Discussion Papers No. 252, March 1999
Statistics Norway, Research Department

Rolf Aaberge

Samling Errors and Cross-Country Comparisons of Income Inequality

Abstract:
The growing interest in cross-national comparisons of income inequality is primarily a result of the establishment of the Luxembourg Income Study (LIS) database and the wide range of studies on income inequality based on LIS data. These studies suffer, however, from a major weakness since sampling errors neither are reported nor taken into account when nations are ranked according to estimates of the Gini coefficient or some alternative measure of inequality. This paper discusses the impact of accounting for sampling error when making comparisons of income inequality across nations.

Keywords: Income inequiality, the Gini coefficient, sampling errors.

JEL classification: C12, D31

Acknowledgement: I would like to thank Anne Skoglund for typing and editing the paper.

Address: Rolf Aaberge, Statistics Norway, Research Department. E-mail: rolf.aaberge@ssb.no

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

The growing attention to cross-national comparisons of income distributions is largely attributable to the establishment of the Luxembourg Income Study (LIS) database, which made it more easy to carry through such comparisons. The majority of the studies based on LIS microdata focuses on income inequality (relative incomes) and does not compare countries with regard to absolute levels of income. This is first of all due to the fact that comparisons of income inequality do not require convertions of all country money incomes to a common comparable money measure.

Although the LIS project deserves much credit for improving the degree of cross-national comparability, significant differences still remain between data for different countries. The strength and the weaknesses of the LIS data have been highlighted and carefully discussed by Atkinson et al. (1995) in a comprehensive study undertaken for OECD. The OECD study extended the conventional studies by simultaneously dealing with several important issues of methodology and should serve as a basis for future research on distribution of income. However, as most other studies on income distribution this study also lacks information on sampling errors. Estimates of sampling errors are essential for judging the significance of ranking countries with regard to income inequality even in cases where non-sampling errors may be quantitatively more important. As will be demonstrated in Section 2 the complete ranking of countries suggested by Atkinson et al. (1995) should be replaced by a ranking of countries in a few groups when sampling errors are taken into account. Section 3 deals with the problem of interpreting changes in the Gini coefficient and the question of whether a change is large or small.

2. Sampling errors and inequality ranking

A major aim of the OECD study by Atkinson et al. (1995) was to compare and rank OECD countries according to the level of income inequality. The results of this study show that a definite ranking was not possible when the comparison was based on cumulative income shares (Lorenz curves). However, complete, but not identical, rankings were provided by using the Gini coefficient and the Atkinson indexes. Although Atkinson et al. recognize the importance of taking sampling errors into account when judging inequality comparisons, they express the following concern (pages 42 and 43): "It would be possible to calculate the sampling errors asociated with the Lorenz curve, and require that one curve be significantly different from another at a specified level of confidence. This focuses on sampling errors and excludes other non-sampling errors which may be quantitatively more important. Differences in definition may lead to sizeable differences in measured shares, and there are other variations not considered. Calculating sampling errors is not given priority, though it certainly warrants fuller attention." Atkinson et al. bring forward differences in measurement techniques and definitions as major sources of non-sampling errors and, moreover, point out that cross-country comparisons also may depend on the choice of the methodological framework. However, the presence of extensive non-sampling errors does not justify the ignorance of sampling errors even in cases where non-sampling errors may be quantitatively more important. Being without estimates of standard errors makes it impossible to judge the significance of rankings which emerge from estimates of Lorenz curves and measures of inequality.

Table 1. Inequality of disposable income per equivalent adult*) in OECD countries. Per cent

Country and year	Gini coefficient
Finland, 1987	20.7
Sweden, 1987	22.0
Norway, 1986	23.4
Belgium, 1988	23.5
Luxembourg, 1985	23.8
Germany, 1984	25.0
Netherlands, 1987	26.8
Canada, 1987	28.9
Australia, 1985	29.5
France, 1984	29.6
United Kingdom, 1986	30.4
Italy, 1986	31.0
Switzerland, 1982	32.3
Ireland, 1987	33.0
United States, 1986	34.1

*) Based on the square root equivalence scale.
Source: Atkinson et al. (1995).

