School of Social Sciences
Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

On the Problem of Inference for Inequality Measures for Heavy-Tailed Distributions

Christian Schluter

No. 1009

ISSN 0966-4246

On the Problem of Inference for Inequality Measures for Heavy-Tailed Distributions

Christian Schluter*
University of Southampton

May 2010

Abstract

The received wisdom about inference problems for inequality measures is that these are caused by the presence of extremes in samples drawn from heavy-tailed distributions. We show that this is incorrect since the density of the studentised inequality measure is heavily skewed to the left, and the excessive coverage failures of the usual confidence intervals are associated with low estimates of both the point measure and the variance. For further diagnostics the coefficients of bias, skewness and kurtosis are derived for both studentised and standardised inequality measures, and the explicit cumulant expansions make also available Edgeworth expansions and saddlepoint approximations. In view of the key role played by the estimated variance of the measure, variance stabilising transforms are considered and shown to improve inference.

Keywords: Inequality measures, inference, statistical performance, asymptotic expansions, variance stabilisation.

JEL classification: C10, D31, D63.

*Department of Economics, University of Southampton, Highfield, Southampton, SO17 1BJ, UK. Tel. +44 (0)2380 59 5909, Fax. +44 (0)2380 59 3858. Email: C.Schluter@soton.ac.uk. http://www.economics.soton.ac.uk/staff/schluter/

1 Introduction

While first order asymptotics for estimators of measures of inequality, such as Generalized Entropy 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 which 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 error errors significantly larger than their nominal rates thus rendering inference unreliable. Davidson and Flachaire (2007) have shown that this 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. In particular, if the distribution is heavy-tailed, samples are likely to contain "extremes" or "outliers", i.e. income realisations from the tail of the distribution which are substantially larger than income realisations associated with the main body of the distribution. The natural intuition, pursued in Cowell and Flachaire (2007), Davidson and Flachaire (2007) and Davidson (2010) 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). We show in this paper that this intuition is incorrect since 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. We also show that this holds for income distributions whose right tail decays faster than a power function, such as the lognormal provided its shape parameter is sufficiently large.

The principal contribution of the paper is the diagnosis of the underlying problem for inference. Denoting $\widehat{I}$ and $\widehat{var}(\widehat{I})$ the standard estimators of the inequality measure and its variance, the problem is made visible via simulations in plots of realisations of $\widehat{var}(\widehat{I})$ against $\widehat{I}$ and identifying those $(\widehat{I}, \widehat{var}(\widehat{I}))$ pairs which are associated with a coverage failure of standard two-sided confidence intervals. Since the actual density of the studentised 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* realisations of both $\widehat{I}$ and $\widehat{var}(\widehat{I})$.

In order to understand better the separate and joint contributions of $\widehat{I}$ and $\widehat{var}(\widehat{I})$ to the inference problem we develop asymptotic expansions for both studentised and standardised (by the theoretical variance) inequality measures. Building on the second-order expansions of Schluter and van Garderen (2009), we now consider third-order expansions. In particular, we derive the bias, skewness and kurtosis coefficients. The principal usefulness of these coefficients is as diagnostic tools. These enable us to quantify the departure from normality of the finite sample distributions, and to highlight the role of the variance estimate $\widehat{var}(\widehat{I})$ in the comparison between the studentised and the standardised inequality measures. While the density of the standardised inequality measure is close to normal and skewness is only modest, the studentised density exhibits significant skewness. This good performance of the standardised inequality measure contrasts starkly with the poor performance of the studentised measure, and also shows that it is not the non-linearity of the inequality index per se, as suggested in Davidson and Flachaire (2007), which contributes to the poor performance but it 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. Given the prominent role of the estimated variance $\widehat{var}(\widehat{I})$ in the inference problem, we consider the merit of using variance stabilising transforms, and show that, in conjunction with the studentised bootstrap, inference does improve.

The plan of paper is as follows. The class of inequality measures considered in this paper is defined in Section 2. Section 3 presents the simulation evidence which shows that the received wisdom about the role of extremes is incorrect. In particular, we show that it is particularly low realisations of both $\widehat{I}$ and $\widehat{var}(\widehat{I})$ which are associated with excessive coverage errors of the usual two-sided confidence intervals. We propose asymptotic expansions for the cumulants of both standardised and studentised 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 of the cumulant coefficients are collected in Appendix A. Given the availability of the cumulant coefficients, it is of interest to investigate the extent to which Edgeworth expansions and saddlepoint approximations yield distributional improvements. This is done in the digression of Section 4.1. Improvements are shown to be available for income distributions with sufficiently fast decaying tails, but in line with established results in the statistics literature in different settings, performance worsens as the magnitude of the moments increase sufficiently. The specific behaviour of the estimated variance $\widehat{var}(\widehat{I})$ suggests the application of variance stabilising transforms. This is done in Section 4.2, which also quantifies the resulting inferential improvements. All proofs are collected in the appendix.

2 Generalized Entropy Indices of Inequality

We consider the popular and leading class of inequality indices, the Generalized Entropy (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

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

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 upper tail 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 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 (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 studentised index

$$S_n = \sqrt{n}\left(\frac{\widehat{I} - I}{\widehat{\sigma}}\right), \quad (2)$$

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 $\sigma = \left[\left(\alpha^2 - \alpha\right)\mu_1^{\alpha+1}\right]^{-1} B_0^{1/2}$ with $B_0 = \alpha^2\mu_\alpha^2\mu_2 - 2\alpha\mu_1\mu_\alpha\mu_{\alpha+1} + \mu_1^2\mu_{2\alpha} -$

$(1-\alpha)^2 \mu_1^2 \mu_\alpha^2$. $\widehat{B}_0$ 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 standardised inequality measure

$$\widetilde{S}_n = \sqrt{n}\left(\frac{\widehat{I}-I}{\sigma}\right). \tag{3}$$

We will distinguish standardised quantities from their studentised counterparts throughout by tildes. Simplifying a little we have thus $S_n = \sqrt{n}\widehat{B}_0^{-1/2}\left[m_\alpha m_1 - \mu_1^{-\alpha}\mu_\alpha m_1^{\alpha+1}\right]$ 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 Gaussian distribution (see e.g. Cowell, 1989), thus $\Pr(S_n \leq x) = \Phi(x) + O\left(n^{-1/2}\right)$ where Φ denotes the Gaussian distribution.

3 Simulation Evidence: the Role of $\widehat{I}$, $\widehat{var}(\widehat{I})$, and the Tail Behaviour of F

We follow the previous literature cited above and consider the two leading parametric income distributions which are regularly used to fit real real-world income data, namely the lognormal $LN(m, sd)$ and the Singh-Maddala distribution $SM(a,b,c)$ whose density is $f(x;a,b,c) = abcx^{b-1}/\left(1+ax^b\right)^{c+1}$. These distributions are skewed to the right, but differ in other ways, such as their right tail behaviour. In particular, the tail of LN decays exponentially fast, whereas Schluter and Trede (2002) have shown that the tail of SM decays like a power function (with right tail index equal to bc).

Generalized Entropy 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 below. Since I is scale invariant, so is σ and thus S_n. The population values are in the lognormal case $I(\alpha; sd) = \left(\alpha^2-\alpha\right)^{-1} \times [\exp(0.5\left(\alpha^2-\alpha\right)(sd)^2) - 1]$, and in the Singh-Maddala case, defined only for $bc > \alpha$, $I(\alpha; b, c) = \left(\alpha^2-\alpha\right)^{-1} c^{-(\alpha-1)} B(1+\alpha/b, c-\alpha/b)$ $/\,[B(1+1/b, c-1/b)^\alpha - 1]$ where $B(\cdot,\cdot)$ denotes the Beta function. The asymptotic variance σ^2 of the inequality measure is always finite in the LN case, but in the SM case we require that $bc > \max\{2, 1+\alpha, 2\alpha\}$.

For the sake of brevity, we illustrate the main insights for two lognormal cases with $sd \in \{.3, .7\}$ and the SM case with $b = 2.8$ and $c = 1.7$,[1] while letting the sensitivity parameter α of the inequality index take on values in $\{.05, 1.05, 2\}$, so we focus essentially on the Theil indices and $I(2)$. We consider samples of size $n = 500$, and repeat the experiments 10,000 times. The simulation exercises are illustrative, and not exhaustive. Complementary simulation evidence is provided in Davidson and Flachaire (2007) and Schluter and van Garderen (2009). Our qualitative conclusions also hold for these other settings.

The first set of experiments simply consists in estimating, using standard kernel density estimators, the actual densities of the studentised inequality measure S_{500} and of the standardised 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. The kernel density estimates for S_{500} in the SM clearly reveal the substantial skewness the density of the studentised 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 Gaussian 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 standardised inequality measure $\widetilde{S}_{500}$ do appear very symmetric. However, the densities also exhibit a greater concentration around 0 than the Gaussian

[1] $SM(.,2.8,1.7)$ and $LN(.,.3)$ are good fitting parametrisations of the income distribution in Germany.

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 Gaussian.

