Statistical Inference via Bootstrapping for Measures of Inequality

by

Jeffrey A. Mills*
Sourushe Zandvakili**

Working Paper No. 136

February 1995

*University of Cincinnati
**The Jerome Levy Economics Institute and University of Cincinnati, Department of Economics, 1200 Crosley Tower, ML 371, Cincinnati, OH 45221

Abstract

In this paper we consider the use of bootstrap methods to compute interval estimates and perform hypothesis tests for decomposable measures of economic inequality. The bootstrap potentially represents a significant gain over available asymptotic intervals because it provides an easily implemented solution to the Behrens-Fisher problem. Two applications of this approach, using the PSID (for the study of taxation) and the NLSY (for the study of youth inequality), to the Gini coefficient and Theil's entropy measures of inequality, are provided. The results suggest that (i) statistical inference is essential even when large samples are available, and (ii) the bootstrap appears to perform well in this setting.

1. Introduction

Measures of inequality are widely used to study income and welfare issues. A major shortcoming of this literature is the lack of statistical measures of relative size. In particular, given a computed value for an inequality measure, is this computed value significantly different from a benchmark (such as complete equality)? Further, in a dynamic setting, as observed inequality changes, is there a statistically significant change in the inequality measure over time? To answer these questions we need interval estimates for these measures.

It is evident from the empirical research on inequality that interval estimation and statistical testing are largely absent at this point in time.[1] Statistical measures are currently available, but all of the existing statistical theory in this area is based on asymptotic approximations.[2] The need for statistical inference with small samples should be obvious, but even for large samples it may be essential to report statistical measures of precision. As Maasoumi (1994) point out, the argument that measures of precision are unnecessary when large samples are available, because central limit theorems ensure convergence of a consistent estimator to the population value, is occasionally contradicted by large standard errors. Also, the rate of convergence may be slow for these statistics.

The problem with constructing interval estimates for any of the measures of inequality used in the literature is that they are all nonlinear functions of a random variable (usually income), and so do not readily lend themselves to standard statistical techniques. Interval estimates are available from asymptotic theory, however, the

[1] Cowell (1989a) is a notable exception.

[2] See, for example, Gastwirth (1974), Gastwirth *et al.* (1986) and Cowell (1989a). Maasoumi (1994) provides a thorough review and some examples of use of asymptotic results.

small sample properties of these intervals are not known. Further, all the decomposable inequality measures used in the literature are bounded (e.g. the Gini coefficient lies in the $[0, 1]$ interval), so that application of standard asymptotic results may lead to estimated intervals that extend beyond the theoretical bounds of a particular measure (e.g. a negative lower bound for Gini).

An alternative method for computing probability intervals is to bootstrap. The bootstrap provides interval estimates drawn from the small sample distribution. These interval estimates have been shown to be superior to asymptotic intervals both theoretically and in a variety of applications.[3] Bootstrap intervals are computationally inexpensive and easy to calculate, the same method applies to all the inequality measures used in the literature, and the bootstrap method automatically takes into account any bounds that apply to a particular measure. Further, since bootstrap intervals computed using the percentile method have a clear Bayesian interpretation, they provide a straightforward solution to the Behrens-Fisher problem of comparing means from two distributions (see section 3).

Given the potential advantages from bootstrapping, it appears worthwhile to consider its use as a tool for statistical inference for inequality measures. In this paper bootstrap methods are used to compute standard errors and probability intervals and to conduct hypothesis testing for two inequality measures – the Gini coefficient and Theil's entropy. We consider two data sets, the PSID, which provides us with a small sample of before and after tax average income within states in the U.S., and the NLSY, from which we extract a relatively large sample of income levels for youths in the U.S. and consider decompositions of Theil's measure based on age groupings.

A brief description of the inequality measures is given in the next section. Section

[3]See, e.g., Burr (1994), Freedman and Peters (1984a,b) and Hall (1992).

3 outlines the bootstrap method. The empirical applications are presented in section 4. Section 5 draws some conclusions.

2. Measures of Inequality

Early attempts to measure inequality led to the use of several *ad hoc* nonparametric descriptive statistics. The most popular of these is the Gini coefficient. Though it has been shown to be inferior to the more recently developed axiom based measures, it has a number of advantages over other *ad hoc* measures and it is still widely used in empirical studies. Further, as Cowell (1989a) points out, while there are good reasons to restrict attention to decomposable measures, the Gini falls within this class, though only in a limited sense.[4]

The Gini is defined as one half of the relative mean difference, which is the arithmetic average of the absolute value of the difference between all pairs of incomes. It is given by:

$$G = 1/2n^2\mu_y \sum_{i=1}^{n}\sum_{j=1}^{n} |y_i - y_j| \qquad (2.1)$$

where y_i is the income of the ith individual and μ_y is the sample mean.

G is equal to 1 when inequality is at its maximum and is zero with an equal distribution. This measure does not satisfy the property of full additive decomposability, though less desirable forms of decomposition are possible. Also, this measure is most sensitive with transfers toward the middle of the distribution, and least sensitive toward the two tails. Thus, it should be avoided if activity around the tails is of concern, i.e. tax and transfer analysis. However, due to it's popularity in applied research, and since the statistical properties of the Gini are not known (so that it

[4]See Cowell (1989a) for a thorough discussion.

may possibly be superior to other measures in this regard), we adopt the Gini as one example of an inequality measure in our applications.

