Methodological Issues in the Measurement of Income and Poverty

vorgelegt von
Jan Goebel, M.A.

Fakultät VIII - Wirtschaft & Management
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Dr. rer. oec. –

genehmigte Dissertation

Promotionsausschuss
Vorsitzender: Prof. Dr. Klaus-Dirk Henke
Berichter: Prof. Dr. Gert G. Wagner
Prof. Dr. Christof Helberger

Tag der wissenschaftlichen Aussprache: 05.02.2007

Berlin 2007
D 83

Danksagung

Ich hätte diese Arbeit nicht erstellen können ohne die Hilfe von zahlreichen Menschen (und auch frei verfübaren Programmen). All diesen möchte ich meinen tiefsten Dank aussprechen, dafür das sie mir geholfen haben und mir auch immer Geduld und Vertrauen entgegenbrachten.

Mit dem Wind,
den man selber macht,
lassen sich die Segel nicht füllen.
(*K.H. Waggerl*)

Wer fertig ist,
dem ist nichts recht zu machen;
ein Werdender wird immer
dankbar sein.
(*J.W. von Goethe*)

Summary

The main focus of this thesis are the important methodological problems that arise in the field of poverty and income analysis when using empirical data from large surveys. The problems of measuring poverty addressed here are the aggregation problem and the measurement of poverty over time, considering possible income mobility. In the empirical analysis, the results on the robustness for mean-based and median-based poverty lines as well as for monthly and annual income are presented. With the help of this concept, the differences in the consequences of mobility on poverty status are described for the USA and Germany. The results make clear that the kind of income used (pre- or post-government income) is very important when comparing countries that differ so widely, especially with respect to the characteristics of the welfare state.

Issues of the quality of data for empirical analysis in a wider sense are also discussed in two chapters. The first of these deal with the robustness of imputation methods using a special dataset of the Finnish part of the European Household Panel, in which register information is available for missing values on income questions. The empirical results show that multi-method imputation algorithms are more robust, especially if they allow the user to adjust all settings to the specific problem at hand.

The second chapter of these two deals with the problems of correctly measuring high-income households, carried out here with a special high-income sample from the German Socio-Economic Panel. The analysis also looks at the impact of better coverage of these cases, as well as the effects of integrating them into an ongoing panel survey. Furthermore, the distribution of the high-income households is described with the help of a Pareto distribution for both, the new and old samples.

The last chapter presents an empirical application of the decomposition of inequality as measured with the Gini coefficient by German regions. The regional variation in income inequality and the development over time in Germany since reunification is analyzed. Concluding from the empirical results with respect to post-government income, one must reject the hypothesis that East and West Germany are moving towards a common income distribution. The picture is much different concerning the pre-government income distribution, however: here we see that East Germany already surpassed the Western level of inequality in the early 1990s and that this difference has increased continuously since. Even when enlarging the number and structure of the regions under consideration by splitting the western part into its northern, central and southern components, a clear picture of East Germany as still very different from the rest of the country is seen.

Zusammenfassung

Der Hauptfokus dieser Arbeit liegt auf methodischen Problemen bei der Messung von Einkommen und Armut mit Hilfe von empirischen Daten aus Bevölkerungsumfragen. Im Bereich Armutsmessung wird das Aggregationsproblem sowie die Messung von Armut über die Zeit unter Berücksichtigung von Einkommensmobilität untersucht. In der empirischen Analyse zu diesem Kapitel werden Robustheitsanalysen für den entwickelten Messansatz unter Verwendung verschiedener Armutsgrenzen und Einkommenskonzepten durchgeführt. Mit der Hilfe des eingeführten Armuts-Stabilitäts-Koeffizienten werden die Unterschiede in den Auswirkungen der Einkommensmobilität auf die Persistenz von Armut in Deutschland und den USA verglichen. Es wird deutlich, das die Art des Einkommens (verfügbares oder Markt-Einkommen) entscheidend ist für den Vergleich von Armut für Länder mit starken Unterschieden in der Ausgestaltung des staatlichen Steuer- und Transfersystems.

Darüberhinaus ist Datenqualität im weiteren Sinne Gegenstand von zwei Kapiteln. Eines dieser Kapitel analysiert die Robustheit von Imputationsmethoden mit Hilfe eines speziellen Datensatzes des finnischen Teils des Europäischen Haushaltspanels, der für fehlende Einkommensangaben die entsprechenden Werte aus offiziellen Registern beinhaltet. Sogenannte "multi method" Imputations Strategien zeigen sich dabei als am robustesten, insbesondere wenn sie sehr flexibel auf die jeweiligen Anforderungen vom Nutzer angepaßt werden können.

Das anschließende Kapitel befaßt sich mit dem Problem der korrekten Erfassung von Haushalten mit hohen Einkommen. Die genutzte Datenbasis ist hierbei eine spezielle Teil-Stichprobe des Sozio-oekonomischen Panels. Analysiert werden die Auswirkungen einer besseren Abdeckung dieser Fälle und die Integration in ein laufendes Panel. Die Verteilung der hohen Einkommen wird mit Hilfe einer Pareto Verteilung geschätzt und für die neue Stichprobe und die laufende Panelbefragung verglichen.

Das letzte Kapitel nutzt eine Zerlegung der Ungleichheit, gemessen mit dem Gini Koeffizienten, nach Regionen in Deutschland. Die regionalen Unterschiede der Einkommensungleichheit und deren zeitliche Entwicklung seit der Wiedervereinigung wird dabei dargestellt. Auf Grund der empirischen Ergebnisse muß die Hypothese, dass Ost- und Westdeutschland sich bezüglich der Verteilung und Ungleichheit des verfügbaren Haushaltseinkommen angleichen als widerlegt angesehen werden. Allerdings ist die Situation gänzlich unterschiedlich hinsichtlich der Verteilung des erzielten Markteinkommens. Die gemessene Einkommensungleichheit ist bei diesem Einkommenskonzept bereits in den frühen 90er Jahren höher als in den alten Bundesländern und der Abstand wuchs seitdem kontinuierlich. Auch bei einer weiteren Differenzierung von Westdeutschland in einen nördlichen, zentralen und südlichen Teil bleibt die klare Unterscheidbarkeit Ostdeutschlands bestehen.

Contents

Introduction **1**

1 Measuring Aggregated Poverty **9**

1.1 Notations and definitions 11

1.2 Poverty axioms 11

1.3 Poverty measures 17

1.3.1 Distribution-Insensitive Poverty Measures 19

1.3.2 Sen-Type Poverty Measures 20

1.3.3 Ethical Poverty Measures 24

1.3.4 Distribution-Sensitive Poverty Measures 27

1.4 Summary 33

2 Measuring Poverty over Time **35**

2.1 The measurement of persistent poverty 36

2.1.1 Measuring permanent inequality 38

2.1.2 The measurement of persistent poverty reconsidered . . . 39

2.1.3 Poverty-line-independent measurement of persistent poverty 40

2.2 Empirical Application 42

2.2.1 Different concepts for measuring persistent poverty 42

2.2.1.1 Mean based vs. Median based poverty line . . . 42

2.2.1.2 Annual vs. Monthly Net Household Income . . . 44

2.2.2 Persistent poverty in the USA and Germany 45

2.2.2.1 Post-Government Income 46

2.2.2.2 Pre-Government Income 48

2.2.2.3 Poverty stability with shifting time windows . . 49

2.3 Summary 49

3 The Impact of Item-Nonresponse using a Multiple Imputation Strategy 53

3.1 Concepts for handling missing data by means of imputation . . . 54

3.2 Multiple imputation 57

3.3 Methodological framework 61

3.4 Data 64

3.4.1 Sample structure 64

3.4.2 Income variables 64

3.5 Generating multiple imputations 65

3.5.1 General description of the imputation approach 65

3.5.2 Single method imputation software 68

3.5.2.1 Software for Hot-deck Imputation 68

3.5.2.2 Imputation software 'NORM' 70

3.5.2.3 Imputation software 'aregImpute' 71

3.5.3 Multiple method imputation software 71

3.5.3.1 Imputation software 'IVEware' 71

3.5.3.2 Imputation software 'MICE' 73

3.5.3.3 Imputation software 'MVIS' 75

3.6 Evaluation of imputation methods 76

3.6.1 Hot-Deck Imputation 77

3.6.2 NORM Imputation 78

3.6.3 aregImpute Imputation 79

3.6.4 MICE Imputation 81

3.6.5 MVIS Imputation 82

3.6.6 IVEware Imputation 83

3.7 Impact on parameter estimation 85

3.8 Summary 92

4 Measuring Households with High Incomes – The Case of Germany 95

4.1 The High-Income Sample within the *SOEP* 96

4.1.1 Description of the *SOEP* Samples 96

4.1.2 High-Income Sample G 97

4.2 Distribution of High Income 102

4.3 Description of the income distribution by a parametrical model . 109

4.3.1 Estimation of the Gini coefficient with sparse data 111

4.3.2 The distribution of high incomes and the Pareto Law . . 113

4.4 Summary 119

5 Regional Income Stratification **121**
5.1 Income Distribution in West and East Germany 123
5.2 Emprical Analysis: methods and data 126
5.2.1 The ANOGI (Analysis of Gini) methodology 126
5.2.2 The data 128
5.3 Empirical Results 129
5.3.1 Income inequality decomposition by region: West and East Germany 129
5.3.1.1 Pre-government income 129
5.3.1.2 Post-government income 132
5.3.2 Income inequality decomposition using an extended regional grouping 134
5.3.2.1 Pre-government income 134
5.3.2.2 Post-government income 138
5.4 Summary 141

Conclusion and Discussion **143**

Bibliography **166**

A Imputation Strategies **167**

B High Income Distribution **169**

C Regional Stratification **179**

D List of Tables, Figures, Symbols and Index **181**
Symbols and Abbreviations 181
List of Tables 185
List of Figures 187

Introduction

Income is without doubt one of the main indicators of economic well-being and hence a key measure for the living conditions of individuals in general (e.g. poverty). Reliable income measures are therefore of central interest, not only for researchers but also for policy makers. In the majority of cases information about the distribution of income is derived from large-scale surveys. Thus the problems that arise when dealing with income questions in research can be grouped into three categories: first, the integration of income into a broader theory of welfare; second, the statistical estimation of distributional parameters or of the whole income distribution; and third, the methodological problems in using empirical surveys to obtain valid and reliable information about the *true* income distribution, mostly at a national level. This thesis deals with some specific aspects of the latter two.

The aim of this introduction is twofold. First it acts as a general guide for the reader, giving a broad overview of the Chapters that follow and briefly introducing the problems that will be covered in depth in individual Chapters. Second, given that a thesis dealing with methodological issues in the wider field of income measurement cannot be exhaustive and two of the three aforementioned areas were selected as a focus, it explains the rationale for this selection of specific aspects as most important.

The thesis is structured around the following selected topics. Chapter 1 gives an overview of the measurement of a specific, however very important, parameter of the income distribution: the estimation of poverty. In Chapter 2 a new approach for a consistent measurement of poverty stability over time is proposed. Chapter 3 and 4 deal with the empirical measurement of income itself with the help of large social surveys, and in particular with the impact of imputing item non-response on household income (Chapter 3), and the proper representation of the upper tail of the income distribution (Chapter 4). The last, Chapter 5, uses a new methodology, Analysis of Gini (ANOGI), to address regional stratification in unified Germany with the help of a inequality decomposition.

As stated above, the first two Chapters deal with the measurement of poverty, a very important topic within the analysis of income, as well as in the political and public debate. The World Bank, for example, has recently initiated a major program to ensure that growth in developing countries is pro-poor.[1] Every developing country has to create a strategy to ensure that further economic development is pro-poor and submit a description thereof, a so-called 'Poverty Reduction Strategy Paper' (PRSP) when applying for further funding. Also the European Union (EU) is focusing increasingly on unified social policy and strategies for reducing poverty and social exclusion. At the Lisbon Summit in March 2000, corresponding social policies were formulated as a special focus of the EU, and each member state now has to develop a national action plan to combat poverty and social exclusion.[2] For a more detailed description of the evolution of an equitable triad of economic policy, labor market policy and social policy, as well as the 'open method of coordination', see Krause, Bäcker and Hanesch (2003) and Bartelheimer (2004).

One important development within the EU was the establishment of comparable national reporting systems in response to the requirement that each member state prepare its own national action plan on social inclusion. At the Laeken European Council (2001), a set of indicators for measuring poverty and social exclusion was adopted, the so-called 'Laeken Indicators'. These indicators are based on a study commissioned by the Belgian Presidency of the EU during the second half of 2001, which has been published in Atkinson, Marlier, Nolan and Vandenbroucke (2002).

The Laeken indicators are a set of 18 common statistical measures for social inclusion which are designed to enable the progress of Member States towards the agreed EU objectives to be evaluated and compared. They cover four important dimensions of social inclusion (financial poverty, employment, health and education).[3]

Within Germany, a national reporting system on poverty has also been established in recent years. The 'German Bundestag' (the German National Parliament) asked the German Federal Government in 2000 to prepare regular reports on poverty and wealth. The first of these was published in 2001 and the second in 2005.[4]

For all these facets of the public debate - and particular for establishing a reporting system on poverty - it is crucial to fully understand the practical options and problems in poverty measurement. According to Sen (1981), two main problems can be distinguished when thinking about poverty: first, the *identification step*, and second, the *aggregation step*. Identification deals with the problem of differentiating the population into poor and non-poor, whereas aggregation addresses the problem of finding a measure that can describe the previous identification made in an appropriate way for the entire society. Chapter 1 discusses the possible aggregations and thus provides an extensive overview of the poverty measures proposed in the scientific literature. While this is only

[1] See, for example, the extensive material on the World Bank web page under www.worldbank.org/prsp.

[2] The inclusion of social security in the political program of the EU after the Treaty of Rome (1957) (which was limited to economic domain) was first mentioned in the Turin 'European Social Charta' of 1961. However, poverty and social exclusion were not noted explicitly until the revised Strasbourg version of the Social Charta in 1996. And with the Lisbon Summit of 2000, the EU's legal competence for questions of social policy was codified.

[3] For more details about the Laeken indicators see EUROSTAT (2003) and EUROSTAT (2004).

[4] See BMAS (2001) and BMAS (2005).

a fragment of the research on poverty, it is very important for understanding poverty overall, or better, for differentiating the different measures of poverty measure that exist.

The objective of this Chapter is to provide clarification of the extensive literature on aggregate poverty measures, including recent developments not incorporated into earlier surveys by Foster (1984), Seidl (1988) and Zheng (1997). Since poverty is a complex phenomenon, there exist many different ways to analyze it and this Chapter intends to give researchers some assistance in deciding whether or not a possible measure is appropriate to the case at hand.

All poverty indicators described in Chapter 1 ignore a possible time dimension of poverty, i.e. how long poverty lasts and how frequently households slip into and out of poverty (Walker and Ashworth, 1992). The so-called 'dynamic perspective' on poverty is closely related to the availability of representative panel data (Walker, 1998). The theoretical discussion on poverty and inequality is moving away from a view of poverty as a stable state (von Wiese, 1954; Schäuble, 1984) and toward a differentiation by duration of poverty.[5]

However, the implementation of the time dimension into the measurement of poverty is not clear, and applied studies often only count periods (most often as years) of poverty, based only on a head-count ratio of poverty. More sophisticated methods[6] have not received much attention in the policy debate and do not fit into the framework of aggregated poverty measurement as described in Chapter 1.

The method proposed in Chapter 2 combines the impact of income mobility with the measurement of poverty and can be applied to all concave income-gap-based poverty measures, and is demonstrated with the popular index proposed by Foster, Greer and Thorbecke (1984). The poverty stability coefficient is based on a smoothed income gap over time, similar to the concept of permanent income, and on the inequality stability index proposed by Shorrocks (1978). This new index can show the stability of poverty over time, as well as the impact of a lengthened time period under investigation, i.e. how income mobility over time influences the measurement of poverty.

Regardless of how carefully one has considered different poverty measures before making a decision on one, some empirical data are needed to calculate the chosen indices. Naturally, this is not only true for the measurement of poverty but for all research based on quantitative empirical methods, especially in economics.[7] This data often stems from large-scale social surveys with voluntary participation.[8] To draw valid conclusions from these data, it is extremely important that the survey be representative for the underlying population of

[5]For the discussion about poverty mobility see for example Bane and Ellwood (1986); Rohwer (1992); Piachaud (1992); Leibfried and Voges (1992); Duncan, Gustafsson, Hauser, Schmauss, Messinger, Muffels, Nolan and Ray (1993); Leibfried, Leisering, Buhr, Ludwig, Mädje, Olk, Voges and Zwick (1995).

[6]See e.g. Burgess and Propper (1998) for the use of a panel regression model or Devicienti (2002), who perform a survival analysis.

[7]For a discussion of the importance of high quality data on income in economics, see Atkinson (2002)

[8]Although in official statistics, participation in the survey (most often a census) is usually obligatory, it also has drawbacks such as misreporting and a very narrow catalog of possible questions due to legal regulations.

interest.[9] In other words, garbage in, garbage out. Chapters 3 and 4 deal with this issue.

The use of social surveys to obtain data on the population almost always results in non-complete data, which can in the worst case result in biased estimations. Therefore Chapter 3 concerns itself with the impact of not completely observed data on the analysis of income, and Chapter 4 deals with the representativeness for a group of special interest: households with high income.

In all surveys, two general types of missing data can be distinguished: unit nonresponse and item nonresponse. Unit nonresponse describes the situation where a person (observation unit) refuses to give an interview and consequently does not provide any answer to any of the questions, while item nonresponse describes the situation where the respondent gives an interview but does not answer particular questions (items in the data). Therefore, in the latter case, one observation unit (e.g. a person or household) shows missing data in *some parts* of the survey but also some valid data in others.

Unit nonresponse is dealt with mostly through weighting adjustments, whereas item-nonresponse is often dealt with through imputation, or simply ignored. Imputation describes the process of filling in missing values in a dataset with an estimate of what the values *could be.* The problem of item nonresponse is very widespread, but often not assessed, and most empirical studies use only complete case analysis. The related assumption that the missing values are missing completely at random (MCAR)[10] is very restrictive and unlikely. But even when missing values are imputed, the gaps are often filled in with only one value and the imputed values are often treated as true values. Also, the variance estimates are computed using standard formulas for a specified sample design although this may cause underestimation of the true variance of the estimates.

When using single imputation, one has to correct each variance estimation for each statistical method used, and for the majority of methods, no correction has yet been proposed. As one alternative, Rubin proposed the multiple imputation approach, where standard statistical procedures can be used.[11] The main idea here is that multiple imputation gives m ($m \geq 2$) imputed values for each missing item, considering the case in which the data are missing at random (MAR)[12] we can compute m different estimates y_{MIl} (and $l = 1, \ldots, m$) for example of the population mean, and the multiple estimate y_{MI} is given by $\hat{y}_{MI.} = \frac{1}{m} \sum_{l=1}^{m} y_{MIl}$. The variance can be appropriately calculated by a simple combining rule, which considers the within and between imputation variance.[13] The focus of this Chapter is therefore to compare different imputation strategies, namely multiple imputation methods, for handling the problem of missing items in large real surveys for a complex variable, i.e. net disposable household income.

[9] To be more precise, the data can be considered representative only at a certain probability, *and* only if valid statistical procedures are applied, for example when the correct weighting scheme is applied with respect to the chosen survey design.

[10] According to Little and Rubin (2002) the missing data are called missing completely at random (MCAR) if the probability of the observed missing pattern ('missingness') does not depend on any of the variables whose values are observed or missing.

[11] See Rubin (1987, 1996) and Little and Rubin (2002).

[12] The missing data are called missing at random (MAR) if missingness depends on the observed values but *not* on the variables *with* missing values. For a description of Missing at Random, see section 3.1.

[13] For the exact definition, see equation (3.4) in Chapter 3.

It should be noted explicitly that evaluating the statistical properties of different imputation methods requires either a theoretical basis or a numerical simulation. Neither of these two possibilities are explored in Chapter 3, and it is also not the goal of this Chapter to derive statistical properties for the different imputation methods. Rather an empirical example with real data is given to assess the robustness of the selected imputation software to possible misspecifications. All software used in Chapter 3 is theoretically capable of generating proper imputation, as defined by Rubin (1987), if the underlying assumptions hold. With real data, however, this is often not the case.

For the empirical analysis in this Chapter a special dataset from the year 1996 of the Finnish subsample of the European Community Household Panel (ECHP) was used. The great advantage of this dataset (provided by Statistics Finland) is that it is merged with various corresponding income data from official registers. The Finnish EHCP income variables are nearly complete since the corresponding values were taken from register data. The interviewed persons are normally asked only very few questions about their income. However, in the year 1996 the interviewees were given the whole ECHP questionnaire, including the section on income. This special situation has the great advantage that for this year, the 'true' income data are available from the register for every person in the sample, *whether they refused to answer the corresponding questions or not*. By nature, the income data from the register are complete and do not contain any missing items, but values for the income questions stemming from the interview situation may be missing, as with any other survey. For the evaluation of the imputation methods it is assumed that the register incomes are the true values and serve as benchmarks for the imputed incomes.

This unique dataset (where the true values are known for the missing values) used in Chapter 3 makes it possible to compare the true distribution or true parameters of the distribution with the corresponding results after imputation. Therefore, the robustness of the imputation algorithms can be assessed with a real world example, although it is not possible to test for the assumptions violated.

Whereas the previously described Chapter 3 deals with item nonresponse, Chapter 4 concerns itself with the question of representativeness[14] of the Socio-economic Panel (*SOEP*)[15] in the upper income region. There are two rationales for a special analysis of this group. First, it can help in assessing what impact better coverage of high-income households has on the overall income distribution. Second, it can help check whether the distribution of high incomes can accurately be measured by the *SOEP*, or if the inclusion of a special high-income sample with more cases changes the picture of the distribution in the high-income region.

[14] The term 'representativeness' is also not clearly defined in statistics. The sense in which it is used here is to describe the extent to which more high-income cases allow a better estimation of the underlying assumed Pareto income distribution. For a critical discussion of representativeness analysis see Pötter and Rendtel (1993) in their reply to Hartmann and Schimpl-Neimanns (1992).

[15] The *SOEP* is a wide-ranging representative longitudinal study of private households in Germany, for a more detailed discussion of this dataset see Schupp and Wagner (1995), Burkhauser, Kreyenfeld and Wagner (1997), SOEP Group (2001) and Haisken-DeNew and Frick (2005) as well as the description in Chapter 4, especially section 4.1.1. The *SOEP* data used in this publication were made available by the German Socio-Economic Panel Study (*SOEP*) at the German Institute for Economic Research (DIW), Berlin.

The tails of the income distribution - especially the upper one, because of the typically skewed distribution of income - are often problematic to measure, because these cases, by definition, have a lower probability of being sampled.[16] There is also some indication that households with very low or very high income are more likely to refuse to participate in social survey.[17] Therefore it is often claimed that most large-scale surveys suffer a middle-class bias. There is a great deal of research about the correct coverage of the lower tail of the income distribution (e.g. homeless people), but very little on the upper tail, cf. Isengard (2002). To carry out research with a greater degree of socio-economic differentiation, an oversampling of high-income households is necessary. This necessity was stated in the first 'German Poverty and Wealth Report' (BMAS, 2001, 'Armuts- und Reichtumsbericht'), which describes the available statistical data as inadequate to carry out more detailed analysis on the socio-economic structure of wealthy households. Therefore the *SOEP* group was instructed by the Federal Ministry of Labour an Social Affairs to draw a special sample for high-income households, which was done in 2002.

The aim of Chapter 4 is to give a detailed analysis of the income distribution of this new high-income sample with respect to the overall picture of income inequality that appears with and without the new subsample. More specifically, are high-incomes earners covered by the old *SOEP* subsamples in approximately the same manner in the income distribution as suggested by the new 'special sample'? What is the effect on the measurement of poverty[18] and inequality when the upper tail of the income distribution is more accurately covered? This latter question in particular can be addressed by testing whether the actual distribution of high incomes follows a Pareto distribution, and if the oversampling of high-income cases affects this estimation.

The distribution of high incomes and their development over time heavily influences the inequality measured in a society (Atkinson and Leigh, 2005), and for this reason, the last Chapter of this thesis dealt with an empirical application of a rather new inequality decomposition technique (Analysis of Gini; ANOGI)[19] and additionally considered spatial stratification. Regional disparities in living standards have been the subject of keen attention in recent years, especially in those countries where income inequality has been rising over time or where average incomes vary considerably across regions or provinces (Shorrocks and Wan, 2005). This is true especially for developing countries, but also the case for Germany after the fall of the Wall and in particular since reunification. The differences in economic performance between East and West Germany, and therefore also the cross-regional variation in living standards, are highly relevant policy issues. Even the German constitution states that economic and social policy should be designed with the aim of diminishing regional differences in living circumstances, and this has hardly been accomplished even 16 years after

[16] For a discussion of other problems concerning the measurement of the income distribution tails, see Isengard (2002).

[17] See Frick and Grabka (2004); Frick and Grabka (2005) and Lipsmeier (1993)

[18] The distribution of high incomes can also affect the measurement of poverty if the poverty threshold is set relative to the whole income distribution, as is done when the poverty line is calculated as x % of the overall mean or median. This kind of relative poverty line was also used in the German Poverty and Wealth Report and is proposed for the Laeken indicators.

[19] See Frick, Goebel, Schechtman, Wagner and Yitzhaki (2006).

reunification.[20]

Chapter 5 thus attempts to contribute to a better understanding of the persisting differences in economic performance between East and West Germany and employs a Gini decomposition to analyze stratification as reflected in the personal income distribution. Although the public debate since reunification has been dominated heavily by the comparison between *the East* and *the West*, one should keep in mind that regional variation in economic performance has a long history within West Germany alone, and that even before reunification there existed financial instruments to deal equalization between West German states. At that time, however, the focus of discussion on this issue was on the existence of a possible North-South divide.[21] In order to give an indication of the sensitivity of an East-West comparison, also an inequality decomposition for a more diversified grouping of West Germany is also carried out. West Germany is split up into a northern, a central and a southern part; for the exact classification, see table 5.1.

[20] The basis for this regulation is laid down in articles 106 and 107 of the constitutional law. See http://archiv.bundesregierung.de/bpaexport/artikel/47/45447/multi.htm

[21] See Friedrichs, Häußermann and Siebel (1986) and citation therein.

Chapter 1

Measuring Aggregated Poverty

Poverty is an important topic in the public and political debate, and the approach how poverty is measured can heavily influence the conclusions drawn from empirical analysis. A thorough understanding of poverty measurement is therefore crucial for any empirical investigation. Generally, two principle problems must be distinguished when thinking about poverty. First, the *identification step* and second the *aggregation step* (Sen, 1981). The first problem deals with the question which persons or households are poor? What is the appropriate concept for a poverty line? How should be differentiated between poor and non-poor? Having solved this problem a researcher has to decide how this information should be aggregated and with what kind of measure. To clarify this second step is the goal of this chapter. Throughout it is assumed that the problem of setting a poverty line is already solved, or in other words is exogenously set.[22]

The discussion how poverty should be measured within economics changed since the two classical papers by Watts (1968) and Sen (1976). Watts was the first who gives an economic definition for the measurement of poverty, however without wider attention until his rediscovery in the 1980s and 1990s by Foster (1984) and Zheng (1993). The paper with much more impact on the following research was Sen's article, which constitutes the so called axiomatic approach to poverty measurement. An appropriate poverty index is henceforth not only the counting of persons living in poverty, but has to fulfill certain properties (axioms).

Sen mainly proposed three different axioms a poverty measure has to fulfill. First the focus axiom, which states that a change in the incomes of the non-poor should not alter poverty. The second axiom is like the other way round, a change in the income of a poor person should change the measure, even if the person stays in poverty. This axiom is called the monotonicity axiom. And a transfer of income from a 'richer' poor person to a 'poorer' poor person should also be reflected by a poverty index, as described in the transfer axiom. None of the

[22] One approach to avoid setting one specific poverty line to separate the society into poor and non-poor is for example the fuzzy set approach Cheli and Lemmi (1995); Schaich and Münnich (1996); Betti, Cheli, Lemmi and Verma (2005).

poverty indeces proposed before Sen's paper[23] or used in official and scientific reports – the head count ratio, income gap, and poverty gap – fulfill these axioms simultaneously and motivated Sen to develop a new poverty index. This change in the view on the measurement of poverty initiated a vast amount of scientific literature written by many researches who uses Sen's approach to refine his measure or to propose additional axioms and/or poverty indices.[24]

The distinction between relative and absolute poverty is often discussed in the way how the poverty line is set (Dietz, 1997). Hence the absolute poverty threshold is the level of income or goods a person definitely needs to live and dose not change over time[25], sometimes even called 'hunger line' (Hagenaars, 1987). A Relative poverty line, on the other side, changes with respect to the overall income distribution, often set as 50% of the arithmetic mean or 60% of the median. However, the term relative in poverty measurement does not only affect the definition of the poverty line, but also the resulting poverty measure as such (Zheng, 1994). If poverty remains unchanged when all incomes are multiplied by a constant (scale invariance) then the underlying poverty measure is a relative one. While translation invariance requires that a poverty measure should be unaffected by equal increments to each income. But also other questions are relevant to understand the measurement of poverty. How are the poverty gaps weighted in each measure? It is rank based, by an explicitly formulated welfare function, or even by the distance to the poverty line itself? However the decision which one is more appropriate lies by the researcher himself and what question the analysis exactly addresses, but should also be based on the knowledge about the underlying axioms of the index chosen. Therefore the aim of this chapter is not to find the *best* poverty measure for all circumstances, which clearly does not exist (Foster, 1984).

A poverty measure, as argued by Sen, should satisfy certain properties or axioms and the acceptability of each poverty measure should be evaluated along this axioms. The objective of this chapter is to provide a clarification on the extensive literature of aggregate poverty measures, including recent developments not incorporated in earlier surveys by Foster (1984), Seidl (1988) or Zheng (1997). And as poverty is a complex phenomenon many different ways to analyze it exist, therefore another intention of this chapter is to give researchers some assistance to decide whether a possible measure is appropriate or not.

However, it should be noted that other important issues in poverty research apart from measurement of aggregation exists. This includes, multidimensional poverty measurement, aggregation over time and partial poverty orderings which arise from the multiplicity of poverty measures, poverty lines and equivalence scales.[26]

[23] Of course the exception is the above mentioned article by Watts (1968).

[24] Some of the more prominent articles which refer directly to this work are inter alia Takayama (1979); Pattanaik and Sengupta (1995); Shorrocks (1995); Chakravarty (1997).

[25] Or only adjusted for inflation, e.g. the official poverty line in the USA.

[26] All of this problems are extensively discussed in the literature, the interested reader can found a starting point in the more general survey by Seidl (1988), although this paper is rather old it gives a good overview about the problems in the measurement of poverty. A more comprehensive and more recent publication can be found in Cowell (2003).

1.1 Notations and definitions

In this chapter only discrete income distributions are considered, represented by vectors, $y = (y_1, y_2, \ldots, y_n)$, drawn from an income space described by $D = \cup_{n=1}^{\infty} \mathbb{R}_+^n$, where $y \in D$, and D^n is a set of all n-tuples of elements from D. Without loss of generality, we assume that the elements of y are sorted in nondecreasing order, such that $y_1 \leq y_2 \leq \ldots \leq y_n$.

Because the focus of this chapter is in the aggregation of poverty, it will be assumed that for any given poverty line $z \in D$ and distribution $y \in D$, the population can be separated one to one into poor and non-poor cases. Hence the definition of poor is given by:

Definition 1. For all $y \in D$, the poverty domain is $D_p(z) \equiv \{t \in D | t < z\}$.[27]

Therefore, a person with income y is *poor* if $y_i \in D_p(z)$ and *non-poor* if $y_i \in D/D_p(z)$. The size of the population corresponding to y is given by $n(y)$ (or n), the number of poor persons is $q(y;z)$ (or q), the mean income of the poor is described by $\mu_p(y;z)$ (or μ_p), and accordingly the variance of the incomes of the poor is $\sigma^2(y;z)$.

The first formal definition of a poverty measure goes back to Watts (1968), which is a function of individual income (y_i) and the poverty line (z). Accordingly we can define (cf. Zheng, 1997):

Definition 2. A *poverty measure* is a function $P(y;z) : DxD \rightarrow R_+$ whose value – *poverty value* – indicates the degree of poverty intensity, or *poverty level*, associated with the distribution of y and the poverty line z, where R_+ is the non-negative real number set. Therefore, for a given poverty measure and poverty line, each income distribution is assigned a number – *poverty index*.

1.2 Poverty axioms

Sen was the first who formally proposed axioms that a poverty measure should satisfy (Sen, 1976, 1981). He proposed a set of three axioms which are still the core of poverty measurement today.[28] These axioms are called the *focus axiom*, the *monotonicity axiom*, and the *transfer axiom*.

According to Zheng (1997), who described also several other poverty axioms, the discussed axioms in the scientific literature can be classified into three groups. A group of core axioms, a group of 'implied' axioms and ad-hoc axioms. The group of independent core axioms consists of the following axioms:

[27] This is equivalent to the so called *weak* definition of poverty. The corresponding strong definition of poverty is $D_p(z) \equiv \{t \in D | t \leq z\}$. Empirically, the use of both definitions do not produce any substantial differences. However, theoretically the choice between the two definitions may affect the properties a poverty measure satisfies (see Donaldson and Weymark, 1986; Zheng, 1997).

[28] Beside these three axioms (focus, monotonicity, and transfer axiom) Sen (1976) also proposed several other more specific axioms. Because these axioms provided only the basis for the formulation of the specific measure proposed by Sen and have not been widely recognized, they will not be presented here.

- *Focus axiom,*
- *Symmetry,*
- *Replication invariance,*
- *Continuity,*
- *Increasing poverty line,*
- *Regressive transfer,*
- *Weak transfer sensitivity, and*
- *Subgroup consistency*

The second group includes all those axioms which are also quite reasonable but can be implied by the axioms in the 'core' group and are not very restrictive to the from of the poverty measure. This group include the following axioms:

- *Weak (Strong) monotonicity,*
- *Non-poverty growth,*
- *Minimal (Weak) Transfer,*
- *Monotonicity sensitivity,*
- *Progressive transfer,*
- *Transfer sensitivity,*
- *Normalization,*
- *Decomposability, and*
- *Restricted continuity.*

The third and last group of poverty axioms, classified by Zheng (1997) are axioms that "cannot be justified". This group includes inter alia

- *Poverty growth,*
- *Scale invariance, and*
- *Translation invariance.*

The focus axiom, introduced by Sen (1976), is still unchanged and requires that a poverty measure should be independent of the income distribution of the *non-poor*.[29] If one regards poverty as deprivation, as suggested by Sen, then the focus axiom seems to be perfectly appropriate. But as Foster (1984) and Zheng (1997) annotate this is not the case if one wants to measure poverty as the difficulty of eliminating poverty by redistributing income from the non-poor to the poor, because then the income distribution of the whole population has to be considered. An example of such a measure was proposed by Anand (1977), which expresses the aggregate poverty gap as a ratio of the aggregate income of the non-poor. A recent contribution by Sallila, Hiilamo and Sund (2006) uses the ratio of the mean income of poor households but to the mean income of all households.[30] However, such a measure is not a poverty measure as defined in definition 2, but rather an indicator of the ease of poverty alleviation, as Anand already mentioned.

[29] Note that this does not imply that the *poverty line* has to be set independent from the income distribution of the non-poor persons. Setting a poverty line is totally different to the problem of aggregating poverty, at least theoretically.

[30] Although the authors claim 'to develop an alternative measure for relative poverty', and as their aim they want 'to combine information both on the depth of poverty and the number of people living in poverty', this measure is in the sense used in this chapter not a measure of poverty. The measure does not satisfy the *focus axiom* and is not sensitive to changes in the number of poor people.

Definition 3. *Focus Axiom*: $P(y;z) = P(x;z)$ whenever $y \in D$ is derived by $x \in D$ by an increment to a non-poor person.

Note, that this axiom does not assume that the number of non-poor persons are irrelevant to the poverty measurement, only the income of the non-poor are irrelevant. Because of this axiom, some researchers have used the so called *censored income distribution* instead of the original income distribution. A censored income distribution sets all income above the poverty line to the poverty line itself, or accordingly sets the *poverty gap* to zero.

The second axiom Sen (1976) proposed was the monotonicity axiom which says that a drop (increase) in a poor persons income should increase (decrease) the poverty level. This axiom could be divided into two forms, as formulated by Donaldson and Weymark (1986), i.e., the *weak monotonicity axiom* and the *strong monotonicity axiom*[31]:

Definition 4. *Weak Monotonicity Axiom*: $P(y;z) > P(x;z)$ whenever $y \in D$ is derived from $x \in D$ by a simple decrement to a poor person (leaving all other incomes unchanged).

Definition 5. *Strong Monotonicity Axiom*: $P(y;z) < P(x;z)$ whenever $y \in D$ is derived from $x \in D$ by a simple increment to a poor person (leaving all other incomes unchanged).

Both monotonicity axioms are very plausible, because other things equal, a decrease (increase) in a poor persons income should increase (decrease) the overall poverty level. However, these two axioms are not equivalent, which arises in a situation when the increment of a small amount of income to a poor person lifts her out of poverty.

The third axiom proposed by Sen (1976) was the transfer axiom which requires the poverty measure to be sensitive to the redistribution of income within the poor population. This axiom originate from the literature about income inequality. It first appeared in Dalton (1920), which is now a classical paper about income inequality. He referred to it as the "principle of transfers". Donaldson and Weymark (1986) distinguished four different transfer axioms by differentiate the effects and directions of transfers:

Definition 6. *Minimal Transfer Axiom*: $P(y;z) < P(x;z)(P(y;z) < P(x;z))$ whenever $y \in D$ is derived from $x \in D$ by a *progressive (regressive) transfer* between two poor persons with no one leaving poverty as a consequence of this transfer.

Definition 7. *Weak Transfer Axiom*: $P(y;z) < P(x;z)(P(y;z) < P(x;z))$ whenever $y \in D$ is derived from $x \in D$ by a *progressive (regressive) transfer* with at least the recipient (donor) being poor with no one crossing the poverty line as a consequence of the transfer.

Definition 8. *Regressive Transfer Axiom*: $P(y;z) > P(x;z)$ whenever $y \in D$ is derived from $x \in D$ by a *regressive transfer* with at least the donor being poor.

[31] Donaldson and Weymark (1986) used the terms upward and downward monotonicity, but as Seidl (1988) and Zheng (1997) describe are the terms weak and strong more appropriate, because strong monotonicity implies weak monotonicity.

Definition 9. *Progressive Transfer Axiom*: $P(y;z) < P(x;z)$ whenever $y \in D$ is derived from $x \in D$ by a *progressive transfer* with at least the recipient being poor.

The quintessence of these four transfer axioms is that an equalization of incomes (a transfer from a richer person to a *poor* person) should decrease poverty, while a disequalizing transfer (from a *poor* person to a richer person) should increase the poverty value. The minimal transfer axiom is, by definition, the weakest form among these four axioms while progressive transfer is the strongest form, i.e., *minimal transfer* $\rightarrow$ *weak transfer* $\rightarrow$ *regressive transfer* $\rightarrow$ *progressive transfer*. The difference between the weak forms (minimal and weak transfer) and the strong forms (regressive and progressive transfer) lies in the fact that the transfer makes anyone crossing the poverty line or not.[32]

There are a lot of other axioms proposed in the literature, the more widely recognized axioms with respect to poverty measurement include the *Replication Invariance Axiom*, *Continuity Axiom*, *Symmetry Axiom*, as well as Axioms regarding decomposability, sensitivity and economic growth. Only the more prominent axioms will be presented in the following.

The symmetry axiom proves that "the names" (the order) of income recipients do not matter for measuring poverty. Symmetry does not impose any real restriction as any aggregate "snapshot" measure cannot avoid symmetry. However, if one is interested in measuring chronic or lifetime poverty, this would be a nonsensical requirement.

Definition 10. *Symmetry Axiom*: $P(y;z) = P(x;z)$ whenever $y \in D$ is derived from $x \in D$ by a permutation.

An implication of this axiom is that poverty can be defined over ordered income distributions without any loss of generality. Foster (1984) has shown that when symmetry is assumed, the three axioms proposed by Sen (1976) (focus, monotonicity, and weak transfer) are equivalent to 'requiring that on each set of income distributions with constant q, the poverty measure [...] must be a strictly decreasing, strictly Schur-convex function of the first q incomes'.[33] And because of this, any poverty measure satisfying these four axioms does also ensure that the poverty measure follows the absolute Lorenz criterion for the incomes of the poor.[34]

Kakwani (1980b) criticized the lack of sensitivity to the income level of transfer of the measure invented by Sen (1976). He argues that a convenient poverty measure should be more sensitive to changes among the bottom poor. In order to achieve such a measure he introduced three sensitivity axiom, two on income transfers and one on income increment (decrement).

[32] For a more detailed discussion about the different transfer axioms and their interrellationship see Sen (1981); Thon (1983a); Donaldson and Weymark (1986); Zheng (1997). Kundu and Smith (1983) showed that the transfer axioms can not hold along with the requirement of monotonicity in the fraction of the poor, which is only fulfilled with a head-count based measure.

[33] Foster (1984, p. 220).

[34] The Lorenz criterion says in this context, that an income vector y dominates another income vector x by the absolute Lorenz criterion if the poor persons (q) in y have no less income than the poorest q persons in x, and at least for some persons in q they hold strictly more income.

Definition 11. *Monotonicity Sensitivity Axiom*: $P(y''; z) - P(y; z) > P(y'; z) - P(y; z)$ whenever y' and $y'' \in D$ are derived from $x \in D$ by the same amount of decrement to poor incomes y_i and y_j, respectively, where $y_i < y_j$.

In other words this axioms require that a poverty measure should be more sensitive to a drop in a poor persons income, the poorer this person is. Which is connected to the next axiom Kakwani (1980b) has proposed, the weak transfer axiom. The idea of this axiom is that the poverty estimation should give more emphasis to transfers taking place down in the distribution, ceteris paribus.[35]

Definition 12. *Weak Transfer Sensitivity Axiom*: $P(y; z) > P(y'; z)$ whenever y and $y' \in D$ are derived from $y \in D$ by transferring income $\delta > 0$ from (poor) incomes y_i to y_j and from poor income y_k to y_l respectively with $y_j - y_i = y_l - y_k > \delta$, $y_k > y_i$ with no one crossing the poverty line after the transfers.

The third axiom Kakwani (1980b) proposed with respect to the sensitivity of a poverty measure is the transfer sensitivity axiom, the stronger form of the weak transfer sensitivity axiom.

Definition 13. *Transfer Sensitivity Axiom*: $P(y; z) > P(x; z)$ whenever $y \in D$ is derived from $x \in D$ by a favorable composite transfer (): a progressive transfer of income $\delta > 0$ from y_j to y_i and a regressive transfer of income $\rho > 0$ from y_k to y_l, i.e. $y = x + \delta(\epsilon_i - \epsilon_j) + \rho(\epsilon_l - \epsilon_k)$

The difference to the former weak transfer sensitivity axiom is, that neither the amount of two transfers have to be the same, nor the distance between the two persons involved.

The *replication invariance axiom* was first introduced into poverty measurement by Chakravarty (1983a) and Thon (1983b) from the income inequality literature. Because any two income distributions which differ only by size can be replicated to the same size, and so their inequality and poverty levels can be directly compared. Although this axiom is intuitively appealing, it is surprising that many early proposed poverty measures violate it (including the one introduced by Sen (1976) in his explicitly axiomatic approach).

Definition 14. *Replication Invariance Axiom*: $P(y; z) = P(x; z)$ whenever y is derived from x by a (k-)replication[36].

Another important axiom is the *continuity axiom*, which is also quite reasonable. Given a small change in a poor persons income, we do not expect a huge jump in the poverty level. The formal definition is given by:

Definition 15. *Continuity Axiom*: $P(y; z)$ is continuous as a function of y on D for any given z.

[35] The original versions of the monotonicity sensitivity axiom and the weak transfer sensitivity axiom are given by Kakwani (1980b) in a rank form, but in order to get better comparability the form presented in Zheng (1997) is used. For a discussion with the original form, see Foster (1984) and Cowell (1988).

[36] Where *replication* is defined as: $y \in D$ is derived from $x \in D$ by a (k-)replication if $n(y) = k \cdot n(x)$ and $y = (x, x, \ldots, x)$ for some positive integer k.

Watts (1968) argued that "poverty is not really a discrete condition" and "one does not immediately acquire or shed the afflictions we associate with the notion of poverty by crossing any particular income line". Therefore, "it would be appropriate to maintain the graduation provided by a continuum"[37]

Two very important and popular axiom for poverty measures are subgroup consistency and subgroup decomposability, introduced by Foster et al. (1984) and Foster and Shorrocks (1991).

Definition 16. *Subgroup Consistency Axiom*: $P(y; z) < P(x; z)$ whenever $y = (y', y'') \in D$ is derived from $x = (x', x'') \in D$ with $n(y') = n(y'')$, $n(x') = n(x'')$ and $P(y'; z) < P(x'; z)$, $P(y''; z) < P(x''; z)$.

Definition 17. *Subgroup Decomposability Axiom*: For $y = (y', y'') \in D$ with $n(y) = n(y') + n(y'')$, and

$$P(y; z) = \frac{n(y'}{n(y)} P(y'; z) + \frac{n(y''}{n(y)} P(y''; z) \ .$$

As Foster and Shorrocks (1991) highlighted the subgroup consistency axiom can be compared to the monotonicity axiom. While the latter deals with the individual change in poverty status, the subgroup consistency axiom is about the change in the poverty level of the subgroup, in fact Foster (1984) calls this axiom subgroup monotonicity. If a poverty measure satisfies this axiom an increase in poverty for a given subgroup ceteris paribus leads to an increase in total poverty.[38]

Foster and Shorrocks (1991) argue that this axiom is useful, because policy strategies are typically targeted at specific subgroups or regions of a country. The typical characteristic of a decomposable poverty measure that it can disaggregate the overall poverty into the poverty of specific population subgroups. Every decomposable poverty measure has also to be subgroup consistent and meets the need not only for measuring the effects of group specific poverty alleviation policies, but also to identify particular subgroups which are highly affected by poverty.

The two axioms together ensure that a poverty measure can additively decomposed in a consistent way, which was the primary motivation for the paper of Foster et al. (1984). However, Foster (1984) also note that "for applications that do not involve analyzing subgroup poverty, there is no particular reason for choosing P_α above other measures satisfying the same properties". But almost all researchers insists on "additive separability" in poverty measures (also in inequality measures) as a necessary technical property, but seldom the general question about the appropriateness is asked.

> "Separability is certainly convenient property, and permits us to build up the overall poverty picture from the poverty measures applied to subgroups. The requirement has much cutting power. [...] There remains a more general question as to whether it is sensible to

[37] Watts (1968, p. 325). A approach where this is explicitly assessed is the fuzzy set approach, e.g. Cheli and Lemmi (1995); Betti et al. (2005).

[38] Foster (1984) gives an example where this is not the case for the Sen measure.

> assume that poverty indicators should be combinable in this way, which requires that the view of poverty for particular groups be, in some specific way, insensitive to what happens to other groups, and that the whole picture does not introduce anything other that what is already there in the parts."[39]

Certainly it could be argued that a decomposition do not have to be additively, but has also to take care of the connection between the separated groups. An example for this is the decomposition of the Gini index which needs an additional component, the overlapping of each subgroup with each other, as proposed by Yitzhaki and Lerman (1991).[40] However, until now there is no such an approach available for the analysis of poverty, if it is at all meaningful.

1.3 Poverty measures

The following sections reviews the most "important" poverty measures, in terms of being discussed most frequently in the scientific literature. However, the list is not complete in a way that each poverty measure ever proposed is included, nor is each poverty measure discussed in detail with all implications when applied to empirical data or when evaluate policy strategies. The main purpose is to give an "in-depth overview" of the scientific discussion on how to aggregate poverty using empirical micro-data. Similar to the grouping of the axioms the grouping of the poverty measures is oriented on the very comprehensive survey in Zheng (1997). This grouping consists of four categories, the first one includes distribution insensitive measures, and the second group is made up of measures very closely linked with the Sen measure. The third group consists of ethical poverty measures and the last group includes all distribution sensitive measures.[41] A distribution sensitive poverty measure is a poverty measure that satisfies the *minimal transfer axiom*.

Apart from this grouping the former mentioned core axioms for poverty measurement have a strong impact on the functional form of any poverty measure. Any measure that satisfies the "core" axioms (*symmetry, continuity, replication invariance and subgroup consistency*) must have the following functional form:

$$P(y;z) = F\left(\frac{1}{n}\sum_{i=1}^{q} p\left(y_i;z\right)\right) . \tag{1.1}$$

And with the additional consideration of *regressive transfer*, *weak transfer sensitivity*, and *increasing poverty line* the individual poverty deprivation function p must also satisfy $\frac{\partial p}{\partial y} < 0$, $\frac{\partial^2 p}{\partial y^2} > 0$, $\frac{\partial^3 p}{\partial y^3} < 0$, and $\frac{\partial p}{\partial z} > 0$. For a quick overview of the axioms each measure satisfies or not see table 1.1.

[39] Sen (1992, p. 106.)

[40] More details about this approach can be found in section 5.2.1.

[41] Foster (1984) gives a similar classification of poverty measures into three groups. The first group includes all measures who are more or less direct extensions of the Sen measure (our second group), the second consists of indeces using inequality measures or are closely linked to the methodology drawn from the inequality literature (a subset of our third group). His third group includes measures whose development was motivated by a practical concern and consists only of the new proposed measure by Foster et al. (1984).

Table 1.1: Overview about which axioms are satisfied by each poverty measure

	Class 0			Class 1				Class 2					Class 3					
Axioms and measures	H	I	HI	S	K	T	T_a	C_1	BD	Ch	C_h	K_o	C_2	F	W	H	HD	SCH^a
Focus axiom	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Symmetry	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Replication invariance	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Continuity	$\ominus$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Increasing poverty line	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\checkmark$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Regressive Transfer	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Weak transfer sensitivity	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark^*$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Subgroup consistency	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Weak monotonicity	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Strong monotonicity	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\checkmark$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Minimal transfer	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Weak transfer	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Progressive sensitivity	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Restricted continuity	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Decomposability	$\checkmark$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\ominus$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Non-poverty Growth	$\checkmark$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\ominus$	$\ominus$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Normalization	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

H: Headcount Ratio; I: Income Gap Ratio; HI: Poverty Gap Ratio; S: Sen Measure; K: Kakwani Measure; T: Thon Measure; T_a: Takayama Measure; C_1:Clark et al. ethical measure; BD: Blackorby-Donaldson Measure; CH Chakravarty Ethical Measure; C_h: Chakravarty Measure; K_o: Kockläuner Measure; C_2: Clark et al. Measure; F: Foster et al. Measure; W: Watts Measure; H: Hagenaars Measure; HD: Hagenaars-Dalton Measure.; SCH: Schmid Measure

$\checkmark$: Satisfies the axiom. $\ominus$: Does not satisfy the axiom or need additional restrictions.

*: Satisfies the axiom for $\alpha > 2$.

^{a} Because the proposed class is very general the classification is only done for the specific proposed measure by Schmid (1993), see eq. (1.41)

Source: Based on Zheng, 1997, p. 143 with additional representation.

1.3.1 Distribution-Insensitive Poverty Measures

This class includes three measures, the poverty head-count ratio, the income gap ratio and the poverty gap ratio. Although the measures in this group satisfy the fewest axioms (compared to the other groups of measures) they build up the mainstream approach in measuring poverty. Indeed there are strong tendencies in the scientific research into poverty to use distribution sensitive measures.[42] But the political debate and the process to define how to measure poverty in official statistics stuck with traditional distribution insensitive poverty measures. The so called "Laeken-Indicators" (EUROSTAT, 2004) consist of 18 different indicators, which are mainly variations of the head-count ratio.[43] Distribution sensitiveness is only for the whole distribution with the help of the Gini coefficient as well as with a decile ratio considered.

The simplest method to measure an aggregated poverty level for a population is to simply count the persons with an income under a predefined poverty line z and calculate the proportion. Therefore this measure is called head-count-ratio and can be defined as:

$$\textbf{Head-count ratio}: \quad H(y;z) = \frac{q(y;z)}{n(y)} \ , \tag{1.2}$$

where q is the number of the poor ($y_i < z$) and n is the population size.

Although this measure is very popular[44] and of course important it "has little but its simplicity to recommend it"[45], and the shortcomings are also very obvious. It pays no attention to the fact how far the people are *away* from the poverty line, or in other words how *poor* they are. The second measure in this group that has also been widely used in the traditional literature is the so-called Income gap ratio, which measure the percentage of the average shortfall of the poor persons to the poverty line. Formally defined as:

$$\textbf{Income gap ratio}: \quad I(y;z) = 1 - \frac{\mu_p(y;z)}{z} = \sum_{i \in D(z)} \frac{z - y_i}{qz} \ , \tag{1.3}$$

where μ_p is the mean income of the poor. Whereas the head count ratio H is completely insensitive to how the poor fall short of the poverty line and takes note only of the number below this line, the income gap ratio I is the other way round. It is completely insensitive to the number of poor persons and consider only the average income gap of all poor persons. Evidently a combination of the two is needed which brings us to the definition of the poverty gap ratio, defined as:

$$\textbf{Poverty gap ratio}: \quad HI(y;z) = \frac{q(y;z)}{n(y)} \left(1 - \frac{\mu_p(y;z)}{z}\right) \ , \tag{1.4}$$

[42] This is true in particular for the measure proposed by Foster et al. (1984) with parameter $\alpha = 2$. See for some empirical applications e.g. Rodgers and Rodgers (1991); Hill and Jenkins (2001); Goebel and Kuchler (2003)

[43] There is only one income gap measure included within the Laeken indices (see footnote 46) and distribution sensitiveness is only addressed by choosing different thresholds as poverty line.

[44] Foster (1984) cites early examples of the use of a head count measure dating back to 1797. However, the most prominent early example goes back to Rowntree (1901).

[45] Watts (1968, p. 326)

which is simply the product of the head-count ratio (H) and the income gap (I). Because both incorporated measures are not sensitive to the distribution among the poor population, the poverty gap ratio is still not sensitive for a transfer of income from a poor person to one less poor person (*transfer axiom*, see def. 7).[46] Sen (1976) argued that for the special case in which all poor persons have precisely the same income level below the poverty line the poverty gap ratio "should give us adequate information on the level of poverty"[47]. The poverty gap ratio is therefore the starting point for Sen to derive a measure which is also sensitive for the distribution of income among the poor population, presented in the next section.

1.3.2 Sen-Type Poverty Measures

All measures in this class use a poor persons rank within the poor (or the whole population) as an indicator of relative deprivation. All measures are focused and symmetric, and satisfy the weak and minimal transfer axioms (except one measure). But violate the replication invariance and subgroup consistency.

In his seminal work on the measurement of poverty, Sen (1976) suggested also a new index. The central point of this work was to take note of the inequality in the distribution of income among the poor. Derived from his three main axiom (focus, monotonicity, and transfer axiom) he deduces the following measure which itself has become a starting point for several measures suggest by other researchers.[48]

Sen Measure :

$$\begin{aligned} S(y;z) &= \frac{2}{(q+1)nz}\sum_{i=1}^{q}(z-y_i)(q+1-i) && (1.5) \\ &= H(y;z)\left[I(y;z)+(1-I(y;z))G_p\frac{q}{q+1}\right] \\ S'(y;z) &= H\left[I+(1-I)G_p\right]\ \textit{if}\ n\to\infty && (1.6) \end{aligned}$$

Where G_p is the Gini coefficient[49] of the poor (incomes). Note, that $S'(y;z) > S(y;z)$ and $S(y;z)$ violates the *replication invariance* axiom while $S'(y;z)$ satisfies it, "because the Gini coefficient as Sen (1973a, p.31) defined is replication invariant."[50] With introducing this measure Sen linked the analysis of poverty with that of inequality. However, most poverty measures only implement the connection between the distribution among the poor and poverty measured,

[46] The Laeken indicators include a kind of poverty gap, which does not use the arithmetic mean in formula (1.4), but the median. This has the effect that some specific transfer of income among *can* have an effect, but mostly do not.

[47] Sen (1976, p. 223.)

[48] For an alternative axiomatization of the Sen index see Pattanaik and Sengupta (1995).

[49] The Gini coefficient is defined as

$$G_p = \frac{1}{2\mu_p q^2}\sum_{j=1}^{q}\sum_{i=1}^{q}|y_i - y_j|\ .$$

See also section 5.2.1 on page 126.

[50] Zheng (1997, p. 145).

but do not connect overall inequality and poverty. Although this seems to be obvious, especially in consideration with the focus axiom (see definition 3 on page 13), there are some thoughts about a more general connection[51]as in Thon (1981); Pyatt (1987); Lewis and Ulph (1988); Kakwani (1997). Yitzhaki even states that an appropriate inequality measure includes all necessary information for the analysis of inequality *as well as* poverty. He proposes a decomposition of the Gini index into 'Sen's poverty index, an affluence index (a mirror image of the poverty index), and an index of between-groups (poverty line-affluence line gap) inequality'[52]

A key problem with the measure proposed by Sen is, that it does not satisfy the strong version of the transfer axiom. This has motivated Thon (1979) and Shorrocks (1995) to suggest a slightly changed version of this index. The difference between the Sen and the Thon measure lies in the weighting function, which utilizes the rank of the individuals in a varying way.[53] Whereas Sen uses an ordinal rank function of the poor persons within the poor population, Thon uses the rank of the poor persons among the total population.

$$\textbf{Thon-Measure} : T(y;z) = \frac{2}{(n+1)nz}\sum_{i=1}^{q}(z-y_i)(n+1-i) \tag{1.7}$$

This simple change in the weighting function makes the Thon measure satisfy several more axioms which the Sen measure violates (in particular regressive transfer and continuity). As n and q becomes large, the Thon measure approaches:

$$\begin{aligned} T'(y;z) &= \frac{2}{n^2 z}\sum_{i=1}^{q}(z-y_i)(n+0.5-i) && (1.8)\\ &= H(y;z)\left[S'(y;z)+2(1-H(y;z))I(y;z)\right] . && (1.9)\end{aligned}$$

The last measure belongs to the class of measures that was proposed later also by Thon (1983b) and which satisfies replication invariance.[54]

A closely related problem of the Sen index was addressed by Kakwani (1980b), who also changed the weighting function. He generalized the Sen measure in changing the power of the weighting function from 1 to a parameter k. The impact of an income transfer between two individuals, when using the Sen measure, depends only upon the difference in the rankings of the two persons concerned, i.e. a fixed equidistant transfer between the donor and the recipient, over equal number of units will reveal the Gini index to be equally sensitive no matter where in the distribution these occur.[55] With the weighting function

[51]Feldstein (1998) neglects the connection of overall inequality and poverty in the sense that a rise in the top incomes should neither effect the level of poverty nor the level of inequality. He argues that no one is worth off and therefore the Pareto principle holds. However thinking of this as a rise in money supply this argument implies that there is no connection in the amount of money in circulation and the overall price level. Because if the price level would rise the persons without extra money would be worse off in terms of real income.

[52]Yitzhaki (2002), p. 67.

[53]Please note that this weighting function assigns different importance to the poor cases according to the measured poverty intensity. The *normal* weighting of survey samples because of stratification or any other complex design is totally unaffected by this.

[54]Shorrocks (1995) referred to this measure as the 'modified Sen measure'.

[55]cf. Clark, Hemming and Ulph (1981).

proposed by Kakwani and a parameter $k > 1$ the index will become more sensitive to transfers among those at the bottom of the income distribution.

