

Atkinson, Tony; Brandolini, Andrea; Smeeding, Timothy

Working Paper

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

LIS Working Paper Series, No. 295

Provided in Cooperation with:

Luxembourg Income Study (LIS)

Suggested Citation: Atkinson, Tony; Brandolini, Andrea; Smeeding, Timothy (2001) : Producing Time Series Data for Income Distribution: Sources, Methods and Techniques, LIS Working Paper Series, No. 295, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:

<https://hdl.handle.net/10419/160967>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper No. 295

**PRODUCING TIME SERIES DATA FOR
INCOME DISTRIBUTION: SOURCES,
METHODS AND TECHNIQUES**

**Tony Atkinson
Andrea Brandolini
Timothy M. Smeeding**

January 2002

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

Tony Atkinson (Nuffield College, Oxford)

Andrea Brandolini (Bank of Italy, Rome)

and

Timothy Smeeding (Syracuse University, USA)

January 2001

Corresponding author: Timothy Smeeding, Center for Policy Research, The Maxwell School, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020 USA, fax: +315 4431081, Phone: +315 4439042, e-mail: tmsmeed@maxwell.syr.edu. 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 Paul vanderLaan, Anne Harrison, Michael Ward, and Michael Wolfson for their comments on an earlier draft of this paper. 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.

Abstract

This chapter addresses the important issue of the quality of time series data on income distribution. We hope to suggest both standards and practice patterns that will improve the production and use of time series data on inequality. Thus, we address three groups: primary data producers; secondary data producers (who assemble databanks from primary data sources), and users. We consider measurement error in both theory and practice. Along the way we include a number of practical suggestions for improving the comparability of time series data and trend analyses for income distribution.

1. Introduction and Overview

This chapter concerns itself with methods and sources for producing better time series data for income distribution research, a subject close to Richard Hauser's heart and one to which he has made great contributions, including many in recent years (e.g., Hauser 1996, 1999; Hauser and Becker 1997, 2000). Throughout his career, Richard has been remarkable for his care in the use of micro-data, for his imaginative and meticulous studies of income distribution, and for the generosity with which he has made available his material. He has greatly enriched comparative study of inequality, poverty and social policy, and so we dedicate this entry to him.

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 may appear to be based on more consistent data than are cross-country comparisons because they tend to come from the same producer. However, this presumption is unwarranted if the producer changes definitions, survey practices, or the data source experiences non-random errors which change over time. There are, in fact, many different cases where published time series are not internally consistent, and 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 information about 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. The recent experience of the European Community Household Panel (ECHP) has shown how much care has to be taken to achieve complete harmonization across countries even when there is a clear objective from the start. The Luxembourg Income Study (LIS)—a project to which Richard Hauser has greatly contributed, both by his initial participation in 1983 and also by serving as Chairman of the Executive Board from 1991 to 1997—has made considerable progress in point-in-time cross-national consistency. However, 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 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 chapter addresses three groups of statisticians and researchers:

- ***Time Series Data Originators (Producers)***. The central statistical offices and other 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.
- ***Secondary Time Series Data Suppliers***. 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). Recent producers include Tabatabai (1996), Deininger and Squire (1996), WIDER (1999) and others. LIS provides harmonized data and the possibility for time series comparisons but only for a limited number of years.

- ***Time Series Data Users.*** Researchers and policy analysts who use these time series and who have to make 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 chapter will help set standards for provision of better data, and robustness assessment of time series data. We begin with a conceptual section, based on Gottschalk and Smeeding (2000), which describes some of the issues involved with cross-national comparisons of both levels and trends in inequality.¹ We then consider in turn each of the three groups identified above. This chapter is a natural complement to recent Canberra Group papers by Smeeding et al. (2000) on point in time comparability; Harris (2000) on robustness; and Epland and Jansson (2000) on data presentation issues. For more on the Canberra group, see International Expert Group on Household Income Statistics (2000).

2. Measurement Error and Inequality Comparisons

Cross-time and cross-country comparisons of income inequality must confront the fact that the available data are subject to measurement error. The data fall a long way short of the measure one would ideally use to compare distributions of well-being across countries and over time. What is the impact of using imperfect data to approximate this ideal?

Within-Country Trends

To focus attention on the key elements consider the following simple error components model for the j^{th} percentile in year t :

$$\ln P_t^j = \ln \pi_t^j + m_t^j, \quad (1)$$

$$m_t^j = d_t + v^j + e_t^j, \quad (2)$$

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 ratio $\ln(P_t^{90} / P_t^{10})$ in a single year, which we call the 90/10 or decile ratio for convenience. Since

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

we see right away that measurement error that affects all deciles equally in the year (d_t) cancels.

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 $t+1$:

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

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, under-reporting by respondents at the top or bottom of the distribution will not bias inter-temporal comparisons to the extent that this under-reporting 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 under-reporting

is greater at the 10th than at the 90th percentiles *and* this degree of differential under-reporting also differs across years. This may happen, for instance, if under-reporting is linked to a particular income source, such as investment income, whose relative contribution varies over time and which is a higher proportion of incomes at the top compared to the bottom of the income distribution.

