Bootstrap Inference for Inequality, Mobility and Poverty Measurement[1]

Martin Biewen

University of Essex and University of Heidelberg

This version: February 2001

Abstract.

This paper proposes the use of the bootstrap for the most commonly applied procedures in inequality, mobility and poverty measurement. In addition to simple inequality index estimation the scenarios considered are inequality difference tests for correlated data, decompositions by subgroup or income source, decompositions of inequality changes, and mobility- and poverty index estimation. Besides showing the consistency of the bootstrap for these scenarios, the paper also develops simple ways to deal with longitudinal correlation and panel attrition or non-response. In principle, all the proposed procedures can be handled by the δ-method, but Monte Carlo evidence suggests that the simplest possible bootstrap procedure should be the preferred method in practice, as it achieves the same accuracy as the δ-method and takes into account the stochastic dependencies in the data without explicitly having to deal with its covariance structure. If a variance estimate is available, then the studentized version of the bootstrap may lead to an improvement in accuracy, but substantially so only for relatively small sample sizes. All results incorporate the possibility that different observations have different sampling weights.

JEL-Classification: C14, D31

Keywords: Bootstrap, Inequality, Mobility, Poverty

Correspondence:

Martin Biewen, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, Fax: +44-1206-873151, e-mail: biewen@essex.ac.uk

---

[1] I would like to thank Enno Mammen, Christoph M. Schmidt, Joel Horowitz, Stephen Jenkins, Christian Schluter, Mark Bryan, seminar participants in Heidelberg, Aarhus and at the London School of Economics as well as three anonymous referees for many useful comments and suggestions. Any errors are my own. Partial funding by the European Union through the TMR Network 'Living Standards, Inequality and Taxation' (Contract #ERBFMRXCT980248) is gratefully acknowledged.

1 Introduction

Statistical inference for inequality, mobility and poverty measurement is - if it is carried out at all - usually based on the asymptotic distribution of the respective indices in connection with the δ-method (see e.g. COWELL(1989), NYGÅRD/SANDSTRÖM(1989), THISTLE(1990), KAKWANI(1993) and SCHLUTER(1998)). One difficulty arising in this context are possible dependencies in the data, for example in the form of longitudinal correlation. Inequality, mobility or poverty indices are often estimated using the cross-sections of a panel survey. To test whether the index under consideration has changed in a statistically significant way from one wave to another, it is necessary to take into account the intertemporal covariance structure of incomes. Although standard asymptotic methods are in principle capable of dealing with such difficulties (see DAVIDSON/DUCLOS(1997), WIEGAND(1998), VAN DE GAER ET AL.(1999) and DAVIDSON/DUCLOS(2000)), the resulting estimates usually require cumbersome covariance calculations.

An alternative to the discussed methods is the bootstrap. (For a general discussion of the bootstrap, see HALL(1992) and SHAO/TU(1995).) The bootstrap provides an estimate of the sampling distribution of a given statistic by resampling from the original sample, thus simulating the original sampling procedure. Besides having advantages in small samples under certain circumstances (see section 5), the bootstrap can take into account stochastic dependencies in multivariate data without explicitly dealing with its covariance structure. In the context of inequality measurement, the bootstrap was first applied by MILLS/ZANDVAKILI(1997). They calculate bootstrap confidence intervals for some inequality indices as well as for the components of a decomposition of the THEIL coefficient by subgroup and compare them with intervals obtained by the normal approximation, i.e. the δ-method.

In this paper, the validity of the bootstrap method is shown for these and many other settings in the context of inequality, mobility and poverty measurement. The results cover procedures involving indices which can be expressed in terms of population moments. The most important of these indices are the family of generalized entropy measures, ATKINSON indices, the coefficient of variation, the logarithmic variance, KOLM indices, MAASOUMI/ZANDVAKILI/-SHORROCKS mobility indices, PRAIS mobility indices and FOSTER/GREER/THORBECKE poverty measures.

The measurement scenarios considered are, besides simple inequality estimation, inequality difference tests for correlated data, decompositions by subgroup or income source, decompositions of inequality changes, and mobility- and poverty index estimation. As will be seen, the approach is also capable of dealing with specific forms of non-response or attrition in the context of longitudinal data. In contrast with much of the literature, all results are given for the case of weighted estimation.

In order to compare the performance of different variants of the bootstrap to that of the δ-method, comprehensive Monte Carlo results for the different scenarios and for realistic sample sizes are presented. This is necessary since the theoretical results in the literature (for an overview, see SHAO/TU(1995), pp. 91 -104 and pp. 141 -154) are asymptotic in nature and give no indication of the ranking of the different methods for fixed sample sizes. The simulations also allows one to assess in absolute terms the reliability of these methods for empirical work.

The rest of the paper is organized as follows. Section 2 reviews the most commonly used moment-based inequality measures. The validity of the bootstrap procedure for these measures is verified in section 3. Section 4 then develops extensions of the framework to other scenarios of inequality, mobility and poverty measurement. Section 5 provides Monte Carlo evidence on the performance of different versions of the bootstrap compared to that of the δ-method. Section 6 illustrates some of the techniques discussed using wage data from the German Socio-Economic Panel. Section 7 concludes.

2 Inequality Indices

In this section, it will be assumed that the observed data is of the form $y_i = (w_i, x_i)'$ for $i = 1 \ldots, n$, which can be interpreted as an i.i.d. sample of size n from a joint distribution G of w and x. Let w_i denote the weight of observational unit i and x_i its income. The strategy here is to model sampling from a finite population as i.i.d. draws from a distribution G of weights w and incomes x. The weights are thought to reweighten the cases with a particular value of x. For example, if incomes are equivalent household incomes and weights are household sizes, then equivalent incomes associated with a household size of $w = 2$ count twice as much as incomes associated with a household size of $w = 1$. Let $H(\cdot)$ denote the distribution function of income that results after the weights have been taken into account. The ultimate goal is to estimate functionals of this distribution, e.g. inequality indices. The relative frequency of cases with a particular value of income x in this distribution is then

$$dH(x) = \int_w w\, dG(w,x) / \int_x \int_w w\, dG(w,x), \tag{1}$$

i.e., for all possible values of weights w, the relative frequency of cases with weight w and income x, $dG(w,x)$, is inflated by the weight and the result is normalized by the integral over all weights. In the following, the distribution function $H(\cdot)$ in the usual formulae for inequality, mobility and poverty indices is substituted by the weighted expression given by (1). For example, mean income $\int_x x\, dH(x)$ would be given by $\int_x \int_w wx\, dG(w,x) / \int_x \int_w w\, dG(w,x)$.

The weight w_i may incorporate two considerations (compare COWELL(1989,2000)). First, if observational units are households, the equivalent income of every household member has to be reweighted by household size w_i' in order to obtain a distribution of income accross *individuals*. Second, households may have differing sampling weights $v_{i,n}''$. Note that sampling weights are usually inverse inclusion probabilities, which means that their scale depends on sample size n. For a given sampling design with a fixed sample size n, drawing a sample can then be modeled as drawing from a population (v_n, x) with $v_n = w'v_n''$, where w' denotes household size as above and v_n'' sampling weight. Sampling weights are constructed to gross up the sample to the size of the target population N, in other words, the expected size of the sampled population $E(\sum_{i=1}^n v_{i,n}'')$ is equal to N (subscripts i denote i.i.d. draws from (v_n, x)). The scaling factor implicit in the sampling weights is (N/n), i.e. the ratio of population size and sample size. (The larger the sample, the larger is the probability that a given unit will be included in the sample and the smaller is the sampling weight which is used to gross up this observation to the size of the target population.)

Now consider weights $w_n = w' \, [v_n''(n/N)] = w'w_n''$ which are obtained by replacing sampling weights by relative sampling weights w_n''. (Note that $E(\sum_{i=1}^n w_{i,n}'') = n$. For equal probability sampling $w_{i,n}'' \equiv 1$.) As a consequence of (1), all statistics in this paper are invariant with respect to the scale of the weights so that sampling from (v_n, x) is equivalent to sampling from (w_n, x). The asymptotics in the following sections are therefore valid for any sequence of sampling designs for which relative sampling weights are independent of sample size, i.e. $w_n'' = w''$ a.s., which implies $w_n = w$ a.s. (The idea is to fix a sufficiently rich sampling design and let inclusion probabilities vary proportionately to sample size). The resulting weight for observation i is then denoted as $w_i = w_i'w_i''$[2].

[2]This is done to keep the exposition consistent. As the scale of the weights does not matter numerically, it is not necessary to multiply the sampling weights $v_{i,n}''$ by (n/N) in practice. In fact, weights can be rescaled arbitrarily.

For the following, it is convenient to define the following population moments

$$\mu_\alpha = E(w_i x_i^\alpha), \tag{2}$$

$$\tau_{\alpha,\gamma} = E(w_i x_i^\alpha (\log x_i)^\gamma), \tag{3}$$

$$\eta = E(w_i \exp(\beta x_i)), \tag{4}$$

where the expectation is with respect to the joint distribution G of the observed data (compare COWELL(1989)). Note that μ_0 is the effective population size.