Kakwani-Measure :

$$K(y; z, k) \quad = \quad \frac{q}{nz \sum_{i=1}^{q} i^k} \sum_{i=1}^{q} (z - y_i)(q + 1 - i)^k \tag{1.10}$$

This measure satisfies the rank based sensitivity axiom as proposed by Kakwani (1980b), but as Foster (1984) has shown, for any given k there is a population size n for which $K(y; z, k)$ does not satisfy the transfer sensitivity axioms (both weak and strong versions). Clark et al. (1981) stressed the point that there is no clear rule to select an appropriate value for k and

> "...there is an additional argument in favour of rejecting this approach. Kakwani suggests that k should be chosen according to social preferences regarding transfer sensitivity at different levels of income. There are strong grounds for believing that such preferences should be independent of the particular distribution being considered. The parameter k does not have this property."[56]

Since the Sen measure is based in his distribution sensitiveness on the Gini inequality measure, Takayama (1979) proposed a measure which is from his point of view "a more natural translation of the Gini coefficient from the measurement of inequality into that of poverty" and provides a "more full-blooded representation of the notion of relative deprivation than Sen's measure".[57] The proposed poverty index is the Gini coefficient of the censored income distribution truncated from above the poverty line and can be written as:

Takayama-Measure :

$$T_\alpha(y; z) \quad = \quad 1 + \frac{1}{n} - \frac{2}{\mu_0 n^2} \sum_{i=1}^{n} (n + 1 - i) y_i^* \ , \tag{1.11}$$

where y^* is a *censored income distribution* in which all incomes above the poverty line are set to the poverty line z. The new measure ($T_\alpha(y; z)$) is than simply the Gini coefficient of the censored distribution.

Zheng (1997) has shown that the Takayama measure (T_α) has a connection to the Thon measure (T'), as defined in equation (1.8). When $n(y) \to \infty$ then T_α approaches T' is the Thon measure as defined in equation (1.8)

$$T'_\alpha(y; z) = 1 + \frac{1}{1 + HI}(T' - 1) \ . \tag{1.12}$$

Although most researchers agreed upon the fact that the Takayama measure is a 'smoother' translation of the Gini coefficient, it has been seriously questioned. It is argued that the measure violates every axiom that the Sen measure fails to

[56]Clark et al. (1981, p. 518).
[57]Takayama (1979, p. 748).

satisfy except *continuity* and *replication invariance*. But the major drawback mentioned is the failure to fulfill the *monotonicity axiom*.[58] However, Kakwani (1981) demonstrates "that the monotonicity will be violated by this measure only in an unusual situation when the poverty line strictly exceeds the median income of the distribution, i.e., when the society has considerably more than 50 per cent of its population as the poor".[59] Especially for developed countries this drawback is not a serious constraint and does not rationalize the lack of attention for this measure.

Aside the emphasize on the problematic axiomatization of the proposed measure by Takayama (1979), the concept of a censored income distribution has been widely accepted and has also been used by Shorrocks (1995) to redefine the original Sen measure of poverty. Shorrocks changed the normalization condition[60] by Sen to:

$$P(y;z) = I \text{ whenever } H = 1 \text{ and } G_p = 0 \ .$$

Applying this he was able to propose a 'modified Sen index' which "is not only continuous in individual incomes and consistent with the transfer axiom, but also admits a geometric interpretation in terms of the area beneath a graph called the 'inverse generalized Lorenz curve for normalized poverty gaps' by Jenkins and Lambert (1993)."[61] The proposed measure is defined as:

Shorrocks Measure :

$$\begin{aligned} S_s(y;z) &= \frac{1}{n^2}\sum_{i=1}^{q}(2n-2i+1)\frac{z-y_i}{z} && (1.13) \\ &= (2-H)HI + H^2(1-I)G_p && (1.14) \end{aligned}$$

The measure is closely related to the Takayama measure (T_α, eq. 1.11) and the Thon measure (T in eq. 1.7), however, in contrast, this measures is monotonic and replication invariant. But the most attractive feature of S_s is his geometric interpretation as the area below the poverty gap profile.[62]

Another variation of the Sen index was introduced by Giorgi and Crescenzi (2001), they replaced the Gini index in (1.5) by the Bonferroni inequality index and propose the following index:

Giorgi & Crescenzi Measure :

$$S_{GC}(y;z) = H\left\{I + (1-I)\left[1 - B_p\left(\frac{q-1}{q}\right)\right]\right\} \ . \quad (1.15)$$

And if the number of poor increases this index can be approximated by

$$S_{GC}(y;z) = H\left[I + (1-I)B_p\right] \ . \quad (1.16)$$

[58] See Chakravarty (1983a); Foster and Shorrocks (1991); Shorrocks (1995) and Zheng (1997).
[59] Kakwani (1981, p. 525).
[60] Sen (1976) suggested that a poverty index should be equal the poverty gap ratio when the poor have equal incomes; i.e. $P(y;z) = HI$, whenever $G_p = 0$.
[61] Shorrocks (1995, p. 1226).
[62] For detail on the deprivation profile curve see Shorrocks (1998), and for the closely related Three I's of poverty (TIP) curve, see Jenkins and Lambert (1993, 1997, 1998).

In the above two equation B_p stands for the Bonferroni inequality index among the poor population.[63]

This index retains all the three nice properties of the Sen index. It is sensitive to the head-count ratio, sensitive to how poor the poor are (because of its dependence on the relative gap) and sensitive to the amount of inequality among the poor, because of the use of the Bonferroni index. However it has the additional advantage of satisfying the *Transfer Sensitivity* Axiom, because of the use of the Bonferroni index (see also, Chakravarty and Muliere, 2004). Since the Bonferroni index is not additively decomposable, S_{GC} does not fulfill the *Subgroup Consistency* axiom. Overall this index gives more weight to the inequality component and are sensitive to the low levels of the income distribution then the original Sen measure.

1.3.3 Ethical Poverty Measures

The indices of this class extend Sens approach in another direction, they all either customize the measure of inequality or use a social welfare function. The measures are generally labeled as "ethical poverty measures", or more generally "ethical indices", because of the fact that "these are indices, usually of inequality, that are exact for social evaluation functions. Each index is thus implied by and implies at least one social evaluation function."[64] In other words, ethical poverty indices refer to a representative income or representative income gap (Kockläuner, 2002).

> "This representative income is that level of income which, if given to each poor person, would prove ethically equivalent to the existing income profile of the poor according to a social evaluation function satisfying certain regularity conditions."[65].

As Blackorby and Donaldson note, the concept of the *representative income*[66] of the poor is essential for the formulation of ethical indices. They reinterpreted the Sen measure as a product of the headcount ratio and the percentage shortfall of the representative income of the poor from the poverty line. They show that 'there is a Sen-like poverty index for every social evaluation function defined over the poor' (Blackorby and Donaldson, 1980) and the class of indeces they

[63] The Bonferroni inequality index is defined as

$$B = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{\mu-\mu_i}{\mu} = 1 - \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{\mu_i}{\mu}\ , \quad B \in [0,1]\ ,$$

where μ_i (defined for an income distribution in non-decreasing order) denotes the partial mean of the i lower incomes ($i = 1, \dots, n-1$).

[64] Blackorby and Donaldson (1980, p. 1053).

[65] Chakravarty (1990, p. 147)

[66] The representative income is similar to the 'equally distributed equivalent income' in income inequality literature for the whole population (see Atkinson, 1970). The representative income is the income which, when equally distributed among the poor, leads to the same level of welfare among the poor as the original distribution.

proposed is defined by

Blackorby & Donaldson Measure :

$$BD(y; z, W) = \frac{q(y; z)}{n(y)} \left[1 - \frac{\zeta_p}{z}\right] \quad , \tag{1.17}$$

where ζ_p is the representative income of the poor for an arbitrary (homothetic) social welfare function W. The formulation is very general (the Sen measure and the Kakwani measure are included)[67] and the properties that this class of measures may satisfy critically depends upon the specification of the social welfare function.

If W is completely recursive, continuous, monotonically increasing, S-concave and symmetric, then the measures will be focused and symmetric and satisfy restricted continuity, weak monotonicity and weak transfer. But the measures violate continuity, strong monotonicity, regressive transfer (Chakravarty, 1983a) and subgroup consistency (Zheng, 1997).

Clark, Hemming and Ulph addressed in their paper the transfer sensitivity of the Sen measure as well as the incorporation of the relative deprivation concept by Sen, i.e. it is in some way an approach which combines the criticism addressed by Takayama (1979) and Kakwani (1980b). They pointed out that the Sen measure can be written as a combination of the poverty gap ratio (see eq. 1.4) and the Gini coefficient of the distribution of the poverty gaps. They actually suggest two indices of poverty, in which they implement social evaluation functions for individual poverty gaps, and the inequality in poverty gaps will reflect relative deprivation when aggregated. The first evaluation function used leads to an index defined as

Clark, Hemming & Ulph (1) Measure :

$$CHU_1(y; z, \alpha) = \frac{q}{n} \cdot \frac{g^*}{z} \quad , \tag{1.18}$$

where $g^* = \left[\frac{1}{q}\sum_i (z - y_i)^\alpha\right]^{\frac{1}{\alpha}}$ with $\alpha \geq 1$. As Zheng (1997) showed, this measure does satisfy the *increasing poverty line* axiom, contrary to the discussion in Clark et al. (1981).[68] Note that as $\alpha \to \infty$ the poverty level is determined by the headcount ratio and the Rawlsian maximin justice:[69]

$$CHU_1(y; z, \alpha \to \infty) \to H(y; z) \left(1 - \frac{\min\{y_i\}}{z}\right) \ . \tag{1.19}$$

[67] To derive the original Sen index the representative income of the poor is measured by the Gini social evaluation function and therefore defined as $\zeta_p = \frac{1}{q(z,y)^2} \sum_{i \in D(z)} (2i - 1)y$.

[68] To see this rewrite equation (1.18) as

$$CHU(y; z, \alpha) = \frac{q^{1-\frac{1}{\alpha}}}{n} \left[\sum_{i=1}^{q} \left(1 - \frac{y_i}{z}\right)^\alpha\right]^{\frac{1}{\alpha}}$$

"As z increases, q may also increase (at least not decrease). Therefore, because $\alpha \geq 1$, CHU is necessarily increasing." (Zheng, 1997, p. 148).

[69] cf. Clark et al. (1981) and Zheng (1997).

Kockläuner (2002) used the first proposed measure by Clark et al. (1981) (CHU_1, eq. 1.18) and the extension of Shorrocks (1995) (S_s, eq. 1.13) as a starting point in proposing a new index. He combines both extensions and as a result his poverty index is defined as:

$$\textbf{Kockläuner Measure} : K(y;z) = \frac{1}{z}\left(\frac{1}{n}\sum_{i=1}^{n} g_i^{*\alpha}\right)^{\frac{1}{\alpha}} \tag{1.20}$$

where $g_i^* = z - y_i^*$ and y_i^* is the vector of the censored incomes with all income above the poverty line are set to the poverty line z. The index is dual to the second measure of Clark et al. (CHU_2, eq. 1.29), and has a close relationship to the index proposed by Foster et al. (1984). In comparison with the Foster et al. measure, the index K additionally satisfies the increasing poverty aversion axiom but not the decomposability axiom. Kockläuner sees this as a decision criteria, “when having to decide between the indexes [...], priority has to be given either to a sound ethical foundation and the increasing poverty aversion axiom or to decomposability.”[70]

The second recommendation of Kakwani generalized the Sen measure in a way that the focus axiom is dropped and the whole distribution will be considered. He also replaced the ‘simple’ Gini coefficient with a function of it. The measure will be denoted as SK, because in the literature, it is often called the Sen-Kakwani measure, and of course to discriminate it from the other Kakwani measure in equation (1.10). The measure is defined as

$$\textbf{Kakwani-Measure}: \; SK(y;z,g) = \frac{q}{n\mu}\left[z - \mu_p g(G_p)\right] , \tag{1.21}$$

with $0 \leq g(G_p) \leq 1$ and $g(G_p) = 1$ if $G_p = 0$ and $g'(G_p) < 0$ and where μ is the mean income of the distribution and μ_p is the mean income of the poor. However,

> ”this measure satisfies very few axioms and is not appropriate to be used as a measure of the intensity of poverty. According to this measure, other things being equal, society can reduce poverty level simply making the rich richer, which is contrary to most people’s perceptions of poverty.”[71]

The paper by Chakravarty (1983a) builds directly and explicitly on the work and indices of Takayama (1979, eq. 1.11), Blackorby and Donaldson (1980, eq. 1.17), and Clark et al. (1981, eq. 1.29). He noticed the failure of these measures to satisfy both *continuity* as well as *regressive transfer* concurrently. He proposes the following family of ethical poverty measures, by applying a social welfare function to the censored income distribution and used the percentage gap between the representative income of the censored distribution and the poverty line:

$$\textbf{Chakravarty Measure} \quad C(y;z,W) = \frac{z - \zeta_z}{z} . \tag{1.22}$$

[70] Kockläuner (2002, p. 304).
[71] Zheng (1997, p. 149).

Where ζ_z is the representative income of the censored income distribution. The formulation is a generalization of the Clark et al. (1981) approach, because the social welfare function are not restricted to be strictly recursive. Chakravarty also defines an absolute poverty measure by omitting the denominator in equation (1.22) resulting in a per capita poverty measure.

Vaughan (1987) suggested that poverty indices can be viewed as the loss of welfare which results from the existence of poverty, because the development of poverty measures has followed a 'positive' approach and the welfare interpretation of any poverty measure is 'ex post'. He included the social welfare function, W, directly into the poverty measures and proposed general classes of relative and absolute welfare poverty measures:

Vaughan Measure

$$V_r(y; z, W) = 1 - \frac{W(y)}{W(\tilde{y})} \tag{1.23}$$

$$V_a(y; z, W) = W(\tilde{y}) - W(y) , \tag{1.24}$$

where $\tilde{y}$ is derived from y by setting all poor incomes to the poverty line. To ensure not to violate many poverty axioms the measures defined in (1.23) and (1.24) had to fulfill additional restrictions.

1.3.4 Distribution-Sensitive Poverty Measures

All measures in this class belong to the general category of poverty measures defined in equation (1.1). One feature in common is that all measures fulfill subgroup consistency axiom. Therefore all poverty measures in this class are either decomposable or can be expressed "as increasing transformations of some decomposable poverty measures"[72]. Each member of this class satisfies the basic axioms for a poverty measure (for the Foster, Greer and Thorbecke measure only if $\alpha > 2$).

Eight years before the seminal work of Sen (1976), Watts (1968) had already identified the problem with the use of the headcount ratio when measuring poverty. Especially that "poverty is not really a discrete condition. One does not immediately acquire or shed the afflictions we associate with the notion of poverty by crossing any particular income line."[73] He defines a family's "welfare ratio" w as the ratio of its permanent income, $\mathbf{Y}$, to the appropriate poverty threshold[74] (z), that is,

$$w_i = \frac{\mathbf{Y_i}}{z} . \tag{1.25}$$

In order to get an aggregate measure of poverty, he proposes to sum the logarithms of welfare ratios, weighted by family size, over some part or all of the

[72] Zheng (1997, p. 150).
[73] Watts (1968, p. 325).
[74] In fact Watts (1968) defines the poverty threshold dependent on the family size, place and time.

lower half of the distribution of families, that is,

$$\textbf{Original Watts Measure}: P = \sum_{i \subset L} N_i \log(w_i)\ . \tag{1.26}$$

Where L is the set of subscripts belonging to families with $W \leq W^* \leq \tilde{W}$, N_i is the size of family i, and w_i is the welfare ratio of family i; $\log(x)$ denotes the logarithm of any (positive) number x. W^* is an essentially arbitrary threshold value comparable to the "poverty line". If W is equal 1 only poor peoples welfare ratios are taken into account and P cannot take on positive values. It would have a limiting value of zero if no one lies below the poverty line.

This index has long been neglected in the literature[75] until the 'reinvention'[76] by Blackburn (1989), followed by an axiomatic characterization of Zheng (1993) and the parametric formulation in Muller (2001) as well as the description in Zheng (1997). Although most people have adopted his definition of poverty, the Watts poverty measure was rarely used in empirical studies. However, the presentation of this measure slightly changed in the above mentioned papers, to

$$\textbf{Watts Measure}: W_p(y;z) = \frac{1}{n}\sum_{i=1}^{q}(\ln z - \ln y_i) \tag{1.27}$$

and Zheng (1997) showed the equivalence with

$$W_p(y;z) = H(y;z)(\ln z - \ln \zeta_p) = \ln z - \ln \zeta_z$$

where ζ_p is the representative income of the poor and ζ_z is the representative income of the censored income distribution. Blackburn (1989) showed that the Watts measure can also be written as

$$W_p(y;z) = H(y;z)\left[T_l\left(y;z\right) - \ln\left(1 - I\left(y;z\right)\right)\right] \tag{1.28}$$

where

$$T_l(y;z) = \frac{1}{q}\sum_{i=1}^{q}(\ln \mu_p - \ln y_i)$$

is an inequality measure proposed by Theil (1967). A very appealing property of the Watts index, apart from that the measure satisfy a very reasonable set of axioms (see table 1.1), is the interpretation as the absolute amount of social welfare loss due to poverty.

With their second social evaluation function[77] Clark et al. (1981) end up by replacing the Gini coefficient in the Sen measure with another inequality measure - the Atkinson measure. This second proposed poverty measure is subgroup

[75] A notable exception is Atkinson (1987) in his Walras-Bowley Lecture.

[76] Blackburn was aware of this after his article was completed, but he was the first showing the connection between the Watts measure (or his 'new' measure) and the Theil inequality index.

[77] For a description of the first measure see equation (1.18).

consistent but not decomposable and is defined by

Clark, Hemming & Ulph (2) Measure

$$\begin{aligned} CHU2(y;z,\beta) &= 1-\frac{1}{z}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\min\left\{y_i,z\right\}\right)^{\beta}\right]^{\frac{1}{\beta}} && (1.29)\\ &= 1-\Big(H\left[(1-A)(1-I)\right]^{\beta}+(1-H)\Big)^{\frac{1}{\beta}} && (1.30)\end{aligned}$$

with $\beta < 1$ and A is the well known Atkinson inequality measure.[78]As Clark et al. (1981, p. 522) state CHU_2 "therefore measures the proportionate welfare loss in having individuals with incomes below this, in terms of the proportion of the population which is so affected, the average deviation of their incomes from the poverty line and the inequality in their incomes"

Zheng (1997) shows that $CHU_2(y;z;\beta)$ can be expressed as a monotonic increasing function of a decomposable poverty measure for the different values of the parameter β. For $0 < \beta < 1$,

$$CHU_2(y;z;\beta) = 1-\left[1-C_h(y;z;\beta)\right]^{\frac{1}{\beta}}, \tag{1.31}$$

where $C_h(y;z;\beta)$ is the Chakravarty measure given in (1.36). For $\beta < 0$,

$$CHU_2(y;z;\beta) = 1-\left[1+\bar{C}_2(y;z;\beta)\right]^{\frac{1}{\beta}}, \tag{1.32}$$

in which $\bar{C}_2(y;z;\beta)$ is defined by

$$\bar{C}_2(y;z;\beta) = \frac{1}{n}\sum_{i=1}^{q}\left[\left(\frac{y_i}{z}\right)^{\beta}-1\right]. \tag{1.33}$$

And when β becomes zero,

$$\bar{C}_2(y;z;\beta) = 1-\frac{\left(\prod_{i=1}^{n}\{\min_i(y_i,z)\}\right)^{\frac{1}{n}}}{z} = 1-e^{-W_p(y;z)}, \tag{1.34}$$

[78] The Atkinson Index is an inequality measure that explicitly incorporates normative judgments about the underlying social welfare function Atkinson (1970). The index is derived by calculating the so-called equally-distributed-equivalent income (y_e), which is defined as that level of per capita income which if enjoyed by everybody would make total welfare exactly equal to the total welfare generated by the actual income distribution. The equally-distributed-equivalent income is given by:

$$y_e = \begin{cases} \left(\sum_{i=1}^{n} f(y_i)(y_i)^{(1-e)}\right)^{\frac{1}{(1-e)}} & e>0, e\neq 1\\ \sum_{i=1}^{n} f(y_i)\log y_i & e=1\end{cases}$$

The parameter e reflects the preferences for equality in the society under investigation, and can take values ranging from zero to infinity. When $e > 0$, there is a social preference for equality (or an aversion to inequality). As e rises, society attaches more weight to income transfers at the lower end of the distribution and less weight to transfers at the top. Typically used values of e include 0.5 and 2. The Atkinson Index (A) is then given by: $A = 1-\frac{y_e}{\mu}$ where μ is the mean income. The more equal the income distribution, the closer y_e will be to μ, and the lower the value of the Atkinson Index. For any income distribution, the value of A lies between 0 and 1.

where $W_p(y;z)$ is the Watts poverty measure as defined in equation (1.27).

The relationships noted by Zheng (1997) in (1.31), (1.32) and (1.34) are necessary results as predicted by the theorems proposed in Foster and Shorrocks (1991). Zheng also stated that empirical work often finds $\bar{C}_2(y;z;\beta)$ with negative value of β, or $\bar{C}_2(y;z;\beta)$, to be troublesome as the measure puts 'too much' weight to what happens to the bottom poor. In fact, as $\beta \to -\infty$, $\bar{C}_2(y;z;\beta)$ approaches

$$1 - \frac{\min_i(y_i)}{z}, \tag{1.35}$$

which is the Rawlsian maximin justice.

Chakravarty (1983b) followed Sen's axiomatic approach also in his second proposed measure, which is a decomposable poverty index. Unlike Sen (1976), Chakravarty derived his measure by solving a functional equation, which directly takes the three basic axioms into account. His measure is

$$\textbf{Chakravarty (2) Measure}:$$
$$C_h(y;z;\beta) = \frac{1}{n}\sum_{i=1}^{q}\left[1-\left(\frac{y_i}{z}\right)^{\beta}\right],\ 0<\beta<1 \tag{1.36}$$

$C_h(y;z;\beta)$ also satisfies transfer sensitivity and any higher level of sensitivity axioms. According to Chakravarty, the measure has a clear link to the normalized Theil entropy index when one replaces z with the mean income μ, normalizes the measure and sums over the uncensored income distribution (Chakravarty, 1983b).

Foster et al. (1984) proposed a class of poverty measures out of the practical demand for a decomposable poverty measure without a between group component.[79] Their very popular measure is defined as:

$$\textbf{Foster, Greer and Thorbecke Measure}:$$
$$FGT(y;z,\alpha) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^{\alpha},\ \alpha \geq 0. \tag{1.37}$$

The main difference between the Sen measure and the class of Foster et al. measures is the weighting function. Foster et al. uses the α ($\alpha \geq 0$) power of the income gap ratio between the poverty line and the poor incomes itself instead of the relative rank as done by Sen. The parameter α can be interpreted as an indicator of *aversion* to poverty (Seidl, 1988). If the parameter α is equal to zero, the extent of the poverty gap plays no role (the weighting is 1 for all poor persons), this yields the headcount ratio or poverty incidence, $H(y;z)$ in equation (1.2). If α equals one, the sum of the poverty gaps is taken into account (a linear weighting) and divided by the whole population. This results in an

[79] With the exception of the measure proposed by Chakravarty (1983b), see equation (1.36), all poverty measures after Sen and prior to Foster et al. (1984) are *not* decomposable in the sense defined above.

average poverty gap for the whole population (FGT_1). This is equivalent to $HI(y;z)$, as defined in equation (1.4).

Only when α is greater than one the FGT measure satisfies all axioms that the Chakravarty (1983b) measure satisfies except *weak transfer sensitivity.* Only if α is greater than two the measure satisfy also the *weak transfer sensitivity* axiom.[80]

If one chooses a weighting factor of two, $FGT(y;z,2)$, the resulting measure[81] can also be expressed as a function of the head count (1.2), the income gap (1.3) and an inequality measure, notably the coefficient of variation (Foster et al., 1984):

$$FGT(y;z,2) = H(y;z)\left[I^2(y;z) + (1 - I(y;z))^2\, CV^2(y;z)\right] \ , \tag{1.38}$$

where $CV^2(x,z)$ is the squared coefficient of variation of the poor incomes.[82] A similarity exists between the $FGT(y;z,\alpha)$ measure and the $CHU(y;z,\alpha)$, as Zheng (1997) pointed out:

$$FGT(y;z,\alpha) = [H(y;z)]^{1-\alpha}\,[CHU(y;z,\alpha)]^{\alpha} \ . \tag{1.39}$$

Because of this relationship "[...] one may regard the Foster et al. (1984) measure as essentially the Clark et al. measure applied to the censored distribution."[83]

If $\alpha \geq 0$ and $\alpha \to \infty$ then $FGT(y;z,\alpha) \to \frac{q_0}{n}$ with q_0 denotes the number of zero incomes, while the transformation of the Foster et al. measure, $[FGT(y;z,\alpha)]^{\frac{1}{\alpha}}$, approaches the Rawlsian maximum justice and the well-being of the poorest persons dictates the overall picture of poverty. However, the poverty ranking will be sensitive to the choice of poverty aversion parameter (in defining a particular FGT poverty-measure to use). Tungodden (2005) therefore recommend to report also what he calls the 'Critical Comparison Value', which describes the number of parameters (α) representing possible reversal points in poverty ranking and leaving the evaluation to the final decision-makers.

Schmid (1993) proposed a more general class of poverty measures which is closely related to the FGT. The class includes the measure proposed by Foster et al. (1984) (and therefore also the head-count ratio and the poverty gap) and the measure by Clark et al. (1981). Schmid defines the general class as:

$$\textbf{Schmid Measure}: \quad SCH(y;z;V) = \frac{1}{n}\sum_{i\in D(z)} V\left(\frac{y_i}{z}\right) \ . \tag{1.40}$$

[80] When $\alpha > 1$ this yields to a nonlinear weighting function with greater weight the poorer the persons.

[81] The value 2 for the parameter α is frequently used in empirical research, the resulting measure is often named poverty intensity.

[82] The squared coefficient of variation of the poor incomes is defined as

$$CV^2 = \frac{\sum_{i=1}^{q}(\mu_p - y_i)^2}{q\mu_p^2} \ ,$$

where $\mu_p = \frac{1}{q}\sum_{i=1}^{q} y_i$.

[83] Zheng (1997, p. 158, footnote 56).

If $V()$ is given by $(1 - \frac{y_i}{z})^\alpha$, $SCH(y; z; V)$ reaches the measure given by Foster et al. (1984), and if it is defined as $1 - (\frac{y_i}{z})^\alpha$ the measure is identical to the Clark et al. (1981) measure. The new proposed measure by Schmid uses for $V(y; z)$:

$$V(y; z) = Ae^{\beta \frac{y_i}{z}} \tag{1.41}$$

where $A > 0$ and $\beta > 0$. Although the measure satisfy all major axioms this measure is not used in any empirical analysis so far (to the best of my knowledge). This may depend on the fact that Schmid do neither derive any recommendations about the parameter A and β, nor gives he any help in the interpretation of these parameters. In his simulation study he uses $A = 1$ and $\beta = 2$ without any justification. However, the results a very reasonable and shows that the measure is much more sensitive to an increase of the poverty line than the FGT measure with $\alpha = 2$.

Hagenaars (1987) proposed a Dalton-Type class of poverty measures, which is the relative welfare poverty measure $V_r(y; z, W)$ in equation (1.23) when the social welfare function is utilitarian and y is the censored income distribution. Assuming every person of the society has the same utility function, U, the measures defined are:

$$\begin{aligned} &\textbf{Hagenaars} && \textbf{Measure}: \\ HG(y; z, U) &= && 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} \min\{U(y_i), U(z)\}}{U(z)} && (1.42) \\ &= && 1 - \frac{U(\zeta_z)}{U(z)}\ , && (1.43) \end{aligned}$$

where ζ_z is representative income of the censored income distribution and hence, $U(\zeta_z)$ is the representative individual utility. Members of this class include the measures proposed by Clark et al. (1981), the Blackorby and Donaldson (1980) measure, as well as the measures of Foster et al. (1984).

The properties of this measure depend largely on the specification of $U(x)$. If $U' > 0$, then $HG(x; z, U)$ satisfies the *monotonicity axioms*; if $U' > 0$ and $U'' < 0$, then $HG(x; z, U)$ satisfies *regressive transfer* and if $U''' > 0$, then $HG(x; z, U)$ satisfies *weak transfer sensitivity* (Hagenaars, 1987; Zheng, 1997). When *scale invariance* is maintained, the Hagenaars measure collapses to the Chakravarty measure as shown by Chakravarty (1990). When *translation invariance* instead of *scale invariance* is imposed, there is no satisfactory poverty measure in this class.

The specific poverty measure Hagenaars (1987) gave, is obtained by assuming that the utility function can be specified as $U(y) = ln(y)$. The resulting :

$$H(x; z) = \frac{1}{n}\sum_{i=1}^{q}\left(1 - \frac{\ln x_i}{\ln z}\right) = \frac{q}{n}\left(\frac{\ln z - \ln \bar{\bar{y}}_q}{\ln z}\right)\ , \tag{1.44}$$

where $\bar{\bar{y}}_q$ is the geometric mean[84] income of the poor. An interesting property is,

[84] The geometric mean is defined by:

$$\bar{\bar{y}} = \sqrt[n]{\prod_{i=1}^{n} y_i}\ .$$

that this index is neither relative nor absolute but satisfies the *increasing poverty line* axiom.[85] Shorrocks (1998) showed also the connection with the poverty gap profile, in a way that the indeces can be interpreted as the weighted area under the poverty gap profile curve.

1.4 Summary

Unfortunately, but not unexpected, this chapter does not give an easy answer to the question which poverty measure one should use. And as Foster (1984) notes ‘the choice of a *single* poverty measure involves a certain degree of arbitrariness’. However, the purpose of this chapter was rather to provide an overview and last but not least some clarification of the extensive literature on poverty measurement since the seminal work of Sen (1976). To fully understand each measure it is first of all important to know of the underlying axioms and the interrelationships among them. Based on the set of core axioms the reader gets some further information for an appropriate decision at his hand.

30 years after the work of Sen, it is widely recognized that a measure of poverty intensity needs to be distribution sensitive in addition to only counting the poor and calculating the average shortfall of persons below the poverty line. However, many of the developed measures are not easy to interpret, which avoids a wider perception by more applied poverty researches. From the group of the more advanced poverty measures the only one which was able to take this barrier of ‘easiness’ was the one proposed by Foster et al. (1984), which has a very intuitive understanding and close relationship to the ‘famous’ poverty head count ratio. This is, at least in part, due to the intuitive weighting function of the poverty gaps by itself and not by a derived welfare function.

However, focusing on ‘nearly one’ measure sometimes prevent to notice that for some more specialized questions about poverty other measures may be more applicable. For instance the decomposition by the source of the change in poverty over time can be better accomplished with the modifications of the Sen measures by Shorrocks and Thon[86]. But as poverty is a very complex phenomenon the measurement should not be oversimplified or restricted and as many research questions exists as many specialized measures may needed. Especially more welfare oriented research on poverty can maybe better accomplished with a ethical poverty measure, where the social evaluation function are stated explicitly and the interpretation goes along the loss of welfare due to poverty.

[85] See Zheng (1994) for further info on this topic.
[86] For a description of this decomposition see Xu and Osberg (2001).

Chapter 2

Measuring Poverty over Time

The increasing use of a dynamic perspective in poverty research has meant that more and more researchers distinguish between persistent (or chronic) poverty and transitory poverty. The classification of different durations of poverty by Walker and Ashworth (1992) is based on the widespread method of N-times-X (NTX) measures and the associated use of individual cross-sectional income. The NTX approach just count how many times a persons is observed as poor out of a given time period. Results from research on mobility and inequality indicate, however, that these studies of income at separate points in time fail to take long-term welfare level into account. The analysis of the stability of poverty is also important in more sociological concepts like welfarization and forming of an underclass.[87]

Using the well established concept of permanent income[88] Rodgers and Rodgers (1993) suggested dividing poverty into chronic poverty and transitory poverty.[89] These results were then compared with those of the NTX measurement[90] Their results for the United States indicated that the NTX-method drastically underestimated the proportion of persons in persistent poverty among impoverished as a whole. One advantage of the method from Rodgers and Rodgers is that poverty measures other than the head-count ratio can be used.[91] However, the Rodgers and Rodgers index is not normalized (i.e. it does not ensure that the value calculated lies between zero and one) and require a mean based or absolute poverty line (i.e. the popular median based poverty line does not fir into this framework). The approach discussed here, which is based on measuring permanent inequality as suggested by (Shorrocks, 1978) and a transformation of

[87]See for example Kronauer (1997).
[88]For the use in poverty measurement see Watts (1968) and citations therein.
[89]A similar decomposition was also proposed by Ravallion (1988) and Ravallion (1998). The latter uses also censored conditional quantile estimation methods to get more robust results, with respect to distributional misspecifications of the error term, for the modeling of influential characteristics on transient and chronic poverty.
[90]Other possibilities of directly including the duration of poverty in the poverty index are illustrated for the Sen-Shorrocks-Index by Osberg and Xu (2000) and for the FGT-Index by Weikard (2000).
[91]For another application, see Hill and Jenkins (2001) or Valletta (2006).

the incomes according to the poverty lines, overcome this weakness.

The index introduced by Shorrocks (1978) (the so-called Shorrocks-Stability-Index or Shorrocks-Index) is widely used to compare inequality in m periods with the inequality in the aggregated period to measure income (im-) mobility.[92] The index measures the exact ratio of permanent inequality to total inequality with respect to individual income mobility over time.

The chapter will prove that any additive decomposable[93] poverty measure – here shown for the index introduced by Foster, Greer and Thorbecke (1984) (*FGT*)[94] – corresponds with the axioms used by Shorrocks (1978) and that the Shorrocks index can therefore be used for the analysis of the effects of income mobility for poverty measurement as well.

The chapter starts with a short description of Rodgers & Rodgers as well as Shorrocks basic ideas, then Shorrocks' concept will be extended to the measurement of poverty. Finally, some empirical results will be presented. First only for Germany, where income concepts with different references to the underlying time period (monthly vs. annual income) will be compared. The impact of these differences on poverty stability will be analyzed. And secondly, a cross-national comparison between the degree of poverty stability within Germany and the United States will be undertaken. For Germany data from the German Socio-Economic Panel (*SOEP*) is used, and for the USA data from the Panel Study of Income Dynamics (PSID) as included in the Cross-National Equivalent Files (CNEF).

2.1 The measurement of persistent poverty

A commonly used approach for the measurement of persistent and transitory poverty is the *NTX* method (N times X). It tabulates the proportion of persons with incomes below the poverty line in x out of n time periods, where $0 \leq x \leq n$. The prevalence of persistent versus transitory poverty is then assessed by comparing the proportion of people who where poor in all periods with the proportion of people who were poor in just one or a few periods.[95]

Methods based on permanent income represent an alternative approach: (real) income from several years are simply added up or averaged. This concept assumes implicitly that individuals can perfectly shift and redistribute their income over time, and hence that an individual can compensate for a short-term stint in poverty by a perfect self-insurance. Because inter-temporal redistribution may be costly Rodgers and Rodgers (1993) attempt to take the different interest rates for debt and savings into account. They give the following reason for doing so:

[92] Burkhauser and Poupore (1997), Fabig (1999).

[93] For the discussion of necessary axioms within the measurement of poverty see Zheng (1997) or Bourguignon and Chakravarty (2002) and chapter 1.

[94] This family includes the head-count ratio, the poverty gap and reflects the distribution of income among the poor with an $\alpha > 1$, see equation (2.11)and chapter 1 for a more detailed discussion.

[95] Duncan et al. (1993) and Walker (1998).

> "Given a T-year observation period, our measure of an agent's permanent income, Y^*, is equal to the maximum sustainable annual consumption level that the agent could achieve with his or her actual income stream over the same T years, if the agent could save and borrow at prevailing interest rates."[96]

The disadvantage of this type of approach lies with its failure to reflect individual need and therefore the poverty line. This leads to a situation in which savings and debt exist for all incomes (except for if income was the same in all the years), regardless of the income level. Since poor households often are liquidity constrained, the interest rate should not be applied symmetrically to incomes above and below the poverty line.[97]

In Rodgers and Rodgers approach, an average annual poverty index, which consists of the weighted average of the annual poverty measures for each year, is first defined as:

$$A_p(T) = \sum_{t=1}^{T} w_t P_t \ , \tag{2.1}$$

where $w_t =$ weight at time t, $\sum_{t=1}^{T} w_t = 1$ and $P_t =$ is an additively decomposable poverty measure at time t. Rodgers and Rodgers use the percentage of the cross-sectional population in the total population as weight ($w_t = \frac{n_t}{N}$, where $N = \sum_{t=1}^{T} n_t$ and n_t the number of respondents at time t). They define their population of interest to be all individuals who are present at the end of the income period (Rodgers and Rodgers, 1993, p. 36).

In order to measure chronic poverty, permanent income during the period studied is used and a poverty measure (P) is calculated: They use the head-count ratio and the $FGT_{\alpha=2}$.[98]

$$C_p(T) = P(\mathbf{Y}) \ , \tag{2.2}$$

where $\mathbf{Y} = (Y^*_{T1}, Y^*_{T2} \ldots Y^*_{Ti} \ldots Y^*_{Tn})$. n is the size of the population and Y^*_{Ti} is the permanent income of person i during the period T. The extent of transitory poverty is therefore defined (according to Rodgers and Rodgers) as the difference between the two indices:

$$T_p(T) = A_p(T) - C_p(T) \ . \tag{2.3}$$

It is not necessarily the case that $A_p(T) \geq C_p(T)$, Rodgers and Rodgers comment on this as follows:

> "A positive $T_p(T)$ equals the amount of poverty which is not chronic in an average year. A negative $T_p(T)$ equals the amount of chronic poverty which is temporarily absent in an average year." (Rodgers and Rodgers, 1993, p. 32)

[96] Rodgers and Rodgers (1993), p. 31.

[97] For a more detailed discussion, see Rodgers and Rodgers (1993), p. 36f and Valletta (2006) as well as Jäntti and Danziger (2000).

[98] The head-count ration is defined as $\frac{q}{n}$, where $q =$ number of persons poor and n number of all persons. For the definition of the FGT, see equation 2.11.

and, furthermore,

> "We have found that negative values for $T_p(T)$ occur quite frequently when the poverty index, P, is the head-count ratio [...]. Negative values seldom occur when P is a function of poverty gaps." (Rodgers and Rodgers, 1993, footnote 6)

This means, however, that the percentage of persistent poverty in average poverty is not always between zero and one. The incidence of a negative value does not necessarily indicate that chronic poverty did not appear in a certain year.[99] This value can be attributed to two other factors. One is the weights used: they depend on the size of the population at the individual periods of time. This is particularly the case with the head count ratio, because each persons poverty status is thereby counted using the same weight (in contrast to the $FGT_{\alpha>1}$). The second lies in the use of a fixed poverty line. Because of these two factors the concavity of the index during several periods is no longer ensured. A method of modifying the index so that it ensures a normalized value between zero and one and is independent from the poverty line use, will be outlined in the next two sections.

2.1.1 Measuring permanent inequality

Shorrocks showed that there is a clear connection between individual income mobility and inequality in each period compared to the overall or permanent inequality. Where permanent inequality is defined as the degree of inequality in permanent income, measured as the individual sums of income in each time period. Shorrocks (1978) defines the following conditions for inequality measures to derive his coefficient of income stability:

$$I[\mathbf{Y}] = g\left(\frac{\mathbf{Y}}{\mu}\right), \tag{2.4}$$

where I is any convenient inequality measure, $\mathbf{Y}$ an income vector of length n ($\mathbf{Y} = \{Y_1, Y_2, \ldots, Y_n\}$) and μ the (cross-sectional) mean of incomes over all persons ($\frac{1}{n}\sum_{i=1}^{n} Y_i$).

Using a social welfare function (continuous, strictly monotonic and symmetric) written as a function of $\mathbf{Y}$ ($W = F(\mathbf{Y})$), the welfare level of the equally distributed equivalent income[100] (Y_e) is then equal to the welfare level of the empirical distribution:

$$F(Y_e, Y_e, \ldots, Y_e) = F(\mathbf{Y}) \tag{2.5}$$

and can be written explicity as

$$Y_e = f(\mathbf{Y}) > 0, \tag{2.6}$$

[99] This is a rather unlikely interpretation when one considers that the term chronic does, by definition, mean that a person is *always* poor

[100] "... the equally distributed equivalent income Y_e is that amount of income which, if received by all individuals, generates the same level of social welfare as the distribution $\mathbf{Y}$." Shorrocks (1978), p. 379.

where $f(.)$ is also continuous, strictly monotonic and symmetric. The inequality index is then

$$I = 1 - \frac{Y_e}{\mu} = 1 - \frac{f(\mathbf{Y})}{\mu} = 1 - f\left(\frac{\mathbf{Y}}{\mu}\right), \tag{2.7}$$

if $F(.)$ is assumed to be strictly quasi-concave and homothetic. Therefore $f(.)$ is between $[0,1]$ and linear homogeneous. These steps correspond to the assumption that the inequality measure is mean independent.

Shorrocks (1978) further shows that $g(x) = 1 - f(x)$ is strictly convex: Define $\sigma > 0$ such that $f(x) = \sigma f(x') = f(\sigma x')$ and after certain transformations (for further details, see Shorrocks, 1978, p. 380) one reaches the following conclusion, which is the basic condition for convexity:[101]

$$\begin{aligned} g(\lambda x + (1-\lambda)x') &= 1 - f(\lambda x + (1-\lambda)x') && (2.8)\\ g(\lambda x + (1-\lambda)x') &\leq \lambda g(x) + (1-\lambda)g(x') && (2.9) \end{aligned}$$

Finally, the index of income stability can be written as the inequality of the averaged income divided by the weighted average of the inequality in each period:

$$\mathbf{S}_I = \frac{I(\mathbf{Y})}{\sum_{t=1}^{T} w_t I(y_t)}, \tag{2.10}$$

where $Y = \sum_{t=1}^{T} y_t$, $w_t = \frac{\mu(Y)}{\mu(y_t)}$ and $I(.)$ is a convenient convex inequality measure.[102]

2.1.2 The measurement of persistent poverty reconsidered

The well known poverty measure introduced by Foster et al. (1984) is defined as (see Foster 1984 and Zheng 1997 for a discussion of other poverty measures):

$$\mathbf{FGT}_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z - y_i}{z}\right)^\alpha ;\ \alpha \geq 0 \tag{2.11}$$

where $q =$ is the number of persons poor, $z =$ poverty threshold, $n =$ number of all persons, $y =$ an indicator to describe poverty (i.e. income) and $\alpha =$ a parameter for weighting the individual poverty gap.

Besides other features, the FGT measure is mean independent with the use of a relative mean based or absolute poverty threshold. This means that

$$\mathbf{FGT}(y) = \mathbf{FGT}\left(\frac{y}{\mu}\right). \tag{2.12}$$

To use the FGT with the Shorrocks-Index, the following condition has to be satisfied:

$$\mathbf{FGT}(\lambda \cdot x + (1-\lambda)\cdot x') \leq \lambda \cdot \mathbf{FGT}(x) + (1-\lambda)\cdot \mathbf{FGT}(x'). \tag{2.13}$$

[101] Where $0 \leq \lambda \leq 1$ and describes the range in which the convexity has to be fulfilled.
[102] For a derivation of w_t see Theorem 1 in Shorrocks 1978, p. 381.

In other words, it has to be strictly convex. This will be the case, because "the strict convexity of P_α in the vector of poor incomes for $\alpha > 1$" is given (Foster et al. (1984), p. 763).

This means that the use of a mean based or absolute poverty line allows to simply plug-in the FGT into the Shorrocks index. Assuming these steps, the Shorrocks Index can be similarly applied to measure the extent of persistent poverty and will be defined as follows:

$$\mathbf{S}_P = \frac{P(Y)}{\sum_{t=1}^{T} w_t P(y_t)}, \tag{2.14}$$

where $Y = \sum_{t=1}^{T} y_t$, $w_t = \frac{\mu(Y)}{\mu(y_t)}$ and $P(.)$ is a convenient convex poverty measure, like $FGT_{\alpha>1}$.

The proportion of persistent poverty can therefore be calculated precisely using

1. the concept of permanent income,
2. a convex poverty measure and
3. a relative mean based or absolute poverty line.

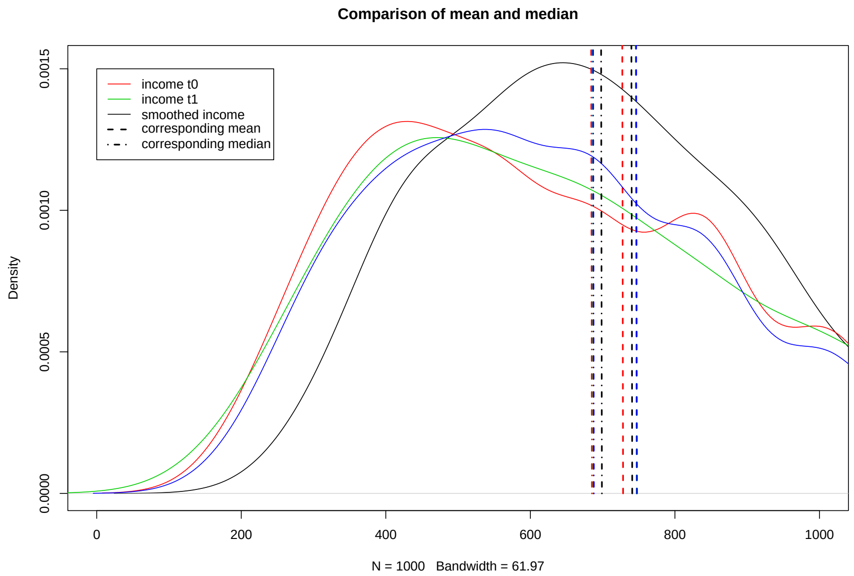
But having to choose a poverty line that is a fraction of the mean income is undoubtedly restrictive. One might to use median based poverty lines for example. Therefore the next section will propose a possibility of a poverty-line-independent form of this index.

2.1.3 Poverty-line-independent measurement of persistent poverty

The calculation of the smoothed income which is the base for the measurement of chronic poverty is a simple average over the income received at different time points. The consequence is that a median based poverty line could be very different between the yearly incomes and the smoothed income, which depends on the skewness of the particular yearly income distribution and individual income mobility.

To illustrate this problem figure 2.1 shows density functions of hypothetical "cross-sectional" income distributions at three points in time (red, green and blue line) and the corresponding "longitudinal" smoothed income distribution (black line), defined as the individual means over the three "cross-sectional" distributions. The vertical dotted lines are the belonging means and medians of these distributions, indicated by the colors. The lines with additional points are the medians, the others are the means. The figure shows that the black mean line (smoothed income) is between the maximum and minimum of the cross-sectional means. In contrast the behavior of the median, the median of the smoothed income lies further to the left, i.e. the median based poverty line of the smoothed income is higher than the cross-sectional (median based) poverty lines. This problem is the greater the more skewed the cross-sectional income distributions and the greater the individual income mobility.

Figure 2.1: Behavior of median and mean using the smoothed income concept



Having in mind that the 'smoothed income' is a theoretical concept one could also use a transformation of the current incomes for describing the 'persistent' poverty situation of individuals. Nearly all poverty measures satisfies the scale invariance axiom and therefore a linear transformation of incomes have no impact on the measured poverty.[103] Also because of the ficticious nature of the 'smoothed income' setting a poverty line for this situation is without any theoretical foundation (except of an absolute poverty line which will be unchanged). The poverty lines of current incomes are undoubtedly equally arbitrary but an alternative is not available or at least not in common use by poverty researchers[104].

Redefining the definition of the 'persistent poverty' situation in a way that it depends on the current distance from the current poverty line in each point in time will overcome the weakness that the Shorrocks measure can only be used with a mean dependent poverty line. Without loss of generality one can redefine the current incomes in the following way[105]:

$$y_t^* = \frac{z_t - y_t}{z_t} \ . \tag{2.15}$$

Note that all incomes are included, not only incomes lower than or equal the poverty line, therefore the new income y_t^* is defined from $-\infty$ to 1. If the transformed income is below zero the household is non-poor and if the

[103] See Foster (1984) and Zheng (1997) as well as chapter 1 for details.

[104] For a definition of a poverty line combining relative and absolute properties see D'Ambrosio, Muliere and Secchi (2002); Madden (2000); Hagenaars and van Praag (1985).

[105] Note that this is *not* equal to the censored distribution of normalized poverty gaps in the usual sense, as used by Takayama (1979) or Shorrocks (1995)

transformed income is between zero and one the household is poor in the current period. Hence this income reflects the relative (negative or positive) distance from the poverty line at a given point in time. The 'smoothed' transformed income can be calculated by

$$Y^* = \frac{1}{T}\sum_{t=1}^{T} y_t^* \ , \tag{2.16}$$

which is simply the average relative distance (also positive or negative) from the time specific poverty lines at a period of time. With the help of this transformation it is no longer necessary to define a poverty line of the smoothed income (including the drawbacks described above), the actual defined thresholds are "responsible" for the persistent poverty state of each person. This procedure does not need any additional assumptions than the ones currently used in poverty research, apart from accepting that current poverty status will effect adjacent poverty.

Using the transformed income the FGT measure will become

$$\mathbf{FGT}_\alpha(y^*) = \frac{1}{n}\sum_{i=1}^{q} (y^*)^\alpha \ ; \ \alpha \geq 0 \tag{2.17}$$

with q is the number of persons with y^* greater than or equal zero.

This means that the use of incomes relative to cross-sectional poverty lines (independent of their definition) allows to use an equivalent to the Shorrocks index in the measurement of chronic poverty. The poverty stability coefficient can therefore defined as follows:

$$\mathbf{PSC} = \frac{P(Y^*)}{\sum_{t=1}^{T} w_t P(y_t^*)}, \tag{2.18}$$

where $Y^* = \frac{1}{T}\sum_{t=1}^{T} y_t^*$, $w_t = \frac{\mu(Y^*)}{\mu(y_t^*)}$ and $P(.)$ is a convenient convex poverty measure, like $FGT_{\alpha>1}$.

The proportion of persistent poverty can therefore be calculated precisely using

1. the concept of smoothed poverty gaps,
2. a convex poverty measure and
3. with any cross-sectional poverty lines.

2.2 Empirical Application

2.2.1 Different concepts for measuring persistent poverty

2.2.1.1 Mean based vs. Median based poverty line

The FGT poverty measure from Foster et al. (1984) was used for calculating the poverty stability coefficient, with $\alpha = 2$ (so called poverty intensity) as well as

Table 2.1: Poverty Stability for West-Germany (1985-2002) with increasing periods for mean and median based poverty line

	Poverty				Inequality
Year	Median based poverty line	Decrease from prev. year	Mean based poverty line	Decrease from prev. year	Net household income
1984	100	–	100	–	100
1985	74.07	25.93	61.83	38.17	87.27
1986	63	11.06	46.86	14.98	81.51
1987	56.39	6.612	38.44	8.419	79.33
1988	52.09	4.304	35.77	2.662	76.16
1989	47.82	4.266	32.93	2.84	74.19
1990	45.51	2.317	31.19	1.743	73.16
1991	41.09	4.413	25.34	5.854	71.76
1992	37.12	3.97	22.16	3.181	70.19
1993	35.99	1.132	21.87	0.2874	69
1994	34.55	1.443	20.62	1.254	68.05
1995	33.74	0.8093	20.01	0.6022	67.56
1996	32.8	0.9354	19.56	0.4579	66.94
1997	31.35	1.456	18.51	1.046	66.5
1998	29.94	1.411	17.4	1.112	66.25
1999	29.16	0.776	17.07	0.3321	66.08
2000	28.18	0.9802	16.72	0.3435	65.81
2001	27.64	0.54	16.61	0.1146	65.39
2002	26.47	1.175	16.25	0.355	64.55

Source: *SOEP* 1984-2002, author's calculations (monthly net household income, modified OECD-Scale).

with $\alpha = 0$ (headcount ratio). However, given the use of the $FGT_{\alpha>0}$ index, the interpretation of the results is somewhat different as compared to the *classical NTX* method which uses the head count ratio. A result of 40% does not mean that 40% of poor persons should be classified as persistently poor, but rather that 40% of *poverty, as measured by the $FGT_{\alpha>0}$ index*, is classified as persistent poverty.

Table 2.1 and 2.2 show the values of the stability indices calculated for a mean based and a median based poverty line.[106]The values for 1984-2002 (Table 2.1) indicate that the clearest reduction of persistent poverty took place in the first three to five years under investigation for both poverty lines considered. Persistent poverty continously decreased after that, but at a slower rate. After ten years there is hardly any reduction to observe. Note that by definition it is not possible that poverty stability increases over time. A comparison of persistent poverty measured on the one hand by a mean based poverty line and

[106]The thresholds were defined according to 50% of mean income and 60% of median income. A *balanced-panel*-Design was used to calculate these values for each of the years listed, in table 2.1 balanced over 1984-2002 and in table 2.2 over 1997-2002. The data are weighted with the *SOEP* longitudinal weighting factors according to the underlying time period. The incomes are real household net incomes adjusted for inflation (base year: 2000), separately for East- and West-Germany and adapted to different household size by means of the modified OECD scale. However, for long periods involving 1990 only persons from West Germany could be included within the calculation leading up to reunification.

on the other hand by a median based poverty line reveals that the *reduction* of persistent poverty is stronger within the first three years when using a mean based poverty threshold, but is lower in the following years. This results in a 16.5% lower persistent poverty in 1985, which rises to a stable difference around 40% from the year 1992 (7th year within the period under investigation) onwards.

Table 2.1 also provides the values for the Shorrocks inequality stability coefficient (Column 6). The reduction of inequality over time, as indicated by these values, is very low compared to the reduction in persistent poverty. The reason is that only the lowest end of the income distribution is considered when looking at poverty reduction, while a decrease in inequality reflects developments within the income distribution as a whole.

2.2.1.2 Annual vs. Monthly Net Household Income

The SOEP provides figures for two different income concepts, on the one side the monthly net household income and on the other side the annual period.[107] Both concepts have pros and cons, however, the most important difference for our setting is the fact that they differ with respect to the underlying reference period.[108] The monthly measure maps only a snapshot for the month of the interview, whereas the annual income of the previous year gives a more comprehensive description of the well-being of households. One can argue, that the poverty stability should be higher for the annual figures, because of the possibility to smooth income over months within one year. In fact, if the annual income were only the sum of 12 monthly income figures, poverty stability should be higher. But if for example some structural differences in income components not covered by the monthly figures exist, this effect could be lowered or even reversed. To what extent the estimation of poverty stability is different for these two concepts is indeed an empirical question.

The results for monthly and annual income are provided in figure 2.2, the lines represent poverty stability over time[109] and the different points are the actual cross-sectional poverty rates using different income concepts and poverty lines for each year respectively. The graphic shows clearly that according to the measurement of poverty stability both concepts are measure nearly the same, especially if the time period gets longer. Only for poverty stability with a median based poverty line a small difference in levels persists. For the cross-sectional poverty rates the annual income shows slightly higher numbers over the whole period under investigation. But looking at the permanence of poverty the annual income figures show a lower stability (or a higher mobility) as compared to the monthly figures.

[107] Both concepts describe net equivalent household income, for both are the same methods applied as described in footnote 106.

[108] See Böheim and Jenkins (2000) for details.

[109] The red lines are for annual income and the black ones are for monthly income. Solid lines represent results based on a mean based poverty line and the dotted ones for a mean based poverty line.

Table 2.2: Poverty Stability for Germany (1997-2002) with increasing periods for monthly and annual net-equivalent income.

Year	Poverty				Inequality
	z = 60%-Median		z = 50%-Mean		
	Monthly income	Annual income	Monthly income	Annual income	Annual income
1997	100	100	100	100	100
1998	72.43	71.76	66.01	63.12	87.11
1999	65.97	65	57.12	56.39	83.5
2000	60.26	59.31	47.13	50.21	80.77
2001	56.89	56.4	44.21	47.87	79.39
2002	53.22	54.3	40.18	46.07	77.45

Source: *SOEP* 1984-2002, author's calculations (real monthly and annual net household income, modified OECD-Scale).

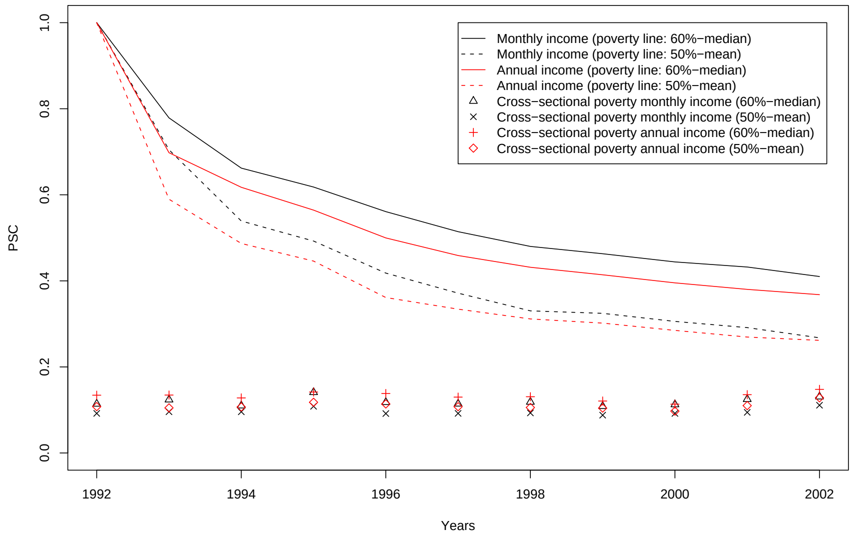


Figure 2.2: Poverty stability for monthly and annual income between 1992 and 2002 in Germany

2.2.2 Persistent poverty in the USA and Germany

For a comparison between Germany and the USA from the view point of welfare economics one is often referred to the categories proposed by Esping-Andersen (1990). Germany is described as a member of the corporatist-conservative welfare regime and the USA is characterized as belonging to the liberal welfare regime according to this categorization. For a more detailed discussion see e.g.

Goodin, Headey, Muffels and Dirven (1999) as well as Krause et al. (2003) and articles therein. Within this framework one would expect on the one hand that poverty mobility is higher in the US than in Germany, e.g. because labor market regulations are much higher in Germany and therefore it is more difficult to get out of unemployment, a major cause of poverty. On the other hand the welfare benefits are more generous in Germany than in USA, which may keep more persons out of poverty. How effectively the welfare systems prevent income poverty can be investigated by comparing index estimates for pre- and post-government income poverty.

The data for this analysis are the Cross-National Equivalent Files (CNEF), which provide special versions of the *SOEP* and of the Panel Study of Income Dynamics (PSID) especially for the use in cross-national analysis.[110]For the comparison the period from 1985 to 1989 as well as the years from 1993 to 2001 are chosen. These years were selected for three reasons.

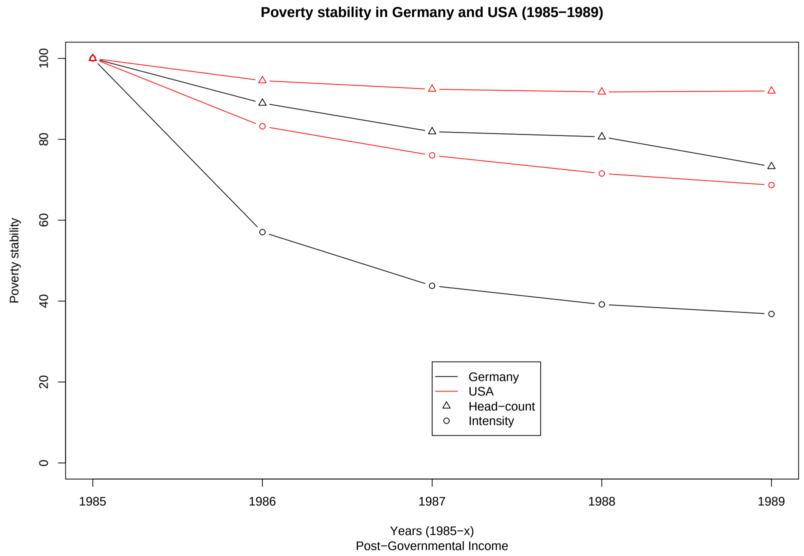
1. The *SOEP* started in 1984, so the first annual income is available in 1985.
2. Because of the unification in Germany in 1990 there is a break. Only for post-1990 period East-Germany can be included, simply because of data availability.
3. From 1997 onwards the PSID has collected data only every second year, therefore the years 1998 and 2000 are also dropped from the German sample.

