Session Number: 3
Session Title: International Standards for Income Distribution Statistics
Paper Number: 5
Session Organizer: Paul van der Laan
Discussant: Michael Wolfson

Paper prepared for the 26th General Conference of
The International Association for Research in Income and Wealth
Cracow, Poland, 27 August to 2 September 2000

Producing Time Series Data for Income Distribution: Sources, Methods, and Techniques

Anthony B. Atkinson (Nuffield College, Oxford)
Andrea Brandolini (Bank of Italy, Rome)
Paul van der Laan (Central Bureau of Statistics, The Netherlands)
Timothy Smeeding (Syracuse University, USA)

For additional information please contact:

Timothy Smeeding
Center for Policy Research
The Maxwell School
Syracuse University
426 Eggers Hall
Syracuse, NY 13244-1020 USA
E-mail: tmsmeed@maxwell.syr.edu
Fax: + 315/ 443-1081
Phone: + 315/ 443-9042

The authors bear entire responsibility for the conclusions of this paper. The paper does not express the official views of any of our sponsoring agencies. The authors thank Esther Gray, Andrej Grodner, and Kati Foley with their help in preparing this manuscript, and the members of the Canberra Group, especially Anne Harrison and Michael Ward, for their comments on an earlier draft of this paper.

This paper is placed on the following websites: www.stat.gov.pl
www.econ.nyu.edu/dept/iariw.

Table of Contents

		Page No.
I.	Introduction and Overview	1
II.	Conceptual Issues for Inequality Comparisons: From the Ideal to the Possible	3
	Ideal Measure	4
	Impact of Measurement Error	4
III.	Formulae for Progress: How the Primary Data Producer Can Help	8
	Documentation	9
	Data Elements: Common Sources of Error	10
	Measures of Inequality	12
	Additional Information	12
IV.	Secondary Data Collections: Pitfalls and Strategies to Overcome Them	13
	Type of Sources	13
	Nature of Summary Statistics	15
	Relationship with Other Secondary Datasets	17
V.	Trends in Income Inequality: The Researchers' and Users' Perspective	18
	Detecting Trends	18
	Significance of Changes	21
	Trends vs. Episodes	22
VI.	Summary and Conclusions	23
	References	25

I. Introduction and Overview

This paper concerns itself with methods and sources for producing better time series data for income distribution research. Cross-time comparisons within a country are not really different, conceptually, from cross-country comparisons at a point in time (Smeeding et al. 2000). The general consistency requirements are exactly the same. However, trend data need a separate treatment for at least two reasons.

First, cross-time comparisons within a country appear to be based—and in several cases are based—on more consistent data than are cross-country comparisons, mainly because they tend to come from the same producer. This is the "originator" (or producer) of the estimate; the party with the broadest knowledge of the data. However, this presumption may be unwarranted if the producer changes definitions, survey practices, or experiences a host of other nonrandom sampling or non-sampling errors which change over time. There are, in fact, many different cases where published time series are not internally consistent. A good general rule is that the longer the time frame, the more likely are non-random differences to occur. One main task is therefore to make the producer and the user aware of these problems, and to ask the producer to be as consistent as possible, to provide overlapping observations when changes are implemented, and to provide historical data on changes in time series.

The second reason is that the story gets much more complicated when we compare trends across countries, because we have to impose—in principle—a double (spatial and temporal) consistency constraint. Double international harmonization across nations *and* over time is the ideal outcome. However, such a project is daunting at this time. The recent experience of the European Community Household Panel (ECHP) has shown the difficulties to achieve complete harmonization across countries even when there is a clear objective from the start. Moreover,

longitudinal panel data is not always the best vehicle for time series estimates because samples may be small and attrition bias may affect the results. Moreover, panel data sets are not usually representative of the national population beyond the base year. That is, due to their basic nature, panels follow a set of persons sampled in a base year, thereby excluding immigrants and emigrants beyond that year.

The Luxembourg Income Study (LIS) has made considerable progress in point-in-time cross-national consistency. However, both LIS harmonization techniques and differences in national surveys made available to LIS at different points in time hamper it from achieving double consistency over time. Hence, one must ask from a practical point of view, what we can be accomplished with existing national time series. Even if we have continuous time series for different countries, is a fixed-effects correction enough to account for the methodological/ definitional differences that are found in these time series?

This paper addresses three groups of statisticians and researchers:

- ***Time Series Data Originators (Producers).*** The CSO's and other survey organizations which collect and process national estimates on income distribution from primary sources (surveys, administrative records, tax data, and other sources). Most rich (OECD) nations fall into this grouping along with the World Bank and other international organizations which collect their own survey data. LIS provides the possibility for time series comparisons but only for a limited number of years.

- ***Secondary Time Series Data***. Groups who use published or computed time series data to make large multi-period and multi-nation databases and who assure some degree of comparability over time (and sometimes across nations). Such producers include Tabatabai (1996), Deininger and Squire (1998), WIDER (1999) and others. These data are typically made available to users without a complete discussion of their strengths and weaknesses (e.g., see Atkinson and Brandolini 1999).

- ***Time Series Data Users***. Those researchers and policy analysts who use these time series and are often far too casual in their assumptions about comparability over time and across nations. Here the enormous effort which goes into model specification and econometric estimation needs to be balanced with equally serious efforts to identify and make use of the best time series datasets, and to understand the biases in many existing data series.

The hope is that this paper will help set standards for provision of better data, and robustness assessment of time series data. The rest of the paper proceeds with specific issues. We begin with a conceptual section which describes some of the issues involved with cross-national comparisons of both levels and trends in inequality.[1] This section provides some clues as to the important sources of measurement errors within and across nations and over time. Next we provide advice for data producers, particularly regarding methodology, documentation, and related issues. Then we move to secondary data series producers, i.e., those who provide an intermediary product which is used by others.[2] Here we feel that standards of comparability need to be improved if they are to provide a better product for end users. We include a series of recommendations for both groups and also for end users of trend data, as they are the ones who ultimately make use of both types of data series.

This paper is a natural complement to Smeeding et al. (2000) on point in time comparability, Harris (2000) on robustness, and Epland and Jansson (2000) on data presentation issues.

II. Conceptual Issues for Inequality Comparisons: From the Ideal to the Possible

Cross-national comparisons of income inequality must confront two major issues. The first is conceptual. What measure would one ideally use to compare distributions of well-being across countries and over time? The second issue moves from the ideal to the possible. What is the impact of using imperfect data to approximate this ideal? While both of these questions would have to be addressed even in a study of a single country in a temporal context they take on a somewhat different role in cross-national studies of level and trend in inequality.

Ideal Measure

The interest in income distribution may be justified either *per se* as a way to see how national product is distributed across people, or indirectly as the best proxy for the distribution of economic well-being. In a strictly utilitarian framework, the ideal measure of well-being would be the lifetime utility of a person. A utility measure should reflect differences in leisure as well as all forms of potential consumption, including home production and publicly provided goods; it should take account of differences in constraints faced both by people living in the same country, and differences in constraints faced by people in different countries; it should account for differences in the ability to smooth income across periods. It is therefore clear that yearly post-tax family income adjusted for family size is, at best, a proxy for this ideal concept. On the other hand, income remains a fundamental determinant of people's well-being in non-utilitarian frameworks, such as Sen's capability approach (Sen 1992).

Whatever the approach to the measurement of income inequality, we must bear in mind that measurement error arises both from differences between the ideal and the measurable, and from reporting error in the measurable. The first difference is, to a large extent, a theoretical question: concepts such as "utility" or "capabilities" are not directly measurable, and go well beyond the domain of statisticians. Reporting error is, on the other hand, something that statisticians can and must keep under control. In this chapter we focus only on this second type of measurement error.

Impact of Measurement Error

The problem of measurement error is endemic to all income distribution studies, whether they focus on a single country or many countries. The question we ask in this section is whether the bias introduced by this measurement error is aggravated in inter-temporal studies. We start

by focusing on differences in inequality within a country at two points in time. We then turn to the impact of measurement error on cross-country differences in trends in inequality.

Within-Country Trends. To focus attention on the key elements consider the following simple error components model for the jth percentile in year *t*:

$$\ln P_t^j = \ln \pi_t^j + m_t^j, \tag{1}$$

$$m_t^j = d_t + v^j + e_t^j, \tag{2}$$

where P_t^j is the measured percentile, π_t^j is the "true" percentile for the relevant measure, m_t^j is measurement error, d_t is a time-specific component that affects all deciles, v^j is a decile-specific component constant over time, and e_t^j is a decile- and time-specific component.

We start by considering the effects of measurement error on estimates of the $\ln(P_t^{90} / P_t^{10})$ in a single year, which we call the 90/10 or decile ratio for convenience. Since

$$\ln\left(P_t^{90} / P_t^{10}\right) = \ln\left(\pi_t^{90} / \pi_t^{10}\right) + \left(v^{90} - v^{10}\right) + \left(e_t^{90} - e_t^{10}\right), \tag{3}$$

we see right away that measurement error that affects all deciles equally in the year (d_t) cancel. For example, consumption of public goods unrelated to decile rank will not bias the 90/10 ratio.

