J. Econ. (2002) Suppl. 9: 261-281

Journal of Economics
Zeitschrift für Nationalökonomie

Bootstrapping Inequality Measures Under the Null Hypothesis: Is it Worth the Effort?

Mark Trede

This paper discusses methods of statistical inference for inequality measures, in particular the nonparametric bootstrap. Standard resampling techniques and a new method for nonparametric resampling under the null hypothesis are discussed. Monte-Carlo simulations show that some bootstrap methods outperform the commonly used normal approximation while other bootstrap methods—including those which are used in most empirical applications—are hardly any better.

Keywords: Statistical Inference, Resampling, Bootstrap.

JEL classification: C12, D31, D63.

1 Introduction

Nonparametric statistical inference for inequality measures is mostly based on the asymptotic normality of moments, quantiles etc.[1] An alternative to approximating the distribution of inequality estimators by a normal distribution is the bootstrap.[2] There is by now a well developed literature concerning the asymptotic properties of bootstrap techniques. It turns out that from a theoretical point of view the bootstrap is at least as appropriate as the normal approximation.[3] However, in the context of inequality measurement two important questions arise concerning the bootstrap:

First, which bootstrap? There are various ways to resample and also various ways to evaluate the resamples. This paper concentrates on the nonparametric bootstrap and sketches parametric methods

[1] See, e.g., Cowell (1995, p. 156), Nygård and Sandström (1981), chap. 10.
[2] Cf. Efron and Tibshirani (1993).
[3] Cf. Hall (1992).

only briefly. In the framework of hypothesis testing, the usual nonparametric bootstrap is based on resamples drawn under the alternative hypothesis. This paper also presents a resampling technique for resampling under the null hypothesis.

Second, is the bootstrap really worth the effort? Bootstrap methods are computationally intensive and there is no guarantee that they actually perform any better than the normal approximation in small or medium samples.

This paper tackles both questions. The structure is as follows: Section 2 introduces notation and briefly reviews the results on the asymptotic normality of inequality measures. Section 3 gives an overview of resampling techniques. Both the parametric and the nonparametric resampling scheme are described—with special emphasis on nonparametric resampling under the null hypothesis. Section 4 discusses in some detail how to do hypothesis tests with the bootstrap resamples. Section 5 compares the performance of various inferential methods by Monte Carlo simulations. Section 6 concludes.

2 Notation and Asymptotics

Let $F(y) = P(Y \leq y)$ be the (unknown) income distribution function. Since many inequality measures cannot cope with negative values we assume that $F(0) = 0$. Let Y_i, $i = 1, \ldots, n$, be an i.i.d. sample from F. The empirical distribution function is

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1\left(Y_i \leq y\right), \tag{2.1}$$

where $1(c) = 1$ if c is true and $1(c) = 0$ otherwise. Scalar inequality measures will be denoted by $I(\cdot)$. The argument will either be the random variable Y itself, its distribution function F or a sample $Y_1, \ldots, Y_n$, depending on which is most convenient. Estimators of $I(\cdot)$ are written as $\hat{I}(\cdot)$.

The objective of statistical inference in the area of inequality measurement is to derive (at least an approximation of) the distribution of $\hat{I}$. Once its distribution is established one can readily construct confidence intervals and perform hypothesis tests. The asymptotic distribution is particularly easily found if the inequality index is estimable by the method of moments. This class of measures includes, for instance, the generalized entropy index, Atkinson's index, the coefficient of variation and Theil's indices (the Gini coefficient is not a function of moments of Y).

Let μ be a vector of moments of Y (or of some transformations of Y) and $\hat{\mu}$ its estimator. Under some mild regularity conditions (which can safely be assumed to hold when we deal with income distributions) $\hat{\mu}$ is asymptotically jointly normally distributed,

$$\sqrt{n}(\hat{\mu} - \mu) \overset{asy}{\sim} N(0, \Sigma_\mu) \tag{2.2}$$

where Σ_μ is the asymptotic covariance matrix which can be consistently estimated from the data. Since inequality measures are differentiable functions of μ their distribution is asymptotically normal as well. The asymptotic variance can be derived by the delta method; let

$$\mathsf{D}I(\mu) := \frac{\partial I(\mu)}{\partial \mu},$$

then

$$\sqrt{n}(I(\hat{\mu}) - I(\mu)) \overset{asy}{\sim} N(0, \sigma^2)$$

where $\sigma^2 = \mathsf{D}I(\mu)'\Sigma_\mu \mathsf{D}I(\mu)$. A consistent estimator of the variance is

$$\hat{\sigma}^2 = \mathsf{D}I(\hat{\mu})'\hat{\Sigma}_\mu \mathsf{D}I(\hat{\mu}). \tag{2.3}$$

The Gini-Index and some related inequality indices are not estimable by the method of moments. However, using results on U-statistics and their asymptotic distribution, it can be shown that these indices are asymptotically normally distributed as well.[4]

Suppose we wish to test the two-sided hypothesis

$$\begin{aligned} H_0 &: I = I_0 \\ H_1 &: I \neq I_0. \end{aligned}$$

Using the test statistic[5] $\sqrt{n}\left|\hat{I} - I_0\right|/\hat{\sigma}$ the asymptotic p-value is, of course,

$$p = 2 - 2\Phi\left(\sqrt{n}\left|\frac{\hat{I} - I_0}{\hat{\sigma}}\right|\right)$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and we reject the null hypothesis if the p-value is less than the prescribed level α. Since the p-value is only asymptotically valid there is in

[4]Cf. Hoeffding (1948).