2.2.2.1 Post-Government Income

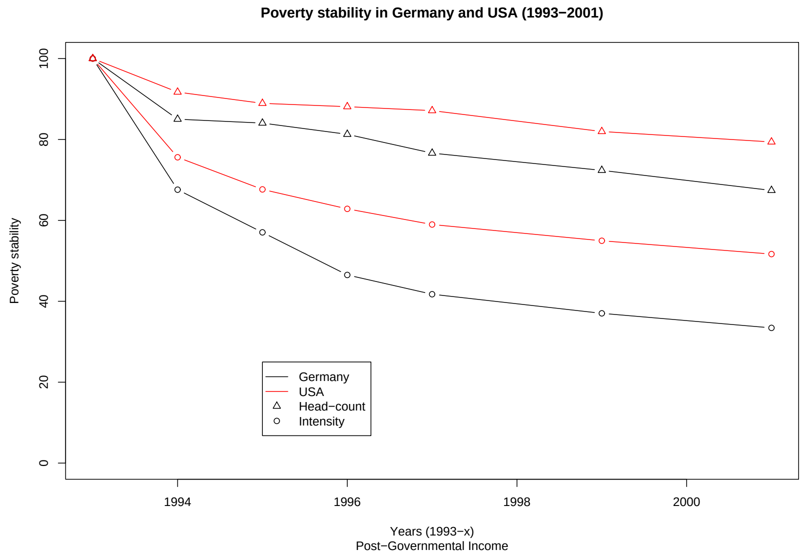
Figure 2.3 shows in the upper part the poverty stability for the period 1985 to 1989 and in the lower part the years from 1993 to 2001. The red lines represent the results for the USA and the black lines for Germany respectively. When the head-count ratio ($FGT_{\alpha=0}$) was used the results are represented with triangles, and for the poverty intensity measure ($FGT_{\alpha=2}$) circles accordingly.

Figure 2.3(a) indicates that there is a major difference between the stability of poverty in the United States and Germany. Especially when using the head count ratio the mobility in Germany out of poverty is clearly higher for post government income. This difference is smaller when comparing the stability of poverty intensity. However, it seems that this differential alleviate over time, because in figure 2.3(b) the poverty mobility has risen in the USA. However, poverty stability in the USA it seems to be relatively high, and theories about the possible formation of an under-class seems to be more relevant for the US than for Germany, when comparing post government income poverty. The more generous welfare state in Germany is much more effective in preventing households from long time poverty than the more liberal one in the USA. These results are perfectly in line with findings from comparative research about income mobility in the USA and Germany, which find that surprisingly mobility is higher in

[110] The income is annual net household income from the previous year, and all methods as describe in footnote 106 are applied. For a more detailed description of the CNEF data, see Burkhauser, Butrica, Daly and Lillard (2001).



(a) 1985 - 1989



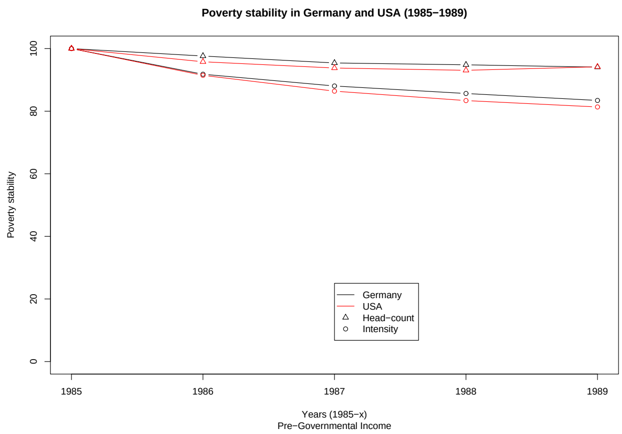
(b) 1993 - 2001

Figure 2.3: Poverty stability in Germany and the USA for post-government income (1985-1989; 1993-2001)

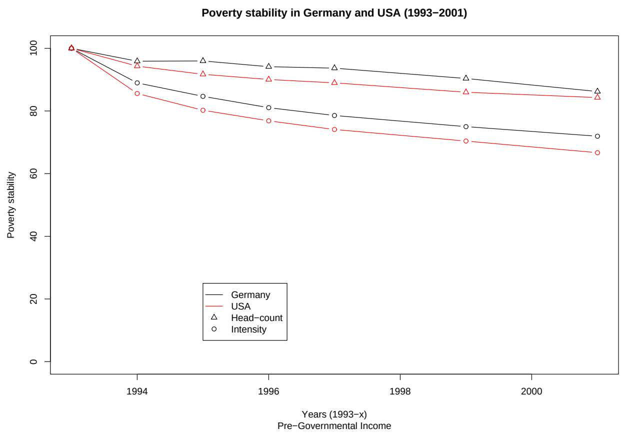
Germany. See for example Burkhauser and Poupore (1997); Burkhauser, Holtz-Eakin and Rhody (1998); Maasoumi and Trede (2001); Jenkins and van Kerm (2003); Schluter and Trede (2003) and van Kerm and Jenkins (2003).

2.2.2.2 Pre-Government Income

A quite different situation appears if we compare the poverty stability by means of pre government income. For the late 1980s there is hardly any difference to find. In both countries poverty stability is much higher when using market incomes in place of income after the tax and transfer system. Even after five years over 90% of poverty intensity has to be called persistent, for both countries. In the 90th there is a slight better situation to state for the USA than in Germany. Presumably this is because of the employment situation in East Germany, especially of growing long-time unemployment. However, given the more liberal job market in the USA the differences found are of minor scale.



(a) 1985 - 1989



(b) 1993 - 2001

Figure 2.4: Poverty stability in Germany and the USA for pre-government income (1985-1989; 1993-2001)

2.2.2.3 Poverty stability with shifting time windows

The drawback when analyzing poverty stability with an increasing time period as above is that the calculation of the index requires a fully balanced panel, which could be a potential source of selection bias. Especially in longer periods (even temporary) unit-nonresponse can bias the results. Therefore in this section the length of the period to determine persistent poverty is fixed to three "observation points". The underlying years differ between three years (1984-1997) and five years (1997-2001) because of the switch in the PSID interviewing time-lag. Figure 2.5 shows the percentage of persistent poverty for a shifting time window of three years from the mid-1980s to the beginning of the 21st century for post- and pre-government annual income of the previous calendar year. The definition of poverty is 60% of the country specific contemporary median income and as poverty measure $FGT_{\alpha=2}$ is used. Differences in the results to the figures above for the first years are because of the different sample design. For the following figures a running three year balanced panel is used as opposed to the definition before, where a balanced sample of individuals present over the whole period was used.

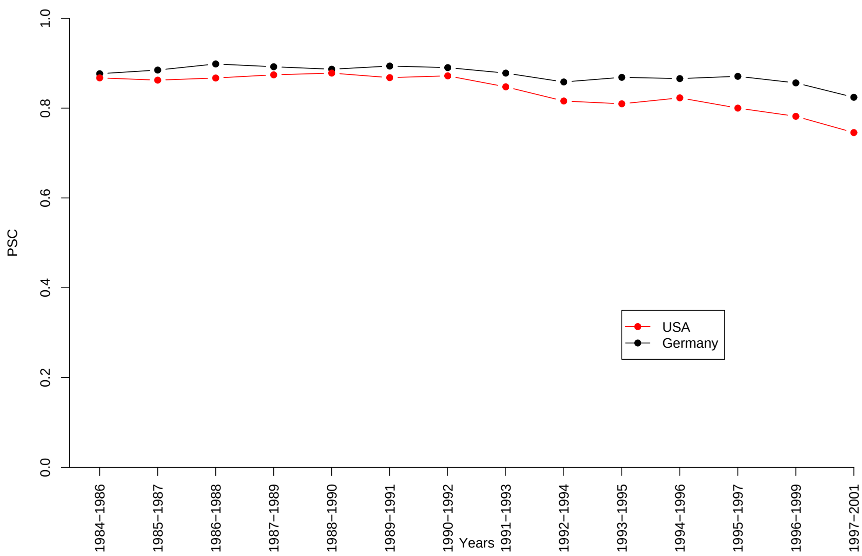
From figure 2.5(a) on the following page it is evident that during the 1980s and the beginning of the 1990s no distinction between the stability of poverty in the USA and Germany can be made when using pre-government income. However, during the mid 1990s there is an emerging discrepancy to notice, which is greatest in the most recent years. This widening gap after the German unification is in line with the above described results for the longer balanced panel.[111]

As already described the picture changes dramatically when switching to post-government income. Now during the 80th there is huge difference between the two countries to state (1984-1986; Germany: approx. 40% and USA: approx. 75%). On the one side this much lower poverty stability in Germany during the early years is continously rising until the mid 90th, and on the other side the high stability observed in the USA at the beginning is also continously declining over the years. This assimilation process peaks into a difference of no more than 10 percentage point in the period 1995-1997. However, there seems to be a tendency of decline in poverty stability in the most recent years in Germany, which cause a slightly greater distance between Germany and the USA.

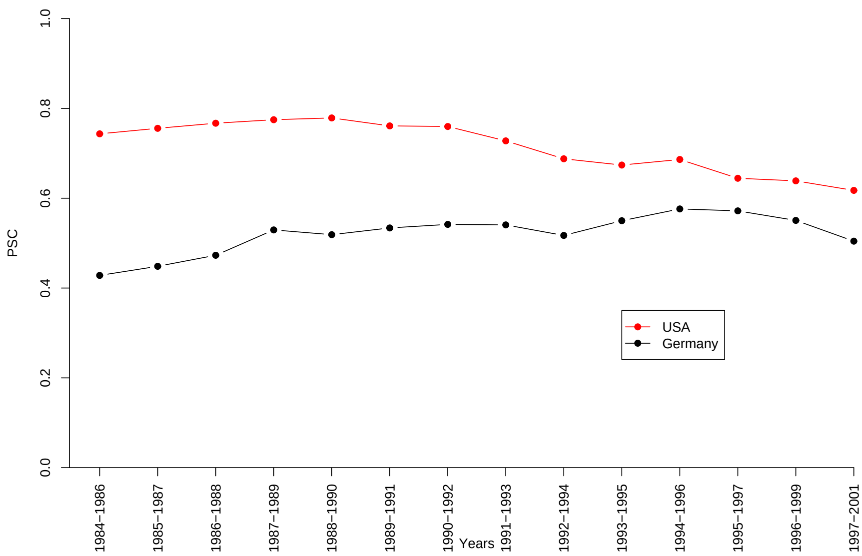
2.3 Summary

This chapter shows that it is convenient to use the Shorrocks approach to measure income immobility or stability not only in the context of inequality, but also with a convex poverty measure, e. g. the FGT with a poverty aversion parameter of $\alpha > 1$. This feature makes it possible to enhance the dynamic description of poverty, because the extent of persistent and transitory poverty can be estimated with the use of longitudinal datasets. With a transformation

[111] For a more detailed description of the rising inequality after unification for pre-governmental income especially in the Eastern part of Germany and the connection to poverty, see Goebel, Krause and Schupp (2005) and chapter 5.



(a) Pre-Government Income



(b) Post-Government Income

Figure 2.5: Poverty stability in Germany and the USA for a shifting time period of three years for Pre- and Post-Government income (1984-2001))

of the current income, in a form that the percentage shortfall of the current poverty line is expressed, also median based poverty lines can be used. A further advantage of this approach to measure permanent poverty, is that Shorrocks' 'framework' is well-known, can be interpreted easily, and is sensitive to the amount of mobility within the population.

In the empirical part robustness for mean and median based poverty lines as well as for monthly and annual income were checked. Whereas the income concept does not have any significant effect, the reduction of poverty stability with a mean based poverty line is higher during the first years of a period and lower thereafter.

With the help of this concept the differences in the consequences of mobility on poverty status could be described for the USA and Germany. It becomes clear that the kind of income used is very important when comparing such different countries. Whereas for post-government income the poverty stability is much higher in the USA than in Germany, this changes dramatically when using pre-government income. For the market income almost no differences between the two states where found, which shows the strong impact of redistribution in the German welfare system. Only during the more recent years a slightly higher stability within Germany can be observed, which could be seen as an indication that long-time unemployment effects the stability of poverty.

Chapter 3

The Impact of Item-Nonresponse using a Multiple Imputation Strategy*

Most surveys suffer from missing data either as the result of questions that do not apply to the corresponding units or because information is not available although valid values do exist. This chapter does not deal with the former problem, i.e. missingness by design, but with compensation methods in the presence of the second type of missingness, often denoted as missingness by nonresponse. Traditionally, a distinction is made between unit nonresponse, where units (e.g. Households of Persons) who are selected into a sample are not observed at all, and item nonresponse where selected units are only partially observed, i.e. only a fraction of the asked questions are answered. The focus of the present chapter is on item nonresponse, although unit nonresponse can be considered as a special case of item nonresponse, and hence the same arguments apply.

Whether missing data are the result of a missing mechanism that is 'good natured', merely leading to smaller sample sizes without any bias[112], or whether analysis results are seriously biased if only the fully observed part of the sample is used depends, among other things, on the missing mechanism, the type of inference adopted and the variables used in the analysis. Clearly, if ignoring the missing mechanism leads to biased inference, some kind of compensation is necessary. However, even if not, the information loss can be severe if e.g. only the completely observed units are analyzed, or the analysis can be quite demanding if it is based on all of the observed data, including the missing values. Therefore,

*This chapter benefits largely from the work with my colleague Martin Spiess within the CHINTEX project. The sections about the concepts for handling missing data and the methodological framework for the evaluation are build up on this work. Earlier versions are published in Spiess and Goebel (2004, 2005).

[112] Nevertheless this does also lead to a loss in power and efficency because of the smaller sample sizes.

even in this case some kind of action to preserve the observed information and to simplify the intended analysis may be desirable.

One way to handle the problem of item nonresponse is to impute for each missing value one (single imputation) or more[113] (multiple imputation) 'plausible values'.[114] Single imputation methods seem to be the preferred method in the context of large public surveys, especially from a data producer point of view. However, like all methods intended to compensate for missing data, the generation of plausible values generally rests upon assumptions which are not testable given the observed data alone. But, if these assumptions are violated then the resulting estimators may be biased. If the assumptions apply and the estimators considered can be expected to be unbiased or consistent, another concern is about the estimation of further aspects of the (asymptotic) distribution of the estimators, e.g. the (asymptotic) variance.

The focus of the present chapter is to compare different strategies, namely multiple imputation methods, to handle the problem of missing items in large surveys for a complex variable, i.e. net disposable household income. More precisely, this chapter is organized as follows.

Section 3.1 briefly recapitulates the concepts of missing mechanisms and ignorability, also some general remarks on imputation methods are given. Section 3.2 gives an overview of the theory of multiple imputation based on Rubin (1987). In section 3.3 within the framework of the theory of multiple imputation, the methods used to compare and evaluate the different strategies to handle missing values are described. Further evaluation of imputation methods is based on the Finnish subsample of the European Community Household Panel (ECHP[115]) together with register information. Therefore, in section 3.4 a description of the data set is given, with main emphasis on the version used for the analysis. Section 3.5 describes the imputation procedures to be compared. The results of the empirical comparison will be presented in section 3.6, and the last section as usually (3.8) summarizes the results and offers some concluding remarks.

3.1 Concepts for handling missing data by means of imputation

Estimations in the presence of missing data are far from being a trivial task. The simplest strategy would be to discard all units with missing items and then use standard software to analyze the completely observed units. This 'case-wise

[113] Another imputation strategy which uses repeated imputation (more than one value) is the so called *fractional imputation*. This method replaces each missing observation with a set of imputed values and assigns a weight to each imputed value. In that case, the number of observations in the data file is augmented to include all the fractionally imputed observations. For a more detailed description see Kim and Fuller (1999), Opsomer and da Silva (2002, 2004), Kim and Fuller (2004) and Durrant (2005).

[114] Although a re-weighting of the data could also a possible strategy in principle, generally weights are used mostly to compensate for unit-nonresponse. The main reason lies in the high multidimensionality of the item-nonresponse problem.

[115] The ECHP is a representative panel survey on the situation of around 60.000 households in 14 European countries, which is co-ordinated by EUROSTAT. The ECHP provides up to eight waves of data (1994-2001); c.f. Clemenceau and Wirtz (2001) and Wirtz and Mejer (2002).

deletion' procedure was for a long time the standard procedure, and even today with more sophisticated methods and software being available this is the most common way for dealing with item-nonresponse. Intuitively it is clear that if the deleted units are systematically different from the completely observed units, then the results will probably lead to biased or, even worse, misleading conclusions.

A fundamental question therefore is, what kind of missing mechanism may have caused the observed pattern of missing and observed data. This pattern of missing and observed data itself can be considered as a realization of a random variable, the probability function of which is called the missing mechanism. According to Little and Rubin (2002) the missing data are called missing completely at random (MCAR) if the probability of the observed missing pattern ('missingness') does not depend on the variables whose values are observed or missing. The missing data are called missing at random (MAR) if missingness depends on the observed values but *not* on the variables *with* missing values. If missingness depends on missing values then the missing data are called not missing at random (NMAR). Precise definitions are given in Rubin (1976) and Rubin (1978); a clear presentation can be found in Little and Rubin (2002).

Whether or not a missing mechanism can be ignored depends not only on the missing data being MCAR, MAR or NMAR but also e.g. on the type of inference adopted or on the specific analysis to be conducted. A detailed treatment of this problem is beyond the scope of this section, but some discussion can be found in Diggle (1993, 1994), Heitjan (1994), Heitjan and Basu (1996), Kenward and Molenberghs (1998), Rubin (1976) and Shih (1992). Generally, if the data are MAR and the parameters to be estimated and those of the missing mechanism are distinct, in the sense that the joint parameter space is the product of the two individual parameter spaces, then the missing mechanism is ignorable for likelihood inferences (see Rubin, 1976). If, as is often the case, the parameters are estimated e.g. using the likelihood but are evaluated from a model based frequentist perspective ('model based frequentist inference'), not conditional on the observed pattern of missing and observed data, then, for large samples, the missing mechanism generally can still be ignored.[116] This latter, classical view is the one adopted throughout this chapter. It should be noted, that the distinctness condition is considered only of minor importance in practice, since inference, although not fully efficient, is still valid from a frequentist point of view if the parameters are not distinct (e.g. Little and Rubin, 2002).

If the decision is to compensate for missing data then, within the context of large surveys, ignorability of the missing mechanism is in general presupposed. With this assumption, explicitely specifying a missing mechanism can be avoided. However, the assumption of ignorability can not be tested without further information and hence is speculative in most cases. Therefore, Horovitz and Manski (1998) (see also Horovitz and Manski, 2000) proposed a method to derive a 'worst case' interval for the parameter of interest without any assumption on the missing mechanism. This method is applied to the ECHP by Nicoletti and Peracchi (2003) in the context of poverty measures. They suggest to base

[116] If the missing data are MAR and the parameters are a priori independent then the missing mechanism is ignorable for Bayesian inference. If the data are MCAR, then the missing mechanism is ignorable for general frequentist inference conditional on the pattern of observed and missing data (see Rubin, 1976, or Little and Rubin, 2002, and the literature given above).

conclusions concerning poverty probabilities on versions of the bounds taking sampling variation into account rather than on point estimates using imputed values. From an analyst's point of view this can be a useful strategy. Further, the bounds may be helpful in identifying suspect estimates based on imputed values, taking on values outside the interval.

However, from the data base constructor's view this may not be a useful way to proceed, since as a consequence, at least the most important estimators and corresponding bounds that may be used with the data set at hand had to be estimated and derived, respectively. Second, if the intervals for several estimators are derived, it is not clear how many violations of the bounds should lead to a rejection of the chosen compensation method.

Despite the above discussion,from a data base constructors point of view, there are good arguments in favor of imputed data sets based on the assumption of ignorability (see e.g. Rubin, 1996; Schafer, 1997).

1. The imputer may have more information available than the analyst that can be used to create the imputations, thus making the ignorability assumption more probable. For example data about the course of the field work or additional information about the interviewer.

2. For a given research question, not all missing data may be the result of the same missing mechanism. For example, some of the missing items may be the result of an ignorable mechanism and others of a non-ignorable mechanism. If the imputations are generated assuming ignorability, then at least part of the imputations are generated according to the correct assumption, and the resulting distortions in the inference can be expected to be smaller than using methods that rely on the MCAR assumption, e.g. analyzing only the completely observed units, or use less information than is available to the imputer.

3. Even if the missing mechanism is not ignorable, compensation methods based on the MAR assumption seem to be better than simple ad hoc methods that make no use of the information in the observed data (e.g. Schafer, 1997). On the other hand, modeling non-ignorable missing mechanisms, generally, rests on untestable assumptions and if misspecified may lead to additional bias.

4. If a user is suspicious about the imputed values then he or she may use them as a starting point to conduct a sensitivity analysis (see e.g. Little and Rubin, 2002; Scharfstein, Rotnitzky and Robins, 1999).

5. Finally,the imputed values may be ignored if an analyst is convinced that the assumption of ignorability is grossly misleading

Many imputation methods have been proposed to create rectangular data sets (see e.g. Little and Rubin, 2002). A basic requirement for an imputation method to be acceptable is that estimators which are unbiased or consistent, and (asymptotically) normally distributed, if there were no missing data are still unbiased or consistent, and (asymptotically) normally distributed, if they are based on observed and imputed values. If, for example, net household income rises over time and for each missing value in year t the observed net household income in $t-1$ is imputed ('last observation carried forward' method), then the mean household income in year t will be systematically underestimated.

Although unbiasedness or consistency generally is of major concern, with imputed data sets the ultimate user needs the necessary information to calculate valid variance estimates for the parameters. According to the different possible variance estimators different kind of information is necessary. Only for some approximate methods and a restricted set of estimators no further information is necessary beyond knowledge of the sample design. Generally, in addition to information about the sample design the analyst should at least know which value is imputed and which is observed ('imputation flag'). Then, for example, jackknife or bootstrap methods may be used (e.g. Fay, 1996; Rao, 1996; Shao, 2000; Yung and Rao, 2000). Other methods require even more information (e.g. Robins and Wang, 2000; Schafer and Schenker, 2000).

3.2 Multiple imputation

Another possible way to provide the necessary information to enable the ultimate user to compute valid variance estimates is by multiple imputation. The concept behind multiple imputation is outlined in figure 3.1. In multiple imputation, for each missing data entry, m imputations ($m >= 2$) are generated by simulated draws from the probability distribution representing the uncertainty of the missing value, given the observed data. The m resulting completed data sets are separately analyzed by an appropriate statistical method for complete data, and the m intermediate results are pooled into a final result by explicit procedures (see below).

This way of imputation ensures that the uncertainty about the missing data is reflected in the imputed data sets. The method of multiple imputation is derived under a Bayesian paradigm but is with mild conditions also valid within a design based frequentist framework for inference (Rubin, 1987, 1996) and should also be valid under a model based frequentist inference (e.g. Meng, 1994).

Basically, multiple imputations are draws from the joint posterior (predictive) distribution of the variables whose values are unobserved given the observed values of all other variables, including the response and sample selection indicators. If, however, the sampling and the response mechanism is ignorable, then conditioning on the sample selection and the response indicators is not necessary.

The variation of the multiple imputations should reflect the uncertainty in two ways, on the one side the fact that there are non-observed values, and on the other side the situation that the model to predict the unobserved values is not completely known. If only the parameters of the model are unknown, then within a Bayesian framework the values of the parameter can be drawn from a suitable prior distribution to generate imputations. If there is doubt about the model, then also the model itself can be subject to variation. If the imputations are drawn (or if they are approximate draws) from a Bayesian posterior distribution of the variables with missing values, then in large samples, the imputation method tends to be proper (under the posited response mechanism and given an

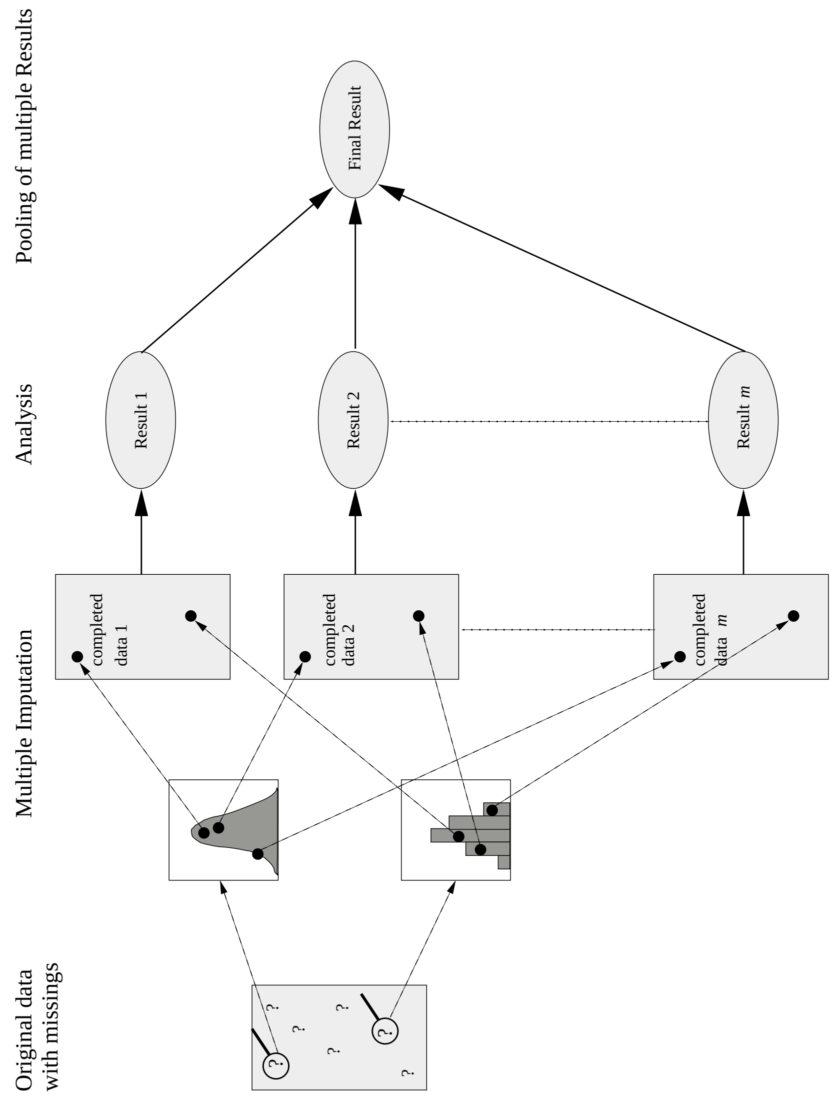


Figure 3.1: Schematic concept of multiple imputations

appropriate model for the data). For a clear treatment and precise definitions see Rubin (1987, chapter 4). Basically, if the inference is valid given the complete data, i.e. the data set without missing values, and the imputation method is proper in the sense of Rubin (1987), then the inference based on the observed and the multiply imputed data is valid under the assumed response mechanism.

If the data base producer and the analyst are distinct entities then, with large public data sets, it is likely that the model used for imputations is uncongenial to the models used by analysts. However, if more information is used to create the imputations than to analyze the data and this information is correct, the analysis based on the completed data set will tend to be more efficient than anticipated, leading to exaggerated confidence intervals and conservative inferences. According to Rubin (1996) these intervals are still smaller than those resulting from other methods which could be used in this situation. There are, of course, other sources of uncongeniality. For example, in contrast to the analyst the imputer may misspecify distributions or structural relationships. Up to now this topic has not received much attention, but results of some studies suggest that this type of uncongeniality usually leads to conservative inferences (e.g. Little and Rubin, 2002). But a few results also show that certain types of uncongeniality may also lead to anti-conservative inferences (e.g. Meng, 2002). Furthermore, Robins and Wang (2000) give an example where the variance estimator given in Rubin (1987) (see eq. 3.4) is – although with a smaller magnitude – downwardly biased even when the imputers model is correct. For several approaches and methods to create multiple imputations see e.g. Little and Rubin (2002) and Schafer (1997).

A general problem is that in data sets with many variables and complicated patterns of missing data a joint predictive distribution of the variables with missing values given the observed values is hard to specify. Therefore, simpler and less formally rigorous methods that approximate draws from this distribution have been proposed. One such method is implemented in the programs IVEware[117] and MICE[118]. An approach based on the assumption of (conditional) joint normality of the variables with missing values is provided by Schafer (1997).[119]

Once the imputations are generated several ($m = 1, \dots, M$) completed data sets are created. Each completed data set can then be analyzed with standard software routines used for data sets in the absence of missing data (Rubin, 1987). Let $\hat{\boldsymbol{\theta}}_m$ be a $(K \times 1)$-estimator and $\hat{\sigma}_{\hat{\boldsymbol{\theta}},m}$ the variance estimator based on the mth completed data set. Then the M estimation results can be combined as follows

[117] See Raghunathan, Lepkowski, Hoewyk and Solenberger (2001a) and for more details see Raghunathan, Lepkowski, Hoewyk and Solenberger (2001b) as well as the IVEware User Guide (Raghunathan, Solenberger and Hoewyk, 2002).

[118] See van Buuren and Oudshoorn (1999, 2000).

[119] A more detailed description of the software used to estimate the imputed values can be found in section 3.5.

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_M &= \frac{\sum_{m=1}^{M} \hat{\boldsymbol{\theta}}_m}{M} \quad , && (3.1)\\
\overline{W}_M &= \frac{\sum_{m=1}^{M} \hat{\sigma}_{\hat{\boldsymbol{\theta}},m}}{M} \quad , && (3.2)\\
\widehat{B}_M &= \frac{1}{(M-1)} \sum_{m=1}^{M} (\hat{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_M)(\hat{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_M)' \quad , && (3.3)\\
\hat{\sigma}_{\hat{\boldsymbol{\theta}}_M} &= \overline{W}_M + (1+M^{-1})\widehat{B}_M \quad , && (3.4)
\end{aligned}
$$

where $\overline{W}_M$ is called the within-imputation variance and $\widehat{B}_M$ the between-imputation variance. Imputation methods that are not proper in the sense of Rubin (1987) often underestimate $\widehat{B}_M$, leading to anti-conservative inferences in the following analyses of the imputed data.

For large samples and a finite number of imputations, inferences for scalar estimators can be based on $\hat{\theta}_M$, $\hat{\sigma}_{\hat{\theta}_M} = \bar{w}_M + (1+M^{-1})\widehat{b}_M$, where $\bar{w}_M$ and $\widehat{b}_M$ are scalar versions of $\overline{W}_M$ and $\widehat{B}_M$, and, quite generally, a t-reference distribution with

$$\nu = (M-1)(1+r_M^{-1})^2 = \frac{M-1}{\hat{\gamma}_M^2} \ , \tag{3.5}$$

degrees of freedom, where

$$r_M = (1+M^{-1})\frac{\hat{b}_M}{\bar{w}_M} \tag{3.6}$$

can be interpreted as an estimate of the relative increase in variance due to the missing data of $\hat{\theta}_M$, and

$$\hat{\gamma}_M = (1+M^{-1})\frac{\hat{b}_M}{\hat{\sigma}_{\hat{\theta}_M}} \tag{3.7}$$

can be interpreted as an estimate of the fraction of missing information about θ due to the missing data. Refinements of these statistics for smaller sample sizes are given e.g. in Little and Rubin (2002).[120]

Inferences about multicomponent parameters can be obtained in a similar way (e.g. Rubin, 1987). In particular, the test statistic

$$\tilde{d}_M = (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_M)' \, \overline{W}_M^{-1} \frac{\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_M}{K(1+r_{K,M})} \ , \tag{3.8}$$

where now

$$r_{K,M} = (1+M^{-1})\mathrm{tr}(\widehat{B}_M \bar{W}_M^{-1})/K \tag{3.9}$$

is an estimate of the average relative increase in variance due to the missing data across the components of $\hat{\boldsymbol{\theta}}_M$ and $\mathrm{tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. Equation (3.8)

[120] See also Barnard and Rubin, 1999.

can be used to test the hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, note that it is an estimated Wald statistic (Little and Rubin, 2002). Li, Raghunathan and Rubin (1999) propose $F_{K,\tilde{\nu}_M}$ as its reference distribution, where

$$\tilde{\nu}_M = \begin{cases} 4 + (K(M-1) - 4)\left[1 + \frac{1-\frac{2}{K(M-1)}}{r_{K,M}}\right]^2 & \text{if } K(M-1) > 4 \\ (M-1)\left(\frac{K+1}{2}\right)(1 + r_{K,M}^{-1})^2 & \text{else.} \end{cases}$$

Even with a few number of imputations and a high fraction of missing data this approach yields fairly good results (Li et al., 1999). Li, Raghunathan and Rubin showed with simulations that the resulting level of significance is nearly identical to the given level in most situations, especially the empirical relevant ones. The loss of testing power with finite M ($M \leq 10$) over the same approach with $M = \infty$ was very small in the majority of cases. This approach gets problematic if $M \leq 2$ or if K is small (Spiess, 2005, p. 310).

3.3 Methodological framework for the evaluation of imputation methods

In this section the methodological procedure for an empirical comparison of the different imputation methods and strategies for handling missing data will be described. Generally, with a given data set an emprical evaluation of the imputation method, e.g. hot-deck imputation, stochastic regression imputation or multiple imputation, is not straightforward. The problem is, that the generation of imputations rests on untestable assumptions about the missing mechanism and therefore, without further information, it is not possible to decide which method is 'better' in some sense. Without knowing the 'truth', one way to proceed is to assume a missing mechanism, conduct analyses and to evaluate the empirical results in the light of theoretical arguments, plausibility, comparisons with external information – e.g. results using similar data sets – assuming that the conditions needed for the analysis if there were no missing data to be valid are satisfied. If, however, for a main variable of interest register information is available that can be considered to be the 'truth' for all units in the sample, then further comparisons are possible.

The derivation of the results in section 3.2 is based on Bayesian theory, but inferences tend to be valid – under mild conditions – also for design and model based frequentist approaches (Meng, 1994; Rubin, 1987). In particular, for model based frequentist inference, the variance estimator in section 3.2 is based on a corresponding decomposition of the mean squared error of the parameter estimator using the multiply imputed data set.[121]

In the following let $\mathbf{x} = (x_1, \ldots, x_N)'$ be a variable vector of interest and $\mathbf{r} = (r_1, \ldots, r_N)'$ a vector of response indicator variables ($n = 1, \ldots, N$). If $r_n = 1$ then the corresponding element in $\mathbf{x}$ will be observed and if $r_n = 0$

[121] For the more general case, where the imputers and the analysts models are uncongenial, an extended definition of 'valid inference' has to be adopted (Meng, 1994; Rubin, 1996).

the corresponding element in $\mathbf{x}$ will be not observed. Let $\mathbf{x}_{obs}$ denote those variables whose values are observed and $\mathbf{x}_{mis}$ those variables whose values are not observed. For simplicity reasons, the mechanism used to select units into the sample is assumed to be ignorable and is therefore ignored in what follows.

For model based frequentist inference and given a weak condition concerning the relative efficiency of the estimators based on the complete and the completed data set, respectively, (for details, see Meng, 1994) which will be presupposed throughout, we have for a scalar estimate

$$\begin{aligned} \mathrm{MSE}(\hat{\theta}_\infty) &= E[(\hat{\theta}_\infty - \theta)^2] && (3.10) \\ &= E[(\hat{\theta}_c - \theta)^2] + E[(\hat{\theta}_\infty - \hat{\theta}_c)^2], && (3.11) \end{aligned}$$

where $\hat{\theta}_c$ is the complete data estimator, i.e. the estimator that would be used in the absence of missing data, $\hat{\theta}_\infty$ is the corresponding estimator based on an infinite number of imputations and the expectation is with respect to $\mathbf{x}$ and $\mathbf{r}$.

The term $E[(\hat{\theta}_c - \theta)^2]$ reflects the variance (mean squared error) of the estimator in the absence of missing data. On the other hand, the term $E[(\hat{\theta}_\infty - \hat{\theta}_c)^2]$ reflects the variation around $\hat{\theta}_c$ due to the missing data and the method used to generate the imputations, where it is assumed throughout that complete data inference using $\hat{\theta}_c$ leads to valid inferences. This latter term can be further decomposed into

$$\begin{aligned} E[(\hat{\theta}_\infty - \hat{\theta}_c)^2] &= E[E[(\hat{\theta}_\infty - \hat{\theta}_c)^2 | \, \mathbf{x}]] \\ &= E[E[(\hat{\theta}_\infty - E(\hat{\theta}_\infty | \, \mathbf{x}) + E(\hat{\theta}_\infty | \mathbf{x}) - \hat{\theta}_c)^2 | \, \mathbf{x}]] \\ &= E[\mathrm{var}(\hat{\theta}_\infty | \, \mathbf{x})] + E[(E(\hat{\theta}_\infty | \, \mathbf{x}) - \hat{\theta}_c)^2], && (3.12) \end{aligned}$$

where $E(\hat{\theta}_\infty | \, \mathbf{x})$ is the expectation of $\hat{\theta}_\infty$ over $\mathbf{r}$ for fixed $\mathbf{x}$. The first term, $E[\mathrm{var}(\hat{\theta}_\infty | \, \mathbf{x})]$, reflects the uncertainty with respect to the imputed data, whereas the second term $E[(E(\hat{\theta}_\infty | \, \mathbf{x}) - \hat{\theta}_c)^2]$ is the expectation of the squared bias of the estimator based on multiply imputed data sets with respect to the complete data estimator.

To see how the terms derived above can be used to compare aspects of multiple imputation methods given a data set were the missing values are nevertheless available, the concept of proper imputation methods (Rubin, 1987) will shortly be reviewed. Loosely speaking, if the complete data inference is valid in a frequentist sense and the multiple imputation method is proper then inferences based on multiply imputed data sets tend to be valid (Rubin, 1987, p. 116 ff.). Since the first condition does not depend on methods to compensate for missing data it is always regarded to be true. Basically, a multiple imputation method is proper, if for an infinite number of imputations the estimators $\hat{\theta}_\infty$, $\overline{w}_\infty$ and $\hat{b}_\infty$, i.e. the scalar versions of (3.1), (3.2) and (3.3) based on an infinite number of imputations, yield valid inferences in a frequentist sense for the complete data statistics $\hat{\theta}_c$ and $\hat{\sigma}(\hat{\theta}_c)$ under the posited response mechanism. An exact definition is given in (Rubin, 1987, p. 118 ff.), where three conditions are listed for a multiple imputation method to be proper.

Of particular interest are parts of the first and second condition (Rubin, 1987, p. 118 f.), namely that

$$
\begin{aligned}
\hat{\theta}_\infty | \mathbf{x} &\sim N(\hat{\theta}_c, \sigma(\hat{\theta}_\infty | \mathbf{x})) \quad \text{and} \\
E(\overline{w}_\infty | \mathbf{x}) &= \hat{\sigma}(\hat{\theta}_c) \, .
\end{aligned}
$$

The second part of the first condition implies that $E(\hat{b}_\infty | \mathbf{x}) = \sigma(\hat{\theta}_\infty | \mathbf{x})$ and the third condition states that $E(\hat{b}_\infty) = \sigma(\hat{\theta}_\infty)$. A necessary condition for a multiple imputation method to be proper is that $E(\hat{\theta}_\infty | \mathbf{x}) - \hat{\theta}_c = 0$, cf. equation (3.12). That is if $E(\hat{\theta}_\infty | \mathbf{x}) - \hat{\theta}_c \neq 0$, then the method used to generate the imputations is not proper.

Given that $E(\hat{\theta}_\infty | \mathbf{x}) - \hat{\theta}_c = 0$ holds for two different multiple imputation methods (and both are proper), the one leading to a smaller variance, $\sigma(\hat{\theta}_\infty | \mathbf{x})$, should be preferred. However, a comparison of two multiple imputation methods based only on estimates of $\hat{b}_{1,\infty}$ (method one) and $\hat{b}_{2,\infty}$ (method two), say, may be very misleading, since e.g. a smaller value for $\hat{b}_{1,\infty}$ could simply reflect that method one does not account for the extra variation due to the missing data. Indeed the underestimation of $\sigma(\hat{\theta}_\infty | \mathbf{x})$ is often the problem with improper multiple imputation methods.

Given a real data set with missing values, where – like in the present case – the missing data are available, there is only one realization of $\mathbf{x}$ and $\mathbf{r}$ and the number of imputations is finite. Therefore, although $E(\hat{\theta}_M | \mathbf{x}) - \hat{\theta}_c$ and $\hat{\sigma}(\hat{\theta}_c)$ are not available, they can be estimated by $\hat{\theta}_M - \hat{\theta}_c$ and $\overline{w}_M$, respectively.

For the data used in this chapter, and described in more detail in section 3.4, register information for the variable 'disposable net household income' from an external source is available, regardless whether the corresponding question is answered or not. This is also a variable of main interest in the ECHP and in the field of income analysis and welfare economics. Therefore, an estimator based on the income variable denoted as x_n is considered. Let M $(m = 1, \ldots, M)$ be the number of imputations, $x^*_{m,n}$ the mth imputation for the nth unit and $x_1, \ldots, x_{N_1}$ be the variables whose values are given in the questionnaire and $x_{N_1+1}, \ldots, x_N$ the variables whose values are not[122]. Then, in chapter 3.6 as an estimator the arithmetic mean $\hat{\theta}_c = N^{-1} \sum_{n=1}^{N} x_n$ will be considered. The bias of $\hat{\theta}_M | \mathbf{x}$ with respect to $\hat{\theta}_c$ can be estimated by

$$
\begin{aligned}
\widehat{\text{bias}} &= \Big(M^{-1} \sum_{m=1}^{M} N^{-1} \Big(\sum_{n=1}^{N_1} x_n + \sum_{n=N_1+1}^{N} x^*_{n,m} \Big) \Big) - N^{-1} \sum_{n=1}^{N} x_n \qquad (3.13) \\
&= (NM)^{-1} \sum_{n=N_1+1}^{N} \sum_{m=1}^{M} (x^*_{n,m} - x_n), \qquad (3.14)
\end{aligned}
$$

and is simply the difference between the arithmetic mean based on the imputations (see eq. 3.1) and the arithmetic mean using the complete sample. Ideally,

[122] Instead of the answers given in the questionnaire, the register information was used for the analysis, see section 3.4.

its expected value given x_n $(n = 1, \ldots, N)$ should be zero. An estimator of the variance of $\hat{\theta}_c$ is simply

$$\widehat{\text{var}}(\hat{\theta}_c) = \frac{1}{N(N-1)} \sum_{n=1}^{N} (x_n - \bar{x})^2, \tag{3.15}$$

where $\bar{x} = \sum_{n=1}^{N} x_n$. The corresponding variance estimator for $\hat{\theta}_M|\,\mathbf{x}$ can be calculated using eq. (3.2). In chapter 3.6 the bias (3.13) and the square root of the variance (3.15) together with the distribution of the differences $x^*_{n,m} - x_n$ will be compared for two different multiple imputation methods in a descriptive way.

3.4 Data

3.4.1 Sample structure

For the empirical analyses data from the year 1996 of the Finnish subsample of the ECHP are used.[123] This section describes the structure of the Finnish European Community Household Panel (ECHP), the sampling procedure and, in more detail, the specific data utilized. The ECHP started in 1994 and Finland joined the ECHP in 1996. The Finnish sample of the ECHP is based on the Income Distribution Survey (IDS). The IDS has a rotating half-sample design, where every incoming sample is drawn, using systematic sampling, from a master sample. This master sample is drawn from the population register that has been stratified by income and socio-economic status.

The first (preliminary) full sample consists of persons belonging (at the end of 1995) to the same dwelling unit as the sample person. However, during the interview the sample persons' household structure is checked. Some persons who are listed as sharing the same dwelling will not be considered part of the household, because they have moved, they were falsely registered there or they may be sub-tenants (Rendtel, Nordberg, Jäntti, Hanisch and Basic, 2004). For non-respondent sample persons or dwelling unit members, the household structure can naturally not be checked. Thus, the analysis relies on dwelling units rather than households.

The sample size for persons with an interview is 8173. The analysis is restricted to the missingness on disposable household income, which is one of the main income concepts used in the analysis of welfare, inequality as well as poverty.

3.4.2 Income variables

The income variable of most interest in this chapter is disposable net household income, which includes all cash income from labour and capital markets, private

[123] I would like to thank Statistics Finland for providing the data. I am also very grateful to Susanna Sandström and Marjo Pyy-Martikainen for their helpful advice using the Finnish data.

and public cash transfers, minus direct taxes. There are a two alternative sources for the income measure: either based on governmental information from income registers or based on asked questions from the interviews.

The special Finnish data set used in the empirical analysis is merged with various corresponding income data registers. The Finnish part of the income variables of the ECHP is nearly completely observed as the corresponding values are taken from register data. The interviewed persons are normally asked only very few questions about their income. However, in the years 1996 and 2000 the interviewees were asked the whole ECHP questionnaire including the section on income. This special situation has the great advantage that for these two years the 'true' income data are available from the register for every person in the sample *and* whether they refused to answer the corresponding questions or not. If not, then additional information is available through the amounts given. By nature, the income from register are complete without any missing items, but values for the income questions may be missing like in any other surveys.

For the evaluation of the imputation methods it is assumed that the register incomes are the true values and serve as benchmarks for the imputed incomes.[124] The conducted imputation task combines the register information together with the information whether the corresponding answer was given or not in the questionnaire. That is, for the evaluation of the imputation methods in later sections, a mixture of survey and register information is used. This strategy is chosen to isolate effects due to missing data problems from those due to error-in-variables problems at the stage of the comparison of imputation methods.

Throughout this chapter the disposable equivalent household income is calculated by using the so-called modified OECD scale, which assigns a weight of one to the first adult, a weight of 0.5 to each individual over the age of 14 and 0.3 to children who are less than 14 years old.[125]

3.5 Generating multiple imputations

3.5.1 General description of the imputation approach

Most evaluation studies of imputation methods are using simulations (e.g., Hoogland and Pannekoek, 2000; Bernaards, Farmer, Qi, Dulai, Ganz and Kahn, 2003), and these kind of simulations often incorporate a very simplified data structure.[126] Normally one variable of interest is missing (by varying the missing

[124] Note however, that household disposable register income mixes interview and register income since household membership is asked in interviews whereas dwelling unit register disposable income is purely based on registers. Differences between the two concepts may thus be due to differences in the two 'household' concepts. While is is customary to assume that register incomes are a more accurate measure of income than interview income, there is no reason to assume that households are more accurately defined in registers than in interviews (Rendtel et al., 2004).

[125] For more details on the discussion about equivalence scales see Atkinson et al. (2002); Burkhauser, Smeeding and Merz (1996); Coniffe (1992); De Vos and Zaidi (1997); Tsakloglou and Panopoulou (1998); Schwarze (2003). For robustness checks also the original equivalence scale was used and no substantial differences in the results were observed.

[126] A notable exception is the EU project EUREDIT (Mesa, Tsai and Chambers, 2000; EUREDIT-Project, 2000) and Wiggins, Ely and Lynch (2000)

mechanism) with all auxiliary variables completely observed. While such a setting is perfect to study the properties and robustness of the investigated imputation methods, it gives only minor informations about the usability in real large scale empirical surveys (like the *SOEP* or ECHP). But when using a real survey the researcher normally has no information about the 'true' values and an evaluation can therefore only consist of a comparison of descriptive statistics with and without imputation.[127] Fortunately the data used in this chapter has on the one hand 'true' values for the most incomes variables and for all persons who ever participated (gross longitudinal sample), and on the other hand realistic missing patterns resulting from nonresponse in a real interview situation.

The general goal of this imputation task is to provide multiply imputed data for the variable 'Monthly Net Disposable Household Income' (thereafter 'household income') as provided by the head of the household in the household questionnaire.[128] The resulting imputations will than be compared with the available 'true' values from the register. As auxiliary variables information about additional components of the household income (like social benefits or housing allowances as well as capital or rental income) and household characteristics are used (to be specified below). However, the household income consists of all income components of any individual currently living in the household (plus some income components at the household level). In the ECHP, fortunately every person aged 16 plus is interviewed and asked for any income received. Because of this fact it is possible to construct an 'artifical' calculated household income from the data provided by the personal interviews, which can be used as a proxy variable (naturally, apart from the income components solely received at the household level).

To calculate this 'special' household income valid values for each income component from each household member are needed.[129] As item-nonresponse also occur on this variables an imputation procedure is also required at the personal level. The auxiliary variables at this layer are personal characteristics and the 'original' household income (as obtained from the household level provided by the head of the household). After this imputation the constructed household income can be calculated without missing values. Therefore the household income is an 'independent' variable (to be imputed) at the household level *and* a dependent variable at the personal level (an auxiliary variable for the imputation of the personal incomes), for the personal income this is true the other way round. Because of this, the overall imputation procedure has to iterate over the household and the personal level for k-times (see figure 3.2), in the present application k has been set to 10.[130] Finally, after the 10th imputation cycle, the multiple imputed datasets at the household level can be estimated (without further changes to imputed values from the personal level). Note however, that even only one imputation is produced within each cycle, the iteration process *for this one* imputation can possibly use repeated imputation, depending on the algorithm the software uses.

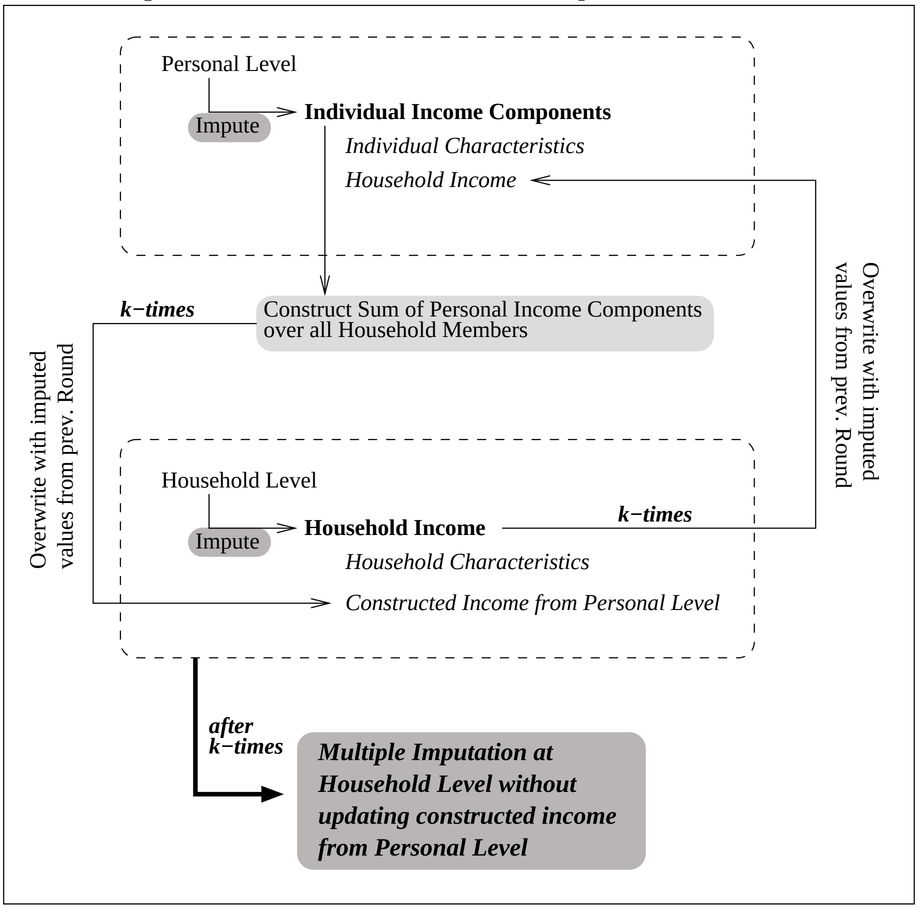
[127] See for example the approach by Frick and Grabka (2005).

[128] For a more detailed description of the dataset used, see section 3.4.

[129] At the personal level 30 different income components are available. For the sum of all components around 5% of cases were missing.

[130] Although 10 is arbitrar, the empirical imputation process has shown that after 5-7 rounds the imputation does not differ substantially anymore between the different rounds. To be on the safe side a bit higher value of ten was used.

Figure 3.2: General Overview of the Imputation Procedure



Variables which have to be imputed are in **regular bold**, auxiliary variables are in *italic*.

The imputation cycle starts at the personal level for 8172 cases and the model uses 50 variables. The fraction of missing is rather low and varies between zero and 1.4%. However the most important variables for personal income, net and gross income from wages and salary, hold the highest proportion of item-nonresponse (1.4% and 1.1% respectively). For the first imputation of the personal components the original household income was filled with the overall mean as starting values, to avoid different imputed values for persons in living in the same household. The imputation model at the household level was conducted for 4139 cases and consists of 15 variables.[131] The fraction of missing values per variable varies widely, from zero up to 23% just for the household income. A list of all variables in both analysis levels is provided in the appendix (see tables A.1 and A.2).

In the following sections a description of each imputation software used is presented. The different programs are divided in two categories, one for

[131] The low number of variables compared with the model at the personal level is due to the fact that less income components exist on the household level.

approaches which use a single method to impute all variables and another for approaches which use different methods according to the type of variable to impute. All imputation methods included are *stochastic* imputation methods, i.e. even for units with the same characteristics different values can be imputed, because of the inclusion of uncertainty. Imputation methods who always produce the same imputed values for observations with the same characteristics are called *deterministic* (Durrant, 2005).

Two rather prominent software packages are not included in this evaluation chapter, the `PROC MI` statement in `SAS` and the 'Missing Value Analysis' (MVA) by SPSS. The SPSS MVA software has been criticized in an article by von Hippel (2004). The MVA procedure provides two options, the first is a regression method that uses only the observed data in the imputations and the second is based on the normal distribution and resembles the first step of the NORM package. Currently none of the implemented methods produces proper imputations . `PROC MI` assumes a multivariate normal distribution and therefore that all variables are continuous. Because the logic is based on Schafers NORM package and IVEware is also available for `SAS` the inclusion of `PROC MI` was not seen as essential.

3.5.2 Single method imputation software

This section gives an overview of the imputation software utilized to create the later evaluated imputations. All imputation methods in this sections are based on one algorithm regardless of the underlying scale level of the variables under investigation.

3.5.2.1 Software for Hot-deck Imputation

Hot-deck imputation is the most commonly used method for item nonresponse. The main principle of the hot-deck method is using the very same data (donors) to provide imputed values for records with missing values.[132] The procedure to find the donor that matches the record with missing values is different according to the particular techniques used.

The matching process is carried out using the so called filter variables, records match if they have (nearly) the same values on the filter variables. Other hot-deck imputation methods include distance function matching or nearest neighbor imputation in which a nonrespondent is assigned the item value of the nearest neighbor.

For the actual imputation the hot-deck procedure within a Stata ado file is used (Mander and Clayton, 1999). A row of data with missing values in any of the variables is defined as a missing line of data, similarly a complete line is one where all the variables contain valid data. The hot-deck procedure replaces the variables in the missing lines with the corresponding values in the complete lines. The missing lines within each stratum of the data described by a list of variables are replaced by lines sampled from the complete lines in the same stratum. The approximate Bayesian bootstrap (ABB) method of Rubin and

[132] In contrast, cold-deck procedures are based on external information.

Table 3.1: Frequency distribution of the cluster variable used with hot-deck (Personal Level)

Value	Frequency	Percent	Cumulative
1	65	0.80	0.80
2	12	0.15	0.95
3	3,342	41.23	42.18
4	686	8.46	50.65
5	411	5.07	55.72
6	391	4.82	60.54
7	712	8.78	69.33
8	710	8.76	78.09
9	1,132	13.97	92.05
10	644	7.95	100.00
Total	8,105	100.00	

Schenker (1986) is used. First a bootstrap sample is drawn with replacement from the complete lines, and second the missing lines are sampled at random (again with replacement) from this bootstrap sample.

A major assumption with the hot-deck procedure is that the missing data are not only either missing completely at random (MCAR) or missing at random (MAR), but also that the probability that a line is missing varies only with respect to the categorical variables specified to identify each stratum. If a dataset contains many variables with missing values then it is possible that many data rows will contain at least one missing value. The hot-deck procedure will not work very well in such circumstances, as explicitly stated by the author of this software.

Table 3.2: Frequency distribution of the cluster variable used with hot-deck (Household Level)

Value	Frequency	Percent	Cumulative
1	452	10.95	10.95
2	405	9.81	20.76
3	396	9.59	30.35
4	403	9.76	40.11
5	401	9.71	49.82
6	402	9.74	59.55
7	409	9.91	69.46
8	399	9.66	79.12
9	410	9.93	89.05
10	452	10.95	100.00
Total	4,129	100.00	

The definition of the different strata should be incorporate as much information as possible, according to the imputation theory. However, when using to much variables the amount of different strata getting inflated and the case numbers in each 'cell' getting very low or become zero. To create comparable imputations, all auxiliary variables used in the other imputation methods (see below) should also used to define the strata for the hot-deck procedure. However doing this the cell size of the majority of the different strata are very small or even empty, because of the huge amount of explanatory variables. Therefore a cluster analysis with all these variables was accomplished beforehand and the resulting clustering was applied to define the strata for the hot-deck imputation procedure. An exemplary descriptive result for one imputation round is presented in the tables 3.1 and 3.2, which show the distribution of the resulting cluster variable used for the definition of the strata within the hot-deck imputation procedure.

3.5.2.2 Imputation software 'NORM'

The 'NORM' software is written by Schafer and utilize the method proposed by Schafer (1997). It draws in principle missing elements of a data matrix under the multivariate normal model and a user-supplied parameter using Maximum-Likelihood and an Expectation-Maximization procedure. The software is available as open-source for S-Plus/R and as a stand-alone program for MS Windows.[133]

The Expectation-Maximization (EM) procedure is a general technique for fitting models to incomplete data based on a process with two steps:

1. the Expectation (E) step in which missing sufficient statistics are replaced by their expected values given the observed data, using estimated values for the parameters; and
2. the Maximization (M) step where the parameters are updated by their maximum-likelihood estimates, given the sufficient statistics obtained from the E-step.

This iterative procedure runs until convergence is obtained. Upon convergence, NORM provided an iteration history, including the observed data likelihood, as well as the final estimates of the means, standard deviations and correlations.

Following the convergence of the EM procedure a Data Augmentation (DA) procedure is applied, for which the EM algorithm provides the starting values. DA is an iterative process which, utilizing the observed data, provides estimates of both the missing data and distributional parameters using a two-step iteration procedure:

1. the Imputation (I) step in which the missing data are imputed by drawing values from the conditional distribution, given the observed values and the parameters;
2. the Posterior (P) step in which new values for the parameters are imputed by drawing them from a Bayesian posterior distribution given the observed data and the most recent estimates for the missing data.

[133] For further details see Schafers web page at http://www.stat.psu.edu/~jls/misoftwa.html

A very handy feature is the possibility to check the convergences process with the help of implemented graphical tools. With the help of a auto correlation function (ACF) and the worst linear function the user can easily check if the parameters of the generated imputations are really uncorrelated over the iteration cycle as they should be.

3.5.2.3 Imputation software 'aregImpute'

The imputation procedure aregImpute is as a function part of the 'HMisc' library for S-PLus or R.[134] The imputation process is based on additive regressions, bootstrapping, and predictive mean matching. Two methods are used to fit the imputation models, 'Alternating conditional expectations' (ace) and 'Additivity and variance stabilisation' (avas). In principle, these methods simultaneously try to find transformations of the target variable and of all of the predictors, to get a good fit assuming additivity. 'ace' maximizes R-squared, and 'avas' attempts to maximize R-squared while stabilising the variance of residuals. And as the according help page states,

> "[...] 'aregImpute' takes all aspects of uncertainty in the imputations into account by using the bootstrap to approximate the process of drawing predicted values from a full Bayesian predictive distribution. Different bootstrap resamples are used for each of the multiple imputations, i.e., for the 'i'th imputation of a sometimes missing variable, 'i=1,2,... n.impute', a flexible additive model is fitted on a sample with replacement from the original data and this model is used to predict all of the original missing and non-missing values for the target variable."[135]

The function also provides the possibility to use predictive mean matching with weighted probability sampling of donors rather than using only the 'closest match'.

