Università di Siena
Centro di Ricerca Interdipartimentale sulla Distribuzione del Reddito

Regional Indicators to reflect social exclusion and poverty VT/2003/43

Final Report

European Commission
Employment and Social Affairs DG
B – 1409 Brussels
Belgium

Project Director
Prof. Achille Lemmi

Research Director
Prof. Vijay Verma

Contributors
Gianni Betti
Achille Lemmi
Anna Mulas
Michela Natilli
Laura Neri
Nicola Salvati
Vijay Verma

Università degli Studi di Siena
C.R.I.DI.RE. (Centro di Ricerca Interdipartimentale sulla Distribuzione del Reddito)
c/o Dipartimento di Metodi Quantitativi, Piazza S.Francesco 8, 53100 Siena (Italy) tel 0577 298604 fax 0577 298626
E-mail: Lemmi@Unisi.It

Acknowledgement

The project research team would like to thank the national statistical institutes of Italy, Romania and Albania, and University of Finance and Management, Poland, for the production and provision of the survey data used; to Eurostat for access to ECHP data and Eurostat Free Dissemination Database, and the regional boundaries database; to Professor Panek for facilitating the handling of the Polish data; to Francesca Ballini for assistance in the preparation of research proposal for this projects, computations for Chapter 9, and especially the maps and material presented in Chapter 14; to Laura Bardone for monitoring and supporting the project; and to other members of the Commission who provided helpful comments on the draft reports.

Contents

List of Tables

Chapter 11

Chapter 12

Chapter 13

Chapter 14

Chapter 15 Statistical Annex

List of Figures

Chapter 6

Chapter 7

Chapter 8

Chapter 13

Chapter 14

Executive Summary

1 Project objectives

As provided in the project Terms of Reference (ToR), the aim of this project is to identify the appropriate methodologies and strategies for the development of indicators of poverty and social exclusion at regional level, the ultimate goal being the development of a coherent and integrated strategy for the incorporation of the regional dimension into the Member States' NAP/incl. This study takes as its point of departure the methodological framework used for defining the indicators of poverty and social exclusion endorsed at Laeken, and more generally, existing methodological research and data in the area of indicators of poverty and social exclusion as well as in the area of regional indicators.

The strategy we recommend for the construction of regional indicators of poverty and deprivation has three fundamental aspects: (a) making the best use of available sample survey data, such as by cumulating and consolidating the information so as to obtain more robust measures which permit greater spatial disaggregation; (b) exploiting to the maximum 'meso' data - such as the highly disaggregated tabulations available in Eurostat Free Data Dissemination (NewCronos) - for the purpose of constructing regional indicators; and (c) using the two sources in combination to produce more precise estimates for regions using appropriate small area estimation (SAE) techniques. We have aimed to address a number of major issues or challenges:

(1) Identifying special features and requirements of the system of indicators of poverty and social exclusion for use at the regional level.

(2) Choosing appropriate units of analysis.

(3) Describing a practical strategy for measuring poverty and social exclusion at the regional level.

(4) Illustrating the recommended strategy concretely, with necessary technical detail on the basis of real statistical data.

(5) Constructing income poverty-related indicators appropriate for the regional level.

(6) Incorporating with increased emphasis non-monetary dimensions of deprivation to complement indicators of income poverty.

(7) Extending indicators normally produced at the national level to the level of regions – going down to NUTS2 level, then to NUTS3 level and even beyond.

2 Choice of appropriate regional indicators

One of the major tasks in our research was to "assess the extent to which indicators of social exclusion and poverty of the type endorsed at Laeken can be applied at the regional level, either using EU-wide or national sources".

Our conclusions are as follows. Simply the introduction of more extensive breakdown is neither possible because of data limitations, nor sufficient in itself. It is necessary to *focus on the more basic among the Laeken set of indicators*. This is because of the substantially increased data requirements when the results have to be disaggregated geographically.

Indicators of income poverty and non-monetary deprivation

For the purpose of regional indicators, the focus has to be primarily on ordinary *poverty rates* for the total population, and possibly for some special groups such as children, the elderly, and youth entering the labour market.

Certain more complex poverty and inequality measures, which are more sensitive to details and irregularities of the empirical income distribution, are less suited for disaggregation to small populations and small samples. Examples are Gini coefficient and decile ratios (S80/S20); to a lesser extent, the same applies to Laeken indicators 'relative median at-risk-of-poverty gap, by age and gender' and 'at-risk-of-poverty rate before social transfers, by age and gender'. For the same reason, emphasis has to be shifted away from the study of trends over time and longitudinal measures to essentially cross-sectional measures. The main simplification we propose is to focus on *longitudinal indicators defined over short time periods* such as over *pairs of adjacent years*. Such measures can be aggregated over suitable time periods, so as to illuminate the more stable aspects of patterns of variation across regions.

On the other hand, there is need for *additions* to the existing indicators developed primarily for application at national level - region-specific indicators able to capture aspects which are essentially regional. It is possible that a more diverse "portfolio of indicators" is required for the purpose of addressing concerns of regional policy and research.

Perhaps the most important of these is simply the *mean income levels* of the regions, the dispersion among which provides a measure of regional disparities. Secondly, it is essential to consider regional poverty rates defined in relation not only to the national poverty line, but also *regional poverty lines*, i.e., determined in relation to the income distribution separately within each region. *General entropy measures* may also be useful because they can be decomposed into within and between region components.

One of the important objectives of this work is to incorporate, with increased emphasis, non-monetary dimensions of deprivation to complement indicators of income poverty. Putting together individual items of information in a statistically valid manner, we have constructed indices reflecting five different dimensions of deprivation: (1) the lack of ability to afford most basic requirements; (2) enforced lack of widely desired possessions; (3) lack of basic housing facilities; (4) serious problems with accommodation; and (5) problems with the neighbourhood and the environment. Using these, the measures of income poverty can all be generalised to *multi-dimensional measures of deprivation*: any-time, persistent or continuous incidence of non-monetary deprivation, and also of combined incidence of income and non-monetary forms of deprivation.

Other Laeken indicators

We do not consider indicators such as 'self defined health status by income level' or ' life expectancy at birth, by gender' to be statistically feasible at the regional level. Indicators 'regional cohesion (dispersion of regional employment rates)' has been criticised for not providing statistically valid or substantively meaningful information for comparisons across countries. However, we consider this indicator useful, and have proposed some refinements to the methodology of its construction..

By contrast, other labour force and education related Laeken indicators are likely to be quite suitable and useful at the regional level. They often come from large data sources. Our statistical work also confirms that infant morality rate (IMR) is often a remarkably good predictor of normal deprivation indicators. This measure is also more easily estimated at any level of aggregation, normally from administrative sources directly. Hence we consider IMR to be a more suitable indicator than, for instance, life expectancy at birth, especially in the regional context.

Area-level indicators

An important question is the extent to which regional deprivation can be defined as a self-contained concept, different from individual deprivation.

The important addition to the set of Laeken indicators would be the incorporation of indicators defined and measured at the area level in order to identify, as it were, the "territorial reality" of the region. These indicators are not necessarily simple aggregations of individual level values. It is this sort of indicators which underpin area-based policies that have become a common part of some governments' approach to tackling social exclusion. We give examples in particular for UK.

One has to be careful with indicators relating to geographical access to services, transport, crime, pollution, etc. Some of these indicators may actually not reflect individual-level deprivation. For instance, poor geographical access to services may simply result from the fact that an area is a rich one and the population can rely on private means and services. Or, depending on the context, it may indicate real deprivation.

3 Choice of units for the construction of regional indicators

The definition and choice of appropriate units to serve as 'regions' for the construction of poverty and related indicators is a most fundamental issue to be considered at the outset.

Indicators for units based on *functional criteria* can be suited for specialised purposes and populations; this type of units are generally less suitable for general purpose use.

There is also a wide scope for the application of indicators based on *urban-rural classification*: policy requirements and objective can be quite distinct according to the types of place. However, there are no universally agreed criteria as to what constitutes 'rural' and 'urban'.

For a number of substantive and practical reasons, we consider geographical-administrative regions, specifically NUTS regions at various level of classification, as the most appropriate choice. NUTS system provides units which are hierarchical and cover the entire population exhaustively without overlap. Most importantly, a great deal of statistical information is already available for NUTS regions which can be used to construct indicators of regional poverty and deprivation, and to obtain 'covariates' which can be used in statistical models to produce more precise regional indicators of poverty and deprivation.

In our numerical illustrations, we have made extensive use in this manner of NewCronos tabulations, which provides many variables of interest to NUTS2 level, and some to NUTS3.

The results of individual computations (using a single specific poverty line for instance) tend to be too sensitive to irregularities in the empirical income distribution based on small regional samples. Our basic recommendation is to "consolidate" the measures - such as in the form of suitably weighted averages over different numerical measures computed for each wave with different poverty lines thresholds.

4 Small Area Estimation (SAE): Approach and Methodology

In this project we have taken the view that rather than discussing the SAE procedures in general terms independently of the actual data situation, it is more useful to *develop and implement the estimation procedures in concrete terms on the basis of the data sources as actually available to us* in the context of this project. Such a practical approach is much more likely to bring out the variety of situations and problems one may actually encounter in the course of producing regional indicators of poverty and deprivation.

This approach is constrained by the data and country-specific knowledge actually available to us in the context of this project. In national implementation of this and similar methodologies, data and knowledge of specifics of the situation can be expected to be more favourable. The same can be expected in relation to the time and resources available for the purpose.

The particular approach we have developed is constrained by the data availability situation. For EU15 countries, we have access to micro data for up to 8 ECHP waves; for Poland and Romania, single waves of similar surveys are available. For the remaining EU Member State and Candidate countries, only a very limited set of indicators published by Eurostat are available to the project team; these indicators are at the national level, and all concern only monetary poverty. For all countries (except Turkey) NewCronos tables are available, though individual items of information may be lacking to varying degrees.

In most surveys, the code to identify NUTS1 regions is present. By contrast, in most ECHP survey data, no code is available to identify NUTS2 or lower level areas. In view of the above data situation, the modelling options we have implemented are as follows.

	Data situation	Type of estimator preferred
	Access to area-coded survey data + Access to area-level correlates +Unclustered samples	Composite (area-level EBLUP)
	Lack of access to area-coded survey data, or access only to country-level survey estimates + Access to area-level correlates	Synthetic (regression-prediction)

In countries where no area-coded survey data are available, we had to resolve to much simpler and cruder modelling. The procedure we have followed is to use the regression coefficients determined from the corresponding EBLUP model, and simply use these coefficients to predict the target variables on the basis of available predictors for the areas concerned from NewCronos tables. The results of such modelling depend critically on how good the available covariates are in predicting the target variables. *The illustrative results presented here should be treated with caution, pending the development of better models on the basis of better data.*

With the type of surveys available (HBS, ECHP, or subsequently EU-SILC), the sample sizes are generally too small to provide useful information for estimation at NUTS4 or NUTS5 level, even after consolidation of the data over a number of years. It is also not possible to go beyond NUTS3 in the type of models developed here using NewCronos data since those data are available with up to NUTS3 breakdown at most. Production of estimates at lower (NUTS4 and NUTS5) levels would require models of a different type. These models involve imputing the required target variables – such as poverty measures –

to areas or to individual households in a large data set such as a population census, essentially on the basis of a regression model fitted from a small-scale survey containing common covariates and the required target variables. We give examples of these two types of models from the UK and Albania, respectively.

5 Illustrative results

In this project, we have considered it important to illustrate the recommended strategy concretely, with necessary technical detail on the basis of real statistical data. We have felt that, for this work to be *useful* to the Commission, it is absolutely necessary to go beyond simply 'offering' a strategy: hopefully this will facilitate others in replicating and improving the approaches we have recommended. This aspect of our work, which has been by far the most intensive and time consuming, is essential for meeting the ToR objective to "identify data gaps and propose a strategy for the development of a system of relevant indicators of social exclusion and poverty at the regional level". Of course, it must be stressed that the numerical results we present throughout this report are more illustrative than 'final' or 'best' estimates. Others with access to more data and more national knowledge can undoubtedly produce more precise results. Below we give a small selection of some pertinent results.

Income poverty and non-monetary deprivation in combination

We define 'manifest deprivation' as a measure of the individual being subject to both forms of deprivation – income and non-monetary – simultaneously. By 'latent deprivation' we mean the incidence of either form.

Figure 8.20 Manifest deprivation rate as a percentage of Latent deprivation rate, against a measure of the level of poverty or deprivation in the country. EU15

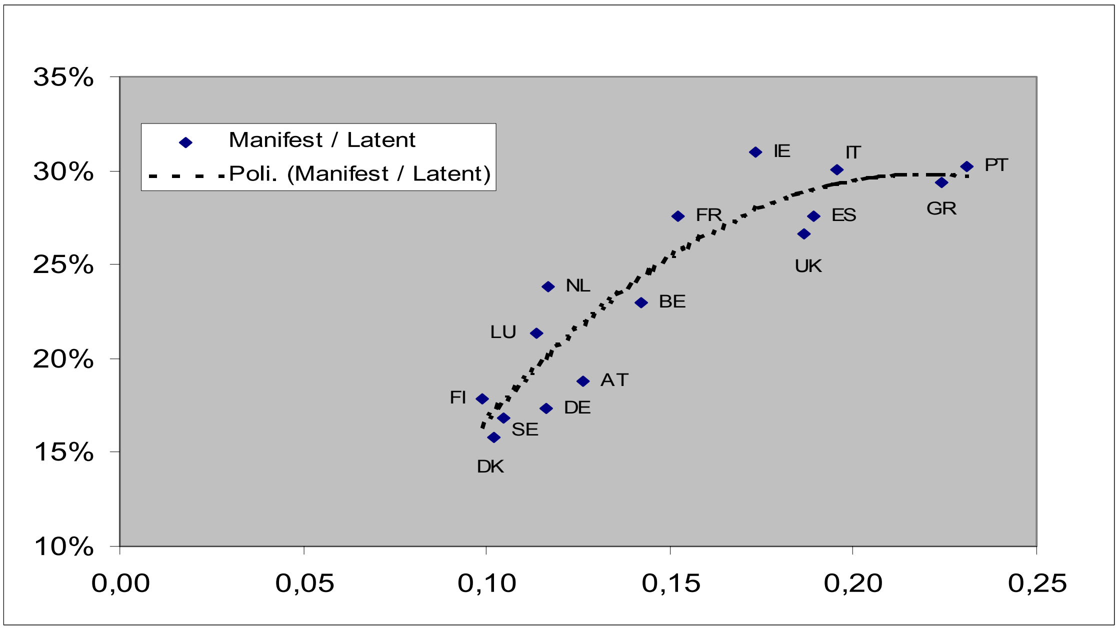


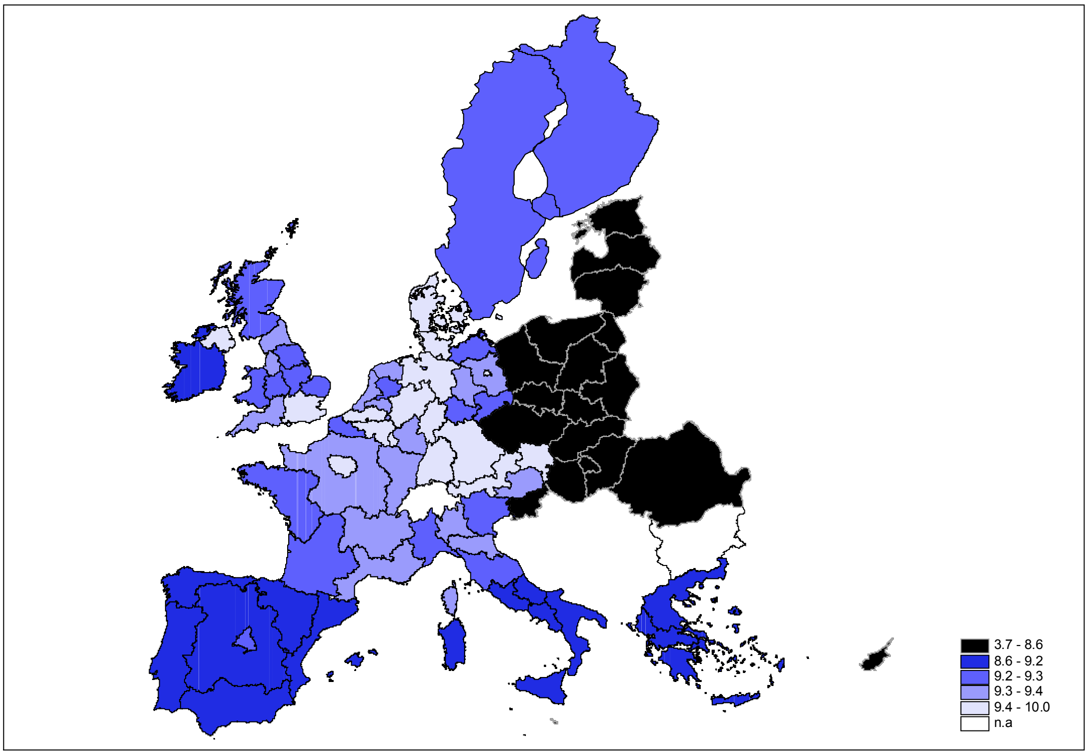
Figure 8.20 shows the manifest deprivation rate as percentage of the latent deprivation rate: this ratio could be interpreted as an index of the degree of overlap of the two forms of deprivation. It is important to highlight that the adoption of a multi-dimensional approach is particularly important when analysing richer countries where different dimensions have

less overlap. On the other hand, there is higher degree of overlap in poorer (and less equal) countries: hence for the individuals involved overall deprivation is more intense in these countries, which is also important. All this underlines the need to supplement monetary indicators by multi-dimensional measures.

Mean level of income

While national level indicators have to be simplified for application at the regional level, there is also the need for *adding* more region-specific indicators able to capture aspects which are essentially regional. Perhaps the most important of these is simply the *mean income levels* of the regions, the dispersion among which provides a measure of regional disparities. Figure 8.5 reports the variable net household equivalent income (in logarithm terms estimated by small-area estimation models) at NUTS1 level. The New Member States have much lower income levels (in PPS) compared to former EU15 countries. More directly relevant are the large regional disparities within countries like Italy and Spain. Note also the higher mean incomes in metropolitan centres (Paris, London).

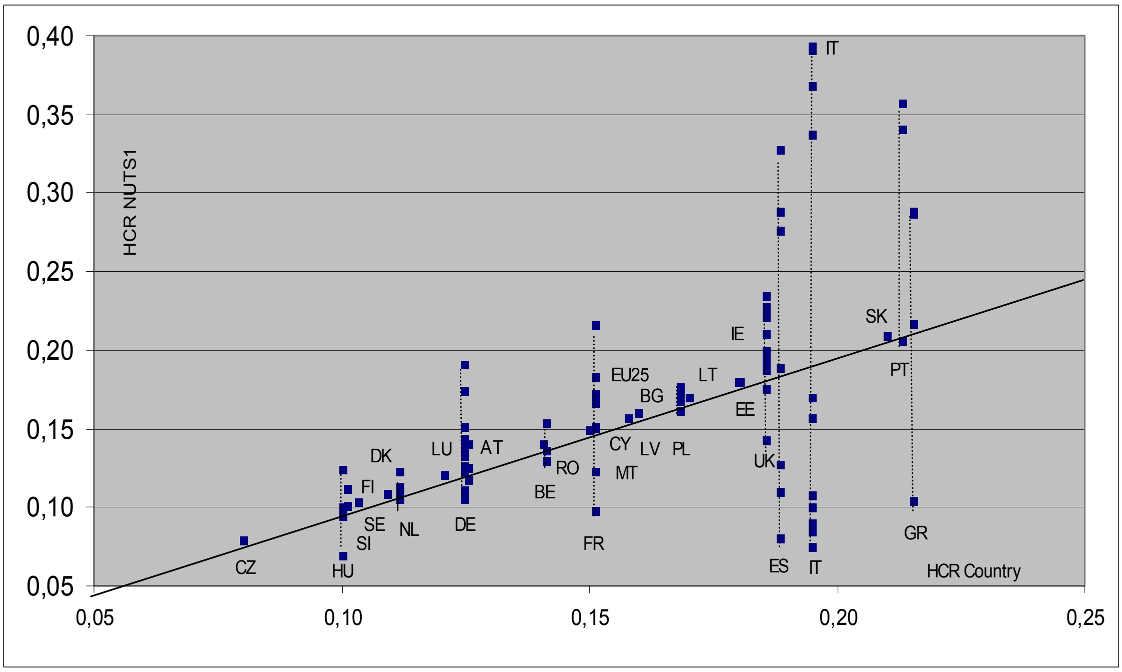
Figure 8.5 Net equivalised income (log). NUTS1 regions



Regional dispersion of poverty rates

Figure 8.4 shows the variation of income poverty rates across NUTS1 regions (poverty rate for each region is computed using *national* poverty line). Note that countries with the highest national poverty rates (towards the right in the figure) also have the most pronounced regional differences in these rates. It is also interesting that most of NMS10 have poverty rates lower than EU15 countries, and generally less regional variation in the rates.

Figure 8.4 Dispersion of Head Count Ratios of NUTS1 regions within countries - country poverty lines



The poverty line level

The choice of the level of poverty line is an important consideration in the construction of indicators suited for regional analysis. By "the level of poverty line" we mean the population level at which the income distribution is pooled for the purpose of defining the poverty line. All poverty related indicators in the Laeken list are based on country poverty lines, meaning that poverty line is always determined on the basis of national income distribution, for instance as 60% of the *national* median income. The income distribution is considered separately at the level of each country, in relation to which a poverty line is defined and the number (and proportion) of poor computed. These numbers may then be pooled over countries to obtain the EU poverty rate (but defined in terms of national poverty lines). Similarly, we may disaggregate the number poor according to the national poverty line in each country by region, and obtain regional poverty rates (but again defined in terms of the national poverty line).

For defining regional indicators, poverty lines defined at the regional level are also relevant. That is, in addition to using national poverty lines, we should consider the income distribution of, for instance, each NUTS2 region separately, and on this basis define a NUTS2 poverty line (generally differing from one region to another), and determine the NUTS2 poverty rate with the poverty line so defined at NUTS2 level. *We recommend regional poverty rates with national and regional poverty lines as two separate indicators.*

Specifically, for regions at NUTS2 level or below, both country level and NUTS2 level poverty lines provide useful information.

Figure 12.1 Head Count Ratio NUTS2 regions (country poverty lines)

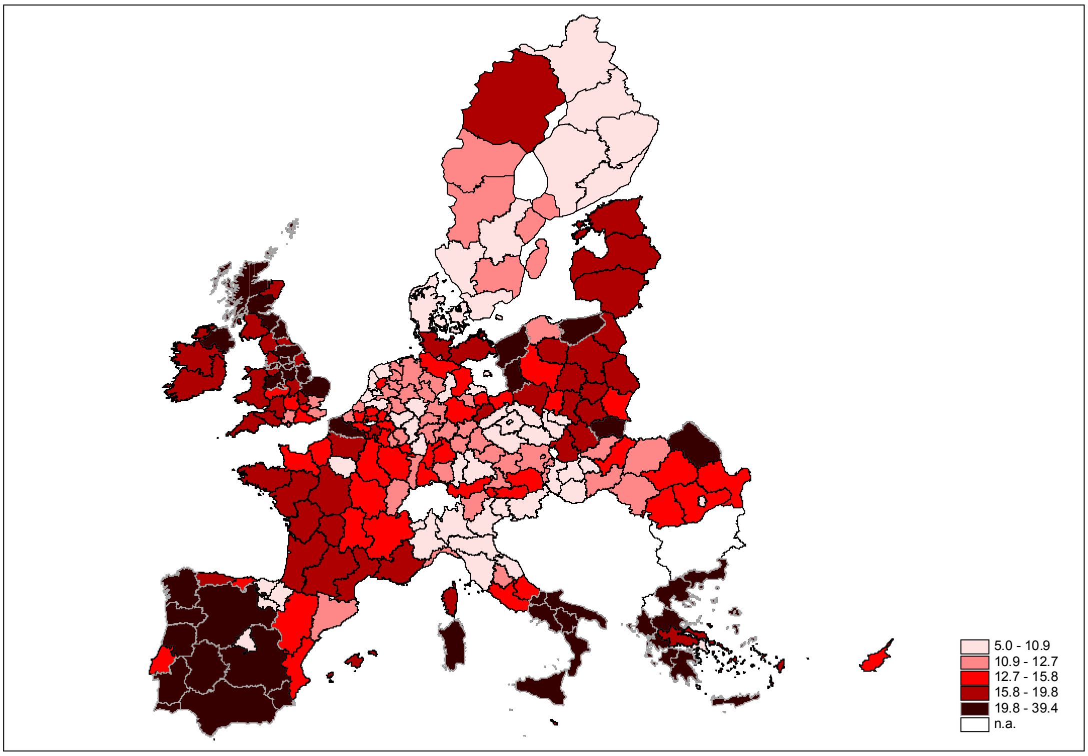


Figure 12.1 shows poverty rates for NUTS2 regions using national poverty lines. For five countries with necessary survey data available at NUTS2 level - Italy, UK, Portugal, Poland, Romania - we have used the more sophisticated small-area estimation (SAE) procedures to construct these rates; for all other countries divided into NUTS2 regions, cruder 'regression-prediction' has been used, and the results should therefore be treated with caution.

For the five countries for which more precise estimates using SAE procedures have been made, Figures 5.9 and 5.10 show the scatter of NUTS2 poverty rates for each country, the first computed using country poverty lines, and the second using NUTS2 poverty lines. It is evident that country level poverty line is almost an "absolute" poverty line for a NUTS2 region, while NUTS2 line is a purely relative line for each region. In countries such as Italy with large variations in mean regional incomes, with country level poverty lines the poverty rates in less well-off regions are inflated and those in more well-off regions reduced; consequently, the scatter of regional poverty rates around the national average is increased. By contrast, using NUTS2 poverty lines gives less scatter, measuring only inequality *within* each region. The difference in the results with the two types of poverty lines is less marked in countries with smaller regional differences in mean income levels.

Poverty rates in NUTS3 regions: Illustration from Italy

Because of special access to micro data from Italian ECHP, provided by ISTAT, and also the above-average size of the available sample, we are able to produce some illustrative estimates of poverty and deprivation at NUTS3 (provincial) level for Italy.

Figure 5.9 Head Count Ratio NUTS2 (using country poverty lines) - dispersions within countries

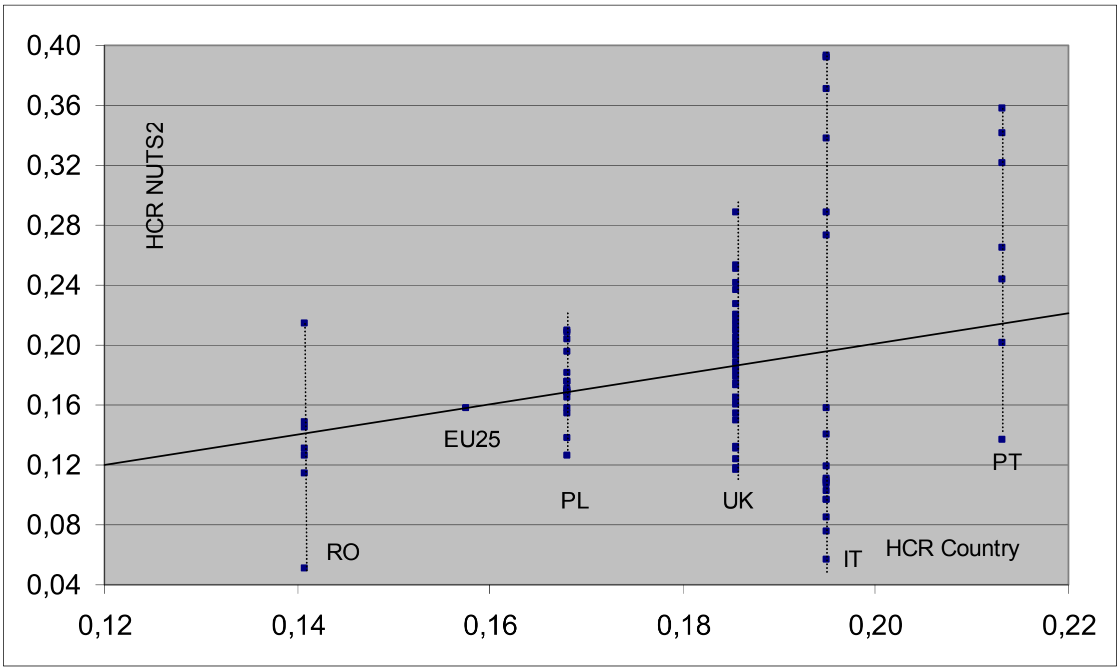
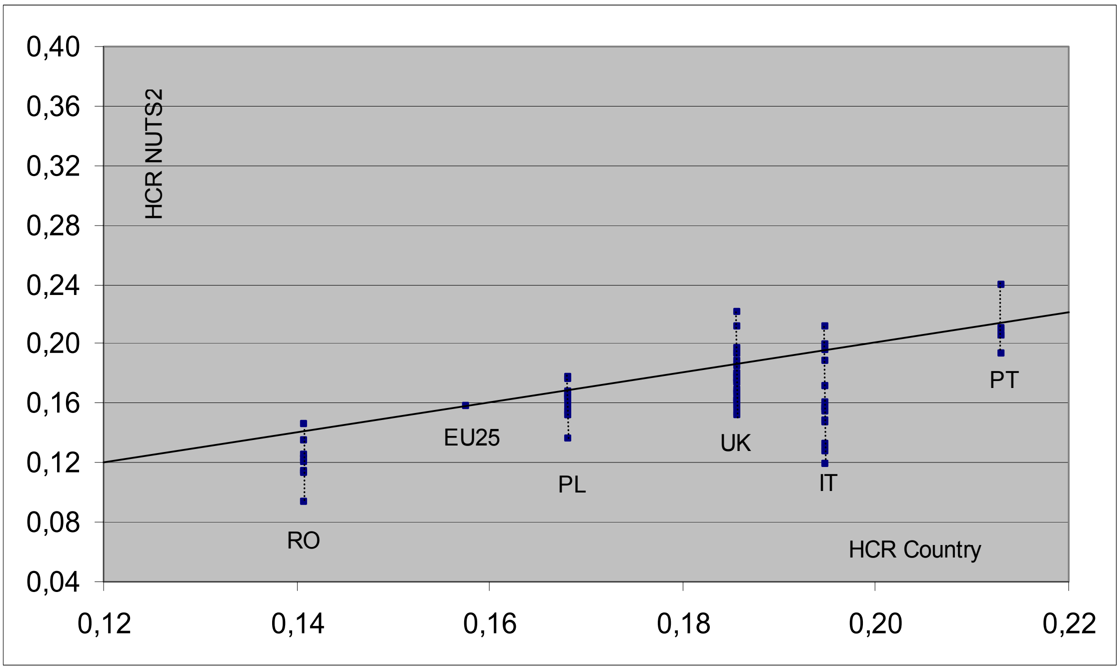
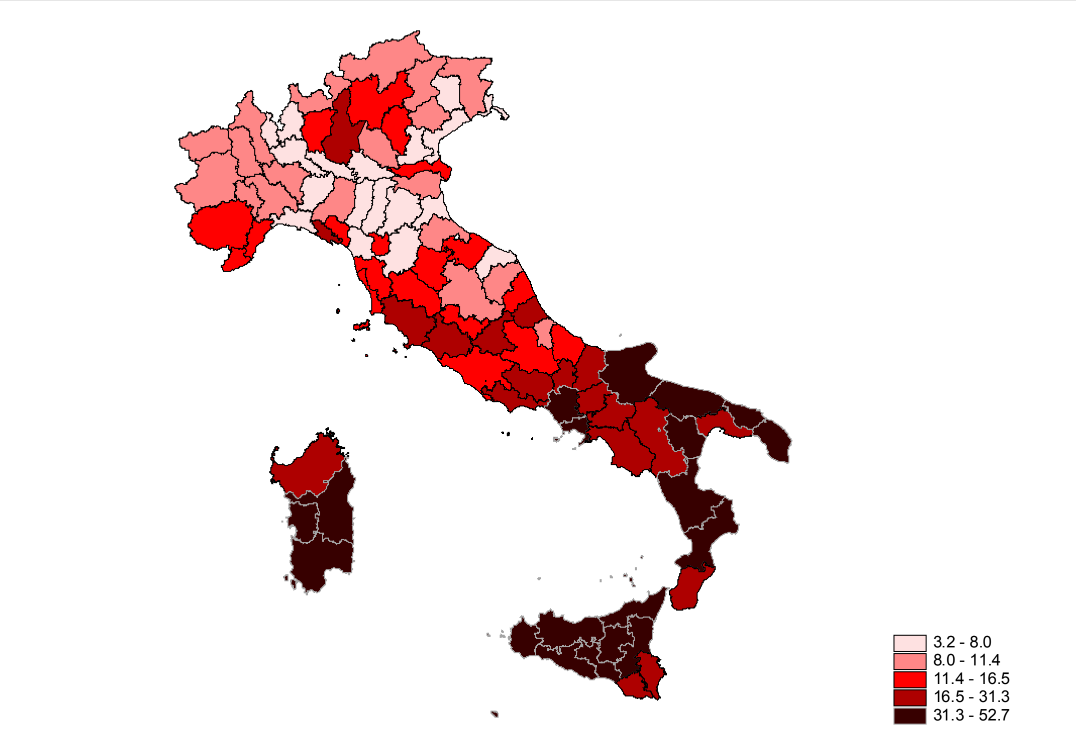


Figure 5.10 Head Count Ratio NUTS2 (using NUTS2 poverty lines) - dispersions within countries



The statistical model we have used is constrained by the limited set of covariates available in NewCronos at NUTS3 level. It is possible in principle to incorporate more specific information on provinces, regions and macro regions from diverse sources in Italy, and hence improve the performance of the model. In practice, all models going down to low area levels such as NUTS3 (or beyond) must be country-specific: it is *not* necessary or useful to seek standardisation at EU level in this regard.

Figure 13.4 Head Count Ratio, NUTS3 regions (country poverty line). Italy



We show results for NUTS3 regions for two poverty line levels: (1) Country poverty line, defined on the basis of pooled income distribution for the whole country; (2) NUTS2 poverty lines, defined on the basis of the income distribution within each NUTS2.

Country poverty line

Using the country poverty line, we note that the highest poverty rates (above 48%) are for Catania, Enna (in Sicilia), Oristano (in Sardegna) and Foggia (in Puglia). The provinces with lowest rates are all in North East, in the same region (Friuli Venezia Giulia) all with a poverty rate below 5%)

There is a clear relationship between the level of income and the poverty rate using country poverty line (Figure 13.7). Lower income levels tend to go with higher poverty rates, and higher income levels with lower poverty rates. The North-South division is prominently obvious in the graph. The Centre (Toscana, Umbria, Marche and Lazio) is more mixed.

NUTS2 poverty lines

While using the country poverty line provides a more or less absolute measure of regional differences in income, the use of NUTS2 poverty lines gives an approximately relative measure. The latter does depend on regional differences in levels of income, but only among NUTS3 regions within the *same* NUTS2. Differences on a larger geographical scale, such as across NUTS2 or higher units, do not matter.

Figure 13.7 Mean equivalised income level versus HCR for NUTS3 regions- country poverty lines. Italy

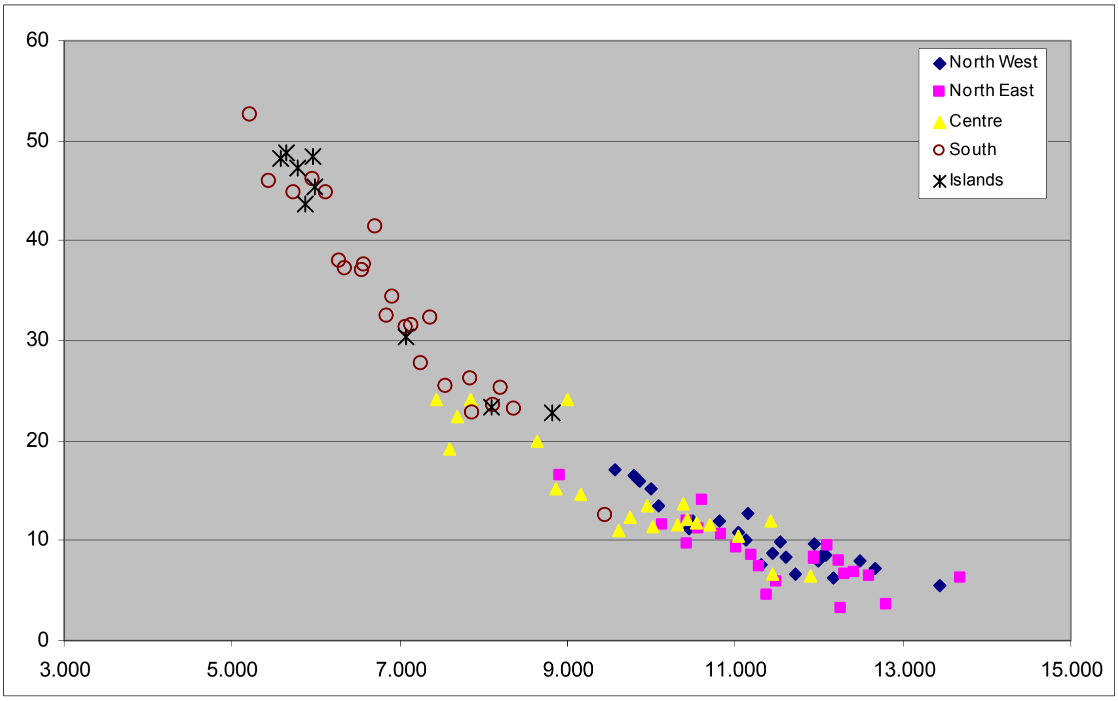
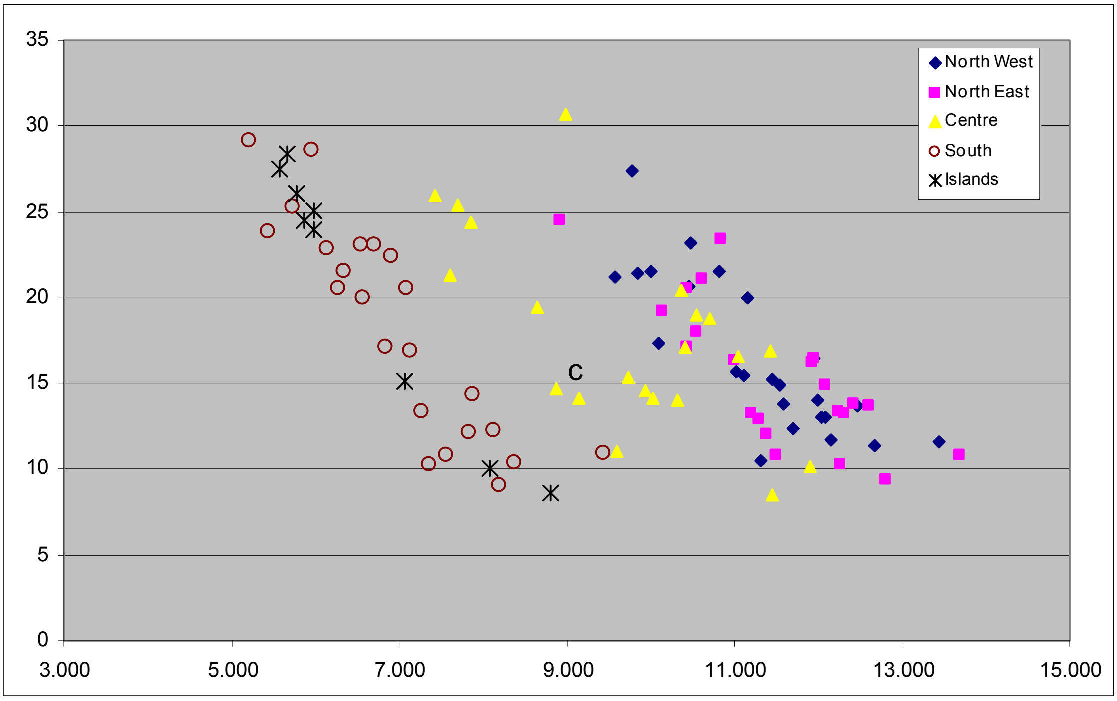


Figure 13.8 Mean equivalised income level versus HCR for NUTS3 regions- NUTS2 poverty line. Italy



The pattern in Figure 13.8 is remarkable. Though it displays an almost purely *relative measure* of income distribution, the North-South divide is clearly seen. A negative HCR-Income Level relationship can be seen within each micro region separately, the series of parallel regression lines moving to the right (towards higher income levels) from South to North.

Chapter 1
Introduction

As provided in the Project Terms of Reference (ToR), the aim of this study is to identify the appropriate methodologies and strategies for the development of indicators of poverty and social exclusion at regional level, the ultimate goal being the development of a coherent and integrated strategy for the incorporation of the regional dimension into the Member States' NAP/incl. This study takes as its point of departure the methodological framework used for defining the indicators of poverty and social exclusion endorsed at Laeken, and more generally, existing methodological research and data in the area of indicators of poverty and social exclusion as well as in the area of regional indicators.

We have aimed to address a number of major issues or challenges:

(1) Identifying special features and requirements of the system of indicators of poverty and social exclusion for use at the regional (subnational) level.

This means to identify whether and how the regional indicators differ from the indicators used primarily at the national level. One of the research tasks identified in the Terms of Reference is to "assess the extent to which indicators of social exclusion and poverty of the type endorsed at Laeken can be applied at the regional level, either using EU-wide or national sources". Some of the primary and secondary Laeken indicators may be suitable for regional application; others may be suitable after modification; while some may not be appropriate for the purpose. In addition, it is also necessary to consider whether there is need for *addition* to the existing "national-level portfolio" so as to capture essential regional aspects. We address these issues throughout this report, but in specific terms in Chapter 2.

(2) Choosing appropriate units of analysis.

It is also important to clarify what is meant by the "regional" level. We address the ToR objective to "analyse the extent to which regional deprivation should be defined as self-contained concept, different from individual deprivation", specifically in Chapter 3.

(3) Describing a practical strategy for measuring poverty and social exclusion at the regional level.

It is noted in the Terms of Reference that the study "need not be constrained by considerations of data availability, but should also offer a strategy for measuring poverty and social exclusion at the regional level". We have taken 'need not be constrained by considerations of data availability' to mean that the strategy for measuring poverty and social exclusion at the regional level should not be defined only in terms of the data which are currently available and accessible. This is so because the situation in this respect may change and, even more importantly, because the identification of better measures may itself provide an impetus for improving the data situation. On the other hand, we have felt that *the choice of particular types of units and indicators must be realistic and potentially realisable*. For instance, there are certain types of data which, by their very nature and complexity, can only be collected on a relatively small scale and/or only infrequently. Such unavoidable constraints must of course be respected in the choice of regional indicators. This realism is essential in order to meet the second clause of the above statement from the Terms of

Reference, namely that the study should "offer a *strategy* for measuring poverty and social exclusion at the regional level". We address these issues throughout this report, but in more specific terms in Chapters 4, 10 and 11.

(4) Illustrating the recommended strategy concretely, with necessary technical detail on the basis of real statistical data.

We have felt that, for this work to be *useful* to the Commission, it is absolutely necessary to go beyond simply 'offering' a strategy. We have tried to develop and illustrate technical details of the offered strategy, and to provide indicative results using real data, as available to us. We hope that this will facilitate others in replicating and improving the approaches we have recommended. This aspect of our work, which has been by far the most intensive and time consuming, is essential for meeting the ToR objective to "identify data gaps and propose a strategy for the development of a system of relevant indicators of social exclusion and poverty at the regional level". Of course, it must be stressed that the numerical results we present throughout this report are more illustrative than 'final' or 'best' estimates. Others with access to more data and more national knowledge can undoubtedly produce more precise results.

(5) Constructing income poverty-related indicators appropriate for the regional level.

We have proposed and constructed income poverty rates and related measures which are more appropriate in the regional context in view of sample size constraints. Both cross-sectional poverty rates and longitudinal measures of persistent poverty have been developed, as presented in Chapters 5, 7 and 8.

(6) Incorporating with increased emphasis non-monetary dimensions of deprivation to complement indicators of income poverty.

This is a part of the response to ToR objective to "explore other approaches to define indicators at regional level which could be more relevant when taking regions as the observation unit, in particular identify the non-monetary dimensions of deprivation and social exclusion that determine regional deprivation (e.g., issues of access to essential services, education, transport, etc.)". We have aimed to extend the monetary poverty indicators hitherto considered (and emphasised, for instance, in the Laeken set) to *non-monetary dimensions of social exclusion and poverty* – in particular to capture physical aspects of living conditions, including the capacity to meet certain basic needs, the possession of various amenities and facilities, and housing and environmental conditions. The data required for many of the non-monetary indicators are in fact more easily and widely collected than monetary data on individuals' income, but the methodological difficulty is the multitude of dimensions involved and the fact that the related information is qualitative (yes-no dichotomies, categories etc), rather than quantitative which is more readily treated in our statistical techniques. In Chapter 6 we provide details and illustrations of a methodology for the incorporation of diverse non-monetary indices in the construction of multi-dimensional indicators of deprivation.

(7) Extending indicators normally produced at the national level to the level of regions – going down to a level such as NUTS2 or even beyond.

A major constraint in such extension is smallness of the sample sizes for the required data, but this is not the only issue. In fact, at least two types of extensions needed to be made in the statistical methodology used. Firstly, we have tried to identify ways to use the available survey data in a more efficient way, with the objective of ameliorating the problems resulting from limited sample sizes when disaggregation to the regional level is required. Secondly, it has been necessary to introduce 'small-area estimation' methods. The basic idea of these procedures is to combine data set of large size but simple content with small-sized but more detailed data sets, so as to draw on the strength of both. A variety of small-area estimation (SAE) methods exist, and the appropriate choice of the method depends on the particular situation – in particular on the type of data available for the purpose. Taking as basis the micro-data actually available to us for the purpose of this research - namely from ECHP for EU15 and similar surveys from Poland and Romania - and using these in conjunction with Eurostat Free Dissemination Database (NewCronos), we have produced *composite estimates* for a number of income and non-monetary indicators using SAE procedures. And we have used parameters estimated from these procedures to produce estimates, albeit rather crude ones, for other countries and regions where no survey data are available to us.

In brief, the recommended strategy for the construction of regional indicators of poverty and deprivation has three fundamental aspects: (a) making the best use of available sample survey data, such as by cumulating and consolidating the data to construct more robust measures which can permit a greater degree of spatial disaggregation; (b) exploiting to the maximum 'meso' data (such as the highly disaggregated tabulations available in NewCronos) for the purpose of constructing indicators for small areas; and (c) using the two sources in combination to produce the best and most complete possible estimates for subnational regions using appropriate small area estimation (SAE) techniques. We consider (a) in Chapters 4-9, (b) in Chapter 10, and (c) in Chapters 11-15.

The scope and terms of reference of this project are quite broad. It is possible that we – like any other team of researchers – have paid, somewhat selectively, more attention to areas in which we feel to have a comparative advantage. In the case of our team, these are statistical methodology and data analysis. We hope, nevertheless, that this research makes a useful contribution to the development by European Commission of regional indicators to reflect social exclusion and poverty.

Chapter 2
Choice of appropriate regional indicators

2.1 Introduction

We begin by the task of identifying special features and requirements of the system of indicators of poverty and social exclusion appropriate for use at the regional level. Specifically, the requirement is to identify whether, and if so in what manner, indicators appropriate for the regional level may differ from the indicators designed primarily for the national level.

The Project Terms of Reference (ToR) specify that the present study will "draw on existing methodological research and data in the area of indicators of poverty and social exclusion as well as the area of regional indicators"; that it will use, as point of departure, "the methodological framework used for the definition of indicators of poverty and social exclusion endorsed at Laeken". One of the research tasks identified in the ToR is to "assess the extent to which indicators of social exclusion and poverty of the type endorsed at Laeken can be applied at the regional level, either using EU-wide or national sources".

Indicators of poverty, deprivation and social exclusion of course have an important territorial dimension, pointing to the need to take account of regional and local differences. In an ideal context, one may seek to give regional breakdown on *all* indicators. That is, one may introduce regional analysis within each of the indicator fields, for instance producing poverty rates by NUTS regions, urban-rural classification, etc. *However, simply the introduction of more extensive breakdown is neither possible because of data limitations, nor sufficient in itself.*

Some of the primary and secondary Laeken indicators may be suitable for regional application; others may be suitable after modification; while some may not be appropriate for the purpose. In addition, it is also necessary to consider whether there is need for *addition* to the existing indicators developed primarily for application at national level - region-specific indicators able to capture aspects which are essentially regional. It is possible that a more diverse "portfolio of indicators" is required for the purpose of addressing concerns of regional policy and research.

2.2 Background

In order to provide some necessary background, we begin by paraphrasing the following pertinent points from the Joint Report on Social Inclusion (European Commission, 2003; Section 10, Use of Indicators), and also the Report on Social Inclusion 2004 covering New Member States. They primarily relate to the construction of indicators at the national level, occasionally also concerning some subpopulations, such as children and minorities. Nevertheless, they are equally pertinent to the development of appropriate indicators at the regional level, and provide the necessary methodological framework and a starting point. Subsequently, with specific reference to individual Laeken indicators, we will discuss ways in which the introduction of the regional dimension may make some fundamental differences in the choice of a "portfolio" of indicators.

Laeken indicators

Indicators are an essential tool in the Open Method of Co-ordination as they help monitor progress towards the common objectives and measure the challenges ahead. The importance of indicators was stressed at Lisbon; the Nice European Council invited Member States and the Commission to develop commonly agreed indicators, and this recommendation was reinforced by the Stockholm European Council in March 2001. The task of developing this set of indicators was undertaken by the Social Protection Committee, and more specifically its Indicators Sub-Group. The Laeken European Council endorsed a first set of 18 indicators of social exclusion and poverty, organised in a two-level structure of primary indicators – consisting of 10 lead indicators covering the broad fields considered to be the most important elements in leading to social exclusion – and 8 secondary indicators – intended to support the lead indicators and describe other dimensions of the problem.

As to methodological principles to guide the selection of indicators, the SPC stressed first that the portfolio of EU indicators should be *balanced across different dimensions* and that common indicators *should address social outcomes rather than the means by which they are achieved*. An indicator should be responsive to policy interventions, and should have a clear and accepted normative interpretation. Also, any indicator should be robust and statistically validated, should be measurable in a sufficiently comparable way across Member States, and should be timely and susceptible to revision.

On the basis of the above methodological principles, the Indicators Sub-Group has continued to refine and consolidate the original list of "Laeken indicators". It highlighted the need to give *children a special focus when analysing the common indicators* and, to this purpose, to have a standard breakdown by age of all the Laeken indicators, whenever relevant and meaningful (and conditional upon statistical reliability); it redefined the indicator of population living in jobless households and added a new indicator of in-work poverty.

The use of the common and national indicators in the NAPs

Common indicators. All NAPs make use of the common indicators. Many Member States draw an extensive analysis of the situation of poverty and social exclusion on the basis of both the common indicators and national indicators supporting them or highlighting aspects relevant to their national situation. The Joint Report on Social Inclusion (2003) provides comment of the specific situation in a number EU15 countries.

National indicators. Several "third-level" indicators have been used in the NAPs. Alongside the definitions of the common indicators, some Member States have used different definitions and/or alternative data sources for measuring and characterising current levels of poverty and social exclusion. For example, Greece and Italy define relative poverty risk on the basis of both income (Laeken definition) and consumption. Some countries refer also to national indicators of absolute poverty (Italy), non-monetary indicators of living conditions (Belgium, Italy, France), or, still, measures of self-perceived poverty or deprivation (Belgium, Italy); Ireland uses the measure of consistent poverty, a combination of relative income and deprivation measures. Such indicators clearly provide useful complementary information to that of relative poverty risk.

Regional indicators. The sub-national dimension of poverty and social exclusion is in some instances (Belgium, Greece, France, Italy) described through a regional breakdown of the common indicators. In particular, Greece interestingly distinguishes between rural and urban regions, highlighting the different nature of poverty and social exclusion in these two areas.

Policy-related indicators. Most Member States used policy-related indicators, which can be more easily integrated within the development of a policy strategy. Examples of these indicators are the number of unemployed or long-term unemployed persons who are assisted by some labour market policy measure, the number of available social housing units and the amount of minimum income benefits. In fact, the distinction between input-related and performance indicators is not always straightforward and some indicators are better qualified as "intermediate output" indicators. Such indicators express on the one hand the policy effort in favour of those at risk of poverty and on the other hand the impact of social policies as well as of the economic context. Benefit dependency indicators – quite largely used in the NAPs – are an example of this type of indicators. Even the Laeken indicator of "early school leavers" can be seen as belonging more to the category of intermediate output indicators than a performance indicator in the strict sense.

Indications for future developments at EU15 level

Looking at the current list of common indicators as a whole, the concept of social exclusion that emerges seems to be related to lack of income, income inequality, lack of employment and lack of an adequate educational attainment level. It is unquestionable that these are some of the key dimensions of social exclusion and poverty, but other important areas – such as health, living conditions and housing - are not yet adequately covered and further efforts need to be devoted to them. Furthermore, it would be important to develop a better understanding of poverty and social exclusion at the subnational level. In all these domains, however, a combination of factors – data as well as institutional differences across the EU – still make it difficult to define common indicators that can be used across all 15 Member States.

Report on Social Inclusion 2004

… The new Member States of the EU were to make use for the first time of the commonly agreed indicators in their NAPs; they were also invited to use third-level indicators defined at the national level to highlight specificities in particular areas not adequately covered by the common indicators (particularly housing), and to help interpret them. … [T]he common indicators are used in order to identify the most vulnerable groups and the extent to which they are vulnerable to poverty and social exclusion. Furthermore, much attention has been devoted to the examination of indicators of exclusion from the labour market. Some countries (Estonia, Hungary, Lithuania, Poland and Slovenia) even set overall quantified targets based on some of the common indicators (relative poverty risk/income distribution; long-term unemployment; life expectancy).

In general, countries have gone beyond the examination of the common indicators by using tertiary indicators … A wealth of quantitative information on the economic, demographic and labour market situation allows understanding the particular context of poverty and social exclusion in the new Member States. Indicators of material deprivation, absolute poverty or living conditions, in particular highlighting housing problems, are prominent in some of the NAPs. Also, additional breakdowns of the common indicators of poverty and social exclusion are used, notably by ethnic/linguistic group or immigration status (for example, Roma are singled out in the Hungarian NAP). The sub-national distribution of poverty and social exclusion is in some instances described through the territorial breakdown of various indicators, both common and tertiary ones. Finally, most Member States used policy-related indicators, which can be more easily integrated within the development of a policy strategy … Due to the missing longitudinal dimension in the underlying data sources, persistent risk-of poverty rates could not be calculated for any new Member State and Candidate Country.

[Just as noted in the 2003 report], the dimensions of social exclusion and poverty that emerge as more clearly depicted are insufficient income, lack of employment and inadequate skills. It is unquestionable that these are key dimensions, but other important areas – such as health, living conditions and housing - are not yet adequately covered and further efforts need to be devoted to exploring them.

2.3 Regional indicators of income poverty: cross-sectional measures

For the development of monetary indicators of poverty for use at the regional level, the starting point is the specified set of Laeken indicators in this area. These indicators are summarised at the end of this section with some necessary technical detail.

Adaptation to the regional level

Henceforth these indicators have been applied at the national level. It is necessary to adapt them for regional application, taking into account any differences in the requirements, but equally important, differences in the practical situation. As in the case of regional adaptation of all other indicators, it is necessary to *focus on the more basic among this set of indicators*. This is because of the substantially increased data requirements when the results have to be geographically disaggregated.

Detailed disaggregation of the indicators by age, gender and other characteristics - simultaneously with disaggregation by geographical region – has to be severely restricted where the information comes from sample surveys of limited size, as is the case in most Member States lacking income registers. Broad classification, such as distinguishing children, elderly persons, and youth entering the labour market may be possible, but even that has to be subsidiary to the need for adequate regional breakdown.

Certain more complex poverty and inequality measures - measures which are more sensitive to details and irregularities of the empirical income distribution - are less suited for disaggregation to small populations and small samples. Indicators such as Gini coefficient and even decile ratios (S80/S20) may be too demanding at say NUTS2 level.

The above considerations apply, though to a lesser extent, to Laeken indicators such as "Indicator 4: Relative median at-risk-of-poverty gap, by age and gender" and "Indicator 13: At-risk-of-poverty rate before social transfers, by age and gender".

The level of income poverty is determined by the chosen poverty line. By choosing different poverty lines, different numerical values are obtained, and to some extent each such figure provides additional information. It is for this reason that Laeken list includes Indicator 11 "Dispersion around the at-risk-of-poverty threshold", meaning poverty rates defined using 40, 50, 60 and 70% of the national mean as poverty line. Incorporation of the effect of choosing different poverty lines is also important when we move down to the regional level. However, again in view of small sample sizes, it is desirable to avoid producing too many separate figures, any real differences between which may be overwhelmed by sampling variability and other errors in the data. Rather, it is more useful to consolidate such separate figures into a single (or at most a very small set of) more robust measure(s) if possible. The ideas explained later (see Chapter 4 in particular) of "consolidating" the measures - such as in the form of suitably weighted averaging over different numerical measures computed with different thresholds and levels of poverty lines - could be applied. In specific terms, a single measure based on suitable consolidation over, say, 50%, 60% and 70% of median poverty lines, would be preferable to separate indicators such as Laeken Indicators 1 and 11.

As a consequence, for the purpose of regional indicators, the focus has to be primarily on ordinary *poverty rates* for the total population, and possibly for some special groups such as children, the elderly, and youth entering the labour market.

On the other hand, this indicator has to be supplemented by other indicators not considered explicitly in the Laeken list. Perhaps the most important of these is simply the *mean income levels* of the regions, the dispersion among which provides a measure of regional disparities. *General entropy measures* such as GE(0) and GE(1) may also be useful because they can be decomposed into within and between region components.

Laeken indicators concerning cross-sectional measures of income poverty[1]

Among the Laeken indicators, the following cross-sectional measures of poverty have been included:

- Indicator 1a: At-risk-of-poverty rate, by age and gender
- Indicator 1b: At-risk-of-poverty rate, by most frequent activity status and gender
- Indicator 1c: At-risk-of-poverty rate, by household type
- Indicator 1d: At-risk-of-poverty rate, by accommodation tenure status
- Indicator 1e: At-risk-of-poverty threshold (illustrative values)
- Indicator 2: Inequality of income distribution S80/S20 income quintile share ratio
- Indicator 4: Relative median at-risk-of-poverty gap, by age and gender
- Indicator 11: Dispersion around the at-risk-of-poverty threshold
- Indicator 13: At-risk-of-poverty rate before social transfers, by age and gender
- Indicator 14: Inequality of income distribution Gini coefficient

New indicator included in Social Inclusion Report 2004 are as follows. There have also been some refinements to the definitions of indicators.

- Poverty risk by work intensity of households
- In-work poverty risk

Between 1994 and 2001 these indicators were calculated for the majority of Member States on the basis of the ECHP. Henceforth EU-SILC is to become the EU reference source for comparative statistics on income distribution and social exclusion at European level. Indicators 1a-1e, 2 and 4 are among the "primary" Laeken indicators, while Indicators 11, 13 and 14 are among the "secondary" indicators. A subset of these indicators is used in the Statistical Annex to the annual Commission report to the Spring European Council, and reported in the Structural Indicators database. Among the above, it includes Indicator 2, and a subset of Indicators 1a and 13 (giving breakdown only by gender). A similar subset has been proposed for use in monitoring the EU Sustainable Development Strategy. Indicator 1a is included in the "shortlist" in the Statistical Annex.

