN(., sd)$.

				S_n				$\widetilde{S}_n$			
α	sd	I	σ	$k_{1,2}$	$k_{3,1}$	$k_{2,2}$	$k_{4,1}$	$\widetilde{k}_{1,2}$	$\widetilde{k}_{3,1}$	$\widetilde{k}_{2,2}$	$\widetilde{k}_{4,1}$
2	0.1	0.005	0.007	−2.30	−6.37	30.37	95.19	−0.71	3.13	−0.67	16.47
2	0.2	0.020	0.031	−2.87	−8.67	52.76	161.41	−0.74	4.13	−1.26	35.50
2	0.3	0.047	0.080	−3.98	−13.14	111.20	331.99	−0.77	6.13	−2.46	95.12
2	0.4	0.087	0.168	−5.96	−21.08	263.86	743.25	−0.82	9.79	−4.70	290.80
2	0.5	0.14	0.33	−9.51	−35.29	694.11	1666.81	−0.87	16.54	−8.82	1013.09
2	0.6	0.22	0.61	−16.22	−62.09	2053.56	3080.70	−0.94	29.57	−16.51	4174.65
0.05	0.3	0.045	0.065	−2.32	−6.44	31.48	97.48	−0.73	3.11	−0.75	16.45
1.05	0.3	0.045	0.068	−2.86	−8.59	53.05	160.13	−0.76	4.01	−1.36	33.59

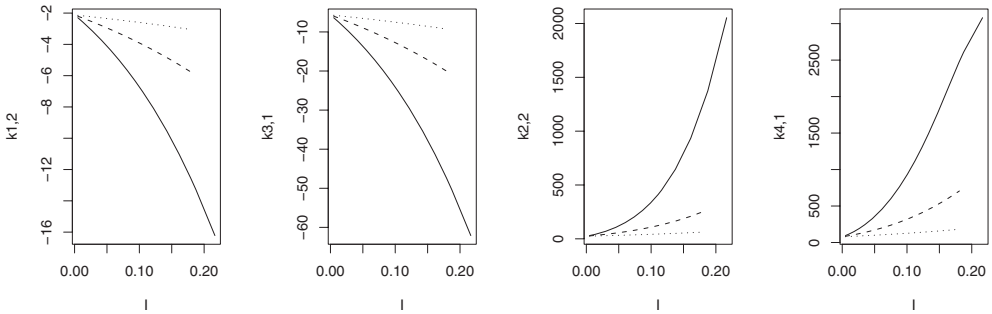

Figure 3. Cumulant coefficients of S_n as functions of I for the lognormal distribution.

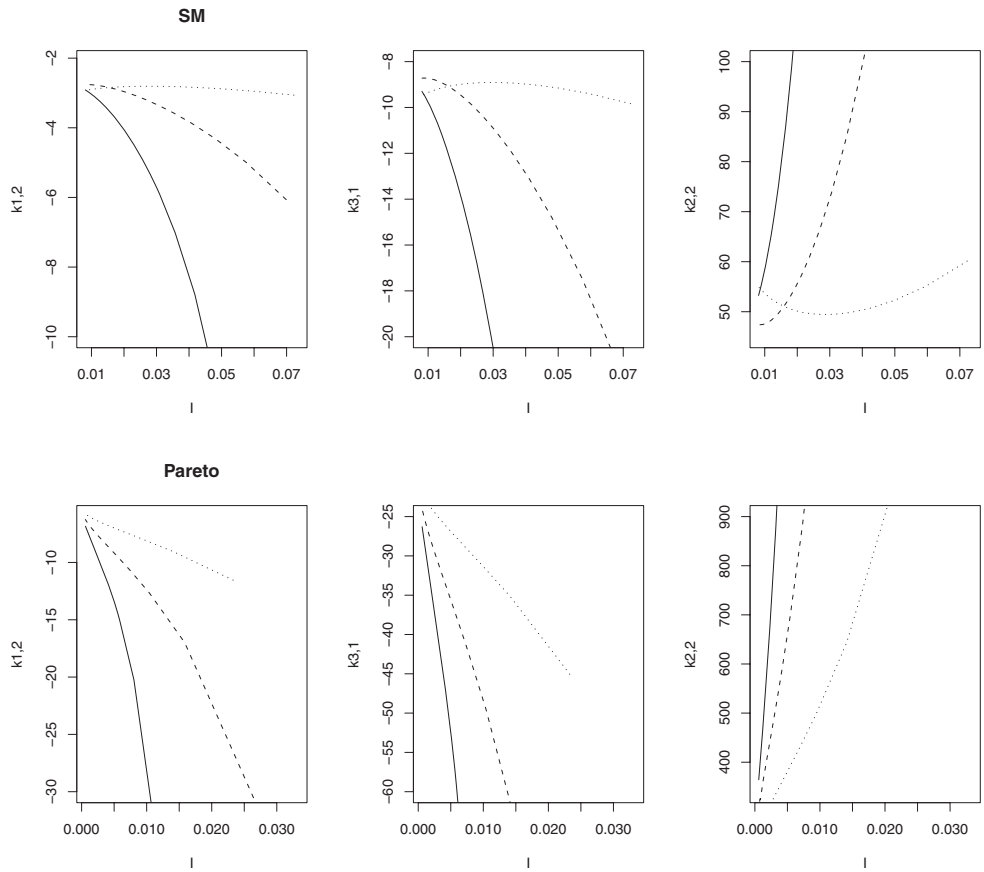

Figure 4. Cumulant coefficients as functions of I for $SM(., b, 1.7)$ and the Pareto distribution.

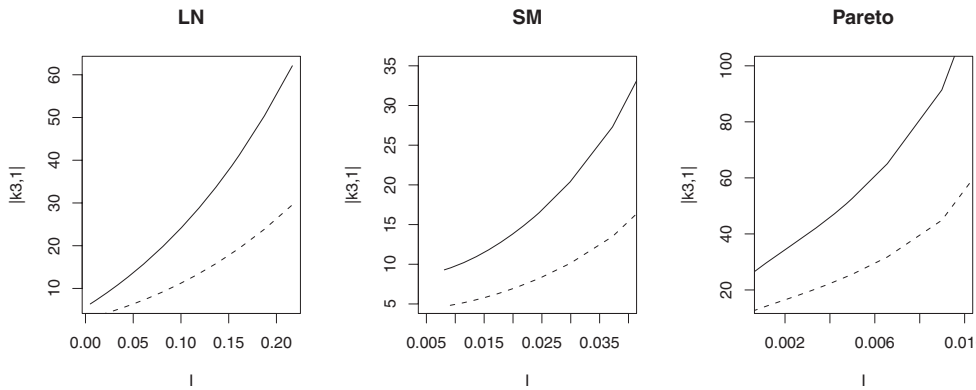

Figure 5. Cumulant coefficients $|k_{3,1}|$ (solid line) and $\widetilde{k}_{3,1}$ (dashed line) as functions of $I(2)$.

4.1.1. The Distortions caused by $\widehat{\sigma}$. The distortions caused by $\widehat{\sigma}$ can also be quantified in terms of the associated cumulant coefficients. Table 1 focuses on the *LN* case, and reports the cumulant coefficients for S_n and $\widetilde{S}_n$. It is evident that the first three cumulant coefficients have substantially smaller magnitudes in the latter case (consistent with Figure 1, the skewness coefficient $\widetilde{k}_{3,1}$ has now the opposite sign). In Figure 5 we focus on the case $I(2)$ and compare the magnitudes of the skewness coefficients $k_{3,1}$ and $\widetilde{k}_{3,1}$, while the tails of the income distribution become heavier. We conclude that the resulting distortions are substantial across all income distributions.