3.5.3 Multiple method imputation software

This section gives an overview of the multi-method imputation software utilized to create the imputations later evaluated. Below multi-method means that the software uses different models depending on the underlying scale level of the dependent variables under investigation.

3.5.3.1 Imputation software 'IVEware'

Basically, the generation of imputations within IVEware[136] (**I**mputation and **V**ariance **E**stimation Soft**ware**) is based on a method that can be considered

[134] The library Hmisc is provided as open source by Frank Harrell, for details see Alzola and Harrell (2004) and the associated web-page http://stat.cmu.edu/S/Harrell/Hmisc.html

[135] See the help page of aregImpute in R or S-Plus or the Hmisc Description in the Internet (`http://cran.r-project.org/src/contrib/Descriptions/Hmisc.html`).

[136] See Raghunathan et al. (2001a) and for more details see Raghunathan et al. (2001b) as well as the IVEware User Guide (Raghunathan et al., 2002).

as an approximation to the Gibbs-sampler[137] in that values for variables whose values are missing are drawn from distributions conditional on all previously imputed and observed values of variables but only some but not all parameter values at each step.

IVEware is available as `SAS` callable application, built on the `SAS` macro language and a set of independent `C` and `FORTRAN` routines, as well as a standalone program for Windows and Linux. To generate the imputations the program uses a multivariate sequential regression approach, in which linear, logistic, Poisson and polytomous regression models are fitted. The program is still a beta version and available at no costs[138].

More specifically, in a first step a conditional model for the variable with the lowest proportion of missing values given all completely observed variables is estimated. For example, for a metric response variable this amounts to estimating a simple linear regression model. Then, using the estimation results and non-informative priors, values for the parameters are drawn from their respective a-posteriori distribution. These are then used to draw values for the variables with missing data from their corresponding predictive distribution, however conditioning only on the value of the corresponding regression parameter and replacing the missing data for this variable. After the values for this variable are filled up with the imputations, it is – besides all variables which are completely observed – used as an explanatory variable in a regression from the variable with the lowest proportion of missing data from the remaining set of variables. This procedure is repeated until all missing values are replaced by imputed values. In the second step the procedure from the first step is repeated with the exception that in each regression all other variables are used as explanatory variables, irrespective of whether their values have been observed or imputed. The second step is repeated until stable imputed values occur or for a pre-specified number of rounds (Raghunathan et al., 2001a).

This approach is different from approaches where a joint distribution of the variables, whose values are missing given all the observed values and parameter values, is specified (e.g. Schafer, 1997, 1999, and section 3.5.2.2) in that only conditional distributions are specified and it is only assumed that a corresponding joint distribution exists. Therefore, it may be possible (as is the case with MICE) that sequences of draws of predicted values generated according to this method may not converge to a stationary distribution. According to Raghunathan et al. (2001a) using real datasets such anomalies may be rather the exception than the rule (c.f. van Buuren and Oudshoorn, 1999). However, to minimize the risk of sequences not converging to any genuine joint probability distribution, methods like the one described above may be used to generate only enough imputations to create a monotone missing data pattern[139] (Little and Rubin, 2002).

An advantage of IVEware is the very comfortable and easy way to produce multiple imputations. The user has to set up a file which includes all necessary

[137] Gibbs sampling is basically an algorithm to generate a sequence of samples from the joint probability distribution of more than one random variables, c.f. Casella and George (1992)

[138] The software can be downloaded from http://www.isr.umich.edu/src/smp/ive/, but only in binary format, i.e. the users can not modify or adapt the code according to their special needs.

[139] A monotone missing data pattern is given when the variables subject to missing data can be arranged in a way that variable q is observed for a particular case whenever variable $q+1$ is also observed for this case, see Little and Rubin (2002) or Durrant (2005).

information about the data and the imputation. For each variable used within the imputation task the variable type has to be specified in the 'setup file'. According to the type, a suitable regression model is used. This could be:

1. a generalized linear regression if the dependent variable is continuous.
2. a logistic regression if the dependent variable is binary.
3. a polynomial or generalized logit regression if the dependent variable is categorical with more than 2 outcome categories.
4. a Poisson loglinear regression if the dependent variable is a count variable.
5. a mixed regression model if the possible outcomes are zero or continuous. Imputation for this variable is done in two stages: first, use a logistic regression to impute '1' or '0', then, if '1' is imputed, the linear regression approach described above is used to impute the corresponding value and replace the value '1'.

For further details about the IVEware, see the references given in footnote 136. Apart from the covariates themselves, interaction terms, restrictions and bounds for each variable can be specified. A drawback in the empirical application is the little control the user has to determine the auxiliary variables for each model, only the number of variables used can be specified or an overall 'correlation level'. Also missing are graphical tools to screen the convergence process or the correlation between the iterations.

3.5.3.2 Imputation software 'MICE'

MICE owes its name to the method it implements, which is called **M**ultivariate **I**mputation by **C**hained **E**quations (MICE). It assumes that, for each incomplete variable, the user specifies a conditional distribution for the missing data given the other data. For example, a logistic regression could be used for incomplete binary variables, polytomous regression for categorical data, and linear regression for numerical data. Under the assumption that a multivariate distribution exists from which these conditional distributions can be derived, MICE constructs a Gibbs sampler from the specified conditionals. This sampler is used to generate multiple imputations (van Buuren and Oudshoorn, 1999).

MICE imputation algorithm: Let $X = (X_1, X_2, \ldots, X_k)$ be a set of k random variables, where each variable may be partially observed due to item-nonresponse, i.e. $X_j = (X_j^{obs}, X_j^{mis})$, with $j = 1, \ldots, k$. The imputation problem is to draw from the unconditional multivariate distribution of X, $P(X)$. Assuming that the missing mechanism is missing at random (MAR), one may repeat the following sequence of a Gibbs sampler iterations (t denote an iteration counter):

For X_1**:** draw imputations X_1^{t+1} from $P(X_1 | X_2^t, X_3^t \ldots X_k^t)$
For X_2**:** draw imputations X_2^{t+1} from $P(X_2 | X_1^t, X_3^t \ldots X_k^t)$
until
For X_k**:** draw imputations X_k^{t+1} from $P(X_k | X_1^t, X_2^t \ldots X_{k-1}^t)$

In other words, the imputations for every variable is each time conditioned on the most recently drawn values of all other variables.[140] van Buuren and Oudshoorn argue that "Rubin and Schafer (1990) show that if $P(X)$ is multivariate normal, then iterating linear regression models like $X_1 = X_2^t\beta_{12} + X_3^t\beta_{13} + \ldots + X_k^t\beta_{1k} + \epsilon_1$ with $\epsilon_1 \sim N(0, \sigma_1^2)$ will produce a random draw from the desired distribution."[141] Schafer (1997) generalized this approach to other multivariate distributions.

According to the documentation of MICE the implemented algorithm differs slightly from the proposed approach by Schafer. The advantage is that there is no need to choose a multivariate model for the entire data set. van Buuren and Oudshoorn (1999) assume "that a multivariate distribution exists, and that draws from it can be generated by iteratively sampling from the conditional distributions. In this way, the multivariate problem is split into a series of univariate problems." The approach is also known as regression switching or variable-by-variable imputation.

Although it is not always ensured whether the multivariate distribution actually exists. It may be possible that the specification of two conditional distributions $P(X_1|X_2)$ and $P(X_2|X_1)$ are incompatible, and therefore no joint distribution $P(X_1, X_2)$ exists. What follows is that there is no distribution for convergence and the algorithm will alternate between isolated conditional distributions. According to van Buuren and Oudshoorn (1999) this is probably more an exception than a rule, at least in the linear case.

For each incomplete variable, one can specify an elementary imputation method. This is the method that the Gibbs sampling algorithm uses for imputing the variable, for example linear or logistic regression. Several elementary imputation methods are available. For numeric data these are: Bayesian linear regression imputation with normal errors, improper linear regression with normal errors, predictive mean matching, and unconditional mean imputation. Logistic regression imputation is used for binary data, and polytomous logistic regression for categorical data with more than two categories. Also, a simple random sample can be taken as imputations. However, this is only useful if the data are supposed to be missing completely at random (MCAR). In empirical application this may be very seldom the case, but if for a subset of variables this could be assumed it can speed up the imputation process significantly.

In addition, users can write their own customized elementary imputation algorithms, and can call these from within the Gibss sampler. This allows for specialized imputation methods for specific variables, e.g. imputation under particular editing constraints.

There is often a need for transformed versions of the (imputed) data, e.g. a logarithm. In the case of incomplete data, one could 1) impute the original, and transform the completed original afterwards, or 2) transform the incomplete original and impute the transformed version. If, however, both the original and the transformed versions are needed within the imputation algorithm, none of these approaches work because one cannot be sure that the transformation is synchronized between the original and transformed versions.

[140] For details, see van Buuren and Oudshoorn (1999).
[141] van Buuren and Oudshoorn (1999, p. 9).

A special built-in elementary imputation method, called passive imputation, ensures the consistency among different transformations of the same imputed data. Passive imputation synchronizes the transformation with the most recently imputed original. The user can specify the transformation function. For example, the formula 'log(income)' searches for a variable called 'income', applies the specified function to the values whenever income is imputed, and stores the result. This mechanism provides a very user friendly way to use dummy variables within a imputation process. In the current implementation, passive imputation is linked to only one original, and therefore it is not possible to define a passive variable that depends on two or more columns of a dataset, for example, as the product of two variables.

In comparison to other software used in this chapter it is worth noticing that MICE is different, because of the very open and modular assembly. It is the only software which includes explicitly the option for the user to introduce his own imputation strategies or methods and therefore tailored to the special imputation task in question. MICE also allows to fully specify each model used within the imputation task and provides some graphical tools for the review of the iteration process. Although this graphical tools are very basic, the user has at least the possibility to implement further diagnostic methods because of the open nature of MICE.

3.5.3.3 Imputation software 'MVIS'

'MVIS' is not a different imputation software, but an implementation of MICE for Stata by Royston (Royston, 2004, 2005), also as open source.[142] However, there are some technical differences or limitations compared to MICE that legitimate the extra use of MVIS. Additionally to these technical differences there are also some distinctions with respect to the ease of use (MVIS is much more comfortable for the end user) and, because of the more widespread use of Stata especially in social sciences, may become more common. All technical differences are in principle restrictions in the flexibility of the original MICE software.[143] The user of MVIS can not implement additional imputation methods for specific variables. Also it is not possible to specify the covariates on a per variable basis, MVIS uses each time all other variables as covariates for the imputation of each variable. This will be the main reason for the large performance difference between MICE and MVIS (see table 3.3 on the following page). The following regression types are supported by MVIS: linear, logistic, multinomial logistic, and ordered logit regression. The final imputation can be done by predictive mean matching or draws from the posterior distribution (default).[144]

[142] For a description of MICE see section 3.5.3.2 on page 73.

[143] It should be noted that this is not mainly due to the porting by Royston, but of problems implementing this features into the very different concept of Stata compared to R or S.

[144] Shortly after the completion of this chapter Royston published a new release of MVIS under the name ice, which is capable of the most features MICE have, in particular the predictor definition for each variable and the passive imputation option for dummy variables and interaction terms.

Name	Program	Author	Computation Time
hot-deck	Stata	Mander & Clayton (1999)	30m
NORM	*R*	Schafer (1996)	20m
aregImpute	*R*	Harrell (1996)	1h10m
IVEware	standalone	Raghunathan, et al. (2001)	40m
MICE	*R*	van Buuren (2000)	1h20m
MVIS	Stata	Royston (2004)	11h30m

Table 3.3: Imputation Software Overview

3.6 Evaluation of imputation methods

In this section the effects of different imputation methods on the estimated bias and the estimated within variance (see chapter 3.3) are compared in order to evaluate the different imputation methods.[145]

The bias of $\hat{\theta}_M | \mathbf{x}$ with respect to $\hat{\theta}_c$, where $\hat{\theta}_M$ is the arithmetic mean of *income* over multiply imputed data sets and $\hat{\theta}_c$ is the complete data estimator, can be estimated by

$$\widehat{\text{bias}} = (NM)^{-1} \sum_{n=N_1+1}^{N} \sum_{m=1}^{M} (x^*_{n,m} - x_n)$$

(cf. (3.13), chapter 3.3). Obviously, $\widehat{\text{bias}}$ is a function of the differences between imputed and 'true' values.

For each imputation method used two figures will be presented. Each first figure displays the smoothed sample distributions of these differences, with Gaussian kernel and bandwidth 30. Each dashed line corresponds to one imputed data set, so that each graph displays 10 distributions of *individual differences* between the imputed and the true value. The straight line corresponds to the distribution of mean differences $M^{-1} \sum_{m=1}^{M} (x^*_{n,m} - x_n)$, i.e., the *individual* mean over the differences between the ten imputed values and the true value.

In the second figure different density functions of the household income distribution as such are shown. The solid black line represents all cases with register income (the benchmark distribution), the solid red line shows the income distribution only for those who answered the question about household income (complete case analysis) and the red dashed line are the cases who refused to answer the question about the disposable household income (in 'normal' data typically not available). The ten dashed grey lines are the distribution of the imputed incomes, each line represents one multiple imputation. These dashed grey lines should be identical to the red dashed line, if the imputed values are identical for all cases with the true values. And finally, the solid green line shows

[145] The evaluation of the imputations methods are only along the possibility to use the software with a real empirical example, in this case the Finnish subsample of the ECHP. All software programs included are capable to generate proper imputation in the sense of Rubin (1987). Therefore any shortfalls described in the following sections do not criticize the software as such but only the appropriateness of the specific method under these particular circumstances.

the income distribution density for all cases using the first imputation for the unobserved incomes (analysis of the imputed/completed data). In other words, the number of cases is identical for the solid black line and the solid green line as well as for the dashed grey lines and the dashed red line. The imputations are so much the better the more equal the dashed lines (red and grey), which includes that the solid black and solid red lines are also close together. Also note that in the second picture, in all imputation settings, the black line (all cases, true income) and the red lines (true values for the observed and unobserved cases) are identical, because these lines are not affected by the imputations. The income in all pictures illustrated is Finnmark divided by thousand.

3.6.1 Hot-Deck Imputation

The distributions of the differences calculated with the value imputed by the hot-deck procedure are shown in figure 3.3. The highest density (the peak of the distribution) is shifted to the left, which gives evidence that the imputation slightly underestimate the income values which have to be imputed. This is also confirmed by a bias around -8, the highest observed over all methods. However the mean over all imputation gives a slight improvement, in a sense that the density at the lower tail of the distribution shrinks most. In other words, the highly underestimated cases where not underestimate that much over all imputations. At this first look the imputations seem to be rather promising from figure 3.3.

However, looking at the second figure (3.4) it is obvious that none of the ten distributions of the values imputed by the hot-deck procedure (dashed light-gray

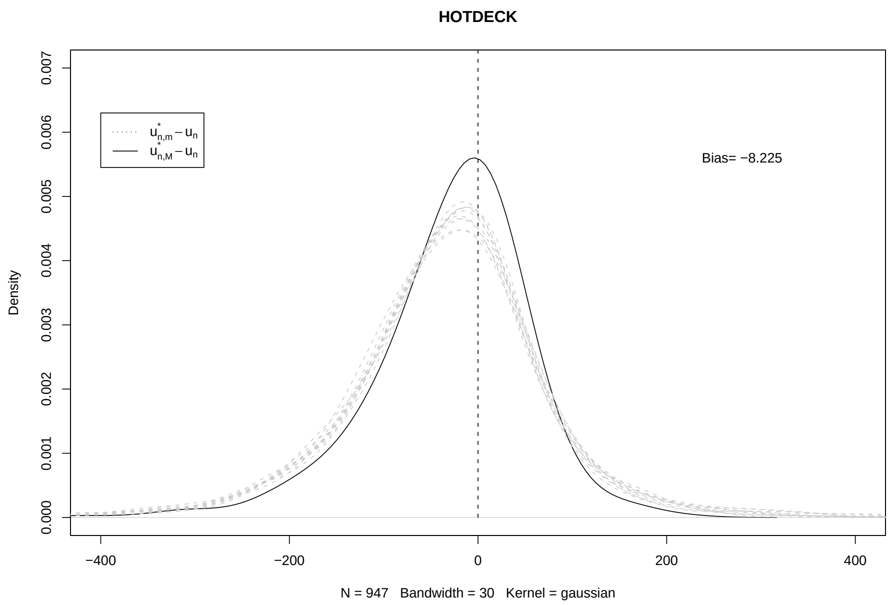


Figure 3.3: Differences between Imputations and Register (hot-deck)

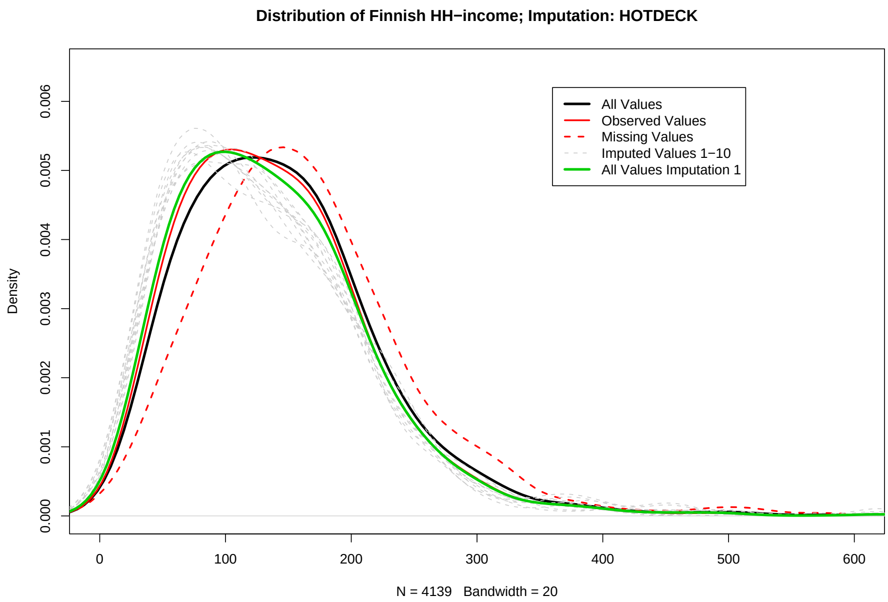


Figure 3.4: Income distribution density using hot-deck imputation

lines) can mimic the distribution of the missing values (dashed red line). The imputed values seem to stuck or even biased towards the 'heavy condensed' parts of the distribution of observed values (green line). This results in a situation where the income distribution of the completed dataset (including the imputed values) is even more biased than the distribution of the complete case analysis. Especially the cases at the lower part of the distribution are overestimated, because of the minor flexibility of the hot-deck approach incorporated in the software used. The missing at random assumption could only account in a restricted way. The cluster analysis only includes the completely observed cases and therefore may neglect some relevant relationships within the data.[146]

3.6.2 NORM Imputation

Figure 3.5 shows the distribution of the differences between the imputed value by the NORM software and the corresponding true values as reported by the official registers. The ten differences distributions seem relatively symmetric around zero and the estimated bias is rather low (-2.1). In comparison to the hot-deck imputation two things are worth mentioning. First, the imputations with the NORM software do not show the underestimation of income, because of the symmetry around zero, but the variance of the individual imputations are much higher (i.e. a higher within imputation variance), the peaks of the distribution are at a lower level. And secondly, the difference between the imputation specific distributions and the distribution of the individual means of all imputed values are more different than in the hot-deck case.

[146]For more details about the generation of the imputations see section 3.5.2.1 on page 68.

Looking at the second picture for the NORM imputations confirms this first impression. The distribution of the imputed values spread wide over the income range. Although the peak is nearly at the peak of the reference distribution (red dashed line), the tails of the distribution are overestimated. Especially the higher incomes are severely overestimated, whereas the lower incomes are only at the very end 'overstaffed'. For the resulting distribution of the completed data this lead to an underestimation of the 'middle class' and a clear overestimation of the upper income regions. Similarly to the hot-deck imputation the distribution of the completed data is even more biased than the complete case analysis. But in contrast to the hot-deck procedure, the NORM imputation shifts the distribution more to the right. As already noted the NORM algorithm assumes a multivariate normal distribution of the underlying posterior distribution and for complex surveys, the violation of this assumption is likely.

3.6.3 aregImpute Imputation

Also the observed bias (-4.2) of the created imputations with the aregImpute function is higher than with the NORM imputation, it is obvious that the within variance is much lower (cf. figure 3.7). However, it is noticeable that the distribution of 'mean' imputation is slightly shifted to the right, even that the 'single' imputations are shifted in the opposite direction. But in comparison to the former imputation strategies the cases are much more concentrated around zero difference. Overall we therefore expect for the distribution of the completed data a smaller bias as compared to the complete case analysis, although the imputed values seem to underestimate the amount of income.

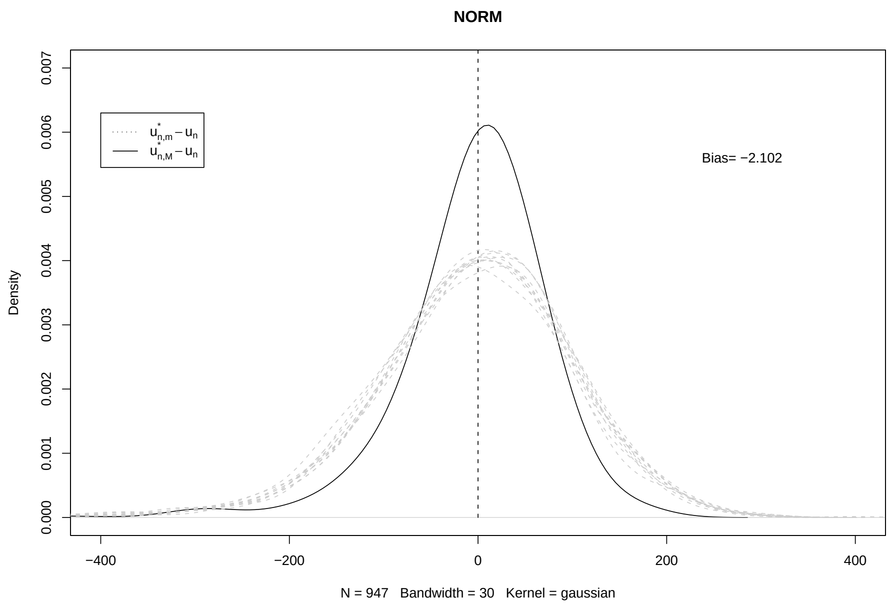


Figure 3.5: Differences between Imputations and Register (NORM)

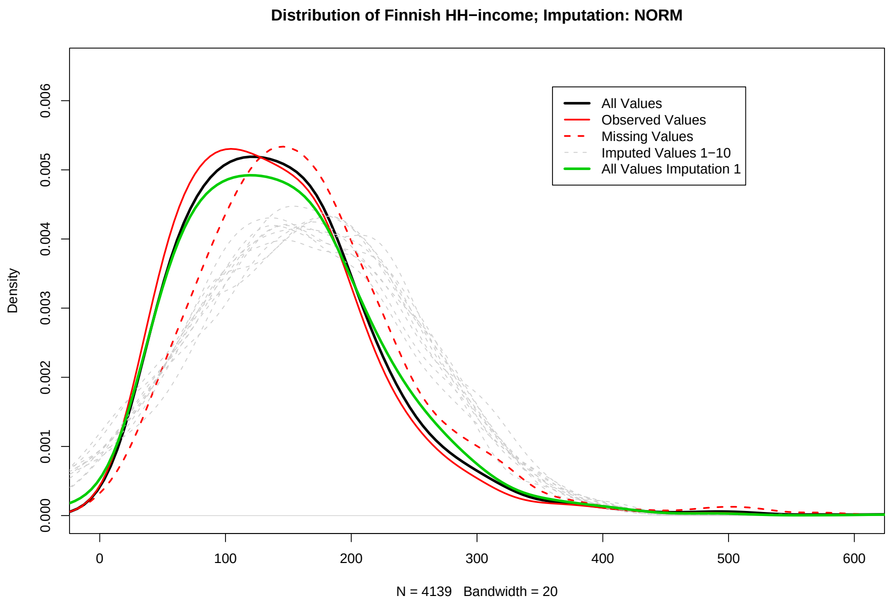


Figure 3.6: Income distribution density using NORM imputation

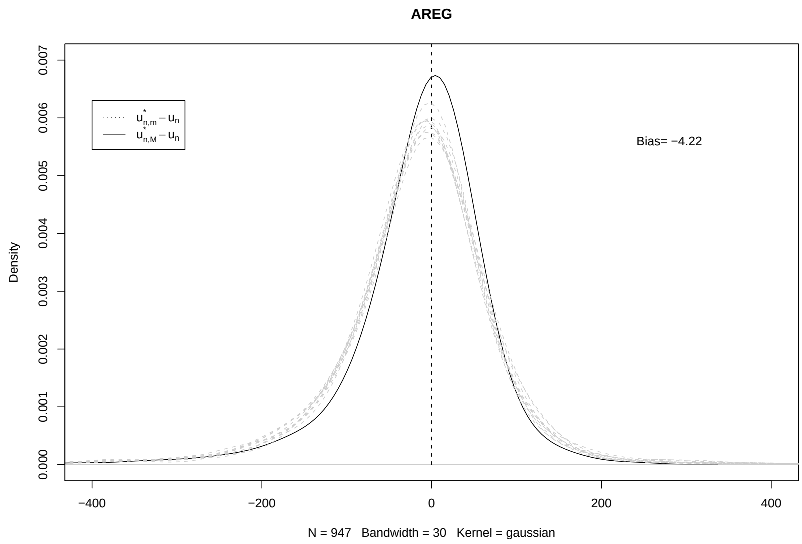


Figure 3.7: Differences between Imputations and Register (aregImpute)

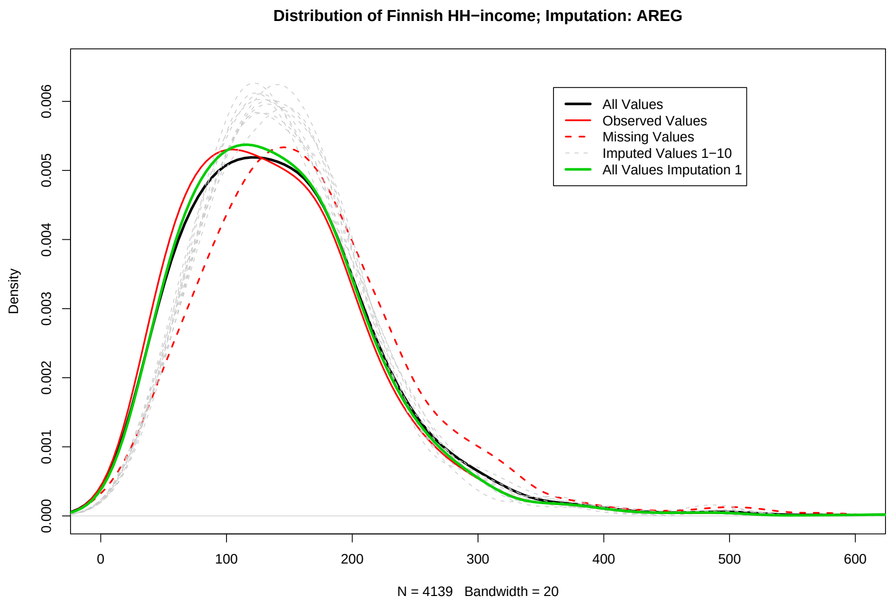


Figure 3.8: Income distribution density using aregImpute Imputation

The impact on the overall distribution of incomes for the aregImpute imputations is shown in figure 3.8. For the distribution of the imputed values only there is a refinement determinable, at least in comparison to the above described imputations with a hot-deck procedure and with the NORM software. Nevertheless, the distribution of the imputed values are shifted to the left, in particular the middle incomes are overestimated. The distribution of the completed data arising from these imputations are almost over the whole range of incomes closer to the true distribution than the resulting distribution of a complete case analysis would be. Only a very narrow range of incomes around the peak of the true distribution are more biased (overestimated) in comparison to the complete case analysis. All in all the aregImpute function has done a fairly good job in restoring the original distribution of the 'true' incomes.

3.6.4 MICE Imputation

The results for the imputations done with the MICE software are presented in the figures 3.9 (the distribution of the differences) and 3.10 (the distribution of the imputed incomes). As is the case for all previous imputations we find a negative bias, for the MICE imputations it is approximately -3.2. The distributions for each multiple imputation look very similar to the graphs presented for the aregImpute function (c.f. figure 3.7), however there are more positive differences, as the smaller bias accordingly shows. The distribution of the individual means over all differences shows a desirable sharp peak very close to zero. In total picture 3.9 is very promising for the resulting distribution of incomes, presented in figure 3.10.

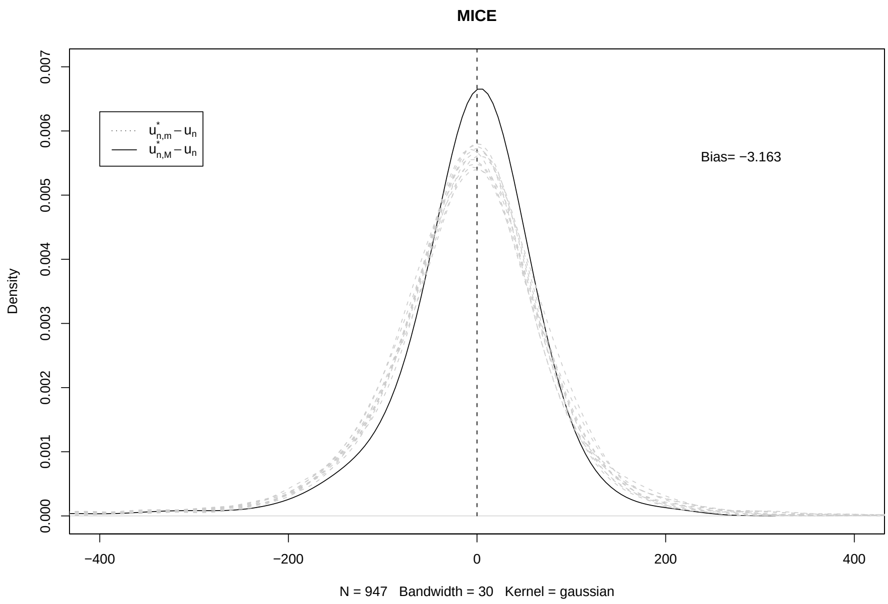


Figure 3.9: Differences between Imputations and Register (MICE)

The distribution of the imputed incomes by MICE are very close to the distribution of the true incomes of the non-respondents. There is only a minor distortion with respect to this reference distribution, with a marginal underestimation (overestimation) of higher (lower) incomes. Although the peak of the reference distribution is nearly reached, due to this mild deformation the imputations are a bit shifted to the left. However, when comparing the estimated curves for the complete case analysis and the completed data by MICE it is evident that the latter is much closer to the distribution of the true incomes of the whole sample. The MICE algorithms really outperform the three previous described imputation strategies in this application with a large scale survey with complex design.

3.6.5 MVIS Imputation

Figures 3.11 and 3.12 display the resulting distributions of income using the MVIS ado files. MVIS is an implementation of the MICE software (results are described in the previous section) and therefore uses the same basic idea of regression switching.[147] The differences between the two software programs concern only the flexibility[148] of the imputation design the user can choose, therefore we expect that the resulting imputations with MVIS are very similar to that of MICE, which performed very well. As this is true for the distribution of differences (figure 3.11), where the shape of the distribution is very similar to

[147] For a more detailed description see sections 3.5.3.2 and 3.5.3.3.

[148] The other side of the coin (at least in this case) is that with less flexibility the user friendliness rises and the software do not ask too much of the user.

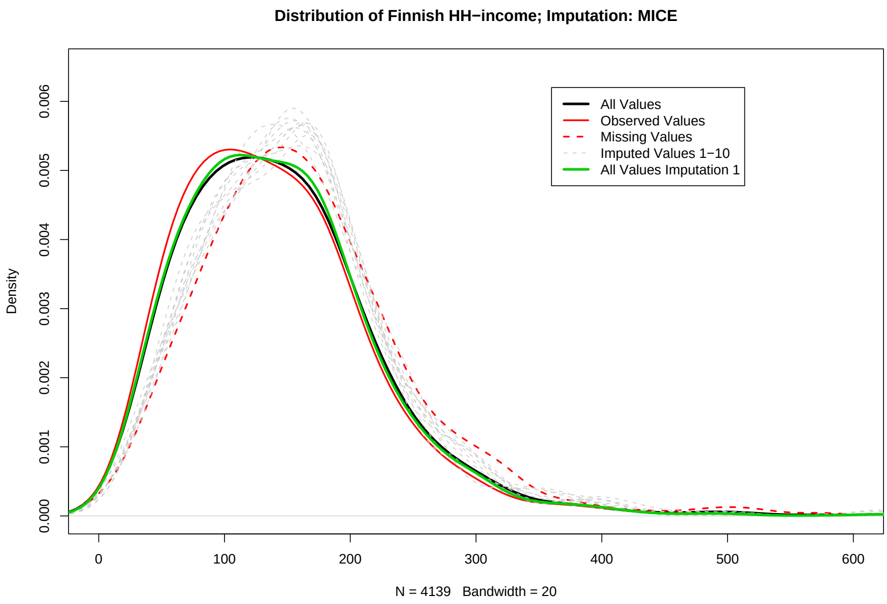


Figure 3.10: Income distribution density using MICE Imputation

that of MICE and the bias is only marginal higher (-3.4 and -3.2 respectively), we can see more differences in the distributions of the actual incomes (figure 3.12).

The distribution of the imputed values are rather close to the reference values, but only when comparing MVIS with one of the single method imputation strategies, like the hot-deck or the NORM procedure. Using the original (MICE) as a yardstick, than we see a higher overestimation of cases with low incomes. However, over the remainder of the income range the distributions of MICE and MVIS are very similar. As a result the income distribution of the completed data is not as close to the 'true' distribution than was the case in the MICE imputation. Nevertheless, the result of the imputation with MVIS is an improvement to the complete case analysis, when deleting all cases with item-nonresponse on the income variable.

3.6.6 IVEware Imputation

The results according to the imputations of IVEware, which is also a multi method approach based on sets of regressions, are presented in figures 3.13 and 3.14. The first obvious thing to mention is the fact that the IVEware performed best with respect to the overall bias measure and achieved the value nearest to one, namely -1.1. But it is also obvious that the variance of each imputation is higher than with the other regression based methods. The peak of the distributions are also a bit shifted to the right, which means a slight overestimation of incomes. However, in spite of the negative bias, this suggests the conclusion that the distribution of the differences are not symmetric.

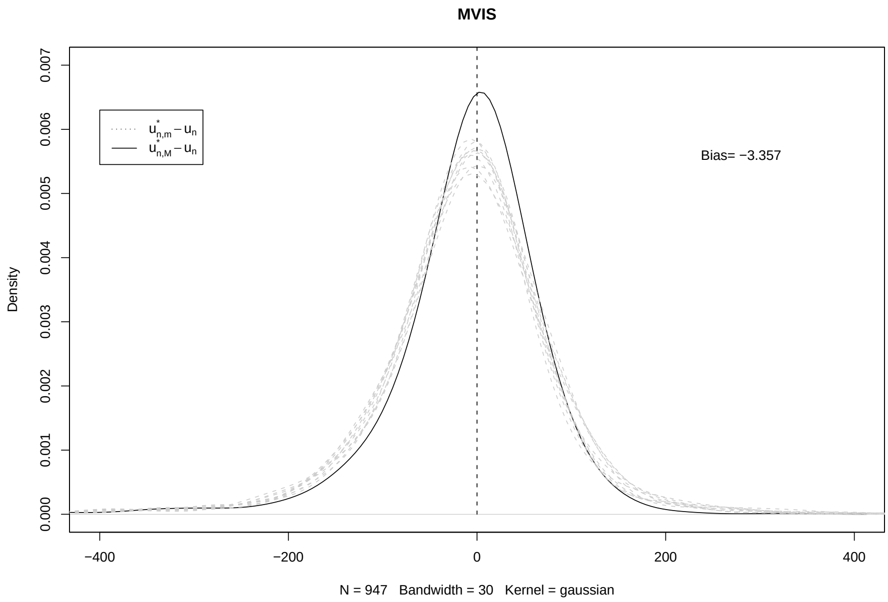


Figure 3.11: Differences between Imputations and Register (mvis)

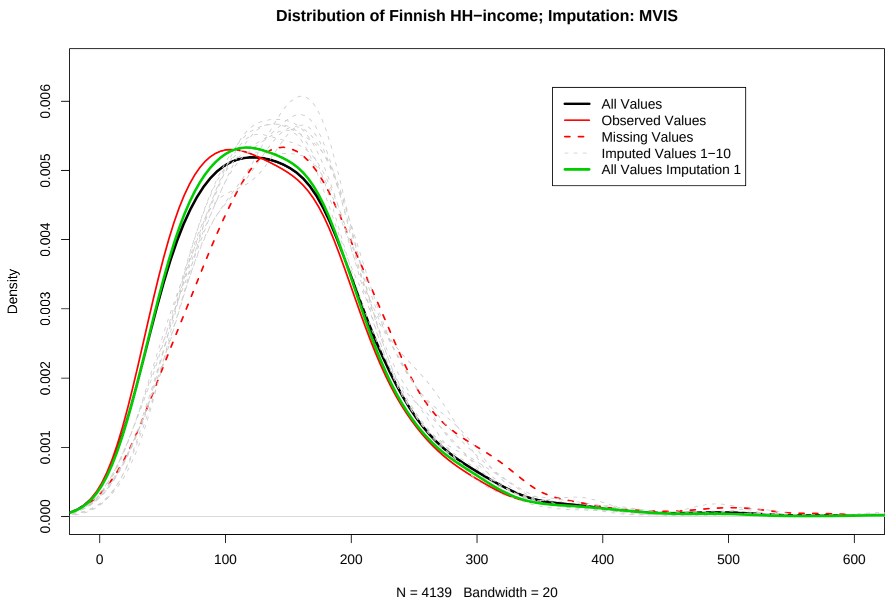


Figure 3.12: Income distribution density using mvis imputation

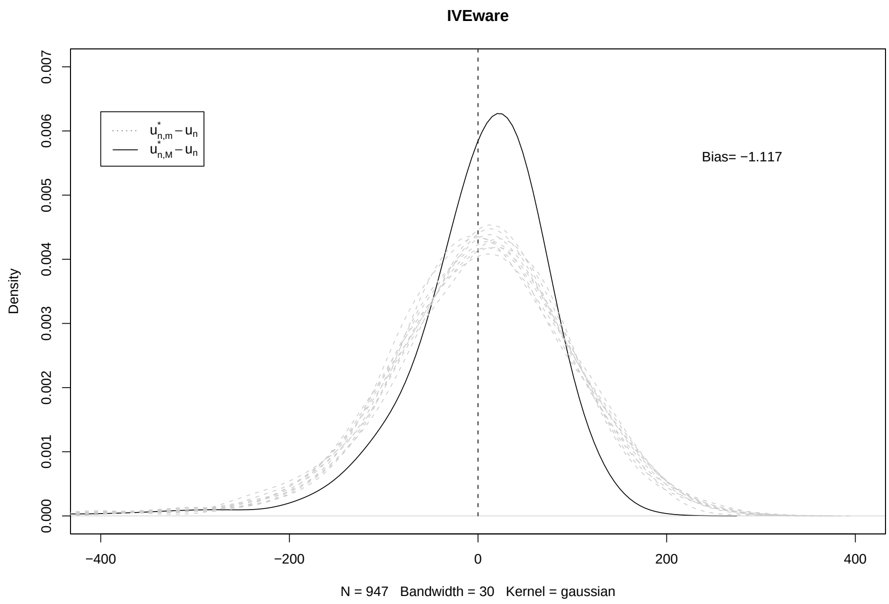


Figure 3.13: Differences between Imputations and Register (IVEware)

The distribution of the imputed incomes with IVEware do not stick to the distribution of the observed cases. Rather the other way round, the imputed cases do severely overestimate the number of high income cases. The resulting distribution of the completed data therefore also overstates the high income ranges. Even the lower tail of the distribution is better represented with the IVEware imputation, in comparison to a complete case analysis, the middle and high income range seems to be severely distorted.

3.7 Impact on the estimation of some important parameters in income analysis

In the previous section the impact of each imputation strategy on the distribution of income was discussed. However, in most empirical analysis only some parameters are of interest and not the whole distribution. Because the set of possible parameters available to analysts is near infinity, only the more important ones with respect to the main focus of the thesis were selected. The parameter chosen are some standards as the arithmetic mean, median, and standard deviation. With respect to the measurement of inequality the Gini, Theil (zero and one), and Atkinson[149] (with parameter ϵ from 0 to 5) index were calculated. For the measurement of poverty the measure by Foster et al. (1984) – FGT – is used with the parameter α running from 0 to 3. With an $\alpha = 0$ this corresponds to

[149] Often used values for ϵ are 0.5 and 2. For more details on the measure, see footnote 78 on page 29.

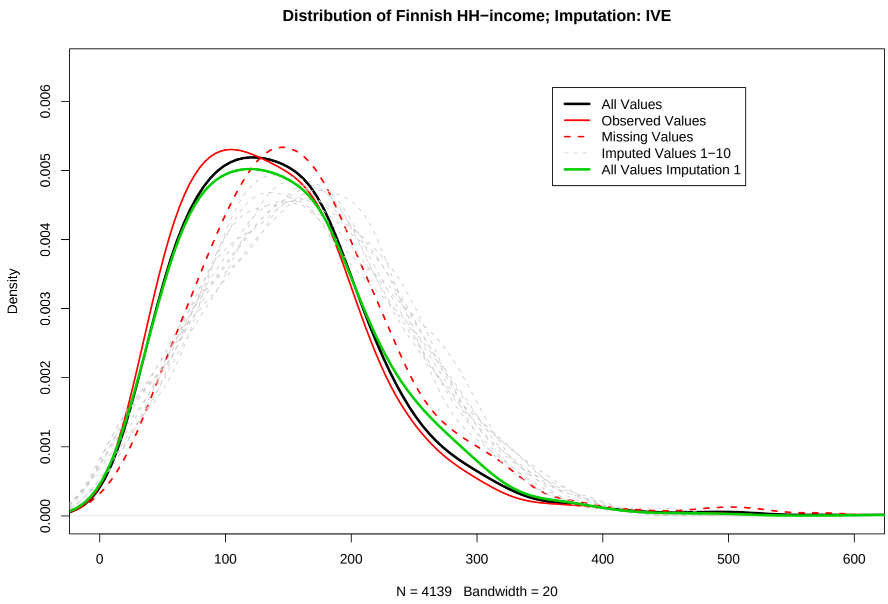


Figure 3.14: Income distribution density using IVEware imputation

the head count ratio and with $\alpha = 1$ to the poverty gap.[150]

Although this kind of analysis is normally done at the personal level, here it is performed at the household level. Because we are only interested in the differences between the true, the complete case, and the imputed case analysis and not in the estimates as such this is negligible. However, the analyzed household income is corrected for differences in the household composition by dividing with the square root of the household size.

The following figures (3.15 to 3.17) show the value of each estimate for the fully observed register case as a straight black line, and for the complete cases analysis (all cases with item-nonresponse on household income are deleted) as a straight dashed line. Each imputation strategy is represented in figure 3.15 by three signs, a circle for the resulting estimate when applying the combination rule[151] by Rubin as well as a plus sign for the minimum and the maximum of the calculated estimate over all multiple imputations. In figures 3.16 (Atkinson inequality measure) and 3.17 (FGT poverty measure) the combined estimate result is presented by a straight red line, because these measures are calculated for a set of different parameter values. Accordingly the minimum and maximum values are displayed as dashed red lines. To clarify the amount of the bias additional bar charts are plotted in the background of each figure. The light grey bars show the differences between the complete case analysis and the resulting estimate with all true cases in percent of the 'true estimate'. The other bars show the percentage bias if the imputed distribution is used, blue bars denote a positive

[150] A more detailed description of the FGT measure is given in section 1.3.4 and the formal definition in equation (1.37).

[151] For the details see equation (3.1).

bias and a negative bias is displayed by red bars. For all presented measures a simple complete case analysis would noticeably underestimate the true sample estimate, so a imputation strategy would be preferable if this underestimation can be reduced.

When choosing the mean or the median as a location parameter of interest, only minor differences between the different imputation strategies can be seen. In particular all imputations, with one exception, shift the mean from the complete case analysis (the dashed black line) more towards the true reference case (the solid black line). The only imputation method where this is not true is the hot-deck imputation, even for such basic parameters. Both values decrease and therefore fade away from the ‘true case’. The best result according to mean and median achieve IVEware, which nearly hits the mark. However, the range of the estimated means and medians over the ten multiple imputations for each of the six imputation methods are rather small.

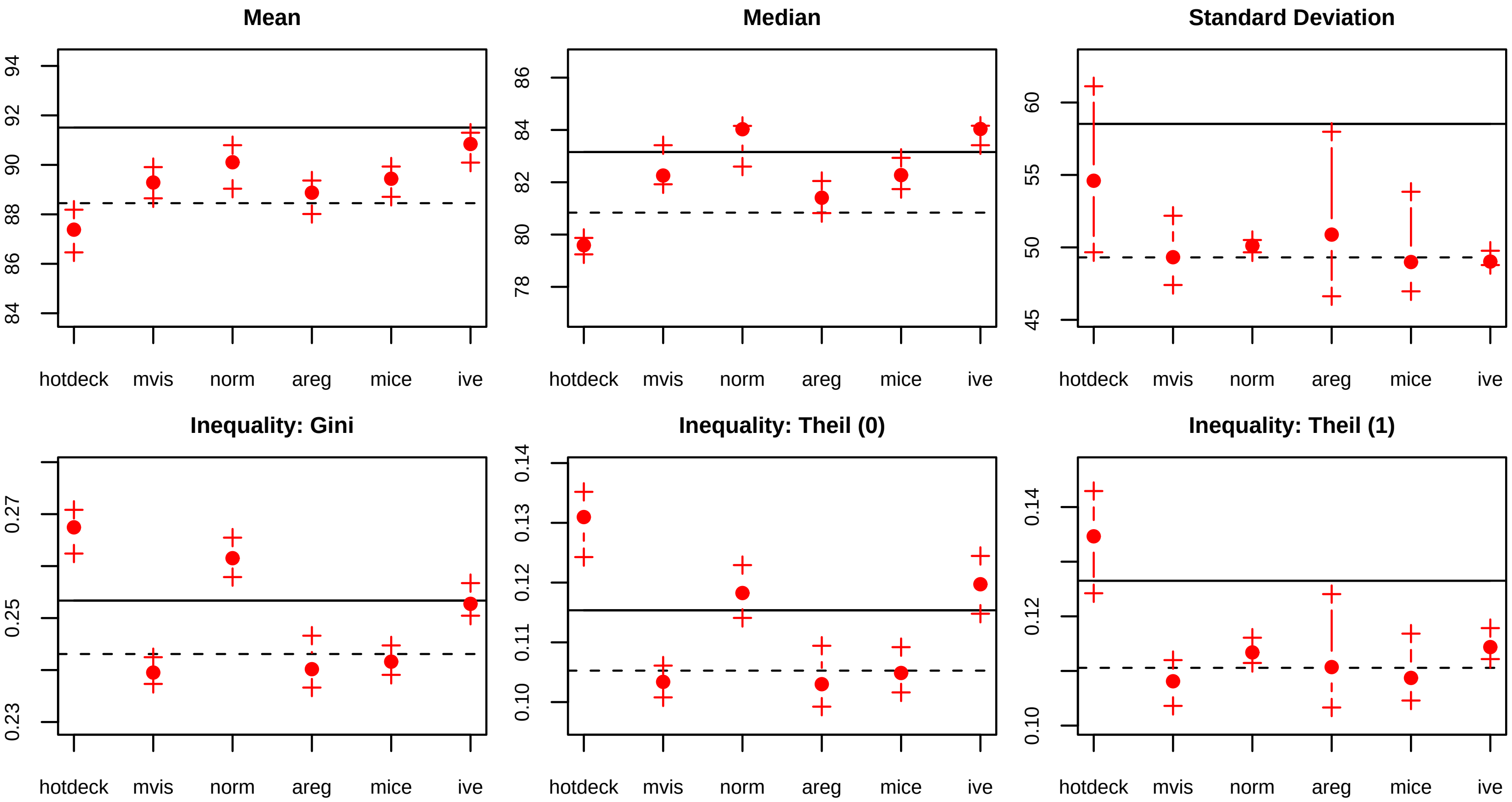


Figure 3.15: Impact of different imputation methods on mean, median, standard deviation, Gini and Theil

This range is much wider compared to the results for the standard deviation, in the upper right corner of figure 3.15, except for NORM and IVEware. Interestingly is the only method who shifts the standard deviation away from the complete case analysis, into the direction of the true case, the hot-deck imputation. All other imputation methods under consideration are 'glued' to the complete case analysis.

The picture for the Gini inequality index differs slightly. Compared with the range of the estimated means for each imputation strategy, the variation between the multiple imputed datasets increases for the hot-deck and aregImpute imputation method, whereas for the other four the variation seems stable. As was the case in the mean comparison, the worst performance can be seen again for the hot-deck imputation, but now with an increase in inequality. Almost identical results with the complete case analysis will be achieved with MICE, MVIS and aregImpute. Only NORM and especially IVEware operate in a corrective way, in which IVEware again nearly hits the line of the reference case. The same holds for the Theil(0) measure, only with the Theil(1) index some differences can be observed. The best performing imputation method would be the hot-deck imputation and all others almost identical to the complete case analysis. Surprisingly is the fact, that NORM and IVEware behave nearly identical with respect to these inequality measures. Very good performance with respect to the Gini and Theil(0) measure and only marginal differences with respect to the complete case analysis when using the Theil(1) index. If an applied researcher had only these results, IVEware or NORM would be the best choices. And the 'runner-up' would be MICE or MVIS.

But looking at the results with the Atkinson index and poverty related indicators gives a completely different picture. Figure 3.16 displays the results for the Atkinson index with the parameter ϵ vary from 0 to 5. Obviously there are two different "groups" of imputation methods with respect to the impact on the Atkinson measure. The one group is robust to the specification of ϵ and consists of hot-deck, MVIS, areg and MICE. The other group is very sensitive to medium values of the parameter ϵ and include NORM and IVEware.

The complete case analysis would underestimate inequality around 5 and 15%, as one can see by the grey bars. Nearly identical results to the complete case analysis will be obtained when using the MVIS or areg imputation strategy. The resulting bias by the imputed datasets with respect to the Atkinson index are very large for the NORM and IVEware software. This is especially the case for values of ϵ between 1.5 and 3.5, which includes the often used value of 2 in empirical inequality research. This bias are up to 80% for the NORM imputation, and up to nearly 100% for the IVEware imputation.

The best and less biased results will be obtained with either the hot-deck or the MICE imputation, which are both capable to reduce the bias of the complete case analysis, at least over parts of the parameter range under investigation (this is also true for the MVIS imputation, but only to a very small extend). However, the hot-deck imputation introduce a greater bias just around the 'popular' value of 2. Only the MICE imputation gives partly a bias reduction, which is not combined with the introduction of an additional bias when the whole range of the parameter is considered. However this bias reduction is rather low and only around parameter values around 3, which are not often used in empirical studies.

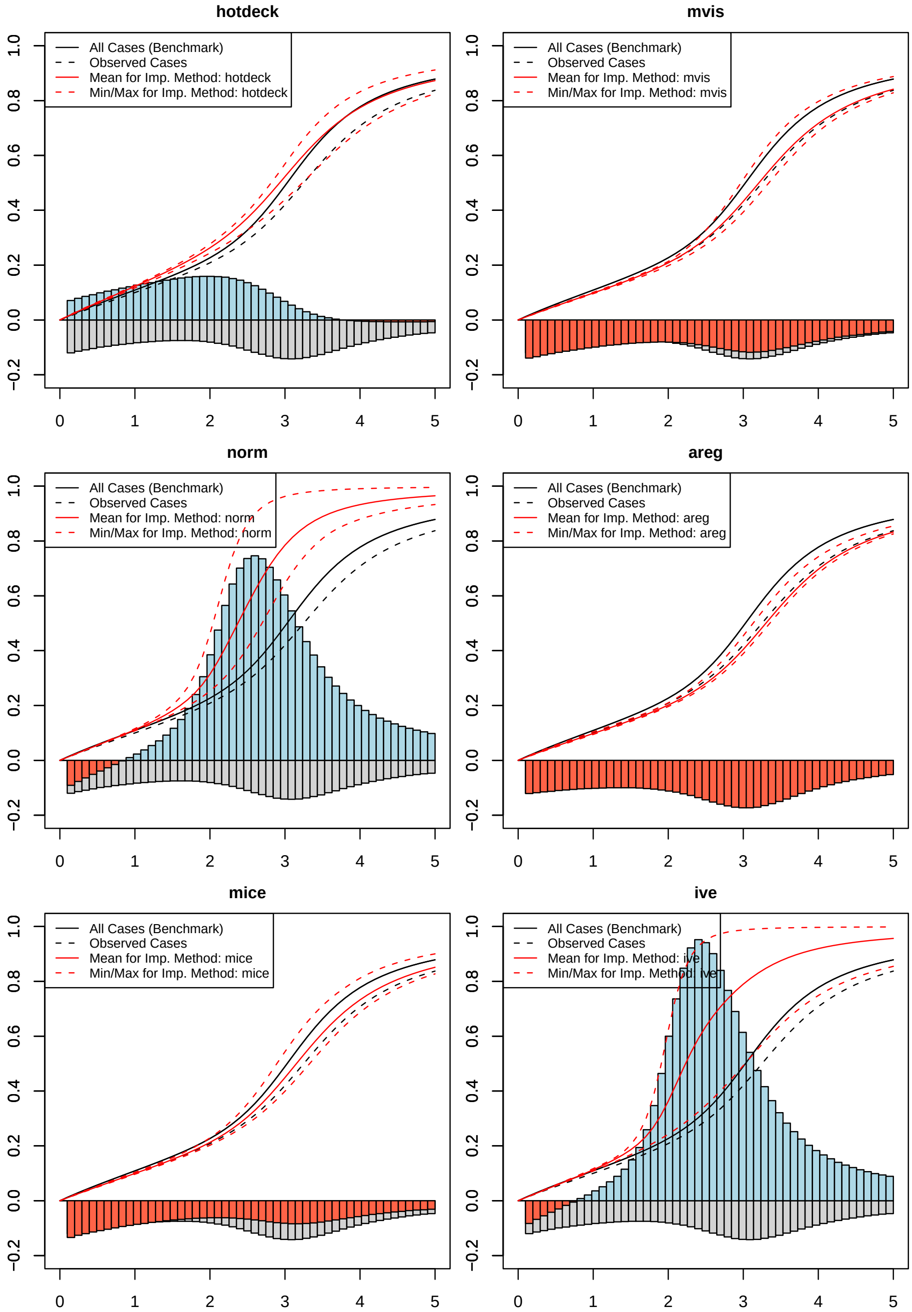


Figure 3.16: Impact of different imputation methods on the Atkinson inequality measure for different values of the parameter ϵ

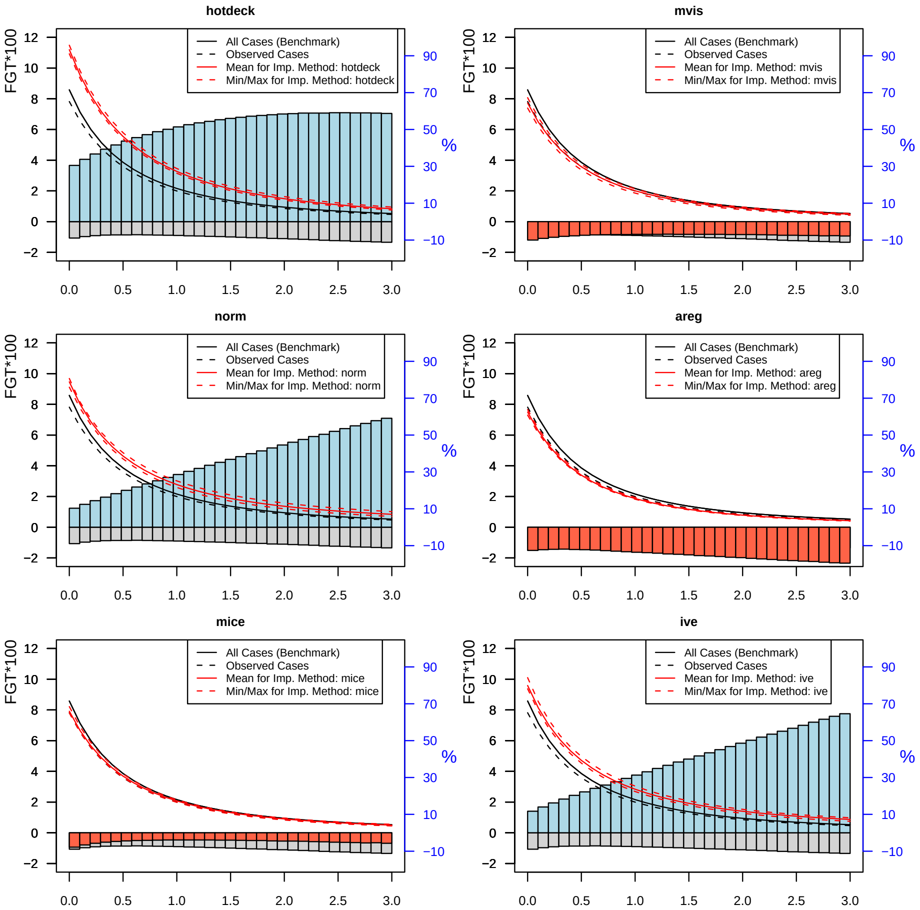


Figure 3.17: Impact of different imputation methods on the FGT poverty measure, for different values of the parameter α

Figure 3.17 presents the same graphs and bars for the FGT poverty measure (Foster et al., 1984) with values for the parameter α from zero to three. The bias for the complete case analysis is uniformly distributed over the whole parameter range, with a small increase for higher α values. The bias lies between 10 and 17%. The imputation methods can easily distinguished into two groups, as was the case with the Atkinson index. The one group of imputation software introduce a high positive bias and consists of hot-deck, NORM, and IVEware. And the other group has a negative bias with respect to the true case and includes MVIS, areg, and MICE.

Far from acceptable from a poverty research point of view is the hot-deck imputation, even for lower α values (like zero which corresponds to the head count ratio) the imputation introduces a bias triply as high as with the complete case analysis. Although the bias for lower α values are not so high for the NORM and IVEware imputation, this methods are also not acceptable with respect to poverty estimation, because of the steadily rising bias up to 60%.

The areg imputation method is nearly identical to the complete case analysis. However, it does not improve the bias for any parameter value, but slightly increase the observed bias. Only two imputation methods reduce the bias when compared to the complete case analysis, these are MICE and the MICE derivate MVIS. And only MICE was capable to reduce the bias over the whole range of α values considered. Therefore it can be noted, that with respect to the estimation of poverty MICE performs best, and second is MVIS, followed by aregImpute.