[1] This subsection is based on Eurostat. Domain: ILC - Income and Living Conditions: The methodology of calculation of certain Laeken indicators of social inclusion and other common cross-sectional indicators derived from EU-SILC.

The Indicators Sub-Group of the Social Protection Committee has continued its work to refine and extend them, and has also developed related statistics under the open method of coordination. Pending formal adoption by the Council, many of these additional statistics have been used in high profile publications such as the Joint Pensions Report 2002, the Joint Inclusion Report 2003, the Social Situation in the EU report 2003, and Social Inclusion report 2004.

This has included the following cross-sectional income indicators:

- Indicator 19: At-risk-of-poverty rate among workers by gender (derived from Indicator 1b)

  Pension indicators:

- P1.01 At-risk-of-poverty rate, by household type (same as Indicator 1c)
- P1.02 At-risk-of-poverty rate, by age and accommodation tenure status (Indicator 1d by age)
- P2.01 Relative median income ratio, by age
- P2.02 Structure of income, by age
- P3.01 Inequality of income distribution S80/S20 income quintile share ratio, by age (Indicator 2 by age)
- P3.04 Dispersion around the at-risk-of-poverty threshold, by age (Indicator 11 by age)

At-risk-of-poverty rate

The *total disposable income* of a household is calculated by adding together the personal income received by all of household members plus income received at household level, after editing, imputation and weighting of the survey data as required. For each person, the *equivalised disposable income* is defined as his/her total household disposable income divided by equivalised household size. The *equivalised household size* is defined according to the 'modified-OECD scale', which gives a weight of 1.0 to the first adult, 0.5 to other household members aged 14 or over and 0.3 to household members aged under 14. Each person in the same household receives the same equivalised disposable income. *At-risk-of-poverty rate (after social transfers)* is defined as the percentage of persons, over the total population, with an equivalised disposable income below the 'at-risk-of-poverty threshold'. The *at-risk-of poverty threshold* is set at 60% of the national median equivalised disposable income. The EU average of the indicator is calculated as a weighted average of the indicators established for each country, where the weighting of countries is done according to the number of persons living in private households in each country.

At-risk-of poverty rate (after social transfers) broken down according to certain characteristics

The 'at-risk-of poverty rate (after social transfers)' broken down by population group defined in terms of a certain characteristic is calculated as the percentage of persons in the population group (over the total population in the same group) with an equivalised disposable income below the 'at-risk-of-poverty threshold'. Apparently, the 'at-risk-of-poverty threshold' is defined with reference to the income distribution of the total population (and not just that of the particular population group being considered).

At-risk-of-poverty threshold (illustrative values)

This is simply the 'at-risk-of-poverty threshold' for the total population, multiplied by equivalised household size of the household type being considered. Values are expressed in PPS, Euro and national currency.

Inequality of income distribution S80/S20 income quintile share ratio

S80/S20 income quintile share ratio: Ratio of the sum of equivalised disposable income received by the 20% of the country's population with the highest equivalised disposable income (top interquintile interval) to that received by the 20% of the country's population with the lowest equivalised disposable income (lowest interquintile interval)

Relative median at-risk-of-poverty gap by age and gender

The difference between the median equivalised disposable income of persons below the at-risk-of-poverty threshold and the at-risk-of-poverty threshold itself, expressed as a percentage of the at-risk-of-poverty threshold. Gender, age breakdown and total.

Dispersion around the at-risk-of-poverty threshold

The percentage of persons, over the total population, with an equivalised disposable income below 40%, 50% and 70% of the national median equivalised disposable income.

At-risk-of-poverty rate before social transfers by age and gender

At-risk-of-poverty rate before social transfers except old-age and survivors' benefits

The 'at-risk-of-poverty rate before social transfers except old-age and survivors' benefits' shows the percentage of persons (over the total population) having an equivalised disposable income before social transfers except old-age and survivors' benefits below the national at-risk-of-poverty threshold'.

At-risk-of-poverty rate before social cash transfers including old-age and survivors' benefits

The 'at-risk-of-poverty rate before social transfers including old-age and survivors' benefits' shows the percentage of persons (over the total population) having an equivalised disposable income before social transfers including old-age and survivors' benefits below the national 'at-risk-of-poverty threshold'.

Inequality of income distribution: Gini coefficient

The Gini coefficient is defined as the relationship of cumulative shares of the population arranged according to the level of equivalised disposable income, to the cumulative share of the equivalised total disposable income received by them.

Poverty risk by work intensity of households

Poverty risk for the total population in different work intensity categories and broad household types. The work intensity of households refers to the number of months that all working-age household members have been working during the income reference year, as a proportion of the total number of months that could theoretically be worked within the household.

In-work poverty risk

Individuals who are classified as employed (distinguishing between wage and salary employment and self-employment) according to the situation of most frequent activity status, and who are at risk of poverty. This indicator is to be analysed according to personal, job and household characteristics.

2.4 Longitudinal indicators of poverty and deprivation

For the development of longitudinal indicators of poverty and deprivation for use at the regional level, the starting point of course is again the specified set of Laeken indicators in this area. These indicators are summarised in the Inset "Laeken indicators concerning longitudinal measures of income poverty" below. The indicators concerned are:

- Indicator 3: At-persistent-risk-of-poverty rate, by age and gender (60% median)
- Indicator 12: At-risk-of-poverty rate anchored at a moment in time[2]
- Indicator 15: At-persistent-risk-of-poverty rate, by age and gender (50% median)

Simplifications

Henceforth these have been applied at the national level. It is necessary to adapt them for regional application, taking into account any differences in the requirements, but equally important, differences in the practical situation. As in the case of regional adaptation of all other indicators, it is necessary to focus on the more basic of the set. This is because of the substantially increased data requirements when the results have to be geographically disaggregated.

Firstly, the disaggregation of the Indicators 3 and 15 by age and gender - simultaneously with disaggregation by geographical region – has to be severely restricted where they information comes from sample surveys of limited size, as is the case in most Member States lacking income registers. Broad classification, such as distinguishing children and old persons, may be possible, but even that has to be subsidiary to the need for adequate regional breakdown.

For the same reason, emphasis has to be shifted away from the study of trends over time and longitudinal measures to essentially cross-sectional measures of the type discussed in preceding section. Furthermore, it is more appropriate to aggregate such measures over suitable time periods, so as to illuminate the more stable aspects of the patterns of variation across regions.

Thirdly, the ideas explained later (see Chapter 4 in particular) of "consolidating" the measures - such as in the form of suitably weighted averaging over different numerical measures computed with different thresholds and levels of poverty lines - should be applied to obtain fewer but more stable measures. The results of individual computations (using a single specific poverty line for instance) may be too sensitive to irregularities in the

[2] We have listed this indicator under 'longitudinal indicators', though it may also be argued that this should be considered merely a cross-sectional indicator. This is because strictly it does not require longitudinal data, i.e., data linked over time for the same individual. However, the indicator does require comparable data at more than one points in time, and in practice such data often happen to be longitudinal.

empirical income distribution based on small regional samples. In specific terms, a single measure based on suitable consolidation over, say, 50%, 60% and 70% of median poverty lines, would be preferable to separate indicators such as Laeken Indicators 3 and 15.

Laeken indicators concerning longitudinal measures of income poverty*

Among the Laeken indicators, the following *longitudinal measures* of poverty have been included:

Indicator 3: At-persistent-risk-of-poverty rate, by age and gender (60% median)

Indicator 12: At-risk-of-poverty rate anchored at a moment in time

Indicator 15: At-persistent-risk-of-poverty rate, by age and gender (50% median)

Indicator 3 is among the "primary" Laeken indicators, while Indicators 12 and 15 are among the "secondary" indicators. A subset of these indicators is used in the Statistical Annex to the annual Commission report to the Spring European Council, and reported in the Structural Indicators database. Among the above, it only includes a subset of Indicator 15, namely:

Indicator 15-subset: At-persistent-risk-of-poverty rate, by gender.

Some of the Laeken indicators are included in the "shortlist" in the Statistical Annex, and a similar subset has been proposed for use in monitoring the EU Sustainable Development Strategy.

The Indicators Sub-Group of the Social Protection Committee has continued its work to refine and extend them, and has also developed related statistics under the open method of coordination. Pending formal adoption by the Council, many of these additional statistics have been used in high profile publications such as the Joint Pensions Report 2002, the Joint Inclusion Report 2003, the Social Situation in the EU report 2003 and others. No addition has been made to the list of longitudinal indicators identified above. The same is true of the additional Pension indicators and to some indicators specified as "Other" developed so far.

Between 1994 and 2001 the Laeken indicators 12 and 15-subset were calculated for the majority of Member States on the basis of the ECHP. Henceforth EU-SILC is to become the EU reference source for comparative statistics on income and living conditions at EU level. This is noted to apply particularly in the context of the Open Method of Co-ordination on Social Exclusion, and for producing structural indicators on social cohesion for the annual spring report to the European Council.

Longitudinal indicators are of course less frequently used in social inclusion and other reports than cross-sectional indicators of poverty and exclusion. Due to the missing longitudinal dimension in the underlying data sources, persistent risk-of poverty rates could not be calculated for any new Member State and Candidate Countries (European Commission, 2004).

Indicator 3 – "At-persistent-risk-of-poverty rate by age and gender (60% median) - is defined as the share of persons with an equivalised disposable income below the at-risk-of-poverty threshold in the current year at least two of the preceding three years. On the basis of EU-SILC, it will be calculated with longitudinal component after 4 years of the panel survey.

Indictor 12 – "At-risk-of-poverty rate anchored at a moment in time" – is defined as follows.

For a given year t, it is defined as the percentage of the population whose equivalised total disposable income in that given year is below a risk-of-poverty threshold calculated in the standard way for the earlier year (t-3) and then up-rated for inflation. Since in general the income reference period is the year preceding the survey year t, the appropriate inflation rate to be applied will be, in general, that for the period (t-4) to (t-1).

Indicator 15 – "At-persistent-risk-of-poverty rate by age and gender (50% median) – is merely a variant of primary indicator 3 using a lower poverty threshold.

*Eurostat. Domain: ILC - Income and Living Conditions: The methodology of calculation of certain Laeken indicators of social inclusion and other common cross-sectional indicators derived from EU-SILC

In view of the above considerations, it would be appropriate to consider somewhat simplified longitudinal indicators in the regional context. One can expect that simpler indicators will be more robust and less demanding on the data available. The main simplification we propose is to focus on *longitudinal indicators defined over a shorter time periods*. Where the available statistical data cover a longer time period, those longitudinal indicators can themselves be averaged over time to obtain more robust measures.

In specific terms, we define and construct in the following illustrations indicators based on the persistence of poverty over *pairs of adjacent years*:

- Persons are *persistently poor* over two consecutive years if, in relation to the poverty line specific to each of the years, they are classified as poor in *both* the years.
- Persons are in *any-time poverty* over two consecutive years if, in relation to the poverty line specific to each of the years, they are classified as poor in *either* of the years.

With a longer reference period of T years, assuming that the necessary time series of data are available, the (T-1) pair-wise persistent or any-time rates can be averaged over time to obtain more stable measures for regional comparisons. The choice of the appropriate reference period T for averaging depends, apart from data availability, on substantive and policy considerations. It is matter of trade-off between temporal and spatial detail. Perhaps a moving average over a 4 or 5-year period may be considered generally appropriate. In our illustrations (see Chapter 7), we have taken T as 8 years where the ECHP data were available, and naturally shorter when the data were lacking.

Note that with the above indicators, defined with reference to only a two-year period, the type of distinction implied between Laeken Indicators 3 and 12 is not likely to be important or useful. Hence only one or the other definition of the "at-risk-of poverty threshold" should suffice (preferably the more conventional one used in Indicator 3).

Additions

One of the important objectives of this work is to incorporate, with increased emphasis, non-monetary dimensions of deprivation to complement indicators of income poverty. We will comment further on this in Sections 2.6 and 2.7 below. Specifically in the context of longitudinal indicators, the following important addition should be considered:

> The measures of income poverty described above can all be generalised to multi-dimensional measures of deprivation of the type discussed in Chapter 6 (any-time, persistent or continuous incidence of supplementary, latent and manifest forms of deprivation).

Some illustrations are given in Chapter 7 (Section 7.3).

Finally, it may also be appropriate where possible to supplement the above 'basic' longitudinal measures with additional indicators – but the latter considered as 'secondary' in relation to the former. A good candidate is the following indicator:

An indicator similar to Laeken Indicator 3 can be useful to identify longer-term poverty. However, for reasons of data constraints for regional estimation noted at the beginning of this section, it would be preferable not to tie the measure to a particular year.

This will be discussed further in Chapter 7 (Section 7.3), where we introduce the concepts of any-time, persistent and continuous poverty with a longer reference period.

2.5 Other Laeken indicators

Apart from "financial poverty" indicators discussed in preceding sections, Laeken indicators cover three additional dimensions of social inclusion: employment, health and education. These highlight the "multidimensionality" of the phenomenon of social exclusion.

The additional indicators are:

- Indicator 5: Regional cohesion (dispersion of regional employment rates)
- Indicator 6: Long term unemployment rate, by gender
- Indicator 7: Persons living in jobless households, by age and gender
  - children (0-17)
  - prime-age adults (18-59)
- Indicator 8: Early school leavers not in education or training, by gender
- Indicator 9: Life expectancy at birth, by gender
- Indicator 10: Self defined health status by income quintile
- Indicator 16: Long term unemployment share, by gender
- Indicator 17: Very long-term unemployment rate, by gender
- Indicator 18: Persons with low educational attainment, by age and gender

New indicator included in Social Inclusion Report 2004 are:

- Low reading literacy performance of pupils

A subset of these indicators are Structural Indicators; some of these are in the short list of structural indicators , and are used in the Statistical Annex to the annual Commission report to the Spring European Council, and reported in the Structural Indicators database. These are:

- Indicator 5 by gender, and Indicators 6-8.

The first of the above is included in the "shortlist" in the Statistical Annex.

A similar subset has been proposed for use in monitoring the EU Sustainable Development Strategy.

Indicator 10: Self defined health status by income quintile

The indicator 10 'self defined health status by income level' was tentatively adopted in Laeken, calculated as the ratio of the proportions in the bottom and top income quintile groups of the population aged 16 and over who classify themselves as in a bad or very bad state of health, the data source being ECHP.

However, it is noted that "Eurostat is still undertaking research into the feasibility and suitability of this indicator, in collaboration with the Indicators Sub-Group of the Social Protection Committee. In the continuing absence of an agreed methodology, this indicator is currently being produced as the proportions of population aged 16 and over who classify themselves as in a bad or very bad state of health in each quintile, instead of as a ratio, for the age group 16-64 and 65+."

In view of the above, the time is not yet ripe for such an indicator to be generally recommended for inclusion as a regional indicator. Perhaps other, more objective health-related indicators can be explored. We think that some indicators of the type identified as "intermediate output indicators" in the Joint Report quoted from in Section 2.2 above should be considered. As noted, such indicators express on the one hand the policy effort in favour of those at risk of poverty and on the other hand the impact of social policies as well as of the economic context. Just as in the case of income variables, simply comparisons of averages across regions, whether within country or across EU, can in themselves be regarded as "deprivation indicators" in the sense they indicate regional disparities.

Indicator 9: Life expectancy at birth, by gender

This indicator is likely to be more complex to construct at the regional level, as it would involve the need for regional life-tables. Also, except perhaps in the largest countries, it may not be considered among high priority indicators.

Statistical work carried out in connection with this project confirms that a different but related indicator, namely infant morality rate (IMR), is often a remarkably good predictor of normal deprivation indicators. This measure is also more easily estimated at any level of aggregation, normally from administrative sources directly. It may well be more suited for regional comparisons than life expectancy at birth, we think.

Indicator 5: Regional cohesion (dispersion of regional employment rates)

This indicator refers to the coefficient of variation of regional employment rates.

In the context of regional indicators, this type of 'regional cohesion' indicator is of more general interest than the above specific form. Such indicators provide a means of summarising different types of information on regional variation. Some important methodological issues in their construction are discussed in the next section.

Other labour force and education related indicators

Indicators 6-8 and 16-18 are likely to be quite suitable and useful at regional levels. They often come from large data sources. Indicators coming from the LFS may be cumulated over time to obtain greater sampling precision if necessary. Most labour force surveys have rotational designs which should permit quite efficient cumulation.

Indicator 'Low reading literacy performance of pupils' refers to the share of 25 years old pupils who are at level 1 or below of the PISA combined reading literacy score.

2.6 Indicators of regional cohesion

As noted above, 'regional cohesion' indicators are of general interest since they provide a means of summarising different types of information on regional variation.

As a specific example, in the Laeken list it has proposed to measure social cohesion (or the lack of it) in a country by the coefficient of variation of regional employment rates.

This indicator has been criticised for not providing statistically valid information for comparison across countries because its magnitude depends on the size and number of regions present in the country.

There also has been some criticism of the indicator from a substantive/policy angle. For instance Atkinson et al (2002) argue that regional disaggregation is primarily a means of interpreting the national position, and hence the forms should be on the comparison of regional with the national aggregate rather than on the comparison of measures of regional dispersion across countries; and also that the relevant objective should be improving the overall national position rather than reducing regional differences.

In our view, while the proposed indicator needs to be improved from a statistical point of view, the above criticism from the substantive/policy is not really valid, or at least is not a fair one. We do need to be able to synthesis the wealth of information contained in the regional breakdown of the common indicators of social inclusion. This argument applies not only to employment rates, but could also be applied to regional disparity in the rates of unemployment, poverty and deprivation etc. Of course, in so far as the currently proposed regional cohesion indicator has some short-comings from a statistical point of view, an attempt should be made to reduce those.

The following suggestion are made towards this objective.

Taking the lowest possible level of regions

Generally the variation in regional population sizes across EU countries seems to reduce as we move down to lower levels of regions. And obviously, the number of regions available increases. (See for instance, Table 3.1, where the mean size of NUTS3 regions is quite stable across counties throughout EU25 with a few exceptions.) Both these factors contribute towards improving the comparability of measures of regional dispersion (or cohesion) across countries. Hence such measures should be constructed using regional units of the lowest level possible for which the required indicators can be produced. For instance, employment related measures (such as employment and unemployment rates) are or can be produced to at least NUTS3 level in most countries. Synthetic estimates to much lower levels (such as LAU2) can be produced using small area estimation procedures of the type described in Chapter 14.

Use of an alternative measure

One proposal is use an indicator formally similar to Laeken indicator 2 "inequality of income distribution: S80/S20 income quintile ratio", except that: (1) the units of analysis are regions rather than individual persons; and (2) the procedure can be applied not only to regional income levels, but to any of a list of indicators such as regional poverty rates, employment rates, unemployment rates, or indeed to summarise regional dispersion in any of the Laeken indicators, which can be reasonably produced at the regional level. The procedure can be applied also to other indicators, such as those which may be constructed from the information compiled in NewCronos. For instance, the procedure has been applied to construct an indicator of regional disparity in GDP/capita for inclusion in the third progress report on cohesion.[3]

[3] We are thankful to Mr Matthew Brooke for providing this information.

Basically the procedure involves the following steps:

(1) the required indicator is computed for each region (such as NUTS2) of the country;

(2) the regions in the country are ranked according to the value of the indicator, say from the smallest to the largest;

(3) population of the regions is cumulated;

(4) in the required ratio S80/S20, S20 refers to the population-weighted average of regional indices for regions in the bottom 20% of the cumulative distribution, and S80 refers to the same in the top 20% of the cumulative distribution.

Some adjustment is required for the fact that cut-off at exactly the required point (eg. 20% of the cumulative population) is not obtained when a unit cuts across that point. This problem may be trivial for the conventional S80/S20 for individual income, but can be serious with a small number of units. The procedure is detailed in the box below.

Let w_i be the population size and y_i the required indicator of region i. With regions arranged according to increasing values of y_i, let $W_i = \sum_{j=1}^{i} w_j \Big/ \sum_{j=1}^{I} w_j$ be the cumulative population share of regions up to i; I is the total number of regions in the population.

Let W_0 be the required cut-off point (such as 20% of the total cumulation). Determine the particular region i which satisfies the relationship

$$W_{i-1} < W_0 \leq W_i .$$

The required estimate S_0 for the "share" of the bottom W_0 of the population is equal (more precisely is proportional) to:

$$S_0 = \sum_{j=1}^{i} w'_j * f_j \Big/ \sum_{j=1}^{i} w'_j$$

where $w'_j = w_j$ for $1 \leq j < i$, and $w'_i = \frac{W_0 - W_{i-1}}{W_i - W_{i-1}} * w_i$.

(Note that with this definition, the denominator of the expression for S_0 equals W_0.)

The procedure is the same for obtaining the share of the *top* W_0%, with only the following modification. We determine i such that $W_{i-1} \leq (1 - W_0) < W_i$. The weight is modified for this particular region i as $w'_i = \frac{W_i - W_0}{W_i - W_{i-1}} * w_i$, and

$$S_0 = \sum_{j=i}^{I} w'_j * f_j \Big/ \sum_{j=i}^{I} w'_j ,$$

Improving the measure of coefficient of variation

As has been noted, comparability of the coefficient of variation (cv) as a measure of regional cohesion is affect by differences among countries in the size and number of regions. Actually, the observed variability in estimated regional rates is also affected by the magnitude of sampling error to which those estimates are subject. In other words, a part of observed variability results merely from sampling error in the estimation of regional indicators which are being compared in computing the cv.

The above mentioned sources of variation need to be taken into account. In the following we propose a more refined (comparable) measure. Only minimal details are presented; the underlying theory may be found in sampling text books, such as Hansen, Hurwitz and Madow (1953).

Consider a particular measure such as regional employment rates. If p is the average rate in a country, variance among individuals in the population is $\sigma^2 = p.(1-p)$.

With the region viewed as a "cluster" of very large size, variability between regional mean rates is approximately $\sigma_\rho^2 = \rho.\sigma^2$, where ρ is the 'intra-cluster correlation coefficient'. This coefficient is a measure of the degree to which elements within the "cluster" are more homogeneous than elements in the population generally. In other words, it provides a measure of regional disparities.

In practice, estimates of regional rates are generally subject to sampling error (se). The observed variability between regional means is larger than the actual variability, and is approximately $\hat{\sigma}_\rho^2 = \sigma_\rho^2 + se^2$.

Parameter ρ varies also with the size (say B) of the "cluster". Other things being equal, the following is an empirical approximation to the reduction in ρ with increase in B (Hansen, Hurwitz and Madow, 1953): $\rho = \rho_0.(B/B_0)^{-0.4}$, where B_0 is merely some standard or 'normalising' constant. We take "size" to mean population size. In principle, the effect simply of variation in mean size of regions is removed in ρ_0. It is this parameter which is of interest in comparing regional cohesion between countries.

From the above equations, we can derive the following:

$$\rho_0 = \left(\frac{B}{B_0}\right)^{+0.4} . \frac{\hat{\sigma}_\rho^2 - se^2}{p.(1-p)}.$$

Note that the coefficient of variation (cv) in the existing measure of (the lack of) regional cohesion is $cv^2 = \hat{\sigma}_\rho^2 / p^2$. We are proposing to replace it by the measure

$$\rho_0 = \left(\frac{B}{B_0}\right)^{+0.4} . \left(cv^2 - \left(\frac{se}{p}\right)^2\right) . \left(\frac{p}{(1-p)}\right).$$

As an illustration, we have carried out a simple simulation as follows.

We created a population with overall p=0.15, divided into a large number (1,024) "clusters" (regions) such that, underlying a lot of random variability, there is some systematic variation in the cluster means along the list. This is akin to the situation in a country with a large number of small regions with the statistic of interest subject to some systematic geographical variation.

By taking groupings of these "clusters", we can simulate the effect of different sizes of "regions", i.e., a situation in which the same population has been divided arbitrarily into different numbers of regions. A "true" measure of regional differences should not be significantly affected by that arbitrary number.

The results in Table 3.3 of this simulation, which cover a very large variation in mean regional sizes, is reassuring.

Table 3.3 Results of simulation concerning the proposed measure of regional cohesion

[1]	assumed proportion p=15%					
[2]	relative size of "region"	1	2	4	8	16
[3]	relative number of "regions"	16	8	4	2	1
[4]	observed population variance	75.0	51.9	38.9	32.0	29.6
[5]	assumed sampling variance*	6.39	3.19	1.60	0.80	0.40
[6]	estimated ρ	.054	.038	.029	.024	.023
[7]	adjusted ρ_0	.026	.025	.025	.027	.034
[8]	variation in adjusted ρ_0 values	1.00	0.94	0.95	1.05	1.29
[9]	Coefficient of variation cv	0.58	0.48	0.42	0.38	0.36
[10]	variation in cv	1.00	0.83	0.72	0.65	0.63

*As to the magnitude of sampling error, we arbitrarily assumed a certain value for the original "regions" (the first column), and then adjusted it for other sizes of regions by assuming that the sample size per region increases proportionately with the increase in regional size. This s based on the realist assumption that there is a given national sample size which is shared among the regions of whatever average size.

It can be seen that the proposed measure ρ_0 is quite constant across regions of very different average sizes. This is as it should be, given that the underlying situation (geographical variavion in the statistic of interest) is essentially the same in the different situations. The only difference is the assumed number and size of regions into which the country has been divided.

It can also be seen from the table that the measure currently used, the coefficient of variation, turns out to be dependent on the assumed number and size of regions.

2.7 Micro-level indicators of non-monetary deprivation

It is clear from Sections 2.3 and 2.4 that for indicators at the regional level, in comparison with those at country level, there has to be less stress on monetary indicators and consequently increased stress on non-monetary dimensions. One of the important objectives of this project is to incorporate, with increased emphasis, non-monetary dimensions of deprivation to complement indicators of income poverty at the regional level.

Here we are concerned with indicators defined at the micro (personal or household level). A number of examples of such indicators are given in Chapters 6 and 8, where we present indicators of non-monetary deprivation constructed at the micro level using ECHP and similar data. These indicators can be consolidated, and also combined with indicators of income poverty, to construct a single or a set of indicators of deprivation (overall or in different dimensions) at the individual level. These can then be aggregated to form regional indicators.

When measures at the regional level are constructed by aggregating information on individual elementary units, two types of measures which can be so constructed should be distinguished:

- Average measures, i.e. ordinary measures such as totals, means, rates and proportions constructed by aggregating or averaging individual values. (Examples: unemployment rate in the area; proportion of the population have a certain characteristic).
- Distributional measures, such as measures of variation, dispersion, distribution etc, among households and persons in the region. A most important example, after income poverty rate, is the non-monetary deprivation rate; such measures depend on the distribution of the whole population.

The patterns of variation and relationship for the two types of measures can be very different from each other, and hence require separate statistical models. Average measures are generally more easily constructed or are available from alternative sources. By contrast, distributional measures tend to be more complex and are less readily available from sources other than complex surveys. At the same time, the distributional measures are often more pertinent in the analysis of poverty and social exclusion.

An important point to note is that, more than at the national level, many measures of averages can also serve as indicators of disparity and deprivation when seen in the regional context. Furthermore, the dispersion of regional means is of direct relevance in the identification of geographical disparity.

Chapter 3
Choice of units for the construction of regional indicators

3.1 The issue

The definition and choice of appropriate units to serve as 'regions' for the construction of poverty and related indicators is a most fundamental issue to be considered at the outset. The Project Terms of Reference identify the following tasks in relation to the choice of appropriate units. To quote, the tasks are to:

- Set the framework for analysis by selecting the appropriate regional unit for analysis. The regional unit should either be defined according to administrative or political criteria (i.e., the NUTS2 regions) or according to more functional criteria;
- Analyse the extent to which regional deprivation should be defined as a self-contained concept, different from individual deprivation;
- Examine the statistical capacity at regional level (bodies responsible for data collection, methods used, etc.) and how it relates to statistical capacity at the national level.

The answer to the first question should be determined in the first instance by how the regional dimension has been incorporated into the National Action Plans for Social Inclusion (NAP/inc).

> "Many NAPs highlight territorial disparities in the extent of social exclusion and poverty. Often, such disparities take the form of an urban/rural divide, but in the case of Hungary and Poland the concern is also expressed in terms of disparities across administrative regions. In fact, only four of the EU-10 States have more than one NUTS2 region – Poland, Hungary, the Czech Republic and Slovakia. In these countries, the indicator of dispersion (coefficient of variation) of employment rates across NUTS2 regions is never as high as in the two EU-15 Member States featuring the highest regional disparities i.e., Italy and Spain. … In their NAPs, Slovenia and Lithuania refer to LFS data below the NUTS2 level pointing to significant regional employment gaps.
>
> The lack of regionally disaggregated data on income and living conditions makes it impossible to develop a comprehensive analysis of the geographical distribution of poverty and social exclusion in the EU. This will be an important area for future development and research. …
>
> … The sub-national distribution of poverty and social exclusion is in some instances described through the territorial breakdown of various indicators, both common and tertiary ones." (European Commission, 2004)

In the Action Plans, extensive reference to the regional dimension appears in a number of countries including Germany, United Kingdom, Italy, Spain, Belgium, Poland, Slovakia, among others. Almost invariably, though by no means exclusively, the reference is to administrative divisions, specifically NUTS regions at various levels. Also, in normal usage, 'regions' are taken to mean geographical-administrative partitions of the country.

Of course, it is also possible to consider ‘functional regions’, such as regions defined in terms of the labour market, production, trade or other economic indicators, or in terms of density and other characteristics of the population distribution (e.g., urban-rural distinction).

All examples above refer to types of units which are geographically based. An additional consideration is the disaggregation according to population subgroups, i.e., groups identified by characteristics of individual households and persons. Examples are children, elderly persons, national minorities, immigrants, … For instance, Laeken indicators already recommend elaborate classification of the indicators by gender, age, economic activity and labour force status, etc. Such subgroups are distinguished from geographically based classifications in that the former tend to be dispersed in the population. The dispersal may be quite uniform, as is the case generally for demographic subgroups, or the groups may be more concentrated or 'segregated'. However, unlike geographically-based classifications, such geographical separation is never complete.

We can take it for given that the term 'region' can refer only to geographically-based units, whatever the actual criteria used to define them. Analysis of distributed subpopulations is important, but in the context of the present project, only as a complement to regional breakdown. In fact, as we have discussed at several points in more detail in this report, there is a serious degree of *competition* between the two types of classifications (geographical regions versus distributed subpopulations). This competition arises primarily from the practical aspects to do with data availability. It is generally not possible to introduce double cross-classification of the available information by geography *and* population subgroup simultaneously.

Hence the main issue is the choice of appropriate type (or types) of geographically-based units for the construction of 'regional' indicators.

3.2 Choice of appropriate geographical units

Perhaps there are three classes of such units which may be considered for the purpose:

- Geographical units based on or defined according to some functional criteria. Examples are Labour Market Regions - such as Sistema Economico Locale (*SEL)* in Tuscany, which are largely but not entirely confined to be within Provinces (NUTS3 regions), but may not take account of administrative divisions below that level. Similarly, it is sometimes useful to consider ecologically defined regions.
- Units defined in terms of the urban-rural classification. The classification often has to be more elaborate than a simple dichotomy.
- And of course, units based on administrative or political criteria, specifically NUTS regions.

Geographical units based on functional criteria

Indicators for geographical units based on functional criteria can be suited for specialised purposes and populations. Different criteria may be required and/or preferred by different users and for different objectives. And all these may be country-, region-, and also domain-specific. Hence we consider that indicators for units defined on the basis of pure functional criteria are generally less suitable for *general* purpose use.

On the other hand, there is also a need for constructing indicators not for general purpose use but for the *specific* policy purpose of the fight against poverty and social exclusion. In this sense, if we are prepared to admit that accumulation of social inclusion disadvantage tends to concentrate geographically as well as on particular socio-economic groups (because of pull factors like land and housing costs, attraction for the peers …), then the use of functional regions for the purpose of area-based policies is perfectly justified. Hence 'functional regions' would be defined as a function of a specific policy need: for example, local market regions in Italy or "zone prioritaires d'éducation" in France have been defined for specific policy purposes; in the UK the concept of 'deprived areas' has been used in order to organise a part of its anti-poverty strategy.

Several among the six "key challenges for social inclusion" noted for NMSs (European Commission, 2004) may involve functionally defined units of analysis. We may mention at least the following three:

> "1. Expand active labour market policies and develop coherent and comprehensive lifelong learning strategies in order to increase labour market integration, especially of the long-term unemployed and groups at high risk of poverty and social exclusion, and ensure more effective and efficient spending in this area so as to improve the quality of provision.
>
> …
>
> 4. Improve access to decent housing and tackle homelessness.
>
> 5. Invest more, and more efficiently, in order to improve the quality of and access to key public services, particularly health and social services, education and training and transport."

The major difficulty in the use of units based on functional criteria is practical. In the *European context*, defining functional regions using the same criteria for 25 countries would be an impossible task. Apart from the current lack of data to construct the required indicators, there is also the lack of an underlying *uniform policy need* to do so. The implication is that it would not only be difficult but may also not be meaningful to try and construct *comparable indicators* based on units defined on functional basis.

Geographical units involving urban-rural classification

There is perhaps a wider scope for the application of indicators based on urban-rural classification. As noted above, many NAPs highlight territorial disparities in the extent of social exclusion and poverty, and often, such disparities take the form of an urban-rural divide. The policy requirements and objective can be quite distinct according to the types of place, and there are many examples of use of indicators based on urban-rural classification.

Unfortunately, there are no universally agreed criteria as to what constitutes 'rural' and 'urban'. We will comment on this further in Section 3.4 below.

Geographical-administrative regions (NUTS and LAUs)

For a number of substantive and practical reasons noted below, we consider geographical-administrative regions, specifically NUTS regions (and LAUs) at various level of classification, as the most appropriate choice. *This by no means precludes those being supplemented* *by the incorporation of other dimensions within geographical-administrative regions so defined.* For instance, NUTS regions at a sufficiently low level can be classified according to whether their character is primarily urban or primarily rural. In fact, indicators can be constructed for geographical-administrative units precisely for the purpose of such classification.

Furthermore, NUTS-based indicators can be enriched by subpopulation analysis to the extent the available data permit their further disaggregation.

Hence except for some more specialised purposes, we believe that NUTS-based regions at various levels of disaggregation are the most suitable units for the general development of regional indicators on poverty and social exclusion.

There are several reasons for this choice, which may be briefly mentioned:

(1) Social policy. NUTS regions are the most commonly used units for the formulation and implementation of social policy. This is clear from the references to such units in NAP/inc of different countries.

(2) Comparability. It is true that NUTS units are not defined in exactly the same way in different Member States, and can differ greatly in size and homogeneity from one country to another. Nevertheless, this territorial system of classification provides a *common framework* which enhances comparability of the resulting statistical information. Inter-country, EU-wide research also benefits from the use of units based on the same system of classification.

(3) Exhaustive and non-overlapping. The NUTS classification covers each country exhaustively. This does not necessarily apply to other types of geographical units. This is so certainly in the case of units defined on the basis of functional criteria. This may also apply to certain systems of urban-rural classification. For instance, sometimes the two types of areas are identified on the basis of characteristics *specific to each* (rather than on the basis of common criteria applied to both), which may guarantee neither exhaustiveness nor that the units are non-overlapping (see example in Section 3.4).

(4) Hierarchical. The NUTS system provides a hierarchical set of units. It is possible to proceed step-by-step from higher to lower units to increase the degree of disaggregation of the indicators. And obviously, data can be linked across different levels. This provides a framework for integration of the information. As will be explained in Chapter 11, we can make use of the hierarchical nature of the units to construct more precise indicators for lower level units.

(5) Communication. The units are well defined and identifiable, and are already widely accepted and used by different users and producers of statistical information.

(6) Links with other information. This is possible simply because a lot of information already exists for this type of units from many different sources. The construction of poverty related regional indicators can benefit from this fact greatly. This type of units have been used in much other research, which the present project seeks to complement and enhance.

(7) Data availability. Here we refer specifically to data availability for the purpose of constructing the required indicators. This is the case at present, and in realistic terms, is likely to remain so in the future. For us as statistician, *this is the major reason for the choice of NUTS regions for the purpose.* The largest comparable compilation of data, covering many

domains at macro - but very detailed - level, namely the tabulations provided in Eurostat Free Dissemination Database (NewCronos), is based on NUTS as the units of classification. Any practical set of indicators must make the maximum use of this data source. The small area estimation methodology we have applied to produce many illustrative indicators with a high degree of disaggregation depends on the combined use of (intensive but relatively small-scale) survey data, and (less-intensive but large-scale) data of the form compiled in NewCronos.

Hitherto, there are only a few examples of the production of social exclusion or poverty indicators at sub-national level, and are mostly confined to NUTS1 level, or even to groupings of NUTS1 regions. In this project we hope to add something significant to previous studies by going down to lower level units to the extent possible. Our target is NUTS2 level in most cases, and further to NUTS3 even if more tentatively. We also review some methodologies for possible disaggregation to still lower levels (Chapter 14).

As noted, the choice of NUTS regions as the units can be supplemented by analysis using other types of units. For instance, using information on population density, available at NUTS3 level in NewCronos tables for instance, it is possible in principle to incorporate the urban-rural dimension, though this may not be feasible in all situations. This task is not easy, however, because most of the other information is available at the most only at NUTS2 level. Another addition we have tried to incorporate is subpopulation analysis within the geographical-administrative regions, in particular poverty and social exclusion among children (persons aged 0-15), and among elderly persons (aged 60+), compared to the general population in the region. Similarly, in the analysis of education and training related variables, one can identify separately persons who have recently completed education or training courses and examine their income and labour market situation compared to the rest of the population, also across regions and countries. (See Chapters 5 and 9).

As to regional statistical capacity, it should be noted that there is no automatic relationship between that and the task of construction of indicators at the regional level. It is possible to have a quite 'centralised' national statistical agency/system, and yet have well developed regional data source (as for example, in Canada). In fact, in such a situation the regional data sources are likely to be more comparable than in a decentralised system. On the other hand, it is also possible that in a more decentralised system the required regional data are available primarily at the regional level, in which case regional statistical capability is of direct concern. The issue is clearly very context-specific. In our experience, whether for constructing good indicators at the national level or at regional levels, the first requirement remains the development of statistical capability at the national level. Real capacity at this level implies suitable arrangements, including decentralisation of that capacity as necessary, for the development of statistics in the regions.

3.3 NUTS regions

Table 3.1 shows the number of NUTS1-NUTS3 regions by country, their average population size, and their availability and identification in ECHP and other survey data.

Table 3.1 Number of NUTS regions tabulated in NewCronos, and in the ECHP survey data where available

			population	NewCronos			ECHP			Average Population			
			2000	Nuts1	Nuts2	Nuts3	Nuts1	Nuts2	Nuts3	Nuts1	Nuts2	Nuts3	
1	DE	Germany	82.195	16	41	439	15			5137	2005	187	DE
2	DK	Denmark	5.337	1	1	15	1	1		5337	5337	356	DK
3	NL	Netherlands	15.926	4	12	40				3981	1327	398	NL
4	BE	Belgium	10.251	3	11	43	3			3417	932	238	BE
5	LU	Luxemburg	436	1	1	1	1	1	1	436	436	436	LU
6	FR	France	58.894	9	26	100	8			6544	2265	589	FR
7	UK	United Kingdom	58.643	12	37	133	10	32		4887	1585	441	UK
8	IE	Ireland	3.787	1	2	8	1			3787	1893	473	IE
9	IT	Italy	57.762	5	20	103	11	20	93	11552	2888	561	IT
10	GR	Greece	10.918	4	13	51	4			2729	840	214	GR
11	ES	Spain	40.263	7	19	52	7			5752	2119	774	ES
12	PT	Portugal	10.226	3	7	30	3	7		3409	1461	341	PT
13	AT	Austria	8.012	3	9	35	3			2671	890	229	AT
14	FI	Finland	5.176	2	5	20				2588	1035	259	FI
15	SE	Sweden	8.872	1	8	21	1			8872	1109	422	SE
16	CY	Cyprus	694	1	1	1				694	694	694	CY
17	CZ	Czech Republic	10.273	1	8	14				10273	1284	734	CZ
18	EE	Estonia	1.370	1	1	5				1370	1370	274	EE
19	HU	Hungary	10.024	3	7	20				3341	1432	501	HU
20	LV	Latvia	2.373	1	1	9				2373	2373	264	LV
21	LT	Lithuania	3.500	1	1	10				3500	3500	350	LT
22	MT	Malta	390	1	1	2				390	390	195	MT
23	PL*	Poland	38.646	6	16	45	6	16		6441	2415	859	PL*
24	SI	Slovenia	1.990	1	1	12				1990	1990	166	SI
25	SK	Slovakia	5.401	1	4	8				5401	1350	675	SK
26	BG	Bulgaria	8.170	1	6	**28**				8170	1362	292	BG
27	RO*	Romania	22.443	1	8	**42**	1	8		22443	2805	534	RO*
Total			481.970	91	267	1287				5296	1805	374	

Blanks: no codes for area identification in survey, or no survey available
* Poland, Romania: national surveys
DE In ECHP, two NUTS1 regions have been coded together as a single unit.
FR Survey excludes overseas departments
UK Survey excludes Northern Ireland
In NewCronos London is a specific NUTS1 region but in ECHP London is just a NUTS2 region (UK55)
UK9, Wales, breakdown in ECHP is not comparable with the NewCronos division
IT The only ECHP survey for which NUTS3 codes provided to us by the NSI
In the survey, the 5 NUTS1 have been divided into 11 part-NUTS1 (the latter corresponds to an earlier classification)
the NewCronos regions can be identified
IE Two regions are identified in ECHP data, but they are not comparable with NUTS1 in NewCronos

The average population size for the 27 EU and Candidate countries shown is around 5.3 million for NUTS1, 1.8 million for NUTS2, and 375 thousand for NUTS3 regions. The units vary considerably in size across the countries. However, generally the range of variation (at least covering a majority of the distribution) declines as we go down the hierarchy. The extreme range of variation is less than 5:1 from the larges to the smallest country average NUTS3 size. Of course the range of variation for individual units would be larger.

Throughout this report, we have identified regions by their standard code, and have provided a full list of region names to NUTS2 for EU25, and for Italy up to NUTS3 because of the special analysis carried out for that country (Chapter 13). This listing appears in Chapter 15 (Statistical Annex).

3.4 Note on urban-rural classification of area units

In this discussion on the choice of units, it is useful to make some further comments on the possibility of using urban-rural classification.

While we consider that by far the most appropriate choice for the present purpose is to use NUTS regions, indicators specifically for urban or rural areas are also widely constructed. For instance, The Urban Audit (European Commission, 2000) has developed a large number of indicators in the context of "benchmarking of quality of life in 58 European countries". In the last data collection round of the Urban Audit (executed in 2003/2005 with reference to year 2001) information on the living conditions in 258 large and medium-sized cities within the European Union and the candidate countries (EU27) was gathered. The analysis of the data was published in 2004 in a book titled: "Urban Audit 2004 Key Indicators on Living Conditions in European Cities". A shorter description of the project was published in "Regions: Statistical Yearbook 2004" (Eurostat, 2004b).[4] The Yearbook provides eh following description of the Urban Audit.

> "In the past, comparing cities in the European Union was fraught with problems due to differences in data collection methods and definitions … The Urban Audit seeks to solve these problems by providing a comprehensive set of urban indicators covering the various aspects of urban life. The audit … covers 258 large (over 250 000 inhabitants) and medium-sized (between 50 000 and 250 000 inhabitants) cities in the enlarged European Union, Bulgaria and Romania (EU-27). … The selected cities are geographically dispersed to ensure a representative sample. The combined population of the 258 cities is 107 million inhabitants, covering more than 20 % of the EU-27 population. This large sample ensures that the Urban Audit can provide much more reliable information about European cities today than was previously available."

In Table 3.2, we provide a list of indicators incorporated, just to give an idea of the range covered. In the context of integrating distressed urban areas, OECD (1998) reports that, while there are "great differences in the use of statistically-based decision-making systems by public authorities, … there has nevertheless been an increase in the use of statistics in the design of urban regeneration policies …".

Rural-urban classification of NUTS3 regions: OECD

The OECD has proposed a definition of rural areas based on the percentage of the population of a region living in local areas (communes, municipalities) classified as rural (OECD, 1994). A local area or commune is classified as rural on the basis of *population density*: if the density is below 150 inhabitants/km^2, the local area is rural. Presumably, other areas are, by default, urban.

---

[4] Eurostat homepage information is also available on the site:
http://www.urbanaudit.org/index.aspx
Data can be accessed through five applications: (1) City Profiles; (2) How do the cities rank?; (3) How does your city compare?; (4) What is the structure of the city?; and (5) Data that can be accessed.

In the second step, NUTS3 regions are classified into three categories as follows:

- o Predominantly rural regions: over 50% of the population lives in rural communes (local areas with less than 150 inhabitants/ km^2)
- o Significantly rural regions: 15 to 50% of the population living in rural communes.
- o Predominantly urban regions: less than 15% of the population living in rural communes.

These definitions have proven to be useful in making international comparisons of rural conditions and trends. They also have the merit of being easily applied. However, it can be argued that the definition has some limitations and can result in counter-intuitive results. Being based on population density alone, the classification of a commune can depend too much on its physical area. The definition does not take into account characteristics of the surrounding area. For instance, localities which are clearly urban but are surrounded by a lot of empty land within the same commune (local area) boundary would get classified as rural.

It is not easy to find a suitable definition to fit all circumstances, as indicated by the following.

As noted by Gallego (n.d.), "this criterion is reasonable but has some limitations and [can] produce unexpected results that depend on the definition of communal boundaries, in particular for large, heterogeneous NUTS3 units, for which attributing the same label of rurality to the whole NUTS3 may be unfair. For instance, the only province in Sicily that turns out to be 'mainly urban' is Ragusa, while Stockholm is 'relatively rural'".

Rural-urban classification of NUTS3 regions: Eurostat

Eurostat has developed an urban-rural classification based on the concept of 'degree of urbanisation' for use in the Labour Force Survey (LFS).[5] The EU Regulation on LFS defines "degree of urbanisation" in column 176. The criteria apply to "local areas", normally LAU level 2 (Local Administrative Units corresponding to communes, municipalities and similar). The three types of area are defined as follows:

- o Code A : Densely-populated area. This is a *contiguous set of local areas*, each of which has a density superior to 500 inhabitants per square kilometre, where the total population for the set is at least 50,000 inhabitants.
- o Code B : Intermediate area. This is a contiguous set of local areas, not belonging to a densely-populated area, each of which has a density superior to 100 inhabitants per square kilometre, and either with a total population for the set of at least 50,000 inhabitants or adjacent to a densely-populated area.
- o Code C : Thinly-populated area. This is a contiguous set of local areas belonging neither to a densely-populated nor to an intermediate area.

A set of local areas totalling less than 100 square kilometres, not reaching the required density, but entirely enclosed within a densely-populated or intermediate area, is to be

[5] This subsection draws on unpublished memorandum (Carlquist, 2004), kindly provided to us by Teodora Brandmueller of Eurostat.

considered to form part of that area. If it is enclosed within a densely-populated area and an intermediate area it is considered to form part of the intermediate area.

For the OECD definition, information on population and land area only is required. However, in contrast to the OECD concept, the LFS concept has a provision about "contiguity" and this requires maps. For instance, so far United Kingdom and Slovakia have to be excluded from the application of the Eurostat concept, because for these areas, input maps are lacking.

Also, unlike the OECD concept, the Eurostat classification in itself does not provide a basis for the classification of *regions* into "urban" or "rural". In order to compare the practical consequences of the OECD and Eurostat definitions, it is necessary to specify some criteria for the "regionalisation" of the latter, to parallel the "predominantly rural", "significantly rural" and "predominantly urban" OECD classification.

This following criteria have been applied in a comparative study by Eurostat. It is noted that this "regionalisation" is unofficial and does not exist elsewhere.

> Local areas are classed into 3 classes (instead of 2 in the OECD concept). The consequence is that some regions do not have an absolute majority of urban or rural local areas. The following decision rule for the classification of the regions has been used ("majority" is in terms of population):

Local areas	Resulting classification of the regional
Absolute majority classified as rural	Rural
Absolute majority classified as urban	Urban
All other cases	Intermediate

> These rules means that if neither urban nor rural local areas account for a majority of the population of the region in which they lie, the region is classified as "intermediate", so as to reflect its "mixed" character.

Comparison of the OECD and Eurostat (LFS) definitions

The following results and comments have been reported from a comparative exercise carried out at Eurostat.

	Eurostat (LFS) definition			
OECD definition	Rural	Intermediate	Urban	total
Rural	307	62	-	369
Intermediate	63	271	47	381
Urban	1	93	230	324
total	371	426	277	1074

Comments:

"In some cases, the reason for the differences is obvious and is a consequence of the way the NUTS regions have been constructed in the various countries. All German city regions (Stadtkreise) which have less than 50,000 inhabitants are classed urban with the OECD method, but may be rural or intermediate (depending on the density of surrounding areas) with the LFS method. 17 German "Stadtkreise" are thus downgraded with the LFS method. This kind of medium-sized towns do not constitute separate NUTS3 regions in any other country.

Another general effect of the comparison is that the "blue banana" of very dense areas that stretches from England over Benelux and western Germany down to Italy is more urban with the OECD concept than with the LFS concept. It can be explained by the fact that OECD looks only at population density, while LFS also considers size of urban areas (the 50,000 threshold for the A and B classes).

On the other side, with the LFS method some peripheral regions that contain a large city are "raised" to urban compared to the OECD method. This effect is particularly visible in Spain and Lithuania and examples can be found also in France, Italy, Greece, and in the Nordic countries. Many of these "raised" regions are almost empty (due to mountains, deep forests or semi-deserts) apart from a large or middle-sized city. The city thus dominates the region completely."

Urban-rural definitions: example from the UK

In the United Kingdom, Office of the Deputy Prime Minister (n.d.) has issued a Users' Guide on urban and rural area definitions. It states that there are *two* basic recommended definitions.

1. Urban Settlement definition based on land use (England & Wales)

"… The basis of the definition is land with an *irreversibly urban use* and it is *independent of administrative area boundaries*. .. The definition appears as computer readable boundaries of all built up settlements with a minimum population of 1,000 and a minimum land area of 20 hectares. The user can choose a settlement size above which land is treated as urban for their purposes. To produce consistency in statistical reporting a cut off population of 10,000 is recommended for general purpose use. ... Using this standard, all settlements of over 10,000 are treated as urban areas. All smaller settlements, together with all other land, are treated as rural areas. .."

2. Administrative Area Classification definition based on socio-economic variables (England)

"This is the .. classification of rural and urban administrative areas based on a range of socio-economic characteristics of the population at local authority and ward [NUTS5] levels. A county level classification, based on the ward level classification, is recommended with the reservation that it should be used only where there is no other choice. *At a high geographical level it is less meaningful to describe an area as urban or rural*, so this definition has limited use.

This definition operates on three geographical levels: ward, local authority and county. … The overall method for the administrative areas classification was designed to identify those areas with certain social and economic conditions that were considered to give them a rural character. By default the remainder are considered urban, providing a classification of all areas as either urban or rural."

These two definitions together do not always provide urban and rural categories which are exhaustive and non-overlapping. An urban-centred definition aims to delineate urban areas, and treats rural areas as a "residual" category. By contrast, a rural-centred definition aims to delineate rural areas, and treats urban areas as a "residual" category. Each definition is of course defined to serve a specific purpose, analysis focused on 'urban' areas for the first definition, and analysis focused on 'rural' areas for the second. But for the purpose of considering both these objectives on an equal and unified footing, a single and consistent system of urban-rural classification would surely be preferable.

Table 3.2 Urban Audit Indicator list: an illustration

de1001v	Total Resident Population
de1002v	Male Resident Population
de1003v	Female Resident Population
de1040v	Total Resident Population 0-4
de1041v	Male Resident Population 0-4
de1042v	Female Resident Population 0-4
de1043v	Total Resident Population 5-14
de1044v	Male Resident Population 5-14
de1045v	Female Resident Population 5-14
de1046v	Total Resident Population 15-19
de1047v	Male Resident Population 15-19
de1048v	Female Resident Population 15-19
de1049v	Total Resident Population 20-24
de1050v	Male Resident Population 20-24
de1051v	Female Resident Population 20-24
de1052v	Total Resident Population 25-54
de1053v	Male Resident Population 25-54
de1054v	Female Resident Population 25-54
de1025v	Total Resident Population 55-64
de1026v	Male Resident Population 55-64
de1027v	Female Resident Population 55-64
de1028v	Total Resident Population 65-74
de1029v	Male Resident Population 65-74
de1030v	Female Resident Population 65-74
de1055v	Total Resident Population 75 and over
de1056v	Male Resident Population 75 and over
de1057v	Female Resident Population 75 and over
de2001v	Residents who are Nationals
de2002v	Residents who are Nationals of other EU Member State
de2003v	Residents who are not EU Nationals
de2004v	Nationals born abroad
de3001v	Total Number of Households
de3002v	One person households (Total)
de3005v	Lone parent households (Total)
de3006v	Lone parent households (Male)
de3007v	Lone parent households (Female)
de3008v	Lone pensioner (above retirement age) households Total
de3009v	Lone pensioner (above retirement age) households Male
de3010v	Lone pensioner (above retirement age) households Female
de3011v	Households with children aged 0 to under 18
de3012v	Nationals that have moved into the city during the last two years
de3013v	EU Nationals that have moved into the city during the last two years
de3014v	Non-EU Nationals that have moved into the city during the last two years
sa1001v	Number of dwellings
sa1004v	Number of houses
sa1005v	Number of apartments
sa1007v	Number of households living in houses
sa1008v	Number of households living in apartments
sa1011v	Households owning their own dwelling
sa1012v	Households in social housing
sa1013v	Households in private rented housing
sa1015v	Number of homeless persons
sa1016v	Average price for an apartment per m2
sa1023v	Average price for a house per m2
sa1017v	Annual rent for social housing per m2
sa1021v	Average annual rent for an apartment per m2
sa1024v	Average annual rent for a house per m2
sa1018v	Dwellings lacking basic amenities
sa1019v	Average occupancy per occupied dwelling
sa1025v	Empty conventional dwellings
sa1026v	Non-conventional dwellings
sa1022v	Average area of living accommodation (m2 per person)
sa2001v	Life expectancy at birth
sa2002v	Male life expectancy at birth
sa2003v	Female life expectancy at birth
sa2004v	Infant Mortality per year
sa2005v	Male Infant Mortality per year
sa2006v	Female Infant Mortality per year
sa2007v	Number of live births per year
sa2008v	Number of live births per year (Male)
sa2009v	Number of live births per year (Female)
sa2013v	Number of deaths per year under 65 due to heart diseases and respiratory illness
sa2014v	Number of deaths per year under 65 due to heart diseases and respiratory illness (Male)
sa2015v	Number of deaths per year under 65 due to heart diseases and respiratory illness (Female)
sa2016v	Total deaths under 65 per year

(table continued)

Table 3.2 (cont.)

ci1005v	National Elections: Total electorate (registered)
ci1006v	National Elections: Total votes counted
ci1007v	City Elections: Total electorate (eligible)
ci1008v	City Elections: Total electorate (registered)
ci1009v	City Elections: Total votes counted
ci1011v	City Elections: Electorate aged less than 25
ci1010v	City Elections: Total votes counted by voters aged less than 25
ci1016v	Total number of elected city representatives
ci1017v	Number of Male elected city representatives
ci1018v	Number of Female elected city representatives
ci2001v	Total Municipality Authority Income
ci2002v	Municipality Authority Income derived from local taxation
ci2003v	Municipality Authority Income transfered from national or regional government
ci2004v	Municipality Authority Income derived from charges for services
ci2005v	Municipality Authority Income derived from other sources
ci2006v	Total Municipality Authority Expenditure
ci2007v	Total number of persons directly employed by the local administration
ci2008v	Number of persons directly employed by the local administration in central administration
ci2009v	Number of persons directly employed by the local administration in education
ci2010v	Number of persons directly employed by the local administration in health and social services
ci2011v	Number of persons directly employed by the local administration in public transport
ci2013v	Number of persons directly employed by the local administration in other
te1001v	Number of children 0-4 in day care
te1002v	Number of children 0-4 in private day care
te1003v	Number of children 0-4 in public day care
te1029v	Number of children 0-4 in other day care e.g. Church
te1005v	Total students registered for final year of compulsory education
te1030v	Students leaving compulsory education without having a diploma
te1017v	Students continuing education after completing compulsory education
te1018v	Male students continuing education after completing compulsory education
te1019v	Female students continuing education after completing compulsory education
te1031v	Students in upper and further education (ISCED level 3-4)
te1032v	Male students in upper and further education (ISCED level 3-4)
te1033v	Female students in upper and further education (ISCED level 3-4)
te1026v	Students in higher education (ISCED level 5-6)
te1027v	Male students in higher education (ISCED level 5-6)
te1028v	Female students in higher education (ISCED level 5-6)
te2016v	Total number of residents qualified at ISCED level 1
te2017v	Number of Male residents qualified at ISCED level 1
te2018v	Number of Female residents qualified at ISCED level 1
te2001v	Total number of residents qualified at ISCED level 2
te2002v	Number of male residents qualified at ISCED level 2
te2003v	Number of female residents qualified at ISCED level 2
te2019v	Total number of residents qualified at ISCED levels 3 and 4
te2020v	Number of male residents qualified at ISCED levels 3 and 4
te2021v	Number of female residents qualified at ISCED levels 3 and 4
te2022v	Total number of residents qualified at ISCED levels 5 and 6
te2023v	Number of male residents qualified at ISCED levels 5 and 6
te2024v	Number of female residents qualified at ISCED levels 5 and 6
en1003v	Average temperature of warmest month
en1004v	Average temperature of coldest month
en1005v	Rainfall (litre/m2)
en1001v	Number of days of rain per annum
en1002v	Total number of hours of sunshine per day
en2001v	Winter Smog: Number of days sulphur dioxide SO2 concentrations exceed 125 microgram/m3
en2002v	Summer Smog: Number of days ozone O3 concentrations exceed 120 microgram/m3
en2003v	Number of days nitrogen dioxide NO2 concentrations exceed 200 microgram/m3
en2005v	Number of days particulate matter PM10 concentrations exceed 50 microgram/m3
en2006v	Concentration of lead Pb in ambient air in microgram/m3
en2007v	Number of residents exposed to outdoor day noise levels above 55 dB(A)
en2008v	Number of residents exposed to sleep disturbing outdoor night noise levels above 45 dB(A)
en2014v	Total carbon dioxide CO2 emissions
en2009v	Total carbon monoxide CO emissions
en2010v	Total methane CH4 emissions
en2011v	Total non-methane volatile organic compounds NVOC emissions
en2012v	Total sulphur dioxide SO2 emissions
en2013v	Total nitrogen dioxide NO2 emissions
en3001v	Total number of annual tests (on all parameters) on drinking water quality
en3002v	Number of annual determinations which exceed the prescribed concentration values
en3003v	Total consumption of water
en3004v	Number of dwellings connected to potable drinking water system
en3006v	Number of dwellings connected to sewerage treatment system
en3008v	Number of water rationing cases, days per year
en3009v	Number of scheduled water cuts, days per year
en4001v	Annual amount of solid waste (domestic and commercial)

3.5 NUTS regions as units of analysis

Area-level indicators

The Project Terms of Reference also call for an analysis of "the extent to which regional deprivation should be defined as a self-contained concept, different from individual deprivation".