The axiomatic approach to the measurement of inequality requires a number of desirable axioms to be satisfied. These axioms are symmetry, decomposability, principle of transfer, mean independence, and rank dominance. The only class of measures satisfying the noted axioms is the Generalized Entropy family of measures. This relationship has been established by, among others. Bourguignon (1979), Shorrocks (1980), Cowell and Kuga (1981), Foster (1983) and Maasoumi (1986). Some well known measures of inequality are special cases of this class of measures.

In light of the intuitive appeal of the axioms adopted, these measures seem superior to the *ad hoc* nonparametric measures. As a second example for our application, we adopt one of Theil's (1967) measures of inequality. which is a member of this family. It is given by:

$$T = \sum_{i=1}^{n} s_i log(ns_i) \tag{2.2}$$

where $s_i = y_i / \sum_{i=j}^{n} y_j$ is the relative share of ith individual's income.

The Theil measure is additively decomposable by population subgroups. This is a desirable property for an inequality measure since. in practice, it is often important to compare inequality both within and between subgroups of the population based on various population characteristics (age, race. education, etc.). Suppose there are ω population subgroups, each with n_j members such that $\sum_j n_j = n$. Then we can write

$$T = \sum_{j=1}^{\omega} s_j^* \ln(\frac{n}{n_j} s_j^*) + \sum_{j=1}^{\omega} s_j^* \sum_{i=1}^{n_j} s_i^j \ln(n_j s_i^j) \tag{2.3}$$

where $s_j^* = \sum_{i=1}^{n_j} y_i^j / \sum_{k=1}^{n} y_k$ is the relative share of income for the jth group, $s_i^j = y_i^j / \sum_{i=1}^{n_j} y_i^j$ is the share of total group income of the ith individual in the jth group.

The first term on the RHS measures between group inequality, the second measures within group inequality. We consider inference for both the aggregate Theil measure (2.2) and for the within and between group components given in (2.3) in our large sample application of the bootstrap.

It is worth clarifying at this point that the bootstrap approach we study is directly applicable to *any* other measure of inequality that has been used in the literature (including nondecomposable measures). We restrict attention to the Gini and Theil measures solely for clarity of exposition.

3. The Bootstrap Method

The bootstrap is a method for recovering the distribution of a statistic by employing simulation methods to approximate the small sample distribution. It has proved superior to asymptotic methods both on theoretical grounds and in a number of studies.[5] In this section we outline the bootstrap method, and suggest its use for obtaining standard errors and probability intervals and for hypothesis testing, for measures of inequality.

Suppose a random sample of size n is observed from a completely unspecified probability distribution, F:

$$X_i = x_i, \ \ X_i \sim F, \ i = 1, ..., n$$

with X_i independent for all i. Let $X = (X_1, X_2, ..., X_n)$ and $x = (x_1, x_2, ..., x_n)$ denote the random sample and its observed realization. Given a specified statistic $H = H(X, F)$, possibly depending on both X and the unknown distribution F, we

[5]See Efron (1979,1982), Hall (1988,1992). Bhattacharya and Qumsiyeh (1989), Freedman and Peters (1984a,b) and Burr (1994).

wish to estimate the sampling distribution of H on the basis of the observed data x.

The bootstrap method is as follows:[6]

1. Construct the sample probability distribution $\hat{F}$, putting mass $1/n$ at each point $x_1, ..., x_n$.

2. With $\hat{F}$ fixed, draw a random sample of size n, *with replacement,* from $\hat{F}$, say $X^* = x^*$. This is the bootstrap sample.

3. Approximate the sampling distribution of H by the *bootstrap distribution* of $H^* = H(X^*, \hat{F})$.

4. The bootstrap distribution is obtained by Monte Carlo approximation. Repeated realizations of X^* are generated by taking random samples of size n from $\hat{F}$, and the histogram of the corresponding values of $H^* = H(x^{*i}, \hat{F})$, for $i = 1, ..., m$ samples of size n, is taken as an approximation to the actual bootstrap distribution.

Given this bootstrap estimate of the sampling distribution of H, we can then calculate standard errors, confidence or probability intervals, and conduct hypothesis testing.

The bootstrap provides a numerical approximation to the distribution of interest, F, that is similar to a high-order Edgeworth expansion (an approximation to a distribution function that involves a series expansion around the Normal distribution). Edgeworth expansions can represent considerable improvements over Normal approximations, and the bootstrap is typically superior to a practically calculable (i.e. short) Edgeworth expansion. Bhattacharya and Qumsiyeh (1989) show that the bootstrap estimate of F outpeforms the short Edgeworth approximation in any L^p metric. Further, Hall (1992) shows that if values of x are of larger order than

[6]See Efron (1979,1982) for a full exposition.

$n^{1/6}$, the Edgeworth series will either not converge or not adequately describe the tail probabilities. For example, when $n = 20$, $n^{1/6} = 1.648$ which is approximately the 5% point of a Standard Normal distribution. The bootstrap on the other hand, provides an accurate approximation to tail probabilities for values of x as large as $o(n^{1/3})$. For $n = 20$, $n^{1/3} = 2.714$, which has an approximate tail probability of 0.003 for the Standard Normal. These crude calculations suggest that the bootstrap provides far more accurate estimates of tail probabilities than asymptotic approximations. [7]

Tail probability values for hypothesis tests with regard to a benchmark value can be calculated directly from the bootstrap distribution in the same manner as probability intervals. Often however, we are more interested in comparing different values of an inequality measure, such as for different points in time (has inequality increased or decreased over time?). This involves comparison of two values of the statistic H_t, each with it's own sampling distribution, F_t, $t = 1, 2$. We suggest the following test for this case, analogous to the comparison of two means from two different samples.

