Luxembourg Income Study Working Paper Series

Working Paper No. 378

Boostrapping the LIS: Statistical Inference and Patterns of Inequality in the Global North

Timothy Moran

February 2005

Luxembourg Income Study (LIS), asbl

Bootstrapping the LIS: Statistical Inference and Patterns of Inequality in the Global North*

Timothy Patrick Moran
State University of New York – Stony Brook, USA

Revised February, 2005

Abstract: The problem of statistical inference has long been associated with quantitative inequality research. Within the last five years, however, significant developments have occurred in both the theory and practice of conducting formal statistical inference with common measures of inequality such as the Gini index. These new techniques involve the use of Monte Carlo, bootstrap resampling plans that seek to recover the standard error and sampling distribution of inequality estimates directly through the empirical distribution of the sample data, thereby facilitating statistical inference via confidence interval estimation and hypothesis testing. Using the income survey of the Luxembourg Income Study (LIS) project, this paper provides an analytical evaluation of the bootstrap procedure in the context of comparative inequality research, and uncovers patterns of distributional change in the global North over the last two decades. While it is now generally accepted that inequality has increased in the United States and United Kingdom during this period, the extent to which other wealthy nations have been able to avoid this trend has generated some debate. The paper presents new evidence to address this discussion, demonstrating along the way how the ability to conduct formal statistical inference with the Gini index provides an effective and important new evaluative tool. The paper provides an informative analysis of current methodological developments in inequality research, and demonstrates how they may be applied in the specific context of the LIS, but also can be used as a practical guide for handling the problems of statistical inference in more general social scientific settings.

---

* This research was supported in part by a generous fellowship from the Luxembourg Income Study Project. I am indebted to David Consiglio for his advice and guidance, from the inception of the project onward. Paul Alkemade, Georges Heinrich, and Martin Biewen provided helpful suggestions at various stages. I gratefully acknowledge and thank the talented staff of the LIS, and especially Thierry Kruten whose expertise and energy significantly improved the quality of this project. Please direct correspondence to the author at Department of Sociology, SUNY - Stony Brook, Stony Brook, NY, USA 11794-4356; timothy.p.moran@stonybrook.edu.

Bootstrapping the LIS: Statistical Inference and Patterns of Inequality in the Global North

Introduction

In the early 1980s researchers in the United States and the United Kingdom began to notice that, after a long period of relative stability, the distribution of income was becoming noticeably more unequal. In 1988 the phenomenon was coined the "great U-turn" by Harrison and Bluestone in recognition that there was something historically unique about rising levels of inequality in wealthy, developed nations. While no single "smoking gun" has been found to explain this trend (Gustafsson and Johansson 1999), debates tend to implicate various forms of economic restructuring (shifting patterns of trade, increased capital and labor mobility, deindustrialization, and the rise of the skill-based economy) that fall under the "globalization" rubric. Since this restructuring characterizes the global North more broadly, academic and popular interpretations have tended to discuss rising inequality as a nearly universal outcome of these processes in the 1980s and 1990s (Friedman 2000; Smeeding 2002). As summarized by Ram (1997:577), "[t]he somewhat cheerless distributional position recently noted for the U.S. seems to characterize most of the postwar developed world."

Others, however, have begun to question the extent to which trends in the United States and United Kingdom were replicated throughout the global North. Observers of these contrasting outcomes argue that technological change has been less skill-based in parts of Europe than in the U.S. and U.K., and that returns to education and skill increased less sharply in these areas (because the supply of skilled workers increased faster), leading to "less of an increase, or even no change" in wage inequality in these countries (Acemoglu 2002:1). Another line of interpretation has emerged around the general idea that politics and political institutions matter. Some argue, for example, that European labor policies and wage-setting institutions mitigate the tendency toward increasing earnings inequality (Acemoglu 2002; Blau and Kahn 1996; Freeman and Katz 1995; Nickell and Bell 1996). In particular, many studies find that labor union density significantly reduces inequality (Alderson and Nielsen 2002; Freeman 1993; Gustafsson and Johansson 1999), and that policy liberalism/leftist government strongly drives the redistribution process (Bradley, et al. 2003; Brady 2003; Kelly 2004).

Using the surveys of the Luxembourg Income Study database (LIS), this analysis brings new methodological considerations to bear on this debate. More specifically, the paper represents the first systematic evaluation of the LIS database using bootstrap inference techniques to interpret levels and trends in inequality across the member countries. The paper is divided into two sections. The first section examines the theoretical and practical aspects of

conducting statistical inference via the new bootstrap procedures, both in the context of inequality analysis generally, and with regard to the specific needs and constraints of the LIS database more specifically. The goal of this section is to provide a methodological guide to regularize application of bootstrap techniques in future inequality and poverty research. The second section presents and discusses the substantive findings of the project, demonstrating how the ability to conduct statistical significance provides an effective and important new tool for evaluating cross-national patterns of income inequality.

I. Bootstrap Resampling Methods for Statistical Inference

The absence of statistical significance has been a glaring shortcoming in comparative research more broadly and in inequality evaluations in particular.[1] Practitioners seeking to quantify inequality have traditionally relied on descriptive devices such as the Lorenz curve, quintile income shares, or various summary indices like the popular Gini index. In the absence of well established and practically useful analytical error theories for these measures (theories that exist for statistics like the mean for example), practitioners are left making evaluations – often between relatively minor differences in measurements – informally based on little more that visual or absolute-numerical comparisons. As Rossi, et al. (2001:905) explain, "[p]oint estimates unaccompanied by any precision measure are the rule rather than the exception, so that tracking the changes of personal distribution through time rests to a large extent on scholars' *a priori* beliefs." As a result, error or random variation in the samples can produce discrepant portraits of distributional change within countries even as scholars build models attempting to explain such changes cross-nationally. Moreover, minor changes in Gini coefficients might not warrant directional conclusions at all, perhaps signaling instead that "no change" in inequality has occurred (a description rarely used by scholars looking for trends). In short, the conclusions we tend to draw in empirical inequality research are "far more tentative than has often been realized" (Mills and Zandvakili 1997:140).

Within the last few years, however, significant developments have occurred in both the theory and practice of conducting formal statistical inference using methods that seek to nonparametrically estimate (at least an approximation of) the sampling distribution of inequality measures such as the Gini index. Pioneered by Efron (1979), these methods employ Monte Carlo, bootstrap resampling plans to evaluate a parameter estimate at reweighted versions of the

empirical probability distribution found in the sample data. The bootstrap procedure seeks to recover the standard error and sampling distribution numerically (as opposed to theoretically) via repeated random samples drawn with replacement from the observed sample distribution, thus allowing the construction of confidence intervals and hypothesis tests on inequality and poverty indicators common in quantitative cross-national research (for general discussions of the bootstrap, see Chernick 1999; Efron 1982; Efron and Tibshirani 1993).

a. The Bootstrap Concept

The bootstrap technique is relatively straightforward, yet analytically powerful. While the mathematical justifications can be quite sophisticated (for more rigorous treatments see Hall 1992, and Shao and Tu 1995), the bootstrap method requires no theoretical calculations, applies the same to any inequality measure currently in use, and is available no matter how mathematically complicated the parameter estimate or its asymptotic standard error may be.

The bootstrap method is conceptually based on the "plug-in principle," where known sample values are taken as simple estimates of the entire population. Say an empirical distribution $\hat{F} \sim \{x_1, x_2, \ldots, x_n\}$ is a known, random *i.i.d* (independent and identically distributed) sample taken to estimate the entire population distribution $F = \{X_1, X_2, \ldots, X_N\}$. Similarly, let $\hat{\theta}$ represent the pont estimate of an unknown population parameter of interest θ, computed from $\hat{F}$. The general idea behind the plug-in principle is to recover the variance of $\hat{\theta}$ through the known distribution $\hat{F}$. This is obviously a straightforward affair when dealing with the mean, yet for most other statistical estimators it is difficult to evaluate the variance of $\hat{\theta}$ theoretically though $\hat{F}$ (see Shao and Tu (1995:10) for an elaboration of this point). The bootstrap solution is to employ an old technique known as the Monte Carlo method to establish a numerical approximation of the variance through repeated simulations – in other words evaluate variance of $\hat{\theta}$ empirically through $\hat{F}$.[2] The nonparametric, one-sample approximation works as follows.[3]

First, randomly draw an independent data set $\left\{x_1^*, x_2^*, \ldots, x_n^*\right\} \overset{iid}{\sim} \hat{F}$ of size *n* with

replacement from the sample distribution, so that each x_i^* represents a bootstrap version of x_i and has probability n^{-1} of being selected. Replacement here refers to each draw not each sample, thus some observations can be counted more than once, some not all in any given resample. Next, evaluate $\hat{\theta}_b^*$, a bootstrap replication of $\hat{\theta}$ computed from the b^{th} bootstrap resample. Independently repeat this procedure a large number of *B* times, each time obtaining bootstrap replications $\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_B^*$. The bootstrap standard error of $\hat{\theta}$ can then be estimated as:

$$\hat{se}_{boot} = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} \left[\hat{\theta}_b^* - \bar{\hat{\theta}}^* \right]^2 \right\}^{1/2} \qquad (1.0)$$

where $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b^*$ is the mean of the parameter estimates obtained in the *B* resamples.

Based on the law of large numbers, as the number of *B* replications approaches infinity the bootstrap standard error (1.0) converges to the estimate of the standard error found in the sample $\hat{F}$.[4]

Several alternative methods have been suggested for assigning confidence intervals to the unknown parameter based on the bootstrap distribution. Hall (1988:927) identifies "an almost bewildering array" of techniques in use today. While it would be impossible to provide a thorough treatment of all aspects of the various bootstrap confidence intervals that have been suggested (see Davison and Hinkley 1997; Hall 1992; Shao and Tu 1995; Trede 2002), only two types of intervals have been used in the specific context of inequality data – standard normal and percentile intervals. Each method has both a primary form and a more technical form that introduced refinements designed to improve the coverage accuracy of the interval.

i. The Standard Normal Approach

Also called the "boot-*t*" or "normal approximation," the standard normal approach for obtaining confidence intervals is interpreted differently throughout the literature but is essentially the same procedure as constructing confidence intervals for sample means. In its primary form, what Mills and Zandvakili (1997:143) call the "naive boot-*t*," the variance of the parameter estimate is evaluated as in (1.0) and the standard student's *t* distribution is used to obtain 1 - α confidence intervals in the traditional manner.

$$\hat{\theta}_{LL} = \hat{\theta} - t_{\alpha/2,n-1}\hat{se}_{boot} \qquad \hat{\theta}_{UL} = \hat{\theta} + t_{\alpha/2,n-1}\hat{se}_{boot} \tag{2.0}$$

with $t_{\alpha/2,n-1}$ denoting the $\alpha/2^{th}$ percentile of the *t* distribution on *n*-1degrees of freedom. This method requires the least amount of computational effort, but also necessitates the parametric assumption that distribution of the parameter estimate is normal. This may be a reasonable expectation as *n* gets large, but the *t* distribution may only be an approximation in small sample situations. Also, the use of the *t*-distribution does not adjust the confidence interval to account for possible skewness in the underlying population or other errors that can result when $\hat{\theta}$ is not the sample mean.

A modified version of the standard normal method seeks to account for these errors by obtaining accurate confidence intervals without having to make normality assumptions. Here the bootstrap is used to estimate the distribution of a "studentized pivot" directly from the data, in essence building a unique *t*-table for the observed data. Following the notation introduced above, let *T* be a studentized statistic whose distribution in nonparametric situations is unknown. Let T^* represent a bootstrap estimate of *T* generated via *B* resamples.

$$T_b^* = \frac{\left(\hat{\theta}_b^* - \hat{\theta}\right)}{\hat{se}_b^*} \tag{2.1}$$

where $\hat{\theta}_b^*$ is the value of $\hat{\theta}$ for the b^{th} bootstrap resample, and $\hat{se}_b^*$ is the estimated standard error of $\hat{\theta}_b^*$ for the b^{th} bootstrap resample. An empirical probability distribution is thus developed with the α^{th} percentile of T_b^* estimated by the value $\hat{t}_\alpha^*$ such that

$$\alpha = \frac{\#\left\{T_b^* \le \hat{t}_\alpha^*\right\}}{B} \tag{2.2}$$

Thus a "bootstrap *t*-table" is constructed directly from the sample distribution, and tail probabilities can be obtained for confidence intervals as in (2.0). The *t*-values are found at the $B \times \alpha^{th}$ and $B \times (1-\alpha)^{th}$ locations of the ordered array of all T_b^* s .[5] For example, to construct a 90% confidence interval for $B = 1000$, simply locate the fifth percentile at the 50th largest value of T_b^* and the ninety-fifth percentile at the 950th largest value. More generally, the lower and upper limits are

$$\hat{\theta}_{LL} = \hat{\theta} - \hat{t}_{1-\alpha/2}^* \hat{se}_b^* \qquad \hat{\theta}_{UL} = \hat{\theta} + \hat{t}_{\alpha/2}^* \hat{se}_b^* \tag{2.3}$$

Efron and Tibshirani (1993) mathematically show that in large samples the coverage of bootstrap-*t* intervals (2.3) are closer than the coverage of intervals based on the *t* table itself, which make sense given that the *t* table applies to all samples and all sample sizes, whereas the bootstrap-*t* table applies only to the given sample. Also notice that unlike (2.0), confidence intervals (2.3) are not necessarily symmetric about $\hat{\theta}$ (because T_b^* is not symmetric about zero as student's *t* is). In (2.3), intervals may be larger below and above $\hat{\theta}$ leading to an important reason for the improvement in coverage of the confidence intervals.