Of course, it must be recognised that indicators of poverty, deprivation and social exclusion may have an important *territorial dimension*, pointing to the need to take account of regional and local differences.

In an ideal context, one may seek to give regional breakdown on *all* indicators. However, simply the introduction of more extensive breakdown is neither possible (because of data limitations), nor sufficient in itself. There are two related additional considerations to be taken into account.

- o The first issue is the extent to which *new types of indicators* have to be added to the 'national list' when the focus moves to the regional level.
- o The second issue is the extent to which this involves the introduction of *new types of units of analysis* – i.e., regions, as distinct from individual persons and households, as the units of analysis.

We have discussed the first question elsewhere in this report (Chapters 2, 5 and 7 for instance). The following remarks concern the second issues, namely regions themselves as units of analysis.

The important addition to the set of Laeken indicators we began with, can be the incorporation of indicators defined and measured only at the area level in order to identify, as it were, the "territorial reality" of the area. These indicators are not necessarily simply aggregations of individual level values, though construction of area indicators through such aggregation is not precluded. It is this sort of indicators which underpin area-based policies which have become a common part of the some governments' approach to tackling social exclusion.

"Just as, for instance, poverty and low education are characteristics of individuals, there are other types of indicators which relate to a population rather than to the individual. Disadvantage may be located in a community and not a property of the particular individuals who live there ..." (Atkinson *et al.*, 2002). It is on the basis of such arguments that deprivation indices are constructed for local government units – such as extensively in the example already given from the UK. "Such territorial indicators have been particularly adopted by national governments that have targeted policy geographically. … Area-based anti-exclusion policy is based on a set of hypotheses about the location of exclusion, and these points to the collection of area-based data. But for the same reason, the collection of household-based indicators is necessary to evaluate the hypothesis on which this policy is based." (ibid).

There are many examples of action programmes targeted specifically and in the first instance at areas, rather than on persons and households identified solely on the basis of their personal characteristics and circumstances. For instance, Indices of Deprivation 2000 published by the UK Department of Environment, Transport and Regions developed at the ward (NUTS5) level six domain indices covering income, employment, health deprivation and disability, education skills and training, housing, and geographical access to services. These indicators were then put together to construct single overall ward-level index of are deprivation using a simple weighting system. We provide further details on the specific indicators, as well as on the methodology of construction in the UK in Chapter 14.

Glennester *et al.* (1999) review the long-running debate about whether area-based policies can make significant impact on poverty and social exclusion. They note that "there is a strong tradition of academic work that argues that this is a misguided strategy." However, "there may be causal factors at work that derive from area-based problems that suggest area-based solutions".

Again in relation to the UK, the study OECD (1998) refers specifically to the Index of Local Conditions (ILC) being the most recent and widely-used of a long series of similar indices in the UK aimed at ranking areas according to their level of deprivation, defined by a range of census variables, covering social, economic, housing and related aspects. Specifically, it covers indicators on unemployment, children in low earning households, overcrowded housing, housing lacking basic amenities, households with no car, children living in unsuitable accommodation, 17-year old not in full-time education, long-term unemployment, income support recipients, low educational attainment, standardise mortality rates, derelict land, and house content insurance premium. *These indicators are combined into a single score for each area.* The index shows whether an *area* has more or less deprivation than the national norm, and enables areas to be ranked. It is possible that such an index constructed around multiple dimensions is better able to take account of longer-term aspects of differential access to education, labour market, housing, health, and general life-chances. (For a description of the methodology involved, see Section 14.2.)

An area-based index on similar lines has been tested in Spain. The index was built at the census enumeration district level in municipalities exceeding a certain population threshold, using three simple but strongly differentiated indicators: rates of unemployment, the proportion of residents without formal education, and the proportion lacking basic amenities. These indicators were combined into a single deprivation score using a Chi-square standardisation method (OECD, 1998).

Aggregation of indicators across domains of deprivation

The issue of regions as units of analysis in their own right is also related to the multi-dimensional nature of deprivation and social exclusion. Two major issues arise in the construction of indicators incorporating this multi-dimensionality.

- The first concerns the extent to which many different dimensions can be consolidated to form a single, or at least a much reduced number of indicators for practical use in policy and research; and where such consolidation is possible, the choice of the methodology and the weights to be given different items or indices of deprivation.
- The second concerns the level of units for which such aggregation should be done.

As to the consolidation of diverse indictors, it is obvious that the complex phenomena of poverty and social exclusion cannot be measured or summarised by a single composite index. As noted by Mickleright (2001), at best such an index could only *complement* a range of separate indices.

Consolidating diverse measures into a single or a very small number of composite indicators can be helpful in summarising the situation. It can facilitate understanding by and communication with the general public. Too many diverse indicators, especially moving in different directions, can complicate and hinder the monitoring of policies and their impact. Consolidating the indicator(s) is also expected to be statistically more robust and stable. These advantages become more important as we move down from national to regional level, where indicators have to be constructed from more sparse and less precise data.

On the other and, there is the danger of summarising too much, and in the process losing important information. In fact, crude aggregation can be seriously misleading. Furthermore, the results may be sensitive to the specific methodology adopted in putting together different items of information, such as the weights used and other assumptions made in the process.

The inclusion of geographical access to services is an interesting one. This indicator is a 'true' area-level indicator, rather than simply an aggregation of individual values for people living in that area (except in the trivial sense that all such individuals have the same value because of the area they live in!). The same would apply to other similar indicators such as on transport, crime, pollution, … It should be noted, however, that some of these indicators may actually not reflect deprivation and hence it is not always appropriate to combine them into a single index with other deprivation indicators. For instance, poor geographical access to services may simply result from the fact that an area is a rich one and the population can rely on private means and services.

Two aspects of the aggregation are involved: aggregation over dimensions of deprivation; and aggregation over individuals or households. Different types of measures are obtained depending on the order in which the two aggregations are applied.

Aggregation first over dimensions for each individual provides a clear picture of the overlap (intensity) of deprivation in different dimensions as experienced by individuals. This requires that the information on individuals can be linkable across dimensions at the micro level. We have discussed this in the context of micro-level indicators of non-monetary deprivation in Section 2.7 and Chapter 6.

First aggregating indicators of a given dimension over individuals in (say) a region, and then over different dimensions for that region, informs us about characteristics and circumstances of the area concerned in terms of geographical concentration (intensity) of deprivation.

Apart from substantive differences, an important practical difference between the two modes of aggregation is that aggregating first over areas and only then over dimensions of deprivation is generally much less demanding on the statistical data required for the construction of indicators.

Some researchers have argued that the first mode of aggregation – over multiple deprivation indicators for each individual and only then over areas – is the really important one, while other have argued the opposite.

In our considered view, both forms of aggregation are useful – each gives additional new information not fully captured by the other. We have considered both in this report.

Finally, the reader should be reminded that area-level indicators can be – and often are – constructed without involving any aggregation of individual values. This may either be because the variables of interest are only definable at the area level (such as availability of facilities or services in the area, population density, other community level or ecological variables), or because no measurements exit for individual persons on the variables concerned (such as indicators constructed using small area estimation techniques of the type discussed in Chapter 14).

In this respect, regional deprivation can indeed be defined as a "self-contained concept, different from individual deprivation".

Chapter 4
Making the best use of sample survey data

4.1 Introduction

The problem of sample size

Poverty and related indicators may be derived from diverse data sources. However, because of the complexity of the information on which such indicators are based (such as detailed income distribution of households and persons in the population), most have to be obtained from intensive surveys (LFS, HBS, LIS, ECHP, EU-SILC etc.). *The major problem in the production of indicators for regions or other small domains is the smallness of sample size available in such surveys.* Generally, adequate sampling precision is available at the national level, as for instance demonstrated by the extensive use of ECHP data for the purpose. The same may apply to indicators at NUTS1 level in some cases, but generally sampling errors may be too large for the results to be useful even at that level. For instance, in a study on measures of well-being and exclusion in Europe's NUTS1 regions using LIS data (Stewart, 2002, 2004), it is clear that the large sampling errors involved often make it difficult to draw clear conclusion in many instances. This is also true in a study on patterns of poverty across European regions (Berthoud, 2004), to a lesser extent only because the results are aggregated over small NUTS1 regions where sample size problems are most critical. As a final example at the EU level, a study on the impact of relative poverty lines (Kangas and Ritakallio, 2004) goes below the national level only to "mega regions", meaning the division of each country into at most two – the richer and the poorer – parts. There are a few studies of regional variations for individual countries, but mostly confined to countries with large samples. For instance, Rodrigues (1999) reports estimates of mean income (per 'equivalent adult') and head-count ratios (poverty rates) by NUTS2 regions in Portugal. It has been possible to do so at this level of detail because Portugal had an exceptionally large sample size in the ECHP, compared to other countries in the project. Even so, the sample size was just large enough to report results at NUTS2 level. Chakravarty and D'Ambrosio (2002) report similar results for Italy, but this time only at NUTS1 level. It is not possible to produce such results at a finer level of detail such as NUTS2 form the source concerned. This is despite the fact that among the countries participating in the ECHP, Italy had the largest sample size.

Our requirement is considerably more demanding than the above: ideally to be able to produce useful regional indicators for NUTS2 regions (or other geographical subpopulations of similar size) in most cases, and even to NUTS3 level in some cases.

To indicate the magnitude of the problem, we show in Chapter 15 (Statistical Annex) the sample sizes for the survey data we have used (averaged over ECHP waves for EU15 countries, and for the national surveys for Poland and Romania). The extreme case seems to be that of Germany, where some regions also at NUTS1 have sample of only 100-200 persons. Unfortunately, information is not available in ECHP data to identify NUTS2 sample sizes, with the exception only of the UK, Portugal and Italy (in the last case on the basis of information specially provided to us by ISTAT). For NUTS2 level, a number of units in UK for instance have only 150-300 sample persons, corresponding to even fewer households which are the real sampling units. In Italy, it can be seen that NUTS3 sample sizes are small, and no sample cases are present for some regions at this level.

The problem will be reduced when indicators can be based on larger surveys such as the LFS, but will still remain. In any case, a survey like the LFS does not contain the type of information required for the construction of poverty and related indicators, and also the LFS micro data are less accessible.

The problem of sample size requires a more sophisticated statistical approach than simply using direct estimates from single rounds of sample surveys of moderate size. This approach needs to have three fundamental aspects:

(1) making the best use of available sample survey data, such as by cumulating and consolidating the data to construct more robust measures which can permit a greater degree of spatial disaggregation;

(2) exploiting to the maximum 'meso' data (such as the highly disaggregated tabulations available in NewCronos[6]) for the purpose of constructing indicators for small areas;

(3) using the two sources in combination to produce the best and most complete possible estimates for subnational regions using appropriate small area estimation (SAE) techniques.

This chapter is concerned with the first aspect.

Description of available data

To explain the statistical procedures for "making the best use of available sample survey data" in concrete terms, it is useful to consider the actual data sets we have utilised for the purpose of this project.

For most of the EU15 countries, eight waves of ECHP are available. The main exception is Sweden, where data corresponding to the last five waves have been specially constructed from alternative sources, and are somewhat less comparable with the others. There are six waves available for Finland, and seven for Austria and Luxembourg. University of Siena also has access to data from the first wave of a three-wave national survey on living conditions in Poland, and from a large household budget survey in Romania. Through another project with the World Bank, we have been able to analyse data from a living standards measurement survey in conjunction with the population census in Albania. See Figure 4.1.

In describing the statistical procedures in this chapter, we will refer primarily to ECHP data.

[6] We have emphasised at various points in this report that "Eurostat Free Dissemination Database" provides a valuable data resource for the construction of regional indicators. This resource or 'window' compiles from diverse databases very detailed tabulations by NUTS regions, which can be utilised in the construction of many of the required indicators of poverty and social exclusion. Please see Chapter 10 for a brief description of its contents.
For convenience, throughout this report we refer to this resource by its older but more widely known name "NewCronos".

Figure 4.1 Survey data available for the present illustrations

Reference year		1993	94	95	96	97	98	99	2000	2001	2002		Area
ECHP wave		1	2	3	4	5	6	7	8				ID
1 DE	Germany	X	X	X	X	X	X	X	X			SOEP, replacing the original ECHP panel from ECHP wave 4	1
2 DK	Denmark	X	X	X	X	X	X	X	X				-
3 NL	Netherlands	X	X	X	X	X	X	X	X			No regional (NUTS) code available in ECHP-UDB data	0
4 BE	Belgium	X	X	X	X	X	X	X	X				1
5 LU	Luxembourg		X	X	X	X	X	X	X			National panel, available from reference year 1994	-
6 FR	France	X	X	X	X	X	X	X	X				1
7 UK	United Kingdom	X	X	X	X	X	X	X	X			BHPS, replacing the original ECHP panel from ECHP wave 4	2
8 IE	Ireland	X	X	X	X	X	X	X	X				0*
9 IT	Italy	X	X	X	X	X	X	X	X			Identification up to NUTS3 available in ECHP-PDB data provided by ISTAT	3
10 GR	Greece	X	X	X	X	X	X	X	X				1
11 ES	Spain	X	X	X	X	X	X	X	X				1
12 PT	Portugal	X	X	X	X	X	X	X	X				2
13 AT	Austria		X	X	X	X	X	X	X			Started from ECHP wave 2	1
14 FI	Finland			X	X	X	X	X	X			Started from ECHP wave 3; no regional (NUTS) code available in data	0
15 SE	Sweden				X	X	X	X	X			Compiled from registers, starting ECHP wave 4 (limited comparability)	0*
23 PL	Poland									X		First of three rounds (other rounds not available or not usable)	2
27 RO	Romania							X				Based on consumption (rather than income); large sample	2
	Albania										X	Living Standard Measurement Survey (LSMS); Population Census	all

Notes on column 'Area ID'

The column indicates whether information is available in the data for the identification of regions

- \- not applicable (no divisions of the country into NUTS regions)
- 0 no area coding in survey data; 0* NUTS1=Country
- 1-3 coding available up to NUTSn regions
- all full idenfication of census enumeration areas

4.2 Consolidation over time

Where the information comes from sample surveys of limited size, as is the case in most Member States lacking income registers, *a trade-off is required between temporal detail and geographical breakdown*. To a large measure, the emphasis has to be shifted away from the study of trends over time and longitudinal measures to essentially *cross-sectional measures aggregated over suitable time periods*, so as to illuminate the more stable aspects of the patterns of variation across regions.

Aggregation of the data over time does not mean that no temporal effects can therefore be examined. The data can still be used to construct, for instance, measures of persistent poverty, long-term unemployment, etc. Also, long-term underlying trends in poverty levels can still be identified, for instance through comparing moving averages.

Panel data

In most of the EU15 surveys, ECHP provides data for 8 waves, covering the reference years 1993-2000 for income and other retrospective measures, or the survey years 1994-2001 for current data. Generally, the waves provide a consistent and comparable series and the *results can be averaged over waves* to increase precision, that is to increase the effective sample size. Of course, the core of the sample is a panel of the same individuals so that the data from different waves are highly correlated. Therefore the effective sample size for estimates averaged over waves is increased less than proportional to the number of waves included. Nevertheless, there is a significant increase in the effective sample size due to real variation in time in the composition of the sample, in characteristics of the individuals and households, and also due to the presence of response variability and other random effects which tend to be balanced out in the averaging process. Also averaging of the results over waves helps to smooth out short-term trends and bring out more clearly the underlying structural relationships, which are of real interest. Hence the measures constructed from averaging over waves tend to be more robust (stable, certainly with less extreme variations) than results based only on one wave. In Chapter 11 we provide quantitative estimates of the gain in precision from averaging of the estimates over waves

Table 4.3 at the end of this chapter shows a sample from a very large set of computations. In each panel of the table, poverty rates computed for individual waves and countries have been consolidated in the margins of the table: EU15 estimates for each wave, and the estimate for each country consolidated over waves. In the calculation of these marginals, a complication arises because of missing cells (no waves for some countries). Because of country and time effects, simply taking averages over available cells is not sufficient. The technical procedure we have used in view of this is described below.

<u>Note</u>

It is worth clarifying that we do not pool microdata across waves or countries for the construction of 'consolidated' or averaged measures. Rather, measures such as the poverty rate must first be computed separately for each country and each wave, on the basis of the specific income distribution. Thereafter, the computed measures are appropriately averaged so as to obtain more stable values. The same applies to 'consolidation over poverty line thresholds' described in Section 4.3 below.

Calculation of marginals

In order to "consolidate" the results over waves, and also over countries to obtain EU15 values, it is necessary to make some adjustments for the fact that not all countries are present in all waves. Taking simple averages over rows and columns of the countryXwave matrix may distort the results because of the missing cells. We have used the following simple adjustment to reduce the effect of any such distortion. Note that the objective was not to impute values for the missing countries individually, but only to calculate the marginal distributions by wave and by country.

We begin by ordering countries according to the number of missing waves so as to get all the missing cells in the left top corner of the table, as shown. Consider a count statistic such as the number of poor persons or the total number of persons. For each non-missing cell (i,j), we compute C_{ij}, the value of the statistic cumulated over all cell to the right of and below this cell and including the cell itself. For waves 4 onwards, data are available for all countries. For wave j<4, let I be the first country in the list for which data are available for this particular wave (in our case, I=5 for j=1; I=3 for j=2; I=2 for j=3). We have imputed the missing value for a cell (i,j), i<I with the reasonable assumption that:

$$\hat{C}_{ij} = \left(\frac{C_{i,4}}{C_{I,4}} \right) . C_{I,j} .$$

Once all the missing cells have been imputed in this way, the required marginals are obtained by "decomulation":

Marginal $c_{.j}$ for waves j=(1-8) over all countries: $c_{.8} = C_{1,8}$; $c_{.j} = C_{1,j} - C_{1,j+1}$.

Marginal $c_{i.}$ for countries j=(1-15) over all waves: $c_{15.} = C_{15,1}$; $c_{i.} = C_{i,1} - C_{i+1,1}$.

For computing a proportion such as a poverty rate, the above procedure is repeated for the count in the numerator and the denominator and then the required proportion is computed. With data from a sample, the "counts" of course refer to appropriately weighted figures.

	j=	1	2	3	4	5	6	7	8
i=		Wave 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8
1	SE	M	M	M					
2	FI	M	M						
3	AT	M							
4	LU	M							
5	DK								
	….								
	….								
15	UK								

4.3 Consolidation over poverty line thresholds

In the standard analysis, as for instance recommended by Eurostat and in Laeken indicators, poverty line is defined as a certain percentage (x%) of the median income of the national population.

By “poverty line threshold” we mean *the choice of different values of ‘x’*.

The Laeken set in fact specifies two related indicators:[7]

[7] Laeken indicators relating to cross-sectional measures of poverty and inequality will be discussed more fully in Chapter 5 below.

Indicator 1: At-risk-of-poverty rate

> This indicator is defined as "the percentage of persons, over the total population, with an equivalised disposable income below the 'at-risk-of-poverty threshold'. The at-risk-of poverty threshold is set at 60% of the national median equivalised disposable income". The indicator is broken down by various characteristics, such as (a) age and gender, (b) most frequent activity status and gender, (c) household type, and (d) accommodation tenure status, in the primary list; and also by (e) gender among workers, (f) work intensity, and as a part of Pension indicators, further by (g) accommodation tenure status and age.

Indicator 11: Dispersion around the at-risk-of-poverty threshold

> This indicator is defined as "the percentage of persons, over the total population, with an equivalised disposable income below 40%, 50% and 70% of the national median equivalised disposable income". It is a 'secondary' indicator, and is broken down by age as a part of Pension indicators.

The substantive objective of introducing indicators of dispersion around the poverty line is to take more fully into accounts differences among countries in the shape at the lower end of the income distribution. Lower thresholds isolate the more severely poor and tend to be more sensitive in distinguishing among countries or other population groups being compared. As the threshold is raised, this sensitivity generally tends to fall: clearly in the extreme case when 'x' is taken as 100% (poverty line equal to the median), the poverty rate in all situations is 50%, by definition!

In addition to the above systematic differences, the results from using different poverty line thresholds are also likely to be affected by irregularities in the empirical income distribution. *It is this consideration which is likely to dominate in the context of constructing regional measures*. Irregularities are larger when the distributions are estimated from smaller samples, as normally is the case for disaggregated estimates by region.

Some gain in sampling precision can be obtained by computing poverty rates using different thresholds (as shown by each row of panels in Table 4.3), and then *taking their weighted average* using some appropriate pre-specified (i.e., constant or external) weights. In Chapter 11, we estimate quantitatively the gains from such consolidation for ECHP samples. This is the strategy we recommend in the construction of regional indicators, in preference to constructing separate indicators for different thresholds. See Chapter 5 for details of the specific application.

4.4 Poverty line levels

By the "level of poverty line" we mean *the population level at which the income distribution is pooled for the purpose of defining the poverty line*.

All poverty related indicators in the Laeken list are based on country poverty lines, meaning that poverty line is always determined on the basis of national income distribution, for instance as 60% of the *national* median income.

Usually, the income distribution is considered separately at the level of each country, in relation to which a poverty line is defined and the number (and proportion) of poor computed. These numbers may then be pooled over countries to obtain the EU poverty rate (but still defined in terms of national poverty lines). Similarly, we may disaggregate the numbers poor according to the national poverty line in each country by region, and obtain regional poverty rates (but again defined in terms of national poverty lines).

It has been argued that even for constructing poverty indicators at the national level, it may be useful to consider in addition poverty lines defined at different levels, such as using a common EU-level poverty line for identifying the poor in each EU country. Especially for defining regional indicators, poverty lines defined at the regional level are also relevant. As Berthoud (2004) puts it: "When national governments measure poverty among their constituent regions, they always use the benchmark derived from the larger (national) unit. Why then should European analysis not use a Europe-wide benchmark? … If the answer to that question is that the concept of exclusion makes sense only in relation to the conditions people see around them, why stop at the country – why not use a regional benchmark? … It would have substantial impact on the estimates of poverty rates in countries with a wide range of inequality between regions." We have already mentioned the paper by Kangas and Ritakallio (2002) which also discusses the issue in some detail.

Hence, as an alternative to using national poverty lines, we may consider the income distribution of, for instance, each NUTS2 region separately, and on this basis define a NUTS2 poverty line (generally differing from one NUTS2 region to another), and determine the number poor and the NUTS2 poverty rate with the poverty line so defined. These numbers can then be aggregated to the national level if desired, to obtain the country poverty rate - but with the poverty line defined in relation to income distribution at NUTS2 level.

The introduction of different poverty line levels may be interpreted as follows. Firstly, it can be seen simply as implying different poverty line thresholds for different areas. Table 4.1 provides an illustration. It shows, for each EU15 country and each NUTS1 region, the implied poverty line threshold when a 70%, 60% or 50% EU15 poverty line is used. Similarly, it also shows for each NUTS1 region the implied poverty line threshold when a 70%, 60% or 50% national poverty line is used. Countries have been arranged according to their mean income level so as to bring out the effect of using the EU poverty line threshold. For instance, a 60% EU poverty line implies national poverty line slightly *above* the median in Portugal, but only a third of the median in Luxembourg.

Different levels for the poverty line can also be seen as implying a different mix of "relative" and "absolute" measures. By *relative measures* we mean those concerning purely the distribution of income, and by *absolute measures* those concerning income levels. For analysis at the country level, the use of national poverty lines provides a relative measure for each county, but the use of a EU poverty line introduces quite a high degree of absoluteness into the measure.

Considering analysis at NUTS2 level, the use of a NUTS2 poverty line provides a relative measure of poverty determined only by the income distribution within the region, independently of the degree of regional disparities in the country. Use of NUTS1 poverty lines introduces an element of "absoluteness" in the sense defined, since the resulting poverty rate in a NUTS2 region now also depends on differences in income levels among NUTS2 regions in the same NUTS1 region. The degree of absoluteness in the measure increases as the poverty line level is raised to country and then to EU level – meaning that increasingly the resulting poverty rates reflects not only the extent of disparity within the region but also the level of its mean income.

In fact we can mix any level of analysis or aggregation with any poverty line level. The former concerns the units for which the measures are computed; the latter refers to the population of which the income distribution is considered in defining the poverty line:

level of analysis/aggregation		**poverty line level**
EU		EU
Country		Country
NUTS1	⟷	NUTS1
NUTS2		NUTS2
NUTS3		NUTS3 ??

The poverty line level chosen can make a major difference to the resulting poverty rates when it is higher than the level of analysis or aggregation. The extent depends on the degree of disparity between the units of analysis. However, we find that the poverty line level chosen often makes only a small difference to the resulting poverty rates when it is the *same as or lower* than the level of analysis or aggregation. For instance, as can be seen from the results in Table 4.3 at the end of this section, while country poverty rates can differ greatly when a EU poverty line is used, the country rates generally differ only marginally whether we use a poverty line defined at the national, NUTS1 or NUTS2 level.

Table 4.1 Poverty line as percentage of the region's median equivalised income (PPS). NUTS1 regions

	Median Equivalised income (PPS)	EU15 Poverty Line			Country Poverty line		
		70%	60%	50%	70%	60%	50%
EU	11,444	70	60	50			
PT	6,737	119	102	85	70	60	50
PT1	6,818	117	101	84	69	59	49
PT2	4,994	160	137	115	94	81	67
PT3	5,269	152	130	109	89	77	64
GR	7,412	108	93	77	70	60	50
GR1	6,293	127	109	91	82	71	59
GR2	6,363	126	108	90	82	70	58
GR3	9,058	88	76	63	57	49	41
GR4	7,278	110	94	79	71	61	51
ES	8,280	97	83	69	70	60	50
ES1	7,745	103	89	74	75	64	53
ES2	9,535	84	72	60	61	52	43
ES3	11,628	69	59	49	50	43	36
ES4	6,780	118	101	84	85	73	61
ES5	9,376	85	73	61	62	53	44
ES6	6,696	120	103	85	87	74	62
ES7	6,411	125	107	89	90	77	65
IT	9,764	82	70	59	70	60	50
IT1	11,612	69	59	49	59	50	42
IT2	12,905	62	53	44	53	45	38
IT3	11,515	70	60	50	59	51	42
IT4	12,598	64	55	45	54	47	39
IT5	10,996	73	62	52	62	53	44
IT6	9,648	83	71	59	71	61	51
IT7	8,932	90	77	64	77	66	55
IT8	7,492	107	92	76	91	78	65
IT9	7,034	114	98	81	97	83	69
ITA	6,925	116	99	83	99	85	70
ITB	7,146	112	96	80	96	82	68
IE	10,058	80	68	57	70	60	50
FI	11,096	72	62	52	70	60	50
SE	11,389	70	60	50	70	60	50
NL	11,990	67	57	48	70	60	50
FR	12,208	66	56	47	70	60	50
FR1	15,402	52	45	37	55	48	40
FR2	11,875	67	58	48	72	62	51
FR3	10,363	77	66	55	82	71	59
FR4	12,422	64	55	46	69	59	49
FR5	11,272	71	61	51	76	65	54
FR6	11,308	71	61	51	76	65	54
FR7	12,112	66	57	47	71	60	50
FR8	11,588	69	59	49	74	63	53

	Median Equivalised income (PPS)	EU15 Poverty Line			Country Poverty line		
		70%	60%	50%	70%	60%	50%
UK	12,213	66	56	47	70	60	50
UK1	11,929	67	58	48	72	61	51
UK2	11,165	72	61	51	77	66	55
UK3	11,048	73	62	52	77	66	55
UK4	11,531	69	60	50	74	64	53
UK5	13,652	59	50	42	63	54	45
UK6	12,127	66	57	47	70	60	50
UK7	11,446	70	60	50	75	64	53
UK8	12,346	65	56	46	69	59	49
UK9	11,335	71	61	50	75	65	54
UKA	11,482	70	60	50	74	64	53
DE	13,117	61	52	44	70	60	50
DE1	13,899	58	49	41	66	57	47
DE2	14,030	57	49	41	65	56	47
DE3	14,226	56	48	40	65	55	46
DE4	11,905	67	58	48	77	66	55
DE5	12,475	64	55	46	74	63	53
DE6	13,799	58	50	41	67	57	48
DE7	14,337	56	48	40	64	55	46
DE8	11,822	68	58	48	78	67	55
DE9	13,127	61	52	44	70	60	50
DEA	13,383	60	51	43	69	59	49
DED	11,595	69	59	49	79	68	57
DEE	12,261	65	56	47	75	64	53
DEF	12,959	62	53	44	71	61	51
DEG	11,167	72	61	51	82	70	59
DEX	12,578	64	55	45	73	63	52
AT	13,403	60	51	43	70	60	50
AT1	14,043	57	49	41	67	57	48
AT2	12,566	64	55	46	75	64	53
AT3	13,262	60	52	43	71	61	51
BE	13,476	59	51	42	70	60	50
BE1	14,500	55	47	39	65	56	46
BE2	13,571	59	51	42	70	60	50
BE3	13,000	62	53	44	73	62	52
DK	14,196	56	48	40	70	60	50
LU	20,518	39	33	28	70	60	50

It is important to clarify that while we recommend (and have implemented) consolidation over waves and poverty line thresholds for the purpose of increasing sampling precision of the estimates, such consolidation or averaging is not meaningful over poverty line levels. This is because different poverty line levels capture different aspects of the situation from absolute to purely relative aspects, and help to separate out within and between regional variations. It is best, therefore, to keep them separate, each defining a different indicator of poverty.

4.5 Aggregation over similar measures

"Aggregation over similar measures" is a very important element of our research strategy in constructing regional indicators of poverty and social exclusion at the level of regions such as NUTS2, or even NUTS3.

Need for combining individual items of information

Generally, the indicators identified hitherto concentrate on the national level, at least in geographical terms. Disaggregation considered in the Laeken indicators mostly concerns population subgroups dispersed in the population such as demographic classes by age and gender, subgroups according to the type of economic activity, household type, and so on.

Also, the indicators refer to very specific measures, such as poverty rates computed at certain fixed level, e.g. 60% of the national median income. While this kind of specificity may be desirable and practical for indicators at the national level, in view of the severe sample size limitations, it is our general recommendation to *replace these by suitable combinations of whole sets of similar indicators*. This is the response to the objective mentioned in the project Terms of Reference, namely to "assess the extent to which indicators of social exclusion and poverty of the type endorsed at Laeken can be applied at the regional level, either using EU-wide or national sources".

This point may be illustrated by the example of computing the conventional poverty rate or HCR, as discussed above and also further in Chapter 5. In place of simply computing HCR in terms of single poverty line 'level' (Country, or NUTS1, or NUTS2, or even EU), and of a single poverty line 'threshold' (e.g., 40%, or 50%, or 60%, or 70% of the median income at that 'level'), we can compute the statistic for 4x4=16 combinations of all these possibilities. Furthermore, repetition over 8 waves gives well over 100 different estimates.

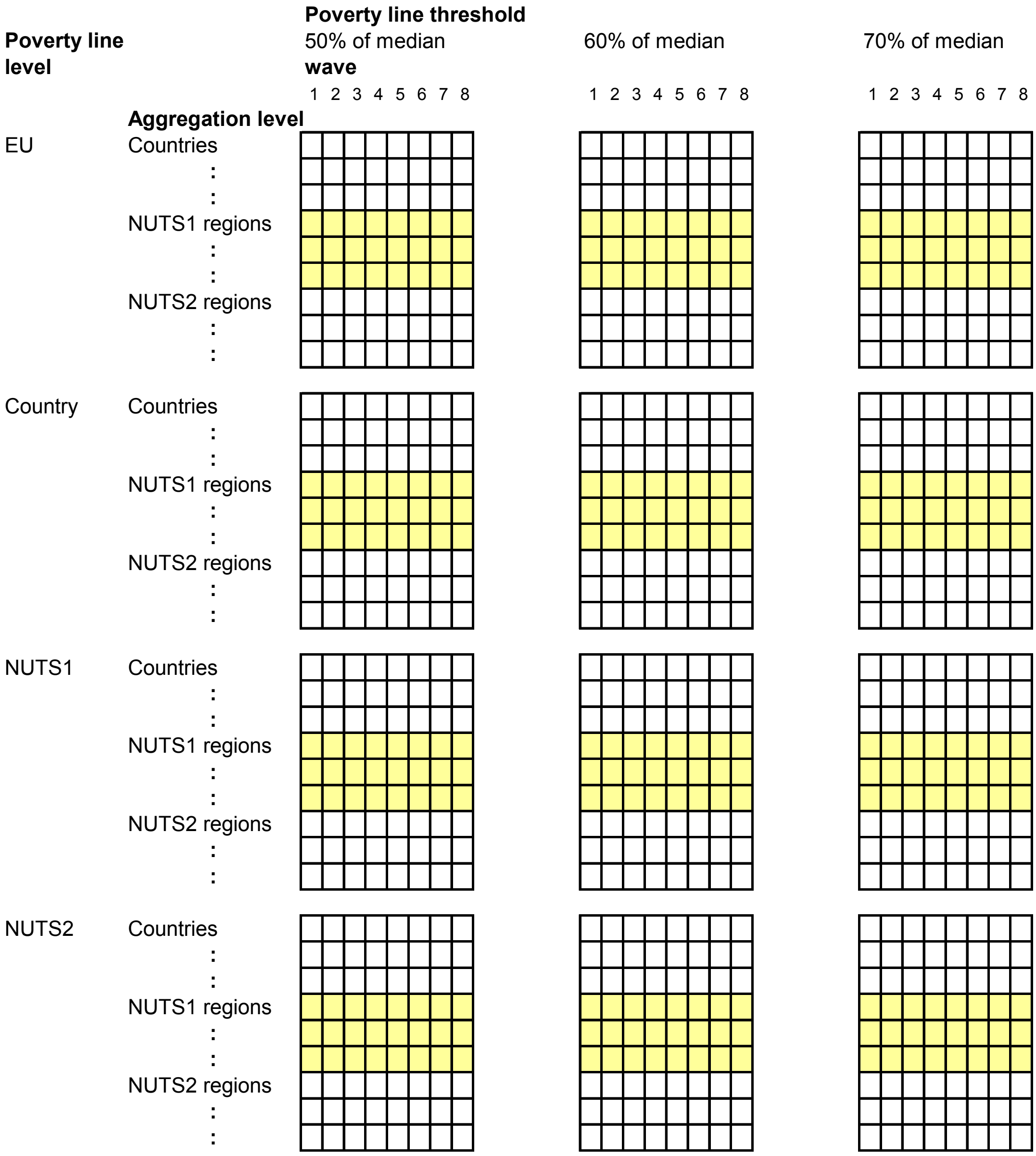
Each of these estimates is then produced for *different levels of aggregation*, such as individual countries, individual NUTS1 regions, individual NUTS2 regions, etc, as the units. The matrix of computations we have performed is indicated schematically in Figure 4.2.

Some combinations, such as an appropriately weighted average, of these different estimates provides greater stability. Individual estimates based on small samples are often subject to large fluctuations due to random or irregular variations in the income distribution, such as near the chosen poverty line.

The possibility of *appropriately averaging over a range of similar measures* is a very important consideration in the choice and construction of regional indicators.

The argument can also be extended to averaging over different types of statistics in so far as they measure the same or similar phenomenon, such as poverty, or inequality, or income level. Such averaging will require appropriate rescaling of the different types of measures before they can be pooled together to provide a single 'consolidated' measure.

Figure 4.2 Structure of computations
- by poverty line level, poverty line threshold, aggregation level, and wave



Working with ranks

In comparative analysis, it is often useful to work with rankings rather than actual values of the indicators being compared. Though rankings generally contain less information than actual numerical values of the estimates, the former have the advantage of providing an automatic means of handling extreme or very different values, of spreading out the estimates more uniformly, and also the advantage of reducing different types of measures to the same scale so as to permit their aggregation or averaging in a straightforward way.

Indicators of consistency or otherwise of different measures can be constructed using rank correlation techniques, as described below. Highly correlated ranks for two measures may indicate that the measures reflect essentially the same underlying dimension in different forms – hence justifying the pooling of the two measures, such as by averaging the two rankings. Low rank correlations would indicate that different measures reflect different dimensions – hence information would be lost in pooling them, or at least that pooling will produce a new type of measure different from either of the measures pooled.

Some researchers have argued against the use of ranking rather than of actual numerical values in the construction of indicators (see for instance, Atkinson *et al.*, 2002). It is true that rank ordering often contains less information than actual values – though this is not inevitably the case. In any case, perhaps the main reason for aversion to the use of ranks is that ranking tends to lead to a "league table" which are unfair, misleading and at best largely meaningless, and which make units (countries, regions, schools or whatever) compete rather than co-operate. It is also asserted that rankings can provide no indication of the performance or changes therein of a group as a whole since, by definition, for a group of size n the average rank is around (n/2) and therefore can never be improved. In our view such arguments are mostly not valid. *All* measures of inequality whether in the condition or the performance of units - including the much used poverty rates - tell us nothing about absolute levels. They are not meant to do so; indeed, to the extent they mix absolute and relative dimensions (e.g., income inequality and actual levels of income), they become less satisfactory measures of disparity.

As noted, with limited sample sizes available for the purpose of regional estimation, individual items of information are likely to be subject to large uncertainties and errors. In such situations, statistically it is better to seek consolidated measures which appropriately put together individual items of information into fewer and more robust indicators. Various statistical techniques such as factor and cluster analyses may be used for this purpose, but often putting together of rankings in various dimensions provides a convenient, effective and acceptable means for the same purpose. Of course, as a rule one should avoid putting together indicators which rank the population very differently.

In putting together data over survey waves and over different poverty line thresholds, we have verified consistency of the rank orderings across the measures. This is indicated by computing rank correlations among them. The technical procedure for rank correlations is described below briefly (Kendall, 1970).

Spearman's rank correlation

Let *i* be the rank (1 to n) of units (regions) according to some standard criterion, and suppose the objective is to compare the overall consistency of ranking of those units according to each of a set of measures (j) against the standard ranking. Let r_{ij} be the rank of unit *i* (i.e. the unit ranked *i* according to the standard) according to criterion j. Define:

$$d_j^2 = \Sigma_i w_i.(r_{ij} - i)^2 / \Sigma_i w_i$$

$$d_{max}^2 = \Sigma_i w_i.(n + 1 - 2 * i)^2 / \Sigma_i w_i$$

where w_i is the relative weight given to unit *i*, such as a function of its population size.

Spearman's rank correlation is given by:

$$\rho_j = 1 - \left(d_j^2 / d_{max}^2 \right).$$

Kendall's rank correlation

With the same notation as above, define:

$$d_{ij} = \text{count}\left(r_{kj} > r_{ij} : k < i \right).$$

Kendall's rank correlation is given by:

$$\tau_j = 1 - \frac{2}{n-1} . \left\{ \Sigma_i w_i . \left(2 * d_{ij} + r_{ij} - i \right) / \Sigma_i w_i \right\}.$$

An example

Table 4.2 provides an illustrative example of the usefulness of these rank correlation measures. Units of analysis are the lowest level (NUTS2 or NUTS1) of regions identifiable in our ECHP data. Columns of the table correspond to different indicators being compared. Indicators 1-9 are poverty rates computed with different poverty line levels and thresholds as specified, and indicators 10-11 are regional median and mean income levels. The distribution of regions according to each of these measures is compared with three distributions taken as the standard in turn. These are shown in rows of the table, and correspond respectively to (A) the mean of ranks corresponding to measures 1-3; (C) the mean of ranks corresponding to measures 10-11; and (B) the mean of the above two. Hence (A) represents a more less purely relative standard (indicating income inequality within each region; (C) corresponds to an absolute standard (indicating differences in income levels among the regions); while (B) reflects the two aspects in combination. The rank correlations computed using the above procedures show how the poverty measures become more absolute as we raise the level at which the poverty line is defined.

Table 4.2 Illustrative example:
Rank correlation between measures of poverty and low mean income level

	Poverty measure (HCR)									Income Level	
Poverty Line Level	NUTS2/NUT1			Country			EU			median	mean
Threshold (%of median)	50%	60%	70%	50%	60%	70%	50%	60%	70%		
	1	2	3	4	5	6	7	8	9	10	11
Spearman rank correlation											
Rank correlation with rank distribution according to											
(A) 'Poverty'. HCR at NUTS2 (av. of cols. 1-3)	0,95	0,99	0,96	0,66	0,61	0,52	0,73	0,72	0,66	0,41	0,34
(B) 'Hardship' (average of rows (1) and (3))	0,79	0,80	0,79	0,78	0,76	0,71	0,96	0,97	0,95	0,83	0,79
(C) Low Income Level (av. of cols.10-11)	0,38	0,35	0,35	0,65	0,66	0,67	0,84	0,87	0,90	0,98	0,98
Kendall rank correlation											
Rank correlation with rank distribution according to											
(A) 'Poverty'. HCR at NUTS2 (av. of cols. 1-3)	0,81	0,91	0,83	0,48	0,43	0,35	0,54	0,53	0,48	0,28	0,24
(B) 'Hardship' (average of rows (1) and (3))	0,60	0,60	0,58	0,61	0,59	0,53	0,83	0,85	0,81	0,65	0,60
(C) Low Income Level (av. of cols.10-11)	0,27	0,24	0,24	0,48	0,49	0,51	0,68	0,71	0,76	0,90	0,91

Results avereged over waves 1-7 of ECHP

Table shows correlation of ranking according to each column with the ranking according to averaged criteria defined in rows (A)-(C).

Table 4.3 Illustration of detailed computations of poverty rates:
for different waves, poverty line thresholds and poverty line levels, with aggregation to the country level.

Poverty line level: COUNTRY

Poverty line threshold	70%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.25	0.25	0.24	0.24	0.23	0.23	0.23	0.23	0.24
PT	0.29	0.29	0.30	0.28	0.28	0.28	0.28	0.28	0.29
GR	0.30	0.29	0.28	0.29	0.28	0.28	0.27	0.28	0.28
ES	0.27	0.27	0.25	0.27	0.26	0.26	0.25	0.27	0.26
IT	0.28	0.28	0.27	0.27	0.26	0.26	0.27	0.28	0.27
IE	0.26	0.28	0.28	0.27	0.28	0.28	0.28	0.29	0.28
FI			0.16	0.17	0.18	0.19	0.20	0.20	0.18
SE				0.17	0.18	0.17	0.18	0.18	0.18
NL	0.20	0.20	0.20	0.19	0.18	0.18	0.18	0.19	0.19
FR	0.25	0.24	0.24	0.23	0.23	0.24	0.24	0.23	0.24
UK	0.27	0.27	0.27	0.27	0.27	0.27	0.26	0.25	0.27
DE	0.21	0.21	0.21	0.19	0.18	0.17	0.17	0.18	0.19
AT		0.21	0.21	0.21	0.21	0.20	0.19	0.19	0.21
DK	0.18	0.17	0.17	0.17	0.19	0.19	0.17	0.19	0.18
BE	0.25	0.25	0.23	0.23	0.22	0.22	0.21	0.21	0.23
LU		0.21	0.20	0.20	0.20	0.22	0.20	0.21	0.21

Poverty line threshold	60%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.17	0.17	0.17	0.16	0.15	0.15	0.15	0.15	**0.16**
PT	0.23	0.23	0.21	0.22	0.21	0.21	0.21	0.20	0.21
GR	0.23	0.22	0.21	0.22	0.21	0.21	0.20	0.20	0.21
ES	0.20	0.19	0.18	0.20	0.18	0.19	0.18	0.19	0.19
IT	0.20	0.20	0.20	0.20	0.18	0.18	0.18	0.19	0.19
IE	0.17	0.19	0.19	0.19	0.19	0.18	0.20	0.21	0.19
FI			0.08	0.08	0.09	0.11	0.11	0.11	0.10
SE				0.09	0.10	0.10	0.11	0.10	0.10
NL	0.10	0.11	0.12	0.10	0.10	0.11	0.10	0.11	0.11
FR	0.17	0.15	0.15	0.15	0.15	0.15	0.16	0.15	0.15
UK	0.19	0.20	0.19	0.18	0.19	0.19	0.19	0.17	0.19
DE	0.14	0.15	0.14	0.12	0.11	0.11	0.11	0.11	0.12
AT		0.13	0.14	0.13	0.13	0.12	0.12	0.12	0.13
DK	0.10	0.10	0.09	0.09	0.12	0.11	0.11	0.11	0.11
BE	0.17	0.16	0.15	0.14	0.14	0.13	0.13	0.13	0.14
LU		0.13	0.12	0.11	0.12	0.13	0.12	0.12	0.12

Poverty line threshold	50%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.11	0.11	0.10	0.10	0.10	0.09	0.09	0.09	0.10
PT	0.17	0.16	0.14	0.14	0.14	0.13	0.14	0.13	0.14
GR	0.17	0.16	0.14	0.16	0.15	0.14	0.14	0.14	0.15
ES	0.13	0.12	0.12	0.14	0.12	0.13	0.12	0.13	0.12
IT	0.14	0.14	0.13	0.13	0.12	0.12	0.12	0.13	0.13
IE	0.05	0.07	0.08	0.09	0.10	0.11	0.13	0.15	0.10
FI			0.04	0.04	0.05	0.05	0.05	0.06	0.05
SE				0.05	0.06	0.05	0.07	0.06	0.06
NL	0.06	0.07	0.07	0.06	0.06	0.06	0.05	0.06	0.06
FR	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.09	0.09
UK	0.14	0.13	0.12	0.10	0.12	0.12	0.11	0.10	0.12
DE	0.10	0.10	0.08	0.07	0.07	0.06	0.06	0.06	0.07
AT		0.07	0.07	0.07	0.07	0.07	0.05	0.06	0.07
DK	0.04	0.04	0.04	0.04	0.06	0.06	0.05	0.07	0.05
BE	0.10	0.09	0.08	0.08	0.08	0.07	0.07	0.06	0.08
LU		0.07	0.05	0.05	0.06	0.06	0.06	0.06	0.06

Poverty line level: EU15

Poverty line threshold	70%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.27	0.27	0.27	0.26	0.26	0.26	0.25	0.25	0.26
PT	0.64	0.61	0.62	0.64	0.64	0.63	0.61	0.60	0.62
GR	0.58	0.56	0.55	0.56	0.54	0.55	0.52	0.55	0.55
ES	0.47	0.50	0.50	0.50	0.50	0.47	0.44	0.43	0.48
IT	0.38	0.37	0.37	0.38	0.38	0.35	0.34	0.37	0.37
IE	0.45	0.39	0.36	0.36	0.32	0.32	0.34	0.30	0.35
FI			0.18	0.19	0.22	0.23	0.25	0.25	0.23
SE				0.18	0.21	0.21	0.24	0.23	0.23
NL	0.16	0.18	0.18	0.16	0.14	0.13	0.15	0.19	0.16
FR	0.21	0.19	0.19	0.21	0.19	0.19	0.22	0.20	0.20
UK	0.24	0.24	0.25	0.21	0.23	0.25	0.24	0.21	0.23
DE	0.14	0.16	0.14	0.12	0.12	0.12	0.11	0.12	0.13
AT		0.12	0.13	0.13	0.15	0.14	0.12	0.12	0.13
DK	0.10	0.08	0.08	0.07	0.09	0.08	0.09	0.10	0.09
BE	0.14	0.14	0.14	0.13	0.13	0.12	0.15	0.16	0.14
LU		0.02	0.01	0.01	0.02	0.02	0.01	0.03	0.02

Poverty line threshold	60%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.20	0.20	0.19	0.19	0.19	0.18	0.18	0.18	**0.19**
PT	0.53	0.51	0.51	0.51	0.52	0.50	0.51	0.50	0.51
GR	0.47	0.46	0.46	0.46	0.44	0.44	0.42	0.45	0.45
ES	0.37	0.39	0.39	0.40	0.39	0.37	0.33	0.32	0.37
IT	0.28	0.28	0.28	0.29	0.28	0.25	0.25	0.28	0.27
IE	0.34	0.29	0.27	0.27	0.23	0.24	0.25	0.23	0.27
FI			0.09	0.10	0.12	0.14	0.16	0.16	0.13
SE			0.00	0.10	0.12	0.12	0.15	0.14	0.13
NL	0.08	0.10	0.11	0.09	0.08	0.08	0.08	0.11	0.09
FR	0.14	0.12	0.12	0.13	0.12	0.12	0.14	0.13	0.13
UK	0.18	0.16	0.18	0.14	0.15	0.17	0.16	0.13	0.16
DE	0.10	0.11	0.08	0.08	0.08	0.07	0.07	0.07	0.08
AT		0.07	0.07	0.08	0.08	0.08	0.06	0.07	0.08
DK	0.05	0.04	0.04	0.03	0.04	0.04	0.05	0.06	0.04
BE	0.08	0.09	0.08	0.08	0.08	0.07	0.09	0.08	0.08
LU		0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01

Poverty line threshold	50%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0.14	0.13	0.12	0.12	0.12	0.12	0.11	0.11	0.12
PT	0.40	0.39	0.40	0.39	0.40	0.38	0.37	0.38	0.39
GR	0.37	0.35	0.33	0.35	0.33	0.33	0.30	0.33	0.34
ES	0.26	0.28	0.26	0.28	0.28	0.25	0.23	0.21	0.26
IT	0.20	0.19	0.19	0.19	0.18	0.17	0.16	0.19	0.18
IE	0.22	0.18	0.16	0.17	0.14	0.15	0.17	0.16	0.17
FI			0.04	0.04	0.06	0.06	0.08	0.08	0.06
SE				0.05	0.07	0.06	0.08	0.07	0.07
NL	0.05	0.07	0.07	0.04	0.05	0.05	0.04	0.06	0.05
FR	0.08	0.06	0.06	0.08	0.07	0.06	0.08	0.07	0.07
UK	0.12	0.10	0.11	0.08	0.09	0.11	0.10	0.08	0.10
DE	0.07	0.08	0.05	0.05	0.05	0.04	0.04	0.04	0.05
AT		0.05	0.04	0.04	0.05	0.05	0.03	0.04	0.04
DK	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.03	0.02
BE	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
LU		0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.00

Table 4.3 (cont.)

Poverty line level	NUTS1																										
Poverty line threshold	70%median									60%median									50%median								
	w1	w2	w3	w4	w5	w6	w7	w8	w1-8	w1	w2	w3	w4	w5	w6	w7	w8	w1-8	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	0,24	0,24	0,23	0,23	0,23	0,22	0,22	0,22	0,23	0,17	0,16	0,16	0,15	0,15	0,15	0,14	0,14	**0,15**	0,11	0,10	0,09	0,09	0,09	0,09	0,08	0,08	0,09
PT	0,29	0,29	0,30	0,29	0,28	0,28	0,27	0,28	0,28	0,22	0,23	0,21	0,22	0,20	0,20	0,20	0,20	0,21	0,17	0,16	0,14	0,15	0,14	0,13	0,13	0,13	0,14
GR	0,29	0,28	0,27	0,28	0,27	0,25	0,26	0,27	0,27	0,22	0,21	0,20	0,21	0,21	0,18	0,19	0,19	0,20	0,16	0,14	0,14	0,14	0,14	0,12	0,12	0,12	0,14
ES	0,26	0,25	0,24	0,26	0,25	0,26	0,25	0,26	0,25	0,17	0,17	0,17	0,18	0,17	0,19	0,17	0,19	0,17	0,11	0,11	0,11	0,12	0,11	0,12	0,10	0,11	0,11
IT	0,26	0,26	0,26	0,26	0,24	0,23	0,22	0,23	0,25	0,19	0,18	0,18	0,18	0,16	0,16	0,15	0,16	0,17	0,12	0,12	0,11	0,11	0,11	0,09	0,09	0,09	0,11
IE	(not applicable)																										
FI																											
SE	(not applicable)																										
NL																											
FR	0,24	0,23	0,22	0,22	0,22	0,23	0,23	0,22	0,23	0,16	0,15	0,14	0,14	0,14	0,14	0,15	0,15	0,15	0,10	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08
UK	0,27	0,27	0,27	0,27	0,28	0,26	0,25	0,25	0,27	0,20	0,19	0,19	0,18	0,19	0,18	0,18	0,16	0,18	0,14	0,12	0,11	0,11	0,12	0,11	0,11	0,10	0,12
DE	0,21	0,22	0,20	0,19	0,19	0,17	0,17	0,18	0,19	0,15	0,15	0,14	0,12	0,11	0,11	0,10	0,10	0,12	0,10	0,10	0,08	0,07	0,07	0,06	0,06	0,06	0,07
AT		0,21	0,21	0,22	0,21	0,21	0,20	0,19	0,21		0,14	0,13	0,13	0,13	0,12	0,12	0,11	0,13		0,08	0,07	0,07	0,07	0,06	0,05	0,06	0,07
DK	(not applicable)																										
BE	0,25	0,24	0,23	0,23	0,22	0,22	0,21	0,21	0,23	0,17	0,16	0,15	0,14	0,14	0,13	0,13	0,13	0,14	0,10	0,10	0,08	0,08	0,08	0,07	0,06	0,06	0,08
LU	(not applicable)																										

Notes NUTS1=Country: DK, LU, IE, SW Data for NUTS1 identification not available NL, FI EU15 figures based on NL, FI taken as single units

Poverty line level	NUTS2																										
Poverty line threshold	70%median									60%median									50%median								
Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8	w1	w2	w3	w4	w5	w6	w7	w8	w1-8	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15																											
PT	0,28	0,29	0,29	0,27	0,29	0,27	0,29	0,28	0,28	0,21	0,22	0,21	0,21	0,20	0,19	0,22	0,21	**0,21**	0,16	0,15	0,14	0,14	0,12	0,13	0,14	0,13	0,14
GR																											
ES																											
IT	0,25	0,24	0,23	0,24	0,23	0,22	0,21	0,22	0,23	0,18	0,17	0,16	0,17	0,15	0,15	0,14	0,15	**0,16**	0,12	0,11	0,10	0,11	0,10	0,09	0,09	0,09	0,10
IE																											
FI																											
SE																											
NL																											
FR																											
UK	0,26	0,27	0,27	0,27	0,28	0,26	0,25	0,25	0,26	0,20	0,19	0,19	0,17	0,19	0,18	0,18	0,16	**0,18**	0,14	0,12	0,11	0,10	0,12	0,11	0,11	0,10	0,11
DE																											
AT																											
DK	(not applicable)																										
BE																											
LU	(not applicable)																										

Notes NUTS2=Country: DK, LU ECHP data for NUTS2 identification available only for: UK, IT, PT EU15 figures with NUTS2 poverty line cannot be computed

Chapter 5
Regional indicators of income poverty

5.1 Choice of indicators

For the development of monitory indicators of poverty for use at the regional level, the starting point of course is the specified set of Laeken indicators in this area. These indicators have been summarised in Chapter 2 with some necessary technical detail.

Henceforth these indicators have been applied at the national level. It is necessary to adapt them for regional application, taking into account any differences in the requirements, but equally important, differences in the practical situation. As in the case of regional adaptation of all other indicators, it is necessary to *focus on the more basic among this set of indicators*. This is because of the substantially increased data requirements when the results have to be geographically disaggregated.

Recommendations on the set of indicators on cross-sectional income poverty have been presented in Section 2.3.

5.2 Income poverty rate

The emphasis in the following is on the most important basic indicator of income poverty, namely the "at-risk-of-poverty rate", "Head Count Ratio" (HCR), or simply the *poverty rate*. In the context of regional analysis, an almost equally important measure is the *mean level of income*, since variations in mean income across regions is itself an important indicator of regional disparities. The following should also be noted, however.

- Firstly, for the construction of consolidated indicators, it is desirable that these results are supplemented by and integrated with results on other measures of income poverty and inequality, as discussed in Section 2.3.
- Secondly, these should be supplemented by a host of non-monetary indicators, such as those described in Chapter 6 below.
- Thirdly, of course, any results based on small samples must be combined with and modelled on the basis of statistical relationships with other variables which are available with sufficient reliability on the basis of large samples or on a complete basis from censuses and administrative sources.

We begin by presentation of some results on HCR computed on the basis of ECHP data.[8]

As in the case of all the results presented in other sections of this report, these are based on an extremely large number of computations. To convey an impression of that volume, some of the results in this sections are shown in detail. These are then used to construct *consolidated measures* of the type we recommend for regional analysis. Even so, it should be noted that this is the first step in the process of constructing reliable regional indicators of poverty and social exclusion.

---

[8] Single wave data for Poland and Romania have also been analysed. The measures in the latter survey are based on consumption, rather than income.

As explained in the preceding chapter, we first have to compute poverty rates separately taking into account four dimensions (Figure 4.2): survey waves; poverty line thresholds; poverty line levels i.e. the level at which the income distribution is considered (EU, country, NUTS1, NUTS2); and the level of aggregation, i.e. the level (EU, country, NUTS1, NUTS2 etc.) at which the results are reported.

Table 5.1 presents just a small part of this set of computations. It shows the "standard" output for national level analysis: poverty rates based on country level poverty lines as 60% of the national median, by country and wave. All computations are performed in terms of individual persons, with the total household annual income equivalised using the modified-OECD scale, and the equivalised income ascribed to each member of the household. The recommended Eurostat methodology is used throughout.

Table 5.1 Conventional poverty rate (Head Count Ratio, HCR): poverty line as 60% of national median

Wave	w1	w2	w3	w4	w5	w6	w7	w8	w1-8
EU15	17,1	17,0	16,5	15,9	15,4	15,5	15,4	15,4	**16,0**
DE	14,4	14,5	14,0	12,0	11,2	10,9	10,7	11,0	12,4
DK	10,3	10,2	9,4	9,3	11,8	11,1	11,5	10,8	10,6
NL	10,0	11,3	11,7	10,5	10,3	10,7	10,4	11,3	10,8
BE	16,7	15,9	15,3	14,2	13,8	12,7	13,0	13,4	14,4
LU		13,2	11,8	11,4	12,2	12,7	11,9	12,5	12,4
FR	16,5	15,5	15,2	14,9	14,7	15,2	15,6	15,5	15,4
UK	19,5	19,6	19,4	17,7	19,1	19,4	18,5	16,5	18,7
IE	16,9	18,7	19,3	19,3	19,1	18,4	20,2	21,4	19,2
IT	20,4	20,4	20,1	19,7	18,0	18,0	18,5	19,3	19,3
GR	23,1	21,6	21,0	21,6	20,8	20,5	19,8	20,4	21,1
ES	19,6	19,0	18,0	20,4	18,2	18,9	18,0	18,8	18,8
PT	22,5	22,9	21,1	21,7	20,8	20,5	20,8	20,1	21,3
AT		13,4	14,0	13,0	12,9	12,2	11,5	12,0	12,8
FI			8,1	8,3	9,4	10,7	10,9	11,4	10,0
SE				9,0	10,1	9,6	11,0	10,4	10,5

Consolidation of the measures across waves

The first step is the consolidation of the results over waves. Simple average of wave-specific poverty rates over waves is taken as an indicator reflecting the *overall situation over the period covered*.[9] This is necessary because of small sample sizes for regional estimation. Generally very high values of the rank correlations between income distributions at different waves provide a justification of such averaging. (Some statistical adjustment is required in such averaging since data for some countries are missing for one or more waves at the beginning of the panel. The procedure used has been described in Section 4.2.)

Measures consolidated over waves are shown in Table 5.1 in the columns marked "W1-8".

[9] Actually the procedure is a little more complicated than taking simple averages. It is necessary to make some adjustments for the fact that not all countries are present in all waves (see Section 4.2).