Now consider the effect of measurement error in a study of changes of inequality over time. The object of interest is the difference in the 90/10 ratio between two years *t* and *t*+1:

$$\begin{aligned}\ln\left(P_t^{90} / P_t^{10}\right) - \ln\left(P_{t+1}^{90} / P_{t+1}^{10}\right) &= \ln\left(\pi_t^{90} / \pi_t^{10}\right) - \ln\left(\pi_{t+1}^{90} / \pi_{t+1}^{10}\right) \\ &\quad + \left(e_t^{90} - e_t^{10}\right) - \left(e_{t+1}^{90} - e_{t+1}^{10}\right).\end{aligned} \tag{4}$$

This illustrates the obvious, but sometimes overlooked point that decile-specific errors that do not vary over time (v^j) do not affect inter-temporal comparisons of percentile ratios in a given country. For example, underreporting by respondents at the top or bottom of the distribution will

not bias inter-temporal comparisons to the extent that this underreporting is consistent across years.

The remaining measurement error in equation (4) reflects differences between years at the 90th and 10th percentiles. Thus, the key measurement of concern to inter-temporal studies is measurement error that differs *both* across deciles *and* across years. For example, estimates of differences in inequality between two years will be biased in as much as income underreporting is greater at the 10th than at the 90th percentiles *and* this degree of differential underreporting also differs across years.

While this simple notation illustrates that certain types of measurement error do not lead to bias in inter-temporal studies, we do not want to leave the impression that measurement error is not potentially important. Measurement error may be reduced by taking differences across years, but the signal to noise ratio may be increased. This can clearly be seen by comparing the signal (S) to noise (N) ratio for estimates of inequality measures $(S/N)_t$ in year t, as given on the right-hand side of equation (3)

$$(S/N)_t = \ln\left(\pi_t^{90}/\pi_t^{10}\right)/\left\{\left(v^{90}-v^{10}\right)+\left(e_t^{90}-e_t^{10}\right)\right\}, \tag{5a}$$

while the signal to noise ratio for differences in these ratios between years t and t+1, as given by the right-hand side of equation (4):

$$(S/N)_t-(S/N)_{t+1} = \left\{\ln\left(\pi_t^{90}/\pi_t^{10}\right)-\ln\left(\pi_{t+1}^{90}/\pi_{t+1}^{10}\right)\right\}/\left\{\left(e_t^{90}-e_t^{10}\right)-\left(e_{t+1}^{90}-e_{t+1}^{10}\right)\right\}. \tag{5b}$$

Comparison of equation (5a) with (5b) shows that while taking differences across years reduces noise (as shown in equation (4)), it may reduce the signal even more. Thus, differences in 90/10 ratios over time, which eliminates decile-specific errors that are constant across years (the v's in equation (3)), reduces the noise but the remaining noise may be large relative to what we are trying to measure, namely the *difference* in 90/10 ratios. Our distinction between measurement error that does and does not affect inter-temporal comparisons is, therefore, not

meant to minimize the importance of measurement error but to focus attention on the relevant source of error.

Cross-Country Comparisons of Trends. Much of the recent literature, and this chapter itself, is focused on differences in trends across countries rather than the trends themselves. Analyzing the biasing source of measurement error for these comparisons requires that we enter country *c* explicitly into equations (1) and (2):

$$\ln P_{ct}^{j} = \ln \pi_{ct}^{j} + m_{ct}^{j}, \tag{6}$$

$$m_{ct}^{j} = d_{c}^{j} + v^{j} + e_{ct}^{j}, \tag{7}$$

$$e_{ct}^{j} = g_{ct} + w_{t}^{j} + f_{ct}^{j}, \tag{8}$$

where d_{c}^{j} is a country-specific time-invariant component that affects differently each decile *j*, g_{ct} is a time-specific component that affects all deciles in country *c*, w_{t}^{j} is a time specific component that has common effects across countries but differential effects across deciles, and f_{ct}^{j} is a component that is time, decile- and country-specific.

The trend in the 90/10 in country *c* is given by

$$\begin{aligned} \ln\left(P_{ct}^{90} / P_{ct}^{10}\right) - \ln\left(P_{c,t+1}^{90} / P_{c,t+1}^{10}\right) &= \ln\left(\pi_{ct}^{90} / \pi_{ct}^{10}\right) - \ln\left(\pi_{c,t+1}^{90} / \pi_{c,t+1}^{10}\right) \\ &\quad + \left(e_{ct}^{90} - e_{ct}^{10}\right) - \left(e_{c,t+1}^{90} - e_{c,t+1}^{10}\right). \end{aligned} \tag{9}$$

Following the logic of the previous section, differences across countries in trends will depend on w_{t}^{j} and f_{ct}^{j} but not on g_{ct}, since the latter is measurement error that differs across time and country but not across deciles. Again, taking inter-temporal differences reduces the absolute level of noise but has an ambiguous effect on the signal-to-noise ratio.

Summary. This section has shown that some but not all sources of measurement error affect inter-temporal inequality comparisons, within a country or across countries, in percentile ratios such as the decile ratio. The following generalizations emerge:

- Measurement error that is independent of decile rank (d_t in case of single country; g_{ct} in case of cross-national comparisons) affects neither level nor trend in inequality in a single country, as well as in cross-national comparisons.
- Measurement error that is time invariant (v^j in case of single country; $d_c^j + v^j$ in case of cross-national comparisons) does not affect inter-temporal comparisons, but affects each year's decile ratio.
- Cross-national comparisons of trends in decile ratios are not affected by measurement error that is either time invariant ($d_c^j + v^j$), or time varying but common across countries (w_t^j).

The difficult issue that is faced by these comparisons is therefore the comparative error structure of data within countries, across countries, and over time. If biases remain constant, errors are liable to be reduced. We conclude that it is incumbent upon both primary and secondary data producers to realize the sources of these errors and to make them known to end users of the data.

III. Formulae for Progress: How the Primary Data Producer Can Help

For decades, National Statistical Offices (CSO's), Census Bureaus, Finance Ministries, Social Security Bureaus, and others have produced time series estimates of income distribution for national audiences. Often these series are obscurely published, or are available as internal memos, not as primary publications.[3] This section of the paper addresses the issues involved in producing these series from a statistical and robustness point of view. It suggests methods for avoiding pitfalls, and it makes recommendations for best practices in this area.

The most important lesson for nations and their CSO's is that greater care need be spent providing these series and their providing statistical properties. Official publications need be made where none now exist. If the task of time series are left to one or another junior statistician, each of whom is asked by an important third party for a time series, different statisticians (or the same statistician following different sets of instructions) may provide different results. And, as these various results appear in the World Bank, OECD, WIDER, UNICEF, United Nations, ILO, or other similar publications, they will be subject to international scrutiny and critique. If considerable effort is devoted to producing one time series or a set of time series of a comparable, accurate, and well documented nature, less time and effort will be spent defending series and reacting to (or criticizing) the way that various bodies make use of producers' estimates.

Documentation

As the chapter on Robustness Assessment Reports (RAR's) suggests, documentation of data quality (sampling and non-sampling errors, imputations, simulations, etc.), income measures, inequality measures, top- or bottom-coding of data, and related technical documentation is the first step toward accurate assessment of data comparability. Trend analyses demand that this same procedure be repeated each year and important changes in survey practices, measurement techniques, etc., be reported every decade (or better, every five years). Both the RAR (Harris 2000) and the LIS Technical Documentation template provide useful examples of one time RAR's. But now we must turn our attention to multiple years RAR's and the issue of comparability over time. The rest of this section of the paper might be seen as a list of elements which are necessary ingredients for such an inter-temporal RAR.

Data Elements: Common Sources of Error

Here we suggest an initial list of cautions for data producers. Other unmentioned differences over time may also affect trend analyses. Hence, this is not an exhaustive list.

Definitions. Every measure of income distribution forces the producer or the analyst to make numerous choices: the reference unit (e.g., inner family, tax unit, household); the adjustment for the size and composition of the reference unit (equivalence scales); the sharing rules among the component of the unit (usually equal division); the welfare weighting of each single unit observation (e.g., persons vs. tax units/households);[4] the definition of income (e.g., post-tax vs. pre-tax after allowing for tax deductions vs. pre-tax before deductions); the time period over which income is measured (week, month, year); and the treatment of the incomes of persons who are present for only part of the period on account of entering or leaving the survey-eligible population.[5] While these aspects are critical for any point in time measure, both one-time choices and, important changes in these choices, are important for trend analyses.

