J Econ Inequal (2008) 6:5–32
DOI 10.1007/s10888-006-9043-7

Ranking inequality: Applications of multivariate subset selection

William C. Horrace · Joseph T. Marchand · Timothy M. Smeeding

Received: 28 September 2005 / Accepted: 31 October 2006 /
Published online: 22 December 2006

Abstract Inequality measures are often presented in the form of a rank ordering to highlight their relative magnitudes. However, a rank ordering may produce misleading inference, because the inequality measures themselves are statistical estimators with different standard errors, and because a rank ordering necessarily implies multiple comparisons across all measures. Within this setting, if differences between several inequality measures are *simultaneously* and statistically insignificant, the interpretation of the ranking is changed. This study uses a multivariate subset selection procedure to make simultaneous distinctions across inequality measures at a pre-specified confidence level. Three applications of this procedure are explored using country-level data from the Luxembourg Income Study. The findings show that simultaneous precision plays an important role in relative inequality comparisons and should not be ignored.

Key words income distribution · inference · poverty · subset selection.

JEL Classifications C12 · C15 · D31 · D63 · I32.

1 Introduction

Comparisons of income distributions are often used to understand how different groups of agents distribute their resources. Indeed, cross-country comparisons of income inequality are common in analyses informing policy-making. Measures of income distribution are

W. C. Horrace (✉) · J. T. Marchand
Department of Economics and Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020, USA
e-mail: whorrace@maxwell.syr.edu

T. M. Smeeding
Luxembourg Income Study and Center for Policy Research,
Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, NY, USA

featured among the indicators of social cohesion agreed upon by the European Union to monitor the performance of member countries and are central to the 2006 World Development Report of the World Bank [39]. From these comparisons, researchers draw conclusions on how equally or unequally the resources of a group are distributed, relative to their comparison groups. Consequently, the subject of inequality is necessarily one of relative measure. One cannot typically draw strong conclusions about a group's inequality, unless it is in comparison to the inequality of another group. By itself, an inequality measure of a particular value or an income distribution of a certain shape may mean little to the observer. Rather, inequality becomes meaningful through comparison of these measures to measures of other groups.[1]

This study introduces a ranking-and-selection procedure known as *multivariate subset selection* to the inequality literature.[2] The selection procedure allows us to make multivariate inferential statements such as: "with a pre-specified probability, some subset of countries (from a larger universe) is best (most equal) in terms of inequality, and some subset of countries is worst (most unequal), relative to the other countries in the sample." By taking into account *multivariate* sampling variability, there can be multiple ties (in a probabilistic sense) for best and worst when ranking inequality estimates. This is in stark contrast to the deterministic outcome that the countries at the extreme ends of the rank ordering are best and worst. This is also an improvement over a series of *univariate* statistical inferences from which one might simply identify pairs of countries as statistically distinguishable in terms of inequality.[3] For example, the country at the top of the ranking may be no different from the next three countries in the ranking, given a certain level of confidence. Multivariate subset selection allows the researcher to use information that has previously been ignored or determined by arbitrary magnitude cutoff rules.[4] This study demonstrates that precision matters in the rank ordering of inequality measures, and that ignoring it can lead to erroneous conclusions. As such, the technique represents a substantial contribution to the inequality literature.

The multivariate subset selection procedure is described and then applied to the latest data from the Luxembourg Income Study (LIS) in three different ways. First, there is a single period analysis, where the magnitudes and precision of the Gini coefficient, the Theil index, and the Varlog index are compared across a cross-section of countries. The subsets that are created are the first-best, second-best, first-worst, and second-worst in terms of inequality using a pre-determined 90% confidence level. The different measures of inequality are compared based on their implications. The differences show that interpretation of the ranked estimators can change once precision is taken into account, and that certain estimators may be better than others in a rank order setting.

Second, a panel of 12 countries is followed across four LIS waves (periods) to track how relative inequality changes over time. This is done for the Gini coefficient only. Again, the subset selection procedure determines the first-best, second-best, first-worst, and second-

[1] Indeed, in economics absolute measures are not often identified; typically only relative measures are. For a more detailed discussion of these issues, see Horrace [22], Atkinson and Bourguignon [2], and Atkinson, Rainwater, and Smeeding [3].

[2] Subset selection has also been applied in the analysis of productive efficiency [23] and labor market wage differentials [22].

[3] Such inference is called "per comparison" inference and ignores the multiplicity implicit in the rank ordering. It is not just that country A is bigger than B; it is that A is bigger than B and bigger than C but, perhaps, smaller than D; a multiple inferential statement.

[4] For instance, the 1% difference in centile shares used by Atkinson, Rainwater, and Smeeding [3] as a rule of thumb.

worst countries. Two different levels of confidence are selected for analysis, 90 and 95%, to demonstrate how the confidence level affects inference on order statistics and cardinality of the best and worst subsets. This analysis also demonstrates that a country's relative position in a rank ordering may change over time but that its rank is not changing in a statistical sense.

Third, in an extension to the second exercise, this same panel is followed over the same four LIS waves to see how relative poverty changes over time. (Technically poverty is *not* a measure of income inequality, but we include it in this research to show that the rank techniques are widely applicable. Additionally, like the income inequality measures, poverty is an economically and empirically relevant measure of social well-being.) The relative poverty measure used is 50% of the median income of the total population. The purpose is to determine whether subset selections based on bootstrapped standard errors are different for the poverty measures than for the inequality measures, since Davidson and Flachaire [17] have argued that bootstrapped standard errors have different levels of accuracy when applied to the different measures.

The findings of this study suggest that the precision of inequality estimators matters for both the measurement and interpretation of relative inequality rankings. For the single period analysis, the ranking differences on the cross-section of countries result from using different measures and different sample sizes between countries. For example, a country that ranks third using one measure and sixth using another measure may still be contained in the same subset for a given level of confidence. For the panel analysis, the ranking differences are based on relative country movements of inequality and different sample sizes between countries. This setting is extremely relevant for policy makers.[5] For example, a country with a rank of 10 in one year and 9 in the next year may be meaningless in a statistical sense, as it could be that neighboring countries in the rankings are getting worse rather than the country in question getting better in a relative sense. Multivariate subset selection helps make these distinctions with statements of confidence. Also, the subset selection technique is a practical alternative to using arbitrary magnitude cutoff rules in rankings.

The paper is organized as follows. Section 2 provides a brief review of the previous literature. The data are described in Section 3. Section 4 details the methodology of the study. The methodology section consists of the construction of the inequality estimators and their bootstrap standard errors, the ranking of these estimates, and the subset selection technique that is applied to the ranking. The empirical results are presented in Section 5. Section 6 extends the analysis to poverty. Concluding remarks are offered in Section 7.

2 Literature review

This research adds to the economics literature on inference for rank statistics and stochastic dominance, especially those using the Lorenz curve. Various procedures have been developed to assess the usefulness of rankings and to determine "ties" when specific estimates appear to be the same [e.g., 8, 9, 16, 18, 19, 31–33, 41]. The issue of consistent and careful ranking is important in other domains, as well. Rankings are made for measures of equality opportunity [30]; in cases where information is incomplete, censored, or

[5] For interpretation, it is worthwhile to point out that part of this discussion is the issue of point estimation versus interval estimation. This is an issue for policy makers, as they usually want a point estimate rather than an interval. The subset selection allows policy makers to 'have their cake and eat it too'. That is, they get point estimates along with the precision of an interval. Subset selection is an intuitive way to incorporate the precision of the point estimates in a ranking setting. Also, policy makers tend to make too much of a relative ranking without viewing the ranking in contexts of statements of confidence.

unavailable [e.g., 15]; where only one or another attribute of economic status is used in the ranking [29]; and where differential weights based on entropy measures are employed [4]. Axiomatic approaches to ranking have also been applied to test for equality using the Gini coefficient and methods for weighting and extending the Gini coefficient [e.g., 12, 20, 38]; for ranking opportunity sets more generally [34]; and for measures of poverty as well as measures of inequality [16]. The inference implied by the subset ranking procedure described in the next section may have implications for a wide range of decision rules and procedures for assessing rank statistics in the income and poverty literature.

Given the ease of computation, scalar inequality indices are often computed, with various measures to choose from [for a review, see 14]. These indices are often ranked using only the magnitudes of the estimators [e.g., 10].[6] The strength of ranking these scalar measures is that it produces a linear, complete, and transitive order. This makes the ranking easy to interpret. There are also no ties between coefficients if taken to the last decimal, so there is a sole best and worst measure unless the magnitudes of the estimates alone are exactly the same, which is highly unlikely.[7]

Rankings of countries (or sets of countries) with respect to inequality are important social indicators for measuring relative well-being at a point in time and over time. They are also important for assessing the effectiveness of tax and benefit policy, and in comparison to other social and economic rankings such as living standards (GDP/capita), literacy, health status, productivity, hours of market work, net foreign investment, and so on. If we are able to consistently and accurately rank nations according to their level of inequality, we may learn much about groups of countries that share common or uncommon characteristics. Additionally, it may serve to improve our understanding of equity-efficiency tradeoffs, growth and equality relations, and numerous other topics of economic importance.

The problem with previous studies that rank inequality measures is that they tend to ignore the precision of the measures (notably their standard errors). When applying the techniques in the current literature, researchers can estimate the standard errors of each of these indices using resampling methods, such as the bootstrap [6, 27, 28, 40]. These standard errors can be used for hypothesis testing when ranking indices by their relative magnitudes, but typically only single comparisons (between two countries, say) are made at a time.[8] Rank statistics imply *simultaneous*, *multiple* (multivariate) inference procedures, which are typically not employed. Moreover, when multiple single comparisons across countries are made, the overall confidence level of the inferential procedure becomes eroded (i.e., 10 tests at the 95% level have an overall confidence level of less than 95%). Subset selection procedures are both simultaneous and multivariate, so overall confidence levels are preserved.

There is also a debate in the literature about these inference procedures: bootstrapping is the commonly accepted and most preferred procedure [8, 9, 16, 27, 31–42]. However, other asymptotic approaches have also been used [e.g., 1, 7]. Finally, the bootstrapping of

[6] See also Shorrocks [35], which argues for using constant scale factors in assessing rankings of nominal and adjusted income.

[7] Stochastic or Lorenz dominance techniques can also be used to obtain a ranking of measures for inequality by directly comparing their curves, either the cumulative distribution function (cdf) or Lorenz, respectively. Using these techniques, there is more of a chance of a tie at the top or bottom of the ranking as one country may not stochastically or Lorenz dominate the one below it. These techniques also expose ties that can exist in the middle of ranking. A Hasse diagram can be drawn to visualize these relationships [see 3, p45, Figure 4.4). These techniques may not, however, produce a linear, complete, and transitive rank ordering.

