EUROPEAN COMMISSION
Community Research

SEVENTH FRAMEWORK PROGRAMME

AMELI

Advanced Methodology for European Laeken Indicators

Deliverable 1.1

State-of-the-art of Indicators on Poverty and Social Exclusion – the Laeken Indicators

Version: 2011

Monique Graf, Andreas Alfons, Christian Bruch, Peter Filzmoser, Beat Hulliger, Risto Lehtonen, Bernhard Meindl, Ralf Münnich, Tobias Schoch, Matthias Templ, Maria Valaste, Ariane Wenger, and Stefan Zins

The project **FP7–SSH–2007–217322 AMELI** is supported by European Commission funding from the Seventh Framework Programme for Research.

http://ameli.surveystatistics.net/

Contributors to deliverable 1.1

Chapter 1: Monique Graf and Ariane Wenger, Swiss Federal Statistical Office

Chapter 2: Maria Valaste and Risto Lehtonen, Department of Mathematics and Statistics, University of Helsinki, Finland.

Chapter 3: Monique Graf, Swiss Federal Statistical Office.

Chapter 4: Ralf Münnich, Stefan Zins, and Christian Bruch, University of Trier, Germany.

Chapter 5: Beat Hulliger and Tobias Schoch, University of Applied Sciences North-Western Switzerland.

Chapter 6: Andreas Alfons, Peter Filzmoser, Matthias Templ, Vienna University of Technology, Austria, Beat Hulliger and Tobias Schoch, University of Applied Sciences North-Western Switzerland, Bernhard Meindl, Statistics Austria.

Main responsibility

Monique Graf, Swiss Federal Statistical Office

Evaluators

Internal expert: Beat Hulliger, University of Applied Sciences North-Western Switzerland.

External expert: Asterios Chatziparadeisis, Ministry of Development, Athens, Greece.

Aim and Objectives of Deliverable 1.1

The main target of the project is to review the state-of-the-art of existing indicators monitoring the multidimensional phenomena of poverty and social exclusion – the Laeken Indicators – including their relation to social cohesion. Special emphasis will be put on methodological aspects of indicators and especially on their impact on policy making. This will include quality aspects as well as mathematical and statistical properties within a framework of a complex survey in the context of practical needs and peculiarities. The relevance and importance of these methods can be drawn from the methodological principals of the Social Protection Committee (SPC, 2001: report on indicators in the field of poverty and social exclusion).

The work package's output will present relevant methodological aspects to be considered within the AMELI project. Special emphasis will be put on practical issues, such as the impact of outliers and possible gaps of statistical adequacy. The study will consider results from other projects in the AMELI area, e. g. the KEI project (http://www.kei.publicstatistics.net/) and other Framework Programms projects focusing on Laeken indicators, indicator methodology or quality measurement.

The work considers three main aspects:

1. State-of-the-art of Laeken Indicators and the corresponding methodology;
2. Review of the latest developments and investigation of its impact on AMELI, including the definition of social cohesion;
3. Impact of AMELI methodology on other indicator areas.

First, Chapter 1 provides a panorama of the context in which the European Union Statistics on Income and Living Conditions (EU-SILC) database has been constituted. The topics treated include the definition of social cohesion, the rationale for the Laeken indicators and the need for a harmonized definition of income. This is followed by a presentation of former international databases with relation to EU-SILC, current results on Laeken Indicators and a review of European research projects having a potential impact on AMELI.

Five accompanying thematic summaries follow, giving a quick overview of the methodology that will be developed in the AMELI project. Chapter 2 presents a review on the state-of-the-art in small area estimation of Laeken Indicators. Chapter 3 gives a summary of the state-of-the-art in parametric income distributions. Afterwards Chapter 4 gives an overview of variance estimation methods. Next, in Chapter 5, there is a summary of the state-of-the-art on robustness methods for Laeken Indicators. Finally, in Chapter 6 a summary of the state-of-the-art on visualization of Laeken Indicators is given.

Contents

Chapter 1

Bibliography on Laeken indicators

1.1 Introduction

This chapter provides a panorama of the context in which the database on Social and Living Conditions in Europe (EU-SILC) has been constituted. The rationale for the Laeken indicators, the need for a harmonized definition of income permitting well founded international comparisons, the technical difficulties related to skew distributions and the existence of extreme values, the way they were addressed in the past, the studies based on former databases like the European Household Panel (ECHP, see Section 1.4.1) constitute a striking picture of all the work that has already been achieved towards a reliable measurement of social cohesion.

The chapter is organized as follows: The context in which the set of indicators has been derived, the Open Method of Coordination (OMC), will be briefly described (Section 1.2.1). The policies within the framework of the OMC in Social Inclusion are outlined in Sections 1.2.2 and 1.2.3. Further, in Section 1.2.4, we give the definition of social cohesion as stated in the European Council's Methodological Guide Council of Europe Publishing (2005). Next, the Laeken indicators are presented. They were devised by the Social Protection Committee and analyzed in the *Statistics in Focus series* devoted to the Laeken indicators and issued by Eurostat.
The methodological studies around the concept of income are presented in Section 1.3. In Section 1.4 other databases related to EU-SILC are described.
Section 1.5 is devoted to current results on Laeken indicators. These results include quality aspects (i.e. comparability, extremes). International comparison studies involving countries outside Europe can be found in Section 1.5.5. We thank Beat Hulliger (FHNW) who is the author of Section 1.5.6 on composite indicators.
In Section 1.6 a list of recent research projects with potential impact on AMELI is provided.

1.2 Rationale of the Indicators

Laeken indicators are an essential element of the Open Method of Coordination (OMC) in the context of social cohesion. This section introduces the OMC in general, then the

OMC in social protection and social inclusion. The definition of social cohesion is the one given by the Council of Europe's Methodological Guide. These instruments are the rationale for Laeken indicators.

1.2.1 Open Method of Coordination

At the Lisbon European Council in March 2000, Lisbon Summit, March 2000 the Union has set itself the strategic goal *to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion.(...) In order to implement this strategy, the existing processes will be improved by introducing a new open method of coordination at all levels, coupled with a stronger guiding and coordinating role for the European Council to ensure more coherent strategic direction and effective monitoring of progress.* In paragraph 37: *Implementing a new open method of coordination*

1. *Implementation of the strategic goal will be facilitated by applying a new open method of coordination as the means of spreading best practices and achieving greater convergence towards the main EU goals. This method, which is designed to help Member States to progressively developing their own policies, involves:*
    - *fixing guidelines for the Union combined with specific timetables for achieving the goals which they set in the short, medium and long terms;*
    - *establishing, where appropriate, quantitative and qualitative indicators and benchmarks against the best in the world and tailored to the needs of different Member States and sectors as a means of comparing best practices;*
    - *translating these European guidelines into national and regional policies by setting specific targets and adopting measures, taking into account national and regional differences;*
    - *periodic monitoring, evaluation and peer review organised as mutual learning processes.*
2. *A fully decentralised approach will be applied in line with the principle of subsidiarity in which the Union, the Member States, the regional and local levels, as well as the social partners and civil society, will be actively involved, using varied forms of partnership. A method of benchmarking best practices on managing change will be devised by the European Commission networking with different providers and users, namely the social partners, companies and NGOs.*

Interesting historical views and comparisons on the OMC's in different contexts can be found in Pochet (2005).

1.2.2 OMC on Social Protection and Social Inclusion

The implementation of the Open Method of Coordination (OMC) in different areas has been supported in a Presidency Note *Follow-up of the Lisbon European Council - the*

ongoing experience of the open method of coordination, Presidency Note, June 2000.
In the context of social inclusion, we find:

- *Policies for combating social exclusion should be based on an open method of coordination combining national action plans and a Commission initiative for co-operation in this field to be presented by June 2000, (in Lisbon Summit Conclusions, paragraph 32);*
- *In social inclusion, priorities, best practices and indicators are being identified in order to prepare national plans;*
- *In Report of the High Level Group on Social Protection relating to the effort of cooperation for the modernisation and the enhancement of social protection, High Level Group on Social Protection, 18.05.2000;*
- *In Proposal for a European Parliament and Council Decision establishing a Community Action Plan to encourage co-operation between the Member States to fight social exclusion, European Commission, 08.05.2000.*

Community Action Programme to combat poverty and social exclusion was set up. In the social inclusion OMC, some funds were made available for NGOs (Non-Governmental Organization) and consequently its *inclusive* approach to civil society has been favourably commented upon. The Commission (COMMISSION OF THE EUROPEAN COMMUNITIES, 2008) issued on 2th July a Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions proposing to reinforce the OMC for social protection and social inclusion and to call it *Social OMC*. Four objectives are highlighted: *(1) increasing political commitment and the visibility of the process; (2) strengthening the positive interaction with other EU policies; (3) reinforcing the analytical tools underpinning the process, with a view to moving towards the definition of quantified targets and enhancing evidence-based policymaking; (4) increasing ownership in Member States, by boosting implementation and enhancing mutual learning.*

1.2.3 Social Protection Committee

The Social Protection Committee (SPC) was made official in the Nice Treaty (26 February 2001). The Indicators' Sub-Group of the Social Protection Committee was created in February 2001 to support the Committee in its work. In particular, the Indicators' Sub-Group (ISG) is charged of the formulation and definition of indicators. In December 2001 the Laeken European Council took place. European Union (EU) Heads of State and Government endorsed a first set of 18 indicators of social exclusion and poverty. Later on they have been refined by the SPC (SOCIAL PROTECTION COMMITTEE, 2001). Each member state was asked to benchmark its situation by producing a two year national action plan. Community Action Programme to combat poverty and social exclusion was set up.

Since then the ISG continued refining and consolidating the Laeken list. In 2002, the ISG started working on pensions - to support the Open Method of Coordination on adequate and sustainable pensions. From January to June 2003, the ISG have had a meeting on pension adequacy indicators, on social inclusion indicators and most of their meetings have been devoted to redefining and widening the Laeken Indicators of social inclusion. Social Protection Committee (2003) contains a summary of the work of the Group during the first six months of 2003. Work on health and long term care indicators started in 2005. Social Protection Committee (2005) focuses on common indicators relating to national strategies for adequate and sustainable pensions.

The ISG proposal (European Commission, 2006) was the basis for the adoption in June 2006 by the SPC of a set of common indicators for the social protection and social inclusion process. It consists of a portfolio of 14 overarching indicators (+11 context indicators) meant to reflect the newly adopted overarching objectives (a) *social cohesion* and (b) *interaction with the Lisbon strategy growth and jobs objectives*; and of three strand portfolios for social inclusion, pensions, and health and long-term care. The list has been revised since then. The updated list and other information on the history of the development of the portfolio can be found in Eurostat (2009). These indicators are an essential element in the Open Method of Coordination to monitor the progress of Member States in the fight against poverty and social exclusion.

1.2.4 Social cohesion and social cohesion indicators

Social inclusion/exclusion are the two facets of social cohesion. Of course numerous approaches to the concept of social cohesion can be elaborated, like definitions based on community bonds, on shared values and a sense of belonging or definitions based on the ability to work together. In order to fit within the context of the OMC, the definition must lead to measurable dimensions.

Berger-Schmitt (2000) analyses the different approaches and draws the conclusion that *social cohesion incorporates mainly two societal goals which can be analytically distinguished: the first dimension concerns the reduction of disparities, inequalities, and social exclusion; the second dimension concerns the strengthening of social relations, interactions and ties. This dimension embraces all aspects which are generally also considered as the social capital of a society.*

In 2005, the Council of Europe issued a methodological guide *Concerted Development of Social Cohesion Indicators* (Council of Europe Publishing, 2005), in which the following definition is proposed:

Definition of Social Cohesion of a modern society *Social cohesion is the society's ability to secure the long term well-being of all its members, including equitable access to available resources, respect for human dignity with due regard for diversity, personal and collective autonomy and responsible participation.*
This definition is based on the four constituent dimensions of human well-being that are essential for the functioning of societies that recognise human rights and democracy as underpinning the way they are organised:

- fair and equal access (to fundamental rights, new information technology, to culture...),

- individual (and collective) dignity,
- the autonomy of the individual,
- the participation in community life.

These principles determine the *quality* of the bonds between individuals and between them and the community to which they belong.

Several chapters of the methodological guide are devoted to the assessment of social cohesion. The first level of assessment (trends in social cohesion) being given by 20 key indicators describing each component of social cohesion.

ATKINSON et al. (2002) assess the strengths and weaknesses of different indicators relevant to social inclusion in Europe, and their usefulness in promoting good practice by member state governments and allowing comparable assessment of social outcomes. The scientific and political basis on which the indicators were selected, and the implications for the future development of policy-making in Europe are reviewed in ATKINSON et al. (2004). The key features of the indicators are described as well as some of the ways in which they can be developed.

1.3 Income

One key element of social cohesion is the income concept. First methodological problems linked to the choice of income components and their estimation are presented, then the estimation of the empirical income distribution is addressed.

1.3.1 Income Concept and Measurement

In 1996 the initiative to organize an International Expert Group on Household Income Statistics was taken by the Australian Bureau of Statistics in order to work on the development of statistics on household economic well-being and particularly on household income. This International Expert Group met for the first time in December 1996 in Canberra, Australia under the auspices of the United Nations Statistical Commission and, taking its name from the venue of the First Meeting, is known as the *Canberra Group*.

The primary objective of the Canberra Group was to enhance national household income statistics by developing standards on conceptual and practical issues related to the production of income distribution statistics and in support of a revision of international guidelines on income distribution statistics. The Group issued a report and recommendations (THE CANBERRA GROUP, 2001), that have been generally followed for EU-SILC.

LAAN (2007) discusses the proposed income concept in EU-SILC in view of the recommendations of the Canberra Group and its appropriateness for policy-making and policy analysis. The priorities concerning the measurement of various income components in EU-SILC are addressed, considering the exhaustiveness as suggested by the Canberra Group, as well as feasibility, current practice and current policy assessments and the need for different income concepts. The paper ends by making recommendations on improving EU-SILC as a relevant source of income data in the EU.

One of the basic requirements for EU-SILC, and also for a large number of micro-simulation models, is to have detailed information on gross income components at household and personal level. However, the information about the incomes of the EU-SILC are sometimes more conveniently observed as net.