As societies evolve, data producers are naturally led to revise their definitions to obtain a better description of reality. For instance, the recent shift of European economies toward a more flexible labor market means that the share of part-year and part-time earners—among whom are people employed with fixed-term contracts—is probably on the rise, and it is more sensitive to the business cycle than are permanent contracts. A change in the treatment of these units may improve the quality of each years' data. But it may also have a related impact on the time series and therefore on the measured changes in income distribution over time.

The definition of income deserves attention too. The composition of households' revenues substantially changes over long periods of time. Sources of income previously unrecorded, or whose imprecise record was a minor problem, may gain sudden significance, generating a discontinuity in the series. A good example is represented by the increasing

importance of investments in the stock market and the ensuing capital gains (or losses), whose effects on the Swedish income distribution are discussed below. On the other hand, the major thrust of this report is to push nations toward common definitions of income. In making these changes, nations will affect their own trend analyses of distribution when the income definition changes. In either cases, it is important, for purposes of trend analysis, that data producers preserve older definitions and continue long time series based on these definitions.

Coverage. A second point concerns itself with changes in the coverage of the data. For instance, presentation of a data series assumes that population groups included in the sample survey do not change over time. Countries with rapid and large waves of immigrants must specify rules for inclusion or exclusion of immigrants to remain the same over time. A special case is represented by longitudinal panel datasets. They provide important information of changes in an individual's economic circumstances over time. However, panels are particularly bad for trend analyses of income distribution if the original sample excludes major changes in the population (emigrants or immigrants) after the original sample is drawn.

Practices. Changes in survey practices are also a common source of differences in time series even when the other components of income distribution measurement do not change at all. For instance, changes in survey questions; ordering of questions; methods of data collection (telephone vs. face-to-face); use of computers (CATI or CAPI) are all liable to be made to improve data quality at a point in time, but also affect the time series from the same data. Imputation methodologies, estimation techniques for different income components and other characteristics of survey reporting might also differ over time. These efforts are to be applauded as they improve the quality of data, but they are also sources of bias in time series that need be reported to analysts in an easily accessible way, with breaks in series where appropriate, and with overlapping of old vs. new techniques to indicate how these series differ.[6]

Measures of Inequality

At the heart of every time series lies a few summary statistics of inequality which are provided by the data-producing agency. The Gini coefficient is the most popular Lorenz-based measure, but others, e.g., Atkinson and Theil, are often used as well. Great strides have been made in analyzing the statistical properties of these measures (e.g., see Cowell 2000), and these analyses have led to various articles concerning the biases in each type of measure. Underlying most of these summary measures is a Lorenz Curve which contains information about the share of income to various population subgroups. The availability of this elementary data, e.g., percentile shares of income, or mean incomes by percentiles, along with data on top- and bottom-coding of incomes, provides the analysts with the raw material to make their own estimates of any of these summary measures and others, e.g., the percentile ratios, depending on the research issues at hand. Such provisions of "elementary" data allow the users to construct their own summary measures of inequality. And in this day of web pages and electronic data transfer, the producer cost for making such data available for the sophisticated user is very much reduced.

Additional Information

If the national authorities or data producers know of important studies which document changes or their effects, or which analyze sources of survey error, these should be made available to the analyst in a summary bibliography. Similarly, if multiple time series of estimates are made using different surveys or by different agencies, reports which compare these are important to report to analysts such that comparisons can be made (e.g., Erikson 1999). For instance, the multiple surveys and similar measures of income which underlie trend estimates for The Netherlands (Figure 1) give the user confidence that the trend in inequality in The Netherlands has been roughly similar regardless of the series or income measure used, while a

Swedish Finance Ministry report helps one sort out different series of changes in the Swedish income distribution since 1990 (Erikson 1999).

IV. Secondary Data Collections: Pitfalls and Strategies to Overcome Them

The first problem for the producer of a "secondary" dataset—i.e., a collection of summary measures of inequality drawn from a number of heterogeneous sources—is to set internal standards for accepting or rejecting estimates. Selection criteria must be based on the features described in the previous sections. For instance, Deininger and Squire (1996) chose the statistics to be included in their dataset by requiring that they be from national household surveys for expenditure *or* for income, that they be representative of the national population, and that all sources of income or expenditure be accounted for, including own-consumption. As with primary data producers, the main duty of a researcher or organization assembling a secondary dataset is to document the origin and characteristics of all estimates included, according to the criteria which they develop and the information made available by the primary data producer.

The producer of a secondary dataset has as his/her goal the production of time series estimates for multiple nations. As a result, he/she is faced with three additional problems: the types of alternative sources, the nature of the summary statistics, and the relationship with other secondary datasets.

Type of Sources

There are two main sources of data: household surveys and administrative archives, of which income tax records are the most important and have historically provided long run time series of continuous data.

Tax records suffer from potentially serious problems:

- incomplete coverage of those with incomes below the tax threshold, a problem which varies over time with the tax base;
- the tendency to under-report certain types of income;
- the definition of taxable income may not correspond to that chosen in studying income distribution;
- the definition of the tax unit may not be appropriate; and
- there may be difficulties in treating part-year units.

For these reasons, tax records are typically used in conjunction with other sources: for example, social security information for non-taxpayers, and information on total incomes from national accounts. Appropriate use of these files almost always involves direct matching of individual files by a personal identifier and, hence, runs up against privacy and confidentiality concerns. In most nations, the individual respondent is required to give his/her "informed consent" before the match takes place.

Household surveys are also subject to many problems. These obviously include sampling error, which in turn depends on the size and structure of the sample. Where the survey is part of a panel, there is sample attrition. Perhaps even more important, most household surveys face problems of differential non-response, particularly item nonreporting and item misreporting (usually underreporting) which reduce the representativeness of the observed sample. This may necessitate grossing-up procedures based on administrative data, census data, or other population, data. The resulting income distribution estimates may be affected by the accuracy of the latter data and by revisions (for example, where decennial census results become available). There are also problems of failure to tailor questions asked to the chosen definitions. These may, as with tax information, mean that there is a need for the adjustment of raw data to exogenous information, such as national accounts. Moreover, procedures employed to "adjust" data for these nonsampling errors may differ over time. In most statistical offices, changes in

adjustments for misreporting errors, nonreporting errors, and other nonsampling errors (e.g., questionnaire changes), may improve the quality of data in a given year compared to a former year. However, while the data gets "better," the time series may become biased. Secondary producers need to be aware of these biases and incorporate information on these changes into their series.

For some income distribution estimates, information may be combined from several sources to yield "synthetic" estimates. For instance, income tax data on higher incomes may be merged with household survey data for the rest of the distribution, drawing on the relative strengths of each data source. These estimates may also be further adjusted using national accounts or administrative data. Taxes and transfers may be calculated using a simulation model and added to survey data. Such a procedure may be required where the original survey does not contain the information, or where the tax information in the survey relates to a different time period from the income information.

Ideally a secondary dataset should include several different time series, if available, and include information which can be used to assess the reliability of the observations and evaluate their relative merits. For example, this could include the sampling errors associated with the Gini coefficients; it could include the proportion of the population covered, in the case of tax records. This is important because the evidence of alternative sources on inter-temporal changes may be contradictory, as discussed below.

Nature of Summary Statistics

The role of secondary datasets is to make accessible and enlarge the range of "ready made" income distribution statistics. This process can take several forms, and it may be helpful to bear in mind the different origin of the "ready made" income distribution statistics contained in secondary sources:

- calculated from individual national micro datasets (e.g., Current Population Survey tapes in the case of the United States), where there may be differences between "original" and "public use" datasets;

- calculated from collections of harmonized micro datasets such as LIS; as again these may differ from those available in the original source;

- calculated from tabulations published by (or supplied by) national sources; here it should be noted that national sources may give differing degrees of detail (e.g., the data published in *Statistical Yearbooks* may have fewer ranges than in a specialized publication on income distribution), and that the published sources may be revised or published in alternative forms (e.g., based on different definitions);

- calculated from tabulations in another secondary dataset;

- summary statistics published by (or supplied by) national sources (e.g., the Gini coefficients published by the U.S. Bureau of the Census);

- summary statistics obtained directly from another secondary dataset producer or the publication of another analyst.

In all cases, the calculations involve decisions being made—such as those discussed in Section III above. There is for example the application of procedures of top-coding. This may happen in the course of the collection of the data, or as a decision of the researcher to reduce the noise that is typically concentrated in the tails of the distribution. Changes in these procedures may significantly affect the comparability of results.[7] At the bottom of the scale, there is the issue of zero or negative incomes, which cause problems for certain summary measures (e.g., the Atkinson and Theil indices, but not the Gini coefficient). These may be bottom-coded, being set to zero or a small positive number, or may be omitted. All of this needs to be documented.

A second example is the procedure for estimating quantile shares and inequality indices when the original data were used in grouped form in primary sources,[8] or were available only in grouped form to researchers. When some kind of statistical procedure, such as the fitting of a parametric Lorenz curve, is followed, results may diverge from those reported in the original sources. It would be advisable, and relatively inexpensive, to include in secondary datasets not

only the recalculated series but also the original statistics. Equally, the upper and lower bounds with grouped data (obtained with different assumptions about the within-class distribution) are readily calculated and should be included.