There are two problems with the more elaborate bootstrap-*t* procedure, however, that all but eliminate its use in applied settings. First, this approach requires an estimate of the standard error $\hat{se}_b^*$ which is the standard deviation of $\hat{\theta}_b^*$ <u>for the bootstrap sample *b*</u>. This figure needs to be estimated either via asymptotic theory or a two-level, nested bootstrap procedure – generate B_1 bootstrap data sets to approximate the distribution, then for each given bootstrap data set, generate B_2 bootstrap data sets to approximate the standard error of the estimate. The former method carries all of the disadvantages of analytical theory cited above and might not be available for a given statistic, while the latter method is always available but implies a substantial increase in computational burden. For example, to estimate both the distribution of T_b^* and the

standard error $\hat{se}_b^*$ by Monte Carlo, we may want B_1 to be as high as 1000 for the former, whereas B_2 can be much lower for the latter, say 200. In this case, however, we would need a total number of bootstrap resamples of B_1B_2 or 200,000. Second, this version of the bootstrap-*t* is not a transformation respecting procedure, in other words it may require variance stabilization measures. Since this procedure depends on the scale of the statistic under consideration, it can return confidence interval limits that are beyond the range of the parameter itself (for example, a lower limit less than zero in the case of the Gini index).

ii. The Percentile Method

The percentile method is similar to the modified bootstrap-*t* in that it seeks to recover the sampling distribution of $\hat{\theta}$ via a bootstrap cumulative distribution function, but instead of taking a studentized pivot the percentile method employs $\hat{\theta}^*$ directly. Let (3.0) represent the cumulative sampling distribution of the parameter estimate $\hat{\theta}$.

$$H_F(x) = P\left\{\hat{\theta}(x_1, x_2, \ldots, x_n, F) \leq x\right\} \tag{3.0}$$

where $\{x_1, x_2, \ldots, x_n\} \overset{iid}{\sim} F$, the true population. Following the same recipe for obtaining bootstrap standard error estimates (1.0), we can easily obtain the bootstrap approximation of $H_F(x)$:

$$\hat{H}_{boot}(x) = \frac{1}{B}\sum_{b=1}^{B} I\left\{\hat{\theta}_b^*\left(x_{1b}^*, x_{2b}^*, \ldots, x_{nb}^*, \hat{F}\right) \leq x\right\} \tag{3.1}$$

where $\{x_{1b}^*, x_{2b}^*, \ldots, x_{nb}^*\} \overset{iid}{\sim} \hat{F}$, b = 1, ... , B are independent bootstrap resamples from the observed sample data. Since in practice we must take a finite number of B replications, I denotes the indicator function of the set which is the same way of saying that

$$\hat{H}_{boot}(x) = \frac{\#\left\{\hat{\theta}_b^* \leq x\right\}}{B} \tag{3.2}$$

The percentile method consists of establishing confidence interval limits directly from the bootstrap distribution (3.1). The method gets its name because the $\alpha/2^{th}$ and $1-\alpha/2^{th}$ empirical percentiles of the bootstrap distribution (3.1) are taken as two-sided α-level percentile confidence interval boundaries, in the same manner as (2.2). Thus no assumptions need to be made about the *a priori* form of the distribution.[6] Once the values of $\hat{\theta}_b^*$ are ordered ascendingly, $\hat{\theta}_{LL}$ becomes the $B \times \alpha^{th}$ value and $\hat{\theta}_{UL}$ is the $B \times (1-\alpha)^{th}$ value of the distribution.

Since (3.2) uses the empirical distribution in place of knowledge of the underlying population distribution, in theory it would tend to underestimate the tails of the distribution of $\hat{\theta}_b^*$ (Efron and Tibshirani 1993). Subsequent improvements have been suggested under more advanced "bias corrected" or "bias reduction" percentile bootstraps, but their adoption has not been widespread in applied situations. Bias-corrected intervals are also determined by percentiles of the bootstrap distribution, but not necessarily the same ones. The percentiles used may be different based on a bias-correction value determined from the proportion of bootstrap statistics $\hat{\theta}^*$ that ended up being less than the original estimate $\hat{\theta}$.[7] If exactly half of the bootstrap estimates are less than or equal to the original estimate, than the bias-correction value is zero and the percentile intervals are taken as above. The bias-correction yields more accurate confidence intervals the extent to which the median of $\hat{\theta}^*$ differs from $\hat{\theta}$. Table 1 summarizes the major advantages and disadvantages of these four confidence interval techniques.

Table 1 about here

iii. Bootstrap Hypothesis Testing

Bootstrap inference can be undertaken either through confidence interval estimation, or by computing a bootstrap *p*-value, *p**, the proportion of bootstrap samples that yield a test statistic more extreme than the actual test statistic computed from the data. Following similar procedures in 2.2 and 3.2 above, a one-tailed test that rejects above the test statistic can be approximated as:

$$p^* = \frac{1}{B}\sum_{b=1}^{B} I\left\{\hat{\theta}_b^* \geq \hat{\theta}\right\} \tag{4.0}$$

where *I* denotes the indicator function of the set.

In the two-sample context, when confidence intervals around two estimates (either between units or within them over time) do not overlap, we can unambiguously assess this difference to be statistically significant (at level α). In the presence of overlap, however, it does not immediately follow that the difference is not statistically significant. In this case, an hypothesis test is required that tests for the difference in parameter estimates.

Following the procedure outlined in Efron and Tibshirani (1993), say we have two *i.i.d.* samples $\boldsymbol{x}$ (size *n*) and $\boldsymbol{y}$ (size *m*) taken to represent populations X and Y, and we want to test the null hypothesis H_o: X = Y. The hypothesis test requires a test statistic $t(s)$ which could be a parameter estimate $\hat{\theta}$ but not necessarily. In the cross-national application presented in this paper, $t(s) = \text{Gini}_x - \text{Gini}_y$, the difference between the inequality estimates say in two countries, or in one country in two time periods. As in conventional hypothesis testing using the mean, we attempt to find a one-sided *p*-value defined as the probability of finding at least that large of a difference when the null hypothesis is true.

$$p - value = P_{H_o}\left\{t(s^*) \geq t(s)\right\} \tag{4.1}$$

where *t(s)* is fixed at its observed value and *t(s*)* is distributed under the assumption that the null hypothesis is true. Since this distribution F_o is unknown, the bootstrap method derives a "plug-in" distribution F_o* in the following manner.

Under the assumption that the two samples are randomly drawn and independent, first draw a bootstrap resample $\boldsymbol{s}_x$* of size *n* with replacement from $\boldsymbol{x}$ so that the probability for selection in $\boldsymbol{s}_x$* is 1/*n*, and calculate the statistic of interest $\mathbf{G}_x$*. Then draw a bootstrap resample $\mathbf{s}_y$* of size *m* in the same manner and again calculate the statistic of interest $\mathbf{G}_y$*. Repeat the procedure a *B* number of times, each time evaluating the difference between the Gini coefficients

$$t(s)^* = \mathbf{G}_x^* - \mathbf{G}_y^* \tag{4.2}$$

Construct a bootstrap confidence interval of $t(s)$*, and the difference between $\mathbf{G}_x$* and $\mathbf{G}_y$* is then said to be statistically significant (at α) if zero is not contained in the interval (*p** values can then be approximated by 4.0 above).

b. The Numerical Performance of the Bootstrap

Soon after Efron's pioneering work, Singh (1981) and Bickel and Freedman (1981) mathematically demonstrated the asymptotic validity of the bootstrap in a large number of situations. Recent work suggests that the procedure is not only valid, but that bootstrap tests will outperform tests based on asymptotic theory in finite samples in that they will commit errors that are of lower order in the sample size n. Performance in this regard can be measured by evaluating the "size distortion," or what Davidson and MacKinnon (1999) term the "error in rejection probability" (ERP), of an inference test – the difference between the nominal level of a test and its actual rejection probability. The current consensus is that for a sample size n, the ERP committed by tests based on the bootstrap is generally reduced by a factor of $n^{-1/2}$ in one-tailed tests and n^{-1} or more in two-tailed tests. In theory, then, the bootstrap provides a higher-order, asymptotic approximation in that the difference between the test's actual and nominal levels decrease more rapidly with increasing sample size than it would if the distribution of a test statistic is obtained via asymptotic theory (Davidson and MacKinnon 1999; Hall 1992; Horowitz 1994, 1997). In addition, several studies now provide evidence from Monte Carlo simulation experiments that the numerical performance of bootstrap tests improve asymptotic approximations (Davison and MacKinnon 1999a; Horowitz 1994, 1997; Shao and Tu 1995). The bootstrap, it is now generally regarded, almost always minimizes the ERP in fixed sample sizes that occur when critical values are obtained via asymptotic theory; and in cases when asymptotic and bootstrap critical values are very different, it is almost certain that the asymptotic values are inaccurate (Davidson and MacKinnon 2000).[8]

A few simulation experiments have been performed in the context of income distribution data, and again all conclude that statistical inference based on bootstrap procedures outperform, often quite dramatically, asymptotic theory (Biewen 2002; Cowell and Flachaire 2002; Mills and Zandvakili 1997; Trede 2002). Given the newness of these methods in inequality applications, the precise accuracy of the bootstrap for the various indicators (or the extent of the improvement over asymptotic theory) is still an empirical matter. Early research in fact has illustrated an apparent contradiction between the theoretical expectations of using the bootstrap with inequality indicators and its numerical performance. For example, Beran (1988) shows that bootstrap inference is unproblematic when the indicator is asymptotically pivotal (i.e., its distribution does not depend on any unknown parameters) which is the case for every inequality measure except the Gini index. Yet Cowell and Flachaire (2002:15) conducted a number of simulation

experiments in what to date is the only direct comparison of the different indicators, and concluded that among the most popular inequality measures, "the bootstrap does well only for the Gini." In addition, all studies have found that confidence intervals around various inequality indices are too narrow (Biewen 2002; Cowell and Flachaire 2002; Davidson and Flachaire 2004).

In an effort to gauge the coverage accuracy of the confidence intervals established in this study, a Monte Carlo simulation experiment was performed on the Finland 2000 survey of the LIS. Following a similar procedure in Mooney (1996), the Gini index for this survey is taken *a priori* to be the "true" population parameter. Drawing $B = 1000$ replications of a five percent subsample (521 households), a bootstrap "sample" Gini index and 95 percent confidence intervals are established. After 600 simulations, the accuracy of the intervals in capturing the "true" Gini index is evaluated via the ERP. Table 2 presents the results of this experiment, indicating that all three methods are very close, and that all three intervals are slightly too narrow, implying that inference tests based on them over-reject the null hypothesis. Said differently, the tests would result in more Type I errors than are nominally allowed. The bias-corrected method (.055) came the closest to yielding the exact nominal alpha level, followed by the normal approximation (.057), and the percentile (.063). Like other studies, then, the finding is that the intervals around the Gini index are close yet too narrow.[9]

Table 2 about here

In sum, the state of knowledge on using the bootstrap with inequality measures indicates that: a) The standard bootstrap is valid with all common indices, and preliminary tests suggest that the Gini index numerically outperforms the others; b) ERP distortions are reduced for all measures when using the bootstrap, thus it now generally regarded as practical method for improving upon distributional approximations based on asymptotic theory; and c) Bootstrap critical values, although better, are still approximations, and the future direction of research in econometrics will explore ways to adjust the bootstrap to further reduce the ERP in income distribution data (see early attempts by Davidson and Flachaire 2004; Xu 2000).

c. Bootstrap Inference With the LIS Database

In the analysis that follows, bootstrap inference techniques are systematically applied to the income surveys of the Luxembourg Income Study (LIS) database. The bootstrap procedures in this study employ resampling plans in the case of an independent and identically distributed (*iid*) sample of fixed size *n* from an unknown population.[10] Inequality is measured by the Gini index as specified according to the methodological recommendations of the LIS project as more fully outlined in Appendix A. The unit of analysis is the household, adjusted for size using an equivalence scale, and income is measured as net, disposable household income subjected to both top and bottom coding. Also based on LIS specifications, the bootstrap resampling procedures are weighted by "person weights" – the product of the household weight (the weights established by the procedure used in the original dataset, not by LIS) and the number of persons in the household.[11]

i. Standard Errors in the LIS Database

In the first part of the analysis, ten recent LIS income surveys were selected on the basis of obtaining a group with varying sample sizes – from one of the smallest surveys (Luxembourg with 1,813 households) to one of the largest (United States with 49,351 households) – to conduct a more detailed evaluation of the various bootstrap methods. Table 3 reports the bias and standard error of the Gini index, and the 95 percent confidence intervals under the standard normal, percentile, and bias-corrected procedures for the ten surveys. To assess the impact of the number of bootstrap replications *B*, the intervals are calculated at 250, 500, and 1000 replications.

Table 3 about here

As seen in Table 3, the number of replications has little impact either on estimating the standard error of the Gini index or in establishing the confidence intervals (confidence interval boundaries established with 500 replications are never more than .001 off the intervals established with 1000 replications, which always translates into a less than 0.5 percent difference).[12] In contrast, however, larger sample sizes did lead to smaller standard errors as would be expected. Looking not only at Table 3 but at all 120 income surveys in the LIS

database, we see that generally: sample sizes more than 10,000 lead to confidence intervals no bigger than one percentage point in the Gini index; samples between 10,000 and 5,000 lead to confidence intervals no bigger than two percentage points; and samples under 5,000 may be three percentage points wide. No confidence interval in the LIS database is wider than three percentage points in the Gini index.

While these findings provide an interesting guide for practitioners, there are exceptions and variations to this rule for samples of various sizes. For example, the width of the confidence intervals for Luxembourg and the Slovak Republic in Table 3 are roughly the same despite big differences in sample size. These variations in sampling error are caused by the differential ability of the bootstrap to consistently capture from sample to sample the full range of incomes, creating greater variation across the resamples. In part this reflects two interrelated characteristics of the underlying survey, features separate from the degree of inequality itself: 1) The shape of the population structure with regard to incomes, specifically the extent of income "clustering" and the importance of the upper tail of the distribution; and/or 2) The ability of the survey instrument to proportionately sample from the entire distribution. Thus we should expect higher standard errors in situations where income are less fluid, or when the when the upper tail is long (relatively higher incomes accrued to relatively fewer households), or if the survey instrument "creates" these characteristics through inadequate sampling. In these situations – in economies less fully monetized, perhaps, or where inequality is very high, or when administrative and other resource issues limit the survey's effectiveness – greater variation will exist from survey to survey in the population, and thus from resample to resample in the bootstrap.

Surveys based on tax records or other national accounts, for example, will proportionately capture the full range of incomes better than those based on household surveys, where the sample space is typically more narrow and sampling the two tails of the distribution (richest and poorest) in particular is more difficult. Szekely and Hilgert (1999) show how income surveys in Latin America systematically can not capture incomes from the richest sectors of society, leading to an upper tail that is consistently under-represented. As a result, they find bootstrap standard errors much greater than those reported here – standard errors for Colombia were higher than any in the LIS despite sample sizes over 100,000. Within the LIS itself, the surveys for Mexico, with sample sizes 10,000 or more, have standard errors as big as surveys in Western Europe one-third that size.