[5]The notation does not distinguish between the test statistic (as a random variable) and its realization. In the same way $\hat{I}$ denotes both the inequality estimator and its realization. The meaning should be clear from the context.

general a difference between p and the true probability that the test statistic will be larger than the realized value. The error is called the size distortion or error in rejection probability, ERP.[6] It is obviously desirable that the distortion vanishes as quickly as possible for $n \to \infty$. The rate of convergence can be discovered by Edgeworth expansions.[7]

Approximate confidence intervals with level $1-\alpha$ are given by

$$CI = \left[\hat{I} + n^{-1/2} u_{\alpha/2} \hat{\sigma}; \hat{I} + n^{-1/2} u_{1-\alpha/2} \hat{\sigma}\right],$$

where u_α is the α-quantile of the standard normal distribution. These confidence intervals, too, are only asymptotically valid. In finite samples, the probability that the interval covers the true value of the inequality index is in general not equal to $1-\alpha$.

3 Resampling Mechanisms

Let $I = I(F)$ be an inequality index for income distribution function F. The "plug-in"-estimator of I is $\hat{I} = I(\hat{F})$ where $\hat{F}$ is the estimated distribution function of Y. Statistical inference is only possible if the distribution function of $\hat{I}$ is known (at least approximately). The previous section dealt with the asymptotic distribution of $\hat{I}$. The small sample distribution is in most cases analytically intractable. The bootstrap is a versatile and attractive way to approximate the distribution of $\hat{I}$. It is also applicable in complex sampling situations where the normal approximation formulae are cumbersome.[8] Bootstrap confidence intervals and test results are often more reliable in finite samples and, in addition, may have better asymptotic properties than normal ones.

Let $Y_1, \ldots, Y_n$ be the original sample from which the estimated distribution function $\hat{F}$ is computed. Let $Y_1^*, \ldots, Y_n^*$ be a pseudo sample (resample) from $\hat{F}$ (see below). Then $I^* = \hat{I}(Y_1^*, \ldots, Y_n^*)$ is the inequality in the resample. The basic idea of the bootstrap methodology is to interpret the original sample as the population and compute the distribution of I^*; in other words, we try to establish the distribution of I^* conditional on the original sample.[9] Given the distribution of I^* we may then draw conclusions about the distribution of $\hat{I}$. In practice, the distribution of I^* is usually derived by resampling. A large

[6] Cf. Davidson and MacKinnon (1999).
[7] See Hall (1992) and Davidson and MacKinnon (1999).
[8] Cf. Biewen (1999).
[9] Cf. Hall (1992).

number, B, of resamples is generated, and each time the inequality index is calculated. We end up with B realizations $I_1^*, \ldots, I_B^*$.

Depending on (a) how much information is given about the true distribution function F and (b) whether we take the restrictions posed by the null hypothesis into account, we can distinguish the following approaches to resampling.

The next three cases do not take into account the null hypothesis. We will hence call them "resampling under the alternative hypothesis".

First, if the distribution function is known up to a finite dimensional parameter vector ρ the parametric bootstrap is appropriate. In this case the estimated distribution function is $\hat{F} = F_{\hat{\rho}}$ where $\hat{\rho}$ is an estimator of the parameter vector ρ. Resamples are drawn from $F_{\hat{\rho}}$.

Second, the smooth bootstrap is placed between the parametric and the nonparametric case. Using kernel density estimation we find the estimated density

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right)$$

where $K(\cdot)$ is a kernel (i.e., a density function) and h the bandwidth.[10] Hence the distribution function is estimated by

$$\hat{F}(y) = \int_{-\infty}^{y} \hat{f}(t)dt.$$

Fortunately, it is not necessary actually to integrate over $\hat{f}$. Drawing resamples from $\hat{F}$ is easy: Draw resamples from the original sample with replacement and add an independent random noise variable with density $K(y/h)/h$.[11]

Third, if no information at all is given about the form of the distribution function F one should use the nonparametric bootstrap. The distribution function is simply estimated by the empirical distribution function, $\hat{F} = \hat{F}_n$. Each resample is drawn from the original sample with replacement. In general, a resample contains some observations more than once and some not at all.

The parametric, semiparametric and nonparametric approaches may also be applied when the restriction $H_0 : I = I_0$ is used for resampling. The next three cases may therefore be termed "resampling under the null hypothesis".

[10]See Silverman (1986).

[11]Cf. Silverman and Young (1987).

Fourth, the modification of the resampling mechanism is easy if there is a parametric model; simply estimate the parameter vector ρ under the restriction given by H_0 and draw resamples from $F_{\hat{\rho}}$ as in the parametric bootstrap case.