In general, the procedures applied in processing the data should be fully documented, and the user allowed as wide a range of choice as possible. It should be noted that choices such as those regarding interpolation method or treatment of zero incomes may be implicit in the adoption of a statistical package, or the formulae applied in the calculations, and that this may affect the conclusions drawn.

Relationship with Other Secondary Datasets

There is a long tradition, in the field of income distribution, of creating secondary datasets.[9] The comparison of such compilations shows that overlapping is sizeable, and suggests some desirable features for a secondary dataset:

- ***Consolidation***. In principle, multiple observations for the same country and the same date are justified where there are differences in definition (for example, household weights vs. person weights), or where there are different methods of calculation (for example, upper and lower bounds for the Gini coefficient). When there is no apparent reason for a difference, multiple observations need to be traced back to their original sources in order to identify the cause. In view of their use in the past, keeping duplicate figures contained in earlier secondary datasets is valuable because it facilitates comparisons, but it should be clear that their status is that of *memorandum items*.

- ***Comprehensiveness***. When other secondary sources are used, the coverage of such sources should be exhaustive. Omitting observations that fail to meet some pre-specified criteria may be convenient, but it may be preferable to include these unsatisfactory observations with a proper cautionary note.

- ***Full documentation***. Precise references and table numbers and a full account of all adjustments made should be given, so that observations in the dataset can be reproduced and their genealogy reconstructed.

- ***Replication***. As secondary datasets become available on-line, their producers are likely to update and revise them, occasionally or on a regular basis. To address replication problems, there should be a numbering of different releases of the datasets, and all versions should be conserved and remain available.

The burden assumed by secondary dataset producers is a huge one. They attempt to overcome all of the theoretical (Section II) and practical (Section III) biases found in "original" datasets. Moreover, they attempt to make these series comparable over time and sometimes across countries. Their task is a most difficult and complicated one. Hence, while we salute these efforts, we also hasten to add that the devil is always in the details of their estimates, so please, do not skimp on the details.

V. Trends in Income Inequality: The Researchers' and Users' Perspective

The previous sections of this paper addressed inequality time series data from the point of view of the CSO (data series originator) and from that of secondary (intermediary) dataset producers. This penultimate section contains issues related to users and presenters of trend data: researchers, social statisticians, policy analysts, and others. It is a rough and ready collection of lessons learned by the authors in writing several papers on this topic (Atkinson 2000; Atkinson and Brandolini 1999; Brandolini 1998; Gottschalk and Smeeding 1997, 1999; Smeeding 2000). The hope is that others may avoid some of the pitfalls we have noted, or even made ourselves.

Detecting Trends

A sampling of the problems that may arise includes:

- ***Two point trends***. In several nations, comparable household income microdata is only available for two periods (e.g., 1980 and 1990 only for Portugal and for Spain from 1994 through this writing). Having two periods permits the user to estimate the change between them, but it may conveys a rather misleading impression of the underlying trend. There is a considerable danger in taking a very small number of years (two as a minimum) to extrapolate long-run trends.

- ***Business cycle effects***. Because of cyclical variations in inequality, trends based on an arbitrary time period (e.g., 1980 to 1995) might fit the business cycle in different periods across nations. If inequality is pro-cyclical—as is the case in the United

States—peak (year) to trough (year) trend estimates are biased downwards; trough to peak trends are biased upwards. The opposite holds if inequality is counter-cyclical. Comparing peak-to-peak or trough-to-trough provides the least biased estimates and this requires a lengthy time series of estimates (e.g., see Burkhauser, Crews, and Jenkins 1998).

- ***Mixing datasets and definitions***. Trend analysis of inequality requires comparing several income definitions and/or several datasets over time (e.g., see Figure 1). In general, mixing cursorily different datasets to form a single trend is not recommended as the trend will reflect *both* the "real" inequality change *and* differences across datasets.

We illustrate all three of these issues in Figure 2, using actual Spanish income distribution data for 1980 and 1990 from LIS, and for 1993 from the ECHP and a hypothetical business cycle. The LIS datasets (based on the Spanish Income and Expenditure Survey) for 1980 and 1990 indicate a downward trend in inequality. When the ECHP is added, inequality increases and the "trend" line through all three points is moderately upwards. The "true trend" line and the "actual" curved inequality trend line are both hypothetical, but illustrate the fact that peak-to-peak or trough-to-trough lines are consistent with the observed trend across the three (mixed) datasets.

- ***Changes in income definitions over time and differences in definitions across datasets***. The availability of multiple sources and definitions poses the problem of discerning inequality trends, when the direction of change is ambiguous.

The Swedish inequality trend is a fine example of the differences in income definitions and reference unit over time and across datasets (Figure 3). Here we combine three sets of data, one from LIS, two from Statistics Sweden's official publications. The LIS trend shows a modest increase in inequality since 1980, but a decline from 1991 to 1995. The LIS dataset biases inequality upwards at any point in time by ignoring the fact that young adults living with others (e.g., parents) share in household economies of scale. This difference should not bias trend estimates of inequality unless living arrangements or numbers of young adults change drastically

over the period in question. This LIS time series is therefore not the best one for measuring the Swedish inequality trend (see Erikson 1999).

The differences between the two Statistics Sweden trends are also problematic. The official income definition used by Statistics Sweden includes realized capital gains (highest line in figure). Capital gains are sensitive to both business cycles and Swedish tax laws. In 1990 there was an abrupt upward shift of the Gini coefficient due to changes in tax laws. This shift produced a discontinuity in their trend data which is "overcome" in Figure 3 by assuming a one-time "fixed effect" and shifting down the new trend to equate with the old in 1990. Therefore the 1990-1997 trend connects with the pre-1990 line in 1990. (The LIS definition in the dotted line not affected because LIS disposable income excludes capital gains.)

The second Statistics Sweden trend line (middle line in Figure 3) keeps the same tax unit definition and other definitions, except it excludes capital gains. The two Statistics Sweden estimates still indicate an upward trend in Swedish inequality. However, the increase in inequality with capital gains (top line) is more drastic and less regular than that found in the series without capital gains (middle line). Hence, when multiple estimates are selected trends may not be very clear. Inequality has risen modestly or rapidly in the 1990s depending on which income definition and data series is selected.

Another good example is provided by Italian data (Figure 4). During the 1980s and until mid 1990s changes in income inequality appear significantly different according to whether they are measured on data from the Income Survey by the Bank of Italy, or from the Expenditure Survey of the Italian Statistical Office (in Figure 4 denoted by SHIW and SHB, respectively). The discrepancy emerges both for changes over shorter periods, and for the overall change over the entire period, with the SHIW showing a tendency toward greater inequality and the SHB the opposite tendency. The LIS series is similar to the other two SHIW series, but with greater rise

in inequality during the 1990s. In this case, however, a detailed assessment of the characteristics of the two sources leads us to dismiss the SHB evidence as less reliable than the SHIW evidence (Brandolini 1999).

In some nations, e.g., the United Kingdom, several different sets of income distribution data can be used to make trend comparisons: tax estimates (Blue Book); Family Expenditure Survey estimates; Family Resources Survey estimates; and British Household Panel Study estimates, each with their own biases. Comparison of alternative time series estimates may help reinforce one another (e.g., The Netherlands), or they may not (Italy). But in any case, the author should use all of the available evidence in making their judgments about which series, sets of series, or combinations of series produce the most reliable estimates.

- ***Interpolated trends***. If one has detailed knowledge of time series (like in the case of Sweden), one can interpolate among the various estimates to produce as "clean" a series as possible. The bold line in Figure 3 comes from Gottschalk and Smeeding (2000) and Smeeding (2000), and combines various estimates into one "preferred" series. Clearly some judgements were made in creating this series, e.g., capital gains treatment, starting and ending point, choice of unit, etc. These should be made clear by the researcher with alternative estimates or series made available to the reader as in Tables 3 and 4.

Significance of Changes

There are no generally accepted standards for labeling significance of inequality changes. In the literature, authors have used clearcut standards, e.g., a "1.0 point change in the Gini" (Atkinson, Rainwater, Smeeding 1995, p. 39), or some fixed changes, e.g., "a 5 to 10 percentage point change" (OECD 1999), or a "3 to 7 percentage point change" (Gottschalk and Smeeding 1999; Smeeding 2000). But these have not been based on formal tests of significance or on standard errors of the estimated summary index. Thus, a problem such as this for users might only be solved by information made available by dataset providers or by the raw data itself. In the absence of raw data authors must fall back on their own standards, or those imposed by the data providers.[10]

Trends vs. Episodes

A fundamental issue in the analysis of inter-temporal changes of income inequality has to do with the different emphasis on "trends" vs. "episodes." So far, we have used the term "trend" as the intuitive notion of "average" long-run change. However, to the extent that measures of income dispersion alternate periods of small and irregular changes with sudden accelerations—be they in the direction of higher or lower inequality—the search of a long-run regularity such as a single trend may be misleading, and it may be better to think in terms of "episodes" when inequality fell or increased (see Atkinson 2000). As the analysis of long-run movements of income inequality is still a relatively unexplored field of research, opinions differ whether the focus should be on sequences of episodes rather than trends. We do not need to take position on such a question here, but two points need to be stressed.