Unbiased estimates of these moments are provided by

$$m_\alpha = n^{-1} \sum_{i=1}^n w_i x_i^\alpha, \tag{5}$$

$$t_{\alpha,\gamma} = n^{-1} \sum_{i=1}^n w_i x_i^\alpha (\log x_i)^\gamma, \tag{6}$$

$$v = n^{-1} \sum_{i=1}^n w_i \exp(\beta x_i). \tag{7}$$

The family of generalized entropy measures with special cases THEIL coefficient and mean logarithmic deviation, the ATKINSON index with inequality aversion parameter ε , the coefficient of variation, the logarithmic variance and the KOLM index with sensitivity parameter β can be written in terms of population moments as

$$I_{GE}^\alpha(G) = (\alpha^2 - \alpha)^{-1} \left[\mu_0^{\alpha-1} \mu_1^{-\alpha} \mu_\alpha - 1\right], \quad \alpha \in \mathbb{R} \setminus \{0,1\} \tag{8}$$

$$I_{Theil}(G) = \tau_{1,1}\mu_1^{-1} - \log(\mu_1 \mu_0^{-1}), \tag{9}$$

$$I_{MLD}(G) = -\tau_{0,1}\mu_0^{-1} + \log(\mu_1 \mu_0^{-1}), \tag{10}$$

$$I_A^\varepsilon(G) = 1 - \mu_0^{-\varepsilon/(1-\varepsilon)} \mu_1^{-1} \mu_{1-\varepsilon}^{1/(1-\varepsilon)}, \quad \varepsilon \geq 0, \quad \varepsilon \neq 1, \tag{11}$$

$$I_A^1(G) = 1 - \mu_0 \mu_1^{-1} \exp(\tau_{0,1}\mu_0^{-1}), \tag{12}$$

$$I_{CV}(G) = \left[\mu_0 \mu_1^{-2} \mu_2 - 1\right]^{\frac{1}{2}}, \tag{13}$$

$$I_{logvar}(G) = \tau_{0,2}\mu_0^{-1} - \log(\mu_1\mu_0^{-1}) \left[2\tau_{0,1}\mu_0^{-1} - (\log \mu_1 \mu_0^{-1})\right], \tag{14}$$

$$I_K^\beta(G) = \beta^{-1} \left[\exp(-\beta \mu_1 \mu_0^{-1}) \mu_0^{-1} \eta - 1\right], \quad \beta > 0, \tag{15}$$

where (9) and (10) are the limiting forms of (8) for $\alpha \to 1$ and $\alpha \to 0$ and (12) is the limiting form of (11) for $\varepsilon \to 1$. For the definition and a discussion of these indices, see COWELL(2000).

Point estimates $I_{GE}^\alpha(G_n)$, $I_{Theil}(G_n)$, $I_{MLD}(G_n)$, $I_A^\varepsilon(G_n)$, $I_{CV}(G_n)$, $I_{logvar}(G_n)$, $I_K^\beta(G_n)$, are obtained by replacing the population moments $\mu_\alpha, \tau_{\alpha,\gamma}$ and η by their sample counterparts

$m_\alpha, t_{\alpha,\gamma}, v$ and k_α. These are the corresponding moments of the sample distribution function

$$G_n(x) = n^{-1} \sum_{i=1}^{n} 1\{y_i \leq x\} \tag{16}$$

($1\{\cdot\}$ denotes the indicator function).

3 Bootstrapping Inequality Indices

This is one of the scenarios considered in MILLS/ZANDVAKILI(1997). For the following, define $y_1^*, y_2^*, \ldots, y_n^*$ to be an i.i.d. bootstrap sample with the $y_i^* = (w_i^*, x_i^*)'$ having the distribution function G_n, which is the empirical distribution of the sample. Then the $I_n = I(G_n)$ is the inequality estimate and $I_n^* = I(G_n^*)$ is its bootstrap counterpart with $G_n^*(x) = n^{-1} \sum_{i=1}^n 1\{y_i^* \leq x\}$ the empirical distribution function of the bootstrap sample $y_1^*, y_2^*, \ldots, y_n^*$. Moreover $I = I(G)$.

Theorem. *Given that the observed data $y_i = (w_i, x_i)'$ takes values so that the given inequality index is defined, the bootstrap distribution function $H_{Boot}(x) = P\{\sqrt{n}(I_n^* - I_n) \leq x | y_1, y_2, \ldots, y_n\}$ can serve as a substitute for the sampling distribution function $H_n(x) = P\{\sqrt{n}(I_n - I) \leq x\}$ for $n \to \infty$, in the sense that* $\sup_x |H_{Boot}(x) - H_n(x)| \xrightarrow{a.s.} 0$*, for*

(i) $I = I_{GE}^\alpha$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i x_i^\alpha)^2) < \infty$ and $\mu_0 \neq 0$,

(ii) $I = I_{Theil}$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i x_i (\log x_i))^2) < \infty$,

(iii) $I = I_{MLD}$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i (\log x_i))^2) < \infty$,

(iv) $I = I_A^\varepsilon$ for $\varepsilon \neq 1$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i x_i^{1-\varepsilon})^2) < \infty$, $\mu_0 \neq 0$ and $\mu_{1-\varepsilon} \neq 0$,

(v) $I = I_A^1$ if $E((w_i)^2 + (w_i (\log x_i)^2) < \infty$,

(vi) $I = I_{CV}$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i x_i^2)^2) < \infty$ and $\left(\mu_0 \mu_1^{-2} \mu_2 - 1\right)^{-\frac{1}{2}} \mu_0 \mu_1^{-2} \neq 0$,

(vii) $I = I_{logvar}$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i (\log x_i))^2 + (w_i (\log x_i)^2)^2) < \infty$,

(viii) $I = I_K^\beta$ if $E((w_i)^2 + (w_i x_i)^2 + (w_i \exp(\beta x_i))^2) < \infty$.

Proof. This can be shown using general results on the consistency of the bootstrap given in BICKEL/FREEDMAN(1981) or SHAO/TU(1995), Theorem 3.1. To apply the latter theorem, stack the appropriate terms of w_i and x_i in a stochastic vector X_i, so that $E(X_i) = \mu$ contains all the population moments required by the respective index. The index can then be written as $I(G) = g(\mu)$ and the estimated index $I(G_n)$ as $g(\bar{X}_n)$ with $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. The function g has to be continuously differentiable at μ with $\nabla g(\mu) \neq 0$, which is in most cases implied by the existence of the respective index. Furthermore, it has to be verified that $E||X_i||^2 < \infty$. Since X_i is a function of y_i, then resampling from G_n implies resampling from $F_n(x) = n^{-1}\sum_{i=1}^{n} 1\{X_i \leq x\}$. This completes the proof. □

The first assumption of the theorem means for example that observed data for I_{Theil}, I_{MLD} or I_{logvar} must not include zero incomes. The assumptions on the existence of particular moments should never constitute a problem in practice, since all moments exist, if G is the discrete distribution function of a finite target population, e.g. the population of a particular country, which always possesses a probability function with a bounded support. The other assumptions are sufficient to ensure that $\nabla g(\mu) \neq 0$. They are far from being necessary and should not be problematic either, since in reality combinations of weights and incomes are always positive "on average". The condition in case of I_{CV} will hold, if $I_{CV} \neq 0$. The assumptions used in the theorem are the same assumptions that allow application of the δ-method (RAO(1973), p.387) to show asymptotic normality of the point estimates $I(G_n)$, which in turn implies their (weak) consistency.

It can be seen that the application of the bootstrap in the present context is unproblematic, as long as the statistics of interest can be written as smooth functions of population moments. The Gini coefficient is the only commonly used inequality index which does not fit into this framework. It is therefore not considered here, although it should be possible to show the consistency of the bootstrap for this index by using general results on U-statistics (SHI(1986)).

In practice, the distribution $H_{Boot}(x) = P\{\sqrt{n}(I_n^* - I_n) \leq x | y_1, y_2, \ldots, y_n\}$ is not known and has to be simulated. For this purpose, $b = 1, \ldots, B$ empirical resamples $y_1^{*b}, y_2^{*b}, \ldots, y_n^{*b}$ are drawn from G_n. Note that this amounts to randomly drawing B times $i = 1, \ldots n$ integers i_i^{*b} from $\{1, \ldots, n\}$ where each integer $1, \ldots, n$ has equal probability mass n^{-1}. Draw y_i^{*b}, $i = 1, \ldots, n$ is then given the value of $y_{i_i^{*b}}$. The empirical distribution of the collection of bootstrap estimates $I(G_n^{*b})$, $b = 1, \ldots, B$ for $B \to \infty$ then serves as the substitute for the theoretical distribution of $I_n^* = I(G_n^*)$ in $\sqrt{n}(I_n^* - I_n)|y_1, y_2, \ldots, y_n$, which replaces the unknown sampling distribution $\sqrt{n}(I_n - I)$ for statistical inference.

4 Advanced measurement procedures

This section demonstrates how the above framework can be extended to various other measurement procedures. In view of the preceding section, the crucial task is to find a representation of these procedures with the property that the statistic in question is expressed as a function of moments of the population distribution G.

4.1 Difference-in-Means Test for Unbalanced Panel Data

As mentioned in the introduction, it is often interesting to test whether inequality has changed in a statistically significant way from one period to another. Two problems arise if panel data is used to carry out such a test. First, the incomes of a particular observational unit will be correlated across time periods and second, income data may be lacking for some units in one of the periods.