In this sense the importance of sample specifications such as population coverage, length of income period, and top coding in accounting for sampling errors are not entirely independent of the impact of non-sampling, or unsystematic, measurement errors on inequality estimates, although practitioners have long recognized the greater potential for the latter. The ability of a survey to proportionateley sample the full range of incomes, however, is not the same as errors resulting from misreporting, imprecision, non-response, and other measurement errors prone to income data. Thus, inadequate representation from the tails of the distribution will not only increase the sampling error associated with inequality estimates in places like Latin America, the consistent under-reporting of non-labor (especially self-employment and investment) income within the richer segments of society also leads to a substantial underestimation of measured inequality in that region.

ii. Comparing Confidence Interval Methods

While Table 3 shows that sampling error, while highly correlated with sample size, can still vary among surveys, it also shows that the method used to construct confidence intervals seems to have very little impact on interval locations. Users of the LIS database can currently access bootstrap standard errors in the on-line user files, thereby facilitating the construction of standard normal confidence intervals. Comparing these intervals to the two other methods most easily calculable within the constraints of the LIS project (the percentile and the bias-corrected), indicates that little difference exists in establishing interval boundaries. There is never more than a .001 difference across the methods, suggesting that the three methods are very close, and that this conformity is consistent as one moves from relatively small samples to relatively large ones. Figure 1 illustrates the close agreement across the three types of intervals.

Figure 1 about here

The close conformity of the confidence interval methods is consistent with previous performance comparisons in the literature. For example, Mills and Zandvakili (1997) calculate bootstrap confidence intervals using both the percentile and standard normal methods for several inequality indices, as well as for the components of a decomposition of the Theil coefficient by subgroup. They find the standard normal intervals compared closely to the percentile intervals in

each case. Biewen (2002) compares the performance of different variants of the bootstrap using Monte Carlo results on 1996 German wage and income data, finding all intervals to be very close (generally differing by only a fraction of a percentage point). For large samples, there was not a practical difference between the methods. Similarly, Van Kerm (2002) employs both hypothetical distributions and a panel study of households in Belgium and, in the context of complex sampling schemes, finds the same small, one percentage point variation in coverage between the various confidence intervals. In Trede's (2002:281) review of the various confidence interval procedures, he concludes that, while performance varied under different conditions, "[t]here is not much evidence on which bootstrap method performs best for inequality measures."

Since comparisons usually result in little functional difference across confidence interval estimates (especially as sample size gets large), the decision regarding which interval to report can be based on more practical considerations. At least four arguments favor the percentile method, and this is the method most often used in the applied literature. Referring back to Table 1, first, the standard normal method involves implicit assumptions (e.g., shape and symmetry of the distribution, an unbounded statistic) not necessary when using the bootstrap distribution directly, and the bootstrap-*t* requires an empirically burdensome estimation of the variance of the bootstrap statistic (either via asymptotic theory or nested bootstrapping).[13] Second, the percentile method is transformation preserving, meaning that "confidence intervals for a transformation of the statistic are identical to the transformed confidence intervals" (Trede 2002:268). Third, it provides asymmetrical and range preserving intervals that will never extend beyond the permissible range of a bounded statistic. Fourth, it is computationally simple to execute, more so than the bias-corrected method, and thus quite accessible in practical applications.

II. Patterns of Inequality in the LIS

As discussed at the outset, while it is now generally accepted that inequality has substantially increased in the United States and United Kingdom during the 1908s and 1990s, whether or not other wealthy nations have been able to avoid this trend has generated some debate. Recent debate has also focused around the extent to which different institutional and political contexts shape distributional outcomes. This section presents new evidence to address these discussions,

demonstrating along the way how the ability to conduct formal statistical inference with the Gini index provides an effective and important new evaluative tool for empirical inequality research.

a. Cross-National Levels of Inequality

In Table 4, the Gini index, bootstrap standard error, and 95 percent confidence intervals are reported for LIS member countries for circa 1990 and circa 2000. The countries are grouped as global North (or the developed wealthy countries of the world), and Eastern Europe (including Russia), and are sorted within groups alphabetically. Cross-national comparisons of inequality levels have traditionally involved linear "rankings" from low inequality to high based on the absolute value of a summary estimate such as the Gini index, thus in circa 2000 comparing Western Europe and other global North nations, and Eastern Europe and other former communist countries in the LIS database the following picture would emerge:

Country	Gini Index
Finland	.247
Slovenia	.249
Belgium	.250
Norway	.251
Sweden	.252
Netherlands	.256
Luxembourg	.260
Germany	.264
Austria	.266
Romania	.277
Poland	.293
Hungary	.295
Canada	.302
Ireland	.323
Italy	.333
United Kingdom	.345
Estonia	.361
United States	.368

In the absence of formal means to assess the statistical significance of these measured differences, our only recourse at this point is to conclude that a countries with a Gini index greater than others have more unequal distributions of income and that these differences are theoretically worthy of explanation.[14]

Table 4 about here

But would our impression of this ranking change in light of statistical inference? Figure 2 plots the 95 percent confidence intervals (using the percentile method) for these Gini estimates (the line in the middle of the interval marks the location of the Gini index within the interval). While rank ordering coefficients as above gives the impression that comparing inequality levels across countries is a straightforward affair, looking at the confidence intervals in Figure 2 allows a more careful interpretation of cross-national differences. As opposed to a linear ranking from low to high, we see groups of countries that show significant levels of inequality between them but not necessarily within them. These differences can be more formally evaluated via bootstrap null hypothesis testing as discussed above, and with access to the microdata, such procedures are readily available in the LIS database. Table 5 presents the results of the bootstrap hypothesis tests conducted on each pair of countries in Figure 2 where confidence intervals overlap.

Figure 2 and Table 5 about here

The exercise reported in Table 5 provides interesting insights into interpreting differences in the Gini index in the cross-national context. For each hypothesis test, three comparisons are provided – the absolute difference in the two Gini indices, the percentage difference, and the one-sided p^*-value. Across the 40 tests, all percent differences bigger than 4.5 percent are statistically significant at $\alpha = .05$. Conversely, all percent differences smaller than 2.5 percent are statistically insignificant, suggesting that countries with Gini indices this close or closer should not be thought off as having different levels of inequality. The smallest statistically significant difference is 3.0 percent between Poland (Gini index = .293) and Canada (.302), due in part to both country's relatively narrow confidence intervals. Illustrating the potential analytical power of these procedures, notice that the same 3.0 percent difference between Ireland (.323) and Italy (.333) is statistically insignificant given the large amount of sampling error associated with the Irish survey. The largest insignificant difference is 4.4 percent between Slovenia (.249) and Luxembourg (.260). Differences between 2.5 and 4.5 percent represent a "grey area" where the size of the standard error and the subsequent width of the two intervals matters – sometimes leading to statistical significance and sometimes not. Obviously this exercise does not provide a substantive interpretation of what these differences mean in terms of distributional variation, but it does give us a basis by which we can evaluate the

extent to which differences in observed Gini coefficients previously considered meaningful are better attributed to error or random chance.

In Table 6, the results of the hypothesis tests are used to create a new, more nuanced ranking system in which each country now has one of three affiliations to the others. Countries in lower case above the box demonstrate significantly lower inequality, those in upper case below the box demonstrate significantly higher inequality, and those in the box demonstrate statistically the same amount of inequality. For example, looking at Luxembourg (LUX), we see that Finland shows significantly lower levels of inequality. Slovenia, Belgium, Norway, Germany, Sweden, the Netherlands, and Austria all have inequality levels that can not be statistically distinguished from sampling error or random chance. The remaining nine countries all show significantly higher levels of inequality than Luxembourg.[15]

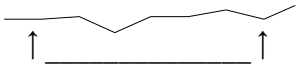
Table 6 about here

b. Distributional Change in the Global North: 1980 – 2000

The bootstrap procedures can also be employed to evaluate longitudinal change. In contexts where we can assume that the bootstrap replications are drawn from independent populations, the testing procedures outlined above are straightforward. In dependent data conditions, however – such as the LIS surveys employing panel designs – bootstrap resampling should be drawn simultaneously (as opposed to separately) so that the resample inherits the correlations of incomes between time t_1 and t_2 of the panel. This task turns out to be difficult in the LIS database where there is no clean way to track households over time. The overall closeness of the bootstrap methods, and the relatively minor effects of various bootstrap corrections found across the econometric literature, however, suggest that these refinements are unlikely to produce very different intervals than reported here (see, for example Trede 2002). Nonetheless the reader should be aware that Ireland, Luxembourg, Germany, and the Netherlands involve tests where both samples come from the same panel and that these tests are likely to be slightly biased in a conservative direction.

With attention to statistical inference, across the 17 nations in the LIS database that constitute the Global north, not one but three patterns of distributional change are evident over the period – a "Continental" pattern, an "Anglo" pattern, and a "Scandinavian" pattern (trends

for other LIS member countries are presented in Appendix B). The three patterns, and the specific trends in inequality found in the constituent countries, are now described in turn. Historical trends in inequality for individual nations are presented in a series of figures that chart the movement in the Gini index inclosed within the bootstrap 95 percent confidence interval. For various sub-periods within the 20-year period (marked by two vertical arrows), the magnitude of changes in the Gini index are evaluated using the following format:



.008 ← Absolute change in the Gini coefficient
2.8 ← Percentage change in the Gini coefficient
.016 ← One-sided *p**-value of the null hypothesis test (if the intervals overlap)

i. The Continental Pattern – Steady States

Looking at the bootstrap results in Figures 3a through 3c, the prevailing trend in Continental Europe is one of little distributional change – or change that at least is not large enough to reach statistical significance. The Continental pattern is characterized by relatively moderate levels of inequality which remain essentially unchanged throughout the period, although in a few cases inequality actually declined.

‣ Exemplars of the Continental Pattern: Seven countries form what can be called the exemplar Continental group – experiencing essentially no change in (or declining) levels of inequality over the period. **France** shows statistically no change in inequality since 1979 (and significantly less inequality than in 1984); **Germany** is just the opposite, showing significantly less (although substantively maybe the same) inequality than in 1973, with no change since 1984; the **Netherlands** shows no change since 1983; **Italy** and **Canada** illustrate the influence of beginning and end points in defining historical trends. Taking 1986 as the starting point for Italy, the trend shows significantly more inequality, but since 1987 Italy shows no change (and also no change since 1993). The last two data points for Canada (1998 and 2000) come from a different survey source than the rest, so the observed uptick in inequality may be attributed in part to survey inconsistencies. However, even using the 2000 figure, Canada shows significantly less inequality than in 1971, a span of some 30 years. Using the 1997 figure, Canada shows no

change since 1975. **Switzerland** and **Spain** have only two data points, but show no change over the 1980s.

▸ A Caveat Country: **Ireland** represents a "hybrid" pattern where overall levels of inequality are more on par with the Anglo pattern (i.e., relatively high), but shows no change in inequality levels since 1987, as is the prevailing trend in the Continental pattern.

Figures 3a - 3c about here

▸ A Caveat Group: Three countries form a caveat group within the Continental pattern. Each basically exhibit the overall pattern, but each also experience a significant increase in inequality over one survey in the 1990s. **Luxembourg** shows no change in inequality from 1985-1994, but an increase in inequality between the 1994 and 1997 surveys means inequality is significantly higher than in 1994. The same is true for neighboring **Belgium** which shows no change from 1985-1992, but a significant increase between the 1992 and 1997 surveys means inequality is significantly higher than in 1992. **Austria** shows a significant increase between the 1987 and 1994 surveys, but shows no change since 1994. Austria's trend may also be partially attributed to changes in survey sources, and in this case the two surveys have noticeably different sample sizes.

Figure 3d about here

ii. The Anglo Pattern – Divergence From the Continent

As discussed in the introduction, the Anglo pattern is now a well known and much analyzed phenomena. Often overlooked, however, is the fact that at the start of the 1980s, levels of inequality are in line with those found in Continental Europe. Beginning sometime in the late-1970s/early-1980s inequality then begins to increase, rapidly in some instances, leading to the continuous separation of these countries from inequality levels found on the continent.

▸ Exemplars of the Anglo Pattern: Three counties in the LIS database illustrate the pattern (see Figure 4). For the **United Kingdom**, inequality levels remain constant until the 1980s, and at the

start of the decade are lower than in France, Canada, Spain, and Switzerland. Inequality then begins to steadily increase beginning in 1979 and continues over the course of the period. Most evidence shows that inequality was also fairly stable in the **United States** prior to the 1980s, and at the start of the decade inequality in the U.S. is lower than Spain and Switzerland. Beginning in 1980, inequality begins to increase and also continues throughout the period (the down-turn seen in the last survey is not statistically significant). The LIS has less data for **Australia**, but here the pattern is also repeated – continuously rising inequality throughout the period.

Figure 4 about here

iii. The Scandinavian Pattern – Convergence to the Continent

The Scandinavian pattern is characterized by relatively low levels of inequality at the start of the period that consistently increase throughout the 1980s and 1990s as they converge to levels of inequality seen on the Continent. This increase, however, still leaves these countries with perhaps the lowest levels of inequality in the world.

▸ Exemplars of the Scandinavian Pattern: Three Scandinavian countries illustrate the pattern (see Figure 5). In **Norway**, **Sweden**, and **Finland**, Gini coefficients at the start of the 1980s are as low as .197, then increase by 18 percent or more throughout the period. Although in Sweden, the Gini index in 2000 is still not higher than in 1967.

Figure 5 about here

iv. Summary of Patterns in the Global North

As opposed to a universal tendency toward rising inequality, the prevailing pattern in the global North (prevailing in that more countries fit this pattern than the other two) is the one found in Continental Europe (and Canada). This pattern is characterized by moderate levels of inequality that are generally maintained throughout the period (with a few countries experiencing declining levels of inequality). The other two patterns are both characterized by rising levels of inequality, but should be described in relation to the Continental pattern – the context of the

Anglo Pattern of rising inequality is divergence from Continental levels of inequality, while the context of the Scandinavian pattern of rising inequality is convergence to these levels. The average Gini indices for the three patterns over time are plotted in Figure 6.

Figure 6 about here

As evidenced here, there are large differences in inequality trajectories for rich nations, and it seems that the tripartite typology of welfare-state regimes famously argued by Esping-Anderson (1990) still has descriptive power today. As Bradley et al. (2003) recently found, the different institutional configurations and political traditions underling Social-Democratic, Christian-Democratic, and Liberal welfare states play a decisive role in determining distributional outcomes. Given that these countries face a common set of socio-economic phenomena – continued de-industrialization, aging populations, immigration from the global South – and experience them from the same advantaged locations within the global hierarchy, the fact that distributional outcomes vary suggests that "globalization" has not usurped the importance of national policy or led to the insignificance of the state; world-economic processes do not compel any single national trajectory. This finding is consistent with recent literature, discussed at the outset, that emphasizes the importance of national political processes in counteracting market driven inequality. Strong leftist government (Bradley, et al. 2003; Brady 2003; Kelly 2004), high levels of democratic participation (Mueller and Stratman 2003), and low public tolerance for inequality (Lambert, Millimet, and Slottje 2003) are all associated with more equal income distributions. As summarized by Smeeding (2002:28), "[t]he overall distribution of income in a country depends on the domestic political, institutional, and economic choices made by those individual countries."

