# Edgeworth expansions and normalizing transforms for inequality measures

Kees Jan van Garderen, Christian Schluter


Edgeworth expansions and normalizing transforms for inequality measures


Kees Jan van Garderen, Christian Schluter



PII: S0304-4076(09)00029-3
DOI: 10.1016/j.jeconom.2008.12.022
Reference: ECONOM 3161

To appear in: *Journal of Econometrics*

Received date: 10 November 2006
Revised date: 17 December 2008
Accepted date: 24 December 2008

Please cite this article as: van Garderen, K.J., Schluter, C., Edgeworth expansions and normalizing transforms for inequality measures. *Journal of Econometrics* (2009), doi:10.1016/j.jeconom.2008.12.022

# Edgeworth Expansions and Normalizing Transforms for Inequality Measures*

Kees Jan van Garderen† University of Amsterdam

Christian Schluter‡ University of Southampton

December 2008

**Abstract**

Finite sample distributions of studentized inequality measures differ substantially from their asymptotic normal distribution in terms of location and skewness. We study these aspects formally by deriving the second order expansion of the first and third cumulant of the studentized inequality measure. We state distribution-free expressions for the bias and skewness coefficients. In the second part we improve over first-order theory by deriving Edgeworth expansions and normalizing transforms. These normalizing transforms are designed to eliminate the second order term in the distributional expansion of the studentized transform and converge to the Gaussian limit at rate $O(n^{-1})$. This leads to improved confidence intervals and applying a subsequent bootstrap leads to a further improvement to order $O(n^{-3/2})$. We illustrate our procedure with an application to regional inequality measurement in Côte d'Ivoire.

**Keywords:** Generalized Entropy inequality measures, higher order expansions, normalizing transformations.

**JEL classification:** C10, C14, D31, D63, I32

*Funding from the ESRC under grant R000223640 is gratefully acknowledged. The research of van Garderen has been supported by a research fellowship of the Royal Dutch Academy of Arts and Sciences. Thanks to our referees for constructive comments which helped to improve the paper substantially. Helpful comments from Andrew Chesher, Russell Davidson, Grant Hillier, Patrick Marsh, and Richard Smith are gratefully acknowledged.

†Corresponding author. Department of Quantitative Economics, Faculty of Economics and Business, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands, Tel. +31-20-525 4220, Email: K.J.vanGarderen@uva.nl, www.ase.uva.nl/pp/kjvangarderen

‡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

Most attention in the statistical literature on inequality measures has focused on the asymptotic properties of their estimators (see e.g. Cowell, 1989, Thistle, 1990, Davidson and Duclos, 1997). Their finite sample properties have rarely been considered. Exceptions are, for instance, Mills and Zandvakili (1997), Biewen (2001) and Davidson and Flachaire (2007) who investigate bootstrap inference, and Maasoumi and Theil (1979) who develop small-sigma approximations. We consider finite sample properties of Generalized Entropy (GE) indices of inequality, which constitute a leading class of inequality indices since it is the only class that simultaneously satisfies the key properties of anonymity and scale independence, and the principles of transfer, decomposability, and population (see e.g. Maasoumi 1999 or Cowell 2000). Studies of industrial concentration, or income studies after decomposition into population subgroups, or cross country or regional comparisons, can easily yield samples of the sizes considered here.

We show that even for relatively large samples standard first order theory provides poor guidance for actual behavior. The distribution of the studentized inequality measure differs substantially from the Gaussian limit in terms of location and skewness. We study the bias (average deviation from zero) and skewness formally in the first part of this paper by deriving the second order expansions of the first three cumulants. We refer to the resulting coefficients of $n^{-1/2}$ as bias and skewness coefficients. This is the first key contribution of this paper. Moreover, the bias and skewness coefficients can be estimated non-parametrically using sample moments without affecting the order of the approximation. In all applications considered below, it is shown that the bias and skewness coefficients times $n^{-1/2}$ are substantial compared to the limit values of zero.

The poor Gaussian approximation has important inferential consequences. The actual coverage error rate of standard confidence intervals differ substantially from their nominal rates. In particular, two-sided symmetric confidence intervals are far too short, leading in some cases to failure rates several times their nominal values. Moreover, actual coverage failures are very asymmetric so that one-sided confidence intervals perform even worse.

Having analyzed these departures from normality of the finite sample distribution of the studentized inequality index, we turn to potential corrections in the second part of the paper. These corrections are based on considering the second order term in the distributional expansion, which is a function of the bias and skewness coefficient derived in the first part of the paper. We consider two approaches. Edgeworth expansions directly adjust the asymptotic approximation by including the $O\left(n^{-1/2}\right)$ term, whereas normalizing transformations of the inequality measure are nonlinear transformations designed to annihilate this term asymptotically. Edgeworth expansions can suffer from negativity of the density and oscillations in the tails and we show that this is indeed a problem for standardized inequality measures.

The focus of the second part of the paper is therefore on normalizing transforms. Our second key contribution is the derivation of normalizing transforms for GE inequality measures. First, we show that the skewness coefficient of a standardized

non-linear transform of the inequality measure is zero if the transform satisfies a crucial differential equation. We further derive the bias coefficient of this transform, so that we obtain a bias-corrected transform which yields the desired asymptotic refinement. Second, we use this general result to compute the normalizing transform for various income distributions and sensitivity parameters of the inequality index, and study their finite sample distributions. We show that these are indeed closer to the Gaussian limit distribution. The associated confidence intervals are much improved in terms of coverage rates and symmetry. A further improvement is obtained asymptotically and in practice by a subsequent application of the bootstrap to the transform. The asymptotic rate of convergence is the same as in Beran's (1987) double bootstrap.

We illustrate the implementation of our procedures in the context of measuring regional income and expenditure inequalities. The motivation for such a study would be the targeting of policy interventions based on inequality differences. Using household data for Côte d'Ivoire, we show that improved confidence intervals based on our methods can be substantially wider and shifted relative to standard first order methods used in the literature.

The organization of this paper is as follows. Section 2 states the class of inequality measures, considers estimation, and states the first order (Normal) approximation. The quality of this Normal approximation in finite samples is studied via simulations in Section 2.3. We consider quantiles of the actual density of the studentized inequality index, and we illustrate the consequences of the departure from normality for inference. In Section 3 we study the problems of bias and skewness formally by deriving the bias and skewness coefficients. These enable us to give the Edgeworth expansion for the GE indices and to derive the normalizing transform. We study the behavior of the transform in Section 4 and Section 5 includes an application of our procedure to the problem of regional inequality in Côte d'Ivoire. Section 6 concludes and the proofs are collected in the Appendix.

# 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 (1977, 1980, 2000).

The class of indices is defined for any real $\alpha$ 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 $\alpha$ 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 $\alpha$. The larger the parameter $\alpha$, the larger is the sensitivity of the inequality index to the upper tail of the income distribution. It is not monotonic in $\alpha$, however.

GE indices constitute a large class which nests some popular inequality measures popular as special cases. If $\alpha = 2$ the index is 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 (see e.g. Hart, 1971). In empirical work on income distributions this value of $\alpha$ is considered large. Two other popular inequality measures are the so-called Theil indices, which are the limiting cases $\alpha = 0$ and $\alpha = 1$ (Theil, 1967). Finally, the Atkinson (1970) index is ordinally equivalent to the GE index.

Although the index is defined for any real value of $\alpha$, 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 $\alpha$.

## 2.1 GE Indices and Income Distributions

Our investigation centers around three parametric income distributions which are regularly used to fit real real-world income data: the Gamma, the Lognormal, and the Singh-Maddala distribution. We use the common shorthand notation $G(r, \lambda)$, $LN(\mu, v^{1/2})$, and $SM(a, b, c)$ to refer to them. Generalized Entropy indices are scale invariant, and thus independent of the scale parameters $\lambda$, $\mu$, and $a$ for the $G$, $LN$, and $SM$ distributions respectively. For notational convenience, we suppress the irrelevant scale parameters.

McDonald (1984) has shown that these three distributions are special cases of the Generalized Beta distribution of the second kind ($GB2$), whose density is given by

$$f(x; a, b, c, d) = \frac{bx^{bd-1}}{a^{bd} B(c, d) \left[1 + (x/a)^b\right]^{d+c}} ,$$

where $B(\cdot, \cdot)$ denotes the Beta function. In particular, $SM$ has density $f(x; a, b, c, 1)$, $G$ has density $\lim_{c\to\infty} f\left(x; c\lambda^{-1}, 1, c, r\right)$, and $LN$ is a special case involving $c \to \infty$ and $b \to 0$. All three distributions are skewed to the right, but differ in other ways, such as their tail behavior. Schluter and Trede (2002), for instance, show that the right tail of the generalized beta distribution can be written as $1 - F(x; a, b, c, d) = g_1 x^{-bc}\left(1 + g_2 x^{-b} + O\left(x^{-2b}\right)\right)$ for some constants $g_1$ and $g_2$ and $x$ large. It follows that $SM$ has a heavy right tail which decays like a power function (with right tail index equal to $bc$), whereas $G$ and $LN$ decay exponentially fast. The left tail of $GB2$ can be written as $F(x, a, b, c, d) = g_3 x^{bd}\left(1 + g_4 x^b + O\left(x^{2b}\right)\right)$ for some constants $g_3$ and $g_4$ and $x$ small. The moments of the distributions are stated in McDonald (1984).

The population inequality index specializes for the different income distributions

to:

$$\begin{array}{ll} \text{Gamma :} & I(\alpha;r) = (\alpha^2-\alpha)^{-1}\left[r^{-\alpha}\Gamma(\alpha+r)/\Gamma(r)-1\right], \\ \text{Lognormal:} & I(\alpha;v) = (\alpha^2-\alpha)^{-1}\left[\exp\left(\frac{1}{2}v(\alpha-1)\alpha\right)-1\right], \\ \text{Singh-Maddala:} & I(\alpha;b,c) = (\alpha^2-\alpha)^{-1}\dfrac{c^{-(\alpha-1)}B\left(1+\alpha/b, c-\alpha/b\right)}{B\left(1+1/b, c-1/b\right)^{\alpha}-1}, \; bc>\alpha. \end{array}$$

We focus on these three income distributions not only because they are quite different, but more importantly because they are regularly used to fit actual real-world income data. For instance, Brachmann et al. (1996) estimate the distributional parameters on German income data for the 1980s and early 1990s. For $G$ they report point estimates $r \in [3.4, 4]$, for $LN$ $v \in [0.28, 0.31]$, and for $SM$ $b \in [2.7, 2.9]$ and $c \in [1.6, 2.1]$. Singh and Maddala (1976) report point estimates of $b \in [1.9, 2.1]$, $c \in [2.5, 3]$ for US income data from the 1960s. For US income data from the 1970s McDonald (1984) reports $r = 2.3$, $v \in [0.48, 0.51]$, $b \in [2.9, 3.76]$ and $c \in [1.8, 2.9]$. For the Lognormal Kloek and van Dijk (1978) find $v \in [0.21, 0.54]$ for different groups of income earners using 1973 Dutch data. Across these studies, we have $r \in [2.3, 4]$, $v \in [0.28, 0.54]$, $b \in [1.9, 3.76]$, and $c \in [1.6, 3]$. Throughout this paper, we use parameter values in these ranges. For further recent examples see Bandourian et al. (2003) who fit these income distribution models for a large number of countries, including the USA, Canada, Taiwan and most European countries for the period 1969-1997.

## 2.2 Estimation and Normal Approximations

In empirical work the inequality measure $I$ needs to be estimated from a sample of incomes denoted $X_i$, $i = 1, \ldots, n$. We follow standard practice and assume that incomes are independently and identically distributed with distribution $F$ and are positive. The measure $I$ is a functional, which maps income distributions into scalars. The commonly used estimator simply uses the Empirical Distribution Function (EDF) $\widehat{F}(x) = n^{-1}\sum_i \mathbf{1}_{(-\infty,x]}(X_i)$, where $\mathbf{1}_{(-\infty,x]}(.)$ denotes the indicator function on the open interval smaller than or equal to $x$,

$$\widehat{I} = I(\widehat{F}).$$

Since $I$ is a function of moments, the EDF-estimator is also referred to in the literature as the method of moments estimator. It is standard practice to obtain the asymptotic variance $\sigma^2 = aVar(n^{1/2}(\widehat{I} - I))$ by the delta method, yielding for $\alpha \neq \{0, 1\}$

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

and to estimate it by an EDF-based estimator, denoted $\widehat{\sigma}^2$.

Inference about the population value $I$ is then based on the studentized measure, defined as

$$S = n^{1/2}\left(\frac{\widehat{I} - I}{\widehat{\sigma}}\right). \tag{2}$$

By standard central limit arguments, $S$ has a distribution that converges asymptotically to the Gaussian distribution (see, inter alia, Cowell, 1989, or Thistle, 1990). Denoting the Gaussian distribution and density by $\Phi$ and $\phi$ respectively, then using order notation, we have

$$\Pr\left(S \leq x\right) = \Phi\left(x\right) + O\left(n^{-1/2}\right). \tag{3}$$

Standard, first order, inference methods only use the first term $\Phi\left(x\right)$ in this approximation.