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 reduced too. 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)

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

with 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):

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

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 eliminate decile-specific errors that are constant across years (the v 's in equation (3)), reduce 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 sources of measurement error for these comparisons requires that we enter country c explicitly into equations (1) and (2):

$$\ln P_{ct}^j = \ln \pi_{ct}^j + m_{ct}^j, \quad (6)$$

$$m_{ct}^j = d_c^j + v^j + e_{ct}^j + h_c, \quad (7)$$

$$e_{ct}^j = g_{ct} + w_t^j + f_{ct}^j, \quad (8)$$

where d_c^j is a country-specific time-invariant component that affects differently each decile j ; h_c is a country-specific, time and decile invariant component (a pure “country” effect which equally affects each decile); g_{ct} is a time-specific component that affects all deciles equally in country c ; w_t^j is a time-specific component that has common effects across countries but differential effects across deciles; and finally, 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

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

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 comparisons in percentile ratios such as the decile ratio, within a country or across countries. 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 the 90/10 ratio in a single country, nor cross-national comparisons.
- Measurement error that is time invariant (v^j in case of single country; $d_c^j + v^j + h_c$ 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 ($d_c^j + v^j + h_c$), 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.

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

For decades, National Statistical Offices, Census Bureaus, Finance Ministries, Social Security Bureaus, and others have produced time series estimates of income distribution for national audiences. In some countries these are published in widely accessible form; in others

they are obscurely published, or are available as internal memos, not as primary publications. This section of the chapter addresses the issues involved in producing these series from a statistical and robustness point of view.

Documentation

The first step is the documentation of data quality (sampling and non-sampling errors, imputations, simulations, etc.) and such key matters as the income definitions, inequality measures, and top- or bottom-coding of data. 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). Harris (2000) and the LIS Technical Documentation template provide useful examples in terms of a Robustness Assessment Report (RAR). The rest of this section of the chapter might be seen as a list of elements which are necessary ingredients for such an inter-temporal RAR. Specific examples of RAR's and their template are available from the LIS website (www.lissy.ceps.lu/Canberra).

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);² 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.³ While these aspects are critical for any point in time measure, both one-time choices and 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 generally 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 year's 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, there are initiatives pushing 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 immigration 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.⁴

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., the Theil index, 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 income-ranked population subgroups. The availability of the “elementary”

data—percentile shares of income, mean incomes by percentiles, percentile values, along with information on top- and bottom-coding of incomes—provides the analysts with the raw material to construct their own summary measures of inequality. 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, then it would be helpful if these are listed 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. See, for instance, the multiple estimates for the Netherlands shown in Figure 1, which could lead the reader to draw different conclusions about the time shape of inequality change. Was inequality rising or falling in the late 1970s/early 1980s? Is there a “step” post 1985 or a continuing trend? An example of the kind of report which is most valuable is that by the Swedish Finance Ministry, which helps one sort out different series of changes in the Swedish income distribution since 1990 (Eriksson and Petterson 2000).

4. The Role of Secondary Data Collections

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 goal of the producer of a secondary dataset is typically the production of time series estimates for multiple nations. As a result, the producer 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 time series of continuous data.

Tax records suffer from potentially serious problems:

- the coverage of those with incomes below the tax threshold may be incomplete, a problem which varies over time with the tax base;
- certain types of income tend to be under-reported;
- 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. Ideally use of these files would involve direct matching of individual files by a personal identifier, but this 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 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, and of item non-reporting and item misreporting (usually under-reporting) by respondents, all of 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 grossing up factors and by revisions (for example, where decennial census results become available). 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. 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 non-sampling errors may differ over time. Changes in adjustments for mis-reporting errors, non-reporting errors, and other non-sampling 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 consistency may be lost. Secondary producers need to be aware of these problems 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. One consequence is that the final estimates may be in the form of tabulated ranges, adjustments being made at that level, and there are no corresponding micro-data.

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 harmonised 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 specialised

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 3 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.⁵ 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 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,⁶ 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. One reason for so doing is that popular methods such as the Pareto interpolation may yield estimates of income shares which lie outside the theoretical

bounds.

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.⁷ 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 and practical problems found in “original” datasets. Moreover, they attempt to make these series comparable over time and across countries. Their task is a most difficult and complicated one. Hence, we salute these efforts, while hastening to add that the devil is always in the details of their estimates, so please do not skimp on the details.

5. Trends in Income Inequality: The Researchers’ and Users’ Perspective

The previous sections of this chapter addressed inequality time series data from the point of view of the central statistical office (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 collection of lessons learned by the authors in writing several papers on this topic (Atkinson 1997 and 2000; Atkinson and Brandolini 2000; Brandolini 1998; Gottschalk and Smeeding 1997 and 2000; Smeeding 2000). The hope is that others may avoid some of the pitfalls we have noted, or even fallen into ourselves.

Detecting Trends

A sampling of the problems that may arise includes:

- **Limited point trends.** In several nations, comparable household income microdata are only available for a small number of periods. Having two periods permits the user to estimate the change between them, but may convey a misleading impression of the underlying trend. There is a considerable danger in taking a very small number of years 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 not allow for the differences in business cycle 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 a more satisfactory measure, but requires a lengthy time series of estimates (e.g., see Burkhauser, Crews, Daly, and Jenkins 1999).

- ***Mixing datasets and definitions.*** Some trend analyses have mixed sources with different income definitions and/or several datasets over time (e.g., see Figure 1). There is a risk that the recorded trend will reflect *both* the “real” inequality change *and* differences across datasets.