Conclusion

This project applied bootstrap resampling techniques to the income surveys of the LIS database in an effort to interpret patterns of distributional change in the global North over the last 20 years. The bootstrap allows us to perform conventional statistical inference and compare alternative distributions using measures that were previously only descriptive devices. These techniques provide analytically powerful methods to estimate the statistical accuracy of

parameter estimates regardless of whether an analytical formula for its standard error is known. The bootstrap does not require assumptions about the underlying form of the data, alleviates the need to work with complex mathematical derivations, and in other cases provides answers when analytical solutions either do not exist or are inappropriate.

The utility of this evaluative tool for comparative research was evident, illustrating in particular that for a sizable number of countries, representing a "Continental" pattern of distributional change, inequality levels at the end of the period were not statistically distinguishable from levels at the start. Where previously even small changes in the Gini index, such as those observed here, would lead scholars to make normative determinations of inequality trends, statistical inference provides new grounds to substantiate conclusions that "no change" in inequality has occurred.[16] It also suggests why others, evaluating inequality estimates in absolute terms, simultaneously argue that inequality has increased universally across the global North. These findings provide empirical support to recent arguments emphasizing the impact of political contexts and collective social forces on distributional outcomes, and in particular that such contexts have mitigated the tendency for rising inequality in Continental Europe.

In the end, the ability to conduct formal statistical inference improves the quality of our empirical descriptions, which in turn helps focus theoretical explanations of distributional variation across countries and over time. The descriptions presented here, for example, can be useful in reorienting future studies of inequality in the global North from specifying the general relationship between globalization and inequality, to uncovering the processes by which specific institutional configurations and power arrangements can a) modify the extent to which different sectors within a population are included/excluded from global-economic processes; and b) shape the distribution of the gains and losses that result from such economic change.

Appendix A
Estimating the Gini Index

Gini coefficients estimated in this study are based on the following methodological recommendations of the LIS.

Definition of Income – The unit of analysis is the household, and income is measured as net disposable household income:

gross income - direct taxes - employee social security contributions

Weights – Used to make adjustments in the relative influence of households in the survey to account for biases in the characteristics of the groups of non-respondents. The LIS does not carry out the weighting procedure, they are constructed by the survey sources themselves, so the exact impact of the weight variable varies throughout the database. The bootstrap procedure in this study resamples using person weights – the household weight multiplied by household size – standardized so that the weighted sample size equals the number of households instead of the population. The LIS standardizes person weights to handle deleted observations that nonetheless have non-zero weights (if one wishes to drop certain observations, for example, than the remaining weights will be adjusted accordingly). This procedure in no way effects the bootstrap results, which are identical with or without weight normalization.

Equivalence Scale – Used to adjust for economies of scale in households with different numbers of people (i.e., because a household's economic needs are related in part to its size). The equivalence scale used in this study is the square root of household size.

Top and Bottom Coding – Used to correct for possible measurement errors in the database. Incomes at the top of the distribution are capped at ten times non-equivalized median income. Incomes at the bottom are limited to one percent equivalized mean income. For comparable cross-national evaluations, bottom and top codes can not be set differently across the countries, the lowest top code and highest bottom code must be applied to all. These particular limits represent the lowest top code (Canada) and highest bottom code (Australia) of all the countries in the landmark LIS study (Atkinson, Rainwater, and Smeeding 1995), and have been applied to all LIS datasets since.

Appendix B

Trends in Inequality in Other LIS Member Countries

Figures 7 and 8 plot observed trends for the six LIS member countries not in the global North group (Slovenia, the Slovak Republic, and the Czech Republic are also LIS member countries but do not have data over time).

Figures 7 and 8 about here

Using the patterns identified in the global North, we see that **Hungary** illustrates the Continental Pattern of moderate and essentially constant inequality levels in the 1990s. **Poland** and **Israel** also show the Continental Pattern, but would belong to the caveat group where inequality increased in one survey. Israel shows constant inequality until 1992, then a significant increase between the 1992 and 1996 surveys. Poland's only increase comes between the 1992 and 1995 surveys. **Taiwan** clearly illustrates the Scandinavian Pattern of low and gradually rising inequality levels.

Russia and **Mexico** have by far the highest levels of inequality in the LIS database, and their trends are unique to the three patterns uncovered here.

Endnotes

1. Until recently, attempts at conducting formal statistical inference in the context of inequality data were limited to theories of asymptotic normality, where mathematical theory is used to analytically derive the limiting behaviors of parameter estimators (as the sample size tends to infinity). While in practice the sample size cannot increase indefinitely, the asymptotic normal approximation is taken as a guide for the "true" behavior of an estimate as sample sizes get reasonably large. Inference making via asymptotic theory has a long tradition in the statistical sciences, even as applied to inequality measurement, yet scholars have long sought alternatives primarily because asymptotic theory often yields poor approximations to the distribution of test statistics in the context of finite samples. Also, no generally accepted asymptotic procedure exists for conducting hypothesis tests for differences in inequality estimates between units or over time, a major shortcoming in applied settings.

2. Technically speaking, the bootstrap and Monte Carlo methods are separate techniques with different traditions. But today this is more or less a semantical distinction. Efron's pioneering contribution in fact was to recognize the benefits of marrying the two techniques, seeing that a "disparate, almost anecdotal" concept such as the bootstrap could be beneficent to all sorts of applications if Monte Carlo simulations were available via increasing computer power (Hall 1992:35). According to Hall (1992:35), the effect of combining these two ideas on statistical practice and theory "has been profound."

3. This procedure is called the "nonparametric bootstrap" because the bootstrap approximation of the standard error is based on the sample distribution, which is taken to be a nonparametric estimate of the unknown population distribution.

4. The simulations can also be used to establish the bias of a functional statistic, approximated as:

$$\hat{\varepsilon} = \frac{1}{B}\sum_{b=1}^{N} \hat{\theta}_b^* - \hat{\theta} = \bar{\hat{\theta}}^* - \hat{\theta}$$

5. If (B x α) is not an integer Efron and Tibshirani (1993:160) suggest the following procedure. Assuming that α is $\leq .5$, let $k = [(B+1 + \alpha)]$, the largest integer $\leq (B+1)\alpha$. The lower and upper limits are then established by taking the k^{th} and $(B+1 - k)^{\text{th}}$ values of $\hat{\theta}_b^*$.

6. As explained by Efron (1982), since $\alpha = \hat{H}_{boot}\left(\hat{\theta}_{LL}\right)$ and $1-\alpha = \hat{H}_{boot}\left(\hat{\theta}_{UL}\right)$, the percentile method confidence interval consists of the central $1 - 2\alpha$ proportion of the bootstrap distribution.

$$\hat{\theta}_{LL} \approx \hat{\theta}_b^{*\alpha} \qquad \hat{\theta}_{UL} \approx \hat{\theta}_b^{*(1-\alpha)}$$

where the lower limit of the confidence interval is the 100 x α^{th} empirical percentile of the distribution of $\hat{\theta}_b^*$ and the upper limit is the 100 x $(1-\alpha)^{\text{th}}$.

7. More formally, Efron and Tibshirani (1993:186) define the bias-correction value as

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left\{\hat{\theta}_b^* < \hat{\theta}\right\}}{B}\right)$$

where Φ^{-1} indicates the inverse function of the standard normal curve. Thus $\Phi^{-1}(.95) = 1.645$. This measures the "median bias" of $\hat{\theta}^*$, or the discrepancy between the median of $\hat{\theta}^*$ and $\hat{\theta}$.

8. In this sense, the bootstrap is a test of the reliability of asymptotic theory and thus should always be preferred. If the asymptotic and bootstrap critical values are similar in a given application, then one should use the bootstrap since it has already been computed. If the two are very different, than the bootstrap "provides an indication that asymptotic approximations are inaccurate" in that particular context (Horowitz 1997:202).

9. Davidson and Flachaire (2004) attribute the over-rejection primarily to the sensitivity of inequality measures to outliers in the upper-tail of the income distribution, which is usually described as being "heavy tailed" (i.e., as incomes get larger the tail decays slowly). Hall (1990) and Horowitz (2000) have shown that the bootstrap performs less well in these situations because a small number of extreme values can greatly influence the bootstrap distribution during resampling. The asymptotically pivotal indicators such as Theil and Atkinson are more sensitive to outliers in the upper tail than the Gini index which tends to be mid-range sensitive (see Cowell and Flachaire 2002 for a full discussion of this issue). While the coverage accuracy of the intervals is not perfect, at least the bias is known and is in a conservative direction. These methods, to paraphrase Chernick (1999:149), do not "get us something for nothing," but instead should be better interpreted as "getting the most from the little that is available."

10. The bootstrap procedures in this study employ resampling plans in the case of an independent and identically distributed (*iid*) sample of fixed size *n* from an unknown population. Although relatively undeveloped in the statistical literature, some have explored extensions of the *iid* bootstrap to data collected under complex sampling designs (Biewen 2002; Rao and Wu 1988; Jäntti 2003). Due to the constraints of working with cross-national aggregations of survey data, however, the LIS does not allow us to evaluate the usefulness of these extensions.

11. Adherence to the LIS guidelines was meant to assure maximum comparability of these findings to other quantitative analyses of the LIS database, and also in an effort to "test" the bootstrap procedures in the context of current LIS procedures.

12. Efron and Tibshirani (1993:52) find that "very seldom are more than $B = 200$ replications needed for estimating a standard error" and "even a small number...say $B = 50$ is often good enough to give a good7 estimate..." To establish bootstrap confidence intervals, most acknowledge that *B* must be much larger (as much as 1000 or more), however the literature offers little formal guidance on establishing *B* for confidence interval estimation.

13. If the bootstrap distribution of $\hat{\theta}_b^*$ is roughly normal (and the central limits theorem tells us that it is as $n \to \infty$), than the standard normal and percentile intervals will nearly agree as they do in the LIS data analyzed here.

14. Even though "Lorenz dominance" is the only unambiguous condition in which this would be true, it's standard practice to make assessments in this manner under conditions of intersecting Lorenz curves as well.

15. Obviously at lower alpha levels less statistical significance exists, meaning that the boxes in Table 6 would contain more nations. The one-sided *p*-values are provided in Table 5 to allow readers to make these assessment as they wish.

16. While the bootstrap procedures are effective tools when one has access to the microdata, the findings presented here are potentially useful to researchers attempting to interpret differences and changes in Gini coefficients in situations where this access is limited or not possible. Specifically the hypothesis tests conducted here suggest the following guidelines when comparing two Gini coefficients over time:

• less than 2.5 percent difference =	conclude that the difference is not statistically significant
• between 2.5 and 5.0 percent difference =	conclusion is ambiguous and likely determined by size of the sampling error, hypothesis test probably required
• more than 5.0 percent difference =	conclude that the difference is statistically significant

References

Acemoglu, Daron. 2002. "Cross-Country Inequality Trends." Luxembourg Income Study Working Paper No. 296.

Alderson, Arthur A., and François Nielsen. 2002. "Globalization and the Great U-Turn: Income Inequality Trends in 16 OECD Countries." American Journal of Sociology 107:1244-99.

Atkinson, Anthony B., Lee Rainwater, and Timothy M. Smeeding. 1995. Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study. Paris: OECD.

Beran, Rudolf. 1988. "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements." Journal of the American Statistical Association 83:687-697.

Bickel, Peter J., and David A. Freedman. 1981. "Some Asymptotic Theory for the Bootstrap." Annals of Statistics 9:1196-1217.

Biewen, Martin. 2002. "Bootstrap Inference for Inequality, Mobility, and Poverty Measurement." Journal of Econometrics 108:317-42.

Blau, Francine D., and Lawrence M. Kahn. 1996. "International Differences in Male Wage Inequality: Institutions Versus Market Forces." Journal of Political Economy 104:791-837.

Bradley, David, Evelyne Huber, Stephanie Moller, François Nielsen, and John D. Stephens. 2003. "Distribution and Redistribution in Postindustrial Democracies." World Politics 55:193-230.

Brady, David. 2003. "The Politics of Poverty: Left Political Institutions, the Welfare State, and Poverty." Social Forces 82:557-88.

Chernick, Michael R. 1999. Bootstrap Methods: A Practitioner's Guide. New York: John Wiley and Sons.

Cowell, Frank A., and Emmanuel Flachaire. 2002. "Sensitivity of Inequality Measures to Extreme Values." EUREQua, University Paris I Panthéon-Sorbonne, Working Paper 2002.04.

Davidson, Russell, and Emmanuel Flachaire. 2004. "Asymptotic and Bootstrap Inference For Inequality and Poverty Measures. Unpublished manuscript.

Davidson, Russell, and James G. MacKinnon. 1999. "The Size Distortion of Bootstrap Tests." Econometric Theory 15:361-376.

---------- 2000. "Improving the Reliability of Bootstrap Tests." Queens Institute for Economic Research, DP #995.

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

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." Annals of Statistics 7:1-26.

---------- 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Efron, Bradley and Robert J. Tibshirani. 1993. An Introduction to the Bootstrap. New York: Chapman and Hall.

Freeman, Richard B., and Lawrence F. Katz. 1995. "Introduction and Summary." In Richard B. Freeman, and Lawrence F. Katz (eds.) Differences and Changes in Wage Structures. Chicago: University of Chicago Press.

Friedman, Thomas L. 2000. The Lexus and the Olive Tree. New York: Farrar, Straus, and Giroux.

Gottschalk, Peter, and Timothy Smeeding. 2000. "Empirical Evidence on Income Inequality in Industrializing Countries." In Anthony B. Atkinson and François Bourguignon (eds.) Handbook of Income Distribution. New York: Elsevier-North Holland.

Gustafsson, Björn, and Mats Johansson. 1999. "In Search of Smoking Guns: What Makes Income Inequality Vary Over Time in Different Countries?" American Sociological Review 64:585-605.

Hall, Peter. 1988. "Theoretical Comparisons of Bootstrap Confidence Intervals" (with discussion). Annals of Statistics 16:925-53.

---------- 1990. "Asymptotic Properties of the Bootstrap for Heavy-Tailed Distributions." Annals of Probability 18:1342-1360.

---------- 1992. The Bootstrap and Edgeworth Expansion. New York: Springer.