Consider the statistic $D = H_1 - H_2$, where H_1 and H_2 are the two values of the inequality measure we wish to compare. The distribution of D can be bootstrapped in the same manner used to obtain distributions for H_1 and H_2. Tail probability values for hypotheses regarding D can be calculated directly from the bootstrap distribution $\hat{F}(D)$.

Note that the interpretation of tail values from the bootstrap distribution as probabilities has a legitimate justification. As Efron (1982) points out, if we take the prior distribution of the density function from F to be a Dirichlet distribution with param-

[7] See Hall (1992) for details.

eter a, and let $a \to 0$ to represent prior ignorance, then the bootstrap distribution is a close (discrete) approximation to the posterior density from Bayesian inference.[8]

The hypothesis test we conduct using the statistic D involves the comparison of means of two distributions, which has become known as the Behrens-Fisher problem. The problem is a very difficult one within the classical hypothesis testing framework and consequently there is no generally accepted classical procedure for this problem. By contrast, the Bayesian procedure is straightforward.[9] The bootstrap method we adopt is a simple implementation of this Bayesian procedure. If the bootstrap performs well in this situation, this represents an important advantage of bootstrap methods over the use of asymptotic interval estimates.

Several alternative methods for calculating bootstrap intervals are available. We use what has become known as the "percentile method" to calculate tail probabilities for several reasons: it performed well in a recent comparison of the different methods by Burr (1994), it is the easiest to compute. and unlike the other methods, it has a clear interpretation as an approximation to a Bayesian posterior probability interval (allowing a straightforward solution to the Behrens-Fisher problem). The other method most often recommended is the "boot-t". The boot-t requires estimation of the standard error by some other method (usually the asymptotic estimate). Burr (1994) does not recommend the boot-t because she found it to be very unstable. As Burr suggests, use of the iterated bootstrap may improve the performance of the boot-t, but this is computationally very expensive. We compare "naive" boot-t intervals (a symmetric interval around the sample estimate of H using the bootstrap estimate of the standard error and standard t tables [see Efron (1982)]) with the per-

[8]See Efron (1982), p.81-82, and Rubin (1981).
[9]See DeGroot (1986) and Jaynes (1976).

centile intervals for a large sample in section 4 and find close agreement. However, we especially prefer the percentile method for small samples because it automatically takes into account bounds on the statistic, whereas the boot-t can lead to confidence intervals that are wider than is theoretically possible.[10]

An important caveat regarding application of the bootstrap is that independence of observations in the sample, X_i, is required for step 2., sampling with replacement, to be valid. This does not necessarily preclude use of the method in dynamic settings however. Provided an independent cross section sample for each time period is used to form the bootstrap distribution for that time period, the dependence between time periods will automatically be taken into account, i.e. the bootstrap distribution obtained in a given period is conditional on the data observed in previous periods if the current realizations are statistically dependent on previous realizations (in this case the bootstrap distribution is an estimate of the Markov transition probabilities $p(X_t|X_{t-1})$. Notice that unconditional inference in this setting would imply a violation of the Likelihood Principle. Examples of this use of the bootstrap are provided in the following section.

4. Empirical Applications

It has often been argued that the large samples typically available for the empirical measurement of inequality obviate the need for statistical hypothesis testing. In what follows we study the usefulness of the bootstrap approach both with small and large samples. We also use the bootstrap with decompositions of the Theil measure, which can lead to inference with very small samples. To allow comparison we compute the asymptotic standard errors for the Gini, Theil and decompositions of Theil using the

[10] See the discussions by Efron, and Buckland, Garthwaite and Lovell following Hall (1988).

results given by Cowell (1989a).[11]

4.1. Empirical Implementation

The empirical studies in this section were carried out using GAUSS 3.0 on a DOS based PC with a 486 50 mhz processor. As one would expect, run times varied considerably depending on the sample size, etc. However, to give some idea of the computational burden involved we observed the following run times. For the small sample study (52 observations for each year), it took approximately 1 min., 30 sec. to produce all the results reported in Tables 1 and 2. Tables 3 and 4 took approximately 2 min., 36 sec., and Table 5 took about 1 min. For the large sample study (4266 observations per year) it took approximately 61 mins., 42 sec. to produce Table 6, and Table 7 took approximately 3 hrs. and 41 mins. The run time for Table 8 was 30 mins, 18 sec. and for Table 9, 3 hrs, 7 mins. Note that Table 9 was computationally less burdensome than Table 7. This is because Gini took far longer than Theil to compute in all cases, despite using the faster algorithm for Gini suggested by Cowell (1989b).

We calculated the bootstrap distribution of each statistic using 500 iterations of the procedure outlined in section 3. As a check, for the small sample study we increased the number of iterations to 2000 and obtained almost identical results as with 500 iterations.

4.2. A Small Sample Study

For our small sample we use data from the Panel Study of Income Dynamics (PSID) for 50 states, the District of Columbia, and those of Americans living outside

[11]We note a typographical error in Cowell's expression for the variance of the between group Theil measure; the RHS of his equation (42) should be $(n_j/nm_{11})(\ln a_j + 1)$.

of the United States. The data consist of 52 observations per year, from 1983 to 1988, on the mean level of income before and after taxes. There are two ways to evaluate the usefulness of hypothesis testing with this data set. First, we compare the observed change in inequality over time. Second, we test the significance of the progressivity in taxes to reduce the after tax inequality in each time period.