The purpose of this section is to examine whether the ranking in Table 1 (Table 4.4 in Atkinson et al., 1995) is affected by sampling errors. The study of Atkinson et al. is based on data sets with sample sizes that very between 2000 and 16000 (see Table A4.3 in Atkinson et al., 1995). Some of the surveys suffer, however, from high non-response rates which may lead to biased estimates of Lorenz curves and measures of inequality even in cases where non-responses are assigned incomes by imputation. The problems caused by use of imputation techniques are of interest in its own right but will not be further discussed here. Thus, it should be noted that we treat imputed incomes as observed incomes in the evaluation of the impact of sampling errors on inequality rankings.

As demonstrated by Goldie (1977) the asymptotic standard deviations of the non-parametric estimators[1] of the cumulative decile shares and the Gini coefficient depend on the shape of the underlying income distribution. The expressions for the asymptotic variances of the cumulative decile shares (empirical Lorenz curve $\hat{L}(\cdot)$) and the empirical Gini coefficent ($\hat{G}$) provided by Aaberge (1982) are given by

(1) $$\operatorname{var}\hat{L}(u) \cong \frac{1}{n\mu^2}\left[\tau^2(u) - 2L(u)\left(\tau^2(u) + \lambda(u,1)\right) + L^2(u)\tau^2(1)\right],$$

(2) $$\operatorname{var}\hat{G} \cong \frac{4}{n\mu^2}\left\{2\int_0^1\int_0^v \left(\tau^2(u) + \lambda(u,v)\right)du\,dv - (1-G)\int_0^1 \left(\tau^2(u) + \lambda(u,1)du\right) + \frac{1}{4}(1-G)^2\tau^2(1)\right\}$$

where $\tau^2(\cdot)$ and $\lambda(\cdot,\cdot)$ are defined by

(3) $$\tau^2(u) = 2\int_0^{F^{-1}(u)}\int_0^y F(x)\left(1 - F(y)\right)dx\,dy,\ 0 \le u \le 1$$

and

(4) $$\lambda(u,v) = \int_{F^{-1}(u)}^{F^{-1}(v)}\int_0^{F^{-1}(u)} F(x)\left(1 - F(y)\right)dx\,dy,\ 0 \le u \le v \le 1,$$

F is the cumulative income distribution with mean μ, Lorenz curve L and Gini coefficient G.

When F is assumed to be a uniform distribution on [0,a] for some maximum income a it follows by straightforward calculation that (1) and (2) are equal to

(5) $$\operatorname{var}\hat{L}(u) \cong \frac{4}{3n}u\left[u(1-u)\right]^2$$

and

(6) $$\operatorname{var}\hat{G} \cong \frac{8}{135n}.$$

---

[1] See Hoeffding (1948) for an alternative derivation of the standard error of the empirical Gini coefficient and Aaberge (1982) who proposed an alternative approach for deriving standard deviations of decile-specific means, cumulative decile shares and summary measures of the Gini type.

Production of unbiased estimates of standard deviations requires access to microdata. As an alternative to estimates based on the LIS data Table 2 provides standard deviations for cumulative decile shares and the empirical Gini coefficient for various sample sizes when the incomes are assumed to be uniformly distributed. The figures in Table 2 give a suggestion of the impact of ignoring sampling errors in cross-country comparisons of inequality even though they cannot be considered as unbiased estimates of the standard deviations of the estimates in Table 1.