Fifth, by using kernel density estimation under constraints[12] the smooth bootstrap may take the null hypothesis into account.[13]

Sixth, as to the nonparametric bootstrap, the resampling mechanism is modified by changing the selection probabilities.[14] The conventional nonparametric bootstrap resamples from the original sample with probabilities $(1/n)$ for each observation. If, for instance, the inequality measure is Theil's index, then the inequality in the original sample is

$$Th = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{\bar{Y}}\right)\ln\left(\frac{Y_i}{\bar{Y}}\right)$$

with $\bar{Y} = \frac{1}{n}\sum Y_i$ (analogously for other indices). Changing the selection probabilities allows us to force inequality in the sample to the value postulated in the null hypothesis. Let $w = (w_1, \ldots, w_n)$ be a vector of weights, $\sum_i w_i = 1$. Then inequality in the sample is

$$Th_w = \sum_{i=1}^{n} w_i \left(\frac{Y_i}{\bar{Y}_w}\right)\ln\left(\frac{Y_i}{\bar{Y}_w}\right),$$

with $\bar{Y}_w = \sum w_i Y_i$. Suitably chosen w_i yield $Th_w = I_0$. There is, however, no unique vector w. An attractive choice is

$$w = \arg\sup \prod w_i \tag{3.1}$$

subject to $\sum w_i = 1$ and $Th_w = I_0$. This can be achieved easily by numerical optimization.[15] The additional computational effort is rather small. The product $\prod w_i$ is also called the empirical likelihood.[16]

The resamples are drawn with selection probabilities w_i; some observations have higher probability to be included in the resample than others. We will write resamples drawn under the null hypothesis as $Y_1^{\#}, \ldots, Y_n^{\#}$. Note that resamples created this way may also be interpreted as being i.i.d. drawings from the weighted empirical distribution function $\hat{F}_w(y) := \sum_i w_i 1(Y_i \leq y)$.

[12]Cf. Hall and Presnell (1999).
[13]See also Silverman and Young (1987).
[14]Cf. Davison and Hinkley (1997), p. 165.
[15]Cf. Hall and LaScala (1990).
[16]Cf. Davison and Hinkley (1997).

In this paper we focus on the nonparametric bootstrap. Both the conventional resampling with selection probabilities $1/n$ and the mechanism for resampling under the null hypothesis are considered.

Having created the resamples one has to decide what to do with them. There are a number of different methods to build confidence intervals and perform hypothesis tests. They are discussed in the next section—mainly in the framework of two-sided hypothesis tests.

4 Bootstrap Methods

Having decided which resampling mechanism to use, a large number B of bootstrap resamples are generated. Let I_b^*, $b = 1, \ldots, B$, denote the inequality values for each resample if resampling is done in the conventional way (i.e., under the alternative hypothesis). Similarly, $I_b^{\#}$, $b = 1, \ldots, B$ are the bootstrapped inequality values when resampling under the null hypothesis. The number of bootstrap replications B should be chosen such that $\alpha(B+1)$ is an integer in order to avoid biases when estimating quantiles.[17] Let $I_{[1]}^* \leq \ldots \leq \hat{I} \leq \ldots \leq I_{[B]}^*$ be the $B+1$ order statistics with $\hat{I}$ appended to the resampled values; for notational convenience we define $I_{B+1}^* = \hat{I}$. Analogously for $I_b^{\#}$.

4.1 The "Other" Percentile Method

The "other" percentile method is widely used but often inappropriate or, in some cases, even wrong. Applications of the other percentile method to inequality measurement are, e.g., Mills and Zandvakili (1997) and Heinrich (1998a, 1998b). The resamples have to be drawn under the alternative; if they are drawn under the null hypothesis the other percentile method breaks down.

A $(1-\alpha)$ confidence interval is given by

$$CI = \left[I^*_{[(B+1)\alpha/2]}; I^*_{[(B+1)(1-\alpha/2)]}\right].$$

To perform a hypothesis test just check whether the confidence interval includes the postulated value I_0. If not, reject the null hypothesis. The p-value of the test is

$$p = \frac{2\min\{\#(I_b^* \geq I_0); \#(I_b^* \leq I_0)\}}{B+1}. \tag{4.1}$$

[17] See Davison and Hinkley (1997), chap. 4.

There are two arguments in favour of the other percentile method.[18] First, it is a transformation preserving method. Confidence intervals for a transformation of the statistic are identical to the transformed confidence intervals (where the transformation is applied to the limits of the confidence interval). Second, the other percentile method is also range preserving. Confidence intervals based on the other percentile method never extend to areas outside the support of the statistic under consideration.

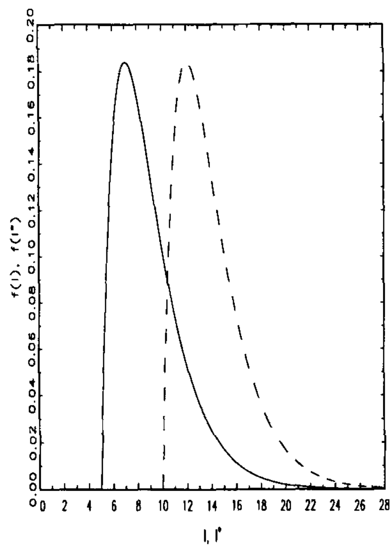

Fig. 4.1: Illustration of the Other Percentile Method.

Figure 4.1 illustrates why the other percentile method may be problematic. The solid line depicts the density function of some estimator $\hat{I}$. Let the true value be $I = 9$. Obviously, the estimator's distribution is skewed: Inequality is often underestimated by a rather small amount and sometimes severely overestimated.[19] Suppose, the estimated value is $\hat{I} = 14$. Then the bootstrap approximation of the distribution of I^* according to the other percentile method would roughly look like the dashed line. The true value $I = 9$ would falsely

[18]See Efron and Tibshirani (1993), chap. 13.6 and 13.7.

[19]This example is merely for illustration. However, many inequality measures display similar skew distributions when the sample size is small.

be judged as impossible since the support of I^* does not include values less than 10.