It has been the norm for most empiricists in the field of income distribution to calculate an index of inequality and provide an interpretation of the results. Most often the calculated results pose more questions than answers. Tables 1 and 2 present the calculated inequality based on Gini and Theil measures of inequality. Inequality in before and after tax income has fluctuated from 1983 to 1988. However, in a given year, post-tax inequality among states has increased regardless of our choice of inequality measure. Casual observation suggests a trend from one year to the next, and from pre- to post-tax income. This observation is inadequate however, since one cannot say whether the observed changes are significant. We provide several relevant statistical measures of precision to address this inadequacy. First, bootstrap standard errors and probability intervals are provided and compared with asymptotic standard errors. Second, hypothesis tests for changes in the calculated inequalities between two periods are conducted (for which there is no asymptotic equivalent).

Our starting point is to look at bootstrap standard errors and .99 and .95 probability intervals for the small sample. Tables 1 and 2 present these statistics for pre- and post-tax income using the Gini and Theil measures of inequality.

A comparison of the bootstrap and asymptotic standard errors reported in Tables 1 and 2 indicate that for the Theil measure they are similar, whereas for Gini there is a substantial difference in these estimates. This leads to the question: which estimates of the standard error are the best? Since the underlying small sample distribution for these statistics is not known, there is no definite yardstick for comparison. For this same reason, it is also difficult to construct a Monte Carlo study to address this issue.

One possible explanation is that the theoretical bounds on Gini have an important effect truncating the tails of the distribution in a small sample. The small sample distribution may also be skewed. If we consider the bootstrap as a numerical evaluation of an Edgeworth expansion, then it can be shown that the third cumulant (skewness) provides the largest gain in the accuracy of the approximation over a Normal approximation.[12] Thus, if the sample distribution is skewed, we would theoretically expect the bootstrap to outperform a Normal approximation, and the more skewed the distribution the greater the difference in the bootstrap and Normal intervals.

As demonstrated in Tables 1 and 2, there is a substantial amount of overlap between years for both measures of inequality and different definitions of income. This is true for both .99 and .95 probability intervals. Thus, one cannot say with any degree of confidence whether the observed changes are significant. For example, measured inequality based on Gini in 1983 and 1984 shows a decline from .176 to .155, but the confidence intervals overlap substantially at 99%, with upper bounds of .248 and .193 and lower bounds of .113 and .107. Comparing changes in Gini from pre-tax income of .176 to post-tax income of .188, one observes a similar overlap. Tests of these hypotheses are needed to make the observed changes from one year to the next, and from pre- to post-tax income, meaningful.

Our solution is to compare the observed changes in inequality from one year to the next, or any succeeding year, and perform hypothesis testing to gauge the significant of such a change. Note that since our statistics are based on a cross section of

[12]See Hall (1992) for details.

observations that can reasonably be assumed to be independent within each period, use of the bootstrap is valid in this case. Tables 3 and 4 show the change in inequality for every pair of years from 1983 to 1988 based on Gini and Theil measures, for both before and after tax income. This is followed by standard errors for the estimate of the observed change, and the probability that such a change is less than zero, p (i.e., $p = Pr(D < 0)$, so that $1 - p = Pr(D \geq 0)$).

Using pre-tax income, and comparing 1983 to all other succeeding years for the Gini and Theil measures (Table 3), shows that only the change from 1983 to 1988 for the Theil measure is significant at the 10% level ($p = 0.065$). All other pair comparisons for 1983 are insignificant at the 10% level. A similar comparison for 1984 to each of the succeeding years shows that the change from 1984 to 1987 is significant at the 5% level for Gini and at the 10% level for Theil. All other paired comparisons are insignificant for 1984. Paired comparisons of 1985 through 1988 show that only the change from 1986 to 1987 is significant (at the 10% level for Gini and at the 5% level for Theil). A comparison of changes from 1986 to 1987 and 1988 suggests that only the observed changes from 1986 to 1987 are significant at the 5% level. The paired comparison of changes from 1987 to 1988 is insignificant.

The results based on after tax income are shown in Table 4. A paired comparison of all possible changes shows that none of the observed changes are significant at the 10% level. The only observed significant change is from 1984 to 1987, for which p is .106.

Often one would like to evaluate the impact of taxes, particularly when taxes are changing over time on the distribution of income. Table 5 shows the changes in inequality from before to after tax income. The only observed significant change is for 1983 based on the Theil measure of inequality, where p is .904. Thus, the observed change is significant at the 10% level. For all other years the observed changes are not significant at the 10% level.

4.3. A Large Sample Study

For our large sample we use the National Longitudinal Survey of Youth (NLSY) 1979-1989, which contains individuals who were 14-21 years of age in 1979. We use data from 1984-89. Thus, in 1984 these individuals were between the ages of 19-26 and 24-31 years of age in 1989. The same individuals are followed over six years. The shorter duration is to minimize the effects of attrition or dropouts. There are 4266 observations for the duration under consideration. Individuals with positive income who are active in the labor market were chosen. Their nominal annual earnings has been adjusted to real earnings to reflect the change in price levels over time using a 1982 base year price. Reported income for these individuals has been top coded by the NLSY. According to the NLSY their method of top coding was changed for 1989. Thus, the top coding procedure has been adjusted to be consistent with previous years. The top coding values used are, $75,001 for 1984, $100,001 after 1984, and was based on average income of those earning more than $100,000 in 1989. A draw back of top coding is its underestimation of inequality.