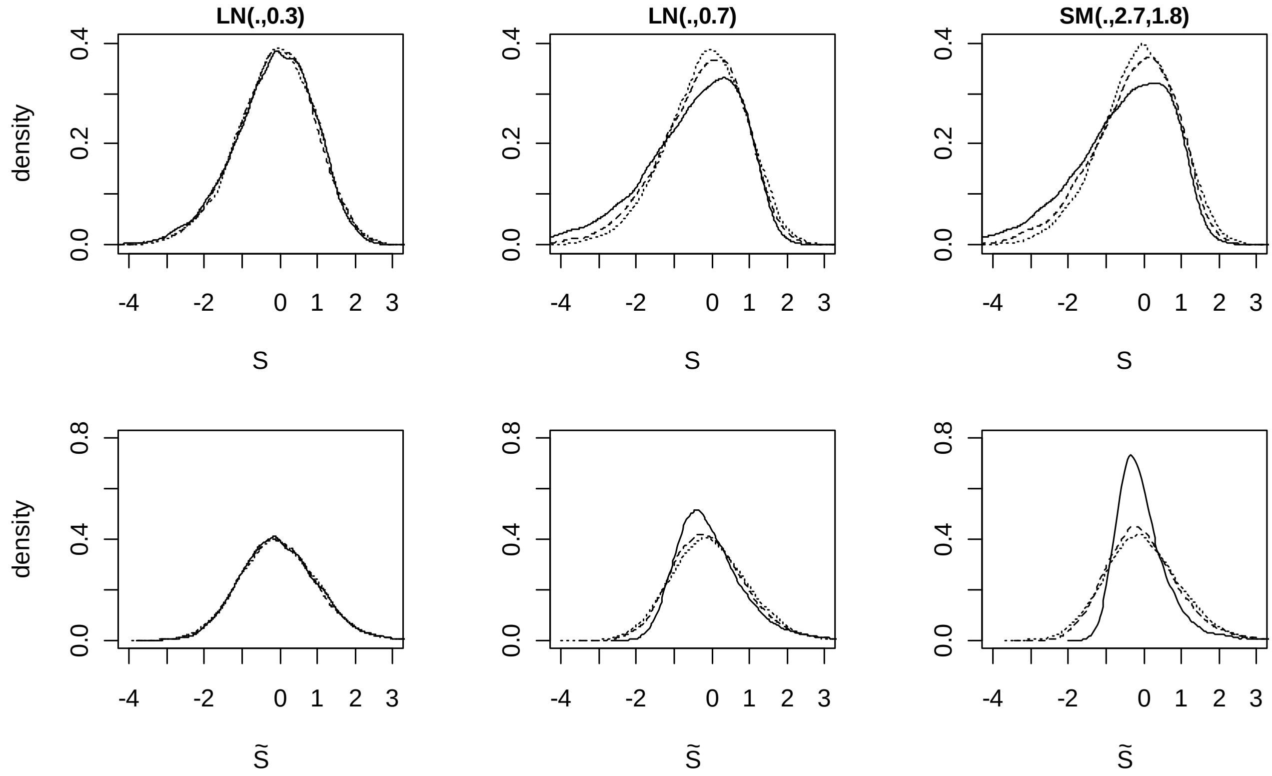

Figure 1: Density estimates for S_{500} and $\widetilde{S}_{500}$. Notes: $\alpha = 2$ (solid line), $\alpha = 1.05$ (broken line), $\alpha = 0.05$ (dotted line).

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 realisations $\widehat{I}$ which are *too small* (contrary to the natural intuition). 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 $\widehat{var}(\widehat{I})$ on which we focus on next.

We turn to the induced inferential problems by considering the actual coverage errors of standard 95% two-sided symmetric confidence intervals. The interplay between $\widehat{I}$, $\widehat{var}(\widehat{I})$, and the coverage errors is examined by simply plotting the $(\widehat{I}, \widehat{var}(\widehat{I}))$ pairs, and by identifying those pairs which are associated with a coverage error. It turns out that both the $LN\,(.,0.7)$ and the SM cases yield qualitatively similar results. In particular, compared to nominal coverage error rate of 5%, the actual total coverage error rate in the lognormal case 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. This is the flip-side of the substantial *left* tail and the heavy skewness of the density of S_{500}. These wrong confidence limits are associated with particularly *low* realisations of both $\widehat{I}$ and $\widehat{var}(\widehat{I})$, which is depicted in the $\widehat{I}$ vs. $n \times \widehat{var}(\widehat{I})$ plot of Figure 2. Given that almost all coverage errors are right rejections, we restrict the depicted range of $\widehat{I}$, and re-label those $(\widehat{I}, n \times \widehat{var}(\widehat{I}))$ 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 $n \times var(\widehat{I})$ exceeds the depicted range.[2]

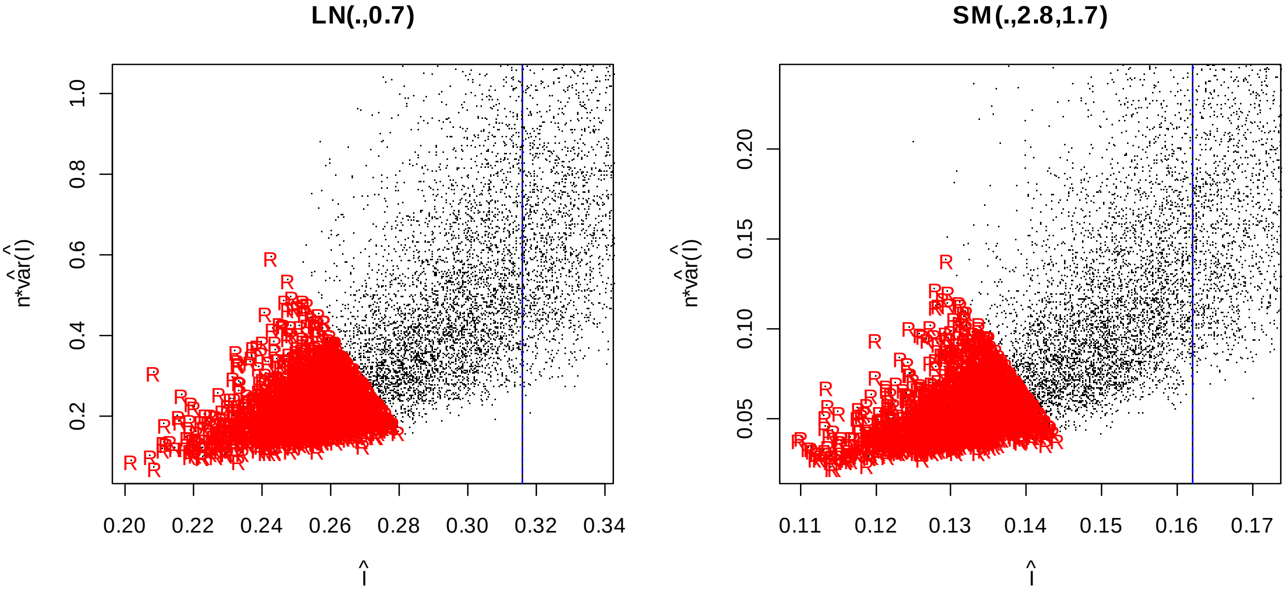

Figure 2: Coverage errors in $\widehat{I}$ vs. $n \times \widehat{var}(\widehat{I})$ plots. Notes: the vertical line corresponds to the population value of I, pairs labelled R correspond to coverage errors on the right of standard 95% two-sided symmetric confidence intervals.

These results are at odds with observations made in the literature and thus constitute an important contribution. In particular, Cowell and Flachaire (2007), Davidson and Flachaire (2007) and Davidson (2010) follow the natural intuition for heavy-tailed income distributions: samples drawn from heavy-tailed income distributions are likely to contain “extremes” or “outliers”, and these extremes are conjectured to be the root cause of the inference problem since most inequality measures are not robust to such extremes. This intuition is incorrect since 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. Moreover, 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. Finally, Davidson and Flachaire (2007) attribute some inferential problems to the non-linearity of the inequality measure. The juxtaposition of the densities of studentised and standardised inequality measure suggests that the problem is the non-linearity of S_n, and in particular the systematic relationship between $\widehat{I}$ and $\widehat{var}(\widehat{I})$.

We proceed to examine the issues of 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 used below as diagnostic tools for the anatomy of the above inference problems.

[2] The population values $(I, 500 \times var(\widehat{I}))$ are (0.316, 1.313) in the lognormal case and (0.162, 0.65) in the SM case.

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

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

Since S_n studentised, 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.[3] In terms of the cumulant generating function of S_n, given by $\exp\left(K_{S_n}(s)\right) = E\left\{\exp\left(sS_n\right)\right\}$, the cumulant coefficients define the second and third order term in the approximation to K_{S_n}, i.e. we have

$$K_{S_n}(s) = \frac{1}{2}s^2 + n^{-1/2}\left(sk_{1,2} + \frac{1}{6}s^3k_{3,1}\right) + n^{-1}\left(\frac{1}{2}s^2k_{2,2} + \frac{1}{24}s^4k_{4,1}\right) + O\left(n^{-3/2}\right). \tag{5}$$

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 studentised and standardised inequality measures. In order to maintain readability of the exposition, these cumulant coefficients are stated explicitly in Appendix A below, since the resulting expressions are lengthy and involve expectations of products of certain mean-zero random variables.

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

$$\Pr\left\{S_n \le x\right\} = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + n^{-1}p_2(x)\phi(x) + O\left(n^{-3/2}\right) \tag{6}$$

with

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

The right hand side of equation (6) 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, its relation to the bootstrap, and in his Section 2.4 a statement of the regularity conditions for the validity of the expansion. 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 (i) X has a proper density function (implying that Cramér's condition is satisfied), and with $\alpha^* = \max\left\{2, \alpha+1, 2\alpha\right\}$ that (ii) $\mu_{3\alpha^*} < \infty$ for the first order expansion which includes the $O\left(n^{-1/2}\right)$ term and $\mu_{4\alpha^*} < \infty$ for the second order expansion which includes the $O\left(n^{-1}\right)$ term. If $\mu_{4\alpha^*} < \infty$ then the regularity condition of footnote 3 with $r = 4$ is satisfied. These moment conditions are satisfied in the lognormal distribution, in case of the SM the admissible parameters must satisfy $bc < (j+2)\alpha^*$ for the Edgeworth expansion of order j. For the standardised inequality measure $\widetilde{S}_n$, α^* appearing in the regularity conditions is replaced by $\max\{1, \alpha\}$.

In view of the inferential problems it can be advantageous to consider non-linear transformations of the inequality index. Asymptotic expansions for such transforms follow immediately after considering the relation between the transform and S_n. Hence the cumulant coefficients can be used again as diagnostic tools to assess whether the transform has had desirable effects. Alternatively, the cumulants coefficients for the transform can be used to inspire specific transformations. For instance, Schluter and van Garderen (2009) consider normalising transformations which seek to annihilate asymptotically the skewness coefficient of the transform. Alternatively, in the light of the role played by $\widehat{\sigma}^2$ we consider the merit of using variance-stabilising transforms below.