The figure suggests that unless the density of I^* is mirrored around $\hat{I}$ statistical inference will be useless. However, the other percentile produces correct results if the distribution under consideration is symmetric. Since the asymptotic distributions of inequality estimators are indeed symmetric the percentile method is at least asymptotically valid for inequality measures.

4.2 Percentile Method

The quantity of interest for the percentile method is $\sqrt{n}(\hat{I} - I_0)$ or, in the case of confidence intervals, $\sqrt{n}(\hat{I} - I)$. The bootstrap resamples are either drawn under the null or the alternative hypothesis. The percentile method has been applied to inequality measures e.g. by Schluter (1996).

4.2.1 Asymmetric Percentile Method

The asymmetric percentile method looks for two values c_1 and c_2 satisfying

$$P\left(-c_2 \leq \sqrt{n}\left(\hat{I} - I_0\right) \leq c_1\right) = 1 - \alpha \tag{4.2}$$

where c_1 and c_2 are chosen such that a probability mass of $\alpha/2$ each is to the left of $-c_2$ and to the right of c_1:

$$P\left(\sqrt{n}\left(\hat{I} - I_0\right) \leq c_1\right) = 1 - \alpha/2 \tag{4.3}$$

$$P\left(\sqrt{n}\left(\hat{I} - I_0\right) \leq -c_2\right) = \alpha/2. \tag{4.4}$$

If the realization of $\sqrt{n}(\hat{I} - I_0)$ is too small or too large we reject the null hypothesis. Of course, the true values of c_1 and c_2 are unknown, they will be estimated by the bootstrap. First consider resampling under the alternative. Then the bootstrap versions of (4.3) and (4.4) are

$$P\left(\sqrt{n}\left(I^* - \hat{I}\right) \leq c_1^*\right) = 1 - \alpha/2 \tag{4.5}$$

$$P\left(\sqrt{n}\left(I^* - \hat{I}\right) \leq -c_2^*\right) = \alpha/2. \tag{4.6}$$

Note that the distribution must be centered around $\hat{I}$, not I_0, since I^* is drawn under the alternative hypothesis. From (4.5) and (4.6) we find that c_1^* and $-c_2^*$ are the $(1-\alpha/2)$ quantile and $\alpha/2$ quantile of $\sqrt{n}(I^* - \hat{I})$, respectively:

$$\begin{aligned} c_1^* &= \sqrt{n}\left(I^*_{[(B+1)(1-\alpha/2)]} - \hat{I}\right) \\ -c_2^* &= \sqrt{n}\left(I^*_{[(B+1)(\alpha/2)]} - \hat{I}\right). \end{aligned}$$

The null hypothesis is rejected if $\sqrt{n}(\hat{I}-I_0) < -c_2^*$ or $\sqrt{n}(\hat{I}-I_0) > c_1^*$, i.e., if either

$$\begin{aligned} 2\hat{I} - I^*_{[(B+1)(\alpha/2)]} &< I_0 \\ \text{or} \quad 2\hat{I} - I^*_{[(B+1)(1-\alpha/2)]} &> I_0. \end{aligned}$$

The p-value of the test is

$$p = \frac{2\min\left\{\#\left(I_b^* - \hat{I} \geq \hat{I} - I_0\right); \#\left(I_b^* - \hat{I} \leq \hat{I} - I_0\right)\right\}}{B+1}.$$

If resampling is done under the null hypothesis, the bootstrap versions of (4.3) and (4.4) are

$$\begin{aligned} P\left(\sqrt{n}\left(I^\# - I_0\right) \leq c_1^\#\right) &= 1-\alpha/2 \\ P\left(\sqrt{n}\left(I^\# - I_0\right) \leq -c_2^\#\right) &= \alpha/2. \end{aligned}$$

In contrast to (4.5) and (4.6) we now have to center around I_0 since resampling is done under the null hypothesis. Obviously,

$$\begin{aligned} c_1^\# &= \sqrt{n}\left(I^\#_{[(B+1)(1-\alpha/2)]} - I_0\right) \\ -c_2^\# &= \sqrt{n}\left(I^\#_{[(B+1)(\alpha/2)]} - I_0\right). \end{aligned}$$

The null hypothesis is rejected if $\sqrt{n}(\hat{I}-I_0) < -c_2^\#$ or $\sqrt{n}(\hat{I}-I_0) > c_1^\#$, i.e., if either

$$\begin{aligned} \hat{I} &< I^\#_{[(B+1)(\alpha/2)]} \\ \text{or} \quad \hat{I} &> I^\#_{[(B+1)(1-\alpha/2)]}. \end{aligned}$$

The p-value of the test is

$$p = \frac{2\min\left\{\#\left(I_b^\# \geq \hat{I}\right); \#\left(I_b^\# \leq \hat{I}\right)\right\}}{B+1}. \tag{4.7}$$

Note that I_0 completely cancels out. The influence of I_0 is solely exerted via the weighted resampling mechanism.

To find a confidence interval for I substitute I_0 by I in (4.2) and solve for it. A confidence interval for I is given by

$$\begin{aligned} CI &= \left[\hat{I} - n^{-1/2}c_1^*; \hat{I} + n^{-1/2}c_2^*\right] \\ &= \left[2\hat{I} - I^*_{[(B+1)(1-\alpha/2)]}; 2\hat{I} - I^*_{[(B+1)\alpha/2]}\right]. \end{aligned}$$

The last equation shows that a long upper tail of I^* will produce a confidence interval with a small lower limit. If resampling is done under the null hypothesis, the whole family of tests generated by varying the value I_0 that corresponds to the test has to be inverted. Though being possible, this approach is highly computer intensive and not dealt with in this paper.