Youth earnings inequality from 1984-1989 is reported in Table 6. The results suggest a decline of overall annual inequality. This decline is observed under two different choices of the inequality measure. The magnitude of the observed inequality is smaller with Theil's measure compared to the Gini coefficient, whereas the Gini has lower standard errors than Theil's for the duration under consideration.

The questions of concern at this point are two-fold. First, is the measured inequality for each period significant? Second, is the change in inequality from one period to the next significant? Gini declines in each of the six periods. However, one cannot say with any confidence that a decline of Gini from .3754 in 1986 to .3699 in 1987 is significant without interval estimates. The same is true with regard to the observed decline of Theil from .2512 in 1984 to .2015 in 1989. Statistical inference is required to make such observations. Bootstrapping allows us to measure standard errors and conduct such tests of statistical significance. As in the small sample study, since the bootstrap is employed using cross section data on individuals for each time period (or pairwise comparison), it is unlikely that the independence assumption is violated.

The two major concerns with respect to measured inequality can be addressed within the context of Table 6. Firstly, it is evident from the calculated upper and lower bounds, that the value of the inequality measures that represents complete equality (zero for both measures) falls well outside the confidence intervals for all periods. This observation is true at both 99% and 95% confidence levels and for both measures of inequality. Thus, we can say with near certainty that the measured inequality is significantly different from complete equality.

Secondly, as the measured inequality declines over time, it is more difficult to say that this decline is significant. As is evident from Table 6, we cannot say that the decline from one period to the next is significant because the bounds from one period to the next overlap. For example, at 99% there is considerable overlap between the lower and upper bounds for Gini for 1986 (0.3645,0.3863) with those for 1987 (0.3600,0.3806). This same pattern is true for Theil's measure of inequality from one period to the next. However, as we extend the duration under consideration to more than two years, the observed decline in inequality becomes significant. Generally, the bounds overlap less the further apart the years under consideration. For example when 1985 and 1987 are considered the .95 probability intervals do not overlap, but the .99 intervals do. As the duration increases, the bounds move further apart; when 1985 is compared with 1988, the .99 probability interval does not overlap at all, so this observed decline is highly significant.

Comparing the bootstrap and asymptotic standard errors we find that with a sample as large as this the asymptotic standard errors are very similar to those from bootstrapping, though they are still slightly larger in some cases. The probability intervals in Table 6 were obtained directly from the bootstrap distribution, so no assumptions are made about the *a priori* form of this distribution. Alternatively, confidence intervals can be calculated by the naive boot-t method using the standard errors from Table 6, relying on standard t tables for the significance point, as suggested by Tukey.[13] We found that these alternative intervals are very similar for our data, for example, for Theil 1989, the estimated standard error is 0.0043. From t tables, the critical t value for a 5% significance level is 1.96. The estimated value of Theil is 0.2015, therefore, a 95% confidence interval based on this standard error is (0.1931,0.2099), which compares closely with the bootstrap interval (0.1934,0.2101).

Use of t tables however, involves implicit assumptions not made when using the bootstrap distribution directly, for example, the t distribution is symmetric, whereas the bootstrap distribution can be skewed (income distributions are typically skewed). A critical problem with use of t tables is that the t distribution assumes that the statistic under study is unbounded. Both the Gini and the Theil measures are bounded above zero (Gini is also bounded below unity), which is automatically taken into account when using the bootstrap distribution directly. This problem also applies to asymptotic standard errors, and so the bootstrap intervals can be considered superior to asymptotic intervals, even in large samples, from a theoretical standpoint.

[13]See Efron (1979).

A statistical test for change in inequality, for each pair of calculated inequalities, is provided in Table 7, where tail probability values, p, for $D = H_1 - H_2$ are provided. It is evident from the paired comparisons between 1984 and 1989 that, with three exceptions, all observed changes are significant at the 5% level based on Gini and Theil measures of inequality. The exceptions are 1984-1985, 1986-1987, and 1988-1989. The first two are significant at the 10% level while the latter is not. So even with a large sample, one has to reserve judgment and base conclusions on statistical inference.

4.4. Decomposition by Population Subgroups

Obviously there are significant differences in income levels among individuals due to differing characteristics, such as age, race, gender and human capital. For policy formulation it is often important to provide some evidence of the degree of inequality both due to these factors, and after these factors have been taken into account. This suggests the need to consider decompositions of the inequality measures.

To evaluate the bootstrap in this context, individuals in the NLSY sample were split into three cohorts based on their age in 1984; 23-26, 27-29 and 30-32. Tables 8 and 9 report the results for the Theil measure decomposed by age groupings. The results suggest that there is very little inequality between these groups, so that most of the observed inequality is due to within group factors.

The bootstrap and asymptotic standard errors compare closely for the within group measure, but the asymptotic s.e's are considerably smaller than the bootstrap s.e's for the between group measure (e.g. 0.0003 compared to 0.0009 for 1987 and 1988) suggesting that the asymptotic s.e's may be biased downwards. Table 9 indicates that there is a significant increase in between group inequality from 1984 to 1988, whereas within group inequality declines significantly during this period. In most cases the change in inequality from one period to the next is significant. There are notable exceptions however. For example, between group inequality does not change significantly from 1987 to 1988 and 1989, and none of the changes are significant at the 10% level from 1988 to 1989.

5. Conclusion

Using the PSID and NLSY income data, bootstrap estimates of standard errors and probability intervals were calculated for the Gini coefficient and Theil's entropy measure of inequality. The bootstrap was also used to perform hypothesis tests regarding the statistical significance of changes in these inequality measures. We find that statistical inference is essential even with a sample of over 4,000 observations.