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

Denote the studentised transform by

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

where $t'(I) \neq 0$, and denote the cumulant coefficients of T_n by $\lambda_{i,j}$.

Proposition 1 *To $O_p\left(n^{-3/2}\right)$ 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 2 *The cumulant coefficients for T_n are*

$$\begin{aligned}
\lambda_{1,2} &= k_{1,2} - \frac{1}{2}\sigma\frac{t''(I_0)}{t'(I_0)} \\
\lambda_{2,2} &= k_{2,2} - \sigma\left(k_{3,1} + 3k_{1,2}\right)\frac{t''(I_0)}{t'(I_0)} + \sigma^2\left(\frac{15}{4}\left[\frac{t''(I_0)}{t'(I_0)}\right]^2 - 2\frac{t'''(I_0)}{t'(I_0)}\right) - \frac{t''(I_0)}{t'(I_0)}\frac{3}{2}c \\
\lambda_{3,1} &= k_{3,1} - 3\sigma\frac{t''(I_0)}{t'(I_0)} \\
\lambda_{4,1} &= k_{4,1} - 2\sigma\frac{t''(I_0)}{t'(I_0)}k_{5,1} - 12\sigma\frac{t''(I_0)}{t'(I_0)}k_{3,1} + 24\sigma^2\frac{t''(I_0)^2}{t'(I_0)^2} - 8\sigma^2\frac{t'''(I_0)}{t'(I_0)} - 21\frac{t''(I_0)}{t'(I_0)}c.
\end{aligned}$$

The term c, 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.

We proceed to examine quantitatively the cumulant coefficients for the distributions considered above.

				S_n				$\widetilde{S}_n$			
α	sd	I	σ	$k_{1,2}$	$k_{3,1}$	$k_{2,2}$	$k_{4,1}$	$\widetilde{k}_{1,2}$	$\widetilde{k}_{3,1}$	$\widetilde{k}_{2,2}$	$\widetilde{k}_{4,1}$
.05	.3	.045	.065	-2.32	-6.44	31.48	97.48	-0.73	3.11	-0.75	16.45
1.05	.3	.045	.068	-2.86	-8.59	53.05	160.13	-0.76	4.01	-1.36	33.59
2	.3	.047	.080	-3.98	-13.14	111.20	331.99	-0.77	6.13	-2.46	95.12
2	.1	.005	.007	-2.30	-6.37	30.37	95.19	-0.71	3.13	-0.67	16.47
2	.2	.020	.031	-2.87	-8.67	52.76	161.41	-0.74	4.13	-1.26	35.50
2	.4	.087	.168	-5.96	-21.08	263.86	743.25	-0.82	9.79	-4.70	290.80
2	.5	.14	.33	-9.51	-35.29	694.11	1666.81	-0.87	16.54	-8.82	1013.09
2	.6	.22	.61	-16.22	-62.09	2053.56	3080.70	-0.94	29.57	-16.51	4174.65

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

Table 1 reports the results for the lognormal case as both shape parameter sd of the income distribution and the sensitivity parameter α of the inequality index vary. These results are consistent with those reported in Section 3. In particular, for the moderate shape parameter value $sd = .3$, the values of the cumulant coefficients for the studentised index S_n are moderate too, leading to only moderate departures from normality when sample sizes of $n = 500$ are considered. The coefficients increase for fixed sd as α increases. As the shape parameter increases and the right tail of the income distribution decreases more slowly, all cumulant coefficients increase. For instance, for $\alpha = 2$ and $sd = .6$ the skewness coefficient is substantial. Note, however, that in this situation σ is substantially larger than I. It is therefore of interest to relate the cumulant coefficients to the value I. This is done in Figure 3, which is an alternative presentation of the results of Table 1. Clearly, all coefficients increase in magnitude as I (equivalently sd) increases, and for given I as α increases.

Turning to the the cumulant coefficients of the standardised inequality measure $\widetilde{S}_n$, both bias and skewness coefficients are substantially smaller in magnitude than for S_n. Consistent with Figure 1, the skewness coefficient $\widetilde{k}_{3,1}$ has now the opposite sign.

Next, we turn to the SM distribution, for which case the same qualitative conclusions hold. Recall that the tail index of this heavy-tailed distribution is bc, so the tail decreases more slowly, or becomes fatter, as bc decreases, and that for the Edgeworth expansion of order j for the distribution of $S_n$to exist that $\mu_{(j+2)\alpha^*}$ be finite. In particular, we consider $SM\,(.,b,4)$ for different values of b and α. Table 2 reports the results. As before, for given b, the cumulant coefficients for S_n increase in magnitude as α increases, and for given α the coefficients increase in magnitude as the tail of the income distribution decreases more slowly (equivalently b decreases or I increases). The coefficients for $\widetilde{S}_n$ are again smaller in magnitude compared to those of S_n, and the sign of the skewness coefficient has changed.

				S_n				$\widetilde{S}_n$			
α	b	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	5	.032	.050	-2.24	-6.53	28.46	91.51	-0.57	3.52	-0.58	23.09
2	4.6	.038	.058	-2.31	-6.73	30.35	93.06	-0.58	3.64	-0.68	26.79
2	4.2	.045	.069	-2.42	-7.14	34.09	97.56	-0.59	3.85	-0.84	33.53
2	4.01	.049	.076	-2.50	-7.45	36.93	101.20	-0.60	4.01	-0.94	38.61
2	3.5	.064	.10	-2.88	-8.85	NA	NA	-0.62	4.70	-1.37	65.42
2	3.01	.09	.14	-3.64	-11.82	NA	NA	-0.65	6.13	-2.17	146.39
.05	2	.20	.31	-2.28	-6.64	30.80	101.14	-0.62	3.32	-0.77	18.67
1.05	2	.18	.27	-3.62	-11.50	88.88	141.20	-0.76	5.64	-2.90	132.83

Table 2: Cumulant coefficients when $X \sim SM(.,b,4)$.

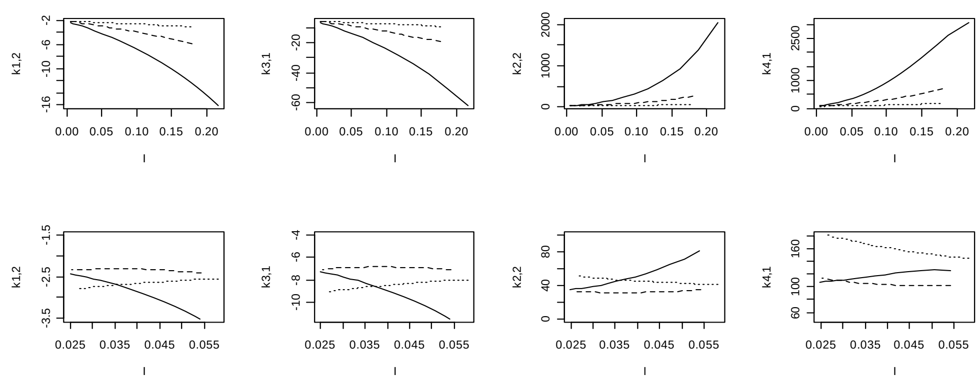

Figure 3: Cumulant coefficients as functions of I. Notes: The top panel depicts the lognormal case, the bottom panel to the SM case with $c = 4$; $\alpha = 2$ (solid line), $\alpha = 1.05$ (broken line), and $\alpha = .05$ (dotted line).

4.1 Digression: The Performance of Edgeworth Expansions and Saddlepoint Approximations

Given the availability of the cumulant coefficients it is of interest to investigate whether the two-term and three-term Edgeworth expansion provide approximations to the actual distribution of S_n which improve on the Gaussian first order approximation. However, the general shortcomings of the Edgeworth expansion are well known: The density expansion is not guaranteed to be positive, and oscillations are can sometimes be observed in the tails, an observation which also applies in our setting for 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 $O\left(n^{-1}\right)$. 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 $K(t) = nK_S\left(tn^{-1}\right)$, and denote its first and second derivatives by K' and K''. The saddlepoint approximation to the density of S_n at x is

$$g(x) = c\left(2\pi K''(s)\right)^{-1/2} \exp\left(K(s) - sx\right), \tag{7}$$

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

$$G(x) = \Phi\left(w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right), \tag{8}$$

with $w = sign(s)\left[2\left(sx - K(s)\right)\right]^{1/2}$ and $v = s\left[K''(s)\right]^{1/2}$. If we denote the pdf of S_n by pdf, then $pdf(x)/g(x) = 1 + O\left(n^{-1}\right)$, 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 (5). 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.

Performance evidence for the various approximations in the lognormal case is reported in Table 3, as both the sensitivity parameter α of the inequality index and the shape parameter sd changes. 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. Both Edgeworth expansion and the saddlepoint approximation do well by contrast when sd is fixed at the moderate value 0.3. For instance, in the case of $\alpha = 2$ when the exact CDF evaluates to .025, the normal approximation is .0069 while the saddlepoint approximation is .030, and turning to the 97.5% quantile, the normal approximation evaluates to .955 while the saddlepoint approximation is .97.

However, the performance of all approximations deteriorates as the tail of the income distribution becomes heavier, which is not unexpected given the sharp increase in the magnitudes of the moments and the cumulant coefficients. As sd increases to .4, while the two-term Edgeworth expansion does remarkably well in the left tail, the Edgeworth density in the extreme right tail does become negative. The extent of the oscillation in the tail becomes more severe as sd increases further.

Similar qualitative and quantitative conclusions follow for the SM distribution, which is not surprising given the similarities of the results Tables 1 and 2 for I, σ^2, and the associated cumulant coefficients for the chosen distributional parameters. For instance, the $LN(, .3)$ and the $SM(., 4.2, 4)$ distributions yield similar values for I and σ^2, as well as for bias and skewness coefficients. It is therefore not surprising that the performance of the Edgeworth and saddlepoint approximations in these two cases are very similar. Precise details are therefore not reported for the sake of brevity. To summarise, both Edgeworth and saddlepoint approximations perform well for moderate parameter values but their performance deteriorates when the speed of tail decay of the income distribution becomes slower. This suggests that it might be preferable to work with suitably transformed inequality measures rather than to seek improved approximations to the actual distribution of the measure.