Setting $x = \Phi^{-1}\left(p\right)$, yields the standard one-sided confidence interval for $I$ with bound $\widehat{I} - \widehat{\sigma} n^{-1/2} \Phi^{-1}\left(p\right)$, and the usual symmetric confidence interval $\widehat{I} - \widehat{\sigma} n^{-1/2} \times \Phi^{-1}\left(\left[1+p\right]/2\right) \leq I \leq \widehat{I} - \widehat{\sigma} n^{-1/2} \Phi^{-1}\left(\left[1-p\right]/2\right)$. Note that the asymptotic coverage rate for the two-sided confidence interval, based on a standard symmetry argument, equals $p + O\left(n^{-1}\right)$. For one-sided confidence intervals we have $p + O\left(n^{-1/2}\right)$.

## 2.3 Quality of the Normal Approximation

We proceed to investigate how well the Normal approximation performs in realistic settings for samples of varying sizes and various income distributions. First we consider coverage failure rates for confidence intervals for $I$ based on the Gaussian quantiles $\pm 1.96$, i.e. we consider one-sided confidence intervals with a nominal error rate of 2.5%, and two-sided confidence intervals with nominal error rate of 5%. Table 1 shows the results for $I\left(2\right)$. L refers to the proportion of $S$ smaller than $-1.96$, R to the proportion larger than 1.96, and T to the total and therefore the two-sided confidence interval.

Four striking facts emerge from the table. First, actual coverage failures for L and T can be a multiple of their nominal sizes and for R can be close to 0. Second, there is a huge asymmetry in the left and right rejection rates. Third, rejection rates vary substantially over income distributions. Finally, even for a sample size of 500, coverage rates can still be a multiple of the nominal sizes. Table 1 shows that standard first order methods based on the Gaussian approximation are very unreliable. Particularly bad is the overstated precision of the one-sided (L) and two-sided (T) confidence intervals.

| | n=50 | | | n=100 | | | n=250 | | | n=500 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | R | T | L | R | T | L | R | T | L | R | T |
| G(.,3) | 12.1 | 0.6 | 12.6 | 9.2 | 0.7 | 9.9 | 6.6 | 0.9 | 7.5 | 5.2 | 1.2 | 6.4 |
| LN(.,0.7) | 29.9 | 0.7 | 30.6 | 24.1 | 0.4 | 24.5 | 17.9 | 0.3 | 18.2 | 14.2 | 0.3 | 14.5 |
| SM(.,2.9,1.9) | 21.0 | 0.4 | 21.3 | 17.6 | 0.3 | 17.8 | 14.0 | 0.3 | 14.3 | 11.9 | 0.3 | 12.2 |

Table 1: Actual coverage failures rates in % Notes: L and R refer to coverage errors on the left and right, T is the total coverage error. Nominal error rates are 2.5 % for L and R and 5% for T. Based on $10^6$ replications.

Changing parameter values of the income distribution does change the absolute and relative performance. Keeping $\alpha = 2$ and $n = 100$, we found in simulations

not reported in this paper for reasons of space that the performance improves in the Lognormal case as $\nu$ decreases (although for $\nu^{1/2} = 0.5$ coverage failure is still 15.7% for the two-sided confidence intervals), worsens in the Singh-Maddala case as $b$ and $c$ decrease, (e.g. when they are decreased to $b = 2.8$ and $c = 1.7$, coverage failure worsens to 22.6%), and does not change very much in the Gamma case as $r$ varies over the range 2.3 to 4, with coverage rates from 10.3% to 9.3% respectively. Finally, changing the significance level does not qualitatively change the results. See Garderen and Schluter (2003) for further results and also Biewen (2001).

The coverage failure will change with the value of $\alpha$. Figure 1 shows the total coverage failure of two-sided confidence intervals as a function of $\alpha$. Failure rates generally improve as $\alpha$ decreases, but still exhibit substantial excess over nominal rates.

Figure 1: approximately here

Given the poor performance and asymmetry in the coverage errors of standard confidence intervals, we investigate the distribution of $S$ through its quantiles. Table 2 reports the results for $\alpha = 2$, whereas Figure 2 considers $\alpha$ varying.

| | | n=100 | | | n=250 | | | n=500 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $\Phi(p)$ | G | LN | SM | G | LN | SM | G | LN | SM |
| 0.025 | -1.96 | -2.97 | -5.26 | -4.03 | -2.60 | -4.34 | -3.53 | -2.40 | -3.81 | -3.29 |
| 0.05 | -1.64 | -2.45 | -4.34 | -3.36 | -2.15 | -3.56 | -2.95 | -1.98 | -3.12 | -2.74 |
| 0.1 | -1.28 | -1.88 | -3.36 | -2.62 | -1.66 | -2.72 | -2.31 | -1.53 | -2.37 | -2.12 |
| 0.25 | -0.67 | -1.01 | -1.90 | -1.50 | -0.87 | -1.49 | -1.29 | -0.80 | -1.27 | -1.18 |
| 0.5 | 0.00 | -0.19 | -0.59 | -0.44 | -0.11 | -0.37 | -0.33 | -0.07 | -0.26 | -0.26 |
| 0.75 | 0.67 | 0.51 | 0.36 | 0.40 | 0.57 | 0.45 | 0.44 | 0.59 | 0.50 | 0.48 |
| 0.9 | 1.28 | 1.05 | 0.95 | 0.94 | 1.11 | 0.97 | 0.95 | 1.14 | 0.99 | 0.97 |
| 0.95 | 1.64 | 1.34 | 1.21 | 1.19 | 1.40 | 1.21 | 1.20 | 1.45 | 1.23 | 1.22 |
| 0.975 | 1.96 | 1.59 | 1.44 | 1.40 | 1.66 | 1.42 | 1.40 | 1.71 | 1.44 | 1.43 |

Table 2: Quantiles of studentised $I(2)$. Notes: As for Table 1.

Figure 2: approximately here

We observe the following. First, the discrepancy between actual and Gaussian quantiles varies substantially across the distributions. It is the smallest for the Gamma case, and the worst for the Lognormal case. Second, the performance worsens as $\alpha$ increases across all income distributions. Third, across the five depicted quantiles, the lower quantiles exhibit the largest deviation from the corresponding

Gaussian quantiles. All empirical quantiles lie below the corresponding Gaussian quantiles. This suggests that the actual distribution is biased, skewed to the left, and that the skewness increases in $\alpha$. Fourth, the deviations decrease naturally as sample size increases, but the improvements are slow. Hence skewness is persistent even in fairly large samples.

# 3 Cumulants, Edgeworth Expansions, and Normalizing Transforms

The simulation study in the previous section has shown that the Normal approximation suffers from substantial bias and skewness problems. In this section we study bias and skewness formally by considering expansions to second order of the first and third cumulant (assuming they exist) of the studentized inequality measures $S$. These expansions are given by

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

The expansion of the second cumulant is $K_{S,2} = 1 + O(n^{-1})$.[1] The terms $k_{1,2}$, and $k_{3,1}$ and other quantities like $\sigma$, depend on $\alpha$ and the income distribution which we leave implicit for notational simplicity throughout. A key contribution of our paper is the derivation of the bias and skewness coefficients $k_{1,2}$ and $k_{3,1}$ for the examined inequality indices. These coefficients are the critical factors in the second order terms in the expansion of the cumulant generating function of $S$,

$$\kappa\left(y\right) = \frac{1}{2}y^2 + n^{-1/2}\left[y\ k_{1,2} + \frac{1}{6}y^3 k_{3,1}\right] + O\left(n^{-1}\right), \tag{5}$$

and the Edgeworth expansion

$$\Pr(S \leq x) = \Phi(x) - n^{-1/2}\left(k_{1,2} + \frac{1}{6}k_{3,1}(x^2-1)\right)\phi(x) + O(n^{-1}). \tag{6}$$

The implied second order expansion of the density follows immediately as

$$pdf\left(x\right) = (1 + n^{-1/2}x[\frac{1}{6}k_{3,1}\left(x^2-3\right) + k_{1,2}])\phi\left(x\right) + O(n^{-1}). \tag{7}$$

See e.g. Hall (1992) for an extensive discussion of Edgeworth expansions, who observes, for instance, that the right hand side of equation (6) does not necessarily converge as an infinite series. Regularity conditions for the validity of the expansion are also stated in Hall (1992, Section 2.4). The GE index is a smooth function of the moments with continuous third derivatives and $\mu_1 > 0$ since we assume incomes to be

[1] By construction $k_{1,1} = 0$ and $k_{2,1} = 1$. Due to the studentization of $S$, $K_{S,3}/K_{S,2}^{3/2} = n^{-1/2}k_{3,1} + O\left(n^{-1}\right)$, and $k_{3,1}$ is therefore also the coefficient of skewness.

positive. This implies that Theorem 2.2 in Hall (1992) applies and hence we require that the income distribution for $X$ satisfies the moment conditions $E\left(X^3\right) < \infty$ and $E\left(X^{3\alpha}\right) < \infty$ and that $X$ has a proper density function (implying that Cramér's condition is satisfied). Note that these moment conditions restrict the admissible parameter values for the Singh-Maddala distribution ($bc > max(3, 3\alpha)$).

Normal approximations only consider the first order term, i.e. $\frac{1}{2}y^2$ in Equation (5) or $\Phi(x)$ in Equation (6), so the higher order term indicates deviation from Normality. Edgeworth expansions directly adjust the asymptotic approximation by including the $O\left(n^{-1/2}\right)$ term. However, Edgeworth expansions can suffer from negativity of the density and oscillations in the tails and we show below that this is indeed a problem for standardized inequality measures.

## 3.1 Normalizing Transforms

Rather than directly adjusting the asymptotic approximation for $S$ by including the $O\left(n^{-1/2}\right)$ term in the approximating density, normalizing transformations of the inequality measure are designed to annihilate this term asymptotically. The resulting distribution of the studentized transformed and bias corrected inequality measure then satisfies $\Phi(x)+O(n^{-1})$, so that, compared to (3) or (6), the order of the approximation has improved.

In our derivations of the required refinement we essentially follow an approach proposed in Niki and Konishi (1986). See also Marsh (2004) for a multivariate extension and Goncalves and Meddahi (2008) who apply the idea to transforming realized volatility . However, there are four important differences. First, we standardize using the empirical quantities, whereas Niki and Konishi use theoretical versions. Second, we deal explicitly with the open issue highlighted by Niki and Konishi (1986, p.377) that the cumulants depend on the true quantity $I$ being estimated. Third, we deal with the complication that income distributions and the inequality measure $I$ often depend on more than one parameter, as in the case of the Singh-Maddala distribution. This implies that $I$ is not an invertible mapping of the parameters and as a consequence a whole family of solutions exist. One can choose any member of this family and we show below what difference this choice makes.

Let $t$ denote a transformation of the inequality measure $I$ with continuous first and second derivatives $t'$ and $t''$, satisfying $t'(\widehat{I}) \neq 0$. The standardized transform defined by

$$T = n^{1/2}\frac{t(\widehat{I}) - t(I)}{\widehat{\sigma}t'(\widehat{I})}, \tag{8}$$

will also be asymptotically Normal, but its cumulants will have changed and depend on the nonlinear transformation $t$. We want to relate the cumulants of $T$ to the cumulants of $S$ and determine $T$ such that the third cumulant vanishes. In order to do so, we first state the basic relation between $S$ and $T$ in the following lemma.

**Lemma 1.**

$$T = S - \frac{1}{2}\frac{t''(I)}{t'(I)}n^{-1/2}\sigma S^2 + O_p\left(n^{-1}\right) \tag{9}$$

Assuming that the distribution of $T$ also admits a valid Edgeworth expansion, it will be of the form

$$\Pr(T \leq x) = \Phi(x) - n^{-1/2}\left(\lambda_{1,2} + \frac{1}{6}\lambda_{3,1}(x^2-1)\right)\phi(x) + O(n^{-1}), \tag{10}$$

where $\lambda_{1,2}$ and $\lambda_{3,1}$ are the coefficients for $n^{-1/2}$ of the first and third cumulant of $T$ respectively. The cumulants of $T$ are naturally related to the cumulants of $S$ since $t$ is a smooth function of $I$. The next lemma states this relationship.

**Lemma 2.**

$$\lambda_{1,2} = k_{1,2} - \frac{1}{2}\sigma\frac{t''(I)}{t'(I)}, \tag{11}$$

$$\lambda_{3,1} = -\frac{3\sigma t''(I) - k_{3,1}t'(I)}{t'(I)}. \tag{12}$$

Our results differ from those stated in Niki and Konishi (1986) because our definition of the standardized transform (8) has $\widehat{\sigma}t'(\widehat{I})$ in the denominator instead of $\sigma t'(I)$ used by Niki and Konishi. The consequence of this is that in expression (9) the second term on the right has a coefficient of -1. Finally, the differences between $\lambda_{1,2}$ and $\lambda_{3,1}$ of Lemma 2, and the results of Niki and Konishi are the negative signs of the second right hand terms. This again is a result of $\sigma t'(I)$ being estimated.

The normalizing transform we seek is a function $t$ that reduces the skewness of $T$ to zero up to second order. It follows from equation (12) that the skewness term $\lambda_{3,1}$ is reduced to zero if the transform $t$ satisfies the differential equation

$$3\sigma t''(I) - k_{3,1}t'(I) = 0, \tag{13}$$

or, assuming $t'(I) \neq 0$,

$$\frac{t''(I)}{t'(I)} = \frac{1}{3}\frac{k_{3,1}}{\sigma}.$$

The formal solution to the differential equation is

$$t(I) = \int \exp\left(\int \frac{1}{3}\frac{k_{3,1}}{\sigma}dI\right)dI.$$

The asymptotic refinement we seek is found by solving the differential equation (13), and making a subsequent direct bias correction based on (11). Note that the differential equation is invariant to affine transformations and the constants of integration are immaterial.