This final report of Eurostat (2004) treats about the income in EU-SILC and more particulary the problem of the conversion. It consists of the following five chapters: (1) describes the structure of the model for net-gross conversion; (2) presents a summary of main results from the micro-simulation system; (3)-(5) document the fiscal systems for France, Italy and Spain as implemented in the model.

The objective of Betti and Verma (2006) is to *develop, test and recommend procedures on how this problem may be overcome. Both in theory and in practice, this is a complex task. The modeling procedure for net-gross transformation can range from the very complex to the very simple.* Rodrigues (2007) discusses the methodology required to model net-gross conversion of Portuguese incomes presented in EU-SILC.

The next three documents are related to different income components in EU-SILC. Törmälehto (2007) explores some conceptual and practical issues related to the measurement of income derived from ownership of financial and non-financial assets. The concept of property income in EU-SILC is reviewed and contrasted to National Accounts' concepts and international recommendations.

Frick et al. (2007) focus on two specific types of non-cash income as these are or will be included in EU-SILC, namely *imputed rent* and *non-cash employee income*. After describing principles and actual implementation of both non-cash components in EU-SILC 2004 and SOEP 2002 (the German Socio-Economic Panel Study), *the more substantive focus of this paper is to analyze incidence and relevance as well as the impact of both non-cash components on the overall income distribution. Section 2 is on imputed rent (IR) and section 3 on non-cash employee income. Section 4 gives preliminary conclusions on future harmonisation of such measures in light of the need for improved cross-national comparability.*

Marco (2007) proposes first a definition of international comparability of income data and emphasises that welfare analyses require comparability at the micro level, i.e. of statistical units within and across countries. Then the paper illustrates the methods adopted for the Italian EU SILC in order to minimize the underestimation of self-employment incomes.

1.3.2 Equivalence scale

Total household income is transformed into an equivalised income (i.e. an income at the individual level) by division by the household size. The commonest household size is the number of household members, but it is not considered as correct scale when thinking of the personal disposable income. It is thus modified into an equivalence scale. The equivalence scale for use in EU-SILC is based on an equivalent household size, defined as $1 + 0.5(n_A - 1) + 0.3n_C$, where n_A is the number of persons in the household aged 14 and more, and n_C is the number of children (less than 13). It is of interest to analyse the sensitivity of various income inequality and poverty measures to the choice of equivalence scale. BUHMANN et al. (1988) review the available equivalence scales on the basis of the Luxembourg Income Study database (LIS, see Section 1.4.3), and provides sensitivity tests.

COULTER et al. (1992) builds on the work by BUHMANN et al. (1988) comparing inequality and poverty across ten countries using the Luxembourg Income Study database. They investigate the sensitivity of their results to the choice of equivalence scale, using a wide range of scales and measures of inequality and poverty for ten countries. Most of the inequality and poverty measures in common use are considered: the inequality measures are the Lorenz curve, the Generalised Entropy family of-indices, the Gini coefficient and the variance of the logs; the poverty measures are those comprising the FOSTER et al. (1984) class of indices.

1.3.3 Income distribution

Income distribution issues in ATKINSON and BOURGUIGNON (2000) are considered *under their theoretical or their empirical side, under a normative or a positive angle, in connection with redistribution policy, in a micro or macro-economic context, in different institutional settings, at various point of space, in a historical or contemporaneous perspective.* As the Editors state it, *there is nowadays no unified theory of income distribution. (...) Rather than a unified theory, the literature offers a series of building blocks with which distribution issues are to be studied.* Table of contents includes: Income distribution and economics; Social justice and distribution in income; Measurement of inequality; Three centuries of inequality in Britain and American; Historical perspectives on income distribution: the case of Europe; Empirical evidence on income inequality in industrial countries; Income poverty in advanced countries; Theories of the distribution of earnings; Theories of persistent inequality and intergenerational mobility; Macroeconomics of distribution and growth; Wealth inequality, wealth constraints and economic performance; The distribution of wealth. Redistribution, Income distribution and development; Income distribution, economic systems and transition.

KERM (2003) offers an exploratory analysis of household income volatility in the 1990s in fourteen EU countries and two future member states, namely Hungary and Poland, using simple summary statistics for average income changes as advocated in Fields and Ok (JET 1996, Economica 1999). The evidence is derived from the data issued by the Consortium of Household Panels for European Socio-Economic Research (CHER, see Section 1.4.2) that contain harmonised data from the ECHP and from a series of independent panel surveys.

The purpose of AABERGE et al. (2007) is to provide a discussion of two basic comparability problems related to the standard definition and measurement of income, with application to distributional assessments and within the context of EU-SILC. First, they focus attention on how to deal with problems of measuring dividends and capital gains that arise when the income reporting behavior is affected by tax changes. The second issue discussed is concerned with comparability of incomes when there is significant non-income heterogeneity between regions within a country.

In Jenkins and Van Kerm (2003) the focus is on relating the inequality trends to intra-distributional movements, i.e. the change from one time period to another of individual income positions within the distribution. The main contribution of this paper is the presentation of a new analytical framework to address jointly the measurement of distributional trends and mobility.

The paper by Gil-Izquierdo (2007) addresses the measurement of low income. Its aim is to provide new evidence related to low-income households and their members in Spain on analyzing the determinants of poverty and social exclusion. For measuring the factors related to situations of low income, the Living Conditions Survey (2004) database (Spain) has been used.

Van Kerm (2007a) addresses the problem of estimation of the two tails of the income distribution. *Micro-data estimates of welfare indices are known to be sensitive to observations from the tails of the income distribution. It is therefore customary to make adjustments to extreme data before estimating inequality and poverty statistics. This paper systematically evaluates the impact of such adjustments on indicators estimated from the EU-SILC.*

A general problem for the assessment of inequality is the top-coding of the very large incomes. Burkhauser et al. (2007) study the measure of US inequality trends with Current Population Survey (CPS) data which are top-coded, and internal non censored Census Bureau data. It addresses the problem of top-coding and its impact on the Gini coefficient and on inequality trends. *Because we have access to the internal CPS data, we have been able to create consistent cell mean values for all top-coded values in all years of internal data made available to us (1975 - 2004) that offer a plausible correction for time inconsistency problems in the public use CPS data when integrated with them.* It is shown that calculating the decile income ratio P90/P10 with public use Current Population Survey data *does not completely obviate the problem of time inconsistency, especially for those interested in trends in the inequality of individuals' size-adjusted household income.*

In Duclos and Araar (2005) the properties of a family of social evaluation functions and inequality indices which merge the features of the family of Atkinson and S-Gini indices are studied. Income inequality aversion is captured by decreasing marginal utilities, and aversion to rank inequality is captured by rank-dependent ethical weights, thus providing an ethically-flexible dual basis for the assessment of inequality and equity. A few of their interesting properties are demonstrated and their application using data from the Luxembourg Income Study is illustrated.

Then in Duclos and Davidson (2006) asymptotic and bootstrap tests are studied for testing whether there is a relation of stochastic dominance between two distributions. These tests have a null hypothesis of non-dominance, with the advantage that, if this null hypothesis is rejected, then all that is left is dominance. This also leads us to define and focus on restricted stochastic dominance, the only empirically useful form of dominance relation that we can seek to infer in many settings. One testing procedure that we consider is based on an empirical likelihood ratio. The computations necessary for obtaining a test statistic also provide estimates of the distributions under study that satisfy the null hypothesis, on the frontier between dominance and non-dominance. These estimates can be used to perform bootstrap tests that can turn out to provide much improved reliability of inference compared with the asymptotic tests so far proposed in the literature. The proposed methodology is illustrated using the LIS data sets5 consisting of the USA (2000), the Netherlands (1999), the UK (1999), Germany (2000) and Ireland (2000).

1.4 International Databases in relation to EU-SILC

1.4.1 ECHP

The ECHP (European Community Household Panel) runs from 1994 to 2001 (ECHP Website). It is a harmonised cross-national longitudinal survey focusing on household income and living conditions. It also includes items on health, education, housing, migration, demographics and employment characteristics. Three central features of the ECHP are:

- The multidimensional character of the topics covered;
- The cross-national comparability of the data;
- The longitudinal dimension.

The EU-SILC was launched in 2004 as a replacement for the ECHP. The motivation was the higher priority that the EU wanted to give to fighting poverty and social exclusion, and to monitor the targets set to do both.

Eurostat (2003) is the ECHP-UDB (User Database) manual and gives the essential information about the construction of the UDB.

Eurostat (2002) provides information about income poverty and social exclusion in each member state and at an aggregate level on the basis of the ECHP. The report presents cross-sectional information at annual intervals during the first waves of this pioneering social survey, exploring the association between monetary and non-monetary aspects of poverty and social exclusion, and develops longitudinal and dynamic analyses.

Eurostat (2005b) analyses the transition between ECHP and EU-SILC. During the transition period until launch of EU-SILC, indicators are being compiled by Eurostat from the best available national sources, harmonised as closely as possible with EU-SILC definitions. There is an unavoidable disruption in the time series of indicators produced. Data collection under the EU-SILC regulations displays some important differences from its predecessor, the ECHP. Similarly, there are important differences between EUSILC and transitional national data sources. The impact of these various differences can be significant, depending on the country and the indicators concerned. Although this paper presents various checks and comparisons which have been made - and further information may become available with the receipt of quality reports - it is impossible to isolate individual causes for all such differences and quantify their impact.

1.4.2 CHER

The Consortium of Household Panels for European Socio-economic Research (CHER) was established in 2000 and terminated in 2003. The CHER database has been set up to carry out analysis of the dynamics of socioeconomic changes in Europe. The project's primary objective is to develop a comparative micro database for longitudinal household studies, by integrating micro data sets from a large variety of independent national panels and from the ECHP. The comparative database contains harmonised and consistent variables and identical data structures for each country included: 14 EU countries, Poland, Hungary, Canada and USA.

Moreover this project can supply objective as well some subjective information on the process of change in various areas of life. A complementary database containing key information about macro data, social security and employment policies improves analysis of social policies. The potential of the CHER database for cross-national research is considerably greater than that of the ECHP. CHER also makes East-West comparisons possible.

The final report is available at CEPS/INSTEAD (2003) (CEPS/INSTEAD: *Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques / International Network for Studies in Technology, Environment, Alternatives, Development*). More information and a publication list on the following Webpage: European Commission.

1.4.3 LIS

The Luxembourg Income Study (LIS) is a non-profit project which produces a cross-national database of micro-economic income data for social science research. The project started in 1983 and is headquartered in Luxembourg. In 2006 the database included data from 30 countries on four continents, with some countries represented for over 30 years.

The LIS database contains anonymised demographic, income and expenditure information at three different levels of analysis (household, person and child). The data has, as far as is practical, been transformed to a structure which make different national data equivalent. For more information, see the LIS Webpage.

1.5 Current results related to Laeken Indicators

This section introduces the Laeken Indicators from the concept of income. Methodologies and strategies for the development of indicators are proposed. After obtaining the list of the overarching indicators, two types of indicators are discerned: monetary and non-monetary indicators. Through these indicators, topics as poverty and social exclusion can be studied and treated.

1.5.1 General analytical papers and results

EUROSTAT (2002) provides a complete overview of the period 1994-1997 for EU15 based on the ECHP. Table of contents includes: conceptual and methodological framework; income distribution and risk of income poverty; the dynamics of income poverty risk; non-monetary or lifestyle deprivation; income poverty risk and lifestyle deprivation; the role of social transfers; country profiles.

GUIO (2005a) gives an introduction to the Laeken indicators and an overview of their development from 1994 to 2003. During the reference period 1994-2001 the ECHP has traditionally been the primary source of data used for the calculation of these indicators in the field of Income, Poverty and Social Exclusion. In 2003, it was decided to replace the ECHP by EU-SILC in order to satisfy new political demands. The first set of micro data and cross-sectional indicators from EU-SILC which covers all the EU25 Member States has been made available in December 2006. During the transition Eurostat launched a collection of indicators derived from national sources. Thus, the indicators cannot be considered to be fully comparable with EU-SILC or between countries during the transition period. However, in spite of this difference of data sources, every harmonisation effort has been made to insure the maximum comparability between definitions and concepts used in the different countries and at the EU level. For the first time in 2003, the precision of the indicators has been computed and a table is provided in the paper.

The aim of the project of LEMMI and VERMA (2003) is to identify the appropriate methodologies and strategies for the development of indicators of poverty and social exclusion at regional level. Small area estimation of both income poverty and deprivation at the NUTS3 level and beyond are tackled.

First results on EU-SILC are given in the 2007 Joint Report on Social Protection and Social Inclusion (COMMISSION OF THE EUROPEAN COMMUNITIES, 2007a), which is composed of the following two documents:

COMMISSION OF THE EUROPEAN COMMUNITIES (2007b) contains 27 countries profiles, that are good examples of the OMC. They aim at providing a synoptic view of key trends, major efforts and challenges ahead in each of the Member States with respect to their policies in the fields of social inclusion, pensions and health and long-term care. They are based on the integrated National Strategies for social protection and social inclusion that Member States have presented in 2006 for the first time.

COMMISSION OF THE EUROPEAN COMMUNITIES (2007c) draws on the material provided by the Member States in their National Reports on Social Protection and Social. The document is divided into two parts. The first part provides an analysis of the social situation across the fields of social inclusion, pensions and health and long-term care. The second part examines the policy strategies presented by the Member States and looks in turn at social inclusion, health and long-term care, and pensions.

The updated list of overarching indicators as well as the algorithmic implementation within EU-SILC under the Open Method of Coordination can be found in EUROSTAT (2009).

The PhD thesis by GOEBEL (2007) addresses different methodological issues in the measurement of income and poverty: (1) aggregation problem; (2) measurement of poverty over time, considering possible income mobility; (3) comparisons between imputation methods; (4) measuring high incomes; (5) regional income stratification.

1.5.2 Monetary and Non-Monetary Indicators

Let us recall that the monetary indicators are relative i.e. based on a threshold defined in relation to the distribution of income within each country (60% of the median income). Four papers present results on monetary Laeken indicators based on ECHP.