[8] For example, see the bootstrap standard errors for LIS key inequality and poverty rate figures at http://www.lisproject.org/keyfigures.htm.

standard errors may perform differently for inequality measures than for poverty measures. See for example, Biewen [6] or Davidson and Flachaire [17]. Since subset selection is based on estimates of standard errors (as we shall see), the accuracy of the procedure may be limited by the accuracy of the standard errors.

3 Data

The data used in this study come from the Luxembourg Income Study (LIS). The LIS database is a collection of household income surveys from various countries. These surveys provide demographic, income, and expenditure information on three different levels: household, person, and child. This study uses the widely accepted data transformations used in the LIS literature. That is, the data are truncated at the top and bottom of the distribution, equivalence scales are implemented, and weights are used. The bottom coding is at 1% of the equivalized mean and the top coding is at ten times the unequivalized median. The equivalence scale used is the square root of the number of persons. The person weight used is the household weight times the number of persons. Double counting of observations is also avoided, as well as missing disposable income and missing weights. Therefore, the observation is the disposable equivalized income of the individual with truncation at the top and bottom of the distribution.[9]

Table 1 displays the countries with their respective years and the sample sizes that are used for the cross-sectional single period analysis. The comparison of inequality measures is based on the latest sample of LIS countries (http://www.lisproject.org/techdoc/datasets.htm.). One observation is used for each of the 29 countries in the sample. These country observations range in years from 1990 to 2002. Their sample sizes range from 2,013 to 49,251 observations. Table 2 displays the countries, years, and sample sizes for the panel analysis on inequality. The panel analysis follows the same set of 12 countries through the four latest LIS waves, which are labeled Waves II through V. This particular set of countries was chosen due to certain criteria, most important of which is that these 12 countries had at least one set of observations in each of the latest four LIS waves. These countries include Australia, Austria, Belgium, Canada, Finland, France, Germany, Ireland, Israel, Italy, Luxembourg, Mexico, Norway, Sweden, Taiwan, United Kingdom, and United States.[10] The country observations range from 1984 to 1987 in Wave II, 1989 to 1992 in Wave III, 1994 to 1998 in Wave IV, and 1999 to 2000 in Wave V.[11]

[9] The LIS definition, while not a complete accounting of income, fits the Canberra Report [11] definition for currently accepted standards of cross-national income distribution measures. The Canberra Report [11] recommends additional research to expand the income definition to include non-cash income and better measures of capital income. Experimental measures of this type are just becoming available in cross-national [21] and national [37] databases. However, the estimates and techniques have not yet been sufficiently tested to be accepted by the international community.

[10] Note the countries that turn out to be the best and worst in a ranking of countries are a function of exactly which countries were included in that ranking. Adding a country or subtracting a country from the set may change the best and worst of the ranking of that set. It should also be noted that Taiwan is technically a region of China (and not a country) according to the United Nations.

[11] The procedure used below only considers the ranking of inequality estimates **within** each wave, so correlation of the inequality measures over time can be ignored. These correlations come from the panel structure of some of the national data sets included in the LIS, in particular the German and Luxembourg cases. In all other cases the wave correlation issue is not problematic. Other data sets include only one panel data wave (Canada) or only short panels where the sample is entirely different after 24 months (United States). The Nordic and Scandinavian nations use samples from register data.

Table 1 Countries, years, and sample sizes for the single period analysis (1990–2002)

Country	LIS Code	Year	n
Australia	AS	1994	6,464
Austria	AT	1997	2,676
Belgium	BE	1997	4,619
Canada	CA	2000	28,970
Czech Republic	CZ	1996	28,131
Denmark	DK	1992	12,829
Estonia	EE	2000	6,062
Finland	FI	2000	10,421
France	FR	1994	11,289
Germany	GE	2000	10,982
Hungary	HU	1999	2,013
Ireland	IE	2000	2,447
Israel	IL	2001	5,787
Italy	IT	2000	7,925
Luxembourg	LX	2000	2,418
Mexico	MX	2002	17,121
Netherlands	NL	1999	4,971
Norway	NW	2000	12,904
Poland	PL	1999	31,375
Romania	RO	1997	32,187
Russia	RL	2000	3,055
Slovak Republic	SK	1996	16,197
Slovenia	SI	1999	3,858
Spain	ES	1990	21,102
Sweden	SW	2000	14,491
Switzerland	CH	1992	6,277
Taiwan	TW	2000	13,801
United Kingdom	UK	1999	24,976
United States	US	2000	49,351

Source: Authors' estimates of LIS data

Note: n=sample size

4 Methodology

4.1 Inequality estimation

For the measurement and panel analyses, the magnitudes and bootstrap standard errors for three inequality indices are calculated by the authors according to the specifications described in the Technical appendix. For the single period analysis, the Gini coefficient, Theil index, and Varlog index are used. The Gini coefficient represents the commonly used inequality measure, the Theil index represents the dispersion measures, and the Varlog index is used as an example of a relatively imprecise inequality measure. Note that these analyses need not be limited to these measures. For the panel analysis, only the Gini coefficient is used.

Suppose there is a country with a given sample of incomes, X, that forms a set of incomes, $X=x_1,\ldots,x_n$, where x_j is the income of an individual and n is the sample size of

Table 2 Countries, years, and sample sizes for the panel analysis by wave

Country	LIS Code	Wave II	n	Wave III	n	Wave IV	n	Wave V	n
Canada	CA	1987	10,987	1991	20,003	1998	31,217	2000	28,970
Finland	FI	1987	11,863	1991	11,748	1995	9,261	2000	10,421
Germany	GE	1984	5,186	1989	4,407	1994	6,374	2000	10,982
Israel	IL	1986	4,997	1992	5,212	1997	5,230	2001	5,787
Italy	IT	1987	8,009	1991	8,175	1995	8,120	2000	7,925
Luxembourg	LX	1985	2,008	1991	1,957	1997	2,514	2000	2,418
Mexico	MX	1984	4,714	1992	10,489	1998	10,889	2000	10,072
Norway	NW	1986	4,969	1991	8,059	1995	10,114	2000	12,904
Sweden	SW	1987	9,516	1992	12,483	1995	16,256	2000	14,491
Taiwan	TW	1986	16,434	1991	16,434	1997	13,701	2000	13,801
United Kingdom	UK	1986	7,174	1991	7,056	1995	6,794	1999	24,976
United States	US	1986	11,577	1991	14,655	1997	50,069	2000	49,351

Source: Authors' estimates of LIS data

Note: n=sample size

individuals within the country. In order to calculate the inequality within this sample, an index must first be calculated. In this study, the inequality index will be denoted with a g, for the Gini coefficient, but will be representative for all of the indices throughout the methodology. The notation can be simplified to $g=g(X)$, where g is now the inequality index as a function of a sample of incomes, X. This inequality calculation can then be applied to more than one sample of incomes, with each sample representing a different country. This yields a set of inequality estimates:

$$g_1, \ldots, g_k \tag{1}$$

where g_i is the inequality estimate of an individual country, and k is the number of countries. Let $G=\{1,...k\}$ be the set of indices for all countries in the sample.

The standard errors for these estimates are then calculated using the bootstrap technique with 100 replications. In what follows this standard error is denoted ω_i. The consistency of the bootstrap standard errors for inequality measures is derived in Biewen [6]. The Monte Carlo study of Davidson and Flachaire [17] shows that the standard bootstrap often fails in inference for income inequality and poverty measures, because it is sensitive to observations (outliers) in the extreme tails of the distribution. Since our data are top and bottom coded, this is less of a factor in our analyses. Biewen [6] conducts a Monte Carlo study where properties of asymptotic and bootstrap confidence intervals are compared based on various sample sizes. Biewen concludes that, for very large datasets, asymptotic and bootstrap methods are very close, but for datasets with as many as 100 observations the 'asymptotic' standard errors have better coverage probability than the standard bootstrap. Simple experiments with unweighted Gini coefficients suggest that the standard errors for our data are approximately equal regardless of the technique used to generate them. However, it is still worth noting that asymptotic standard errors may produce better coverage results in other studies. An empirical comparison between asymptotic and bootstrap standard errors in provided in Mills and Zandvakili [27].

Table 3 Magnitudes, standard errors, and rank for single period analysis

Country	LIS Code	Gini			Theil			Varlog		
		Coef	(s.e.)	Rank	Coef	(s.e.)	Rank	Coef	(s.e.)	Rank
Denmark	DK	0.23647	(0.00267)	1	0.10318	(0.00364)	1	0.32406	(0.01319)	16
Slovak Republic	SK	0.24073	(0.00403)	2	0.10455	(0.00510)	3	0.25540	(0.01074)	5
Finland	FI	0.24742	(0.00268)	3	0.11409	(0.00381)	6	0.20895	(0.00525)	1
Slovenia	SI	0.24942	(0.00417)	4	0.10542	(0.00398)	4	0.24909	(0.00987)	4
Belgium	BE	0.25018	(0.00360)	5	0.10392	(0.00335)	2	0.28164	(0.01803)	9
Norway	NW	0.25077	(0.00338)	6	0.12981	(0.00550)	12	0.27802	(0.00983)	8
Sweden	SW	0.25151	(0.00273)	7	0.11608	(0.00350)	8	0.26779	(0.00903)	7
Netherlands	NL	0.25618	(0.00413)	8	0.11517	(0.00458)	7	0.36694	(0.02009)	18
Czech Republic	CZ	0.25884	(0.00267)	9	0.12050	(0.00356)	9	0.21676	(0.00413)	3
Luxembourg	LX	0.25964	(0.00493)	10	0.11248	(0.00475)	5	0.21387	(0.00782)	2
Germany	GE	0.26360	(0.00299)	11	0.12058	(0.00376)	10	0.28403	(0.01357)	11
Austria	AT	0.26597	(0.00518)	12	0.12228	(0.00619)	11	0.29679	(0.01986)	13
Romania	RO	0.27721	(0.00244)	13	0.14107	(0.00388)	13	0.26018	(0.00377)	6
France	FR	0.28832	(0.00319)	14	0.14849	(0.00474)	14	0.28343	(0.00687)	10
Poland	PL	0.29306	(0.00203)	15	0.15645	(0.00284)	17	0.29299	(0.00463)	12
Hungary	HU	0.29496	(0.00727)	16	0.15496	(0.00897)	16	0.30049	(0.02218)	14
Taiwan	TW	0.29628	(0.00234)	17	0.15138	(0.00308)	15	0.30346	(0.00532)	15
Canada	CA	0.30175	(0.00299)	18	0.16017	(0.00422)	19	0.41126	(0.01167)	20
Spain	ES	0.30308	(0.00265)	19	0.15698	(0.00371)	18	0.34488	(0.00714)	17
Switzerland	CH	0.30705	(0.00472)	20	0.17942	(0.00691)	21	0.89161	(0.05603)	29
Australia	AS	0.31085	(0.00422)	21	0.16289	(0.00515)	20	0.58140	(0.02933)	26
Ireland	IE	0.32326	(0.01082)	22	0.19014	(0.01848)	22	0.38067	(0.02558)	19
Italy	IT	0.33295	(0.00501)	23	0.19635	(0.00721)	23	0.44915	(0.01710)	21
United Kingdom	UK	0.34489	(0.00212)	24	0.21059	(0.00343)	25	0.50051	(0.01022)	23
Israel	IL	0.34641	(0.00407)	25	0.20672	(0.00617)	24	0.46314	(0.02014)	22
Estonia	EE	0.36074	(0.00524)	26	0.22842	(0.00851)	26	0.55423	(0.02082)	24
United States	US	0.36809	(0.00183)	27	0.24350	(0.00297)	27	0.57112	(0.00863)	25
Russia	RL	0.43436	(0.00652)	28	0.33226	(0.01092)	28	0.80985	(0.03334)	27
Mexico	MX	0.49094	(0.00604)	29	0.43442	(0.01192)	29	0.87020	(0.02119)	28