CDF	$\alpha = .05$, $sd = .3$ norm	EE(2)	EE(3)	saddle	$\alpha = 1.05$, $sd = .3$ norm	EE(2)	EE(3)	saddle	$\alpha = 2$, $sd = .3$ norm	EE(2)	EE(3)	saddle
0.01	0.00	0.02	0.04	0.05	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.04
0.10	0.02	0.06	0.12	0.14	0.00	0.01	0.04	0.07	0.00	0.03	0.10	0.19
1.00	0.38	0.83	1.09	1.09	0.20	0.57	0.89	0.92	0.13	0.55	1.12	1.23
2.50	1.29	2.27	2.65	2.63	0.99	2.08	2.66	2.64	0.69	1.99	3.04	3.03
5.00	3.04	4.59	5.04	5.02	2.81	4.73	5.44	5.40	2.06	4.49	5.86	5.75
10.00	7.71	9.93	10.40	10.40	6.97	9.70	10.47	10.46	5.95	9.72	11.29	11.14
25.00	22.43	24.93	25.30	25.36	21.37	24.41	25.07	25.16	20.41	24.58	26.07	26.15
50.00	47.76	49.99	50.02	50.12	46.75	49.31	49.39	49.59	46.73	49.94	50.15	50.67
75.00	72.70	75.14	74.82	74.95	71.62	74.50	73.96	74.23	71.62	75.41	74.19	74.88
90.00	87.94	90.39	89.94	90.02	87.09	90.16	89.40	89.55	87.05	91.37	89.76	89.88
95.00	93.15	95.29	94.82	94.83	92.50	95.29	94.51	94.49	92.64	96.63	95.04	94.72
97.50	96.16	97.88	97.42	97.39	95.62	97.95	97.19	97.08	95.52	98.94	97.42	96.96
99.00	98.09	99.31	98.90	98.85	97.75	99.47	98.77	98.65	97.78	100.30	98.91	98.56
99.90	99.73	100.09	99.86	99.88	99.56	100.20	99.77	99.78	99.60	100.50	99.63	99.74
99.99	99.96	100.05	99.96	99.98	99.86	100.10	99.88	99.94	99.81	100.40	99.70	99.87

CDF	$\alpha = 2$, $sd = .1$ norm	EE(2)	EE(3)	saddle	$\alpha = 2$, $sd = .2$ norm	EE(2)	EE(3)	saddle	$\alpha = 2$, $sd = .4$ norm	EE(2)	EE(3)	saddle
0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.05
0.10	0.01	0.03	0.07	0.09	0.01	0.03	0.08	0.12	0.00	0.00	0.03	0.22
1.00	0.27	0.61	0.83	0.84	0.31	0.83	1.22	1.24	0.06	0.40	1.27	1.72
2.50	1.33	2.31	2.68	2.67	1.16	2.37	2.97	2.95	0.42	1.91	3.97	4.09
5.00	3.29	4.87	5.31	5.29	2.76	4.67	5.38	5.34	1.47	4.65	7.55	7.29
10.00	8.24	10.48	10.93	10.94	6.70	9.41	10.18	10.16	4.97	10.47	14.03	13.41
25.00	23.12	25.59	25.93	25.99	21.01	24.05	24.72	24.81	19.15	25.39	29.33	28.79
50.00	48.51	50.72	50.74	50.83	46.67	49.22	49.31	49.51	46.34	50.74	51.37	52.61
75.00	72.41	74.82	74.53	74.65	72.12	75.00	74.46	74.73	71.48	76.92	73.68	75.21
90.00	87.93	90.35	89.92	90.00	86.96	90.05	89.30	89.45	86.32	92.84	88.83	88.55
95.00	93.23	95.34	94.89	94.90	92.91	95.66	94.89	94.87	91.76	98.03	94.20	93.09
97.50	95.99	97.72	97.28	97.25	95.59	97.94	97.19	97.08	94.80	100.40	96.80	95.53
99.00	98.20	99.37	98.98	98.93	97.70	99.45	98.76	98.63	97.01	101.50	98.23	97.28
99.90	99.62	100.10	99.81	99.82	99.45	100.20	99.73	99.73	99.38	101.30	98.96	99.25
99.99	99.91	100.10	99.94	99.97	99.83	100.20	99.87	99.92	99.76	100.80	99.10	99.63

Table 3: Performance evidence for normal Edgeworth and saddlepoint approximations. Notes: Income are generated by $LN\left(.,sd\right)$, and all CDFs*100. CDF is the "exact" CDF based on 10,000 replications of S_{500}, all approximations are evaluated at the quantiles determined by the exact CDF, normal is the normal CDF, "EE" refers to the two and three-term Edgeworth expansions given by (6), "saddle" is the approximation to the saddlepoint approximation given by (8).

4.2 Variance Stabilising Transforms

Our results have suggested that an important role for the inference problems is played by the estimated variance of the inequality measure and the systematic relation between $\widehat{I}$ and $\widehat{var}(\widehat{I})$. We proceed to examine this relation further. Figure 2 suggests that for the considered income distributions, the relation could be approximately exponential. This is confirmed in Figure 4 Column 1, which plots $\log(\widehat{var}(\widehat{I}))$ vs. $\widehat{I}$, and further depicts a non-parametric estimate based on smooth splines, which is approximately linear. This suggests the application of a variance stabilising transform. Consider

$$h\left(\widehat{I}\right)=\int^{t}\frac{d\widehat{I}}{\left[var\left(\widehat{I}\right)\right]^{1/2}}. \tag{9}$$

Then h is such that, by the delta method, $var(h(\widehat{I}))=1$ asymptotically. If $\log(\widehat{var}(\widehat{I}))=\gamma_1+\gamma_2\widehat{I}+error$, this suggests the transform

$$t\left(I\right)=-\left(\frac{2}{\gamma_2}e^{-\gamma_1/2}\right)\exp\left(-\frac{\gamma_2}{2}I\right). \tag{10}$$

The cumulant coefficients for this transform follow immediately from Corollary 2 with $t''\left(I\right)/t'\left(I\right)=-\gamma_2/2$ and $t'''\left(I\right)/t'\left(I\right)=\left(\gamma_2/2\right)^2$.

Lemma 3 *If the coefficients of the odd cumulants of* S_n *are negative, the even ones are positive and* $\gamma_2>0$*, then the transform (10) reduces both bias, skewness and kurtosis for sufficiently small* γ_2*.*[4]

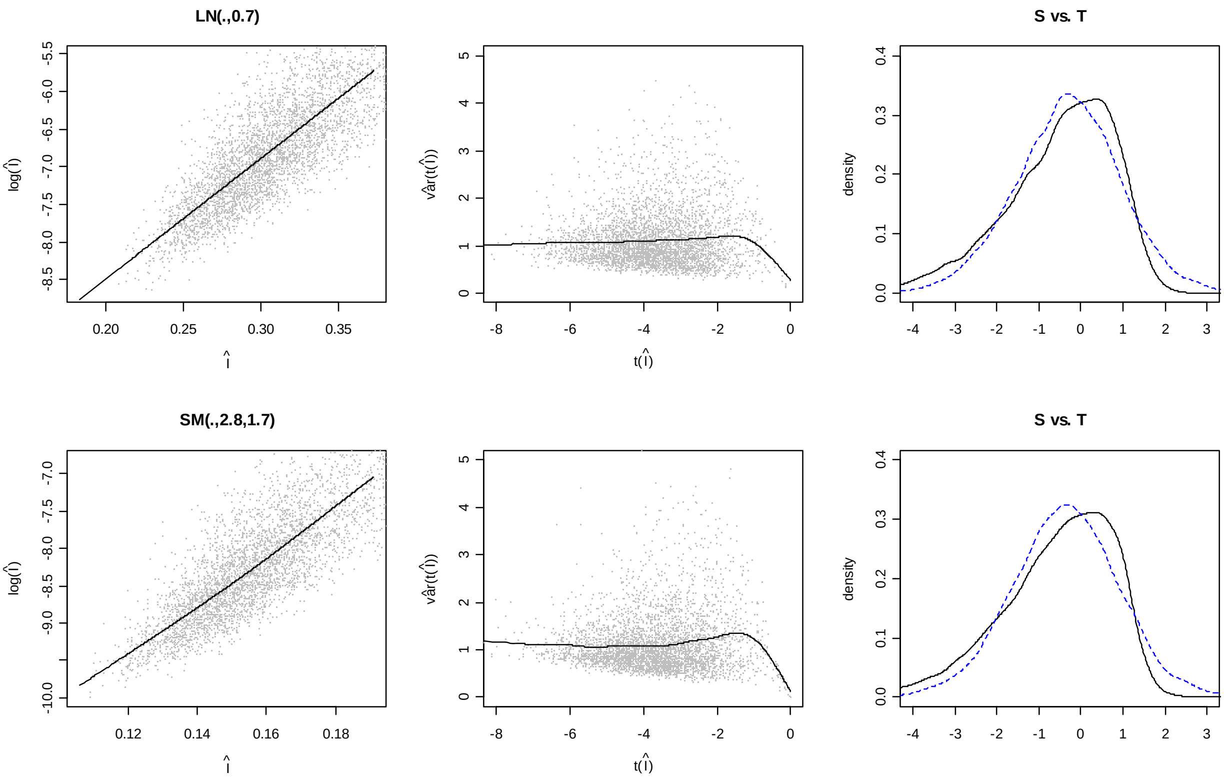

Figure 4: Aspects of variance stabilisation.