**Proposition 3.** *If the transform t satisfies the differential equation (13), then*

$$\Pr\left(T - n^{-1/2}\lambda_{1,2} \leq x\right) = \Phi(x) + O(n^{-1}). \tag{14}$$

The bias correction can be applied using the $\sqrt{n}$-consistent estimator $\widehat{\lambda}_{1,2}$ based on the EDF, giving rise to the following result.

**Corollary 4.** $\Pr\left(T - n^{-1/2}\widehat{\lambda}_{1,2} \leq x\right) = \Phi(x) + O(n^{-1})$.

The implication of this corollary is that with the estimated bias correction we obtain second order correct one-sided confidence intervals.

Another consequence of this proposition is that a further order of magnitude in the accuracy can be gained by bootstrapping the transformed statistic as in the following corollary:

**Corollary 5.** *Let* $z^*_{1-p}$ *denote the quantile of the bootstrap distribution of the bias-corrected studentized transform* $T_{bc} = n^{1/2}\frac{t(\widehat{I})-t(I)}{\widehat{\sigma}t'(\widehat{I})} - n^{-1/2}\widehat{\lambda}_{1,2}$, *and let* $\widehat{I}_p = t^{-1}(t\left(\widehat{I}\right) - n^{-1/2}\widehat{\sigma}\left(z^*_{1-p} + n^{-1/2}\widehat{\lambda}_{1,2}\right)t'\left(\widehat{I}\right))$ *denote the* $p$ *upper confidence limit. Then*

$$\Pr\left\{I \leq \widehat{I}_p\right\} = 1 - \Pr\left\{T_{bc} \leq z^*_{1-p}\right\} = p + O\left(n^{-3/2}\right).$$

The one-sided bootstrap confidence interval thus achieves the same asymptotic rate as Beran's (1987) pre-pivoted, or double bootstrap.

# 4 Expansion and Transforms for Studentized Inequality Measures

Central to the theory in the previous section is knowledge of the bias and skewness coefficients $k_{1,2}$ and $k_{3,1}$ for the studentized inequality measures. We will now state the first key result of the paper, namely:

**Proposition 6.** *Assuming the expectations* $E(X^3)$ *and* $E(X^{3\alpha})$ *exist, then the bias and skewness coefficients for the studentized inequality measures are given by*

$$\begin{aligned} k_{1,2} &= \left(B^{-1/2}M_2 - \frac{1}{2}B^{-3/2}M_5\right)\cdot\left(1 - 2\cdot\mathbf{1}_{(0,1)}(\alpha)\right), \\ k_{3,1} &= B^{-3/2}\left(M_4 + 6M_1M_3 - 3M_5\right)\cdot\left(1 - 2\cdot\mathbf{1}_{(0,1)}(\alpha)\right), \end{aligned}$$

*where* $\mathbf{1}_{(0,1)}$ *is the indicator function on the interval* $(0,1)$ *and*

$$\begin{aligned} B &= \mu_1^2\mu_{2\alpha} + \alpha^2\mu_\alpha^2\mu_2 - 2\alpha\mu_\alpha\mu_1\mu_{\alpha+1} - (1-\alpha)^2\mu_\alpha^2\mu_1^2, \\ M_1 &= \mu_1\mu_{\alpha+1} - \mu_1^2\mu_\alpha - \alpha\mu_\alpha\mu_2 + \alpha\mu_\alpha\mu_1^2, \\ M_2 &= \mu_{\alpha+1} - \mu_1\mu_\alpha - \frac{1}{2}\alpha(\alpha+1)\frac{\mu_\alpha}{\mu_1}\left(\mu_2 - \mu_1^2\right), \\ M_3 &= \mu_1\left(\mu_{2\alpha} - \mu_\alpha^2\right) - \frac{1}{2}\alpha(\alpha+1)\mu_\alpha\left(\mu_{\alpha+1} - \mu_1\mu_\alpha\right) - \alpha\mu_\alpha M_2, \end{aligned}$$

$$
\begin{aligned}
M_4 &= \mu_1^3\left(\mu_{3a} - 3\mu_{2\alpha}\mu_\alpha + 2\mu_\alpha^3\right) \\
&\quad -3\alpha\mu_\alpha\mu_1^2\left(\mu_{2\alpha+1} - \mu_{2\alpha}\mu_1 - 2\mu_{\alpha+1}\mu_\alpha + 2\mu_\alpha^2\mu_1\right) \\
&\quad +3\left(\alpha\mu_\alpha\right)^2\mu_1\left(\mu_{\alpha+2} - 2\mu_{\alpha+1}\mu_1 + 2\mu_\alpha\mu_1^2 - \mu_\alpha\mu_2\right) \\
&\quad -\left(\alpha\mu_\alpha\right)^3\left(\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3\right), \\
M_5 &= 2\left(\mu_1^2\mu_{2\alpha} + \left(\alpha^2-\alpha\right)\mu_1\mu_\alpha\mu_{\alpha+1} - \alpha^3\mu_2\mu_\alpha^2 + (\alpha-1)(1-\alpha)^2\mu_1^2\mu_\alpha^2\right)\left(\mu_{\alpha+1} - \mu_1\mu_\alpha\right) \\
&\quad +\alpha^2\mu_1\mu_\alpha^2\left(\mu_{\alpha+2} - \mu_2\mu_\alpha\right) \\
&\quad +2\mu_1\left(\alpha^2\mu_\alpha\mu_2 - \alpha\mu_1\mu_{\alpha+1} - (1-\alpha)^2\mu_1^2\mu_\alpha\right)\left(\mu_{2\alpha} - \mu_\alpha^2\right) \\
&\quad +2\alpha^2\mu_1\mu_\alpha^2\left(\mu_{\alpha+2} - \mu_1\mu_{\alpha+1}\right) - 2\alpha\mu_1^2\mu_\alpha\left(\mu_{2\alpha+1} - \mu_{a+1}\mu_\alpha\right) \\
&\quad -\alpha\mu_1^2\mu_\alpha\left(\mu_{2\alpha+1} - \mu_1\mu_{2\alpha}\right) + \mu_1^3\left(\mu_{3\alpha} - \mu_\alpha\mu_{2\alpha}\right) \\
&\quad -2\alpha\mu_\alpha\left(\mu_1\mu_{2\alpha} - \alpha\mu_\alpha\mu_{\alpha+1} - (1-\alpha)^2\mu_1\mu_\alpha^2\right)\left(\mu_2 - \mu_1^2\right) \\
&\quad -\alpha^3\mu_\alpha^3\left(\mu_3 - \mu_1\mu_2\right),
\end{aligned}
$$

*and $\mu_\alpha$ is the $\alpha$-th moment of the income distribution $F$.*[2]

All moments $\mu_\alpha$ exist under the regularity assumptions we made for the existence of the Edgeworth expansion above. Appendix B presents simulation support for the expressions in Proposition 6. The term $B$ directly relates to the variance and is therefore bounded away from zero. This leads to the following:

**Corollary 7.** *The coefficients $k_{1,2}$ and $k_{3,1}$ can be estimated $\sqrt{n}$-consistently using corresponding sample moment for the theoretical moments in Proposition 6.*

The second order Edgeworth expansion for the studentized inequality index is given by (6) with $k_{1,2}$ and $k_{3,1}$ given in Proposition 6.

We proceed by analysing the behavior of the cumulants, Edgeworth densities, and normalizing transforms in turn. Each case will be introduced by considering the Gamma distribution and $I(2)$, since this is an attractive example that allows simple explicit analytical solutions. This will be our leading example.

## 4.1 Bias and Skewness Coefficients

We examine the behavior of the bias and skewness coefficients for the income distributions studied earlier, namely the Gamma, Lognormal, and Singh-Maddala distributions. These parametric distributions give the population moments used in Proposition 6 for the calculation of the coefficients. The aim is to gain insight into the poor performance of the Normal approximation by examining the magnitude of the coefficients. This will help to predict when first order methods can be expected to be poor. By varying $\alpha$ we also highlight the statistical implications of a particular choice of $\alpha$ that is usually made by researchers on the basis of economic, rather than statistical considerations.

[2]The R and Mathematica computer code for these expressions are available from the authors upon request.

The interpretation of the individual contributions is made plain in the derivation contained in Appendix A.1.

#### 4.1.1 The Gamma Distribution

We first consider our leading example. Substituting the theoretical moments of the Gamma distribution in Proposition 6, we obtain after some simplification:

$$k_{1,2} = -\frac{3}{\sqrt{2}}\frac{r+3}{\sqrt{r\,(r+1)}} \quad \text{and} \quad k_{3,1} = -\frac{8}{\sqrt{2}}\frac{r+4}{\sqrt{r\,(r+1)}}. \tag{15}$$

Both cumulants decrease in magnitude as $r$ increases. In terms of the parametrization of Table 1, we have with $r = 3$, $k_{1,2} = -3.67$, and $k_{3,1} = -11.43$, so that for samples of size 100 bias and skewness are of moderate size.

#### 4.1.2 Other Values of the Sensitivity Parameter $\alpha$ and Income Distributions

Figure 3 depicts the contour plots of $k_{1,2}$ and $k_{3,1}$ as we vary the sensitivity parameter $\alpha$ of the inequality index and the relevant parameters of the income distribution.

The coefficients share important features across all three income distributions. (i) All bias and skewness coefficients are negative (ii) except for small values of $\alpha$, bias and skewness increase in magnitude as $\alpha$ increases (iii) bias and skewness decrease with $r$ in the Gamma case[3] and with $b$ in the Singh-Maddala case (holding $c$ constant and chosen so that all relevant moments exist), and they increase with $sd$ in the Lognormal case. (iv) the numeric values of the coefficients are of quantitative importance, as dividing them by the square root of the sample sizes considered in Section 2.3 yields values which are large relative to the Gaussian limit values of zero. This explains the poor behavior of the normal approximation observed in Section 2.3.

Finally we note on the basis of Figure 3 that problems increase as $\alpha$ increases. Hence, we recommend choosing $\alpha$ as low as possible within the set of economically acceptable $\alpha$'s.

[3]The Gamma case given by (15) corresponds to the top-most horizontal section of the first panel of Figure 3. Further analytical solutions are for instance:

a. In the Gamma case, fixing $r = 3$ and varying $\alpha$ instead, we obtain

$$\begin{aligned} k_{3,1} \;=\; & -\frac{2}{\sqrt{3}}\left(-\left(\alpha^2+3\right)\Gamma^2\left(3+\alpha\right)+6\Gamma\left(3+2\alpha\right)\right)^{-3/2}\times \\ & \left(\Gamma^3\left(3+\alpha\right)\left(18\alpha^2+\alpha^3+3\alpha^4+18\right)+36\Gamma\left(3+3\alpha\right)\right. \\ & \left.-18\Gamma\left(3+2\alpha\right)\Gamma\left(3+\alpha\right)\left(2\alpha^2+3\right)\right)\times\left(1-2\cdot\mathbf{1}_{(0,1)}\left(\alpha\right)\right). \end{aligned} \tag{16}$$

Skewness increases in magnitude for $\alpha > 0.5$. Obviously, this expression and (15) coincide for $\alpha = 2$ and $r = 3$.

b. In the Lognormal case with $sd = v^{1/2}$ and $\alpha = 2$ fixed we obtain

$$\begin{aligned} k_{3,1} \;=\; & \left(4e^{v}-4e^{2v}+e^{4v}-1\right)^{-3/2}\times \\ & \left(-2e^{12v}+12e^{8v}+12e^{6v}-48e^{5v}-36e^{4v}+136e^{3v}-96e^{2v}+24e^{v}-2\right). \end{aligned}$$

Skewness increases in magnitude with $sd$.

Figure 3: about here

## 4.2 Edgeworth Densities

We use our leading example of the Gamma distribution with $I(2)$ to highlight the main issues involved in the Edgeworth approximation. Other income distributions lead to similar qualitative conclusions and are not reported.

Using (15) and (7), we obtain the second order approximation

$$pdf(x) = \phi(x)\left[1 - n^{-1/2}x\frac{\sqrt{2}}{6\sqrt{r(r+1)}}\left(4x^2(4+r) - 3r - 21\right)\right] + O\left(n^{-1}\right).$$

Figure 4 depicts the actual density of the studentized inequality index obtained by simulation, and two versions of the Edgeworth density for the case $r = 3$ and $n = 100$. The first Edgeworth density uses the theoretical coefficients, and the second uses the empirical version based on sample moments and Corollary (7). This empirical version can be thought of as the mean Edgeworth expansions when averaged over simulation iterations. By the linearity of the density functions in terms of the $k_{1,2}$ and $k_{3,1}$, this simply equals the Edgeworth density evaluated at the averaged estimates of $k_{1,2}$ and $k_{3,1}$.

Figure 4: about here

Both approximations are an improvement over the Normal approximation in that they capture the skewness of the distribution. Several features are noteworthy, however. First, the Edgeworth density is not guaranteed to be positive and we see that the right hand tail actually becomes negative, although less so for the empirical than for the theoretical Edgeworth approximation. Second, the right tail of the theoretical Edgeworth density decays too quickly.

The third problematic feature is oscillations in the tails of the approximation. A graph for $r = 0.6$, shows this more clearly, but is not included for reasons of space. The problem of oscillation, which is also present for the other income distributions studied here, is well known (see e.g. Niki and Konishi, 1986), and motivates the search for a normalizing transform.