We illustrate all three of these issues in Figure 2, using Spanish income distribution data for 1980 and 1990 from LIS (based on the Spanish Income and Expenditure Survey) and for 1993 from the ECHP, and business cycle trends (loosely based on GDP and unemployment figures from the OECD). The LIS datasets indicate a downward trend in inequality between 1980 and 1990, but we must observe that the business cycle reached a trough in 1990 (though its value then was higher than ten years earlier). When the ECHP is added, inequality increases in 1993 relative to the previous observations, but this occurs when the business cycle is heading toward a peak, suggesting a possible cyclical effect. On the other hand, we are drawing such conclusion from mixing two different datasets, so that the change could come from differences in the underlying data themselves.

- ***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 official publications by Statistics Sweden. The LIS data show 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 and this affects trend estimates of inequality where living arrangements or numbers of young adults change over the period in question. This LIS time series is therefore not the best one for measuring the Swedish inequality trend (see Eriksson and Petterson 2000). The differences between the two Statistics Sweden series 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, the direction of change 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 regard 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 judgements about which series, sets of series, or combinations of series produce the most reliable estimates. 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.

Significance of Changes

There are no generally accepted standards for labeling significance of inequality changes. In the literature, authors have used clear-cut 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” (Förster 2000), or a “3 to 7 percentage point change” (Gottschalk and Smeeding 2000; 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.⁸

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 1997). 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 are 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 one starting 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 chapter. The method is to calculate the annual percentage change in the Gini coefficient (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 Gini coefficients 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 change 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 6). 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 4). 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.

6. 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 examine 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. Issues of data quality—issues in which Richard Hauser has taken a keen professional interest and has made numerous contributions—are therefore of great significance. In this chapter, we have indicated some of the necessary 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.⁹ We thank pioneers like Richard Hauser for the care and insight they have shown in making use of income microdata. We can only endeavor to follow his path, as we hope this paper has shown. There is clearly much room for improvement in our time series data on income distribution.

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 No. 3. Helsinki, Finland: United Nations World Institute for Development Economics Research.
- Atkinson, A.B. and A. Brandolini. 2000. "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries," Banca d'Italia, Temi di discussione, No. 379, October. Forthcoming in *Journal of Economic Literature*.
- Atkinson, A.B., L. Rainwater, and T.M. Smeeding. 1995. *Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study*. Paris: Organisation for Economic Co-operation and Development.
- Brandolini, A. 1998. "A Bird's-Eye View of Long-Run Changes in Income Inequality," unpublished manuscript. Rome: 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(2) (September): 183-239.
- Burkhauser, R.V., A.C. Crews, M.C. Daly, and S.P. Jenkins. 1999. "Testing the Significance of Income Distribution Changes Over the 1980's Business Cycle: A Cross-National Comparison," *Journal of Applied Econometrics*, 14(3) (May/June): 253-272.
- Canberra Group. 2000. *Papers and Final Report of the Fourth Meeting on Household Income Statistics, Differdange, Luxembourg, 15-17 May 2000*. Differdange, Luxembourg: Luxembourg Income Study at CEPS/INSTEAD. <<http://www.lis.ceps.lu/Canberra>>.
- 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. Bourguignon (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(3) (September): 565-591.
- Epland, J. and K. Jansson. 2000. "Data Presentation Issues," unpublished manuscript, Canberra Group Meeting. Luxembourg: Luxembourg Income Study, May.
- Eriksson, I. and T. Petterson. 2000. "Income Distribution and Income Mobility—Recent Trends

- in Sweden.” In R. Hauser and I. Becker (eds.), *The Personal Distribution of Income in an International Perspective*. Berlin: Springer, pp. 1589-175.
- Flora, P. 1983. *State, Economy, and Society in Western Europe 1815-1975: A Data Handbook in Two Volumes*. Frankfurt am Main; Chicago: Campus Verlag; St. James Press.
- Förster, M. 2000. “Trends and Driving Factors in Income Distribution and Poverty in the OECD Area,” *Labour Markets and Social Policy*, Occasional paper #42. Paris: Organisation for Economic Co-operation and Development, August.
- Gottschalk, P. and T.M. Smeeding. 1997. “Cross-National Comparisons of Earnings and Income Inequality,” *Journal of Economic Literature*, 35: 633-687.
- Gottschalk, P. and T.M. Smeeding. 2000. “Empirical Evidence on Income Inequality in Industrialized Countries”. In A. B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*. New York: Elsevier-North Holland Publishers, pp. 261-308.
- Harris, G. 2000. “Quality Assessment Reports (RAR) for Micro and Macro,” unpublished manuscript, Canberra Group Meeting. Luxembourg: Luxembourg Income Study, May. Available at <<<http://www.lissy.ceps.lu/Canberra.html>>>
- Hauser, R. 1996. “Vergleichende Analyse der Einkommensverteilung und der Einkommensarmut in den alten und neuen Bundesländern von 1990 bis 1995.” In I. Becker and R. Hauser (eds.), *Einkommensverteilung und Armut in Deutschland von 1962 bis 1995*, Arbeitspapier Nr 9, EVS-Projekt, Universität Frankfurt am Main.
- Hauser, R. 1999. “Personelle Primär- und Sekundärverteilung der Einkommen unter dem Einfluss sich ändernder wirtschaftlicher und sozialpolitischer Rahmenbedingungen – eine empirische Analyse auf der Basis der Einkommens- und Verbrauchstichproben 1973-1993,” *Allgemeines Statistisches Archiv*.
- Hauser, R. and I. Becker. 1997. “The Development of the Income Distribution in the Federal Republic of Germany during the 1970s and 1980s.” In P. Gottschalk, B. Gustafsson, and E. Palmer (eds.), *Changing Patterns in the Distribution of Economic Welfare: An International Perspective*. Cambridge, MA: Cambridge University Press, pp. 184-219.
- Hauser, R. and I. Becker (eds.). 2000. *The Personal Distribution of Income in an International Perspective*. Berlin: Springer-Verlag.
- 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.
- 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.