The bootstrap provides an alternative method of inference to asymptotic standard errors. We find that the bootstrap estimates are easy to compute and compare favourably to asymptotic standard errors. There is also an extensive theoretical literature showing that the bootstrap improves upon asymptotic intervals under very general conditions.[14] Further, theoretical bounds on the various inequality measures are automatically taken into account when using the percentile method to compute intervals. The percentile method also has a clear Bayesian interpretation which, in light of the evidence suggesting superiority of Bayesian intervals (particularly when faced with the Behrens-Fisher problem), is comforting.[15]

Of the two inequality measures considered, the Theil measure is generally considered superior on theoretical grounds. We find further reason to prefer the Theil

[14] See Hall (1992) and Bhattacharya and Qumsiyeh (1989).
[15] See, for example, Jaynes (1976) and DeGroot (1986).

measure for empirical studies: it is somewhat less computationally burdensome and, more importantly, the bootstrap estimates of standard errors compare more closely with the asymptotic estimates in the small sample study, suggesting that the small sample distribution of the Gini may be considerably different from the Normal.

The fact that decomposable inequality measures are nonparametric, highly nonlinear functions of the observed data, coupled with the fact that there is no generally accepted asymptotic procedure for conducting the hypothesis tests performed in section 4, leads us to conclude that the bootstrap may be the only currently viable method for statistical inference with regard to changes in inequality measures over time.

Further work needed involves comparison of the performance of the different bootstrap methods, though a review of previous studies of the bootstrap suggests that potential gains in accuracy are likely to be of second order at best.[16]

[16] *c.f.* Burr (1994).

References

Bhattacharya, R. N. and M. Qumsiyeh (1989): "Second Order L^p-Comparisons Between the Bootstrap and Empirical Edgeworth Expansion Methodologies," *Annals of Statistics*, 17, 160-169.

Bourguignon, F. (1979): "Decomposable Income Inequality Measures," *Econometrica*, 47, 901-20.

Cowell, F.A. (1989a): "Sampling Variance and Decomposable Inequality Measures," *Journal of Econometrics*, 42, 27-41.

Cowell, F.A. (1989b): "Analysis of Income Distribution Using Microcomputer Technology," *Research on Economic Inequality*, 1, 249-267.

Cowell, F.A., and K. Kuga (1981): "Inequality Measurement: An Axiomatic Approach," *Journal of Economic Theory*, 15, 287-305.

DeGroot, M. H. (1986): *Probability and Statistics. Second Edition.* Addison-Wesley.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jacknife." *Annals of Statistics,* 7, 1-26.

Efron, B. (1982): *The Jacknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

Foster, J. E. (1983): "An Axiomatic Characterization of the Theil Measure of Income Inequality," *Journal of Economic Theory*, 31, 105-121.

Freedman, D. A. and S. C. Peters (1984a): "Bootstrapping an Econometric Model: Some Empirical Results." *Journal of Business and Economic Statistics,* 2, 150-158.

Freedman, D. A. and S. C. Peters (1984b): "Some Notes on the Bootstrap in Regression Problems." *Journal of Business and Economic Statistics,* 2, 406-409.

Gastwirth, J. L. (1974): "Large Sample Theory of Some Measures of Inequality." *Econometrica, 42,* 191-196.

Gastwirth, J. L., T. K. Nayak, and A. N. Krieger (1986): "Large Sample Theory for the Bounds on the Gini and Related Indices from Grouped Data." *Journal of Business and Economic Statistics,* 4, 269-273.

Hall, P. (1988): "Theoretical Comparison of Bootstrap Confidence Intervals (With Discussion)," *Annals of Statistics,* 16, 927-985.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

Jaynes, E. T. (1976): "Confidence Intervals vs. Bayesian Intervals," in *Foundations* *of Probability Theory, Statistical Inference, and Statistical Theories of Science,* W. L. Harper and C. A. Hooker, Eds., D. Reidel Publishing Co., Dordrecht, Holland.

Maasoumi, E. (1986): "The Measurement and Decomposition of Multi-Dimensional Inequality," *Econometrica*. 54, 991-97.

Maasoumi, E. (1994): "Empirical Analysis of Inequality and Welfare," forthcoming in the *Handbook of Applied Microeconomics*, P. Schmidt and H. Pesaran (eds.).

Rubin, D. B. (1981): "The Bayesian Bootstrap." *Annals of Statistics,* 9, 130-134.

Shorrocks, A.F. (1980): "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48, 613-625.

Theil, H. (1967). *Economics and Information Theory.* North-Holland: Amsterdam.

TABLE 1: BOOTSTRAP STANDARD ERRORS AND CONFIDENCE INTERVALS PSID BEFORE TAX INCOME DATA (SMALL SAMPLE)

	1983	1984	1985	1986	1987	1988
Gini	.176	.155	.161	.159	.183	.171
S.E.	.029	.018	.021	.022	.022	.021
A.S.E.	.039	.029	.030	.031	.036	.031
99% U.B.	.248	.193	.209	.210	.238	.217
99% L.B.	.113	.107	.112	.107	.120	.128
95% U.B.	.224	.182	.194	.193	.227	.203
95% L.B.	.124	.120	.123	.120	.138	.138
Theil	.063	.041	.045	.046	.063	.051
S.E.	.022	.010	.012	.013	.015	.012
A.S.E.	.025	.010	.012	.014	.018	.012
99% U.B.	.124	.067	.076	.084	.105	.083
99% L.B.	.020	.020	.020	.019	.026	.029
95% U.B.	.106	.057	.067	.069	.095	.071
95% L.B.	.025	.025	.026	.024	.036	.033