DENNIS and GUIO (2003a) and GUIO and DENNIS (2004b) provide an overview of the monetary indicators adopted in Laeken, which have all been calculated on the basis of the ECHP. These two publications focus on nine income indicators in the Laeken list. Indicators are only provided at the level of the total population and for the ECHP UDB 1999, respectively ECHP UDB 2003, wave 8: 1994-2001. In GUIO and DENNIS (2003) the monetary Laeken indicators are calculated for Acceding and Candidate Countries on the basis of national statistical sources and in GUIO and DENNIS (2004a) for new Member States and Candidate Countries on the basis of national statistical sources.

JÄNTTI (2007) considers the EU-SILC in comparative income distribution research. Comparisons are made with the LIS and the ECHP. Several parameters are computed (Gini, percentiles, percentiles ratio and interquartile range standardised by the median), as well as Lorenz curves for different income concepts.

BRANDOLINI (2006) investigates the measurement of income distribution in supranational entities: *Greater social cohesion is an explicit goal of the European Union. Progress is monitored considering the performance in each member country on the basis of national indicators; EU-wide estimates of inequality and poverty play no role. Yet this is a basic information to evaluate the progress of the Union toward grater social cohesion. This paper examines the methodological requirements of this evaluative exercise, and provides the first estimates of inequality and poverty in the enlarged European Union as if it was a single country.*

The non-monetary indicators are more *absolute*, like e.g. *economic strain*, enforced lack of durable goods and problems with housing, the unemployment.

DENNIS and GUIO (2003b) provides an overview of the non-monetary indicators adopted at Laeken, for each Member State and the EU as a whole. First, the relationship between employment and social inclusion is addressed, then the indicators (Persons living in Jobless Households, Long-term unemployment rate, Long-term unemployment share, Very long-term unemployment rate, Regional cohesion, Early school leavers not in education or training, Persons with low educational attainment, Life expectancy at birth and Self-defined health status by income level) are presented for a single year, either at the level of the total population or with a breakdown by gender or age.

1.5.3 Poverty and Social Exclusion

GUIO (2005b) compares material deprivation (defined as *the enforced lack of a combination of items depicting material living conditions, such as housing conditions, possession of* *durables, and capacity to afford basic requirements.*) with the monetary indicators that are relative. The results of the analysis is that deprivation and income poverty have different dimensions. *It seems therefore preferable, at this stage, to present the monetary and non-monetary measures separately, with each containing information crucial to enhance our understanding of poverty and social exclusion.*

The paper by GUIO and BARDONE (2005) discusses *the possible approaches to measure the overlap between work and poverty and describes the methodological choices that have been retained at the level of the EU to define common indicators and variables to measure this overlap. It reviews the available evidence, mainly relating to the 15 old EU Member States (on the basis of the ECHP) but also for the new Member States, for which data are available on a non-strictly comparable basis.* It advocates a complementary household approach to in-work poverty: *(...) since poverty incidences are strongly influenced by household structures and household employment patterns, working poverty must be analysed not only through personal and occupational characteristics but also through household characteristics. (...) By adding a household approach, we look at the economic well-being of all the people, including children and other dependants, living in households where there is some work.*

The aim of WARD (2007) is to examine the information provided in the EU-SILC on differences in the access of women and men to employment, and accordingly to income from employment, and the effect which children have in this regard. The additional concern is to consider indicators which can potentially be derived from the data collected by the survey to throw light on the participation of women in the labour market relative to men and on the ease or difficulty which they have in reconciling childcare responsibilities with the pursuit of a working career.

The problem of child poverty assessment has attracted a lot of interest. The paper of AASSVE et al. (2006) investigates how the extent of childbearing among couples in Europe affects their level of economic well being. By implementing, they do a propensity score matching procedure in combination with a difference-in-difference estimator. Using data from ECHP, they compare how the impact of childbearing on wellbeing varies among countries. Several measures for wellbeing are used, including poverty status and various deprivation indices that take into account the multi-dimensionality of individuals' assessment of wellbeing.

The paper by EUROPEAN COMMISSION (2008a) is divided into three parts: (1) An evaluative review of child poverty and social exclusion in the EU. (2) A description of the most important data sources used by countries for monitoring their policies in the field of child poverty and child well-being; a description of the main types of indicators used in the policy monitoring systems; an in-depth review of the main governance and child well-being monitoring arrangements in 8 selected Member States. (3) Conclusions and proposal for a set of Social Protection Committee recommendations. They aimed at better monitoring and assessing child poverty and well-being at EU and country levels.

In WHELAN et al. (2007), a measure of social class based on the new European Socio-economic Classification (ESeC)[1] for the first wave of EU-SILC data and look at the relationship between social class, income poverty and deprivation is developed.

[1]The ESeC schema distinguishes a relatively small set of classes that are distinctive in terms of their employment relations.

1.5.4 Social Protection

Following the adoption in 2006 by the Council of the Commission's Communication on the streamlining of the OMC on Social Protection and Social Inclusion, it seems of importance to mention studies within the framework of social protection. All of the following documents are taken from Eurostat Webpage on Social Protection, Expenditure and Receipts.

This first Eurostat publication (Eurostat, 2008) relates to social protection data in Europe. It includes two parts: 1) Expenditure and receipts 1997-2005; 2) Detailed tables for each country: *detailed breakdown of social benefits by function, 1996-2005.* All tables contained are compiled from data supplied by the national statistical offices or the ministries in charge of social protection statistics. The 2005 data available are shown in this publication.

Kubitza (2004) examines the pensions expenditure and beneficiaries through different items like the stabilisation of the expenditure on pensions, the domination of the old-age pensions compared to the pension expenditure, or the fact that the economic development, age structure of population and pension reforms determine the trend of pension expenditure.

Abramovici (2003) compares the cash family benefits in Europe. It is observed that the cash benefits vary widely between the countries which explains why the national legislations on cash family benefits are not uniform. An increase of the family benefits in cash is observed despite the decline in the birth rate.

Abramovici (2002)'s focus is on expenditure on cash benefits and on benefits in kind. The relative shares of cash benefits and benefits in kind depend on the specific characteristics of the social protection functions themselves, but also reflect differences in how social protection is organised and the influence of structural factors.

Marlier and Cohen-Solal (2000) analyses social benefits in the European Union on the basis of the 1996 wave of the ECHP. The results cover the 13 countries for which the required data are now available: all Member States of the EU except Finland and Sweden.

1.5.5 International Comparisons on Income and Inequality

In this section, papers on international comparisons involving countries outside Europe are collected. A series of studies have been issued by the World Bank on the income comparisons worldwide. Different definitions of individual income can be used. Inequality can also be measured via the expenditures. Milanovic and Yitzhaki (2002) gives a decomposition of the total inequality between the individuals in the world, by continent and by regions. Here, the Yitzhaki's Gini decomposition is used. It allows for an exact breakdown of the Gini.

Milanovic (2002a) derives world income or expenditure distribution of individuals for 1988 and 1993. It is the first paper to calculate world distribution for individuals based entirely on household surveys from 91 countries, and adjusted for differences in purchasing power parity between countries. The Gini index is used to calculate the inequality (see Milanovic and Yitzhaki, 2002).

MILANOVIC (2002b) introduces three different concepts of world or inter-national inequality. The first uses unweighted countries' GDPs per capita, the second, population weighted GDPs per capita, the third, combines inter-national and internal income distribution to derive *true* world income distribution. The third concept of world inequality is based on incomes or expenditures calculated from household surveys and is consequently available for a much shorter time period.

Later, MILANOVIC (2005) addresses the problem of how to measure global inequality among individuals, and shows that inequality is shaped by complex forces often working in different directions. He analyses income distribution worldwide using, for the first time, household survey data from more than 100 countries. The main approaches to the problem are introduced, a more accurate way of measuring inequality among individuals is proposed, and the relevant policies of first-world countries and nongovernmental organisations are discussed.

The OECD published a study by FÖRSTER and PEARSON (2002) on income distribution and poverty in the OECD area. The authors set up international comparisons of income distribution and poverty for OECD countries and make comparisons in time. They give a broad picture of the trends from the mid-seventies to the mid-nineties.

1.5.6 Composite Indicators for social cohesion and the Laeken Indicators

The OECD Handbook on Composite Indicators (OECD and JRC, 2008) does not propose concrete composite indicators and there is no reference to the Laeken indicators. However, it is clear that the methodology proposed in the handbook could be applied to the Laeken indicators taking as a basis of the indicators the streamlined social inclusion portfolio proposed by the Indicators' Subgroup of the Social Protection Committee (ISG-SPC) (EUROPEAN COMMISSION, 2006, 2008b). The OECD handbook mentions outliers and their impact in several places but does not dedicate a topic to it. Presentation of indicators is discussed in Chapter 1.10. In particular also graphical displays are encouraged and some examples, bar charts and line charts, are presented.

Probably the best known example of composite indicators are the set of indicators published in the Human Development Reports (see, e.g., UNITED NATIONS DEVELOPMENT PROGRAMME, 2007). Technical note 1 (UNITED NATIONS DEVELOPMENT PROGRAMME, 2007, p. 355) explains the calculation of the Human Development-indicators, which are all composite. There has been considerable debate about these indicators because they are composite but also due to shortcomings in the scales, the reporting and in the quality of the underlying data MACREDIE et al. (2001).

The ISG of the SPC does not discuss the use of composite indicators (cf. the ISG website) in their above mentioned papers. There seems to be a concern about the number of indicators and this is probably one of the reasons why the ISG proposes the overarching indicators as core elements of their system. There is clearly a need for multivariate presentation of the set of indicators.

1.6 European Research Projects with Potential Impact on AMELI

5th Framework Programme research projects are collected in the following document Eurostat (2007b) issued by EPROS (European Plan of Research in Official Statistics). This document is a useful source of information. The projects below deal on topics of interest for the AMELI project.

1.6.1 EUREDIT

The main objective of the EUREDIT project (The Development and Evaluation of New Methods for Editing and Imputation) is to investigate and evaluate methods for automatic editing and imputation. Specifically:

- Develop a methodological evaluation framework and develop evaluation criteria for data editing and imputation;
- Produce a standard collection of data sets;
- Establish a baseline by evaluating currently used methods;
- Develop and evaluate a selected range of new techniques;
- Compare and evaluate the different methods and establish best methods for different types of data;
- Disseminate the best methods via a single computer package and publications.

Recommendations have been published. For more information about the EUREDIT project, see EUREDIT Webpage and Eurostat (2007b).

1.6.2 Chintex

Chintex is the acronym of *The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel - Implications on Data Quality*. This European research project, carried out from January 2000 until July 2003, focuses on harmonisation of the data and statistical data quality. The main focus of Chintex is on the analysis of advantages and shortcomings of the different harmonisation strategies. This includes an empirical description and the development and checking of statistical procedures aimed at a methodological improvement of the results of ex-post harmonisation.

Information on Chintex can be found in Eurostat (2007b) and in more details in the final report of the project Ehling et al. (2004) and the website page Chintex The papers in the report are divided into five domains: 1) Conversion as a Tool of Ex-post Harmonisation of Panel Data; 2) Income Measurement by Register and Survey Information; 3) Non-response in European Panel Surveys; 4) Panel Attrition; 5) Imputation and Weighting.

1.6.3 DACSEIS

The main goal of the DACSEIS project (Data Quality in Complex Surveys within the New European Information Society) is to analyse variance estimation and related methods that allow for various sources of errors. The specific objectives are:

- Standardisation and harmonisation of variance estimation methods used to calculate sampling errors. This is achieved with special emphasis on different national surveys within the European statistical system;
- A catalogue of instructions and criteria to be compiled, allowing the user to choose the most effective procedures for variance estimation for complex sample designs;
- Inspection of relevant methods, analyzed and evaluated with respect to their applicability to complex surveys;
- The scrutiny and evaluation of standard software packages for survey sampling, with special emphasis on the implementation of the above mentioned variance estimation methods for complex surveys;
- Dissemination of the results and the estimation methods to NSIs and other users; this includes the delivery of a recommended practice manual as well as the source and pseudo codes of all relevant estimation procedures.

Additional information on the project can be obtained under DACSEIS Webpage, MÜNNICH (2001) and EUROSTAT (2007b).

1.6.4 EURAREA

The goal of the EURAREA project (Enhancing Small Area Estimation Techniques to Meet European Needs) is to investigate methods for Small Area Estimation and their application. It was funded under the Fifth Framework (FP5) Programme of the European Union.

More precisely, the aim is to evaluate the effectiveness of standard estimation techniques for small areas (synthetic estimators, GREGs and composite estimators). The studies carried out before, based on sampling designs with equal selection probabilities only, were extended to the development of the estimation theory to other sampling plans more similar to those applied in official statistics. Finally, all the theory developed has been implemented in a public use SAS IT application. The project focuses the research mainly on four topics:

1. The use of ancillary information from the past;
2. The use of ancillary information from other geographical areas;
3. The adaptation of standard estimators to complex sampling designs, in other words, with the use of unequal probabilities and in particular, with the selection of conglomerates;

4. The obtaining of estimations for cross-classifications.

One can find more information on EURAREA Webpage and in the document Eurostat (2007b).

1.6.5 KEI

The aim of the KEI project (Knowledge Economy Indicators) is to develop and improve indicators for the knowledge economy, including the analysis of aggregation issues and the use of composite indicators. The project covers indicators from 30 European countries (the EU-25 plus Romania, Bulgaria, Iceland, Norway and Switzerland) and six non-European countries (the US, Japan, India, China, Australia and Canada).

The KEI project reviews existing concepts and definitions of the knowledge-based economy and its key components. Main thematic areas in relation to the Lisbon and Barcelona objectives are developed and used for a classification of existing indicators. Moreover data and indicator quality issues are explored.

Composite indicators are analysed in detail using both statistical and participatory approaches, including the use of multi-criteria methods, aggregation and weighting techniques, decomposition methods, and an evaluation of analytical and presentational techniques.

Simulation methods are employed extensively to investigate the robustness of indicators and the conclusions based on them. The study evaluates the quality and accuracy of indicators and the underlying data and assesses the innovative use of additional information to improve indicator quality. For more information see KEI Website.