Consolidation over poverty line thresholds

By Poverty line thresholds we mean different percentages of the median income taken to be the poverty line. Three commonly referred to threshold have been used in our illustrations:[10]

- Poverty line as 50% of the median income
- Poverty line as 60% of the median income
- Poverty line as 70% of the median income

Lower the threshold, sharper are likely to be the differentials between countries and between regions.

The actual estimates of course vary considerably among the three chosen thresholds (50%, 60% and 70% of the median), but the pattern of variation is quite consistent across domains (countries and regions). This can be checked by using the rank correlation procedure as explained in Section 4.5.

Consequently, it is reasonable to consolidate the numerical results across poverty line thresholds to impart them more stability. However, in place of taking the simple average of figures for the three thresholds, we propose a weighted average on the basis of the following considerations.

- Lower thresholds are more distinguishing among domains, as noted above, but the HCR values corresponding to them are numerically lower. It is therefore desirable to increase the weight given to them.
- At the EU level the available sample sizes are very large, and therefore consolidation across thresholds is hardly necessary. It is desirable to define the consolidated measure such that at EU level it exactly coincides with the standard threshold of 60% of the median.[11]
- The weights given to rates for different poverty line thresholds are determined by these rates at the EU level, which are based on a sample of very large size. Hence the weights can be considered as constants, essentially free of sampling error. In fact, the gain in precision from averaging over poverty line thresholds arises from the last-mentioned consideration.

Consider a particular domain (D), such as a country or a region. Head count ratios are computed for the domain using different poverty line thresholds – $H_{D,50}$, $H_{D,60}$ and $H_{D,70}$ with poverty lines defined as, respectively, 50%, 60% and 70% of the medina income. We define the consolidated head-count ratio $H^{(D)}$ for a particular domain (D) as

$$H^{(D)} = H_{D,50} w_{EU,50} + H_{D,60} w_{EU,60} + H_{D,70} w_{EU,70},$$

where the weights do not change over domains but are determined (separately for each poverty line level, EU, Country, NUTS1, NUTS2) by estimates of head-count ratio for EU:

[10] Even though Laeken Indicator 11 refers also to 40% of the median, we have not included this threshold in the consolidation because of its values are often small and lack stability, and would receive rather large weights according to the consolidation procedure we have used.

[11] Since most of the detailed survey data available are for EU15 countries, we have used the procedure described here only with reference to EU15. In principle it can be extended to all EU.

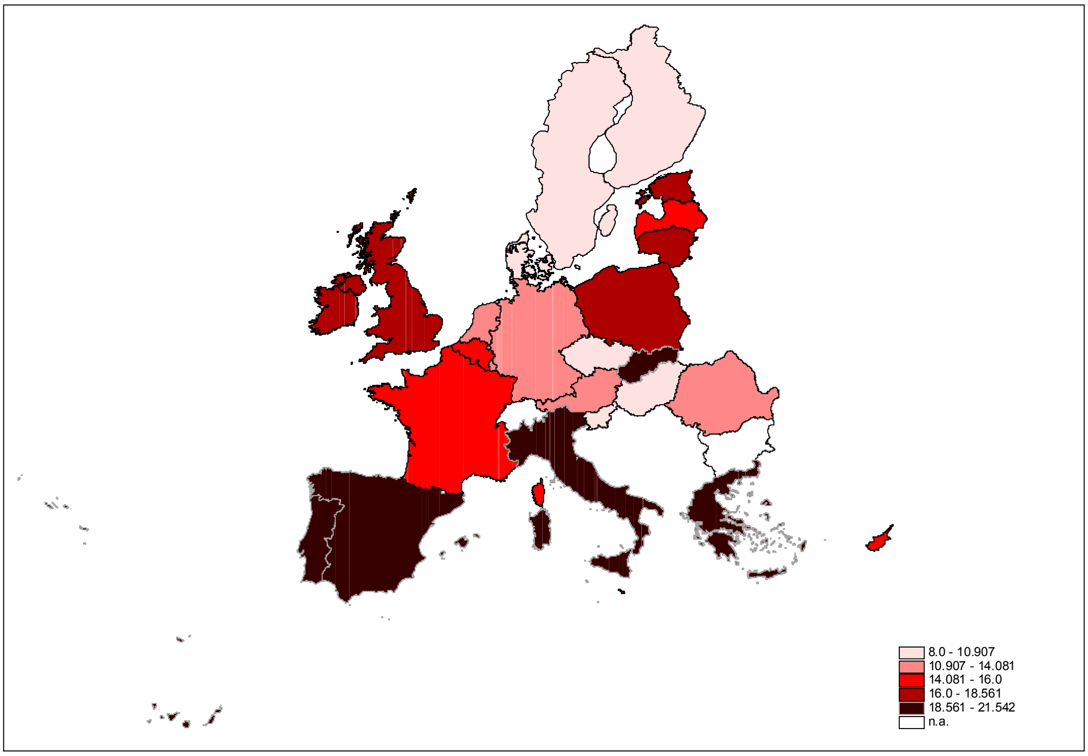
$$w_{EU,50} = \frac{1}{3}.\left(\frac{H_{EU,60}}{H_{EU,50}}\right), \qquad w_{EU,60} = \frac{1}{3}, \qquad w_{EU,70} = \frac{1}{3}.\left(\frac{H_{EU,60}}{H_{EU,70}}\right).$$

Obviously, when these weights are applied to EU as a whole, we have $H^{(EU)} = H_{EU,60}$, the standard definition of HCR.

Figure 5.1 shows HCR's for the EU25 countries and Romania, consolidated as described above. For EU15 countries where up to 8 ECHP waves are available, the consolidation is over the waves and over three poverty line thresholds (50%, 60% and 70% of the national median). In the case of Poland and Romania only a single wave is available; hence the consolidation is only over the above mentioned poverty line thresholds using EU15 weights. For the New Member States, the data are taken from NewCronos (also available in Statistics in Focus published by Eurostat) and are generally based on household budget surveys. No "consolidation" is involved in the case of NMSs. In all cases, country poverty lines have been used, i.e., poverty lines determined on the basis of income distribution within each country. The results in Figure 5.1 are only at county level to illustrate the procedure; poverty rates disaggregated by region will be presented later (especially in Chapter 8).

In the map in Figure 5.1 the 26 countries are grouped into clusters of 5 (6 in the lowest poverty rate group). Portugal, Spain, Italy, Greece and Slovakia are in the worst group (highest poverty rates), while the three Scandinavian countries, Hungary, Czech Republic and Slovenia are in the group with the lowest poverty rates. In some countries such as Italy and Slovakia, the high poverty rates also reflect pronounced regional differences. The concern with regional differences can be seen from their National Action Plan (NAPinc).

Figure 5.1 Head Count Ratio, consolidated over waves and poverty line thresholds where possible - country poverty lines. EU25 and Romania



Poverty line levels

As note in Section 4.4, "level of poverty line" refers the population level at which the income distribution is pooled for the purpose of defining the poverty line. All poverty related indicators in the Laeken list are based on country poverty lines, meaning that poverty line is always determined on the basis of national income distribution, for instance as 60% of the *national* median income. The same is true of the figures presented in Figure 5.1. It is necessary to consider other levels of poverty line, especially for the construction of poverty rates at the regional level. We have considered four poverty line levels as follows.

- <u>EU15 poverty line</u>, that is line determined on the basis of pooled income distribution for all EU15 countries. Country data are pooled in proportion to their population size. This means that sample weights as coded in ECHP data for each country are rescaled such that the total of individual weights summed over the sample is proportional to the country's population size. Note that these computations have to be performed separately for each wave, since the income levels across waves are not measured on the same scale and the distributions cannot be pooled across waves. Within each wave, the national figures expressed in PPS can be considered comparable and therefore can be pooled.
- <u>Country level poverty lines</u>, determined on the basis of income distribution separately within each country. This is the most commonly used measure in practice. Within each wave, the national figures may be expressed in PPS, but it is not necessary that the scaling be comparable across countries. The data requirement is therefore somewhat

less stringent than that using EU poverty line. Within-country comparability in the scaling of income amounts is of course required.

- o NUTS1 level poverty lines, determined on the basis of income distribution separately within each NUTS1 region.
- o NUTS2 level poverty lines, determined on the basis of income distribution separately within each NUTS2 region. This measure is more demanding of the data because income distribution has to be considered for each NUTS2 region separately; when based on a sample survey, it may be subject to large variation and irregularities. But in another sense, defining the poverty line at a lower level is *less* demanding of the data, since it is *not* affected by the fact that PPS values used are determined only at the country level and may not reflect variations in the cost of living among regions within the country.

In principle one could also consider poverty line defined at lower levels such as NUTS3. However sample size considerations would normally preclude that in practice.

As noted in Section 4.4, different poverty line levels (EU, country, NUTS1 or NUTS2 level poverty lines) can be seen as different poverty line thresholds (x% of the median) applied to different regions. For instance, in a country or region with median income 1.2 times the EU median, a poverty line at 60% of EU median is the same as a poverty line at 50% of national median. Similarly, for a richer region, a poverty line at 60% of the national median corresponds to a lower percentage of the median of that region.

Conventionally, most poverty indicators are defined on the basis of the national poverty line in each country. However, the use of different poverty lines to bring out different aspects would enrich analysis and enhance policy use of the results.

The use of EU-level poverty line can help illustrate the wide variety of living standards across the EU, especially with the expansion to EU25. It is true that so far there is no policy objective justifying the need for such a threshold: the objective stated in Lisbon and in other Council documents is to improve living standards and make the necessary steps to eradicate poverty in each Member State; apparently, it is not to make living standards of all Member State converge to a common level. However, with increasing integration in the Union, and increasing need for European level analysis, the use of EU-wide benchmarks becomes increasingly justified and rewarding.

Regional poverty lines provide useful additional information by focusing on *within region* disparities. This is particularly important in countries with large differences in income levels across regions. Indicators based on regional poverty lines become more important with devolution of formulation and implementation of social inclusion policies.

Poverty indicators at NUTS2 and lower levels

For poverty indicators at NUTS2 and lower levels, perhaps the most relevant poverty lines are country-level and NUTS2-level lines. In Table 5.2 we have reported consolidated poverty rates at country level, with poverty lines defined at two levels: country-level, based on the income distribution within each country; and NUTS2-level, based on the separate income distribution within each NUTS2 region.[12] The NUTS2 poverty line results can be produced only for the five countries where we have information of NUTS2 from the national surveys (namely, Italy, UK, Portugal, Poland and Romania).

With a country level poverty line, the resulting poverty rate reflects both disparities between individuals within each region and also inequality between regions. NUTS2-level poverty line removes the latter effect, and hence provides a purer measure of relative poverty within each region.

Table 5.2 Poverty rate (Head Count Ratio, HCR):
Effect of the chosen poverty line level (averaged over 8 ECHP waves)

Poverty line: Level	COUNTRY				NUTS2			
Threshold	70%median	60%median	50%median	consolidated	70%median	60%median	50%median	consolidated
EU15	23,7	16,0	9,9	16,0				
FI	18,4	10,0	4,9	10,1				
DK	17,9	10,6	5,1	10,3	(not applicable)			
SE	18,1	10,5	6,2	10,9				
NL	19,1	10,8	6,0	11,2				
LU	20,7	12,4	6,1	12,1	(not applicable)			
DE	19,2	12,4	7,4	12,4				
AT	20,5	12,8	6,8	12,5				
BE	22,7	14,4	7,9	14,1				
FR	23,6	15,4	8,7	15,1				
IE	27,9	19,2	10,0	18,0				
UK	26,7	18,7	11,7	18,6	26,3	18,2	11,4	18,2
ES	26,2	18,8	12,4	18,9				
IT	27,1	19,3	12,9	19,5	22,9	15,7	10,2	15,7
PT	28,5	21,3	14,4	21,3	28,3	20,9	13,7	20,9
GR	28,4	21,1	15,1	21,5				
PL	23,0	16,8	9,6	16,8	22,9	16,1	9,5	16,1
RO*	22,4	14,1	7,9	14,1	21,9	13,8	7,4	13,8

* HCR based on household consumption data

5.3 Relative and absolute measures

This section explores in-depth an important issue arising in the modelling of poverty and social exclusion indicators based on sample surveys of limited size. (Small area modelling and estimation procedures are described in Chapters 11-14.) Good models can be developed only on the basis of a clear identification of patterns of variation and relationships between the available variables.

Among these, the relationship between measures of *disparity and relative deprivation* on the one hand, and of *average levels of income and other resources* on the other, are likely to be of utmost importance. The former constitute a bulk of the direct indicators of poverty and social exclusion, but are typically available only from small samples with large sampling

[12] In both cases though, the results have been aggregated to the country level in Table 5.2. This is for illustrative purpose; some similar results disaggregated to lower levels will be presented subsequently.

error. The latter are more likely to be available from alternative, larger and more reliable sources, or at least are more directly related to a wide range of auxiliary indicators available from such sources. Often the direct relationships between indicators of disparity and relative deprivation and those auxiliary indicators may be weak, but become clearer when mediated through measures of income levels.

In specific terms, this section aims to empirically explore relationships of the type

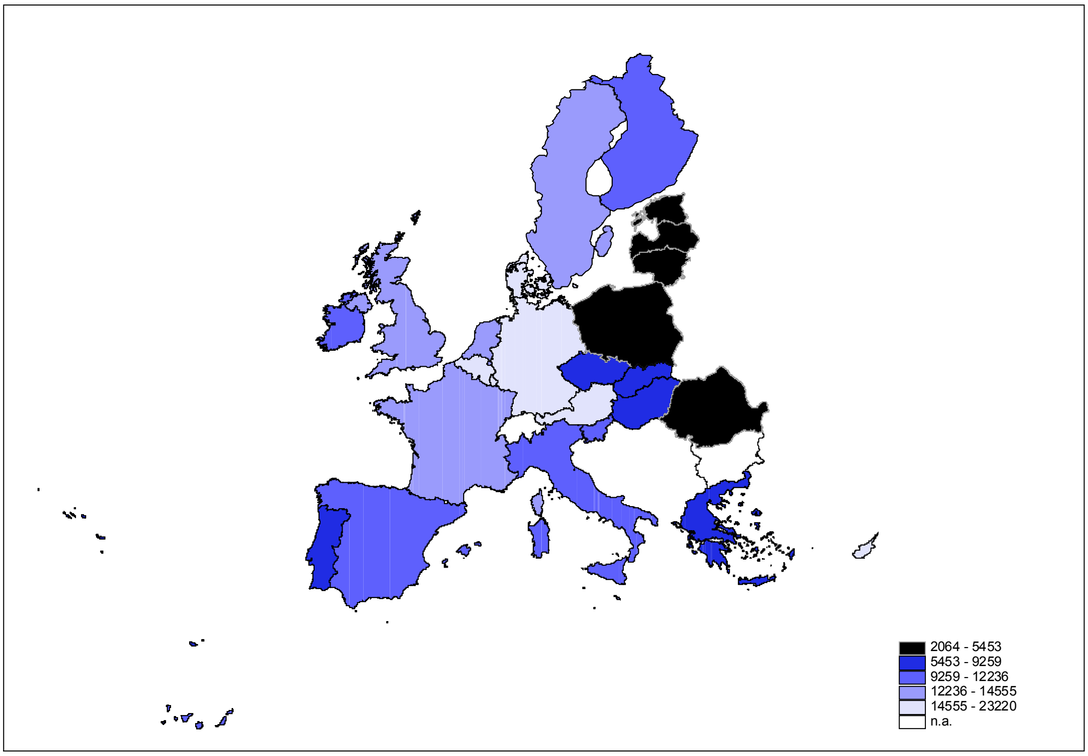
Mean disposable household per capita income	→	poverty rate
Mean net equivalised household income	→	poverty rate

Mean disposable household per capita income as tabulated in NewCronos provides a good example of indicators which are available at NUTS2 level with sufficient reliability from a source of large size.[13]

We can also construct a similar variable, namely net equivalised household income from ECHP surveys for EU15. For NMS10 countries, median income values have been published in *Statistics in Focus* by Eurostat, an assuming a likely value of 1.15 for mean:median ratio, these figures can be converted to approximate mean values. The mean values can then be modelled using small area estimation methodology for subnational regions. These estimates of mean income can be expected to be related to the required direct indicators of poverty and social exclusion which are available only from small samples, possibly through correlated variables. The results are shown in Figure 5.2 at the country level. Comparing with Figure 5.1, there is some negative correlation between the level of income and the level of poverty, but separately for old and new Member States.

[13] We can obtain approximate equivalised income by multiplying the per capita income by some average measure of equivalised household size. A simple measure of equivalised size we have used is to take it as square-root of the actual household size, following Atkinson *et al.* (1995).

Figure 5.2 Mean net equivalised income in PPS. EU25 and Romania



The rest of this section provides some graphs which we have constructed in the process of exploring the relationships between levels of income and poverty rates.

Figures 5.3 and 5.4 show the relationship of the poverty rate with income level estimates from surveys and from the tabulations available in NewCronos.

Figure 5.3 reports in the x-axis the mean net household equivalised income in PPS, while the y-axis reports the Head Count Ratio calculated using country poverty lines. (Note. The value of income for Luxembourg is well above the EU mean; for this reason we have simply reported the label rather than the actual value.)

A similar pattern is evident in Figure 5.4 where the x-axis reports the mean disposable per capita income derived from NewCronos tables.

Figure 5.3 Mean net equivalised income in PPS (from surveys), vs. Head Count Ratio. EU25, Romania and Bulgaria

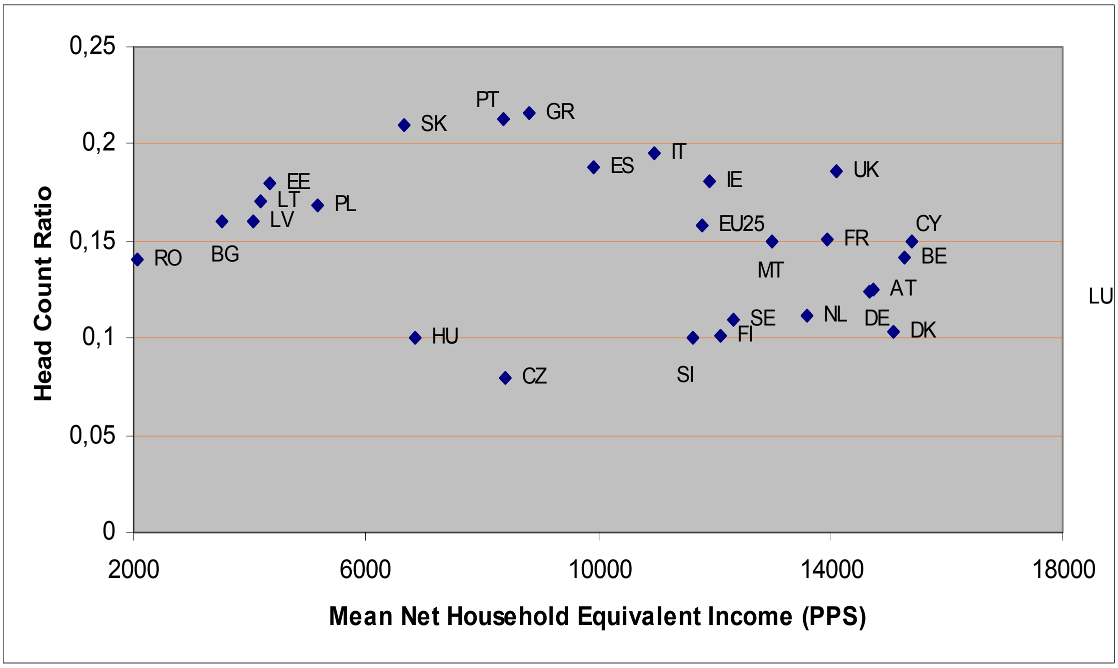
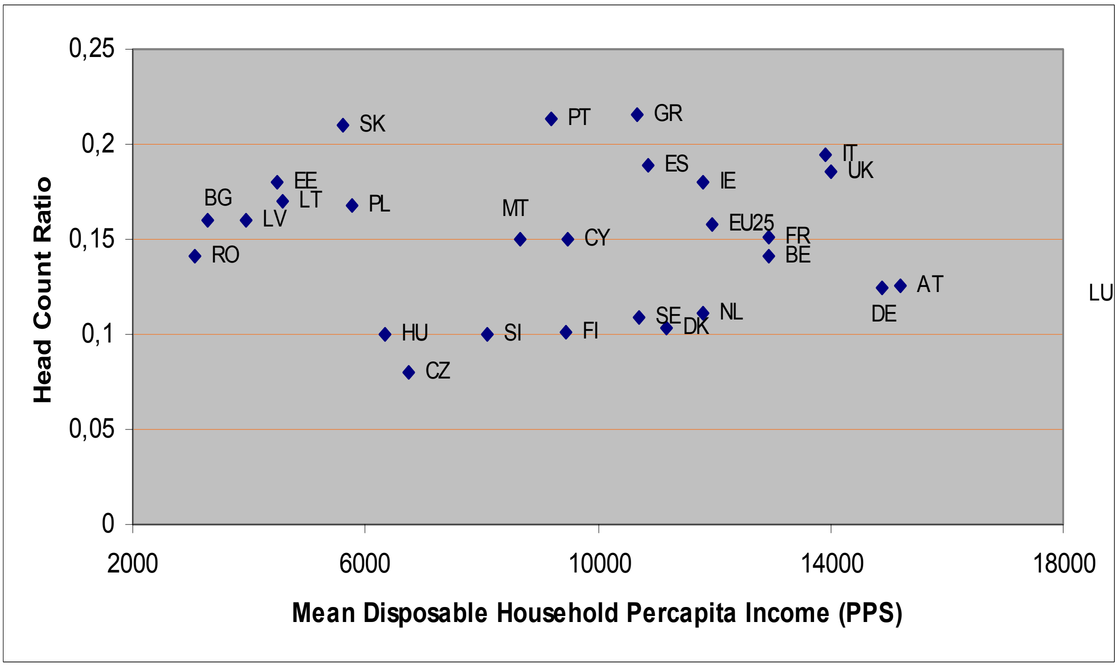


Figure 5.4 Mean disposable per capita income (from NewCronos tables) vs. Head Count Ratio. EU25, Romania and Bulgaria



It is interesting to analyse the relationship between disposable per capita income and Head Count Ratio at NUTS1 and NUTS2 level. Different countries show different levels of homogeneity among regions. For instance Figures 5.5 and 5.6 show how disparity between

NUTS1 regions is small in United Kingdom and Poland, while Italy and Spain show a large variation among NUTS1 regions.

Figure 5.5 Mean disposable per capita income (from NewCronos tables) vs. Head Count Ratio. United Kingdom and Poland NUTS1 regions

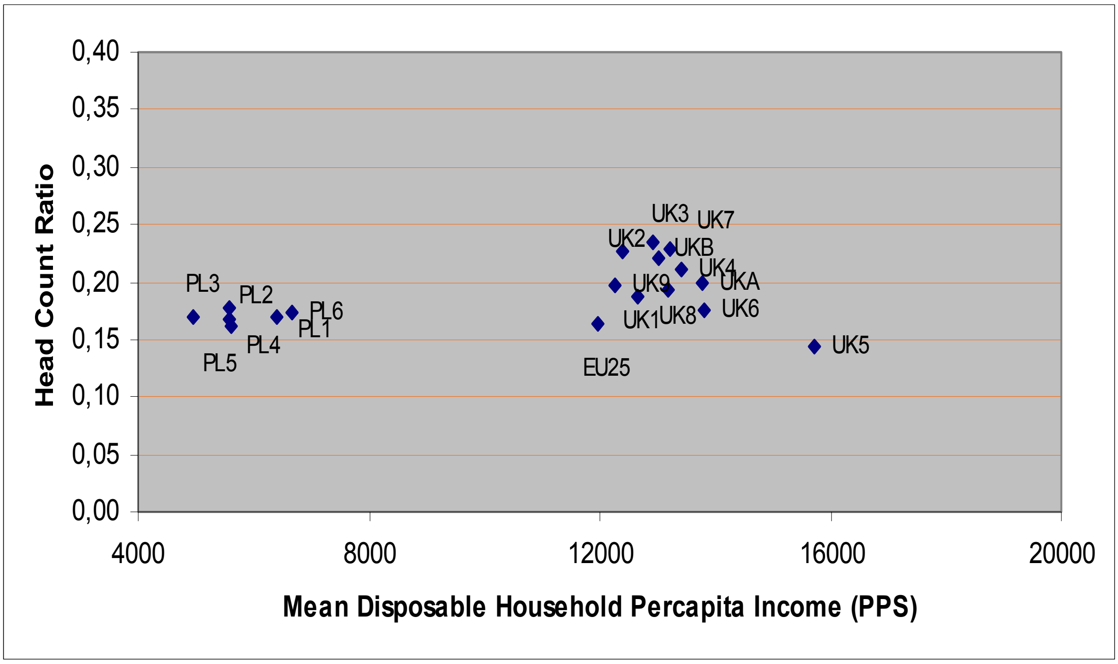
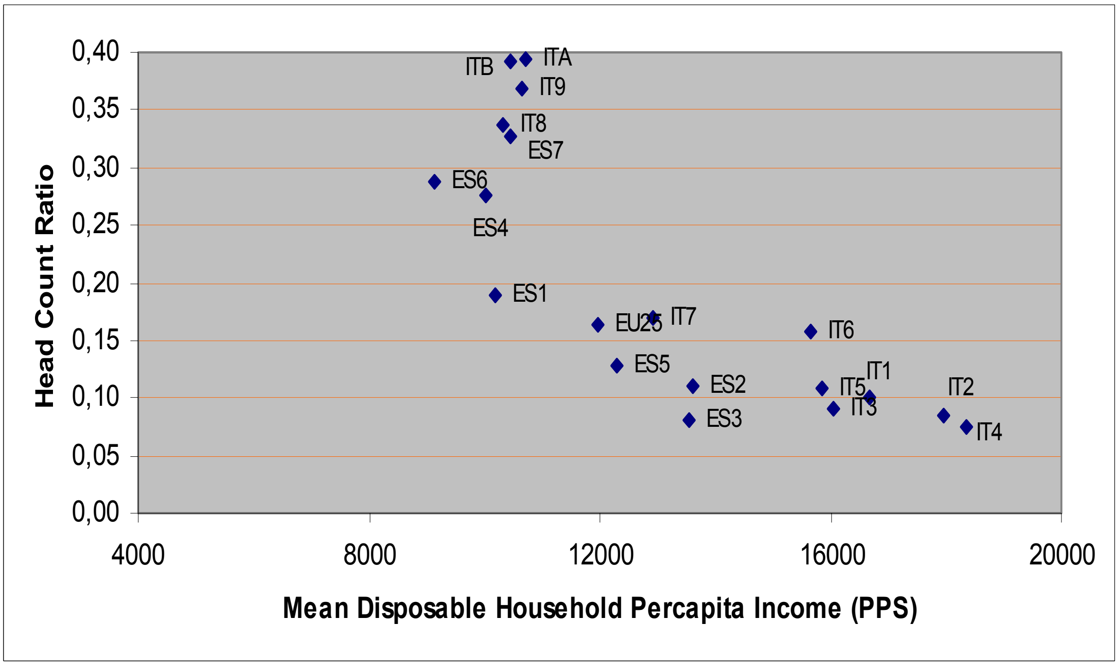


Figure 5.6 Mean disposable per capita income (from NewCronos tables) vs. Head Count Ratio. Italy and Spain NUTS1 regions



The same trend is evident when disaggregating to NUTS2 level; Figures 5.7 and 5.8 clearly show the different patterns in United Kingdom and Italy.

Figure 5.7 Mean disposable per capita income (from NewCronos tables) vs. Head Count Ratio. United Kingdom NUTS2 regions

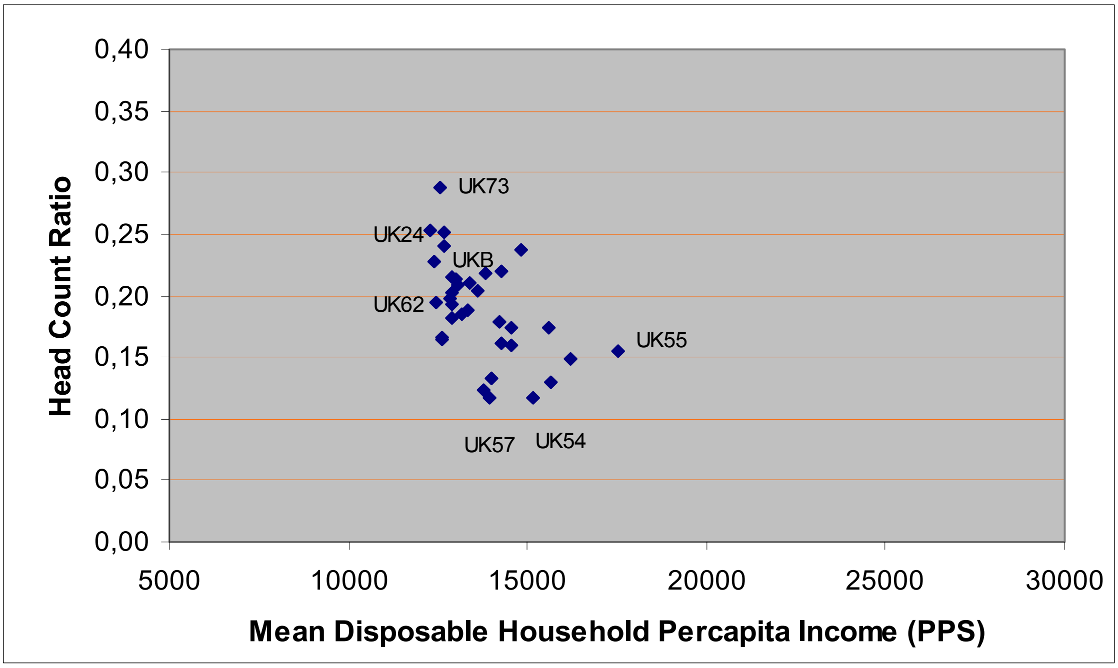
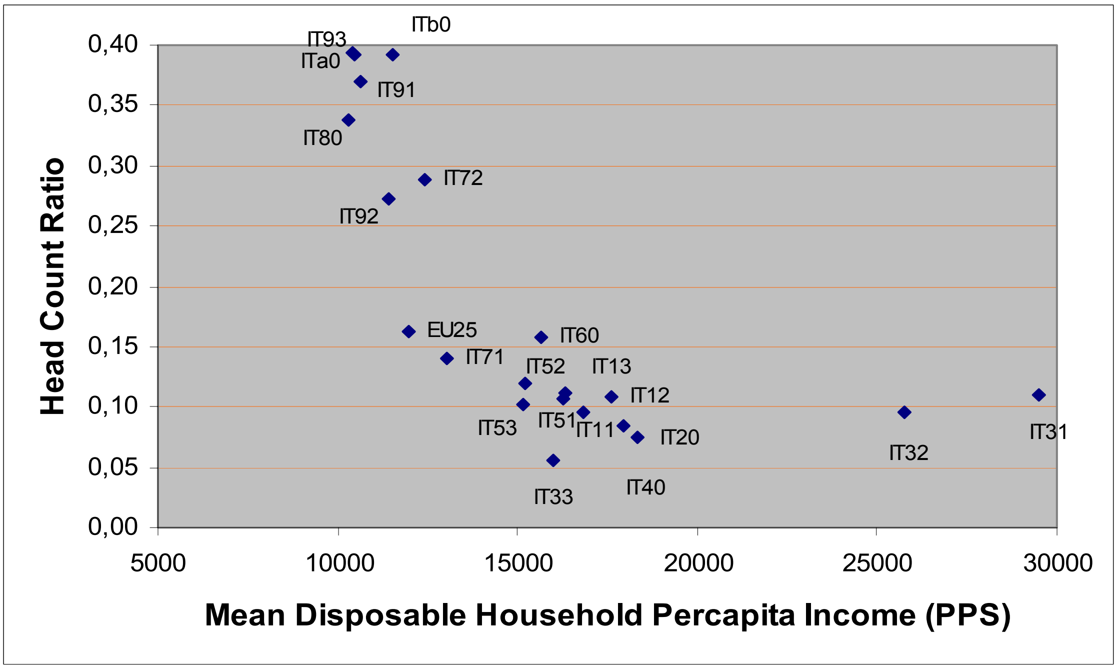


Figure 5.8 Mean disposable per capita income (from NewCronos tables) vs. Head Count Ratio. Italy NUTS2 regions

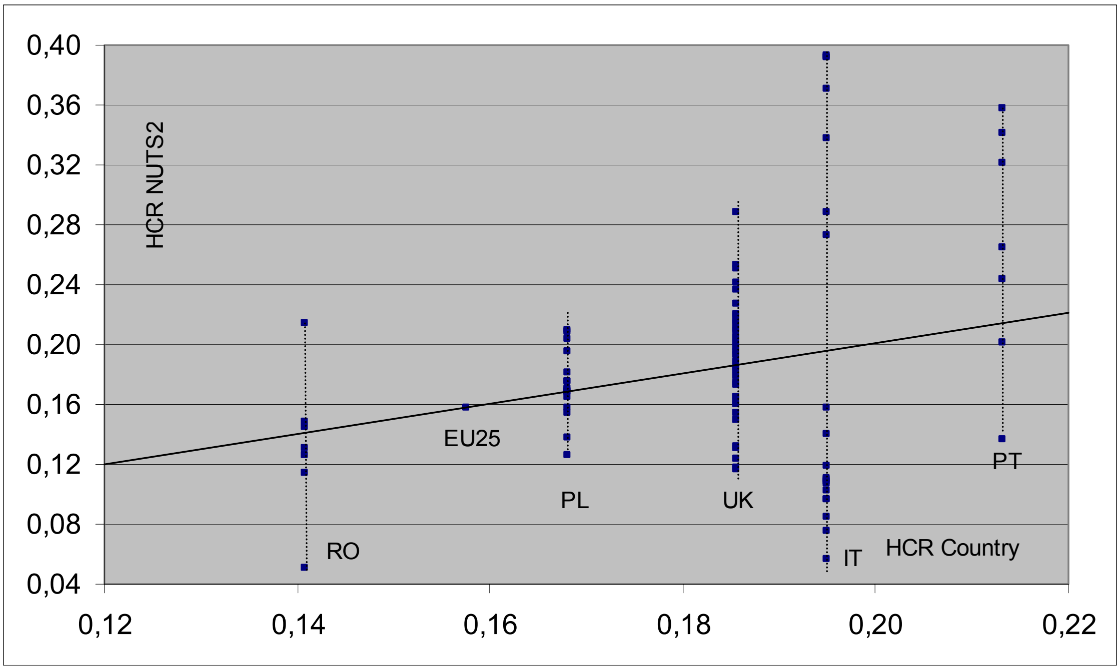


Country versus NUTS2 poverty line

The choice of the level of poverty line is an important consideration in the choice of indicators suited for regional analysis. As noted earlier, for regions at NUTS2 level or below, both country level and NUTS2 level poverty lines provide useful information.

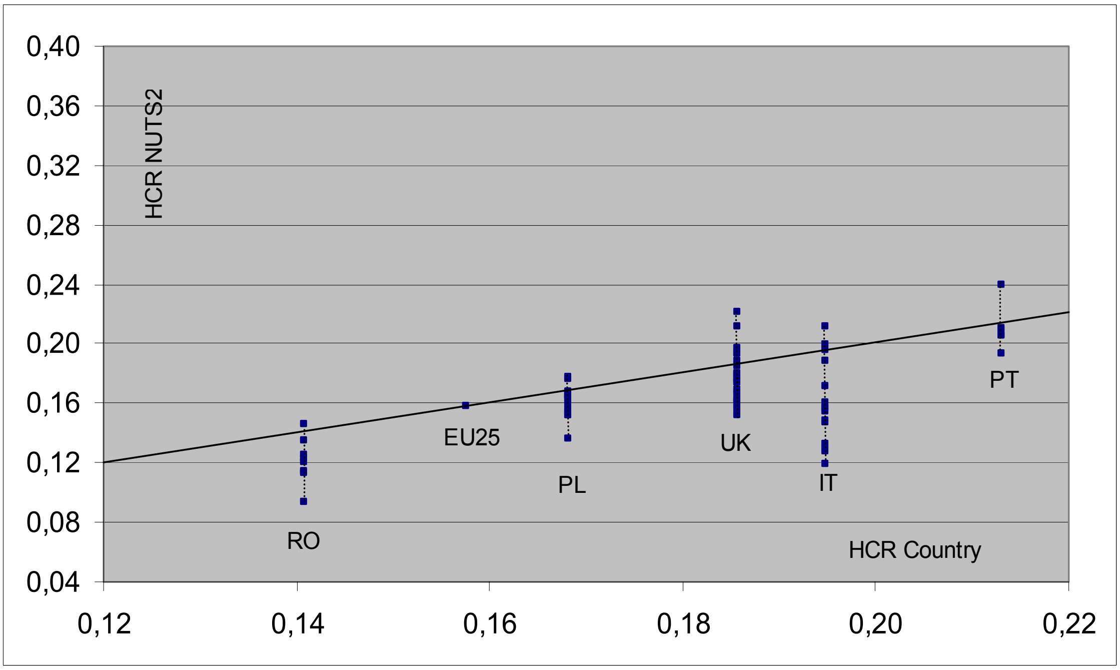
Figures 5.9 and 5.10 show the scatter of NUTS2 poverty rates for each country, the first computed using country poverty line, and the second using NUTS2 poverty line. It is evident that country level poverty line is like an "absolute" poverty line for the NUTS2 regions, while NUTS2 line is a purely relative line for each region. In countries such as Italy with large variations in mean regional incomes, the poverty rates in less well-off regions are inflated and in more well-off regions reduced with country level poverty lines. Consequently, the scatter of regional poverty rates around the national average is increased. By contrast, using NUTS2 poverty line gives less scatter, being a measure only of inequality within each region. The difference in the results with the two types of poverty lines is less marked in countries with smaller regional differences in mean income levels.

Figure 5.9 Head Count Ratio NUTS2 (using country poverty lines) - dispersions within countries



Notes. Because of limited data availability, NUTS2 poverty lines can be computed only for the five countries shown. Countries are placed along the x-axis according to country HCR (based on country poverty lines). For each country, the graph shows the scatter among NUTS2 regions of HCRs, defined using country poverty lines.

Figure 5.10 Head Count Ratio NUTS2 (using NUTS2 poverty lines) - dispersions within countries



Notes. As in Table 5.9, countries are placed along the x-axis according to country HCR (based on country poverty lines). However, HCR for each NUTS2 region is computed using a poverty line based on the income distribution of that region.

5.4 Subpopulation poverty

Choice of geographical/administrative units - regions - as units of interest does not preclude poverty and social exclusion analysis in terms of other types of units. Poverty and social exclusion indicators can and should be constructed for other types of units, and also for particular subpopulations.

The breakdown by age, gender, economic activity, household type etc. required in Laeken indicators has to be curtailed when geographical disaggregation is also required. In the construction of regional indicators, it is appropriate to focus on selected groups such as children, elderly persons, and possibly young persons entering the labour market.

In this section, some results on poverty rates (HCR's calculated on the basis of country poverty lines) for children (persons aged 0-15) and elderly persons (aged 60+) are compared with those among the general population. Table 5.3 shows the HCR's among children and elderly persons by country, also comparing those with the corresponding rates for EU15 as a whole. Figures are consolidated as before: first cross waves (detailed figures by wave are not shown for reasons of space), then across the three poverty thresholds. Variations across countries are displayed in Figures 5.11 and 5.12. These computations have been made only for EU15 countries because of data limitations.

Table 5.3 Poverty rate among children and elderly persons. EU15

	Children			Old persons			Consolidated		
Poverty line	50%	60%	70%	50%	60%	70%	Children	Old persons	all persons
EU15	12,2	19,7	28,7	9,9	17,6	27,2	19,7	17,6	16,0
FI	2,1	5,5	13,5	4,5	14,2	28,7	6,1	13,6	10,1
DK	2,6	4,9	10,5	9,2	24,7	41,1	5,4	22,5	10,3
SE	4,5	9,9	18,8	2,2	7,1	17,6	10,0	7,5	10,9
NL	7,2	13,6	23,6	3,4	6,0	18,2	13,8	7,9	11,2
LU	8,6	18,4	29,0	3,6	8,5	17,6	17,4	8,8	12,1
DE	8,4	14,5	23,8	8,2	14,0	21,1	14,8	14,1	12,4
AT	7,5	14,8	25,8	9,3	20,1	29,7	14,9	18,6	12,5
BE	7,6	14,0	22,7	11,9	22,4	35,1	14,0	22,1	14,1
FR	9,0	16,9	26,4	9,7	17,4	27,0	16,5	17,4	15,1
IE	13,3	24,4	33,7	11,1	26,7	40,0	23,0	24,1	18,0
UK	18,1	27,6	36,5	11,1	21,0	33,9	27,3	20,9	18,6
ES	16,9	24,6	32,8	8,4	17,0	25,9	24,8	16,2	18,9
IT	15,9	23,5	32,3	9,9	16,5	25,6	23,7	16,8	19,5
PT	17,4	25,4	33,9	22,3	32,5	41,4	25,5	32,9	21,3
GR	12,3	18,4	26,4	23,8	31,9	40,4	18,8	33,4	21,5

Country poverty line. Children: persons aged under 16. Old persons: aged 60+.

Figure 5.11 Head Count Ratio among Children (0-15). EU15

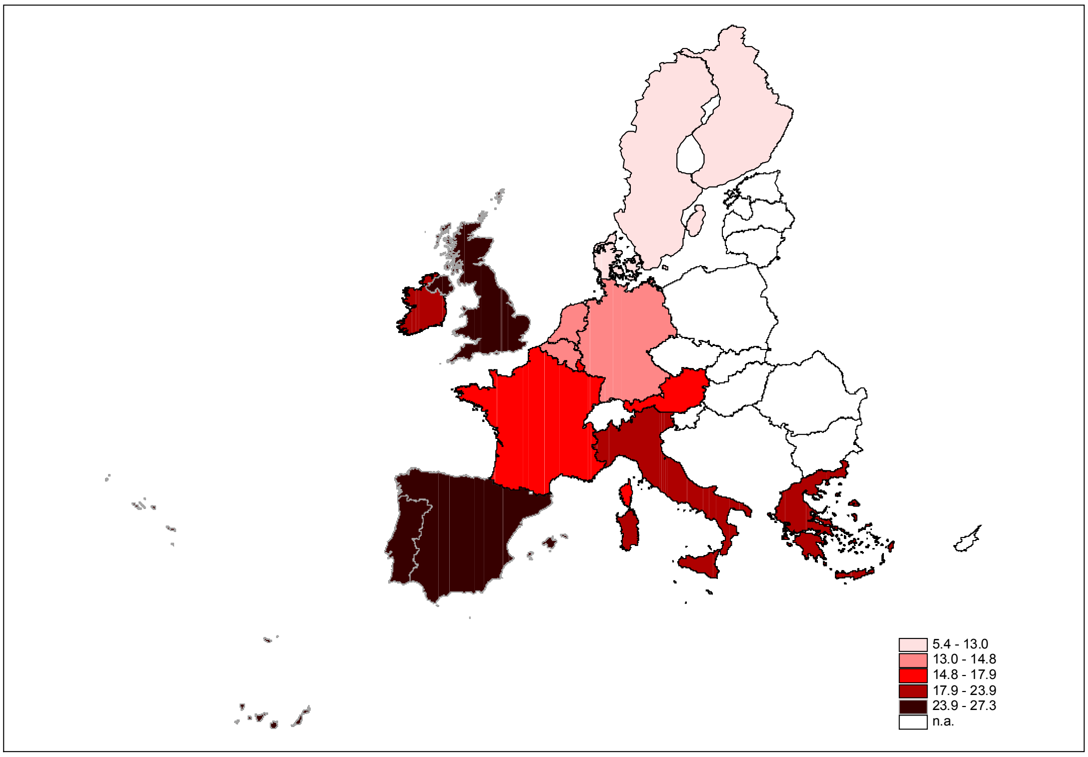


Figure 5.12 Head Count Ratio among elderly persons (60+). EU15

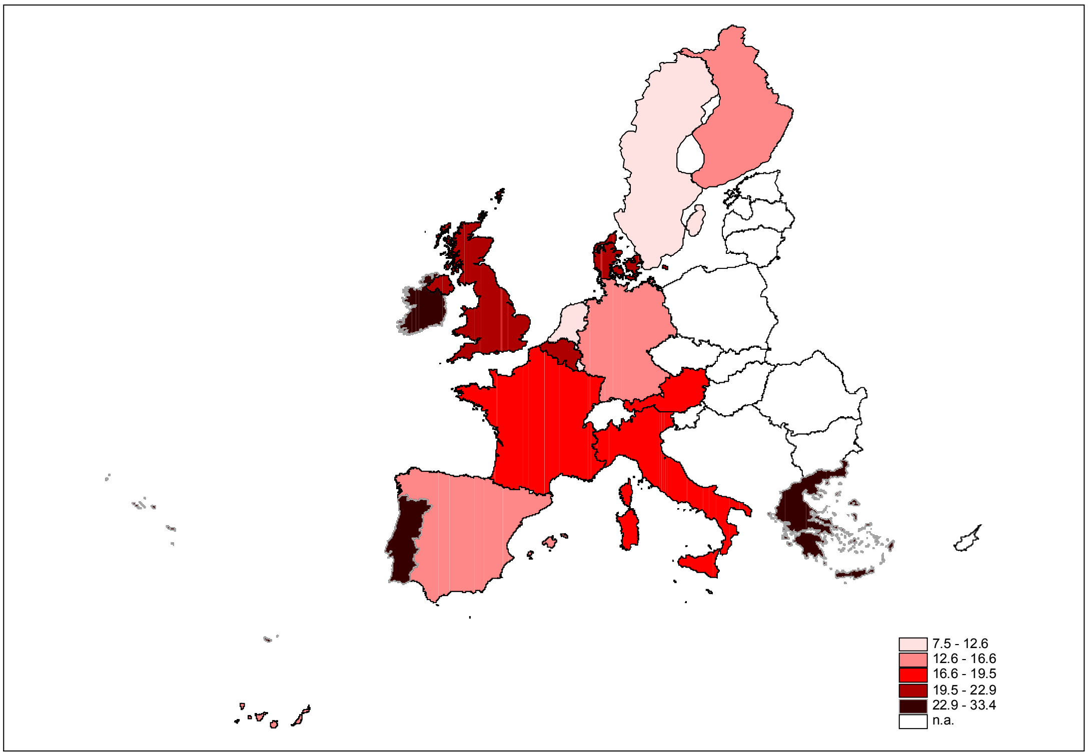
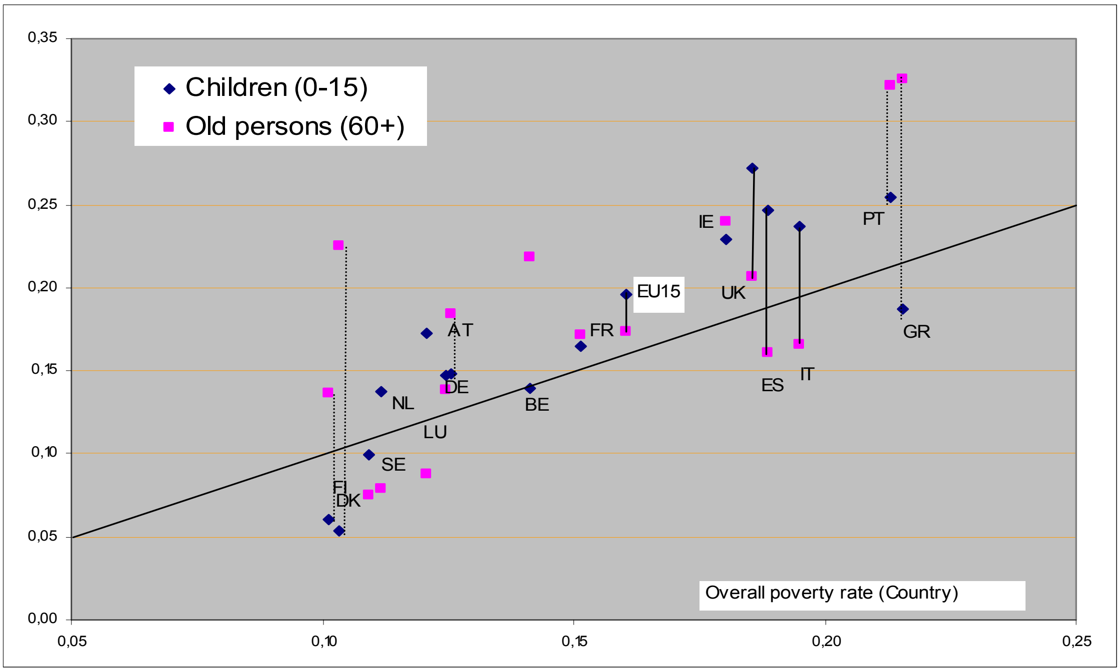


Figure 5.13 Comparison of child and elderly person poverty rates with the rates for the population as a whole. EU15



Notes. Countries are placed along the x-axis according to country HCR. The diagonal line corresponds to this rate for the total population with which the rates for children and elderly persons in the country can be compared. Country poverty lines are used throughout.

In Figure 5.13 the HCRs for children and elderly persons are displayed against the overall population HCRs; in most countries the situation of children in fact appears to be less disadvantaged compared to the general population, with the UK being a major exception.

By contrast, elderly persons fare worse than the general population in a number of countries, with big differences in Greece and Portugal, also in Ireland, and smaller differences in Belgium, Austria, Denmark and Finland.

In comparison with the *general population*, the situation of children and elderly persons by regions may be briefly summarise as follows. The detailed results are presented in Chapter 8 below. In that chapter, Figures 8.9 to 8.13 report the regional variation in the *relative position* of children and of the elderly in the five major EU15 countries: Germany (Figure 8.9), France (Figure 8.10), United Kingdom (Figure 8.11), Italy (Figure 8.12) and Spain (Figure 8.13). By relative position we mean the (NUTS1/Country) ratio of poverty rates for the given subpopulation. This focuses on regional differences, abstracting the effect of national differences in child or elderly poverty rates.

Among these countries we can highlight seven NUTS1 region where the relative position of children appears to be much worse than the relative position of the elderly: Hamburg, Mecklenburg-Vorpommern, Thüringen in Germany, Nord - Pas-de-Calais in France, Sicilia and Sardegna in Italy, while the reverse is the case for East Anglia in the United Kingdom.

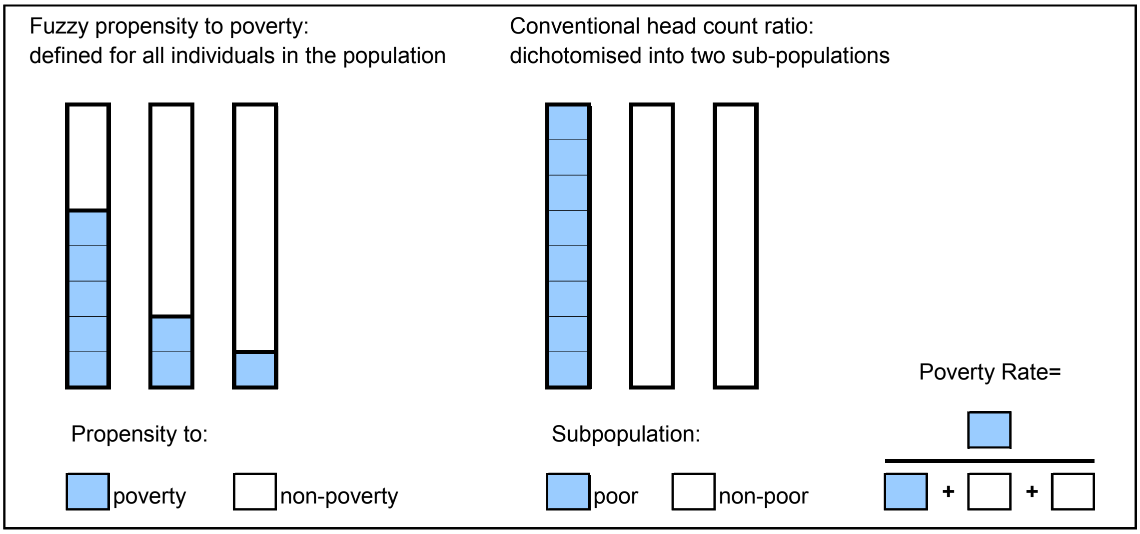
5.5 The propensity to income poverty: Fuzzy monetary

An alternative view of monetary poverty

In place of the conventional classification of the population into a simple "poor/non-poor" or "deprived/non-deprived" dichotomy, in this study we also propose measures treating poverty and deprivation as a matter of degree: in principle all individuals are subject poverty or deprivation, but to varying degrees (some much more than others). That degree, say 1 for the poorest or the most deprived to 0 for the richest or the least deprived, is determined by the individual's rank in the income distribution, and the individual's share in the total income received by the population. The concepts can be extended to cover non-monetary aspects of living standards, in the form of what we have termed "fuzzy supplementary" measures. These are considered in the next chapter. There we also describe indicators of income poverty and non-monetary deprivation in combination.

Figure 5.14 illustrates the basic idea of treating poverty and deprivation as a matter of degree, replacing the conventional classification of the population into a simple dichotomy. In principle all individuals in a population are subject poverty or deprivation, but to varying degrees. We say that each individual has a certain *propensity* to poverty or deprivation, the population covering the whole range [0,1]. The conventional approach is a special case of this, with the population dichotomised as {0,1}: those with income below a certain threshold are deemed to be poor (i.e. are all assigned a constant propensity=1); others with income at or above that threshold are deemed to be non-poor (i.e. are all assigned a constant propensity=0).

Figure 5.14 The basic idea of poverty or deprivation as a matter of degree: comparison with the conventional poor/non-poor dichotomy



There are several advantages of treating poverty and deprivation as a matter of degree, applicable to all members of the population, rather than as simply a "yes-no" state.

1. Further insight into the relative income situations of individuals and groups can be obtained by incorporating into the poverty rates a measure of the actual levels of incomes received, particularly at the lower end of the income distribution.
2. Non-monetary deprivation depends on forced non-access to various facilities or possessions determining the basic conditions of life. An individual may have access to some but not to others. Hence non-monetary deprivation is inherently a matter of degree, and some *quantitative approach* such as the present one is essential.
3. The combined analysis, considering income poverty and non-monetary deprivation simultaneously, is greatly facilitated by treating each dimension as a matter of degree. The need to divide the population into numerous discrete groups - as would normally be required in the conventional analysis, especially in the longitudinal context - is avoided.
4. Equally important is the potential of this approach in studying poverty (or more generally, deprivation in multiple dimensions) in the *longitudinal* context. The conventional approach measures mobility simply in terms of movements across some designated poverty line, and does not reflect the actual magnitude of the changes affecting individuals at all points in the distribution. Consequently, the degree of mobility of persons near to the chosen line tends to be over-emphasised, while that of persons far from that line largely ignored.
5. We can expect the resulting measures to be more precise. The sampling error of a distribution is lower than that of a dichotomy with values concentrated at the two end points. We can also expect the measures to be less sensitive to local irregularities in the income distribution curve, and to the particular choice of the poverty threshold.

The last mentioned is a particularly important point in the context of constructing regional indicators of poverty and inequality given smallness of the available sample sizes.

The propensity to income poverty, defined as Fuzzy Monetary (FM) associated with each individual *j* is related to the person's rank and share in the equivalised income distribution of the population, defined at whatever level (EU, country, NUTS1, NUTS2 …). The model used is as follow[14].

First we construct an income index:

$$V_j = \sum_{i=j+1}^{n} v_i \text{ , j = 1 to n-1; } V_n = 0,$$

where $v_i = y_i \Big/ \sum_{i=2}^{n} y_i$ is the share of total equivalised income (y_i) received by individual of rank i in the ascending income distribution. V_j varies from $V_1 = 1$ for the poorest, to $V_n = 0$ for the richest individual. It is the share of the total equivalised income received by all individuals less poor than the person concerned.

Actually, in order to take into account possible negative incomes, tied rankings (i.e. individuals in the same household, etc…) and weighting, more precisely the formula used is the following:

$$V_j = \frac{\sum_i w_i (y_i - y_1) | i \in K : y_K > y_i}{\sum_i w_i (y_i - y_1) | i \in K : y_K > y_1}.$$

Corresponding to the income index, the propensity to income poverty is defined as: $FM_j = (V_j)^{\alpha / H^{(D)}}$, where $H^{(D)}$ is HCR for a particular domain (D), computed separately for each wave but consolidated over different poverty line thresholds as explained in Section 5.2.

We have determined the three parameters α (one for each level of income distribution, country, NUTS1 and, where available in the data, NUTS2) such that for the *European population as a whole,* the mean of the index FM_j is equal to the proportion poor (HCR) according to the conventional approach. In the above equation α is divided by $H^{(D)}$, since we have found it empirically that this form of the equation results in very stable values of α for different domains despite differences in their head count ratios.

In fact, the α values are very stable irrespective of the income distribution level chosen.

α values	Country =2,27	NUTS1 =2,28	NUTS2 =2,30

Figures 5.15 and 5.16 show the estimated "fuzzy monetary" poverty rates at NUTS1 and NUTS2 levels.[15]

---

14 For a fuller description of the methodology and some empirical results, see Betti and Verma (1999, 2002).

15 For countries where the required information is available, these figures have been obtained using small area estimation (SAE) procedures described in Chapters 11-13. These small area estimation procedures are illustrative, and by no means the best that can be developed with better data and knowledge at the national level. These limitation should be kept in mind. Simpler (and hence cruder) prediction based on regression coefficients obtained in the SAE model have been used for other countries, but are not included in these figures.