To cope with this situation define the observed data to be of the form

$$y_i = (w_{1i}, x_{1i}, c_{1i}, w_{2i}, x_{2i}, c_{2i})',$$

with w_{ti}, x_{ti} the weight and the income of observation i in year t as above, and

$$c_{ti} = 1 \Leftrightarrow i \text{ provides income data in year } t\text{, and } c_{ti} = 0 \text{ otherwise.} \tag{17}$$

Now define the conditional moments for year t corresponding to (2) to (4)

$$\begin{aligned} \mu_\alpha^t &= E(w_{ti} x_{ti}^\alpha | c_{ti} = 1) = E(c_{ti} w_{ti} x_{ti}^\alpha)/P\{c_{ti} = 1\} = E(c_{ti} w_i x_{ti}^\alpha)/E(c_{ti}), && (18) \\ \tau_{\alpha,\gamma}^t &= E(w_{ti} x_{ti}^\alpha (\log x_{ti})^\gamma | c_{ti} = 1) = E(c_{ti} w_{ti} x_{ti}^\alpha (\log x_{ti})^\gamma)/E(c_{ti}), && (19) \\ \eta^t &= E(w_{ti} \exp(\beta x_{ti}) | c_{ti} = 1) = E(c_{ti} w_{ti} \exp(\beta x_{ti}))/E(c_{ti}), && (20) \\ & && (21) \end{aligned}$$

using the fact that c_{ti} is a dummy variable. The inequality index I^t in question for each year $t \in \{1, 2\}$ can then be written as a function of the moments μ_α^t, $\tau_{\alpha,\gamma}^t$ and η^t as before. The sampling distribution of interest is

$$\sqrt{n}\left[(I_n^2 - I^2) - (I_n^1 - I^1)\right]. \tag{22}$$

To apply SHAO/TU's Theorem, define X_{ti}, $t \in \{1, 2\}$ such that $E(X_{ti})$ contains all the population moments required by the respective inequality index. The population difference can then be expressed as

$$I = I^2 - I^1 = g(E(X_{2i})) - g(E(X_{1i})) = g(\mu^2) - g(\mu^1) \tag{23}$$

where $g(\cdot)$ is defined by (8) to (15). The estimated difference

$$I_n = I_n^2 - I_n^1 = g(\bar{X}_{2n}) - g(\bar{X}_{1n}) \tag{24}$$

can again be written in terms of the sample counterparts $\bar{X}_{2n}$ and $\bar{X}_{1n}$ of the population moments (18) to (20), e.g. $E(c_{ti})$ is estimated by $n^{-1}\sum_{i=1}^{n} c_{ti}$. The same is true for the bootstrap estimate of the difference

$$I_n^* = I_n^{*2} - I_n^{*1} = g(\bar{X}_{2n}^*) - g(\bar{X}_{1n}^*) \tag{25}$$

An application of the theorem shows that $\sqrt{n}\,[(I_n^{2*} - I_n^2) - (I_n^{1*} - I_n^1)]$ can serve as the substitute for the sampling distribution $\sqrt{n}\,[(I_n^2 - I^2) - (I_n^1 - I^1)]$, which is that of $\sqrt{n}\,[(I_n^2 - I_n^1)]$ under $H_0:\ I^2 = I^1$.

Note that the resampling procedure automatically takes into account the covariance structure of $(x_{1i}, x_{2i})'$ as well as the stochastic patterns of non-response or attrition (c_{1i}, c_{2i}) in the intertemporal population. The test can also be applied to compare pre-tax vs. post-tax income as in VAN DE GAER ET AL.(1999) or inequality in two independent populations $t \in \{1, 2\}$, e.g. a cross-country comparison. In the latter case, $(w_{1i}, x_{1i})'$ will be independent of $(w_{2i}, x_{2i})'$.

4.2 Decompositions by Subgroup

This is one of the scenarios appearing in MILLS/ZANDVAKILI(1997). For this section, let the observational data take the form

$$y_i = (w_i, x_i, d_{1i}, d_{2i}, \ldots, d_{Ji})',$$

where w_i, x_i are the weights and the incomes as before and

$$d_{ji} = 1 \Leftrightarrow\ i \text{ belongs to subgroup } j \in \{1, \ldots, J\}, \text{ and } d_{ji} = 0 \text{ otherwise.} \tag{26}$$

Analogous to the last section, define

$$\mu_\alpha^j \quad = \quad E(w_i x_i^\alpha | d_{ji} = 1) = E(d_{ji} w_i x_i^\alpha)/E(d_{ji}) \tag{27}$$

$$\tau_{\alpha,\gamma}^j \quad = \quad E(w_i x_i^\alpha (\log x_i)^\gamma | d_{ji} = 1) = E(d_{ji} w_i x_i^\alpha (\log x_i)^\gamma)/E(d_{ji}). \tag{28}$$

It is assumed that the subgroups are disjoint and that all subgroups taken together constitute the population. The class of additively decomposable inequality measures is

$$I_{GE}^{\alpha} = \sum_{j=1}^{J} I^{W,j,\alpha} + I^{B,\alpha}, \quad \alpha \in \mathbb{R}, \tag{29}$$

where

$$I^{W,j,\alpha} = p_j r_j^{\alpha} I_{GE}^{\alpha,j} \tag{30}$$

is the within-group contribution to overall inequality of subgroup j,

$$p_j = \mu_0^{-1} E(d_{ji} w_i) \tag{31}$$

is the population share of subgroup j,

$$r_j = \mu_1^j \mu_0 \mu_1^{-1} (\mu_0^j)^{-1} \tag{32}$$

is the relative income of subgroup j, and

$$\begin{aligned}
I^{B,\alpha} &= (\alpha^2 - \alpha)^{-1} \left[\sum_{j=1}^{J} (r_j^{\alpha} - 1) p_j \right], \text{ for } \alpha \in \mathbb{R} \setminus \{0, 1\}, && (33)\\
I^{B,1} &= \sum_{j=1}^{J} r_j (\log r_j) p_j, \text{ for } \alpha \to 1, && (34)\\
I^{B,0} &= -\sum_{j=1}^{J} (\log r_j) p_j, \text{ for } \alpha \to 0 && (35)
\end{aligned}$$

are the between-group component of overall inequality (see SHORROCKS(1980) or COWELL(2000)). Inequality within subgroup j, i.e. $I_{GE}^{\alpha,j}$ is computed according to (8) to (10) with the population moments μ_α, $\tau_{\alpha,\gamma}$ replaced by those of the respective subpopulation μ_α^j, $\tau_{\alpha,\gamma}^j$. (Note that the $E(d_{ji})$'s eventually cancel out in the calculations.)

The statistics of interest are the contribution of the between-group component to overall inequality $I_n^{\alpha} = I_{GE}^{\alpha}(G_n)$

$$s^{B,\alpha} = I_n^{B,\alpha} / I_n^{\alpha}, \tag{36}$$

the corresponding share of the within-group component of subgroup j

$$s^{j,\alpha} = I^{W,j,\alpha} / I_n^{\alpha}, \tag{37}$$

and the contribution of within-group inequality

$$s^{W,\alpha} = \sum_{j=1}^{J} s^{j,\alpha}. \tag{38}$$

Two differences to the related result in COWELL(1989), p.34) are worth noting. First, the bootstrap makes it easy to consider statistics like $s^{B,\alpha}$ instead of $I_n^{B,\alpha}$, which lead to very complicated covariance calculations, if the δ-method is used. Second, in contrast to the cited result, where group sizes are fixed, through the stochastic d_{ji}'s, the procedure proposed here also takes into account sampling variability coming from the fact that the subgroup sizes vary when sampling from a target population.

4.3 Decomposition of the Change in Inequality

MOOKHERJEE/SHORROCKS(1982) derive the following decomposition of the *change* in inequality measured by $I = I_{MLD}$

$$\begin{aligned} I^2 - I^1 = \Delta \quad \approx \quad & \sum_{j=1}^{J} \overline{p_j} \Delta I^j \\ & + \sum_{j=1}^{J} \overline{I^j} \Delta p_j \\ & + \sum_{j=1}^{J} \left(\overline{r_j} - \overline{\log r_j} \right) \Delta p_j \\ & + \sum_{j=1}^{J} \left(\overline{p_j r_j} - \overline{p_j} \right) \Delta \left(\log \mu_{1,1}^{j} (\mu_{1,0}^{j})^{-1} \right) \\ & = \Delta_W + \Delta_{SW} + \Delta_{SM} + \Delta_M \end{aligned} \tag{39}$$

where the bars indicate averages over the two periods and Δ the difference operator. In this decomposition, Δ_W stands for the contribution of changing levels of within-group inequality to the change in overall inequality, Δ_{SW} and Δ_{SM} are the corresponding contributions of changes in population shares, while Δ_M captures the contribution of changes in mean incomes. The four sums can be further split up, in order to obtain the contributions of each subgroup to the respective component.

Although this procedure has been extensively applied in the empirical literature (see e.g. JENKINS(1995), JÄNTTI(1997) and GRABKA ET AL.(1999)), to the knowledge of the author, no one ever has calculated standard errors. The reason is that it has been unclear how the numerous dependencies in the data could be modeled, especially in the case of panel data, i.e. with longitudinal correlation and ongoing attrition. But even then, the application of the δ-method is probably impractical as the number of moments and the size of their covariance matrix become too large very quickly.

To apply the above developed framework to this scenario, let the observational data take the form

$$y_{ti} = (w_{ti}, x_{ti}, c_{ti}, d_{t1i}, d_{t2i}, \ldots, d_{tJi})'$$

and

$$y_i = (y_{1i}', y_{2i}')'.$$

To incorporate the time dimension, define

$$\mu_\alpha^{j,t} = E(w_{ti}x_{ti}^\alpha | (c_{ti}, d_{tji}) = (1,1)) = E(c_{ti}d_{tji}w_{ti}x_{ti}^\alpha)/E(c_{ti}d_{tji}), \tag{40}$$

$$\tau_{\alpha,\gamma}^{j,t} = E(w_{ti}x_{ti}^\alpha (\log x_{ti})^\gamma | (c_{ti}, d_{tji}) = (1,1)) = E(c_{ti}d_{tji}w_{ti}x_{ti}^\alpha (\log x_{ti})^\gamma)/E(c_{ti}d_{tji}). \tag{41}$$

Again, all parts of (39) to (40) can be written with the help of population moments. For $I^{j,t}$ and

$$r_j^t = \mu_1^{j,t}\mu_0^t(\mu_1^t\mu_0^{j,t})^{-1}, \tag{42}$$

the corresponding conditional moments are taken. Moreover,

$$p_j^t = (E(c_{ti}w_i))^{-1} E(c_{ti}d_{tji}w_i). \tag{43}$$

The statistics of interest are

$$\frac{\Delta_W}{\Delta}, \frac{\Delta_{SW}}{\Delta}, \frac{\Delta_{SM}}{\Delta} \text{ and } \frac{\Delta_M}{\Delta}.$$

This procedure simultaneously takes into account sampling variability due to variability in income x_{ti}, attrition or non-response c_{ti}, membership in subgroup j as well as to transitions between subgroups d_{tji}.