Harrison, Bennett, and Barry Bluestone. 1988. The Great U-Turn. New York: Basic Books.

Horowitz, Joel L. 1994. "Bootstrap-based Critical Values for the Information Matrix Test." Journal of Econometrics 61:395-411.

---------- 1997. "Bootstrap Methods in Econometrics: Theory and Numerical Performance." In David M. Kreps and Kenneth F. Wallis (eds.) Advances in Economics and Econometrics: Theory and Applications, Vol. 3. Cambridge: Cambridge University Press.

Jäntti, Markus. 2003. "Estimating Standard Errors and Confidence Intervals for the Gini Coefficient in a Stratified Rotating Half-Panel Design." Statistics Finland (mimeo).

Kelly, Nathan J. 2004. "Does Politics Matter? Policy and Government's Equalizing Influence in the United States." American Politics Research 32:264-84.

Lamberrt, Peter J., Daniel L. Millimet, and Daniel Slottje. 2003. "Inequality Aversion and the Natural Rate of Subjective Inequality." Journal of Public Economics 87:1061-90.

Mills, Jeffrey A., and Sourushe Zandvakili. 1997. "Statistical Inference Via Bootstrapping for Measures of Inequality." Journal of Applied Econometrics 12:133-150.

Mueller, Dennis C., and Thomas Stratmann. 2003. "The Economic Effects of Democratic Participation." Journal of Public Economics 87:2129-55.

Nickell, Stephen, and Brian Bell. 1996. "The Collapse in Demand for the Unskilled and Unemployed Across the OECD." Oxford Review of Economic Policy 11:40-62.

Ram, Ratti. 1997. "Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World." Southern Economic Journal 64:565-83.

Rao, J.N.K., and C.F.J. Wu. 1988. "Resampling Inference With Complex Survey Data." Journal of the American Statistical Association 83:231-41.

Rossi, Nicola, Gianni Toniolo, and Giovanni Vecchi. 2001. "Is the Kuznets Curve Still Alive? Evidence from Italian Household Budgets, 1881-1961." Journal of Economic History 61:904-25.

Shao, Jun, and Dongsheng Tu. 1995. The Jackknife and Bootstrap. New York: Springer.

Singh, Kesar. 1981. "On the Asymptotic Accuracy of Efron's Bootstrap." Annals of Statistics 9:1187-1195.

Smeeding, Timothy. 2002. "Globalization, Inequality, and the Rich Countries of the G-20: Evidence from the Luxembourg Income Study." Luxembourg Income Study Working Paper No. 320.

Trede, Mark. 2002. "Bootstrapping Inequality Measures Under the Null Hypothesis: Is it Worth the Effort?" Journal of Economics 9:262-81.

Van Kerm, Philippe. 2002. "Inference on Inequality Measures: A Monte Carlo Experiment." Journal of Economics 9:283-306.

Xu, Kuan. 2000. "Inference for Generalized Gini Indices Using the Iterated-Bootstrap Method." Journal of Business and Economic Statistics 18:223-227.

Table 1. Bootstrap Methods for Constructing Confidence Intervals

Method	Advantages	Disadvantages
Standard Normal	• Computationally easy and most familiar	• May lose accuracy for statistics other than the mean • Provides symmetric intervals • Intervals may not be range preserving • Requires distributional assumptions
Bootstrap-*t*	• Corrects for disadvantages of standard normal method	• Requires standard error estimates (nested bootstrap is computationally burdensome) • May require variance-stabilization • Intervals may not be range preserving
Percentile	• Requires no standard error estimates • Requires no variance stabilizing • Intervals are range preserving • Computationally easy	• Caution in small sample sizes • May require bias correction procedures
Bias-Corrected Percentile	• Corrects for disadvantages of percentile method	• Added computational burden • Adjustments may not be necessary in large sample sizes

Table 2. Simulation Analysis of CI Performance Using Finland 2000, 5% Subsample

Gini parameter = 0.2474; n = 521 households; *B* reps = 1,000; Monte Carlo trials = 600

Confidence Interval Method	Median Lower-Bound	Median Upper-Bound	Pseudo α-Level*	ERP**
Standard Normal	0.2220	0.2703	0.057	0.007
Percentile	0.2231	0.2709	0.063	0.013
Bias-Corrected	0.2252	0.2734	0.055	0.005

* Proportion of trials in which the interval fails to capture 0.2474.
** Difference between actual and nominal probabilities of rejection.

Table 3. Survey Descriptions and 95% Confidence Intervals, Subsample of Ten LIS Surveys: Replication Analysis

	Survey Year	Sample Size	Gini Index	Number of Replications	Bias	Std. Error	Standard Normal LL	Standard Normal UL	Percentile LL	Percentile UL	Bias Corrected LL	Bias Corrected UL
Luxembourg (LUX)	1994	1,813	0.235	250	0.0002	0.0049	0.226	0.245	0.226	0.246	0.226	0.246
				500	-0.0002	0.0050	0.225	0.245	0.225	0.245	0.226	0.246
				1000	-0.0004	0.0050	0.226	0.245	0.225	0.245	0.226	0.245
Hungary (HUN)	1999	2,013	0.295	250	-0.0001	0.0066	0.282	0.308	0.281	0.310	0.282	0.311
				500	-0.0004	0.0073	0.281	0.309	0.281	0.309	0.281	0.310
				1000	-0.0004	0.0073	0.281	0.309	0.281	0.309	0.282	0.310
Russia (RUS)	2000	3,055	0.434	250	-0.0002	0.0066	0.421	0.447	0.421	0.448	0.420	0.447
				500	-0.0003	0.0067	0.421	0.448	0.421	0.447	0.422	0.447
				1000	-0.0003	0.0069	0.421	0.448	0.421	0.448	0.421	0.449
Netherlands (NLD)	1994	5,146	0.257	250	-0.0000	0.0035	0.251	0.264	0.251	0.265	0.250	0.265
				500	0.0004	0.0034	0.251	0.264	0.252	0.265	0.250	0.264
				1000	0.0000	0.0034	0.251	0.264	0.251	0.264	0.251	0.264
Italy (ITA)	1995	8,120	0.342	250	0.0001	0.0045	0.333	0.351	0.333	0.351	0.333	0.351
				500	-0.0002	0.0048	0.332	0.351	0.333	0.352	0.333	0.353
				1000	-0.0001	0.0048	0.332	0.351	0.332	0.351	0.333	0.352
Finland (FIN)	2000	10,421	0.247	250	0.0002	0.0033	0.241	0.254	0.241	0.254	0.241	0.254
				500	-0.0001	0.0031	0.241	0.254	0.241	0.254	0.241	0.254
				1000	-0.0000	0.0029	0.242	0.253	0.242	0.253	0.242	0.253
Sweden (SWE)	2000	14,491	0.251	250	0.0001	0.0026	0.246	0.257	0.247	0.257	0.246	0.256
				500	0.0001	0.0026	0.246	0.257	0.247	0.257	0.247	0.258
				1000	-0.0001	0.0026	0.247	0.257	0.246	0.257	0.246	0.257
Slovak Rep. (SVK)	1996	16,197	0.241	250	-0.0001	0.0040	0.233	0.249	0.233	0.249	0.234	0.249
				500	0.0003	0.0043	0.232	0.249	0.233	0.250	0.233	0.250
				1000	-0.0000	0.0041	0.233	0.249	0.233	0.250	0.233	0.250
United Kingdom (GBR)	1999	24,976	0.345	250	0.0000	0.0021	0.341	0.349	0.340	0.349	0.340	0.349
				500	0.0000	0.0023	0.340	0.349	0.340	0.349	0.340	0.349
				1000	-0.0000	0.0023	0.340	0.349	0.340	0.349	0.340	0.349
United States (USA)	2000	49,351	0.368	250	-0.0001	0.0020	0.364	0.372	0.364	0.372	0.364	0.372
				500	0.0000	0.0020	0.364	0.372	0.365	0.372	0.365	0.372
				1000	0.0000	0.0020	0.364	0.372	0.364	0.372	0.364	0.372

Figure 1. Standard Normal, Percentile, and Bias Corrected 95% Confidence Intervals; Subsample of Ten Surveys

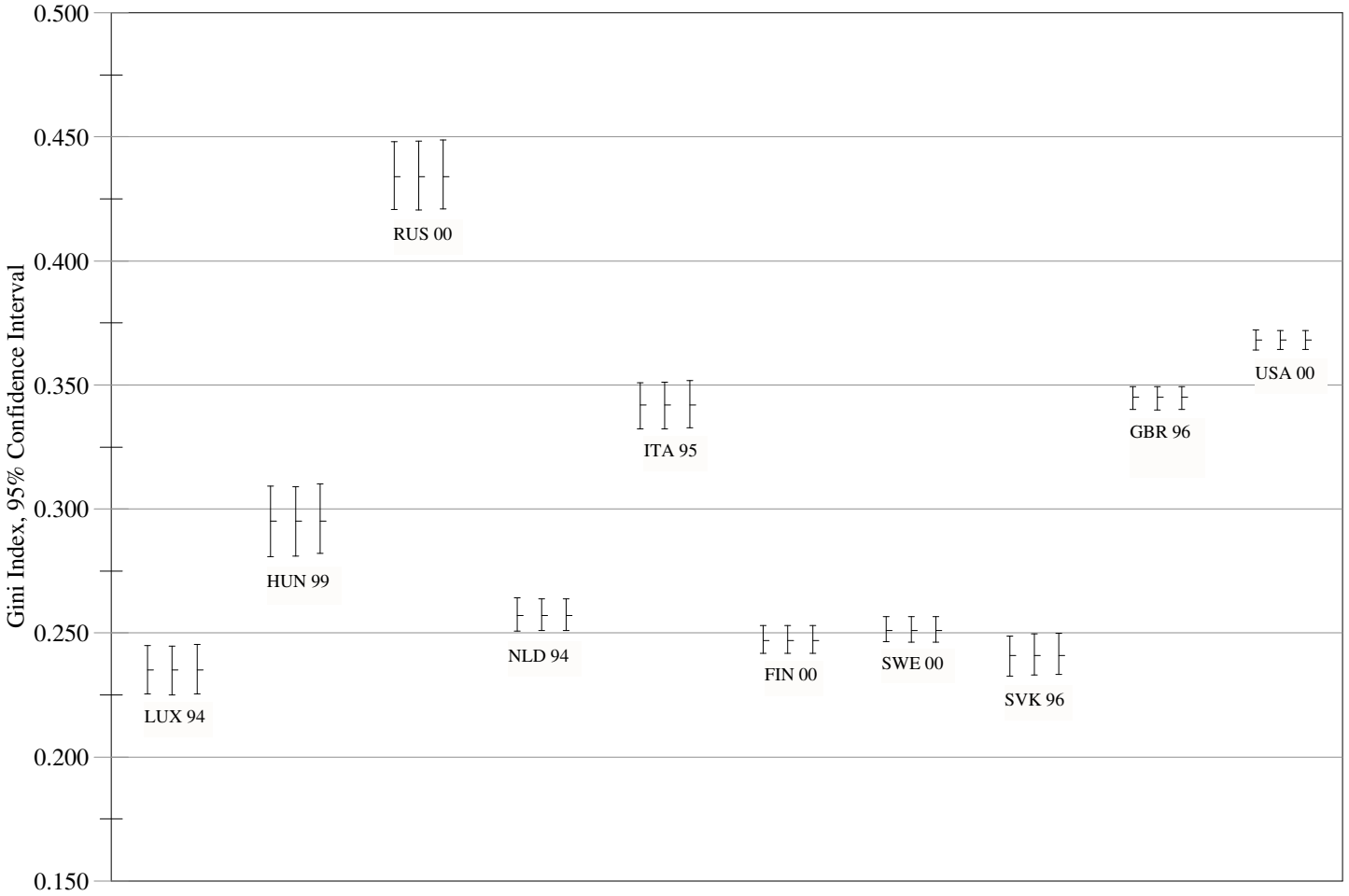


Table 4. Income Inequality in the LIS Database

		c1980				c1990				c2000			
		Gini	Standard	95% CI		Gini	Standard	95% CI		Gini	Standard	95% CI	
	Code	Index	Error	LL	UL	Index	Error	LL	UL	Index	Error	LL	UL
Global North													
Australia	AUS	.281	.0020	.277	.285	.304	.0023	.299	.308	.311*	.0037	.304	.318
Austria	AUT					.227	.0025	.222	.232	.266	.0056	.255	.277
Belgium	BEL					.232	.0040	.224	.240	.250	.0036	.243	.257
Canada	CAN	.284	.0023	.280	.288	.281	.0030	.275	.287	.302	.0028	.297	.308
Denmark	DEN					.236	.0026	.232	.242				
Finland	FIN					.210	.0016	.207	.213	.247	.0033	.241	.254
France	FRA	.288	.0053	.278	.299	.287	.0033	.280	.294	.288*	.0029	.282	.294
Germany	DEU	.244	.0048	.234	.253	.257	.0057	.246	.269	.264	.0032	.258	.270
Ireland	IRE					.328	.0057	.317	.339	.323	.0011	.304	.346
Italy	ITA					.290	.0044	.282	.299	.333	.0048	.324	.342
Luxembourg	LUX					.239	.0063	.226	.252	.260	.0051	.250	.270
Netherlands	NLD	.260	.0083	.246	.278	.266	.0051	.256	.276	.256	.0043	.248	.265
Norway	NOR	.223	.0025	.219	.228	.231	.0042	.224	.240	.251	.0034	.244	.257
Spain	ESP	.318	.0025	.314	.323	.303	.0027	.298	.309				
Sweden	SWE	.197	.0022	.193	.201	.229	.0022	.225	.234	.252	.0025	.246	.256
Switzerland	CHE	.309	.0050	.300	.319	.307	.0055	.297	.318				
United Kingdom	GBR	.270	.0029	.264	.275	.336	.0039	.329	.344	.345	.0023	.340	.349
United States	USA	.301	.0023	.297	.306	.336	.0025	.331	.341	.368	.0020	.364	.372
Group Average		*.270*	*.004*	*.264*	*.278*	*.272*	*.004*	*.265*	*.280*	*.288*	*.003*	*.280*	*.296*
Group Std. Dev.		*.035*	*.002*	*.035*	*.036*	*.040*	*.001*	*.040*	*.040*	*.039*	*.001*	*.039*	*.039*
Russia and Eastern Europe													
Czech Rep.	CZE					.207	.0024	.202	.212	.259^	.0022	.254	.263
Estonia	EST									.361	.0050	.352	.371
Hungary	HUN					.283	.0071	.269	.297	.295	.0075	.280	.309
Poland	POL					.274	.0037	.266	.281	.293	.0019	.289	.297
Romania	ROM									.277	.0023	.273	.282
Russia	RUS					.395	.0049	.385	.405	.434	.0069	.420	.447
Slovak Rep.	SVK					.189	.0018	.186	.193	.241^	.0040	.233	.249
Slovenia	SVN									.249	.0039	.242	.257
Group Average						*.270*	*.004*	*.262*	*.277*	*.301*	*.004*	*.293*	*.309*
Group Std. Dev.						*.073*	*.002*	*.070*	*.075*	*.061*	*.002*	*.059*	*.063*
Other LIS Countries													
Israel	ISR	.303	.0057	.292	.314	.305	.0038	.298	.313	.346	.0043	.338	.355
Mexico	MEX	.448	.0077	.432	.462	.467	.0054	.457	.478	.491	.0067	.478	.504
Taiwan	TWN	.267	.0019	.263	.271	.271	.0019	.268	.275	.296	.0023	.292	.301
Group Average		*.339*	*.005*	*.329*	*.349*	*.348*	*.004*	*.341*	*.355*	*.363*	*.004*	*.369*	*.387*
Group Std. Dev.		*.078*	*.002*	*.073*	*.082*	*.086*	*.001*	*.083*	*.088*	*.084*	*.002*	*.079*	*.086*
Overall Average		*.285*	*.004*	*.278*	*.293*	*.280*	*.004*	*.273*	*.288*	*.302*	*.004*	*.294*	*.310*
Overall Std. Dev.		*.056*	*.002*	*.053*	*.057*	*.060*	*.002*	*.059*	*.061*	*.060*	*.002*	*.059*	*.061*