Figure 4 depicts the results of applying transform (10), in which the coefficients (γ_1,γ_2) were estimated by a simple regression of $\log(\widehat{var}(\widehat{I}))$ on $\widehat{I}$ using the simulated data.[5] In practice, the estimates can be obtained by a preliminary bootstrap. Column 2 of the Figure plots $\widehat{var}(t(\widehat{I}))$

[4]For instance, for the lognormal case $LN(.,.6)$ and $\alpha=2$ reported in the last row of Table 1, we have with γ_2 =22.2 obtained from regressing $\log(\widehat{var}(\widehat{I}))$ on $\widehat{I}$: $\lambda_{1,2}$ =-12.7, $\lambda_{3,1}$ =-41.2, and $\lambda_{2,2}$ =1536.17. Note that $\lambda_{2,2}$ is not expected to be zero, since neither $var(\widehat{I})$ is known exactly, nor is (10) the exact variance stabilising transform (9). $\lambda_{4,1}$ is not available since $k_{5,1}$ is not available. Hence these cumulant coefficients for T_n are smaller in magnitude than the corresponding coefficients for S_n.

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

on $t(\widehat{I})$, and also plots a non-parametric estimate based on smooth splines. It is evident that the transform does indeed stabilise the variance since the estimated curve is approximately equal to 1 except for a small number of observations in the sparse right tail.[6] Column 3 of the Figure shows simple kernel density estimates of the densities of the studentised 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. Qualitatively similar pictures obtain for different values of α and different income distribution parameters.

	$LN(.,.7)$			$SM(.,2.8,1.7)$		
	L	R	T	L	R	T
			$\alpha = 2$			
Normal approximation	0.47	13.8	14.27	0.27	15.18	15.45
Studentised bootstrap	1.09	8.03	9.12	0.93	9.49	10.42
Stud. var-stab bootstrap	0.70	6.85	7.54	0.48	8.32	8.80
			$\alpha = 1.05$			
Normal approximation	0.94	7.54	8.48	0.75	7.91	8.66
Studentised bootstrap	1.82	4.67	6.49	1.85	5.14	6.99
Stud. var-stab bootstrap	1.51	4.29	5.80	1.39	4.72	6.12
			$\alpha = 0.05$			
Normal approximation	1.40	4.90	6.30	1.22	5.08	6.31
Studentised bootstrap	2.36	3.18	5.54	2.31	3.26	5.57
Stud. var-stab bootstrap	2.16	3.02	5.18	2.09	3.05	5.13

Table 4: Coverage error rate of nominal 95 % two-sided confidence intervals. Notes: The nominal error rate is 5 %. "Stud. var-stab. bootstrap" is the studentised bootstrap for the variance stabilising transform given by eq. (10). "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 repetions, in each repetitions B=999 bootstrap sample were drawn, the sample size is n=500.

Table 4 consider the bootstrap evidence for the performance of the studentised variance stabilising transform, and benchmarks this against the normal (first order) approximation and the performance of the studentised bootstrap. The focus is again on the coverage error rates for two-sided confidence intervals with nominal error rate 5%. The table breaks 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). In each bootstrap simulation, samples of size 500 were drawn $B = 999$ times, and the experiments were repeated $R = 100,000$ times. The poor quality of the normal approximation has been discussed extensively above. Consistent with Davidson and Flachaire (2007), the studentised bootstrap improves on this but for large α the discrepancy between nominal and actual coverage behaviour is still substantial. For instance, in the SM case with α the actual error rate is still twice the nominal rate. The variance stabilising transform improves performance further. In line with the previous results, all methods improve as α falls, and as the right tail of the income distribution decays more quickly.

5 Conclusions

We have considered the well-known inference problem for inequality measures when incomes are generated by distributions with sufficiently slowly decaying tails. The received wisdom is that these are caused by the presence of extremes in the sample. We have shown that this is incorrect in so far as the coverage failures of the usual two-sided confidence intervals are associated with particularly low realisations of both $\widehat{I}$ and $\widehat{var}(\widehat{I})$. To understand better the separate and joint contributions of both

[6] In the SM case, the non-parametric curve falls below 1 around -.75, and about 1% of the simulated data lie to the right of this point; in the depicted LN case, the respective numbers are -.9 and 0.7%.

estimators, we have derived the bias, skewness and kurtosis coefficients for both the standardised and studentised inequality measures.

Given the availability of the cumulant coefficients, it is of interest to consider two and three-term Edgeworth expansions as well as saddlepoint approximations. These do lead to distributional improvements compared to the normal approximation, but performance worsens as the sensitivity parameter of the inequality index increases and the parameters of the income distribution changes so that its right tail decays more slowly.

As the diagnostic tools point to the prominent role of the estimated variance $\widehat{var}(\widehat{I})$ for poor inference, we have considered the merit of applying variance stabilising 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 considered in Davidson (2010), which is further discussed in Schluter (2010).

A Cumulant Coefficients for Inequality Measures

This section states the cumulant coefficients $(k_{1,2}, k_{3,1}, k_{2,2}, k_{4,1})$ for the studentised and standardised inequality measures. We associate the following mean-zero stochastic terms with the studentised inequality measure S_n :

$$\begin{aligned}
Y_{1,i} &= (X_i - \mu_1) \qquad (11) \\
Y_{2,i} &= \mu_1 (X_i^\alpha - \mu_\alpha) - \alpha\mu_\alpha (X_i - \mu_1) \\
Y_{3,i} &= (X_i^\alpha - \mu_\alpha) - \frac{1}{2}\alpha(\alpha+1)\mu_1^{-1}\mu_\alpha (X_i - \mu_1) \\
Y_{4,i} &= 2\left[\mu_1\mu_{2\alpha} - \alpha\mu_\alpha\mu_{\alpha+1} - (1-\alpha)^2\mu_1\mu_\alpha^2\right](X_i - \mu_1) + \alpha^2\mu_\alpha^2\left(X_i^2 - \mu_2\right) \\
&\quad +2\left[\alpha^2\mu_2\mu_\alpha - \alpha\mu_1\mu_{\alpha+1} - (1-\alpha)^2\mu_1^2\mu_\alpha\right](X_i^\alpha - \mu_\alpha) \\
&\quad -2\alpha\mu_1\mu_\alpha\left(X_i^{\alpha+1} - \mu_{\alpha+1}\right) + \mu_1^2\left(X_i^{2\alpha} - \mu_{2\alpha}\right) \\
Y_{5,i} &= -\frac{1}{6}(\alpha-1)\alpha(\alpha+1)\mu_1^{-2}\mu_\alpha (X_i - \mu_1) \\
Y_{6,i} &= Y_{1,i} \\
Y_{7,i} &= \left(\mu_{2\alpha} - (1-\alpha)^2\mu_\alpha^2\right)(X_i - \mu_1) - \left(2\alpha\mu_{\alpha+1} + (1-\alpha)^2 4\mu_1\mu_\alpha\right)(X_i^\alpha - \mu_\alpha) \\
&\quad -2\alpha\mu_\alpha\left(X_i^{\alpha+1} - \mu_{\alpha+1}\right) + 2\mu_1\left(X_i^{2\alpha} - \mu_{2\alpha}\right) \\
Y_{8.i} &= (X_i^\alpha - \mu_\alpha) \\
Y_{9,i} &= \left(\alpha^2\mu_2 - (1-\alpha)^2\mu_1^2\right)(X_i^\alpha - \mu_\alpha) + \alpha^2 2\mu_a\left(X_i^2 - \mu_2\right) - 2\alpha\mu_1\left(X_i^{\alpha+1} - \mu_{\alpha+1}\right)
\end{aligned}$$

Theorem 4 *The $n^{-1/2}$ bias and skewness cumulant coefficients for S_n are*

$$k_{1,2} = B_0^{-1/2} E(Y_{1,i}Y_{3,i}) - \frac{1}{2}B_0^{-3/2}E(Y_{2,i}Y_{4,i}) \qquad (12)$$

$$k_{3,1} = B_0^{-3/2}\left[E\left(Y_{2,i}^3\right) + 6E(Y_{1,i}Y_{2,i})E(Y_{2,i}Y_{3,i}) - 3E(Y_{2,i}Y_{4,i})\right]. \qquad (13)$$

The n^{-1} coefficients for S_n are

$$\begin{aligned}
k_{2,2} &= 2\left[B_0^{-1}E(Y_1Y_2Y_3) - \frac{1}{2}B_0^{-2}E(Y_2Y_2Y_4)\right] \qquad (14) \\
&\quad +2B_0^{-1}E\left[Y_{1,i}Y_{1,i}Y_{2,i}Y_{5,i}\right]_{3\ terms, 2+2\ components} \\
&\quad -2B_0^{-2}E\left[Y_{1,i}Y_{2,i}Y_{3,i}Y_{4,i}\right]_{3\ terms, 2+2\ components} \\
&\quad -B_0^{-2}E\left[Y_{2,i}Y_{2,i}Y_{6,i}Y_{7,i}\right]_{3\ terms, 2+2\ components} \\
&\quad -B_0^{-2}E\left[Y_{2,i}Y_{2,i}Y_{8,i}Y_{9,i}\right]_{3\ terms, 2+2\ components} \\
&\quad +B_0^{-3}E\left[Y_{2,i}Y_{2,i}Y_{4,i}Y_{4,i}\right]_{3\ terms, 2+2\ components} \\
&\quad +B_0^{-1}E\left[Y_{1,i}Y_{1,i}Y_{3,i}Y_{3,i}\right]_{3\ terms, 2+2\ components}
\end{aligned}$$

and

$$k_{4,1} = e_{4,1} - 4k_{1,2}k_{3,1} - 6k_{2,2} \qquad (15)$$

with

$$
\begin{aligned}
e_{4,1} &= B_0^{-2} E\left(Y_{2,i}^4\right) - 3 \\
&+4B_0^{-2} E\left(Y_{2,i}Y_{2,i}Y_{2,i}Y_{1,i}Y_{3,i}\right)_{10\ terms,2+3\ components} \\
&-2B_0^{-3} E\left(Y_{2,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{4,i}\right)_{10\ terms,2+3\ components} \\
&+4B_0^{-2} E\left(Y_{1,i}Y_{1,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{5,i}\right)_{15\ terms,2+2+2\ components} \\
&-8B_0^{-3} E\left(Y_{1,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{3,i}Y_{4,i}\right)_{15\ terms,2+2+2\ components} \\
&-2B_0^{-3} E\left(Y_{2,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{6,i}Y_{7,i}\right)_{15\ terms,2+2+2\ components} \\
&-2B_0^{-3} E\left(Y_{2,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{8,i}Y_{9,i}\right)_{15\ terms,2+2+2\ components} \\
&+3B_0^{-4} E\left(Y_{2,i}Y_{2,i}Y_{2,i}Y_{2,i}Y_{4,i}Y_{4,i}\right)_{15\ terms,2+2+2\ components} \\
&+6B_0^{-2} E\left(Y_{1,i}Y_{1,i}Y_{2,i}Y_{2,i}Y_{3,i}Y_{3,i}\right)_{15\ terms,2+2+2\ components}
\end{aligned}
$$