1.6.6 OPUS

The key goal of the OPUS (Optimising the Use of Partial Information in Urban and Regions Systems) project is to develop and demonstrate new methods for the coherent combination and use of data from disparate, cross-sectoral sources, and so to contribute to improved decision- making in the public and private sectors within Europe. The research focuses on developing an innovative methodology, incorporating statistical and database systems, see Eurostat (2007c).

In order to achieve this overall aim, the key scientific objective of the project is to develop a generic statistical framework for the optimal combination of complex spatial and temporal data from survey and non-survey sources. The framework should be sufficiently abstract to be applicable in a wide range of potential socio-economic domains. It is demonstrated in transport pilot applications in London and Zurich, in smaller-scale feasibility studies in a number of other cities and regions and in a feasibility study in the health sector. For more information, see the OPUS Webpage.

Chapter 2

Summary of the State-of-the-Art in SAE for Laeken indicators

2.1 Introduction

This is a short review of recent studies on small area estimation methods in a framework of estimating the poverty rate. This review is based on papers of Verma et al. (2005), Fabrizi et al. (2005), Fabrizi et al. (2007) and D'Alò et al. (2006).

2.2 Data and Regions

In three of the reviewed papers (Verma et al., 2005; Fabrizi et al., 2005, 2007) the estimation of the poverty rate was based on the European Community Household Panel (ECHP) survey data. The ECHP survey was designed to provide reliable estimates for large regions in the countries (NUTS1). D'Alò et al. (2006) used data collected with the Consumer Expenditure Survey (CES) for estimating the Italian poverty rate. The CES data was available at NUTS2 level.

For the purpose of constructing regional indicators auxiliary information was used. New-Cronos ('Eurostat Free Dissemination Database") provided a valuable data source for Verma et al. (2005). Fabrizi et al. (2005) used as covariate the regional unemployment rate estimates produced by the Italian Statistical Institute. Also D'Alò et al. (2006) used auxiliary information provided by ISTAT. Fabrizi et al. (2007) used as covariates from ECHP data for which means were available from the 2001 Italian Census results.

2.3 Construction of indicators

Fabrizi et al. (2005) focused on the most popular financial poverty indicators endorsed by Laeken European Council: the Per-Capita Income, the Poverty Threshold, the At-risk-of-poverty rate based on a national Poverty Threshold, the Gini coefficient.

D'ALÒ et al. (2006) aim was to estimate the relative poverty rate which was defined as the relative number of households whose equalised household consumer expenditure is below the level of the poverty line.

The target variable of FABRIZI et al. (2007) was disposable, post-tax household income. They applied the widely-used modified OECD scale to equivalize income.

2.4 Small Area Estimation approach

2.4.1 Verma et al. (2005)

VERMA et al. (2005) applied two types of estimators depending on available data. If access to area-coded survey data was available, area level EBLUB *composite estimators* was used. In the absence of area coded survey data, the *synthetic estimator* was applied.

Regional estimates may be required for two distinct purposes: (1) providing the best possible indicators for each individual area and (2) providing the best possible indicators capturing the variability between areas. For this two objectives the optimal SAE procedure may not be the same. Because of the effect of sampling error, which increases with decreasing size of samples in the areas, the **direct** (survey based) **estimators** tend to in generally provide overestimates of the variability among small areas. Also for the same reason produce distribution tends to be more extreme than the actual distribution. Conversely, **synthetic estimators** tend to underestimate the true variability. Solution is using composite estimators. These are compromise between (being optimally weighted combinations of) direct and synthetic estimators and they are more likely to reflect the true variability than either of the two. (VERMA et al., 2005)

VERMA et al. (2005) does not attempt to incorporate temporal or spatial autocorrelations like in a large project such as Eurarea. However, a major positive feature of the approach is that the modelling strategy is designed to be **hierarchical**. At the beginning direct survey estimates without any modelling have been used. In the report is stated: "…the predictive power of the model at the regional level to be substantially improved when the target variables as well as the covariates are expressed in terms of their values at the preceding higher level." This method provides the effect of the difficult-to-qualify institutional and historical factors, common to the country and its regions, to be abstracted.

For NUTS1 region i, all target variables and all covariates in the model are expressed in the form of the ratio $R_i = Y_i/Y_0$, where (Y_i, Y_0) refer to the actual values of the variables, respectively, for NUTS1 i and its country. Similarly, in going from NUTS1 region i to its NUTS2 region j, the model variables are expressed in the form $R_{ij} = Y_{ij}/Y_i$ and similarly etc. R values can also weighted by using the appropriate population weights for the regions. Same type of idea are extended to the modelling of subpopulation and modeling separately for different parts of large country.

According to the availability of data for the target variables and the access to area-coded survey data for each country, three different types of SAE models have been estimated:

- SAE Model 1: estimated on the ratio NUTS1/Country;

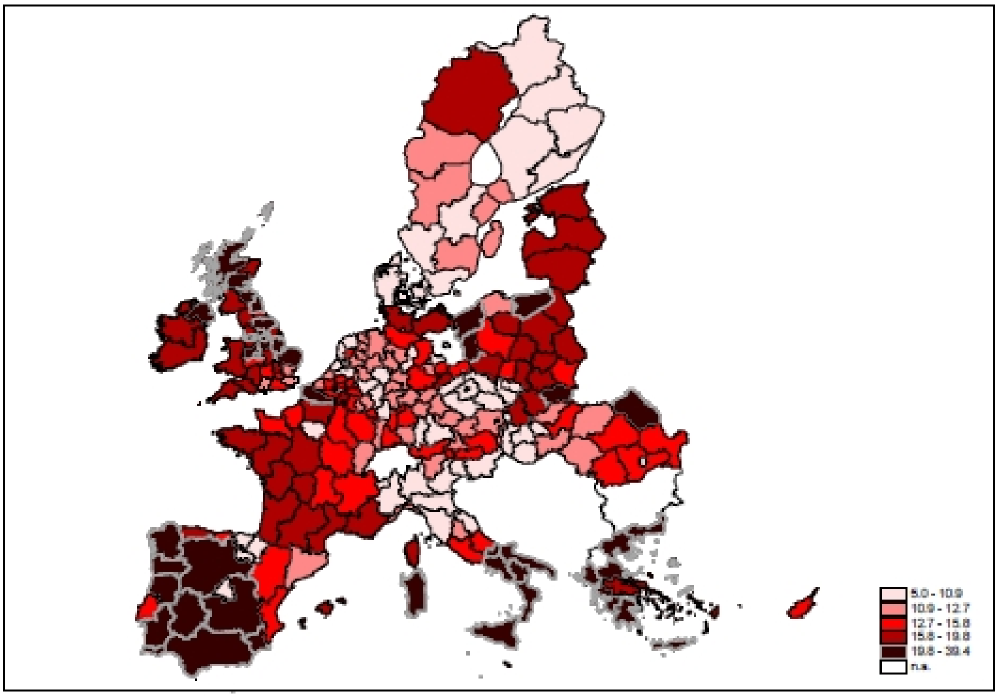


Figure 2.1: Country poverty lines for NUTS2 regions. VERMA et al. (2005, p. 201).

- SAE Model 2: estimated on the ratio NUTS2/ NUTS1;
- SAE Model 3: estimated on the ratio NUTS3/ NUTS2.

Because NUTS3 information was not available in any other survey, model 3 is estimated for Italy only.

Figure 2.1 demonstrates some results. If necessary survey data was available at NUTS2 level, poverty rates was estimated using small area estimation. Otherwise cruder regression-prediction was used.

The approach that VERMA et al. (2005) used for calculating standard errors for the estimated variables at regional level and going beyond NUTS levels, if the necessary information for the purpose was available, had following features:

1. The standard error of any statistics is broken down into a number of factors which represent some aspect(s) of the complexity of the sampling design and the estimation procedure (stratification and clustering, weighting, aggregation over surveys, etc.)
2. The factor effects can be taken as multiplicative because many of these factors act more or less independently of each other
3. To make the result more robust, the each parameter is averaged over dimensions as appropriate.

2.4.2 Fabrizi et al. (2005)

At first Fabrizi et al. (2005) estimated the direct estimates of the indicators. To estimate a multivariate small area model they needed the sampling correlational matrix between the different poverty indicators. To estimate these variances and covariances a bootstrap algorithm was applied. The correlation was large in almost all cases. Large correlation should make multivariate model based estimators more effective in reducing the variability of direct estimators than those based on simpler univariate model (Fabrizi et al., 2005).

Next Fabrizi et al. (2005) applied multivariate hierarchical Bayesian approach to estimate the financial poverty indicators at NUTS2 level. The method preserves the relationship between different indicators and disparities over areas. Because timely covariate information was not available, the regional unemployment rate produced by the Italian Statistical Institute was used as covariate.

$\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{ik}, \ldots, \theta_{iK}$ is the vector of K parameters of interested for the i-th small area $(i = 1, \ldots, m)$ and $\hat{\boldsymbol{\theta}}_i$ the corresponding vector of survey estimates. The model containing area specific auxiliary data is

$$\hat{\boldsymbol{\theta}}_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{v}_i + \boldsymbol{e}_i,$$

where $\boldsymbol{X}_i^T$ the covariates, $\boldsymbol{\beta}$ is the regression coefficient vector, $\boldsymbol{v}_i$ is the area-specific random effects and $\boldsymbol{e}_i$ is the sampling error.

Comparing two models—multivariate models vs. univariate model—multivariate model seemed preferable than the univariate one on the whole. Fabrizi et al. (2005) also noted that the consideration of direct estimates showing a low variability can be very helpful in a multivariate context. In fact these estimates can "give strength" to the estimates less reliable by means of their sampling correlation.

Fabrizi et al. (2005) dealt with the problem of over-shrinkage. Model based estimators such EBLUPs and Bayes estimators tend to "shrink" estimates towards the synthetic component, introducing bias. The solution was to implement a multivariate constraint on the first two moments of the estimates. Fabrizi et al. (2005) extended the univariate methodology of constraining procedure proposed by Ghosh and Maiti (1999) to multivariate case.

2.4.3 D'Aló et al. (2006)

D'Alò et al. (2006) compared the performances of a limited subset of the standard small area estimators (direct, GREG, synthetic and EBLUB estimators) which have been studied in the EURAREA project. Also an EBLUB based on unit level mixed model with spatially correlated area effects have been considered.

A composite estimator which was used also in the EURAREA project was described as

$$\hat{\bar{Y}}_d^{EB\_B} = \hat{\gamma}_d \hat{\bar{Y}}_d + (1 - \hat{\gamma}_d) - \overline{\boldsymbol{X}}_d^T \hat{\boldsymbol{\beta}},$$

where $\hat{\gamma}_d = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_d)$ and $\hat{\sigma}_e^2$ is the pooled estimator of the individual variance.

Model	AARB%	ARMSE%	AEFF$_{Dir}$%	AEFF$_{Sect}$%	RESD%
DIR	0.0	0.787	100.0	-	15.6
COMP	2.7	0.552	70.1	-	-9.8
GREG	0.2	0.543	68.2	-	10.0
SMM	2.3	0.377	47.7	100.0	-8.7
*MM*1	3.1	0.358	45.3	95.0	2.4
*MM*2	2.4	0.427	54.1	113.4	-4.4
*MM*3	2.6	0.380	48.3	101.2	4.7
*MM*4	2.6	0.429	54.2	113.6	-8.0
*MM*5	2.9	0.318	40.4	84.7	-7.7

Figure 2.2: Performance indicators (auxiliary information is available). FABRIZI et al. (2007, p. 194).

D'ALÒ et al. (2006) had two different sets of auxiliary information in their simulation study. First set of auxiliary variables consisted of counts of household members cross-classified by sex and age. The second set consisted of the classification of households by means of some socio-economic variables.

Simulation study results showed that the spatial estimator gives the best results. The spatial estimator seemed to be quite robust to the choice of auxiliary variables.

subsectionFabrizi et al. (2007) On the fourth paper the aim was to compare the performance of different estimators in the controlled environment of a simulation exercise. The ECHP survey data were treated as the pseudo-population.

FABRIZI et al. (2007) chose as covariates from ECHP data only those for which area means were available from the 2001 Italian census results. The chosen covariates are the percentage of adults, the percentage of employed, the percentage of unemployed, the percentage of people with a high/medium/low level of education in the household, household typology (presence of children, presence of aged people, etc.), the number of rooms percapita and the tenure status of the accommodation (rented, owned etc.).

In simulation study FABRIZI et al. (2007) used direct estimator, the design-based estimators (composite and the generalized regression estimator) and the five EBLUB estimators which was derived from a general linear mixed model. The general linear mixed model was described as follows:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}_1 v_1 + \cdots + \boldsymbol{Z}_s v_s + \boldsymbol{e}, \tag{2.1}$$

where $\boldsymbol{y} = \{y_{dti}\}$ is the n-vector of sample observations, β is a $p \times 1$ vector of fixed effects, v_j is a $q_j \times 1$ vector of random effects $(j = 1, \ldots, s)$, $\boldsymbol{e} = \{e_{dti}\}$ is a vector of errors, $\boldsymbol{X}$ is a matrix of covariates, $\boldsymbol{Z} = \{z_{dti}^T\}$ is a $n \times q_j$ matrix of incidence of the j^{th} random effect.

All five models, $MM1$–$MM5$, can be viewed as special cases of the general mixed model (2.1) with different specifications time and area effects. Also a cross-sectional linear mixed model (SMM) was applied. Furthermore random error variance linear models was considered.

In figure 2.2 some of simulation study results are presented. In general simulation study showed, EBLUP estimators derived from unit level linear mixed model specifications that “borrow strength over time”, as well as over areas, provide a significant gain in efficiency compared with both the direct estimator and with other commonly used design based estimators such as the optimal composite estimator and the GREG estimator (Fabrizi et al., 2007).

Fabrizi et al. (2007) compared three different MSE estimators associated with EBLUP estimators. Same measures, AARB and ARMSE, was used in order to compare the performance of the three MSE estimators. They noted that the jackknife estimator provides the best results being correct, on average, over the areas and thus more robust to any departure from the standard assumptions of linear mixed models.