Source: Authors' estimates of LIS data

Note: for formulas of Gini, Theil, and Varlog, see Technical appendix

4.2 Magnitude ranking

The inequality estimates can now be ranked by country according to their respective magnitudes. This is done to gauge the relative ordering of inequality between countries in a given time period. These estimates then form the rank statistic:

$$g_{[1]} \leq \ldots \leq g_{[k-1]} \leq g_{[k]} \tag{2}$$

where the bracketed subscripts represent the rank ordering of inequality estimates. Note that the lowest estimate is at the top of the ranking (most equal) and the largest is at the bottom of the ranking (least equal). Also, this rank ordering will always be linear, complete, and transitive. Linearity means that that the relationships between inequality estimates can always be represented linearly.[12] Completeness means that all of the relationships between the estimates are defined. Transitivity means that if x is better than y, and y is better than z, then x is better than z.

Table 3 presents the magnitude ranking results of the Gini, Theil, and Varlog measures for comparison. The Spearman's rank relation coefficient is calculated for the ranking relationship between each measure. Using the magnitude ranking, the Gini and Theil rankings have a 0.976 correlation measure, which means they are 97.6% correlated in ranking. The Gini and Varlog measures have a 83.1% correlation between magnitude rankings and the Theil and Varlog measures have a 82.3% correlation.

Table 4 shows the magnitude ranking results of the panel for all four waves using the Gini index. The correlation coefficient, Spearman's rank relation coefficient, has been calculated for the magnitude ranking between each of the waves. Moving from Wave II to III, there is a 95.1% rank relation. Moving from Wave III to IV, it is a 98.6% rank relation, and moving from Wave IV to Wave V, we have a 97.2% rank relation. So, the magnitude rankings are not that different mainly because the same measure is used over waves, rather than different measures in one time period as in the single period analysis.

4.3 Multivariate subset selection

Given a pre-specified inferential error rate, $\alpha \in (0,0.5)$, define the following non-empty subsets: $S^{\alpha}_{1B} \subset G$ and $S^{\alpha}_{1W} \subset G$, where S^{α}_{1B} is "the subset of the first-best at confidence level 1-α" and S^{α}_{1W} is "the subset of the first-worst at confidence level 1-α". That is:

$$\Pr\left\{[k] \in S^{\alpha}_{1W}\right\} \geq 1-\alpha \tag{3a}$$

$$\Pr\left\{[1] \in S^{\alpha}_{1B}\right\} \geq 1-\alpha \tag{3b}$$

Equivalently:

- with probability at least 1-α, the subset S^{α}_{1W} contains the indices of the first-worst inequality measures, which means that countries in S^{α}_{1W} are the least equal, or most unequal, countries in terms of income in G,
- with probability at least 1-α, the subset S^{α}_{1B} contains the indices of the first-best inequality measures, which means that the countries in S^{α}_{1B} are the most equal in terms of income in G.

Now, the countries in S^{α}_{1B} and S^{α}_{1W} are removed from the sample, so that we are left with the subset:

$$G^* = G - \left(S^{\alpha}_{1B} \cup S^{\alpha}_{1W}\right) \tag{4}$$

so that $G^* \subset G$ contains indices of all counties that were neither first-best nor first-worst in terms of inequality. Let us assume that G^* is non-empty (this has no effect on what follows). Let the cardinality of G^* be $k^*<k$ so that $G^*=\{1^*,...,k^*\}$, and the ranked inequality measures of the countries in G^* are:

$$g_{[1^*]} \leq \dots \leq g_{[k^*-1]} \leq g_{[k^*]} \tag{5}$$

[12] A Hasse diagram represents an example of a set of relationships which are not necessarily linear.

Table 4 Magnitudes, standard errors, and rank for Gini panel analysis

Country	LIS Code	Wave II			Wave III			Wave IV			Wave V		
		Gini	s.e.	Rank	Gini	s.e.	Rank	Gini	s.e.	Rank	Gini	s.e.	Rank
Finland	FI	0.20856	(0.00173)	1	0.20964	(0.00158)	1	0.21671	(0.00268)	1	0.24742	(0.00268)	1
Sweden	SW	0.21771	(0.00202)	2	0.22912	(0.00230)	2	0.22133	(0.00224)	2	0.25151	(0.00273)	3
Norway	NW	0.23287	(0.00330)	3	0.23124	(0.00388)	3	0.23766	(0.00343)	3	0.25077	(0.00338)	2
Luxembourg	LX	0.23658	(0.00405)	4	0.23895	(0.00613)	4	0.25994	(0.00498)	4	0.25964	(0.00493)	4
Germany	GE	0.26826	(0.00579)	5	0.25739	(0.00576)	5	0.27251	(0.00504)	5	0.26360	(0.00299)	5
Taiwan	TW	0.26850	(0.00209)	6	0.27129	(0.00187)	6	0.29561	(0.00238)	6	0.29628	(0.00234)	6
Canada	CA	0.28286	(0.00363)	7	0.28118	(0.00277)	7	0.30486	(0.00260)	7	0.30175	(0.00299)	7
United Kingdom	UK	0.30321	(0.00316)	8	0.33612	(0.00362)	11	0.34424	(0.00375)	10	0.34489	(0.00212)	9
Israel	IL	0.30762	(0.00325)	9	0.30546	(0.00389)	9	0.33565	(0.00418)	8	0.34641	(0.00407)	10
Italy	IT	0.33193	(0.00504)	10	0.29024	(0.00412)	8	0.33791	(0.00523)	9	0.33295	(0.00501)	8
United States	US	0.33506	(0.00300)	11	0.33581	(0.00256)	10	0.37237	(0.00203)	11	0.36809	(0.00183)	11
Mexico	MX	0.44773	(0.00753)	12	0.48523	(0.00542)	12	0.49364	(0.00503)	12	0.49094	(0.00604)	12

Source: Authors' estimates of LIS data

Define non-empty subsets $S_{2B}^{\alpha} \subset G^*$ and $S_{2W}^{\alpha} \subset G^*$, where S_{2B}^{α} is "the subset of the second-best at confidence level 1-α" and S_{2W}^{α} is "the subset of the second-worst at confidence level 1-α". That is:

$$\Pr\left\{[k^*] \in S_{2W}^{\alpha}\right\} \geq 1-\alpha \tag{6a}$$

$$\Pr\left\{[1^*] \in S_{2B}^{\alpha}\right\} \geq 1-\alpha \tag{6b}$$

Equivalently:

– with probability at least 1-α, the subset S_{2W}^{α} contains the indices of the second-worst inequality measures, which means that countries in S_{2W}^{α} are the least equal, or most unequal, countries in terms of income in G^*,
– with probability at least 1-α, the subset S_{2B}^{α} contains the indices of the second-best inequality measures, which means that the countries in S_{2B}^{α} are the most equal in terms of income in G^*.

If we are willing to assume normality of the income inequality measures g_i (or at least asymptotic normality of the usual functions of g_i), and if we are willing to assume independence of the g_i, then the subsets can be defined as follows:

$$S_{1W}^{\alpha} = \left\{s : g_s - g_j + t_{v_1}^{\alpha}\left(\omega_s^2 + \omega_j^2\right)^{1/2} \geq 0; \forall j \neq s\right\} \tag{7a}$$

$$S_{1B}^{\alpha} = \left\{s : g_j - g_s + t_{v_1}^{\alpha}\left(\omega_s^2 + \omega_j^2\right)^{1/2} \geq 0; \forall j \neq s\right\} \tag{7b}$$

for j and s in G, and

$$S_{2W}^{\alpha} = \left\{s : g_s - g_j + t_{v_2}^{\alpha}\left(\omega_s^2 + \omega_j^2\right)^{1/2} \geq 0; \forall j \neq s\right\} \tag{8a}$$

$$S_{2B}^{\alpha} = \left\{s : g_j - g_s + t_{v_2}^{\alpha}\left(\omega_s^2 + \omega_j^2\right)^{1/2} \geq 0; \forall j \neq s\right\} \tag{8b}$$

for j and s in G^*. The $t_{v_1}^{\alpha}$ is a critical value from (k-1)-dimensional, independent t distribution with v_1 degrees of freedom and diagonal variance matrix with typical elements $\left(\omega_s^2 + \omega_j^2\right)$ for $j \neq s$, such that $\Pr\left\{\max_j t_j \leq t_{v_1}^{\alpha}\right\} = 1-\alpha$.[13] Similarly, the $t_{v_2}^{\alpha}$ is a critical value from (k^*-1)-dimensional, independent t distribution with v_2 degrees of freedom and diagonal variance matrix with typical elements $(\omega_s^2 + \omega_j^2)$ for $j \neq s$, such that $\Pr\{\max_j t_j \leq t_{v_2}^{\alpha}\} = 1-\alpha$. Discussions of these probability integrals can be found in Horrace and Schmidt [23], Horrace and Keane [24], and Horrace [22]. Under an independence assumption, the multi-dimensional probability integrals reduce to one-dimensional integrals that are readily calculable in the *GAUSS* programming language or in *Mathematica* [see 23 for details].