[7]As in Schluter and van Garderen (2009), we have used the shorthand notation $E\left[Y_AY_BY_CY_D\right]_{3\text{ terms},2+2\text{ components}} := E\left(Y_AY_B\right)E\left(Y_CY_D\right)+E\left(Y_AY_C\right)E\left(Y_BY_D\right)+E\left(Y_AY_D\right)E\left(Y_BY_C\right)$ etc. The second subscript indicates (i) how many expectations are to be multiplied, and (ii) how many random variables enter each expectation, the first subscript ("terms") refers to the total number of unordered permutations of these terms. All of the enumerated expectations are easy to compute[8] but expositionally prohibitively expensive to write out explicitly in terms of the underlying population moments $\mu_1, .., \mu_{4\alpha}$. Nor would these expressions be analytically insightful for most purposes. Some analytical insight is possible though. For instance, the inequality measure I is scale invariant, so is σ and hence S_n and so are the cumulant coefficients $(k_{1,2}, k_{3,1}, k_{2,2}, k_{4,1})$. It can be verified that the cumulant coefficients stated in Theorem 4 indeed satisfy this property. Theorem 4 extends the results of Schluter and van Garderen (2009) to include the coefficients $k_{2,2}$ and $k_{4,1}$.

Turning to the standardised inequality measure, we have the following result:

Theorem 5 *The cumulant coefficients for the standardised inequality measure* $\widetilde{S}_n = n^{1/2}\left(\widehat{I} - I\right)/\sigma$ *are given by Theorem 4 with the stochastic terms replaced by* $\widetilde{Y}_{1,i} = Y_{1,i}$, $\widetilde{Y}_{2,i} = Y_{2,i}$, $\widetilde{Y}_{4,i} = \widetilde{Y}_{6,i} = \widetilde{Y}_{7,i} = \widetilde{Y}_{8,i} = \widetilde{Y}_{9,i} = 0$ *and*

$$
\begin{aligned}
\widetilde{Y}_{3,i} &= \frac{1}{2}\alpha(\alpha+1)\mu_\alpha\mu_1^{-1}(X_i - \mu_1) - \alpha(X_i^\alpha - \mu_\alpha) \\
\widetilde{Y}_{5,i} &= \frac{1}{2}\alpha(\alpha+1)\mu_1^{-1}(X_i^\alpha - \mu_\alpha) - \frac{1}{6}\alpha(\alpha+1)(\alpha+2)\mu_\alpha\mu_1^{-2}(X_i - \mu_1).
\end{aligned}
$$

[7]Of course, the validity of the stated cumulant coefficients has been tested extensively using simulation studies. Details are available on request.

[8]The R code is available from the author.

B Proofs and Derivations

Proof of Proposition 1.

Expanding $t(I)$ about I_0 to third order, factorising $t'(I_0)$ and using the definition of S, and expanding $[t'(I)]^{-1}$ to second order similarly yields

$$\begin{aligned} T &= \left[1 - n^{-1/2}\widehat{\sigma}\frac{t''(I_0)}{t'(I_0)} + \left(\frac{t''(I_0)^2}{t'(I_0)^3} - \frac{1}{2}\frac{t'''(I_0)}{t'(I_0)^2}\right)n^{-1}\widehat{\sigma}^2 S^3\right] \times \\ & \left[S + \frac{1}{2}n^{-1/2}\widehat{\sigma}\frac{t''(I_0)}{t'(I_0)}S^2 + \frac{1}{6}n^{-1}\widehat{\sigma}^2\frac{t'''(I_0)}{t'(I_0)}S^3\right] + O_p\left(n^{-3/2}\right). \end{aligned}$$

The claim follows after multiplying out and collecting terms of the same order.

Proof of Corollary 2. Recall that $E\left(S^2\right) = 1 + n^{-1}k_{2,2} + O\left(n^{-3/2}\right)$, $E\left(S^3\right) = n^{-1/2}e_{3,1} + O\left(n^{-3/2}\right)$, $E\left(S^4\right) = 3 + n^{-1}e_{4,2} + O\left(n^{-3/2}\right)$, $E\left(S^5\right) = n^{-1/2}e_{5,1} = n^{-1/2}\left(k_{5,1} + 10k_{3,1} + 15k_{1,2}\right)$, and $E\left(S^6\right) = 15 + O\left(n^{-1}\right)$ with $e_{3,1} = k_{3,1} + 3k_{1,2}$ and $k_{4,2} = e_{4,1} - 4k_{1,2}k_{3,1} - 6k_{2,2}$. From Proposition 1 we have using $\widehat{\sigma} = \sigma + O_p\left(n^{-1/2}\right)$

$$T = S - \frac{1}{2}n^{-1/2}\widehat{\sigma}\frac{t''(I_0)}{t'(I_0)}S^2 + n^{-1}\sigma^2\left[\frac{1}{2}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{1}{3}\frac{t'''(I_0)}{t'(I_0)}\right]S^3. \tag{16}$$

(i) Derivation of $\lambda_{1,2}$. Take expectations of (16) and use $\widehat{\sigma} = \sigma + O_p\left(n^{-1/2}\right)$ to obtain

$$\begin{aligned} E(T) &= E(S) - \frac{1}{2}n^{-1/2}\sigma\frac{t''(I_0)}{t'(I_0)}E\left(S^2\right) + O\left(n^{-1}\right) \\ &= n^{-1/2}k_{1,2} - \frac{1}{2}n^{-1/2}\sigma\frac{t''(I_0)}{t'(I_0)} + O\left(n^{-1}\right). \end{aligned}$$

Therefore $\lambda_{1,2} = k_{1,2} - \frac{1}{2}n^{-1/2}\sigma\frac{t''(I_0)}{t'(I_0)}$.

(ii) Derivation of $\lambda_{2,2}$. Square (16), take expectations and ignore higher order terms to obtain

$$E\left(T^2\right) = E\left(S^2\right) - n^{-1/2}\frac{t''(I_0)}{t'(I_0)}E\left(S^3\widehat{\sigma}\right) + n^{-1}\sigma^2\left(\frac{5}{4}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{2}{3}\frac{t'''(I_0)}{t'(I_0)}\right)E\left(S^4\right) + O\left(n^{-3/2}\right).$$

The expression includes the term $E\left(S^3\widehat{\sigma}\right)$ so that the estimation error of order $n^{-1/2}$ for the standard deviation needs to be accounted for. Specifically, let

$$\widehat{\sigma} = \sigma + n^{-1}\sum_i W_i + O_p\left(n^{-1}\right), \tag{17}$$

where W_i are mean-zero random variables. This expression obtains using the methods employed in the proof of theorem 4. In particular, recall that $\widehat{\sigma} = \left(\alpha^2 - \alpha\right)^{-1} m_1^{-(\alpha+1)}\widehat{B}_0^{1/2}$, centre the sample moments, Taylor expand both $m_1^{-(\alpha+1)}$ and $\widehat{B}_0^{1/2}$, and multiply out. From this proof it also follows that $S^3 = S_0^3 + 3S_0^2 S_{1/2} + O_p\left(n^{-3/2}\right)$ with $S_0 = B_0^{-1/2}n^{-1/2}\sum_i Y_{2,i}$ and $Y_{2,i}$ defined in (11). Then

$$n^{-1/2}E\left(S^3\widehat{\sigma}\right) = n^{-1/2}E\left(S^3\sigma\right) + n^{-1}c_2 + O\left(n^{-3/2}\right)$$

with $c_2 = B_0^{-3/2}E\left(Y_2Y_2Y_2W\right)_{3\text{ terms},2+2} = B_0^{-3/2} \times 3E\left(Y_2^2\right)E\left(Y_2W\right) = 3B_0^{-1/2}E\left(Y_2W\right)$ since $E\left(Y_2^2\right) = B_0$. Now let

$$c = B_0^{-1/2}E\left(Y_2W\right)$$

so that $c_2 = 3c$. The claim then follows after using the above expressions for the moments and hence the cumulant coefficients of S.

It remains to state the mean-zero random variable W explicitly. The expansion of $\widehat{B_0}$ yields $\widehat{B_0} = B_0 + n^{-1}\sum Z_i + O_p\left(n^{-1}\right)$, with $Z_i = \left[\mu_{2\alpha}2\mu_1 - 2\alpha\mu_\alpha\mu_{\alpha+1} - (1-\alpha)^2\mu_\alpha^2 2\mu_1\right](X_i - \mu_1) + \alpha^2\mu_\alpha^2\left(X_i^2 - \mu_2\right) + \left[2\alpha^2\mu_\alpha\mu_2 - 2\alpha\mu_1\mu_{\alpha+1} - (1-\alpha)^2\mu_1^2 2\mu_\alpha\right]\left(X_i^\alpha - \mu_\alpha\right) - 2\alpha\mu_1\mu_\alpha\left(X_i^{\alpha+1} - \mu_{\alpha+1}\right) + \mu_1^2\left(X_i^{2\alpha} - \mu_{2\alpha}\right)$. The Taylor expansions and multiplications finally yield $W_i = \left(\alpha^2 - \alpha\right)^{-1} B_0^{-1/2}\mu_1^{-(\alpha+1)} \times \left[\frac{1}{2}Z_i - (\alpha+1)\mu_1^{-1}B_0(X_i - \mu_1)\right]$. It can be verified that the term c is indeed scale-invariant.