4.2. Size corrections

Given the availability of the cumulant coefficients, it is now possible to correct the size of standard one-sided confidence intervals.[8] Let w_β denote the β-level quantile of S_n and z_β the β-level Gaussian quantile given by $\Phi(z_\beta) = \beta$. Then inverting the Edgeworth expansion (4.3) yields the Cornish-Fisher expansion of w_β in terms of z_β,

$$w_\beta = z_\beta - n^{-1/2} p_1(z_\beta) + n^{-1}\left[p_1\left(z_\beta\right) p_1'(z_\beta) - \frac{1}{2} z_\beta p_1(z_\beta)^2 - p_2(z_\beta)\right] + O_p(n^{-3/2}). \tag{4.4}$$

Hence, since $\Pr\{S_n \le w_\beta\}$ equals β to the stated order, using w_β instead of z_β yields a size correction of the usual one-sided confidence intervals.

Figure 6 ($\alpha = 2$ (solid line), $\alpha = 1.05$ (broken line) and $\alpha = 0.05$ (dotted line). The horizontal line pertains to the Gaussian quantile $z_{0.05} = -1.65$) depicts the size correcting quantiles w_β based on the second term correction for size $\beta = 0.05$, $n = 500$, and the three inequality measures. Given the substantial skewness of the densities of S_n seen in Figure 1, it is clear that these w_β will be substantially larger in magnitude than the Gaussian quantile z_β. Figure 6 quantifies the extent of this and reveals the increase as the tails of the income distributions decay more slowly. Finally we note that convergence in n of w_β

[8]Rothenberg (1988) has considered this in a regression context. I am grateful to a referee for suggesting this.

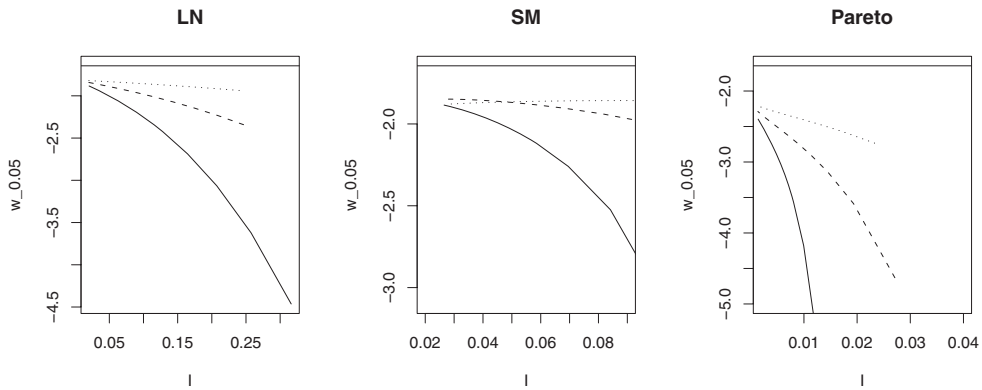

Figure 6. $w_{0.05}$ vs. I.

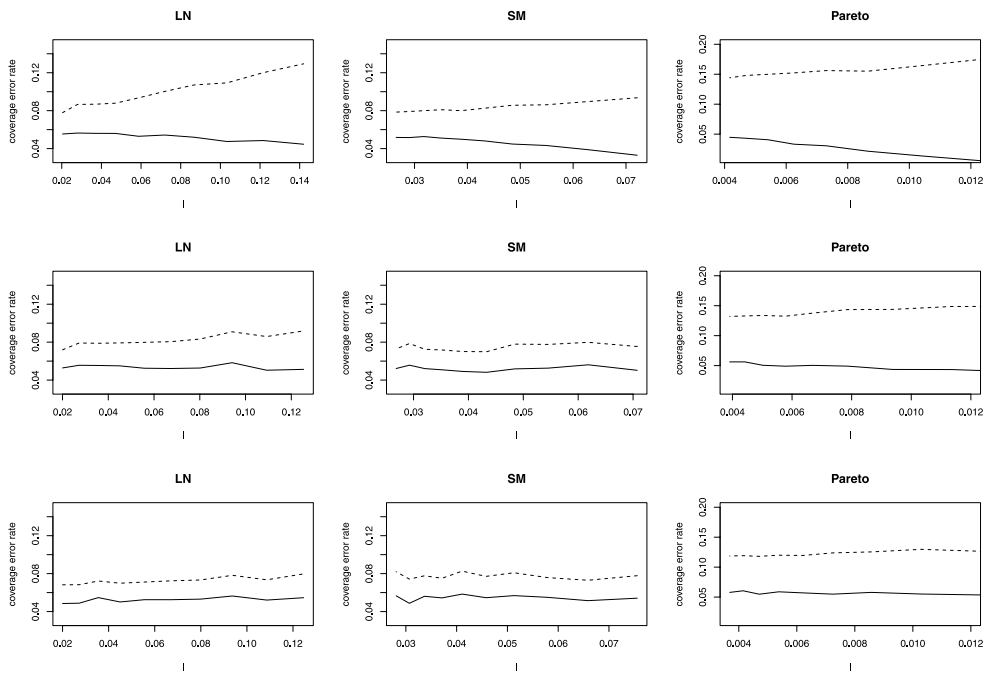

Figure 7. Coverage error rates of nominal size 0.05 one-sided confidence interval.

to z_β is fairly slow. For instance, in the LN case with $sd = 0.5$, w_β has values -2.9, -2.5 and -2.3 for sample size 250, 500 and 1000, respectively.

4.2.1. Performance Evidence. Performance evidence for the size correction as a function of I (and thus of tail heaviness) is reported in Figure 7, (solid line refers to the size corrected CI based on $w_{0.05}$ given by equation (4.4), the dotted line refers to the standard CI based on the Gaussian quantile $z_{0.05} = -1.65$, based on samples of size $n = 500$, and

$R = 10{,}000$ repetitions), for $I(2)$ (row 1), $I(1.05)$ (row 2) and $I(0.05)$ (row 3). The standard confidence intervals based on the Gaussian quantile perform poorly across all settings. The size correction does very well for α equal to 0.05 and 1.05. For $\alpha = 2$ it does also well for moderate values of the distributional parameter, but when the income distribution tail becomes sufficiently heavy, the Cornish Fisher approximation starts to over-correct the size. However, for inference, this over-correction is far less problematic than the excessively short Gaussian confidence interval.

4.3. Saddlepoint approximations

Edgeworth expansions are well known to suffer some shortcomings which limit their usefulness in some practical applications: The density expansion is not guaranteed to be positive, and oscillations can sometimes be observed in the tails. It turns out that these observations also apply in our inequality measurement setting when the income distribution exhibits sufficiently heavy tails. The problems in the tails are disturbing for inference, since it is precisely the tail areas that are typically of interest for inference. By contrast, the expansion is usually good around the mean, in which case it is easily seen that the accuracy of the pdf expansion improves to $O(n^{-1})$. This suggests to Escher-tilt the Edgeworth expansion of the density, which leads to the saddlepoint approximation (Daniels, 1954, see e.g. Reid, 1988, for a survey). The new approximation is guaranteed to be positive and exhibits improved tail behaviour since the approximation error turns out now to be relative rather than absolute.

Recall the cumulant generating function K_S of S_n, let $K(t) = nK_S(tn^{-1})$, and denote its first and second derivatives by K' and K''. The saddlepoint approximation to the density of S_n at x is

$$g(x) = c(2\pi K''(s))^{-1/2} \exp(K(s) - sx), \tag{4.5}$$

where the saddlepoint s satisfies the saddlepoint equation $K'(s) = x$. The saddlepoint approximation is rescaled to integrate to 1 which determines the constant c. The approximation to the distribution function of S is

$$G(x) = \Phi\left(w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right), \tag{4.6}$$

with $w = sign(s)[2(sx - K(s))]^{1/2}$ and $v = s[K''(s)]^{1/2}$. If we denote the pdf of S_n by pdf, then $pdf(x)/g(x) = 1 + O(n^{-1})$, so the approximation error is relative rather than absolute (the case of Edgeworth expansions).

The cumulant generating function of S_n is not known in practice. We therefore approximate $K_S(s)$, following Easton and Ronchetti (1986), to order $n^{-3/2}$ by using the approximation (4.2). This leads to an approximation to the saddlepoint approximation which is of the same order. The approximate solution to the saddlepoint equation $K'(s) = x$ is guaranteed to be unique since the approximation to K' is a cubic in s.