4.4 Decomposition by Income Source

In the case of inequality decomposition by income source the observed sample takes the form

$$y_i = (w_i, \tilde{x}_{1i}, \tilde{x}_{2i}, \ldots, \tilde{x}_{Si})'$$

where $\tilde{x}_{si}$ denotes income from source $s \in \{1, \ldots, S\}$ and household income is

$$x_i = \sum_{s=1}^{S} \tilde{x}_{si}. \tag{44}$$

The statistics of interest are

$$s_s = \text{ the contribution of income source } s \text{ to overall inequality.}$$

Every inequality index suggests its own formula for s_s. For example, the squared coefficient of variation, the THEIL index and the mean logarithmic deviation yield

$$s_s^{CV} = \left[\mu_0^{-1} E(w_i \tilde{x}_{si} x_i) - (\mu_0^{-1} E(w_i \tilde{x}_{si}))(\mu_0^{-1}\mu_1)\right]\left[\mu_0^{-1}\mu_2 - \mu_0^{-2}\mu_1^2\right]^{-1}, \quad (45)$$

$$s_s^{Theil} = \left[\mu_1^{-1} E(w_i \tilde{x}_{si}(\log x_i)) - \mu_1^{-1} E(w_i \tilde{x}_{si})(\log \mu_0^{-1}\mu_1)\right] / I_{Theil}, \quad (46)$$

$$s_s^{MLD} = \left[\mu_0^{-1} E(w_i \tilde{x}_{si} x_i^{-1})(\log \mu_0^{-1}\mu_1) - \mu_0^{-1} E(w_i \tilde{x}_{si} x_i^{-1}(\log x_i))\right] / I_{MLD}. \quad (47)$$

(See SHORROCKS(1983). The formulae there have been modified to incorporate the sampling weights). Resampling from $y_1, y_2, \ldots, y_n$ automatically captures the correlation between different income sources $\tilde{x}_{si}$.

4.5 Mobility Measurement

Many mobility indices have been proposed in the literature (see FIELDS/OK(1999) for an overview) but two main forms are used in empirical work. The first index was developed by SHORROCKS(1978) and MAASOUMI/ZANDVAKILI(1986). The sampling scenario for this case is

$$y_i = (w_i, \tilde{x}_{1i}, \tilde{x}_{2i}, \ldots, \tilde{x}_{Mi})',$$

where $\tilde{x}_{mi}$ denotes income received in period $m \in \{1, \ldots, M\}$ and

$$x_i = S_i(\tilde{x}_{1i}, \tilde{x}_{2i}, \ldots, \tilde{x}_{Mi}) \quad (48)$$

is aggregated income over M periods with aggregation function

$$S_i(\tilde{x}_{1i}, \tilde{x}_{2i}, \ldots, \tilde{x}_{Mi}) = \begin{cases} \left[\sum_{m=1}^{M} \alpha_m \tilde{x}_{mi}^{-\beta}\right]^{-\frac{1}{\beta}}, & \beta \in \mathbb{R} \setminus \{0, 1\}, \\ \prod_{m=1}^{M} \tilde{x}_{mi}^{\alpha_m}, & \beta = 0, \\ \sum_{m=1}^{M} \alpha_m \tilde{x}_{mi}, & \beta = 1, \end{cases} \quad (49)$$

where α_m are period weights with $\sum_{m=1}^{M} \alpha_m = 1$. The MAASOUMI/ZANDVAKILI mobility index is then defined as

$$M_{MZ} = 1 - I/(\sum_{m=1}^{M} \alpha_m I_m), \quad (50)$$

where I is a member of the generalized entropy measures I_{GE} representing inequality in aggregated income x_i and I_m the inequality index calculated from the moments of income received in period m. The weights w_i apply to all periods. Similarly, the SHORROCKS mobility index

$$M_S = 1 - I/(\sum_{m=1}^{M} E(w_i \tilde{x}_{mi})\mu_1^{-1} I_m), \quad (51)$$

where I is a strictly convex inequality index such as I_{GE}, I_{Theil}, I_{MLD}, I_A or I_{CV}, uses the aggregation function $x_i = \sum_{m=1}^{M} \tilde{x}_{mi}$ and weights period inequalities I_m by period mean income relative to aggregated mean income.

A second type of mobility index goes back to PRAIS(1955). This index is based on the main diagonal of a transition matrix $P = [p_{kl}]$, describing the transition probabilities between $m = 1, \ldots, M$ states. The index is defined as

$$M_P = (M-1)^{-1} \left(M - \sum_{m=1}^{M} p_{mm} \right), \tag{52}$$

where p_{kl} denotes the probability of the transition $k \to l$. This index is equal to the so-called eigenvalue index $M_E = (M-1)^{-1}(M - \sum_j |\lambda_j|)$, if the eigenvalues λ_j of P are real and non-negative.

Let the observed sample be

$$y_i = (w_i, d_{11i}, d_{12i}, \ldots, d_{1Mi}, d_{21i}, d_{22i}, \ldots, d_{2Mi})'$$

with

$$d_{tmi} = 1 \Leftrightarrow \; i \text{ is in state } m \text{ at time } t, \text{ and } d_{tmi} = 0 \text{ otherwise.} \tag{53}$$

Using this notation, the presence of the transition $k \to l$ can be expressed as $d_{1ki}d_{2li}$. The weighted transition probability is then

$$p_{kl} = E(w_i d_{1ki} d_{2li})/E(w_i d_{1ki}), \tag{54}$$

i.e. a function of moments of the population distribution function G. Substituting population moments by sample moments yields the mobility estimate

$$M_{P,n} = (M-1)^{-1} \left(M - \sum_{m=1}^{M} \left[(\sum_{i=1}^{n} w_i d_{1mi} d_{2mi})/(\sum_{i=1}^{n} w_i d_{1mi}) \right] \right). \tag{55}$$

Asymptotic normality and consistency of this estimate directly follows from the δ-method.

The limiting distribution of the PRAIS-index has been explicitly derived for unweighted data by SCHLUTER(1998). Similar formulae for the SHORROCKS/MAASOUMI/ZANDVAKILI index do not exist, although variances have been estimated in that case in MAASOUMI/TREDE(1999) who also provide the partial derivatives necessary for the application of the δ-method. In the case, where explicit formulae for the limiting distribution do not exist, the implementation of the above described bootstrap procedure is much simpler than that of the δ-method, as the program code used to obtain the point estimates can be re-used to calculate estimates of the sampling distribution.

An important extension of the results in this section is to introduce dummy variables for time and sample inclusion as in section 4.1 in order to construct a test of mobility *differences*. This is a highly relevant scenario, since mobility estimates of the described form always require panel data, which affects inference about intertemporal mobility changes via longitudinal correlation and attrition. An extension to a decomposition of mobility into within- and between group components as in MAASOUMI/ZANDVAKILI(1990) is also possible.

4.6 Poverty Measurement

As a last application, consider the measurement of poverty using the FOSTER ET AL.(1984) class of decomposable poverty indices. A weighted version of these indices can be defined as

$$P_{FGT} = \mu_0^{-1} E\left(w_i \left(\frac{z - x_i}{z}\right)^a 1\{x < t\}\right), \quad a > 1, \tag{56}$$

where w_i denote weights and x_i incomes as in section 2, z the poverty line and a a sensitivity parameter. To see that the bootstrap works in this case, reformulate the original data $(w_i, x_i)'$ as (w_i, dep_i) with individual deprivation dep_i defined as

$$dep_i = \left(\frac{z - x_i}{z}\right)^a 1\{x < t\},$$

so that P_{FGT} can be estimated as smooth function of sample moments

$$P_{FGT,n} = m_0^{-1}\left(n^{-1}\sum_{n=1}^{n} w_i dep_i\right). \tag{57}$$

The asymptotic distribution of $P_{FGT,n}$ has been studied by KAKWANI(1993) and BISHOP ET AL.(1995). An extension of the bootstrap to the poverty decompositions described in FOSTER ET AL.(1984) or intertemporal poverty comparisons using the ideas of section 4.2 is straightforward. The approach can also be extended to other additive poverty indices (see SEIDL(1988) for a survey of poverty measurement).

5 Monte Carlo Evidence

The aim of this section is to compare the performance of the bootstrap in the above scenarios with that of the δ-method, i.e. the normal approximation. The focus will be on the coverage probability of confidence intervals, since that is what practioners are interested in. Besides the simple bootstrap procedure discussed above, the comparison will also include the studentized bootstrap or bootstrap-t. The idea of the studentized bootstrap is to apply the bootstrap to the studentized version of the statistic $I_n = g(\bar{X}_n)$, i.e. $(I_n - I)/\hat{\sigma}_n$, where $\mu = E(X_i)$ is the vector that collects the required population moments and $I = g(\mu)$ the true parameter. As in the case of the δ-method, this requires an estimate of the standard deviation of I_n, e.g.

$$\hat{\sigma}_n = \left[\frac{1}{n}\nabla g(\bar{X}_n)'\hat{\Sigma}\nabla g(\bar{X}_n)\right]^{\frac{1}{2}} \tag{58}$$

where $\hat{\Sigma}$ is a consistent estimate of $\Sigma = \mathrm{var}(X_i)$.