3.8 Summary

Handling missing data is still far away from being a trivial task and none of the software packages under investigation can handle this problem automatically without serious intervention of the user and adoption to his specific research question. However, stuck with the complete case analysis is in most cases not appropriate, especially with large scale survey data.[152] Similar problems can be seen for the structure of item-nonresponse on the household income question in the Finnish part of the ECHP. The distribution of the denier is very different from the distribution of those persons who were willing to answer (cf. the estimated kernel densities of the respective distributions in figure 3.10). To assume for this case missing completely at random (MCAR) seems highly unreasonable. Even for the estimation of some easy parameters of this distribution the complete case analysis lead to a serious bias, as shown in section 3.7. Therefore the need for an appropriate imputation strategy is obvious in the present case. Note that even for a dataset were it is reasonable to assume MCAR with respect to one variable an imputation should be considered, because the loss of cases in a multivariate analysis could lead to a severe loss of power. Also the multivariate structure of the missing mechanism can be not MCAR for all variables under investigation, and assuming MAR in most cases needs an imputation task.

Fortunately, the Finnish survey includes additional register information (served as 'true' incomes) from outside the survey. This unique situation was utilized to evaluate different imputation methods with respect to their power to minimize this observed bias. To avoid a mixture with error in variables only the true income was used and for cases with item-nonresponse on the household income question deleted. The comparison has been made with six different imputation programs: hot-deck (Stata), MVIS (Stata), NORM (R/S-Plus, standalone), aregImpute (R), MICE (R/S-PLus), IVEWare (SAS, standalone). This software cover a wide range of currently available imputation methods as well as statistical analysis packages.

From the results of the empirical evaluation two main conclusion can be drawn. First, imputing a rather complex variable within a large survey cannot properly be down with a simple imputation method like hot-deck. For all evaluation criteria this method performed worse and even introduced additional bias when compared to a simple complete case analysis. Secondly, the decision which method is most appropriate is even in such a comfortable situation -

[152] In particular because surveys facing a rising share of unit- and item-nonresponse in recent years, at least in Germany (Frick et al., 2006)

where the 'truth' for all missing values is known - enormously dependent on the criterion used and no real best imputation method can be identified.[153]

Overall it seems that the more flexible the imputation software is, and therefore adjustable to the specific needs of the imputations task at hand, the more robust are the imputations with respect to the introduction of an additional bias. Especially the NORM software, which assumes a multivariate normal distribution of the data, is too restricted for a complex survey. But also with IVEware, a very comprehensive imputation package, some problematic situations appeared. Whereas the estimation of the mean and Gini were best with this software, the distribution preservation and especially the poverty estimation were biased.

The most robust imputation was achieved with the help of the MICE package, the resulting distribution of the income was very much improved when compared to the complete case analysis. However, the chosen parameters for evaluation were improved noticeable only for the poverty estimation. In the case of the Gini and the mean no real difference to the complete case analysis could be observed.

Overall it can be concluded that the multi method strategies based on regression models perform best when dealing with a complex survey. Further research should definitely test the impact of including longitudinal information, which normally improve considerably the estimation of the imputed values (Spiess and Goebel, 2004, 2005). But how this can accurately be done with the household as analysis level is an open question, because of the possible instability of households.

When the perspective is one of data producer versus data user some additional arguments may be relevant. For a data user the software should be user friendly and should at best fit into the statistical package the researcher normally use. At least the imputation task and the substantive analysis should fit into a common framework. However, this judgment is highly subjective and can only be done by the researcher himself. The easiest software to use was from my point of view the aregImpute function and MVIS.

Taking up the position of a data producer, arguments like flexibility, reliability and robustness becoming more relevant. In such a situation the MICE framework seems to accomplish this requirements best.[154] Basically IVEware is also capable for such an environment, but the lack of diagnostic tools and 'openness' affects such an adoption. The main problem is that the source code for IVEware is not available, which gives the data producer no possibility to adopt the imputation algorithms or model building procedures to his special needs. Also the assembly chosen for IVEware is not as modular as that of MICE, therefore the imputer can not plug-in some additional methods or diagnostic tools for his specific imputation task.

[153] Williams and Bailey (1996) can also not recommend one solution for all imputation tasks.

[154] It should be noted that the successor of MVIS, ice, may also fit in such an environment, because many restrictions to the flexibility of MVIS are now obsolete.

Chapter 4

Measuring Households with High Incomes – The Case of Germany

To analyze the distribution of income, not only are accurate measures or estimation methods needed, but also accurate income data itself (Atkinson and Leigh, 2005). There is consensus within the scientific community that the tails of the income distribution are usually under-represented in sample surveys. As a result of this methodological problem, almost all surveys suffer from a middle-class bias which leads to an underestimation of income inequality. Whereas there has been a longstanding academic debate about the lower tail (which is at risk of poverty) dealing with numerous specific issues (measurement, coverage, formation and elimination), there has been comparably little discussion or research about the upper tail (with a high chance of wealth). To get a thorough picture of the complete distribution as well as to analyze poverty and affluence correctly it is crucial that the whole distribution of incomes be measured as accurately as possible. As a result the improvement of the coverage of high incomes by large scale surveys has emerged as a central research issue in recent years.

The first "German Poverty and Wealth Report" (BMAS, 2001, "Armuts- und Reichtumsbericht") asserts that the available statistical data is not sufficient to derive a comprehensive picture of wealth. Therefore the German Government[155] authorized the *SOEP* group to conduct a special "high-income sample" and to integrate this new sample (sample G) into the existing samples A to F of the ongoing Socio-Economic Panel study *SOEP*.

The aim of this chapter is to give a comprehensive analysis of this sample with respect to the overall picture of income inequality derived both with and without it. More specifically, are recipients of high incomes covered by the old subsamples in approximately the same manner with respect to the income distribution as in the new "special sample"? How is the measurement of poverty and inequality affected when the upper tail of the income distribution is covered

[155] To be specific it was the "Bundesministerium für Gesundheit und Soziale Sicherung" (Ministry of Health and Social Security).

more accurately? In particular, I address the question of whether the actual distribution of high incomes follows a Pareto distribution and if the oversampling of high income cases affects this estimation.

I will therefore first briefly describe the different subsamples of the *SOEP* and in particular the high-income sample G. Then i will show the impact of adding a "high-income" sample on some key parameters of poverty and inequality. After that I will concentrate on the high-income cases, defined by an income of more than 200% (300% respectively) of the contemporary mean, comparing the differences in the distribution between samples A-F and G. Finally I will present the results of fitting a Pareto power law distribution to the empirical data.

4.1 The High-Income Sample within the *SOEP*

4.1.1 Description of the *SOEP* Samples

The *SOEP* consists of several specific subsamples, named A to G.[156] The main random subsample A of the *SOEP* was collected in 1984 and included around 4,500 households. Subsample B, again a random sample, contains the five major groups of labor migrants to the Federal Republic of Germany, which were oversampled with a total of 1,400 households, also in 1984. The oversampling allows separate analyses for this group because of a sufficiently large number of cases.

Due to German unification, a subsample for East Germany was added in June 1990 (sample C), just before the conclusion of currency, economic, and social union in Germany, consisting of about 2,200 households. Shortly before and after the fall of the Wall, immigration from Eastern Europe to West Germany increased heavily but could not be covered by a longitudinal survey, thus necessitating a special immigrant sample. This subsample D was collected in 1994-1995 for about 500 households with persons who immigrated since 1984. In 1998, a supplementary random sample (E) was started as a methodological test sample to fulfill the following aims: (1) stabilization of the number of observations in the *SOEP* for cross-sectional and longitudinal analysis, (2) enabling analysis of "panel effects", and (3) allowing for analysis of representativeness. With sample E, the methodological basis was established for a sample that would almost double the original sample size and allow the analysis of the changes for relatively small groups of the population. The first wave of subsample F in 2000 consisted of 10,890 adult respondents and 2,993 children living in 6,052 households. An analysis of the similarities with respect to the underlying income distribution of these two samples can be found in Frick et al. (2006). The high-income sample that was added in 2002 will be described in the following section.

An overview of the sample sizes actually used in this chapter is given in Table 4.1. The number of cases are different from the above-mentioned figures for two reasons: first because of panel attrition since the start of each sample, causing the numbers to decrease, and second because the population under analysis is a

[156] For a more detailed description of the various subsamples of the *SOEP* see Haisken-DeNew and Frick (2005).

selection of the total number of cases. The following analysis is conducted at the individual level, as in most welfare economic research.[157] The population covered includes all persons living in private households with a cross-sectional design for the years 2002 through 2004, and including a positive weighting factor.[158]

Table 4.1: SOEP Sample sizes over time (individual level, including children, 2002-2004)

	Sample							
Year	A	B	C	D	E	F	G	Total
2002	6970	2201	4144	642	1662	10491	2813	28923
2003	6857	2109	4115	652	1602	9862	2526	27723
2004	6677	2023	4073	614	1568	9410	2418	26783

Source: *SOEP* 2002 to 2004, author's calculations, not weighted.

4.1.2 High-Income Sample G

The main problem when drawing a sample of high-income earners in Germany is that, as is the case in most countries[159], no official register of incomes exists from which a random sample can be drawn. Therefore the *SOEP* group utilized another very large sample in which the information about household income had been observed previously and the respondent had agreed to further interviews. The "universe" for this draw were the cumulative results of more than 50 telephone surveys with around 100,000 telephone interviews.

To draw a sample from this gross "population" several selection criteria were used, the most important of which was monthly household income, which must be above 3,855 €.[160] In the first wave, a net sample of 1,224 households was achieved. Of these, only households with an income higher than 4,500 € were included in the gross sample for the following waves, which corresponds to the increase in the income threshold. In 2003, the second wave, subsample G consisted of 1,006 households with successful interviews. The response rate for this sample was not lower but in fact higher than for simple random samples, like subsample F. This could be seen as an indication of the first-wave households' willingness to participate in the survey. For a detailed description of subsample G and the underlying sampling method, see Schupp, Gramlich, Isengard, Pischner, Wagner and von Rosenbladt (2003) and Schupp, Gramlich, Pischner, Wagner and von Rosenbladt (2005).

The following analysis uses as an income measure the monthly net household income as reported by the household head in the household questionnaire. All

[157] See chapter 2 and 5

[158] The actual selection is based on the variables `snetto`, `tnetto`, or `unetto` as well as `sphrfag`, `tphrfag`, or `uphrfag` greater than zero and the corresponding variables `spop`, `tpop`, or `upop` equal one or two, which also includes children under the age of 16. A more detailed Table of the sample sizes can be found on the SOEPinfo server for households and persons.

[159] A special case is Finland, where such register information does exist. For an analysis of this data and the use of such true information for the evaluation of income imputation strategies, see chapter 3 on page 53.

[160] The decision was based on the "pure" household income, not corrected by an equivalence scale.

income measures are deflated to prices of 2000 and corrected by household size and structure with the help of the modified OECD equivalent scale. All calculations are made on weighted estimations using the weight variables provided by the *SOEP* group.[161] Because the income threshold for the selection into sample G was based on non-equivalent household income, not all households in sample G has to be rich *a priori*, not even for the first wave.[162] Therefore the selection of rich households by means of a cut-off value is defined relative to the *overall mean income* in this analysis. As cut-off values, 200% (high income) and 300% (very high income) of the mean net equivalent household income were used.

When comparing the mean and median for each sample, as presented in Table 4.2, the expected picture is seen. Subsample A ("West German households") has the highest mean of the "old" samples, because it consisted mainly of West German citizen. And the samples for East Germany and foreigners as well as migrants have lower values. Samples E and F, which are random samples for the whole population living in private households in Germany, are in-between. Both mean and median for subsample G are dramatically higher. In all years under investigation, the values are more than twice the values in the other samples. Obviously the sampling procedure successfully reached households with high incomes, even when correcting for household size and age structure.

Table 4.2: Mean and median by sample (2002-2004)

Sample	Mean			Median		
	2002	2003	2004	2002	2003	2004
A	1546	1597	1569	1425	1479	1407
B	1232	1242	1190	1123	1148	1082
C	1334	1343	1352	1264	1265	1278
D	1286	1298	1281	1189	1239	1203
E	1428	1510	1460	1304	1343	1317
F	1424	1456	1441	1320	1347	1324
G	3769	3399	3328	3081	2927	2924
Total	1557	1572	1549	1366	1389	1360

Source: *SOEP* 2002 to 2004, author's calculations.

However, because of the income threshold used to select a household into sample G, the inclusion into the *SOEP* universe changed the distribution only in the upper tails, even when using an equivalent income concept. This can be seen in Figure 4.1, where the quantiles with and without subsample G are plotted within one Figure. The black lines represent Samples A-F and the

[161] Because household net income was used as a selection criterion for sample G, this criterion was also utilized to establish an integrated weighting scheme by the *SOEP* group. Therefore the weights for the old samples were only adjusted above the income threshold for selecting cases into sample G. This integration assured that the estimated population numbers of high-income cases as well as the distribution below this threshold remained constant. As a consequence, the median for the first wave (2002), for example, is identical regardless of the use of sample G or not, as can be seen in Table 4.3 and Figure 4.1. For a more detailed description of the integration of sample G into the existing weighting scheme of the *SOEP*, see Pischner (2004).

[162] For the following waves possible income mobility can also lead to the presence of middle-income or even poor households within subsample G.

red line displays the values for the high-income sample. There are hardly any differences below the 90% quantile. Only at the very top of the distribution (99% quantile) is a real distinction between the two estimation observable. This means the common procedure does not really lead to a biased picture of the income distribution; the number of cases in the upper tail might be too small for a meaningful analysis of this group according to sociodemographic characteristics, for example, but the overall distribution is not biased.

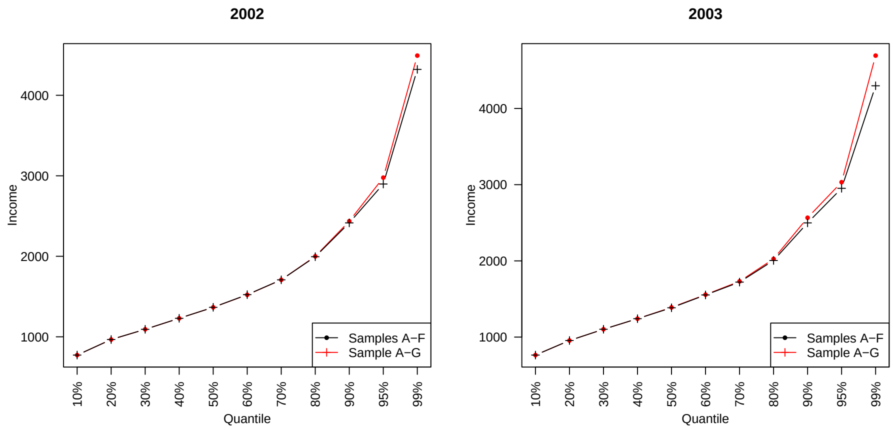


Figure 4.1: Quantiles with and without high-income sample (all cases)

In Table 4.3 some frequently used indices on income distribution are given for the *SOEP* with and without the high-income sample. For the mean, Gini, Theil (0 and 1) and FGT (0, 1 and 2), we also give the lower and upper bound for a 7% significance level. The confidence intervals are calculated using the random groups approach with the corresponding variable provided by the *SOEP* group, because standard errors based on textbook formulae does not correctly account for the clustered structure of *SOEP*.[163] Numbers in **boldface** are significantly different on a 93% confidence level and an additional asterisk denotes significant differences at a 99.2% level.[164]

As could be expected, the inclusion of more cases in the high-income group and therefore better estimation of this tail of the distribution does not significantly change the estimated level of poverty. Here poverty is measured with the Foster et al. index with an α of zero, one, and two, corresponding to the head-count ratio, poverty gap ratio and poverty intensity. The poverty line is set to 60% of the contemporary median over all samples.

Whereas the median is constant[165] (see Figure 4.1), the mean rises when sample G is included. A significant difference, albeit a relatively small one, can only be found in the years 2002 and 2003. Most differences can be found in

[163] A description of the random groups approach can be found in Rendtel (1995, ch. 4).
[164] The complete Table with the confidence bands for the corresponding lower α level can be found in the Appendix in Table B.1.
[165] Which must be the case for the first period by definition because of the inclusion of sample G in the weighting scheme of the *SOEP* (see footnote 161 and for more details, Pischner 2004).

Table 4.3: Differences in distribution parameters with and without the high-income sample

Year	Samples A-F			Samples A-G		
	Lower bound	Estimate	Upper bound	Lower bound	Estimate	Upper bound
			Median			
2002	1355	1366	1380	1357	1366	1381
2003	1351	1385	1418	1351	1389	1432
2004	1338	1358	1380	1327	1360	1393
			Mean			
2002	1505	1528	1544	1546	1557	1568
2003	1532	1543	1557	1556	1572	1592
2004	1487	1517	1552	1522	1549	1577
			Gini			
2002	0.2483	**0.2576**	0.2640	0.2661	**0.2706**	0.2773
2003	0.2556	**0.2594**	0.2638	0.2641	**0.2699**	0.2748
2004	0.2551	0.2590	0.2616	0.2633	0.2693	0.2733
			Theil (0)			
2002	0.1053	**0.1126**	0.1181	0.1210	**0.1266**	0.1327
2003	0.1105	0.1136	0.1200	0.1177	0.1237	0.1288
2004	0.1093	**0.1121**	0.1152	0.1182	**0.1221**	0.1274
			Theil (1)			
2002	0.1118	**0.1141***	0.1182	0.1324	**0.1442***	0.1515
2003	0.1106	**0.1151**	0.1207	0.1247	**0.1329**	0.1511
2004	0.1108	**0.1154**	0.1195	0.1240	**0.1331**	0.1439
			FGT ($\alpha = 0$)			
2002	0.1120	0.1164	0.1316	0.1120	0.1164	0.1312
2003	0.1174	0.1333	0.1426	0.1159	0.1335	0.1412
2004	0.1040	0.1252	0.1327	0.1122	0.1225	0.1300
			FGT ($\alpha = 1$)			
2002	0.0246	0.0283	0.0325	0.0242	0.0283	0.0324
2003	0.0255	0.0294	0.0339	0.0257	0.0297	0.0346
2004	0.0232	0.0273	0.0320	0.0241	0.0268	0.0312
			FGT ($\alpha = 2$)			
2002	0.0088	0.0106	0.0131	0.0087	0.0106	0.0131
2003	0.0082	0.0106	0.0134	0.0084	0.0107	0.0138
2004	0.0071	0.0094	0.0111	0.0073	0.0092	0.0108

Source: *SOEP* 2002 to 2004, author's calculations.
Significantly different estimates at a 93% confidence level are typed in bold face, an asterisk (*) denotes significantly different at a 99.2% level.

the measurement of income inequality.[166] The three chosen inequality measures are not equally sensitive to changes in the income distribution. The Theil (0), or MLD (Mean Logarithmic Deviation), is most sensitive to the lower tail of the distribution, whereas the Theil (1) measure is sensitive to both tails of the distribution. The Gini coefficient is most sensitive to changes in the income distribution with the highest density, normally the middle-income range.

Exactly this behavior is mirrored in the estimated coefficients for the years 2002 through 2004 for the *SOEP* with and without the high-income sample. The Theil (0) measure is only significantly different for the starting year of the new sample, the Gini for the first two years, and the most sensitive measure to top incomes, the Theil (1), for all years.[167] Overall it can be observed that the changes in the parameters presented in Table 4.3 are minor in scale, and major significant differences can only be found for a top-sensitive inequality measure in the first year of inclusion. When a higher confidence level is chosen (99.2%), only the difference in 2002 for the Theil (1) measure remains significantly different (marked in Table 4.3 with an asterisk).[168]

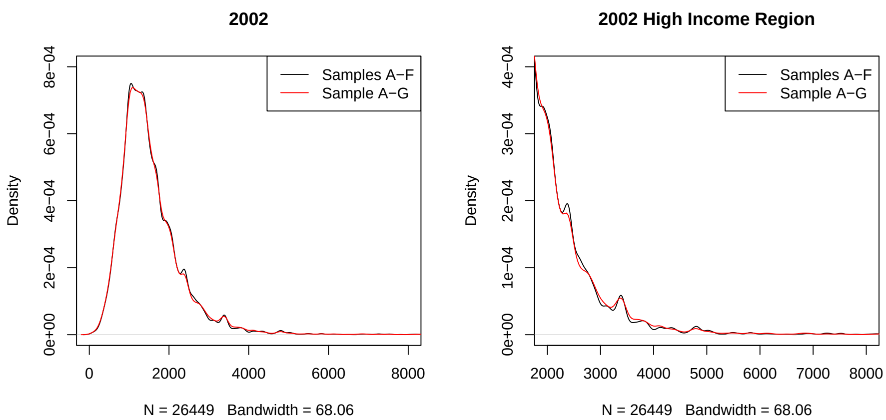


Figure 4.2: Non-parametrical kernel density with and without high-income sample (all cases)

This picture of nearly no differences in the overall sample due to the inclusion of the new high-income sample is confirmed by a non-parametric kernel density estimation, which can be found in Figure 4.2. The red line represents all *SOEP* samples, where as the black line is the distribution estimated when using only cases from samples A-F. The two estimated distributions are so close together

[166]The improved coverage of high-income cases and therefore a decrease in the weights assigned to these cases does not lead automatically to an increase of inequality. This can be illustrated by a simple example. Consider four persons drawn in a sample with the incomes $y = \{10, 22, 300, 500\}$ and an equal frequency weight of 10. The resulting Gini would be 0.5252. If we now introduce more people to the upper income group, our income vector may change to $y = \{10, 22, 300, 350, 450, 500\}$, but the weights vector also changes to $w = \{10, 10, 5, 5, 5, 5\}$. Now the resulting Gini coefficient would be 0.5177, which is clearly lower than in the first case.

[167]For the effect of the first interview in an ongoing panel on the distribution of income and possible behavior changes, see Frick et al. (2006).

[168]The confidence bands for all measures with a lower α level can be found in the Appendix, in Table B.1.

that the different lines are hard to distinguish. There is clear evidence that the overall distribution is not changed by the integration of the high income sample. Even when the Figure is zoomed to the relevant high-income area, as shown in the right-hand picture, no real differences can be seen. These findings are stable over time, which is confirmed by the estimated densities for the years 2003 and 2004 in Figure B.1 in the Appendix.

However, it is likely that a comparison between samples A-F and sample G restricted to persons living in rich households would yield to differences in the underlying income distribution. Especially the limited number of cases in the old samples may have led to a biased estimation of the high-income range, which was one reason for the special high-income sample, which was intended to increase this number of cases. The question of biased estimation is addressed in the next section.

4.2 Distribution of High Income

Since the *SOEP* has no cut-off value when asking for income like some other surveys, we should find rich households possibly wealthy individuals in each of the older samples. However, because the probability of such high incomes is rather low, the number of cases should be considerably lower than in the special high-income sample. Table 4.4 shows the number of rich persons for each sample in absolute numbers as well as the percentage of persons living in households with a net equivalent household income more than twice the overall mean. Note that the case numbers in the Table are not weighted because the actual possible number of cases for further analysis is of interest here. However, the threshold (200% of mean income) was calculated with the appropriate weights. The thresholds used are, for the 200% classification, €3,134 in 2002, €3,145 in 2003, and €3,099 in 2004. The respective figures for the 300% differentiation are €4,670, €4,717 and €4,649.

The Table shows that in 2002 around five percent of sample A were high-income earners, whereas only around one percent of the “foreigner” and East German samples (B and C). Because Samples E and F cover all three of these groups, their value is in-between, at around four percent. The immigrant sample D has a surprisingly high fraction of high-income cases, namely 3.53% in 2002. The high-income sample has – as expected – a very high percentage of cases above the defined threshold, around 46%. These numbers are almost completely stable over time, with the exception of samples B and D, which may give some indication of widespread problems of integration.

These findings are further accentuated when using 300% of the contemporary mean income as threshold for wealth, as shown in Table 4.5. None of the older samples has even one percent of wealthy cases. For Sample G, it is clear that the structure of the sample is completely different: even with this high threshold, more than 12% of the cases are still defined as (very) rich. In contrast, we find less than 10 very rich individuals in Samples B, C and D. For sample B, wealthy people with an income above 300% are almost nonexistent. Although the distinction between foreigner and native is not identical with the sample distinction, it seems that the region of very high incomes is dominated by persons

Table 4.4: High-income cases by sample and year (more than 200% of contemporary mean)

High Income	A	B	C	D	E	F	G
				2002			
FALSE	6724	2195	4170	1040	1560	9851	1919
TRUE	343	30	44	38	69	385	1604
% TRUE	4.85	1.35	1.04	3.53	4.24	3.76	45.53
				2003			
FALSE	6596	2124	4128	1031	1484	9286	1429
TRUE	338	17	75	36	65	305	1109
% TRUE	4.87	0.79	1.78	3.37	4.2	3.18	43.7
				2004			
FALSE	6415	2019	4113	987	1474	8898	1329
TRUE	328	14	73	29	64	369	1132
% TRUE	4.86	0.69	1.74	2.85	4.16	3.98	46

Source: *SOEP* 2002 to 2004, author's calculations, not weighted.

with (West) German citizenship. These findings are perfectly in line with those of Schupp et al. (2003), who observed a significant underrepresentation of foreigners in income regions above €5,113. Only 2.5% of these high incomes recipients were foreigners, whereas the overall share of foreigners was around 6.7%.[169] In the following, only those cases are included that are wealthy, either according to the 200% (high income) or the 300% threshold (very high income).

Table 4.5: High-income cases by sample and year (more than 300% of contemporary mean)

High Income	A	B	C	D	E	F	G
				2002			
FALSE	7000	2223	4210	1070	1613	10149	3072
TRUE	67	2	4	8	16	87	451
% TRUE	0.95	0.09	0.09	0.74	0.98	0.85	12.8
				2003			
FALSE	6874	2138	4191	1061	1539	9512	2176
TRUE	60	3	12	6	10	79	362
% TRUE	0.87	0.14	0.29	0.56	0.65	0.82	14.26
				2004			
FALSE	6695	2032	4178	1008	1526	9199	2172
TRUE	48	1	8	8	12	68	289
% TRUE	0.71	0.05	0.19	0.79	0.78	0.73	11.74

Source: *SOEP* 2002 to 2004, author's calculations, not weighted.

Table 4.6 shows the mean and median as well as some inequality measures for

[169] Please note that this distinction was drawn on the basis of pure household income not corrected for household size and structure, because the underlying interest was mainly driven by the sampling procedure chosen for sample G. However, using an equivalence scale would intensify this structure, because households within the foreigner population are larger on average.

the high-income cases in Samples A-F and Sample G for the years 2002 through 2004. As above, also the lower and upper bounds of the 93% confidence interval are shown, and significant differences are typeset in **bold**. In contrast to Table 4.3, clear differences can be seen. The mean income of the wealthy in Sample G is much higher for all years compared to the older samples. However, because of the lower number of cases, only the years 2002 and 2004 can be characterized as significantly different.

Table 4.6: Differences in the distribution of high income between Samples A-F and Sample G (More than 200% of contemporary mean)

Year		Samples A-F			Sample G	
	Lower Bound	Estimate	Upper Bound	Lower Bound	Estimate	Upper Bound
				Mean		
2002	3845	**3969**	4061	4681	**5011**	5561
2003	3839	4054	4424	4363	4830	5345
2004	3670	**3975**	4200	4318	**4680**	5047
				Gini		
2002	0.0926	**0.1236**	0.1416	0.2071	**0.2663**	0.3149
2003	0.0927	**0.1156**	0.1247	0.1472	**0.2201**	0.2787
2004	0.0908	**0.1340**	0.1497	0.1677	**0.2141**	0.2601
				Theil (0)		
2002	0.0172	**0.0304**	0.0392	0.0865	**0.1408**	0.1791
2003	0.0149	0.0272	0.0392	0.0350	0.0965	0.1364
2004	0.0187	**0.0372**	0.0467	0.0513	**0.0911**	0.1324
				Theil (1)		
2002	0.0195	**0.0365**	0.0477	0.1346	**0.2159**	0.2620
2003	0.0160	0.0352	0.0518	0.0376	0.1441	0.1925
2004	0.0213	0.0465	0.0654	0.0642	0.1336	0.2091

Source: *SOEP* 2002 to 2004, author's calculations.
Significantly different estimates at a 93% confidence level are typed in bold face.

The inequality measures calculated give clear indications that the distribution of these two groups are different. For all years, the Gini coefficient is significantly different and much higher for sample G, in fact, nearly twice as high. The difference in the estimated inequality between samples A-F and G is even higher when using the entropy-based measures proposed by Theil (1967), with a magnitude of 3 to 6. However, not all of these differences were classified as significantly different according to the random groups approach. This is the case because the Theil measures are more sensitive to the tails of the distribution, and here only the high-income region is under investigation. Therefore the distribution is heavily skewed, and at the upper tail only very few cases are left that heavily influence these measures, especially the Theil (1) measure.

We see similar results when using the higher threshold (300% of mean income) to differentiate between rich and non-rich, as shown in Table 4.7. A higher mean income for all years as well as a substantially higher level of inequality. However, when relying on statistical differences, only for the first year can we state that

Table 4.7: Differences in the distribution of high income between Samples A-F and Sample G (More than 300% of contemporary mean)

Year	Samples A-F			Sample G		
	Lower Bound	Estimate	Upper Bound	Lower Bound	Estimate	Upper Bound
			Mean			
2002	5509	**6026**	6417	7183	**8924**	11340
2003	5296	5759	6788	5727	7403	9328
2004	5953	6663	7213	6777	7795	9561
			Gini			
2002	0.0796	**0.1354**	0.1521	0.2246	**0.3330**	0.3638
2003	0.0645	**0.1238**	0.1904	0.0969	**0.2566**	0.3258
2004	0.1098	0.1720	0.1853	0.1742	0.2621	0.3193
			Theil (0)			
2002	0.0137	**0.0334**	0.0440	0.1122	**0.1899**	0.2646
2003	0.0085	0.0390	0.0733	0.0159	0.1263	0.2253
2004	0.0201	0.0501	0.0665	0.0558	0.1240	0.1893
			Theil (1)			
2002	0.0138	**0.0388**	0.0504	0.1696	**0.2541**	0.3540
2003	0.0092	0.0541	0.0816	0.0171	0.1799	0.3347
2004	0.0208	0.0576	0.0788	0.0619	0.1702	0.2772

Source: *SOEP* 2002 to 2004, author's calculations.
Significantly different estimates at a 93% confidence level are typed in bold face.

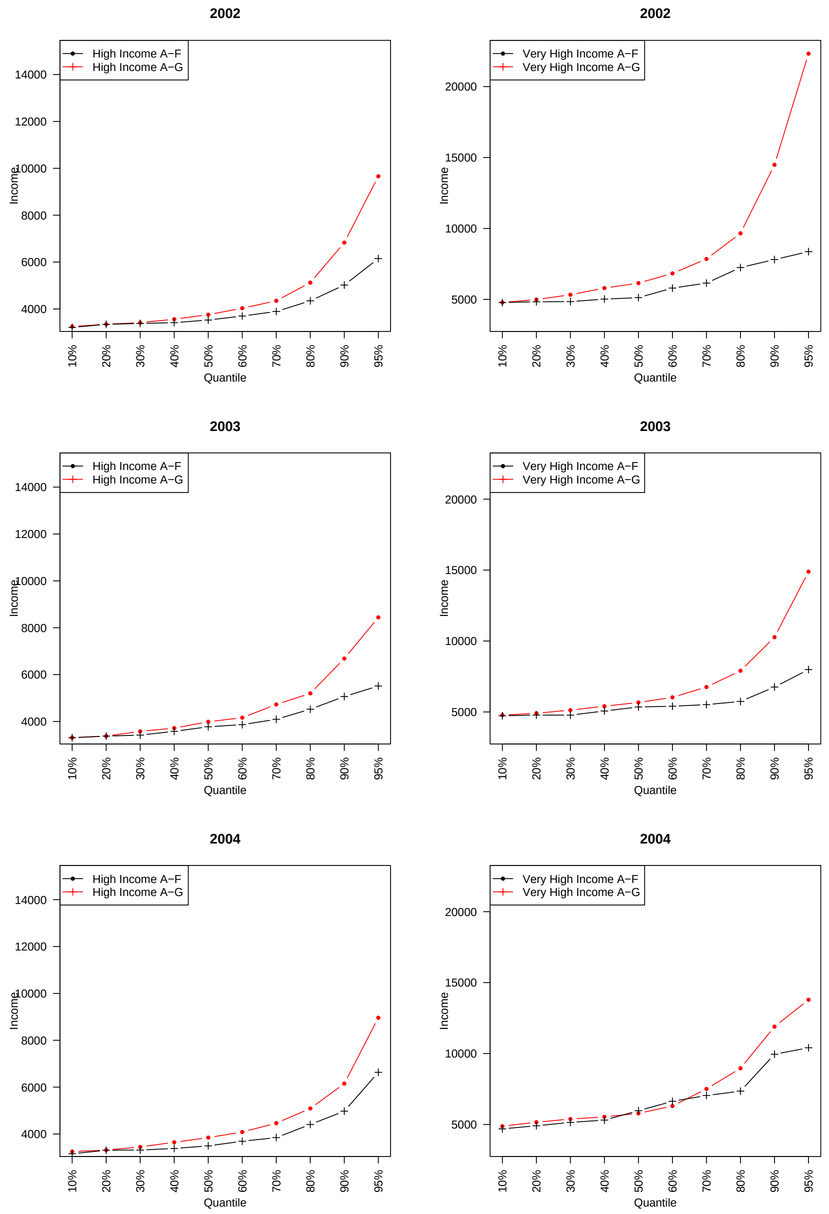
the estimated parameters are different. This is the case mainly because of the low number of cases left in the analysis and the corresponding inflation of the confidence intervals. For Samples A-F, the random groups are below 30 cases in seven out of 8 groups, which is often seen as the minimum number of cases needed for a reliable estimation.

The values of the quantiles plotted in Figure 4.3 (page 107) for the years 2002 through 2004 for the high incomes (Figure 4.3(a)) as well as for the very high incomes (Figure 4.3(b)) show clear differences between the two differentiated samples of wealthy people. Only for up to the median do we find similar values for Samples A-F and Sample G. For the upper half of the plotted quantiles, a widening gap between the two samples can be seen. The differences are more accentuated when using the higher threshold of 300% of mean income as a definition of wealth. But what also emerges is the fact that the discrepancies at the upper tail become narrower over time, especially for the very high income definition.

These findings of differently shaped distributions in Samples A-F and G are confirmed by the non-parametric kernel density estimations for each year within the period under investigation, presented in Figure 4.4 (page 108). As in the Figures above, the black lines correspond to Samples A-F and the red lines to the new Sample G. The estimated densities clearly show with the older samples much more density is concentrated at the left hand side of the plotted income scale. Whereas the red line is nearly throughout the upper income region above the black line, which means that the distribution of sample G is more widespread over the upper income range.

Applying the analysis of Gini, described later in section 5.2.1, allows a more heterogenous picture to emerge. The results are given in the Appendix in Tables B.4 and B.5. For the 200% high-income group, a convergence process can only be observed from 2002 to 2003, but with divergence in the year thereafter. However, when looking at the results for the above 300% income group, a convergence between Samples A-F and Sample G can be observed. The differences in the mean income, the group-specific Gini and the mean rank diminish over time. Between-group inequality shrinks and the overlapping index nearly reaches one, i.e., perfect overlapping of the two distributions under investigation.

However, income inequality measures are not robust to extreme cases, especially when the underlying data are sparse as it is often the case for high incomes (as is also the case here). Therefore the conclusions drawn from simple sample inequality indexes can be misleading (Victoria-Feser, 1999). The more appropriate and robust way to compare the distribution of high incomes in Samples A-F and Sample G is therefore to apply a parametrical model to the data and compare the inequality indices resulting from the parameter estimates of the model used (Victoria-Feser, 1999; Cowell and Victoria-Feser, 1996). This approach is described and applied in the next sections.



(a) High Income (200%) (b) Very High Income (300%)

Figure 4.3: Quantiles of high-income cases for Samples A-F vs. G

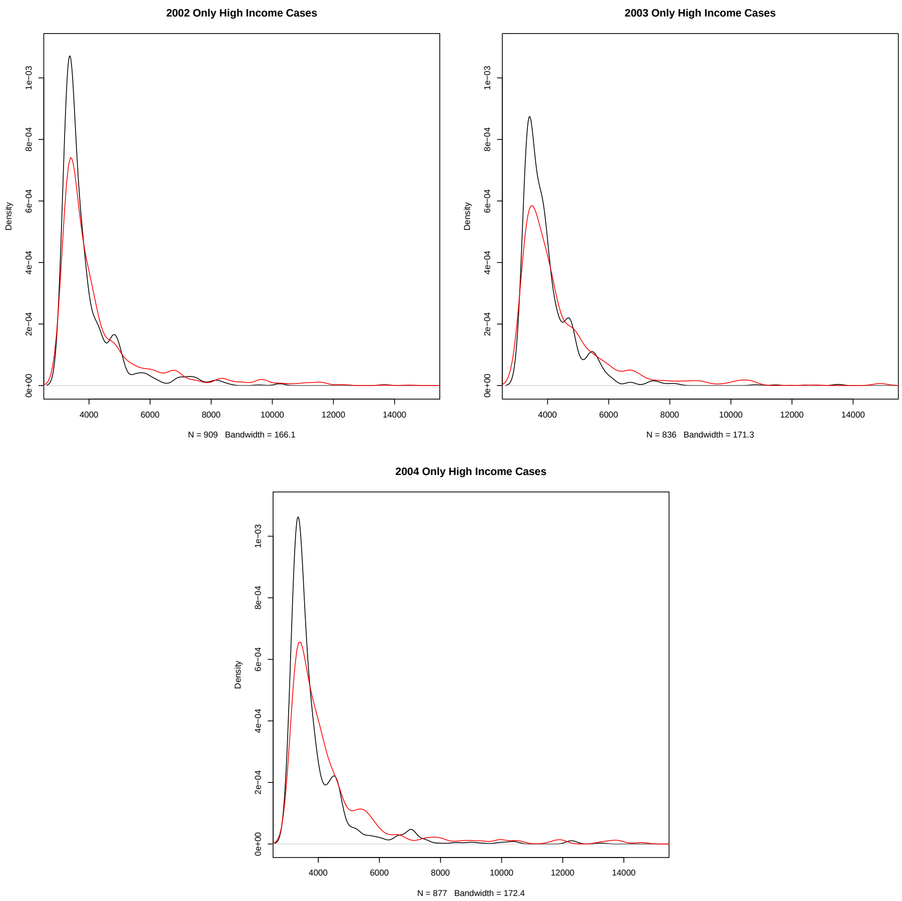


Figure 4.4: Non-parametric kernel density for high-income cases (Samples A-F and Sample G, 200% of mean income)

4.3 Description of the income distribution by a parametrical model

There is a long and still unsolved discussion in the scientific literature about which parametrical model best describes the distribution of incomes. One of the first suggestions was made in 1897, when Pareto presented the following relationship for the distribution of wealth

$$P(Y \geq y) \propto y^{-\alpha} \ . \tag{4.1}$$

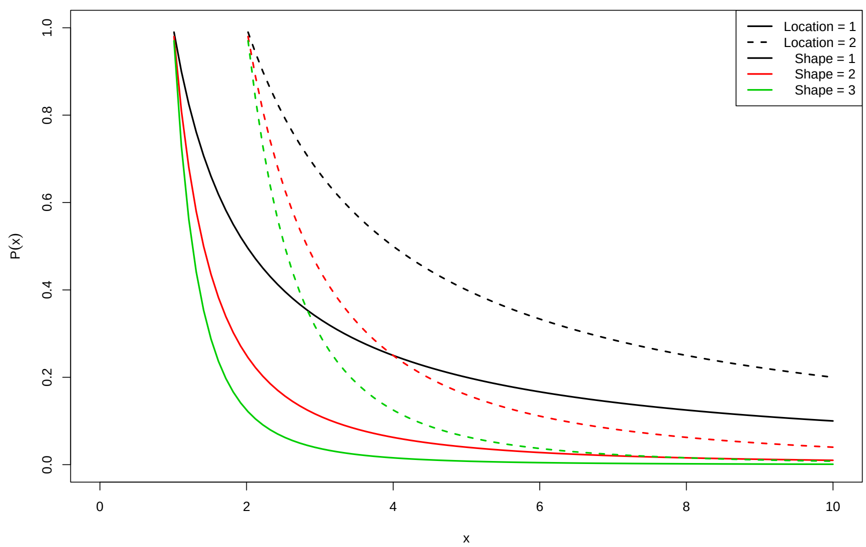


Figure 4.5: Pareto distribution with different location and shape parameters

Here y denotes income and $P(Y \geq y)$ the number of persons with an income larger than or equal to y. The exponent α, the shape parameter, is often called the Pareto index. This corresponds to what is observed when plotting the logarithm of the number of persons with an income above a certain threshold against the logarithm of the income itself: a series of points close to a straight line. This power law behavior is nowadays known as the *Pareto law* and normally restricted to the upper incomes. The appearance of the Pareto distribution with different shape parameters is shown in Figure 4.5.

There is a clear connection between the shape parameter of the Pareto distribution and the resulting Gini coefficient. Moothathu (1985) pointed out the following relationship:

$$G(\alpha) = \frac{1}{2\alpha - 1} \, , \tag{4.2}$$

where G denotes the Gini coefficient and α the shape parameter of the underlying Pareto distribution. From this follows that the more pronounced the shape of the estimated Pareto distribution, the lower the corresponding Gini coefficient (see Figure 4.6).

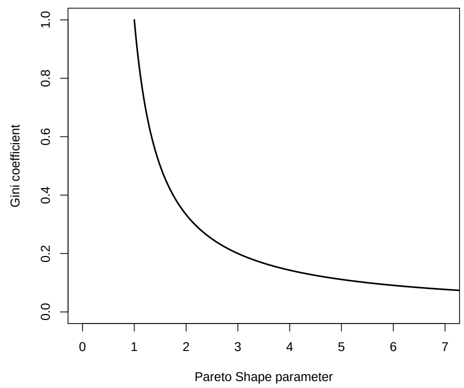


Figure 4.6: Shape parameter of the Pareto distribution and corresponding Gini index

Gibrat (1931) seems to have been the first to point out that Paretos theory only holds for the top 1% of incomes. The incomes of the remaining 99% of the empirical distribution follow a log-normal distribution; that is, the logarithms of the incomes follow a normal, symmetrical distribution.[170] The probability density function is defined as:

$$p(y) = \frac{1}{y\sigma\sqrt{2\pi}} \cdot e^{\left[-\left(\frac{\ln y - \mu}{2\sigma}\right)^2\right]} \tag{4.3}$$

where $0 < y < \infty$, and where μ and σ are the mean and the standard deviation of the normal distribution. The value of the fraction $\beta = 1/\sqrt{2\sigma^2}$ is called the Gibrat index; if β has low values (large variance of the global distribution), the personal income is more unevenly distributed.

Mandelbrot (1960) invented the weak Pareto-Lévy-Law. Using his notation where u denotes income, this weak law means that the $P(Y)$ only "behaves like" $(\frac{u}{u^0})^{-\alpha}$, as $u \to \infty$. He showed that the Pareto distribution fits the distribution of the high incomes very well, concluding that there is little question of the validity of the Pareto law where "sufficiently large values of u are concerned" (Mandelbrot, 1960, p. 81). Krämer and Ziebach (2002) state that most of the "purported 'disproofs' of 'the' Pareto law" apply to the original strong variant only, and even assert that almost all distribution obey the weak Pareto law in the upper tail.

Singh and Maddala (1976) proposed a generalization of the Pareto and the Weibull distribution, a distribution used mostly in analyses of equipment failures. The proposed distribution fits more accurately than the Pareto and log-normal distributions if the overall income range is used.[171] Also two forms of the generalized Beta distribution were proposed by McDonald (1984), who fits this distribution to income data for the USA. These distributions include many other proposed models as special or limiting cases, such as the generalized Gamma, Singh-Maddala, log-normal, Weibull and Pareto distribution.

[170] Clementi and Gallegati (2005b) found precisely this for the Italian income distribution.
[171] Later this distribution was named the Singh-Maddala distribution after its inventors.

Bandourian (2000) and Bandourian, McDonald and Turley (2002) fitted various of the above-mentioned income distribution models to the data of the Luxembourg Income Study (LIS) and found that the more general the distribution is, the more accurately one can describe the empirical distribution. The best fitting distributions are in the class of the generalized beta distribution, when the fit is compared to the overall distribution and not restricted to the upper tail.

However, there are clear indications in the literature that the upper tail of the income distribution follows a Pareto distribution although the complete distribution can be better approximated by more general, flexible models. Drăgulescu and Yakovenko (2001) showed for the income distribution of the USA and Great Britain "that above 100 k£ the data follow a power law with the exponent equal to 1.9" and "that below 100 k£ the data is very well fitted by an exponential distribution". Ishikawa (2004) analyzed the distribution of high-income companies in Japan and found a good relation to the Pareto distribution. Clementi and Gallegati (2005b) stated that the upper tail of the Italian income distribution is consistent with the Pareto law and estimated shape parameters between 2.09 and 3.3 over the years 1977 to 2002, but excluded the top 1.4% as outliers.

In another paper the same authors (Clementi and Gallegati, 2005a) also analyzed data from the *SOEP* and found that for Germany between 1990 and 2002, the shape parameter varied between 2.42 and 3.96. They fitted a combination of a log-normal and a Pareto Law distribution and determined the cut-off values with the help of explorative data analysis tools like Quantile-Quantile Plots (QQ-Plots). With the help of these thresholds, they classified the upper 1-3% of the income distribution to follow a Pareto distribution. For their analysis they used annual net household income as provided by the *SOEP* group in the CNEF files. Klass, Biham, Levy, Malcai and Solomon (2006) showed that according to the list published in Forbes, the wealth distribution of the 400 richest people in the world follow a Pareto power law distribution with a shape parameter of 1.49.

4.3.1 Estimation of the Gini coefficient with sparse data

Based on the discussion above, it seems appropriate to assume that the region of high incomes follows a Pareto Law distribution. But why is is important to take this detour, given that micro-data at a personal level are provided within the *SOEP*? Because the number of households with high incomes relative to the overall mean income is by definition low, the correspondingly very low sampling probability results in sparse data for this region. The problem that inequality indices are only estimates from sampled income data and not robust to outlier and sparse data has been discussed by Nygård and Sandström (1989), Cowell and Victoria-Feser (1996), Victoria-Feser and Ronchetti (1997), and Cowell and Flachaire (2002). The overall recommendation is to first estimate a parametrical model for the underlying (assumed) income distribution and calculate the inequality index from the estimated parameter resulting from the chosen model.

The reason for this lies in the introduction of a bias to the inequality parameter when sparse data are used directly, even when the wider confidence bands, due to smaller samples, are taken into account. The resulting bias of this estimate is highly dependent on the underlying inequality in the society (and naturally to the sample size, converging to no bias for a total sample).

This phenomenon can easily be shown in a small simulation study. Assume that the true distribution of income over a specific threshold u follows a Pareto distribution with a shape parameter α. Let us further randomly draw k times n cases out of a (stable) population of N persons. Each time, one can directly calculate the Gini coefficient or estimate the parameter α of the underlying Pareto distribution. The bias and mean squared errors for this estimations are defined as

$$\Delta = E(\hat{\theta}_k) - \theta \tag{4.4}$$

$$MSE = E\left[\left(\hat{\theta}_k - E(\hat{\theta}_k)\right)^2\right] , \tag{4.5}$$

where θ denotes the true parameter and $\hat{\theta}$ the sample parameter estimated k times.

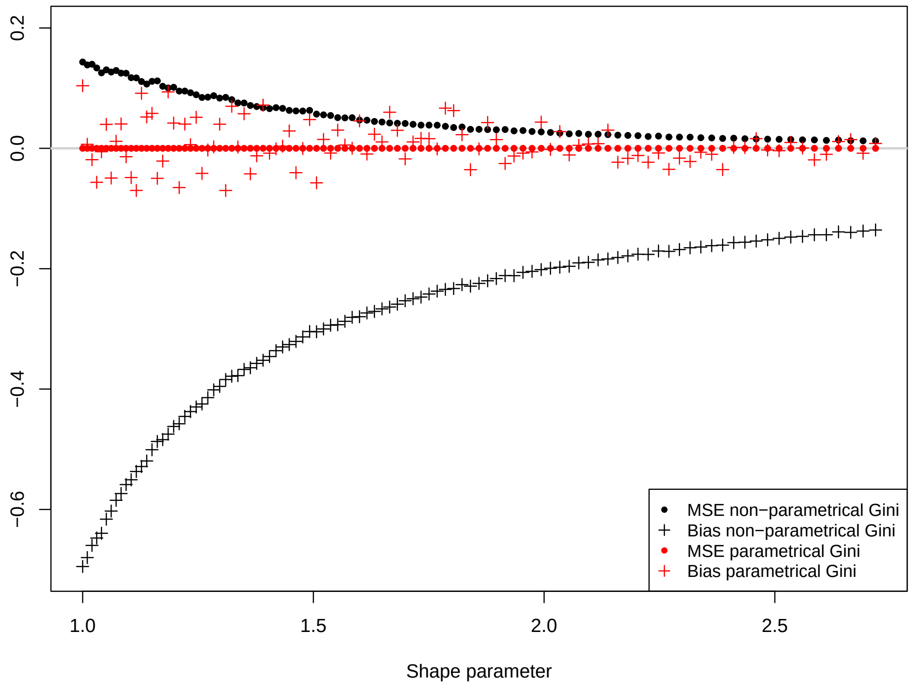


Figure 4.7: Bias and MSE for the estimated Gini index using different shape parameter of the Pareto distribution

For the simulation, the population size N is set to 200,000, the sample size n to 500, and the location parameter u is assumed to be 4,000. For each shape parameter α running from zero to one, the sample is drawn $k = 200$ times. The resulting mean squared error and bias for the direct estimation and the parametrical estimation can be seen in Figure 4.7. The direct estimations from the data for the Gini coefficient are severely biased (black points and plus signs), especially if the inequality level is high, as is normally the case for the high-income region. The resulting Gini from the parametrical distribution is much more robust (red points and plus signs), shows no evident bias, and has a much lower mean squared error. This difference between the two methods is again more important the more unequal the distribution of incomes in the population is.

When fitting a parametrical model to empirical data, it is important to obtain an impression of the model fit. McDonald (1984), Bandourian (2000), and Bandourian et al. (2002) use the sum of squared errors (SSE) and sum of absolute errors (SAE) as goodness-of-fit criteria as well as the χ^2-test for the categorized data. However, all of these studies fit different distributions to the same (grouped) data, which means that there is no need to deal with the number of "data points" used for the estimations. Because the sum of absolute or squared errors necessarily increases when the number of cases increases, this kind of criterion cannot be used in a setting like the one here. A way to derive this kind of goodness-of-fit test is to divide by N (the number of data points or people), which leads to the mean squared error. Also reported for each estimation are the log-likelihood, the degrees of freedom, and the correlation coefficient of the empirical data with the fitted distribution model.

4.3.2 The distribution of high incomes and the Pareto Law

This section describes the estimation results when using a parametrical model for the description of the high-income distribution in Germany. Assuming that the underlying distribution following the Pareto power law. The shape parameter is estimated by Maximum-Likelihood with the help of the VGAM package by Yee (2006) within the statistical program *R* (R Development Core Team, 2005).

Before discussing the estimated shape parameter α and the corresponding Gini coefficient, I will first briefly describe the fit to the different distributions, and thereafter will come back to the fitted parameters. An overview of some critical key indicators for the model fit are presented in Table 4.8, namely the Log-Likelihood (LL) with the corresponding degree of freedom (df), the mean squared error (MSE) and the correlation coefficient (cor) of the predicted values with the empirical observed cases. In comparing these numbers it becomes obvious that generally the model fit is better for the high incomes than for the very high incomes, which is most likely due to the heavily reduced number of cases. An indication of this can be seen in the comparison of the separate model fits with the fit for all samples together. Whereas for the high incomes the model fit improves when estimating samples A-F and G separately (higher LL), this is not true for the very high incomes. The reduction in sample size is more important than the additional flexibility of two possibly different shape parameters.

For incomes above the 200% mean threshold, the fit is best for sample G throughout all years under investigation. This holds for all criteria listed in Table 4.8, for a higher Log-Likelihood and correlation as well as for a lower mean squared error. This is not the case for the very high incomes, where the LL is higher in all years for the samples A-F but the MSE is lower and the correlation is higher for Sample G in two out of three years, namely 2003 and 2004. However, again it should be noted that especially for Samples A-F the case

Table 4.8: Overview: Model fit of the estimated Pareto distributions

	More than 200% mean				More than 300% mean			
	LL	MSE	cor	df	LL	MSE	cor	df
				2002				
Sample A-G	-26966822	0.0022	0.9910	2361	-6437892	0.0048	0.9924	599
Sample A-F	-23871211	0.0018	0.9925	859	-5014089	0.0039	0.9831	171
Sample G	-23257146	0.0010	0.9960	1304	-6910817	0.0051	0.9915	377
				2003				
Sample A-G	-26426620	0.0033	0.9910	1866	-6466908	0.0119	0.9830	503
Sample A-F	-22192104	0.0011	0.9969	788	-4393339	0.0041	0.9919	154
Sample G	-20876315	0.0010	0.9982	1077	-6904980	0.0020	0.9979	348
				2004				
Sample A-G	-28371537	0.0029	0.9904	1909	-5586318	0.0059	0.9896	409
Sample A-F	-23723442	0.0016	0.9938	819	-3825322	0.0083	0.9779	131
Sample G	-23088667	0.0013	0.9990	1089	-6238908	0.0017	0.9917	277

Source: *SOEP* 2002 to 2004, author's calculations. LL= Log-Likelihood; MSE= Mean-Square-Error; cor= Correlation; df= Degrees of Freedom

numbers are extremely low in the very high income area. However, overall, the Pareto power law model seems to fit rather well to the empirical distributions of the high-income ranges.

The fitted results can be investigated in more detail when looking at the figures provided for each estimation. The comparison of the empirical and the parametrical distributions for the high-income range can be found in Figures 4.8 (2002), B.2 (2003, see Appendix), and B.2 (2004, see Appendix). The corresponding figures for the very high-income range are 4.9 (2002), B.4 (2003, see Appendix), and B.5 (2004, see Appendix).

The following figures present the estimated Pareto distributions against the observed empirical distributions for the years 2002 through 2004 (the figures for the years 2003 and 2004 can be found in the Appendix B on page 169). For each year, the estimation has been done for all samples (A-G), for the older samples (A-F) and separately for the high-income Sample G. The left-hand figures always show as the black line the distribution of the empirical data and as the red line the resulting Pareto distribution from the maximum-likelihood estimation. The corresponding right-hand figure shows for each estimation a QQ-Plot, which plots the empirical data against the estimated distribution. The grey diagonal line indicates identical values and therefore the more the red line corresponds with the diagonal line the better the model's fit. The dotted grey lines are the confidence bands at a 93% significance level for the estimated parameter.

Figure 4.8 shows the estimation results for the high-income group in the year 2002, i.e. more than 200% of the contemporary mean net household income. For all models it can be said that the Pareto distribution fits the data rather well. However the fit is slightly better for Sample G, especially in the "lower income area" the estimated Pareto distribution slightly underrates the cases for Samples A-F.[172] This area is much better fitted for Sample G, where the

[172] Remember that $P(\cdot)$ denotes the probability for persons with an equal or higher income, therefore a lower $P(\cdot)$ value indicates that there are less persons left with a higher income or, conversely, more persons with a lower income.

empirical distribution lies within the confidence bands over almost the entire income range.

This is also mirrored in the diagrammed QQ-Plots on the right-hand side of the figure. The confidence bands enclose the diagonal at nearly every point, whereas the estimated distribution for Samples A-F shows some distortions. These differences, also at a minor scale, in the distribution between samples A-F and G seem to decrease over time, as can be seen in the respective pictures for the years 2003 (Figure B.2) and 2004 (Figure B.3) in the Appendix.

The situation is a bit different concerning the very high-income cases as illustrated in Figure 4.9. First of all it is evident that all confidence bands of the model fits are wider due to the sharp drop in observations compared to the 200% income threshold. Although the fit seems reasonably good, because nearly all of the empirical distributions lie within the confidence bands, pronounced jumps within the distributions can be observed. For sample G the estimated curve lies mostly above the empirical values due to some very high incomes. This corresponds to the high quantile values for sample G's very high income group in Figure 4.3(b).

The model fits over the three years are fairly stable (for the very high income-cases they even improve over time). Therefore, i will skip the discussion of the results for the years 2003 and 2004 and the present corresponding figures only in the Appendix. See Figures B.2 and B.4 for the year 2003 as well as B.3 and B.5 for the year 2004.

Table 4.9 presents the shape coefficients of the estimated Pareto distribution for Samples A-F and Sample G for the years 2002 to 2004. It is noticeable that in all years, the shape coefficient for Sample G is lower than the corresponding parameter for Samples A-F. The shape parameters for Samples A-F are also a bit higher than the values estimated by Clementi and Gallegati (2005a) for Germany. However, the latter use data from 1990 to 2001, which showed the clear trend of a rising α for the upper tail of the income distribution. Additionally the differences are, at least in part, due to the different income concepts used. Clementi and Gallegati analyzed annual household income in contrast to the monthly household income used here. The annual income normally has a higher inequality level, which corresponds to a lower α or shape parameter.

When using a 7% significance level (as shown in Table 4.9 with the lower and upper bounds of the corresponding point estimator) we see significant differences in the parametrical shape of the distribution of high incomes between Samples A-F and Sample G for all years under investigation, but with a declining trend. As Table B.2 illustrates, significant differences are only stable for the first year when using a more rigorous significance level of around 1%. It comes as a bit of a surprise that also for the very high income region only differences for the first year are observable, regardless of the selected significance level.[173]

Table 4.10 shows the Gini coefficients directly from the empirical data ("Empirical Gini") and the corresponding Gini coefficients of the estimated Pareto distributions ("Gini"), using equation (4.2) as proposed by Moothathu (1985). Also reported in the Table are the corresponding lower and upper bounds at

[173]For a possible "first wave" effect on income measurement, see Frick et al. (2006).