4.3.1. Performance Evidence. Performance evidence for the saddlepoint approximations in the lognormal case is reported in Table 2, as both the sensitivity parameter α of the inequality index and the shape parameter sd and thus tail heaviness increases. All approximations are evaluated at the quantiles determined by the 'exact' (i.e. simulated) CDF of S_{500}.

The tail accuracy of the normal approximation is poor, and decreases as α increases for fixed sd and as sd increases for fixed α. By contrast, the saddlepoint approximation does well for the

Table 2. Performance evidence for saddlepoint approximations in the $LN(., sd)$ case.

	$\alpha = 2$						$sd = 0.3$					
	$sd = 0.1$		$sd = 0.2$		$sd = 0.4$		$\alpha = 0.05$		$\alpha = 1.05$		$\alpha = 2$	
CDF	norm	saddle	norm	saddle	norm	saddle	norm	saddle	norm	saddle	norm	saddle
0.01	0.00	0.03	0.00	0.03	0.00	0.04	0.00	0.02	0.00	0.02	0.00	0.02
0.10	0.02	0.13	0.01	0.12	0.00	0.20	0.01	0.12	0.01	0.17	0.00	0.21
1	0.31	0.95	0.20	0.92	0.06	1.73	0.27	0.87	0.29	1.18	0.14	1.24
2.50	1.35	2.71	0.96	2.58	0.43	4.12	1.11	2.37	1.02	2.70	0.77	3.21
5	3.16	5.14	2.88	5.51	1.44	7.21	3.18	5.19	2.49	4.97	2.24	6.03
10	7.44	10.07	7.31	10.84	4.65	12.95	7.65	10.34	6.61	10.06	6.02	11.23
25	22.44	25.32	21.90	25.69	19.00	28.61	22.90	25.80	21.10	24.92	20.80	26.50
50	48.37	50.69	46.90	49.76	45.60	51.95	48.20	50.55	46.40	49.23	46.50	50.41
75	72.26	74.50	72.20	74.83	72.10	75.76	73.10	75.32	71.70	74.29	71.90	75.18
90	87.64	89.73	87.90	90.32	86.40	88.61	87.60	89.70	86.90	89.39	86.60	89.49
95	93.04	94.74	93.00	94.98	91.80	93.11	93.00	94.73	92.30	94.34	92.10	94.27
97.50	96.04	97.29	96.10	97.50	95.00	95.71	96.10	97.33	95.10	96.70	95.20	96.68
99	97.99	98.78	97.90	98.77	97.20	97.43	98.20	98.94	97.80	98.71	97.60	98.43
99.90	99.67	99.85	99.60	99.83	99.40	99.24	99.60	99.82	99.30	99.65	99.60	99.72
99.99	99.84	99.94	99.90	99.94	99.80	99.68	100.00	99.99	100.00	99.99	99.90	99.90

Note: CDF is the 'exact' CDF based on 10,000 replications of S_{500}, all CDFs x 100, all approximations are evaluated at the quantiles determined by the exact CDF, normal is the normal CDF, 'saddle' is the approximation to the saddlepoint approximation given by (4.6).

moderate values of *sd* considered. For instance, in the case of $\alpha = 2$ and $sd = 0.7$ when the exact CDF evaluates to 0.025, the normal approximation is 0.008 while the saddlepoint approximation is 0.03, and turning to the 97.5% quantile, the normal approximation evaluates to 0.95 while the saddlepoint approximation is 0.97. However, the performance of all approximations deteriorate as the tail of the income distribution becomes heavier.

Similar qualitative and quantitative conclusions follow for the *SM* distribution and the Pareto distribution. For instance, the *LN*(, 0.3) and the *SM*(., 4.2, 4) distributions yield similar values for I and σ^2, as well as for bias and skewness coefficients. Table 3 reports the results for this case. (Details for the Pareto distribution are not reported for the sake of brevity). To summarize, the saddlepoint approximation performs well for moderate parameter values but its performance deteriorates markedly when the speed of tail decay of the income distribution becomes slower. Rather than to seek improved approximations to the actual distribution of the measure, the next section shows that it is preferable to work directly with suitably transformed inequality measures.

5. VARIANCE STABILIZING TRANSFORMS

An alternative approach to improving inference is to consider non-linear transformations of the inequality measure, which are designed to address directly the root cause of the inference problem. Specifically, our results of the previous sections suggest that an important role is played by the estimated variance of the inequality measure and the systematic relationship between $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$. This suggests the application of a variance stabilizing transform.[9]

Figure 2 suggests that for the considered income distributions, the relation could be approximately exponential, so that $\log(\widehat{\text{var}}(\widehat{I}))$ is approximately linear in $\widehat{I}$. This conjecture is confirmed in Figure 10, Column 1, which plots $\log(\widehat{\text{var}}(\widehat{I}))$ vs. $\widehat{I}$ for several income distributions, and further depicts a non-parametric estimate based on smooth splines, which is approximately linear. This approximate linearity can be shown explicitly for the heavy-tailed Pareto distribution $P(\lambda)$ and $\alpha = 2$ as follows. We have $I \equiv I(2) = \frac{1}{2}\frac{1}{\lambda(\lambda-2)}$ with variance $\sigma^2 = \frac{2}{\lambda^3}\frac{(\lambda-1)^4}{(\lambda-2)^3(\lambda^2-7\lambda+12)}$. Inverting I, then substituting out λ in σ^2 and taking logs yields

$$v(I) = \ln\left(\frac{(2I+1)^2}{2I^2}\left[\frac{7}{8I^3} + \frac{1}{16I^4} - \frac{5}{8I^3}\sqrt{\frac{1}{2I}+1}\right]^{-1}\right). \tag{5.1}$$

Then $v(I)$ is approximately linear in the relevant range as depicted in Figure 8 (The solid curve depicts $v(I)$, the straight dashed line connects the two endpoints).

The variance stabilizing transform is of the form

$$h(I) = \int_0^I \frac{du}{[\sigma^2(u)]^{1/2}}, \tag{5.2}$$

where $\sigma^2(I)$ denotes the variance as a function of I. By the delta method $var(h(I)) = 1$ asymptotically. By the above reasoning, $\log(\widehat{\text{var}}(\widehat{I})) = \gamma_1 + \gamma_2\widehat{I} + error$, so the specific

[9]An alternative normalizing transformation is considered in Schluter and van Garderen (2009), which is designed to annihilate asymptotically the skewness coefficient of the transform. This transform, however, does not exploit the systematic relation between $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$.

Table 3. Performance evidence for saddlepoint approximations for $SM(., b, 4)$.

	$\alpha = 2$						$b = 4.2$					
	$b = 4.6$		$b = 4.4$		$b = 4$		$\alpha = .05$		$\alpha = 1.05$		$\alpha = 2$	
CDF	norm	saddle	norm	saddle	norm	saddle	norm	saddle	norm	saddle	norm	saddle
0.01	0.00	0.04	0.00	0.00	0.00	0.04	0.00	0.02	0.00	0.01	0.00	0.02
0.10	0.02	0.16	0.01	0.11	0.01	0.13	0.01	0.14	0.01	0.09	0.01	0.12
1	0.33	1.00	0.38	1.12	0.34	1.11	0.25	1.02	0.35	0.97	0.34	1.06
2.50	1.21	2.53	1.32	2.74	1.24	2.73	0.99	2.52	1.16	2.34	1.29	2.74
5	3.32	5.38	3.28	5.38	3.35	5.65	2.72	5.08	3.06	4.90	3.24	5.42
10	8.07	10.79	7.80	10.60	7.53	10.52	7.04	10.23	7.47	9.93	7.79	10.66
25	22.50	25.40	22.60	25.50	22.48	25.68	22.80	26.07	22.80	25.41	23.00	26.05
50	47.90	50.18	47.30	49.60	47.96	50.38	47.40	49.83	48.40	50.50	48.30	50.63
75	73.50	75.65	72.90	75.10	73.05	75.31	72.50	74.96	72.90	74.98	73.00	75.28
90	87.80	89.93	88.20	90.30	87.55	89.77	87.50	89.96	87.90	89.90	88.10	90.29
95	93.20	94.95	93.50	95.20	92.98	94.85	93.40	95.30	93.20	94.89	93.20	94.98
97.50	96.20	97.48	96.30	97.60	95.87	97.31	96.20	97.52	96.00	97.25	96.00	97.38
99	98.30	99.05	98.20	99.00	97.80	98.76	97.90	98.79	98.10	98.89	98.00	98.85
99.90	99.70	99.88	99.70	99.90	99.60	99.84	99.70	99.84	99.60	99.83	99.60	99.85
99.99	100.00	99.99	99.90	100.00	99.87	99.96	99.90	99.96	99.90	99.94	99.90	99.97

Note: As for Table 2.