Chapter 3

Summary of the State-of-the-Art in Parametric Income Distributions

3.1 Introduction

Parametric income distributions have been studied for over 100 years, starting with V. Pareto. Needless to say that this overview will not be exhaustive. We sketch the following themes: modelling of income distributions, estimation methods and inequality indices. More information will be found in the work packages *Estimation* and *Robustness*.

3.2 Statistical size distributions

Income distributions are a particular instance of size distributions, e.g. distributions on positive continuous random variables. Negative incomes are either discarded, or modeled separately (the resulting overall income distribution is then a mixture of the positive and negative income distributions), or modeled together with the positive incomes by introducing a shift parameter. When modeling income distributions, two different viewpoints can be taken: 1. find the best fitting distribution to empirical data; 2. find an economic rationale for the particular distribution, e.g. by means of a differential equation. Of course the two viewpoints are not antinomic.

Three books on income distributions and inequality indices are of great value:
Kleiber and Kotz (2003) is a reference book on statistical size distributions. It contains a encyclopedic bibliography on the derivation of the different types of distributions as well as on empirical applications. One huge difficulty that is overcome with the help of Kleiber and Kotz's book is the terminology they have unified and clarified.

In Chotikapanich (2008), seminal and recent papers on income distributions and Lorenz curve are collected to honour the memory of Camilo Dagum. The first chapter is reprinted from Dagum (1977). Dagum postulates that *the final choice a particular model may be governed by its capacity to account fairly well to a set of economic, econometric, stochastic and mathematical properties.* Denoting the income cumulative distribution function by $F(x)$ and defining the income elasticity of the distribution function as

$d \log[F(x) - \alpha]/d \log x$, i.e. as the ratio of the relative change in $(F(x) - \alpha)$ over the relative change in income x, Dagum proposes four assumptions: 1. unimodality of the income distribution; 2. existence of a small percentage of individuals with negative income; 3. positive income range in the interval $[x_0, \infty[$, where $x_0 \geq 0$ and $F(x_0) = \alpha < 1$; 4. the elasticity is a decreasing function of F. Dagum's distribution arises from a particular specification of the elasticity. In the 3rd chapter, reprinted from McDonald (1984), a unified view of many income distributions is provided, among them, one prevailing family of income distributions is the Generalized Beta distribution of the Second Kind (GB2), which encompasses Dagum's as a special case. See Workpackage *Estimation* for more information on the GB2.

In Atkinson and Bourguignon (2000), income distributions are described from a more econometric point of view. The book starts with a review of existing economic theories seeking to explain the distribution of income. Chap.1: relation between the idea of social justice and the analysis of income distribution(A.Sen); Chap.2: basis for comparing different distributions and measuring inequality (F. Cowell); Chap. 3 and 4: historical perspectives; Chap. 5: empirical evidence on income inequality in industrialized countries (P. Gottschalk and T. Smeeding); Chap.6: income poverty in advanced countries, definitions of poverty and equivalence scales (M. Jäntti and S. Danziger); Chap. 7: theories of the distribution of earnings (D. Neal and S. Rosen); Chap. 14: income distribution, economic systems and transition (J. Flemming and J. Mickelwright). The rest of the book is of a less measurement nature (i.e. purely economic).

3.3 Estimation Methods

Parametric income distributions can be used to model the whole income range, or in a semi-parametric way to model the tails of the income distribution.

3.3.1 Parametric Approach

Different fitting methods have been proposed in the literature. The method to choose depends on the type of data available: grouped or micro data. Until recently, it seems that researchers only had access to grouped data. Non-linear least squares are recommended by Dagum (1977) who used grouped data; McDonald (1984) prefers maximum likelihood (ML), which is the prevailing method nowadays, especially with micro data. ML is used in Jenkins (2007) for the GB2 (see also Section 3.4). Victoria-Feser (2000) shows that robust techniques can play a useful role in income distribution analysis and should be used in conjunction with classical estimation methods.

McDonald and Butler (1987) apply generalized mixture distributions to unemployment duration. Yu et al. (2004) present wage distributions via bayesian quantile regression. In the context of capital asset pricing model, McDonald (1989) estimates regression coefficient using partially adaptive techniques and a generalized t (GT) distribution for the error term. The idea is put further to any type of regression with positive variables in McDonald and Butler (1990) and Butler et al. (1990).

3.3.2 Semi-Parametric Approach

The data available for estimating welfare indicators are often incomplete: they may be censored or truncated. Furthermore, for robustness reasons, researchers sometimes use trimmed samples. COWELL and VICTORIA-FESER (2003) derive distribution free asymptotic variances for wide classes of welfare indicators not only in the complete data case, but also in the important cases where the data have been trimmed, censored or truncated. In NEOCLEOUS and PORTNOY (2008), the partially linear Censored Regression Quantile (CRQ) model combines semi-parametric estimation for censored data with quantile regression techniques, and uses B-splines for the estimation of the non-linear term. An application to administrative unemployment data from the German Socio-Economic Panel Survey is presented.

In BURKHAUSER et al. (2008), trends in US income inequality are analyzed with special emphasis on top income shares. On comparing with estimates from administrative data, they conclude that the trend is linked to the top-coding (for confidentiality reasons) of the CPS data. They show that their CPS estimates of trends in top income shares match the estimates of trends reported on the basis of administrative records, except for within the top 1% of the distribution. Thus, they argue that, if income inequality in the USA has increased substantially since 1993, such increases are confined to this very highest income group.

VAN KERM (2007b) considers extreme incomes and the estimation of poverty and inequality indicators from EU-SILC. Social indicators are known to be sensitive to the presence of extreme incomes at either tail of the income distribution. It is therefore customary to make adjustments to extreme data before estimating such statistics. Thus it is important to evaluate the impact of such adjustments and assess how much resulting cross-country comparisons are affected by alternative adjustments. The paper presents the results of a large scale sensitivity analysis considering both simple, classical adjustments and a more sophisticated approach based on modeling parametrically the tails of the income distribution. A Pareto distribution was used as the parametric tail model. An inverse Pareto distribution was used for the lower tail.

3.4 Inequality Indices

One of the aims of parametric estimation is to obtain simple explicit formulas for inequality measures. This was an important aspect in the earlier literature. With the development of computational tools, closed formulas are not a priority anymore. GRAF (2007) computes the parametric expressions for the monetary Laeken indicators in the case of the lognormal and the Fisk distribution (a special case of the GB2), but there is no systematic account so far on parametric expressions for more general distributions, except for the Gini index. Rather than on Laeken indicators, the emphasis is on inequality measures like the generalized Gini, the Pietra index, the variance of logarithms, the Atkinson measure and the generalized entropy coefficients, see (KLEIBER and KOTZ, 2003, Chap. 2). A recent reference is JENKINS (2007) where a general class of inequality indices is derived for the GB2 income distributions, thereby providing a full range of top-sensitive and bottom-sensitive measures. An examination of British income inequality in 1994/95 and 2004/05 illustrates the analysis.

Chapter 4

Overview of Variance Estimation Methods

4.1 Introduction

The intention of this chapter is to give a short overview of variance estimation methods for survey estimates. The estimators of interest cover, on the one hand, the classical parameters of interest, totals, means, and proportions, and on the other hand special non-linear statistics. In practice, ratio estimators are widely used as non-linear estimators. In the context of poverty and social exclusion, we focus on special statistics such as the At-risk-of Poverty Rate (ARPR), the GINI coefficient (GINI), or the Quintile Share Ratio (QSR).

These two subjects in variance estimation lead to different kinds of rather complex problems which can be addressed by suitable approximations. In the first matter the problem arises from the fact that even for a linear statistic, e.g. the sample mean, exact variance estimation methods can be calculated mainly in some rare simple cases. An overview of sampling designs can be drawn from Lohr (1999) or Wolter (2007). Further, the DACSEIS project gives a wide overview of the methodology (cf. http://www.dacseis.de).

The present overview focuses on the treatment of non-linear statistics. In order to treat non-linear statistics, two methods are applied. The first method is based on linearization techniques such as Taylor-based methods. This can be generally described as that an artificial variable is constructed and the variance of the statistic in question is approximated via the variance of the total of this artificial variable. This concept will be briefly introduced in section (4.2).

An alternative to the above described procedure are resampling methods which can also be used to derive the variance of non-linear statistics. Section (4.3) shall give a general overview of different concepts of resampling methods and their applicability.

A deeper presentation and elaboration of the entire methodology can be drawn from Workpackage 3 of the AMELI project.

4.2 Linearization

When estimating measures of inequality or poverty, one is typically confronted with estimators that are non-linear in the observations. Sampling variances of non-linear estimators are usually not available in a closed form. On way to circumvent this problem is to approximate the estimator by a linear function of the observation. As mentioned in section (4.1) the approximations in place do not allow for variance estimation alone. They only produce a linear form of the statistic of interest θ. Then sampling design based methods for variance estimation, as indicated above, are needed to estimate the variance of the linear approximation (cf. Wolter, 2007, p. 286).

One way of linearization is to employ the principles of Taylor series expansions, with which it is possible to handle all functions that are continuously differentiable up to order two (cf. Andersson and Nordberg, 1994). This includes regular functions of totals like, for instance, ratios, regression and correlation coefficients. However, this method cannot be applied to the linearization of most measures for inequality and poverty. So is the GINI based on rank statistics and the ARPR includes the calculation of a income quantile (cf. Osier, 2009).

A more general framework based on the concept of influence functions has been presented by Deville (1999), which uses methods from robust statistics (Hampel et al., 1986). Deville (1999) employs a general class of non-linear statistics for a finite population $\mathcal{U}$ based on a measurement functional. An alternative to the influence function approach might be the use of estimating equations as proposed by Kovacevic and Binder (1997).

The bottom line of linearization is to construct a linearized variable $z_i \; \forall \; i \in s$, such that the variance of the estimated total of this variable z is an approximation to the variance of the estimator of interest $\hat{\theta}$, (cf. Deville, 1999):

$$\mathrm{V}(\sum_{i \in s} z_i \cdot w_i) \approx \mathrm{V}(\hat{\theta}) \; ,$$

where w_i is some survey weight associated with the ith sampling element, i.e. π_i^{-1}. Thus, as long as it is possible to estimate the variance of a total for a given design, it is also feasible to estimate the variance of a complex non-linear estimator. How to construct the z_i s for different statistics can be drawn from the literature or in the AMELI deliverable D3.2. For the ARPR see, e.g., Deville (1999) or Osier (2009), for the GINI see, e.g, Kovacevic and Binder (1997), and for the QSR, see, e.g. Hulliger and Münnich (2006).

4.3 Resampling Methods

An alternative approach to variance estimation is the application of resampling methods. Their general characteristic is to draw (sub-)samples from a given population or the original sample and to calculate the population parameter of interest from each sample. The variance estimation is then based on the distribution of the several estimates. Resampling methods differ in the way they generate the subsamples and with this in the treatment of complex survey designs (cf. MÜNNICH, 2005, p. 69). Below, balanced repeated replication, the jackknife and the bootstrap are presented in brief.

In the basic model of balanced repeated replication each stratum consists of two elements. In each stratum only one unit is drawn and with this half a sample is taken which leads to 2^H possible replications in H strata. To reduce the complexity, a balanced set of half samples is selected instead of all possible half samples. Such samples can be obtained by using a balance matrix which can be received with a Hadamard matrix (for more details see MÜNNICH, 2008, p. 325 f., SHAO et al., 1998, p. 822). In practice, strata generally contain more than two elements and the basic model has to be extended. One possibility is to divide randomly the PSUs in each stratum into two groups of sizes $n_{l,1} = \lfloor n_l/2 \rfloor$ and $n_{l,2} = n_l - n_{l,1}$ and to take a balanced half sample of these groups (cf. WOLTER, 1985, p. 130 ff., DAVISON and SARDY, 2004, p. 18 f., RAO and SHAO, 1996, p. 343 f., MÜNNICH, 2008, p. 326, MÜNNICH, 2005, p. 77, SHAO et al., 1998, p. 824).

In case of the (delete-1) jackknife n replications are generated by omitting once at the time each (ultimate) sampling element after another. The population parameter of interest is computed on the base of these replications and the variance estimation is done on these several estimates (cf. MÜNNICH, 2008, p. 326, SHAO and TU, 1995, p. 4 ff.). Besides its possible computational effort, which may be large, the delete-1 jackknife could also lead to inconsistent variance estimators when the statistic is non-smooth. For such estimators the delete-d jackknife can be used (cf. SHAO and WU, 1989, p. 1176). This extension of the ordinary jackknife is characterized by omitting d elements simultaneously from the original sample (for more details see SHAO and TU, 1995, Section 2.3 and 5.2). As a possible variant of the delete-d jackknife the delete-a-group Jackknife can be used. In this case the sample is split into disjoint and exhaustive groups/subsets and the replications are created by deleting every group/subset once (cf., e.g. SHAO and TU, 1995, p. 195 f.).

The third resampling method presented is the bootstrap. In practice, especially the Monte Carlo bootstrap is of great importance. Here, subsamples of size n are drawn from the original sample by simple random sampling with replacement. This procedure is repeated B times where B is a constant set by the user. The variance estimates are derived from distribution of the different values of the interesting population parameter which is calculated from the set of subsamples (cf. MÜNNICH, 2008, p. 326, SHAO and TU, 1995, p. 9 ff.). Modified versions of the bootstrap like the without replacement bootstrap (cf. DAVISON and SARDY, 2004, Section 6.2), the with-replacement bootstrap (cf. MCCARTHY and SNOWDEN, 1985, p. 4, DAVISON and SARDY, 2004, Section 6.3), the rescaling bootstrap (cf. RAO and WU, 1988) or the mirror-match bootstrap (cf. SITTER, 1992a) are given in the literature to overcome some problems of the Monte Carlo bootstrap.

Chapter 5

Summary of the State-of-the-Art on Robustness Methods for European Laeken Indicators

5.1 Introduction

Preliminary remark: In this overview we confine ourselves to those Laeken indicators, that are estimated with respect to the European Union Statistics on Income and Living Conditions (EU-SILC) data.