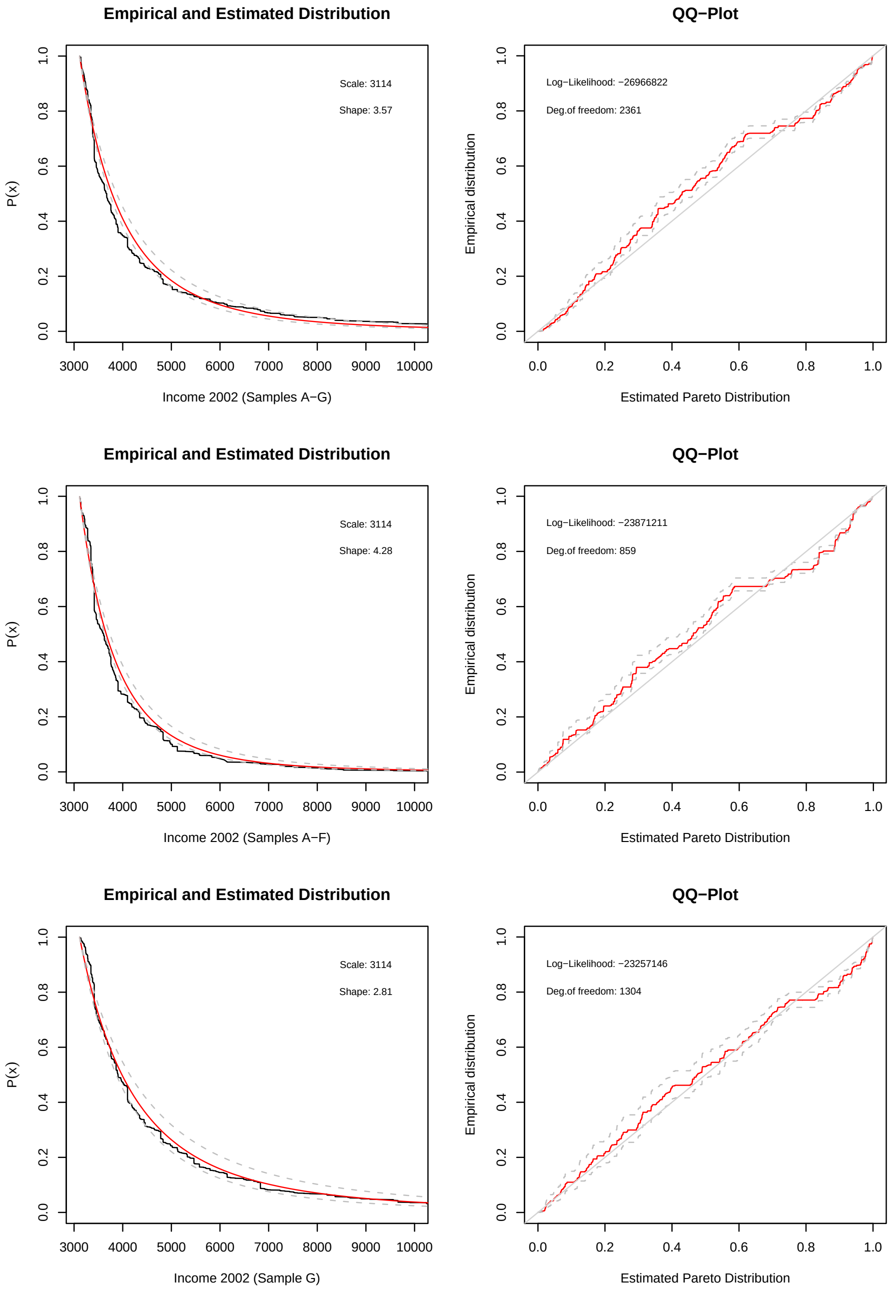


Figure 4.8: Estimated and empirical distribution of high incomes (Year 2002)

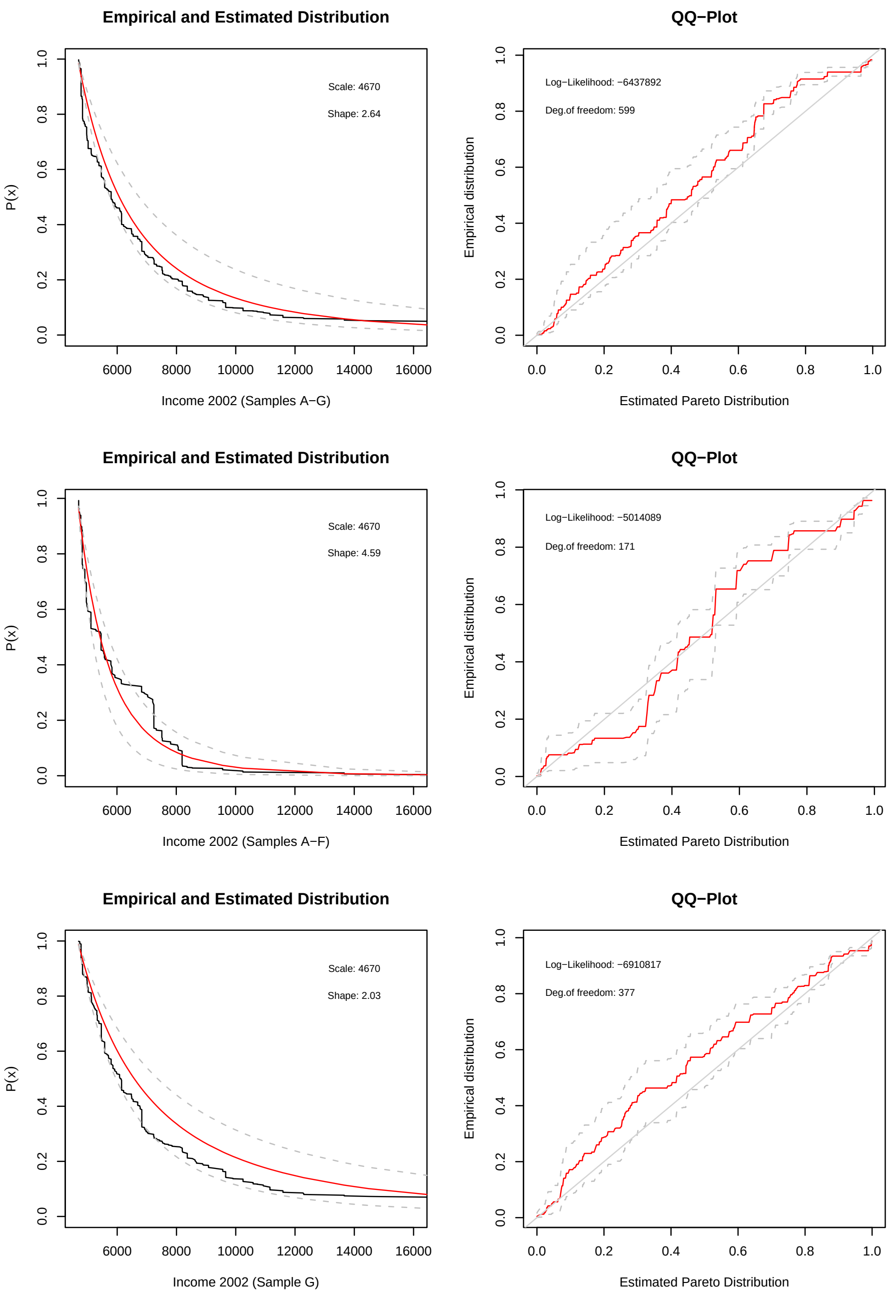


Figure 4.9: Estimated and empirical distribution of very high incomes (Year 2002)

Table 4.9: Overview: Shape coefficients of the estimated Pareto distributions

	Samples A-F			Sample G		
Year	Lower bound	Shape	Upper bound	Lower bound	Shape	Upper bound
	High Income (200% Mean)					
2002	3.80	4.28	4.54	2.42	2.81	3.19
2003	3.50	4.15	5.03	2.46	2.92	3.25
2004	3.66	4.51	5.89	2.68	3.04	3.58
	Very High Income (300% Mean)					
2002	3.45	4.59	6.90	1.52	2.03	2.84
2003	4.15	5.22	7.56	1.84	2.95	5.11
2004	2.59	3.49	4.49	2.02	2.52	3.18

Source: *SOEP* 2002 to 2004, author's calculations.

Table 4.10: Overview: Gini coefficients of the estimated Pareto distributions

	Samples A-F				Sample G				%Δ*
Year	Empirical Gini	Lower bound	Gini	Upper bound	Empirical Gini	Lower bound	Gini	Upper bound	
	High-Income (200% Mean)								
2002	0.12	0.10	**0.13**	0.16	0.27	0.17	**0.22**	0.27	40
2003	0.12	0.11	0.14	0.18	0.22	0.13	0.21	0.30	30
2004	0.13	0.08	0.12	0.18	0.21	0.16	0.20	0.24	0
	Very High-Income (300% Mean)								
2002	0.14	0.08	**0.12**	0.17	0.33	0.20	**0.33**	0.54	-11
2003	0.12	0.07	0.11	0.22	0.26	0.09	0.20	0.38	36
2004	0.17	0.08	0.17	0.36	0.26	0.14	0.25	0.49	11

Source: *SOEP* 2002 to 2004, author's calculations. (Significance Level: 0.8%)
* %Δ is the percentage reduction in the difference between the Gini for samples A-F and sample G when switching from the direct to the parametrically estimated Gini.

a 99% confidence level for the "parametrical Gini". The last column refers to the reduction in the difference between the Gini for Samples A-F and Sample G when using the parametrical estimation with a Pareto function instead of the direct estimation of the Gini over the observed data.[174]

The differences between the empirical Gini and the parametrical Gini do not follow a clear trend and vary between 0 and 0.06 for the very high-income group of sample G in 2003 (around 23%). However, the reduction in the differences between the two samples when using the parametrical approach are, in 4 out of 6 cases, positive, and vary between 11 and 40%. Only for the very high-income group in 2002 is a "negative reduction" observed. Although the differences in the empirical Ginis between the two samples are very large, only in the first year can significant differences be stated securely. The use of the more robust parametrical approach in the presence of sparse data and the very low number of cases result

[174] The results for a more modest confidence level (93%) can be found in the Appendix in Table B.3.

in rather wide confidence bands. Irrespective of statistical significance, there is strong evidence that the inequality within the high-income region of sample G exceeds the corresponding inequality level in the older *SOEP* sample, but with a declining trend.

4.4 Summary

This chapter has analyzed the impacts of integrating a special high-income sample into an ongoing panel survey, namely the addition of Sample G ("High-Income Sample") in 2002 to the German Socio-Economic Panel Study, which started in 1984. First I addressed the impact for estimates using the overall distribution, given the more accurate measure of the upper tail. The latter aspect benefits from the situation that normally the tails of the income distribution consists only of a few cases (by definition). The special high-income sample gives the opportunity, on the one hand, to analyze in more detail the distribution of this upper tail, and on the other, to rate the possible bias of a distribution estimation based only on very sparse data.

When comparing the overall differences due to the inclusion of a special high-income sample, only minor changes were found. The inclusion of Sample G in the *SOEP* survey does not lead to any remarkable changes in the overall distribution, and only very top sensitive measures like Theil (1) are different in the first year of sample G's inclusion. Therefore it can be said that the estimation of the *overall* distribution with *SOEP* seems to be quite good even before the inclusion of a high-income sample.

But when comparing only the high-income range between Samples A-F and Sample G, some remarkable differences appear, especially for the very high incomes. However, most of the differences observed at first sight, when only comparing point estimates based directly on the empirical data, do not hold when using more appropriate model-based methods and taking statistical sample properties into account.

For the distribution of the high incomes, it can stated that for both samples, a Pareto power law distribution seems appropriate. The fits of the estimated Pareto distributions are very good, regardless of the year and sample under investigation. However, there is clear indication that the inequality within the high-income region is higher in sample G than in the older samples A-F, or correspondingly with a lower shape parameter of the Pareto distribution. These differences, although only significant in the first year, seem to decline over time. As Frick et al. (2006) pointed out, such differences in the first wave of a panel with a decline in further years can also be an indication for a learning process and confidence building between the interviewer and the interviewee.

To really assess a possible bias of the high-income distribution in *SOEP*, a much larger survey with more "rich cases" would be required. A possible data source for such an analysis would be the 10% sample of all tax records ("Lohn- und Einkommensteuerstatistik") of the Federal Statistical Office in Germany, with a sample size of over 2.5 million cases. However, currently only data for the year 1998 are available and the next release which covers the year 2001, should be published in the second half of 2006.

The most demanding task for the *SOEP* is to keep the very rich people out of the non-response category, given the existing indications that high-income recipients are more likely not to answer income questions (see chapter 3) and item-nonresponse is often an indication for later unit-nonresponse.[175] The distributions of the older samples also seem to support these findings because most of the people in this rich group are pile up at the "lower" end of the wealth scale. But how to tackle this problem, which indeed confronts all surveys, currently remains altogether unclear and will therefore constitute one of the main challenges for future research.

[175] See Frick and Grabka (2004) and Frick and Grabka (2005)

Chapter 5

Regional Income Stratification - A Gini Decomposition Approach

The motivation for this chapter is twofold: first, it attempts to contribute to a fuller understanding of the persisting differences in economic performance between East and West Germany since the fall of the Berlin Wall. A second and equally important motivation is of a methodological nature: it employs a Gini decomposition as a means of analyzing stratification as reflected in the income distribution.

The difference in economic performance between East and West Germany can be operationalized by various relevant indicators such as unemployment rates, labor productivity and the dependency on public transfers, all of which impact on the income distribution in terms of both market income (from labor and capital) and disposable income. Such cross-regional variation in living standards is a highly relevant policy issue. According to the German constitution, economic and social policy should be targeted at diminishing regional differences in living circumstances. Various instruments and agreements between the federal government and the state governments (*Bundesländer*) are attempting to deal with this issue. A very important financial instrument is the "*Länderfinanzausgleich*". However, one should keep in mind that even aside from the current inequalities between East and West, regional variation in economic performance had a long history within West Germany alone. Agriculture, for example, was a crucially important economic sector in the state of *Bayern* for several decades after WW II, until this state's successful industrial modernization got underway. In the state of *Nordrhein-Westfalen*, on the other hand, the decline of the formerly successful monostructure of the mining and metal industries began to generate huge adjustment costs. These developments spurred extensive discussion of differences in economic performance between the North and the South of Germany (cf. Friedrichs et al., 1986; Geppert, 1999). This discussion has largely disappeared from the policy agenda since the fall of the Wall, and today, East-West comparisons dominate the debate on regional variation in Germany.[176]

From a micro-economic perspective, it is interesting to see the extent to which such regional variation is reflected in the personal distribution of income. This is why, for this empirical analysis, a Gini decomposition is applied for detecting stratification in a given society with respect to the distribution of income. In this special case, it is analyzed whether the observable regional differences in the income distribution in fact also mirror stratification of German society. Decomposing inequality in economic well-being requires additive inequality indices such as the Theil Index (Theil, 1967), but it has long been argued that one of the most commonly used indices for inequality analysis, the Gini index, cannot be adequately decomposed in an additive manner. However, using the covariance-based formula of the Gini coefficient, Lerman and Yitzhaki (1984); Yitzhaki and Lerman (1991) propose a decomposition approach which yields the obligatory between- and within-group components as well as an "overlapping" index for the different sup-populations. This is a very helpful tool for interpreting decomposition results with respect to income stratification. This method (Yitzhaki, 1994) is applied together with a jackknife estimation of confidence bands (Frick et al., 2006).

Welfare economists are interested mainly in the income distribution *after* government intervention, i.e. re-distribution after receipt of public transfers and after paying taxes and contributions to the social security system. However, given the massive monetary transfers from West to East Germany, it is important to find out the capacities for self-sustenance of the populations of East and West Germany, respectively. This also makes it important to look at the distribution of income *prior* to government intervention, i.e. market incomes stemming from both factor income (labor and capital) and private transfers (including private pensions).

One of the central findings is that the distribution of East German market incomes has changed drastically over the period under investigation, starting from a predictably low level in the early 1990s and rising in recent years to much higher levels of inequality compared to West Germany. The development of post-government income presents a different picture: here significantly lower and more equally distributed incomes in the East over the complete period can be find. Between-group inequality decreased over the first years of transition, but this process came to a halt in the mid-1990s. Overall, there is no convincing evidence of increasing regional convergence in post-government income levels and inequality. The question arises whether the policy goal of equalizing regional differences in income levels and income distribution is a realistic one, or whether regional stratification should simply be accepted as a currently unavoidable byproduct of economic evolution.

The chapter is structured as follows: After briefly discussing the literature on income distribution and regional income variation in Germany (Section 5.1), Section 5.2 describes the applied decomposition methodology and the data for the empirical analysis. The empirical application is given in Section 5.3, and the final section draws methodological and substantive conclusions.

[176] See Lammers (2003) for a discussion of an emerging North-South variation within East Germany.

5.1 Income Distribution in West and East Germany

Macroeconomic data, such as national accounts statistics, may enable comparisons of regional differences in absolute or per capita welfare levels, and are often used to analyze processes of regional convergence or divergence (e.g. in 1991, per capita GDP in East Germany was 33% of Western per capita GDP, and rose to "only" 63% in 2003). Analyses along this line looking at the economic performance of various German regions have been conducted by Geppert (1999) and Lammers (2003). But macroeconomic data does not provide an adequate foundation for analyzing trends in the regional variation of income inequality, while micro-data does. German databases that provide the basis for this kind of study include the EVS ("Income and Expenditure survey") and the German Socio-Economic Panel Study (SOEP) which also form the empirical basis of a huge body of literature on the evolution and distribution of pre- and post-government income (and its components) in West and East Germany.[177] The following is a summary of central findings:

Krause (2003) examines trends in income inequality and poverty dynamics in East and West Germany up to the year 2000, and finds an increase in East German inequality in the first half of the 1990s, a more stable picture in the mid-1990s, and a trend towards increasing inequality in both parts of Germany at the end of the century. Frick, Goebel, Grabka, Krause, Schäfer, Tucci and Wagner (2005) show that disposable income inequality in East Germany is consistently lower, but that market incomes - starting from a predictably low level of inequality at the beginning of the transition process - have been more unequally distributed in the East than in the West since the mid-1990s. According to Goebel et al. (2005), this picture is consistent whether the analysis is based on equivalized household pre-government income or individual labor income; in any case, the increase in inequality is driven by both increasing unemployment and widening wage dispersion. Bird, Frick and Wagner (1998) find evidence that the former GDR elites fared well over the first years of transition, maintaining an income advantage of about 10%. Bishop, Formby and Zeager (2001) show that in West Germany, low-income households (below the median income level) bore an above-average share of the costs of unification and the 1992-93 recession.

Focusing on market incomes and analyzing individual labor income, Hunt (2001) identifies rapid wage growth of more than 80% for East Germany over the period 1990-1996, with the biggest gainers being women and the better educated. According to Biewen (2001), the increase in income inequality in East Germany during the first half of the 1990s was due to rising unemployment, decreasing female labor market participation, and a widening income structure. Brenke (2005) stresses the relevance of differential changes in the demographic compositions of East and West German households since the fall of the Wall: East German households are, on average, shrinking faster with respect to household size due in particular to decreasing fertility and the consequentially declining

[177] See Becker, Frick, Grabka, Hauser, Krause and Wagner (2002) for a more detailed discussion of the impact of survey characteristics when comparing distribution results based on SOEP with those based on EVS due to the latter's quota sampling design, misrepresentation of foreigners, and non-coverage of top-income households.

share of families with dependent children. According to Brenke (2005), aging, together with increasing unemployment, is linked to the growing importance of (social) transfer income in East Germany.

Decomposing the Theil(0) inequality measure, Grabka, Schwarze and Wagner (1999) try to disentangle the effects which unification and migration exerted on the German pre- and post-government income distribution over the 1990s. They conclude that migration from East to West reduced overall German income inequality. Büchel and Frick (2001) analyze the participation of various population subgroups in the income redistribution process induced by the tax and transfer system during the mid-1990s. Comparing relative income positions before and after government intervention, they find that East Germans as a whole as well as specific immigrant groups significantly benefit from re-distribution.

While nearly all these analyses focus on differences between East and West Germany, Berthoud (2004) uses ECHP data to look at regional variation of income inequality and poverty across and within EU member states and their regions. For the regional differentiation, he refers to the level of NUTS1[178], which for Germany is defined by the 16 federal states or *Bundesländer*. An important finding from a German point of view is the very low degree of inequality - in cross-national terms - between regions: only 2.2% of overall inequality in Germany is attributed to between-region inequality, while this share is approximately 3 to 5 times higher in France, Spain, and Italy. These findings are in line with those presented by Stewart (2002), who uses LIS data to show that variability of poverty rates at German NUTS1-level around 1990 (West Germany only) is much lower than in Italy, France, Spain and the UK. However, between-region variability clearly increases when including East German federal states in the mid-1990s[179], a result which is confirmed by the EUROMOD-based analysis using 1998 income data from Mercader-Prats and Levy (2004). The latter also find a negative correlation between market income inequality and regional economic performance and consequently, regions showing weak performance will reap above-average gains from re-distribution.[180] Obviously the definition of regions is significant here, and looking at income inequality the county level, for example, "produces" much more between-regional variation than is the case at higher aggregated regional levels.

The findings of Loikkanen, Riihelä and Sullström (2003) for Finland demonstrate that redistribution by the welfare state induced by taxes and public transfers decreases regional variation and inequality. Surprisingly, the joint effect of the Finnish economic crisis of the early 1990s and the welfare state's redistribution did not become visible in the applied measures of regional differences. Förster, Jesuit and Smeeding (2002), who use LIS-data for four Central and Eastern European countries, reveal the extent to which intra-country inequality

[178] NUTS is the acronym for Nomenclature des unités territoriales statistiques.

[179] Based on log GDP per capita information for 110 regions in the EU-12 (applying a mix of NUTS-0, NUTS-1, and NUTS-2) Pittau (2005) identifies a convergence between poorer and richer European regions during the late 1970s and 1980s. In the mid-1990s however, a small group of very rich regions seems to have emerged, mostly large metropolitan areas including the German city-state of *Hamburg*.

[180] This is exemplified by the federal state of Sachsen-Anhalt in East Germany: this region occupies the 88th position (out of 100 regions) with respect to market income inequality, and is ranked 3rd after redistribution by the tax and transfer system (Mercader-Prats and Levy, 2004, p. 19).

Table 5.1: Grouping of federal states, population size and GDP

	East	West		
Grouping 1	Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen	Baden-Württemberg, Bayern, Hessen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Bremen, Hamburg, Niedersachsen, Schleswig-Holstein		
		South	Central	North
Extended Grouping 2	Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen	Baden-Württemberg, Bayern, Hessen	Nordrhein-Westfalen, Rheinland-Pfalz, Saarland	Bremen, Hamburg, Niedersachsen, Schleswig-Holstein
Population size (%)				
1991	22,4	34,2	27,9	15,5
2003	20,5	35,4	28,1	16,0
Gross Domestic Product (%)				
1991	11,0	41,4	30,2	17,3
2003	14,8	41,3	27,5	16,4

Source: Statistisches Bundesamt; own calculations.

is masked by national-level analyses. This appears especially important for those transition economies where Socialist central planning had often created regional concentrations of certain industries, possibly producing lasting regional disparities in macro-economic performance. The transition into more market-oriented structures may have revealed or accentuated this variation.

To target this kind of within-country variation, this chapter applies a new stratification method based on the decomposition of the Gini coefficient, which offers the advantage of producing three components: (1) the region-specific contribution to overall inequality in Germany, (2) the inter-regional contribution, and (3) overlapping information defined by the degree to which a given region's income distribution overlaps with the overall distribution (as well as with the distribution of any other region of interest). In order to give some indication of the sensitivity of inequality results with respect to the choice and number of regions, at first only two regions are defined, West and East Germany, (focusing on the current political debate) and than compared with the results obtained from a more diversified grouping of four regions by splitting West Germany into North, Central, and South (looking at the issue of a North-South divide). Certainly the choice of these regions is somewhat arbitrary, but it is driven by the interest in expanding public awareness in Germany beyond a purely West-East perspective to a broader view of regional variation in income levels and inequality (see also the recommendations by the SVR, 2004, cipher 617).

5.2 Emprical Analysis: methods and data

5.2.1 The ANOGI (Analysis of Gini) methodology

The ANOGI (ANalysis Of GIni) technique can be seen as the equivalent to ANOVA (ANalysis Of VAriance) performed with the Gini coefficient. To measure inequality, the Gini index is used as represented by the covariance formula according to Lerman and Yitzhaki (1984):

$$G = \frac{2cov(y, F(y))}{\mu} \tag{5.1}$$

Here, the Gini is twice the covariance between income y and rank $F(y)$ standardized by mean income μ.[181] The Gini of the entire population, G_u, can be decomposed as:

$$G_u = \sum_{i=1}^{n} s_i G_i O_i + G_b \tag{5.2}$$

where s_i denotes the share of income on overall income for group i, O_i is the overlapping index of the entire population by sub-population i (to be explained below), G_i represents the Gini of sub-population i, and G_b is the between-group inequality component.

The between-group inequality G_b as defined in Yitzhaki and Lerman (1991) is:

$$G_b = \frac{2cov(\mu_i, \bar{F}_{ui})}{\mu_u} \tag{5.3}$$

Hence G_b is twice the covariance between the mean income of each sub-population and the sub-populations' mean rank in the overall population, divided by overall expected income. That is, each sub-population is represented by its mean income, and the mean rank of its members in the overall distribution. The term G_b equals zero if either the mean incomes or the mean ranks are equal for all sub-populations. In extreme cases, G_b can be negative, which occurs when the mean income is negatively correlated with mean rank.[182]

The within-group inequality, $s_i G_i O_i$, therefore consists of three components (rather than only two, when decomposing other inequality measures or when applying ANOVA), of which the overlapping index is the most important for measuring stratification. The formal definition of the overlapping index is given by:

$$O_i = O_{ui} = \frac{cov_i(y, F_u(y))}{cov_i(y, F_i(y))} \tag{5.4}$$

[181] Note that the relative version of Gini is used here, which is most commonly used in the income distribution literature.

[182] For a more detailed discussion of between-group inequality and the relation to the overlapping index as well as alternative specifications of G_b, see Frick et al. (2006). See Dickey (2001) for an empirical application to earnings inequality in the UK using an alternative decomposition of the Gini-coefficient following Pyatt (1976).

where, for convenience, the index u is omitted and cov_i gives the covariance according to distribution i, i.e.

$$cov_i(y, F_u(y)) = \int (y - \mu)(F_u(y) - \bar{F}_{ui}) f_i(y) dy \tag{5.5}$$

where $\bar{F}_{ui}$ is the expected rank of sub-population i in the overall population (all observations of sub-population i are assigned their ranks within the union and $\bar{F}_{ui}$ represents the expected value).[183] Note that the numerator in (5.4) is the covariance between y and its rank, had it been ranked within the entire population, while in the denominator, the ranking is within sub-population i itself. The overlap as defined in (5.4) can be further decomposed to identify the overlapping of sub-population i and all sub-populations that comprise the union. In other words, total overlapping of sub-population i, that is O_i, is composed by the overlapping of all sub-populations (including group i itself) by group i. This further decomposition of O_i is:

$$\begin{aligned} O_i &= \sum_j p_j O_{ji} = p_i O_{ii} + \sum_{j \neq i} p_j O_{ji} \\ &= p_i + \sum_{j \neq i} p_j O_{ji} \quad , \end{aligned} \tag{5.6}$$

where $O_{ji} = \frac{cov_i(y, F_j(y))}{cov_i(y, F_i(y))}$ is the overlapping of group j by group i. From this follows that O_{ji} is equal to zero if no member of distribution j lies within the range of distribution i, which means that group i is a perfect stratum. On the other hand, if over the range of distribution i, the shape of the distribution of group j is similar to the shape of distribution i, then O_{ji} is equal to 1, and of course by definition, O_{ii} in any case is equal to 1. O_{ji} is bounded from above by 2. This maximum value will be reached if all observations belonging to distribution j that are located in the range of i are concentrated at the mean of distribution i.[184] O_{ji} and O_{ij} are connected, in a way that, generally spoken[185], the higher the overlapping index O_{ji}, the lower O_{ij} will be. That is, the more group j is included in the range of distribution i, the less distribution i is expected to be included in the range of j. Therefore O_{ji} is an index that measures the extent to which population j is included in the range of group i.[186]

The interpretation of the overlapping index is as the inverse of stratification. Stratification is a concept used mostly by sociologists, as used in the definition of Lasswell (1965):

> "In its general meaning, a stratum is a horizontal layer, usually thought of as between, above or below other such layers or strata.

[183] It is worth noting that the O_i is a kind of a Gini correlation. See Schechtman and Yitzhaki (1987, 1999) for the properties of Gini correlations.

[184] Note, however, that for a given distribution i the upper limit can be lower than 2 (for details see Schechtman, 2005).

[185] Note that the indices O_{ji} and O_{ij} are not inter-related by a simple relationship. However, it is clear that the indices of overlapping are not independent.

[186] A discussion of the estimation with grouped and weighted data is given in Lerman and Yitzhaki (1989), and for a description of the jackknife estimation see Yitzhaki (1991) and Frick et al. (2006).

> Stratification is the process of forming observable layers, or the state of being comprised of layers. Social stratification suggests a model in which the mass of society is constructed of layer upon layer of congealed population qualities."

According to Lasswell, perfect stratification is achieved when all observations of each population (in this application the population living in different German regions) are found in a specific range of income, and the ranges of the income distribution of the various sub-populations do not overlap. Yitzhaki (1982) links stratification to the concept of relative deprivation, arguing that more stratified societies can tolerate even greater inequalities than non-stratified ones.

One rarely finds perfect stratification in real life, and for this reason an index describing the degree of stratification is required. The index of overlapping is actually an index describing the extent to which the different sub-populations are stratified. For this analysis, this property plays an important role because it tells us whether East and West Germany (according to different groupings) represent different income strata even 14 years after unification.

5.2.2 The data

The empirical application makes use of representative micro-data for private households from the German Socio-Economic Panel Study (SOEP, cf. Wagner, Burkhauser and Behringer, 1993, and Haisken-DeNew and Frick, 2005). We analyze annual pre- and post-government income (previous calendar year) available for all years between 1992-2003 (actually representing the income distribution in the period 1991-2002 as gathered from the population living in the period 1992-2003).[187] Following the recommendations by the Canberra-Group (2001), the income measures include imputed rental values for owner-occupied housing (cf. Frick and Grabka, 2004). Given the multitude of income components incorporated in the aggregated pre- and post-government income measures (factor income from labor and capital, public and private transfers, public and private pensions, etc.) these income constructs are both rather complex to generate; this is especially true for the simulation of direct taxes and social security contributions. All income measures are corrected for missing data due to item-non-response by means of longitudinal and cross-sectional imputation (cf. Frick and Grabka, 2005).

In order to adjust income for differences in household size and age composition, a common international equivalence scale, the modified OECD scale is applied.[188] All income measures are deflated to prices of 2000 including a correction for purchasing power differences between West and East Germany.

[187] Income measures for 1990 and 1991 are not included in this analysis due to the introduction of the common currency in 1990 and comparability problems of East and West German incomes immediately after unification (cf. Hauser, Frick, Mueller and Wagner, 1994).

[188] The modified OECD equivalence scale gives a weight of 1 to the household head, a weight of 0.5 to other adult household members above age 14, and a weight of 0.3 to all children up to 14 years of age.

5.3 Empirical Results

This section provides empirical results on the decomposition of the Gini coefficient for annual pre- and post-government income measures for different German regions (East and West Germany, the latter also being split into North, Central and South).

With reference to the theoretical considerations in Section 5.2.1 on the ANOGI methodology, any significant variation between regions (here: West and East Germany) can be interpreted as an indication of stratification. In other words, no regional stratification is given if all parameters of interest were the same for all regions (i.e. no statistically significant differences apply):

Mean income:	$\mu_{West} = \mu_{East}$
Mean rank:	$F_{West} = F_{East} = 0.5$
Gini coefficient:	$G_{West} = G_{East}$
Overlapping index:	$O_{West} = O_{East} = 1$
Between-group inequality:	$G_b = G_{bp} = 0$

Based on the heavy transfers from West to East Germany, the baseline hypothesis must be that income distribution differentials which may have existed when the Berlin Wall fell diminish over time and eventually disappear. The empirical analysis will show that this is not true at all for post-government income and that it is only true for the overlapping of the pre-government income, because the shape of this income distribution in East Germany developed in a rather specific way.

For the sake of illustration[189], the results of the Gini-decomposition (ANOGI) are presented as time series by groups in graphical form using separate figures for

(a) Mean income (μ_i),
(b) Gini index (G_i) with between and within inequality shares ($\frac{G_b}{G_u}$) and ($\frac{1-G_b}{G_u}$), respectively,
(c) Overlapping component (O_i), and
(d) Mean rank (F_i)

Confidence bands[190] are also indicated for the group-specific Gini, the shares of between and within inequality as well as for the overlapping index.

5.3.1 Income inequality decomposition by region: West and East Germany

5.3.1.1 Pre-government income

Pre-government income levels in West Germany generally mirror the development of the business cycle[191] (see Figure 5.1(a)). Over the whole period, inequality,

[189] All results are available in tabulated form on request.

[190] The confidence bands shown are defined by two times the respective standard errors, based on a jack-knife procedure.

[191] Burkhauser, Cutts, Daly and Jenkins (1999) argue that when comparing time trends on inequality measures, one needs to properly consider the state of the business cycle, i.e., one should compare "peak to peak" and "trough to trough" years.

G_i, increases at a moderate pace, but again in line with the business cycle; i.e. there are years that even show a minor decrease in inequality. As is to be expected, pre-government income inequality in East Germany in the early years of transition was significantly lower than in the West, however, inequality steadily increased and passed the West German level as early as the mid-1990s (see also Biewen, 2001; Hunt, 2001). Market income inequality in the East is still rising at a clearly higher pace in more recent years (see Figure 5.1(b))[192]. This process is mainly driven by massive (increasingly long-term) unemployment (see Frick et al., 2005). East German pre-government income *levels* (as measured by mean and ranks) cannot close the gap to West Germany; again the process of catching up had already stopped in 1995 and mean ranks, F_i, (see Figure 5.1(d)) have remained very stable at about 0.41 for almost 10 years.

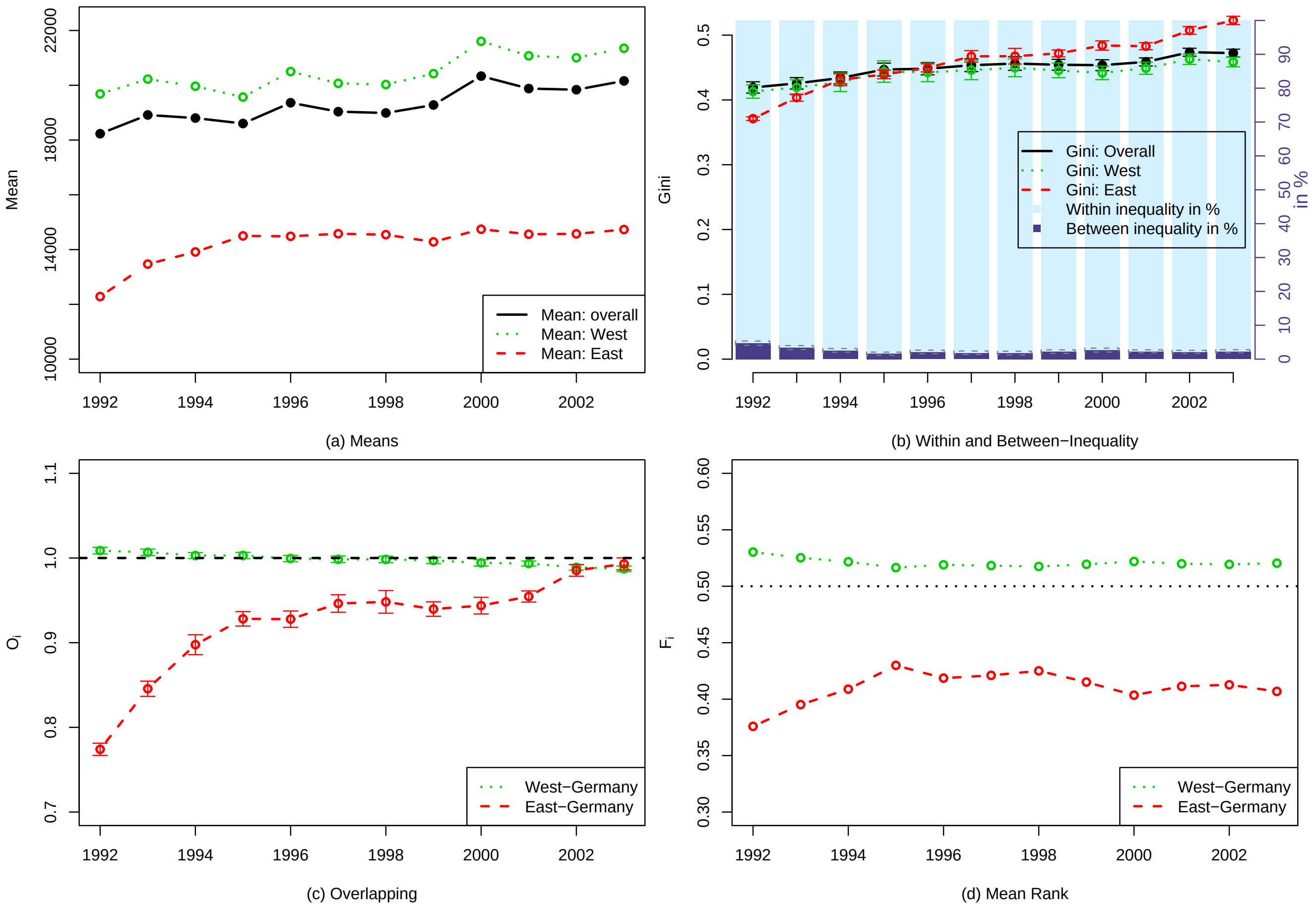
The contribution of between-region *inequality*, G_b, was significantly reduced in those early years (1992: 4.7%) and reached its minimum in the mid-90s (1995: 2.0%) when pre-government income inequality in the East matched that in the Western part of the country. However, the between-group contribution has increased slightly in more recent years in line with East Germany's skyrocketing inequality (2000: 2.7%). According to the overlapping indicator, O_i, East Germany was clearly a pre-government income stratum on its own over the first 10 years of the unification process. In 1992, the O_i for East Germany was as low as 0.7741 but rapidly developed to 0.9282 in 1995. Since then, although at a lower pace, this measure further approached the value of one, and in 2003, the pre-government income distribution in both parts of Germany overlapped almost perfectly ($O_{West} \approx .9873$ versus $O_{East} \approx .9930$, see Figure 5.1(c)).[193] This result, however, must be interpreted together with the consistently lower income levels and increasing income inequality in East Germany: i.e., those East Germans who do have a paid job (which is by far the most important source of pre-government income) "reach into" the West German distribution. However, a large group of East German individuals have very low or even zero market incomes[194] as well due to unemployment or early retirement schemes.

[192] In 1992, the East German Gini was .3711 as compared to .4129 in West Germany. In 2003, the corresponding values were .5227 and .4584, respectively.

[193] In this decomposition for only two groups, the presenting of the results for the group-by-group overlapping index, O_{ji} is ommited, given that O_i qualitatively resembles O_{ji}. Note that O_i is the weighted sum of the group-specific O_{ji} with O_{ii} being equal to one.

[194] See Figure C.1 (Appendix) for kernel density estimations of pre- and post-government income distributions by region for 1992, 1997 and 2003. For the case of pre-government income, it can clearly be seen that the share of zero incomes in East Germany increases dramatically after unification.

Figure 5.1: ANOGI-Results on pre-government income: East vs. West Germany, 1992-2003



(a) Means

(b) Within and Between−Inequality

(c) Overlapping

(d) Mean Rank

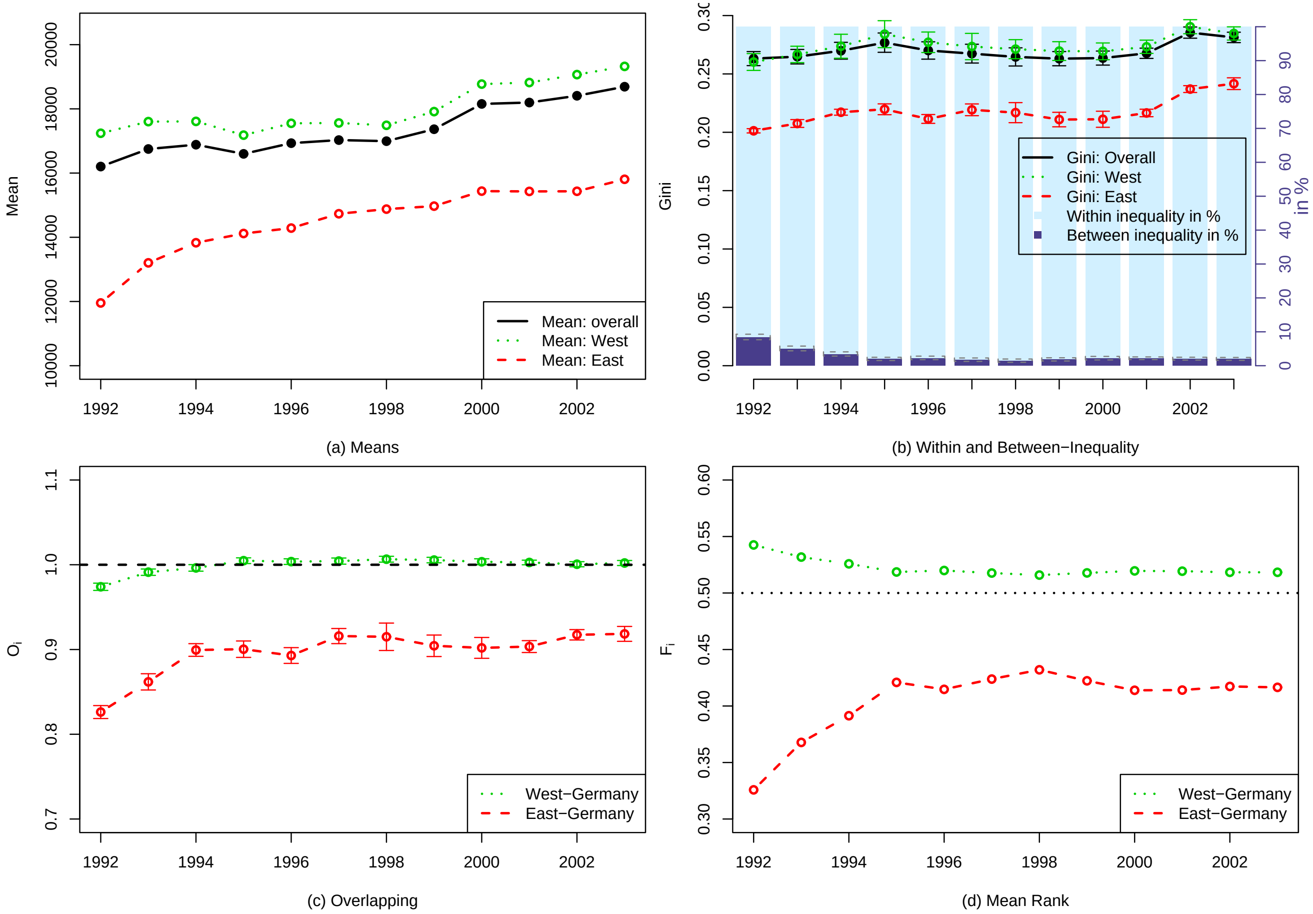
5.3.1.2 Post-government income

Post-government income levels in East Germany increased significantly over the first half of the 1990s, steadily closing the gap to the Western levels (see Figure 5.2). However, as could be observed for pre-government income, this process came to a halt around 1995. Inequality, as measured by the Gini coefficient, G_i, remained consistently and significantly lower in the East compared to the West. This process mirrored the development of the business cycle, although to a less pronounced degree than was the case for pre-government income. This merely reflects the fact that public transfers are effectively performing their stabilizing function, especially the unemployment assistance schemes which appear to be more important in East Germany given the extraordinarily high unemployment rates there (almost 20% and as such about twice as high as in the West). For West Germany, a mildly u-shaped trajectory in the inequality development since 1995 ($G_i = .2841$) is find, with another local maximum reached in 2002 ($G_i = .2904$). The decrease found here in recent years does not appear in the East, resulting in a somewhat narrowed regional inequality gap in 2003, but the difference remains statistically significant ($G_{West} = .2847$ versus $G_{East} = .2416$).

In line with these results, mean ranks do not show any relevant changes since 1995 in West and East Germany ($F_{East} \approx .42$). The overlapping index identifies East Germany to remain significantly different throughout the period under investigation, i.e., the East still forms an income stratum on its own (2003: $O_{East} \approx .9184$).

Concluding from the findings in Section 5.3.1, one can derive that the German welfare state continuously and significantly reduces market-induced inequality in a very effective way. Although for pre-government income, the overlapping index indicates that the market-induced income distributions in East and West now overlap because of the high degree of inequality in East Germany, this is not at all true of disposable or post-government income.

Figure 5.2: ANOGI-Results on post-government income: East vs. West Germany, 1992-2003



5.3.2 Income inequality decomposition using an extended regional grouping

In a second step the differentiation is extended to allow for more regional variation within West Germany. One may argue that the differences between East and West Germany do not come as a surprise, since such differences also arise within West Germany alone if it is divided in an appropriate way.[195] Certainly any such regional grouping is based on some normative decisions. Given the federal organization of Germany on the one hand and the availability of data at the federal state level as well as the identification of these regional entities in the data on the other hand, the grouping chosen in this chapter is based on federal states (NUTS1-level). In the context of the discussion about a "North-South divide" within West Germany, the grouping of the federal states was done into northern, central, and southern states (see Table 5.1 above).

Using this extended grouping, the substantive group-specific results described in Section 5.3.1 above for East Germany (Gini G_i, mean income and rank F_i, and overlapping O_i) will remain unchanged, while the results for West Germany will now be derived from three measures: the northern, central, and southern parts of West Germany. However, the contribution of between-group inequality, G_b, as well as the group-by-group overlapping index, O_{ji}, may very well be subject to change. It is not only relevant to find out the degree to which these three West German regions deviate from each other, but also to see whether the East German results come closer to at least oney of the western regions.. If this were the case, then the hypothesis of East Germany forming an income stratum on its own would be falsified.[196]

5.3.2.1 Pre-government income

With only one exception in the year 2000, there is a consistent picture of pre-government incomes being higher on average in the southern part of West Germany than in the central and the northern federal states, which is perfectly in line with the discussion about the North-South divide (see Figure 5.3). According to mean rank, all three western regions show above-average values throughout the entire observation period, although in most recent years FSouth improved, while it clearly worsened for the northern and central groups. Despite of this development, the average East German income still falls far short of the lowest of these three reference values. It should be noted that this overall development at the micro-level perfectly matches the regionally disaggregated macro-information on GDP as given in Table 5.1.

There is no clear trend with respect to market income inequality across the West German groups - all of them remain rather close and it is only in the early 1990s and during the very recent years that the South has shown

[195] According to Stewart (2002) between-region variation of the poverty-rate in West-Germany as measured at NUTS1-level (federal states) is rather low, especially when compared to other large EU-countries.

[196] For sake of clarity of the presentation, all figures in this section do not show results for Germany as a whole, which are given in Section 5.3.1 above. By definition, these do not change with the number of groups distinguished.

significantly lower inequality than the central and northern regions. But as was true when comparing the eastern result to West Germany overall, East Germans inequality in pre-government income is lower in the early years of transition and significantly higher since the late 1990s. In 2003, G_{East} reached .5227 and the "closest" Western value was given by the northern region with $G_{North} = .4795$. This finding is confirmed by the fact that between-group inequality does not significantly change when using four rather than only two regions for the decomposition analysis.

With respect to the overlapping index, O_i, the conclusion is that, starting in 2002, the distribution in East Germany began extending into the range of the corresponding West German O_i. Nevertheless, the mean rank in the East has remained significantly lower and the increase in inequality has accelerated in recent years.

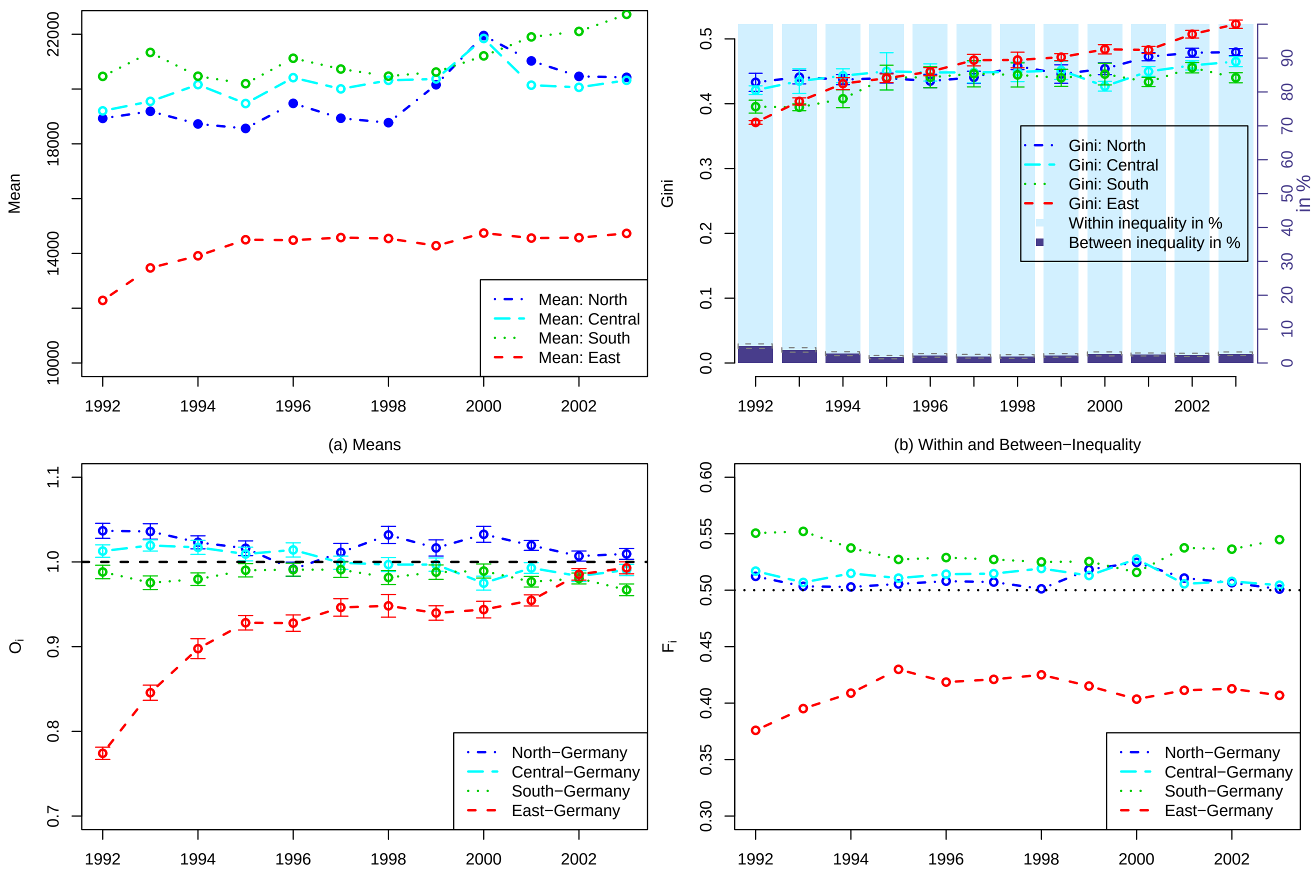
Figure 5.4 presents the decomposition results with respect to the overlapping indices for each group in terms of the respective other groups, namely O_{ji}.[197] In contrast to the above mentioned results for O_i, where each group with the entire population is compared, such a group-by-group comparison is not affected by the relative size of the various groups.[198] Following this consideration, Figure 5.4 includes for each of the regions considered in the analysis (North, Central, South, East) the corresponding overlapping indices with the respective three other regions.

Throughout the entire period under investigation, East Germany (Figure 5.4(d)) formed a distinct income stratum with respect to all three western regions, except for last year, when a significant deviation was found only in comparison to the North. None of the other regions formed an income group with respect to the East in the first years following unification. For the South, this changed starting in 1996 (Figure 5.4(c)), for the Central part in 1998 (Figure 5.4(b)), and finally for the Northern part as well in 2002 (Figure 5.4(a)). Since then each of the other regions has shown a distinctively different distribution from the East German distribution. There is a more heterogeneous picture within West Germany: While over the early 1990s, only South Germany formed a group with respect to North and Central, a convergence can be observed during the mid-1990s, a period with less inequality and more similar mean ranks among the three western regions. Starting in 1998, however, the overlapping results indicate that the South and Central regions also form distinct income groups with respect to the northern part of the country. This may be taken as an indication of a North-South divide.

[197] Results for O_{ji} are presented as time series showing statistically significant deviations from 1 as solid dots, otherwise no annually value is depicted (broken line).

[198] Note that the overlapping index for two groups i and j may not be symmetrical (see Section 5.2).

Figure 5.3: ANOGI-Results on pre-government income: Extended regional grouping, 1992-2003



(a) Means

(b) Within and Between−Inequality

(c) Overlapping

(d) Mean Rank

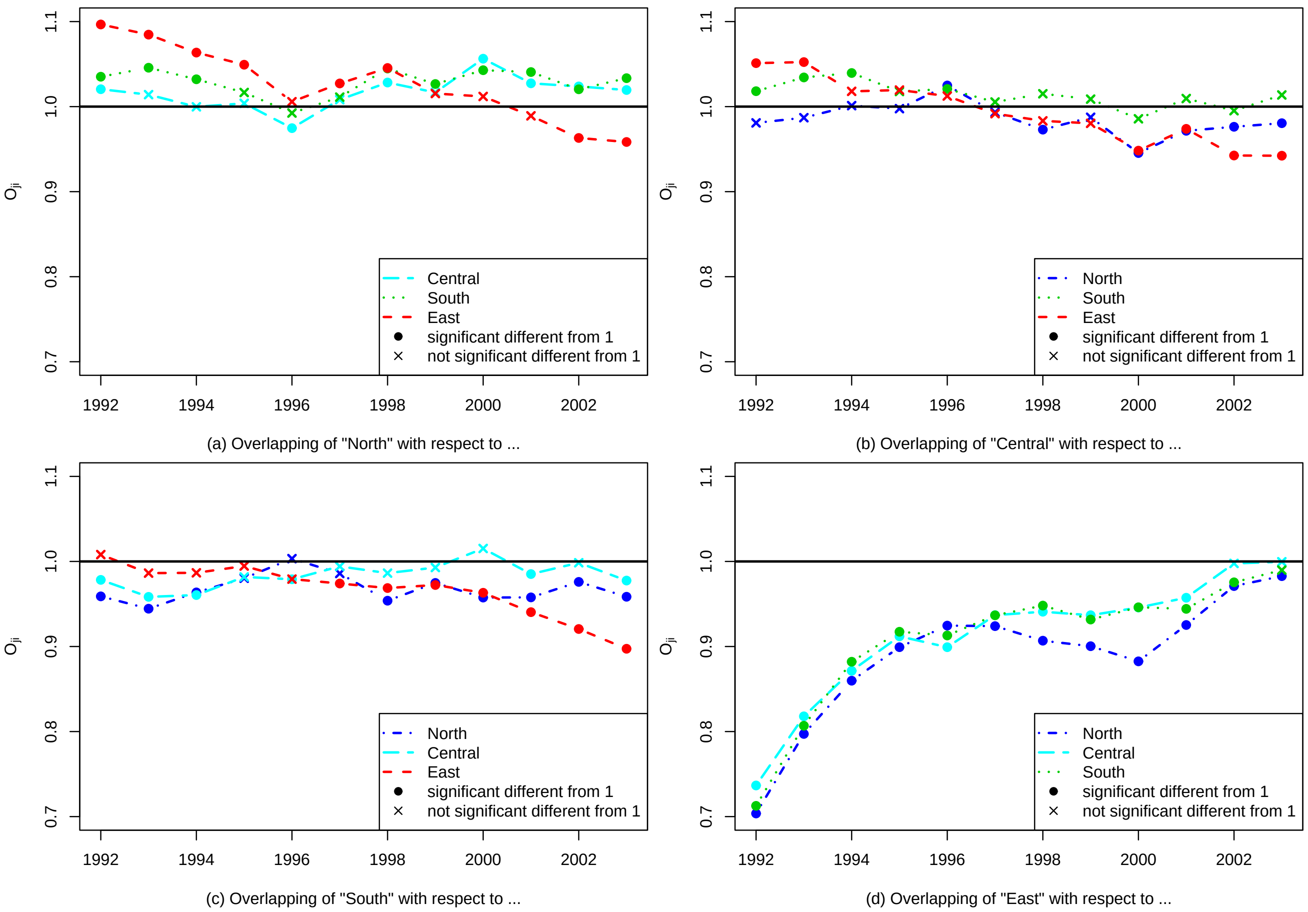


Figure 5.4: Overlapping for each group with each other group (O_{ji}): Pre-Government Income, 1992-2003

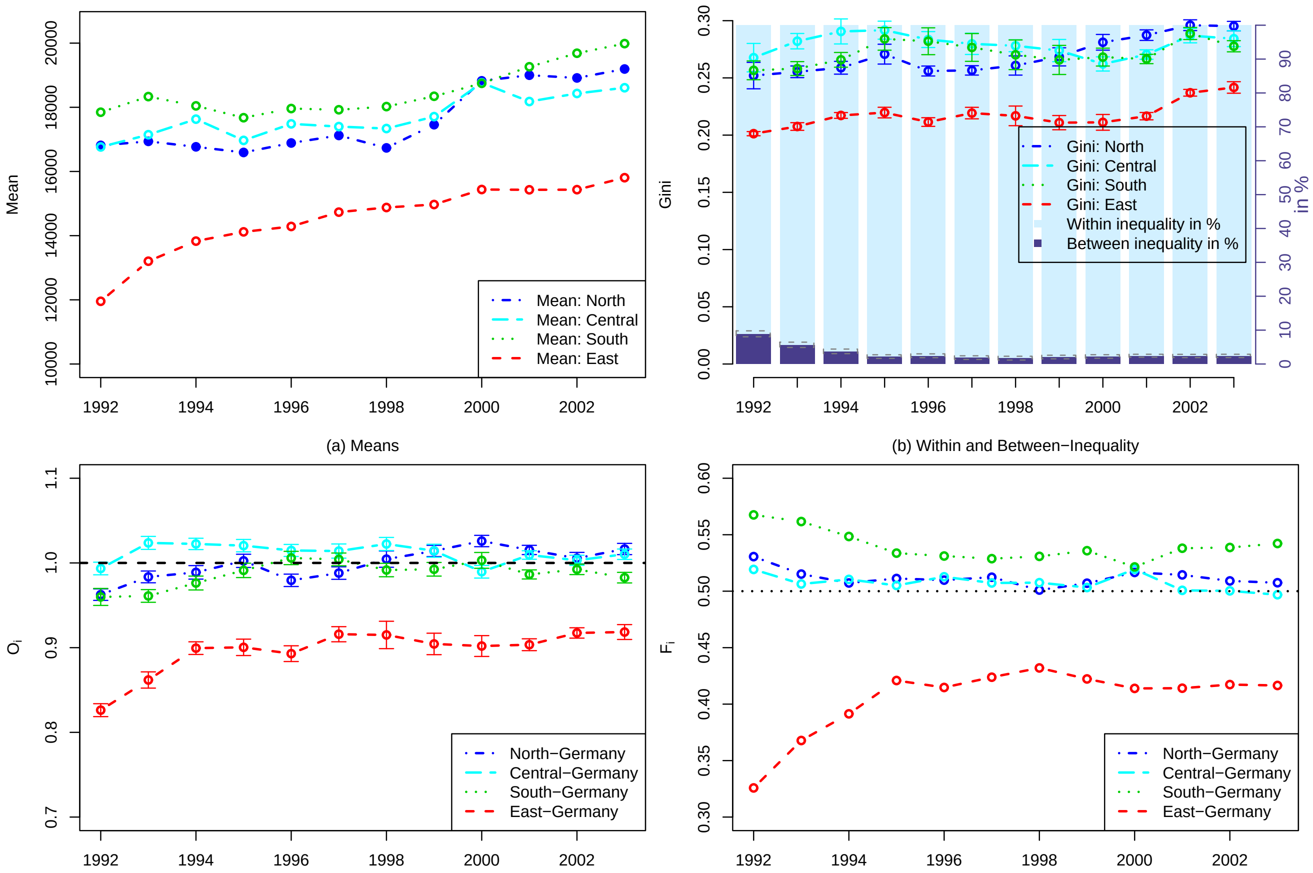
5.3.2.2 Post-government income

Given the results on pre-government income, it comes as no surprise that a more diversified regional grouping in West Germany also does not yield significant changes (for the West-East comparison) when the dependent variable is post-government income (see Figure 5.5). Income levels in the southern part of West Germany are in principle higher than in the central and northern parts, and all of them are clearly above the average eastern income. This is also confirmed by the mean rank, F_i. The ordering of West German regions with respect to income inequality (G_i) changed in the late 1990s, when the North became the region with highest inequality after being below average for the first half of the period under investigation. The more important finding is again that all group-specific decomposition components - inequality, average income, mean ranks, overlapping - for the East German disposable income remains far below any of the three West German regions.

Even with the more differentiated grouping of West Germany, the overlapping index with the overall distribution, O_i (Figure 5.5(c)), as well as the group specific overlapping, O_{ji} (see Figure 5.6(d)) for East Germany remains significantly different from all three reference regions. And again, the conclusion of the above empirical analysis is, that with respect to regional stratification, the East still forms an income stratum on its own.