## 4.3 Transforms for Inequality Measures

We turn to deriving the normalizing transforms for the three distributional cases discussed above. The relative simplicity of the Gamma case with $I(2)$ yields an explicit analytical solution. Typically, however, the transform is computed using numerical techniques which we develop and implement below. We then provide a

systematic discussion of the properties of the transforms across income distributions and sensitivity parameters.

#### 4.3.1 The Gamma Case Revisited

Using the formulas for $k_{1,2}$ and $k_{3,1}$ in (15), and using the fact that $I(2) = (2r)^{-1}$, the differential equation (13) becomes

$$\frac{t''(I)}{t'(I)} = -\frac{4}{3}\frac{1}{I}\frac{1+8I}{1+2I}.$$

Integration yields the exact solution (using constants of integration 0 and 1)

$$\begin{aligned} t(I) &= -3I^{-1/3} + \frac{140}{81}2^{1/3}\ln\left(I^{1/3}+2^{-1/3}\right) \\ &\quad -\frac{70}{81}2^{1/3}\ln\left(I^{2/3}-I^{1/3}2^{-1/3}+2^{-2/3}\right) - \frac{140}{81}3^{1/2}2^{1/3}\arctan 3^{-1/2}\left(2^{4/3}I^{1/3}-1\right) \\ &\quad -\frac{118}{27}\frac{I^{2/3}}{1+2I} - \frac{16}{9}\frac{I^{2/3}}{(1+2I)^2} - \frac{2}{3}\frac{I^{2/3}}{(1+2I)^3}. \end{aligned} \tag{17}$$

The transformation for this particular case is depicted in the second panel of Figure 5. As this figure also collects the transformations for other values of the sensitivity parameter $\alpha$, and other income distributions, we postpone its discussion.

#### 4.3.2 Numerical Solutions

The Gamma example with $\alpha = 2$ is special for two reasons. First, because the simple form of the differential equation. Second, there is a simple invertible relation between the inequality index and the parameters of the distribution. This relation is no longer trivial when $\alpha \neq 2$, and in general there is no analytically tractable relation for other distributions. It is possible however to obtain a numerical inverse and to solve the differential equation (13) numerically. This involves three steps:

1. Using the general formulas for the cumulants $k_{1,2}$, $k_{3,1}$, and $\sigma^2$ we can use the theoretical moments from a specific income distributions to express the cumulants $k_{1,2}$, $k_{3,1}$, and $\sigma^2$ in terms of parameters from the income distribution.

2. We express the cumulants $k_{1,2}$ and $k_{3,1}$ in terms of $I$. This requires the inverse of $I$ which we calculate numerically. The inverse can be determined if $I$ depends on one parameter only. This obviously holds for one parameter families, but also for the Log-Normal and Gamma distributions, because $I$ is scale invariant. This invariance property is inherited by the cumulants, so that $k_{1,2}$ and $k_{3,1}$ depend only on the shape parameter and we can express them as functions of $I$.
   For other distributions we can choose a one dimensional path for the parameters such that $I$ becomes an invertible function. This can be achieved by imposing the right number of restrictions ($d-1$, if $I$ depends on $d$ income parameters), or

by making the $d$ income distribution parameters a function of only one parameter. In our Singh-Maddala distribution we show what difference the restriction makes by first holding $b$ fixed, such that $I$ is an invertible function of $c$ only, and then holding $c$ fixed.

3. We solve the differential equation (13) numerically.[4]

The solutions depend, in general, on the sensitivity parameter $\alpha$, which is a fixed known constant, the true underlying distribution, and the restriction chosen in step 2.[5]

Since no analytical solutions are available in general, we display the transforms graphically in Figures 5 and 6. Figure 5 shows the transforms for different values of the sensitivity parameter $\alpha$ and, in each panel, we consider a different income distribution. In Figure 6 we compare the transforms across income distributions when $\alpha$ is fixed at 2. For the Singh-Maddala income distribution, we display the solutions for the two restrictions, keeping $b$ fixed and $c$ fixed, separately.

The figures have been generated by setting the initial conditions of the differential equation such that the solutions cross the horizontal axis at the same point at an angle of 45 degrees. Recall that the transforms are invariant to affine transformations and hence choosing the constants of integration in this way is inconsequential. In order to relate the curvature of the transforms to the untransformed case, we have also depicted the 45-degree lines, which represent the identity transforms. Consequently, if no transformation was required, the solution would coincide with this 45 degree line.

Figure 5 : about here

Consider Figure 5 first. We see that for all the distributions the transforms change substantially as we vary $\alpha$. The transform for $\alpha = 2$ is the one most curved in all four cases. This implies that $\alpha = 2$ requires, informally speaking, the biggest amount of transformation to obtain a standard Normal distribution. Recall from the previous simulations that $\alpha = 2$ is the most troublesome case. For $\alpha = 0$, the least amount of adaptation is required in the Lognormal and Singh-Maddala cases, whereas for the Gamma case, $\alpha = 0.5$ is slightly flatter than $\alpha = 0$. The behavior of the transform is therefore not monotonic in $\alpha$. For the Singh-Maddala distribution we see quite a difference between the case where we hold $b$ constant and where we hold $c$ fixed. This is related to the fact that $I$, and the gradient of $I$ with respect to $b$ change much faster

[4] The Mathematica and R code is available from the authors upon request.

[5] A complimentary approach is to consider Box-Cox transforms as in Goncalves and Meddahi (2008) in the context of realized volatility modelling. Solving differential equation (13) above provides a solution for the optimal Box-Cox parameter in the sense that, within this class of transformations and without accounting for the fact that $k_{3,1}$ and $\sigma$ depend on $I$, the optimal Box-Cox parameter $\beta^* = 1 + I\frac{1}{3}\frac{k_{3,1}}{\sigma}$ yields $\lambda_{3,1} = 0$. $\beta^*$ could be estimated using our Proposition 6. We are grateful to a referee for alerting us to this.

than with respect to $c$. This has a further consequence that the domain of definition of the numerical transform is much broader for $c$ fixed than for $b$ fixed. This can be seen from the graph e.g. for $\alpha = 2$ the numerical transform is only calculated for $I$ between approximately 0.055 and 0.095 for $b$ fixed and between 0.03 and 0.28 for $c$ fixed.

Figure 6 : about here

Next, we compare the transforms in Figure 6 across income distributions when $\alpha = 2$. The figure shows that the transforms vary substantially between income distributions. The transform for the Singh-Maddala case with $b$-fixed is the most extreme, whereas the Gamma case is more moderate. The Gamma transform is therefore conservative and will not be optimal for specific parametric alternatives, but will correct for certain amount of skewness and will be superior to first order asymptotics. In the next section we show that the improvement is substantial.

#### 4.3.3 Densities of the Normalized Transform and Confidence Intervals

We examine the extent to which the transformed statistics are distributed closer to a standard Normal distribution in samples of size $n = 100$, and the inferential improvements that this entails. We simulate the densities of the studentized inequality measure $S$, and the studentized bias corrected transform $T$ based on Corollary 4. The top panel in Figure 7 considers the Gamma case for which we derived an exact solution for the transform in Equation (17). The actual finite sample density of $S$ departs substantially from the limiting density (the same $S$ has been considered in Figure 4). In particular, the density is skewed to the left and is biased. The transform succeeds in substantially reducing the skewness, and the bias correction shifts the density to the right. The resulting density for $T$ is much closer to the Gaussian density.

Figure 7 : about here

The two lower panels in Figure 7 also show the resulting distributions for the Lognormal and Singh-Maddala cases. The graph shows that the distributions of $T$ are again closer to the Gaussian distribution than that of $S$, with the exception of the transform based on the Singh-Maddala distribution holding $b$ fixed. This was the most extreme transform in Figure 6. The transforms actually seem to overcompensate the skewness as the distribution of $T$ is now skewed to the right. The actual quantiles of the distributions are shown in Table 3. For the Singh-Maddala case we use the case holding $c$ fixed.[6] The table also shows the results of using the Gamma transform, $T_\gamma bc$, when in fact the distribution is Lognormal or Singh-Maddala. The Gamma

[6] The choice of restriction used to establish an invertible relation between the index and parameters is not innocuous, however, since holding $b$ fixed instead of $c$ leads to a coverage that is far worse

transform was chosen for two reasons: it is a conservative choice which will correct without being excessive (such as the Singh-Maddala case with $b$ fixed, last panel Figure 7). Second, it has an explicit analytic solution with domain of definition the whole positive real line.

| | | G(.,3) | | LN(.,$\sqrt{0.1}$) | | | SM(.,3.5,3.0) | | |
|---|---|---|---|---|---|---|---|---|---|
| $p$ | $\Phi(p)$ | S | Tbc | S | Tbc | $T_\gamma$bc | S | Tbc | $T_\gamma$bc |
| 0.025 | -1.96 | -2.98 | -2.04 | -3.17 | -2.05 | -2.29 | -2.91 | -1.90 | -2.09 |
| 0.05 | -1.64 | -2.45 | -1.75 | -2.59 | -1.77 | -1.93 | -2.39 | -1.64 | -1.79 |
| 0.10 | -1.28 | -1.89 | -1.42 | -1.97 | -1.44 | -1.53 | -1.83 | -1.34 | -1.43 |
| 0.25 | -0.67 | -1.02 | -0.83 | -1.06 | -0.86 | -0.86 | -0.99 | -0.82 | -0.82 |
| 0.50 | 0.00 | -0.18 | -0.13 | -0.19 | -0.17 | -0.10 | -0.19 | -0.20 | -0.11 |
| 0.75 | 0.67 | 0.51 | 0.60 | 0.51 | 0.58 | 0.64 | 0.51 | 0.53 | 0.62 |
| 0.90 | 1.28 | 1.04 | 1.33 | 1.03 | 1.36 | 1.30 | 1.03 | 1.36 | 1.30 |
| 0.95 | 1.64 | 1.33 | 1.83 | 1.32 | 1.91 | 1.72 | 1.31 | 2.00 | 1.73 |
| 0.975 | 1.96 | 1.59 | 2.28 | 1.56 | 2.45 | 2.10 | 1.56 | 2.75 | 2.11 |

Table 3: Quantiles for $I(2)$ and $n = 100$. Notes: based on $10^6$ replications.

.

The table reiterates the improvements shown in Figure 7, and in addition shows that the Gamma transform gives reasonable improvements even if the true distribution is not Gamma. As regards to the left tail in the Lognormal and Singh-Maddala case, the Gamma transform is not sufficiently curved, but for the right tail the lesser curvature of the Gamma transform actually gives an improvement over the corresponding transforms. We quantify the net effect by considering the confidence intervals next.

Table 4 reports the coverage errors of one- and two-sided confidence intervals with nominal 2.5% coverage failures in the tails, using the Gaussian critical values of $\pm 1.96$. Across all distributions the transforms improve the overall failure rates, bringing the actual much closer to their nominal values. Moreover the transforms result in far more symmetric failure rates than standard first order methods. Applying the Gamma transform in the Lognormal case leads to performance almost the same as using the Lognormal transform. For the Singh-Maddala case it actually leads to a more symmetric coverage errors and a net improvement.

with 19% coverage failure (of which 2% points are caused by the fact that $\hat{I}$ falls outside the domain of definition of the numerical transform). This implies that care needs to be taken with this choice. One option would be to use a path in the parameter space for which the change in the index $I$ is maximized. This requires a further solution to a differential equation, which is not difficult, but would complicate the exposition. The path holding $c$ fixed is closer to this direction than holding $b$ fixed. The optimal direction is (0.97,0.23) when $b = 3.5$ and $c = 3.0$ and (0.99,0.16) when $b = 3.5$ and $c = 3.5$, hence close to holding $c$ fixed and little gain is therefore expected from determining the optimal restriction.

| | G(.,3) | | LN(.,$\sqrt{0.1}$) | | | SM(.,3.5,3.0) | | |
|---|---|---|---|---|---|---|---|---|
| | S | Tbc | S | Tbc | $T_\gamma$bc | S | Tbc | $T_\gamma$bc |
| L | 9.2 | 3.0 | 10.1 | 3.2 | 4.8 | 8.5 | 2.1 | 3.4 |
| R | 0.7 | 4.1 | 0.6 | 4.7 | 3.2 | 0.6 | 5.2 | 3.3 |
| T | 9.9 | 7.1 | 10.8 | 7.9 | 8.0 | 9.1 | 7.3 | 6.7 |

Table 4: Actual coverage failures rates of confidence intervals for $I(2)$ and $n = 100$. Notes: based on $10^6$ replications.

.

#### 4.3.4 Bootstrap

Applying the transform is theoretically attractive and computationally cheap and results in substantial improvements as shown in the previous section. Further improvements can be obtained by applying an additional bootstrap to the studentized transform as in Marsh (2004).[7] The idea is that the transform is more symmetrically distributed and bootstrapping converges faster for symmetric distributions. Corollary 5 has made this gain explicit.

Table 5 reports the coverage errors of one and two-sided bootstrap confidence intervals with nominal 2.5% coverage failures in the tails, and hence directly compares to Table 4. We see that in all cases bootstrapping improves coverage rates. In the Lognormal case, for instance, the overall coverage rate is reduced from 7.9 % to 5.8%. For comparisons we have also applied a bootstrap to the studentized inequality index $S$. The bootstrap of the transform gives better results than those for $S$. The extend of the improvements mirrors those reported in Marsh (2004, p. 981) in a different setting.[8] See also Biewen (2001) and Davidson and Flachaire (2007) who also investigate bootstrap methods.

| | G(.,3) | | LN(.,$\sqrt{0.1}$) | | | SM(.,3.5,3.0) | | |
|---|---|---|---|---|---|---|---|---|
| | S | T | S | T | $T_\gamma$ | S | T | $T_\gamma$ |
| L | 1.8 | 2.0 | 1.6 | 1.6 | 1.9 | 1.8 | 0.5 | 1.9 |
| R | 4.6 | 4.2 | 5.0 | 4.2 | 4.9 | 3.9 | 2.6 | 3.7 |
| T | 6.4 | 6.2 | 6.6 | 5.8 | 6.8 | 5.7 | 3.1 | 5.6 |

Table 5: Actual coverage failures rates of confidence intervals based on the bootstrap for $I(2)$ and $n = 100$. Notes: based on $10^6$ replications; in each iteration $R = 999$ bootstrap samples were drawn.