Figure 5.15 Fuzzy Monetary deprivation rates for NUTS1 regions. EU15, Poland and Romania

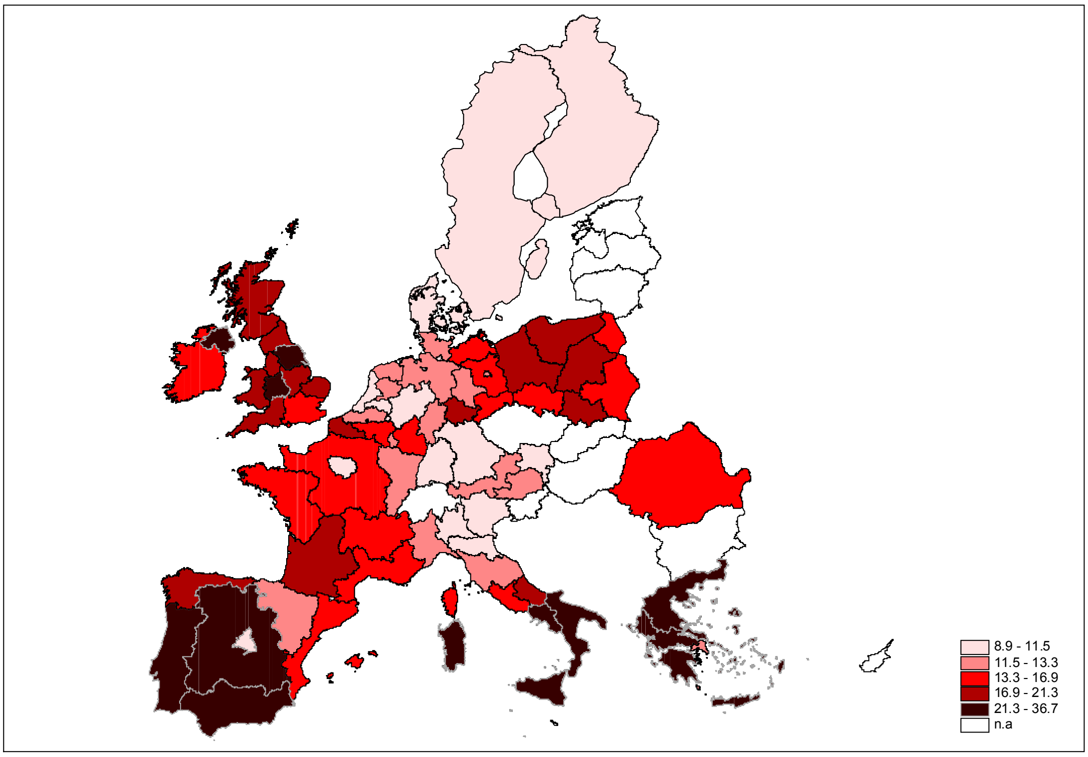
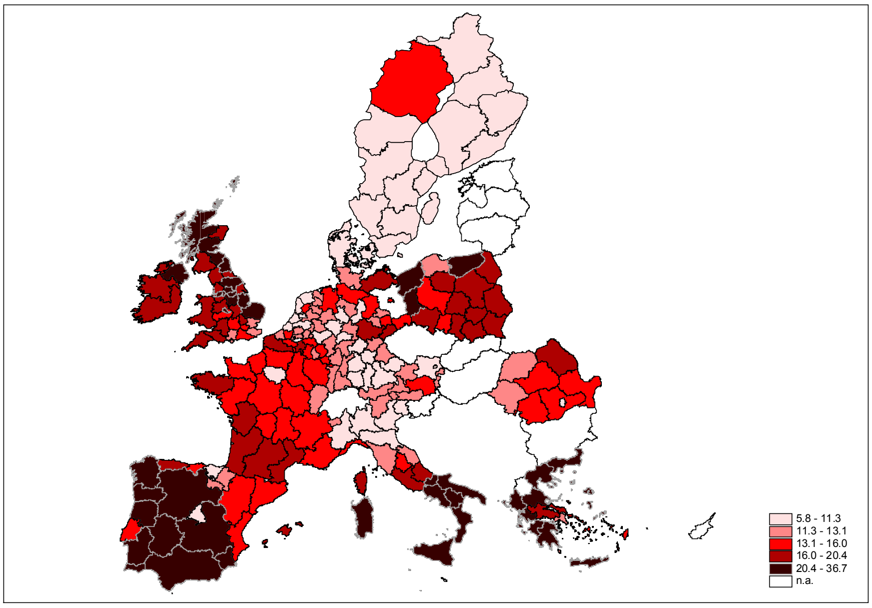


Figure 5.16 Fuzzy Monetary deprivation rates for NUTS2 regions. EU15, Poland and Romania



Chapter 6
Multidimensional deprivation: non-monetary indicators

6.1 Introduction

Deprivation and social exclusion are multi-dimensional. Two major issues arise in the construction of indicators incorporating this multi-dimensionality:

- The first concerns the extent to which many different dimensions can be consolidated to form a single, or at least a much reduced number of indicators for practical use in policy and research; and where such consolidation is possible, the choice of the methodology and the weights to be given different items or indices of deprivation.
- The second concerns the level of units for which such aggregation should be done.

As noted in Chapter 3, two aspects of the aggregation are involved: aggregation over dimensions of deprivation; and aggregation over individuals or households. Different types of measures are obtained depending on the order in which the two aggregations are applied. Aggregation first over dimensions for each individual provides a clear picture of the overlap (intensity) of deprivation in different dimensions as experienced by individuals. First aggregating indicators of a given dimension over individuals in an area (say a region), and then over different dimensions for that region, informs us about characteristics and circumstances of the area concerned in terms of geographical concentration (intensity) of deprivation in different dimensions.

Both forms of aggregation are useful – each gives additional new information not fully captured by the other. Apart from substantive differences, an important practical difference between the two modes of aggregation is that the first one (i.e., the aggregation of indicators of deprivation in different dimensions for the same individual) is generally much more demanding on the statistical data required for the construction indicators. For one thing, it requires data on different dimensions to be linked at the micro level.

It is with the methodology of such aggregation of dimensions of deprivation at the micro level that the present chapter is concerned. We begin with the incorporation of additional, non-monetary dimensions of deprivation supplementing income poverty. Then, measures are constructed combing the income and non-monetary aspects.

6.2 Variables and dimensions of non-monetary deprivation

In addition to the level of monetary income, the standard of living of households and persons can be described by a host of indicators, such as housing conditions, possession of durable goods, the general financial situation, perception of hardship, expectations, norms and values. Quantification and putting together of a large set of non-monetary indicators of living conditions involves a number of steps, models and assumptions.

Firstly, from the large set which may be available, a selection has to be made of indicators which are substantively meaningful and useful for the purpose. For our analysis using the rich ECHP data, and also from the national survey in Poland, a subset of the available indicators was selected. The most important determining factor in the choice of the set of items for analysis was an assessment – based on a detailed examination of variations in frequency distributions across countries and background knowledge of national situations – of the extent to which an item could be meaningfully included to construct reasonably comparable indicators of deprivation. Generally, the result of the exercise has been to include a majority of so-called 'objective' indicators on non-monetary deprivation, such as the possession of material goods and facilities and physical conditions of life, at the expense of what may be called 'subjective' indicators such as self-assessment of the general health condition, economic hardship and social isolation, or the expressed degree of satisfaction with various aspects of work and life. These latter type of indicators tend to be more culture-specific and hence less comparable across countries and regions.

Secondly, it is useful to identify the underlying dimensions and to group the indicators accordingly. Taking into account the manner in which different indicators cluster together (possibly differently in different national situations) adds to the richness of the analysis; ignoring such dimensionality can in fact result in misleading conclusions. In the present analysis we have used the indicators shown in the inset, grouped into five dimensions as in the official 2nd Social Report on Poverty, Income and Social Exclusion (Giorgi and Verma, 2002) published by Eurostat.

6.3 Constructing indicators of non-monetary deprivation

Putting together categorical indicators of deprivation for individual items to construct composite indices requires decisions about assigning numerical values to the ordered categories and the weighting and scaling of the measures. When aggregating several indicators at macro level, an early attempt to choose an appropriate weighting system was due to Ram (1982), using Principal Component Analysis, which was also adopted by Maasoumi and Nickelsburg (1988). Nolan and Whelan (1996) adopted Factor Analysis for evaluating a weighting system to be used at micro level.

Individual items indicating non-monetary deprivation often take the form of simple 'yes/no' dichotomies (such as the presence or absence of enforced lack of certain goods or facilities). However, some items may involve more than two ordered categories, reflecting different *degrees* of deprivation.

Follows a brief description of the methodology we have used for quantifying the degree of deprivation, taking into account its multidimensional nature.

Consider the general case of c=1 to C ordered categories of some deprivation indicator, with c=1 representing the most deprived and c=C the least deprived situation. Let c_j be the category to which individual j belongs. Cerioli and Zani (1990), assuming that the rank of the categories represents an equally-spaced metric variable, assigned to the individual a deprivation score as (a) below:

(a) $$d_j = (C - c_j)/(C - 1),\ 1 \le c_j \le C\,.$$

Cheli and Lemmi (1995) proposed an improvement by replacing the simple ranking of the categories with their *distribution function* in the population:

(b) $$d_j = \{1 - H(c_j)\}/\{1 - H(1)\}.$$

Dimensions and items of non-monetary deprivation

1 Basic non-monetary deprivation – these concern the lack of ability to afford most basic requirements:

Keeping the home (household's principal accommodation) adequately warm.
Paying for a week's annual holiday away from home.
Replacing any worn-out furniture.
Buying new, rather than second hand clothes.
Eating meat chicken or fish every second day, if the household wanted to.
Having friends or family for a drink or meal at least once a month.
Inability to meet payment of scheduled mortgage payments, utility bills or hire purchase instalments.

2 Secondary non-monetary deprivation – these concern enforced lack of widely desired possessions ("enforced" means that the lack of possession is because of lack of resources):

A car or van.
A colour TV.
A video recorder.
A micro wave.
A dishwasher.
A telephone.

3 Lacking housing facilities – these concern the absence of basic housing facilities (so basic that one can presume all households would wish to have them):

A bath or shower.
An indoor flushing toilet.
Hot running water.

4 Housing deterioration – these concern serious problems with accommodation:

Leaky roof.
Damp walls, floors, foundation etc.
Rot in window frames or floors.

5 Environmental problems – these concern problems with the neighbourhood and the environment:

Shortage of space.
Noise from neighbours or outside.
Dwelling too dark/not enough light.
Pollution, grime or other environmental problems caused by traffic or industry.
Vandalism or crime in the area.

Note that the above two formulations for d_j are identical in by far the most common case – that of a dichotomous indicator (C=2), giving a dichotomous m.f. d_j= 1 (deprived) or d_j= 0 (non-deprived).

The procedure for aggregating over a group of item is also the same for the two formulations: a weighted sum is taken over items (k): $f_j = \Sigma W_k . d_{k,j} / \Sigma W_k$, where the W_k are item-specific weights, taken as $W_k = \ln\left(1/\bar{d}_k\right)$. For dichotomous indicators, $\bar{d}_k$, the mean of d_j values for item k, simply equals the proportion deprived of that item.

In our approach here, we use the above framework with some important refinements as follows. In fact, we construct non-monetary indicators in exactly the same way as the income indicator described in Chapter 5 .

(1) We begin by selecting the items to be included in the index or indices of deprivation on substantive grounds, and grouping the items into 'dimensions' as described in Section 6.1. Deprivation scores ($d_{k,j}$) are assigned to ordinal categories of each item as in (a).

(2) The weights to be given to items are determined *within each dimension* (group of items) separately as described below in Section 6.5. With these weights, a deprivation score is determined for each dimension (δ: 1,…, Δ), and also for the overall situation covering all the indicators:

$$S_{\delta,j} = \Sigma_{k\in\delta} W_{k.}.(1-d_{k,j})/\Sigma_{k\in\delta} W_{k.} \qquad S_j = \Sigma_k W_{k.}.(1-d_{k,j})/\Sigma_k W_{k.}$$

Note that S is a 'positive' score indicating lack of deprivation; thus it is akin to income in the study of income poverty (Section 5.5).

(3) As in the Fuzzy Monetary approach, we define the individual's propensity to non-monetary deprivation as the *share* of the total "non-deprivation" assigned to all individuals less deprived than the person concerned. It varies from 1 for the most deprived, to 0 for the least deprived individual. The particular form below has been chosen so as to take into account tied rankings, which are much more frequent for items with a few categories, compared to the case of continuous variables like income.

$$V_{(FS)\delta,j} = \frac{\sum_i w_i S_{\delta,i} \mid i \in I : S_{\delta,I} > S_{\delta,j}}{\sum_i w_i S_{\delta,i} \mid i \in I : S_{\delta,I} > S_{\delta,1}} \; ; \quad V_{(FS)j} = \frac{\sum_i w_i S_i \mid i \in I : S_I > S_j}{\sum_i w_i S_i \mid i \in I : S_I > S_1} .$$

This formulation differs from (b) above in two respects. Firstly, it can be expected to be somewhat more sensitive to the actual levels of deprivation than the normalised distribution function H_j, the latter being only the *number* of individuals less deprived than the person concerned. Secondly, this index is constructed *after* aggregation over individual items, and therefore does not reduce to a mere dichotomy as does (b) for individual items.

(4) Finally, corresponding to income poverty, the propensity to non-monetary deprivation is defined as: $FS_j = \left(V_{(FS)j}\right)^{\alpha/H_{C,w}}$. As before, we have estimated a single value of parameter α on the basis of information pooled over countries and waves so as to match the overall average FS rate with the overall *monetary poverty rate*.[16]

To summarise, the procedure outlined above permits us to summarise diverse items of information on the living conditions of individuals and households in terms of a single quantitative index of non-monetary deprivation, or of 'life-style deprivation' as some researchers prefer to call it. For the sake of comparability across countries and regions within EU, and also because of limitations in the availability of micro data, the items incorporated into the proposed index are largely confined to material conditions of life. Other aspects, such as employment conditions of the household (especially identifying workless households), the type and characteristics of the household (especially identifying

[16] The parameter has been estimated on the basis of data for EU15 countries only, but the procedure can be easily extended to include all EU countries when micro data become available for them.
As before, when data are so pooled, variables V_j and FS_j must still be defined separately for each country and survey wave, since it is the position of each person within the *national distribution at a given time* which is of interest. Note that the denominator of the exponent is the country and wave specific *income poverty rate*, $H_{c,w}$. Incidentally, with our data the value found was α=0.99 (compared with α=2.27 for FM, noted earlier).

lone-parent and female-headed households), the main source of the household's income (especially identifying those relying on social transfers), etc., can – and should – also be incorporated *where data on different dimensions of deprivation can be linked at the micro level*. Like the other items already included in our index, such additional items are also measured relatively easily and are suitable for inclusion in a fairly large-scale surveys. They can also be cumulated over time to increase the available sample size, and hence permit greater regional disaggregation.

More difficult to integrate into such a framework are subjective indicators of satisfaction with conditions of life in various dimensions, other attitudinal items, subjective assessment of health status, etc. The main reason for this difficulty is the lack of comparability of the information because of its cultural and contextual specificity – not only across countries but also across major and minor regions and across different population subgroups within the same country. This does not mean that such information should be ignored: on the contrary, it can have great descriptive value. It is our view, however, that it is not suited for wholesale integration into a single index (or even a set of composite indices) of the type we have constructed above. Perhaps it is best to retain the specificity of such individual items.

Returning to the proposed overall "FS" index, we stress some of its important aspects:

(1) Being based on items which are simply dichotomous ("yes-no") or at best consist only of a few ordered categories, the scaling of the resulting quantitative index is necessarily arbitrary. We have proposed the non-monetary deprivation index to be benchmarked to the widely used and well-understood income poverty rate (HCR). In order to retain flexibility, it is desirable to impose such benchmarking at a high level of aggregation – such as in terms of average values over a period of time (such as over a number of survey waves), and over a group of countries such as EU15, or even EU25 when comparable survey data for all become available. Furthermore, by incorporating the country-specific HCR into the functional form for this index (see above), it has been empirically found that the above benchmarking automatically makes the magnitude of the resulting monetary and non-monetary measure approximately equal at the country level as well.

(2) As already noted, an index of non-monetary deprivation should be supplemented by a set of indicators of deprivation in specific dimensions and aspects which are not suitable for incorporation into a single, overall index.

(3) An index of non-monetary deprivation – even though it summarises a range of indicators of living conditions – should be developed and analysed in its own right, separately from measures of income poverty. The same applies in reverse to measures of income poverty in relation to non-monetary measures.[17]

(4) Nevertheless, it is also useful to combine the two types of measures in order to study the extent to which they overlap or are disjoint, as described in Section 6.6 below. An important difference from some previous approaches is that we first

[17] According to one view, "poverty" is understood to be essentially income poverty – non-monetary deprivation in this view being a *consequence* of poverty. An alternative view sees "poverty" as deprivation in terms of living conditions, with low income as a *cause* - indicating being "at-risk-of-poverty" as in Laeken terminology. Perhaps it is more appropriate to view the two as equally important dimensions of social disadvantage.

construct deprivation measures separately for income and non-monetary domains – whether the latter includes all individual items or only items specific to a particular dimension of deprivation – and integrate the two domains later using an appropriate methodology. This is to ensure that the income poverty – which obviously is a primary determinant of overall deprivation – retains its influence in the analysis. Including it simply as one of the many items of deprivation would not ensure that.

(5) At the same time, it is useful to decompose the overall index of non-monetary deprivation into underlying dimensions, as noted in Section 6.2 above. The procedure we have used in outlined in the next section.

6.4 Dimensions of non-monetary deprivation.

Separate indices have been constructed for five dimensions or domains of non-monetary deprivation. These dimensions are: (1) basic non-monetary deprivation; (2) secondary non-monetary deprivation; (3) lacking housing facilities; (4) housing deterioration; and (5) environmental problems

Since sufficient numbers of individual items are not available for separate dimensions in all countries, dimension-specific propensities have been estimated simply by using the same value parameter α as determined for the overall (all item) deprivation index:

$$FS_{\delta,j} = \left(V_{(FS)\delta,j}\right)^{\alpha/H_{C,w}}.$$

This means that the average value of the deprivation index is allowed to vary from one dimension to another, reflecting the relative prevalence of each. As noted, this choice is dictated by limited data availability on individual dimensions – a constraints which becomes all the more important as we move down to the regional level.

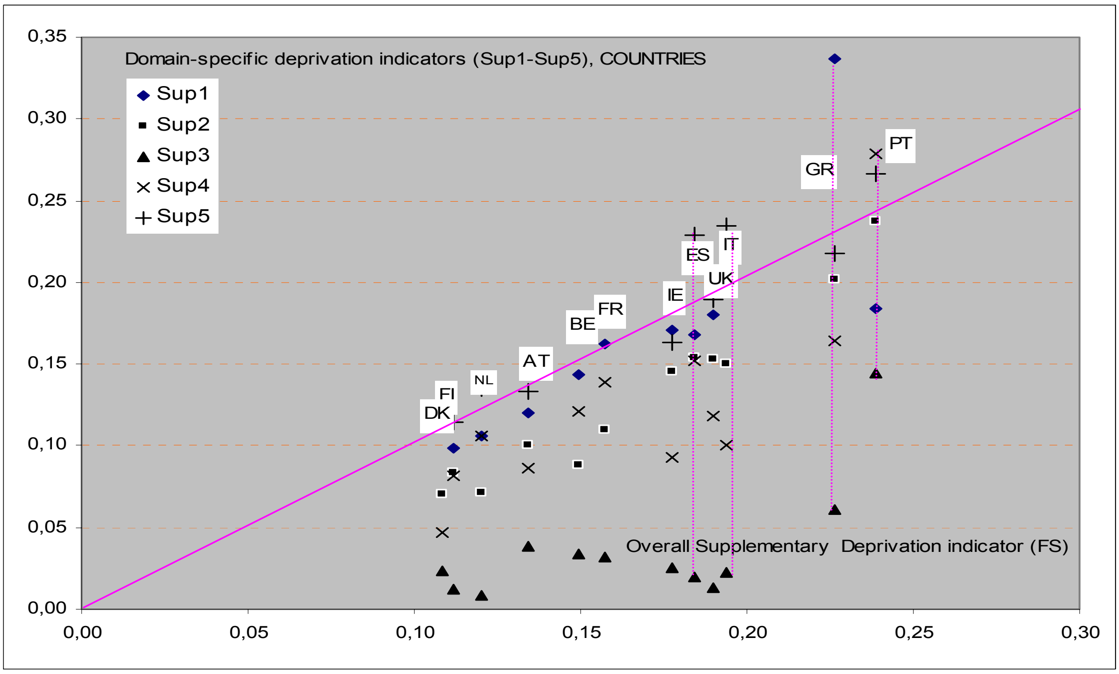
Table 6.1 shows the overall index of non-monetary deprivation, and indices in particular dimensions by country. Figure 6.1 shows the dimensional indices in individual dimensions plotted against the overall non-monetary index of deprivation.[18]

[18] As in most tables, countries are arranged according to national monetary poverty rate (HCR) as shown in **Table 5.2** above. These rates have been consolidated over ECHP waves and over three different poverty line thresholds as explained in the previous chapter. The fuzzy supplementary measures are scaled to equal the corresponding monetary measures at EU15, as explained earlier.

Table 6.1 Non-monetary deprivation: overall index and indices in specific dimensions or domains

	Non-monetary deprivation: overall index									Indices in specific dimensions				
ECHP wave	1	2	3	4	5	6	7	8	mean	FSUP1	FSUP2	FSUP3	FSUP4	FSUP5
FI			9,8	9,3	10,7	12,1	12,3	11,0	10,9	12,5	7,0	2,3	4,7	12,6
DK	11,1	10,6	10,6	9,9	12,8	12,8	11,2	10,6	11,2	9,9	8,3	1,2	8,1	11,5
SE				9,9	10,8	10,7	11,8	11,3	10,9					
NL	11,4	12,8	12,5	11,7	11,0	11,3	11,3	14,2	12,0	10,6	7,1	0,9	10,6	13,6
LU		11,5	7,7	7,7	12,9	13,1	11,6	12,9	11,1					
DE	13,6	13,7	11,9	7,5	11,3	7,7	11,2	10,5	10,9					
AT		14,5	14,2	14,3	14,0	13,0	12,0	12,1	13,4	12,0	10,0	3,8	8,7	13,3
BE	17,3	16,7	15,6	15,3	14,5	13,2	14,1	12,9	15,0	14,4	8,8	3,4	12,1	17,1
FR	17,2	16,1	16,0	15,6	15,3	15,2	15,3	15,4	15,8	16,2	10,9	3,2	13,8	18,7
IE	16,4	17,4	18,0	17,9	17,9	18,3	18,5	17,8	17,8	17,1	14,5	2,5	9,3	16,4
UK	21,4	21,1	19,6	17,8	18,6	18,2	18,1	16,8	19,0	18,0	15,3	1,3	11,9	19,0
ES	19,2	19,0	17,7	19,3	17,7	18,8	17,8	18,0	18,4	16,8	15,4	2,0	15,2	22,9
IT	20,4	20,4	19,5	19,7	18,4	17,9	18,9	19,5	19,4	21,2	15,0	2,3	10,1	23,5
PT	25,0	26,1	24,1	24,2	23,6	22,1	23,1	22,8	23,9	18,4	23,7	14,4	27,9	26,6
GR	25,1	23,7	22,9	23,6	22,1	21,2	21,0	21,6	22,6	33,7	20,2	6,1	16,5	21,8

Figure 6.1 Non-monetary deprivation: domain-specific deprivation indicators in relation to the overall indicator



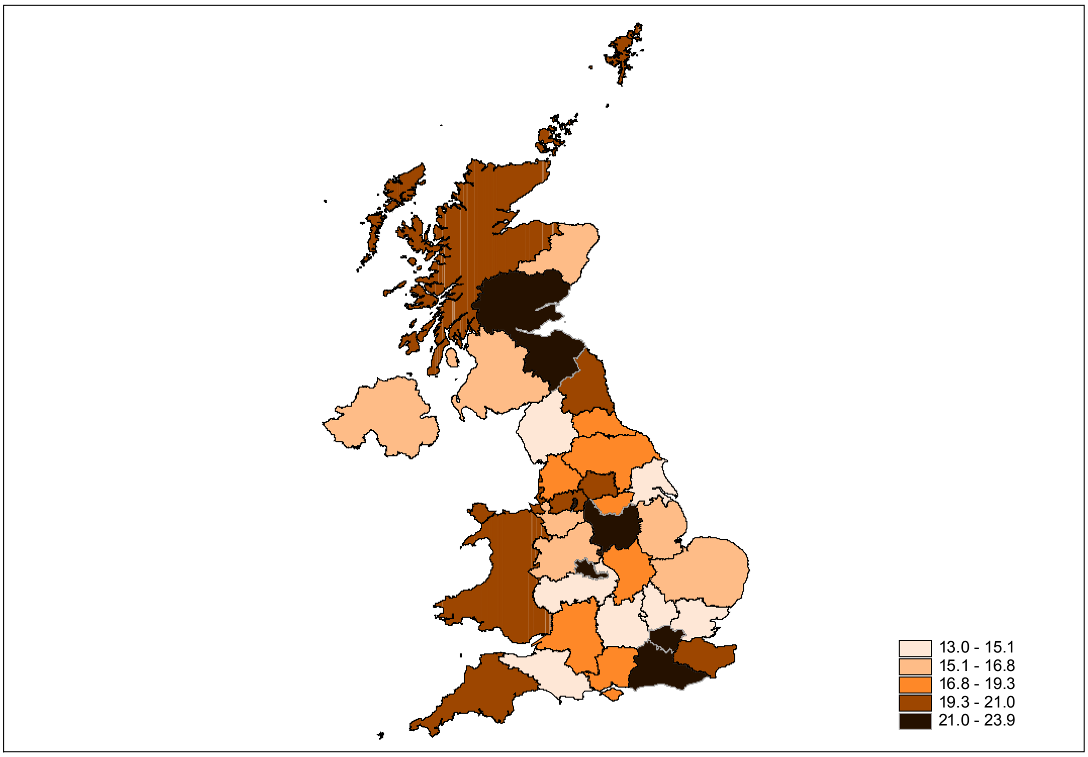
Generally the values are quite stable over waves, justifying taking an average over waves to obtain more stable results. (The results show rather marked year-to-year fluctuations in the case of Luxembourg and Germany, however. This is due to small sample sizes or possibly due to data errors.)

The pattern across countries for different domain indicators is not the same. Differentials across countries are often more pronounced in dimension-specific indicators than in the overall non-monetary deprivation index.

Illustrative figures are presented only at the country level here. This is to illustrate the meaning and magnitude of the indicators involves. The real interest of course is in indicators disaggregated to the regional level, such as to NUTS2. Some more results will be presented in Chapter 8.

In country data sets where code for the identification of regions is available, direct estimate for the index can be made from the survey data. These can then be improved with small area estimation techniques (see Chapter 11). As an illustration, values for the non-monetary index estimated in this way for the United Kingdom are displayed in Figure 6.2. As in all other such maps, the map shows quintiles of the available units (in this case NUTS2 regions in the UK) according to the level of the index.

Figure 6.2 Overall non-monetary ('Fuzzy Supplementary') deprivation rates. United Kingdom. NUTS2 regions



6.5 Technical note: weights for the aggregation over items

As noted above, in our procedure the weights to be given to items in the aggregation are determined *within each dimension* (group of items) separately. Also, the set of weights W_k are taken to be item-specific; for a given item they are common to all individuals (j) in the country. For each item and each country, the weights are in fact determined wave by wave and then averaged over waves. The weighting procedure is based on the following statistical considerations taking into account how the items are distributed in the population.

Firstly, the weight is determined by the variable's power to differentiate among individuals in the population, that is, by its dispersion. We take this as proportional to the coefficient of variation of deprivation score $d_{k,j}$ for the variable concerned: $w_k^a \propto cv_k$. This means that for a small proportion (say, d), the weight varies inversely to the square-root of the proportion. Thus deprivations which affect only a small proportion of the population, and hence are likely to be considered more critical, get larger weights; while those affecting large proportions, hence likely to be regarded as less critical, get smaller weights. Note, however, that the contribution of the deprived individuals to the average values of deprivation in the population resulting from the item concerned turns out to be directly proportional to the square-root of d. In other words, deprivations affecting a smaller proportion of the population are treated as more intense at the individual person's level but, of course, their contribution to the average level of deprivation in the population as a whole is correspondingly smaller.

Secondly, in order to avoid redundancy, it is necessary to limit the influence of those characteristics that are highly correlated with the others included in the analysis. Even for the overall index, it is reasonable to consider this correlation separately within each of the five dimensions of deprivation identified earlier, i.e., the weight of variable k in deprivation dimension δ is taken as the inverse of an average measure of its correlation with all the variables in that dimension. Thus the results are not affected by arbitrary inclusion or exclusion of items highly correlated with other items in the set. An average measure of the correlation is computed as:

$$w_k^b \propto \left(\frac{1}{1+\sum_{k'=1}^{K} \rho_{k,k'} \mid \rho_{k,k'} < \rho_H} \right) x \left(\frac{1}{\sum_{k'=1}^{K} \rho_{k,k'} \mid \rho_{k,k'} \geq \rho_H} \right),$$

where $\rho_{k,k'}$ is the correlation between the two indicators. In the first factor of the equation, the sum is taken over all indicators whose correlation with the variable k is less than a certain value ρ_H (determined, for instance, by dividing the ordered set of correlation values at the point of the largest gap). The sum in the second term always includes the case k' = k, since that correlation coefficient is 1.0. The motivation for this model is that (i) w_k^b is not affected by the introduction of variables entirely uncorrelated with k; (ii) it is only marginally affected by small correlations; but (iii) is reduced proportionately to the number of highly correlated variables present.[19]

To surmise, the weight given to an item is directly proportional to the variability of the item in the population and inversely proportional to its correlation with other items in the deprivation dimension to which it belongs. The final weight is taken as the product of the two factors: $W_k \propto w_k^a . w_k^b$. The scaling of the weights can be arbitrary, though scaling them to sum to 1.0 may be convenient.

[19] In practice we have mostly found that, on the basis of the 'largest gap' criterion, the second factor involves only the variable itself (i.e., is reduced to 1), so that the weight of a variable is simply the inverse of the average of correlations with all the variables (including the variable concerned itself).

6.6. Income poverty and non-monetary deprivation in combination

If individuals are subject both to income poverty and non-monetary deprivation simultaneously, their overall deprivation is more intense. Similarly, if they are subject to only one of the two, their deprivation is, in relative terms, less intense.

In our numerical illustrations, we have developed and implemented these measures in terms of "fuzzy propensities" to income poverty and non-monetary deprivation. The two measures - FM_j propensity to income poverty, and FS_j the overall non-monetary deprivation propensity - may be combined to construct composite measures which indicate the extent to which the two aspects of income poverty and non-monetary deprivation overlap for the individual concerned. These measures are as follows.

M_j manifest deprivation, representing the propensity to both income poverty and non-monetary deprivation simultaneously. One may think of this as the '**m**ore intense' degree of deprivation.

L_j latent deprivation, representing the individual being subject to at least one of the two, income poverty and/or non-monetary deprivation; one may think of this as the '**l**ess intense' degree of deprivation.

Once the propensities to income poverty (FM_j) and non-monetary deprivation (FS_j) have been defined at the individual level (j), the corresponding combined measures are obtained in a straightforward way, which can then be aggregated to produce the relevant averages or rates for the population. The manifest deprivation propensity of individual j is the intersection (the smaller) of the two measures FM_j and FS_j:

$$M_j = \min(FM_j, FS_j).$$

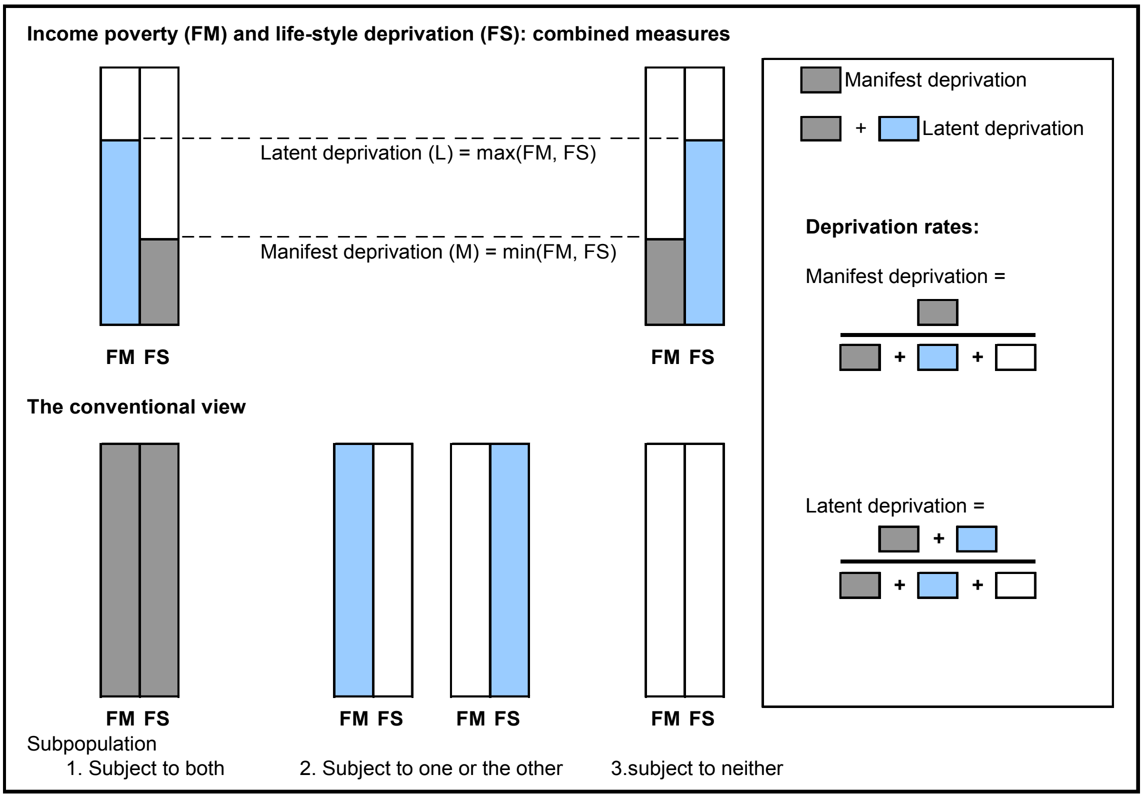
Similarly, the latent deprivation propensity of individual *j* is the union (the larger) of the two measures FM_j and FS_j:

$$L_j = \max(FM_j, FS_j).$$

The composite measures defined above can also be constructed using conventional rather than "fuzzy set" concepts. In fact the former are merely a special case of the latter, perhaps easier to comprehend but somewhat less satisfactory in other respects. Consider persons with $FM_j \leq \overline{FM}$ as *income poor*, and those with $FS_j \leq \overline{FS}$ as subject to non-monetary deprivation ($\overline{FM}, \overline{FS}$ being the mean values of FM_j, FS_j, respectively). Then those satisfying both the conditions are subject to, what we have called, manifest deprivation. Those satisfying either of these two conditions are subject to latent deprivation as defined above.

The relationship between fuzzy and conventional conceptualisations is illustrated in Figure 6.3.

Figure 6.3 Manifest and Latent measures: illustration of the "fuzzy" versus the "conventional" concept



The upper part of the figure shows individuals subject to income (or fuzzy monetary, FM) and non-monetary (or fuzzy supplementary, FS) deprivation to varying degrees. The larger of these equals the degree of the person's latent deprivation, and the smaller of the two equals the degree of manifest deprivation. The corresponding rates for a population are the averages of these individual level degrees. The lower part of the figure illustrates the conventional view in which individuals are classified into non-overlapping categories, and a rate is computed as the proportion of persons in a particular category.

As an illustration, Table 6.2 shows indices of income poverty and non-monetary deprivation in combination for Portugal for NUTS2 regions.[20] These are displayed in Figure 6.4 Also shown in the table are indices for each of the five dimensions of non-monetary deprivation. A few remarks may be made on these results.

The first two columns on Table 6.2 concern monetary poverty, being respectively the 'conventional' and 'fuzzy' measures of the poverty rate. The latter measure is scaled to be identical to the former at EU15 level as explained earlier, and this also ensure their approximate equality at the country level. The third column is the "fuzzy supplementary" measure of overall non-monetary deprivation, and is also scaled to match the preceding two measures at EU15 level.

The combination measures indicate the degree to which income poverty and non-monetary deprivation overlap for individuals in the population. Latent deprivation measures the presence of either and manifest deprivation measures the simultaneous presence of both. As shown in Figure 6.4, the two measures of overlap are placed symmetrically around the

[20] All figures are computed with reference to the *country poverty line*, defined as 60% of the country median.

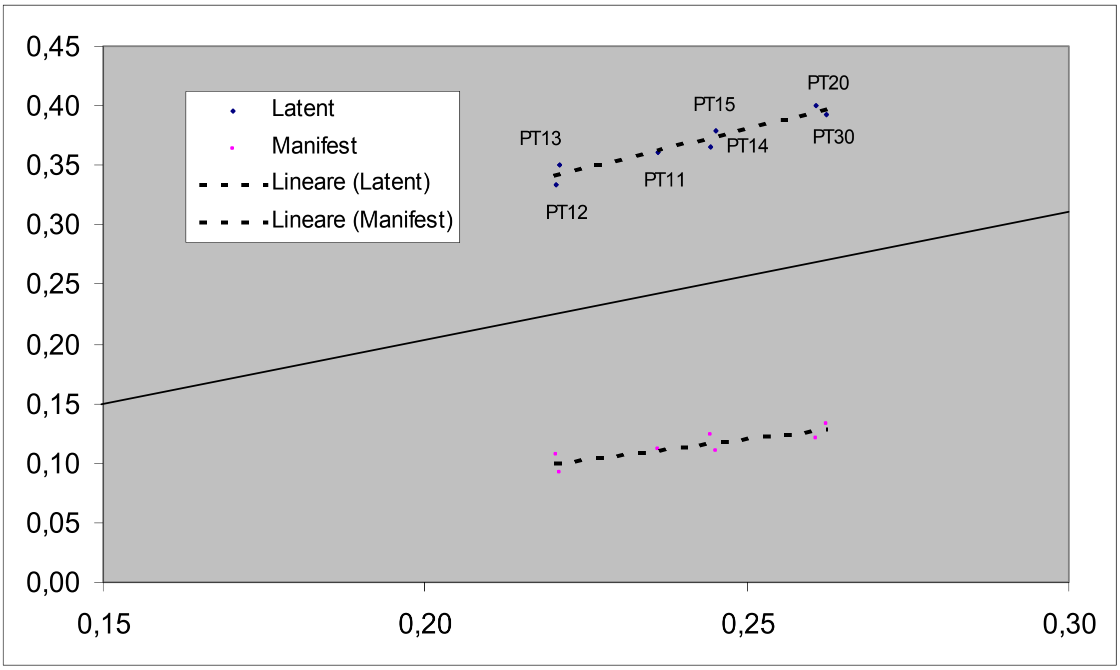
income poverty (or non-monetary deprivation) rate, and the pattern of results is very regular. This is the case generally, as will be seen in Chapter 8 where more detailed results are presented.

Table 6.2 Indices of income poverty and non-monetary deprivation in combination: An illustration from Portugal - NUTS2 regions

	Income poverty		Non-monetary deprivation	In combination		Non-monetary deprivation dimension (1-5)				
				Either ('latent')	Both ('manifest)					
	HCR_C	**FM_C**	**FS_C**	**LAT_C**	**MAN_C**	**FSUP-1**	**FSUP-2**	**FSUP-3**	**FSUP-4**	**FSUP-5**
Region										
PT	22,2	22,3	23,9	35,5	10,7	18,4	23,7	14,4	27,9	26,6
PT1	20,6	21,7	24,1	35,2	10,6	17,6	23,4	14,4	28,0	27,5
PT11	21,2	22,5	25,5	36,9	11,5	19,0	30,4	18,6	29,3	23,9
PT12	28,0	28,1	18,3	34,1	11,1	17,1	25,2	17,9	24,5	15,9
PT13	14,3	16,0	26,8	35,5	9,4	16,7	15,7	5,3	28,6	40,1
PT14	25,7	26,5	22,8	37,0	12,6	18,8	26,3	29,3	29,0	14,3
PT15	33,9	32,5	18,1	38,5	11,4	16,3	20,0	20,4	24,6	15,9
PT20	35,7	34,3	18,1	40,2	12,3	34,4	29,2	11,8	30,7	8,2
PT30	34,2	33,2	19,4	39,3	13,4	32,3	30,9	16,0	21,6	11,5

The direct survey estimates were consolidated over 8 ECHP waves. The table actually shows 'composite estimates', constructed by combining direct survey estimates with small area modelling (Chapters 11 and 12).

Figure 6.4 Latent and Manifest deprivation measures at NUTS2. Portugal



In this graph, y-axis shows Manifest and Latent deprivation rates; x-axis is the mean of these two rates, which in principle approximately equals the income poverty rate, and also equals the non-monetary deprivation rate which is scaled to equal the income poverty rate. As in Table 6.2, composite estimates have been used.

Figures 6.5 and 6.6 display Latent and Manifest deprivation rates across NUTS2 regions of the United Kingdom. These should be compared with Figure 6.2 for 'Fuzzy Supplementary' indicators. The pattern in that figure in fact approximates the mean of Figures 6.5 and 6.6.

Figure 6.5 Latent deprivation rates. United Kingdom NUTS2 regions

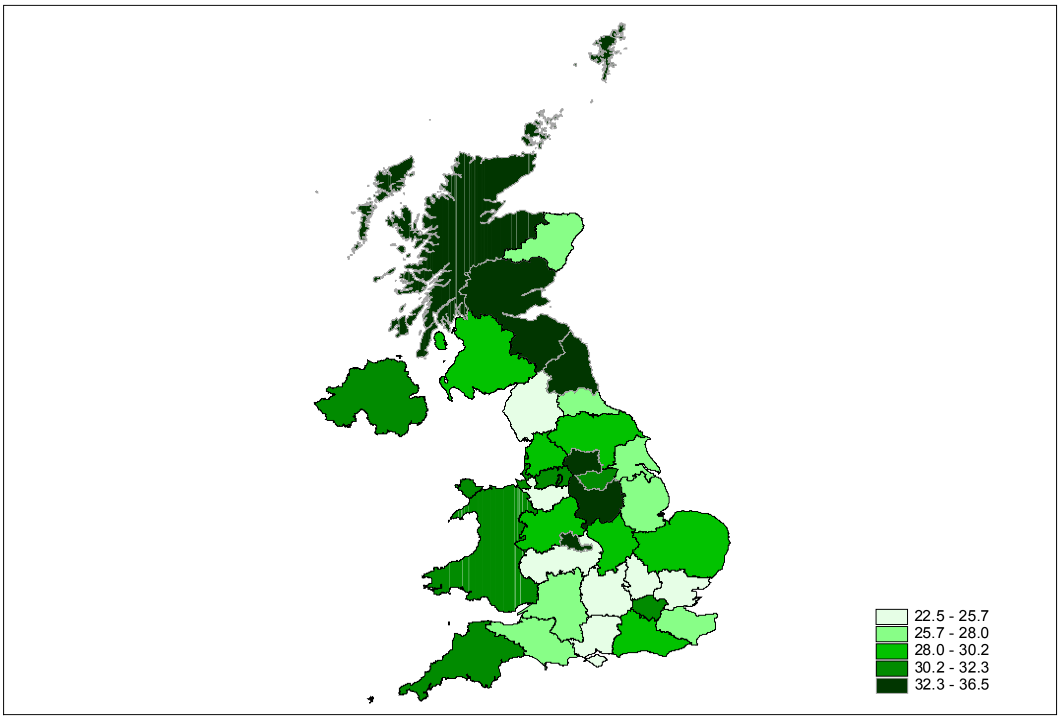
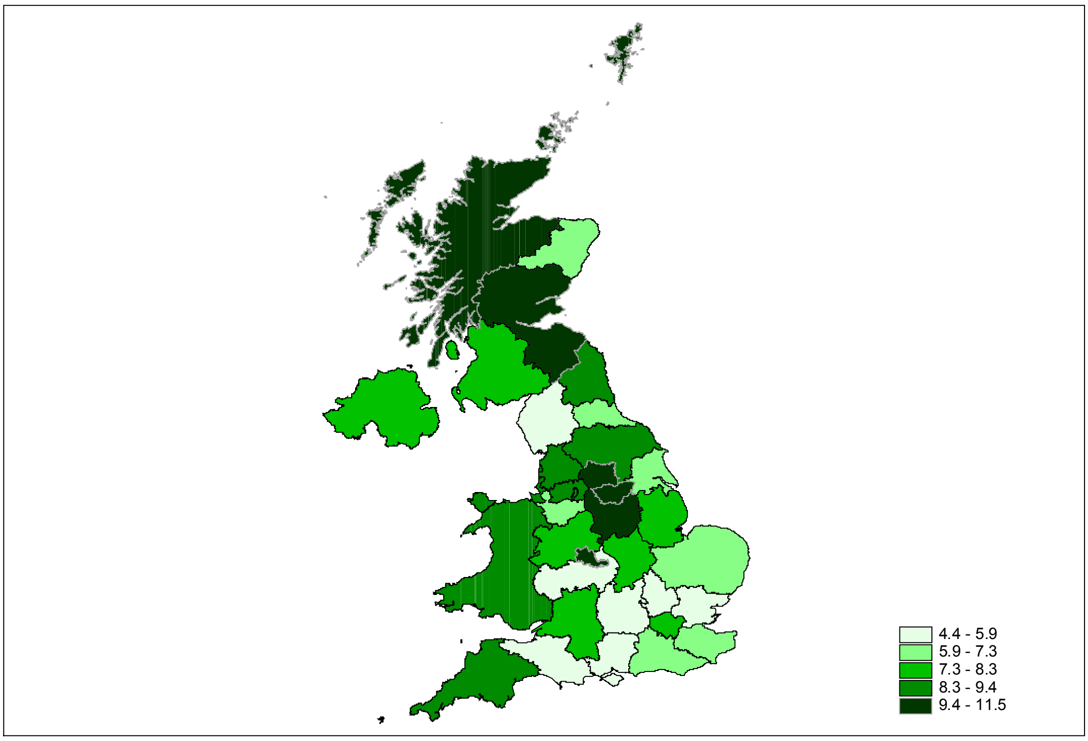


Figure 6.6 Manifest deprivation rates. United Kingdom NUTS2 regions



Chapter 7
Longitudinal indicators of poverty and deprivation

7.1 Limitations: trade-off between spatial and temporal details

The effect of poverty on a person or a household is directly related to the period they are poor. If people's experience of marginality and want is only temporary, their life-chances will be less seriously impaired. A persistent risk of poverty, on the other hand, is more likely to be associated with the erosion of resources and a qualitatively different experience of deprivation.

In the longitudinal dimension, indicators may be designed to capture the experience of poverty and deprivation at any time during a period, or persistently or continuously over the period.

For the development of longitudinal indicators of poverty and deprivation for use at the regional level, the starting point of course is the specified set of Laeken indicators in this area. These indicators have been summarised in Section 2.4. Recommendations on the set of indicators on longitudinal income poverty (as well as concerning longitudinal non-monetary deprivation) have been presented there.

To summarise, we need to consider somewhat simplified longitudinal indicators in the regional context. One can expect that simpler indicators will be more robust and less demanding on the data available. The main simplification we propose is to focus on *longitudinal indicators defined over a short time period*. Where the available statistical data cover a longer time period, those longitudinal indicators can themselves be averaged over time to obtain more robust measures. In our illustrations below, we have averaged this indicator over up to 8 waves where ECHP data were available.

In specific terms, we define and construct in the following illustrations indicators based on the persistence of poverty over *pairs of adjacent years*:

- Persons are *persistently poor* over two consecutive years if, in relation to the poverty line specific to each of the years, they are classified as poor in *both* the years.
- Persons are in *any-time poverty* over two consecutive years if, in relation to the poverty line specific to each of the years, they are classified as poor in *either* of the years.

The measures of income poverty described above can all be generalised to multi-dimensional measures of deprivation of the type discussed in Chapter 6 (any-time, persistent or continuous incidence of supplementary, latent and manifest forms of deprivation). Some illustrations are given in Section 7.3.

It may be appropriate where possible to supplement the above 'basic' longitudinal measures with additional indicators – but the latter considered as 'secondary' in relation to the former.

(1) An indicator similar to Laeken Indicator 3 can be useful to identify longer-term poverty. However, for reasons of data constraints for regional estimation, it would be preferable not to tie the measure to a particular year. This is discussed further in

Section 7.3 below, where we introduce the concepts of any-time, persistent and continuous poverty with a longer reference period.

(2) The measures can be constructed using the 'traditional' approach, which classifies the population into distinct, non-overlapping categories for the purpose of constructing the longitudinal indices. More refined measures can be constructed using the fuzzy-set approach introduced in Chapter 5.

(3) Generally for regional analysis, the direct estimates from a survey need to be made more precise through modelling and small area estimation procedures, such as those presented in Chapters 11-14 below.

It is important to note that *all results presented in this chapter are direct estimates from survey data*. This is in contrast to the modelled results for most other measures in other chapters.

7.2 Persistent poverty over adjacent pairs of years

Some results are presented below on the 'basic' indicators on the persistence of poverty over *pairs of adjacent years*. As noted, we define as *persistently poor* over two consecutive years persons classified as poor in *both* the years. Persons poor in *either* of the years are defined to be in *any-time poverty* over the two consecutive years.

We have computed these measures to NUTS2 level in countries where the survey data provide the necessary area code for the identification of regions. Unfortunately, with ECHP data this has been possible only in the case of three surveys: Italy, the United Kingdom and Portugal. (The necessary code is available in the national surveys for Poland and Romania; however, no longitudinal poverty measures could be computed for those countries because usable data were available to us for only one wave in each country.) The results have been averaged over 8 years where the ECHP data were available, and for fewer years when the data were lacking.

For reasons of time, we have not applied the small area estimation (SAE) procedures to these results, as has been done for a number of other measures presented in this report. In any case, the national level results presented in Tables 7.1 and 7.2 are not in any way affected by this since SAE is required only when we move down to regions. Actually, with the sample sizes available in our data especially when the results are aggregated over many years, the introduction of SAE generally does not affect the final results greatly even at NUTS1 level.

Table 7.1 Percentages in persistent poverty over two adjacent years – by country. EU15

	In persistent poverty over pairs of adjacent years:								consolidated cross-sectional
	yr 1-2	yr 2-3	yr 3-4	yr 4-5	yr 5-6	yr 6-7	yr 7-8	mean	HCR*
FI			4,9	5,1	6,6	7,1	7,1	6,2	10,1
DK	5,2	5,0	4,9	5,2	7,2	6,7	5,3	5,6	10,3
SE				0,8	1,0	1,1	1,3	1,1	10,9
NL	6,0	6,3	6,5	5,7	5,6	5,9	6,5	6,1	11,2
LU		8,3	8,1	9,0	9,0	9,2	8,5	8,7	12,1
DE	8,4	8,1	7,7	6,9	7,1	6,8	7,4	7,5	12,4
AT		7,5	7,9	8,5	8,0	7,2	7,2	7,7	12,5
BE	9,8	9,7	9,3	8,7	8,0	7,9	9,2	8,9	14,1
FR	9,3	9,9	9,4	9,5	10,1	10,4	10,3	9,8	15,1
IE	11,8	13,8	13,3	12,5	13,4	13,7	15,0	13,4	18,0
UK	12,5	12,6	11,6	12,0	13,0	12,7	11,3	12,2	18,6
ES	12,3	10,9	11,8	12,2	11,3	11,4	11,8	11,7	18,9
IT	12,7	12,8	12,3	12,3	12,0	12,8	13,7	12,7	19,5
PT	16,1	16,0	14,9	15,7	15,3	14,9	15,0	15,4	21,3
GR	13,2	14,1	14,3	13,4	14,0	14,2	14,4	13,9	21,5

* see Table 5.2

Countries are arranged according to the average cross-sectional income poverty rate presented in Chapter 5. (Generally, this ordering is followed throughout this report.) It can be seen that overall the pattern is very consistent over time and also across countries. Particularly interesting is the pattern in the last three columns of Table 7.2 which summarises the results. In Greece for instance, 28% of the population is in the state of poverty[21] for at least one of each pair of adjacent years; on the average, around 21% in the state of poverty at a given time; and 14% of these are in poverty persistently over the two years. These are the highest figures observed for EU15. The figures may be substantially lower in other countries, but the relative patterns are generally quite similar.

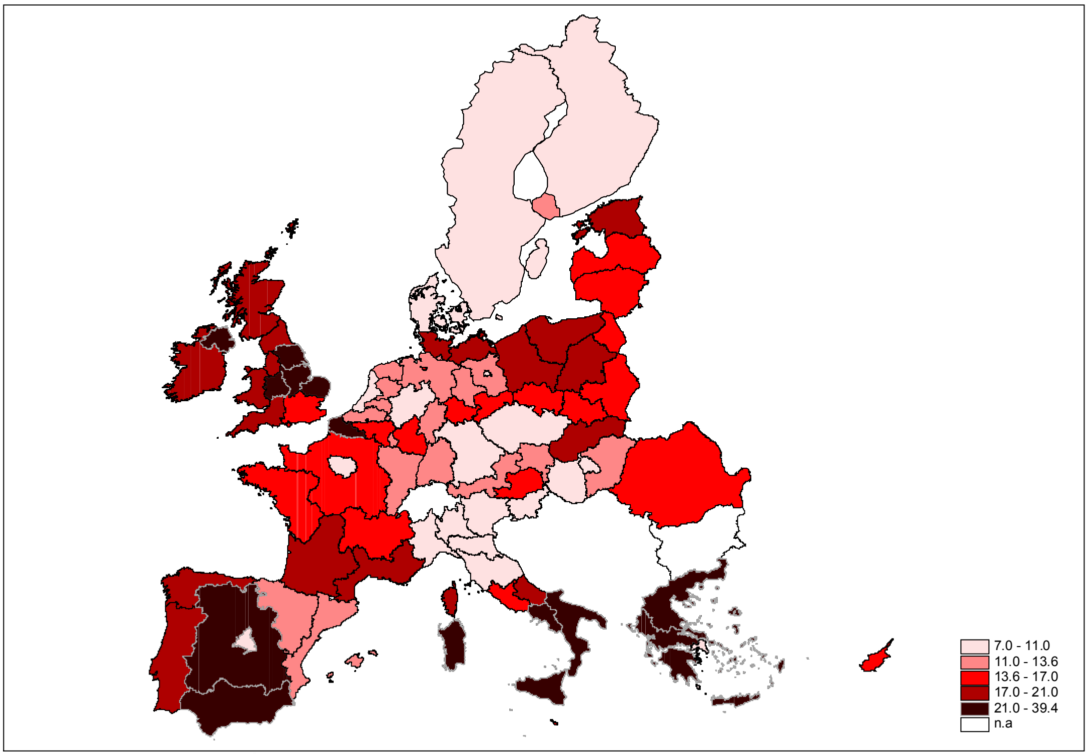
Table 7.2 Percentages in poverty any time over two adjacent years – by country. EU15

	In "any-time poverty" over pairs of adjacent years:								consolidated cross-sectional	persistent over two
	yr 1-2	yr 2-3	yr 3-4	yr 4-5	yr 5-6	yr 6-7	yr 7-8	mean	HCR*	years**
FI			11,0	12,0	13,0	14,2	14,9	13,0	10,1	6,2
DK	13,8	13,6	13,4	15,2	15,2	14,7	16,4	14,6	10,3	5,6
SE				17,5	18,2	19,1	19,6	18,6	10,9	1,1
NL	14,0	15,7	15,6	13,9	14,0	14,2	14,2	14,5	11,2	6,1
LU		16,7	14,1	14,9	15,0	15,8	14,8	15,2	12,1	8,7
DE	19,6	19,1	17,7	15,4	14,5	14,1	13,6	16,3	12,4	7,5
AT		18,8	18,3	16,6	16,9	16,5	15,9	17,2	12,5	7,7
BE	22,1	21,4	19,9	18,2	17,5	16,8	16,9	19,0	14,1	8,9
FR	21,5	19,8	19,7	18,7	19,5	19,9	20,3	19,9	15,1	9,8
IE	22,8	24,2	25,7	25,3	25,3	25,6	26,5	25,0	18,0	13,4
UK	24,9	25,0	24,2	24,0	24,5	23,4	22,3	24,0	18,6	12,2
ES	27,2	26,3	26,2	26,4	25,6	24,9	25,2	26,0	18,9	11,7
IT	27,4	26,4	26,8	25,0	23,5	23,2	24,2	25,2	19,5	12,7
PT	29,6	27,8	28,0	28,3	26,6	26,6	26,8	27,7	21,3	15,4
GR	30,4	28,1	28,0	28,6	28,3	26,3	26,1	28,0	21,5	13,9

* see Table 5.2 ** see Table 7.1

21 Unless otherwise specified, poverty is defined as equivalised income below 60% of the national median.

Figure 7.1 Head Count Ratio NUTS1 regions (cross-sectional, averaged over waves, country poverty line)



Figures 7.2 and 7.3 show the rates of persistent and any-time poverty measured over pairs of adjacent years in the same manner as above, but at NUTS1 level. For comparison, Figure 7.1 gives the averaged cross-sectional poverty rates. The remarkable thing is the high level of consistency of the patterns across these three types of measures (any-time, cross-sectional, and persistent poverty rates over 2-year periods). In other word, the *relative* position of EU15 regions is very similar whichever measure is considered.

Tables 7.3 and 7.4 show the rates of persistent and any-time poverty measured over pairs of adjacent years in the same manner as above, but at NUTS2 level. We have taken the results from the UK for illustration.

It is much more important to note here that *these figures are based on direct survey estimates*. Given the smallness of the sample size available for most NUTS2 areas, it is expected that figures fluctuate from one pair of years to another, especially for small regions (by design, the UK sample is distributed proportional to population size). However, the pattern becomes much more stable when averages are taken over several pairs of years – in the case of most ECHP surveys with T=8 waves, over (T-1)=7 pairs of consecutive years. This can be seen clearly from the last column of Table 7.4 which shows variation over NUTS2 regions of the ratio of persistent to any-time poverty rate. This ratio varies in the narrow range of 0.33-0.67 around its overall mean of 0.50, with a coefficient of variation of only 15%. As will be described in Chapter 11, cumulation over time – even in the case of a panel survey such as the ECHP – can result in a substantial reduction in sampling error. The gains can be expected to be even larger with a rotational design such as the one proposed for EU-SILC 'integrated survey' (Verma, 1981), which a majority of EU countries have opted for.