The confidence intervals $\mathcal{C}_n$ considered here are the one-sided intervals (OS)

$$\begin{aligned}
\mathcal{C}_n^{\text{NOR,OS}} &= \left[-\infty; I_n - \hat{\sigma}_n \Phi^{-1}(\alpha)\right], && (59)\\
\mathcal{C}_n^{\text{HB,OS}} &= \left[-\infty; I_n - n^{-\frac{1}{2}} H_{Boot}^{-1}(\alpha)\right], && (60)\\
\mathcal{C}_n^{\text{BT,OS}} &= \left[-\infty; I_n - \hat{\sigma}_n G_{Boot}^{-1}(\alpha)\right], && (61)
\end{aligned}$$

for the normal approximation (NOR), the simple bootstrap (HB, also called hybrid bootstrap interval) and the bootstrap-t (BT), with

$$\begin{aligned}
H_{Boot}(x) &= P\{\sqrt{n}(I_n^* - I_n) \le x | X_1, X_2, \ldots, X_n\}, && (62)\\
G_{Boot}(x) &= P\{(I_n^* - I_n)/\hat{\sigma}_n^* \le x | X_1, X_2, \ldots, X_n\} && (63)
\end{aligned}$$

where $I_n^* = g(\bar{X}_n^*)$,

$$\hat{\sigma}_n^* = \left[\frac{1}{n}\nabla g(\bar{X}_n^*)'\hat{\Sigma}^*\nabla g(\bar{X}_n^*)\right]^{\frac{1}{2}} \tag{64}$$

and the stars indicate statistics computed from the i.i.d. resample X_i^*. The confidence level of these intervals is $1 - \alpha$ and $\Phi(\cdot)$ denotes the standard normal distribution function.

Similarly, consider the following two-sided intervals (TS)

$$\begin{aligned}
\mathcal{C}_n^{\text{NOR,TS}} &= \left[I_n - \hat{\sigma}_n \Phi^{-1}(1 - \frac{\alpha}{2}); I_n - \hat{\sigma}_n \Phi^{-1}(\frac{\alpha}{2})\right], && (65)\\
\mathcal{C}_n^{\text{HB,TS}} &= \left[I_n - n^{-\frac{1}{2}} H_{Boot}^{-1}(1 - \frac{\alpha}{2}); I_n - n^{-\frac{1}{2}} H_{Boot}^{-1}(\frac{\alpha}{2})\right], && (66)\\
\mathcal{C}_n^{\text{BT,TS}} &= \left[I_n - \hat{\sigma}_n G_{Boot}^{-1}(1 - \frac{\alpha}{2}) I_n - \hat{\sigma}_n G_{Boot}^{-1}(\frac{\alpha}{2})\right]. && (67)
\end{aligned}$$

Asymptotically, the intervals $\mathcal{C}_n^{\text{BT,OS}}, \mathcal{C}_n^{\text{NOR,TS}}, \mathcal{C}_n^{\text{HB,TS}}, \mathcal{C}_n^{\text{BT,TS}}$ are more accurate than the intervals $\mathcal{C}_n^{\text{NOR,OS}}$ and $\mathcal{C}_n^{\text{HB,OS}}$, because the former have a convergence rate of $P\{I \in \mathcal{C}_n\} - (1 - \alpha) = O(n^{-1})$, whereas the latter achieve only the convergence rate $P\{I \in \mathcal{C}_n\} - (1 - \alpha) = O(n^{-\frac{1}{2}})$

(see HALL(1986), HALL(1988), SHAO/TU(1995), p.150). All of these confidence intervals can be used to carry out a test of the hypothesis $H_0: \; I = I_0$ with type I error

$$P\{\text{reject } H_0|H_0\} = P\{I_0 \notin \mathcal{C}_n|H_0\} = P\{I \notin \mathcal{C}_n|I = I_0\} \overset{n\to\infty}{\longrightarrow} \alpha.$$

The differences in the coverage probabilities for one-sided intervals and the fact that the *distribution function* of the studentized bootstrap statistic converges to the true sampling distribution at a higher rate than the normal approximation, is the reason why the studentized bootstrap (not the simple bootstrap) is expected to have advantages in small samples (see BABU/SINGH(1983), SHAO/TU(1995), p. 94, HOROWITZ(2000)).

The parameters for the following experiments were chosen so that the simulated distributions closely mimic corresponding data from the West German subsample of the German Socio-Economic Panel (GSOEP) for the year 1996 or the years 1996 and 1997, if the measurement scenario involves two periods. The simulations use the random generator of PARK and MILLER with BAYS-DURHAM shuffle for uniformly distributed variates and the BOX-MULLER method for standard normal variates (the algorithms were taken from PRESS ET AL.(1992), pp. 280 and 289). Simulations were carried out for both $1 - \alpha = 0.95$ and $1 - \alpha = 0.9$, but only the numbers for $1 - \alpha = 0.95$ will be reported here[3]. The performance of 90% intervals was very similar to that of the 95% intervals, but the former generally achieved a slightly lower accuracy.

5.1 Experiment 1: Inequality Indices

Inspection of the data taken from the GSOEP shows that the joint distribution of the sampling weights multiplied by household size $w_i' w_i'' = w_i$ and net household income equivalized by a OECD equivalence scale x_i is approximately log-normal $\Lambda(\lambda, \Omega)$ with

$$\lambda = E\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 9.1 \\ 7.7 \end{pmatrix}, \quad \Omega = \text{var}\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 0.9025 & 0.01425 \\ 0.01425 & 0.277954 \end{pmatrix}.$$

Using the formula for moments of the log-normal distribution (Johnson/Kotz(1972), p. 20)

$$\mu_\alpha = \exp\left(\lambda_w + \alpha\lambda_x + \frac{1}{2}(\Omega_{ww} + 2\alpha\Omega_{wx} + \alpha^2\Omega_{xx})\right), \tag{68}$$

one can derive inequality in the population as

$$I_{GE}^{\alpha} \;=\; (\alpha^2 - \alpha)^{-1}\left[\exp\left(\frac{1}{2}\left[\alpha^2\Omega_{xx} - \alpha\Omega_{xx}\right]\right) - 1\right], \text{ for } \alpha \in \mathbb{R} \setminus \{0, 1\}, \tag{69}$$

[3]These results as well as the source code of the Monte Carlo simulations (the programs were written in C) are available from the author on request.

$$I_{MLD} = I_{Theil} = \frac{1}{2}\Omega_{xx}, \tag{70}$$

$$I_A = 1 - \exp\left(-\frac{\varepsilon}{2}\Omega_{xx}\right), \text{ for } \varepsilon \geq 0 \tag{71}$$

(compare COWELL/JENKINS(1998) and VAN KERM(2000). Note that the distribution of weights does not matter for inequality.) This results in the true values shown in table 1, which are in accordance with empirical estimates for Germany (see e.g. BIEWEN(2000)).

Table 1. True values for experiment 1: inequality indices.

index	true value
$\mu_1\mu_0^{-1}$	2574.02
I_{GE}^{-1}	0.1602127281
I_{MLD}	0.1389770000
I_{Theil}	0.1389770000
I_{GE}^{2}	0.1602127281
$I_A^{0.5}$	0.0671291387
I_A^{1}	0.1297519561
$I_A^{1.5}$	0.1881709577
I_A^{2}	0.2426683420

Note: $\mu_1\mu_0^{-1}$ is the (weighted) mean income

Table 2 shows the Monte Carlo results for experiment 1. The numbers suggest that in terms of coverage accuracy, the bootstrap-t is better than the normal approximation, which in turn is slightly better than the simple bootstrap. However, for samples of $n = 5000$ observations, all three methods are already very close and generally differ by only a fraction of a percentage point. In fact, for large sample sizes there does not seem to be any practical difference between the three methods, except for the index I_{GE}^2, for which the difference amounts to up to two percentage points.

Table 2. Experiment 1: inequality indices,
coverage probabilities of 95% confidence intervals,
1000 bootstrap replications, 5000 Monte Carlo replications.

index	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
normal approximation intervals $\mathcal{C}_n^{NOR,TS}$ *(two-sided)*					$\mathcal{C}_n^{NOR,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9212	0.9392	0.9442	0.9506	0.8978	0.9238	0.9300	0.9454
I_{GE}^{-1}	0.8718	0.9256	0.9284	0.9420	0.8340	0.8988	0.9044	0.9312
I_{MLD}	0.8678	0.9264	0.9284	0.9454	0.8284	0.8926	0.9030	0.9302
I_{Theil}	0.8346	0.9044	0.9156	0.9404	0.7976	0.8708	0.8798	0.9174
I_{GE}^{2}	0.7816	0.8652	0.8820	0.9180	0.7386	0.8128	0.8372	0.8908
$I_A^{0.5}$	0.8566	0.9154	0.9228	0.9442	0.8180	0.8852	0.8964	0.9270
I_A^{1}	0.8696	0.9272	0.9296	0.9452	0.8332	0.8954	0.9042	0.9306
$I_A^{1.5}$	0.8760	0.9294	0.9292	0.9456	0.8400	0.9040	0.9090	0.9332
I_A^{2}	0.8762	0.9298	0.9320	0.9418	0.8484	0.9052	0.9102	0.9326
hybrid bootstrap intervals $\mathcal{C}_n^{HB,TS}$ *(two-sided)*					$\mathcal{C}_n^{HB,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9072	0.9340	0.9426	0.94820	0.8828	0.9194	0.9264	0.9428
I_{GE}^{-1}	0.8550	0.9166	0.9238	0.94040	0.8256	0.8934	0.8998	0.9282
I_{MLD}	0.8476	0.9198	0.9252	0.94360	0.8276	0.8890	0.9012	0.9298
I_{Theil}	0.8140	0.8960	0.9110	0.93800	0.7976	0.8688	0.8796	0.9144
I_{GE}^{2}	0.7532	0.8498	0.8742	0.91420	0.7386	0.8100	0.8358	0.8858
$I_A^{0.5}$	0.8344	0.9100	0.9206	0.94260	0.8212	0.8840	0.8952	0.9254
I_A^{1}	0.8518	0.9188	0.9272	0.94460	0.8342	0.8940	0.9042	0.9310
$I_A^{1.5}$	0.8604	0.9232	0.9276	0.94320	0.8468	0.9028	0.9098	0.9340
I_A^{2}	0.8620	0.9236	0.9290	0.94160	0.8574	0.9076	0.9116	0.9328
bootstrap-t intervals $\mathcal{C}_n^{BT,TS}$ *(two-sided)*					$\mathcal{C}_n^{BT,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9324	0.9392	0.9420	0.9450	0.9358	0.9380	0.9398	0.9424
I_{GE}^{-1}	0.9238	0.9398	0.9414	0.9410	0.9334	0.9378	0.9430	0.9448
I_{MLD}	0.9226	0.9380	0.9368	0.9416	0.9352	0.9370	0.9444	0.9454
I_{Theil}	0.9122	0.9290	0.9310	0.9408	0.9376	0.9408	0.9414	0.9424
I_{GE}^{2}	0.8746	0.9086	0.9170	0.9328	0.9324	0.9424	0.9434	0.9380
$I_A^{0.5}$	0.9186	0.9354	0.9358	0.9416	0.9364	0.9422	0.9432	0.9440
I_A^{1}	0.9232	0.9394	0.9368	0.9416	0.9360	0.9376	0.9446	0.9454
$I_A^{1.5}$	0.9270	0.9404	0.9424	0.9406	0.9346	0.9388	0.9460	0.9450
I_A^{2}	0.9274	0.9406	0.9428	0.9416	0.9326	0.9384	0.9456	0.9468

Note: $\mu_1\mu_0^{-1}$ is the (weighted) mean income

In absolute terms, all three methods perform very well for sample size $n = 5000$ with, in most cases, none or only two to three percentage points discrepancy between actual and nominal coverage probabilities. Consistent with the asymptotic results, two-sided intervals are more accurate their one-sided counterparts. Also note that there seems to be a relationship between the sensitivity parameters α and ε and the coverage accuracy. Precision increases with increasing ε and decreasing α, respectively.