[7] Thanks to a referee for pointing this out.

[8] We have also investigated the studentized double bootstrap, following the implementations in Hinkley and Shi (1989) and Davison and Hinkley (1997, section 5.6). In view of the computational cost we have repeated the experiment 10,000 times, drawing in each iteration R=999 bootstrap samples, and M=249 bootstrap subsamples. The results are similar to the single studentized bootstrap (and the results for the single bootstrap are virtually same as those reported in Table 5). In particular, we obtain total coverage failures of 6.8%, 7.3%, and 6.3% for the G, LN and SM case respectively.

# 5 An Application to Regional Inequality Comparisons in Côte d'Ivoire

This section illustrates the application of our procedures in the context of regional comparisons of income and expenditure inequality. Inequality measures are often used in determining the targeting of policy interventions, and targeting might be based on observed differences in measured inequality across regions. Such policy interventions by governments or the World Bank often involve large sums of money. The accuracy of the inequality estimates should be taken into account in the targeting decision, and we have shown above that, for sample sizes and distributions considered in this paper, confidence intervals based on the Normal approximation tend to be too short and mislocated.

In this example we use household income and expenditure data from the Côte d'Ivoire Living Standards Survey (CILSS) which forms part of the World Bank's Living Standard and Measurement Study (LSMS).[9] The data have been used extensively in applied work, see e.g. Deaton (1998) and references therein.

The five regions are Abidjan (1), Other Cities (2), East Forest (3), West Forest (4), and Savanna (5). Given concerns about gross measurement error and outliers, we follow standard practice and drop the 2.5% most extreme income observations. This results in sample sizes by region of 288, 314, 353, 232, 295 households respectively.

The actual income and expenditure densities are captured reasonably well by the parametric densities studied in this paper. Figure 8, depicting the estimated densities for region 4, is a representative example. The figure juxtaposes the non-parametric kernel density estimate with the fitted Singh-Maddala, Lognormal, and Gamma densities. All parametric densities capture the features of the actual data, and the Singh-Maddala density is in closest agreement with the non-parametric estimate.

Figure 8 : about here

For inequality measure $I(2)$ Table 6 reports the results by region. The point estimates are reported in column 2 and exhibit substantial variation across the five regions. The confidence limits (CL) of the standard 95% equi-tailed confidence interval based on the normal approximation are reported next in columns 3 and 4. These confidence intervals vary, of course, in terms of location, but also in length. For instance, the confidence interval for region 5 is nearly three times as long as the one for region 1. Also, most confidence intervals overlap in case of income: the confidence interval for region 1 overlaps with all others except for region 3. We turn to our procedure. We transform the inequality measure using the Gamma transform and apply a subsequent studentized bootstrap. We have argued above that the Gamma transform is a conservative choice. Columns 6 to 8 report the results, and show substantially

[9] The income and expenditure data are for the year 1985, generated by the World Bank, and given in Central African Francs (CAFs). Extensive documentation is posted at http://www.worldbank.org/LSMS/country/ci/ci85docs.html.

different confidence limits. In particular, the confidence limits shift to the right in all cases, and in all but one case the confidence intervals widen. The widening is most substantial for region 5. Since the confidence intervals widen we do not expect and do not observe a reversal of the regional inequality ranking.

| | | normal approximation | | | transform + bootstrap | | |
|---|---|---|---|---|---|---|---|
| Region | $\widehat{I}(2)$ | lower CL | upper CL | length | lower CL | upper CL | length |
| | | | | income | | | |
| 1 | 0.259 | 0.222 | 0.296 | 0.074 | 0.226 | 0.301 | 0.075 |
| 2 | 0.339 | 0.291 | 0.387 | 0.095 | 0.297 | 0.390 | 0.092 |
| 3 | 0.440 | 0.363 | 0.517 | 0.154 | 0.370 | 0.534 | 0.164 |
| 4 | 0.311 | 0.250 | 0.372 | 0.123 | 0.257 | 0.395 | 0.138 |
| 5 | 0.348 | 0.250 | 0.446 | 0.196 | 0.264 | 0.569 | 0.306 |
| | | | | expenditure | | | |
| 1 | 0.147 | 0.127 | 0.166 | 0.039 | 0.128 | 0.167 | 0.039 |
| 2 | 0.200 | 0.172 | 0.229 | 0.057 | 0.175 | 0.232 | 0.057 |
| 3 | 0.281 | 0.227 | 0.334 | 0.107 | 0.231 | 0.351 | 0.120 |
| 4 | 0.201 | 0.153 | 0.250 | 0.097 | 0.159 | 0.260 | 0.101 |
| 5 | 0.194 | 0.147 | 0.241 | 0.094 | 0.152 | 0.263 | 0.111 |

Table 6: Confidence intervals for income and expenditure inequality in Côte d'Ivoire by region.

# 6 Conclusions

The finite sample distribution of the studentized inequality measure is not located at zero and is substantially skewed. In the first part of the paper we have derived general nonparametric bias and skewness coefficients based on cumulant expansions. We have shown that these coefficients are of interest in their own right, and they are also the key quantities for the Edgeworth expansions and the normalizing transforms considered in the second part of the paper. Edgeworth expansions directly adjust the asymptotic approximation by including the $O\left(n^{-1/2}\right)$ term, a function of the bias and skewness coefficients. In contrast, normalizing transformations of the inequality measure are designed to annihilate this term asymptotically. The observed problems for the Edgeworth expansion of negativity of the density and tail oscillation have led us to derive and construct the normalizing transforms in the second part of this paper. We have shown that the finite sample distributions of these transforms are much closer to the Gaussian distribution. This results in improved inference, for example in the coverage rates of confidence intervals. We have shown that applying a subsequent bootstrap yields a further improvement for inference, both in theory and in practice. Asymptotically we obtained approximation error of order $n^{-3/2}$ for confidence intervals, which is the same rate as Beran's (1987) pre-pivoted bootstrap. In practice we found actual coverage errors of the new confidence intervals close to the nominal rate.

We have illustrated our procedure with an application to regional inequality measures in Côte d'Ivoire. We showed that the resulting confidence intervals can be substantially different from those based on the Normal approximation commonly used in the applied literature.

## References


Atkinson, A. B. (1970), "On the Measurement of Inequality", *Journal of Economic Theory*, 2, 244–263.

Biewen, M. (2001), "Bootstrap Inference for Inequality, Poverty and Mobility Measurement", *Journal of Econometrics,* 108, 2, 317-342.

Bandourian, R. and J.B. McDonald, and R.S. Turley (2003), "A Comparison of Parametric Models of Income Distribution Across Countries and Over Time", *Estadistica*, 55, 135-152, Luxembourg Income Study Working Paper No. 305. , www.lisproject.org/publications/liswps/305.pdf.

Beran, R (1987) "Prepivoting to Reduce Level Error of Confidence Sets", *Biometrika*, 74, 457-468.

Brachmann, K. and A. Stich, and M. Trede (1996) "Evaluating Parametric Income Distribution Models", *Allgemeines Statistisches Archiv*, 80, 285–298.

Cowell, F. A. (1977), Measuring Inequality, (First. Ed.), Oxford: Phillip Allan.

Cowell, F. A. (1980), "On the Structure of Additive Inequality Measures", *Review of Economic Studies*, 47, 521-531.

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 and F. Bourguignon (Eds.), Handbook of Income Distribution, Chapter 2. Amsterdam: North Holland.

Davidson, R. and J.-Y. Duclos (1997), "Statistical Inference for the Measurement of the Incidence of Taxes and Transfers", *Econometrica*, 65, 1453-1465.

Davidson, R. and Flachaire (2007), "Asymptotic and Bootstrap Inference for Inequality and Poverty Measures", *Journal of Econometrics*, 141, 141-166.

Davison, A.C. and D. V. Hinkley (1997), Bootstrap Methods and Their Application, Cambridge University Press.

Deaton, A. (1998), *The Analysis of Household Surveys*, John Hopkins University Press, Baltimore and London.

Gonc¸alves, S. and N. Meddahi (2008) "Box-Cox Transforms for Realized Volatility", Januray 2008 *Mimeo, Université de Montréal.*

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

Hinkley, D.V. and S. Shi (1989), "Importance Sampling and the Nested Bootstrap", *Biometrika*, 76, 435-446.

Hart, P.E. (1971), "Entropy and Other Measures of Concentration", *Journal of the Royal Statistical Society, Ser.A* 134, 73-85.

Kloek, T. and H. van Dijk (1978), "Efficient Estimation of Income Distribution Parameters", *Journal of Econometrics* 8, 61-74.

Maasoumi, E. (1997), "Empirical Analyses of Inequality and Welfare", in M.H.Pesaran and P.Schmidt (Eds.) Handbook of Applied Econometrics, Chapter 5. Oxford: Blackwell.

MAASOUMI, E. AND H. THEIL (1979), "The Effects of the Shape of the Income Distribution on Two Inequality Measures", *Economics Letters*, 4, 289-291.
MARSH, P. (2004), "Transformations for Multivariate Statistics", *Econometric Theory,* 20, 963-987.
MCDONALD, J.B. (1984), "Some Generalized Functions for the Size Distribution of Income", *Econometrica*, 52, 647–663.
MILLS, A.M. AND S. ZANDVAKILI (1997), "Statistical Inference Via Bootstrapping for Measures of Inequality", *Journal of Applied Econometrics,* 12, 133-150.
NIKI, N., AND S. KONISHI (1986), "Effects of Transformations in Higher Order Asymptotic Expansions", *Annals of the Institute of Statistical Mathematics,* 38, 371-383.
SCHLUTER, C. AND M. TREDE (2002), "Tails of Lorenz Curves", *Journal of Econometrics*,109, 1, 151-166.
SCHLUTER, C. AND K.J. VAN GARDEREN (2003), "Improving Finite Sample Confidence Intervals for Inequality and Poverty Measures", *UvA-Econometrics Working Paper*, 2003/02, http://www.fee.uva.nl/ke/UvA-Econometrics.
SINGH, S.K. AND G.S. MADDALA (1976), "A Function for Size Distribution of Incomes", *Econometrica,* 44, 963–970.
THEIL, H. (1967), *Economics and Information Theory*, Amsterdam: North-Holland.
THISTLE, P.D. (1990), "Large Sample Properties of Two Inequality Indices", *Econometrica,* 58, 725-728.

# A Proofs

**Proof of Lemma 1**.

Expand $T(\widehat{I})$ about $I$ to second order and use the definition of $S$ to obtain

$$\frac{n^{1/2}}{\widehat{\sigma}}\frac{t(\widehat{I})-t(I)}{t'\left(\widehat{I}\right)}=S-\frac{1}{2}\frac{t''(I)}{t'(I)}n^{-1/2}\hat{\sigma}S^2+O\left(\left\|\hat{I}-I\right\|^2\right),$$

Now using the fact that $\hat{\sigma}$ and $\hat{I}$ are $\sqrt{n}$ consistent estimators the result follows.

**Proof of Lemma 2.**

Taking expectations of (9) using $E(S)=n^{-1/2}k_{1,2}+O\left(n^{-3/2}\right)$, and $E(S^2)=1+O(n^{-1})$ leads to

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

Also $E(T^2)=1+O(n^{-1})$. We have

$$T^3=S^3-\frac{3}{2}\frac{t''(I)}{t'(I)}n^{-1/2}\sigma S^4+O_p\left(n^{-1}\right).$$

Taking expectations, noting that $E(S^4)=3+O_p\left(n^{-1/2}\right)$, yields

$$E\left(T^3\right)=E\left(S^3\right)-\frac{9}{2}\frac{t''(I)}{t'(I)}n^{-1/2}\sigma+O\left(n^{-1}\right),$$

with $E(S^3)=n^{-1/2}\left[k_{3,1}+3k_{1,2}\right]$. Therefore $E(T^3)-3E(T^2)E(T)+2\left(E(T)\right)^3=n^{-1/2}\lambda_{3,1}+O(n^{-1})$ with

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

**Proof of Proposition 3.**

Note that $\lambda_{3,1}=0$ by construction. Considering the Edgeworth expansion Equation (10) for $T$ at $x+n^{-1/2}\lambda_{1,2}$, expanding it about $x$ and collecting terms of the same order yields the stated result.