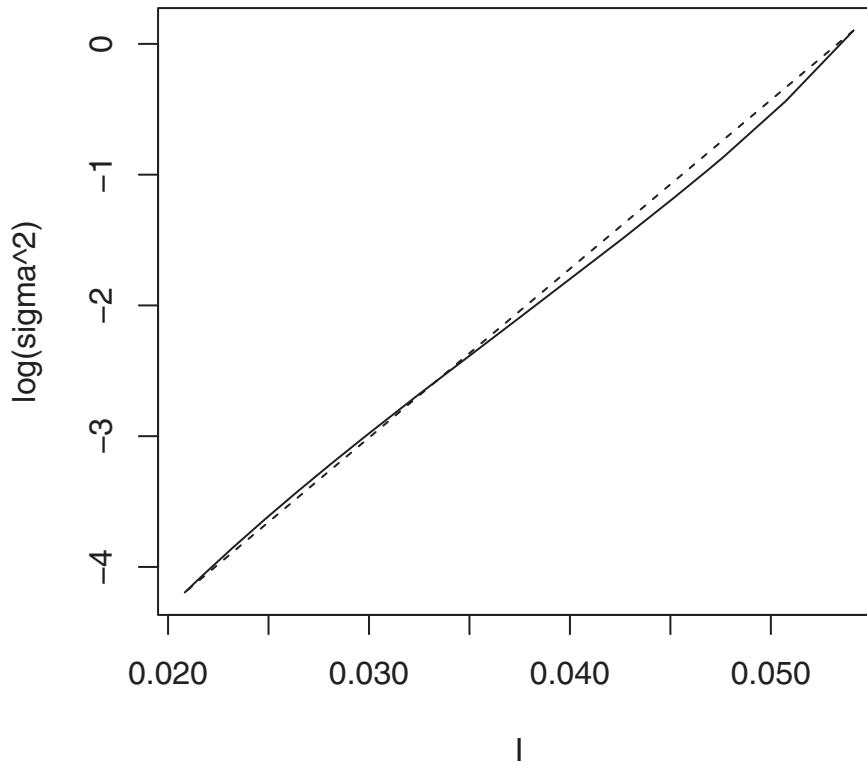

Figure 8. $v(I)$ versus $I(2)$ for the Pareto distribution.

transform is

$$t(I) = -\left(\frac{2}{\gamma_2} e^{-\gamma_1/2}\right) \exp\left(-\frac{\gamma_2}{2} I\right). \tag{5.3}$$

We proceed to examine the properties of this transform in terms of its cumulant coefficients before proceeding to examine its actual performance.

5.1. Asymptotic expansions for transforms

We present some general results before specializing them to the specific transform (5.3). Consider any non-linear transform $t(.)$ with $t'(I) \neq 0$, denote the studentized transform by

$$T_n = n^{1/2} \frac{t(I) - t(I_0)}{\widehat{\sigma} t'(I)},$$

and denote the cumulant coefficients of T_n by $\lambda_{i,j}$.[10]

[10]The term c in Corollary 5.1, which depends on α and certain moments but is scale-invariant, is induced by the estimation error of $\widehat{\sigma}$ of order $n^{-1/2}$, and is defined explicitly in the proof of this corollary.

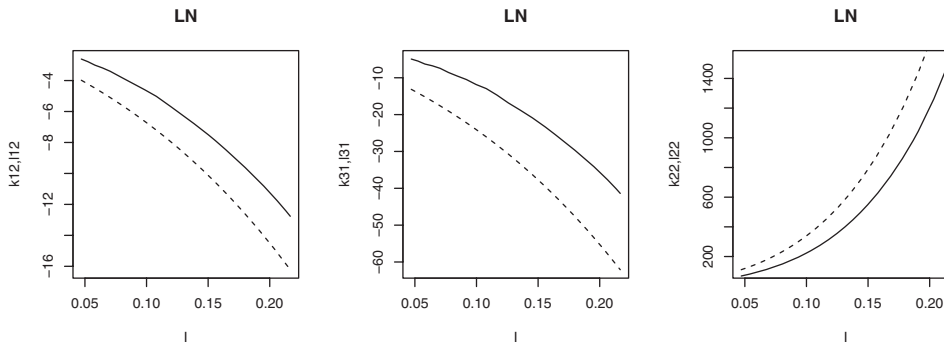

Figure 9. Cumulant coefficients of T_n and S_n in the LN case.

PROPOSITION 5.1. *To $O_p(n^{-3/2})$, we have*

$$T_n = S_n - \frac{1}{2}n^{-1/2}\widehat{\sigma}\frac{t''(I_0)}{t'(I_0)}S_n^2 + n^{-1}\widehat{\sigma}^2\left[\frac{1}{2}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{1}{3}\frac{t'''(I_0)}{t'(I_0)}\right]S_n^3.$$

COROLLARY 5.1. *The cumulant coefficients for T_n are*

$$\lambda_{1,2} = k_{1,2} - \frac{1}{2}\sigma\frac{t''(I_0)}{t'(I_0)}$$

$$\lambda_{2,2} = k_{2,2} - \sigma(k_{3,1} + 3k_{1,2})\frac{t''(I_0)}{t'(I_0)} + \sigma^2\left(\frac{15}{4}\left[\frac{t''(I_0)}{t'(I_0)}\right]^2 - 2\frac{t'''(I_0)}{t'(I_0)}\right) - \frac{t''(I_0)}{t'(I_0)}\frac{3}{2}c$$

$$\lambda_{3,1} = k_{3,1} - 3\sigma\frac{t''(I_0)}{t'(I_0)}$$

$$\lambda_{4,1} = k_{4,1} - 2\sigma\frac{t''(I_0)}{t'(I_0)}k_{5,1} - 12\sigma\frac{t''(I_0)}{t'(I_0)}k_{3,1} + 24\sigma^2\frac{t''(I_0)^2}{t'(I_0)^2} - 8\sigma^2\frac{t'''(I_0)}{t'(I_0)} - 21\frac{t''(I_0)}{t'(I_0)}c.$$

The cumulant coefficients for the specific transform (5.3) follow immediately from Corollary 5.1 with $t''(I)/t'(I) = -\gamma_2/2$ and $t'''(I)/t'(I) = (\gamma_2/2)^2$. We have the following result:

LEMMA 5.1. *If the coefficients of the odd cumulants of S_n are negative, the even ones are positive and $\gamma_2 > 0$, then the transform (5.3) reduces both bias, skewness and kurtosis for sufficiently small γ_2.*

The lemma is illustrated in Figure 9 (the solid curve depicts $\lambda_{i,j}$, the dashed line depicts $k_{i,j}$) for the LN case (qualitative similar results obtain for the other two distributions and are therefore not depicted). It is evident that the transformation has reduced the three cumulant coefficients significantly in magnitude.