Estimates of welfare indices (i.e. poverty and inequality measures) based on survey data are known to be sensitive toward outlying observations from the tails of the income distribution (Van Kerm, 2007b; Cowell and Victoria-Feser, 2002). The presence of only a few extreme incomes can seriously distort the estimate of a statistic. This is particularly problematic for many indicators of inequality (viz. Gini coefficient, Quintile Share Ratio) which are not robust to the presence of data contamination at one or both ends of the distribution. Cowell and Victoria-Feser (1996b) showed the property of non-robustness – in the sense of an unbounded influence function – for a broad class of inequality measures. This formally means that a single observation, provided it is sufficiently large (or small), can drive the estimated inequality indicator arbitrarily large (or small). Results from empirical sensitivity analysis, provided by Van Kerm (2007b), clearly depict that the inequality measures can be severely distorted.

By contrast, poverty indicators (e.g. At-Risk-of-Poverty-Rate) are considered robust provided the poverty line is exogenously determined (Cowell and Victoria-Feser, 1996a) or is derived from a robustly estimated statistic, as it is the case for all Laeken indicators with regard to poverty measurement. In the case of an endogenously determined poverty line (e.g. ARPR), the influence function of the poverty indicator depends on two things: first of all, the determinants of the influence function of the indicator itself (viz. ARPR) and secondly the sensitivity of the poverty line (ARPT) to contamination. Since both components are bounded, the overall effect of an arbitrarily large observation will have bounded influence, too. Empirical evidence for the behaviour of the poverty measures toward the presence extreme incomes can be found by Van Kerm (2007b). Notice, the aforementioned properties rely on the assumption, that the incomes are positive or at least bounded from below (Cowell and Victoria-Feser, 2002). The occurrence of a large amount of negative income values needs to be treated separately. In summary, aside from the two inequality measures, Gini coefficient and Quintile Share Ratio, the monetary Laeken indicators – computed on basis of SILC data – can be considered as robust in the terms of a bounded influence function.

There are two non-monetary Laeken indicators which may have robustness problems. The coefficient of variation of the regional unemployment rates may, of course, be influenced very much by single outlying regions. This seems to be of less importance because the regional employment rates themselves will be closely monitored when they are established. The second indicator is the number of peoples living in jobless households. There is clearly a danger that households with a large and possibly incorrect number of members living in them will influence this estimate unduely. Theoretically the influence function of this estimator is unbounded. However, the number of persons in a household is a concept with a relatively clear range of admissible values which will certainly be controlled in an editing and imputation phase and therefore the robustness problem seems to minor for this indicator.

5.2 Univariate Methods

For the non-robust indicators it is therefore customary to make adjustments to extreme data. Several univariate procedures to robustify welfare indices were proposed in the literature. In this report emphasis is put on the following three methods:

- Parametric fitting of personal-income distribution models by means of robust estimation (Victoria-Feser and Ronchetti, 1994). The welfare indices are computed from the robustly estimated income model.

- Parametric tail modelling: A parametric model (usually Pareto distribution) is fitted to the upper tail (above a defined threshold (Dupuis and Victoria-Feser, 2006), (Danielsson et al., 2000) and (Beirlant et al., 2002)) of the empirical distribution using robust methods (Cowell and Victoria-Feser, 2007), (Vandewalle et al., 2007) and (Brazauskas and Serfling, 2000, 2001). Alternatively, Cowell and Flachaire (2007) showed that the ordinary maximum likelihood estimator produces quite reasonable results, either. The welfare measures are then computed using a combination of the empirical distribution function (up to the threshold) and the parametric estimated upper tail (Cowell and Victoria-Feser, 2007) or by replacing the extreme income by draws from the parametric upper tail model (Van Kerm, 2007b).

- Trimming a selected proportion of the data from the tail(s). The welfare measures are computed with the trimmed sample (Cowell and Victoria-Feser, 2003, 2006)

Although all the methods are concerned with extreme income/ outliers, their underlying concept, definition and treatment of outlying (extreme) observations are quite heterogeneous. Hence, they require a particular amount of assumptions that must be made before the analysis (e.g. define parametric income models). In addition they do not, except the trimming method, explicitly account for the requirements of complex samples.

Chapter 6

Summary of the State-of-the-art in Visualisation of the European Laeken Indicators

6.1 Introduction

In the first part of this summary, we will provide an overview of commonly used visualisation techniques for indicators of poverty and social exclusion. The second part is a short review of current software and recent papers on exploratory data analysis with emphasis on highlighting missing or imputed values. Since *European Union Statistics on Income and Living Conditions* (EU-SILC) data contain missing values, identifying the missing data mechanisms and using proper imputation methods is essential for obtaining high quality results when estimating indicators.

6.2 Visualisation of Indicators

Visualisation of indicators for the general public is a domain with diverse but not too many activities. The OECD World Forum on Key Indicators (OECD, 2004) treated among others issues of visualisation. Many ideas have been put forward for visualisation, among others gapminder (see `http://www.gapminder.org`). Gapminder is a very sophisticated visualisation but it is not a general visualisation tool since it is taylor made for specific data.

The International Conference on “Visualising and Presenting Indicator Systems”, 14 - 16 March 2005, Federal Statistical Office, Neuchátel, was an occasion where some ideas were presented (Swiss Federal Statistical Office, 2005). Hulliger (Hulliger, 2005b,a) presented his ideas about depicting variance in graphical displays. His procedures are available in form of R-functions. Figure 6.1 shows the change of RD expenditures of the industry. Nevertheless there is no coherent methodology for presentation of indicators to the public. However, much of the work by statisticians in general about graphical displays in statistics is applicable to indicators as well (see, e.g., Robbins, 2005).

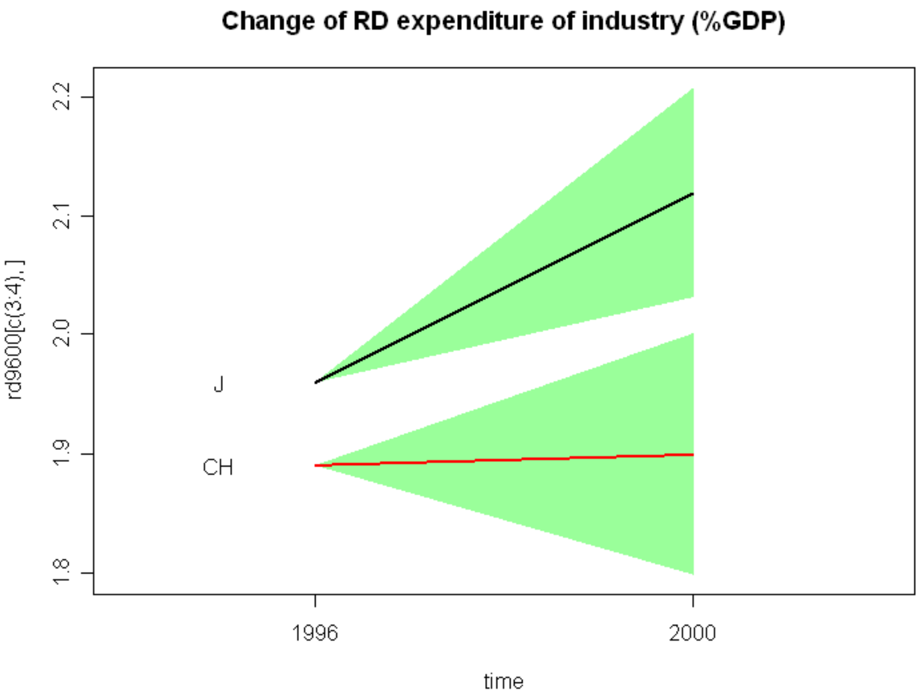
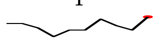


Figure 6.1: Change in time plot.

Hulliger presented work on sustainability indicators and their evaluation at the Q2008 conference (Hulliger et al., 2008) and an internal report of this work in collaboration with the Swiss Federal Statistical Office will be available soon.

In recent Eurostat publications, Laeken indicators are most frequently visualized using barplots (e.g., Eurostat, 2005a). Related indicators, e.g., the *quintile share ratio* and the *Gini coefficient* for measuring the inequality of the income distribution, are sometimes presented as interleaved or juxtaposed bars in the same plot. An example is shown in Figure 6.2. Note that the use of different scales in one plot should be avoided, though. Juxtaposed barplots are also used to visualize the effect of social transfers on the *at-risk-of-poverty rate* (e.g., Eurostat, 2007a).

However, the availability of regional information further allows the indicators to be displayed in maps. While the generation of informative maps can be very sophisticated, their use may be highly appealing to policy makers. Verma et al. (2005) use colored maps to visualize regional indicators (see Figure 6.3 for an example), as well as other information about the regions. They categorize the values of the indicators into five levels, each corresponding to a different level on a contiguous color scale. The regions are then plotted in the color corresponding to the value of the indicator. Furthermore, regions for which no values are available are plotted in white.

Another possibility is to visualise Laeken indicators using sparklines (Tufte, 2001). Sparklines are specific types of information graphics - usually barplots  or polygons  - which are designed to be small in size and possess high data density. The main advantage of sparklines is that those graphs can directly be included in the text, thus showing the variation of the data without the need for explicit graphics.

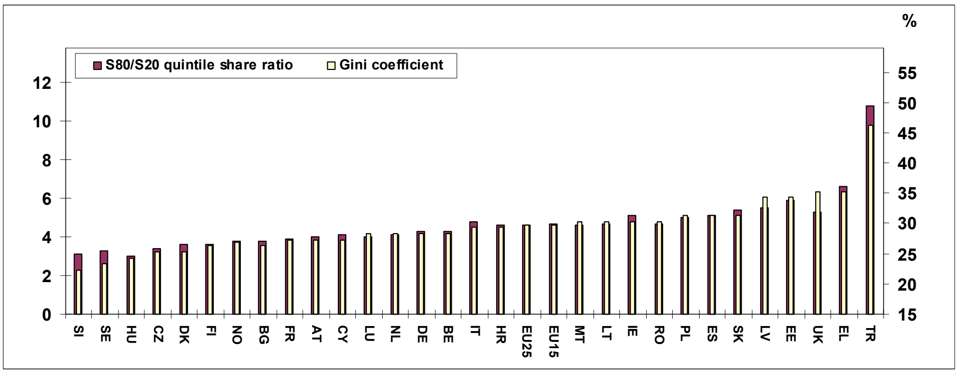


Figure 6.2: Example barplot from Eurostat (2005a) (Figure 4).

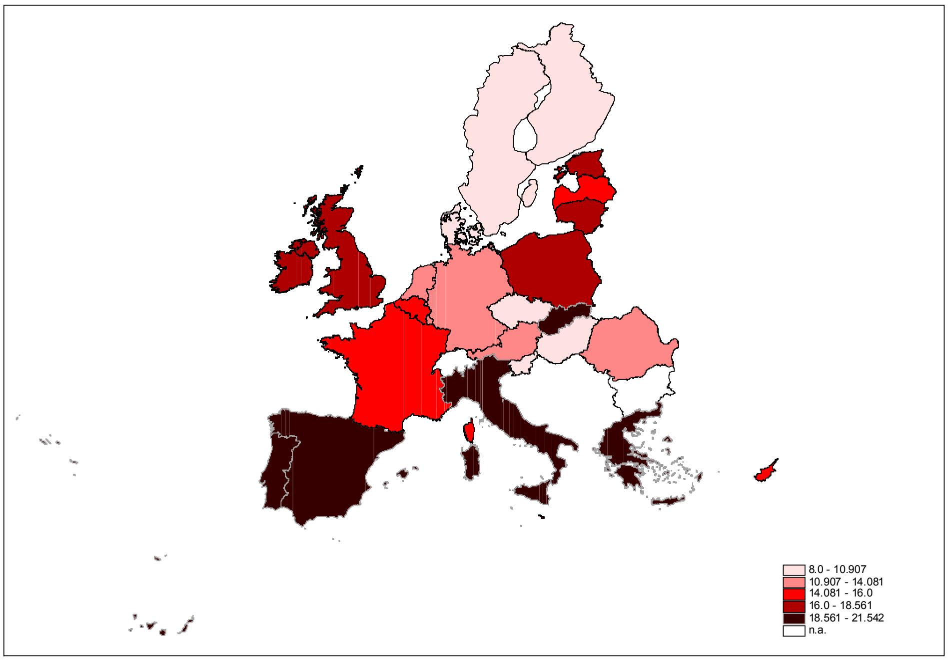


Figure 6.3: Example map from Verma et al. (2005) (Figure 5.1) showing consolidated poverty rates.

Often sparklines are used together as elements of a small multiple (Tufte, 1990). A small multiple is a series of similar graphs.

Thus, using sparklines together with tables leads to *graphical tables* which do not only show current values but are enhanced by graphically showing past information. Currently, an R-package for the creation of sparklines in formats that can be directly used in documents and web-pages is developed at Statistics Austria. All sparkplots in this section have been created using a prototype version of this software.

Examples for a graphical table featuring sparklines is given in table 6.1, another example

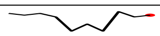
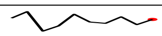
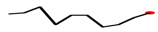
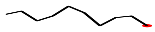
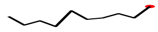
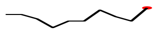
Area	Gini Coeff.	(cur|min|max)	Area	Gini Coeff.	(cur|min|max)
Area1		[35 | 26 | 40]	Area5		[28 | 23 | 36]
Area2		[37 | 29 | 44]	Area6		[37 | 28 | 43]
Area3		[35 | 27 | 41]	Area7		[23 | 17 | 34]
Area4		[32 | 29 | 40]	Area8		[34 | 19 | 35]

Table 6.1: Graphical table with sparklines showing current values and past information.

Area	Gini Coeff.	(cur|min|max)	Area	Gini Coeff.	(cur|min|max)
Area1		[35 | 26 | 40]	Area5		[28 | 23 | 36]
Area2		[37 | 29 | 44]	Area6		[37 | 28 | 43]
Area3		[35 | 27 | 41]	Area7		[23 | 17 | 34]
Area4		[32 | 29 | 40]	Area8		[34 | 19 | 35]

Table 6.2: Graphical table with sparkbars showing current values and past information.

with sparkbars is given in table 6.2.