**Proof of Proposition 5.**

Let $T_{bc}=T-n^{-1/2}\widehat{\lambda}_{1,2}$. By Corollary 4

$$K(x)\equiv\Pr\left\{T_{bc}\leq x\right\}=\Phi(x)+O(n^{-1}).$$

It follows that the three-term Edgeworth expansion and Cornish-Fisher expansion are

$$K(x)=\Phi(x)+n^{-1}q(x)\phi(x)+O\left(n^{-3/2}\right)$$

and $K^{-1}(1-\alpha)=z_{1-\alpha}-n^{-1}q(z_{1-\alpha})+O_p\left(n^{-3/2}\right)$ where $z_{1-\alpha}$ denotes the $(1-\alpha)$ quantile of $\Phi$. Let $\widehat{K}$ denote the bootstrap distribution. Then $\widehat{K}^{-1}(1-\alpha)=$

$K^{-1}(1-\alpha)+O_p\left(n^{-3/2}\right)$ since $\widehat{q}(x)=q(x)+O_p\left(n^{-1/2}\right)$. We have

$$\begin{aligned}
\Pr\left\{I\leq\widehat{I}_\alpha\right\} &= 1-\Pr\left\{T_{bc}\leq z^*_{1-\alpha}\right\}\\
&= 1-\left[\Phi\left(z^*_{1-\alpha}\right)+n^{-1}q\left(z^*_{1-\alpha}\right)\phi\left(z^*_{1-\alpha}\right)\right]+O\left(n^{-3/2}\right)\\
&= 1-\left[\Phi\left(z_{1-\alpha}-n^{-1}q\left(z_{1-\alpha}\right)\right)+\right.\\
&\quad \left. n^{-1}q\left(z_{1-\alpha}-n^{-1}q\left(z_{1-\alpha}\right)\right)\phi\left(z_{1-\alpha}-n^{-1}q\left(z_{1-\alpha}\right)\right)\right]+O\left(n^{-3/2}\right)\\
&= 1-\left[\Phi\left(z_{1-\alpha}\right)-n^{-1}q\left(z_{1-\alpha}\right)\phi\left(z_{1-\alpha}\right)+n^{-1}q\left(z_{1-\alpha}\right)\phi\left(z_{1-\alpha}\right)\right]+O\left(n^{-3/2}\right)\\
&= \alpha+O\left(n^{-3/2}\right).
\end{aligned}$$

See also Hall (1992) p. 122 ff.

## A.1 Proof of Proposition 6

Proposition 6 is derived in several steps. First, we derive an asymptotic expansions of the studentized inequality measure $S$. As a compact notation, we use $S_q$ to denote a term of an expansion of $S$ which is of order in probability $n^{-q}$. The desired stochastic expansion of $S$ is of the form

$$S=S_0+S_{1/2}+O_p\left(n^{-1}\right). \tag{18}$$

We determine the terms $S_0$ and $S_{1/2}$. We then derive the bias and skewness coefficients $k_{1,2}$ and $k_{3,1}$ by considering expectations of powers of $S$. We only consider the case $|\alpha|>1$ explicitly. For $|\alpha|<1$, the coefficients of the expansions need to be multiplied by $-1$ since $\alpha(\alpha-1)<0$ but $(\alpha^2(\alpha-1)^2)^{1/2}>0$.

### A.1.1 The Stochastic Expansion of $S$

Recall our notation for population and sample moments, $\mu_\alpha(F)=\int y^\alpha dF(y)$ and $m_\alpha=\mu_\alpha(\widehat{F})$. The basic technique in the derivation is to center and expand sample moments. For instance, we have $m_1^{-\alpha}=\left(\mu_1+n^{-1}\sum(X_i-\mu_1)\right)^{-\alpha}=\mu_1^{-\alpha}-\alpha\mu_1^{-\alpha-1}\left(n^{-1}\sum(X_i-\mu_1)\right)+O_p\left(n^{-1}\right)$. For this technique it is convenient to define the following stochastic quantities:

$$\begin{aligned}
Y_1 &= (X-\mu_1),\\
Y_2 &= \mu_1\left(X^\alpha-\mu_\alpha\right)-\alpha\mu_\alpha(X-\mu_1),\\
Y_3 &= \left(X^\alpha-\mu_\alpha\right)-\alpha(\alpha+1)\mu_\alpha\mu_1^{-1}(X-\mu_1)/2,\\
Y_4 &= 2\left(\mu_1\mu_{2\alpha}-\alpha\mu_\alpha\mu_{\alpha+1}-(1-\alpha)^2\mu_1\mu_\alpha^2\right)(X-\mu_1)\\
&\quad +\alpha^2\mu_\alpha^2\left(X^2-\mu_2\right)\\
&\quad +2\left(\alpha^2\mu_\alpha\mu_2-\alpha\mu_1\mu_{\alpha+1}-(1-\alpha)^2\mu_1^2\mu_\alpha\right)\left(X^\alpha-\mu_\alpha\right)\\
&\quad -2\alpha\mu_1\mu_\alpha\left(X^{\alpha+1}-\mu_{\alpha+1}\right)+\mu_1^2\left(X^{2\alpha}-\mu_{2\alpha}\right).
\end{aligned} \tag{19}$$

with specific elements for observation $i$ written like $Y_{1,i}=(X_i-\mu_1)$, etc.

We derive the stochastic expansion of $S = n^{1/2}\left(\widehat{I} - I\right)/\widehat{\sigma}$ in four steps.

First, write out the numerator

$$n^{1/2}(\widehat{I} - I) = n^{1/2}\left[\alpha^2 - \alpha\right]^{-1}\mu_1^{-\alpha}m_1^{-\alpha}\left[\mu_1^{\alpha}m_{\alpha} - \mu_{\alpha}m_1^{\alpha}\right].$$

Second, consider the asymptotic variance by applying the delta-method

$$\sigma^2 = aVar(n^{1/2}(\widehat{I} - I)) = \frac{1}{(\alpha^2 - \alpha)^2}\frac{1}{\mu_1^{2\alpha+2}}B_0,$$

with

$$B_0 = \left[\alpha^2\mu_{\alpha}^2\mu_2 - 2\alpha\mu_1\mu_{\alpha}\mu_{\alpha+1} + \mu_1^2\mu_{2\alpha} - (1-\alpha)^2\mu_1^2\mu_{\alpha}^2\right]. \tag{20}$$

The variance is estimated by using the corresponding sample moments. Denote the estimate of $B_0$ by $\widehat{B}_0$. Then combining the results from steps 1 and 2 yields

$$S = n^{1/2}\widehat{B}_0^{-1/2}\left[m_{\alpha}m_1 - \mu_1^{-\alpha}\mu_{\alpha}m_1^{\alpha+1}\right].$$

Third, consider the expansion $\widehat{B}_0 = B_0 + B_{1/2} + O_p\left(n^{-1}\right)$. We have

$$\begin{aligned}\widehat{B}^{-1/2} &= \left[B_0 + B_{1/2} + O_p\left(n^{-1}\right)\right]^{-1/2}, \\ &= B_0^{-1/2} - \frac{1}{2}B_0^{-3/2}B_{1/2} + O_p\left(n^{-1}\right).\end{aligned}$$

The term $B_{1/2}$ is derived by centering and collecting terms of the same order. It then follows that $B_{1/2} = [n^{-1}\sum_i Y_{4,i}]$

Fourth, consider the term $\left[m_{\alpha}m_1 - \mu_1^{-\alpha}\mu_{\alpha}m_1^{\alpha+1}\right]$ by expanding the functions of the sample moments. Putting everything together and collecting terms of the same order, it follows that $S = S_0 + S_{1/2} + O_p\left(n^{-1}\right)$ with

$$S_0 = n^{1/2}B_0^{-1/2}\left[n^{-1}\sum Y_{2,i}\right], \tag{21}$$

$$\begin{aligned}S_{1/2} &= n^{1/2}B_0^{-1/2}\left[n^{-1}\sum_i Y_{1,i}\right]\left[n^{-1}\sum_j Y_{3,j}\right] \\ &\quad -n^{1/2}\frac{1}{2}B_0^{-3/2}\left[n^{-1}\sum_i Y_{2,i}\right]\left[n^{-1}\sum_k Y_{4,k}\right].\end{aligned}$$

### A.1.2 The Asymptotic Bias Term $k_{1,2}$

Taking expectations of the individual terms of (18) immediately yields, because of centering, $E(S_0) = 0$, and $E(S_{1/2}) = n^{-1/2}(B^{-1/2}E(Y_1Y_3) - 0.5B^{-3/2}E\left(Y_2Y_4\right))$. Since $E\left(S\right) = n^{-1/2}k_{1,2} + O\left(n^{-1}\right)$ it follows immediately that

$$k_{1,2} = B_0^{-1/2}E(Y_1Y_3) - \frac{1}{2}B_0^{-3/2}E\left(Y_2Y_4\right), \tag{22}$$

with $E(Y_1Y_3) = M_2$ and $E\left(Y_2Y_4\right) = M_5$ stated explicitly in Proposition 1.

### A.1.3 The Asymptotic Skewness Term $k_{3,1}$

In order to derive the asymptotic skewness term, we first need to obtain an expansion of the third moment of $S$. We take expectations of

$$S^3 = \left(S_0 + S_{1/2} + O_p\left(n^{-1}\right)\right)^3 = S_0^3 + 3S_0^2 S_{1/2} + O_p\left(n^{-1}\right).$$

by considering the constituent parts separately.

1. $E\left(S_0^2 S_{1/2}\right) = n^{3/2} B_0^{-3/2} E\left(n^{-4}\sum_i \sum_j \sum_k \sum_l Y_{2,i} Y_{2,j} Y_{1,k} Y_{3,l}\right) +$
   $-0.5 n^{3/2} B_0^{-5/2} \times E\left(n^{-4}\sum_i \sum_j \sum_k \sum_l Y_{2,i} Y_{2,j} Y_{2,k} Y_{4,l}\right)$.

   Since we are only interested in the $O\left(n^{-1/2}\right)$ term, we conclude that

$$\begin{aligned} E\left(S_0^2 S_{1/2}\right) &= n^{-1/2} B_0^{-3/2} \times \\ &\quad \left[E\left(Y_2^2\right) E\left(Y_1 Y_3\right) + 2E\left(Y_1 Y_2\right) E\left(Y_2 Y_3\right) - \frac{3}{2} E\left(Y_2 Y_4\right)\right] \\ &\quad + O\left(n^{-1}\right), \end{aligned}$$

   after noting that $E\left(Y_2^2\right) = B_0$.

2. Consider $S_0^3 = n^{3/2} B_0^{-3/2} n^{-3} \left(\sum Y_{2,i}\right)^3$. Hence $E(S_0^3) = n^{-1/2} B_0^{-3/2} E(Y_2^3) + O\left(n^{-1}\right)$.

In summary

$$\begin{aligned} E(S^3) &= n^{-1/2} B^{-3/2} \times \\ &\quad \left(E(Y_2^3) + 3\left[E\left(Y_2 Y_2\right) E\left(Y_1 Y_3\right) + 2E\left(Y_1 Y_2\right) E\left(Y_2 Y_3\right) - \frac{3}{2} E\left(Y_2 Y_4\right)\right]\right) \\ &\quad + O\left(n^{-1}\right). \end{aligned} \tag{23}$$

Finally, since $K = n^{-1/2} k_{31} + O\left(n^{-3/2}\right)$, and $K_3 = E\left(S^3\right) - 3E\left(S^2\right) E\left(S\right) + 2\left(E\left(S\right)\right)^3$, using $E\left(S^2\right) = 1 + O\left(n^{-1}\right)$, (22) and (23) we conclude that

$$k_{3,1} = B_0^{-3/2}\left[E(Y_2^3) + 6E\left(Y_1 Y_2\right) E\left(Y_2 Y_3\right) - 3E\left(Y_2 Y_4\right)\right], \tag{24}$$

where $E(Y_2^3) = M_4$, $E\left(Y_1 Y_2\right) = M_1$, and $E\left(Y_2 Y_3\right) = M_3$ are stated explicitly in Proposition 1.

# B Simulation Evidence for $k_{1,2}$ and $k_{3,1}$

This section provides a comparison of the population bias and skewness coefficients $k_{1,2}$ and $k_{3,1}$ as defined in Proposition 6 and simulated $k$-statistics. The experiments are designed as follows. We draw $R$ independent samples of size $n$ from income distribution $F$ and index the iteration by subscript $r$ with $r = 1, 2, ..., R$. The resulting studentized inequality measure is denoted $S_r$.

The scaled $k$-statistics are defined as follows. Consider the first cumulant of $S$ for which we have $K_1 = n^{-1/2}k_{1,2} + O\left(n^{-3/2}\right)$ or

$$k_{1,2} = n^{1/2}K_1 + O\left(n^{-1}\right).$$

The cumulant $K_1$ is simulated using the $k$-statistic $\widehat{K}_1 = R^{-1}\sum_r S_r$ with $\widehat{K}_1 = K_1 + O_p\left(R^{-1/2}\right)$. Therefore

$$\begin{aligned} k_{1,2}^{sim} &\equiv n^{1/2}\widehat{K}_1, \\ &= k_{1,2} + O\left(n^{-1}\right) + O_p\left(n^{1/2}R^{-1/2}\right). \end{aligned}$$

Similarly

$$\begin{aligned} k_{3,1}^{sim} &\equiv n^{1/2}\widehat{K}_3, \\ &= k_{3,1} + O\left(n^{-1}\right) + O_p\left(n^{1/2}R^{-1/2}\right), \end{aligned}$$

where $\widehat{K}_3 = rR^{-1}\sum_r \left(S_r - R^{-1}\sum_r S_r\right)^3$ with correction factor $r = R^2/[(R-1)(R-2)] \to 1$ which ensures unbiasedness of this $k$-statistic.

Figure 9 depicts both $k_{1,2}$ and $k_{1,2}^{sim}$, and $k_{3,1}$ and $k_{3,1}^{sim}$ as functions of the sensitivity parameter $\alpha$ of the inequality measure for the Singh-Maddala SM(.,3.5, 3.5) income distribution. The simulated values are based on $n$=$10^3$ and $R$=$10^6$ replications. The simulated values are in good agreement with the theoretical values. We have repeated these experiment for various income distributions and arrive at similar conclusions.