The overlapping indices for each group in terms of the respective other groups (Figure 5.6) from a western point of view indicate that only in the very first years after unification did the western regions form distinct income strata with respect to the East. However, this changed rather soon and since 2000 only the South again forms a group with respect to the East. Within West Germany, there is a much more homogeneous picture when using the group specific overlapping indices. The central part does not form a group at all over the whole period, and the northern part was only a group with respect to Central Germany in the first half of the period under investigation. Solely the South seems to become more stratified with respect to the North in the more recent years after a process of assimilation during the mid-90s, in line with the pre-government income results presented in Section 5.3.2.1.

Figure 5.5: ANOGI-Results on post-government income: Extended regional grouping, 1992-2003



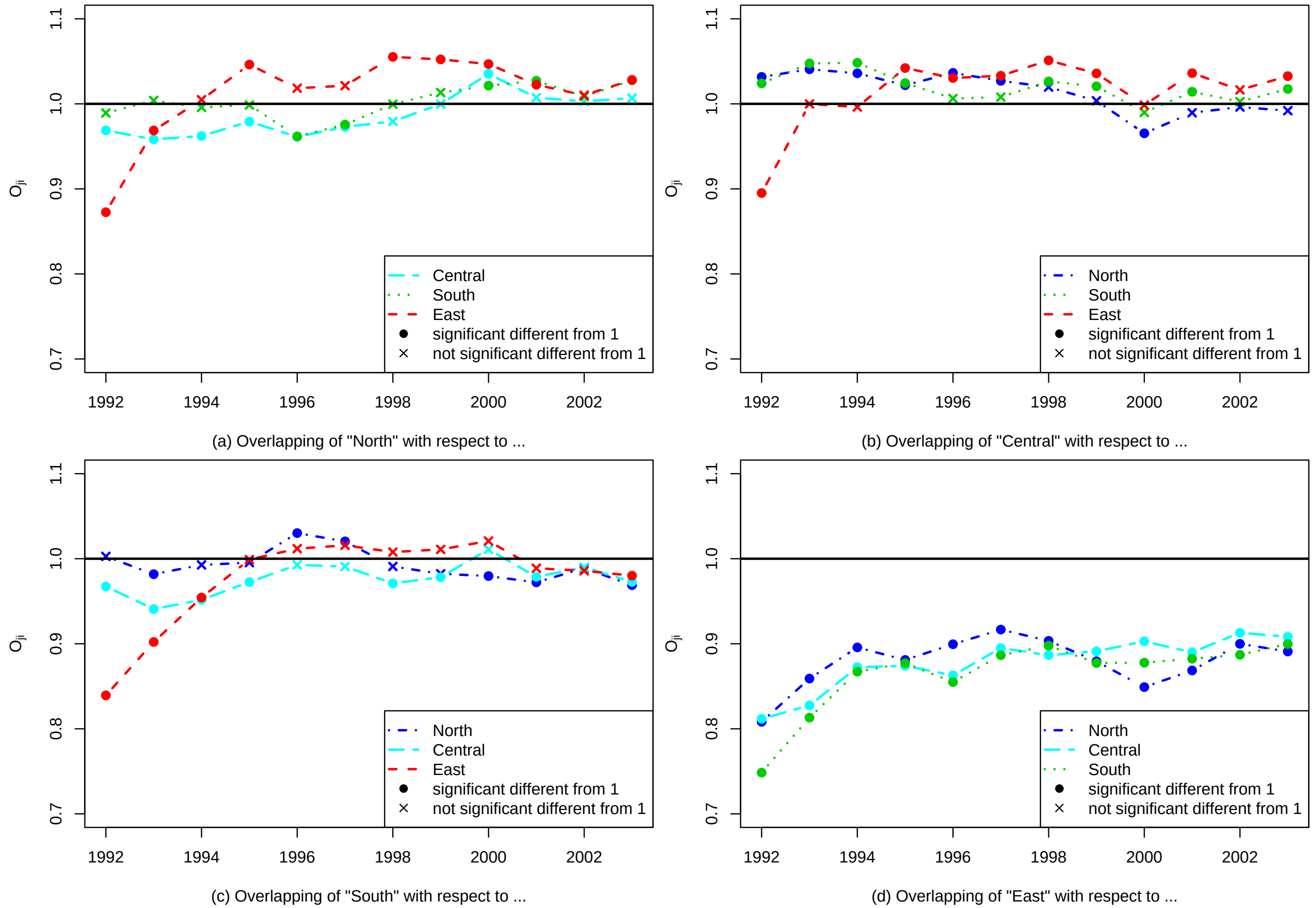
(a) Means

(b) Within and Between−Inequality

(c) Overlapping

(d) Mean Rank

Figure 5.6: Overlapping for each group with each other group (O_{ji}): Post-Government Income, 1992-2003



(a) Overlapping of "North" with respect to ...

(b) Overlapping of "Central" with respect to ...

(c) Overlapping of "South" with respect to ...

(d) Overlapping of "East" with respect to ...

5.4 Summary

The unique advantage of ANOGI is that it provides an additional term which reflects the overlap between the distributions of two or more interesting groups or strata formed by various German regions.

Concluding from the empirical results with respect to post-government income, one must reject the hypothesis that East and West Germany are moving towards a common income distribution.[199] After a "promising" start over the first half of the 1990s with increasing levels of income among East Germans, but accompanied by rising inequality, this process appears to have stopped in the mid-1990s without major changes since then. The picture is quite different for pre-government incomes, which are heavily dominated by labor income for East Germany, while for the West German population, capital gains are a more relevant issue. Mostly driven by massive unemployment and the lack of counteracting capital income, market-income inequality in East Germany already surpassed the Western level in the early 1990s and this difference has continuously increased. The huge inequality of market incomes in East Germany results in East Germany no longer being a stratum on its own with respect to the overlapping of pre-government incomes: very low (zero) as well as (some few) remarkably high market incomes yield an income distribution overlap with that in West Germany. However, the average East German market incomes (as well as the respective mean rank) are still far lower than in West Germany.

Enlarging the number and structure of the regions under consideration by splitting the western part into its northern, central and southern components also reveals a certain degree of regional variation *within* West Germany, with the South being in a somewhat more favorable position with respect to market and disposable income. If regions in West Germany (in particular the South) create income strata on their own at all, then only at a much smaller scale and not persistently over time. There is, however, a clear picture of East Germany still being quite different from the rest of the country, irrespective of any western regional grouping.[200]

Overall, clear indications of post-government income stratification are found.[201] On the one hand, this may be taken as underlining the need for a continuation of transfers from West to East in the context of the new *Solidarity Pact II*, which just started in 2005. However, instead of arguing about the need to counter

[199] As such, the results provide reason for disappointment among those who wish to see Willy Brandt's message come true: "*now what belongs together will grow together*" (Original quote: "*Jetzt wächst zusammen, was zusammen gehört.*") Commentary about the fall of the Berlin Wall by Willy Brandt, German chancellor 1969-1974 and mayor of Berlin, on November 10, 1989 .

[200] These findings are perfectly in line with Colavecchio, Curranand and Funke (2005) who analyze GDP per capita at the county level ("Kreis", n=439) derived from aggregated statistics by the Arbeitskreis VGR. They are also able to identify variation within East Germany, however: in 2001 about 80% of Eastern counties still belonged to the poorest of three income categories, while only 8% could be found in the richest category. The results are based on micro-data with a finite number of observations, which cannot deal with such a high level of regional disaggregation.

[201] It is not clear at this point to what extent these processes are influenced by regional mobility of East Germans moving to the western part of the country and vice versa - however, given tendency towards selective mobility, this issue may be taken up in future extension of this chapter.

any existing differences with even more and higher transfers, politicians and the public may have to start discussing whether one should become more willing to accept regional differences which might become the basis for endogenous growth in the less advantaged regions. This process is not limited solely to the East-West discussion, since there has been a tendency in recent years toward pre-government income stratification in South (West) Germany as compared to both other Western regions, as well as of Central (West) Germany as compared to the Northern part.

Conclusion and Discussion

The main focus of this thesis is on the important methodological problems that arise in the field of poverty and income analysis when using empirical data from large surveys. The measurement of poverty was addressed in two Chapters, one on the aggregation problem of poverty and one on the measurement of poverty over time under consideration of possible income mobility. Issues of the quality of data for empirical analysis in a wider sense were dealt with as well in two Chapters, the first regarding the robustness of imputation methods and second regarding the problems of correctly measuring high income households. The last Chapter presented an empirical application of the decomposition of inequality as measured with the Gini coefficient by German regions.

Foster (1984) states that 'the choice of a *single* poverty measure involves a certain degree of arbitrariness'. Chapter 1 of the present work, however, was intended as an overview and clarification of the extensive literature on poverty measurement since the seminal work of Sen (1976), not least because it is important to know the underlying axioms and the interrelationships among them to fully understand each measure. Based on the set of core axioms presented in Chapter 1 and an overview of all axioms fulfilled by each measure shown in table 1.1, the reader thus has at his disposal adequate information to make an informed and an appropriate decision.

Thirty years since the work of Sen, it is widely recognized that a measure of poverty intensity needs to be distribution sensitive in addition to simply counting the poor and calculating the average shortfall from the poverty line. However, many of the existing measures are difficult to interpret, which limits their accessibility for applied poverty research and their practical use for policy-makers. Among the more advanced poverty measures, the only one which has been able to transcend this barrier was that proposed by Foster et al. (1984), which can be understood very intuitively and has a close relationship to the well-known poverty head-count ratio. This is, at least in part, due to the intuitive weighting function of the poverty gaps by themselves, not via a derived welfare function.

However, focusing on a single measure sometimes distracts from the fact that for some more specialized questions of poverty, other measures may be more applicable. For instance the decomposition by source of the change in poverty over time can be better accomplished with the modifications of the Sen measures by Shorrocks and Thon.[202] However, poverty is a very complex phenomenon whose measurement should not be oversimplified or restricted, and many specialized measures may needed to answer the many different research questions. Especially more welfare-oriented research on poverty may be better accomplished with an ethical poverty measure, where the social evaluation function is stated explicitly and the interpretation is based on the loss of welfare due to poverty.

While Chapter 1 discussed currently available measures for the aggregation of poverty, all of which ignore a possible time dimension, Chapter 2 proposed a new way of measuring the impact of income mobility over time on measured poverty. It was shown here that Shorrocks' approach (Shorrocks, 1978) to measuring the stability of income inequality is also suitable for the measurement of poverty stability with a convex poverty measure, e.g. the FGT with a poverty aversion parameter of $\alpha > 1$. This makes it possible to enhance the dynamic description of poverty, because the extent of persistent and transitory poverty can be estimated with the use of longitudinal datasets. With a transformation of current income, in a form that expresses the shortfall from the current poverty line as a percentage, median-based poverty lines can be used as well. A further advantage of this approach to measure permanent poverty is that Shorrocks' 'framework' is well-known, can be interpreted easily, and is sensitive to the amount of mobility within the population.

In the empirical analysis, the robustness for mean and median-based poverty lines as well as for monthly and annual income were presented. Whereas the income concept does not have any significant effect, the reduction of poverty stability with a mean-based poverty line is higher during the first years of a period and lower thereafter.

With the help of this concept, the differences in the consequences of mobility on poverty status could be described for the USA and Germany. It becomes clear that the kind of income used is crucial when comparing two countries that differ so widely, especially with respect to the characteristics of the welfare state. Whereas poverty stability is much higher in the USA than in Germany for post-government income, this changes dramatically when looking at pre-government income. For market income, almost no differences between the two countries were found, which shows the strong impact of redistribution in the German welfare system. Only during more recent years could a slightly higher stability within Germany be observed, which could be seen as an indication that long-term unemployment affects the stability of poverty. For Germany in particular, further research should incorporate a regional distinction, such as the one used in Chapter 5 to disentangle the possible impact of the varying unemployment by region.[203]

Every empirical analysis relies heavily on the quality of the underlying data, regardless how carefully crafted its theoretical basis. One important aspect is unit or item non-response, i.e. where either entire cases are missing or where some

[202] For a description of this decomposition see Xu and Osberg (2001).

[203] The regional grouping used in Chapter 5 was between West and East Germany and in a further step a more diversified grouping within West Germany (North, Central and South) was analyzed.

information on some cases is missing. For unit non-response, the data producer often makes corrections with the help of an appropriate weighting scheme. But the problem of item non-response is often left up to the user. Handling missing data is still far from simple, however, and none of the software packages used in Chapter 3 deal with this problem automatically without serious intervention by the user and adaptation to his or her specific research question. At the same time, limiting the analysis to the complete cases is in most circumstances also not appropriate, especially with large-scale survey data.[204]

Similar problems can be seen in the structure of item-nonresponse on the household income question in the Finnish part of the ECHP. The income distribution of the non-respondent is very different from the income distribution of those persons who were willing to answer (cf. the estimated kernel densities of the respective distributions in figure 3.10). The assumption of missing completely at random (MCAR) seems highly unreasonable in this case. Even for the estimation of some easy parameters of this distribution, the complete case analysis leads to a serious bias, as shown in section 3.7. Therefore the need for an appropriate imputation strategy is obvious in the present case. Note that even for a dataset where it is reasonable to assume MCAR with respect to one variable, imputation should be considered given that the loss of cases in a multivariate analysis could lead to a severe loss of power. Also the multivariate structure of the missing mechanism can be not MCAR for all variables under investigation, and assuming MAR in most cases also requires imputation.

Fortunately, the Finnish survey includes additional register information (served as 'true' incomes) from outside the survey. This unique situation was utilized to evaluate the ability of different imputation methods to minimize this observed bias. To avoid a mixture with error in variables, only true income was used, and for cases with item-nonresponse on the household income question, this value was deleted. The results were compared using six different imputation programs: hot-deck (Stata), MVIS (Stata), NORM (R/S-Plus, standalone), aregImpute (R), MICE (R/S-PLus), IVEWare (SAS, standalone). These software packages cover a wide range of currently available imputation methods as well as statistical analysis packages.

From the results of the empirical evaluation, two main conclusions can be drawn. First, imputing a rather complex variable within a large survey cannot be carried out properly with a simple imputation method like hot-deck imputation as incorporated in a Stata ado file.[205] For all evaluation criteria, this method performed worse and even introduced additional bias when compared to a simple complete case analysis. Second, even in the relatively 'comfortable' situation where the truth about all missing values is known, the decision as to which method is most appropriate depends heavily on the criterion used, and often there is no one best imputation method for all situations.[206]

Overall, it seems that the more flexible the imputation software is, and therefore adjustable to the specific needs of the imputation task at hand, the

[204] In particular because most surveys have been showing rising unit- and item-nonresponse in recent years, at least in Germany (Frick et al., 2006)

[205] However, this does not mean that more advanced hot-deck methods are incapable of imputing complex problems. Good results for a hot-deck based imputation were reported by Durrant (2005), who programmed some sophisticated hot-deck imputation methods in R.

[206] Williams and Bailey (1996) are also unable to recommend one solution for all imputation tasks.

more robust the imputations with respect to the introduction of an additional bias. The NORM software, in particular, which assumes a multivariate normal distribution of the data, is too restricted for a complex survey. But also with IVEware, a very comprehensive imputation package, some problematic situations appear. Whereas the estimation of the mean and Gini were best with this software, the distribution preservation and especially the poverty estimation were biased.

The most robust imputation was achieved with the help of the MICE package. Here, the distribution of income was significantly improved when compared to the complete case analysis. However, the chosen parameters for evaluation were improved noticeably only for the poverty estimation. In the case of the Gini and the mean, no real difference from the complete case analysis could be observed.

The conclusion is therefore that multi-method strategies based on regression models perform best when dealing with a complex survey. Further research should definitely test the impact of including longitudinal information, which normally improves the estimation of the imputed values considerably (Spiess and Goebel, 2004, 2005). But how this can be done accurately with the household as the analysis level is an open question, because of possible instability of households over time. Further research should therefore analyze the changes within households over time in depth, identify which events are most crucial for income mobility, and describe how this can be accurately modeled in a more dynamic imputation setting.

Chapter 4 analyzed the impact of the integration of a special high-income sample (sample G) into an ongoing panel survey, the *SOEP* (samples A-F), which leads to an oversampling of households with high incomes and enables researchers more in-depth analysis of the socio-economic structure of this group. When comparing the overall differences due to the inclusion of a special high-income sample, only minor changes were found. The inclusion of sample G into the *SOEP* survey does not lead to any remarkable changes in the overall distribution, and only very top-sensitive measures like Theil (1) are different in the first year of the inclusion of sample G. Therefore it can be said that the estimation of the *overall* distribution with the *SOEP* seem to be good, even before the inclusion of a high-income sample.

A comparison of samples A-F with sample G resulted in some remarkable differences, when only the high income range (more than 200% of mean income) was considered, and this is even more true for the very high incomes (more than 300%). However, most of these differences observed, initially, when only comparing point estimates based directly on the empirical data, do not hold when using more appropriate model-based methods and taking statistical sample properties and the sparse data situation into account.

For the distribution of the high incomes, it can be stated that for both samples, a Pareto power law distribution seems appropriate. The estimated Pareto distributions fit well, regardless of the year and the sample under investigation. However, there are clear indications that the inequality within the high income region is higher in sample G than in the older samples A-F, and thus a lower shape parameter of the Pareto distribution. These differences, which are only significant in the first year, seem to decline over time. One reason is that the higher the income, the lower the probability of drawing such cases, given the very low number of high-income households. And as Frick et al. (2006) pointed out, differences in the income distribution in the first wave of a panel that decline in further years can also indicate a learning process regarding the questionnaire and confidence-building between the interviewer and the interviewee.

To really assess a possible bias of the high-income distribution within the *SOEP* a much larger survey with more 'rich cases' is required. A possible data source for this analysis could be the 10% sample of all tax records ("Lohn- und Einkommensteuerstatistik") of the Federal Statistical Office in Germany, with a sample size of over 2.5 million cases. However, this currently only contains data for the year 1998, although the next release covering the year 2001 should be available in the near future.

The last Chapter dealt with the regional variation in income inequality and the development over time in Germany since reunification. The method used was 'Analysis of Gini', a decomposition of income inequality as measured by the Gini Coefficient (Frick et al., 2006). The unique advantage of ANOGI is that it provides an additional term which reflects the overlap between the distributions of two or more interesting groups or strata formed in the present application by various German regions.

Concluding from the empirical results with respect to post-government income, one must reject the hypothesis that East and West Germany are moving towards a common income distribution. After a promising start in the first half of the 1990s, when income levels among East Germans were increasing but also accompanied by rising inequality, this process appears to have stopped in the mid-1990s without major changes since. The picture is quite different for pre-government incomes, which are heavily dominated by labor income for East Germany, while for the West German population, capital gains are a more relevant issue. Driven largely by massive unemployment and the lack of counteracting capital income, market-income inequality in East Germany already surpassed Western levels in the early 1990s and this difference has increased continuously. The huge inequality of market incomes in East Germany results in that this region is no longer a stratum on its own with respect to the overlapping of pre-government incomes: very low (zero) as well as (just a few) remarkably high market incomes yield an income distribution overlap with West Germany. However, average East German market incomes (as well as the respective mean ranks) are still far lower than in West Germany.

Enlarging the number and structure of the regions under consideration by splitting the western part into its northern, central and southern components also reveals a certain degree of regional variation *within* West Germany, with the South being in a somewhat more favorable position with respect to market and disposable incomes. If West Germany regions (the South in particular) do form income strata on their own, then only at a much smaller scale and not persistently over time. There is, however, a clear picture of East Germany still being quite different from the rest of the country, irrespective of any western regional grouping.

Overall, clear indications of post-government income stratification are found. On the one hand, this may be taken as evidence of the need to continue financial transfers from West to East in the context of the new *Solidarity Pact II*, which just started in 2005. However, instead of arguing for the need to equalize existing differences with even more and higher transfers, policy-makers and the public may have to start discussing whether one should begin to accept regional differences as the basis for endogenous growth in the less advantaged regions. These issues are not limited solely to the East-West discussion, since there has been a tendency in recent years toward pre-government income stratification in South (West) Germany as compared to both other Western regions, as well as of Central (West) Germany as compared to the Northern part.

For further research, important improvements could be made by including non-cash income, whereas the present work has used only cash income throughout. As Garfinkel, Rainwater and Smeeding (2006) state, ‘close to more than a half of welfare state transfers consists of in kind benefits (such as health insurance, education, child care, elder care and other services)’. Non-cash income - such as fringe benefits and payment in kind (especially social transfers in kind) - as well as owner-occupied housing and household production are important to correctly estimate the whole income distribution, in particular when doing poverty or inequality research.

Non-cash income components are even more important in cross-national studies, as performed in section 2.2.2. This analysis only takes account of direct taxes and transfers of cash income. However, ‘from a theoretical point of view, a measure that counts in kind transfers and indirect taxes is superior to the conventional measure of cash disposable income as a measure of a household’s standard of living’ (Atkinson and Bourguignon, 2000b).

Empirical studies usually rely only on monetary income and overcoming this limitation is at the root of the new European program ‘Accurate Income Measurement for the Assessment of Public Policies’ (acronym: AIM-AP), financed by the EU Fifth Framework[207] started in 2006. The first of three linked projects deals explicitly with non-cash income and its possible impact on poverty and inequality research. Ignoring additional income components apart from monetary income can lead to an insufficient approximation of the resources at an individual’s disposal and therefore produce imperfect targeting and misallocations of resources (cf. AIM-AP, 2006). It is to be hoped that the results of this project will be integrated into new or ongoing surveys, in such a way so that further research can better account for non-cash income components.

[207] Entitled ‘Improving the Human Research Potential and the Socio-Economic Knowledge Base’.

Bibliography

AIM-AP: *Accurate Income Measurement for the Assessment of Public Policies*. Technical report, European Commission: Sixth Framework Programmme, 2006. Annex I - 'Description of Work'.

Alzola, Carlos and Frank Harrell: *An Introduction to S and The Hmisc and Design Libraries*. Freely available electronic book, 2004. pdf.

Anand, Sudhir: *Aspects of Poverty in Malaysia*. The Review of Income and Wealth, volume 23(1), 1–16, March 1977.

Atkinson, Anthony B.: *On the Measurement of Inequality*. Journal of Economic Theory, volume 2(3), 244–263, September 1970.

— *On the measurement of poverty*. Econometrica, volume 55(4), 749–764, July 1987.

— *Data Matter*, July 2002. Revised version of Alfred Cowles lecture given at the Australasian Meetings of the Econometric Society, Auckland, July 2001, pdf.

Atkinson, Anthony B. and François Bourguignon, eds.: *Handbook of Income Distribution*, volume 1. Elsevier, Amsterdam, New York, 2000a.

Atkinson, Anthony B. and François Bourguignon: *Introduction: Income Distribution and Economics*. In Atkinson and Bourguignon (2000a), 1–58.

Atkinson, Anthony B. and Andrew Leigh: *The Distribution of Top Incomes in New Zealand*, October 2005. pdf.

Atkinson, Anthony B.; Eric Marlier; Brian Nolan and Frank Vandenbroucke: *Social Indicators: The EU and Social Inclusion*. Oxford University Press, April 2002.

Bandourian, Ripsy: *Income Distributions: A Comparison across Countries and Time*. Working Paper 231, Luxembourg Income Study, April 2000. pdf.

Bandourian, Ripsy; James B. McDonald and Robert S. Turley: *A Comparison of Parametric Models of Income Distribution. Across Countries and Over Time*. Working Paper 305, Luxembourg Income Study, 2002. pdf.

Bane, Mary Jo and David T. Ellwood: *Slipping into and out of poverty: The dynamics of spells*. Journal of Human Resources, volume 21(1), 1–23, 1986.

Barnard, John and Donald B. Rubin: *Small-sample Degrees of Freedom with Multiple Imputation*. Biometrika, volume 86(4), 948–955, 1999.

Bartelheimer, Peter: *Teilhabe, Gefährdung und Ausgrenzung als Leitbegriffe der Sozialberichterstattung*. SOFI-Mitteilungen, volume 32, 47–61, 2004.

Becker, Irene; Joachim R. Frick; Markus M. Grabka; Richard Hauser; Peter Krause and Gert G. Wagner: *A Comparison of the Main household Income Surveys for Germany: EVS and SOEP*. In Richard Hauser and Irene Becker, eds., *Reporting on Income Distribution and Poverty. Perspectives from a German and a European Point of View*, 55–90. Springer, Berlin, 2002.

Bernaards, Coen A.; Melissa M. Farmer; Karen Qi; Gareth S. Dulai; Patricia A. Ganz and Katherine L. Kahn: *Comparison of Two Multiple Imputation Procedures in a Cancer Screening Survey*. Journal of Data Science, volume 1(3), July 2003. pdf.

Berthoud, Richard: *Patterns of poverty across Europe*. Policy Press, Bristol, 2004.

Betti, Gianni; Bruno Cheli; Achille Lemmi and Vijay Verma: *On the Construction of Fuzzy Measures for the Analysis of Poverty and Social Exclusion*. Technical report, Paper presented at the International Conference to Honour Two Eminent Social Scientists Gini and Lorenz University of Siena 23-26 May, 2005. pdf.

Biewen, Martin: *Measuring the effects of Socio-Economic Variables on the Income Distribution: An Application to the East German Transition Process*. The Review of Economics and Statistics, volume 83(1), 185–190, 2001.

Bird, Edward J.; Joachim R. Frick and Gert G. Wagner: *The Income of Socialist Upper Classes during the Transition to Capitalism - Evidence from Longitudinal East German Data*. Journal of Comparative Economics, volume 26(2), 211–225, 1998.

Bishop, John A.; John P. Formby and Lester A. Zeager: *The distributional impact of unification and the 1992-93 recession on West German households*. Economics of Transition, volume 9(2), 515–532, 2001.

Blackburn, McKinley L.: *Poverty Measurement: An Index Related to a Theil Measure of Inequality*. Journal of Business & Economic Statistics, volume 7(4), 475–481, 1989.

Blackorby, Charles and David Donaldson: *Ethical indices for the measurement of poverty*. Econometrica, volume 48(4), 1053–1060, May 1980.

BMAS: *Lebenslagen in Deutschland. Der erste Armuts- und Reichtumsbericht der Bundesregierung*. Technical report, Bundesministerium für Arbeit und Soziales, Bonn, 2001. pdf.

— *Lebenslagen in Deutschland. Der 2. Armuts- und Reichtumsbericht der Bundesregierung*. Technical report, Bundesministerium für Arbeit und Soziales, Berlin, 2005. pdf.

Böheim, René and Stephen P. Jenkins: *Do current income and annual income measures provide different pictures of Britain's income distribution?* ISER Working Papers. University of Essex, Colchester, 2000. pdf.

Bourguignon, François and Satya R. Chakravarty: *Multi-dimensional poverty orderings*. Working Paper 2002-22, DELTA, Paris, 2002.

Brenke, Karl: *Income Growth in German Households: East Germany Falls Behind*. Weekly Report 15, DIW Berlin, 2005.

Büchel, Felix and Joachim R. Frick: *Income Composition and Redistribution in Germany - The Role of Ethnic Origin and Assimilation*. In Elke Holst; Dean R. Lillard and Thomas A. DiPrete, eds., *Proceedings of the 2000 Fourth International Conference of Socio-Economic Panel Study Users*, volume 70 of *Vierteljahrshefte zur Wirtschaftsforschung*, 135–145. Duncker & Humblot, Berlin, 2001.

Burgess, Simon M. and Carol Propper: *An Economic Model of Houshold Income Dynamics, with an Application to Poverty Dynamics among American Women*. Discussion Paper CASE/9, Centre for Analysis of Social Exclusion. London School of Economics, London, July 1998.

Burkhauser, Richard; Douglas Holtz-Eakin and Stephen Rhody: *Mobility and inequality in the 1980s: a cross-national comparison of the United States and Germany*. In Jenkins, Kapteyn and van Praag (1998), 111–175.

Burkhauser, Richard V.; Barbara A. Butrica; Mary C. Daly and Dean R. Lillard: *The Cross-National Equivalent File: A product of cross-national research*. In Irene Becker; Notburga Ott and Gabriele Rolf, eds., *Soziale Sicherung in einer dynamischen Gesellschaft (Social insurance in a dynamic society). Festschrift für Richard Hauser zum 65. Geburtstag*, 354–376. Campus, Frankfurt/Main, New York, 2001. pdf.

Burkhauser, Richard V.; Amy Crews Cutts; Mary C. Daly and Stephen P. Jenkins: *Testing the significance of income distribution changes over the 1980s business cycle: a cross-national comparison*. Journal of Applied Econometrics, volume 14(3), 253–272, June 1999.

Burkhauser, Richard V.; Michaela Kreyenfeld and Gert G. Wagner: *The German Socio-Economic Panel - A Representative Sample of Reunited Germany and its Parts*. In *Proceedings of the 1996 Second International Conference of the German Socio-Economic Panel Study Users*, Vierteljahrshefte zur Wirtschaftsforschung, 66:1, 7–16. DIW, Berlin, 1997.

Burkhauser, Richard V. and John G. Poupore: *A Cross-National Comparison of Permanent Inequality in the United States and Germany*. The Review of Economics and Statistics, volume 79(1), 10–17, 1997.

Burkhauser, Richard V.; Timothy M. Smeeding and Joachim Merz: *Relative Inequality and Poverty in Germany and the United States using alternative equivalence Scales*. The Review of Income and Wealth, volume 42(4), 381–400, December 1996.

Canberra-Group: *Expert Group on Household Income Statistics*. Final report and recommendations, LIS, Ottawa, 2001. pdf.

Casella, George and Edward I. George: *Explaining the Gibbs sampler*. The American Statistician, volume 46(3), 167–174, August 1992.

Chakravarty, Satya Ranjan: *Ethically flexible measures of poverty*. Canadian Journal of Economics, volume 16(1), 74–85, February 1983a.

— *A new Index of Poverty*. Mathematical Social Sciences, volume 6(3), 307–313, December 1983b.

— *Ethical Social Index Numbers*. Springer, Berlin, Heidelberg, 1990.

— *On Shorrocks' reinvestigation of the Sen poverty index*. Econometrica, volume 65(5), 1241–1242, September 1997.

Chakravarty, Satya Ranjan and Pietro Muliere: *Welfare indicators: a review and new perspectives. 2. Measurement of poverty*. Metron - International Journal of Statistics, (2), 247–281, LXII 2004. `http://ideas.repec.org/a/mtn/ancoec/040206.html`.

Cheli, Bruno and Achille Lemmi: *A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty*. Economic Notes, volume 24(1), 115–134, 1995.

Clark, Stephen; Richard Hemming and David Ulph: *On indices for the measurement of poverty*. The Economic Journal, volume 91(362), 515–526, June 1981.

Clemenceau, Anne and Christine Wirtz: *Europäisches Haushaltspanel 'Newsletter' 01/01*. Statistik kurzgefaßt, volume 3-14, 2001. EUROSTAT.

Clementi, Fabio and Mauro Gallegati: *Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States*. ArXiv Physics e-prints, April 2005a. pdf.

— *Power Law Tails in the Italian Personal Income Distribution*. Physica A: Statistical and Theoretical Physics, volume 350(2-4), 427–438, May 2005b.

Colavecchio, Roberta; Declan Curranand and Michael Funke: *Drifting Together Or Falling Apart? The Empirics Of Regional Economic Growth In Post-Unification Germany*. Working Paper 1533, CESifo, September 2005. pdf.

Coniffe, Denis: *The non-constancy of equivalence scale*. The Review of Income and Wealth, volume 38(4), 429–443, December 1992.

Cowell, Frank A.: *Poverty Measures, Inequality and Decomposability*. In Dieter Bös; Manfred Rose and Christian Seidl, eds., *Welfare and Efficiency in Public Economics*, 149–166. Springer, Heidelberg, 1988.

Cowell, Frank A., ed.: *The Economics of Poverty and Inequality*, volume 158 of *The international library of critical writings in economics*. Elgar, Cheltenham, April 2003.

Cowell, Frank A. and Emmanuel Flachaire: *Sensitivity of Inequality Measures to Extreme Values*. Discussion Paper DARP 60, London School of Economics, London, March 2002.

Cowell, Frank A. and Maria-Pia Victoria-Feser: *Robustness Properties of Inequality Measures*. Econometrica, volume 64(1), 77–101, January 1996.

Dalton, Hugh: *The Measurement of the Inequality of Incomes*. Economic Journal, volume 30(119), 348–361, September 1920.

D'Ambrosio, Conchita; Pietro Muliere and Pierceare Secchi: *The endogenous poverty line as a change-point in the income distribution. The Principle of transfers revisited*, August 2002. Paper prepared for the 27 th General Conference of The International Association for Research in Income and Wealth.

De Vos, K. and M. A. Zaidi: *Equivalence Scale Sensitivity of Poverty Statistics for the Member States of the European Community*. The Review of Income and Wealth, volume 43(3), 319–333, 1997.

Devicienti, Francesco: *Poverty Persistence in Britain: A Multivariate Analysis using the BHPS, 1991-1997*. In P. Moyes; C. Seidl and A Shorrocks, eds., *Inequalities: Theory, Experiments and Applications*, Journal of Economics, Supplement 9. Springer, Wien, New York, 2002.

Dickey, Heather: *Regional Earnings Inequality in Great Britain: A Decomposition Analysis*. Regional Studies, volume 35(7), 605–612, 2001.

Dietz, Berthold: *Soziologie der Armut*. Campus, Frankfurt/Main, New York, 1997.

Diggle, Peter J.: *On Informative and Random Dropouts in Longitudinal Studies (letter to the editor)*. Biometrics, volume 49, 947–949, 1993.

— *Estimation with Missing Data (letter to the editor)*. Biometrics, volume 50, 580, 1994.

Donaldson, David and John A. Weymark: *Properties of fixed-population poverty indices*. International Economic Review, volume 27(3), 667–688, October 1986.

Drăgulescu, Adrian and Victor M. Yakovenko: *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States*, Mar 2001.
URL http://arxiv.org/abs/cond-mat/0103544

Duncan, Greg J.; Björn Gustafsson; Richard Hauser; Günther Schmauss; Hans Messinger; Ruud Muffels; Brian Nolan and Jean-Claude Ray: *Poverty dynamics in eight countries*. Journal of Population Economics, volume 6(3), 215–234, 1993.

Durrant, Gabriele B.: *Imputation Methods in the Social Sciences: A Methodological Review*. Methods Review Papers NCRM/002, ESRC National Centre for Research Methods, Southampton, June 2005. pdf.

Esping-Andersen, Gsta: *The three worlds of welfare capitalism*. Princeton University Press, Princeton, New Jersey, 1990.

EUREDIT-Project: *(WP3) Evaluation criteria for statistical editing and imputation*. Technical Report T001.04, EUREDIT Project, 2000.

EUROSTAT: *Laeken Indicators - Detailed Calculation Methodology* -. DOC. E2/IPSE/2003, European Commission, Directorate E: Social statistics, Unit E-2: Living conditions, Luxembourg, 2003. Working Group 'Statistics on Income, Poverty & Social Exclusion, pdf.

— *Statistics On Income, Poverty & Social Exclusion (IPSE) and EU/SILC (Statistics On Income And Living Conditions): Methodology of Calculation of Common cross-sectional EU Indicators*. Doc. IPSE/65/04, EUROSTAT, 2004.

Fabig, Holger: *Einkommensdynamik im internationalen Vergleich. Eine empirische Mobilitätsanalyse mit Panel-Daten*, volume 39 of *Wirtschaftswissenschaft*. Campus, Frankfurt/Main; New York, 1999.

Fay, Robert E.: *Alternative Paradigms for the Analysis of imputed Survey Data*. Journal of the American Statistical Association, volume 91(434), 490–498, June 1996.

Feldstein, Martin: *Income Inequality and Poverty*. Working Paper 6770, National Bureau of Economic Research, Cambridge, October 1998.
URL `www.nber.org/papers/w6770`

Förster, Michael; David Jesuit and Timothy M. Smeeding: *Regional Poverty and Income Inequality in Central and Eastern Europe: Evidence from the Luxembourg Income Study*. Working Paper 324, Luxembourg Income Study, July 2002. pdf.

Foster, James E.: *On Economic Poverty: A Survey of Aggregate Measures*. Advances in Econometrics, volume 3, 215–251, 1984.

Foster, James E.; Joel Greer and Erik Thorbecke: *A class of decomposable poverty measures*. Econometrica, volume 52(3), 761–766, 1984.

Foster, James E. and Anthony F. Shorrocks: *Subgroup consistent poverty indices*. Econometrica, volume 59(3), 687–709, May 1991.

Frick, Joachim R.; Jan Goebel; Markus M. Grabka; Peter Krause; Andrea Schäfer; Ingrid Tucci and Gert G. Wagner: *Zur langfristigen Entwicklung von Einkommen und Armut in Deutschland: Starke Reduktion der arbeitsmarktbedingten Ungleichheit durch sozialstaatliche Maßnahmen*. Wochenbericht 4, DIW Berlin, 2005.

Frick, Joachim R.; Jan Goebel; Edna Schechtman; Gert G. Wagner and Shlomo Yitzhaki: *Using Analysis of Gini (ANOGI) for Detecting Whether Two Subsamples Represent the Same Universe The German Socio-Economic Panel Study (SOEP) Experience*. Sociological Methods & Research, volume 34(4), 427–468, 2006.

Frick, Joachim R. and Markus M. Grabka: *Missing Income Data in the German SOEP: Incidence, Imputation and its Impact on the Income Distribution*. Discussion Papers 376, DIW Berlin, April 2004. pdf.

— *Item-non-response on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility*. Allgemeines Statistisches Archiv, volume 89(1), 49–62, 2005.

Friedrichs, Jürgen; Hartmut Häußermann and Walter Siebel, eds.: *Süd-Nord-Gefälle in der Bundesrepublik?* Westdeutscher Verlag, Opladen, 1986.

Garfinkel, Irwin; Lee Rainwater and Timothy M. Smeeding: *A Reexamination of Welfare States and Inequality in Rich Nations: How In Kind Transfers and Indirect Taxes Change the Story*. Journal of Policy Analysis and Management, 2006. Forthcoming.

Geppert, Kurt: *Süd-Nord-Gefälle eingeebnet? Zur räumlichen Wirtschaftsentwicklung in Westdeutschland*. Wochenbericht 66(3), DIW Berlin, 1999.

Gibrat, Robert: *Les Inégalités économiques*. Recueil Sirey, Paris, 1931.

Giorgi, Giovanni Maria and Michele Crescenzi: *A proposal of poverty measures based on the Bonferroni inequality index*. Metron - International Journal of Statistics, volume LIX(3-4), 3–16, 2001. pdf.

Goebel, Jan; Peter Krause and Jürgen Schupp: *Growth in Unemployment Raises Poverty Rates: Most low-wage earnings constitute supplement to primary household income*. Weekly Report 10, DIW Berlin, April 2005.

Goebel, Jan and Birgit Kuchler: *Incidence and Intensity of Smoothed Income Poverty in European Countries*. Journal of European Social Policy, volume 13(4), 357–369, November 2003.

Goodin, Robert E.; Bruce Headey; Ruud Muffels and Henk-Jan Dirven: *The real worlds of welfare capitalism*. University Press, Cambridge, 1999.

Grabka, Markus M.; Johannes Schwarze and Gert G. Wagner: *How unification and immigration affected the German income distribution*. European Economic Review, volume 43(4-6), 867–878, 1999.

Hagenaars, Aldi: *A Class of Poverty Indices*. International Economic Review, volume 28(3), 583–607, October 1987.

Hagenaars, Aldi J. M. and Bernard M. S. van Praag: *A synthesis of poverty line definitions*. The Review of Income and Wealth, volume 31(2), 139–154, June 1985.

Haisken-DeNew, John P. and Joachim R. Frick, eds.: *Desktop Companion to the German Socio-Economic Panel (SOEP), Version 8.0 - Update to Wave 21*. DIW Berlin, 2005. pdf.

Hartmann, Peter H. and Bernhard Schimpl-Neimanns: *Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage*. Kölner Zeitschrift für Soziologie und Sozialpsychologie, volume 44(2), 315–340, 1992.

Hauser, Richard; Joachim R. Frick; Klaus Mueller and Gert G. Wagner: *Inequality in Income: A Comparison of East and West Germans before Reunification and during Transition*. Journal of European Social Policy, volume 4(4), 277–295, 1994.

Heitjan, Daniel F.: *Estimation with Missing Data (letter to the editor)*. Biometrics, volume 50, 580, 1994.

Heitjan, Daniel F. and Srabashi Basu: *Distinguishing "Missing at Random" and "Missing Completely at Random"*. The American Statistician, volume 50(3), 207–213, 1996.

Hill, Martha S. and Stephen P. Jenkins: *Poverty among British children: Chronic or transitory?* In Bruce Bradbury; Stephen P. Jenkins and John Micklewright, eds., *The Dynamics of Child Poverty in Industrialised Countries*, chapter 7, 174–195. Cambridge University Press, 2001.

Hoogland, Jeffrey and Jeroen Pannekoek: *Evaluation of SPSS Missing Value Analysis 7.5*. Technical report, AutImp, February 2000.

Horovitz, Joel L. and Charles F. Manski: *Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations*. Journal of Econometrics, volume 84, 37–58, 1998.

— *Nonparametric Analysis of Randomized Experiments With Missing Covariate and Outcome Data*. Journal of the American Statistical Association, volume 95(449), 77–84, 2000.

Hunt, Jennifer: *Post-Unification wage growth in East Germany*. The Review of Economics and Statistics, volume 83(1), 190–195, February 2001.

Isengard, Bettina: *Machbarkeitsstudie zur Erhebung einkommensschwacher und einkommensstarker Haushalte im Sozio-Oekonomischen Panel (SOEP)*. Materialien 17, DIW Berlin, Juni 2002. pdf.

Ishikawa, Atushi: *Pareto law and Pareto index in the income distribution of Japanese companies*, Oct 2004. pdf.

Jäntti, Marcus and Sheldon Danziger: *Income Poverty in Advanced Countries*. In Atkinson and Bourguignon (2000a), chapter 6, 309–378.

Jenkins, Stephen P.; Arie Kapteyn and Bernard M. S. van Praag, eds.: *The Distribution of Welfare and Houshold Production: International Perspectives*. University Press, Cambridge, 1998.

Jenkins, Stephen P. and Peter J. Lambert: *Ranking income distributions when needs differ*. The Review of Income and Wealth, volume 39(4), 337–356, 1993.

— *Three 'I's of poverty curves, with an analysis of UK poverty trends*. Oxford Economic Papers, volume 49, 317–327, 1997.

— *Ranking Poverty Gap Distributions: Further Tips for Poverty Analysis*. In D. J. Slottje, ed., *Research on Economic Inequality*, volume 8, 31–38. JAI Press, 1998.

Jenkins, Stephen P. and Philippe van Kerm: *Trends in income inequality, Pro-Poor income growth and income mobility.* Working Paper 2003-27, University of Essex ISER, Colchester, 2003. pdf.

Kakwani, Nanak C.: *Income Inequality and Poverty.* Oxford University Press, New York, 1980a.

— *On a class of poverty measures.* Econometrica, volume 48(?), 437–446, 1980b.

— *Note on a New Measure of Poverty.* Econometrica, volume 49(2), 525–526, 1981.

— *Inequality, Welfare and Poverty: Three Interrelated Phenomena.* Discussion Paper 97/18, School of Economics, Sydney, 1997.

Kenward, Mike G. and Geert Molenberghs: *Likelihood Based Frequentist Inference When Data Are Missing at Random.* Statistical Science, volume 13(3), 236–247, 1998.

Kim, Jae Kwang and Wayne Fuller: *Fractional hot deck imputation.* Biometrika, volume 91(3), 559–578, 2004.

Kim, Jae-Kwang and Wayne A. Fuller: *Jackknife Variance Estimation after Hot Deck Imputation.* In *Proceedings of the Survey Research Methods Section*, 825–830. American Statistical Association, 1999. pdf.

Klass, Oren S.; Ofer Biham; Moshe Levy; Ofer Malcai and Sorin Solomon: *The Forbes 400 and the Pareto wea lth distribution.* Economics Letters, volume 90(2), 290–295, February 2006.

Kockläuner, Gerhard: *Revisiting two poverty indexes.* Allgemeines Statistisches Archiv, volume 86(3), 299–305, 2002.

Krämer, Walter and Thorsten Ziebach: *The weak Pareto law and regular variation in the tails1.* Working paper, University Dortmund, October 2002. pdf.

Krause, Peter: *Income, poverty and dynamics in Germany.* In Krause et al. (2003), 93–116.

Krause, Peter; Gerhard Bäcker and Walter Hanesch, eds.: *Combating poverty in Europe Europe: The German welfare regime in practice.* Ashgate, Aldershot, 2003.

Kronauer, Martin: *"Soziale Ausgrenzung" und "Underclass": Über neue Formen der gesellschaftlichen Spaltung.* Leviathan, volume 25(1), 28–49, 1997.

Kundu, Amitabh and Tony E. Smith: *An Impossibility Theorem on Poverty Indices.* International Economic Review, volume 24(2), 423–434, June 1983.

Lammers, Konrad: *Süd-Nord-Gefälle in West- und Ostdeutschland?* Wirtschaftsdienst, volume 11, 736–739, 2003.
URL http://www.hwwa.de/Forschung/Publikationen/Wirtschaftsdienst/2003/wd_docs2003/wd0311lammers.pdf

Lasswell, Thomas E.: *Class and Stratum: An introduction to concepts and research.* Houghton Mifflin, Boston, Massachusetts, 1965.

Leibfried, Stephan; Lutz Leisering; Petra Buhr; Monika Ludwig; Eva Mädje; Thomas Olk; Wolfgang Voges and Michael Zwick: *Zeit der Armut. Lebensläufe im Sozialstaat*. Suhrkamp, Frankfurt/Main, 1995.

Leibfried, Stephan and Wolfgang Voges: *Vom Ende einer Ausgrenzung? — Armut und Soziologie*. In Stephan Leibfried and Wolfgang Voges, eds., *Armut im modernen Wohlfahrtsstaat*, Sonderheft 32 der Kölner Zeitschrift für Soziologie und Sozialpsychologie, 9–33. Westdeutscher Verlag, Opladen, 1992.

Lerman, Robert I. and Shlomo Yitzhaki: *A Note on the Calculation and Interpretation of the Gini Index*. Economics Letters, volume 15, 363–368, 1984.

— *Improving the accuracy of estimates of the Gini coefficients*. Journal of Econometrics, volume 42, 43–47, 1989.

Lewis, Geoffrey W. and David T. Ulph: *Poverty, Inequality and Welfare*. The Economic Journal, volume 98(390), 117–131, 1988. Supplement: Conference Papers.

Li, K. H.; Trivellore E. Raghunathan and Donald B. Rubin: *Large-Sample Significance Levels From Multiply Imputed Data Using Moment-Based Statistics and an F Reference Distribution*. Journal of the American Statistical Association, volume 86(416), 1065–1073, 1999.

Lipsmeier, Gero: *Zur Repräsentation des unteren Einkommensbereiches im Sozio - ökonomischen Panel (SOEP)*. Arbeitspapier des DFG-Projektes "Versorgungsstrategien privater Haushalte im unteren Einkommensbereich" 10, Universität Bielefeld, April 1993. WordPerfect.

Little, Roderick J. A. and Donald B. Rubin: *Statistical Analysis with Missing Data*. Wiley & Sons, 2nd edition, 2002.

Loikkanen, Heikki; Marja Riihelä and Risto Sullström: *Regional income convergence and inequality in boom and bust. Results from micro data in Finland 1971-2000, and especially during the 1990s*. ERSA conference papers 514, European Regional Science Association, June 2003. Paper presented at the WIDER Development Conference: Inequality, Poverty and Human Well-Being, 30-31 May 2003, Helsinki, Finland.
URL http://www.ersa.org/ersaconfs/ersa03/cdrom/papers/514.pdf

Maasoumi, Esfandiar and Mark Trede: *Comparing Income Mobility in Germany and the United States Using Generalized Entropy Mobility Measures*. The Review of Economics and Statistics, volume 83(3), 551–559, August 2001.

Madden, David: *Relative or absolute poverty lines: A new approach*. The Review of Income and Wealth, volume 46(2), 181–199, June 2000.

Mandelbrot, Benoit: *The Pareto-Levy Law and the Distribution of Income*. International Economic Review, volume 1(2), 79–106, May 1960.

Mander, Adrian P. and David Clayton: *HOTDECK: Stata module to impute missing values using the hotdeck method*. Boston College Department of Economics, 1999. Revised 2005.

McDonald, James B: *Some Generalized Functions for the Size Distribution of Income*. Econometrica, volume 52(3), 647–663, May 1984.

Meng, Xiao-Li: *Multiple-Imputation Inferences with Uncongenial Sources of Input (with discussion)*. Statistical Science, volume 9(4), 538–573, 1994.

— *A Congenial Overview and Investigation of Muliple Imputation Inferences under Uncongeniality*. In R. M. Groves; D. A. Dillman; J. L. Eltinge and Roderick J. A. Little, eds., *Survey Nonresponse*, 343–356. Wiley & Sons, New York, 2002.

Mercader-Prats, Magda and Horacio Levy: *The role of tax and transfers in reducing personal Income Inequality in Europe's regions: Evidence from EUROMOD*. Working Paper EM9/04, EUROMOD, December 2004.

Mesa, D. M.; P. Tsai and R. L. Chambers: *Using Tree-Based Models for Missing Data Imputation: An Evaluation Using UK Census Data*, October 2000. Working Paper of the AUTIMP EU Project.

Moothathu, T. S. K.: *Sampling distribution of Lorenz curve and Gini index of the pareto distribution*. Sankhyā: The Indian Journal of Statistics, Series B, volume 47(2), 247–258, 1985.

Muller, Christophe: *The Properties of the Watts Poverty Index under Lognormatlity*. Economics Bulletin, volume 9(1), 1–9, 2001.

Nicoletti, Cheti and Franco Peracchi: *Imputation procedures and the quality of income information in the ECHP*, October 2003. Paper presented at the Workshop on Item-Nonresponse and Data Quality in Large Social Surveys.

Nygård, Fredrik and Arne Sandström: *Income Inequality measures based on sample surveys*. Journal of Econometrics, volume 42, 81–95, 1989.

Opsomer, Jean D. and Damião Nóbrega da Silva: *Estimation after Fractional Imputation under a Nonparametric Response Mechanism*. online article, May 2002.
URL http://citeseer.csail.mit.edu/dasilva02estimation.html

— *Properties of the weighting cell estimator under a nonparametric response mechanism*. Survey Methodology, volume 30(1), 45–55, 2004. pdf.

Osberg, Lars and Kuan Xu: *Poverty Durations and Poverty Measurement*. Paper for the Canadian Economics Association 2000 Meeting, 2000. pdf.

Pattanaik, Prasanta K. and Manimay Sengupta: *An alternative axiomatization of Sen's poverty measure*. The Review of Income and Wealth, volume 41(1), 73–80, March 1995.

Piachaud, David: *Wie mißt man Armut?* In Stephan Leibfried and Wolfgang Voges, eds., *Armut im modernen Wohlfahrtsstaat*, KZfSS, Sonderheft, 63–87. Westdeutscher Verlag, Opladen, 1992.

Pischner, Rainer: *Preliminary Estimate of the Integrated Cross-sectional Weighting (Sub-samples A - G) for 2003*. Technical report, German Socio-Economic Panel, DIW Berlin, May 2004. pdf.

Pittau, Maria Grazia: *Fitting Regional Income Distributions in the European Union*. Oxford Bulletin of Economics and Statistics, volume 67(2), 135–161, April 2005.

Pötter, Ulrich and Ulrich Rendtel: *Über Sinn und Unsinn von Repräsentativitätsstudien*. Allgemeines Statistisches Archiv, volume 77, 260–280, 1993.

Pyatt, Graham: *On the interpretation and disaggregation of Gini coefficients*. Economic Journal, volume 86(342), 243–255, 1976.

— *Measuring Welfare, Poverty and Inequality*. The Economic Journal, volume 97(386), 459–467, June 1987.

R Development Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.
URL http://www.R-project.org

Raghunathan, Trivellore E.; J. M. Lepkowski; John Van Hoewyk and Peter W. Solenberger: *A Multivariate Technique for Multiply Imputing Missing Values. Using a Sequence of Regression Models*. Survey Methodology, volume 27(1), 85–95, 2001a.

— *A Multivariate Technique for Multiply Imputing Missing Values. Using a Sequence of Regression Models*. Working Paper 81, Institute for Social Research, University of Michigan, 2001b. pdf.

Raghunathan, Trivellore E.; Peter W. Solenberger and John Van Hoewyk: *IVEware: Imputation and Variance Estimation Software*. User guide, Institute for Social Research, University of Michigan, March 2002. pdf.

Rao, J. N. K.: *On Variance Estimation With Imputed Survey Data*. Journal of the American Statistical Association, volume 91(434), 499–506, June 1996.

Ravallion, Martin: *Expected Poverty Under Risk-Induced Welfare Variability*. Economic Journal, volume 98(393), 1171–1182, December 1988.

— *Poverty lines in theory and practice*. working paper 133, Living Standards Measurement Study LSMS; The World Bank, Washington, D.C., July 1998.

Rendtel, Ulrich: *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität Frankfurt Campus*. Campus, Frankfurt/Main, New York, 1995.

Rendtel, Ulrich; Leif Nordberg; Markus Jäntti; Jens Hanisch and Edin Basic: *Report on quality of income data*. Working Paper 21, CHINTEX, February 2004. pdf.

Robins, James M. and Naisyin Wang: *Inference for Imputation Estimators*. Biometrika, volume 87(1), 113–124, 2000.

Rodgers, Joan R. and John L. Rodgers: *The Measurement of Chronic and Transitory Poverty: with Application to the United States*. Economics Working Paper Archive 55, Levy Economics Institute, The, June 1991.
URL http://ideas.repec.org/p/lev/wrkpap/55.html

— *Chronic Poverty in the United States*. Journal of Human Resources, volume 28(3), 25–54, 1993.

Rohwer, Götz: *Einkommensmobilität und soziale Mindestsicherung. Einige Überlegungen zum Armutsrisiko*. In Stephan Leibfried and Wolfgang Voges, eds., *Armut im modernen Wohlfahrtsstaat*, KZfSS, Sonderheft, 367–379. Westdeutscher Verlag, Opladen, 1992.

Rowntree, Benjamin Seebohm: *Poverty: A study of town life*. Macmillan, London, 1901.

Royston, Patrick: *MICE: Stata module for multiple imputation of missing values*. Statistical Software Components, Boston College Department of Economics, October 2004.
URL http://ideas.repec.org/c/boc/bocode/s446602.html

— *Multiple imputation of missing values: Update of ice*. Stata Journal, volume 5(4), 527–536, 2005.

Rubin, Donald B.: *Inference and missing data*. Biometrika, volume 63(3), 581–592, 1976.

— *Multiple Imputations in sample surveys – A phenomenological Bayesian Approach to nonrespones*. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 20–28, 1978.

— *Multiple Imputation for Nonresponse in Surveys*. Wiley & Sons, New York, 1987.

— *Multiple Imputation After 18+ Years*. Journal of the American Statistical Association, volume 91(434), 473–489, June 1996.

Rubin, Donald B. and Nathaniel Schenker: *Multiple imputation for interval estimation from simple random samples with ignorable nonresponse*. Journal of the American Statistical Association, volume 81(394), 366–374, 1986.

Sallila, Seppo; Heikki Hiilamo and Reijo Sund: *Rethinking relative measures of poverty*. Journal of European Social Policy, volume 16(2), 107–120, 2006.

Schafer, J. L. and N. Schenker: *Inference With Imputed Conditional Means*. Journal of the American Statistical Association, volume 95(449), 144–154, 2000.

Schafer, Joseph L.: *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London, 1997.

— *NORM: Multiple imputation of incomplete multivariate data under a normal model, Version 2.03*. Software for windows, Pennsylvania State University, 1999.
URL http://www.stat.psu.edu/jls/misoftwa.html

Schaich, Eberhard and Ralf Münnich: *Der Fuzzy-Set Ansatz in der Armutsmessung. A Fuzzy Set Approach to Poverty Measurement*. Jahrbücher für Nationalökonomie und Statistik, volume 215(4), 444–469, 1996.

Scharfstein, Daniel O.; Andrea Rotnitzky and James M. Robins: *Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models*. Journal of the American Statistical Association, volume 94(448), 1096–1146, December 1999. With discussion.

Schäuble, Gerhard: *Theorien, Definitionen und Beurteilung der Armut*. Sozialpolitische Schriften. Duncker und Humblot, Berlin, 1984.

Schechtman, Edna: *Stratification: Measuring and Inference*. Communications in Statisics: Theory and Methods, volume 34(11), 2133–2145, 2005.

Schechtman, Edna and Shlomo Yitzhaki: *A Measure of Association Based on Gini's Mean Difference*. Communications in Statisics: Theory and Methods, volume A16(1), 207–231, 1987.

— *On the proper bounds of the Gini correlation*. Economics Letters, volume 63, 133–138, 1999.

Schluter, Christian and Mark Trede: *Local versus global assessments of mobility*. International Economic Review, volume 44(4), 1313–1335, November 2003.

Schmid, Friedrich: *A general class of poverty measures*. Statistical Papers, volume 34(3), 189–211, 1993.

Schupp, Jürgen; Tobias Gramlich; Bettina Isengard; Rainer Pischner; Gert G. Wagner and Bernhard von Rosenbladt: *Repräsentative Analyse der Lebenslage einkommensstarker Haushalte*. Forschungsprojekt, Bundesministeriums für Gesundheit und Soziale Sicherung, 2003. pdf.

Schupp, Jürgen; Tobias Gramlich; Rainer Pischner; Gert G. Wagner and Bernhard von Rosenbladt: *Repräsentative Analyse der Lebenslage einkommensstarker Haushalte Erbschaft, soziale Herkunft und spezielle Lebenslagenindikatoren*. Forschungsprojekt, Bundesministeriums für Gesundheit und Soziale Sicherung, 2005. pdf.

Schupp, Jürgen and Gert G. Wagner: *The German Socio-Economic Panel: A database for Longitudinal Comparisons*. Innovations, volume 8(1), 95–108, 1995.

Schwarze, Johannes: *Using Panel Data on Income Satisfaction to Estimate Equivalence Scale Elasticity*. The Review of Income and Wealth, volume 49(3), 359–372, March 2003.

Seidl, Christian: *Poverty Measurement: A Survey*. In Dieter Bös; Manfred Rose and Christian Seidl, eds., *Welfare and Efficiency in Public Economics*, 71–147. Springer, Berlin, Heidelberg, 1988.

Sen, Amartya K.: *Poverty: An ordinal approach to measurement*. Econometrica, volume 44(2), 219–231, 1976.

— *Poverty and Famines. An Essay on Entitlement and Deprivation*. Clarendon Press, Oxford, 1981.

— *Inequality Reexamined*. Russell Sage Foundation Publications, 1992.

Shao, J.: *Cold Deck and Ratio Imputation.* Survey Methodology, volume 26(1), 79–85, 2000.

Shih, Weichung Joe: *On Informative and Random Dropouts in Longitudinal Studies (Letter to the Editor).* Biometrics, volume 48, 970–972, 1992.

Shorrocks, Anthony F.: *The measurement of mobility.* Econometrica, volume 46(5), 1013–1024, 1978.

— *Revisiting the Sen Poverty Index.* Econometrica, volume 63(5), 1225–1230, September 1995.

— *Deprivation Profiles and Deprivation Indices.* In Jenkins et al. (1998), 250–267.

Shorrocks, Anthony F. and Guanghua Wan: *Spatial Decomposition of Inequality.* Journal of Economic Geography, volume 5(1), 59–81, January 2005.

Singh, S. K. and G. S. Maddala: *A Function for Size Distribution of Incomes.* Econometrica, volume 44(5), 963–970, September 1976.