First, the conclusion on trends depends crucially on the choice of the initial and ending periods. Take the case of Sweden (Figure 3). The pattern is one of falling inequality until 1980 and then rising inequality since then, faster in the 1980s than in the 1990s. Hence, beginning a time series of Swedish inequality in 1975 produces a very different pattern than from 1980 or 1990. The long-run movement of inequality can be obscured by different presentations of data time series.

Second, an apparently common trend across nations may disguise very different patterns of shorter period changes. As an example, consider the "summary bar chart" in Figure 5, which is based on various sources of time series data summarized in Gottschalk and Smeeding (2000) and updated for this paper. The method is to calculate the annual percentage change in the Gini (from the first to the last data year) and to also calculate the absolute change year-to-year (from the first to the last year). The technique overcomes comparisons based on different years of data (long series for some, shorter for others). It also allows for comparisons of percents (absolute

change) and percentages (relative change) which are quite different because the base Ginis vary by a factor of roughly two-to-one across nations at any point in time, e.g., about 0.222 for Sweden (1995), and 0.375 for the United States (1997) (Smeeding 2000, Table 1).

The shortcoming of this method is that the bar chart smoothes over periods of change where inequality first falls then rises. For instance, Figure 5 indicates a small but very similar changes in Italian inequality (1979-1995), and in Canadian inequality (1979-1996). In fact, the Canadian pattern is just that—very little change since 1969 (Figure 4). Conversely, Italian inequality fluctuated considerably between 1979 and 1995, and distinct episodes of falling and rising inequality were submerged within one summary trend number (Figure 6). The lesson is that both assessing percentage changes *and* showing the actual pattern of change add to our knowledge because trends and episodes of inequality are not always the same. Moreover, it needs to be noted that difference between beginning and end points is meaningful only when the trend exists, as it may be impossible to reduce a complex time series to U or inverse U shapes alone.

VI. Summary and Conclusions

Increasingly economists and social policy analysts are focusing attention on the long-run trend in income inequality. The availability of 20 to 40 years or more of estimates in many nations are making it possible for analysts to analyze the determinants and consequences of long periods of distributional change, e.g., the relationship between inequality and growth, trends in world income inequality and related issues. The future will bring more, not fewer, uses of such data, and policy discussions of national governments and international bodies may be heavily influenced by such trends and analyses of trends.

In this light, this paper and those which preceded it (e.g., Atkinson and Brandolini 1999) must be seen as first steps in setting standards for time series analysis. The real long-run value of these efforts depends on whether primary income inequality time series producers such as national statistical offices and secondary producers such as ILO, WIDER, and the World Bank, pay attention to the cautions and suggestions made above.[11] There is much room for improvement in our time series data on income distribution.

Endnotes

1. Section II relies on the analysis in Gottschalk and Smeeding (2000) transformed into a time series context.

2. Section IV draws on the paper by Atkinson and Brandolini (1999).

3. For instance, see the list of sources in Atkinson (2000), Smeeding (2000), Atkinson and Brandolini (1999), Brandolini (1998), Atkinson, Rainwater, and Smeeding (1995), Gottschalk and Smeeding (1997, 2000), OECD (1999), and others who have collected time series estimates of income inequality measures from various sources.

4. Notice that welfare weighting is a different issue from re-weighting sample data to allow for differential sampling or non-response.

5. If the period is a year, their income can be excluded, included without any adjustment, or included after being annualised. Some attention to this issue has been paid by the CBS in The Netherlands, whose changes in the treatment of part-year units limit the continuity of the published series, and by the Central Statistical Office in the United Kingdom, which showed that in 1978/79 the exclusion of part-year incomes led to a reduction in the Gini coefficient for income before tax of 2 percentage points (Central Statistical Office 1981, p. 86, Table E).

6. Additional comment on using fixed effect adjustments to interpolate series is found below.

7. For instance, in the United States several income items are recorded with a pre-set upper limit (see Bureau of the Census 1998, p. B7, footnote 3). According to Ryscavage (1995, p. 55), "increasing the upper limits, or top codes, in 1993 ... had a significant impact both on the Gini index and on the shares of aggregate income received by various quintiles of the distribution ...". See also Smeeding (2000, Table 2) on how these changes affect the gini.

8. This is the case of the United States gini coefficients (see Bureau of the Census 1998, p. A1).

9. Among earlier collections, trend data were reported in United Nations Economic Commission for Europe (1957, 1967), Kravis (1962), Kuznets (1963), Paukert (1973), Jain (1975), Sawyer (1976), United Nations Economic Commission for Asia and the Pacific (1979), United Nations (1981). Historical data for Western European countries were assembled by Flora (1987) on the basis of income tax statistics. The most recent secondary collections are the ones compiled by Tabatabai (1996) at ILO, Deininger and Squire (1996) at the World Bank, and Cornia et al. at WIDER (1998). These secondary datasets are by far the largest and relatively more documented. At the same time, they

are not exempt of problems, as discussed at length in Atkinson and Brandolini (1999) with reference to that assembled by Deininger and Squire.

10. For instance, the U.S. Bureau of the Census (1999) provides information on the significance of changes in income distribution together with appendices that contain the formulae for standard errors that are used to make these estimates.

11 In fact, recent work by the Inter-American Development Bank which improves comparability for Latin American Data is an optimistic sign (Szekely and Hilgert 1999).

References

Atkinson, A.B. 1997. "Bringing Income Distribution in From the Cold," *Economic Journal*, 107: 297-321.

Atkinson, A.B. 2000. "Is Rising Income Inequality Inevitable? A Critique of the Transatlantic Consensus," WIDER Annual lecture #3. Helsinki, Finland: United Nations World Institute for Development Economics Research.

Atkinson, A.B. and A. Brandolini. 1999. "Promise and Pitfalls in the Use of 'Secondary' Data-sets: Income Inequality in OECD Countries," mimeo. Oxford, UK: Nuffield College, Oxford University.

Brandolini, A. 1998. "A Bird's-Eye View of Long-Run Changes in Income Inequality," mimeo, Research Department. Rome, Italy: Bank of Italy.

Brandolini, A. 1999. "The Distribution of Personal Income in Post-War Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality," *Giornale degli Economisti e Annali di Economia*, 58: 183-239.

Central Statistical Office. 1981. "The Distribution of Income in the United Kingdom, 1978/79," *Economic Trends*, 327: 82-91.

Cowell, F. 2000. "Measurement of Inequality." In A.B. Atkinson and F. Bourgignon (eds.), *Handbook of Income Distribution*. New York: Elsevier-North Holland Publishers, pp. 87-166.

Deininger, K. and L. Squire. 1996. "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10: 565-591.

Epland, J. and K. Jansson. 2000. "Data Presentation Issues," mimeo, Canberra Group Meeting, LIS, Luxembourg, May.

Erikson, I. 1999. "Income Distribution Trends in Sweden." Stockholm: Swedish Ministry of Finance, June.

Flora, P. 1987. *State, Economy and Society in Western Europe 1815-1975, Volume II*. Frankfurt: Campus Verlag.

Gottschalk, P. and T.M. Smeeding. 2000. "Empirical Evidence on Income Inequality in Industrialized Countries." In A. B. Atkinson and F. Bourgignon (eds.), *Handbook of Income Distribution*. New York: Elsevier-North Holland Publishers, pp. 261-308.

Gottschalk, P. and T.M. Smeeding. 1997. "Cross-National Comparisons of Earnings and Income Inequality," *Journal of Economic Literature*, XXXV (June): 633-687.

Harris, G. 2000. "Quality Assessment Reports (RAR) for Micro and Macrodata," mimeo, Canberra Group Meeting, LIS, Luxembourg, May.

Jain, S. 1975. *Size Distribution of Income. A Compilation of Data*. Washington, DC: The World Bank.

Kravis, I.B. 1962. *The Structure of Income. Some Quantitative Essays*. Philadelphia: University of Pennsylvania Press.

Kuznets, S. 1963. "Quantitative Aspects of the Economic Growth of Nations: VIII. Distribution of Income by Size," *Economic Development and Cultural Change*, 11(2): 1-80.

Organisation for Economic Co-operation and Development (OECD). 1999. "Trends in Income Distribution and Poverty in the OECD Area," DEELSA/ELSA/WPI(99) 15, Social Policy Studies Division. Paris: OECD, September 30.

Paukert, F. 1973. "Income Distribution at Different Levels of Development: A Survey of Evidence," *International Labour Review*, 108: 97-125.