Figure 7.2 Percentages in persistent poverty over two adjacent years. NUTS1 regions

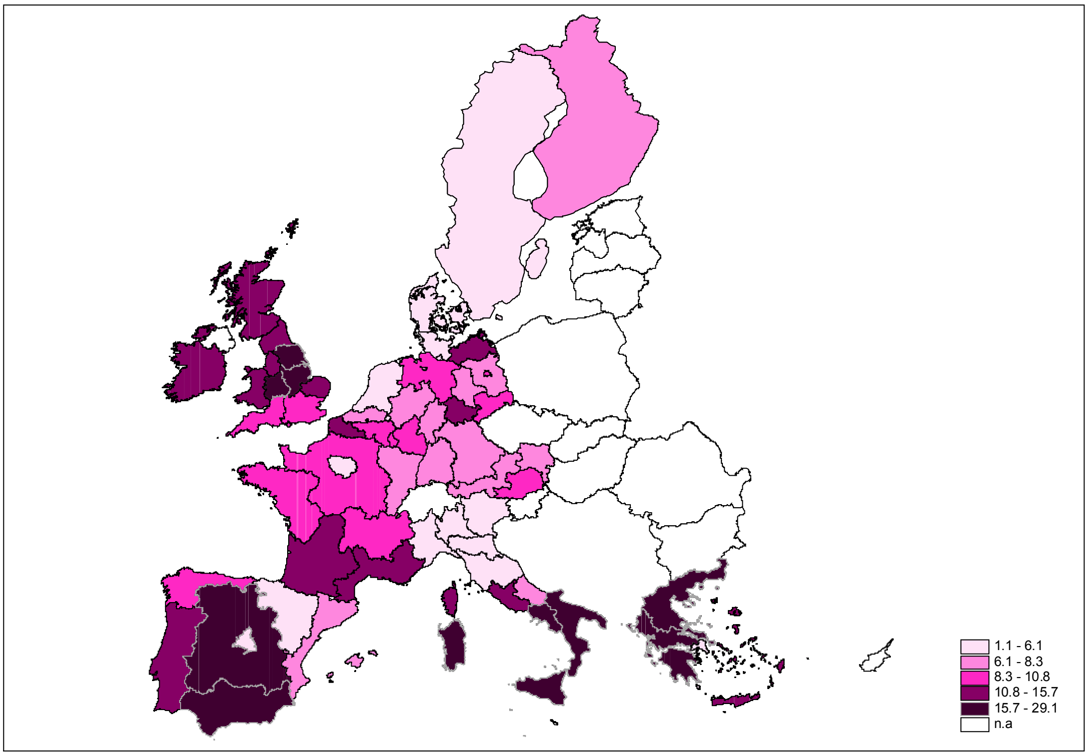


Figure 7.3 Percentages in poverty any time over two adjacent years. NUTS1 regions

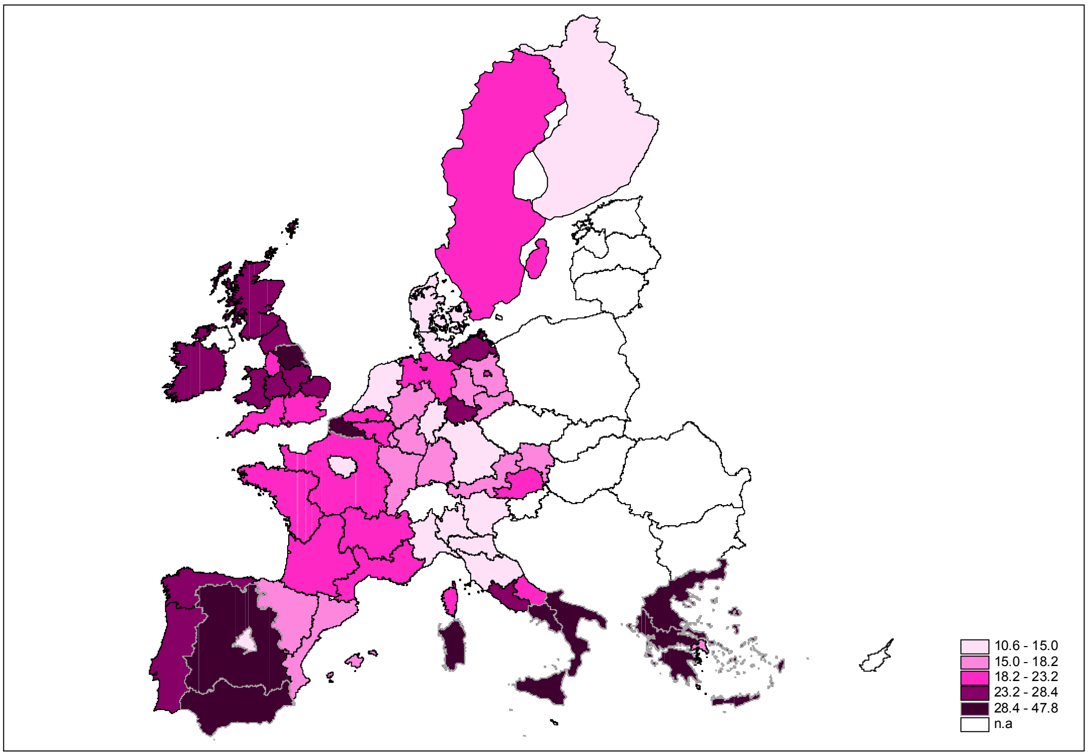


Table 7.3 Percentages in persistent poverty over two adjacent years - United Kingdom. NUTS2 regions

	In persistent poverty over pairs of adjacent years:							
	yr 1-2	yr 2-3	yr 3-4	yr 4-5	yr 5-6	yr 6-7	yr 7-8	mean
UK11	10	13	9	9	4	11	10	9
UK12	13	12	13	14	16	13	9	13
UK13	12	15	16	21	19	19	16	17
UK21	16	15	16	21	13	11	8	14
UK22	15	12	12	15	18	18	19	16
UK23	22	20	24	19	22	17	17	20
UK24	17	22	16	14	19	16	10	16
UK31	17	21	20	19	20	17	17	19
UK32	7	10	10	14	17	13	11	12
UK33	6	3	2	14	13	21	20	11
UK40	15	17	16	14	15	13	14	15
UK51	9	10	12	12	19	17	4	12
UK52	10	9	6	5	6	7	4	7
UK53	10	8	12	10	8	13	7	10
UK54	7	6	4	6	7	7	10	7
UK55	12	10	8	10	9	9	9	9
UK56	12	8	7	8	8	6	13	9
UK57	5	3	4	3	8	5	7	5
UK61	10	12	11	10	12	10	11	11
UK62	16	11	6	9	5	11	11	10
UK63	10	13	11	10	10	14	7	11
UK71	6	9	10	5	7	5	10	7
UK72	10	16	13	18	14	13	20	15
UK73	22	24	23	23	22	23	20	22
UK81	12	11	8	8	10	10	7	9
UK82	14	14	11	12	13	13	11	13
UK83	17	9	6	8	10	11	6	10
UK84	14	16	10	9	15	8	9	12
UK91	13	11	12	18	20	24	20	17
UK92	12	8	11	10	9	7	7	9
UKA1	15	13	14	14	16	14	12	14
UKA2	9	10	11	9	16	18	13	12
UKA4	15	13	8	11	14	19	5	12

7.3 Secondary longitudinal indicators

As noted in Section 7.1, we recommend to focus on *longitudinal indicators defined over a short time period*, specifically persistent and any-time poverty defined in relation to pairs of consecutive years, with the provision that where the available statistical data cover a longer time period, those longitudinal indicators can themselves be averaged over time to obtain more robust measures. It was also noted that it would be appropriate where possible to supplement the above 'basic' longitudinal measures with additional 'secondary' indicators.

Table 7.4 Percentages in poverty any time over two adjacent years - United Kingdom. NUTS2 regions

	In "any-time poverty" over pairs of adjacent years: yr 1-2	yr 2-3	yr 3-4	yr 4-5	yr 5-6	yr 6-7	yr 7-8	mean [1]	persistent [2]*	ratio [2] / [1]
UK11	22	22	20	14	16	22	19	19	9	0,49
UK12	19	24	25	27	25	27	20	24	13	0,54
UK13	25	29	30	31	32	25	24	28	17	0,60
UK21	24	25	25	26	27	20	18	23	14	0,61
UK22	19	24	28	28	30	30	32	27	16	0,58
UK23	33	38	37	36	32	29	32	34	20	0,59
UK24	33	32	29	30	32	26	22	29	16	0,56
UK31	31	34	35	30	32	33	29	32	19	0,58
UK32	19	25	24	26	25	22	19	23	12	0,52
UK33	21	14	19	24	27	29	32	24	11	0,48
UK40	28	28	26	28	27	30	27	28	15	0,54
UK51	22	22	24	25	32	26	27	25	12	0,47
UK52	23	23	22	21	18	16	18	20	7	0,33
UK53	25	24	25	21	17	18	19	21	10	0,46
UK54	17	18	19	16	16	14	15	17	7	0,39
UK55	24	22	20	20	19	16	17	20	9	0,48
UK56	21	17	15	13	12	18	19	16	9	0,53
UK57	13	18	13	16	16	17	15	15	5	0,34
UK61	22	23	25	22	22	23	21	23	11	0,48
UK62	27	25	28	22	22	20	29	25	10	0,41
UK63	26	22	18	17	22	26	25	22	11	0,47
UK71	12	18	20	16	12	13	18	16	7	0,47
UK72	24	26	28	29	26	23	30	27	15	0,55
UK73	34	34	37	33	34	36	27	33	22	0,67
UK81	23	24	21	19	25	26	18	22	9	0,42
UK82	25	25	22	26	25	23	23	24	13	0,52
UK83	27	18	17	21	24	23	16	21	10	0,46
UK84	29	22	22	22	31	23	21	24	12	0,47
UK91	32	31	30	28	29	31	29	30	17	0,56
UK92	27	24	20	19	18	21	21	21	9	0,42
UKA1	27	28	29	32	34	30	25	29	14	0,47
UKA2	22	25	18	28	31	27	25	25	12	0,48
UKA4	34	23	23	16	25	30	16	24	12	0,51
									mean	0,50
									minimum	0,33
									maximum	0,67
									cv	15%

* from Table 7.3

Continuous, persistent, any-time poverty

On lines similar to Laeken Indicator 3, it is useful to introduce the concepts of continuous, persistent and any-time poverty defined over a longer period of time. The choice of the appropriate reference period depends, apart from data availability, on substantive and policy considerations. It is a matter of trade-off between temporal and spatial detail. It would be appropriate to follow the 4-year reference period used in Laeken Indicator 3. However, for reasons noted earlier, for the purpose of regional analysis, it would be preferable not to tie the measure to a particular year.

We propose 'secondary' longitudinal income poverty measures in the following terms.

- For a chosen reference period (such as 4 consecutive years) *continuous poverty* refers to poverty experienced during *all* the years during the period.

- o Over the same period, *any-time* poverty refers to poverty experienced for *one or more* years.
- o Finally we adopt the following definition of *persistent* poverty: it refers to poverty during *at least a majority* of the T years, i.e. for at least T_p years, where T_p = int(T/2) + 1 (i.e. the smallest integer strictly larger than T/2). For instance, for a 4 or 5 year period, persistent refers to poverty for at least 3 years; for T=6 or 7 years, it refers to poverty for at least 4 years; for T=8 years as in ECHP for most countries, it would refer to poverty for at least 5 years.

These measures can be constructed using the 'traditional' approach, which classifies the population into distinct, non-overlapping categories for the purpose of constructing the longitudinal indices. Suppose that for each of T years, a person has been classified as poor (i.e. with equalised income below the poverty line specific to each year) in exactly t years. Then the person's longitudinal status is determined as follows:

Whether the person is in:	YES	NO
Continuous poverty	$t = T$	$t < T$
Persistent poverty	$t \geq \text{int}(T/2)+1$	$t < \text{int}(T/2)+1$
Any-time poverty	$t \geq 1$	$t = 0$

It is in this manner, using the 'traditional' approach, that the numerical illustrations presented below have been constructed.

More refined measures can be constructed using the fuzzy-set approach introduced in Chapter 5. Briefly, it is as follows.

Consider a period of T time points. For each time i there are two complementary cross-sectional sets, "poor" and "non-poor", with membership functions for any individual as a_i and $(1-a_i)$ respectively.

For any number of periods with propensities to poverty (or more general form of deprivation – see below) as (a_i), any value in the range 0 to 1:

- o the propensity to continuous poverty is defined as $C_i=\min(a_i)$;
- o the propensity to any-time poverty is defined as $A_i=\max(a_i)$;
- o the propensity to persistent poverty is given by $\min_P(a_i)$, meaning the P^{th} smallest value in the set, where $P=T-T_p+1$. Continuous and any-time poverty are merely special cases of this with, respectively, (T_P = T, P = 1) and (T_P = 1, P = T).

Obviously, the 'traditional' approach is merely a special case of the above: with function (a_i) taking one of only two values, 0 or 1, for any individual. In either case, average over the population of these individual functions gives the corresponding poverty rate.

It may be pointed out that persistent poverty as defined above is a variant on Laeken Indicator 3, in that in determining whether a person is in persistent poverty over (say) 4 year, it treats all the years in the same way: whether it is the 'current' year or any of the preceding three years. We believe that the choice we have made is more suitable for regional indicators, where it is often necessary to take averages over time, in the form of moving averages for instance.

Generalisation to multi-dimensional measures

The measures of income poverty described above can all be generalised to multi-dimensional measures of deprivation of the type discussed in Chapter 6: giving any-time, persistent or continuous incidence of income, supplementary, latent and manifest forms of deprivation.

In Chapters 5 and 6 above we have described five main measures; in addition to H_j, the conventional income poverty index {0,1}, these include the measures: FM_j, fuzzy income poverty; FS_j, non-monetary deprivation; M_j, manifest deprivation, representing the propensity to both income poverty and non-monetary deprivation simultaneously; and L_j, latent deprivation, representing the individual's propensity to being subject to at least one of the two, income poverty and/or non-monetary deprivation. In addition, the propensity to non-monetary deprivation can be analysed separately in its various dimensions, such as the five dimensions (FSup1-FSup5) identified in Chapter 6. Any of these diverse measures can be studied in the time dimension: both in the cross-sectional and the longitudinal contexts.

Some useful results

Table 7.5 analyses the levels of income poverty in the time dimension. The results are based on a 'balanced panel' consisting of individuals present in all the 8 waves of ECHP. (EU15 countries for which fewer than 8 ECHP are available have been excluded in this analysis.) Only country level results are presented here for illustration and explaining the measures and procedures. As in the case of other indicators discussed earlier, regional measures can be similarly constructed. However, generally they would be subject to large sampling errors, and therefore should be improved through small area estimation methodology.

For the conventional and fuzzy measures separately, four types of rates are shown:

(1) the rate of poverty experienced at any time (at least 1 year) during the 8-year period;

(2) the average of cross-sectional poverty rates over the period;

(3) the persistent poverty rate, meaning poverty for a majority (5 or more out of 8) of the years; and

(4) the continuous poverty rate over the entire period.

The rates sharply decline from (1) to (4): taking a simple average over courtiers, from 35-37% for any-time poverty to only 3-4% for continuous poverty.

Table 7.5 Conventional and fuzzy measures of longitudinal monetary poverty rates

	Conventional head-count ratio (H)				Fuzzy monetary (FM)				Ratio (H / FM)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
3 DK	28.0	9.1	4.5	0.9	23.9	8.0	4.4	1.0	1.17	1.14	1.02	0.90
4 NL	26.8	8.8	5.6	0.6	26.8	9.8	6.9	1.6	1.00	0.90	0.82	0.38
6 DE	26.6	10.2	6.2	1.6	26.2	10.5	6.9	2.3	1.01	0.97	0.89	0.67
8 BE	32.9	12.6	8.5	2.3	29.4	12.0	8.7	2.8	1.12	1.05	0.98	0.82
9 FR	33.4	13.4	9.3	2.3	30.4	13.0	9.3	3.5	1.10	1.04	1.00	0.66
10 UK	39.3	16.5	11.1	2.5	36.9	16.5	12.2	4.2	1.07	1.00	0.90	0.60
11 ES	46.5	18.8	13.3	2.8	43.7	19.4	14.8	4.3	1.06	0.97	0.90	0.65
12 IE	45.2	19.6	15.0	4.4	38.7	17.3	13.6	4.4	1.17	1.13	1.10	1.00
13 IT	42.6	18.7	13.2	3.0	40.2	19.2	14.5	4.8	1.06	0.97	0.91	0.63
14 GR	47.7	22.3	15.6	4.7	45.4	23.1	17.6	6.5	1.05	0.96	0.89	0.73
15 PT	46.5	21.6	18.5	5.3	44.5	22.5	19.6	8.1	1.04	0.96	0.95	0.66
simple average	37.8	15.6	11.0	2.8	35.1	15.6	11.7	4.0	1.08	1.00	0.94	0.70

Poverty/deprivation rates:
(1) Anytime: propensity to poverty/deprivation for at least 1 out of the 8 years of ECHP
(2) Cross-sectional: rate averaged over 8 waves
(3) Persistent: propensity to poverty/deprivation for at least 5 out of the 8 years
(4) Continuous: propensity to poverty/deprivation over all the 8 years of ECHP

Note: The results are for a 'balanced panel', i.e., for the population present in all 8 ECHP waves.

Noteworthy from a methodological point, however, is the difference in the performance of the conventional and the fuzzy approaches, especially concerning the estimated incidence of continuous poverty. It appears that movements in and out of poverty tend to be somewhat over-estimated (and hence the persistent or continuous poverty rates under-estimated) with the conventional approach, presumably because it gives too much weight even to small movements across the poverty line.

For this reason, the 'fuzzy approach' may be preferable in constructing multi-dimensional and longitudinal measures of poverty and deprivation.

Table 7.6 illustrates the usefulness of the present (fuzzy set) methodology in dealing with the double complexity of *longitudinal* analysis of *multi-dimensional* measures. Latent and manifest deprivation measures are constructed for each time taking into account the degree of overlap between income and non-monetary aspects at the micro level. These measures are then studied longitudinally taking into account their degree of persistence over time, again at the micro level.

The most intense deprivation is reflected in the last column, M(4), of the table. The rates are under 0.5% in Denmark, Netherlands and Germany, and at the other end 1.5-2.5% in Italy, Greece and Portugal for the continuous experience of income poverty simultaneously with non-monetary deprivation. By contrast, the experience of one or the other form of deprivation at some time during the 8-year period, column L(1), varies between 40% in Denmark to 60% in Portugal.

Table 7.6 Fuzzy supplementary, latent and manifest measures of longitudinal deprivation rates

	Latent (L)				Fuzzy supplementary (FS)				Manifest (M)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
3 DK	40.2	15.5	9.7	2.1	27.9	9.8	5.0	0.8	8.4	2.2	0.5	0.1
4 NL	43.1	16.7	12.2	4.3	31.4	10.7	7.0	2.0	11.6	3.7	2.4	0.4
6 DE	43.4	17.7	11.5	4.1	32.0	10.0	4.3	1.1	10.2	2.9	1.1	0.2
8 BE	48.7	20.8	15.1	6.0	36.2	13.2	8.2	2.6	11.9	4.3	2.6	0.6
9 FR	45.3	21.7	17.0	7.0	32.8	14.4	10.6	3.5	14.3	5.7	3.8	1.1
10 UK	55.3	27.0	21.2	8.6	41.7	17.3	11.8	3.8	18.1	6.8	4.2	1.1
11 ES	59.3	29.0	23.4	7.8	43.4	17.4	11.6	2.1	21.1	7.7	4.9	0.8
12 IE	52.8	26.2	22.5	8.7	39.5	16.8	13.2	3.7	20.0	7.9	5.8	1.4
13 IT	56.3	29.1	23.3	9.6	42.2	18.4	13.0	4.1	20.3	8.5	5.7	1.5
14 GR	64.8	34.9	28.6	11.4	51.3	22.4	16.1	4.1	25.1	10.7	6.9	1.7
15 PT	61.0	34.8	32.6	15.3	46.1	22.6	20.0	7.5	25.2	10.3	8.1	2.4
simple average	51.8	24.9	19.7	7.7	38.6	15.7	11.0	3.2	16.9	6.4	4.2	1.0

Manifest propensity to both fuzzy monetary (FM) and non-monetary (FS) deprivation
Latent propensity to either form of deprivation (FM and/or FS)
Poverty/deprivation rates: (1) Anytime; (2) Cross-sectional; (3) Persistent; (4) Continuous

Note: The results are for a 'balanced panel', i.e., for the population present in all 8 ECHP waves.

Chapter 8
Poverty and deprivation: illustrative results

8.1 Introduction

In this chapter we report some illustrative results for complementing and integrating those already included in the previous Chapters 4 to 6. Sections cover four main topics: income poverty, subpopulation analysis, non-monetary deprivation, and monetary and non-monetary deprivation in combination.

8.2 Income poverty

In the first set of maps (Figures 8.1 to 8.3) we want to compare the percentages of poor individuals (Head Count Ratio) for poverty lines calculated at three different levels: Europe, countries and NUTS1 regions.

For most countries other than EU15 we did not have direct access to national survey data, and therefore it was not possible to calculate the HCR's using an European line or NUTS1 regional poverty lines. For this reason the three maps cover only EU15 countries. Figure 8.2 reports the same variable (HCR calculated using country poverty lines) as Figure 5.1 in Chapter 5: but here only the EU15 Countries are considered, therefore the colouring of countries in this map is different from 5.1. It is included here to facilitate comparison with the other figures presented.

It is interesting to compare Figures 8.1, 8.2 and 8.3. They could be seen as illustrating three different concepts:

a) Absolute poverty, Figure 8.1;
b) Relative poverty, or inequality among regions, Figure 8.2;
c) Inequality within regions, Figure 8.3.

A poverty line calculated at EU15 could be seen as a threshold that changes little when one single country has a big change from one year to the next one; for this reason for each country it could be seen approximately as a fixed poverty line. It is 'absolute' in the sense that the mean of the actual level of income in a country (compared to the EU15 mean) is a major determinant of the country's poverty rate.

Portugal, Spain and Greece, followed by Italy, Ireland and United Kingdom, are the countries showing the highest poverty rates; by contrast, Luxembourg, Denmark and Austria are subject to the lowest poverty rates determined on the basis of a common EU15 poverty line.

Numbers reported in Figure 8.1 also show very large differences among EU15 countries in the HCRs (range 1–52 %), in part reflecting variations in national income levels.

When adopting a relative (i.e., national) poverty line, calculated on the basis of the income distribution separately in each country, the three countries with the highest poverty rates continue to be Portugal, Greece and Italy. In fact Italy clearly shows a very high inequality among regions which is also reflected by the HCR based on the national poverty line.

The three better-off countries are also the three most equal, and they all belong to Scandinavia: Denmark, Sweden and Finland.

Calculating the poverty lines at NUTS1 level (Figure 8.3) we can explore more clearly the degree of inequality within regions. In this case the three countries with the highest poverty rates are Portugal, Greece and – not surprisingly – United Kingdom; in the UK the high inequality is highlighted by using the NUTS1 poverty line threshold.

Again, the three most equal countries all belong to Scandinavia: Denmark, Sweden and Finland.

Figure 8.1 Head Count Ratio - EU15 poverty line

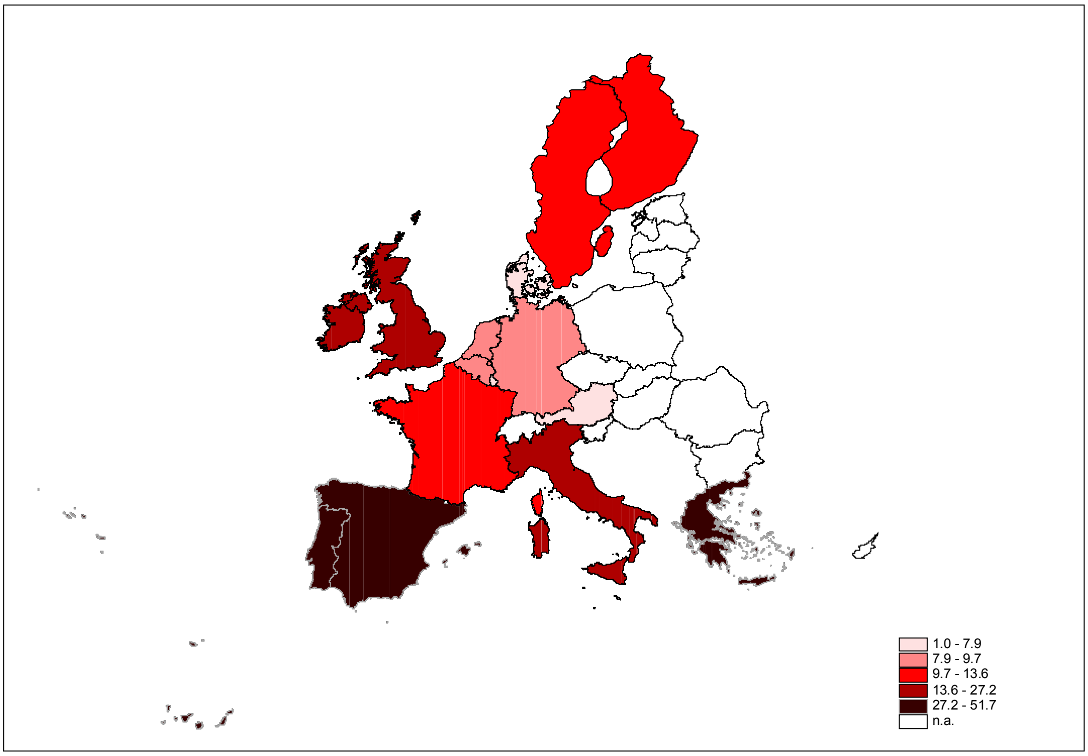


Figure 8.2 Head Count Ratio - country poverty lines. EU15

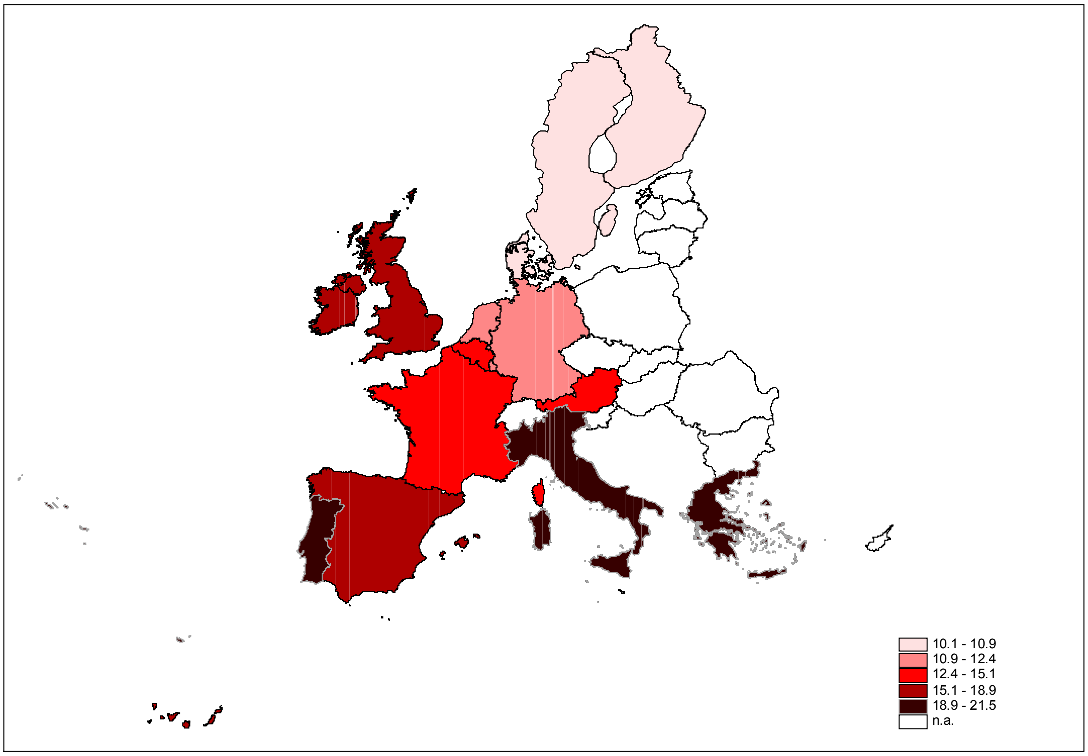


Figure 8.3 Head Count Ratio - NUTS1 poverty lines. EU15

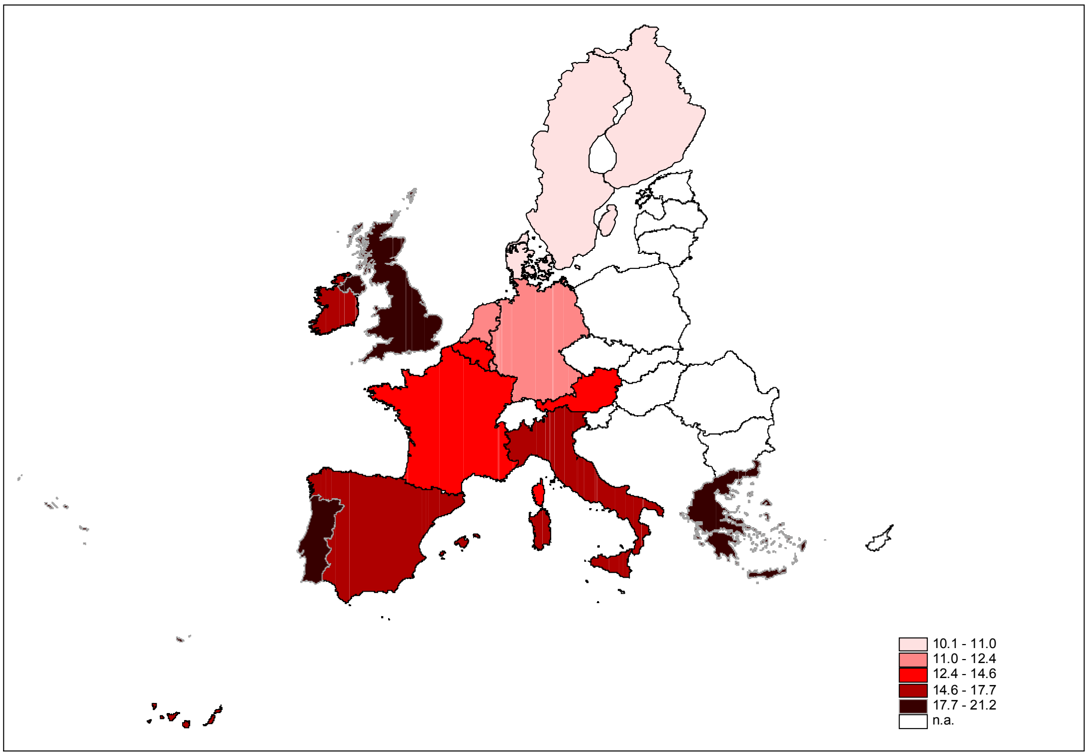


Figure 8.4 Dispersion of Head Count Ratios of NUTS1 regions within countries - country poverty lines

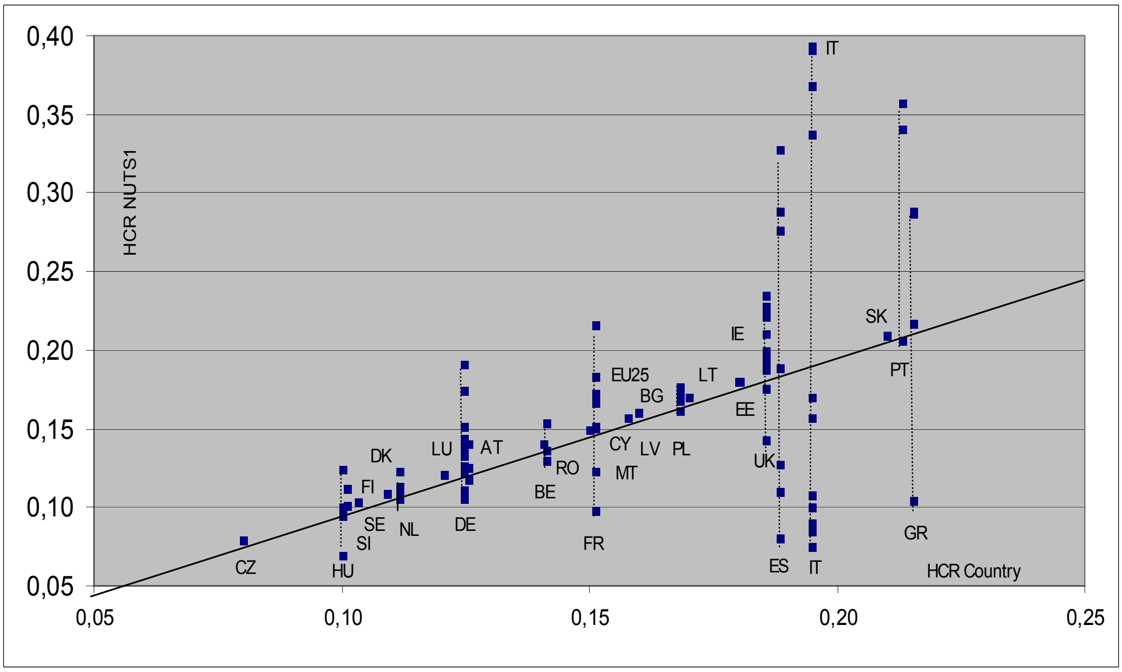


Figure 8.4 shows the relationships between countries and NUTS1 regions, according to the Head Count Ratio based on *national poverty lines*. First of all it is important to emphasise that most of NMS10 countries have poverty level lower than EU15 countries.[22] This fact certainly points to more equity in the income distribution.

It would have been interesting to have the possibility to calculate a EU25 base poverty line in order to better compare the countries: this should be one possible new indicator to be adopted in order to better compare the new countries.

Coming back to the comparisons among NUTS1 regions, it is clear how in Italy, Greece and Spain there is a large variation among NUTS1 regions. This is much less evident in the United Kingdom, confirming the ranking of countries showed by figure 8.2.

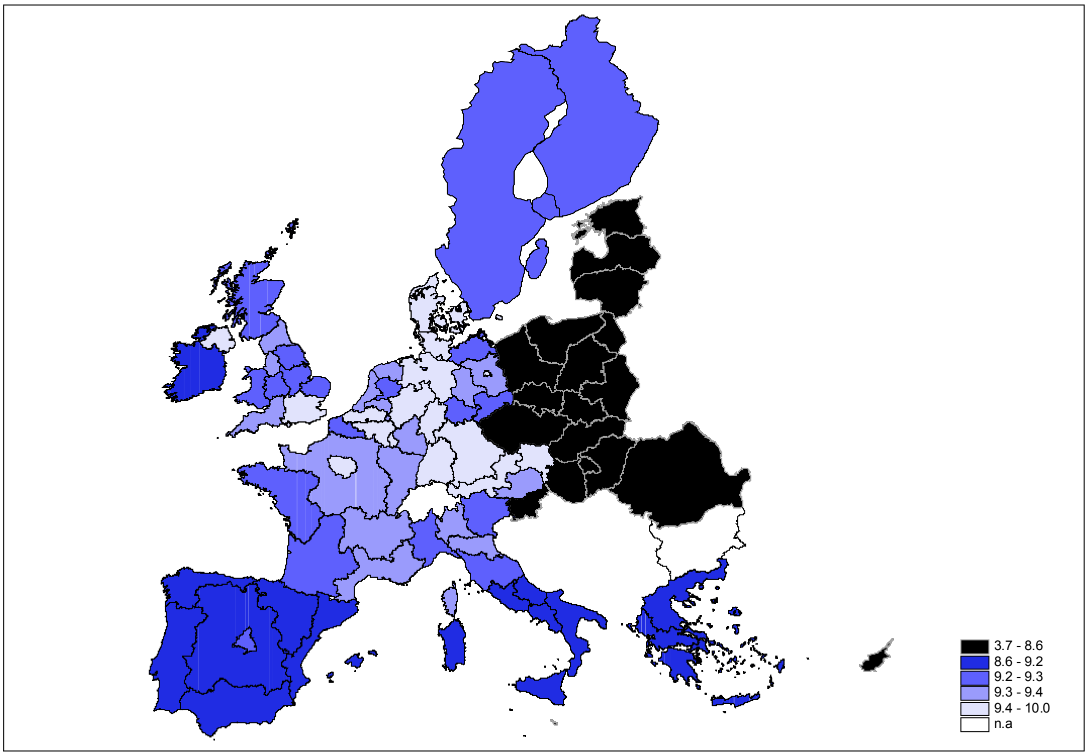
Figure 8.5 reports the variable net household equivalised income (in logarithm terms) estimated by SAE models at NUTS1 regions level (see Chapter 12 for details). It is clear how the new NMS10 countries have much lower income levels (in PPS) compared to former EU15 countries.

This would mean that if we calculated an EU25 poverty line we would expect very high values of the HCR in the new Member States, much higher than those obtained using national poverty lines.

---

[22] It should be mentioned that most of the results from NMS10 countries are based on published results from household budget surveys, and NUTS1 breakdown has been obtained on the basis of relatively simple regression-prediction methodology explained in Chapters 11 and 12.

Figure 8.5 Net equivalised income (log). NUTS1 regions



8.3 Subpopulation analysis

This analysis has been conducted on EU15 countries only. Here we have chosen country poverty lines, and we have reported NUTS1 figures for HCR among children and old persons.

Observing Figure 8.6 (Children), the very large differences among regions of Italy are particularly noteworthy. Regions in North and Central Italy are among the top 20% of EU15 NUTS1 regions with the lowest poverty rates; by contrast, regions in South Italy are among the bottom 20% regions with the highest poverty rates. Incidently, it may also be noted the former also have the lowest birth rates in Europe. One interesting question arises: do those few children tend to be concentrated in the richer households?

Most of the Southern Italian NUTS1 regions and Lazio are in the group with the highest poverty rates, along with two Spanish NUTS1 regions (Centro and Sur) and three British NUTS1 regions (Yorkshire and Humberside, East Midlands, West Midlands).

Analysing the situation among the elderly (Figure 8.7), all regions of Greece and Portugal, as well as Southern Italy, Sud-Ouest in France and West Midlands in UK are in the bottom group with the highest poverty rates among elderly persons.

Some Northern Italian regions, two regions in Spain (Noreste and Este), the Netherlands and Sweden are in the top group with the lowest poverty rates among elderly persons.

Figure 8.6 Head Count Ratio for children (country poverty lines). NUTS1 regions

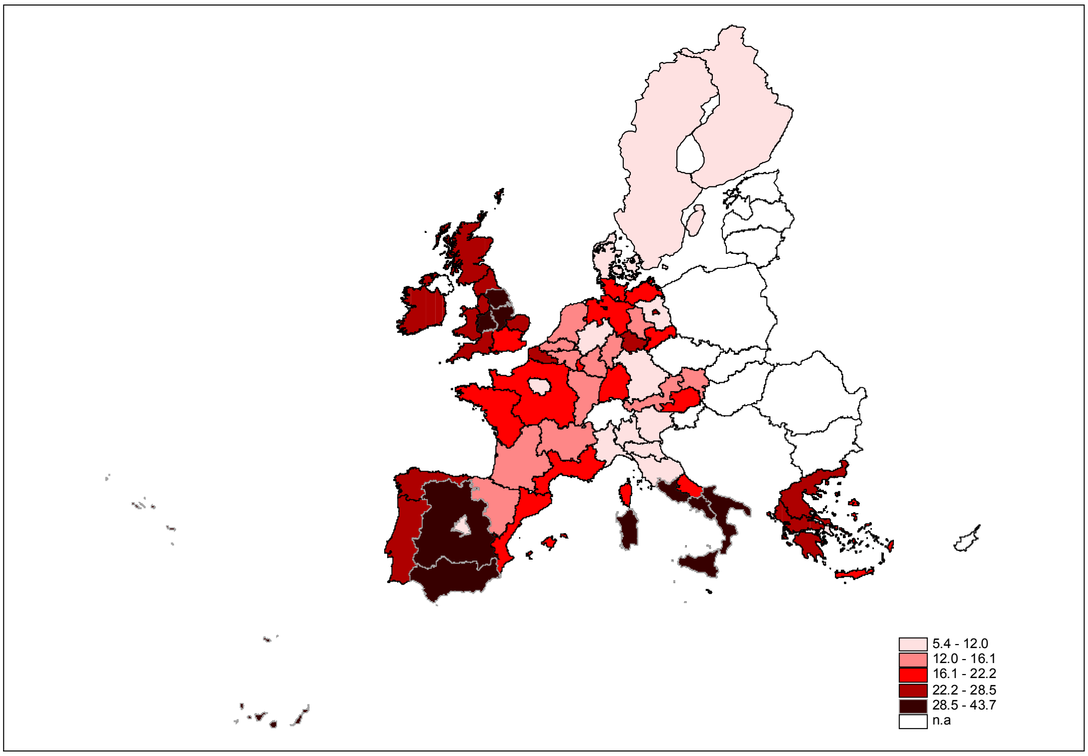


Figure 8.7 Head Count Ratio for elderly persons (country poverty line). NUTS1 regions

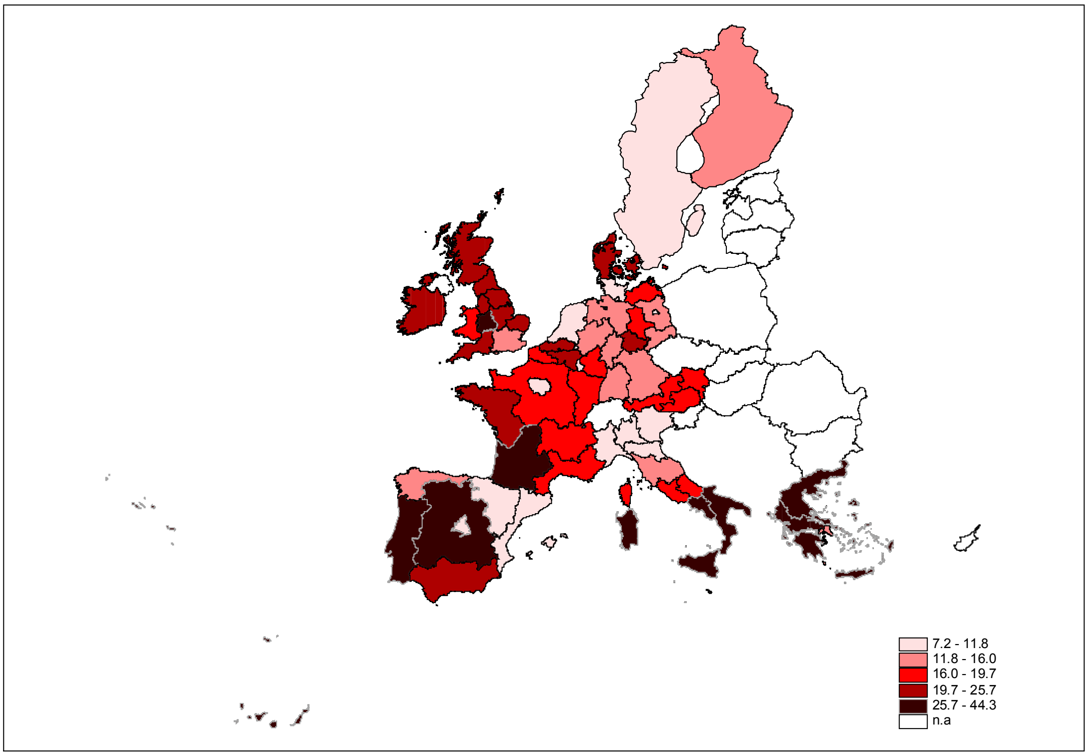
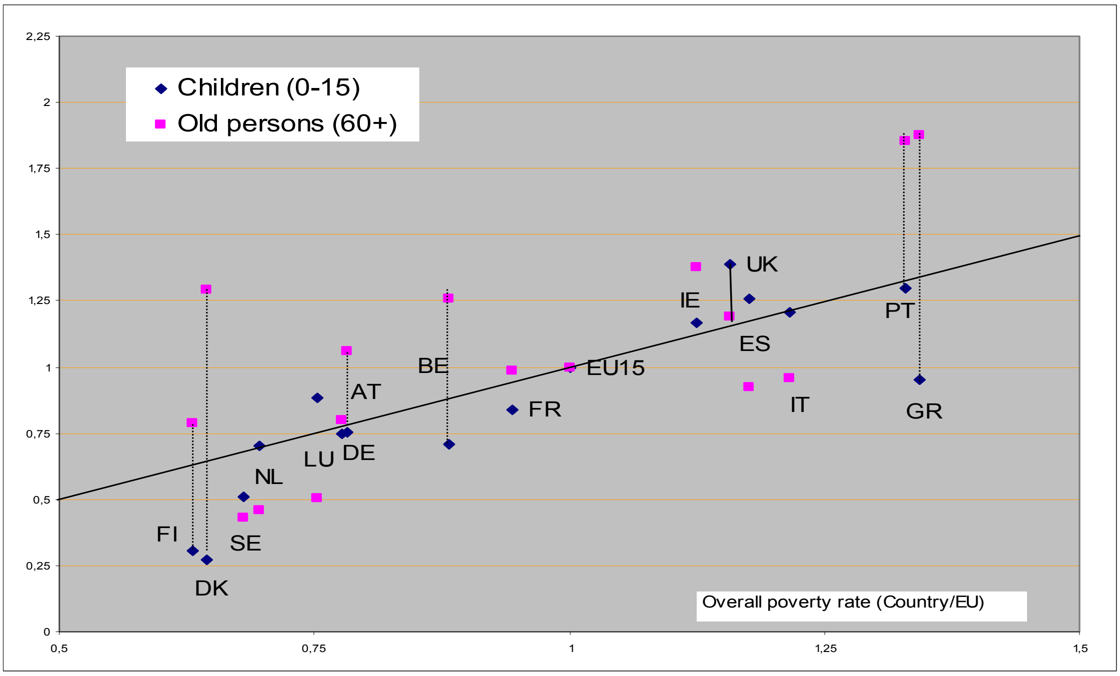


Figure 8.8 Ratio of (country / EU) poverty rates: comparison of children and elderly persons with the general population (country poverty lines). EU15



The figure shows along x-axis the ratio of country to EU15 poverty rate for the whole population. The y-axis shows the same ratio for children and for elderly persons. Hence, by definition, EU15 is represented by point (1,1). A point above the 45° line means that the situation of a subpopulation in the country, relative to the situation of the same subpopulation in EU15, is worse compared to the same comparison made for the total population.

Figure 8.8 shows ratios (country / EU) poverty rates for children and elderly persons in comparison with the same ratio for the general population. From the figure two main points should be pointed out:

1) In some countries (Italy, Sweden, the Netherlands), for both children and the elderly, the relative situation (compared to the corresponding figure at EU15 level) is better than the relative situation when the same comparison is made for the whole population.

2) In some countries (Finland, Denmark, Belgium, Portugal and Greece) the relative situation (in the above sense, compared to the overall EU15 situation) of children is much worse than the relative situation of the elderly.

Figures 8.9 to 8.13 report the relative (ratio NUTS1 /country) position of children and elderly persons in the five largest EU15 countries: Germany (Figure 8.9), France (Figure 8.10), United Kingdom (Figure 8.11), Italy (Figure 8.12), and Spain (Figure 8.13).

Among these countries we can highlight seven NUTS1 regions where the relative position of children is *much worse* than the relative position of the elderly: Hamburg, Mecklenburg-Vorpommern, Thüringen in Germany, Nord-Pas-de-Calais in France, Sicilia and Sardegna in Italy. By contrast, in East Anglia in the United Kingdom (UK4), the relative position of children is better than the relative position of elderly persons.

Figure 8.9 Ratio of (NUTS1 / country) poverty rates: comparison of children and elderly with the general population. Germany

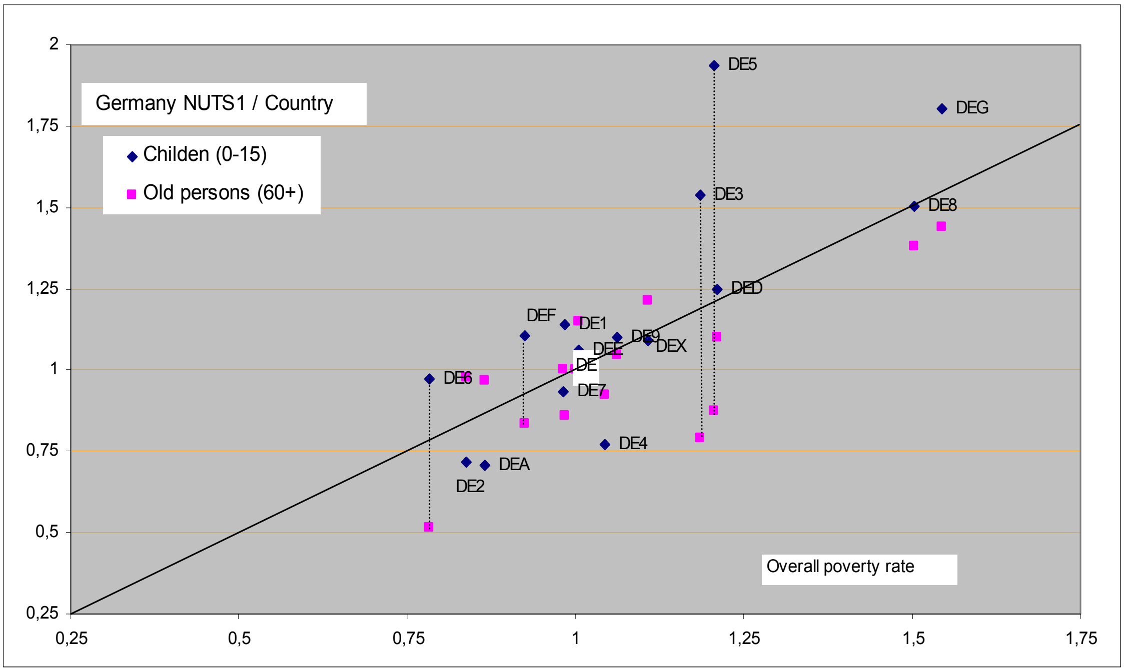
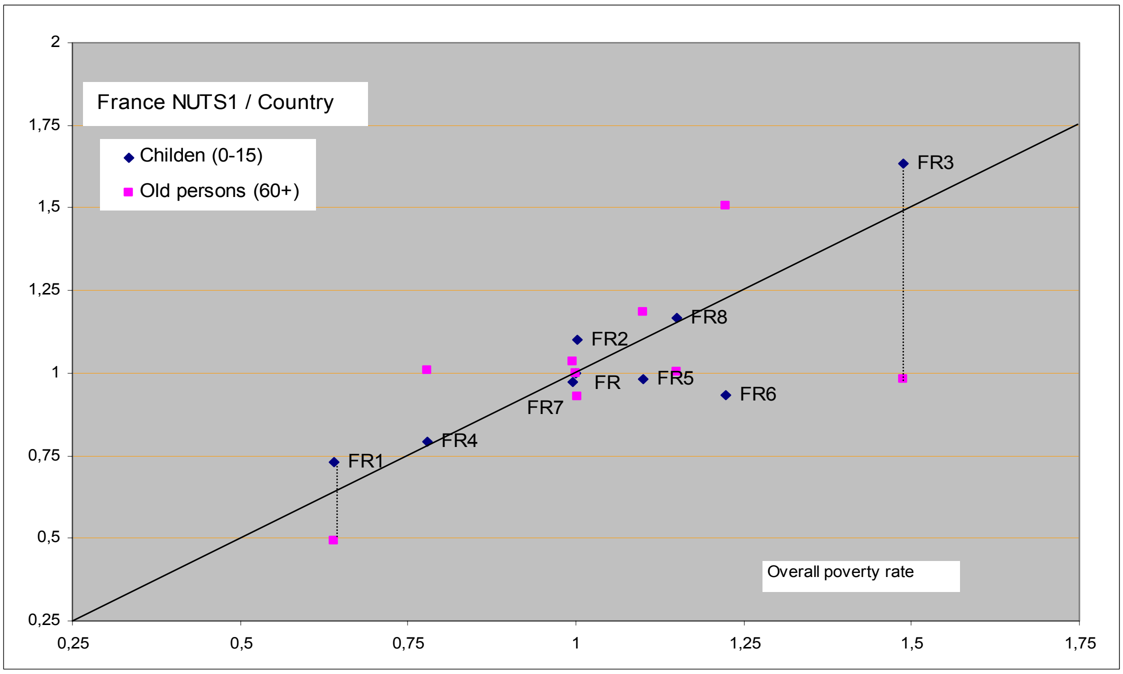


Figure 8.10 Ratio of (NUTS1 / country) poverty rates: comparison of children and elderly with the general population. France



Notes to Figures 8.9-8.13. The figures along x-axis show the ratio of NUTS1 region to country poverty rate for the whole population. The y-axis shows the same ratio for children and for elderly persons. By definition, the whole country is represented by point (1,1). A point above the 45° line means that the situation of a subpopulation in the region, relative to the situation of the same subpopulation in the whole country, is worse compared to the same for the total population.

Figure 8.11 Ratio of (NUTS1 / country) poverty rates: comparison of children and elderly with the general population. United Kingdom

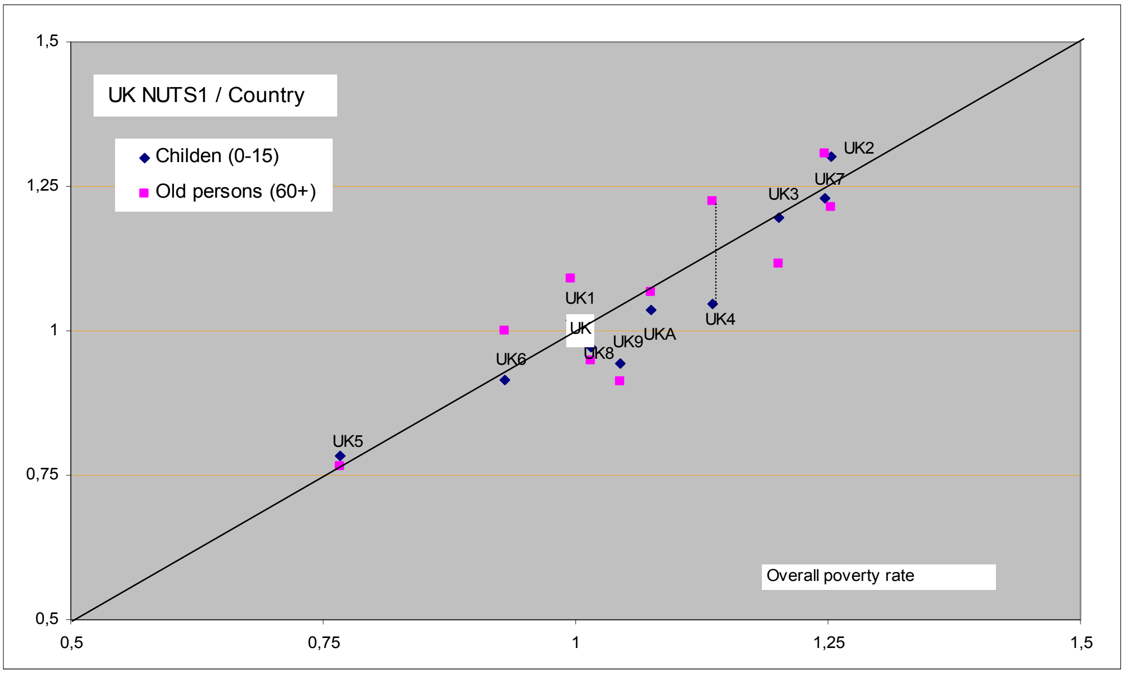


Figure 8.12 Ratio of (NUTS1 / country) poverty rates: comparison of children and elderly with the general population. Italy

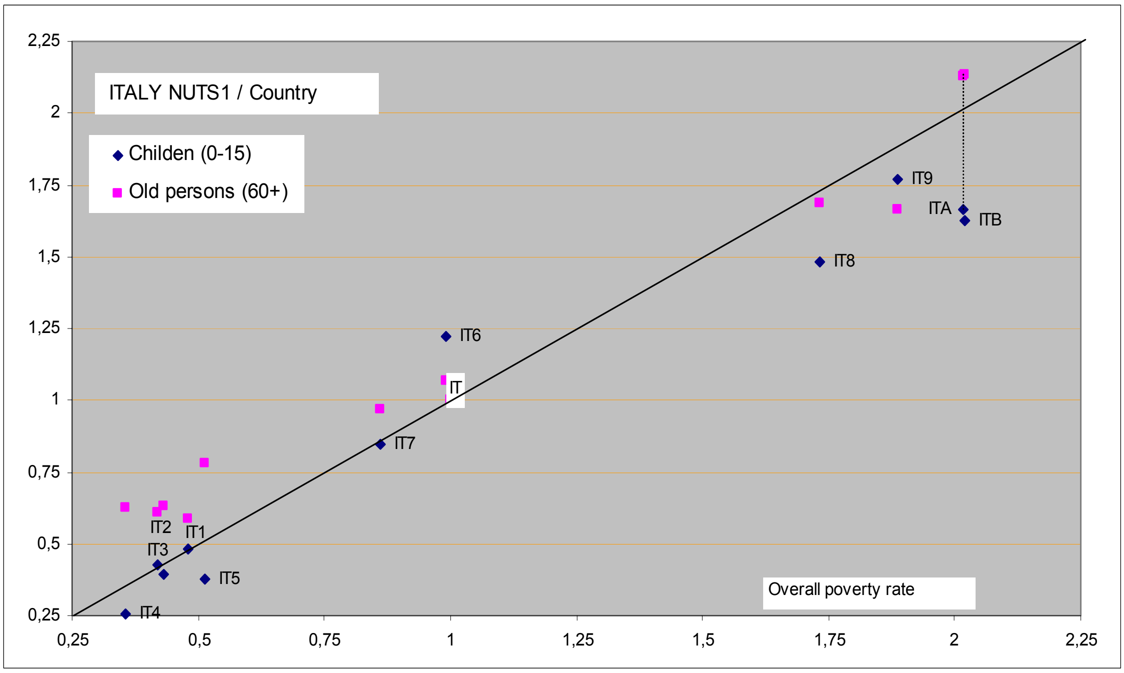
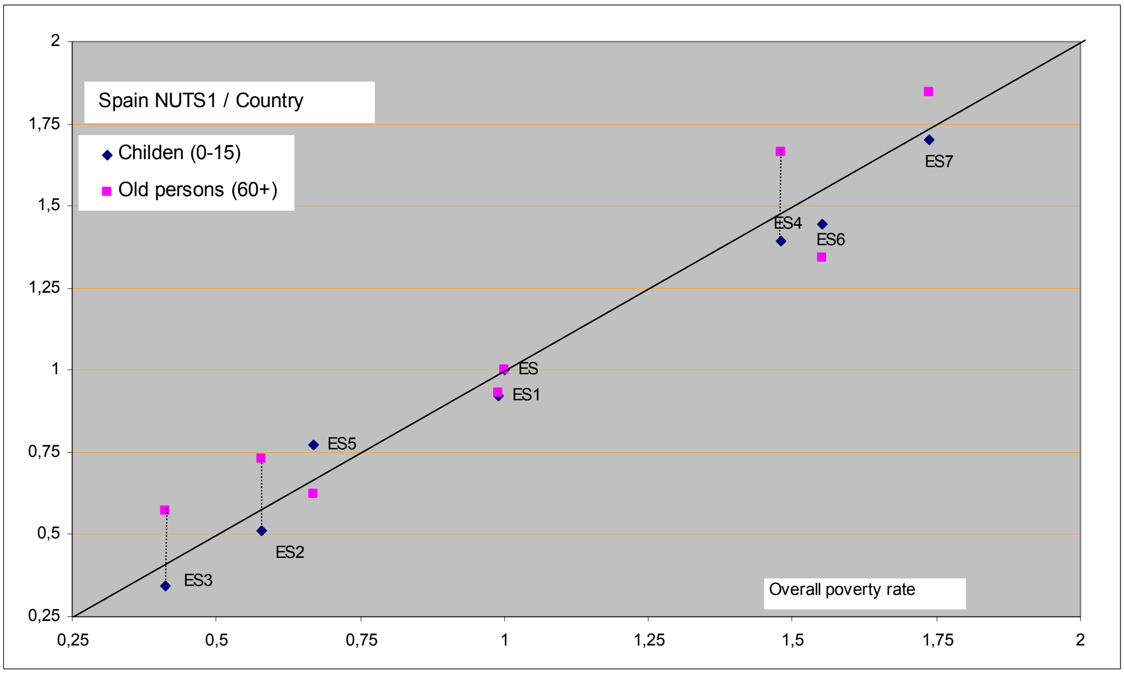


Figure 8.13 Ratio of (NUTS1 / country) poverty rates: comparison of children and elderly with the general population. Spain



8.4 Fuzzy supplementary deprivation rates

The analysis conducted in this section is based on the EU15 countries and Poland;. Although the original indicators in the Polish survey are of a different nature from the ECHP- EU15 countries, it is interesting to try to make comparisons between them. For some EU15 countries it was not possible to calculate the supplementary deprivation indices disaggregated into the five dimensions described in Chapter 6; for Poland only dimension 3 was not included in the analysis.