4.2.2 Symmetric Percentile Method

The symmetric percentile method determines a single value c satisfying

$$P\left(\sqrt{n}\left|\hat{I} - I_0\right| \leq c\right) = 1 - \alpha. \tag{4.8}$$

When resampling under the alternative the bootstrap version is

$$P\left(\sqrt{n}\left|I^* - \hat{I}\right| \leq c^*\right) = 1 - \alpha,$$

hence c^* is the $(1-\alpha)$ quantile of $\sqrt{n}\left|I^* - \hat{I}\right|$. Reject the null if $\sqrt{n}\left|\hat{I} - I_0\right|$ exceeds c^*. The p-value is

$$p = \frac{\#\left(\left|I_b^* - \hat{I}\right| \geq \left|\hat{I} - I_0\right|\right)}{B+1}.$$

When bootstrapping under the null hypothesis the bootstrap version of (4.8) is

$$P\left(\sqrt{n}\left|I^{\#} - I_0\right| \leq c^{\#}\right) = 1 - \alpha,$$

hence, $c^{\#}$ is the $(1-\alpha)$ quantile of $\sqrt{n}\left|I^{\#} - I_0\right|$. The null is rejected if $\sqrt{n}\left|\hat{I} - I_0\right|$ exceeds $c^{\#}$. The p-value is

$$p = \frac{\#\left(\left|I_b^{\#} - I_0\right| \geq \left|\hat{I} - I_0\right|\right)}{B+1}.$$

A confidence interval for I is

$$CI = \left[\hat{I} - n^{-1/2}c^*; \hat{I} + n^{-1/2}c^*\right].$$

4.3 Percentile-t Method

The exact distribution of $\hat{I}$ is unknown for finite samples. However, if the null hypothesis is true the distribution of

$$\sqrt{n}\left(\frac{\hat{I} - I_0}{\hat{\sigma}}\right) \tag{4.9}$$

converges to a standard normal distribution and, hence, does not depend on any unknown parameters. Therefore, (4.9) is asymptotically pivotal. Exploiting this property may improve the performance of the bootstrap. If the resamples are drawn under the alternative, the bootstrap version of (4.9) is

$$\sqrt{n}\left(\frac{I^* - \hat{I}}{\sigma^*}\right)$$

where σ^* is the standard deviation estimated from the resample (by Equation (2.3)). In the same way, the bootstrap version of (4.9) is $\sqrt{n}\left(I^{\#} - I_0\right)/\sigma^{\#}$ when resampling is done under the null hypothesis.

4.3.1 Asymmetric Percentile-t Method

Analogously to (4.3) and (4.4) we look for values c_1 and c_2 such that

$$P\left(\sqrt{n}\left(\frac{\hat{I} - I_0}{\hat{\sigma}}\right) \leq c_1\right) = 1 - \alpha/2 \tag{4.10}$$

$$P\left(\sqrt{n}\left(\frac{\hat{I} - I_0}{\hat{\sigma}}\right) \leq -c_2\right) = \alpha/2. \tag{4.11}$$

The corresponding bootstrap equations (when resampling under the alternative) are

$$\begin{aligned} P\left(\sqrt{n}\left(\frac{I^* - \hat{I}}{\sigma^*}\right) \leq c_1^*\right) &= 1 - \alpha/2 \\ P\left(\sqrt{n}\left(\frac{I^* - \hat{I}}{\sigma^*}\right) \leq -c_2^*\right) &= \alpha/2. \end{aligned}$$

Obviously, c_1^* and $-c_2^*$ are the $(1-\alpha/2)$ quantile and $\alpha/2$ quantile, respectively, of $\sqrt{n}(I^* - \hat{I})/\sigma^*$. These quantiles are estimated as follows: Generate B resamples; compute the asymptotic pivots

$$\tau_b^* = \sqrt{n}\left(\frac{I_b^* - \hat{I}}{\sigma_b^*}\right), \qquad b = 1, \dots, B, \tag{4.12}$$

where σ_b^* is the (asymptotic) standard deviation computed from the b-th resample. Let $\tau_{B+1}^* \equiv \hat{\tau} = \sqrt{n}(\hat{I} - I_0)/\hat{\sigma}$. Let $\tau_{[b]}^*, b = 1, \dots, B+1$ be the order statistics. Then $c_1^* = \tau^*_{[(B+1)(1-\alpha/2)]}$ and $c_2^* = -\tau^*_{[(B+1)(\alpha/2)]}$. The p-value of the test is

$$p = \frac{2\min\left\{\#\left(\tau_b^* \geq \tau_{B+1}^*\right); \#\left(\tau_b^* \leq \tau_{B+1}^*\right)\right\}}{B+1}$$

If the resamples are drawn under the null hypothesis the bootstrap versions of (4.10) and (4.11) are

$$\begin{aligned} P\left(\sqrt{n}\left(\frac{I^\# - I_0}{\sigma^\#}\right) \leq c_1^\#\right) &= 1 - \alpha/2 \\ P\left(\sqrt{n}\left(\frac{I^\# - I_0}{\sigma^\#}\right) \leq -c_2^\#\right) &= \alpha/2. \end{aligned}$$