6.3 Exploring Microdata

Exploratory data analysis is an important first step in every statistical analysis. Proper visualisation tools may help to gain insight on the distribution of the data and possible data problems like missing values at the same time. EU-SILC data are very complex and typically contain numerous data problems. Therefore, the availability of such software is essential for our purpose.

6.3.1 Highlighting Missing Values

Visualisation of missing values is a crucial task as the EU-SILC data contain a high amount of missing values in certain variables, e.g., the individual income components. In order to select an appropriate imputation method, possible patterns of the missing values need to be identified first. Visualisation can be a very powerful tool for the detection of the missing data mechanisms.

Nevertheless, visualisation tools for missing values are rarely or not at all implemented in SAS, SPSS, STATA or even R (R Development Core Team, 2008). Through interaction, missing values can be highlighted in GGobi (Cook and Swayne, 2007) and Mondrian (Theus, 2002). Note that only a few plots are available in GGobi, though. Users of older *Mac OS* operating systems on machines based on the PowerPC architecture can also use MANET (Unwin et al., 1996; Theus et al., 1997) for interactive visualization of data with missings.

Although the interactivity of GGobi, Mondrian and MANET is an advantage, each of these pieces of software has its disadvantages as well. Moreover, for our purpose it is desired that visualisation of missings, imputation and analysis can all be done from within R, without the need of additional software. Thus, an R-package for interactive visualization of missing values is created during the AMELI project and a prototype of this software has been already developed (Templ and Filzmoser, 2008).

6.3.2 Highlighting Imputed Values

In order to avoid meaningless results for imputed data, the quality of the imputations needs to be evaluated. Abayomi et al. (2008) proposed different graphical tools for that purpose (univariate and bivariate plots). They compare the density of the observed part of the data with the density of the imputed part. In addition, they use multiple scatterplots in which the imputed values are highlighted by a different color. However, Abayomi et al. (2008) only address the case of continuous variables. Scatterplot matrices, parallel coordinate plots Wegman (1990) and ternary diagrams (Aitchison, 1986) where the imputed values are highlighted are shown can be found in Templ et al. (2009).

Chapter 7

Summary and Outlook

The present text gives an overview of the state-of-the-art in poverty measurement focusing on the European measures for poverty and social exclusion. Extensions and latest developments are covered in detail in the different workpackages and deliverables of the AMELI project.

Recent advances were made in the ESSnet project for the analysis of EU-SILC (Net-SILC; http://epp.eurostat.ec.europa.eu/portal/page/portal/essnet/eu_silc). The outcome of the project was published in European Commission (2010). Finally, the Sample Project (http://www.sample-project.eu) also focuses on small area estimation methods for poverty measures.

Bibliography

Aaberge, R., **Fjaerli, E.**, **Langorgen, A.** and **Mogstad, M.** (**2007**): *Comparability of Income Data Across Households/Individuals and Over Time*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 59–74.

Aassve, A., **Mazzuco, S.** and **Mencarini, L.** (**2006**): *An empirical investigation into the effect of childbearing on economic wellbeing in Europe*. Statistical Methods & Applications, 15, pp. 209–227.

Abayomi, K., **Gelman, A.** and **Levy, M.** (**2008**): *Diagnostics for multivariate imputations*. Journal of the Royal Statistical Society: Series C (Applied Statistics), 57 (3), pp. 273–291.

Abramovici, G. (**2002**): *Social Protection: Expenditure on Cash Benefits and on Benefits in Kind*. Eurostat: Statistics in Focus: Population and Social Conditions, Theme 3, 16.

Abramovici, G. (**2003**): *Social Protection: Cash Family Benefits in Europe*. Eurostat: Statistics in Focus: Population and Social Conditions, Theme 3, 19.

Aitchison, J. (**1986**): The Statistical Analysis of Compositional Data. Wiley, New York.

Andersson, C. and **Nordberg, L.** (**1994**): *A Method for Variance Estimation of Non-Linear Functions of Totals in Surveys - Theory and Software Implementation*. Journal of Official Statistics, 10 (4), pp. 395 – 405.

Atkinson, A. B. and **Bourguignon, F.** (editors) (**2000**): Handbook of income distribution. Elsevier.

Atkinson, A. B., **Marlier, E.** and **Nolan, B.** (**2004**): *Indicators and Targets for Social Inclusion in the European Union*. Journal of Common Market Studies, Vol. 42, pp. 47–75.

Atkinson, T., **Cantillon, B.**, **Marlier, E.** and **Nolan, B.** (**2002**): Social Indicators: The EU and Social Inclusion. Oxford University Press.

Beirlant, J., **Dierckx, G.**, **Guilliou, A.** and **Starica, C.** (**2002**): *On exponential Representations of Log-Spacings of Extreme Order Statistics*. Extremes, 5.

Berger-Schmitt, R. (**2000**): *Social Cohesion as an Aspect of the Quality of Societies: Concept and Measurement*. EuReporting Working Paper No. 14, Centre for Survey Research and Methodology (ZUMA).

Betti, G. and **Verma, V.** (**2006**): *Gross-Net Conversion Issues.* Statistics Finland International Conference on *Comparative EU Statistics on Income and Living Conditions: Issues and Challenges*, Helsinki.

Brandolini, A. (**2006**): *Measurement of Income Distribution in Supranational Entities: The Case of the European Union.* Working Paper Series 452, Luxembourg Income Study.

Brazauskas, V. and **Serfling, R.** (**2000**): *Robust and Effcient Estimation of the Tail Index of a Single-Parameter Pareto Distribution.* North American Actuarial Journal, 4.

Brazauskas, V. and **Serfling, R.** (**2001**): *Robust Estimation of Tail Parameters for Two-Parameter Pareto and Exponential Models via Generalized Quantile Statistics.* Extremes, 3.

Buhmann, B., **Rainwater, L.**, **SChmaus, G.** and **Smeeding, T. M.** (**1988**): *Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates across Ten Countries using the Luxembourg Income Study (LIS) Database.* Review of Income & Wealth, 34 (2), pp. 115–142.

Burkhauser, R., **Feng, S.**, **Jenkins, S.** and **Larrimore, J.** (**2008**): *Estimating Trends in US Income Inequality using the Current Population Survey: the Importance of Controlling for Censoring.* Technical report, Institute for social and Economic research.

Burkhauser, R. V., **Feng, S.** and **Jenkins, S. P.** (**2007**): *Using the P90/P10 Index to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults* . Working Paper Series 72, Society for the Study of Economic Inequality (ECINEQ).

Butler, R., **McDonald, J.**, **Nelson, R.** and **White, S.** (**1990**): *Robust and Partially Adaptive Estimation of Regression Models.* The review of economics and statistics, 72, pp. 321–327.

CEPS/INSTEAD (**2003**): *CHER Consortium of Household Panels for European Socio-economic Research: Final Report.* Technical report, CEPS/INSTEAD, Luxembourg.

Chotikapanich, D. (editor) (**2008**): Modeling Income Distributions and Lorenz Curves. Springer: Economic Studies in Equality, Social Exclusion and Well-Being, Vol. 5, doi: 10.1007/978-0-387-72796-7.

Commission of the European Communities (**2007**a): *Joint Report on Social Protection and Social Inclusion 2007: Social inclusion, Pensions, Healthcare and Long Term care.* Commission Staff Working Document, Commission of the European Communities.

Commission of the European Communities (**2007**b): *Joint Report on Social Protection and Social Inclusion: Country Profiles.* Commission Staff Working Document SEC(2007) 272, Commission of the European Communities.

Commission of the European Communities (**2007**c): *Joint Report on Social Protection and Social Inclusion: Supporting Document.* Commission Staff Working Document SEC(2007) 329, Commission of the European Communities.

Commission of the European Communities (**2008**): *A Renewed Commitment to Social Europe: Reinforcing the Open Method of Coordination for Social Protection and Social Inclusion*. Technical report COM(2008) 418 final, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

Cook, D. and **Swayne, D.** (**2007**): Interactive and Dynamic Graphics for Data Analysis: With R and GGobi. New York: Springer.

Coulter, F. A. E., **Cowell, F. A.** and **Jenkins, S. P.** (**1992**): *Equivalence Scale Relativities and the Extent of Inequality and Poverty* . The Economic Journal, Vol. 102 (414), pp. 1067–1082.

Council of Europe Publishing (**2005**): Concerted development of Social Cohesion Indicators: Methodological guide. Council of Europe Publishing.

Cowell, F. A. and **Flachaire, E.** (**2007**): *Income distribution and inequality measurement: The problem of extreme values*. Journal of Econometrics, 141, pp. 1044–1072.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**1996**a): *Poverty measurement with contaminated data: A robust approach*. European Economic Review, 40, pp. 1761–1771.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**1996**b): *Robustness properties of inequality measures*. Econometrica, 64 (1), pp. 77–101.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**2002**): *Welfare rankings in the presence of contaminated data*. Econometrica, pp. 1221–1233.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**2003**): *Distribution-free Inference for Welfare Indices under Complete and Incomplete Information*. Journal of Economic Inequality, 00, pp. 1–29.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**2006**): *Distributional Dominance With Trimmed Data*. Journal of Business & Economic Statistics, 24 (3), pp. 291–300.

Cowell, F. A. and **Victoria-Feser, M.-P.** (**2007**): *Robust stochastic dominance: A semi-parametric approach*. Journal of Economic Inequality, 5, pp. 21–37.

Dagum, C. (**1977**): *A New Model of Personal Income Distribution: Specification and Estimation*. Economie Appliquée, 30, pp. 413–437.

D'Alò, M., **Consiglio, L. D.**, **Falorsi, S.** and **Solari, F.** (**2006**): *Small area estimation of the Italian poverty rate*. Statistics in transition, pp. 771–784.

Danielsson, J., **de Haan**, **Peng, L.** and **de Vries, C.** (**2000**): *Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation*. Journal of Multivariate Analysis.

Davison, A. C. and **Sardy, S.** (**2004**): *Resampling Methods for Variance Estimation*. Technical report, DACSEIS deliverable D5.1, http://www.dacseis.de.

Dennis, I. and **Guio, A.-C.** (**2003**a): *Poverty and Social Exclusion in the EU after Laeken-Part 1.* Eurostat: Statistics in Focus: Population and Social Conditions, Theme 3, 8.

Dennis, I. and **Guio, A.-C.** (**2003**b): *Poverty and Social Exclusion in the EU after Laeken-Part 2.* Eurostat: Statistics in Focus: Population and Social Conditions, Theme 3, 9.

Deville, J.-C. (**1999**): *Variance estimation for complex statistics and estimators: Linearization and residual techniques.* Survey Methodology, 25 (2), pp. 193– 203.

Duclos, J.-Y. and **Araar, A.** (**2005**): *An Atkinson-Gini Family of Social Evaluation Functions: Theory and Illustration Using Data from the Luxembourg Income Study.* Working Paper Series 416, Luxembourg Income Study.

Duclos, J.-Y. and **Davidson, R.** (**2006**): *Testing for Restricted Stochastic Dominance.* Working paper series, Luxembourg Income Study.

Dupuis, D. and **Victoria-Feser, M.** (**2006**): *A robust prediction error criterion for Pareto modelling of upper tails.* The Canadian Journal of Statistics, 34 (4).

Ehling, M., **Rendtel, U.** and **al.** (**2004**): Harmonisation of Panel Surveys and Data Quality: Chintex. Statistiches Bundesamt, Wiesbaden.

European Commission (**2006**): *Portfolio of Overarching Indicators and Streamlined Social Inclusion, Pensions, and Health Portfolios.* Technical report, European Commission. Social Protection and Social Integration.

European Commission (**2008**a): Child Poverty and Well-Being in the EU: Current status and way forward. Luxembourg: Office for Official Publications of the European Communities, 246 pp.

European Commission (**2008**b): *Portfolio of Overarching Indicators and Streamlined Social Inclusion, Pensions, and Health Portfolios. April 2008 Update.* Technical report, European Commission. Social Protection and Social Integration.

European Commission (**2010**): Income and living conditions in Europe. European Commission.

Eurostat (**2002**): *European Social Statistics: Income, poverty and social exclusion, 2nd Report (Data 1994-1997 (ECHP))* . Detailed Tables, Eurostat.

Eurostat (**2003**): ECHP UDB Manual, European Community Household Panel Longitudinal Users' Database, Waves 1 to 7, Survey years 1994 to 2000. European Commission. Directorate E: Social and regional statistics and geographical informationsystem.

Eurostat (**2004**): *Income in EU-SILC: Net/Gross/Net Conversion. Report on Common Structure of the Model; Model Description and Application to the ECHP Data for France, Italy and Spain.* Final Report, European Commission. Working Group on Statistics on Income and Living Conditions (EU-SILC), Luxembourg.

Eurostat (**2005**a): *Income poverty and social exclusion in the EU25.* Statistics in Focus, 13/2005.

Eurostat (**2005**b): *The Continuity of Indicators during the Transition between ECHP and EU-SILC*. Working Paper and Studies, European Commission.

Eurostat (**2007**a): Europe in figures: Eurostat yearbook 2006-07. Eurostat.

Eurostat (**2007**b): *European Plan of Research in Official Statistics (EPROS): Main conclusions from the activities in the 5th Framework Programme*. Methodologies and Working Papers, European Commission.

Eurostat (**2007**c): *Optimising the Use of Partial Information in Urban and Regions Systems* . Methodologies and Working Paper p.163-167, European Commission.

Eurostat (**2008**): European Social Statistics: Social Protection, Expenditure and Receipts, Data 1997-2005 . European Communities, Eurostat Statistical Books.

Eurostat (**2009**): *Algorithms to compute Overarching indicators based on EU-SILC and adopted under the Open Method of Coordination (OMC)*. Technical report, European Commission. Directorate F: Social Statistics and Information Society.

Fabrizi, E., **Ferrante, M. R.** and **Pacei, S.** (**2005**): *Estimation of poverty indicators at sub-national level using multivariate small area models*. Statistics in transition, pp. 587–608.

Fabrizi, E., **Ferrante, M. R.** and **Pacei, S.** (**2007**): *Small area estimation of average household income based on unit level models for panel data*. Survey Methodology, pp. 187–198.