(iii) Derivation of $\lambda_{3,1}$. Cube (16), use $\widehat{\sigma} = \sigma + O_p\left(n^{-1/2}\right)$, take expectations and ignore higher order terms to get

$$\begin{aligned} E\left(T^3\right) &= E\left(S^3\right) - n^{-1/2}\frac{3}{2}\sigma\frac{t''(I_0)}{t'(I_0)}E\left(S^4\right) + O\left(n^{-1}\right) \\ &= n^{-1/2}\left[e_{3,1} - \frac{9}{2}\sigma\frac{t''(I_0)}{t'(I_0)}\right] + O\left(n^{-1}\right). \end{aligned}$$

Hence $e_{3,1} - \frac{9}{2}\sigma\frac{t''(I_0)}{t'(I_0)} = \lambda_{3,1} + 3\lambda_{1,2}$, and solving for $\lambda_{3,1}$ using the result for $\lambda_{1,2}$ obtained in step (i) proves the claim.

(iv) Derivation of $\lambda_{4,1}$. Raise expression (16) to the fourth power, take expectations and ignore higher order terms to get

$$E\left(T^4\right) = E\left(S^4\right) - n^{-1/2}2\frac{t''(I_0)}{t'(I_0)}E\left(S^5\widehat{\sigma}\right) + n^{-1}\sigma^2\left(\frac{7}{2}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{4}{3}\frac{t'''(I_0)}{t'(I_0)}\right)E\left(S^6\right) + O_p\left(n^{-3/2}\right).$$

As in the proof of part (ii), use expansion (17) and $S^5 = S_0^5 + 5S_0^4S_{1/2} + O_p\left(n^{-3/2}\right)$. Then

$$\begin{aligned} E\left(T^4\right) = \; & E\left(S^4\right) - n^{-1/2}2\frac{t''(I_0)}{t'(I_0)}\sigma E\left(S^5\right) + n^{-1}\sigma^2\left(\frac{7}{2}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{4}{3}\frac{t'''(I_0)}{t'(I_0)}\right)E\left(S^6\right) \\ & - n^{-1}2\frac{t''(I_0)}{t'(I_0)} \times 15c \\ & + O_p\left(n^{-3/2}\right), \end{aligned}$$

since $B_0^{-5/2}n^{-7/2}E\left(\sum_i\sum_j\sum_k\sum_l\sum_m\sum_n Y_{2,i}Y_{2,j}Y_{2,k}Y_{2,l}Y_{2,m}W_n\right) = B_0^{-5/2}n^{-1/2}E\left(Y_2Y_2Y_2Y_2Y_2W\right)_{15\text{ terms},2+2+2} = 15B_0^{-1/2}n^{-1/2}E\left(Y_2W\right) = n^{-1}15c$ because $E\left(Y_2^2\right) = B_0$. Therefore

$$E\left(T^4\right) = 3 + n^{-1}\left[e_{4,1} - 2\sigma\frac{t''(I_0)}{t'(I_0)}e_{5,1} + \sigma^2 15\left(\frac{7}{2}\frac{t''(I_0)^2}{t'(I_0)^2} - \frac{4}{3}\frac{t'''(I_0)}{t'(I_0)}\right) - 30\frac{t''(I_0)}{t'(I_0)}c\right] + O_p\left(n^{-3/2}\right).$$

Denoting the coefficient of n^{-1} by γ we have $\lambda_{4,2} = \gamma - 4\lambda_{1,2}\lambda_{3,1} - 6\lambda_{2,2}$. The claim follows using the above expressions for $e_{4,1}$ and $e_{5,1}$, and the use of results for $\lambda_{1,2}$, $\lambda_{3,1}$, and $\lambda_{2,2}$ in steps (i)-(iii) after some simplifications.

Proof of Lemma 3. We have

$$\begin{aligned} \lambda_{1,2} &= k_{1,2} + \frac{1}{2}\sigma\frac{\gamma_2}{2} \\ \lambda_{3,1} &= k_{3,1} + 3\sigma\frac{\gamma_2}{2} \\ \lambda_{2,2} &= k_{2,2} + \sigma\frac{\gamma_2}{2}\left[k_{3,1} + 3k_{1,2} + \frac{7}{4}\sigma\frac{\gamma_2}{2}\right] + \frac{3}{4}\gamma_2 c \\ \lambda_{4,1} &= k_{4,1} + \sigma\gamma_2\left[k_{5,1} + 6k_{3,1} + 8\sigma\frac{\gamma_2}{2}\right] + \frac{21}{2}\gamma_2 c. \end{aligned}$$

With $\gamma_2 > 0$ and $(k_{1,2}, k_{3,1}, k_{5,1})$ all negative and $(k_{2,2}, k_{4,1})$ all positive, the statement follows for sufficiently small γ_2.

B.1 Proof of Theorem 4

$$\begin{aligned} S_n &= \sqrt{n}\left(\frac{\widehat{I}-I}{\widehat{\sigma}}\right) \\ &= \sqrt{n}\widehat{B}_0^{-1/2}\left[m_\alpha m_1 - \mu_1^{-\alpha}\mu_\alpha m_1^{\alpha+1}\right] \end{aligned}$$

Centering the sample moments, $m_\alpha = \mu_\alpha + n^{-1}\sum\left(X_i^\alpha - \mu_\alpha\right)$, and Taylor expanding the power functions yields to a stochastic expansion of the general form

$$S_n = S_{n,0} + S_{n,1/2} + S_{n,1} + O_p\left(n^{-3/2}\right)$$

where $S_{n,\beta}$ denotes a term that is $O_p\left(n^{-\beta}\right)$. These stochastic terms are defined in terms of the primitive mean-zero stochastic terms $Y_{1,i},..,Y_{9,i}$ given in (11) as follows:

$$\begin{aligned} S_{n,0} &= B_0^{-1/2}n^{-1/2}\sum_i Y_{2,i} \\ S_{n,1/2} &= B_0^{-1/2}n^{-3/2}\sum_i\sum_i Y_{1,i}Y_{3,i} - \frac{1}{2}B_0^{-3/2}n^{-3/2}\sum_i\sum_i Y_{2,i}Y_{4,i} \\ S_{n,1} &= B_0^{-1/2}n^{-5/2}\sum_i\sum_i\sum_i Y_{1,i}Y_{1,i}Y_{5,i} \\ &\quad -\frac{1}{2}B_0^{-3/2}n^{-5/2}\sum_i\sum_i\sum_i Y_{1,i}Y_{3,i}Y_{4,i} \\ &\quad -\frac{1}{2}B_0^{-3/2}n^{-5/2}\sum_i\sum_i\sum_i Y_{2,i}Y_{6,i}Y_{7,i} \\ &\quad -\frac{1}{2}B_0^{-3/2}n^{-5/2}\sum_i\sum_i\sum_i Y_{2,i}Y_{8,i}Y_{9,i} \\ &\quad +\frac{3}{8}B_0^{-5/2}n^{-5/2}\sum_i\sum_i\sum_i Y_{2,i}Y_{4,i}Y_{4,i} \end{aligned}$$

B.2 Moments and Cumulant Coefficients of Order $n^{-1/2}$

We have

$$K_1 = E\left(S_{n,1/2}\right),$$

hence it is immediate that

$$k_{1,2} = B_0^{-1/2}E\left(Y_{1,i}Y_{3,i}\right) - \frac{1}{2}B_0^{-3/2}E\left(Y_{2,i}Y_{4,i}\right). \tag{18}$$

Next, we have, noting $E\left(S_n^2\right) = 1 + O\left(n^{-1}\right)$, that

$$\begin{aligned} E\left(S_n^3\right) &= E\left(S_{n,0}^3\right) + 3E\left(S_{n,0}^2 S_{n,1/2}\right) + O\left(n^{-3/2}\right) \\ K_3 &= E\left(S_n^3\right) - 3E\left(S_n\right) + O\left(n^{-3/2}\right) \end{aligned}$$

since $E\left(Y_{2,i}^2\right) = B_0$ and it turns out that $E\left(S_{n,0}^2 S_{n,1}\right) = O\left(n^{-3/2}\right)$ and $E\left(S_{n,0}S_{n,1/2}^2\right) = O\left(n^{-3/2}\right)$. We have $E\left(S_{n,0}^3\right) = n^{-1/2}B_0^{-3/2}E\left(Y_{2,i}^3\right)$ and $E\left(3S_{n,0}^2 S_{n,1/2}\right) = n^{-1/2}3B_0^{-3/2}[E\left(Y_{2,i}^2\right)E\left(Y_{1,i}Y_{3,i}\right) + 2E\left(Y_{1,i}Y_{2,i}\right)E\left(Y_{2,i}Y_{3,i}\right) - \frac{3}{2}E\left(Y_{2,i}Y_{4,i}\right)]$ so

$$k_{3,1} = B_0^{-3/2}\left[E\left(Y_{2,i}^3\right) + 6E\left(Y_{1,i}Y_{2,i}\right)E\left(Y_{2,i}Y_{3,i}\right) - 3E\left(Y_{2,i}Y_{4,i}\right)\right]. \tag{19}$$

B.3 Moments and Cumulant Coefficients of Order n^{-1}

$$
\begin{aligned}
S_n^2 &= S_{n,0}^2 + 2S_{n,0}S_{n,1/2} + 2S_{n,0}S_{n,1} + S_{n,1/2}^2 + O_p\left(n^{-3/2}\right) \\
S_n^4 &= S_{n,0}^4 + 4S_{n,0}^3 S_{n,1/2} + 4S_{n,0}^3 S_{n,1} + 6S_{n,0}^2 S_{n,1/2}^2 + O_p\left(n^{-3/2}\right).
\end{aligned}
$$

Take expectations and consider the individual contributions. Note that $E\left(S_{n,0}^2\right) = 1$ since $E\left(Y_{2,i}^2\right) = B_0$.