Notes: All Gini indices are based on net, disposable household income, see Apendix A

Reported confidence intervals are percentile method

*1994 ; ^1996

Figure 2. Gini Coefficients and 95% Confidence Intervals for Levels of Inequality, circa 2000

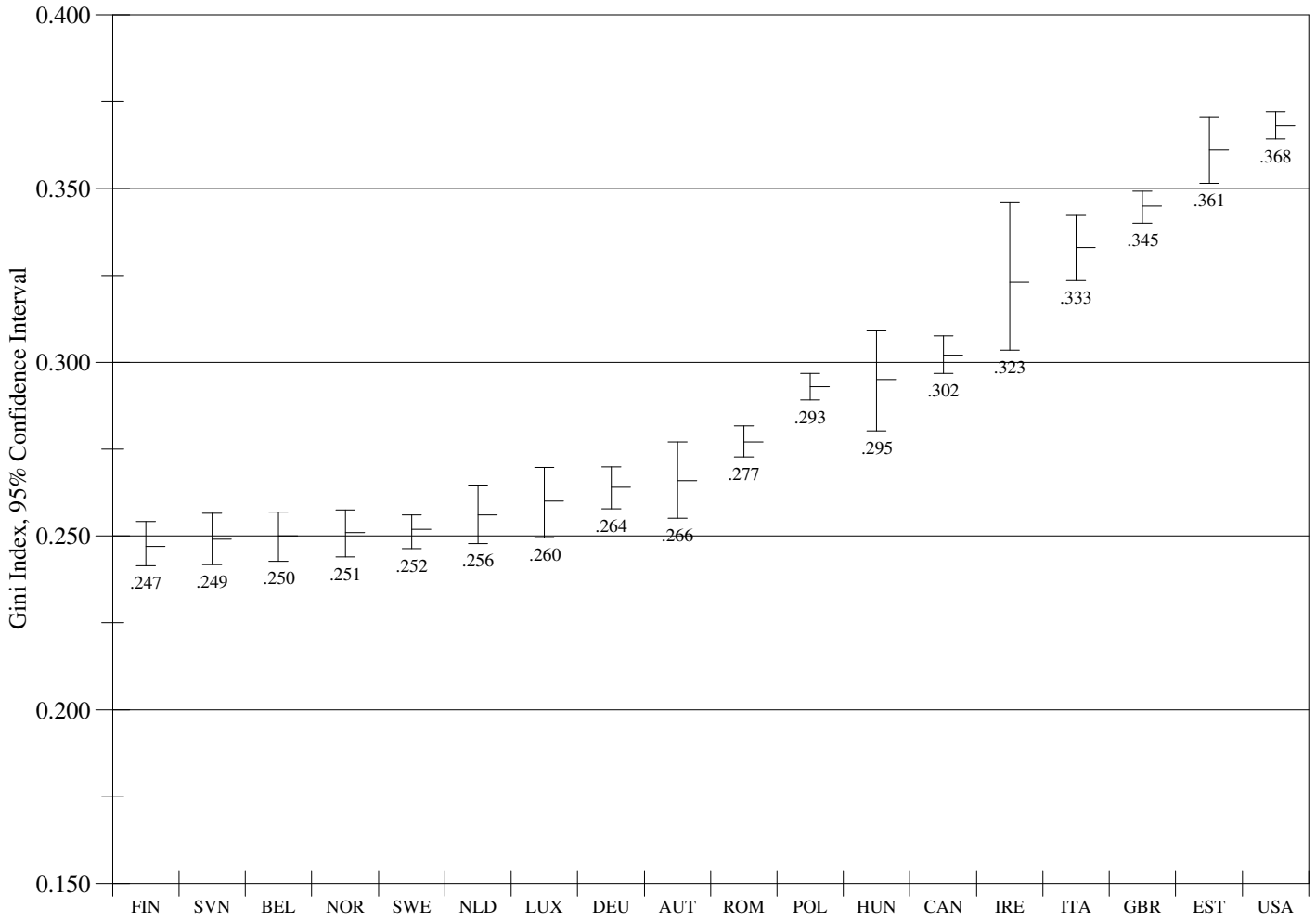


Table 5. Null Hypothesis Tests for Differences in Inequality Levels, circa 2000

Absolute Diff. Percentage Diff. One-Sided p*	FIN .247	SVN .249	BEL .250	NOR .251	SWE .252*	NLD .256	LUX .260	DEU .264	AUT .266	ROM .277	POL .293	HUN .295	CAN .302	IRE .323	ITA .333	GBR .345	EST .361	USA .368
FIN																		
SVN	.002 0.8 .367																	
BEL	.003 1.2 .294	.001 0.4 .432																
NOR	.004 1.6 .259	.002 0.8 .391	.001 0.4 .423															
SWE	.005 2.0 .160	.003 1.2 .329	.002 0.8 .351	.001 0.4 .405														
NLD	.009 3.6 .045	.007 2.8 .124	.006 2.4 .153	.005 2.0 .171	.004 1.6 .194													
LUX	**.013** **5.3** **.018**	.011 4.4 .042	.010 4.0 .067	.009 3.5 .067	.008 3.2 .074	.004 1.5 .313												
DEU	**.017** **6.4** **o**	**.015** **5.7** **o**	**.014** **5.3** **o**	**.013** **4.9** **o**	**.012** **4.5** **.001**	.008 3.0 .067	.004 1.5 .238											
AUT	**.019** **7.1** **o**	**.017** **6.8** **.007**	**.016** **6.4** **.008**	**.015** **5.6** **.012**	**.014** **5.6** **.006**	.010 3.8 .072	.006 2.3 .181	.002 0.8 .369										
ROM									**.011** **4.1** **.022**									
POL																		
HUN										**.018** **6.5** **.006**	.002 0.7 .422							
CAN											**.009** **3.0** **.007**	.007 2.3 .189						
IRE													**.021** **6.5** **.023**					
ITA														.010 3.0 .189				
GBR														**.022** **6.4** **.024**	**.012** **3.6** **.014**			
EST																		
USA																	.007 1.9 .081	

* Sweden's Gini index is .2515

Notes: Blank cells or "o" in p* cell indicate that confidence intervals overlap

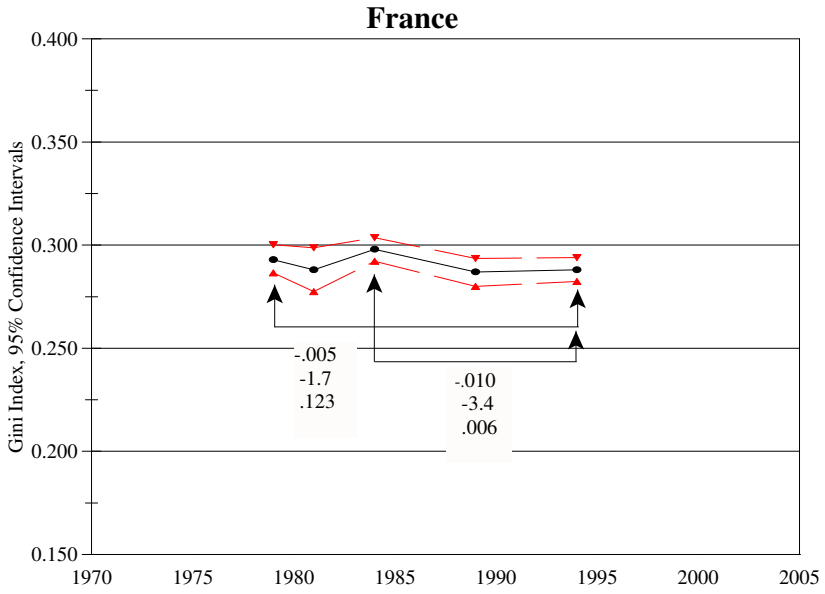
Bolded numbers indicate that the difference is statistically significant at $\alpha = .05$

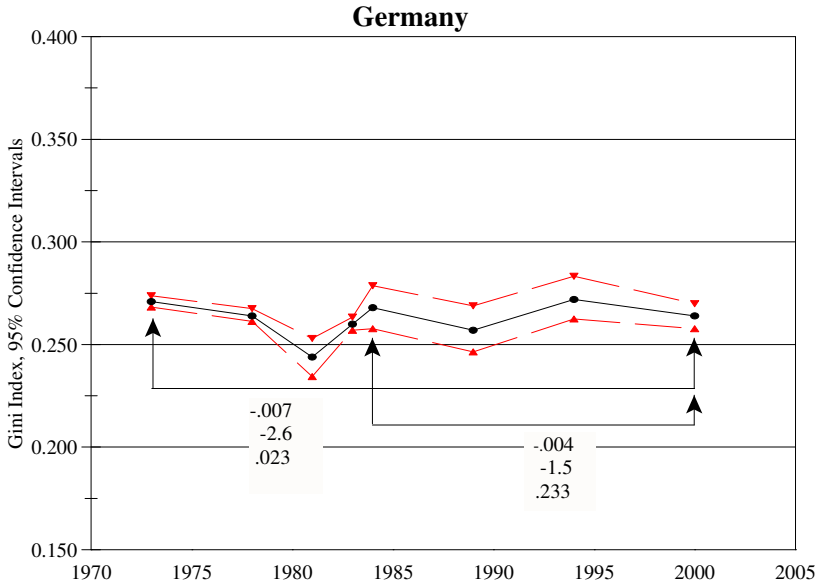
Table 6. Ranking Levels of Inequality using Statistical Significance, circa 2000

FIN	SVN	BEL	NOR	SWE	NLD	LUX	DEU	AUT
Fin	Fin	Fin	Fin	Fin	Fin	*fin*	*fin*	*fin*
Svn	Svn	Svn	Svn	Svn	Svn	Svn	*svn*	*svn*
Bel	Bel	Bel	Bel	Bel	Bel	Bel	*bel*	*bel*
Nor	Nor	Nor	Nor	Nor	Nor	Nor	*nor*	*nor*
Swe	Swe	Swe	Swe	Swe	Swe	Swe	*swe*	*swe*
Nld	Nld	Nld	Nld	Nld	Nld	Nld	Nld	Nld
LUX	Lux	Lux	Lux	Lux	Lux	Lux	Lux	Lux
DEU	DEU	DEU	DEU	DEU	Deu	Deu	Deu	Deu
AUT	AUT	AUT	AUT	AUT	Aut	Aut	AUT	Aut
ROM	ROM	ROM	ROM	ROM	ROM	ROM	ROM	ROM
POL	POL	POL	POL	POL	POL	POL	POL	POL
HUN	HUN	HUN	HUN	HUN	HUN	HUN	HUN	HUN
CAN	CAN	CAN	CAN	CAN	CAN	CAN	CAN	CAN
IRE	IRE	IRE	IRE	IRE	IRE	IRE	IRE	IRE
ITA	ITA	ITA	ITA	ITA	ITA	ITA	ITA	ITA
GBR	GBR	GBR	GBR	GBR	GBR	GBR	GBR	GBR
EST	EST	EST	EST	EST	EST	EST	EST	EST
USA	USA	USA	USA	USA	USA	USA	USA	USA

ROM	POL	HUN	CAN	IRE	ITA	GBR	EST	USA
fin	*fin*	*fin*	*fin*	*fin*	*fin*	*fin*	*fin*	*fin*
svn	*svn*	*svn*	*svn*	*svn*	*svn*	*svn*	*svn*	*svn*
bel	*bel*	*bel*	*bel*	*bel*	*bel*	*bel*	*bel*	*bel*
nor	*nor*	*nor*	*nor*	*nor*	*nor*	*nor*	*nor*	*nor*
swe	*swe*	*swe*	*swe*	*swe*	*swe*	*swe*	*swe*	*swe*
nld	*nld*	*nld*	*nld*	*nld*	*nld*	*nld*	*nld*	*nld*
lux	*lux*	*lux*	*luc*	*lux*	*lux*	*lux*	*lux*	*lux*
deu	*deu*	*deu*	*deu*	*deu*	*deu*	*deu*	*deu*	*deu*
aut	*aut*	*aut*	*sut*	*aut*	*aut*	*aut*	*aut*	*aut*
Rom	*rom*	*rom*	*rom*	*rom*	*rom*	*rom*	*rom*	*rom*
POL	Pol	Pol	*pol*	*pol*	*pol*	*pol*	*pol*	*pol*
HUN	Hun	Hun	Hun	*hun*	*hun*	*hun*	*hun*	*hun*
CAN	CAN	CAN	Can	*can*	*can*	*can*	*can*	*can*
IRE	IRE	IRE	IRE	Ire	Ire	*ire*	*ire*	*ire*
ITA	ITA	ITA	ITA	Ita	Ita	*ita*	*ita*	*ita*
GBR	GBR	GBR	GBR	GBR	GBR	Gbr	*gbr*	*gbr*
EST	EST	EST	EST	EST	EST	EST	Est	Est
USA	USA	USA	USA	USA	USA	USA	Usa	Usa