[13] The degrees of freedom are those associated with any estimate of the variance of g_s-g_j. Ultimately, the large sample sizes in our analyses allow us to appeal to an infinite degrees of freedom assumption, and the multivariate t critical values become multivariate normal critical values. The t is used here to make the discussion more general.

Table 5 Subset selection for Gini in single period analysis

Country	LIS Code	Gini			1st Crit Val	1st Subsets	2nd Crit Val	2nd Subsets
		Coef	(s.e.)	Rank	90%	90%	90%	90%
Denmark	DK	0.23647	(0.00267)	1	2.544	1B	*	*
Slovak Republic	SK	0.24073	(0.00403)	2	2.416	1B	*	*
Finland	FI	0.24742	(0.00268)	3	2.543		2.506	2B
Slovenia	SI	0.24942	(0.00417)	4	2.402		2.367	2B
Belgium	BE	0.25018	(0.00360)	5	2.456		2.420	2B
Norway	NW	0.25077	(0.00338)	6	2.477		2.441	2B
Sweden	SW	0.25151	(0.00273)	7	2.538		2.501	2B
Netherlands	NL	0.25618	(0.00413)	8	2.406		2.371	2B
Czech Republic	CZ	0.25884	(0.00267)	9	2.544		2.507	
Luxembourg	LX	0.25964	(0.00493)	10	2.334		2.299	2B
Germany	GE	0.26360	(0.00299)	11	2.514		2.477	
Austria	AT	0.26597	(0.00518)	12	2.312		2.278	
Romania	RO	0.27721	(0.00244)	13	2.565		2.528	
France	FR	0.28832	(0.00319)	14	2.495		2.459	
Poland	PL	0.29306	(0.00203)	15	2.599		2.562	
Hungary	HU	0.29496	(0.00727)	16	2.150		2.119	
Taiwan	TW	0.29628	(0.00234)	17	2.574		2.537	
Canada	CA	0.30175	(0.00299)	18	2.514		2.478	
Spain	ES	0.30308	(0.00265)	19	2.546		2.509	
Switzerland	CH	0.30705	(0.00472)	20	2.352		2.318	
Australia	AS	0.31085	(0.00422)	21	2.398		2.362	
Ireland	IE	0.32326	(0.01082)	22	1.940		1.914	
Italy	IT	0.33295	(0.00501)	23	2.326		2.292	
United Kingdom	UK	0.34489	(0.00212)	24	2.593		2.555	
Israel	IL	0.34641	(0.00407)	25	2.411		2.376	
Estonia	EE	0.36074	(0.00524)	26	2.307		2.273	
United States	US	0.36809	(0.00183)	27	2.615		2.577	
Russia	RL	0.43436	(0.00652)	28	2.204		2.172	2W
Mexico	MX	0.49094	(0.00604)	29	2.241	1W	*	*

Source: Authors' estimates of LIS data

Note: 1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

Consider the statement in Eq. 7a. This equation says, "select country s from G to be in contention for the first-worst (in S^{α}_{1W}), if the difference g_s-g_j is non-negative (≥0) for all $j \neq s$ after adjusting by the statistical tolerance $t^{\alpha}_{v_1}(\omega_s^2 + \omega_j^2)^{1/2}$". That is, designate country s as having high income inequality if its income inequality is consistently larger than all other countries *in a statistical sense* (after adjustment for sampling variability). Similarly, Eq. 7b says, "select country s from G to be in contention for the first-best (in S^{α}_{1B}), if the difference g_j-g_s is non-negative (≥0) for all $j \neq s$ after adjusting by the statistical tolerance." Similar

Table 6 Subset selection for Theil in single period analysis

Country	LIS Code	Theil			1st Crit Val	1st Subsets	2nd Crit Val	2nd Subsets
		Coef	(s.e.)	Rank	90%	90%	90%	90%
Denmark	DK	0.10318	(0.00364)	1	2.542	1B	*	*
Belgium	BE	0.10392	(0.00335)	2	2.562	1B	*	*
Slovak Republic	SK	0.10455	(0.00510)	3	2.440	1B	*	*
Slovenia	SI	0.10542	(0.00398)	4	2.519	1B	*	*
Luxembourg	LX	0.11248	(0.00475)	5	2.465	1B	*	*
Finland	FI	0.11409	(0.00381)	6	2.530	1B	*	*
Netherlands	NL	0.11517	(0.00458)	7	2.477	1B	*	*
Sweden	SW	0.11608	(0.00350)	8	2.551		2.453	2B
Czech Republic	CZ	0.12050	(0.00356)	9	2.547		2.450	2B
Germany	GE	0.12058	(0.00376)	10	2.533		2.437	2B
Austria	AT	0.12228	(0.00619)	11	2.368		2.289	2B
Norway	NW	0.12981	(0.00550)	12	2.414		2.330	2B
Romania	RO	0.14107	(0.00388)	13	2.525		2.430	
France	FR	0.14849	(0.00474)	14	2.465		2.377	
Taiwan	TW	0.15138	(0.00308)	15	2.579		2.477	
Hungary	HU	0.15496	(0.00897)	16	2.209		2.140	
Poland	PL	0.15645	(0.00284)	17	2.594		2.490	
Spain	ES	0.15698	(0.00371)	18	2.537		2.441	
Canada	CA	0.16017	(0.00422)	19	2.502		2.409	
Australia	AS	0.16289	(0.00515)	20	2.437		2.351	
Switzerland	CH	0.17942	(0.00691)	21	2.323		2.247	
Ireland	IE	0.19014	(0.01848)	22	1.856		1.803	
Italy	IT	0.19635	(0.00721)	23	2.305		2.230	
Israel	IL	0.20672	(0.00617)	24	2.369		2.289	
United Kingdom	UK	0.21059	(0.00343)	25	2.556		2.458	
Estonia	EE	0.22842	(0.00851)	26	2.233		2.163	
United States	US	0.24350	(0.00297)	27	2.586		2.483	
Russia	RL	0.33226	(0.01092)	28	2.117		2.053	2W
Mexico	MX	0.43442	(0.01192)	29	2.074	1W	*	*

Source: Authors' estimates of LIS data

Note: 1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

statements are forthcoming for membership in the subset of second-worst (S^{α}_{2W}) and second-best (S^{α}_{2B}) when selection is from G^* and the degrees of freedom are v_2.[14]

The probability statements in Eqs. 3a, 3b and 6a, b are extremely useful. They inform our understanding of the significance of the ranking for each of the inequality measures,

[14] As pointed out by a referee, the second-best and -worst subsets do not *precisely* control for the overall error rate of the exercise, since they do not take into consideration the error rate associated with the first-best and -worst subsets. They are, however, a reasonable approximation.

Table 7 Subset selection for Varlog in single period analysis

Country	LIS Code	Varlog Coef	(s.e.)	Rank	1st Crit Val 90%	1st Subsets 90%	2nd Crit Val 90%	2nd Subsets 90%
Finland	FI	0.20895	(0.00525)	1	2.622	1B	*	*
Luxembourg	LX	0.21387	(0.00782)	2	2.566	1B	*	*
Czech Republic	CZ	0.21676	(0.00413)	3	2.643	1B	*	*
Slovenia	SI	0.24909	(0.00987)	4	2.518		2.443	2B
Slovak Republic	SK	0.25540	(0.01074)	5	2.498		2.422	2B
Romania	RO	0.26018	(0.00377)	6	2.649		2.571	2B
Sweden	SW	0.26779	(0.00903)	7	2.538		2.464	2B
Norway	NW	0.27802	(0.00983)	8	2.519		2.445	2B
Belgium	BE	0.28164	(0.01803)	9	2.342		2.251	2B
France	FR	0.28343	(0.00687)	10	2.587		2.514	
Germany	GE	0.28403	(0.01357)	11	2.435		2.354	2B
Poland	PL	0.29299	(0.00463)	12	2.634		2.558	
Austria	AT	0.29679	(0.01986)	13	2.307		2.212	2B
Hungary	HU	0.30049	(0.02218)	14	2.264		2.164	2B
Taiwan	TW	0.30346	(0.00532)	15	2.621		2.546	
Denmark	DK	0.32406	(0.01319)	16	2.443		2.363	
Spain	ES	0.34488	(0.00714)	17	2.581		2.508	
Netherlands	NL	0.36694	(0.02009)	18	2.303		2.207	
Ireland	IE	0.38067	(0.02558)	19	2.205		2.098	
Canada	CA	0.41126	(0.01167)	20	2.477		2.400	
Italy	IT	0.44915	(0.01710)	21	2.361		2.272	
Israel	IL	0.46314	(0.02014)	22	2.302		2.206	
United Kingdom	UK	0.50051	(0.01022)	23	2.510		2.435	
Estonia	EE	0.55423	(0.02082)	24	2.289		2.192	2W
United States	US	0.57112	(0.00863)	25	2.547		2.473	2W
Australia	AS	0.58140	(0.02933)	26	2.145		2.031	2W
Russia	RL	0.80985	(0.03334)	27	2.086	1W	*	*
Mexico	MX	0.87020	(0.02119)	28	2.282	1W	*	*
Switzerland	CH	0.89161	(0.05603)	29	1.834	1W	*	*

Source: Authors' estimates of LIS data

Note: 1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

while accounting for sampling variability captured in the bootstrapped standard errors. Notice that the probability statement $\Pr\left\{\max_j t_j \le t_{v_1}^{\alpha}\right\} = 1-\alpha$ implies that $t_{v_1}^{\alpha}$ is decreasing in α. That is, $t_{v_1}^{.05}$ is greater than $t_{v_1}^{.10}$. Therefore, as our inferential confidence level gets larger (α gets smaller), the statistical tolerance of the probability statements, $t_{v_1}^{\alpha}\left(\omega_s^2+\omega_j^2\right)^{1/2}$, gets larger (as one would expect). Consequently, as the confidence level increases, $g_s - g_j + t_{v_1}^{\alpha}\left(\omega_s^2+\omega_j^2\right)^{1/2}$ is more likely to be positive for each s, and, therefore, the cardinality of S_{1W}^{α} will increase; there will be more countries in contention for the first-worst at higher confidence levels. Therefore, at higher confidence levels the inference will be "less sharp" in

Table 8 Subsets of the 1st best, 2nd best, 2nd worst, and 1st worst for measures at 90% confidence level

	Gini	Theil	Varlog
1st Best	Denmark Slovak Republic	Belgium Denmark Finland Luxembourg Netherlands Slovak Republic Slovenia	Czech Republic Finland Luxembourg
2nd Best	Belgium Finland Luxembourg Netherlands Norway Slovenia Sweden	Austria Czech Republic Germany Norway Sweden	Austria Belgium Germany Hungary Norway Romania Slovak Republic Slovenia Sweden
2nd Worst	Russia	Russia	Australia Estonia United States
1st Worst	Mexico	Mexico	Mexico Russia Switzerland

Source: Authors' estimates of LIS data

Note: This table is a summary of the results of Tables 5, 6, 7

the sense that we cannot differentiate bad countries from good countries with a high probability. At a low enough confidence level, S^{α}_{1W} will reduce to a singleton, so that a single country can be designated as first-worst in income inequality at the 1-α level (a lower probability). A similar relationship holds for a subset of the first-best; as the confidence level increases more countries will be in contention for the first-best, and as the confidence level decreases, S^{α}_{1B} will reduce to a singleton, so that a single country can be designated as first-best in income inequality at the 1-α level.