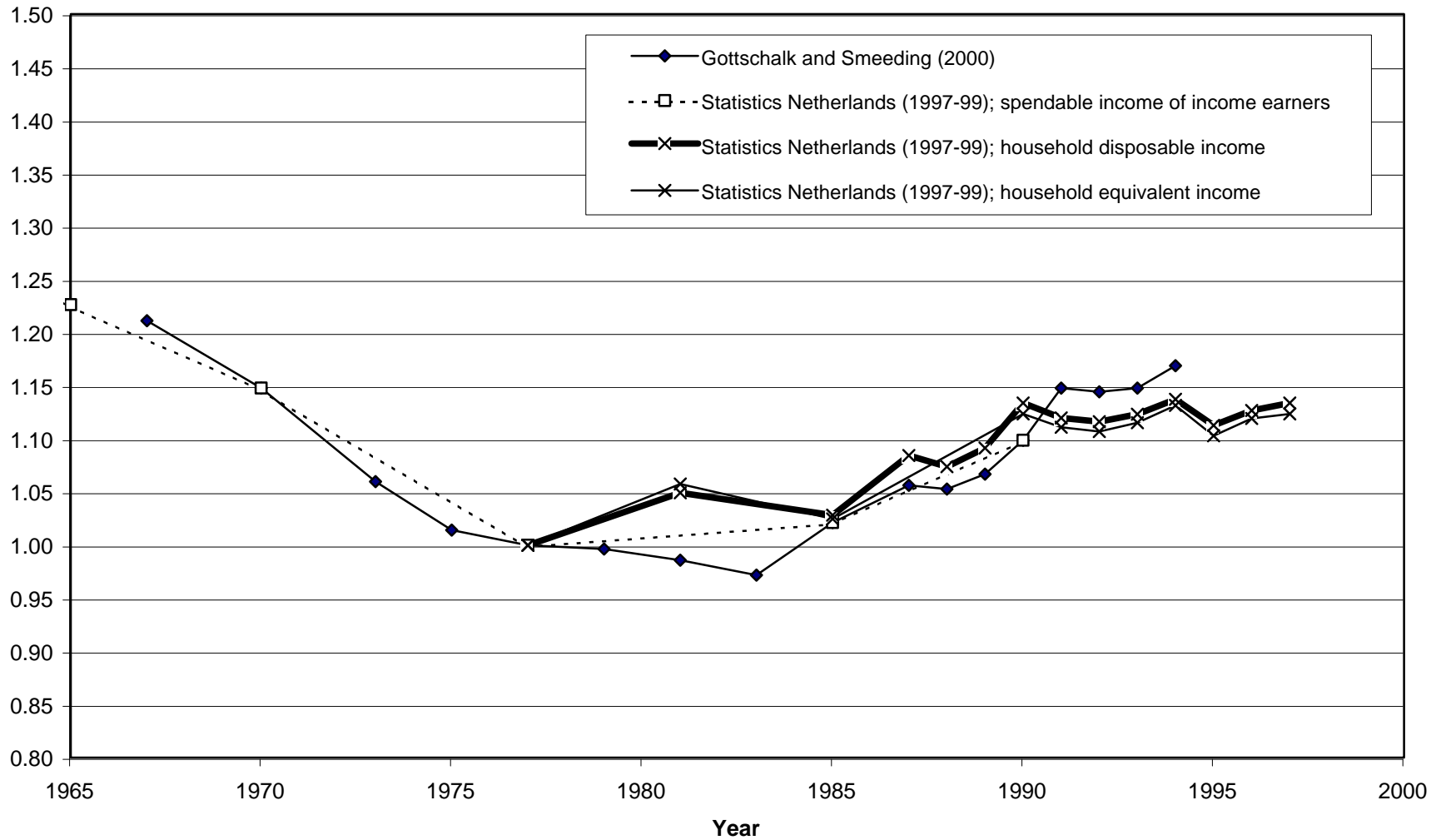
- Paukert, F. 1973. "Income Distribution at Different Levels of Development: A Survey of Evidence," *International Labour Review*, 108: 97-125.
- Ryscavage, P. 1995. "A Surge in Growing Income Inequality?" *Monthly Labor Review*, 118(8): 51-61.
- Sawyer, M. 1976. "Income Distribution in OECD Countries". *OECD Economic Outlook. Occasional Studies*. Paris: Organisation for Economic Co-operation and Development.
- 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: Springer-Verlag, pp. 205-224.
- Smeeding, T., M. Ward, I. Castles, and H. Lee. 2000. "Cross-Country Comparisons of Inequality," unpublished manuscript, Canberra Group Meeting. Luxembourg: Luxembourg Income Study, May. Available at <<http://www.lissy.ceps.lu/Canberra.html>>
- Szekely, M. and M. Hilgert. 1999. . "What's Behind the Inequality We Measure: An Investigation Using Latin American Data for the 1990s," Research Department Working Paper #409. Washington, DC: Inter-American Development Bank, December.
- Tabatabai, H. 1996. *Statistics on Poverty and Income Distribution: An ILO Compendium of Data*. Geneva: International Labor Office.
- 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.
- 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," *Economic Survey of Europe in 1965, Part II*. Geneva: United Nations.
- World Institute for Development Economics Research (WIDER). 1999. *World Income Inequality Database*, Beta 3. Helsinki, Finland: World Institute for Development Economics Research, November 8.