Figure 3a. The Continental Pattern of Inequality





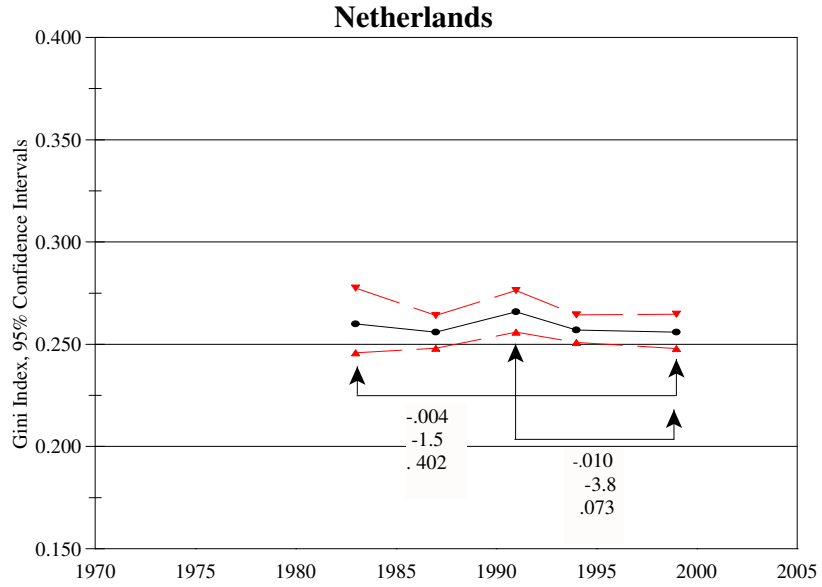


Figure 3b. The Continental Pattern of Inequality

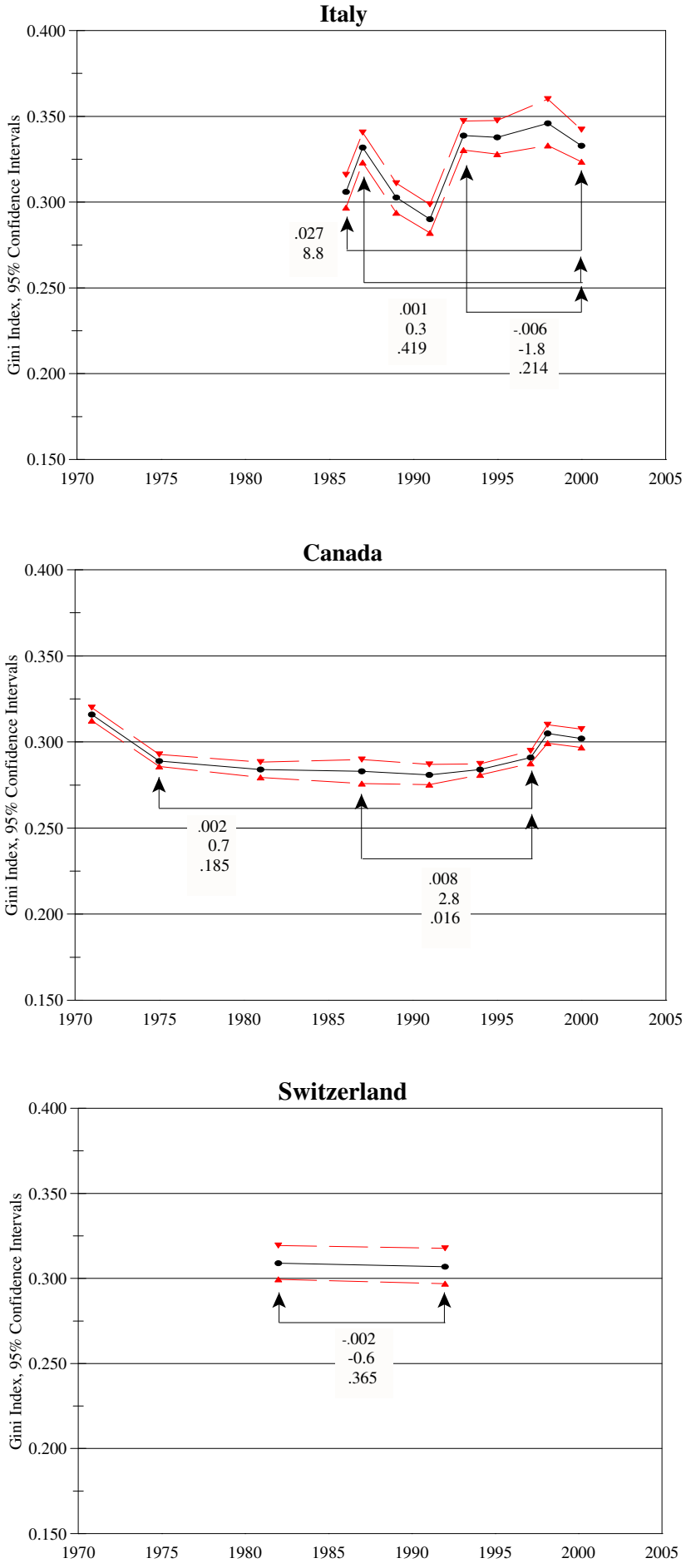
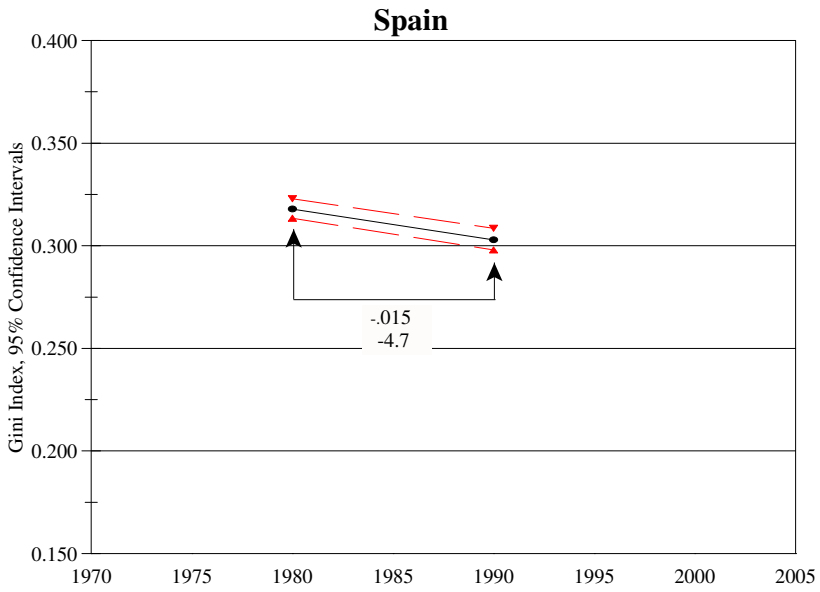


Figure 3c. The Continental Pattern of Inequality



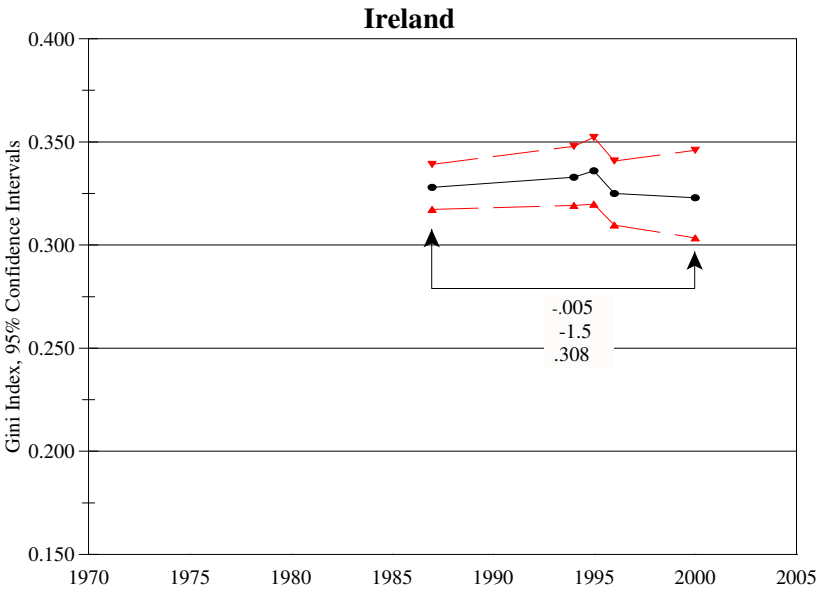
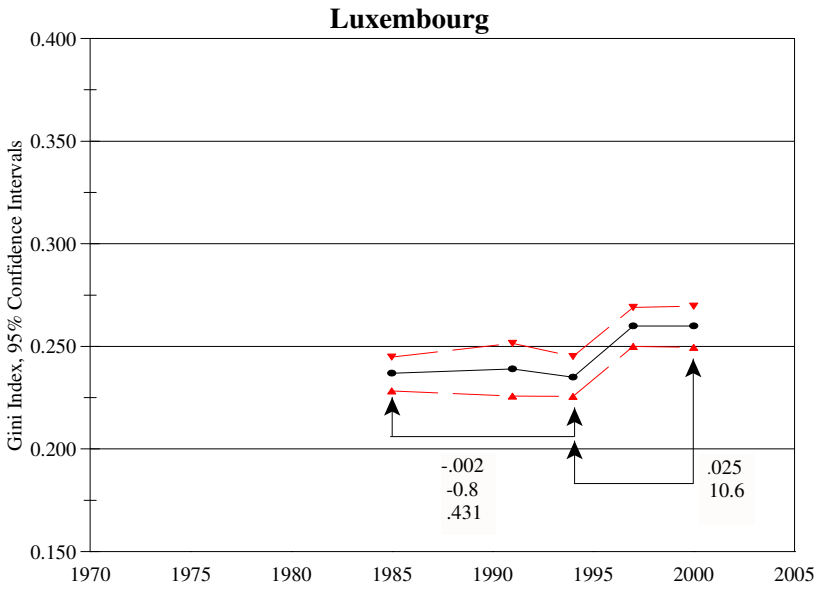
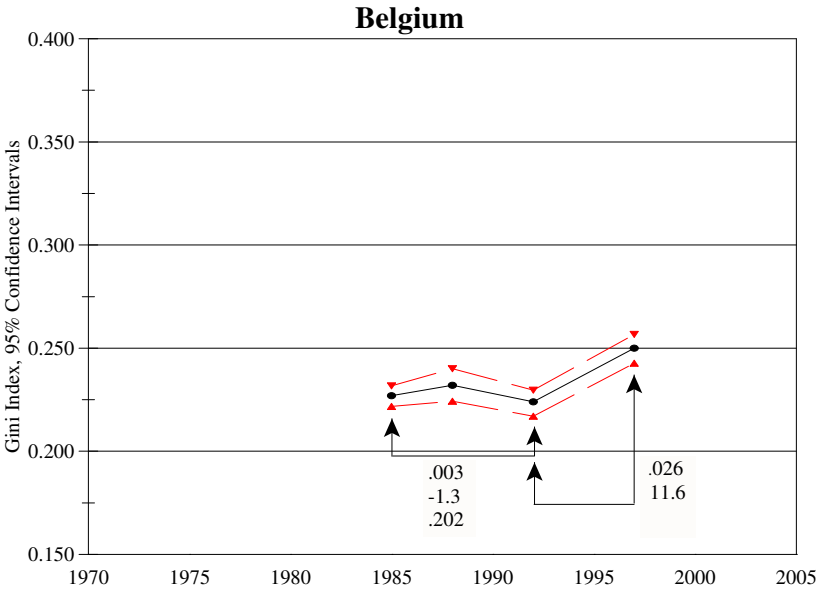