SOEP Group: *The German Socio-Economic Panel (SOEP) after more than 15 years - Overview.* In Elke Holst; Dean R. Lillard and Thomas A. DiPrete, eds., *Proceedings Of The 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000)*, volume 70 of *Vierteljahrshefte zur Wirtschaftsforschung*, 7–14. Berlin, 2001.

Spiess, Martin: *Analyse von Längsschnittdaten mit fehlenden Werten: Grundlagen, Verfahren und Anwendungen.* Habilitations-schrift, Universität Bremen, 2005. pdf.

Spiess, Martin and Jan Goebel: *A Comparison of Different Imputation Rules.* In Manfred Ehling and Ulrich Rendtel, eds., *Harmonisation of Panel Surveys and Data Quality -CHINTEX-*, 293–316. Federal Statistical Office, Wiesbaden, 2004.

— *On the effect of item nonresponse on the estimation of a two-panel-waves wage equation.* Allgemeines Statistisches Archiv, volume 89(1), 63–74, 2005.

Stewart, Kitty: *Measuring Well-Being and Exclusion in Europe s Regions.* CASEpaper 53, Centre for Analysis of Social Exclusion, London School of Economics, 2002.

SVR: *Erfolge im Ausland - Herausforderungen im Inland.* Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, Jahresgutachten 2004/05, November 2004.

Takayama, Noriyuki: *Poverty, Income Inequality, and their Measures: Professor Sen's Axiomatic Approach Reconsidered.* Econometrica, volume 47(3), 747–759, May 1979.

Theil, Henri: *Economics and Information Theory.* North-Holland, Amsterdam, 1967.

Thon, Dominique: *On Measuring Poverty.* The Review of Income and Wealth, volume 25(4), 429–440, 1979.

— *Income inequality and poverty: Some problems.* The Review of Income and Wealth, volume 27(2), 207–210, 1981.

— *A Note on a Troublesome Axiom for Poverty Indices.* The Economic Journal, volume 93(369), 199–200, March 1983a.

— *A poverty measure.* Indian Economic, volume 30, 55–70, 1983b.

Tsakloglou, Panos and Giota Panopoulou: *Who are the poor in Greece? Analysing poverty under alternative concepts of resources and equivalence scales.* Journal of European Social Policy, 213–236, 1998.

Tungodden, Bertil: *Poverty measurement: the critical comparison value.* Social Choice and Welfare, volume 25(1), 75–84, 2005.

Valletta, Robert G.: *The Ins and Outs of Poverty in Advanced Economies: Government Policy and Poverty Dynamics in Canada, Germany, Great Britain, and the United States.* The Review of Income and Wealth, volume 52(2), 261–284, June 2006.

van Buuren, Stef and C. G. M. Oudshoorn: *Flexible multivariate imputation by MICE.* TNO report PG/VGZ/99.054, TNO Prevention and Health, Leiden, Oktober 1999. pdf.

— *Multivariate Imputation by Chained Equations. MICE V1.0 User's manual.* TNO report PG/VGZ/00., TNO Prevention and Health, Leiden, June 2000.

van Kerm, Philippe and Stephen P. Jenkins: *Linking income mobility and inequality: a re-assessment of US and German inequality trends*, May 2003. Paper prepared for the UNU/WIDER Conference on Inequality, Poverty and Human Well-being in Helsinki, pdf.

Vaughan, R. N.: *Welfare Approaches to the Measurement of Poverty.* The Economic Journal, volume 97, 160–170, 1987. Supplement: Conference Papers.

Victoria-Feser, Maria-Pia: *Comment.* In Jacques Silber, ed., *Handbook of Income Inequality Measurement*, Recent economic thought series, chapter 8, 260–267. Kluwer, Boston, Dordrecht, London, 1999.

Victoria-Feser, Maria-Pia and Elvezio Ronchetti: *Robust Estimation for Grouped Data.* Journal of the American Statistical Association, volume 92(437), 333–340, March 1997.

von Hippel, Paul T.: *Biases in SPSS 12.0 Missing Value Analysis.* The American Statistician, volume 58(2), 160–164, 2004.

von Wiese, Leopold: *Über die Armut.* Kölner Zeitschrift für Soziologie und Sozialpsychologie, volume 6, 42–62, 1954.

Wagner, Gert G.; Richard V. Burkhauser and Frederike Behringer: *The English Language Public Use File of the German Socio-Economic Panel Study.* Journal of Human Resources, volume 28(2), 429–433, 1993.

Walker, Robert: *Rethinking poverty in a dynamic perspective.* In Hans-Jürgen Andreß, ed., *Empirical poverty research in a comparative perspective*, 29–49. Ashgate, Aldershot, 1998.

Walker, Robert and Karl Ashworth: *Time: The Forgotten Dimension of Poverty*. Briefings. CRSP (Centre for Research in Social Policy), Loughborough, 1992.

Watts, Harold: *An Economic Definition of Poverty*. In D. P. Moynihan, ed., *On understanding Poverty*, chapter 11, 316–329. Basic Books, New York, 1968.

Weikard, Hans-Peter: *Zur Zeitdimension in der Armutsmessung*. Schmollers Jahrbuch, Journal of Applied Social Science Studies, volume 120(1), 25–39, 2000.

Wiggins, R. D.; M. Ely and K. Lynch: *A comparative evaluation of currently available software remedies to handle missing data in the context of longitudinal design and analysis*. Paper, 5^{th} International Conference on Social Science Methodology, Cologne, 2000. pdf.

Williams, Todd R. and Leroy Bailey: *Compensating for Missing Wave Data in the Survey of Income and Program Participation (SIPP)*. In *Proceedings of the Survey Research Methods Section*, 305–310. American Statistical Association, 1996. pdf.

Wirtz, Christine and Lene Mejer: *The European Community Household Panel (ECHP)*. Schmollers Jahrbuch, Journal of Applied Social Science Studies, volume 122(1), 143–154, 2002. pdf.

Xu, Kuan and Lars Osberg: *How to Decompose the Sen-Shorrocks-Thon Poverty Index: A Practioner's Guide*. Journal if Income Distribution, volume 10(1-2), 77–94, 2001.

Yee, Thomas W.: *VGAM: Vector Generalized Linear and Additive Models*, 2006. R package version 0.6-8.
URL `http://www.stat.auckland.ac.nz/~yee`

Yitzhaki, Shlomo: *Relative Deprivation and Economic Welfare*. European Economic Review, volume 17, 99–113, 1982.

— *Calculating Jackknife Variance Estimators for Parameters of the Gini Method*. Journal of Business & Economic Statistics, volume 9(2), 235–239, April 1991.

— *Economic distance and overlapping of distributions*. Journal of Econometrics, volume 61, 147–159, 1994.

— *Do we need a separate poverty measurement?* European Journal of Political Economy, volume 18, 61–85, 2002.

Yitzhaki, Shlomo and Robert I. Lerman: *Income stratification and income inequality*. The Review of Income and Wealth, volume 37(3), 313–329, 1991.

Yung, W. and J. N. K. Rao: *Jackknife Variance Estimation Under Imputation for Estimators Using Poststratification Information*. Journal of the American Statistical Association, volume 95(451), 903–915, 2000.

Zheng, Buhong: *An axiomatic characterization of the Watts poverty index*. Economics Letters, volume 42(1), 81–86, September 1993.

— *Can a poverty index be both relative and absolute?* Econometrica, volume 62(6), 1453–1458, November 1994.

— *Aggregate poverty measures.* Journal of Economic Surveys, volume 11(2), 123–162, 1997.

Appendix A

Additional Information on the Evaluation of Imputation Strategies

Table A.1: List of Variables available at household level in the Finnish subsample of the ECHP used for the imputation models

Variable	Label	% Missing
r03hseqn	Household Number	0.00
hinc96	Disposable HH Income	23.02
snkid14	Number of children under 14	0.00
snkid6	Number of children under 6	0.00
snkid3	Number of children under 3	0.00
hhgr96	Household size	0.00
hhsavings	Household savings	0.12
hhvermiet	HH-income from renting property	0.05
hhincsat	HH satisfaction with income	2.39
hhsozh	Income from social assistance payments	0.00
hjwoge	Housing benefits	1.96
nemphh96	Number of employed persons in HH	0.00
psum	Sum of pers. income comp. of all HH members	0.00
partresp	Number of persons without personal interview	0.00

Table A.2: List of Variables available at personal level in the Finnish subsample of the ECHP used for the imputation models

Variable	Label	% Missing
hnro	Personal identification number	0.00
r03hseqn	Household number	0.00
r03work	Working 15+ hours	0.00
occup96	Occupation	0,13
age96	Age	0.00
educ96	Highest completed education	0.00
gend96	Sex	0.00
mast96	Present marital status	0.00
workh96	Working hours per week	0,58
hinc96	Disposable hh income	25.37
nwage96	Net income	1.37
gwage96	Gross income	1.13
ueb96	Unemployment benefits	0.00
xinc96	Any extra payments	0.00
citiz96	Has local citizenship	0.00
finsat96	Satisfaction with financial situation	0.24
auempbenf	Amount unempl. benefits	0.20
auempassi	Amount unempl. assistance	0.15
atrainallo	Amount training allowance	0.05
arehaballo	Amount rehab. alowance	0.00
aouemptrain	Amount other uempl. benefits	0.00
aoldagepen1	Amount oldage pension (first pillar)	1.03
aoldagepen2	Amount oldage pension (second pillar)	0.35
aoldagepen3	Amount oldage pension (third pillar)	0.44
aearlretire	Amount early retirement	0.70
aoldo	Amount other oldage benefits	0.13
awidowpen1	Amount widows pension (first pillar)	0.35
awidowpen2	Amount widows pension (second pillar)	0.02
awidowpen3	Amount widows pension (third pillar)	0.01
aowidow	Amount other widows pension	0.01
aorphanpen	Amount orphans pension	0.05
achildallo	Amount child allowance	0.04
acareallo	Amount care allowance	0.01
amaternallo	Amount maternity allowance	0.09
abirthallo	Amount birth allowance	0.04
aunmarmothal	Amount unmarried mothers allowance	0.00
aofam	Amount other family benefits	0.04
asickhelp	Amount sick help benefits	0.17
aosick	Amount other sick benefits	0.11
aaccidcomp	Amount accidents compensation	0.02
ainvalidpen	Amount invalid pension	0.38
aoinvalid	Amount other invalid pension	0.02
astudyhelp	Amount study benefits	0.09
aobenf	Amount other benefits	0.02
aperssup	Amount personal support	0.07
ysavings	Income savings	0.20
berufstell	Activity status	0.00
hstell	Head of household	0.00
snkid14	Number of children under the age of 14	0.00
snkid6	Number of children under the age of 6	0.00
snkid3	Number of children under the age of 3	0.00

Appendix B

Additional Information on the distribution of High Incomes

Table B.1: Differences in distributions parameters with and without high-income sample (Significance Level: 0.8%)

Year	Samples A-F			Samples A-G		
	Lower bound	Estimate	Upper bound	Lower bound	Estimate	Upper bound
			Median			
2002	1352	1366	1393	1352	1366	1394
2003	1348	1385	1432	1351	1389	1432
2004	1327	1358	1407	1327	1360	1407
			Mean			
2002	1496	1528	1548	1536	1557	1576
2003	1514	1543	1580	1531	1572	1594
2004	1486	1517	1569	1517	1549	1601
			Gini			
2002	0.2421	0.2576	0.2711	0.2568	0.2706	0.2780
2003	0.2405	0.2594	0.2693	0.2597	0.2699	0.2814
2004	0.2470	0.2590	0.2673	0.2572	0.2693	0.2810
			Theil (0)			
2002	0.0977	0.1126	0.1222	0.1137	0.1266	0.1340
2003	0.09561	0.11357	0.12376	0.11286	0.12367	0.13861
2004	0.1008	0.1121	0.1227	0.1093	0.1221	0.1350
			Theil (1)			
2002	0.09731	0.11407	0.12436	0.12842	0.14420	0.15474
2003	0.09552	0.11513	0.13163	0.12018	0.13286	0.15185
2004	0.1023	0.1154	0.1303	0.1145	0.1331	0.1476
			$FGT\ (\alpha = 0)$			
2002	0.09836	0.11644	0.13598	0.09655	0.11642	0.13569
2003	0.1128	0.1333	0.1702	0.1134	0.1335	0.1719
2004	0.09751	0.12517	0.15801	0.09347	0.12247	0.15377
			$FGT\ (\alpha = 1)$			
2002	0.0230	0.0283	0.0340	0.0228	0.0283	0.0341
2003	0.0237	0.0294	0.0377	0.0236	0.0297	0.0390
2004	0.0206	0.0273	0.0344	0.0196	0.0268	0.0350
			$FGT\ (\alpha = 2)$			
2002	0.0086	0.0106	0.0137	0.0085	0.0106	0.0138
2003	0.0078	0.0106	0.0142	0.0078	0.0107	0.0145
2004	0.0070	0.0094	0.0121	0.0069	0.0092	0.0124

Source: *SOEP* 2002 to 2004, author's calculations.

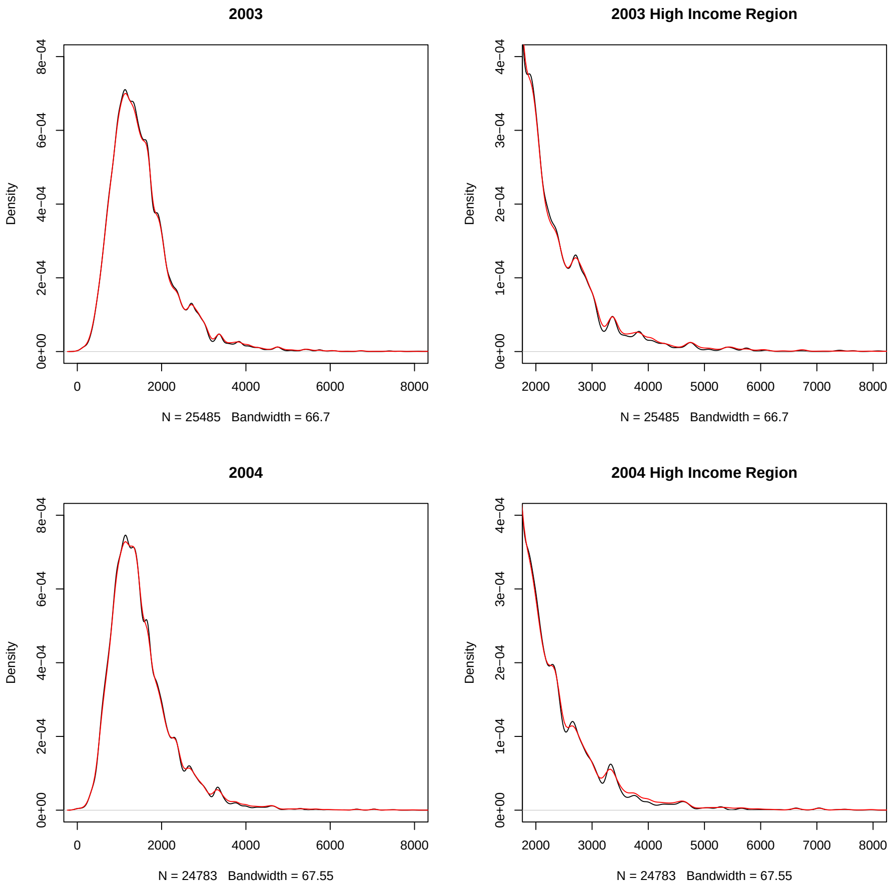


Figure B.1: Non-parametric kernel density with and without high-income sample

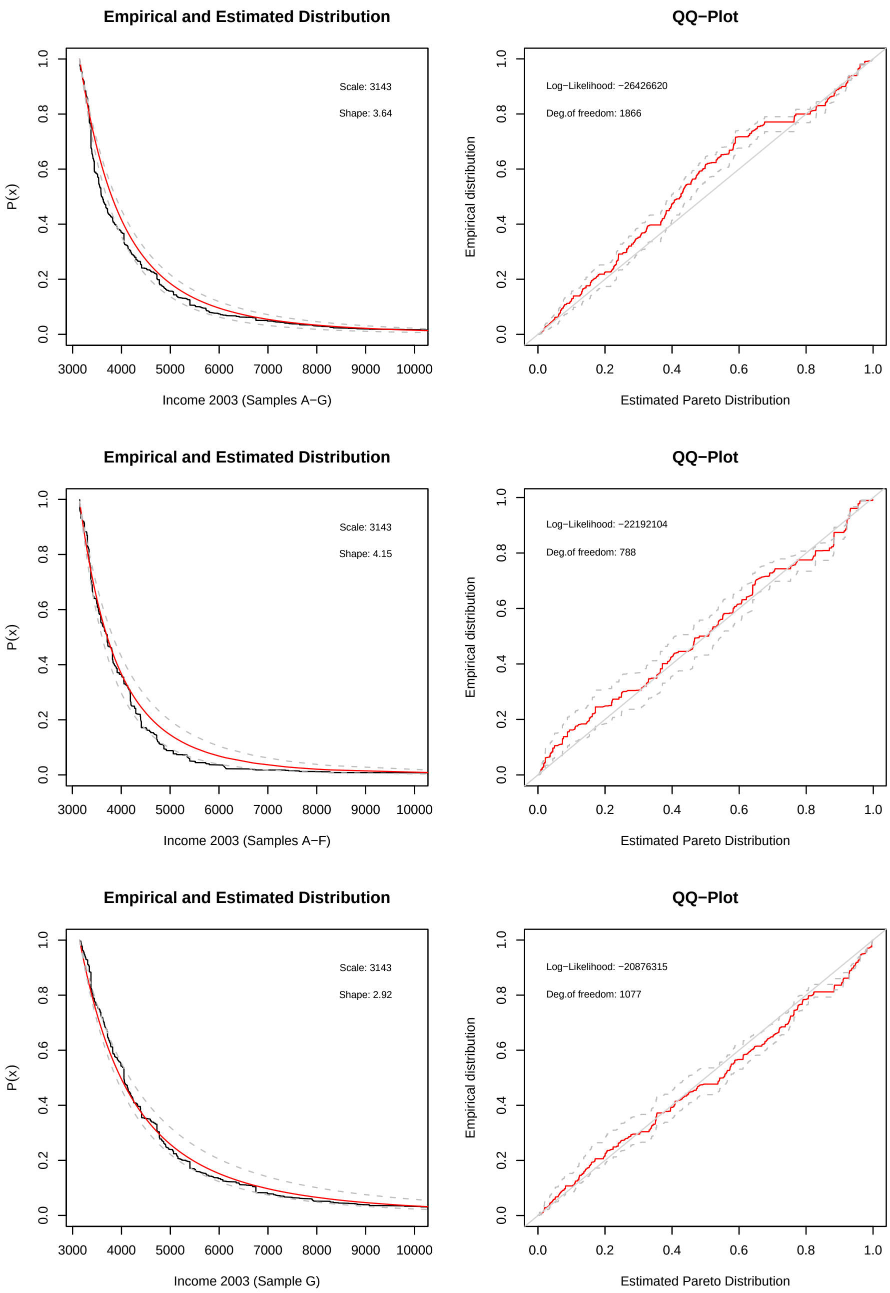


Figure B.2: Estimated and empirical distribution of high incomes (year: 2003)

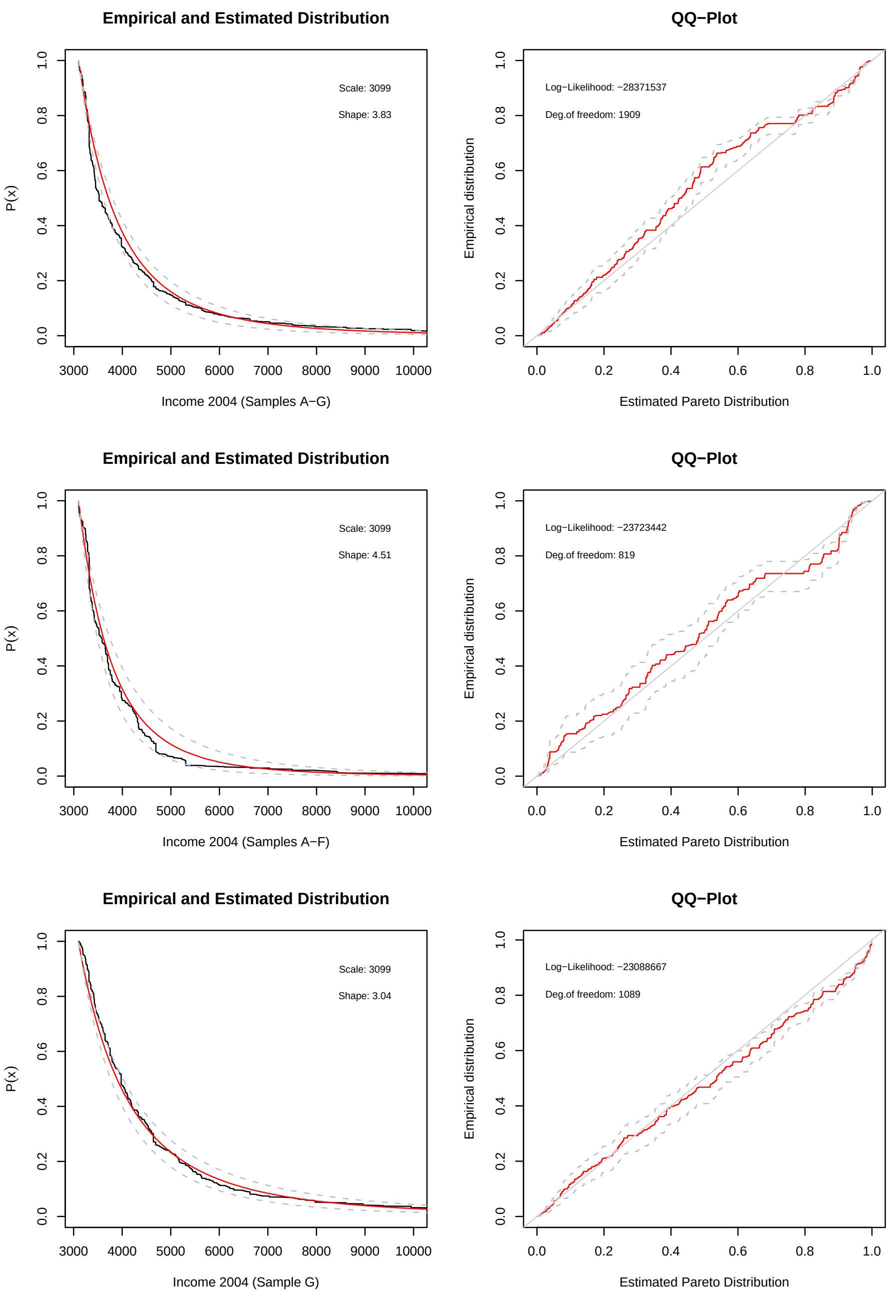


Figure B.3: Estimated and empirical distribution of high incomes (year: 2004)

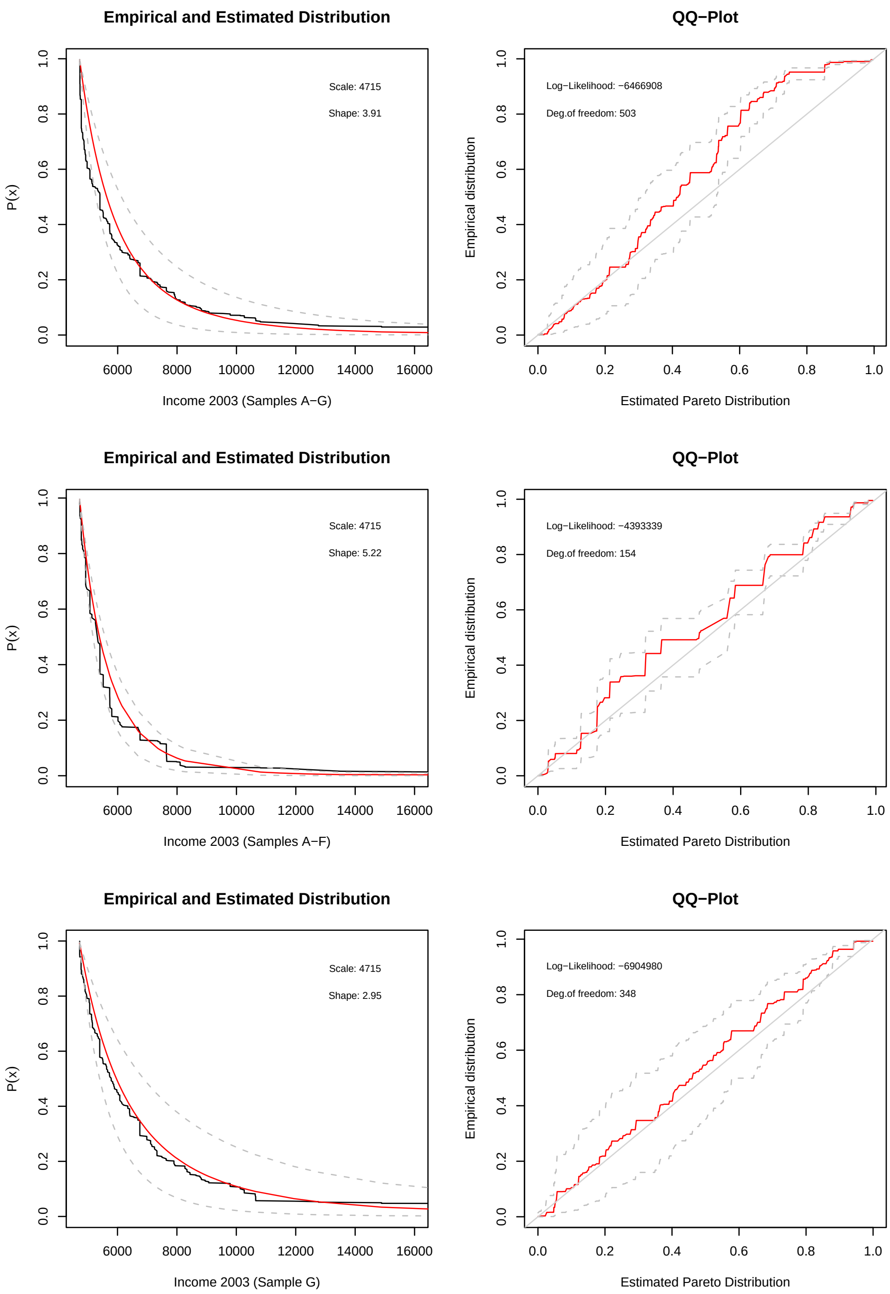


Figure B.4: Estimated and empirical distribution of very high incomes (year: 2003)

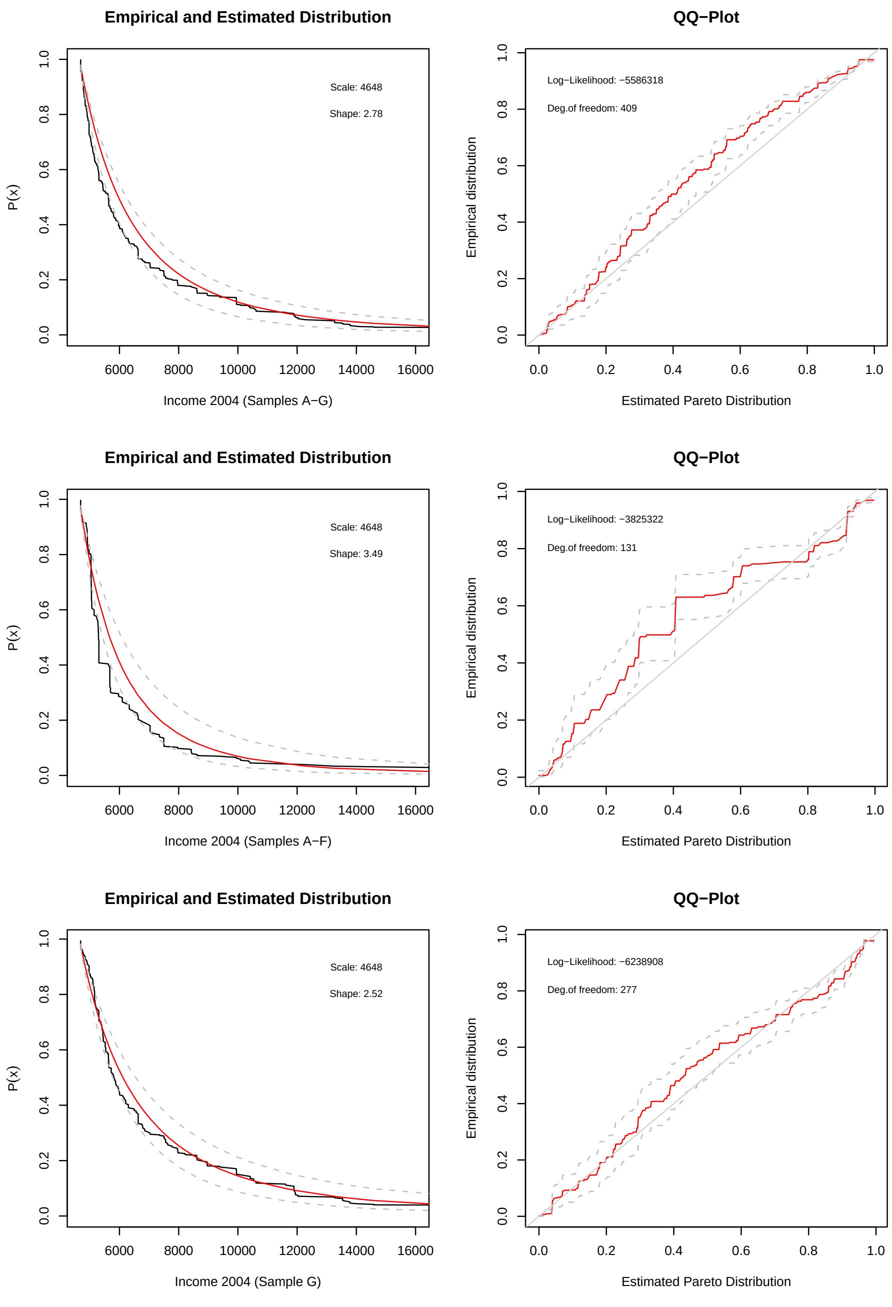


Figure B.5: Estimated and empirical distribution of very high incomes (year: 2004)

Table B.2: Shape coefficients of the estimated Pareto distributions (Significance Level: 0.8%)

	Samples A-F			Sample G		
Year	Lower bound	Shape	Upper bound	Lower bound	Shape	Upper bound
	High income (200% Mean)					
2002	3.55	4.28	5.52	2.36	2.81	3.37
2003	3.25	4.15	5.06	2.17	2.92	4.36
2004	3.25	4.51	6.91	2.62	3.04	3.60
	Very high income (300% Mean)					
2002	3.38	4.59	7.00	1.42	2.03	3.02
2003	2.80	5.22	7.94	1.81	2.95	6.17
2004	1.90	3.49	7.16	1.52	2.52	4.06

Table B.3: Gini coefficients of the estimated Pareto distributions (Significance Level: 7%)

	Samples A-F				Sample G			
Year	Empirical Gini	Lower Bound	Gini	Upper Bound	Empirical Gini	Lower Bound	Gini	Upper Bound
	High Income (200% Mean)							
2002	0.12	0.12	0.13	0.15	0.27	0.19	0.22	0.26
2003	0.12	0.11	0.14	0.17	0.22	0.18	0.21	0.26
2004	0.13	0.09	0.12	0.16	0.21	0.16	0.20	0.23
	Very High Income (300% Mean)							
2002	0.14	0.08	0.12	0.17	0.33	0.21	0.33	0.49
2003	0.12	0.07	0.11	0.14	0.26	0.11	0.20	0.37
2004	0.17	0.13	0.17	0.24	0.26	0.19	0.25	0.33

Table B.4: Analysis of Gini for high-income earners by sample (2002 to 2004)

			Year 2002				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	*SGO*
Samples A-F	0.4459	3969	0.4518	0.3893	0.1236	0.9965	0.048
Sample G	0.5541	5011	0.5388	0.6107	0.2663	0.9586	0.1559
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	0.994					
Sample G	0.907	1.000					
Gini	Overall	Between					
	0.2137	0.0099					
			Year 2003				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	*SGO*
Samples A-F	0.5501	4054	0.4641	0.5065	0.1156	0.9725	0.0569
Sample G	0.4499	4830	0.5439	0.4935	0.2201	0.976	0.106
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	0.939					
Sample G	0.956	1.000					
Gini	Overall	Between					
	0.1699	0.0070					
			Year 2004				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	*SGO*
Samples A-F	0.5512	3975	0.448	0.5105	0.134	1.0336	0.0707
Sample G	0.4488	4680	0.5639	0.4895	0.2141	0.9166	0.096
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	1.075					
Sample G	0.849	1.000					
Gini	Overall	Between					
	0.1762	0.0094					

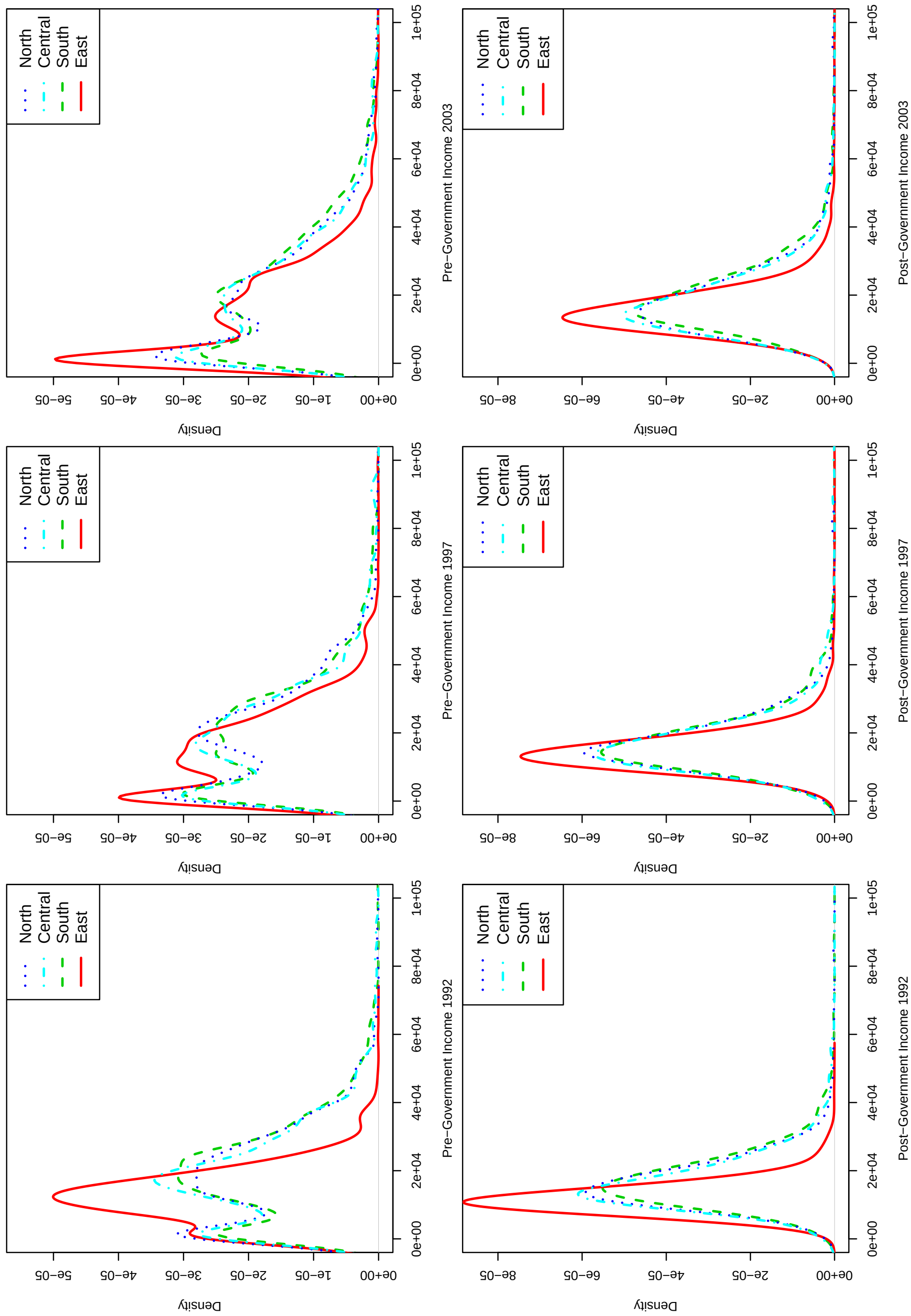
Table B.5: Analysis of Gini for very high income earners by sample (2002 to 2004)

			Year 2002				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	SGO
Samples A-F	0.3354	6026	0.4039	0.2541	0.1354	0.962	0.0331
Sample G	0.6646	8924	0.5485	0.7459	0.333	0.9455	0.2348
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	0.943					
Sample G	0.838	1.000					
Gini	Overall	Between					
	0.2914	0.0235					
			Year 2003				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	SGO
Samples A-F	0.4198	5759	0.4114	0.3601	0.1238	1.0399	0.0464
Sample G	0.5802	7403	0.5641	0.6399	0.2566	0.9151	0.1502
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	1.069					
Sample G	0.798	1.000					
Gini	Overall	Between					
	0.2148	0.0182					
			Year 2004				
Group	%	Mean	F_{0i}	S_i	G_i	O_i	SGO
Samples A-F	0.3991	6663	0.4706	0.3621	0.172	1.0098	0.0629
Sample G	0.6009	7795	0.5195	0.6379	0.2621	0.9867	0.1649
O_{ji}	Samp. A-F	Samp. G					
Samples A-F	1.000	1.016					
Sample G	0.967	1.000					
Gini	Overall	Between					
	0.2314	0.0036					

Appendix C

Additional Information on Regional Stratification

Figure C.1: Kernel Density Estimates by region and year (Bandwith=2500)



Appendix D

List of Tables, Figures, Symbols and Index

List of Symbols and Abbreviations

$\mathbf{Y}$	Permanent or Smoothed Income
μ	Mean
$\overline{W}_M$	Within-Imputation Variance
σ	Variance
$\widehat{B}_M$	the Between-imputation variance
ζ_p	Representative income of the poor for an arbitrary social welfare function W
ζ_z	Representative Income of the censored income distribution
B	Bonferroni inequality index
BD	Poverty Measure proposed by Blackorby and Donaldson (1980)
C	Poverty Measure proposed by Chakravarty (1983a)
C_h	Poverty Measure proposed by Chakravarty (1983b)
CHU_1	Poverty Measure proposed by Clark et al. (1981)
CHU_2	Poverty Measure proposed by Clark et al. (1981)
CV	Squared Coefficient of Variation
FGT	Poverty Measure proposed by Foster et al. (1984)
FGT	Poverty Measure proposed by Foster et al. (1984)
G	Gini Inequality Measure
H	Head count ratio
HI	Poverty Gap ratio
I	Income gap ratio
K	Poverty Measure proposed by Kakwani (1980b)
n	Number of Persons, Population Size
P	General Poverty Measure
S	Poverty Measure proposed by Sen (1976)
S_s	Poverty Measure proposed by Shorrocks (1995)
S_{GC}	Poverty Measure proposed by Giorgi and Crescenzi (2001)
SK	Poverty Measure proposed by Kakwani (1980a)
T	Poverty Measure proposed by Thon (1979)
T_α	Poverty Measure proposed by Takayama (1979)

U	Utility Function
V_a	Absolute Poverty Measure proposed by Vaughan (1987)
V_r	Relative Poverty Measure proposed by Vaughan (1987)
W	Social Welfare Function
W_p	Poverty Measure proposed by Watts (1968)
x or y	Vector, typically of Incomes
z	Poverty Line
ANOGI	Analysis of Gini
CNEF	Cross-National Equivalent Files
DA	Data Augmentation
ECHP	European Community Household Panel
EM	Expectation-Maximization
EU	European Union
ice	MICE implementation for Stata by Royston (Successor of MVIS)
IDS	Income Distribution Survey
IVE or IVEware	Imputation and Variance Estimation Software by Raghunathan et al.
LIS	Luxembourg Income Study
MAR	Missing At Random
MCAR	Missing Completely At Random
MICE	Multivariate Imputation by Chained Equations, Imputation Software by van Buuren and Oudshoorn
MSE	Mean Squared Error
MVIS	MICE implementation for Stata by Royston
NMAR	Not Missing At Random
NORM	Imputation Software written by Schafer
NTX	N Times X Approach: Measures the proportion of persons with incomes below the poverty line in x out of n time periods.
OECD	Organisation for Economic Co-operation and Development
PRSP	Poverty Reduction Strategy Paper
PSID	Panel Study of Income Dynamics
SOEP	Socio-Economic Panel

List of Tables

1.1 Overview about which axioms are satisfied by each poverty measure 18

2.1 Poverty Stability for West-Germany (1985-2002) with increasing periods for mean and median based poverty line 43

2.2 Poverty Stability for Germany (1997-2002) with increasing periods for monthly and annual net-equivalent income. 45

3.1 Frequency distribution of the cluster variable used with hot-deck (Personal Level) 69

3.2 Frequency distribution of the cluster variable used with hot-deck (Household Level) 69

3.3 Imputation Software Overview 76

4.1 SOEP Sample sizes over time (individual level, including children, 2002-2004) 97

4.2 Mean and median by sample (2002-2004) 98

4.3 Differences in distribution parameters with and without the high-income sample 100

4.4 High-income cases by sample and year (more than 200% of contemporary mean) 103

4.5 High-income cases by sample and year (more than 300% of contemporary mean) 103

4.6 Differences in the distribution of high income between Samples A-F and Sample G (More than 200% of contemporary mean) . . 104

4.7 Differences in the distribution of high income between Samples A-F and Sample G (More than 300% of contemporary mean) . . 105

4.8 Overview: Model fit of the estimated Pareto distributions 114

4.9 Overview: Shape coefficients of the estimated Pareto distributions 118

4.10 Overview: Gini coefficients of the estimated Pareto distributions 118

5.1 Grouping of federal states, population size and GDP 125

A.1 List of Variables available at household level in the Finnish subsample of the ECHP used for the imputation models 167

A.2 List of Variables available at personal level in the Finnish subsample of the ECHP used for the imputation models 168

B.1 Differences in distributions parameters with and without high-income sample (Significance Level: 0.8%) 170

B.2 Shape coefficients of the estimated Pareto distributions (Significance Level: 0.8%) 176

B.3 Gini coefficients of the estimated Pareto distributions (Significance Level: 7%) 176

B.4 Analysis of Gini for high-income earners by sample (2002 to 2004) 177

B.5 Analysis of Gini for very high income earners by sample (2002 to 2004) 178

List of Figures

2.1 Behavior of median and mean using the smoothed income concept 41

2.2 Poverty stability for monthly and annual income between 1992 and 2002 in Germany 45

2.3 Poverty stability in Germany and the USA for post-government income (1985-1989; 1993-2001) 47

2.4 Poverty stability in Germany and the USA for pre-government income (1985-1989; 1993-2001) 48

2.5 Poverty stability in Germany and the USA for a shifting time period of three years for Pre- and Post-Government income (1984-2001)) 50

3.1 Schematic concept of multiple imputations 58

3.2 General Overview of the Imputation Procedure 67

3.3 Differences between Imputations and Register (hot-deck) 77

3.4 Income distribution density using hot-deck imputation 78

3.5 Differences between Imputations and Register (NORM) 79

3.6 Income distribution density using NORM imputation 80

3.7 Differences between Imputations and Register (aregImpute) . . . 80

3.8 Income distribution density using aregImpute Imputation 81

3.9 Differences between Imputations and Register (MICE) 82

3.10 Income distribution density using MICE Imputation 83

3.11 Differences between Imputations and Register (mvis) 84

3.12 Income distribution density using mvis imputation 84

3.13 Differences between Imputations and Register (IVEware) 85

3.14 Income distribution density using IVEware imputation 86

3.15 Impact of different imputation methods on mean, median, standard deviation, Gini and Theil 88

3.16 Impact of different imputation methods on the Atkinson inequality measure for different values of the parameter ϵ 90

3.17 Impact of different imputation methods on the FGT poverty measure, for different values of the parameter α 91

4.1 Quantiles with and without high-income sample (all cases) 99

4.2 Non-parametrical kernel density with and without high-income sample (all cases) 101

4.3 Quantiles of high-income cases for Samples A-F vs. G 107

4.4 Non-parametric kernel density for high-income cases (Samples A-F and Sample G, 200% of mean income) 108

4.5 Pareto distribution with different location and shape parameters 109

4.6 Shape parameter of the Pareto distribution and corresponding Gini index 110

4.7 Bias and MSE for the estimated Gini index using different shape parameter of the Pareto distribution 112

4.8 Estimated and empirical distribution of high incomes (Year 2002) 116

4.9 Estimated and empirical distribution of very high incomes (Year 2002) 117

5.1 ANOGI-Results on pre-government income: East vs. West Germany, 1992-2003 131

5.2 ANOGI-Results on post-government income: East vs. West Germany, 1992-2003 133

5.3 ANOGI-Results on pre-government income: Extended regional grouping, 1992-2003 136

5.4 Overlapping for each group with each other group (O_{ji}): Pre-Government Income, 1992-2003 137

5.5 ANOGI-Results on post-government income: Extended regional grouping, 1992-2003 139

5.6 Overlapping for each group with each other group (O_{ji}): Post-Government Income, 1992-2003 140

B.1 Non-parametric kernel density with and without high-income sample 171

B.2 Estimated and empirical distribution of high incomes (year: 2003) 172

B.3 Estimated and empirical distribution of high incomes (year: 2004) 173

B.4 Estimated and empirical distribution of very high incomes (year: 2003) 174

B.5 Estimated and empirical distribution of very high incomes (year: 2004) 175

C.1 Kernel Density Estimates by region and year (Bandwith=2500) . 180

Index

A

additive regressions 71
aggregation step 9
AIM-AP 148
AIM-AP (2006) 148
Alzola and Harrell (2004) 71
Anand (1977) 12
Approximate Bayesian bootstrap (ABB) 68
aregImpute 71, 76
Atkinson (1970) 24, 29
Atkinson (1987) 28
Atkinson (2002) 3
Atkinson and Bourguignon (2000b) 148
Atkinson and Leigh (2005) 6, 95
Atkinson et al. (2002) 2, 65
Atkinson Measure 28

B

Bandourian (2000) 111, 113
Bandourian et al. (2002) . . . 111, 113
Bane and Ellwood (1986) 3
Barnard and Rubin (1999) 60
Bartelheimer (2004) 2
Becker et al. (2002) 123
Bernaards et al. (2003) 65
Berthoud (2004) 124
Betti et al. (2005) 9, 16
bias 112
biased inference 53
Biewen (2001) 123, 130
Bird et al. (1998) 123
Bishop et al. (2001) 123
Blackburn (1989) 28
Blackorby and Donaldson (1980) . 24, 26, 32
BMAS (2001) 2, 6, 95
BMAS (2005) 2
bootstrapping 71
Bourguignon and Chakravarty (2002) 36
Brenke (2005) 123, 124
Burgess and Propper (1998) 3
Burkhauser and Poupore (1997) . . 36, 47
Burkhauser et al. (1996) 65
Burkhauser et al. (1997) 5
Burkhauser et al. (1998) 47
Burkhauser et al. (1999) 129
Burkhauser et al. (2001) 46
Böheim and Jenkins (2000) 44
Büchel and Frick (2001) 124

C

Canberra-Group (2001) 128
Casella and George (1992) 72
Censored Income Distribution 13, 22
Chakravarty (1983a) . . . 15, 23, 25–27
Chakravarty (1983b) 30–32
Chakravarty (1990) 24, 32
Chakravarty (1997) 10
Chakravarty and Muliere (2004) . . 24
Cheli and Lemmi (1995) 9, 16
chronic poverty 35
Clark et al. (1981) 21, 22, 25–29, 31, 32
Clemenceau and Wirtz (2001) 54
Clementi and Gallegati (2005a) . 111, 115
Clementi and Gallegati (2005b) . 110, 111
CNEF 36, 46, 111
Colavecchio et al. (2005) 141
Coniffe (1992) 65
Cowell (1988) 15
Cowell (2003) 10
Cowell and Flachaire (2002) 111
Cowell and Victoria-Feser (1996) 106, 111
Cross-National Equivalent Files . . *see* CNEF

D

D'Ambrosio et al. (2002) 41
DA 70
Dalton (1920) 13
Data Augmentation .... *see* DA
Deprivation 12
deterministic imputation method . 68
Devicienti (2002) 3
De Vos and Zaidi (1997) 65
Dickey (2001) 126
Dietz (1997) 10
Diggle (1993) 55
Diggle (1994) 55
Donaldson and Weymark (1986) . 11, 13, 14
Drăgulescu and Yakovenko (2001) 111
Duncan et al. (1993) 3, 36
Durrant (2005) 54, 68, 72, 145

E

ECHP 54, 55, 63–65
EM 70
Esping-Andersen (1990) 45
EUREDIT-Project (2000) 65
European Community Household Panel *see* ECHP
EUROSTAT (2003) 2
EUROSTAT (2004) 2, 19
Expectation-Maximization .. *see* EM

F

Fabig (1999) 36
Fay (1996) 57
Feldstein (1998) 21
FGT .... *see* Poverty Measures
Foster (1984) . 3, 9, 10, 12, 14–17, 19, 22, 33, 39, 41, 143
Foster and Shorrocks (1991) 16, 23, 30
Foster et al. (1984) . 3, 16, 17, 19, 26, 27, 30–33, 36, 39, 40, 42, 85, 91, 99, 143
fractional imputation 54
Frick and Grabka (2004) . 6, 120, 128
Frick and Grabka (2005) .. 6, 66, 120, 128
Frick et al. (2005) 123, 130
Frick et al. (2006) 6, 92, 96, 101, 115, 119, 122, 126, 127, 145, 147
Friedrichs et al. (1986) 7, 121
Förster et al. (2002) 124

G

Garfinkel et al. (2006) 148
generalized Beta distribution 110
Geppert (1999) 121, 123
German Socio-Economic Panel ... *see* SOEP
Gibrat (1931) 110
Gibrat index 110
Gini Coefficient 20, 22, 28
Giorgi and Crescenzi (2001) 23
Goebel and Kuchler (2003) 19
Goebel et al. (2005) 49, 123
Goodin et al. (1999) 46
Grabka et al. (1999) 124

H

Hagenaars (1987) 10, 32
Hagenaars and van Praag (1985) . 41
Haisken-DeNew and Frick (2005) .. 5, 96, 128
Hartmann and Schimpl-Neimanns (1992) 5
Hauser et al. (1994) 128
Heitjan (1994) 55
Heitjan and Basu (1996) 55
Hill and Jenkins (2001) 19, 35
Hmisc 71
Hoogland and Pannekoek (2000) . 65
Horovitz and Manski (1998) 55
Horovitz and Manski (2000) 55
hot-deck 68–70, 76–78
Hunt (2001) 130

I

identification step 9
ignorability 54
imputation flag 57
Isengard (2002) 6
Ishikawa (2004) 111
item nonresponse 53
IVEware 59, 68, 71–73, 76

J

Jenkins and Lambert (1993) 23
Jenkins and Lambert (1997) 23
Jenkins and Lambert (1998) 23
Jenkins and van Kerm (2003) 47
Jäntti and Danziger (2000) 37

K

Kakwani (1980a) 26
Kakwani (1980b) .. 14, 15, 21, 22, 25
Kakwani (1981) 23
Kakwani (1997) 21
Kenward and Molenberghs (1998) 55
Kim and Fuller (1999) 54
Kim and Fuller (2004) 54
Klass et al. (2006) 111
Kockläuner (2002) 24, 26

Krause (2003) 123
Krause et al. (2003) 2, 46
Kronauer (1997) 35
Krämer and Ziebach (2002) 110
Kundu and Smith (1983) 14

L

Lammers (2003) 122, 123
Lasswell (1965) 127
Leibfried and Voges (1992) 3
Leibfried et al. (1995) 3
Lerman and Yitzhaki (1984) 122, 126
Lerman and Yitzhaki (1989) 127
Lewis and Ulph (1988) 21
Li et al. (1999) 61
Linux 72
Lipsmeier (1993) 6
Little and Rubin (2002) ... 4, 55, 56, 59–61, 72
Loikkanen et al. (2003) 124

M

Maasoumi and Trede (2001) 47
Madden (2000) 41
Mandelbrot (1960) 110
Mander and Clayton (1999) 68
MAR . *see* missing at random, 56, 69, 73
MCAR .... *see* missing completely at random, 56, 69, 74
McDonald (1984) 110, 113
mean squared error 112
Meng (1994) 57, 61, 62
Meng (2002) 59
Mercader-Prats and Levy (2004) 124
Mesa et al. (2000) 65
MICE 59, 73, 75, 76
missing at random 4, 55
missing completely at random . 4, 55
missing mechanism 53, 55, 56
missing mechanisms 54
missingness by design 53
monotone missing data pattern ... 72
Moothathu (1985) 110, 115
Muller (2001) 28
multiple imputation 54
MVIS 75, 76

N

N-times-X approach 35
Nicoletti and Peracchi (2003) 55
NMAR ... *see* not missing at random
non-cash income 148
NORM 68, 70–71, 76
not missing at random 55
Nygård and Sandström (1989) ... 111

O

observed data 53
Opsomer and da Silva (2002) 54
Opsomer and da Silva (2004) 54
Osberg and Xu (2000) 35

P

Panel Study of Income Dynamics *see* PSID
Pareto distribution 110
Pareto index 109
Pareto law 109
Pareto-Lévy-Law 110
Pattanaik and Sengupta (1995) 10, 20
pattern of missing 55
permanent income .... *see* smoothed income
Piachaud (1992) 3
Pischner (2004) 98, 99
Pittau (2005) 124
Poverty Axioms 11–17
- Continuity Axiom 15
- Focus Axiom 11, 13
- Minimal Transfer Axiom 13
- Monotonicity Axiom 11, 13
- Monotonicity Sensitivity Axiom 15
- Progressive Transfer Axiom . 14
- Regressive Transfer Axiom .. 13
- Replication Invariance Axiom 15
- Strong Monotonicity 13
- Subgroup Consistency Axiom 16
- Subgroup Decomposability Axiom 16
- Symmetry Axiom 14
- Transfer Axiom 11
- Transfer Sensitivity Axiom .. 15
- Weak Monotonicity Axiom .. 13
- Weak Transfer Axiom 13
- Weak Transfer Sensitivity Axiom 15

Poverty Deprivation Function 17
Poverty Gap 13
poverty gap ratio 25
Poverty Level 11
Poverty Measure 9, 10, 17–33
- Blackorby & Donaldson (1980) 25
- Chakravarty (1983) 26
- Chakravarty (1983:2) 30
- Clark, Hemming & Ulph (1981:1) 25
- Clark, Hemming & Ulph (1981:2) 29
- Definition 11
- Distribution-Insensitive 19
- Distribution-Sensitive 17, 27

Ethical Poverty Measures 24
Foster, Greer & Thorbecke (FGT, 1984) 30
Giorgi & Crescenzi (2001) ... 23
Hagenaars (1987) 32
Head-count ratio 19
Income gap ratio 19
Kakwani (1980a) 22
Kakwani (1980b) 26
Kockläuner (2002) 26
Poverty gap ratio 19
Schmid (1993) 31
Sen (1976) 20
Sen-Kakwani ...... *see* Kakwani (1980b)
Sen-Type Poverty Measures . 20
Shorrocks (1995) 23
Subgroup Consistent Poverty Measures 27
Takayama (1979) 22
Thon (1979) 21
Vaughan (1987) 27
Watts (1968) 28
poverty stability 3
Poverty Value 11
predictive mean matching 71
PSID 36, 46
Pyatt (1976) 126
Pyatt (1987) 21
Pötter and Rendtel (1993) 5

Q

QQ-Plot 111, 114

R

R Development Core Team (2005) 113
Raghunathan et al. (2001a) 59, 71, 72
Raghunathan et al. (2001b) ... 59, 71
Raghunathan et al. (2002) 59, 71
random variable 55
Rao (1996) 57
Ravallion (1988) 35
Ravallion (1998) 35
Rendtel (1995) 99
Rendtel et al. (2004) 64, 65
Representative Income 24, 27
Robins and Wang (2000) 57, 59
Rodgers and Rodgers (1991) 19
Rodgers and Rodgers (1993) .. 35–38
Rohwer (1992) 3
Rowntree (1901) 19
Royston (2004) 75
Royston (2005) 75
Rubin (1976) 55
Rubin (1978) 55
Rubin (1987) . 4, 5, 54, 57, 59–63, 76, 86
Rubin (1996) 4, 56, 57, 59, 61
Rubin and Schenker (1986) 68

S

Sallila et al. (2006) 12
SAS 68, 72
Schafer (1997) 56, 59, 70, 72, 74
Schafer (1999) 68, 72
Schafer and Schenker (2000) 57
Schaich and Münnich (1996) 9
Scharfstein et al. (1999) 56
Schechtman (2005) 127
Schechtman and Yitzhaki (1987) 127
Schechtman and Yitzhaki (1999) 127
Schluter and Trede (2003) 47
Schmid (1993) 18, 31, 32
Schupp and Wagner (1995) 5
Schupp et al. (2003) 97, 103
Schupp et al. (2005) 97
Schwarze (2003) 65
Schäuble (1984) 3
Seidl (1988) 3, 10, 13, 30
Sen (1976) 9–15, 20, 21, 23, 27, 30, 33, 143, 144
Sen (1981) 2, 9, 11, 14
Sen (1992) 17
Shao (2000) 57
Shih (1992) 55
Shorrocks (1978) 3, 35, 36, 38, 39, 144
Shorrocks (1995) .. 10, 21, 23, 26, 33, 41, 144
Shorrocks (1998) 23, 33
Shorrocks and Wan (2005) 6
Shorrocks-Index 36
Shorrocks-Stability-Index .... *see* Shorrocks-Index
Singh and Maddala (1976) 110
Singh-Maddala distribution 110
single imputation 54
smoothed income 27, 35, 36
smoothed income gap 3
SOEP 5, 6, 36, 43, 45, 46, 66, 95–105, 111, 114, 118–120, 146, 147, 170
Subsamples 96
SOEP Group (2001) 5
Spiess (2005) 61
Spiess and Goebel (2004) 53, 93, 146
Spiess and Goebel (2005) 53, 93, 146
SPSS 68
Stewart (2002) 124, 134

stochastic imputation method 68
SVR (2004) 125

T

Takayama (1979) 10, 22, 23, 25, 26, 41
Theil (1967) 28, 104, 122
Thon (1979) 21, 23, 33, 144
Thon (1981) 21
Thon (1983a) 14
Thon (1983b) 15, 21
transitory poverty 35
Tsakloglou and Panopoulou (1998) 65
Tungodden (2005) 31

U

unit nonresponse 53

V

Valletta (2006) 35, 37
van Buuren and Oudshoorn (1999) 59, 72–74
van Buuren and Oudshoorn (2000) 59
van Kerm and Jenkins (2003) 47
Vaughan (1987) 27
Victoria-Feser (1999) 106
Victoria-Feser and Ronchetti (1997) 111
von Hippel (2004) 68
von Wiese (1954) 3

W

Wagner et al. (1993) 128
Walker (1998) 3, 36
Walker and Ashworth (1992) ... 3, 35
Watts (1968) 9–11, 16, 19, 27, 30, 35
Weibull distribution 110
Weikard (2000) 35
Wiggins et al. (2000) 65
Williams and Bailey (1996) .. 93, 145
Windows 72
Wirtz and Mejer (2002) 54

X

Xu and Osberg (2001) 33, 144

Y

Yee (2006) 113
Yitzhaki (1982) 128
Yitzhaki (1991) 127
Yitzhaki (1994) 122
Yitzhaki (2002) 21
Yitzhaki and Lerman (1991) 17, 122, 126
Yung and Rao (2000) 57

Z

Zheng (1993) 9, 28
Zheng (1994) 10, 33
Zheng (1997) 3, 10–15, 17, 18, 20, 22, 23, 25–32, 36, 39, 41