Table 2. Standard deviations of cumulative decile share estimtates and Gini coefficients estimates when incomes are assumed to be uniformly distributed. Per cent

Population share	Cumulative decile share	Sample size				
		1000	2000	3000	4000	5000
10	1	0.10	0.07	0.06	0.05	0.05
20	4	0.26	0.18	0.15	0.13	0.12
30	9	0.42	0.30	0.24	0.21	0.19
40	16	0.55	0.39	0.32	0.28	0.25
50	25	0.65	0.46	0.37	0.32	0.29
60	36	0.68	0.48	0.39	0.34	0.30
70	49	0.64	0.45	0.37	0.32	0.29
80	64	0.52	0.37	0.30	0.26	0.23
90	81	0.31	0.22	0.18	0.16	0.14
Gini coefficient	33.3	0.77	0.54	0.44	0.39	0.34

Focusing on a specific decile, for example the 50 percent decile, we find that the standard deviation of an estimated difference between two countries is given by $\sqrt{2}\cdot 0.46=0.65$ when the estimate is based on 2000 observations. Thus, requiring 95 percent level of confidence we have that the observed difference must exceed $1.96\cdot 0.7=1.3$ to be significant, where 1.96 is the 0.975-fractile of the standard normal distribution. Our example discusses a two-sample test problem. Atkinson et al., however, discuss multiple comparisons involving 16 countries which imply 120 pair-wise comparisons[2]. Therefore, a multiple comparison method that aims at discovering significant differences is required. Note that the asymptotic distributions of the estimators are given by the normal distribution. The 95 percent confidence coefficient corresponding to the simultaneous confidence intervals (120 intervals) is equal to 3.53, which means that the probability of making at least one false statement is at most 0.05. Now, assuming 2000 observations for each country the observed differences between the 50 percent decile shares must exceed $3.53\cdot 0.65=2.3$ to be claimed significant.

[2] See Table 4.3 in Atkinson et al. (1995).

Note that Atkinson et al. use a thumb rule that claims that differences greater than 1 percentage point between cumulative decile shares are "significant". However, if significant differences between Lorenz curves are to be stated, empirical Lorenz curves need to be supplemented by simultaneous confidence bands.

To judge the significance of the ranking which is suggested by the estimates of the Gini coefficient given in Table 1 we rely on equation (6). When the income distribution F is a uniform distribution (on [0,a]) it follows from Table 2 that the Gini coefficient (multiplied by 100) is equal to 33.3 and that its standard deviation increases from 0.34 to 0.44 as the sample size decreases from 5000 to 3000. Note that the Gini coefficient of the uniform distribution is larger than the Gini coefficients in most OECD countries. This result does, however, not necessarily imply that the standard deviations of the observed Gini coefficients in Table 1 are lower than the standard deviation of the Gini coefficient of the uniform distribution when estimates are based on equal number of observations. As an illustration Table 3 provides Gini coefficients and corresponding standard deviations based on Norwegian data from 1985-1994.

Table 3. Inequality of disposable incomes per equivalent person[*)] in Norway 1985-1994. Per cent

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
Sample size	2652	4975	3393	3423	3475	6046	8072	8104	3522	12799
Gini coefficient	23.6	23.6	23.8	23.0	25.6	24.2	24.5	24.4	25.3	26.5
Standard deviation of (estimated) Gini coefficient	0.61	0.35	0.50	0.40	0.76	0.43	0.60	0.49	0.89	0.49

[*)] Based on the square root scale.

By comparing the results in Table 3 with the standard deviations of the "uniform" Gini coefficient in Table 2 for equal sample sizes we find that the standard deviations of the observed Gini coefficients are larger, and in some cases considerably larger, than the standard deviations of the "uniform" Gini coefficient even though the observed Gini coefficients are significantly lower than the "uniform" Gini coefficient. This is due to the fact that the standard deviation of the Gini coefficient depends on the form of the distribution function F. The essential difference between observed income distributions and the uniform [0,a] distribution is that the latter has a less heavy right tail since the incomes are spread uniformly across [0,a].

The estimates of Table 1 do not yield an unambiguous ranking of Gini coefficients when sampling errors are taken into account. Based on 3000 observations for each country the multiple comparison method suggests that the estimated differences between Gini coefficients must exceed $3.49 \cdot 0.44 \cdot \sqrt{2} = 2.2$ percentage points to be claimed statistical significant. Note that the size of the net sample is smaller than 3000 observations for some countries and larger for others. The result of this multiple comparison between 15 countries shows that 77 out of 105 pair-wise comparisons give significant differences. Table 4 displays the detailed results. As suggested by the results in Table 3, the standard deviation used in this exercise is probably downward biased which means that there may be fewer than 77 significant differences.