5 Empirical results

5.1 Single period analysis

The subset selection results for each measure are shown in Tables 5, 6 and 7 and then summarized in Table 8. Table 5 shows the results for the Gini index. The Gini coefficients range from the most equal with a Gini of 0.23647 (Denmark) to the least equal with a Gini of 0.49094 (Mexico), while the bootstrap standard errors range from the most precise of 0.00183 (United States) to the least precise of 0.01082 (Ireland). The first subset selection produces a first-best subset of two countries (Denmark and Slovak Republic, denoted "1B" in the table) and a first-worst subset of one country (Mexico, denoted "1W" in the table). These three countries are then dropped and the second subset selection procedure is performed. The second subset selection procedure produces a large second-best subset of

Table 9 Subset selection results of Gini rankings for Wave II

Country	LIS Code	Gini			1st Crit Val	1st SS	2nd Crit Val	2nd SS	1st Crit Val	1st SS	2nd Crit Val	2nd SS
		Coef	(s.e.)	Rank	90%	90%	90%	90%	95%	95%	95%	95%
Finland	FI	0.20856	(0.00173)	1	2.302	1B	*	*	2.574	1B	*	*
Sweden	SW	0.21771	(0.00202)	2	2.283		2.215	2B	2.561		2.497	2B
Norway	NW	0.23287	(0.00330)	3	2.183		2.116		2.482		2.421	
Luxembourg	LX	0.23658	(0.00405)	4	2.120		2.052		2.429		2.367	
Germany	GE	0.26826	(0.00579)	5	1.985		1.912		2.312		2.246	
Taiwan	TW	0.26850	(0.00209)	6	2.278		2.210		2.557		2.493	
Canada	CA	0.28286	(0.00363)	7	2.155		2.088		2.459		2.398	
United Kingdom	UK	0.30321	(0.00316)	8	2.194		2.128		2.491		2.430	
Israel	IL	0.30762	(0.00325)	9	2.187		2.120		2.485		2.424	
Italy	IT	0.33193	(0.00504)	10	2.041		1.970	2W	2.361		2.297	2W
United States	US	0.33506	(0.00300)	11	2.208		2.142	2W	2.502		2.441	2W
Mexico	MX	0.44773	(0.00753)	12	1.869	1W	*	*	2.209	1W	*	*

Source: Authors' estimates of LIS data

Note: SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

seven countries (denoted "2B" in the table) and a sole country in the second-worst subset (Russia denoted "2W" in the table).

The salient feature of the selection procedure is that the first-best and second-best subsets contain more than one country, because the differences in magnitudes of the coefficients are relatively small. When precision of the estimators is taken into account, these small differences become indistinguishable. Also, at the bottom of the ranking, the first-worst and second-worst subsets only contain one country due to the large differences in magnitudes at the bottom of the rank order. Another interesting point is that the second-best subset includes Luxembourg but does not include the Czech Republic, even though Luxembourg (ranked 10th) is lower in the ranking than the Czech Republic (ranked ninth). This is due to the relative precision of the Czech Republic coefficient, shown by its smaller bootstrap standard error (0.00267), as compared with the ranking of its neighbors, the Netherlands (0.00413) and Luxembourg (0.00493). Therefore, with 90% probability, we can say that the Czech Republic is not in the second-best subset while Luxembourg is.

Table 6 shows the single period analysis for the Theil index. The coefficients range from the most equal of 0.10318 (Denmark) to the least equal of 0.43442 (Mexico), while the bootstrap standard errors range from the most precise of 0.00284 (Poland) to the least precise of 0.01848 (Ireland). The first subset selection procedure produces a first-best subset of seven countries with no breaks and a first-worst subset of one country. A 'break' is defined as a country that is excluded from a subset of its nearest neighbors in the rank order. After discarding those subsets, the second subset selection procedure finds five

Table 10 Subset selection results of Gini rankings for Wave III

Country	LIS Code	Gini Coef	Gini (s.e.)	Gini Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Finland	FI	0.20964	(0.00158)	1	2.309	1B	*	*	2.579	1B	*	*
Sweden	SW	0.22912	(0.00230)	2	2.262		2.199	2B	2.545		2.485	2B
Norway	NW	0.23124	(0.00388)	3	2.133		2.077	2B	2.440		2.388	2B
Luxembourg	LX	0.23895	(0.00613)	4	1.958		1.906	2B	2.289		2.241	2B
Germany	GE	0.25739	(0.00576)	5	1.985		1.933		2.313		2.264	
Taiwan	TW	0.27129	(0.00187)	6	2.292		2.226		2.567		2.504	
Canada	CA	0.28118	(0.00277)	7	2.225		2.165		2.516		2.459	
Italy	IT	0.29024	(0.00412)	8	2.113		2.058		2.424		2.372	
Israel	IL	0.30546	(0.00389)	9	2.133		2.077		2.440		2.387	
United States	US	0.33581	(0.00256)	10	2.242		2.181	2W	2.529		2.471	2W
United Kingdom	UK	0.33612	(0.00362)	11	2.155		2.098	2W	2.459		2.405	2W
Mexico	MX	0.48523	(0.00542)	12	2.010	1W	*	*	2.335	1W	*	*

Source: Authors' estimates of LIS data

Note: SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

countries in the second-best subset and one country in the second-worst subset. Again, the top of the rank order is much closer in magnitudes than the bottom of the rank order.

Table 7 shows the results for the Varlog measure. Note that the Varlog index is the least precise measure of inequality used in this study. It is included to show how the subset selection method produces vastly different results depending on the characteristics of the measure. The Varlog coefficients range from the most equal of 0.20895 (Finland) to the least equal of 0.89161 (Switzerland). The bootstrap standard errors range from the most precise of 0.00377 (Romania) to the least precise at 0.05603 (Switzerland). The first subset selection procedure produces three countries in the first-best subset and also three in the first-worst subset. The second subset selection procedure produces nine countries in the second-best and three countries in the second-worst. Note that with the Varlog measure, ties are produced at the bottom of the ranking, unlike the other two measures. The second subset is most interesting. According to the magnitude ranking, both France and Poland should be included in the second-best subset, but they are not. This is due to the relative precision of their estimates compared with their neighbors in the ranking. So, it can be said, with 90% confidence, that these two countries do not belong in the second-best subset, whereas that cannot be said for Germany, Austria, or Hungary.

Table 8 summarizes the countries in the first-best, second-best, first-worst, and second-worst subsets for each of the three inequality measures. Note that Denmark is in the first-best subset for both the Gini and Theil, but not in either the first-best or second-best for the

Table 11 Subset selection results of Gini rankings for Wave IV

Country	LIS Code	Gini Coef	(s.e.)	Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Finland	FI	0.21671	(0.00268)	1	2.238	1B	*	*	2.527	1B	*	*
Sweden	SW	0.22133	(0.00224)	2	2.272	1B	*	*	2.553	1B	*	*
Norway	NW	0.23766	(0.00343)	3	2.175		2.084	2B	2.476		2.389	2B
Luxembourg	LX	0.25994	(0.00498)	4	2.043		1.999		2.365		2.289	
Germany	GE	0.27251	(0.00504)	5	2.038		1.999		2.361		2.286	
Taiwan	TW	0.29561	(0.00238)	6	2.261		2.158		2.545		2.446	
Canada	CA	0.30486	(0.00260)	7	2.244		2.144		2.532		2.436	
Israel	IL	0.33565	(0.00418)	8	2.110		2.025		2.422		2.341	
Italy	IT	0.33791	(0.00523)	9	2.023		1.999		2.347		2.273	
United Kingdom	UK	0.34424	(0.00375)	10	2.147		2.059		2.453		2.369	
United States	US	0.37237	(0.00203)	11	2.286		2.179	2W	2.563		2.461	2W
Mexico	MX	0.49364	(0.00503)	12	2.039	1W	*	*	2.361	1W	*	*

Source: Authors' estimates of LIS data

Note: SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

Varlog. Note that Finland is in the first-best subset for both the Theil and Varlog, but is only in the second-best subset using the Gini. Also, examining the subsets of the worst, Russia is in the second-worst subset for the Gini and Theil, but moves into the subset of the first-worst, along with Mexico and Switzerland, using the Varlog.

To conclude the single period analysis, the subset selection procedure shows that different inequality measures produce differing magnitude rankings as shown by the Spearman's rank relation coefficient. In addition, when the precision of these estimates is taken into consideration, as given by the bootstrap standard error, we can say with high probability which countries are no different from one another at the top and bottom of the rankings. Sometimes this allows countries to be excluded in the best and worst subsets ('breaks'), even though they may look similar given the magnitude of their coefficients.

5.2 Panel analysis

The subset selection results are shown for the Gini measure for each LIS Wave in Tables 9, 10, 11 and 12 presents the Gini subset selection results for Wave II of the LIS data. The magnitudes range from 0.20856 (Finland) to 0.44773 (Mexico). The standard errors range from 0.00173 (Finland) to 0.00753 (Mexico). The first-best is Finland, with the second-best being Sweden. The first-worst is Mexico, with the second-worst being Italy and the United States. The rest of the countries are contained in the middle subset. Note that using critical values at either the 90 or 95% confidence level, yield the same subset results.