Then, analogously to (4.12),

$$\tau_b^\# = \sqrt{n}\left(\frac{I_b^\# - I_0}{\sigma_b^\#}\right), \qquad b = 1, \dots, B,$$

are the asymptotic pivots. Let $\tau_{B+1}^\# = \sqrt{n}\left(\hat{I} - I_0\right)/\hat{\sigma}$ where $\hat{\sigma}$ is computed by (2.3). Again, $c_1^\#$ and $c_2^\#$ are quantiles of $\tau_b^\#$, $b = 1, \dots, B+1$. The p-value is

$$p = \frac{2\min\left\{\#\left(\tau_b^\# \geq \tau_{B+1}^\#\right); \#\left(\tau_b^\# \leq \tau_{B+1}^\#\right)\right\}}{B+1}.$$

A confidence interval for I is given by

$$CI = \left[\hat{I} - c_1^* n^{-1/2}\hat{\sigma}; \hat{I} + c_2^* n^{-1/2}\hat{\sigma}\right].$$

4.3.2 Symmetric Percentile-t Method

Similarly to (4.8) we look for a value c such that

$$P\left(\sqrt{n}\left|\frac{\hat{I}-I}{\hat{\sigma}}\right| \leq c\right) = 1-\alpha.$$

Under the alternative the bootstrap version is

$$P\left(\sqrt{n}\left|\frac{I^{*}-\hat{I}}{\sigma^{*}}\right| \leq c^{*}\right) = 1-\alpha,$$

i.e., c^{*} is the $(1-\alpha)$ quantile of $\sqrt{n}\left|I^{*}-\hat{I}\right|/\sigma^{*}$. The quantile is estimated using

$$\tau_b^{*} = \sqrt{n}\left|\frac{I_b^{*}-\hat{I}}{\sigma_b^{*}}\right|, \qquad b=1,\ldots,B,$$

as described above. The p-value of the test is

$$p = \frac{\#\left(\tau_b^{*} \geq \tau_{B+1}^{*}\right)}{B+1}.$$

Under the null hypothesis the bootstrap version is

$$P\left(\sqrt{n}\left|\frac{I^{\#}-I_0}{\sigma^{\#}}\right| \leq c^{\#}\right) = 1-\alpha,$$

i.e., $c^{\#}$ is the $(1-\alpha)$ quantile of $\sqrt{n}\left|I^{\#}-I_0\right|/\sigma^{\#}$. The quantile is estimated using

$$\tau_b^{\#} = \sqrt{n}\left|\frac{I_b^{\#}-I_0}{\sigma_b^{\#}}\right|, \qquad b=1,\ldots,B,$$

as described above. The p-value of the test is

$$p = \frac{\#\left(\tau_b^{\#} \geq \tau_{B+1}^{\#}\right)}{B+1}.$$

A $(1-\alpha)$ confidence interval is given by

$$CI = \left[\hat{I} - c^{*}n^{-1/2}\hat{\sigma}; \hat{I} + C^{*}n^{-1/2}\hat{\sigma}\right].$$

5 Simulations

An important argument in favour of the percentile-t bootstrap method is that the difference between the nominal and the actual size of the hypothesis tests (i.e., the error in rejection probability, ERP) converges to zero more quickly compared to the other methods.[20] However, the rate of convergence alone is not a sufficient criterion for evaluating methods of hypothesis testing. A quick rate of convergence is useless if the error is still large for the given sample size.

This section compares the other percentile method, the symmetric and asymmetric percentile methods, the symmetric and asymmetric percentile-t methods and the normal approximation in terms of the size of the hypothesis tests when the sample size is small ($n = 50$) or medium ($n = 500$). Following Davidson and MacKinnon (1999) we will display ERP plots. These plots show the difference between nominal size α and the actual size of the test as a function of α.

The Monte-Carlo simulations consist of six steps.

1. Specify an income distribution function F. The main features of empirical distributions are skewness, unimodality and a thick upper tail. As shown by Brachmann et al. (1996) a reasonably good fit to empirical income distributions can be achieved by the Singh-Maddala distribution. All simulations in this paper are based on the distribution function

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

with parameters $a = 100$, $b = 2.8$ and $c = 1.7$. The density function is shown in Figure 5.1. Apart from scale this distribution closely mirrors the monthly net income distribution of households in Germany.

2. Compute the true value of the inequality measure. Inequality is measured by Theil's index in these simulations. The true value I_0 is:[21]

$$\begin{aligned} I_0 &= Th(F) \\ &= \int_0^\infty \left(\frac{y}{\mu}\right) \ln \left(\frac{y}{\mu}\right) dF(y) \\ &= \frac{\psi\left(\frac{1}{b}+1\right) - \psi\left(c - \frac{1}{b}\right)}{b} - \ln\left(cB\left(\frac{1}{b}+1, c-\frac{1}{b}\right)\right) \\ &= 0.1401 \end{aligned}$$

[20] Cf. Hall (1992).

[21] Cf. Schader and Schmid (1988).