Figure 9: about here

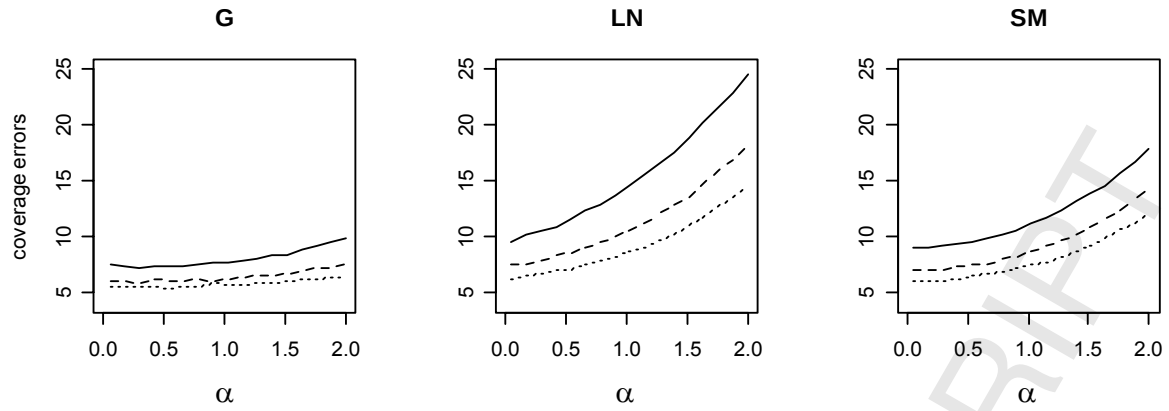


Figure 1: Actual coverage error rates of two sided symmetric Gaussian confidence intervals for $I(\alpha)$ for nominal rate of 5%, for varying  and different sample sizes. Notes: Income distributions as for Table 1. The solid line refers to samples of size 100, the dashed lined to sizes 250, and the dotted line to sizes 500. All simulations are based on $10^6$ replications.

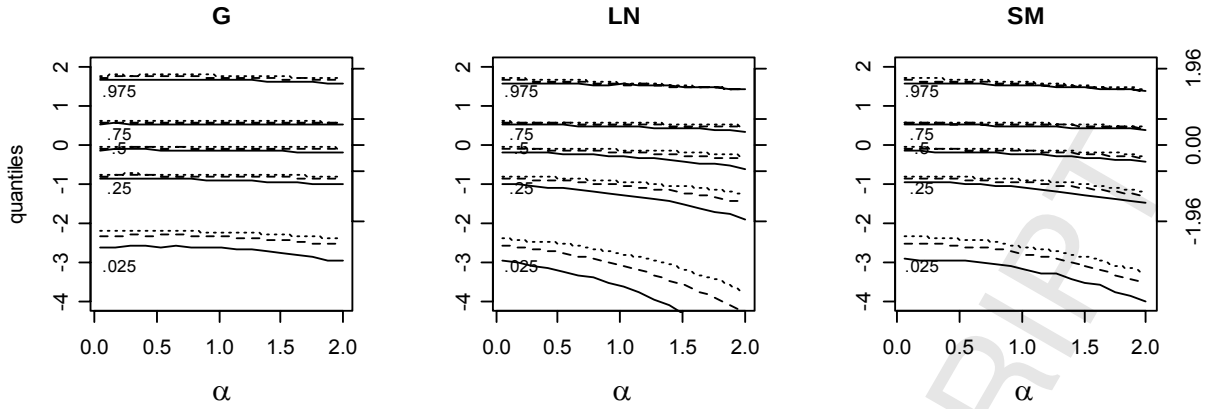


Figure 2: Quantiles of the studentised inequality measure $I(\ )$ for varying $\alpha$ and samples of different sizes. Notes: As for Figure 1.

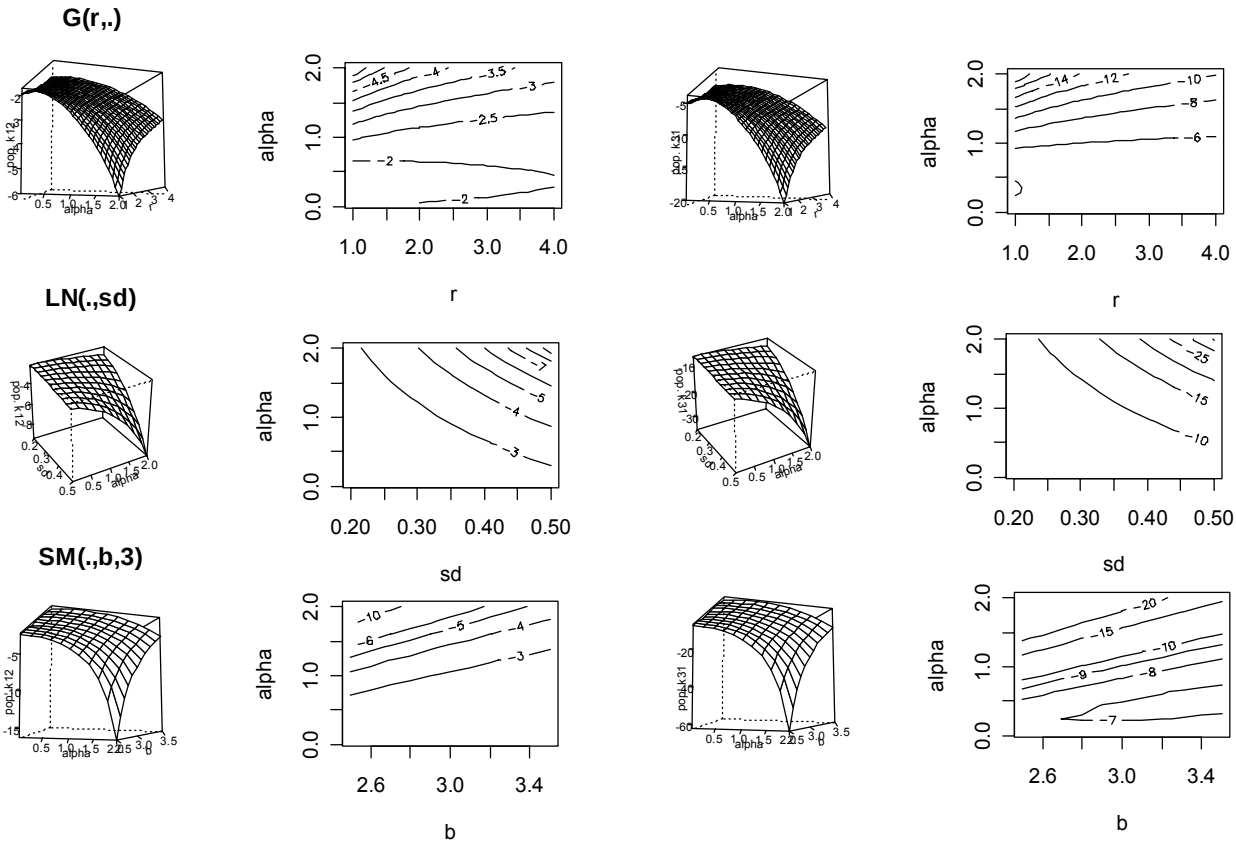


Figure 3: Bias and skewness coefficients $k_{1,2}$ and $k_{3,1}$ as functions of relevant parameters of the income distributions, and of the sensitivity parameter of the inequality index.

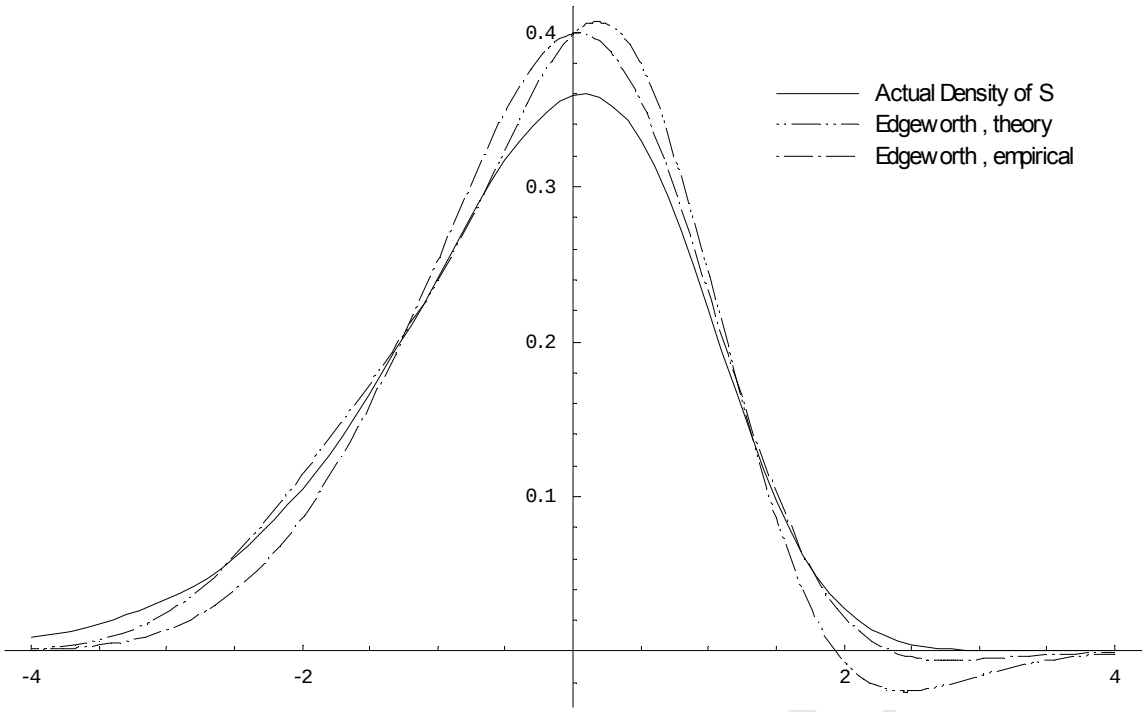


Figure 4: Density estimates. Notes: The income distribution is G(3,.), the sensitivity parameter of the inequality index is $\alpha = 2$, and the sample size is $n = 100$. The solid line depicts the simulated density of S, the first dashed line (– · · –) depicts the Edgeworth density based on the theoretical $k_{1,2}$ and $k_{3,1}$, and the second dashed line (– · –) is the Edgeworth density based on the estimated $k_{1,2}$ and $k_{3,1}$. Kernel density estimates based on $10^5$ replications.

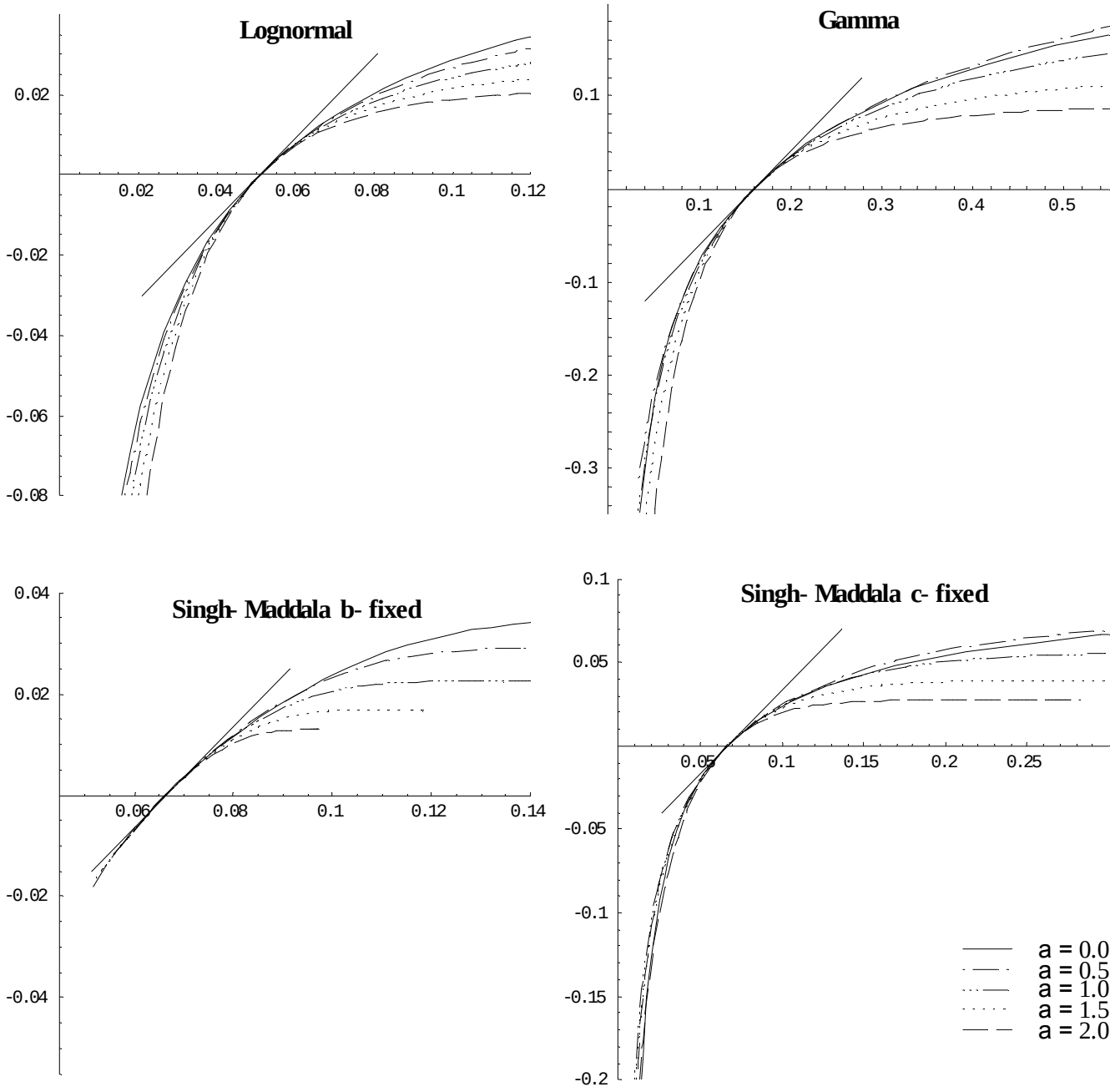


Figure 5: Transforms for different distributions and α's. Notes: On the horizontal axis is I (itself a function of parameters), and on the vertical axis is $t(I)$. The Singh-Maddala case at the bottom requires a restriction and $b = 3.5$ is chosen in the left hand panel and $c = 3.5$ in the right hand panel.