We have $E\left(2S_{n,0}S_{n,1/2}\right) = n^{-1}2\left[B_0^{-1}E\left(Y_1Y_2Y_3\right) - \frac{1}{2}B_0^{-2}E\left(Y_2Y_2Y_4\right)\right]$. Next, we have

$$
\begin{aligned}
S_{n,1/2}^2 &= B_0^{-1}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{1,i}Y_{1,i}Y_{3,i}Y_{3,i} \\
&\quad -B_0^{-2}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{1,i}Y_{2,i}Y_{3,i}Y_{4,i} \\
&\quad +\frac{1}{4}B_0^{-3}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{2,i}Y_{2,i}Y_{4,i}Y_{4,i}
\end{aligned}
$$

The generic summands of the expectation are $E\left(\sum_i\sum_i\sum_i\sum_i Y_{A,i}Y_{B,i}Y_{C,i}Y_{D,i}\right)$, and the $O\left(n^{-1}\right)$ terms are the 3 distinct pairs, i.e.

$$E\left(Y_AY_B\right)E\left(Y_CY_D\right)+E\left(Y_AYC\right)E\left(Y_BY_D\right)+E\left(Y_AY_D\right)E\left(Y_BY_C\right) := E\left[Y_AY_BY_CY_D\right]_{3\text{ terms},2+2\text{ components}}$$

We use the notion on the RHS as a convenient shorthand (rather than tensor notation). Similarly,

$$
\begin{aligned}
2S_{n,0}S_{n,1} &= 2B_0^{-1}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{1,i}Y_{1,i}Y_{2,i}Y_{5,i} \\
&\quad -B_0^{-2}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{1,i}Y_{2,i}Y_{3,i}Y_{4,i} \\
&\quad -B_0^{-2}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{2,i}Y_{2,i}Y_{6,i}Y_{7,i} \\
&\quad -B_0^{-2}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{2,i}Y_{2,i}Y_{8,i}Y_{9,i} \\
&\quad +\frac{3}{4}B_0^{-3}n^{-3}\sum_i\sum_i\sum_i\sum_i Y_{2,i}Y_{2,i}Y_{4,i}Y_{4,i}
\end{aligned}
$$

Therefore

$$K_2 = 1 + n^{-1}k_{2,2} + O\left(n^{-2}\right)$$

with

$$
\begin{aligned}
k_{2,2} &= 2\left[B_0^{-1}E\left(Y_1Y_2Y_3\right) - \frac{1}{2}B_0^{-2}E\left(Y_2Y_2Y_4\right)\right] \\
&\quad +2B_0^{-1}E\left[Y_{1,i}Y_{1,i}Y_{2,i}Y_{5,i}\right]_{3\text{ terms},2+2\text{ components}} \\
&\quad -2B_0^{-2}E\left[Y_{1,i}Y_{2,i}Y_{3,i}Y_{4,i}\right]_{3\text{ terms},2+2\text{ components}} \\
&\quad -B_0^{-2}E\left[Y_{2,i}Y_{2,i}Y_{6,i}Y_{7,i}\right]_{3\text{ terms},2+2\text{ components}} \\
&\quad -B_0^{-2}E\left[Y_{2,i}Y_{2,i}Y_{8,i}Y_{9,i}\right]_{3\text{ terms},2+2\text{ components}} \\
&\quad +B_0^{-3}E\left[Y_{2,i}Y_{2,i}Y_{4,i}Y_{4,i}\right]_{3\text{ terms},2+2\text{ components}} \\
&\quad +B_0^{-1}E\left[Y_{1,i}Y_{1,i}Y_{3,i}Y_{3,i}\right]_{3\text{ terms},2+2\text{ components}}.
\end{aligned}
\tag{20}
$$

Next, consider the fourth moment of S_n, and proceed component by component. $S_{n,0}^4 = B_0^{-2}n^{-2}\sum_i\sum_i\sum_i\sum_i Y_{2,i}Y_{2,i}Y_{2,i}Y_{2,i}$.

Then $E\left(S_{n,0}^4\right) = B_0^{-2} E\left[Y_2 Y_2 Y_2 Y_2\right]_{3 \text{ terms},2+2 \text{ components}} + n^{-1} B_0^{-2} E\left(Y_{2,i}^4\right) = 3 + n^{-1} B_0^{-2} E\left(Y_{2,i}^4\right)$ since $E\left[Y_2 Y_2 Y_2 Y_2\right]_{3 \text{ terms},2+2 \text{ components}} = 3\left[E\left(Y_2^2\right)\right]^2$ and $E\left(Y_2^2\right) = B_0$

$$E\left(S_{n,0}^4\right) = 3 + n^{-1} B_0^{-2} E\left(Y_{2,i}^4\right)$$

$$\begin{aligned} E\left(4S_{n,0}^3 S_{n,1/2}\right) &= 4B_0^{-2} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{1,i} Y_{3,i}\right)_{10 \text{ terms},2+3 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{4,i}\right)_{10 \text{ terms},2+3 \text{ components}} \end{aligned}$$

$$\begin{aligned} 4S_{n,0}^3 S_{n,1} &= 4B_0^{-2} E\left(Y_{1,i} Y_{1,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{5,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{1,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{3,i} Y_{4,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{6,i} Y_{7,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{8,i} Y_{9,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad +\frac{3}{2} B_0^{-4} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{4,i} Y_{4,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \end{aligned}$$

$$\begin{aligned} E\left(6S_{n,0}^2 S_{n,1/2}^2\right) &= 6B_0^{-2} E\left(Y_{1,i} Y_{1,i} Y_{2,i} Y_{2,i} Y_{3,i} Y_{3,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad -6B_0^{-3} E\left(Y_{1,i} Y_{2,i} Y_{2,i} Y_{3,i} Y_{2,i} Y_{4,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \\ &\quad +\frac{3}{2} B_0^{-4} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{4,i} Y_{4,i}\right)_{10 \text{ terms},2+2+2 \text{ components}} \end{aligned}$$

Therefore

$$\begin{aligned} E\left(S_n\right)^4 &= E\left(S_{n,0}^4\right) + 4E\left(S_{n,0}^3 S_{n,1/2}\right) + 4E\left(S_{n,0}^3 S_{n,1}\right) + 6E\left(S_{n,0}^2 S_{n,1/2}^2\right) + O\left(n^{-3/2}\right) \\ &= 3 + n^{-1} e_{4,1} + O\left(n^{-2}\right) \end{aligned}$$

with

$$\begin{aligned} e_{4,1} &= B_0^{-2} E\left(Y_{2,i}^4\right) - 3 \\ &\quad +4B_0^{-2} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{1,i} Y_{3,i}\right)_{10 \text{ terms},2+3 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{4,i}\right)_{10 \text{ terms},2+3 \text{ components}} \\ &\quad +4B_0^{-2} E\left(Y_{1,i} Y_{1,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{5,i}\right)_{15 \text{ terms},2+2+2 \text{ components}} \\ &\quad -8B_0^{-3} E\left(Y_{1,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{3,i} Y_{4,i}\right)_{15 \text{ terms},2+2+2 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{6,i} Y_{7,i}\right)_{15 \text{ terms},2+2+2 \text{ components}} \\ &\quad -2B_0^{-3} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{8,i} Y_{9,i}\right)_{15 \text{ terms},2+2+2 \text{ components}} \\ &\quad +3B_0^{-4} E\left(Y_{2,i} Y_{2,i} Y_{2,i} Y_{2,i} Y_{4,i} Y_{4,i}\right)_{15 \text{ terms},2+2+2 \text{ components}} \\ &\quad +6B_0^{-2} E\left(Y_{1,i} Y_{1,i} Y_{2,i} Y_{2,i} Y_{3,i} Y_{3,i}\right)_{15 \text{ terms},2+2+2 \text{ components}}. \end{aligned}$$

Finally since $K_4 = E(S^4) - 4E(S^3)E(S) - 3\left(E(S^2)\right)^2 + 12E(S^2)\left(E(S)\right)^2 - 6\left(E(S)\right)^4$ it follows that with $K_4 = n^{-1}k_{4,1} + O(n^{-3/2})$, we have

$$\begin{aligned} k_{4,1} &= e_{4,1} - 4k_{1,2}\left[k_{3,1} + 3k_{1,2}\right] - 6k_{2,2} + 12\left(k_{1,2}\right)^2 \\ &= e_{4,1} - 4k_{1,2}k_{3,1} - 6k_{2,2}. \end{aligned} \tag{21}$$

References

COWELL, F.A. (1989), “Sampling variance and decomposable inequality measures”, *Journal of Econometrics*, 42, 27-41.

COWELL, F.A. (2000), “Measurement of inequality”, in: Atkinson, A.B., Bourguignon, F. (Eds.), *Handbook of Income Distribution*. North Holland, Amsterdam (Chapter 2).

COWELL, F. A. AND E. FLACHAIRE (2007), “Income Distribution and Inequality Measurement: The Problem of Extreme Values”, *Journal of Econometrics*, 141, 1044-1072.

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*, Vol. 25, No. 4., 631-650.

DAVIDSON, R. AND E. FLACHAIRE (2007), “Asymptotic and bootstrap inference for inequality and poverty measures”, *Journal of Econometrics*, 141, 141-166.

DAVIDSON, R. (2010), “Inference for Inequality Measures for Heavy-Tailed Distributions”, paper presented in the Econometrics Journal sponsered special session on the "econometrics of inequality" of the Royal Economics Society conference 2010.

EASTON G. S. AND E. RONCHETTI (1986), “General saddlepoint approximations with applications to L-statistics”, *Journal of the American Statistical Association*, Vol. 81, No. 394, 420-430.

HALL, P. (1992), *The Bootstrap and Edgeworth Expansions*, Springer Verlag.

REID, N (1988), “Saddlepoint methods and statistical inference”, *Statistical Science*, Vol. 3, No. 2., 213-227.

SCHLUTER, C. AND M. TREDE (2002), “Tails of Lorenz curves”, *Journal of Econometrics*, 109, 151-166.

SCHLUTER, C. AND K.J. VAN GARDEREN (2009), ”Edgeworth Expansions and Normalizing Transforms for Inequality Measures”, *Journal of Econometrics*, 50, 16-29.

SCHLUTER, C. (2010), ”Discussion of Donald et al. (2010) and Davidson (2010)”, presented in the Econometrics Journal sponsered special session on the "econometrics of inequality" of the Royal Economics Society conference 2010.