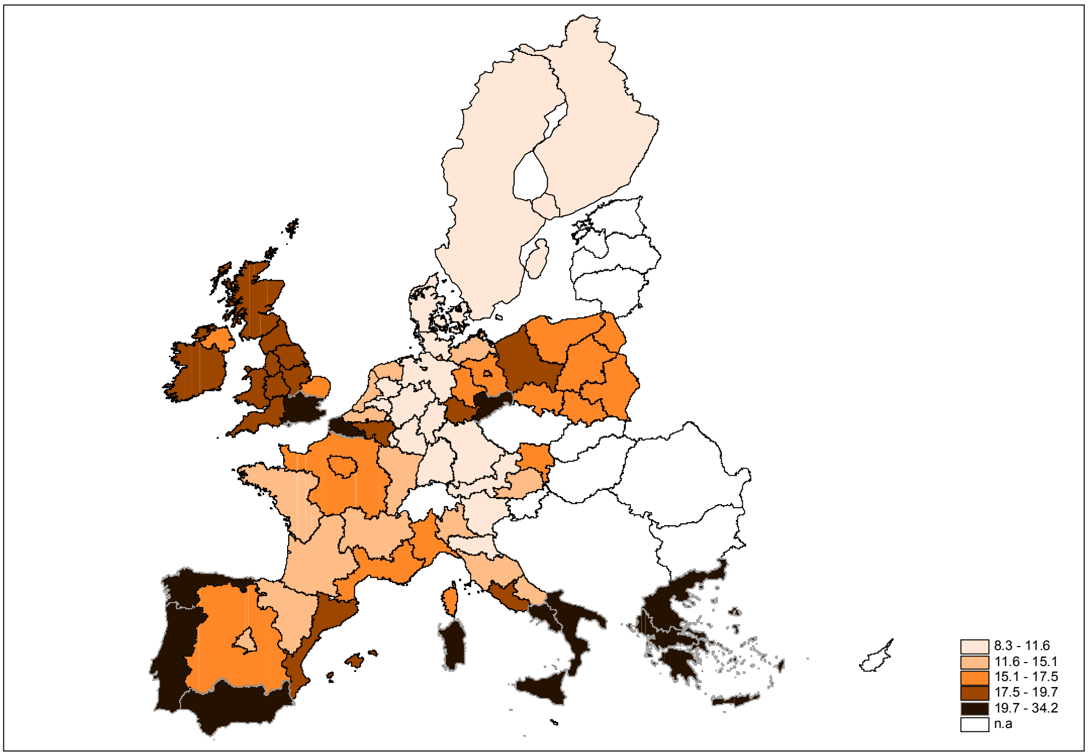
Figure 8.14 shows the Fuzzy Supplementary overall deprivation rates disaggregated at NUTS1 regions; the worst situation appears to be in the following regions: the whole Greece and Portugal, the entire Southern Italy, Noroeste and Sur in Spain, the Région Bruxelles-capitale in Belgium, East Midlands in UK, and Sachsen in Germany.

On the other hand, most of former Western Germany, the whole Scandinavia, Luxembourg, Oost-Nederland and Zuid-Nederland in the Netherlands, Triveneto (Nord Est) and Emilia-Romagna in Italy are the regions with better conditions.

The situation of most deprived regions is much more marked when observing the Basic non-monetary deprivation index (Dimension 1): here the whole of Greece, Southern Italy and Southern Spain show the highest deprivation indices.

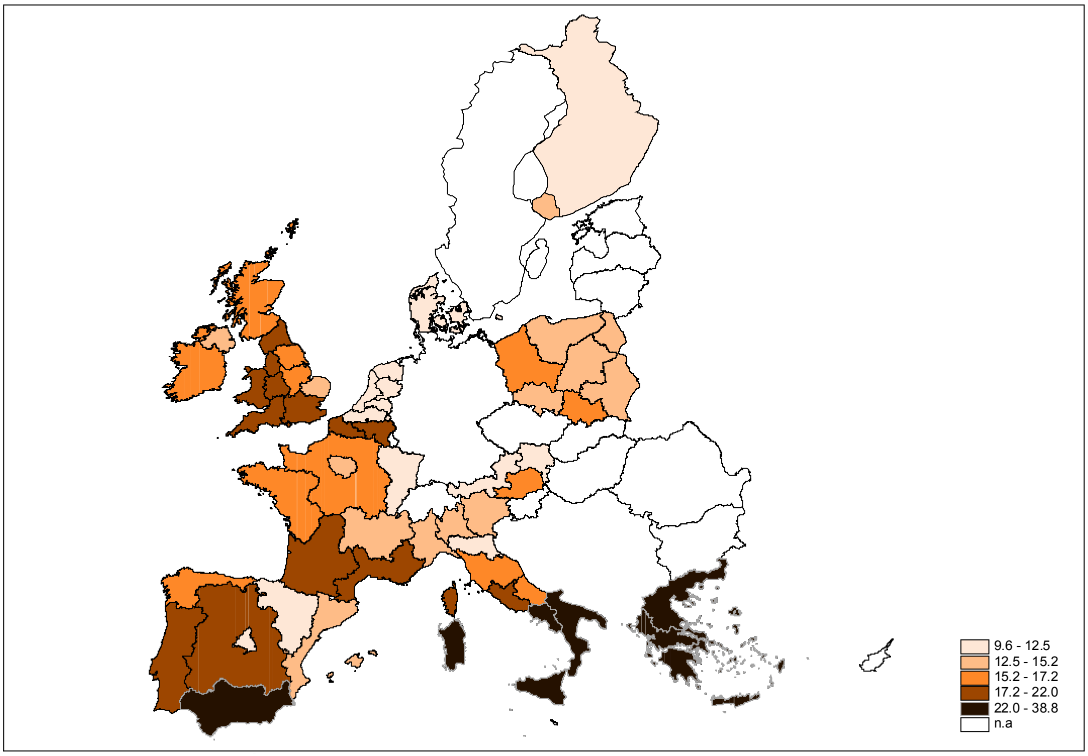
It is important to note that from Figures 8.15 up to Figure 8.19 data for Germany, Sweden and Luxemburg are not reported; this has to be remembered when we identify the better off NUTS1 in Europe. (In fact Germany, Sweden and Luxemburg would have been good candidates for that!).

Figure 8.14 Fuzzy Supplementary: overall deprivations rates. NUTS1 regions



Considering the overall non-monetary deprivation rates (Figure 8.14), the NUTS1 regions showing the greatest levels of inequality do in fact also have the greatest levels of inequality in monetary income (as shown in the statistical tables in Chapter 15). The whole of Portugal and Greece, southern Spain and southern Italy have the highest deprivation rates (with the maximum of 34% in Canarias). The NUTS1 with lowest rates are all in Germany.

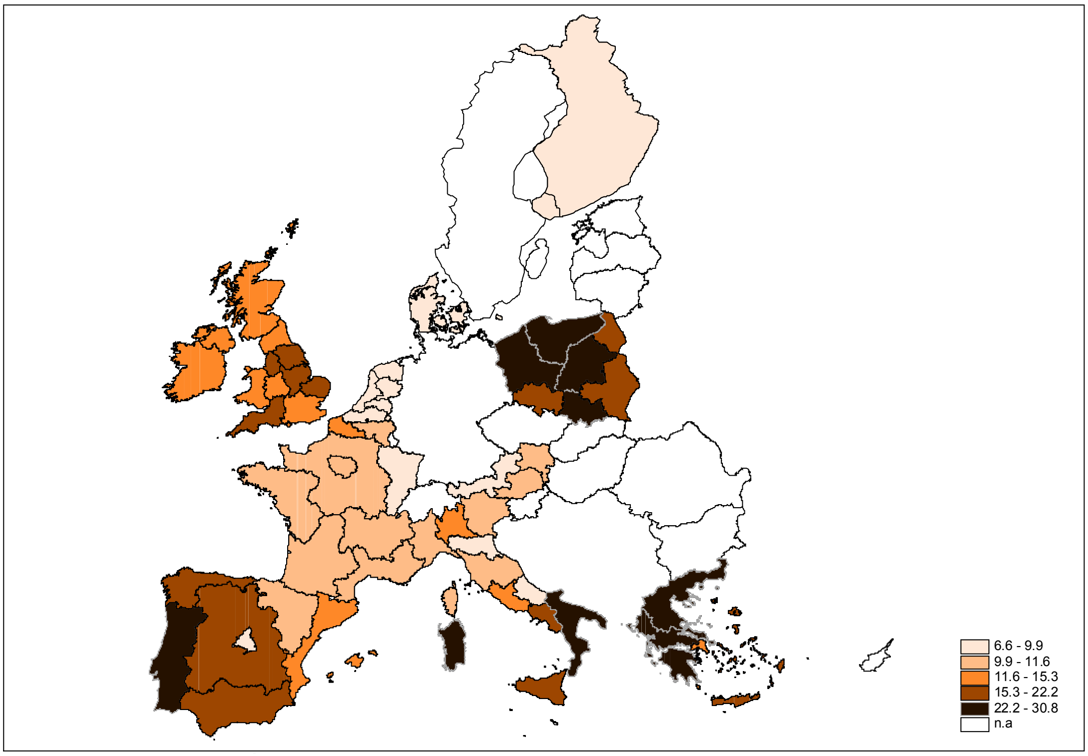
Figure 8.15 Fuzzy Supplementary, Dimension 1 (general life style deprivation). NUTS1 regions



According to Basic non-monetary deprivation index (Dimension 1) the situation shown in Figure 8.15 is quite interesting. The least deprived NUTS1 regions (always considering that we miss 3 countries) are Comunidad de Madrid (in Spain), Zuid Nederland and Noord Nederland. The most deprived areas are in South Italy (Sud and Sardegna) and in Greece (Kentriki Ellada), with a maximum index of 38.8 in Sud.

The situation according to the Secondary non-monetary deprivation index (Dimension 2), reported in Figure 8.16, shows deprived areas present in Poland as well as Greece, southern Italy and the whole Portugal. Again, even if it is not very meaningful to rank the "best" region because of missing information for Germany, Sweden and Luxemburg, we find that the Netherlands (Zuid Nederland) and the entire Finland have a very low index of deprivation.

Figure 8.16 Fuzzy Supplementary, Dimension 2 (secondary life style deprivation rates). NUTS1 regions



As mentioned above, for Poland was not possible to calculate Fuzzy Supplementary deprivation indices for Dimension 3 (lacking housing facilities). For this reason and because of lack of comparability with the other dimensions here we have not reported the corresponding map for Dimension 3. Anyway, it is interesting to highlight how unexpectedly the following four NUTS1 appear quite deprived: Nord-Pas-de-Calais and Su-Ouest in France, the Région Bruxelles-capitale in Belgium, and Ostösterreich in Austria.

The picture of Europe is quite different when observing the Dimension 4 in Figure 8.17 (housing deterioration); the most deprived regions are well concentrated in Poland, Greece, Portugal and Noroeste in Spain; on the other hand most of Central and Northern Italy, the whole Austria and Finland perform very well for this dimension.

Finally the fifth considered Dimension (Environmental problems) clearly highlights the bad situation of most European large cities: the NUTS1 containing the cities of Paris, Rome, Madrid, Lisbon, Brussels, Barcelona, Naples seem to show high environmental problems.

This fact is more evident when disaggregating the map at NUTS2 level (Figure 8.19): here the bad situation of London, Warsaw, Milan, Turin, Oporto, Birmingham and other large cities is also evident.

Figure 8.17 Fuzzy Supplementary, Dimension 4 (housing deterioration rates). NUTS1 regions

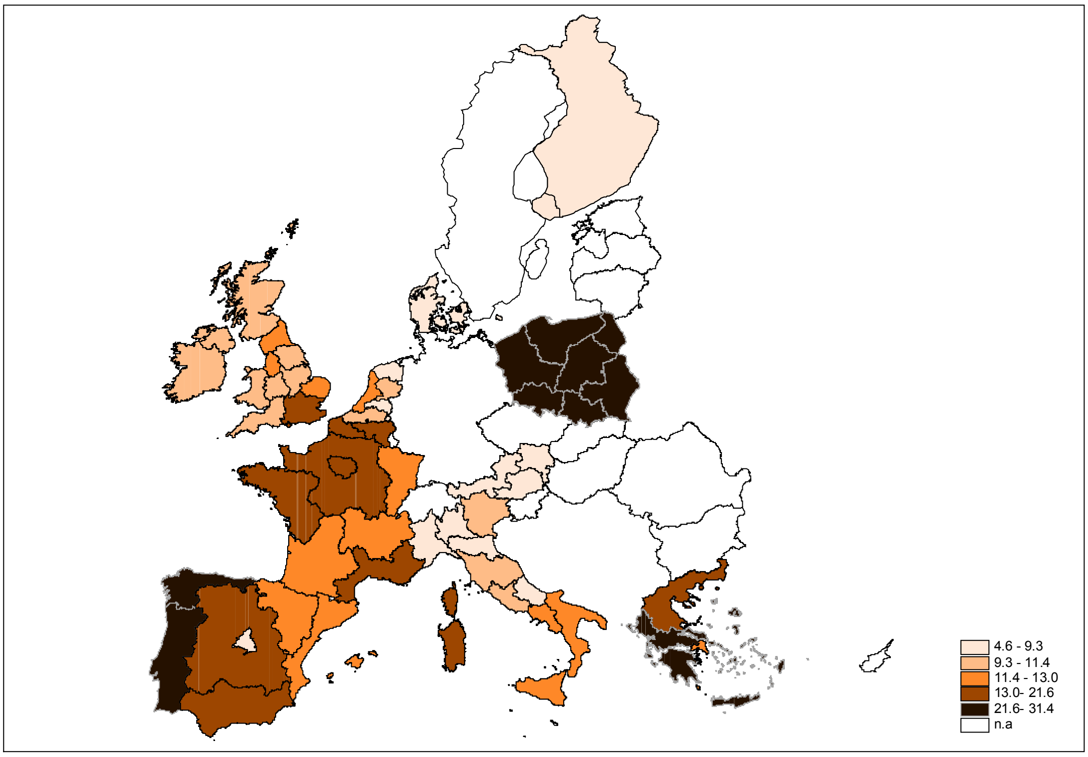


Figure 8.18 Fuzzy Supplementary, Dimension 5 (Environmental Problems). NUTS1 regions

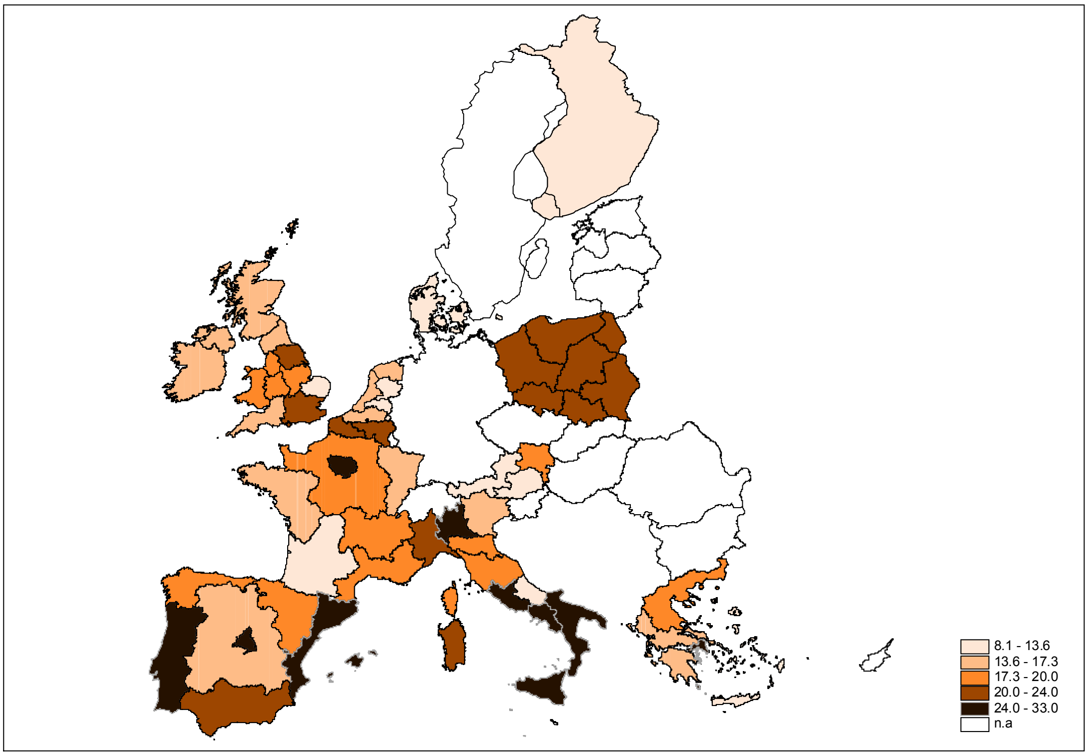
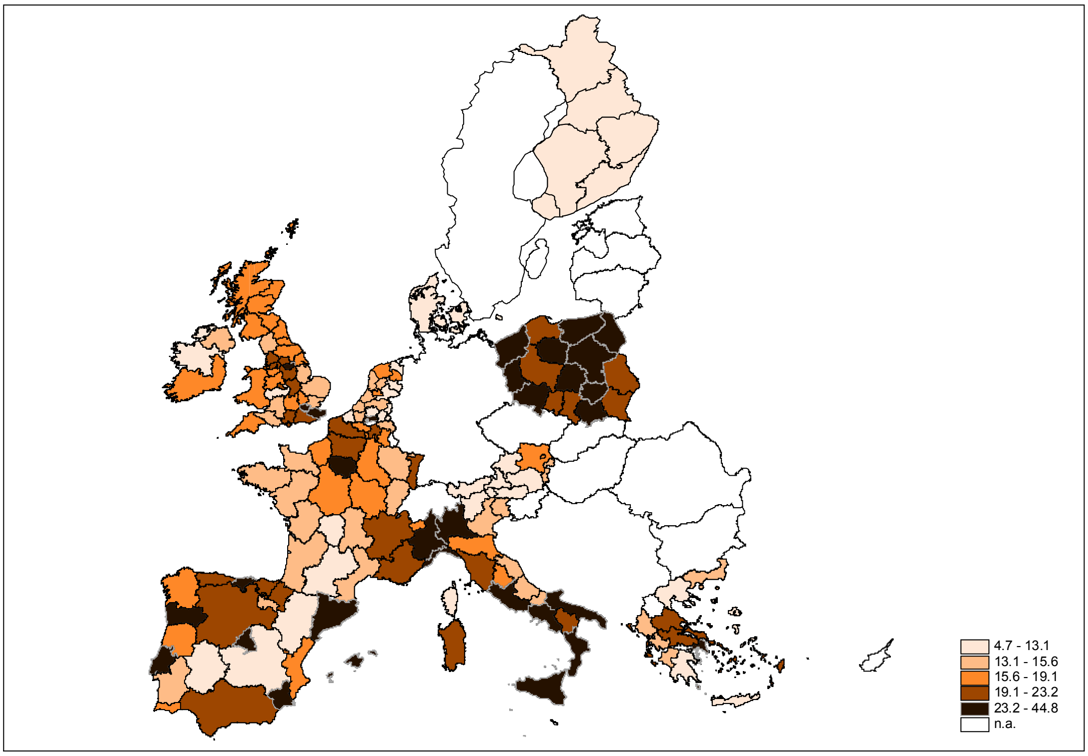


Figure 8.19 Fuzzy Supplementary, dimension 5 (Environmental Problems). NUTS2 regions



8.5 Latent and manifest deprivation indices

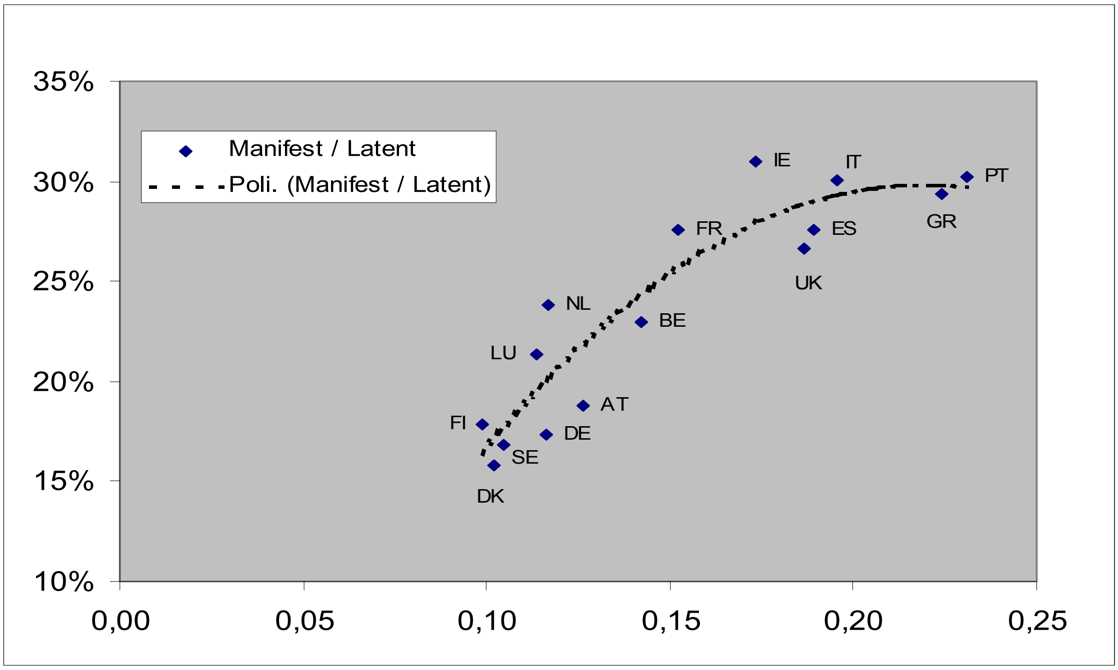
This section reports empirical results on income poverty and non-monetary deprivation in combination: the Latent and Manifest deprivation rates introduced in Chapter 6.

Figure 8.20 shows the manifest deprivation index as percentage of the latent: it can be interpreted as an index of the degree of overlap (or intersection), at the level of individual persons, between income poverty and non-monetary deprivation.

In theory, this ratio varies from 0 to 1. When there is no overlap (i.e., when the subpopulation subject to income poverty is entirely different from the subpopulation subject to non-monetary deprivation), Manifest deprivation rate and hence the above mentioned ratio equals 0. When there is complete overlap (i.e., when exactly the same subpopulation subject to both to income poverty and to non-monetary deprivation), the Manifest and Latent deprivation rates are the same and hence the above mentioned ratio equals 1.

It is important to highlight that there is a *higher* degree of overlap between income poverty and non-monetary deprivation at the level of individual persons in *poorer* countries, and a lower degree of overlap in richer countries. This leads to the conclusion that the adoption of a multi-dimensional approach is particularly important when analysing richer countries where different dimensions have less overlap. Therefore in this cases the adoption of a supplementary indicator as a complement to the monetary one is justified, because it has an added value! On the other hand, because of the higher degree of overlap in poorer (and less equal) countries, the overall deprivation is more intense for the subpopulations involved, which is also important. All this underlines the need to supplement Laeken monetary indicators by multi-dimensional measures.

Figure 8.20 Manifest deprivation rate as a percentage of Latent deprivation rate, against a measure of the level of poverty or deprivation in the country. EU15

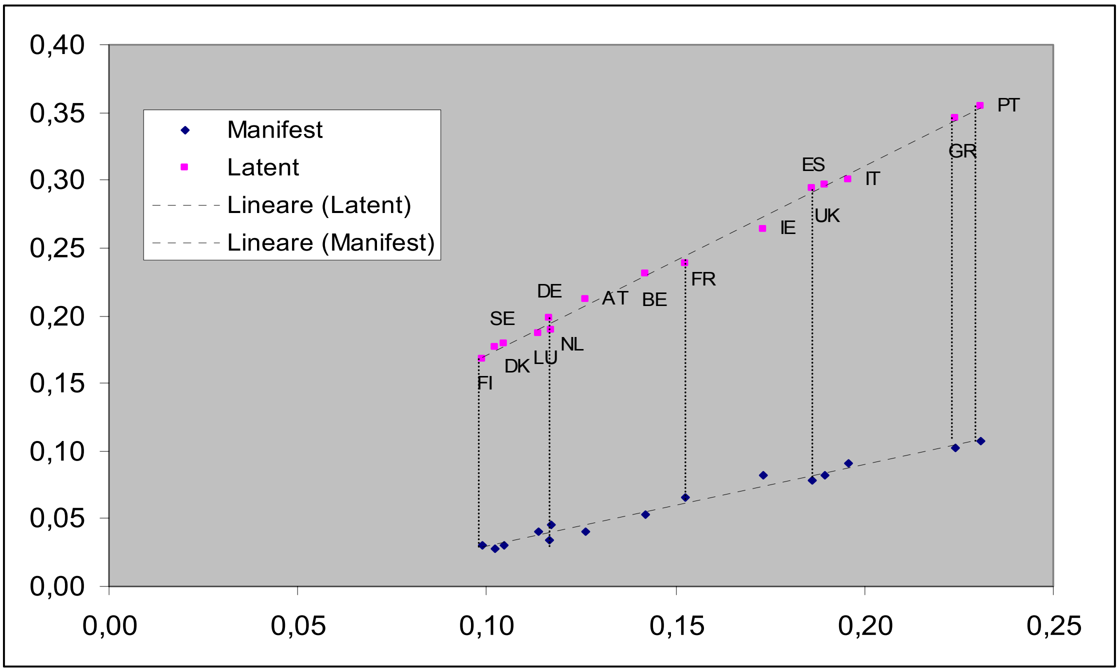


In this graph, y-axis is the ratio of Manifest to Latent deprivation rate; x-axis is the mean of these two rates, which in principle approximately equals the income poverty rate, and also equals the non-monetary deprivation rate which is scaled to equal the income poverty rate.[23] The ratio can be interpreted as an index of the degree of overlap, at the level of individual persons, between income poverty and non-monetary deprivation. Please see text for a description of the measure.

[23] Actually, this theoretical equivalence is modified in our results as a consequence of small area estimation procedures (see Chapter 11) applied independently to the different variables involved.

This trend, concerning the relationship of the income level with the degree of overlap between income poverty and non-monetary deprivation at the level of individual persons, is also seen in Figure 8.21 that reports values of Manifest and Latent rates (underlined by the regression line) at the country level.

Figure 8.21 Manifest and latent deprivation rates- variation by national poverty rate



In this graph, y-axis shows Manifest and Latent deprivation rates; x-axis is the mean of these two rates, which in principle approximately equals the income poverty rate or the non-monetary deprivation rate.

This trend is also present when the results are disaggregated to NUTS1 level (Figure 8.22, Italy highlighted) and NUTS2 level (Figure 8.23, United Kingdom only).

Finally Figures 8.24 and 8.25 report maps of Latent and Manifest measures disaggregated to NUTS1 regions level. The picture is remarkably similar – though not identical - for the two measures.

Figure 8.22 Manifest and latent deprivation rates- variation by NUTS1. All countries, Italy highlighted

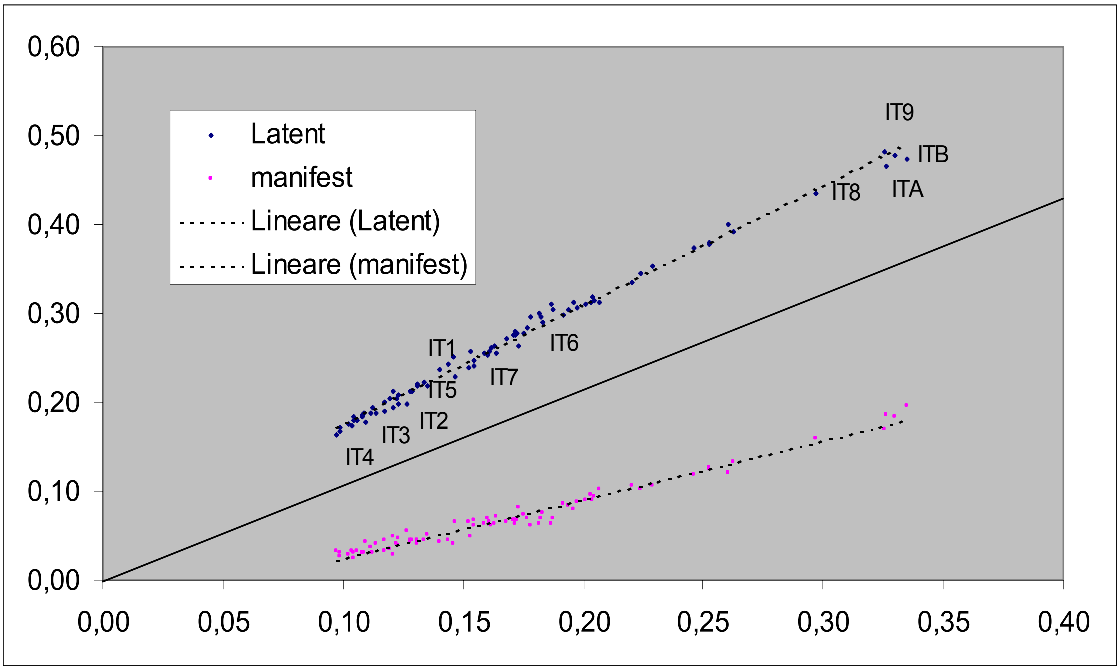


Figure 8.23 Manifest and latent deprivation rates- variation by NUTS2. UK only

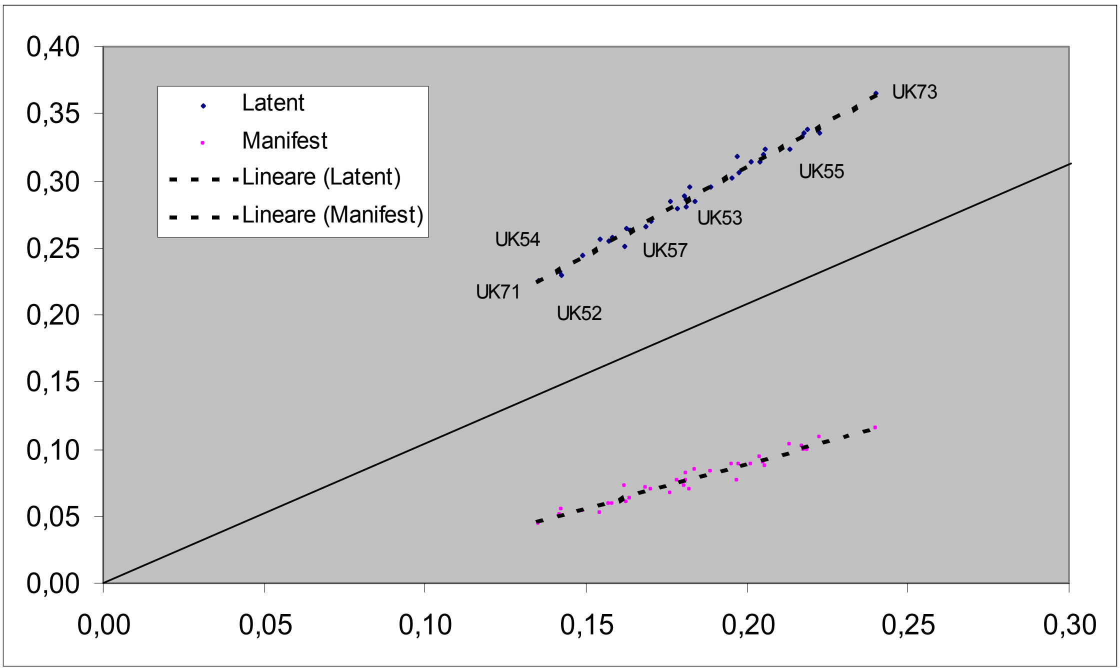


Figure 8.24 Latent deprivation rates. NUTS1 regions

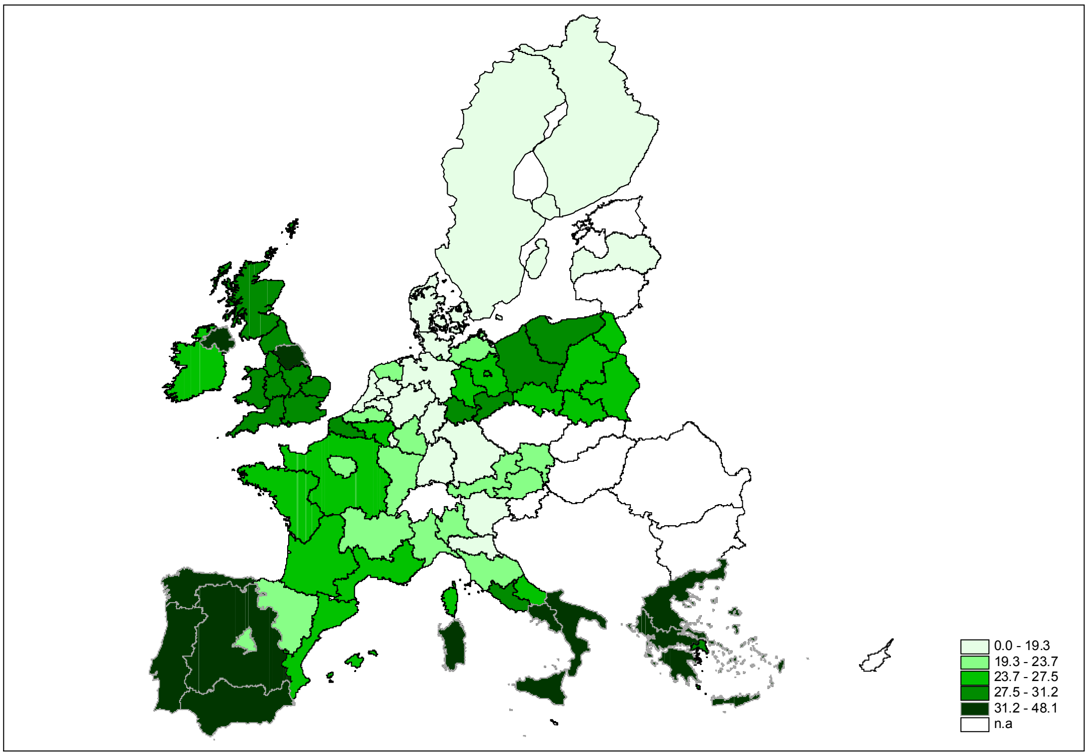
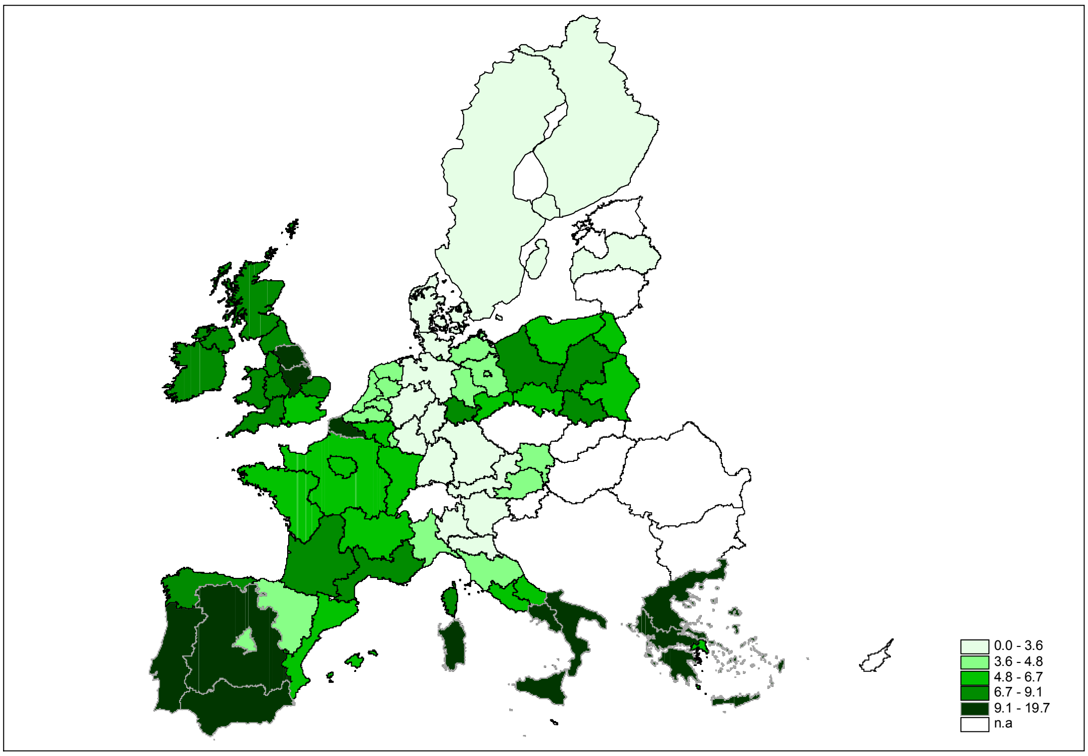


Figure 8.25 Manifest deprivation rates. NUTS1 regions



Chapter 9
Employment and education

Education and labour and their relationship are very important indicators for describing living conditions at national and regional level. They are specially important for young persons (or persons who have recently finished education); also the transition from education to the labour market (in the literature *school-to-work transition*) is itself an important indicator describing the situation of such persons. Differentials in this situation by age, gender, education, income, etc.. are important in order to identify the segments of the population that are in a disadvantage position.

The emphasis in the present project is on regional rather than subpopulation breakdown of the indicators on poverty and social exclusion. Nevertheless, it is necessary to keep in view the need to provide measures on the situation of particularly important subpopulations. At least three such populations of special interest can be identified: children, elderly persons, and young persons during the period of "school-to-work" transition.

Statistical analysis of the last mentioned subpopulation is more complex, and also more demanding on the data. This chapter describes certain aspects of school-to-work transition by analysing the employment situation of individuals as a function of the time elapsed since the completion of education or training.

There have been a number of studies analysing school-to-work transition at the EU15 level. A series of *Statistics in Focus* published by Eurostat, for instance, summarise the main results of some research, covering issues such as general indicators on school-to-work transition, association between social origin and educational attainment, and labour market effects of job mismatches (Eurostat, 2003 a, b and c). The basic approach in these studies has been to construct various indicators based on retrospective information on the time of first leaving continuous education, and current information on status and characteristics of the person's economic activity – expressing the status of activity as a function of the time elapsed since leaving continuous education. In this approach, essentially cross-sectional (though in part retrospective) information is interpreted as if it pertains to real cohorts.

Much of this comparative analysis of school-to-work transitions in EU15 countries has been based on the EU15 Labour Force Survey, the 2000 round of which incorporated a special module to collect information on the subject. By contrast, the analysis presented here is based on the European Community Household Panel, since for the present project we do not have access to the Labour Force Survey micro data. The ECHP has the advantage of providing more detailed and pertinent indicators. However, by using the ECHP data by themselves, severe sample size problems often arise; for this reason we present results for subpopulations only at national level.

The methodology proposed here shows some new (and hopefully interesting) ways of analysing and presenting the information. Our basic approach is to use the longitudinal data to identify, at the time of each ECHP wave, the person's most recent exit from education or training, and study this in relation to the person's current employment situation and other characteristics as a function of the time elapsed since that exit. Hence, in form at least, our approach is similar to that of earlier studies based on the LFS referred to above, though there are considerable differences in substantive content and statistical methodology resulting from the use of data of different types.

Section 9.1 describes the data source and the methodology for constructing the indicators of interest. Sections 9.2 and 9.3 describe, respectively, the education and employment situation. In Section 9.4 we analyse the relationship between educational level and employment situation, with the aim to answer questions such as the following. How do those who have recently undergone a school-to-work transition differ from the general population in the level of education or training received and in their employment situation? The employment situation of those with low levels of education/training can be expected to be worse than that of the better educated and trained. But how big are these differences? Do they relate to national differences in the overall employment situation?

Section 9.5 treats the important relationship between school-to-work transition and income level: this is extremely interesting when analysing poverty and living conditions at national or regional level. The main question to which we seek an answer is the following. After having completed a course of education or training, does the employment situation depend - even for a given educational level achieved – on income level of the individual's household or family? How do any differences between poor and rich vary by the level of education or training achieved?

Section 9.6 treats the employment situation after the completion of the education/training period in a longitudinal perspective – the employment situation seen as a function of the time elapsed since the exit from education/training and of the level of education and training completed and other characteristics of the individual.

9.1 Statistical Methodology

As noted, hitherto most empirical analyses of school-to-work transitions at the EU level have relied on the Labour Force Survey, specifically the special module on the subject incorporated into a round of the LFS. Primarily because of reasons to do with data availability, this analysis uses by contrast the panel data from the ECHP data. Because of differences in the nature and scope of the LFS and ECHP datasets, both the actual measures and the statistical methodology in the proposed analysis differ somewhat from those of previous studies, in particular the series of *Statistics in Focus* published by Eurostat.

It is instructive to clarify the similarity and differences in analytical structures of the more conventional LFS-based "cross-sectional" approach, and the ECHP-based "panel" approach used here. In essence the first approach is as follows. Information obtained for each individual in the survey pertains to two points in time: (i) current information on the person's employment situation (referring essentially to the same point in time for all individuals in the survey); and (ii) retrospective information on completion of continuous education (referring to different points in time for different individuals). Considering "period effects" over the short time period of interest to be secondary, the information can be relocated in time and viewed as if (ii) refers to the same moment in time for all units, thus defining a single cohort in terms of completion of continuous education. Data (i) are then seen as giving the employment situation of members of this cohort at various points in time following the completion of full-time education. Each individual in the sample provides one 'event' or 'unit' for analysis.

In the ECHP-based panel approach used here, the essential difference from the above is that the data from a number of waves of the panel are pooled, so that each individual provides as many events or units of analysis as the number of waves of the panel for which the individual appears in the survey. These events can be defined only through longitudinal linkage of the micro data. This approach permits pooling of the data over time (waves), which is highly desirable because of the smallness of sample sizes for individual years. Different events pertaining to the same individual are of course correlated, but each also provides a measure of independent information, thus increasing the effective sample size available for analysis. *This increase is considerable when the data are cumulated over as many as 8 survey waves.*

As before, each analysis unit is defined by information at two point in time: (i) information on the employment situation, current to each wave; and (ii) information on the most recently completed education or training as observed at each wave. Note that in our ECHP-based analysis variable (ii) refers to the date of the most recently completed education or training, rather than the date of completion of full-time education as in the EU-LFS analysis published by Eurostat. This is dictated by the nature of the data available in ECHP. We may again view this information as if pertaining to a real cohort defined by (ii), namely the completion of education/training, with (i) giving the subsequent employment situation of the cohort at various points in time. Information (i) is unique to each event (defined by the point of observation), but for a given individual in the sample, (ii) will be different for two events only if the individual has completed some relevant education/training course in the interval between the two points of observation, as recorded in the sequence of panel interviews. In summary, we use time since most recent exit from education/training as the reference variable, and study characteristics of the employment situation as a functions of the interval since that time. No other time or longitudinal dimension need appear, thus greatly simplifying data treatment and analysis. This information is treated as if it reflected the experience of real cohorts defined solely by the length of the time interval (ii) less (i), irrespective of the particular wave from which the information comes.

The interval since the most recent exit from education training is the primary determinant of current employment situation in our analysis of the school-to-work transition. For much of the analysis, we divide the population into two segments: persons with exit-to-observation interval of up to 5 years (defined as *juniors*), and the majority of the population with intervals longer than 5 years (defined as *seniors*). The analysis excludes persons for whom this interval cannot be defined, i.e. those still in education/training at the observation time. Persons outside the working ages (taken here as 16-59) are also excluded throughout.

The junior population classified by age provides a most important indicator of differences among the countries of the spread of education/training to higher ages ("life-long learning"). Several patterns can be identified.[24] (a) Countries where a majority (50-60%) of those aged 25-29 have had recent education/training; these include Germany, Belgium, Denmark, and Finland in EU15. In these countries, the mean age of the junior group exceeds 30 years, indicating continued education/training to higher ages. In Finland, in fact 20-30% have been in recent education/training even at the highest ages up to 55-59. (b) In France and Spain, a third (30-35%) of the 25-29s have been in recent education/training. Among the remaining countries, this proportion is mostly in the range 20-25%; among the last mentioned, (c) high levels of participation are nevertheless sustained to high ages (10%

[24] Detailed classifications by age are not presented in this report for reasons of space.

or more even among those aged 35-39) in Italy, Austria and the UK, while (d) there is little sustained education/training beyond age 25-29 in Portugal, Ireland, and particularly in Greece. In Greece, the mean age of the junior group is the lowest, at only about 23 years, indicating a relative lack of continued education/training to higher ages.

9.2 The education situation

Firstly, it is instructive to note how the population is distributed over various categories used in the analysis. Levels vary very much among countries. Tertiary level of education is common in Germany, Denmark, Belgium, the UK and Finland (at least 40% of units), while percentages are much lower (10% or under) in Italy, Portugal and Austria.

Some results for selected age groups are shown in Table 9.1. Three broad levels have been defined in terms of ISCED (International Standard Classification of Education). The educational and training levels have been clearly higher in northern countries such as Belgium, Finland, Denmark and Germany; this is also true of the UK except for the inflated number at Low level, and Austria is characterised by high values for the Middle level. Among the other countries, the contrast between the patterns in Italy and Spain is noteworthy. In Spain - in a way similar to France, Greece, and to a lesser extent Ireland – there has been a considerable expansion of education and training at High levels. Italy, in a way similar to Portugal, seems to have largely missed this historical improvement in the achieved levels of education and training. Portugal in fact remains in the least favoured situation among EU15 countries. This contrast between the Italian and Spanish situations has also been noted by Iannelli and Soro-Bonmadì (2001) based on EU-LFS data for 1996.

Table 9.1 Distribution by level of education, for selected age classes

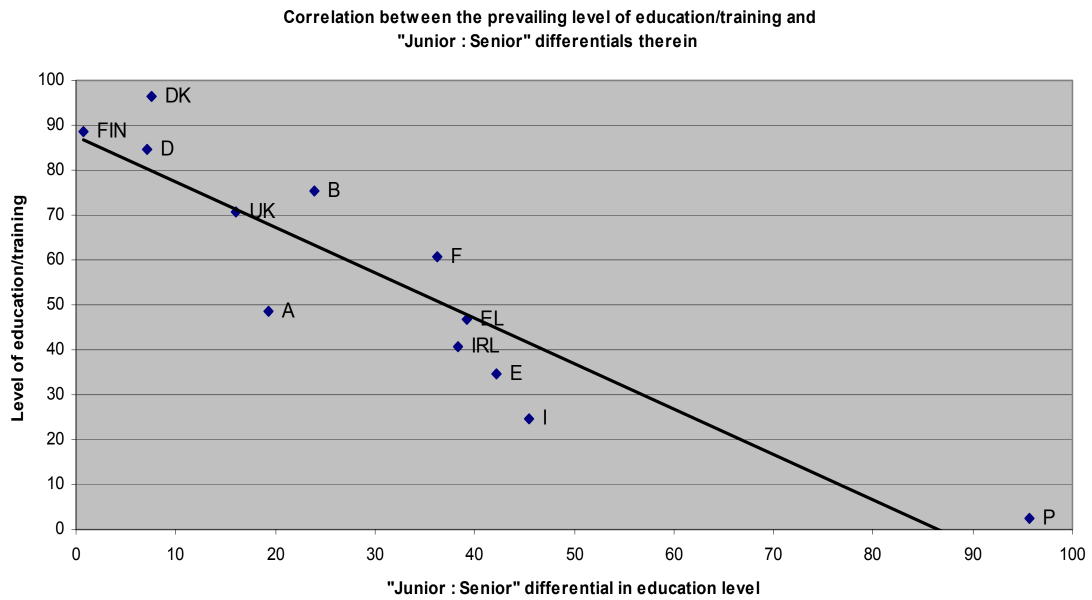
age group	25-29			30-34			55-59		
Level completed	High	Middle	Low	High	Middle	Low	High	Middle	Low
D-SOEP	35	28	37	61	11	28	67	3	30
DK	43	53	4	50	46	4	46	43	11
B	55	32	12	50	34	16	28	33	38
F	41	42	17	30	48	22	15	30	55
UK-BHPS	54	19	26	50	21	29	33	21	46
IRL	28	44	28	21	43	36	12	29	59
I	12	53	35	14	45	41	5	21	74
EL	33	42	25	33	36	31	10	15	76
E	38	26	36	32	24	45	11	11	78
P	13	25	62	11	17	72	5	4	91
A	10	78	11	12	72	16	4	60	36
FIN	54	44	3	57	39	5	38	34	28

Averaged over waves; includes persons still in education or training; High=isced 5-7, Low=isced 0-2.

Below is a brief description of a methodology which can illustrate important aspects of these patterns clearly and simply. Specifically, we want to understand how the level of education and training of recent exits (juniors) differs from that of the rest of the population.

Consider a set of population groups classified by level of education/training completed into three categories: say, with percentages Ll, L2 and L3, from the highest to the lowest level. A 'score' computed as a weighted sum of these proportions, with more weight given to higher levels, would be indicative of differences in the overall level of education/training among the population groups. For this purpose, we have used a score defined as $S = 2 * L_1 + L_2 = 100 + (L_1 - L_3)$.[25] These scores were classified into 36 categories by country (total, male, female), and the resulting set rescaled to the range (0-100) over the categories in the form of an index $Y_i = 100 * (S_i - S_{min}) / (S_{max} - S_{min})$ for category i, where the max and min values are defined over the range of categories being compared. A similar index can be constructed to summarise differentials in the levels of education and training between (1) juniors who have recently (taken here as within past 5 years) completed a relevant education or training course, and (2) seniors who have not done so. For each of these subgroups in each i, we can compute education level scores (S1i, S2i) as defined above, and take their ratio or difference, respectively Ri=(S1i/S2i) or Di=100-(S2i-S1i), as a score measuring the (1):(2) differential. The ratio generally provides a more sensitive measure, but the difference measure is preferable when small denominators or negative quantities are involved. As before, it is convenient to normalise these measures such as $X_i = 100 * (R_i - R_{min}) / (R_{max} - R_{min})$, with the min and max values defined over the range of i values of interest. Higher values of X indicate that the educational position of the junior group is better (or less disadvantaged) compared to the senior group.

Figure 9.1 Correlation between educational level and "Junior : Senior" differential



The results reported in Figure 9.1 indicate a very large negative correlation (-0.9) between (Yi,Xi), implying a substantial narrowing of the large national differences (in Yi) when we consider the junior population. This reflects the historical trend of narrowing national differentials among EU countries, but it is important to note that this is also reflective of self-selectivity of the junior group in relation to education/training. This self-selectivity tends to be stronger where the undertaking of education/training tends to be concentrated

---

25 Summarisation of detailed information in the form of a single score such as the above is a very useful device when the results have to be disaggregated to the regional level.

at lower ages (e.g. Portugal; see Section 9.1), and weaker when that is spread out over a wide age range (e.g., Finland). In any case, the important point here is that national differences in the achieved educational level are much smaller among the junior groups than the more long-standing, overall national differences.

9.3 The Employment Situation

How do those who have recently undergone a school-to-work transition differ from the general population in their employment situation? What kind of differentials exist within and between countries? We describe the employment situation in terms of the following five indicators.[26] With e, s and u, respectively, as the proportions employed, self-employed and unemployed in the working age population (z) (taken here as aged 16-59),

%I inactivity rate	% economically inactive of the working age population, =1 - (e+s+u)/z
%U unemployment rate	% unemployed of the economically active population, =u/(e+s+u)
%S self-employment rate	% self-employed of the working population, =s/(e+s)
%P part-time work rate	% working part-time (p) of the working population, =p/(e+s)
%T temporary employment rate	% of the employed population working without a permanent contract and/or working part-time (t), =t/e

As in the previous section, in order to consolidate the information, we compute normalised indices of the overall levels and of junior:senior differentials in the employment situation. For each employment situation indicator, such as %U, we can construct a score $S_i = 100*(U_{max} - U_i)/(U_{max} - U_{min})$, normalised to the range (0-100) over the population categories. As before, max and min values above are defined over the set of population categories of interest in the analysis, such as the 36 population categories, by country (total, male, female). [27]

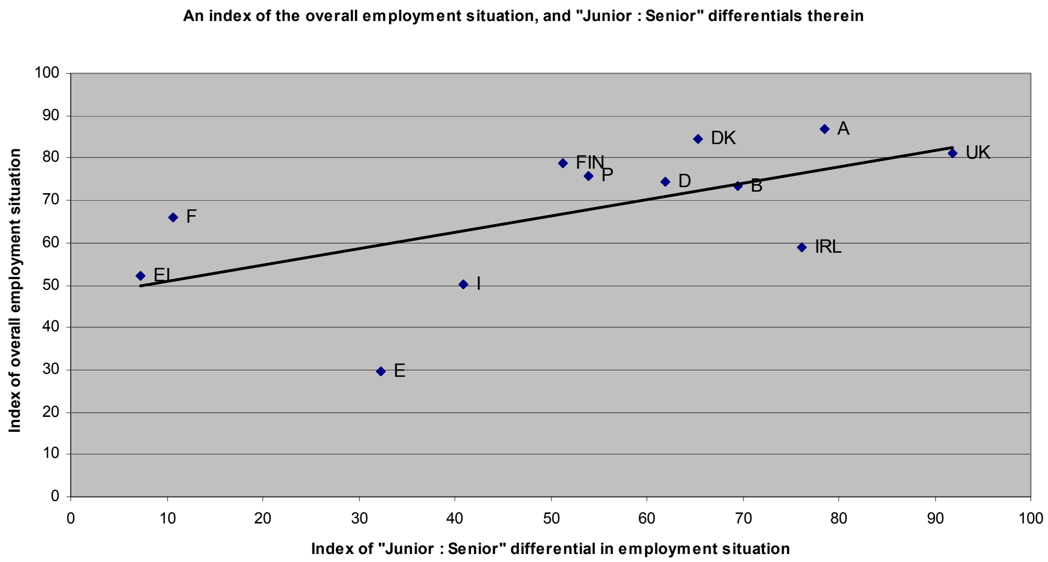
Next, in order to summarise the overall employment situation of groups, we use a weighted index combining the different employment measure (k): $Y_i = \Sigma_k W_k.S_{ki}, \quad \Sigma_k W_k = 1$. The choice of the weights is necessarily subjective. However, clearly unemployment (U) in the present context is the most important indicator; the next in importance is perhaps the indicator of the absence of full-time work with a regular or permanent contract (T) for those who have obtained employment. Consequently, we have give a weight of 0.5 to SU, 0.2 to ST, and 0.1 to each of the other score SI, SS and SP. The combined measure Yi is an index of the overall employment situation of population category i. A similar index can be constructed to summarise differentials in the levels of education and training between (1) juniors who have recently completed a relevant education or training course, and (2) seniors who have not done so. For these two subgroups in each i, we compute weighted employment situation scores (S1i, S2i) and take a measure of their difference, Di=100-(S2i-S1i), as a score measuring the “J-S” differential. As before, it is convenient to normalise these measures such as $X_i = 100*(D_i - D_{min})/(D_{max} - D_{min})$, with the min and max

[26] In ECHP, economic activity has been recorded using two concepts: using the standard ILO definitions, and in terms of self-declared status. We have computed the results for both these types of measures. Generally, the differences in the conclusions are small, especially in relation to the most important indicator (%U), and therefore we mostly have reported results only for the ILO measures.

[27] Note that while the original rates (U etc) indicate a negative (unfavourable) situation, the corresponding scores as defined above are positive (favourable): a score of 100 means the most favoured category, and 0 the least favoured.

values defined over the range of i values of interest. Higher values of X indicate that the employment situation of the junior group is better (or less disadvantaged) compared to the senior group. The results in Figure 9.2 indicate a positive correlation (0.5-0.6) between (Yi,Xi). This implies that in countries where the overall employment situation is already less favourable - Italy, Greece and especially Spain - the *disadvantage of the junior group tends to be accentuated* due to the adverse junior:senior differential in the employment situation. More detailed analysis (not reported here) indicates that this pattern applies equally for males and females separately.

Figure 9.2 Overall employment situation, and "Junior : Senior" differentials therein



9.4 Employment Situation and Educational Level

The employment situation of those with low levels of education/training can be expected to be worse than that of the better educated and trained. But how big are these differences? Do they relate to national differences in the overall employment situation? Figure 9.3 explores the association between the level of education and the employment situation in general, preliminary terms. The analysis is restricted to the junior population of interest in the study of school-to-work transition. Using the same methodology (and weights) as above, we compute weighted score (Si), and then the index of overall employment situation (Yi) in each country (range 0 least favourable, to 100 most favourable among the countries). Similarly computed weighted scores for the two extreme groups by level of education, S1i (highest) and S3i (lowest), are differenced to obtain a score summarising employment situation differentials by level of education, Di=100-(S3i-S1i); these scores are normalised to corresponding indices Xi as before.

Higher values of X indicate greater equality, i.e. smaller differentials, in the employment situation among groups with different levels of education/training. The results indicate a negative but weak correlation (-0.3) between the overall employment situation (Yi) of the juniors in a country or population group i, and the differential in this situation by the level of education/training (Xi). It is more instructive to note that in fact *countries form fairly distinct clusters*. For instance, the UK and Ireland are characterised by good overall employment situation (index=high) among recent exits, but among them there are marked differences by the level of education/training (low index of "equality" – meaning that the poorly educated/trained do much worse); by contrast in France and Italy, the overall employment situation among recent exits is only moderately good, but the situation is more equitable as concerns persons with different levels of education/training.

Cluster	Index of overall employment situation (Y)	Index of "equality" by level of education (X)
A, FIN, P	High	High
DK, D, B	High	Medium
UK, IRL	High	Low
I, F	Medium	High
EL, E	Low	Medium

The identification of such "clusters" is an important aspect of the analysis, given the diverse situation in EU countries, accentuated with EU enlargement. It is likely to be all the more important in the analysis of regional patterns, which may differ according to such clusters of countries and regions.