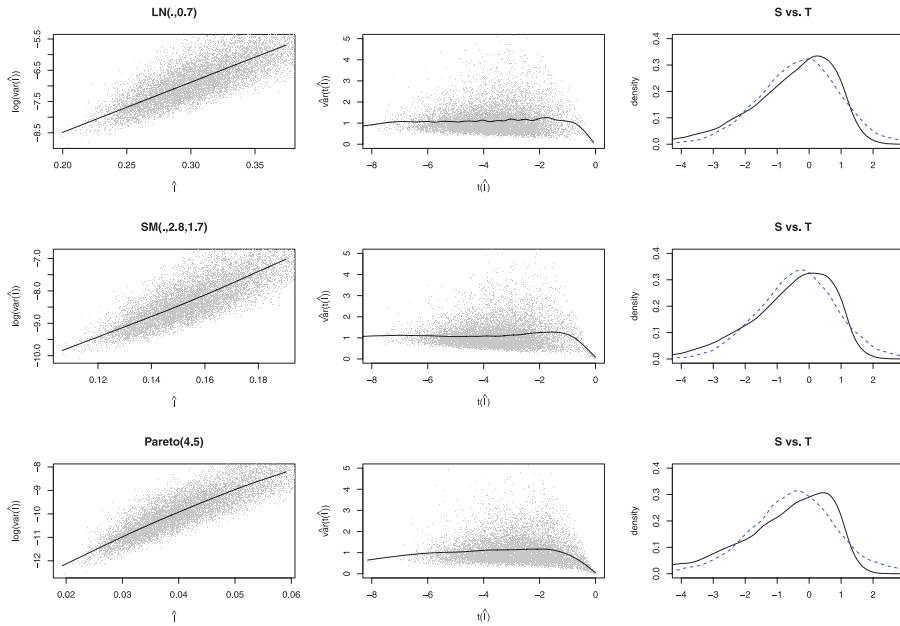

Figure 10. Aspects of variance stabilization for $I(2)$.

5.2. *Performance evidence*

Figure 10 depicts the results of applying transform (5.3), in which the coefficients (γ_1, γ_2) were estimated by a simple regression of $\log(\widehat{\text{var}}(\widehat{I}))$ on $\widehat{I}$ using the simulated data.[11] In practice, the estimates can be obtained by a preliminary bootstrap. Column 2 of the Figure plots $\widehat{\text{var}}(t(\widehat{I}))$ on $t(\widehat{I})$, and also plots a non-parametric estimate based on smooth splines. It is evident that the transform does indeed stabilize the variance since the estimated curve is approximately equal to 1 except for a small number of observations in the sparse right tail.[12] Column 3 of the Figure shows simple kernel density estimates of the densities of the studentized S_{500} (solid line) and T_{500} (dashed line). The density of the transform is more symmetric and the skewness problem has been much reduced. Hence we expect performance improvements for inference when the transform is followed by an application of the bootstrap. Qualitatively similar pictures obtain for different values of α and different income distribution parameters.

Tables 4 to 6 consider detailed bootstrap evidence for the performance of the studentized variance stabilizing transform, and benchmark this against the normal (first order) approximation and the performance of the studentized bootstrap, as the tail of the income distribution becomes progressively more heavy. The experiments are conducted for samples of sizes 250, 500 and 1000, across LN, SM and Pareto distributions (we also include the case $\lambda = 2.5$ and $\alpha = 2$ when

[11]The estimates of (γ_1, γ_2) in the $LN(., .7)$ case are $(-11.7, 15.9)$, in the $SM(., 2.8, 1.7)$ case $(-13.6, 34.1)$, and in the $P(4.5)$ case $(-13.9, 98.8)$.

[12]In the SM case, the non-parametric curve falls below 1 around -0.75, and about 1% of the simulated data lie to the right of this point; in the depicted LN case, the respective numbers are -0.9 and 0.7%.

Table 4. Coverage error rate of nominal 95% two-sided confidence intervals for $I(2)$.

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
									$n = 250$									
		$LN(., .3)$			$LN(., .4)$			$LN(., .5)$			$LN(., .6)$			$LN(., .7)$			$LN(., .8)$	
Normal app.	1.09	6.62	7.71	0.87	8.31	9.18	0.72	10.73	11.45	0.60	13.64	14.24	0.54	17.14	17.68	0.49	21.06	21.55
Stud. boot.	2.10	3.94	6.04	1.82	4.90	6.72	1.47	6.29	7.76	1.12	7.97	9.09	0.78	9.98	10.76	0.51	12.12	12.63
Var-stab boot.	1.77	3.59	5.35	1.43	4.41	5.84	1.09	5.49	6.58	0.83	6.92	7.75	0.65	8.50	9.16	0.52	10.21	10.73
		$SM(., 3.6, 1.7)$			$SM(., 3.4, 1.7)$			$SM(., 3.2, 1.7)$			$SM(., 3.0, 1.7)$			$SM(., 2.8, 1.7)$			$SM(., 2.6, 1.7)$	
Normal app.	0.46	11.76	12.22	0.42	12.83	13.25	0.37	14.11	14.48	0.33	15.73	16.06	0.29	17.66	17.95	0.26	20.36	20.61
Stud. boot.	1.24	7.13	8.37	1.11	7.83	8.94	0.98	8.69	9.67	0.84	9.71	10.55	0.67	10.99	11.66	0.54	12.59	13.12
Var-stab boot.	0.80	6.24	7.04	0.71	6.84	7.55	0.61	7.52	8.13	0.51	8.38	8.88	0.42	9.44	9.86	0.34	10.78	11.12
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.31	13.21	13.52	0.28	14.24	14.51	0.24	15.85	16.09	0.17	19.17	19.34	0.08	28.87	28.95	0.41	62.63	63.04
Stud. boot.	1.34	6.33	7.67	1.24	6.81	8.05	1.06	7.64	8.70	0.75	9.45	10.20	0.28	14.80	15.08	0.05	37.18	37.22
Var-stab boot.	0.60	5.04	5.63	0.53	5.41	5.94	0.43	6.03	6.45	0.28	7.38	7.66	0.10	11.27	11.37	0.77	25.15	25.91
									$n = 500$									
		$LN(., .3)$			$LN(., .4)$			$LN(., .5)$			$LN(., .6)$			$LN(., .7)$			$LN(., .8)$	
Normal app.	1.19	5.48	6.66	0.93	6.84	7.76	0.70	8.68	9.37	0.53	11.00	11.53	0.47	13.8	14.27	0.37	17.41	17.77
Stud. boot.	2.25	3.50	5.74	2.05	4.23	6.28	1.73	5.19	6.92	1.37	6.51	7.87	1.09	8.03	9.12	0.72	9.95	10.66
Var-stab boot.	2.02	3.28	5.30	1.68	3.89	5.57	1.28	4.70	5.98	0.94	5.73	6.66	0.70	6.85	7.54	0.48	8.39	8.88
		$SM(., 3.6, 1.7)$			$SM(., 3.4, 1.7)$			$SM(., 3.2, 1.7)$			$SM(., 3.0, 1.7)$			$SM(., 2.8, 1.7)$			$SM(., 2.6, 1.7)$	
Normal app.	0.48	9.90	10.38	0.44	10.80	11.23	0.38	12.00	12.37	0.32	13.39	13.71	0.27	15.18	15.45	0.21	17.46	17.66
Stud. boot.	1.57	6.27	7.84	1.43	6.85	8.28	1.28	7.54	8.82	1.12	8.37	9.49	0.93	9.49	10.42	0.74	10.85	11.59
Var-stab boot.	1.00	5.64	6.64	0.87	6.10	6.97	0.74	6.68	7.42	0.61	7.41	8.02	0.48	8.32	8.80	0.36	9.36	9.73