Table 4. Comparison[*)] of countries according to Gini coefficients

	Fin.	Swe.	Nor.	Bel.	Lux.	Ger.	Net.	Can.	Aus.	Fra.	UK	Ita.	Swi.	Ire.	USA
Finland			+	+	+	+	+	+	+	+	+	+	+	+	+
Sweden						+	+	+	+	+	+	+	+	+	+
Norway							+	+	+	+	+	+	+	+	+
Belgium							+	+	+	+	+	+	+	+	+
Luxembourg							+	+	+	+	+	+	+	+	+
Germany								+	+	+	+	+	+	+	+
Netherlands									+	+	+	+	+	+	+
Canada													+	+	+
Australia													+	+	+
France													+	+	+
U.K.														+	+
Italy															+
Switzerland															
Ireland															
U.S.A.															

[*)] A + means that the country in the row has a lower Gini coefficient than the country in the column.

Table 4 shows that Finland has lower Gini coefficient than all other countries except for Sweden which together with Norway, Belgium and Luxembourg form the group with second lowest inequality. Germany and Netherlands fall in an intermediate position with lower Gini coefficients than 8 and 7 of the remaining countries. Canada, Australia and United Kingdom have lower Gini coefficients than Switzerland, Ireland and U.S.A. which form the group with highest income inequality.

As indicated by Atkinson et al. the effects from non-sampling errors on inequality rankings may be quantitatively more important than the effects from sampling errors. In that case accounting for sampling as well as non-sampling errors may rise the requirement of a “significant” difference from 2.2 to 5 percentage points. Then it follows from Table 1 that solely 50 out of the 105 pair-wise comparisons give unambiguous conclusions.

3. Interpretation of changes in the Gini coefficient

The results for Norway reported in Table 3 shows that the Gini coefficient rose by approximately 12 per cent from 1985 to 1994. The question of whether this increase in inequality is large or small is separate from that of whether the increase is statistically significant. While the meaning and importance of a 12 per cent increase in the level of income is readily understood the interpretation of a 12 per cent increase in the Gini coefficient is less obvious. Thus, it will be useful to have an intuitive appealing method which helps to clarify the importance of a certain change in the Gini coefficient. To this end Aaberge (1997) proposed a method that involves a hypothetical intervention of a tax/transfer reform. It follows by employing this method that a rise of 12 per cent in the Gini coefficient from 1985 to 1994 is equivalent to the effect of introducing an equal-sized lump-sum tax of 12 per cent of the mean income in 1985 and then redistributing the collected tax revenue as transfers where each person receives 12 per cent of her/his income in 1985. The losers of this hypothetical reform are people with incomes below the mean income in 1986, whilst people with incomes above the mean receive more in transfers than they pay in tax. The mean equivalent income (based on the square root scale) in 1985 was approximately 100 000 NOK. Thus, a person with an 1985 equivalent income equal to 50 000 NOK would loose 6 000 NOK of this hypothetical intervention. By contrast, a person with 1985 equivalent income equal to 200 000 NOK would gain 12 000 NOK.

References

Aaberge, R. (1982): Om måling av ulikskap. Rapporter 82/9, Statistics Norway (In Norwegian).

Aaberge, R. (1997): Interpretation of changes in rank-dependent measures of inequality. *Economic Letters*, **55**, 215-219.

Atkinson, A.B., L. Rainwater and T. Smeeding (1995): Income Distribution in OECD Countries. *Sosial Policy Studies* No. 18, OECD, Paris.

Goldie, C.M. (1977): Convergence theorems for empirical Lorenz curves and their inverses. *Advances Applied Probability*, **9**, 765-791.

Hoeffding, W. (1948): A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, **19**, 293-325.