Ryscavage, P. 1995. "A Surge in Growing Income Inequality?" *Monthly Labor Review*, 118(8): 51-61.

Sawyer, M. 1976. "Income Distribution in OECD Countries." *OECD Economic Outlook. Occasional Studies*. Paris: Organisation for Economic Co-operation and Development.

Sen, A. 1992. *Inequality Reexamined*. Cambridge, MA: Harvard University Press.

Smeeding, T. 2000. "Changing Income Inequality in OECD Countries: Updated Results from the Luxembourg Income Study (LIS)." In R. Hauser and I. Becker (eds.), *The Changing Distribution of Income*. Berlin, Germany: Springer-Verlag, in press.

Smeeding, T., M. Ward, I. Castles, and H. Lee. 2000. "Cross-Country Comparisons of Inequality," mimeo, Canberra Group Meeting, LIS, Luxembourg, May.

Szekely, M. and M. Hilgert. 1999. "What's Behind the Inequality We Measure: An Investigation Using Latin American Data for the 1990s," mimeo, Inte-American Development Bank, December.

Tabatabai, H. 1996. *Statistics on Poverty and Income Distribution: An ILO Compendium of Data*. Geneva: International Labour Office.

United Nations. 1981. "A Survey of National Sources of Income Distribution Statistics (First Report)," *Statistical Papers*, Series M, No. 72. New York: United Nations.

United Nations Economic Commission for Asia and the Pacific. 1979. *Economic and Social Survey of Asia and the Pacific, 1978*. Bangkok: United Nations.

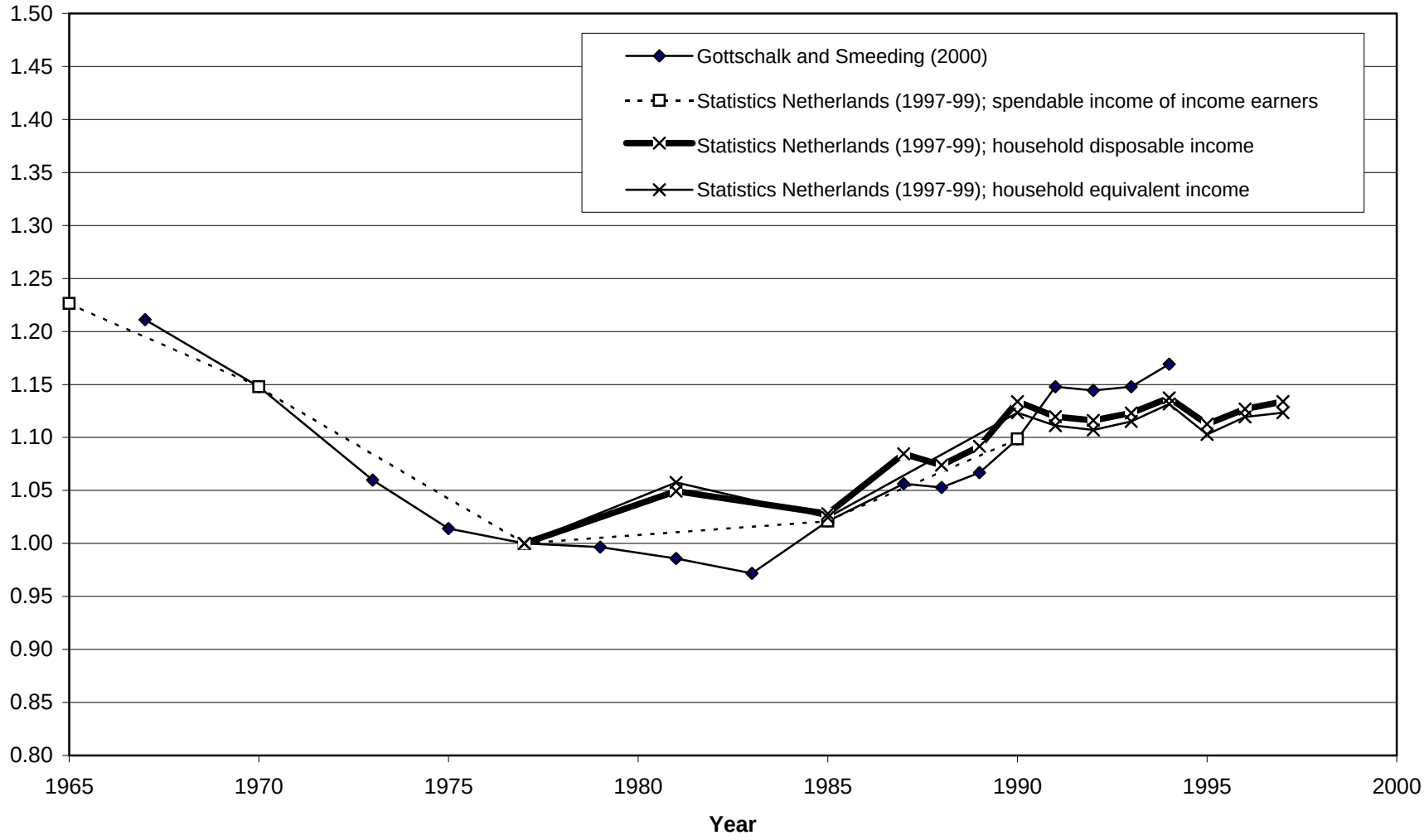
United Nations Economic Commission for Europe. 1957. "Income Distribution in Western Europe." In *Economic Survey of Europe, 1956*, part III. Geneva: United Nations.

United Nations Economic Commission for Europe. 1967. *Incomes in Post-War Europe: Study of Policies, Growth and Distribution*. Geneva: United Nations.

U.S. Bureau of the Census. 1998. "Money Income in the United States: 1997 (With Separate Data on Valuation of Noncash Benefits)," *Current Population Reports*, Series P-60, No. 200. Washington, DC: Government Printing Office.

World Institute for Development Economics Research (WIDER). 1999. *World Income Inequality Database*, Beta 3. Helsinki, Finland: WIDER, November 8.

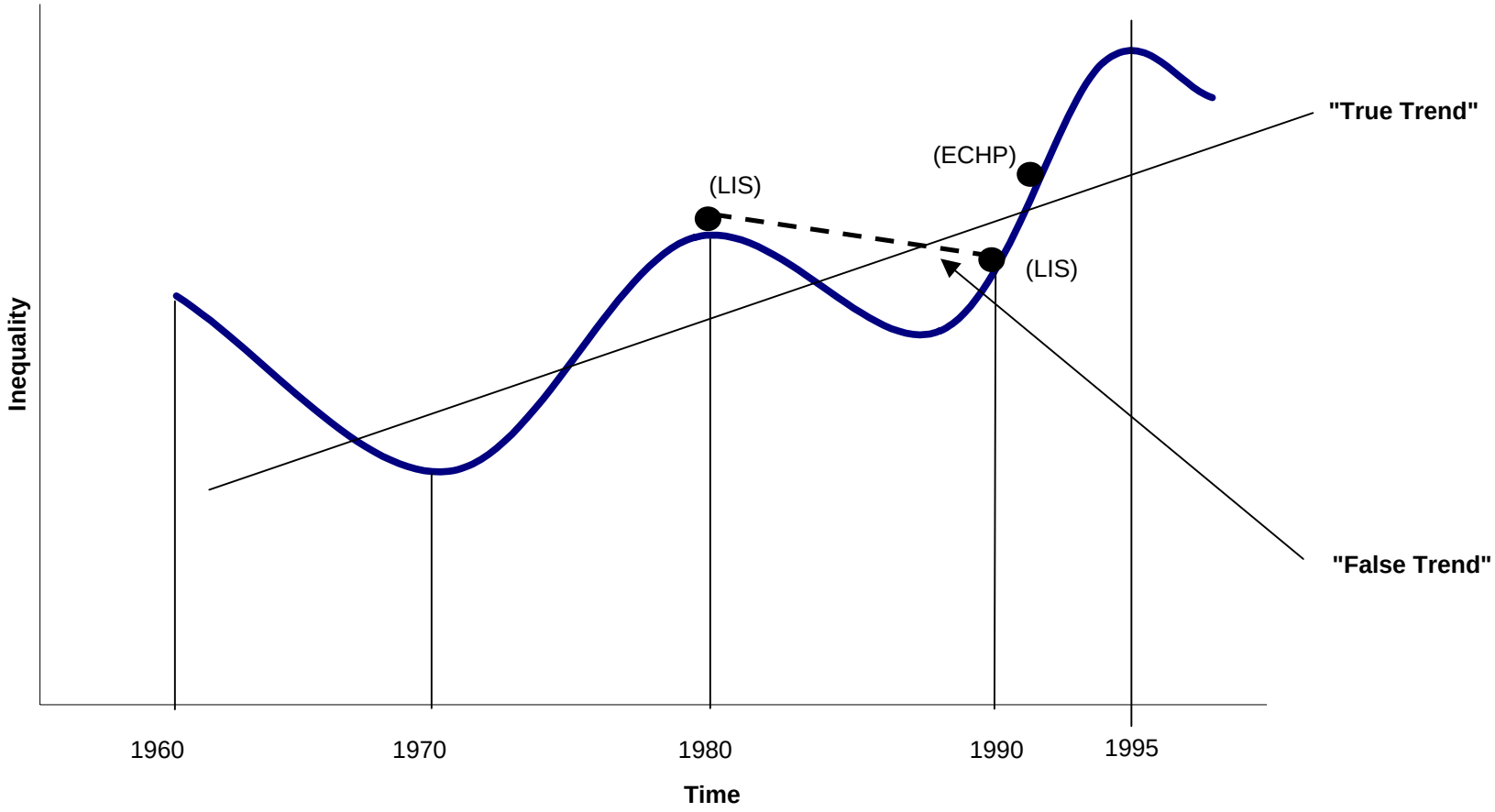
Figure 1.
Trend in Income Inequality: Gini Coefficients (1977=1) in Netherlands



Sources: Gottschalk and Smeeding (2000); Central Bureau of Statistics, Netherlands (1999); Atkinson and Brandolini (1999).