Continued

Table 4. Continued

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.38	10.44	10.82	0.34	11.30	11.64	0.27	12.71	12.98	0.18	15.74	15.92	0.06	24.94	25.00	0.16	59.47	59.63
Stud. boot.	1.68	5.39	7.07	1.55	5.80	7.35	1.34	6.54	7.87	1.02	8.10	9.12	0.40	13.27	13.67	0.03	36.40	36.43
Var-stab boot.	0.85	4.46	5.30	0.73	4.82	5.55	0.57	5.38	5.95	0.36	6.58	6.94	0.11	10.35	10.45	0.31	25.08	25.39
									$n = 1,000$									
		$LN(.,.3)$			$LN(.,.4)$			$LN(.,.5)$			$LN(.,.6)$			$LN(.,.7)$			$LN(.,.8)$	
Normal app.	1.43	4.66	6.09	1.19	5.56	6.75	0.90	6.94	7.84	0.63	8.94	9.57	0.46	11.39	11.85	0.32	14.29	14.61
Stud. boot.	2.48	3.18	5.65	2.39	3.73	6.12	2.16	4.47	6.63	1.75	5.44	7.18	1.39	6.72	8.11	1.02	8.17	9.19
Var-stab boot.	2.32	3.06	5.38	2.06	3.54	5.60	1.70	4.13	5.83	1.22	4.87	6.09	0.84	5.91	6.76	0.57	6.93	7.50
		$SM(.,3.6,1.7)$			$SM(.,3.4,1.7)$			$SM(.,3.2,1.7)$			$SM(.,3.0,1.7)$			$SM(.,2.8,1.7)$			$SM(.,2.6,1.7)$	
Normal app.	0.53	8.32	8.85	0.46	9.10	9.56	0.43	10.09	10.51	0.36	11.57	11.93	0.28	13.17	13.45	0.19	15.00	15.19
Stud. boot.	1.70	5.46	7.16	1.58	5.89	7.46	1.58	6.53	8.11	1.39	7.38	8.77	1.18	8.35	9.53	0.96	9.47	10.43
Var-stab boot.	1.12	4.97	6.09	1.00	5.35	6.35	0.93	5.90	6.83	0.74	6.59	7.32	0.56	7.31	7.87	0.41	8.27	8.67
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.57	8.29	8.85	0.50	8.98	9.48	0.40	10.06	10.46	0.23	12.77	13.00	0.07	21.37	21.45	0.06	57.14	57.20
Stud. boot.	1.99	4.60	6.59	1.88	4.97	6.85	1.67	5.45	7.12	1.32	7.05	8.38	0.61	11.84	12.44	0.02	35.05	35.07
Var-stab boot.	1.30	4.01	5.32	1.15	4.29	5.44	0.91	4.62	5.54	0.59	5.88	6.47	0.14	9.48	9.62	0.13	24.96	25.09

Note: The heaviness of the tail of the income distribution increases across the panels from left to right. The nominal error rate is 5 %. 'Stud. var-stab. bootstrap' is the studentized bootstrap for the variance stabilizing transform given by eq. (5.3). 'L' are rejections on the left of the confidence interval, 'R' are rejections on the right and 'T' are the total coverage error rates [%]. Based on R=100,000 repetitions, in each repetitions B=999 bootstrap samples were drawn.

Table 5. Coverage error rate of nominal 95% two-sided confidence intervals for $I(1.05)$.

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
									$n = 250$									
		$LN(.,.3)$			$LN(.,.4)$			$LN(.,.5)$			$LN(.,.6)$			$LN(.,.7)$			$LN(.,.8)$	
Normal app.	1.31	5.52	6.83	1.19	6.24	7.44	1.07	7.20	8.27	0.99	8.31	9.30	0.92	9.70	10.62	0.90	11.40	12.30
Stud. boot.	2.24	3.28	5.52	2.12	3.72	5.85	1.96	4.29	6.24	1.73	4.98	6.70	1.51	5.79	7.30	1.26	6.79	8.05
Var-stab boot.	2.06	3.05	5.11	1.90	3.43	5.34	1.69	3.94	5.63	1.49	4.56	6.05	1.28	5.24	6.52	1.12	6.08	7.20
		$SM(.,3.6,1.7)$			$SM(.,3.4,1.7)$			$SM(.,3.2,1.7)$			$SM(.,3.0,1.7)$			$SM(.,2.8,1.7)$			$SM(.,2.6,1.7)$	
Normal app.	0.78	7.27	8.05	0.75	7.61	8.35	0.71	8.01	8.72	0.67	8.54	9.21	0.71	9.36	10.07	0.58	10.09	10.67
Stud. boot.	1.82	4.31	6.13	1.72	4.55	6.27	1.62	4.87	6.50	1.52	5.24	6.77	1.53	5.81	7.34	1.29	6.24	7.52
Var-stab boot.	1.44	3.87	5.31	1.37	4.10	5.47	1.27	4.38	5.64	1.17	4.70	5.87	1.15	5.21	6.37	0.94	5.59	6.53
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.32	11.94	12.26	0.30	12.51	12.80	0.26	13.36	13.62	0.21	14.90	15.10	0.16	18.90	19.05	0.23	29.64	29.87
Stud. boot.	1.47	5.59	7.06	1.39	5.87	7.26	1.28	6.38	7.66	1.08	7.27	8.35	0.67	9.37	10.04	0.20	15.36	15.55
Var-stab boot.	0.69	4.52	5.21	0.62	4.73	5.35	0.54	5.06	5.60	0.44	5.74	6.18	0.26	7.38	7.64	0.36	11.68	12.04
									$n = 500$									
		$LN(.,.3)$			$LN(.,.4)$			$LN(.,.5)$			$LN(.,.6)$			$LN(.,.7)$			$LN(.,.8)$	
Normal app.	1.42	4.46	5.87	1.29	4.95	6.24	1.14	5.60	6.75	1.05	6.47	7.52	0.97	7.51	8.48	0.88	8.77	9.66
Stud. boot.	2.35	2.90	5.25	2.28	3.21	5.49	2.16	3.60	5.77	2.00	4.10	6.09	1.81	4.68	6.49	1.61	5.45	7.07
Var-stab boot.	2.21	2.76	4.97	2.11	3.05	5.16	1.92	3.36	5.27	1.71	3.79	5.51	1.53	4.30	5.83	1.33	4.96	6.28
		$SM(.,3.6,1.7)$			$SM(.,3.4,1.7)$			$SM(.,3.2,1.7)$			$SM(.,3.0,1.7)$			$SM(.,2.8,1.7)$			$SM(.,2.6,1.7)$	
Normal app.	0.98	6.09	7.07	0.92	6.38	7.30	0.85	6.77	7.62	0.76	7.25	8.01	0.69	7.82	8.51	0.62	8.57	9.19
Stud. boot.	2.09	3.92	6.02	2.03	4.11	6.14	1.97	4.36	6.33	1.89	4.65	6.55	1.76	5.06	6.82	1.63	5.52	7.15
Var-stab boot.	1.78	3.62	5.40	1.68	3.82	5.50	1.61	4.04	5.65	1.48	4.30	5.78	1.34	4.65	5.98	1.17	5.06	6.22