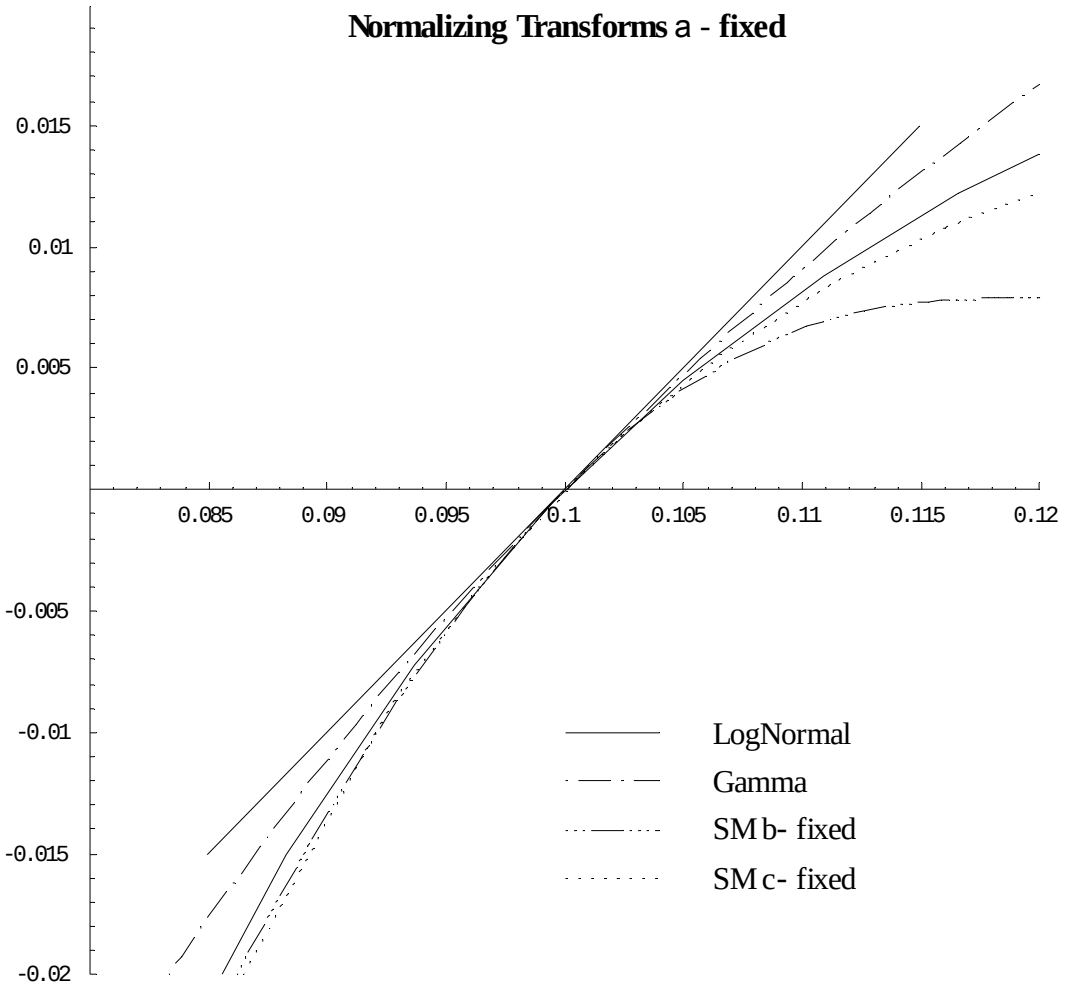


Figure 6: Normalizing Transforms for different distributions and fixed $\alpha = 2$. Notes: On the horizontal axis is I (itself a function of parameters), and on the vertical axis is $t(I)$. For the two Singh-Maddala cases $b = 3$ and $c = 3$ fixed are chosen.

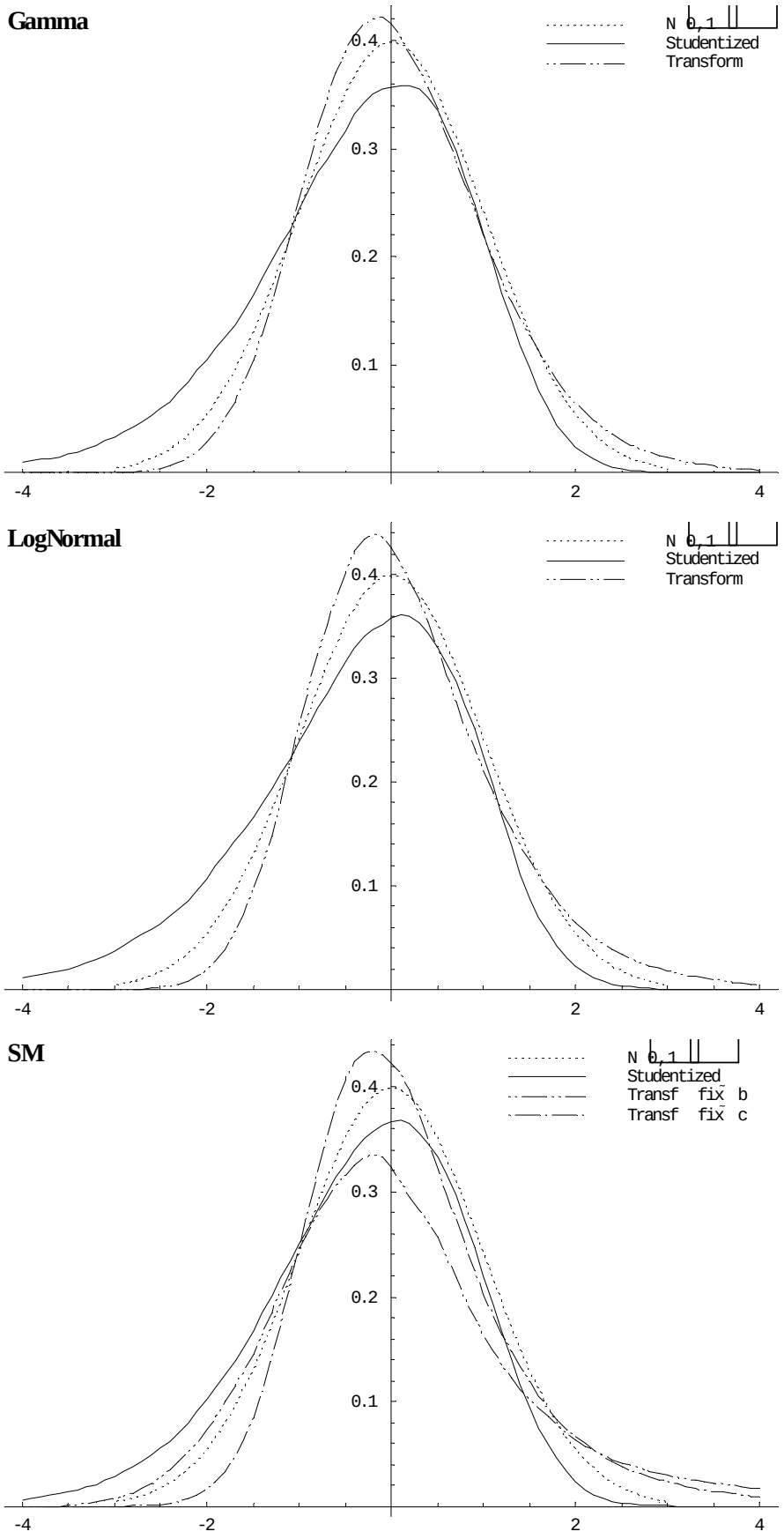


Figure 7: Simulated densities of $S$ and $T$ for various income distributions. Notes: Sample size is $n$ = 100. Parameter values: Gamma: r = 3, Lognormal: $v^2 = 0.1$, Singh-Maddala: $b$ = 3.5, $c$ = 3. Kernel density estimates based on $10^5$ replications.

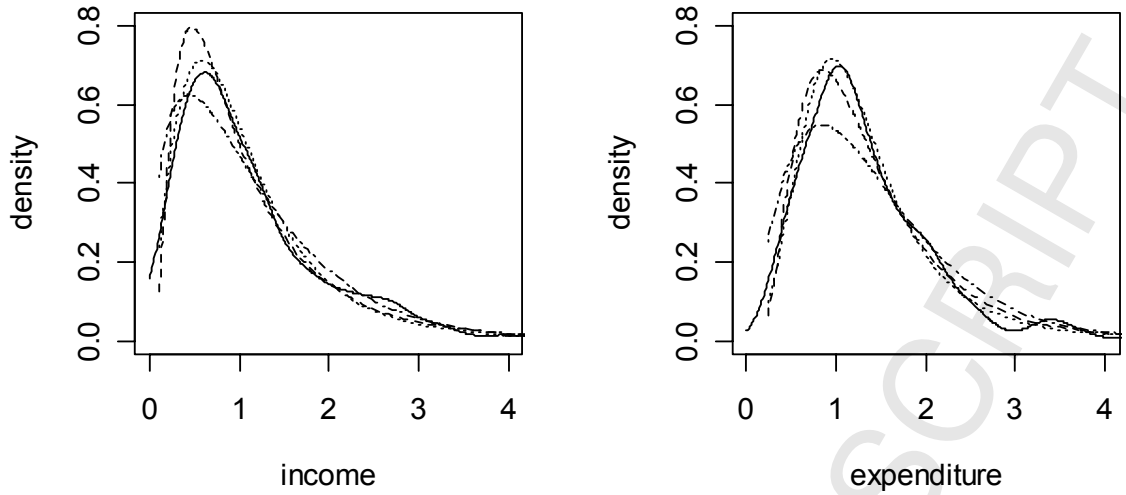


Figure 8: Income and expenditure densities for households in region 4. Notes: Incomes and expenditures are in Central African Francs divided by $10^6$. Non-parametric kernel estimate (solid line), fitted SM density (· · · ), fitted LN density ( – –), and fitted G density ( – · – ). The maximum likelihood point estimates for the income distributions are SM($1.4\times10^6$,1.86,1.92), LM(13.7,0.78), G($1.6,1.3\times10^{-6}$), and for expenditure are SM($1.7\times10^6$,2.49,1.96), LM(13.99,0.58), G($2.47,1.8\times10^{-6}$).

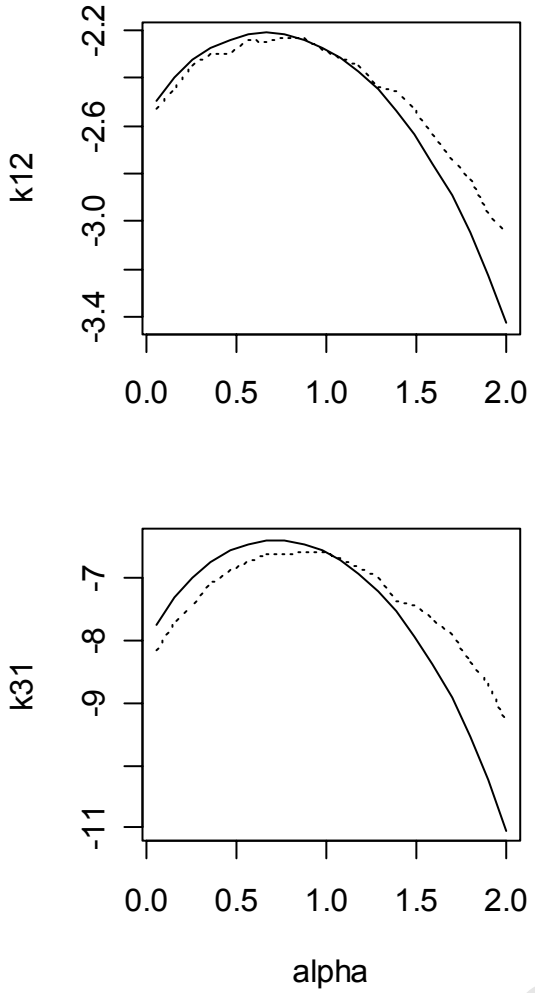


Figure 9: Theoretical bias and skewness coefficients and simulated k-statistics for the SM(.,3.5,3.5) income distribution. Notes: The solid lines depict the population coefficients $k_{1,2}$ and $k_{3,1}$ as a function of $\alpha$. The dashed lines are the simulated k-statistics with $n = 1000$ and $R = 10^6$ repetitions.