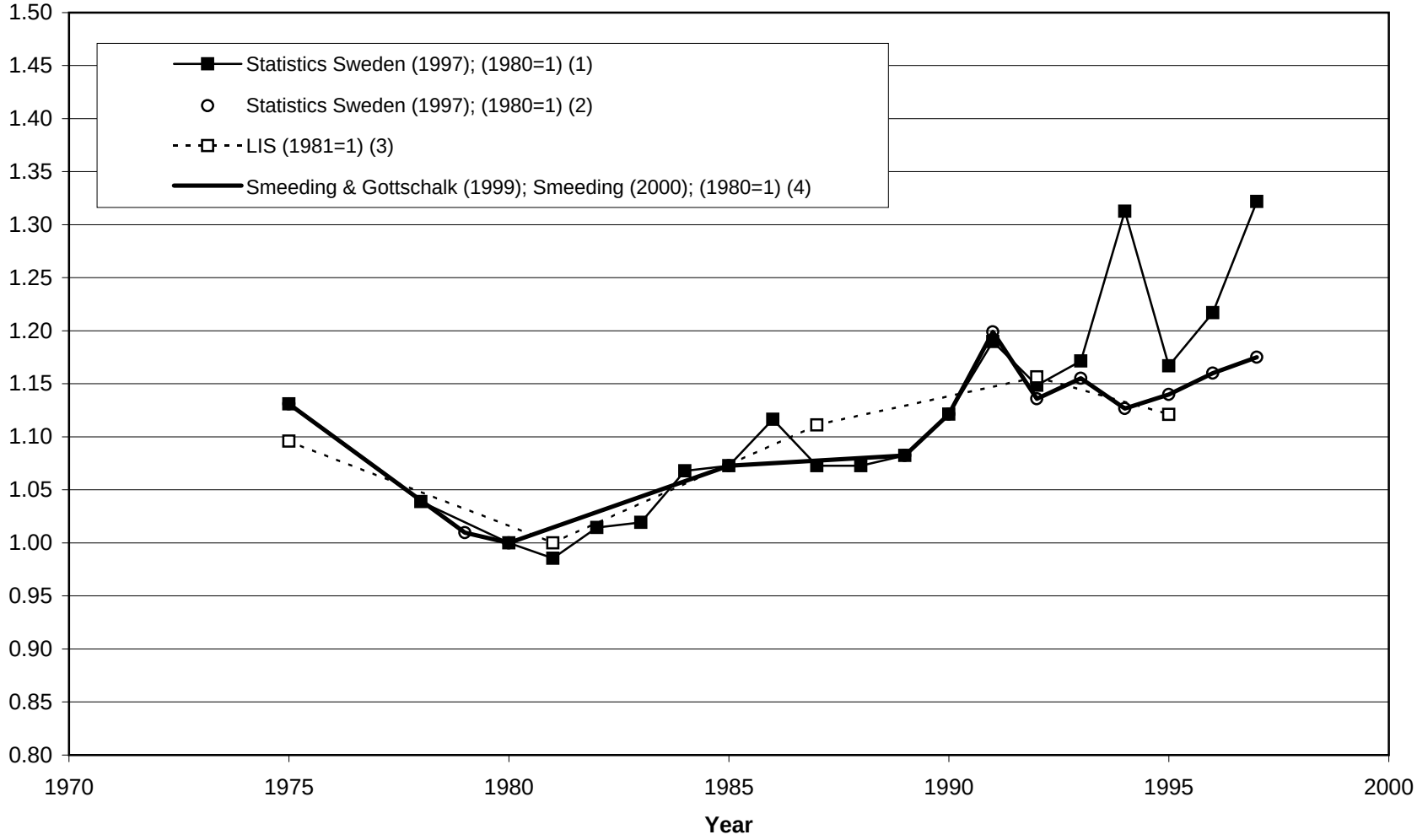
Figure 2.
Inequality in Spain: An Illustration of Three Pitfalls

(a) The Danger of Making a Trend Estimate from Only Two Points
(b) Peak to Peak; Trough to Trough
(c) Mixing Datasets (LIS and ECHP)



Sources: LIS Database (Gottschalk and Smeeding, 2000) and the European Community Household Panel (ECHP) database.

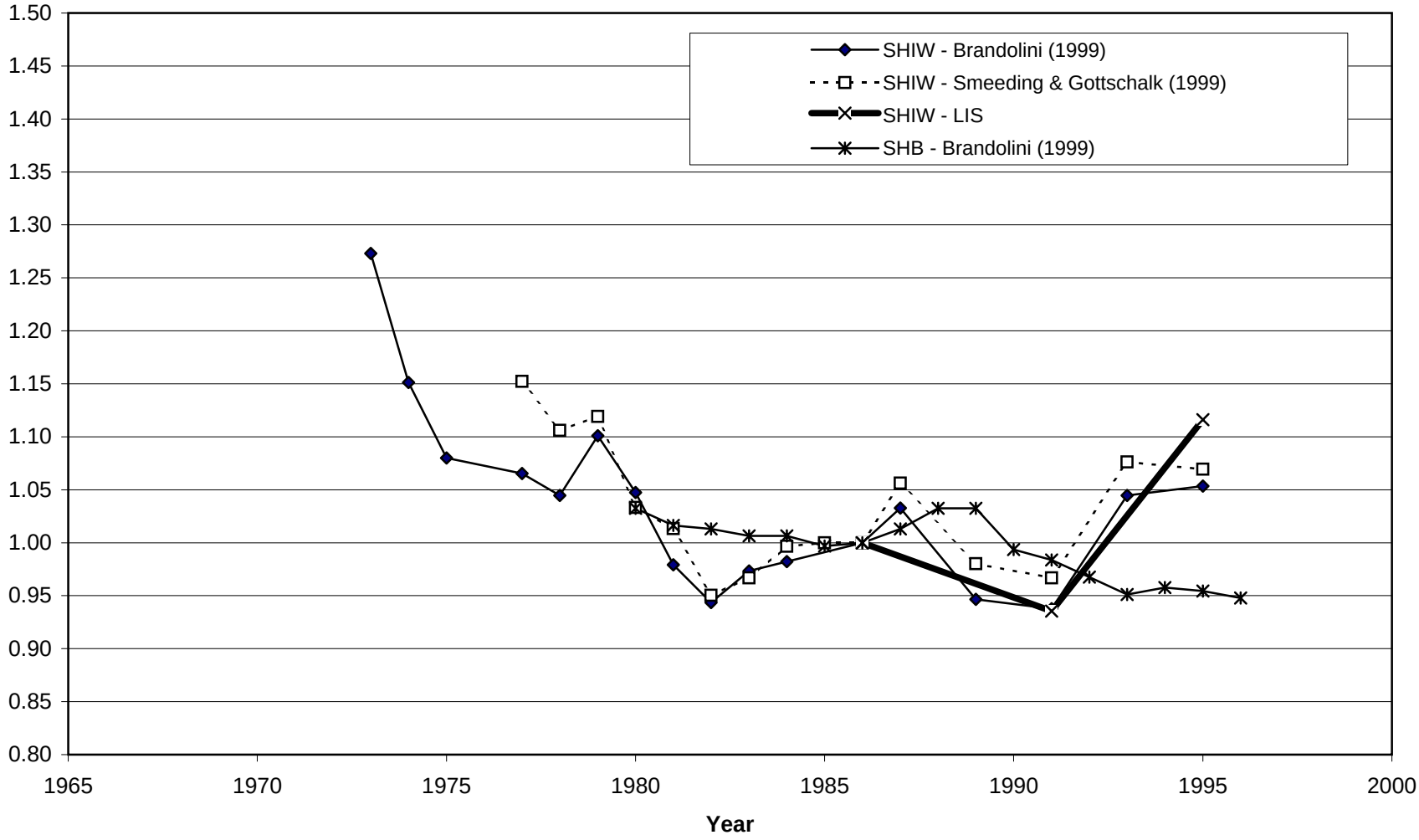
Figure 3.
Trends in Income Inequality: Gini Coefficients in Sweden



Notes: See text for complete explanation. - (1) Series with capital gains. - (2) Series without capital gains. - (3) LIS series with 1995 value based on complete households; earlier values based on tax filing units. - (4) Interpolated series based on judgement of authors.