Continued

Table 5. Continued

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.49	9.12	9.61	0.44	9.59	10.03	0.38	10.32	10.71	0.30	11.67	11.97	0.18	15.34	15.53	0.14	25.32	25.46
Stud. boot.	1.86	4.61	6.47	1.79	4.83	6.62	1.69	5.23	6.92	1.49	5.98	7.46	1.02	7.94	8.96	0.34	13.62	13.96
Var-stab boot.	1.12	3.89	5.02	1.04	4.10	5.14	0.91	4.37	5.28	0.71	4.92	5.62	0.39	6.39	6.77	0.23	10.72	10.95
									$n = 1,000$									
		$LN(., .3)$			$LN(., .4)$			$LN(., .5)$			$LN(., .6)$			$LN(., .7)$			$LN(., .8)$	
Normal app.	1.69	3.91	5.60	1.54	4.28	5.81	1.39	4.80	6.19	1.23	5.48	6.71	1.08	6.26	7.34	0.96	7.24	8.20
Stud. boot.	2.52	2.76	5.28	2.47	2.95	5.42	2.42	3.20	5.62	2.27	3.61	5.88	2.12	4.07	6.19	1.90	4.59	6.49
Var-stab boot.	2.44	2.69	5.13	2.34	2.84	5.18	2.26	3.05	5.31	2.06	3.42	5.48	1.84	3.83	5.67	1.59	4.24	5.83
		$SM(., 3.6, 1.7)$			$SM(., 3.4, 1.7)$			$SM(., 3.2, 1.7)$			$SM(., 3.0, 1.7)$			$SM(., 2.8, 1.7)$			$SM(., 2.6, 1.7)$	
Normal app.	1.12	5.24	6.35	1.05	5.48	6.52	0.99	5.79	6.78	0.92	6.19	7.11	0.81	6.70	7.51	0.72	7.40	8.12
Stud. boot.	2.36	3.64	6.00	2.31	3.82	6.13	2.24	4.05	6.29	2.15	4.28	6.43	2.07	4.60	6.67	1.96	4.99	6.94
Var-stab boot.	2.06	3.46	5.52	1.99	3.63	5.62	1.89	3.82	5.71	1.77	4.04	5.80	1.63	4.32	5.95	1.47	4.62	6.09
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.66	7.32	7.98	0.62	7.69	8.31	0.55	8.33	8.88	0.44	9.48	9.92	0.26	12.51	12.76	0.11	21.90	22.01
Stud. boot.	2.15	4.04	6.19	2.08	4.23	6.31	2.00	4.55	6.54	1.84	5.16	7.00	1.39	6.88	8.26	0.54	12.19	12.73
Var-stab boot.	1.58	3.62	5.20	1.47	3.75	5.22	1.32	3.99	5.31	1.06	4.48	5.54	0.62	5.81	6.43	0.19	9.82	10.01

Note: As for Table 4.

Table 6. Coverage error rate of nominal 95% two-sided confidence intervals for $I(0.05)$.

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
									$n = 250$									
		$LN(.,.3)$			$LN(.,.4)$			$LN(.,.5)$			$LN(.,.6)$			$LN(.,.7)$			$LN(.,.8)$	
Normal app.	1.40	4.82	6.21	1.35	4.99	6.34	1.30	5.27	6.58	1.26	5.64	6.90	1.23	6.01	7.24	1.19	6.48	7.67
Stud. boot.	2.42	2.86	5.28	2.36	2.95	5.31	2.29	3.11	5.41	2.22	3.33	5.55	2.14	3.60	5.74	2.01	3.96	5.97
Var-stab boot.	2.23	2.67	4.90	2.19	2.77	4.95	2.11	2.92	5.02	2.03	3.11	5.14	1.91	3.34	5.24	1.81	3.61	5.42
		$SM(.,3.6,1.7)$			$SM(.,3.4,1.7)$			$SM(.,3.2,1.7)$			$SM(.,3.0,1.7)$			$SM(.,2.8,1.7)$			$SM(.,2.6,1.7)$	
Normal app.	1.07	5.89	6.96	1.07	5.96	7.03	1.05	6.04	7.09	1.04	6.12	7.16	1.00	6.40	7.41	1.02	6.24	7.26
Stud. boot.	2.27	3.33	5.60	2.24	3.37	5.61	2.23	3.42	5.65	2.22	3.47	5.69	2.10	3.65	5.75	2.15	3.56	5.71
Var-stab boot.	1.95	3.05	5.00	1.94	3.08	5.02	1.90	3.12	5.03	1.88	3.18	5.06	1.78	3.36	5.14	1.82	3.25	5.07
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.43	10.76	11.19	0.42	10.98	11.40	0.40	11.40	11.80	0.37	12.20	12.57	0.30	14.00	14.31	0.23	18.77	19.00
Stud. boot.	1.66	5.15	6.81	1.63	5.28	6.92	1.58	5.51	7.09	1.47	5.87	7.34	1.24	6.76	8.00	0.75	9.28	10.03
Var-stab boot.	0.90	4.19	5.08	0.85	4.28	5.13	0.80	4.41	5.22	0.72	4.71	5.43	0.56	5.40	5.96	0.35	7.26	7.61
									$n = 500$									
		$LN(.,.3)$			$LN(.,.4)$			$LN(.,.5)$			$LN(.,.6)$			$LN(.,.7)$			$LN(.,.8)$	
Normal app.	1.72	4.00	5.73	1.66	4.10	5.76	1.62	4.27	5.89	1.54	4.48	6.02	1.47	4.80	6.27	1.39	5.14	6.52
Stud. boot.	2.57	2.63	5.20	2.56	2.74	5.30	2.54	2.81	5.35	2.49	2.97	5.46	2.43	3.16	5.59	2.36	3.35	5.71
Var-stab boot.	2.49	2.53	5.02	2.45	2.62	5.07	2.41	2.69	5.10	2.34	2.83	5.18	2.28	3.00	5.28	2.18	3.16	5.34
		$SM(.,3.6,1.7)$			$SM(.,3.4,1.7)$			$SM(.,3.2,1.7)$			$SM(.,3.0,1.7)$			$SM(.,2.8,1.7)$			$SM(.,2.6,1.7)$	
Normal app.	1.28	4.82	6.10	1.27	4.88	6.14	1.25	4.96	6.20	1.23	5.04	6.26	1.18	5.19	6.38	1.14	5.28	6.42
Stud. boot.	2.30	3.07	5.37	2.30	3.09	5.39	2.29	3.14	5.43	2.27	3.20	5.47	2.25	3.23	5.48	2.19	3.41	5.60
Var-stab boot.	2.12	2.88	5.00	2.11	2.92	5.03	2.08	2.96	5.04	2.03	3.01	5.04	2.01	3.02	5.03	1.96	3.19	5.15

Continued

Table 6. Continued

	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T	L	R	T
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.56	8.24	8.80	0.54	8.43	8.98	0.51	8.74	9.25	0.45	9.36	9.81	0.37	10.97	11.34	0.24	15.31	15.55
Stud. boot.	1.98	4.18	6.15	1.94	4.29	6.23	1.89	4.47	6.36	1.79	4.78	6.57	1.55	5.57	7.11	0.97	8.02	8.99
Var-stab boot.	1.31	3.54	4.85	1.25	3.60	4.86	1.18	3.76	4.94	1.05	4.01	5.06	0.81	4.67	5.48	0.46	6.47	6.93
									$n = 1{,}000$									
		$LN(., .3)$			$LN(., .4)$			$LN(., .5)$			$LN(., .6)$			$LN(., .7)$			$LN(., .8)$	
Normal app.	1.81	3.61	5.42	1.77	3.70	5.47	1.72	3.80	5.52	1.65	3.95	5.60	1.57	4.19	5.76	1.47	4.49	5.96
Stud. boot.	2.47	2.64	5.11	2.46	2.69	5.14	2.42	2.73	5.16	2.42	2.84	5.25	2.39	2.95	5.34	2.36	3.12	5.48
Var-stab boot.	2.43	2.59	5.02	2.40	2.61	5.01	2.35	2.67	5.02	2.33	2.75	5.08	2.29	2.85	5.14	2.24	3.00	5.24
		$SM(., 3.6, 1.7)$			$SM(., 3.4, 1.7)$			$SM(., 3.2, 1.7)$			$SM(., 3.0, 1.7)$			$SM(., 2.8, 1.7)$			$SM(., 2.6, 1.7)$	
Normal app.	1.55	4.12	5.67	1.53	4.17	5.70	1.50	4.23	5.74	1.47	4.31	5.78	1.42	4.52	5.93	1.37	4.56	5.93
Stud. boot.	2.46	2.80	5.27	2.47	2.82	5.29	2.45	2.89	5.34	2.46	2.95	5.41	2.41	3.09	5.50	2.40	3.18	5.58
Var-stab boot.	2.36	2.71	5.07	2.35	2.74	5.09	2.34	2.79	5.13	2.33	2.85	5.17	2.27	3.01	5.28	2.26	3.04	5.30
		$P(12)$			$P(10)$			$P(8)$			$P(6)$			$P(4)$			$P(2.5)$	
Normal app.	0.79	6.54	7.33	0.77	6.73	7.50	0.71	7.01	7.71	0.63	7.54	8.17	0.49	8.84	9.33	0.27	12.66	12.93
Stud. boot.	2.22	3.70	5.91	2.20	3.78	5.97	2.15	3.90	6.05	2.07	4.17	6.23	1.86	4.84	6.70	1.23	6.97	8.20
Var-stab boot.	1.71	3.31	5.02	1.66	3.38	5.04	1.59	3.45	5.04	1.48	3.67	5.15	1.15	4.20	5.35	0.63	5.89	6.52

Note: As for Table 4.