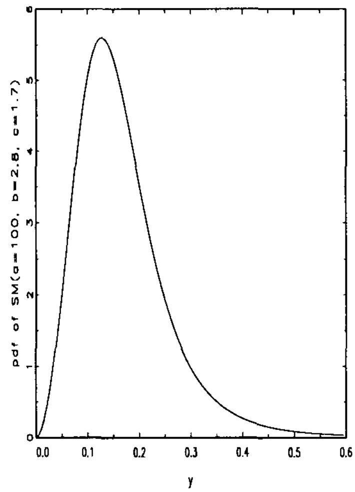

Fig. 5.1: Density of the Singh-Maddala Distribution.

where $\psi(\cdot)$ is the digamma function and $B(\cdot,\cdot)$ the beta function:[22]

3. Draw a sample $Y_1, \ldots, Y_n$ of size n from F.

4. Based on the sample compute and store the p-value of the test under consideration. The null hypothesis is $H_0 : I = 0.1401$ against $H_1 : I \neq 0.1401$. The bootstrap tests are based on $B = 999$ replications. Power is not considered in this paper.

5. Repeat steps 3 and 4 a large number of times, $N = 100\,000$.

6. Display the ERP plot.

The Monte-Carlo simulations are very computer intensive. For example, if we want to graph the ERP plot of the percentile-t method, then the inequality index has to be computed $N \times B \approx 10^8$ times. Thus, efficient and fast programming is indispensable. Notice that the short-cut for Monte-Carlo simulations of bootstrap tests suggested by Davidson and MacKinnon (1999) is not applicable here because $\hat{I}$ and I^* are not independent.

[22] Cf. Abramowitz and Stegun (1972).

Figures 5.2 to 5.6 show the ERP plots of the normal approximation, the other percentile method, the symmetric and asymmetric percentile method and the symmetric and asymmetric percentile-t method, respectively. Each plot contains four lines: The two bottom lines are the plots for sample size $n = 500$, the two upper ones are for $n = 50$. The most important part of the plots is on the left hand side, in particular the ERP at $\alpha = 0.05$ and $\alpha = 0.1$ (indicated by vertical dotted lines). The ERP is positive in all plots meaning that the null hypothesis is rejected too often, sometimes far too often. Remember that a positive ERP is more harmful than a negative one (which indicates that the test is conservative). Of course, we see that increasing the sample size brings down the error.

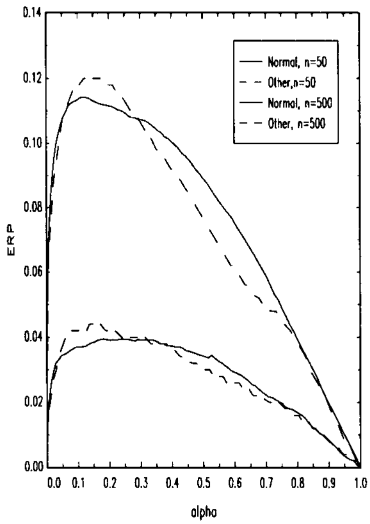

Fig. 5.2: ERP Plot for the Normal Approximation and the Other Percentile Method.

Figure 5.2 compares the normal approximation and the other percentile method. The difference is not large, for $\alpha = 0.05$ they are virtually identical. For $\alpha = 0.1$ the other percentile method is even somewhat worse than the normal approximation. Hence the use of the other percentile method in empirical applications is not advisable when better methods are available.

In Figures 5.3 to 5.6 solid lines represent bootstrapping under the null hypothesis whereas dashed lines give plots for bootstrapping under the alternative. As to the percentile method (Figures 5.3 and 5.4) we find that resampling under the null hypothesis decreases the ERP, sometimes substantially so. Only for $\alpha = 0.1$ and $n = 500$ there is hardly any difference when the asymmetric method is used. Note that the superiority of resampling under the null hypothesis sometimes vanishes for larger values of α (which are, however, of no practical importance). The percentile methods perform slightly worse than the normal approximation when resampling is done under the alternative.

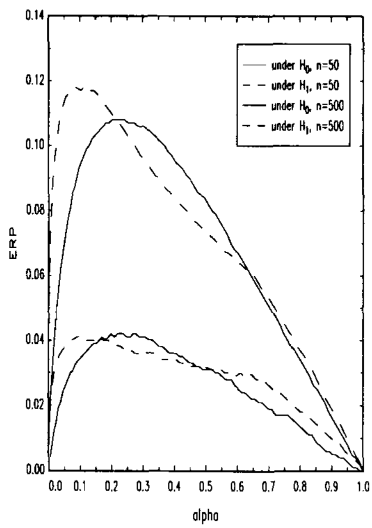

Fig. 5.3: ERP Plot for the Symmetric Percentile Method.

Concerning the percentile-t methods (Figures 5.5 and 5.6) their ERP plots are generally lower than for the other methods. This is an important lesson to be drawn from the simulations: Pivoting does improve the bootstrap, at least when we are dealing with inequality measures. This is in contrast to Mills and Zandvakili (1997) and may, in fact, not be true for other applications such as bootstrapping correlation coefficients.[23] A drawback of the percentile-t methods is that the asymptotic standard errors have to be calculated for each resample. For complex sampling designs this might be too difficult or time consuming.

[23]See Efron and Tibshirani (1993), chap. 12.5.

Resampling under the null hypothesis is much less advantageous for the percentile-t method than it is for the percentile method. Figure 5.6 shows that resampling under the null hypothesis actually deteriorates the performance.

Summarizing, we find that pivoting decreases the ERP; that bootstrapping under the null hypothesis is better than under the alternative hypothesis for the percentile method; that the other percentile method is roughly as inaccurate as the normal approximation.