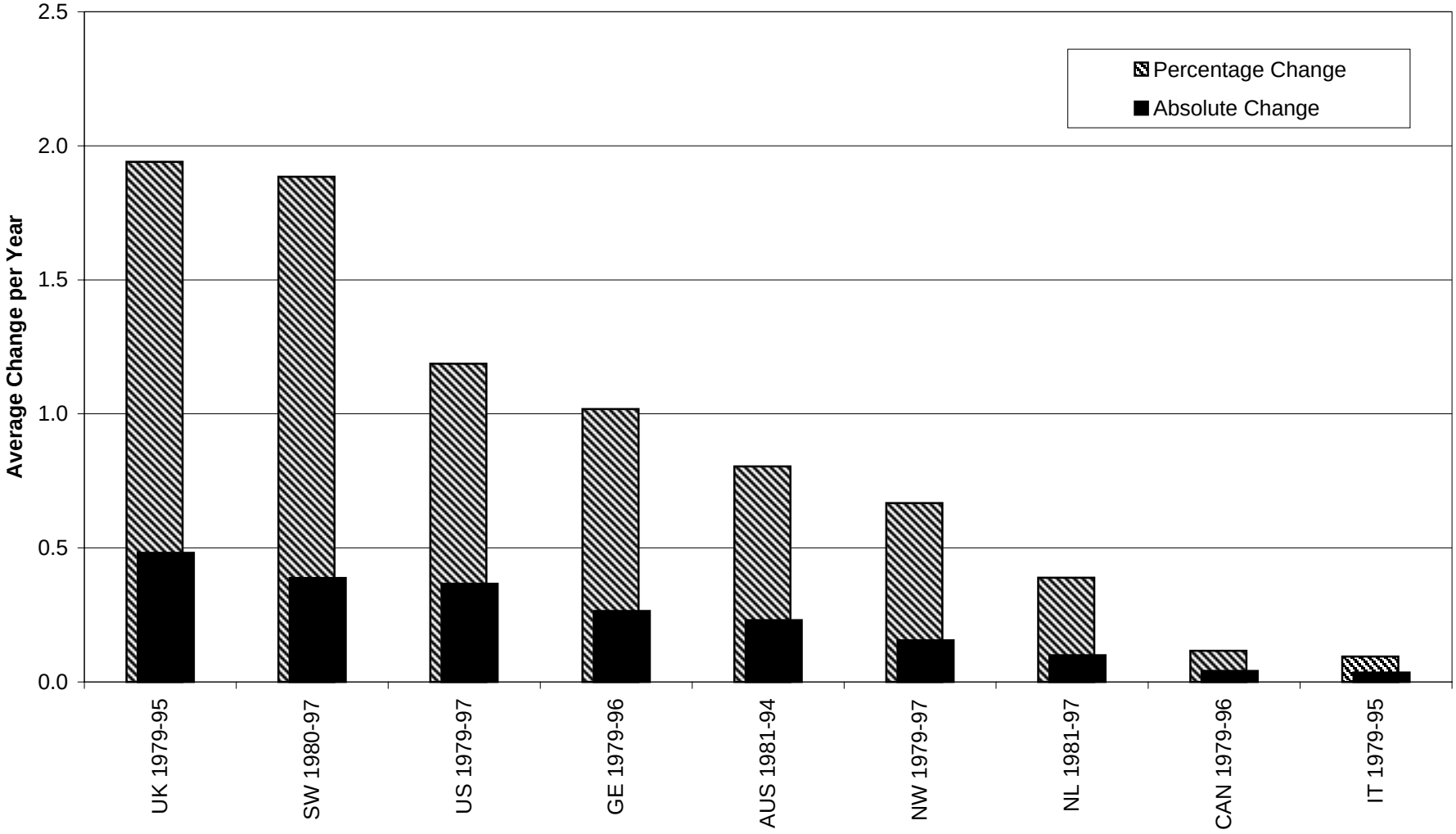
Sources: Statistics Sweden (1997); LIS (2000); Gottschalk and Smeeding (2000).

Figure 4.
Trends in Income Inequality: Gini Coefficients (1986=1) in Italy



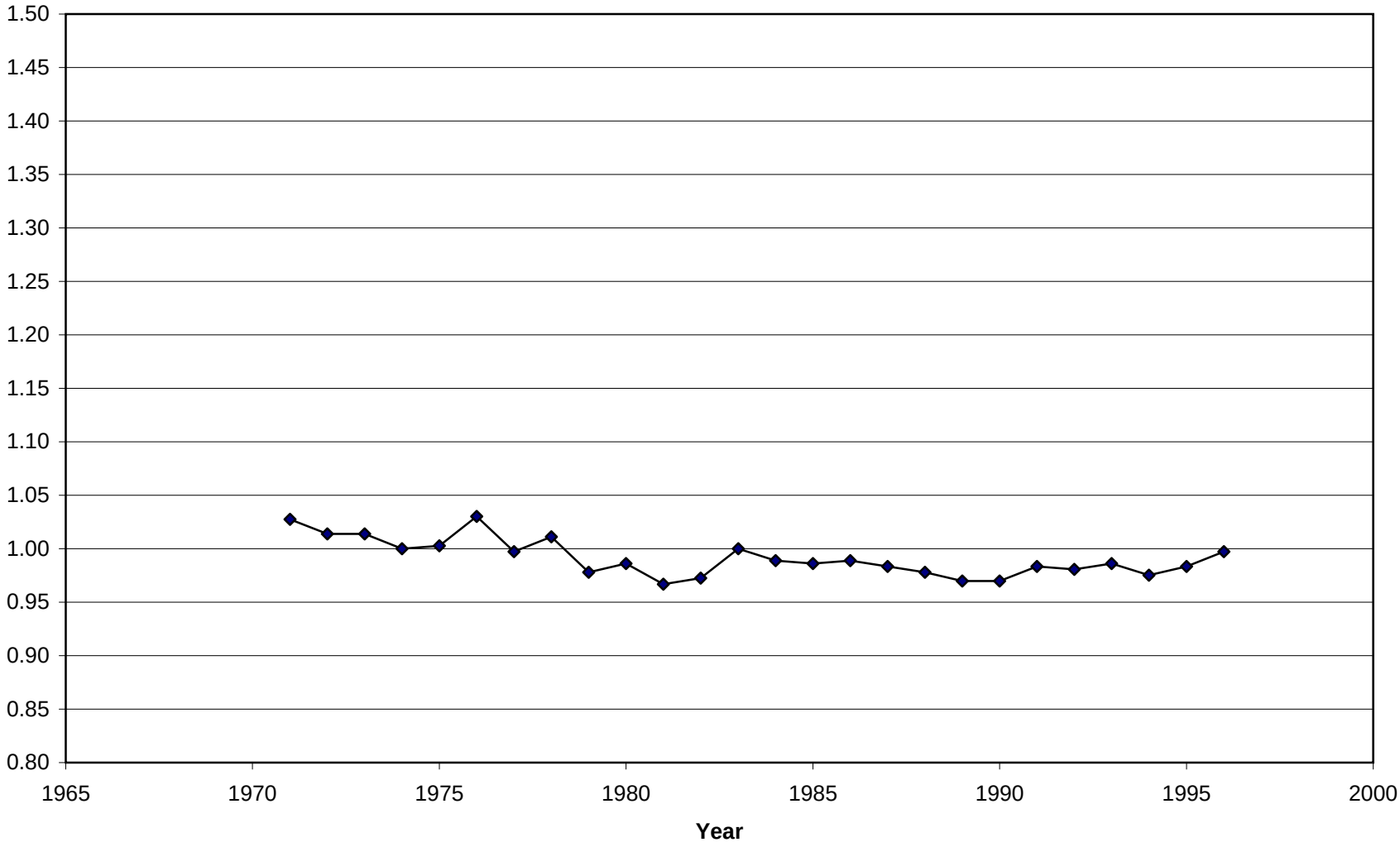
Sources: Brandolini (1998); Gottschalk and Smeeding (2000); LIS (2000).

Figure 5.
Trends in Income Inequality (Gini coefficients).
Percentage Change per Year and Absolute Change per Year: 1979-97



Source: Gottschalk and Smeeding (2000) and authors' calculations.

Figure 6.
Trend in Income Inequality: Gini Coefficients (1983=1) in Canada



Source: Statistics Canada (1999) as found also in Atkinson (2000) and in Gottschalk and Smeeding (2000).