Förster, M. and **Pearson, M.** (**2002**): *Income Distribution and Poverty in the OECD Area: Trends and Driving Forces*. OECD Economic Studies No. 34, 2002/I, OECD.

Foster, J., **Greer, J.** and **Thorbecke, E.** (**1984**): *A class of Decomposable Poverty Measures*. Econometrica, 52, pp. 761–766.

Frick, J. R., **Goebel, J.** and **Grabka, M. M.** (**2007**): *Assessing the Distributional Impact of Imputed Rent and Non-Cash Employee Income in Micro-Data*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 119–134.

Gil-Izquierdo, M. (**2007**): *Characterization of Poverty in Spain*. Proceedings of the 56th Session of the International Statistical Institute, Lisboa, Portugal.

Goebel, J. (**2007**): Methodological Issues in the Measurement of Income and Poverty. Ph.D. thesis, Universität Berlin.

Graf, M. (**2007**): *Use of Distributional Assumptions for the Comparison of four Laeken Indicators on EU-SILC Data*. 56th Session ISI 2007.

Guio, A.-C. (**2005**a): *Income Poverty and Social Exclusion in the EU25*. Eurostat: Statistics in Focus: Population and Social Condition, 13.

Guio, A.-C. (**2005**b): *Material Deprivation in the EU*. Eurostat: Statistics in Focus: Population and Social Condition, 21.

Guio, A.-C. and **Bardone, L.** (**2005**): *In-Work Poverty: New commonly agreed indicators at the EU level*. Eurostat: Statistics in Focus: Population and Social Condition, 5.

Guio, A.-C. and **Dennis, I.** (**2003**): *Monetary Poverty in EU Acceding and Candidate Countries*. Eurostat: Statistics in Focus: Population and Social Condition, Theme 3, 21.

Guio, A.-C. and **Dennis, I.** (**2004**a): *Monetary Poverty in New Member States and Candidate Countries*. Eurostat: Statistics in Focus: Population and Social Condition, 12.

Guio, A.-C. and **Dennis, I.** (**2004**b): *Poverty and Social Exclusion in the EU*. Eurostat: Statistics in Focus: Population and Social Condition, 16.

Hampel, F., **Ronchetti, E.**, **Rousseeuw, P.** and **Stahel, W.** (**1986**): Robust statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, New York: John Wiley & Sons Inc., ISBN 0-471-82921-8, the approach based on influence functions.

Hulliger, B. (**2005**a): *Displays of indicators and their accuracy*. Conference on Visualising and Presenting Indicator Systems, march 2009.
URL http://www.bfs.admin.ch/bfs/portal/en/index/themen/21/11/visu/02.parsys.0011.downloadList.00111.DownloadFile.tmp/hulliger.pdf

Hulliger, B. (**2005**b): *On displaying indicators and their accuracy*. Knowledge Economy Indicators Conference, Tübingen: The KEI Project, invited paper.

Hulliger, B., **Lussmann, D.**, **Kohler, F.**, **Mayerat, A.-M.** and **de Montmollin, A.** (**2008**): *Evaluation of Indicators on Environment and Sustainability*. European Conference on Quality in Official Statistics, Rome: Eurostat and ISTAT.

Hulliger, B. and **Münnich, R.** (**2006**): *Variance Estimation for Complex Surveys in the Presence of Outliers*. Proceedings of the American Statistical Association, Survey Research Methods Section, pp. 3153–3161.

Jäntti, M. (**2007**): *The EU-SILC in Comparative Income Distribution Research: Design and Definitions in International Perspective*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 313–334.

Jenkins, S. P. (**2007**): *Inequality and the GB2 Income Distribution*. ECINEQ Working Paper Series, Society for the Study of Economic Inequality.

Jenkins, S. P. and **Van Kerm, P.** (**2003**): *Linking Income Mobility and Inequality: A Re-Assessment of US and German Inequality Trends*. WIDER Conference on Inequality, Poverty and Human Well-Being, Helsinki.

Kerm, P. V. (**2003**): *An anatomy of Household Income Volatility in European Countries*. Technical report, European Countries in European Commission under the Transnational Access to major Research Infrastructures programme.

Kleiber, C. and **Kotz, S.** (**2003**): Statistical Size Distributions in Economics and Actuarial Sciences. Hoboken, NJ: John Wiley & Sons.

Kovacevic, M. S. and **Binder, D. A.** (**1997**): *Variance Estimation for Measures of Income Inequality and Polarization - The Estimating Equations Approach.* Journal of Official Statistics, 13 (1), pp. 41 – 58.

Kubitza, A. (**2004**): *Pensions in Europe: Expenditure and Beneficiaries.* Eurostat: Statistics in Focus: Population and Social Condition, 8.

Laan, P. V. D. (**2007**): *The Income Concept in EU-SILC: Relevance, Feasibility, Challenges.* Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 47–57.

Lemmi, A. and **Verma, V.** (**2003**): *Regional Indicators to Reflect Social Exclusion and Poverty (VT/2003/43): Final Report.* Technical report, European Commission. Employment and Social Affairs DG.

Lohr, S. (**1999**): Sampling: Design and Analysis. Pacific Grove: Duxbury Press.

Macredie, I., **Diop, L.**, **Holt, T.**, **Hulliger, B.**, **Oudhof, K.** and **Sastry, N.** (**2001**): *An assessment of the statistical criticism made of the Human Development Report, 1999.* **Commission, U. S.** (editor) Report of the Friends of the Chair to the Statistical Commission.

Marco, M. D. (**2007**): *Self-Employment Incomes in the Italian EU-SILC: Measurement and International Comparability.* Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 145–156.

Marlier, E. and **Cohen-Solal, M.** (**2000**): *Social Benefits and their Redistributive Effect in the EU.* Eurostat: Statistics in Focus: Population and Social Condition, Theme 3, 9.

McCarthy, P. J. and **Snowden, C. B.** (**1985**): The bootstrap and finite population sampling. Data Evaluation and Methods Research. Series 2, no. 95, Vital and health statistics, U.S. Dept. of Health, Education, and Welfare.

McDonald, J. (**1984**): *Some Generalized Functions for the Size Distribution of Income.* Econometrica, 52 (3), pp. 647–663.

McDonald, J. (**1989**): *Alternative Beta Estimation for the Market Model using Partially Adaptive Techniques.* Communications in Statistics Theory and Methods, 16, pp. 4032–4058.

McDonald, J. B. and **Butler, R. J.** (**1987**): *Some Generalized Mixture Distributions with an Application to Unemployment Duration.* The Review of Economics and Statistics, 69, pp. 232–240.

McDonald, J. B. and **Butler, R. J.** (**1990**): *Regression Models for Positive Random Variables.* Journal of Econometrics, 43, pp. 227–251.

Milanovic, B. (**2002**a): *True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone.* The Economic Journal, 112, pp. 51–92.

Milanovic, B. (**2002**b): Worlds Apart: Inter-National and World Inequality 1950-2000.

Milanovic, B. (**2005**): Worlds Apart: Measuring International and Global Inequality. Princeton University Press.

Milanovic, B. and **Yitzhaki, S.** (**2002**): *Decomposing World Income Distribution: Does the World Have a Middle Class?* . Review of Income and Wealth, Serie 48 (Number 2), pp. 155–178.

Münnich, R. (**2005**): *Datenqualität in komplexen Stichprobenerhebungen*, unpublished thesis, university Tübingen.

Münnich, R. (**2008**): *Varianzschätzung in komplexen Erhebungen.* Austrian Journal of Statistics, 37 (3&4), pp. 319 – 334.

Münnich, R. (**2001**): *Data Quality in Complex Surveys whithin the New European Information Society (DACSEIS).* Proceedings of the 2001 ETK/NTTS Conference in Hersonissos, Crete.

Neocleous, T. and **Portnoy, S.** (**2008**): *A Partially Linear Censored Quantile Regression Model for Unemployment Duration.* Working Paper 2008-07, IRISS, Luxembourg.

OECD (**2004**): *OECD World Forum on Key Indicators.* May 2009.

OECD and **JRC** (**2008**): Handbook on Constructing Composite Indicators, METHODOLOGY AND USER GUIDE. OECD.

Osier, G. (**2009**): *Variance estimation for complex indicators of poverty and inequality using linearization techniques.* Survey Research Methods, 3 (3), pp. 167 – 195.

Pochet, P. (**2005**): The Open Method of Co-ordination and the Construction of Social Europe: A Historical Perspective. Brüssel, Belgien: Peter Lang.

R Development Core Team (**2008**): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rao, J. N. K. and **Shao, J.** (**1996**): *On balanced half-sample variance estimation in stratified random sampling.* Journal of the American Statistical Association, 91 (433), pp. 343 – 348.

Rao, J. N. K. and **Wu, C. F. J.** (**1988**): *Resampling inference with complex survey data.* Journal of the American Statistical Association, 83 (401), pp. 231 – 241.

Robbins, N. (**2005**): Creating More Effective Graphs. Wiley.

Rodrigues, C. F. (**2007**): *Income in EU-SILC-Net/Gross Conversion Techniques for Building and using EU-SILC Databases.* Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 159–172.

Shao, J., **Chen, Y.** and **Chen, Y.** (**1998**): *Balanced repeated replication for stratified multistage survey data under imputation*. Journal of the American Statistical Association, 93 (442), pp. 819 – 831.

Shao, J. and **Tu, D.** (**1995**): The jackknife and bootstrap. New York: Springer.

Shao, J. and **Wu, C. F. J.** (**1989**): *A general theory for jackknife variance estimation*. The Annals of Statistics, 17 (3), pp. 1176 – 1197.

Sitter, R. R. (**1992a**): *A resampling procedure for complex survey data*. Journal of the American Statistical Association, 87, pp. 755 – 765.

Social Protection Committee (**2001**): *Report on Indicators in the Field of Poverty and Social Exclusion*. Technical report, European Comission.

Social Protection Committee (**2003**): *Mid Year Report from the Indicators Sub-Group to the Social Protection Committee*. Technical report, European Comission.

Social Protection Committee (**2005**): *Common Indicators Relating to National Strategies for Adequate and Sustainable Pensions: Second Progress report of the Indicators Sub-Group of the SPC to the Social Protection Committee*. Progress Report, European Comission.

Swiss Federal Statistical Office (**2005**): *Visualising and Presenting Indicator Systems*.

Templ, M. and **Filzmoser, P.** (**2008**): *Visualization of missing values using the R-package VIM*. Research report CS-2008-1, Department of Statistics and Probability Theory, Vienna University of Technology.

Templ, M., **Filzmoser, P.** and **Hron, K.** (**2009**): *Robust Imputation of Missing Values in Compositional Data Using the R-Package robCompositions*. Conference on New Techniques and Technologies for Statistics.

The Canberra Group (**2001**): *Expert Group on Household Income Statistics*. Final Report and Recommendations, The Canberra Group, Ottawa.

Theus, M. (**2002**): *Interactive Data Visualization using Mondrian*. Journal of Statistical Software, 7 (11), pp. 1–9, ISSN 1548-7660.
URL http://www.jstatsoft.org/v07/i11

Theus, M., **Hofmann, H.**, **Siegl, B.** and **Unwin, A.** (**1997**): *MANET - Extensions to Interactive Statistical Graphics for Missing Values*. In New Techniques and Technologies for Statistics II, pp. 247–259, IOS Press.

Törmälehto, V.-M. (**2007**): *Measurement of Property Income in EU-SILC*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 99–115.

Tufte, E. (**1990**): Envisioning Information. Graphics Press.

Tufte, E. (**2001**): The Visual Display of Quantitative Information. Graphics Press.

United Nations Development Programme (**2007**): Human Development Report 2007/2008 Fighting climate change: Human solidarity in a divided world. 1 UN Plaza, New York, New York, 10017, USA: UNDP.

Unwin, A., **Hawkins, G.**, **Hofmann, H.** and **Siegl, B.** (**1996**): *Interactive Graphics for Data Sets with Missing Values: MANET*. Journal of Computational and Graphical Statistics, 5 (2), pp. 113–122.

Van Kerm, P. (**2007**a): *Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC*. IRISS-C/I Working Paper.

Van Kerm, P. (**2007**b): *Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC*. IRISS-C/I Working Paper.

Vandewalle, B., **Beirlant, J.**, **Christmann, A.** and **Hubert, M.** (**2007**): *A Robust Estimator for the Tail Index of Pareto-type Distributions*. Elsevier Science.

Verma, V., **Betti, G.**, **Lemmi, A.**, **Mulas, A.**, **Natilli, M.**, **Neri, L.** and **Salvati, N.** (**2005**): *Regional indicators to reflect social exclusion and poverty. Final report. Project VT/2003/45*. Technical report, Centro di Ricerca Interdipartimentale sulla Distribuzione del Reddito, Università di Siena.

Victoria-Feser, M.-P. (**2000**): *Robust Methods for the Analysis of Income Distribution, Inequality and Poverty*. International Statistical Review, 68, 3, pp. 277–293.

Victoria-Feser, M.-P. and **Ronchetti, H.** (**1994**): *Robust Methods for Personal-Income Distribution Models*. The Canadian Journal of Statistics, 22,2, pp. 247–258.

Ward, T. (**2007**): *The Differential Access of Women and Men to Employment and Income: Evidence from the EU-SILC*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 231–246.

Wegman, E. (**1990**): *Hyperdimensional data analysis using parallel coordinates*. Journal of the American Statistical Association, 85, pp. 664–675.

Whelan, C. T., **Watson, D.** and **Maitre, B.** (**2007**): *Social Class Variation in Income Poverty, Deprivation and Consistent Poverty: An Analysis of EU-SILC*. Comparative EU Statistics on Income and Living Conditions : Issues and Challenges. Proceedings of the EU-SILC conference (Helsinki, 6-8 November 2006), pp. 261–278.

Wolter, K. M. (**1985**): Introduction to variance estimation. New York: Springer.

Wolter, K. M. (**2007**): Introduction to variance estimation. New York: Springer.

Yu, K., **Van Kerm, P.** and **Zhang, J.** (**2004**): *Bayesian Quantile Regression: An application to the wage distribution in 1990s Britain*. Technical report, CEPS/INSTEAD, G.-D. Luxembourg.