Table 12 Subset selection results of Gini rankings for Wave V

Country	LIS Code	Gini Coef	Gini (s.e.)	Gini Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Finland	FI	0.24742	(0.00268)	1	2.222	1B	*	*	2.514	1B	*	*
Norway	NW	0.25077	(0.00338)	2	2.158	1B	*	*	2.462	1B	*	*
Sweden	SW	0.25151	(0.00273)	3	2.217	1B	*	*	2.510	1B	*	*
Luxembourg	LX	0.25964	(0.00493)	4	2.020		1.878	2B	2.344	1B	*	*
Germany	GE	0.26360	(0.00299)	5	2.194		2.038	2B	2.492		2.283	2B
Taiwan	TW	0.29628	(0.00234)	6	2.251		2.090		2.537		2.327	
Canada	CA	0.30175	(0.00299)	7	2.194		2.038		2.492		2.283	
Italy	IT	0.33295	(0.00501)	8	2.013		1.871		2.338		2.138	
United Kingdom	UK	0.34489	(0.00212)	9	2.269		2.106		2.551		2.340	
Israel	IL	0.34641	(0.00407)	10	2.095		1.947		2.409		2.204	
United States	US	0.36809	(0.00183)	11	2.291		2.125	2W	2.566		2.355	2W
Mexico	MX	0.49094	(0.00604)	12	1.932	1W	*	*	2.266	1W	*	*

Source: Authors' estimates of LIS data

Note: SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

Table 10 presents the Gini subset results for Wave III. The magnitudes range from 0.20964 (Finland) to 0.48523 (Mexico). The standard errors range from 0.00158 (Finland) to 0.00613 (Luxembourg). The first-best is again Finland, with the second-best being Sweden, Norway, and Luxembourg. The first-worst is again Mexico, with the second-worst being the United States and United Kingdom. These results are the same using either level of confidence (0.90 or 0.95).

Table 11 presents the Gini results for Wave IV. The magnitudes range from 0.21671 (Finland) to 0.49364 (Mexico). The standard errors range from the most precise 0.00203 (United States) to the least precise 0.00523 (Italy). The first-best is Finland and Sweden, with the second-best being Norway. The first-worst is Mexico, with the second-worst being the United States. Again, the results for both critical values are the same.

Table 12 presents the Gini subset selection results for Wave V. The magnitudes range from 0.24742 (Finland) to 0.49094 (Mexico). The standard errors range from 0.00183 (United States) to 0.00604 (Mexico). The first-worst is Mexico and second-worst is the United States at both critical values (0.90 and 0.95). However, the first-best and second-best subsets differ by critical value. At the 90% level, the first-best is Finland, Norway, and Sweden, while the second-best is Luxembourg and Germany. At the 95% level, the first-best is Finland, Norway, Sweden, and Luxembourg, with the second-best being Germany, with Luxembourg moving between subsets according to the level of confidence.

Table 13 brings the subset selection results for the four waves of Gini measures together in one table, in order to compare which countries are moving in and out of the subsets over

Table 13 Subsets of the 1st best, 2nd best, 2nd worst, and 1st worst for Gini at 90% confidence level

	Wave II	Wave III	Wave IV	Wave V	
				(at 90%)	(at 95%)
1st Best	Finland	Finland	Finland Sweden	Finland Norway Sweden	Finland Luxembourg Norway Sweden
2nd Best	Sweden	Luxembourg Norway Sweden	Norway	Germany Luxembourg	Germany
2nd Worst	Italy United States	United Kingdom United States	United States	United States	United States
2nd Worst	Mexico	Mexico	Mexico	Mexico	Mexico

Source: Authors' estimates of LIS data

Note: This table is a summary of the results of Tables 9, 10, 11, and 12

Different critical values only reported when the 90% and 95% differ

time. In this table, Finland is always in the subset of the first-best and Mexico is always in the subset of the first-worst. However, Finland is joined in the subset of the first-best by more countries over time, which would lead one to conclude that countries are catching up to Finland in equality (in a relative sense). It could also be that Finland is moving down, however. For instance, Sweden is contained in the second-best subset for the first two waves, but moves up to the first-best subset in the last two waves. At the bottom, the United States is consistently contained in the second-worst subset over the four waves, but it only stands alone in the last two waves. In the first wave it is joined by Italy, and then in the second by the United Kingdom.

Table 14 Magnitudes, standard errors, and rank for relative poverty panel analysis

Country	LIS Code	Wave II			Wave III			Wave IV			Wave V		
		RelPov	(s.e.)	Rank	RelPov	(s.e.)	Rank	RelPov	(s.e.)	Rank	RelPov	(s.e.)	Rank
Taiwan	TW	5.2	(0.20)	1	6.5	(0.22)	5	9.1	(0.26)	6	9.1	(0.27)	6
Luxembourg	LX	5.3	(0.55)	2	4.7	(0.71)	1	6.2	(0.61)	2	6.0	(0.78)	2
Finland	FI	5.4	(0.23)	3	5.7	(0.23)	2	4.2	(0.23)	1	5.4	(0.35)	1
Norway	NW	7.2	(0.37)	4	6.4	(0.43)	4	6.9	(0.30)	4	6.4	(0.24)	3
Sweden	SW	7.5	(0.33)	5	6.7	(0.27)	6	6.6	(0.21)	3	6.5	(0.22)	4
Germany	GE	7.9	(0.47)	6	5.8	(0.45)	3	8.2	(0.43)	5	8.3	(0.35)	5
United Kingdom	UK	9.1	(0.46)	7	14.6	(0.59)	10	13.4	(0.53)	8	12.5	(0.29)	8
Italy	IT	11.2	(0.61)	8	10.4	(0.47)	8	14.1	(0.60)	10	12.7	(0.59)	9
Canada	CA	11.4	(0.45)	9	11.0	(0.34)	9	12.8	(0.29)	7	11.4	(0.29)	7
Israel	IL	11.7	(0.59)	10	10.2	(0.57)	7	13.5	(0.61)	9	15.6	(0.63)	10
United States	US	17.8	(0.43)	11	17.5	(0.36)	11	16.9	(0.19)	11	17.0	(0.22)	11
Mexico	MX	20.8	(0.89)	12	20.6	(0.64)	12	22.1	(0.49)	12	21.6	(0.63)	12

Source: Authors' estimates of LIS data

Note: Relative Poverty Rate is 50% of median of the total population for all LIS Waves

Table 15 Subset selection results of relative poverty rankings for Wave II

Country	LIS Code	Rel Pov Coef	(s.e.)	Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Taiwan	TW	5.2	(0.20)	1	2.311	1B	*	*	2.580	1B	*	*
Luxembourg	LX	5.3	(0.55)	2	2.097	1B	*	*	2.411	1B	*	*
Finland	FI	5.4	(0.23)	3	2.298	1B	*	*	2.571	1B	*	*
Norway	NW	7.2	(0.37)	4	2.217		2.079	2B	2.510		2.379	2B
Sweden	SW	7.5	(0.33)	5	2.242		2.100	2B	2.530		2.395	2B
Germany	GE	7.9	(0.47)	6	2.151		2.020	2B	2.456		2.332	2B
United Kingdom	UK	9.1	(0.46)	7	2.157		2.026		2.461		2.337	
Italy	IT	11.2	(0.61)	8	2.059		1.933		2.378		2.260	
Canada	CA	11.4	(0.45)	9	2.164		2.032		2.467		2.342	
Israel	IL	11.7	(0.59)	10	2.071		1.946		2.389		2.271	
United States	US	17.8	(0.43)	11	2.177		2.044	2W	2.478		2.352	2W
Mexico	MX	20.8	(0.89)	12	1.896	1W	*	*	2.235	1W	*	*

Source: Authors' estimates of LIS data

Note: Relative Poverty Rate is 50% of median of the total population for all LIS Waves

SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

To conclude, the panel analysis for inequality shows that looking at magnitude alone tells us only a partial story, especially in movements across waves. Because the concept of relative inequality is important, we can say with high probability which countries are in the top and bottom of the ranking, and also which are the 'runners-up' to those top and bottom subsets. This gives researchers of inequality a first look at how cutoffs in relative movements can be established by the precision of the estimators rather than by arbitrary magnitude cutoff rules.

6 Extension to poverty

For the relative poverty panel analysis, the relative poverty measure is 50% of the median for the total population. The magnitudes of this estimator come directly from LIS Key Figures, along with their respective bootstrap standard errors. It is best to use these widely accepted measures, which enables us to compare our results to previous LIS studies that use similar measures. There is an issue here of whether poverty is relative or absolute in nature, though this mainly depends on how you are viewing it. In this study, international comparisons of poverty are used, so poverty is relative in nature.

Table 14 presents the magnitude ranking results for the 12 countries over the four waves according to the relative poverty measure. The Spearman's rank relation coefficient is again calculated. The relative poverty rankings are shown to change much more from wave to

Table 16 Subset selection results of relative poverty rankings for Wave III

Country	LIS Code	Rel Pov			1st Crit Val	1st SS	2nd Crit Val	2nd SS	1st Crit Val	1st SS	2nd Crit Val	2nd SS
		Coef	(s.e.)	Rank	90%	90%	90%	90%	95%	95%	95%	95%
Luxembourg	LX	4.7	(0.71)	1	1.975	1B	*	*	2.304	1B	*	*
Finland	FI	5.7	(0.23)	2	2.292	1B	*	*	2.567	1B	*	*
Germany	GE	5.8	(0.45)	3	2.148	1B	*	*	2.454	1B	*	*
Norway	NW	6.4	(0.43)	4	2.162	1B	*	*	2.466	1B	*	*
Taiwan	TW	6.5	(0.22)	5	2.298		2.071	2B	2.571	1B	*	*
Sweden	SW	6.7	(0.27)	6	2.270		2.047	2B	2.551		2.290	2B
Israel	IL	10.2	(0.57)	7	2.065		1.857		2.383		2.160	
Italy	IT	10.4	(0.47)	8	2.134		1.921		2.442		2.209	
Canada	CA	11.0	(0.34)	9	2.225		2.006		2.516		2.266	
United Kingdom	UK	14.6	(0.59)	10	2.052		1.844		2.372		2.150	
United States	US	17.5	(0.36)	11	2.211		1.993	2W	2.505		2.258	2W
Mexico	MX	20.6	(0.64)	12	2.019	1W	*	*	2.343	1W	*	*

Source: Authors' estimates of LIS data

Note: Relative Poverty Rate is 50% of median of the total population for all LIS Waves

SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

wave than the inequality results for the panel. From Wave II to III, the Spearman's rank relation coefficient is 0.839. From Wave III to IV, it is 0.888. From Wave IV to V, it is 0.986. We can see that the rankings differ more in the first waves than in the latter waves.