Sample size = 52

S.E. = bootstrap standard error of the estimate
A.S.E. = asymptotic standard error of the estimate
U.B. = Upper Bound
L.B. = Lower Bound

TABLE 2: BOOTSTRAP STANDARD ERRORS AND CONFIDENCE INTERVALS PSID AFTER TAX INCOME DATA (SMALL SAMPLE)

	1983	1984	1985	1986	1987	1988
Gini	.188	.168	.174	.173	.186	.178
S.E.	.036	.018	.020	.020	.025	.021
A.S.E.	.044	.029	.029	.030	.034	.031
99% U.B.	.261	.214	.214	.216	.248	.221
99% L.B.	.115	.116	.125	.126	.127	.126
95% U.B.	.239	.200	.210	.203	.222	.207
95% L.B.	.127	.133	.139	.139	.141	.142
Theil	.074	.047	.052	.051	.062	.053
S.E.	.031	.010	.011	.012	.015	.011
A.S.E.	.032	.011	.011	.012	.015	.012
99% U.B.	.143	.076	.079	.084	.103	.079
99% L.B.	.024	.023	.025	.027	.028	.028
95% U.B.	.122	.065	.068	.071	.085	.072
95% L.B.	.027	.030	.033	.033	.036	.034

Sample size = 52

TABLE 3: HYPOTHESIS TESTS FOR GINI AND THEIL MEASURES PSID BEFORE TAX INCOME DATA (SMALL SAMPLE)

			1984	1985	1986	1987	1988
1983	Gini	D	.0217	.0145	.0170	.0065	.0049
		S.E.	.0272	.0287	.0293	.0250	.0252
		p	.745	.610	.640	.410	.522
	Theil	D	.0218	.0171	.0166	.0002	.0120
		S.E.	.0223	.0227	.0236	.0212	.0208
		p	.745	.660	.637	.465	.065
1984	Gini	D		.0071	.0046	.0282	.0167
		S.E.		.0147	.0116	.0168	.0177
		p		.275	.282	.025	.202
	Theil	D		.0046	.0051	.0215	.0097
		S.E.		.0083	.0060	.0111	.0091
		p		.242	.172	.067	.147
1985	Gini	D			.0025	.0210	.0095
		S.E.			.0109	.0127	.0177
		p			.512	.070	.277
	Theil	D			.0004	.0168	.0051
		S.E.			.0068	.0083	.0101
		p			.410	.032	.270
1986	Gini	D				.0236	.0120
		S.E.				.0128	.0160
		p				.037	.240
	Theil	D				.0164	.0046
		S.E.				.0079	.0088
		p				.020	.295
1987	Gini	D					.0115
		S.E.					.0168
		p					.692
	Theil	D					.0117
		S.E.					.0113
		p					.815

D = Difference in inequality measure (changes in Gini and Theil respectively)
S.E. = bootstrap standard error of D
p = probability that D is less than zero

TABLE 4: HYPOTHESIS TESTS FOR GINI AND THEIL MEASURES PSID AFTER TAX INCOME DATA (SMALL SAMPLE)

			1984	1985	1986	1987	1988
1983	Gini	D	.0198	.0132	.0150	.0018	.0110
		S.E.	.0309	.0315	.0319	.0248	.0310
		p	.678	.618	.656	.536	.582
	Theil	D	.0268	.0231	.0234	.0128	.0214
		S.E.	.0248	.0280	.0283	.0221	.0297
		p	.726	.690	.692	.658	.678
1984	Gini	D		.0066	.0047	.0179	.0088
		S.E.		.0120	.0162	.0186	.0149
		p		.304	.358	.160	.300
	Theil	D		.0037	.0034	.0140	.0054
		S.E.		.0062	.0081	.0109	.0081
		p		.290	.314	.106	.270
1985	Gini	D			.0018	.0113	.0022
		S.E.			.0134	.0156	.0127
		p			.584	.294	.434
	Theil	D			.0002	.0103	.0017
		S.E.			.0071	.0094	.0069
		p			.532	.174	.426
1986	Gini	D				.0131	.0040
		S.E.				.0188	.0138
		p				.250	.412
	Theil	D				.0105	.0019
		S.E.				.0107	.0075
		p				.188	.444
1987	Gini	D					.0091
		S.E.					.0077
		p					.774
	Theil	D					.0085
		S.E.					.0077
		p					.878

TABLE 5: HYPOTHESIS TESTS FOR GINI AND THEIL MEASURES PSID BEFORE AND AFTER INCOME TAX DATA (SMALL SAMPLE)

		1983	**1984**	**1985**	**1986**	**1987**	**1988**
Gini	D	.0118	.0137	.0131	.0138	.0034	.0058
	S.E.	.0104	.0115	.0130	.0136	.0106	.0057
	p	.866	.880	.854	.870	.662	.848
Theil	D	.0117	.0066	.0056	.0049	.0008	.0023
	S.E.	.0080	.0061	.0077	.0080	.0085	.0032
	p	.904	.854	.774	.758	.474	.760

TABLE 6: BOOTSTRAP STANDARD ERRORS AND CONFIDENCE INTERVALS NLSY INCOME DATA (LARGE SAMPLE)