Endnotes

1. For instance, see the list of sources in Atkinson (2000), Atkinson and Brandolini (2000), Atkinson, Rainwater, and Smeeding (1995), Brandolini (1998), Gottschalk and Smeeding (1997, 2000), Förster (2000), Smeeding (2000), and others who have collected time series estimates of income inequality measures from various sources.
2. Notice that welfare weighting is a different issue from re-weighting sample data to allow for differential sampling or non-response.
3. 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).
4. Additional comment on using fixed effect adjustments to interpolate series is found below.
5. For instance, in the United States several income items are recorded with a pre-set upper limit (see U. S. 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.
6. This is the case of the United States Gini coefficients (see U. S. Bureau of the Census 1998, p. A1).
7. 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 (1983) 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 and co-authors at WIDER (1999). 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 (2000) .

8. 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.
9. 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).

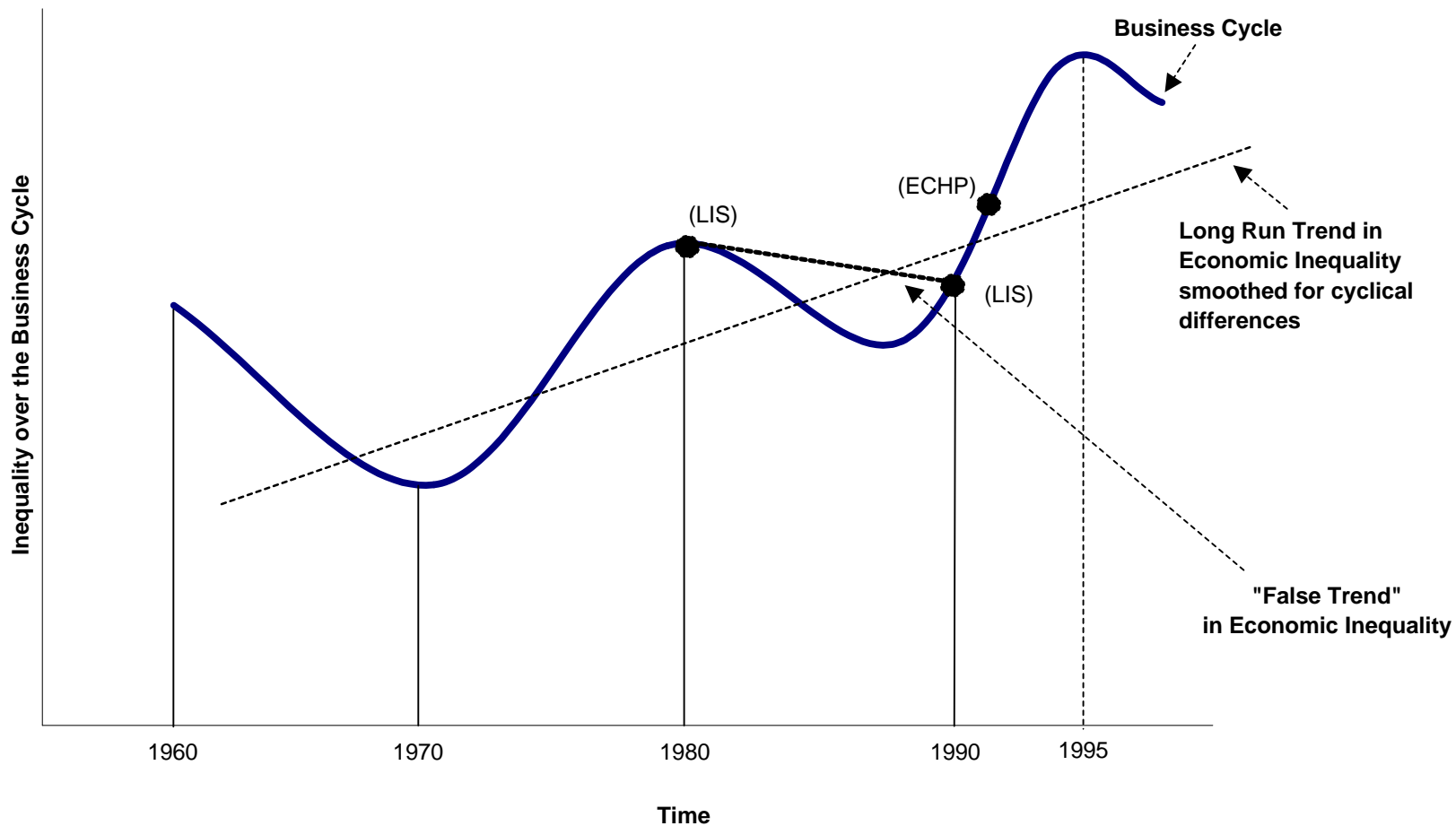
Figure 1.
Trend in Income Inequality: Gini Coefficients (1977=1) in Netherlands