The subset selection technique is again applied to the panel data. The results for relative poverty are contained in Tables 15, 16, 17 and 18. Table 15 shows the subset results for Wave II. The magnitudes range from 5.2 (Taiwan) to 20.8 (Mexico). The standard errors range from 0.20 (Taiwan) to 0.89 (Mexico). The first-best subset contains three countries: Taiwan, Luxembourg, and Finland. The second-best subset contains Norway, Sweden, and Germany. The first-worst subset contains Mexico, with the second-worst being the United States. These results are the same at both the 90 and 95% levels of confidence.

Table 16 contains the relative poverty results for Wave III. The magnitudes range from 4.7 (Luxembourg) to 20.6 (Mexico). The standard errors range from the most precise 0.22 (Taiwan) to the least precise 0.71 (Luxembourg). The subset of the first-worst contains Mexico, while the second-worst contains the United States. This is true at either the 90 or 95% confidence level. However, the subsets of the first-best differ at the two confidence levels. At the 90% level, the first-best subset contains Luxembourg, Finland, Germany, and Norway, whereas at the 95% level, Taiwan is also included among the first-best. Taiwan's poverty measure is estimated with high precision (standard error of 0.22). At the 90% level the smaller critical value (2.298) combines with the small standard error to allow us to infer with at least 90% probability that Taiwan is not really first-best. However, the larger critical

Table 17 Subset selection results of relative poverty rankings for Wave IV

Country	LIS Code	Rel Pov Coef	(s.e.)	Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Finland	FI	4.2	(0.23)	1	2.273	1B	*	*	2.553	1B	*	*
Luxembourg	LX	6.2	(0.61)	2	1.994		1.946	2B	2.320		2.274	2B
Sweden	SW	6.6	(0.21)	3	2.286		2.218	2B	2.563		2.499	2B
Norway	NW	6.9	(0.30)	4	2.223		2.161	2B	2.514		2.455	2B
Germany	GE	8.2	(0.43)	5	2.124		2.069		2.433		2.379	
Taiwan	TW	9.1	(0.26)	6	2.253		2.188		2.537		2.476	
Canada	CA	12.8	(0.29)	7	2.231		2.168		2.520		2.460	
United Kingdom	UK	13.4	(0.53)	8	2.050		1.999		2.369		2.320	
Israel	IL	13.5	(0.61)	9	1.994		1.946		2.320		2.274	
Italy	IT	14.1	(0.60)	10	2.001		1.952		2.326		2.280	
United States	US	16.9	(0.19)	11	2.298		2.229	2W	2.571		2.506	2W
Mexico	MX	22.1	(0.49)	12	2.079	1W	*	*	2.394	1W	*	*

Source: Authors' estimates of LIS data

Note: Relative Poverty Rate is 50% of median of the total population for all LIS Waves

SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

value (2.571) at the 95% confidence level makes Taiwan statistically indistinguishable from the other countries in the first-best subset, even though its measure is estimated with very high precision. This is an interesting result.

Table 17 shows the analysis on Wave IV of the LIS panel data. The magnitudes range from 4.2 (Finland) to 22.1 (Mexico). The standard errors range from 0.19 (United States) to 0.61 (Luxembourg). The first-best is Finland. The second-best subset contains Luxembourg, Sweden, and Norway. The first-worst is Mexico and the second-worst is the United States.

Table 18 shows the subset selection results for Wave V. The magnitudes range from 5.4 (Finland) to 21.6 (Mexico). The standard errors range from 0.22 (Sweden) to 0.78 (Luxembourg). At the 90% confidence level, Finland and Luxembourg are contained in the first-best subset, whereas at the 95% confidence level, Norway is also included with Finland and Luxembourg. At the 90% confidence level, Norway and Sweden are included in the second-best subset, whereas at the 95% confidence level, only Sweden is in the second-best subset. The first-worst is Mexico, with the second-worst being the United States, and this is true at both confidence levels.

Table 19 displays all subset results for the relative poverty panel analysis. It is useful to compare the poverty results in Table 19 to the Gini results in Table 13. It can be seen that the countries that tend to be the worst in terms of inequality (Gini) also tend to be the worst in terms of poverty, and these results are fairly consistent over time (i.e., Mexico and the United States are always first-worst and second-worst, respectively, and they are almost

Table 18 Subset selection results of relative poverty rankings for Wave V

Country	LIS Code	Rel Pov Coef	(s.e.)	Rank	1st Crit Val 90%	1st SS 90%	2nd Crit Val 90%	2nd SS 90%	1st Crit Val 95%	1st SS 95%	2nd Crit Val 95%	2nd SS 95%
Finland	FI	5.4	(0.35)	1	2.185	1B	*	*	2.483	1B	*	*
Luxembourg	LX	6.0	(0.78)	2	1.894	1B	*	*	2.230	1B	*	*
Norway	NW	6.4	(0.24)	3	2.267		2.136	2B	2.548	1B	*	*
Sweden	SW	6.5	(0.22)	4	2.280		2.151	2B	2.558		2.401	2B
Germany	GE	8.3	(0.35)	5	2.185		2.044		2.483		2.324	
Taiwan	TW	9.1	(0.27)	6	2.246		2.112		2.532		2.374	
Canada	CA	11.4	(0.29)	7	2.231		2.095		2.520		2.362	
United Kingdom	UK	12.5	(0.29)	8	2.231		2.095		2.520		2.362	
Italy	IT	12.7	(0.59)	9	2.010		1.856		2.334		2.172	
Israel	IL	15.6	(0.63)	10	1.984		1.829		2.311		2.149	2W
United States	US	17.0	(0.22)	11	2.280		2.151	2W	2.558		2.401	2W
Mexico	MX	21.6	(0.63)	12	1.984	1W	*	*	2.311	1W	*	*

Source: Authors' estimates of LIS data

Note: Relative Poverty Rate is 50% of median of the total population for all LIS Waves

SS=subset

Crit Val=critical value

1B=country is in the first-best subset of Eq. 3b

1W=country is in the first-worst subset of Eq. 3a

2B=country is in the second-best subset of Eq. 6b

2W=country is in the second-worst subset of Eq. 6a

always the sole-possessors of this distinction over time). Finland is always in the subset of first-best. For inequality (Gini in Table 13), the cardinality of the first-best in monotonically non-decreasing in time. Moreover, once a country enters the first-best subset, it stays there over time. The results are less consistent for the poverty measure in Table 19. While Finland is always in the first-best subset of poverty, it is only in sole-possession of this distinction in Wave IV. In preceding and successive waves it is not alone. In particular, Luxembourg is in the first-best subset for relative poverty in all waves *except* Wave IV.

Why does this inconsistency exist in the first-best poverty rankings over time, but not in the first-best inequality rankings? This is not a simple question to answer. Because there are many features of the inference that are simultaneously changing over waves, over countries, and over measures, exact ceteris paribus comparisons are not possible. That being said, it *does* appear that the critical values at any particular confidence level remain relatively constant across countries, waves, and measures. For example, compare the values in the "1st Crit Val 90%" column in Tables 9, 10, 11, 12 and Tables 15, 16, 17, 18; they are all approximately the same. Therefore, most of the difference in the dynamics of the measures is probably due to either differences in the magnitudes of each measure across countries or differences in the bootstrapped standard errors of each measure across countries. This latter possibility may be linked to the arguments made by Davidson and Flachaire [17] who suggest that bootstrapped standard errors have different levels of accuracy for inequality measures than for poverty measures. Perhaps this is what is driving the different results across the Gini and the relative poverty measures.

Table 19 Subsets of the 1st best, 2nd best, 2nd worst, and 1st worst for relative poverty at 90% confidence level

	Wave II	Wave III		Wave IV	Wave V	
		(at 90%)	(at 95%)		(at 90%)	(at 95%)
1st Best	Finland Luxembourg Taiwan	Finland Germany Luxembourg Norway	Finland Germany Luxembourg Norway Taiwan	Finland	Finland Luxembourg	Finland Luxembourg Norway
2nd Best	Germany Norway Sweden	Sweden Taiwan	Sweden	Luxembourg Norway Sweden	Norway Sweden	Sweden
2nd Worst	United States	United States	United States	United States	United States	Israel United States
1st Worst	Mexico	Mexico	Mexico	Mexico	Mexico	Mexico

Source: Authors' estimates of LIS data

Note: This table is a summary of the results of Tables 15, 16, 17, 18

Different critical values only reported when the 90% and 95% differ

7 Conclusions

This study has applied a subset selection procedure to the analysis of rank statistics in an income and poverty study. For the single period analysis, we have shown that precision matters when ranking different estimators and that estimators differ in their ranking interpretation under different selection procedures. For the panel analysis, using a subset selection procedure improves our understanding of the relative movement of countries in and out of various inequality ranks for a given level of confidence. If lowering inequality or poverty is of interest to policy makers, then understanding which set of countries is performing the best (in a statistical sense) is obviously important. New policies can be fashioned after those of countries that are performing particularly well.

The multivariate subset selection procedures do have empirical limitations that are worth noting. The rankings of the countries implied by the inference will be a function of both the selected index (i.e., Varlog, Gini, or Theil) and the selected error rate, α. In the former instance, reporting ranks of several indices may produce subsets with very little overlap across the different measures. The usefulness of these procedures may be called into question. In the latter instance, it is standard practice in the economics literature to look for significance at the 95% level and then, failing this, the 90% level. This provides guidance for selecting the error rate for these procedures: report two sets of subsets, one at α=0.05 and another at α=0.10, and compare the results. In either case, less restrictive methods based on stochastic dominance may be preferred.

There may be other unexplored applications of subset selection procedures in the inequality literature. For example, subset selection may be applied to a single country in the LIS data that has multiple observations across the waves. Subset selection would be applied here to the set of *years* of a given country, in order to see which years were best and worst and how the inequality situation of the country has changed or not changed over time. Because inequality does not change much over time in a given country, however, there may not be enough variation to produce interesting results (i.e., the years are simultaneously indistinguishable in a statistical sense). Another potential problem is that the estimate of inequality in 1 year may be correlated with the same measure in subsequent years. The selection procedure described herein assumes zero correlation across measures, and this would need to be incorporated into the new procedure. This is not necessarily difficult to do as long as consistent estimates of the correlations are available. How one would estimate these correlations remains to be seen.