Perhaps the most important conclusion for practice is that the use of the studentized bootstrap only pays off for relatively small sample sizes of up to 500 observations, which are rather rare in this context. This also means that further refinements of the bootstrap, as considered for example by TREDE(2000), do not seem to be worth the additional effort. Finally, all intervals are too narrow implying that a null hypothesis is rejected too often if a hypothesis test is carried out.

5.2 Experiment 2: inequality differences

The data $(w_{1i}, x_{1i}, w_{2i}, x_{2i})'$ for this experiment were drawn from a four-dimensional log-normal distribution $\Lambda(\lambda, \Omega)$ with

$$\lambda = E\begin{pmatrix} \log w_{1i} \\ \log x_{1i} \\ \log w_{2i} \\ \log x_{2i} \end{pmatrix} = \begin{pmatrix} 9 \\ 7.75 \\ 9.02 \\ 7.77 \end{pmatrix}$$

and

$$\Omega = \mathrm{var}\begin{pmatrix} \log w_{1i} \\ \log x_{1i} \\ \log w_{2i} \\ \log x_{2i} \end{pmatrix} = \begin{pmatrix} 0.9801 & 0.011682 & 0.99 & 0.0099 \\ 0.011682 & 0.27053924 & 0.01179948 & 0.189918 \\ 0.99 & 0.01179948 & 1.0625 & 0.01000039 \\ 0.0099 & 0.189918 & 0.01000039 & 0.22635 \end{pmatrix}.$$

This implies the population values for the true intertemporal inequality and mean income differences shown in table 3.

Table 4 gives coverage probabilities for one-sided and two-sided intervals. One-sided intervals may be of interest here, since one might want to test hypotheses like whether inequality has significantly risen. In the setup simulated here, all intervals perform very well, even for the smallest sample size of $n = 100$. For the more interesting case $n = 5000$, differences between actual and nominal coverage probabilities generally do not exceed two or three percentage points. The relationship between the sensitivity parameters and coverage accuracy found in experiment 1 also holds in this scenario.

Table 3. True values for experiment 2: inequality differences.

index	true difference
$\mu_1 \mu_0^{-1}$	-10.13540744
I_{GE}^{-1}	-0.02832826643
I_{MLD}	-0.02209462000
I_{Theil}	-0.02209462000
I_{GE}^{2}	-0.02832826643
$I_A^{0.5}$	-0.01038207633
I_A^{1}	-0.01951400027
$I_A^{1.5}$	-0.02750895393
I_A^{2}	-0.03447098865

Note: $\mu_1 \mu_0^{-1}$ is the (weighted) mean income

The experiment shows that asymptotic results can be misleading in small samples, as most of the one-sided intervals are more precise than their two-sided counterparts. Moreover, it shows that the bootstrap-t does not necessarily lead to a more precise coverage probability, although again, all three methods are practically equivalent for large sample sizes. Differences generally lie within the range of one percentage point.

As an illustration, table 5 shows how statistical inference on intertemporal inequality differences can go wrong, if longitudinal correlation is not accounted for. The intervals were calculated on the false assumption that the moments of the two periods are uncorrelated. The confidence intervals are far too wide, as the positive correlation of the two inequality estimates reduces the variability of their difference.

Table 4. Experiment 2: inequality differences,
coverage probabilities of 95% confidence intervals
1000 bootstrap replications, 5000 Monte Carlo replications.

index	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
normal approximation intervals $\mathcal{C}_n^{NOR,TS}$ *(two-sided)*					$\mathcal{C}_n^{NOR,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9354	0.9470	0.9488	0.9476	0.9456	0.9506	0.9498	0.9486
I_{GE}^{-1}	0.9280	0.9514	0.9478	0.9514	0.9636	0.9602	0.9612	0.9574
I_{MLD}	0.9188	0.9476	0.9432	0.9484	0.9510	0.9530	0.9548	0.9554
I_{Theil}	0.9062	0.9416	0.9400	0.9486	0.9482	0.9562	0.9554	0.9558
I_{GE}^{2}	0.8832	0.9294	0.9358	0.9452	0.9468	0.9600	0.9644	0.9628
$I_{A}^{0.5}$	0.9140	0.9454	0.9430	0.9486	0.9462	0.9548	0.9544	0.9536
I_{A}^{1}	0.9178	0.9462	0.9438	0.9502	0.9436	0.9510	0.9530	0.9538
$I_{A}^{1.5}$	0.9186	0.9452	0.9462	0.9518	0.9426	0.9490	0.9550	0.9526
I_{A}^{2}	0.9182	0.9468	0.9476	0.9496	0.9380	0.9486	0.9548	0.9528
hybrid bootstrap intervals $\mathcal{C}_n^{HB,TS}$ *(two-sided)*					$\mathcal{C}_n^{HB,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9338	0.9484	0.9492	0.9470	0.9432	0.9528	0.9496	0.9496
I_{GE}^{-1}	0.9154	0.9500	0.9484	0.9520	0.9560	0.9578	0.9640	0.9586
I_{MLD}	0.9006	0.9412	0.9436	0.9516	0.9356	0.9496	0.9570	0.9560
I_{Theil}	0.8722	0.9340	0.9374	0.9500	0.9288	0.9506	0.9552	0.9586
I_{GE}^{2}	0.8450	0.9180	0.9322	0.9462	0.9262	0.9552	0.9610	0.9638
$I_{A}^{0.5}$	0.8832	0.9392	0.9396	0.9526	0.9280	0.9498	0.9540	0.9556
I_{A}^{1}	0.8890	0.9390	0.9418	0.9526	0.9258	0.9462	0.9528	0.9548
$I_{A}^{1.5}$	0.8878	0.9382	0.9424	0.9504	0.9228	0.9456	0.9540	0.9532
I_{A}^{2}	0.8868	0.9380	0.9432	0.9506	0.9180	0.9422	0.9534	0.9544
bootstrap-t intervals $\mathcal{C}_n^{BT,TS}$ *(two-sided)*					$\mathcal{C}_n^{BT,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9302	0.9358	0.9334	0.9400	0.9358	0.9380	0.9398	0.9424
I_{GE}^{-1}	0.9042	0.9328	0.9308	0.9416	0.9334	0.9378	0.9430	0.9448
I_{MLD}	0.9070	0.9324	0.9306	0.9378	0.9352	0.9370	0.9444	0.9454
I_{Theil}	0.9004	0.9278	0.9198	0.9296	0.9376	0.9408	0.9414	0.9424
I_{GE}^{2}	0.8710	0.9134	0.9112	0.9210	0.9324	0.9424	0.9434	0.9380
$I_{A}^{0.5}$	0.9066	0.9310	0.9264	0.9350	0.9364	0.9422	0.9432	0.9440
I_{A}^{1}	0.9122	0.9342	0.9314	0.9390	0.9360	0.9376	0.9446	0.9454
$I_{A}^{1.5}$	0.9138	0.9372	0.9344	0.9406	0.9346	0.9388	0.9460	0.9450
I_{A}^{2}	0.9184	0.9378	0.9356	0.9416	0.9326	0.9384	0.9456	0.9468

Note: $\mu_1\mu_0^{-1}$ is the (weighted) mean income

Table 5. Experiment 2: inequality differences, coverage probabilities of 95% confidence intervals not taking into account longitudinal correlation, 5000 Monte Carlo replications.

index	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
normal approximation $\mathcal{C}_n^{NOR,TS}$ *(two-sided)*					$\mathcal{C}_n^{NOR,OS}$ *(one-sided)*			
$\mu_1\mu_0^{-1}$	0.9992	1.0000	1.0000	0.9996	0.9980	0.9992	0.9992	0.9990
I_{GE}^{-1}	0.9858	0.9942	0.9948	0.9976	0.9920	0.9932	0.9928	0.9958
I_{MLD}	0.9856	0.9924	0.9950	0.9966	0.9870	0.9938	0.9912	0.9948
I_{Theil}	0.9808	0.9910	0.9934	0.9936	0.9852	0.9926	0.9916	0.9938
I_{GE}^{2}	0.9740	0.9846	0.9880	0.9900	0.9842	0.9926	0.9934	0.9926
$I_A^{0.5}$	0.9834	0.9922	0.9952	0.9956	0.9852	0.9936	0.9914	0.9936
I_A^{1}	0.9850	0.9924	0.9950	0.9968	0.9866	0.9932	0.9910	0.9946
$I_A^{1.5}$	0.9854	0.9924	0.9950	0.9976	0.9868	0.9926	0.9910	0.9952
I_A^{2}	0.9854	0.9944	0.9946	0.9974	0.9858	0.9918	0.9918	0.9958