	1984	1985	1986	1987	1988	1989
Gini	.3975	.3901	.3754	.3699	.3539	.3513
S.E.	.0037	.0038	.0038	.0038	.0036	.0038
A.S.E.	.0036	.0038	.0038	.0037	.0036	.0035
99% U.B.	.4074	.3996	.3863	.3806	.3632	.3605
99% L.B.	.3879	.3795	.3645	.3600	.3442	.3411
95% U.B.	.4052	.3975	.3827	.3779	.3606	.3583
95% L.B.	.3893	.3822	.3677	.3625	.3464	.3441
Theil	.2569	.2504	.2322	.2245	.2050	.2015
S.E.	.0047	.0051	.0048	.0045	.0042	.0041
A.S.E.	.0046	.0051	.0048	.0045	.0041	.0040
99% U.B.	.2700	.2630	.2444	.2373	.2155	.2120
99% L.B.	.2446	.2360	.2205	.2126	.1942	.1898
95% U.B.	.2668	.2600	.2413	.2342	.2129	.2101
95% L.B.	.2466	.2402	.2228	.2156	.1964	.1934

Sample size = 4266

TABLE 7: HYPOTHESIS TESTS FOR GINI AND THEIL MEASURES
NLSY INCOME DATA (LARGE SAMPLE)

			1985	1986	1987	1988	1989
1984	Gini	D	.0074	.0221	.0276	.0436	.0461
		S.E.	.0037	.0039	.0042	.0043	.0042
		p	.020	.0	.0	.0	.0
	Theil	D	.0065	.0247	.0324	.0519	.0553
		S.E.	.0049	.0049	.0053	.0053	.0052
		p	.090	.0	.0	.0	.0
1985	Gini	D		.0147	.0202	.0362	.0387
		S.E.		.0037	.0037	.0043	.0042
		p		.0	.0	.0	.0
	Theil	D		.0182	.0259	.0454	.0489
		S.E.		.0049	.0049	.0056	.0053
		p		.0	.0	.0	.0
1986	Gini	D			.0054	.0215	.0240
		S.E.			.0036	.0039	.0041
		p			.072	.0	.0
	Theil	D			.0077	.0271	.0306
		S.E.			.0047	.0048	.0050
		p			.054	.0	.0
1987	Gini	D				.0160	.0185
		S.E.				.0033	.0037
		p				.0	.0
	Theil	D				.0194	.0229
		S.E.				.0040	.0044
		p				.0	.0
1988	Gini	D					.0035
		S.E.					.0030
		p					.224
	Theil	D					.0034
		S.E.					.0035
		p					.172

TABLE 8: BOOTSTRAP STANDARD ERRORS FOR AGE GROUP DECOMPOSITIONS NLSY INCOME DATA (LARGE SAMPLE)

	1984	1985	1986	1987	1988	1989
Theil	.2569	.2504	.2322	.2245	.2050	.2015
S.E.	.0047	.0051	.0048	.0045	.0042	.0041
A.S.E.	.0046	.0051	.0048	.0045	.0041	.0040
BTheil	.0004	.0012	.0020	.0034	.0034	.0027
S.E.	.0003	.0005	.0007	.0009	.0009	.0008
A.S.E.	.0003	.0003	.0003	.0003	.0003	.0003
WTheil	.2566	.2493	.2302	.2211	.2016	.1988
S.E.	.0047	.0051	.0049	.0045	.0041	.0041
A.S.E.	.0046	.0050	.0048	.0045	.0041	.0039

Sample size = 4266

BTheil = between group Theil measure
WTheil = within group Theil measure

TABLE 9: HYPOTHESIS TESTS FOR AGE GROUP DECOMPOSITIONS

			1985	1986	1987	1988	1989
1984	Theil	D	.0065	.0248	.0324	.0519	.0554
		S.E.	.0048	.0052	.0054	.0054	.0053
		p	.066	.0	.0	.0	.0
	BTheil	D	-.0008	-.0016	-.0030	-.0031	-.0024
		S.E.	.0004	.0006	.0009	.0009	.0008
		p	.990	.998	1.000	1.000	1.000
	WTheil	D	.0073	.0264	.0355	.0550	.0578
		S.E.	.0048	.0052	.0054	.0054	.0053
		p	.046	.0	.0	.0	.0
1985	Theil	D		.0183	.0260	.0454	.0489
		S.E.		.0048	.0055	.0052	.0054
		p		.0	.0	.0	.0
	BTheil	D		-.0008	-.0022	-.0022	-.0015
		S.E.		.0005	.0008	.0008	.0007
		p		.9481	.0	.998	.988
	WTheil	D		.0191	.0282	.0477	.0504
		S.E.		.0048	.0054	.0052	.0053
		p		.0	.0	.0	.0
1986	Theil	D			.0077	.0272	.0306
		S.E.			.0049	.0049	.0052
		p			.058	.0	.0
	BTheil	D			-.0014	-.0015	-.0008
		S.E.			.0008	.0008	.0007
		p			.968	.978	.852
	WTheil	D			.0091	.0286	.0314
		S.E.			.0049	.0049	.0053
		p			.028	.0	.0
1987	Theil	D				.0195	.0229
		S.E.				.0043	.0043
		p				.0	.0
	BTheil	D				-.0000	.0007
		S.E.				.0008	.0008
		p				.542	.210
	WTheil	D				.0195	.0223
		S.E.				.0043	.0043
		p				.0	.0
1988	Theil	D					.0035
		S.E.					.0036
		p					.158
	BTheil	D					.0007
		S.E.					.0006
		p					.126
	WTheil	D					.0028
		S.E.					.0036
		p					.214