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

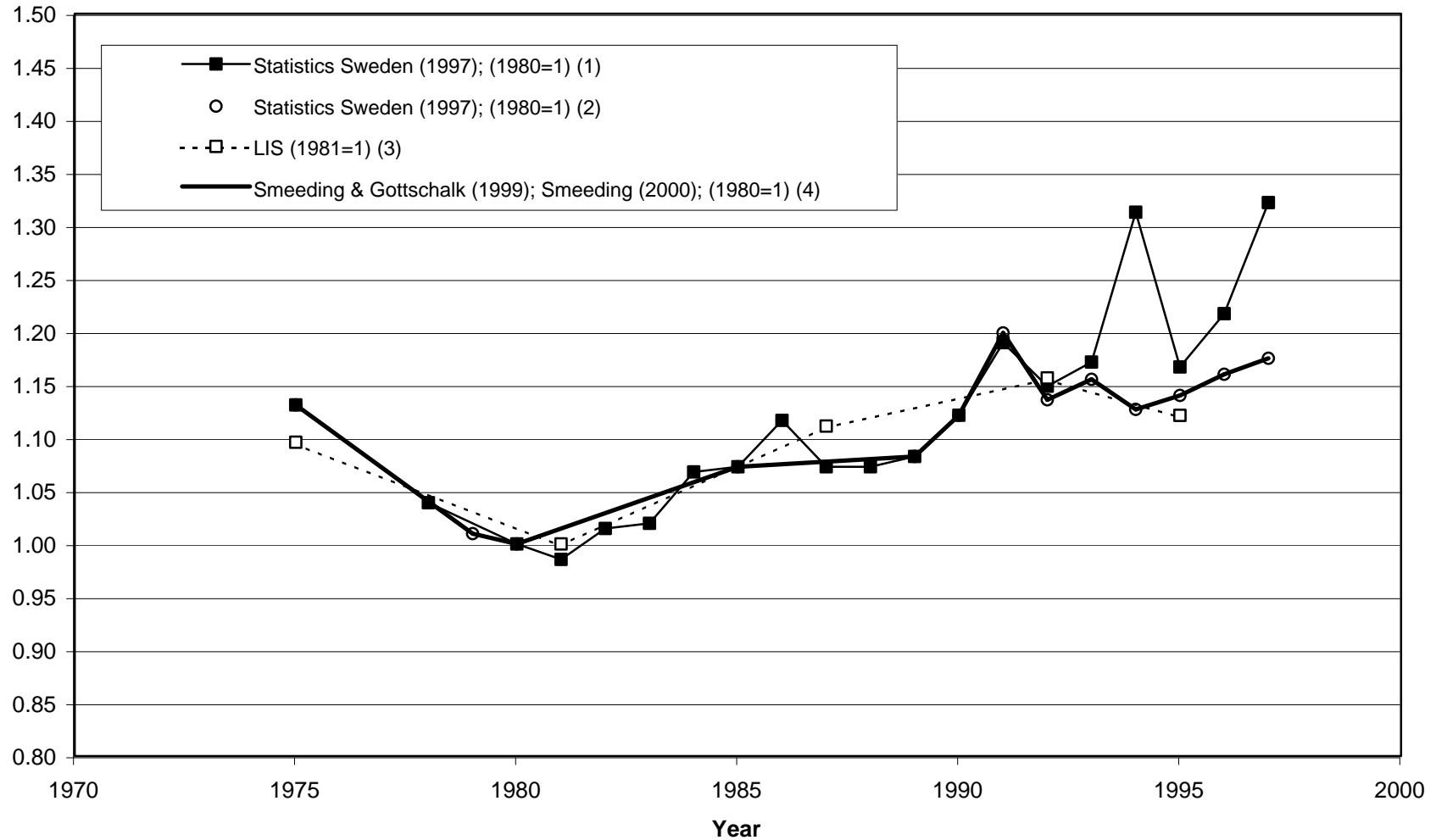
Figure 2. Inequality in Spain: An Illustration of Three Pitfalls

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



Source: LIS database (Gottschalk and Smeeding, 2000) and European Community Household Panel (ECHP) database; authors' estimates of business cycle patterns.