Note: $\mu_1\mu_0^{-1}$ is the (weighted) mean income

5.3 Experiment 3: subgroup decomposition

The data in this section mimics the joint distribution of the sampling weights multiplied by household size and equivalent income and nationality in the 1996 wave of the GSOEP. The data generating process is a mixture of two log-normal distributions with the following parameters. The distribution of the foreigner subsample is well represented by

$$\lambda_1 = E\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 8.05 \\ 7.5 \end{pmatrix}, \quad \Omega_1 = \mathrm{var}\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 0.1300825 & -0.0539 \\ -0.0539 & 0.161225 \end{pmatrix},$$

and that of the German subsample by

$$\lambda_2 = E\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 9.04 \\ 7.8 \end{pmatrix}, \quad \Omega_2 = \mathrm{var}\begin{pmatrix} \log w_i \\ \log x_i \end{pmatrix} = \begin{pmatrix} 0.593125 & -0.01905 \\ -0.01905 & 0.250225 \end{pmatrix}.$$

The two distributions were mixed with probabilities $P\{\text{foreigner}\} = P\{(d_{1i} = 1\} = 0.2001443$ (in the notation of section 4.2) and $P\{\text{German}\} = 1 - P\{\text{foreigner}\}$, implying population shares $p_1 = 0.0845910272$, $p_2 = 1 - p_1$ (foreigners were oversampled in the GSOEP) and relative incomes $r_1 = 0.703081774$, $r_2 = 1.027437592$. This results in the true parameters shown in table 6.

Table 6. True values for experiment 3: subgroup decomposition.

statistic	true value
$I^{W,1,0}$	0.0068190941
$I^{W,2,0}$	0.1145291051
$I^{B,0}$	0.0050216783
$I^{W,1,1}$	0.0047943808
$I^{W,2,1}$	0.1176715550
$I^{B,1}$	0.0045063105

$I^{W,j,\alpha}$= within-group component of group j

$I^{B,\alpha}$= between-group inequality

For exact definitions, see section 4.2.

Table 7 shows the Monte Carlo results for experiment 3. The general ranking places the bootstrap-t first, the normal approximation second and the ordinary bootstrap third. The differences between the normal approximation and the ordinary bootstrap are generally very small, except for the between-group components $I^{B,0}$ and $I^{B,1}$. The overall performance of all three intervals for large sample sizes is reasonably good, but not quite as good as in experiments 1 and 2. For $n = 5000$, actual and nominal coverage probabilities can still differ by up to five percentage points for two-sided and up to ten percentage points for one-sided intervals. In general, the two-sided intervals are more accurate than the one-sided ones. Again, precision seems to decrease with increasing parameter α.

Note the poor performance of the normal approximation and the simple bootstrap for very small sample sizes. However, the use of the studentized bootstrap leads only in samples of 500 to 1000 observations to substantial improvements in coverage accuracy. For $n = 5000$, the improvement amounts to not more than one to two percentage points for two-sided intervals and between two and four percentage points for one-sided intervals.

Table 7. Experiment 3: subgroup decomposition,
coverage probabilities of 95% confidence intervals,
1000 bootstrap replications, 5000 Monte Carlo replications.

index	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
normal approximation intervals $\mathcal{C}_n^{NOR,TS}$ *(two-sided)*					$\mathcal{C}_n^{NOR,OS}$ *(one-sided)*			
$I^{W,1,0}$	0.6782	0.8390	0.8774	0.9212	0.6384	0.7924	0.8396	0.8952
$I^{W,2,0}$	0.8624	0.9228	0.9318	0.9396	0.8274	0.8966	0.9108	0.9310
$I^{B,0}$	0.8166	0.8922	0.9114	0.9256	0.7876	0.8548	0.8828	0.8928
$I^{W,1,1}$	0.6812	0.8370	0.8796	0.9220	0.6376	0.7964	0.8372	0.9020
$I^{W,2,1}$	0.8410	0.9076	0.9230	0.9370	0.8012	0.8786	0.8938	0.9210
$I^{B,1}$	0.8186	0.8944	0.9138	0.9260	0.7882	0.8582	0.8836	0.8944
hybrid bootstrap intervals $\mathcal{C}_n^{HB,TS}$ *(two-sided)*					$\mathcal{C}_n^{HB,OS}$ *(one-sided)*			
$I^{W,1,0}$	0.6508	0.8192	0.8626	0.9152	0.6310	0.7848	0.8350	0.8892
$I^{W,2,0}$	0.8480	0.9160	0.9286	0.9390	0.8290	0.8956	0.9090	0.9286
$I^{B,0}$	0.7060	0.8248	0.8734	0.9090	0.6964	0.8042	0.8446	0.8758
$I^{W,1,1}$	0.6434	0.8192	0.8602	0.9188	0.6236	0.7874	0.8264	0.8972
$I^{W,2,1}$	0.8258	0.9008	0.9186	0.9372	0.8060	0.8778	0.8926	0.9226
$I^{B,1}$	0.7100	0.8322	0.8798	0.9132	0.7026	0.8076	0.8512	0.8800
bootstrap-t intervals $\mathcal{C}_n^{BT,TS}$ *(two-sided)*					$\mathcal{C}_n^{BT,OS}$ *(one-sided)*			
$I^{W,1,0}$	0.9116	0.9206	0.9324	0.9380	0.8772	0.8970	0.9138	0.9318
$I^{W,2,0}$	0.9188	0.9374	0.9366	0.9430	0.9010	0.9302	0.9386	0.9430
$I^{B,0}$	0.9390	0.9584	0.9462	0.9346	0.9306	0.9384	0.9274	0.9112
$I^{W,1,1}$	0.9106	0.9176	0.9316	0.9382	0.8760	0.8990	0.9130	0.9340
$I^{W,2,1}$	0.9054	0.9274	0.9334	0.9396	0.8860	0.9134	0.9260	0.9356
$I^{B,1}$	0.9402	0.9564	0.9444	0.9354	0.9296	0.9376	0.9272	0.9106

$I^{W,j,\alpha}$= within-group component of group j,
$I^{B,\alpha}$= between-group inequality,
For exact definitions, see section 4.2.

Further experiments were conducted to examine the performance of confidence intervals for SHORROCKS and PRAIS mobility indices[4]. Again, differences between actual and nominal coverage probabilities were so small that the three methods can be regarded as performing equally well. In particular, using the studentized bootstrap does not seem to pay off for sample sizes beyond $n = 500$.

[4]Results for these experiments are available from the author on request.

6 Empirical Illustration

This section illustrates some of the techniques discussed for wage data from the German Socio-Economic Panel (GSOEP) (for details see BURKHAUSER ET AL.(1997)). All wages are monthly gross wages (from the previous year) for West Germany. Observations are weighted by their inverse inclusion probabilities in a given year, which were estimated by the providers of the survey. Wage inequality refers in all three examples to the subpopulation of wage earners, excluding zero wages. As the main conclusion from the previous section is that the simple bootstrap procedure is generally as good as any of the other methods, but much easier to implement, only hybrid bootstrap intervals $\mathcal{C}_n^{HB,TS}$ are reported.

Table 8. Change in wage inequality, West Germany, 1986 - 1996,
95% hybrid bootstrap intervals, 19601 observations (see text for details).

index	difference 1996 - 1986	lower bound	upper bound
$\mu_1\mu_0^{-1}$	1364.35	1248.47	1494.74
I_{GE}^{-1}	-0.0234	-0.0679	0.0163
I_{MLD}	-0.0136	-0.0358	0.0063
I_{Theil}	-0.0139	-0.0443	0.0149
I_{GE}^{2}	-0.0241	-0.1379	0.0841
$I_A^{0.5}$	-0.0062	-0.0172	0.0037
I_A^{1}	-0.0111	-0.0291	0.0053
I_A^{2}	-0.0164	-0.0477	0.0118
I_{CV}	-0.0365	-0.2028	0.1146
I_{logvar}	-0.0340	-0.0904	0.0186

Note: $\mu_1\mu_0^{-1}$ is the (weighted) mean wage

Source: German Socio-Economic Panel (GSOEP), weighted data.

The first example is the test of whether inequality in the distribution of wages has significantly changed from 1986 to 1996. The test takes into account the panel structure of the data[5] as well as possible non-response (for this test, all individuals appearing in either of the years were included, but zero wages or non-respondents were set to $c_{ti} = 0$ in the notation of section 4.1). The results in table 8 show that wage inequality fell during the period. However, this decrease was not statistically significant, as zero is contained in the confidence intervals. On the other hand, the mean wage $\mu_1\mu_0^{-1}$ increased in a statistically significant way.

[5] Actually, the GSOEP is a stratified multi-stage survey. Aspects of sample design other than its longitudinal nature and the sampling weights have not been taken into account for this illustrative exercise.

Table 9. Decomposition of wage inequality, foreigners vs. Germans, West Germany, 1996, 95% hybrid bootstrap intervals, 4616 observations.

statistic	estimate	lower bound	upper bound
$s^{1,-1}$	0.0753	0.0534	0.0939
$s^{2,-1}$	0.9208	0.9010	0.9438
$s^{W,-1}$	0.9961	0.9936	0.9996
$s^{B,-1}$	0.0039	0.0004	0.0065
$s^{1,0}$	0.0626	0.0469	0.0762
$s^{2,0}$	0.9313	0.9180	0.9471
$s^{W,0}$	0.9939	0.9899	0.9990
$s^{B,0}$	0.0061	0.0010	0.0101
$s^{1,1}$	0.0502	0.0344	0.0638
$s^{2,1}$	0.9431	0.9303	0.9574
$s^{W,1}$	0.9933	0.9888	0.9986
$s^{B,1}$	0.0067	0.0014	0.0113
$s^{1,2}$	0.0352	0.0144	0.0509
$s^{2,2}$	0.9595	0.9438	0.9798
$s^{W,2}$	0.9947	0.9912	0.9994
$s^{B,2}$	0.0053	0.0005	0.0088

$s^{1,\alpha}$ = contribution of inequality among foreigners,
$s^{2,\alpha}$ = contribution of inequality among Germans,
$s^{W,\alpha}$ = contribution of inequality within subgroups,
$s^{B,\alpha}$ = contribution of inequality between subgroups,
For exact definitions, see Section 4.2.