Figure 3d. The Continental Pattern of Inequality





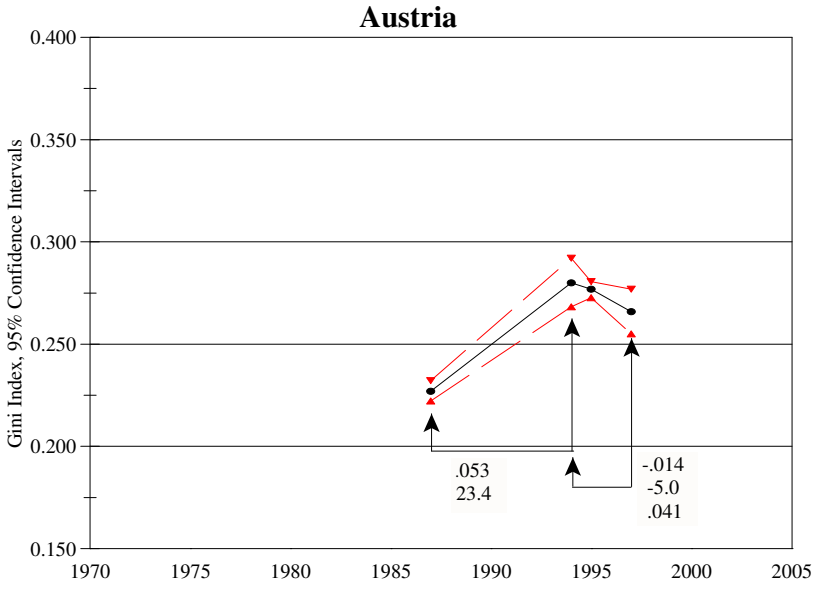


Figure 4. The Anglo Pattern of Inequality

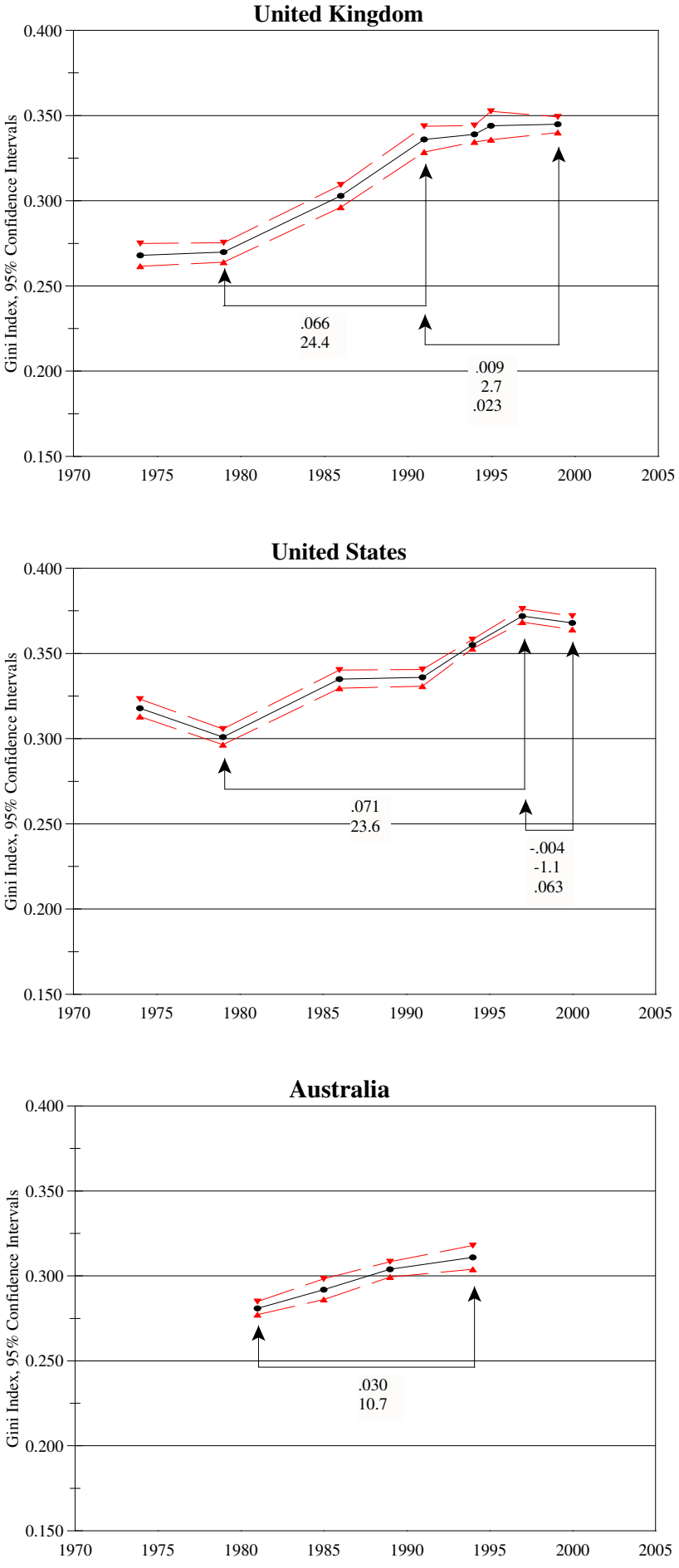


Figure 5. The Scandinavian Pattern of Inequality

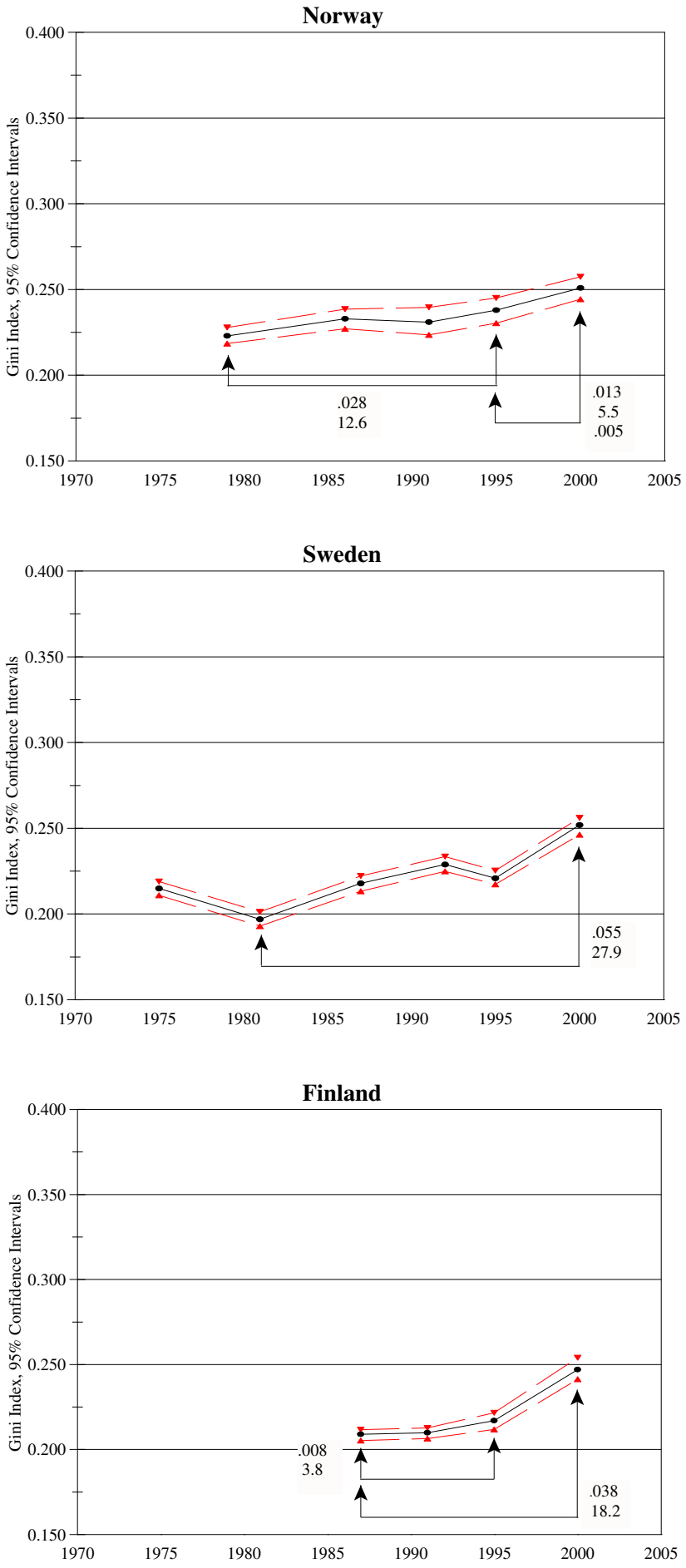


Figure 6. Patterns of Distributional Change in the Global North, 1980 - 2000

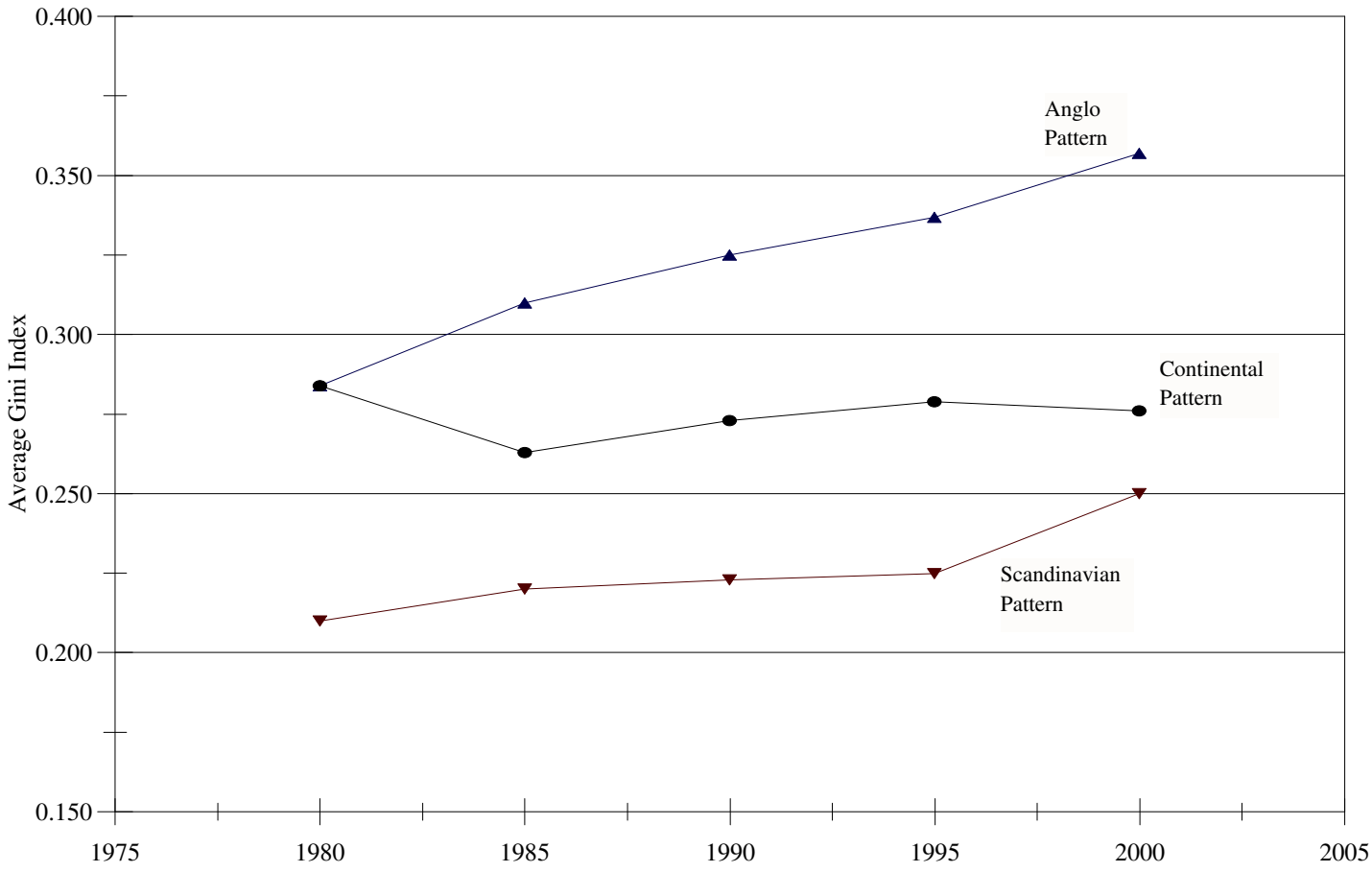
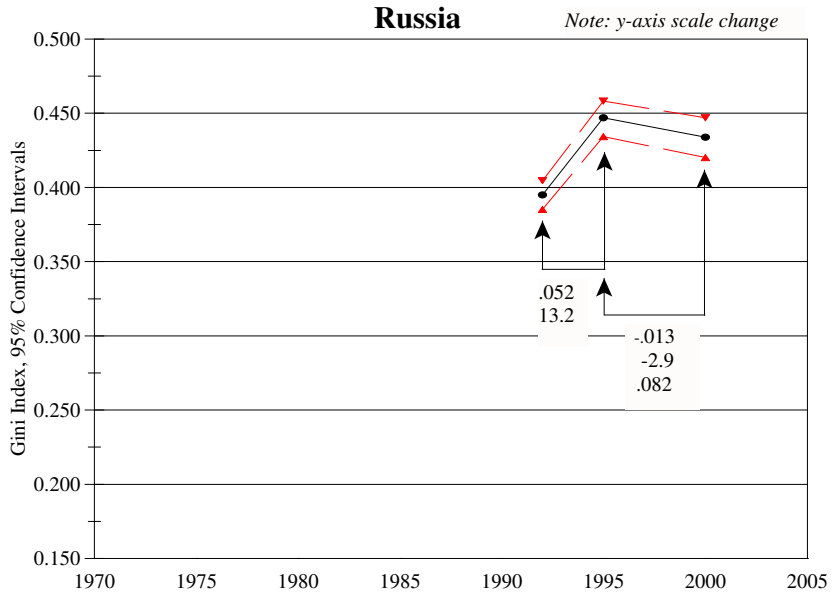
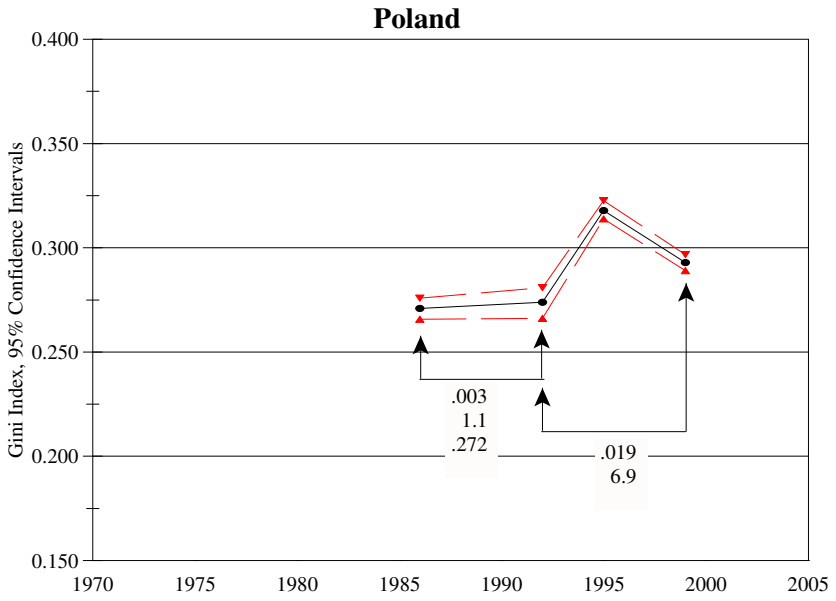


Figure 7. Russia and Eastern Europe





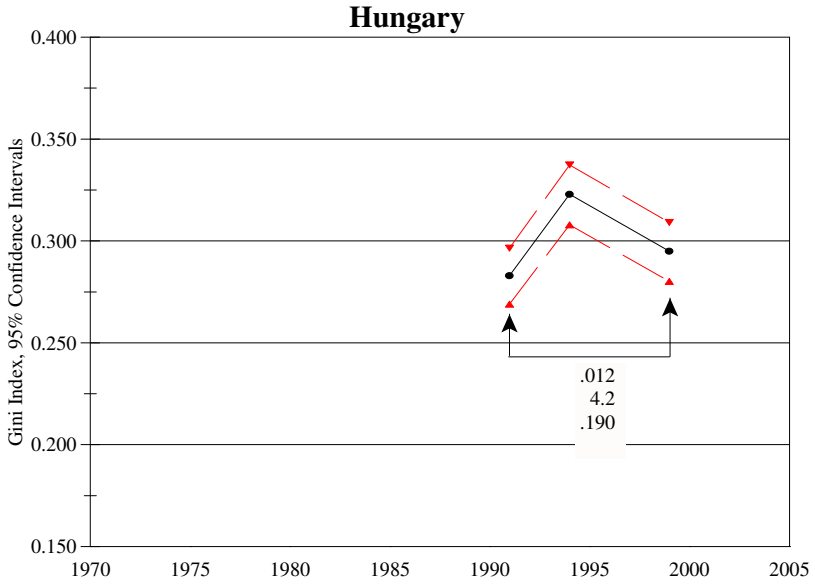


Figure 8. Other LIS Member Countries