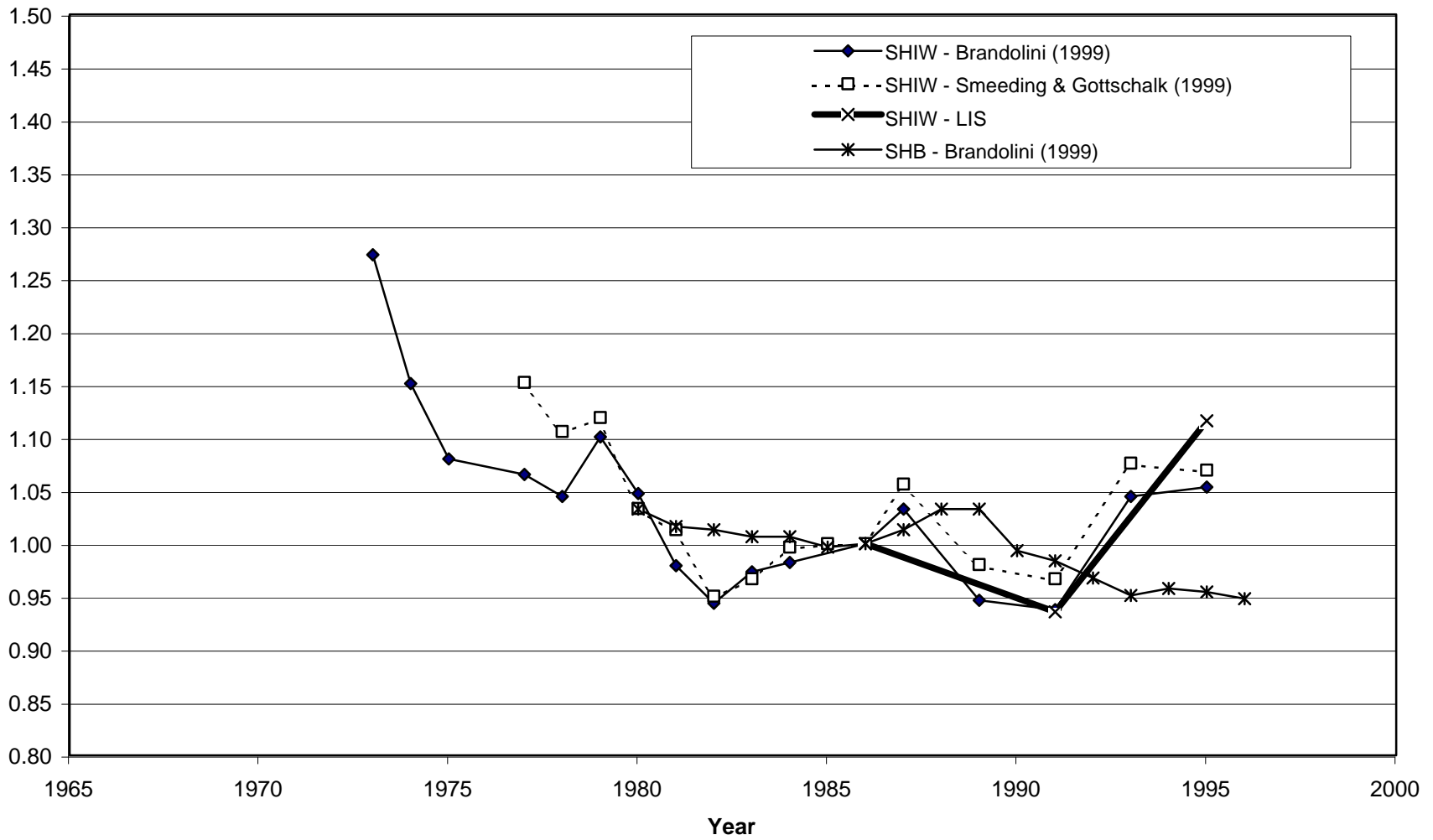
Figure 3.
Trends in Income Inequality: Gini Coefficients in Sweden



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.

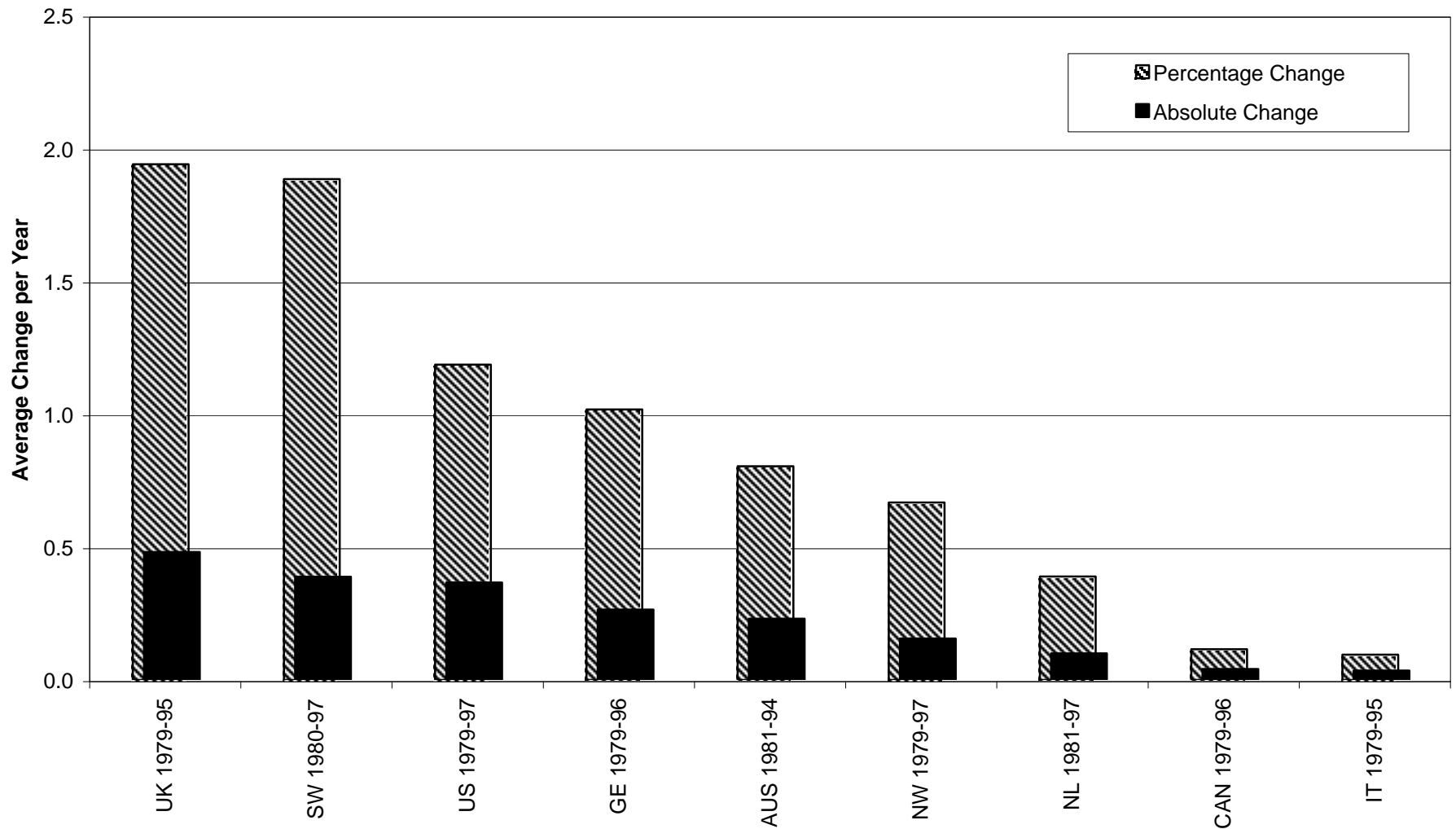
Sources: Statistics Sweden (1997); LIS (2000); Gottschalk and Smeeding (2000).