Source: German Socio-Economic Panel (GSOEP), weighted data.

The second application is a decomposition of wage inequality with respect to nationality. The population is partitioned into the two subgroups foreigners and Germans. The magnitude of the contribution of the between-group component $s^{B,\alpha}$ in table 9 suggests that wage differences between foreigners and Germans do not play an important role in explaining overall wage inequality.

7 Discussion and Conclusion

This paper showed that the bootstrap provides a valid procedure for statistical inference in many situations of inequality, mobility and poverty measurement. Statistical inference in this field is often encumbered by special features of the data used such as longitudinal correlation, panel attrition or non-response. The paper proposed simple ways to deal with these features by reformulating them as multivariate problems. In principle, all of the proposed procedures can be handled by conventional asymptotic methods, but two reasons seem to favor the use of a bootstrap procedure.

Firstly, Monte Carlo results suggest that for realistic population distributions and sample sizes, confidence intervals based on the simplest possible bootstrap procedure achieve the same coverage accuracy as intervals based on the conventional normal approximation. In some cases, a higher coverage accuracy can be obtained by using the studentized version of the bootstrap. However, as experiment 2 shows, this is not always the case.[6] Moreover, in large samples of several thousand observations, all three methods already perform so well that the difference between actual and nominal coverage accuracy is usually in the range of only one or two percentage points, i.e. small from a practical point of view.

Secondly, this equivalence for realistic sample sizes implies that practical and computational aspects move to the foreground. Both the normal approximation and the studentized bootstrap require an estimate of the asymptotic variance of the estimated statistic. If this statistic is a function of p moments then this amounts to estimating a $p \times p$ variance-covariance matrix and a $p \times 1$ vector of partial derivatives. In many cases, this renders the application of these two methods infeasible. For example, in the MOOKHERJEE/SHORROCKS(1982) decomposition of section 4.3, an intertemporal decomposition with $J = 5$ subgroups would require $p = 2(6J + 3) = 66$ moments, i.e. $(p + 1)p/2 = 2211$ distinct elements in the estimated variance-covariance matrix. But even in the cases, where a closed form expression for the asymptotic variance exists, the simple bootstrap is much easier to implement, as the program code used to compute the point estimates - which has to be programmed anyway - can be re-used to obtain estimates from the resample. Computing time may be a consideration, but should be no problem if compiled program code is used.

These points strongly suggest that the simple bootstrap should be the preferred method in practice, as it does not a require a variance estimate and is in most cases as accurate as the other methods. If a variance estimate is available then the studentized bootstrap should be chosen, as it generally leads to an improvement in accuracy.

[6]Compare the discussion between HOROWITZ(2001A,B) and MAASOUMI(2001) on the use of aymptotically pivotal statistics in econometrics.

8 References

BABU, G.J. and SINGH, K. (1983): Inference on means using the bootstrap, *The Annals of Statistics 11*, pp. 999 - 1003.

BICKEL, P.J. and D.A. FREEDMAN (1981): Some Asymptotic Theory for the Bootstrap, *The Annals of Statistics 9*, pp. 1196-1217.

BIEWEN, M. (2000): Income Inequality in Germany during the 1980s and 1990s, *Review of Income and Wealth 46(1)*, pp. 1 - 19.

BISHOP, J.A., K.V. CHOW and B. ZHENG (1995): Statistical Inference and Decomposable Poverty Measures, BULLETIN OF ECONOMIC RESEARCH 47:4, pp. 329 - 340.

BURKHAUSER, R.V., M. KREYENFELD and G.G. WAGNER (1997): The German Socio-Economic Panel: A Representative Sample of Reunited Germany and its Parts, *DIW Vierteljahreshefte zur Wirtschaftsforschung 1/97*, pp. 7-15.

COWELL, F. (1989): Sampling Variance and Decomposable Inequality Measures, *Journal of Econometrics 42*, pp. 27-41.

COWELL, F. (2000): *Measurement of Inequality*, in: ATKINSON, A.B. and F. BOURGUIGNON(ed.): *Handbook of the Income Distribution*, Elsevier, Amsterdam.

COWELL, F. and S. P. JENKINS(1998): Estimating Welfare Indices: Household Weights and Sample Design, *mimeographed*, London School of Economics.

DAVIDSON, R. and J.-Y. DUCLOS (1997): Statistical Inference for the Measurement of the Incidence of Taxes and Transfers, *Econometrica 65*, pp. 1453-1465.

DAVIDSON, R. and J.-Y. DUCLOS (2000): Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica 68*, pp. 1435-1465.

FIELDS, G.S. and E.A. OK (1999): The Measurement of Income Mobility: An Introduction to the Literature, in: SILBER, J. (ed.): *Handbook of income inequality measurement*, Kluwer Academic Publishers, Norwell, pp. 557 - 596.

FOSTER, J., J. GREER and E. THORBECKE: A class of decomposable poverty indices, *Econometrica 52*, pp. 761 - 765.

GRABKA, M. M., J. SCHWARZE and G. G. WAGNER (1999): How unification and immigration affected the German income distribution, *European Economic Review (43)4-6*, pp. 867-878

HALL, P. (1986): On the bootstrap and confidence intervals, *Annals of Statistics 11*, pp. 1156 - 1174.

HALL, P. (1988): Theoretical comparisons of bootstrap confidence intervals, *Annals of Statistic 16*, pp. 924 - 953.

HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer, New York etc.

HOROWITZ, J.L. (2000): The Bootstrap, forthcoming in: J.J. HECKMAN and E.E. LEAMER: *Handbook of Econometrics Vol.5*, Amsterdam: Elsevier.

HOROWITZ, J.L. (2001A): The bootstrap and hypothesis tests in econometrics, *Journal of Econometrics 100*, pp. 37-40.

HOROWITZ, J.L. (2001B): Response to comments of Esfandiar Maasoumi, *Journal of Econometrics 100*, pp. 97-98.

JÄNTTI, M. (1997): Inequality in Five Countries in the 1980s: The Role of Demographic Shifts, Markets and Government Policies, *Economica 64*, pp. 415 - 440.

JENKINS, S. P. (1995): Accounting for Inequality Trends: Decomposition Analyses for the UK, 1971 - 86, *Economica 62*, pp. 29-63.

JOHNSON, N.L. and S. KOTZ (1972): *Continuous multivariate distributions*, Wiley & Sons, New York.

KAKWANI, N. (1993): Statistical Inference in the Measurement of Poverty, *Review of Economics and Statistics 75*, pp. 632 - 639.

MAASOUMI, E. (2001): On the relevance of first-order asymptotic theory to economics, *Journal of Econometrics 100*, pp 83-86.

MAASOUMI, E. and M. Trede (1999): Comparing Income Mobility in Germany and the US using Generalized Entropy Mobility Measures, *Discussion Papers in Statistics and Econometrics 2/99*, University of Cologne.

MAASOUMI, E. and S. ZANDVAKILI (1986): A class of generalized measures of mobility with applications, *Economics Letters 22*, pp. 97 - 102.

MILLS, A.M. and S. ZANDVAKILI (1997): Statistical Inference Via Bootstrapping for Measures of Inequality, *Journal of Applied Econometrics Vol.12*, pp. 133-150.

MOOKHERJEE, D. and A.F. SHORROCKS (1982): A decomposition analysis of the trend in UK income inequality, *Economic Journal 92*, pp.886 - 902.

NYGÅRD, F. AND A. SANDSTRÖM (1989): Income Inequality Measures Based on Sample Surveys, *Journal of Econometrics 42*, pp. 81-95.

PRAIS, S.J. (1955): Measuring Social Mobility, *Journal of the Royal Statistical Society A 118*, pp. 56-66.

PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING and B.P. FLANNERY (1992): *Numerical Recipes in C*, Cambridge University Press, Cambrigde.

SCHLUTER, C. (1998): Statistical inference with mobility indices, *Economics Letters 59*, pp. 157-162.

SEIDL, C. (1988): Poverty Measurement: A Survey, in: BÖS, D., M. ROSE and C. SEIDL (eds.): *Welfare and Efficiency in Public Economics*, Springer, Berlin/New York.

SHAO, J. and D. Tu (1995): *The Jacknife and the Bootstrap*, Springer, New York etc.

SHI, X. (1986): A note on bootstrapping U-statistics, *Chinese Journal of Applied Probability and Statistics 2*, pp. 144 - 148.

SHORROCKS, A.F. (1978): Income Inequality and Income Mobility, *Journal of Economic Theory 19*, pp. 376-393.

SHORROCKS, A.F. (1980): The class of additively decomposable inequality measures, *Econometrica 48*, pp. 613-625.

SHORROCKS, A.F. (1983): The Impact of Income Components on the Distribution of Family Incomes, *Quarterly Journal of Economics*, pp. 311-326.

THISTLE, P.D. (1990): Large sample properties of two inequality indices, *Econometrica 58*, pp. 725-728.

TREDE, M. (2000): Bootstrapping Inequality Measures under the Null Hypothesis: is it worth the effort?, *mimeographed*, University of Cologne.

VAN DE GAER, D., N. FUNNELL and T. MCCARTHY (1999): Statistical inference for two measures or inequality when incomes are correlated, *Economics Letters 64*, pp. 295 -300.

VAN KERM, PH. (2000): Inference on inequality measures: a Monte Carlo Experiment, *mimeographed*, University of Namur.

WIEGAND, J. (1998): Pictures of a Revolution: Household on Low Incomes in East and West Germany after Reunification, *mimeographed*, University College London.