Another potential application is to compare the subsets of this procedure with the subsets produced by the techniques of stochastic or Lorenz dominance. The ranking of the scalar measures produces a linear ranking that leaves one country at the top of the rank ordering and one country at the bottom (when the coefficients are not rounded). When precision is accounted for with a subset selection procedure, it produces a subset at the top (all countries at the top that cannot be distinguished from one another) and a subset at the bottom (all countries at the bottom that cannot be distinguished from one another). The stochastic and Lorenz dominance procedures also produce a subset at the top (all countries that are not dominated by any other country) and bottom (all countries that are dominated by all other countries). One may think that the subsets created by these different techniques might be related somehow. However, there is no mathematical reason for these subsets to be the same. While this issue is not addressed in this research, it may be an area worth exploring in greater detail.

Acknowledgments The authors would like to thank Andrea Brandolini, John Bishop, Julie Koch, Lukas Moller, Jeff Thompson, and Beyza Ural for their helpful comments on the first draft, Mary Santy and Martha Bonney for excellent editorial assistance, Helen Connolly for technical assistance, and three anonymous referees for their suggestions. The authors would also like to thank the Luxembourg Income Study country sponsors for their support of the LIS database and Russell Davidson for an early conversation regarding this study.

Technical appendix

The Gini, Theil, and Varlog indices and their respective bootstrap standard errors were calculated using a Stata program called *ineqerr* from Jolliffe and Krushelnytskyy [25]. The definitions used are as follows.

The weighted Gini coefficient used in this study follows the formula:

$$G = 1 + \frac{1}{N} - \frac{2}{N^2 \mu_w} \sum_{h=1}^{H} w_h \overline{\rho}_h M_h \tag{9}$$

where N is the weighted sample size (or the number of individuals in the sample when the weight is household size), μ_w is the weighted average value of M, w_h is the weight that adjusts the measure to reflect inequality of individuals and not households, $\overline{\rho}_h$ is the average rank of all the individuals in household h ranging from 1 to H, and M_h is the measure of welfare which is sorted in descending order so that M_1 is the richest individual and M_H is the poorest individual.

The Theil entropy measure used in this study is:

$$T = (1/H) S_h \left(M_h / \overline{M} \right) \log \left(M_h / \overline{M} \right) \tag{10}$$

where H is the sample size, S_h is the income share, M_h is income, and $\overline{M}$ is mean income.

The Variance of Logs (Varlog) formula is:

$$V = \frac{1}{H}\sum\nolimits_h \left[\ln M_h - \overline{\ln M}\right]^2 \tag{11}$$

where the terms are defined as above, except the mean is now of the logarithm of incomes.

Asymptotic standard errors for these indices can be found in Beach and Kaliski [5], Cowell [13], Kovacevic and Binder [26], and Mills and Zandvakili [27].

This study also the uses the Spearman's rank relation coefficient. It is calculated with the formula:

$$\left(R^2\right) = 1 - \frac{6\sum d^2}{n^3 - n} \tag{12}$$

where d is the difference between the two rankings and n is the size of the sample.

References

1. Andres, R., Calonge, S.: Inference on income inequality and tax progressivity indices: U-Statistics and bootstrap methods. (working paper). Presented at the first meeting of the society for the study of economic inequality, Palma de Mallorca, Spain, 20–22 July 2005
2. Atkinson, A.B., Bourguignon, F.: Income distribution and economics. In: Atkinson, A.B., Bourguignon, F. (eds.) Handbook of Income Distribution 1, pp. 1–58. Elsevier, New York, NY (2000)
3. Atkinson, A.B., Rainwater, L., Smeeding, T.M.: Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study. OECD, Paris, France (1995)
4. Bandyopadhyay, S., Cowell, F., Flachaire, E.: Inequality measurement and goodness-of-fit. (working paper). Presented at the first meeting of the society for the study of economic inequality, Palma de Mallorca, Spain, 20–22 July 2005
5. Beach, C.M., Kaliski, S.F.: Lorenz curve inference with sample weights: an application to the distribution of unemployment experience. Appl. Stat. **35**(1), 38–45 (1986)
6. Biewen, M.: Bootstrap inference for inequality, mobility, and poverty measurement. J. Econom. **108**, 317–342 (2002)
7. Bishop, J., Chow, K.V., Zeager, L.: Statistical inference for interdistributional Lorenz curves. (working paper). Presented at the first meeting of the society for the study of economic inequality, Palma de Mallorca, Spain, 20–22 July 2005
8. Bishop, J., Formby, J., Smith, W.: Tests for Lorenz dominance: a study of inequality in nine countries. Economica **58**, 461–477 (1991) (November)
9. Bishop, J., Formby, J., Zheng, B.: Inference tests of Gini-based tax progressivity indices. J. Bus. Econ. Stat. **16**, 322–330 (1998) (July)
10. Buhmann, B., Rainwater, L., Schmaus, G., Smeeding, T.M.: Equivalence scales, well-being, inequality and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) Database. Rev. Income Wealth **34**, 115–142 (1988)
11. Canberra Report: Expert Group on Household Income Distribution Statistics: Final Report and Recommendations. Statistics Canada, Ottawa, Canada (2001)
12. Chateauneuf, A., Moyes, P.: Lorenz non-consistent welfare and inequality measurement. J. Econ. Inequality **2**(2), 61–87 (2004) (August)
13. Cowell, F.A.: Sampling variance and decomposable inequality measures. J. Econom. **42**, 27–41 (1989)
14. Cowell, F.A. "Measurement of inequality", In A.B. Atkinson and F. Bourguignon (eds.) *Handbook of Income Distribution* 1. (2000), New York: Elsevier-North Holland, 87-166.
15. Cowell, F.A., Victoria-Feser, M.-P.: Distribution-free inference for welfare indices under complete and incomplete information. J. Econ. Inequality **1**, 191–219 (2003) (December)
16. Davidson, R., Duclos, J.Y.: Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica **68**, 1435–1465 (2000) (November)
17. Davidson, R., Flachaire, E.: Asymptotic and bootstrap inference for inequality and poverty measures. (working paper). Presented at the first meeting of the society for the study of economic inequality, Palma de Mallorca, Spain, 20–22 July 2005

18. Duclos, J.-Y., Makdissi, P.: Restricted and unrestricted dominance for welfare, inequality, and poverty orderings. J. Public Econ. Theory **6**, 145–164 (2004) (February)
19. Duclos, J.-Y., Makdissi, P.: Sequential stochastic dominance and the robustness of poverty orderings. Rev. Income Wealth **51**, 63–87 (2005) (March)
20. Farina, F., Peluso, E., Savaglio, E.: Ranking opportunity sets in the space of functions. J. Econ. Inequality **2**, 105–116 (2004) (August)
21. Garfinkel, I., Rainwater, L., Smeeding, T.M.: A reexamination of welfare state and inequality in rich nations: how in-kind transfers and indirect taxes change the story. J. Policy Anal. Manage. **25**(4), 897–918 (2006)
22. Horrace, W.C.: On the ranking uncertainty of labor market wage gaps. J. Popul. Econ. **18**, 181–187 (2005)
23. Horrace, W.C., Schmidt, P.: Multiple comparisons with the best, with economic applications. J. Appl. Econ. **15**, 1–26 (2000)
24. Horrace, W.C., Keane, T.P.: Ranking and selection of motor carrier safety performance by commodity. Accident Anal. Prev. **36**, 953–960 (2004)
25. Jolliffe, D., Krushelnytsky, B.: Bootstrap standard errors for indices of inequality: INEQERR. Stata Tech. Bull. **51**, 1–6 (1999)
26. Kovacevic, M.S., Binder, D.A.: Variance estimation for measures of income inequality and polarization. J. Off. Stat. **13**(1), 41–58 (1997)
27. Mills, J., Zandvakili, S.: Statistical inference via bootstrapping for measures of inequality. J. Appl. Econ. **12**, 133–150 (1997)
28. Moran, T.P.: Bootstrapping the LIS: statistical inference and patterns of inequality in the global north. Luxembourg Income Study Working Paper, No. 378 (2005) (February)
29. Mosler, K.: Restricted Lorenz dominance of economic inequality in one and many dimensions. J. Econ. Inequality **2**, 89–103 (2004) (August)
30. Peragine, V.: Ranking income distributions according to equality of opportunity. J. Econ. Inequality **2**, 11–30 (2004) (April)
31. Ryu, H., Slottje, D.: Two flexible form approaches for approximating the Lorenz Curve. J. Econom. **72**, 251–274 (1996)
32. Ryu, H., Slottje, D.: Parametric approximations of the Lorenz Curve. In: Silber, Jacques (ed.) Handbook of Economic Inequality Measurement, pp. 291–314. Kluwer, Amsterdam, The Netherlands (1999)
33. Sarabia, J., Castillo, Slottje, E., Slottje, D.: An ordered family of Lorenz Curves. J. Econom. **91**, 43–60 (1999)
34. Schechtman, E., Yitzhaki, S.: A family of correlation coefficients based on the extended Gini index. J. Econ. Inequality **1**, 129–146 (2003) (August)
35. Shorrocks, A.: Inequality and welfare evaluation of heterogenous income distributions. J. Econ. Inequality **2**, 193–218 (2004) (December)
36. Smeeding, T.M., Saunders, P., Coder, J., Jenkins, S., Fritzell, J., Hagenaars, A.J.M., Hauser, R., Wolfson, M.: Poverty, inequality, and family living standards impacts across seven nations: the effect of noncash subsidies for health, education, and housing. Rev. Income Wealth **39**, 229–256 (1993)
37. Smeeding, T.M., Thompson, J.: Income from wealth and income from labor: stocks, flows and more complete measures of well-being mimeo. Center for Policy Research, Syracuse University (August, 2006)
38. Weymark, J.A.: Generalized Gini indices of equality of opportunity. J. Econ. Inequality **1**, 5–24 (2003) (April)
39. World Bank: The World Development Report. World Bank, Washington, DC (2006)
40. Xu, K.: Inference for generalized Gini indices using the iterated-bootstrap method. J. Bus. Econ. Stat. **18**, 223–227 (2000)
41. Xu, K., Osberg, L.: A distribution-free test for deprivation dominance. Econom. Rev. **17**, 415–429 (1998)
42. Zandvakili, S.: Analysis of sex-based inequality: use of axiomatic approach in measurement and statistical inference via bootstrapping. In: Cinar, M. (ed.) Research in Middle East Economics The Economics of Women and Work in the Middle East and North Africa 4, pp. 75–91 (2001)