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



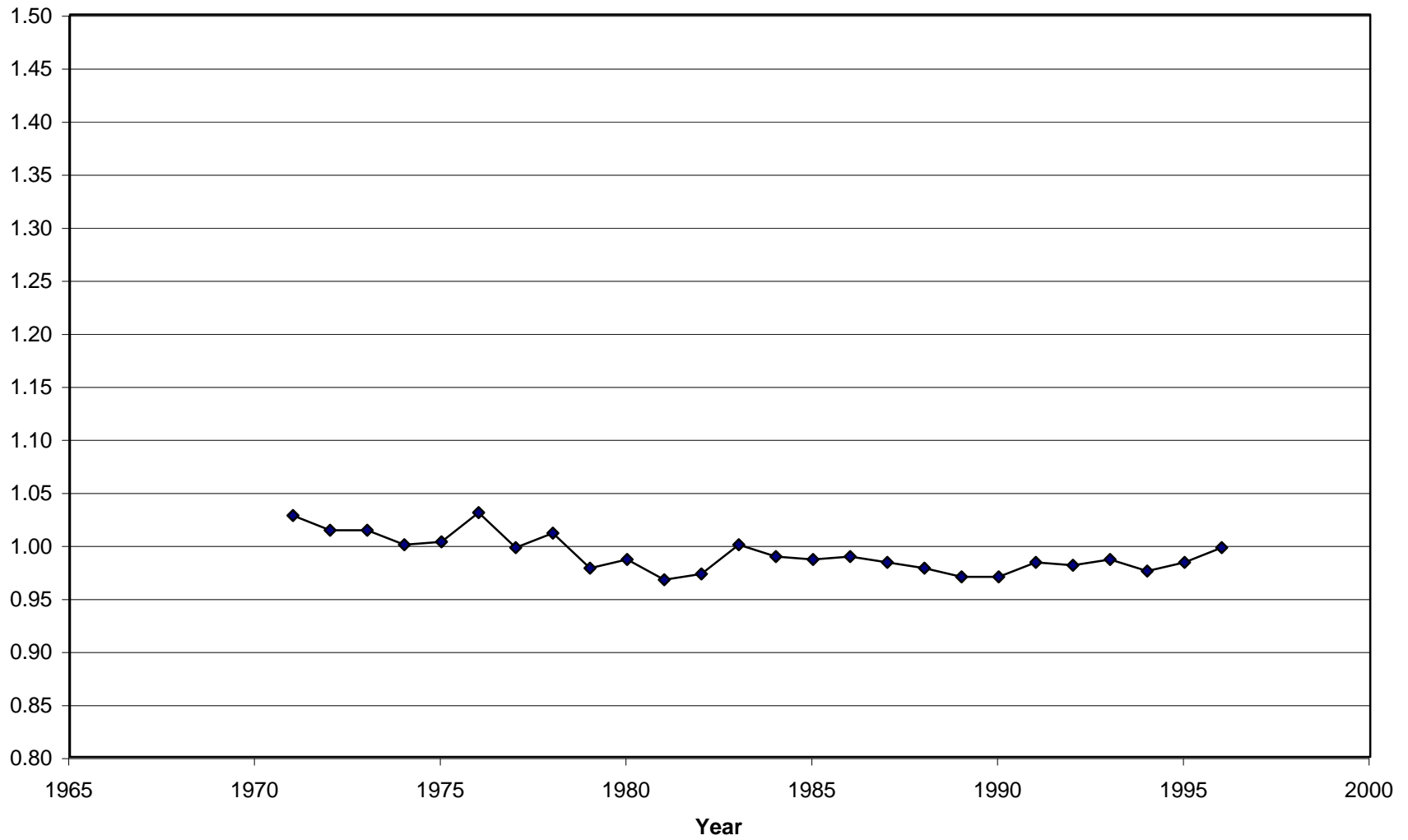
Sources: Brandolini (1998); Gottschalk and Smeeding (2000); LIS (2000).

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).