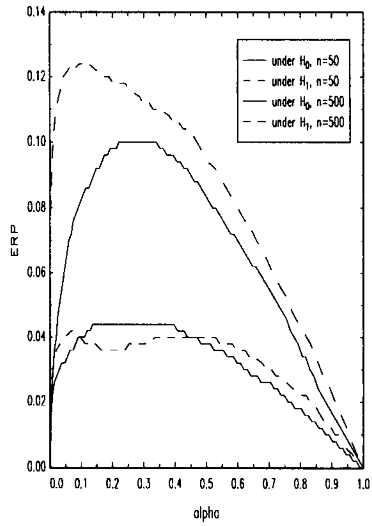

Fig. 5.4: ERP Plot for the Asymmetric Percentile Method.

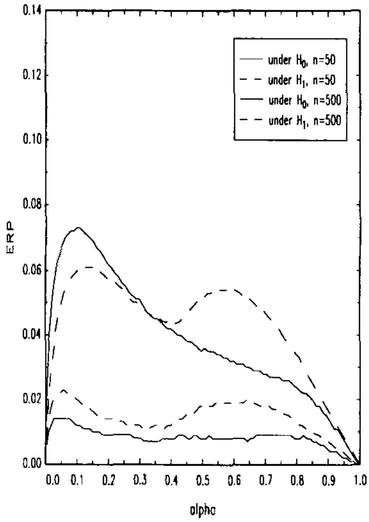

Fig. 5.5: ERP Plot for the Symmetric Percentile-t Method.

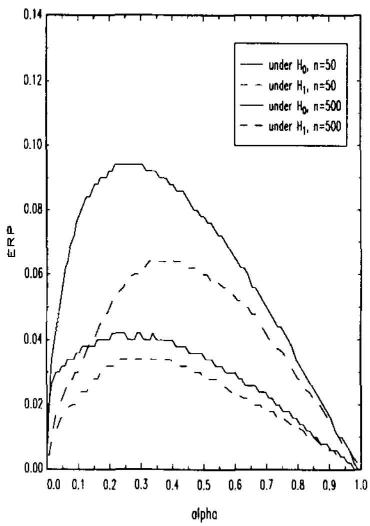

Fig. 5.6: ERP Plot for the Asymmetric Percentile-t Method.

6 Conclusion

Inequality measurement is mostly based on random samples. Thus methods of statistical inference are necessary to build confidence intervals and perform hypothesis tests. Most studies use the asymptotic distribution of the inequality index as an approximation even in small samples. There is not much evidence on which bootstrap method performs best for inequality measures.

This paper discusses resampling schemes and bootstrap methods, and demonstrates that they can be more reliable than the normal approximation in smaller samples. However, there are substantial differences in performance: The other percentile method which is mostly used in empirical studies performs badly. The ERP of the percentile method decreases considerably when resampling under the null hypothesis is used, while the percentile-t method performs better with resampling under the alternative.

The main finding of this paper is that using the bootstrap can improve statistical inference for inequality measures a lot. However, care must be taken as to which bootstrap method is used: The other percentile method works as poorly as the normal approximation. Resampling under the null hypothesis can improve the percentile method substantially.

References

Abramowitz, M., and Stegun, I. A. (eds.) (1972): *Handbook of Mathematical Functions.* New York: Dover Publications.

Biewen, M. (1999): "Bootstrap Inference for Inequality and Mobility Measurement." Technical Report 286, Universität Heidelberg, Department of Economics.

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

Cowell, F. A. (1995): *Measuring Inequality*, 2nd edn. London: Prentice Hall.

Davidson, R., and MacKinnon, J. G. (1999): "The Size Distortion of Bootstrap Tests."*Econometric Theory* 15: 361–376.

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

Efron, B., and Tibshirani, R. J. (1993): *An Introduction to the Bootstrap.* New York: Chapman and Hall.

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

Hall, P., and La Scala, B. (1990): "Methodology and Algorithms of Empirical Likelihood."*International Statistical Review* 58: 109–127.

Hall, P., and Presnell, B. (1999): "Density Estimation Under Constraints."*Journal of Computational and Graphical Statistics* 8: 259–277.

Heinrich, G. A. (1998a): "Changing Times, Testing Times: A Bootstrap Analysis of Poverty and Inequality Using the PACO Data Base." Technical Report, CERT, Department of Economics, Heriot-Watt University, Edinburgh.

Heinrich, G. A. (1998b): "Ageing Gracefully? A Bootstrap Analysis of Poverty Among Pensioners Using Evidence from the PACO Database." Technical Report 2039, Centre for Economic Policy Research, London.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution."*Annals of Mathematical Statistics* 19: 293–325.

Mills, J. A., and Zandvakili, S. (1997): "Statistical Inference via Bootstrapping for Measures of Inequality."*Journal of Applied Econometrics* 12: 133–150.

Nygård, F., and Sandström, A. (1981): *Measuring Income Inequality.* Stockholm: Almquist and Wiksell.

Schader, M., and Schmid, F. (1988): "Zur Messung der relativen Konzentration aus gruppierten Daten."*Jahrbücher für Nationalökonomie und Statistik* 204: 437–455.

Schluter, C. (1996): "Statistical Inference for Inequality Indices: The Role of Sample Size." Working Paper, STICERD, London School of Economics, London.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Silverman, B. W., and Young, G. (1987): "The Bootstrap: To smooth or not to smooth?"*Biometrika* 74: 469–479.

Address of author: Mark Trede, Seminar für Wirtschafts- und Sozialstatistik, Universität zu Köln, D-50923 Köln, Germany.