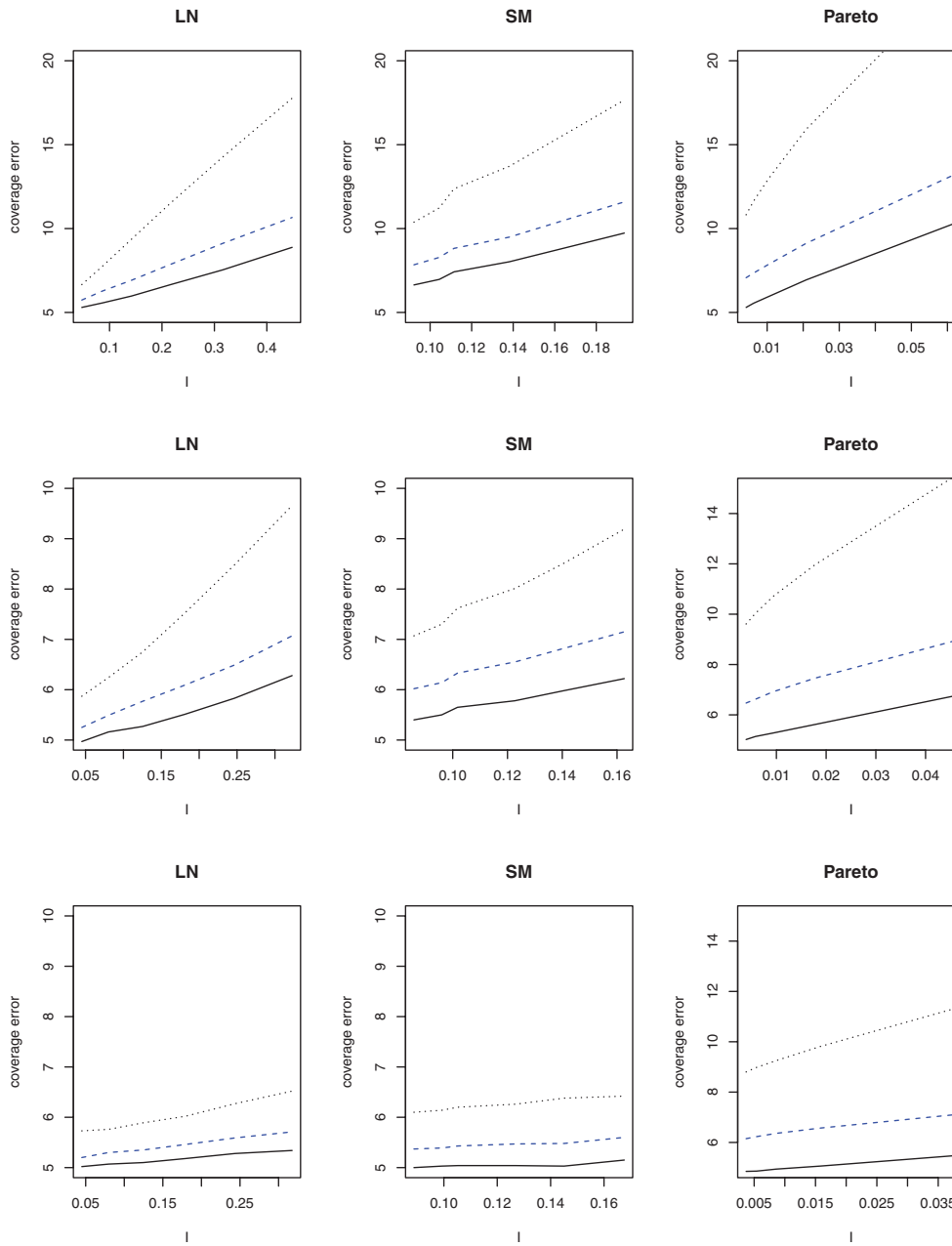

Figure 11. Coverage errors of nominal 95% two-sided confidence intervals as functions of I.

σ^2 does not exist). Each table considers one value of α . The focus is on the coverage error rates for two-sided confidence intervals with nominal error rate 5%, based on drawing $B = 999$ bootstrap samples of size n and repeating the experiment $R = 100,000$ times. The tables break down the total coverage error rates (T), into rejections on the left (L) of the confidence interval, i.e. when the population value lies to the left of the lower confidence limit, and into rejections on the right (R). Figure 11 (Row 1: $\alpha = 2$, row 2: $\alpha = 1.05$, row 3: $\alpha = 0.05$, based on Tables 4 to 6 for $n = 500$, normal approximation (dotted line), studentized bootstrap (dashed line), studentized bootstrap of the variance stabilizing transform (solid line)) summarizes this evidence across the αs for $n = 500$, and depicts the total coverage errors as functions of I.

The poor quality of the normal approximation has been discussed extensively above. Consistent with Davidson and Flachaire (2007), the studentized bootstrap improves on this but for $\alpha = 2$ the discrepancy between nominal and actual coverage behaviour is still substantial even for samples of size 1000. For instance, in the *SM* case with α the actual error rate is still twice the nominal rate. The variance stabilizing transform improves performance further across all income distributions and αs.

6. CONCLUSIONS

We have considered the inference problem for inequality measures when incomes are generated by distributions with sufficiently slowly decaying tails, and have demonstrated how the severity of the inference problem responds to the exact nature of the right tail of the income distribution. In particular, it has been shown that the coverage failures of the usual two-sided confidence intervals are associated with particularly low realizations of both $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$. To understand better the separate and joint contributions of both estimators, we have derived and examined quantitatively the bias, skewness and kurtosis coefficients for both the standardized and studentized inequality measures. Exploiting the uncovered systematic relationship between $\widehat{I}$ and $\widehat{\text{var}}(\widehat{I})$, we have proposed variance stabilizing transforms. Such transforms are shown to lead to improved inference, and could be used as inputs in more sophisticated bootstrap methods. The diagnosis of the inference problem and the suggested avenues for remedies complement the methods surveyed in Davidson (2011), and is further discussed in Schluter (2011).

ACKNOWLEDGMENTS

Thanks to Oliver Linton and the referees whose constructive comments have helped to improve this paper considerably.

REFERENCES

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics 108*, 317–42.

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics 42*, 27–41.

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson, F. Bourguignon (Eds.), *Handbook of Income Distribution*, 87–166. Amsterdam: North-Holland.

Cowell, F. A. and E. Flachaire (2007). Income Distribution and Inequality Measurement: The Problem of Extreme Values. *Journal of Econometrics 141*, 1044–72.

Cowell, F. A. and M.-P. Victoria-Feser (1996). Robustness Properties of Inequality Measures. *Econometrica 64*, 77–101.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics 25*, 631–50.

Davidson, R. (2012). Statistical inference in the presence of heavy tails. *Econometrics Journal 15*, C31–C53.

Davidson, R. and E. Flachaire (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics 141*, 141–66.

Easton, G. S. and E. Ronchetti (1986). General saddlepoint approximations with applications to L-statistics. *Journal of the American Statistical Association 81*, 420–30.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansions*. Berlin: Springer.

Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science 3*, 213–27.

Rothenberg, T. (1988). Approximate power functions for some robust tests of regression coefficients. *Econometrica 56*, 997–1019.

Schluter, C. (2012). Discussion of S.G. Donald et al. and R. Davidson. *Econometrics Journal 15*, C54–C57.

Schluter, C. and M. Trede (2002). Tails of Lorenz curves. *Journal of Econometrics 109*, 151–66.

Schluter, C. and K. J. van Garderen (2009). Edgeworth Expansions and Normalizing Transforms for Inequality Measures. *Journal of Econometrics 150*, 16–29.