Figure 9.3 Employment situation and education level differentials, for "Juniors"

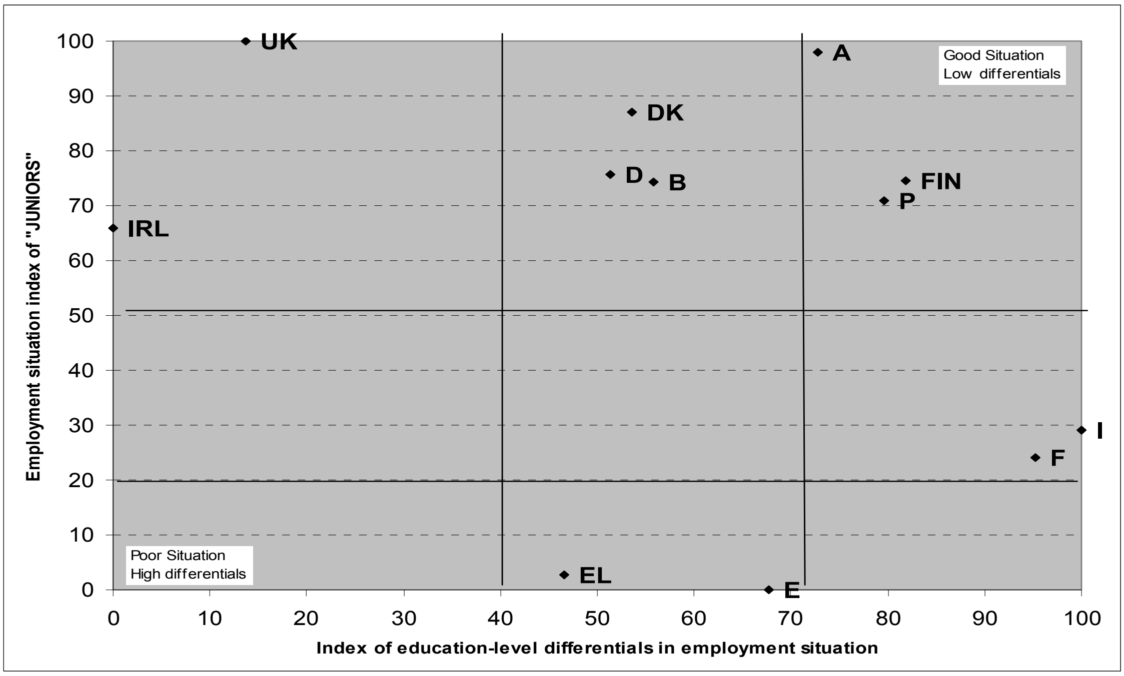
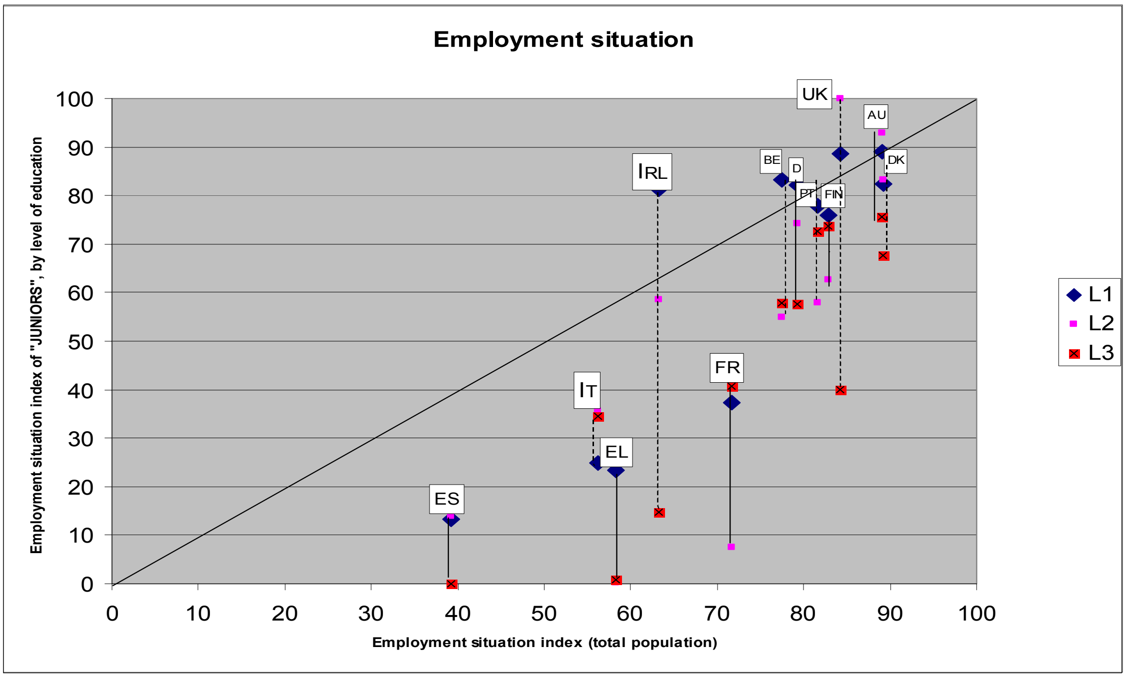


Figure 9.4 shows in more detail the information summarised in the previous graph. As before, Y-axis shows an index reflecting the overall employment situation of persons who have completed a relevant education/training course within past 5 year – but separately for the three levels of qualification (from L1 the highest, to L3 the lowest). The same index for the total population of each country is shown on X-axis. Hence the actual employment situation of the juniors by education/training level can be seen against that of the general population in each country. Points below the 45^0 line imply relative disadvantage of the former, as in the case of Spain, Greece, Italy, France, and also, the least educated/trained in Ireland and the UK.

Figure 9.4 Relative employment situation of "Juniors" by level of education/training



For a given country, the spread of the index from the highest to the lowest level, (L1-L3), indicates the magnitude of the disparity by education/training level among the juniors. The UK and Ireland are distinguished by the largest disparity of this type, as already seen in Figure 9.3. In fact, the highest education/training level does not always go with the best employment situation, as for instance in the case of Italy and France.[28] The other country-clusters can be identified as in the previous graph. In a number of countries, while the average employment situation of the juniors is less favourable than the overall national situation, better educated/trained juniors nevertheless tend to do better than the general population.

9.5 Poor: Rich Differences

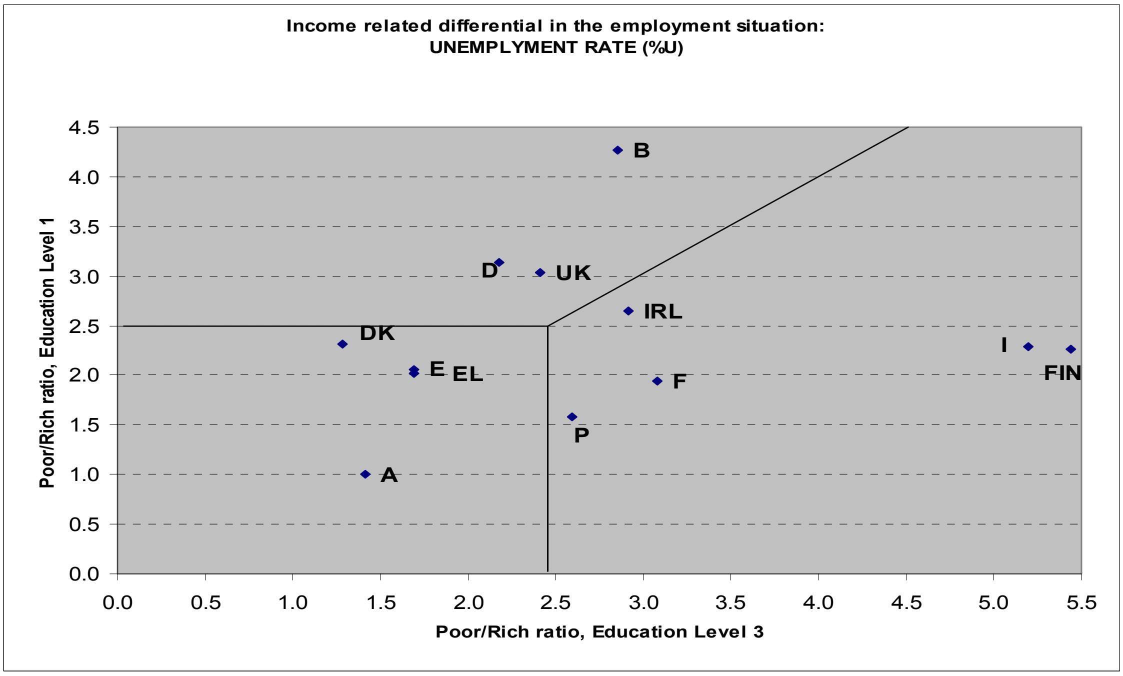
After having completed a course of education or training, does the employment situation depend - even for a given educational level achieved – on the individual's income level? How do any differences between poor and rich vary by the level of education or training achieved? The analysis in this section is restricted to the junior population. To identify income differentials, we have ranked these individuals within each country according to the level of their equivalised household income, and taken the bottom 25% as the "poor" and the remaining 75% as "rich" (or non-poor) for the purpose of this comparison.

[28] Of course such a "reversed pattern" is possible and may be real, but smallness of the sample sizes available should also be kept in view. Such a pattern implies a more favourable relative position of the less educated.

There are indeed large Poor:Rich differentials in all countries, in all indicators of the employment situation of individuals who have recently completed an educational/training qualification and have potentially entered the labour market; and furthermore, this applies in most cases when the level of education/training completed is controlled. This is as may be expected, but the *magnitude of the differentials remains remarkable*. For instance, the unemployment rate among persons in school-to-work transition as defined above is 4 times higher for individuals from poorer households in Ireland and Belgium, 3-4 times higher in UK, Finland and Italy, 2-3 times in Germany, France and Portugal, and below 2 times in only the remaining countries, with the lowest value (1.6 times) in Austria; the simple average over 12 countries is 2.7 times. The overall pattern is essentially the same within each level of education, though individual figures at the country level are subject to fluctuations due to smallness of the sample sizes.

Figure 9.5 compares the Poor:Rich differences in the unemployment rates by country for the highest and the lowest levels of education/training achieved (L1 and L3 respectively). The situation of counties falls into three clusters in terms of the Poor:Rich ratio in unemployment rates: (a) ratio under 2.5 at all education levels – Austria, Greece, Spain and Denmark, with Portugal at the margin; (b) income differentials more extreme at the highest educational level (L1) – Belgium, Germany and the UK; and (c) income differentials more extreme at the lowest educational level (L3) – Finland, Italy and France, with Ireland at the margin.

Figure 9.5 Income related differentials in unemployment rates, for "Juniors"



Similar differentials are observed in other employment related indicators. At all levels of education and training, inactivity rate among juniors from poorer households is more than twice as high as that for richer individuals in Belgium, Ireland Germany and Finland. By contrast, France, Italy, Spain and Greece form a group with the smallest income differential in activity rates at all educational levels. In the UK, Denmark and Portugal, income differentials in activity rates are sharpest at the highest level of education.

Concerning self-employment rates, these among the poor in school-to-work transition are over three times higher in Portugal, and twice as high in Italy and France, compared with self-employment rates among non-poor persons. This applies equally at all levels of education/training. Austria, Finland and the UK are characterised by sharp income differentials at lower levels of education/training.

9.6 Dynamic aspects

As seen in the preceding section, the employment situation of junior is considerable worse than that of the general working-age population. This applies not only in relation to employment, but also in relation to other indicators of the employment situation. It applies generally across countries, by gender, by level of education, and more forcefully in the case of persons from poorer households. Now we examine the employment situation following completion of an education/training course from a more dynamic perspective, as a function of the time elapsed since last exit from education/training. How does the employment situation look after 1, 2, 3… years ? Does it begin to resemble that of the general population after a relatively long period such as 5 year?

Unemployment following exit from education/training

Let us first look at the overall pattern – the picture obtained by taking a simple (unweighted) average over the EU15 countries covered. During the first year following the completion of education/training (i.e., after an average duration of 6 months), around 15% of males and 25% of females are reported to be unemployed. Similar female : male differences in the unemployment rate also exist in the general population, but the here the actual unemployment level is *two and a half (2.5) times* higher than the general level. Even after an elapse of 4-5 years following the completion of education/training, unemployment rates remain 50% higher than those at the general level.

Figure 9.6 Incidence of unemployment following exit from education/training (EU15)

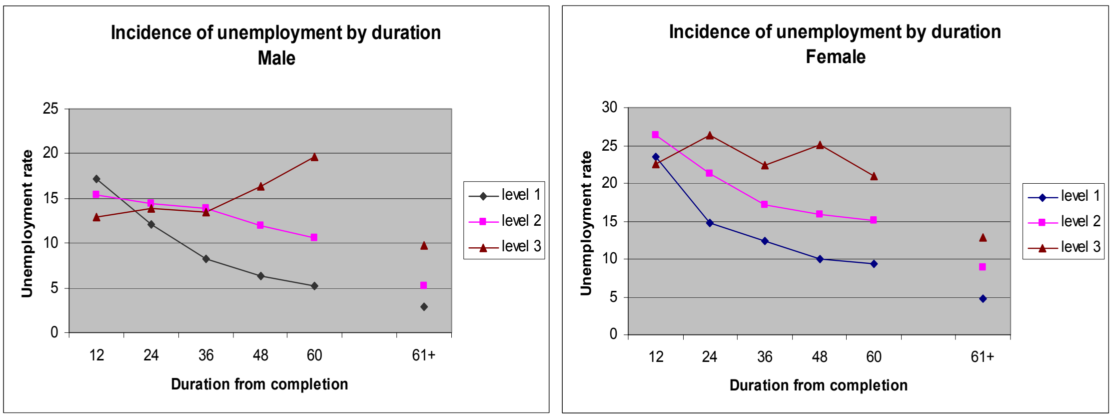


Figure 9.6 shows the pattern over time by gender and level of education/training, again averaged over countries. These provide a useful indication of the situation despite the limitation of such simple averaging. For the highest education/training group, there is a consistent and substantial improvement over time, though the unemployment levels for the juniors still remain notably higher than those prevailing for the population as a whole. The pattern holds for both males and females. For the intermediate level education/training

group, the improvement with time is somewhat less marked, but still quite significant and consistent – especially among females. By contrast, *for those with the lowest level of education/training, there is little consistent improvement with time for females, and for males the situation appears even to get worse!* The differences with the general population at the same (lowest) education/training level remain very large even after 5 years. This may well reflect a worsening historical trend for the least educated – especially among males, though the generally high levels of unemployment among females should be kept in mind in this examination of patterns and trends.

It is also necessary to keep in view differences between national situations. Detailed examination, using the survey data alone, is limited by the small sample sizes available. Nevertheless, we have found a remarkable similarity in the situation of many countries. In relation to inactivity and self-employment rates as well, while there are some national differences, the pattern seems to be generally quite similar across countries.

Identification of disadvantaged groups

The index of overall employment situation scaled uniformly across countries in the EU reflects the position of each category at the EU level. The classification categories (by country, level of education and duration since completion of education/training, etc.) can be raked according to the employment situation score. In this way, treating the *classification categories as units of analysis*, we can determine say the median score, and identify particularly disadvantaged groups as those with a score below a certain threshold, such as 50%, 60% or 70% of the EU median score. Furthermore, we can see which categories are at the margins and enter or leave the state of "severe employment disadvantage" as the threshold is changed. We have applied this procedure to identify disadvantaged groups in Figure 9.7.

Figure 9.7 Categories in "severely adverse employment situation"

Education Level 1								Level 2							Level 3							all levels						
Months since completed	12	24	36	48	60	.1-60	61+	12	24	36	48	60	.1-60	61+	12	24	36	48	60	.1-60	61+	12	24	36	48	60	.1-60	61+
DE																												
DK																												
BE																												
FR	X							X	X	X	L		X		X							X						
UK																												
IE																X	X	X	L	L								
IT	X	L						X										X	X			X						
GR	X	X				L		X	X	X	L	L	X			X	X	X	X	X		X	X	X			X	
ES	X	X	L			X		X	L	X	X	X	X		X	X	X	X	X	X	L	X	X	X	X	X	X	
PT								L																				
AT																												
FI																												

X with "employment situation score" below 50% of the median score, EU-wide **L** score between 50-60% of the median

This illustrates a mode of analysis and presentation which can be applied equally to regions or other geographical domains as the units, in order to identify the most disadvantaged ones.

9.7 Concluding remarks

The results discussed in this chapter indicate very clearly that even after the elapse of a considerable amount of time (up to 5 years) following an exit from education and training, the employment situation of the concerned groups remains much worse than that of the general population in the same country. Furthermore, some categories are much more disadvantaged than others as concerns the school-to-work transition.

Just as these differentials vary across countries, they can be expected to vary across regions and other geographic divisions. Though the primary focus of the empirical results presented here has been on subgroup analysis – subgroups defined in terms of demographic and other characteristics which make them geographically dispersed in the national population – the same approach can be extended to disaggregation by geographic regions and urban-rural classification of localities. Such geographical disaggregation can of course be severely limited by small sample sizes in the available data sets. The reliability of the results can be improved by combining the survey data with other large-scale (but usually less intensive and also less specific) data sources using small-domain estimation methodology of the type described in later chapters. We believe that one of the particularly fruitful approaches for this purpose can be the fairly simple logistic regression methodology which imputes to a large data source the complex employment situation indicators for recent exits from education/training obtained from a rich but relative small data source, on the basis of common demographic and social indicators measured in both the sources. Chapter 14 briefly reviews the use of such methodology in the United Kingdom (see for instance, Gordon, 1995), and the use of a more sophisticated model applied in Albania (Betti, Ballini and Neri, 2003).

Chapter 10
Area-level indicators from NewCronos (Eurostat Free Dissemination Database)

10.1 NewCronos as a resource for regional indicators

The NewCronos window (now termed "Eurostat Free Dissemination Database") provides a valuable data resource for the construction of regional indicators.

It is important to note at the outset that Eurostat Free Dissemination Database or NewCronos is not in itself a source of original data, but represents a compilation of information from a diversity of sources, presented in the form of very detailed tabulations. In reference to the information we have used in this chapter and in the small area estimation models in Chapters 11-13, for convenience we refer to "NewCronos" as a "data source". It may be preferable to indicate, in each case, the original data source. We are not able to do so because of lack of full information. In any case, it should be possible for any user interested in particular data items to obtain this information from Eurostat.

We believe that this resource has hitherto been under-utilised, and that there is a great potential for more thorough exploitation of the information which already exists. While direct indicators of regional poverty and living conditions are generally not available with sufficient regional breakdown in NewCronos, several exceptionally positive aspects of the resource need to be appreciated. Some of these become even more important as we move down from the national to the regional level.

1. A wide range of subject-matter areas are covered in the very detailed tabulations provided. These can be utilised to construct many direct indicators pertaining to poverty and living conditions, as well as to obtain many more variables correlated with direct indicators.
2. Detailed break-down – especially for variables correlated with direct indicators of interest – is available, mostly to NUTS2, and in a few cases to NUTS3 level.
3. NewCronos is a dynamic resource, in principle regularly updated as new or improved data become available.
4. Of course, the timeliness, statistical quality and comparability of NewCronos depends on the original data sources from which the information is derived. But *the very process of bringing those data into a unified framework through a centralised operation can be expected to enhance data quality in all its dimensions* – coherence, consistency, completeness, transparency, and also comparability.
5. The data base is accessible and convenient to use, and most importantly, it has recently been placed in the public domain as Eurostat Free Dissemination Database.

REGIO domain

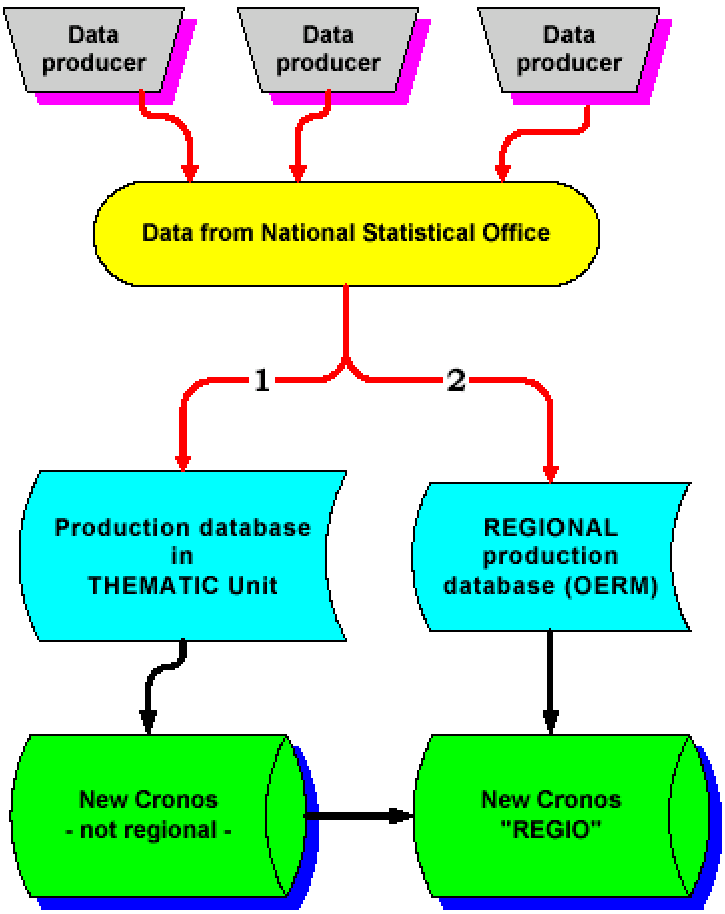
For our purposes we used the REGIO domain.

The following description is taken from European Regional Statistics: Reference Guide (Eurostat, 2004 edition).

Eurostat's regional statistics cover the principal aspects of the economic and social life of the European Union: demography, economic accounts, employment, unemployment, and so on. The concepts and definitions used are as close as possible to those used by Eurostat for the production or collection of statistics at national level.

Comparable regional statistics, a major part of the European Statistical System, are used for a wide range of purposes, *inter alia* for allocating structural funds in a rational and coherent way.

The standard model for the data flow has been as follows (see the diagram): first, the data from various national sources are compiled in the National Statistical Office of each country and then sent to the thematic units of Eurostat, who validate the data. This data set is then loaded into NewCronos by the thematic unit in question. The Regional Statistics Section copies this information from the thematic domain into the REGIO domain of NewCronos. This is option 1. Alternatively, data may be sent directly to the regional team of Eurostat and then, after validation, loaded into the REGIO domain of NewCronos (option 2).



Source: Eurostat (2004a), *European Regional Statistics - Reference Guide*.

Table 10.1 A brief summary of the contents of the REGIO domain

Agriculture	The agricultural collection of the REGIO database contains several variables such as: agricultural accounts, structure of agricultural holdings, land use, some agricultural production, etc.
Demographic Statistics	In general the statistics refer to the resident population of each country. In accordance with this concept, persons normally resident in a country but temporarily absent on business, holiday, etc., are included in the total population figure, whilst foreigners temporarily resident in the country for similar reasons are excluded. Nationality is not taken into consideration when this concept is applied, and foreigners whose usual place of residence is in that country are included along with the citizens of that country.
Migration statistics	The regional migration datasets provide the national figures corresponding to the in and out movements within the country. No distinction is made between national and non-national residents but movements are differentiated depending on whether or not they involve the crossing of national borders
Economic accounts	The regional accounts are compiled in accordance with the European System of Integrated Economic Accounts (ESA), which should be referred to for the definition of the aggregates.
Education statistics	There are two major sources for data on education at regional level: a) The regional tables of the UOE data collection: Data are collected using EU specific tables included as a supplement for EU countries in the joint UNESCO-OECD-Eurostat data collection on education. The UOE data collection covers primarily the "regular" school and university system. Data included in the REGIO data base concern: Pupils and students and Non-national students in tertiary education by citizenship. There are two sets of tables presenting data collected on the basis of two different versions of the International Standard Classification of Education (ISCED) of 1976 and 1997. The version of ISCED used is already indicated in the title of each table. b) The EU Labour Force Survey (Data are collected through the LFS concerning the highest level of education attained (educational attainment) as well as on recent or current participation of the population in education and training.)
Environment statistics	Environment covers three major environmental domains: water uses, waste water management and municipal and hazardous waste management. Each domain is largely inspired by the the joint OECD/Eurostat questionnaire on the State of the Environment
Regional Labour Market	The results of the Labour Force Survey (LFS) refer exclusively to private households. The Community LFS is now carried out in all four quarters and almost all EU and AC countries. Therefore in 2003, Eurostat in co-operation with National Statistical Institutes implemented a major reform of regional labour market statistics, switching from second quarter LFS results to LFS annual averages (calculated from 1999 onwards). As LFS, like all surveys, is based on a sample of population
Science and Technology (R&D, patents)	*Definition of R&D* : Research and Development includes creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications *R&D expenditure* : R&D expenses are all funds used for the realisation of R&D. They include current expenses such as employment costs or expenditures on materials, plus capital expenditure on, for example, buildings or equipment. Regional data on R&D, at NUTS levels 1 and 2, are supplied by Member States, generally on the base of national surveys. Some Member States cannot supply a regional breakdown for all R&D expenses *R&D personnel* : R&D personnel includes all persons employed directly on R&D sectors plus any supplying direct services to R&D such as manager, administrative staff and office staff *R&D sectors* : The structure of the sectors in the R&D domain differs in one major point from the sectorial structure of National Accounts. Due to the special importance of Universities and Technical Colleges, the sector "government" of National Accounts is split in two: "government sector" and "Higher education sector". The latter includes not only all universities, colleges of technology and other institutes of post-secondary education , but also all research institutes, experimental stations and clinics operating under the direct control, administrated by or associated with higher education establishments *Patents* : A patent is a legal title of industrial property granting its owner the exclusive right to exploit an invention commercially for a limited area and time. Patent data provide a measure of R&D output.
Structural business statistics	The SBS (structural business statistics) describes the activity of businesses in the European Union. The regulation applies to all market activities (except agriculture) normally included in industry, construction, the distributive trades and services
Health statistics	Causes of death Health personnel Health infrastructure (hospital beds)
Tourism statistics	This collection on regional tourism statistics contains data on 1) The capacity of collective tourist accommodation (number of establishments, number of bedrooms, number of bedplaces) and 2)Occupancy in collective accommodation establishments (arrivals and nights spent, broken down into residents and non-residents).
Transport and energy statistics	Energy Net production of electrical energy is measured as it leaves the power station, i.e. after deduction of consumption for auxiliary services and losses in the power station transformers. Transport The concepts used for drawing up Community data on transport are summarized in the Transport Statistical Yearbook published by Eurostat

Table 10.1 provides a brief summary of the contents of the REGIO domain (from Eurostat (2004a), *European Regional Statistics - Reference Guide*).

In Table 10.2 a summary is provided of the type of data available in various dimensions which appear pertinent for the construction of regional indicators on poverty and related aspects. The table also indicates the level of regional breakdown in most countries, distinguishing between EU15, NMS-10, and Candidate Countries. Mostly the breakdown is to NUTS2 level. Data available to NUTS3 are highlighted in Table 10.3.

Table 10.2
Summary of useful regional indicators which can be derived from NewCronos

	EU15	NMS10	Candidate	Notes
Demographic statistics				
Population density	Nuts3	Nuts3	Nuts3	
Crude birth rate and crude death rate	Nuts3	Nuts3	Nuts3	
Regional scenarios on population by sex and age groups (NUTS95)	Nuts2			[1]
Infant mortality	Nuts2	Nuts2	Nuts2	
Economic accounts				
Gross domestic product - ESA95	Nuts3	Nuts3	Nuts3	
Secondary distribution of income account of households	Nuts2	Nuts2	Nuts2	[2]
Balance of primary income, net (uses)	Nuts2	Nuts2	Nuts2	
Social benefits other than social transfers in kind (resources)	Nuts2	Nuts2	Nuts2	
Other current transfers, received (resources)	Nuts2	Nuts2	Nuts2	
Current taxes on income, wealth, etc. (uses)	Nuts2	Nuts2	Nuts2	
Social contributions (uses)	Nuts2	Nuts2	Nuts2	
Other current transfers, paid (uses)	Nuts2	Nuts2	Nuts2	
Disposable income, net (uses)	Nuts2	Nuts2	Nuts2	
Income of households - ESA95	Nuts2	Nuts2	Nuts2	
Education statistics				
Number of students by sex and age - (ISCED97)	Nuts2	Nuts2		[3]
Structural business statistics				
Structural business statistics by economic activity	Nuts2	Nuts2	Nuts2	
Wages and Salaries	Nuts2	Nuts2	Nuts2	
Number of persons employed	Nuts2	Nuts2	Nuts2	
Growth rate of employment (%)	Nuts2	Nuts2	Nuts2	
Investment per person employed	Nuts2	Nuts2	Nuts2	
Health statistics				
Causes of death by region - Crude death rate	Nuts2	Nuts2	Nuts2	
Causes of death by region - Crude Death Rate (3 years average)	Nuts2	Nuts2	Nuts2	
AIDS (HIV-disease)	Nuts2	Nuts2	Nuts2	
Accidents	Nuts2	Nuts2	Nuts2	
Homicide, assault	Nuts2	Nuts2	Nuts2	
Health personnel - Absolute numbers and rate per 100.000 inhabitants	Nuts2	Nuts2	Nuts2	[4]
Hospital beds - Absolute numbers and rate per 100.000 inhabitants	Nuts2	Nuts2	Nuts2	[4]
Infectious diseases - Reported cases and incidence rates per 100.000 inhabitants	Nuts2	Nuts2	Nuts2	
Annual regional statistics				
Regional data (according to Nuts 2003)	Nuts2	Nuts2		[5]
Regional labour market				
Economically active population by sex and age	Nuts3	Nuts3	Nuts3	
Economic activity rates by sex and age	Nuts2	Nuts2	Nuts2	
Economically active population by sex, age and highest level of education attained	Nuts2	Nuts2	Nuts2	
Regional employment - LFS series				
Employment by sex and age	Nuts2	Nuts2	Nuts2	
Employment by professional status	Nuts2	Nuts2	Nuts2	
Employment by full-time/part-time and sex	Nuts2	Nuts2	Nuts2	
Employment by sex, age and highest level of education attained	Nuts2	Nuts2	Nuts2	
Employment and commuting among NUTS level 2 regions	Nuts2	Nuts2	Nuts2	
Employment rates by sex and age	Nuts2	Nuts2	Nuts2	
Dispersion of regional (NUTS level 2) employment rates of age group 15-64	C	C	C	[6]
Average number of usual weekly hours of work in main job (full-time)	Nuts2	Nuts2	Nuts2	
Regional unemployment - LFS adjusted series				
Unemployment rates by sex and age	Nuts3	Nuts3	Nuts3	
Dispersion of regional (NUTS levels 2 and 3) unemployment rates	C	C	C	[6]
Long-term unemployment (12 months and more)	Nuts2	Nuts2	Nuts2	
Regional socio-demographic labour force statistics - LFS series				
Number of households by degree of urbanisation of residence	Nuts2	Nuts2	Nuts2	[7]
Life-long learning - participation of adults aged 25-64 in education and training	Nuts2	Nuts2	Nuts2	

NOTES

C Available only at the country level
NutsX Maximum break down available to Nuts 'X' (1, 2 or 3) level
[1] The indicators distinguish three scenarios: Low, Base and High; these are available only for EU15.
[2] From this table we can construct the indicator "net to gross ratio" for disposable income.
[3] Laeken Indicators 8 and 18 can be derived at Nuts2 level for EU Member States.
Indicator 8: "Early school leavers not in education or training, by gender"
Indicator 18: "Persons with low educational attainment, by gender"
[4] High presence of missing data in NewCronos table as available at present.
[5] These tables are the same as those in domain Structural Business Statistics, except that here we also have NACE breakdown.
[6] Laeken Indicator 5 can be constructed at country level, based on Nuts2 employment rates.
Indicator 5: "Regional cohesion (dispersion of regional employment rates)"
Dispersion of Unemployment rates can also be considered a useful indicator of 'regional cohesion'.
[7] Degree of urbanisation may perhaps also be considered as a variable for geographical disaggregation.

Table 10.3 Indicators available to NUTS3 level

Demographic statistics	Population density
Demographic statistics	Crude birth rate and crude death rate
Economic accounts	Gross domestic product - ESA95
Regional labour market	Economically active population by sex and age; activity rates
Regional unemployment - LFS adjusted series	Unemployment rates by sex and age

There are three main forms in which variables derived from NewCronos can be utilised for the present purpose.

1. Predictors. A large number of measures, potentially closely related to direct indicators of poverty and deprivation can be constructed. In conjunction with direct indicators obtained from more intensive surveys, these measures can be used as "covariates" or "regressors" to produce more precise indicators using small area estimation (SAE) procedures. In Chapters 10 and 11 we develop and illustrate some such procedures in detail, with applications covering almost all EU and Candidate countries. In Chapter 15 (Statistical Annex), the actual estimates we have produced in this illustrative exercise for a number of indicators on poverty and deprivation are presented, going down to NUTS2 level for all countries, and to NUTS3 level for Italy.

2. Direct deprivation indicators. In addition, some statistics in NewCronos can serve, in their own right, as direct indicators pertaining to poverty and living conditions. In fact, the scope for such use is likely to be greater in the context of regional indicators, compared to that in the national context. This is because *measures of levels* – which are more abundantly available in NewCronos than the generally more complex distributional measures - *can themselves serve as indicators of disparity when compared across regions.*

3. Intermediate output variables. As was noted in Chapter 2, reference is made in the Joint Report on Social Exclusion (European Commission, 2003) to "intermediate output" indicators, which tend to be related to policy. It has been noted that, in fact, the distinction between input-related and performance indicators is not always straightforward and some indicators are better qualified as "intermediate output" indicators. Such indicators express on the one hand the policy effort in favour of those at risk of poverty and on the other hand the impact of social policies as well as of the economic context. Many such indicators can be constructed from NewCronos tabulations.

In the rest of this chapter we provide some examples of such indicators.

10.2 Some illustrations

Figures 10.1 – 10.5 show the following indicators disaggregated to NUTS1 level:

(1) Disposable income in PPS per capita

(2) The ratio of net-to-gross disposable income

(3) Unemployment rates

(4) Long-term unemployment rates

(5) Infant mortality rates

This is a selection from a larger number of similar indicators which may be extracted from NewCronos. These are also among the "covariates" we have used in small area estimation presented in Chapters 11-13. Generally, (1) and (2) tend to be negatively related to poverty rates, and (5) is a strong predictor of the level of poverty and deprivation. The relationship of (3) and (4) to poverty levels is more variable.

In each map, the NUTS2 regions in the up to 26 countries covered are divided into quintiles, from the most deprived group (darkest colour) to the least deprived (lightest colour).

Figure 10.1 Disposable income in PPS per capita. NUTS2 regions

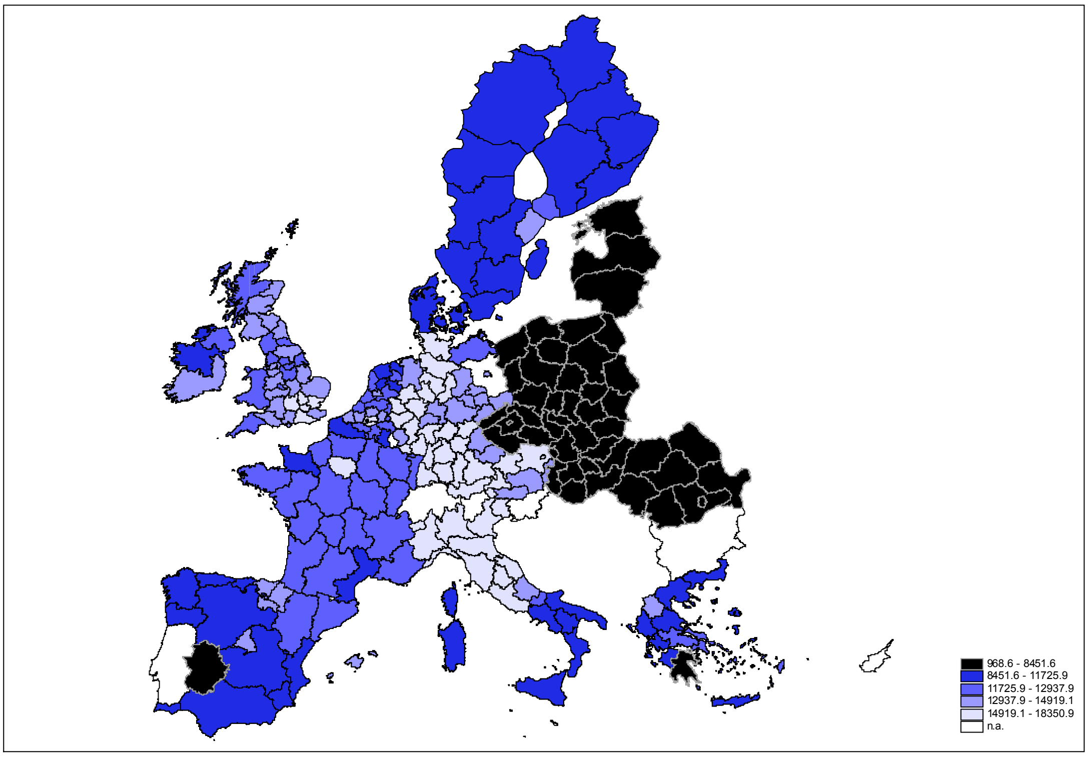


Figure 10.2 The ratio of net-to-gross disposable income. NUTS2 regions

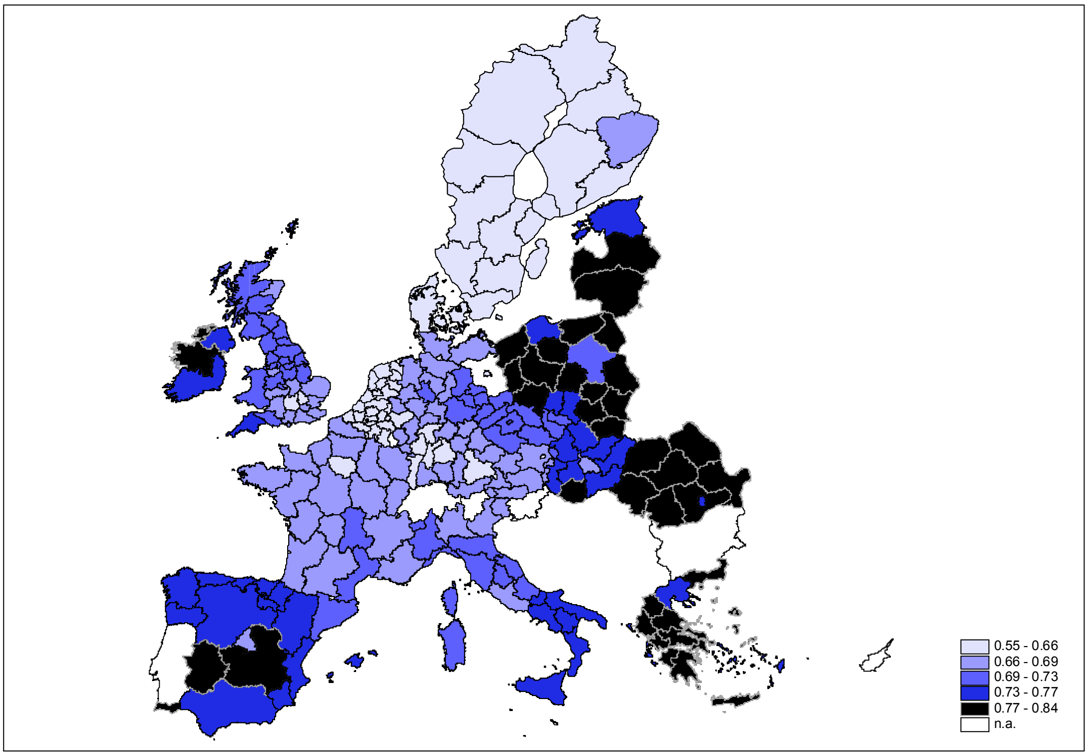


Figure 10.3 Unemployment rates. NUTS2 regions

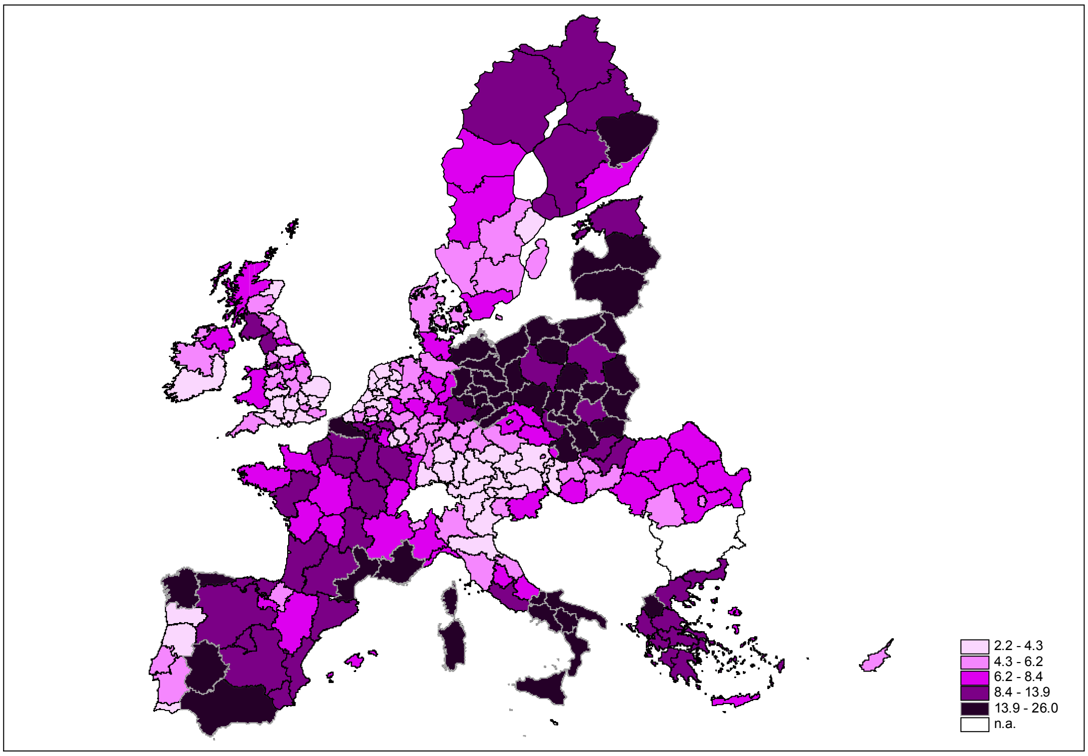
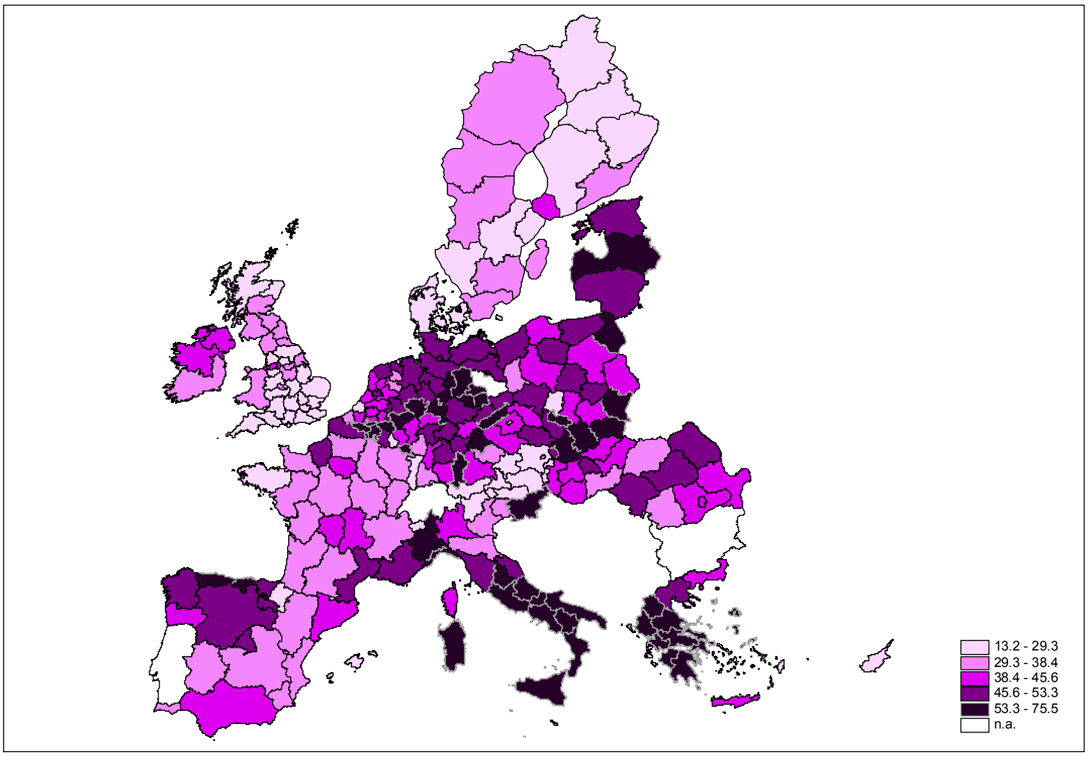
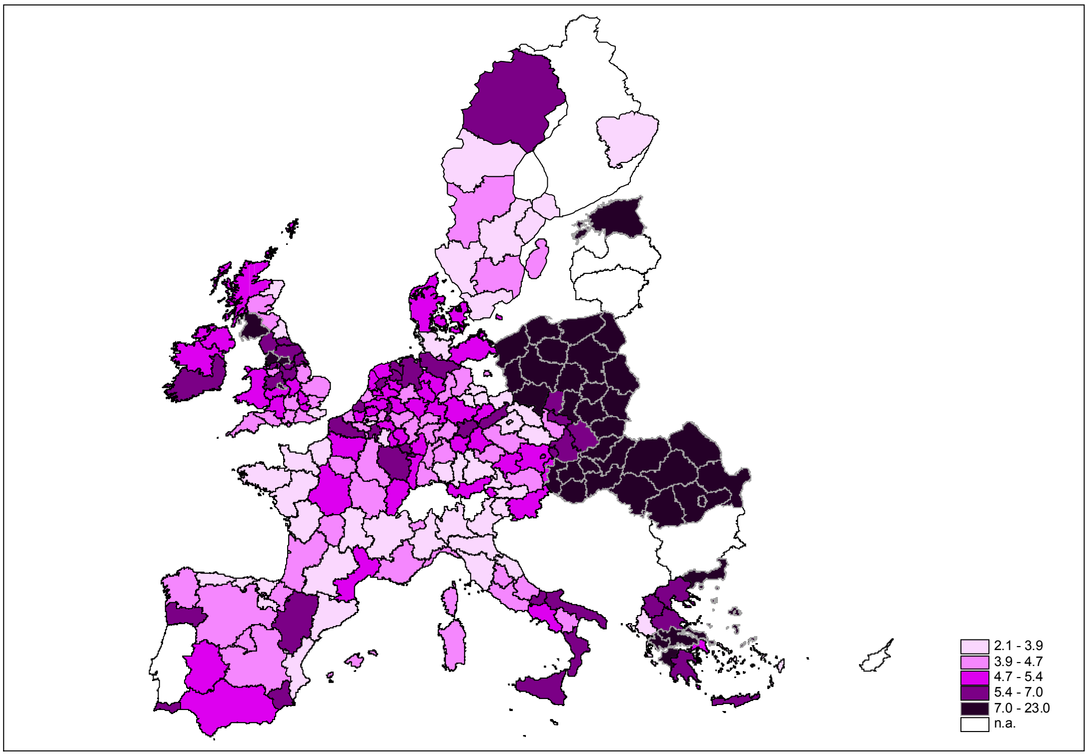


Figure 10.4 Long-term unemployment rates. NUTS2 regions



Note. Long-term unemployment rate is defined as the ratio of persons unemployed for 12 months or longer to the total number of persons unemployed in the region.

Figure 10.5 Infant mortality rates. NUTS2 regions



Chapter 11
Small Area Estimation (SAE): Approach and Methodology

11.1 Introduction

The strategy recommended in this research for the construction of regional indicators of poverty and deprivation has three fundamental aspects:

- making the best use of available sample survey data, such as by cumulating and consolidating the data to construct more robust measures which can permit a greater degree of spatial disaggregation;
- exploiting to the maximum 'meso' data (such as highly disaggregated tabulations available in NewCronos) for the purpose of constructing indicators for small areas;
- using the two sources in combination to produce the best and most complete possible estimates for subnational regions using appropriate small area estimation (SAE) techniques.

The last mentioned is the concern of the rest of this report.

There is a wide variety of SAE techniques available, and the field is rapidly expanding. The suitability and efficiency of a particular technique depends on the specific situation and on the nature of the statistical data available for the purpose.

It is, of course, not possible within the framework of this project to develop and evaluate SAE models for diverse poverty and related indicators in the specific situation of individual countries. Nor would it be appropriate to make such an attempt, given that the knowledge and experience of national statisticians and other researchers about the specific possibilities and limitations in their own country can be expected to be superior than those of the present project team.

Reference must be made to the work of the "Eurarea" project on "Enhancing Small Area Estimation Techniques to Meet European Needs", the final report of which became available in August 2004 (Eurarea, 2004). As stated, the aim of this project was "to provide European Statisticians, particularly government statisticians, with the information they needed to assess and use a range of small area estimation techniques, including techniques incorporating recent theoretical advances." Several classes of small area estimators were investigated and evaluated under that project, in particular: (1) direct estimators; (2) area level synthetic estimators; (3) generalised regression estimators (GREG); and (4) composite ("EBLUB") estimators.

The Eurarea project did not consider, perhaps in view of practical constraints in a multi-country exercise, more complex estimation procedures, such as "Empirical Bayes" and "Hierarchical Bayes" approaches. All the methods considered in the Eurarea project assumed a situation in which unit level information on the target variables (in our case, various measures of poverty, deprivation and social exclusion) is provided by data from a sample survey; then there are auxiliary variables ("covariates") which are known for the target areas (such as all NUTS2 areas in each EU country in our case).

It is important to keep in view that regional estimates may be required for two rather distinct purposes. (1) One is the objective of providing the best possible indicators for each individual area – the aim being to place the area concerned in the context of other areas. Area-specific indicators are required for, for instance, designing policies and monitoring the situation in the area concerned. (2) The second purpose is to provide the best possible indicators capturing the *variability* between areas, such as between NUTS2 regions of each country. Such variability (or more precisely, its lack) is, for instance, the basis of indicators of "regional cohesion", such as Laeken Indicator 5 on the dispersion of regional employment rates. Allocation of certain regional funds may also be determined on the basis of the 'tail' (extreme) end of the distribution of characteristics of administrative of other regional units.

The optimal SAE procedure for the above two objectives – the best estimates for individual areas and the best estimates for their dispersion – may not be the same. It is generally the case that direct (survey based) estimators tend to provide overestimates of the variability among small areas; this is because of the effect of sampling error, which increases with decreasing size of samples in the areas. In other words, direct estimates tend to be more scattered than the true underlying values and, for the same reason, produce distributions which are more extreme than the actual distribution. By contrast, synthetic estimators tend to underestimate the true variability. Composite estimators, being a compromise between (being optimally weighted combinations of) direct and synthetic estimators, are more likely to reflect the true variability than either of the two. This advantage of the composite estimators becomes more marked as we move to smaller domains with smaller sample sizes.

11.2 SAE Models: Theoretical view

In the literature small area models are classified as: (i) *area level random effect models* (Fay and Herriot, 1979), which are used when auxiliary information is available only at area level (such as the prevailing unemployment rate); (ii) *nested error unit level regression models*, used if unit specific covariates (such as the individual's or the household's employment situation) are available at unit level (Battese *et al.*, 1988).

One of the results confirmed through extensive simulations in the Eurarea project is that "area level synthetic estimates tend to produce better results than their unit-level counterparts". This is because regression coefficients calculated at unit level do not always correctly reflect the relationship between the area-level averages involved in the synthetic estimator. The same would apply to composite estimators incorporating the synthetic estimators.

Area Level Random Effect Models

The adopted approach deals with area level random effect models*,* because in the application data are available at area level only.

The equation estimated in the model is given by (11.1) below, and equation used to produce the estimates is (11.3), with the sampling variance for the parameter estimated externally.

Area level random effect models relate small area direct estimators to domain specific covariates, considering the random area effects as independent. The basic area level model includes random area specific effects, and the area specific covariates, $\mathbf{x}_i=(x_{i,1}, x_{i,2}, \ldots, x_{i,p})$, are related to the target parameters θ_i (totals, means, proportion, etc.) as follows:

$$\theta_i = \mathbf{x}_i\beta + z_i v_i \quad with\ i = 1...m \tag{11.1}$$

where z_i are known positive constants, β is the regression parameters vector $px1$, v_i are independent and identically distributed random variables with 0 mean and variance σ_v^2. Moreover it is assumed that the direct estimators $\hat{\theta}_i$ are available and design unbiased:

$$\hat{\theta}_i = \theta_i + e_i \tag{11.2}$$

where e_i are independent sampling errors with zero mean and known variance ψ_i. Combining the above two equations, the obtained model is:

$$\hat{\theta}_i = \mathbf{x}_i\beta + z_i v_i + e_i \quad with\ i = 1...m$$

which is a special case of the linear mixed models with a diagonal covariance structure. The mxm covariance matrices of v_i and e_i are respectively:

$$G = \begin{bmatrix} \sigma_v^2 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_v^2 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_v^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \sigma_v^2 \end{bmatrix} = \sigma_v^2 I$$

and

$$R = \begin{bmatrix} \psi_1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \psi_2 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \psi_i & \cdots & 0 \\ \vdots & \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \psi_m \end{bmatrix} = diag(\psi_i).$$

It follows that the covariance matrix of $\hat{\theta}$ is:

$$V = R + ZGZ^T.$$

The **Best Linear Unbiased Predictor** (BLUP) estimator of θ_i is:

$$\tilde{\theta}_i(\sigma_v^2) = \mathbf{x}_i\hat{\beta} + b_i^T GZ^T V^{-1}(\hat{\theta}_i - \mathbf{x}_i\hat{\beta}) \qquad (11.3)$$

where b_i^T is a *mx1* vector (0,0,…,0,1,0,..,0) with 1 referred to the *i*-th area and β are estimated by generalized least square as $\hat{\beta} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1}\hat{\theta}$.

The BLUP estimator is a weighted average of the design-based estimator and the regression synthetic estimator:

$$\tilde{\theta}_i(\sigma_v^2) = \gamma_i\hat{\theta}_i + (1-\gamma_i)\mathbf{x}_i\hat{\beta}$$

where: $\gamma_i = \dfrac{\sigma_v^2}{\sigma_v^2 + \psi_i}$ is a weight (or 'shrinkage factor') which assumes values in the range [0-1]. This parameter measures the uncertainty in modelling θ_i. (Gosh and Rao, 1994). Mathematical details for the BLUP estimators are available in Handerson (1950).

The mean square error of the BLUP estimator depends on the variance parameter σ_v^2 and it is:

$$MSE[\tilde{\theta}_i(\sigma_v^2)] = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2)$$

with

$$g_{1i}(\sigma_v^2) = b_i^T(G - GZ^T V^{-1} G)b_i = \sigma_v^2 z_i^2 \psi_i(\sigma_v^2 z_i^2 + \psi_i)^{-1} = \gamma_i\psi_i$$

and

$$g_{2i}(\sigma_v^2) = (\mathbf{x}_i - b_i^T GZ^T V^{-1}X)(X^T V^{-1} X)^{-1}(\mathbf{x}_i - b_i^T GZ^T V^{-1} X)^T =$$

$$= (1-\gamma_i)^2\mathbf{x}_i\left[\frac{\sum_{i=1}^m \mathbf{x}_i^T\mathbf{x}_i}{(\sigma_v^2 z_i^2 + \psi_i)}\right]^{-1}\mathbf{x}_i^T$$

where the second term g_{2i} is due to the estimate β (Rao, 2003).

In practice the variance parameter σ_v^2 is unknown and it is replaced by its estimator $\hat{\sigma}_v^2$, so that a two stage estimator $\tilde{\theta}\left(\hat{\sigma}_v^2\right)$ is obtained; it is called Empirical BLUP (EBLUP). The EBLUP estimator has the following properties: *i)* it is unbiased for θ, *ii)* $E\left[\tilde{\theta}\left(\hat{\sigma}_v^2\right)\right]$ is finite, *iii)* $\hat{\sigma}_v^2$ is any translation invariant estimator of σ_v^2 (Kackar and Harville, 1984).

Assuming normality, the variance of the random effects can be estimated either by Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) methods. The MSE of the EBLUP seems to be insensitive to the choice of the estimator $\hat{\sigma}_v^2$. Under normality of the random effects:

$$MSE[\tilde{\theta}_i(\hat{\sigma}_v^2)] = MSE[\tilde{\theta}_i(\sigma_v^2)] + E[\tilde{\theta}_i(\hat{\sigma}_v^2) - \tilde{\theta}_i(\sigma_v^2)]^2$$

where the last term is obtained by an approximation because it is generally intractable.

The approximated form of the mean square error is given by:

$$MSE[\tilde{\theta}_i(\hat{\sigma}_v^2)] \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2)$$

where g_{2i} and g_{3i} are of lower order than g_{1i}.

An approximately unbiased estimator of this mean square error is computing as:

$$mse[\tilde{\theta}_i(\hat{\sigma}_v^2)] \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2)$$

if, as in our present application, σ_v^2 is estimated by REML.[29]

11.3 SAE modelling strategy adopted

The approach

In this project we have taken the view that rather than discussing the SAE procedures in general terms independent of the actual data situation, it is more useful to *develop and implement the estimation procedures in concrete terms on the basis of the data sources as actually available to us* in the context of this project. Such a practical approach is much more likely to bring out the variety of situations and problems one may actually encounter in the course of producing regional indicators of poverty and deprivation.

In one important sense, however, our application and discussion of the methodology and approaches lacks realism. This is because the approach is constrained by the data and country-specific knowledge actually available to us in the context of this project. *In national implementation of this and similar methodologies, data and knowledge of specifics of the situation can be expected to be more favourable.* The same can be expected in relation to the time and resources available for the purpose.

Data availability

The type of data available for poverty analysis at the regional level seems to generally preclude the use of unit (household or person) level models. This is because often the covariates are available only at a more aggregate level such as small areas rather than at the unit level. For instance, the situation which we have explored here for the illustrative construction of small area estimates, direct information on poverty-related information at the micro (unit) level comes from intensive and small scale surveys such as the ECHP. These indicators can be aggregated to areas such as NUTS regions where the latter contain some sample units. On the other side, the rich body of correlates which may 'explain' poverty-related characteristics of the areas come from NewCronos. The two sources can be combined to produce composite estimates, provided that (1) the survey data contain information for the identification of the area to which each unit belongs, and (2) the aggregate data on the correlates is available for all the areas in the population of interest.

The structure of the SAE methodology we have adopted is determined by the nature of the data as available to us presently. We believe that many aspects of the data situation and the

[29] For the application of these procedures, we have used the computer programs developed by Nicola Salvati, University of Pisa.

proposed approach to modelling are of wider and lasting interest for the production of regional indicators on poverty and deprivation. Nevertheless, it should be stressed that the *primary concern in this report is to provide detailed and concrete illustrations of the recommended approach, rather than to produce the 'best' or 'final' values for specific regional indicators.*

Table 11.1 shows the data situation. The SAE structure adopted in view of that data situation will be shown in Table 11.3.

Table 11.1 Data availability by NUTS regions

			Country	NUTS1	NUTS2	NUTS3
A	**EU15**					
1	DE	Germany	**a**	**a**	**b**	**b**
2	DK	Denmark	**a**	**-**	**-**	**b**
3	NL	Netherlands	**a**	**b**	**b**	**b**
4	BE	Belgium	**a**	**a**	**b**	**b**
5	LU	Luxembourg	**a***	**-**	**-**	**-**
6	FR	France	**a**	**a**	**b**	**b**
7	UK	United Kingdom	**a**	**a**	**a**	**b**
8	IE	Ireland	**a**	**-**	**b**	**b**
9	IT	Italy	**a**	**a**	**a**	**a**
10	GR	Greece	**a**	**a**	**b**	**b**
11	ES	Spain	**a**	**a**	**b**	**b**
12	PT	Portugal	**a**	**a**	**a**	**b**
13	AT	Austria	**a***	**a**	**b**	**b**
14	FI	Finland	**a***	**b**	**b**	**b**
15	SE	Sweden	**a***	**-**	**b**	**b**
B	**NMS10**					
16	CY	Cyprus	**c**	**-**	**-**	**-**
17	CZ	Czech Republic	**c**	**-**	**c**	**c**
18	EE	Estonia	**c**	**-**	**-**	**c**
19	HU	Hungary	**c**	**c**	**c**	**c**
20	LV	Latvia	**c**	**-**	**-**	**c**
21	LT	Lithuania	**c**	**-**	**-**	**c**
22	MT	Malta	**c**	**-**	**-**	**c**
23	PL	Poland	**b**	**a**	**a**	**b**
24	SI	Slovenia	**c**	**-**	**-**	**c**
25	SK	Slovakia	**c**	**-**	**c**	**c**
C	**Candidate**					
26	BG	Bulgaria	**c**	**c**	**c**	**c**
27	RO	Romania	**b**	**-**	**a**	**b**
28	TR	Turkey#	**c**			

Key:

Country		NUTS2	
a	ECHP 8 waves	-	not applicable (N2=N1=country)
a*	ECHP 5-7 waves	a	N2 code available in survey
b	Similar survey, 1-2 waves	b	N2 code n.a. in survey
c	only some published indicators	c	no survey available
NUTS1		**NUTS3**	
-	not applicable (N1=country)	-	not applicable (N3=N2=N1=country)
a	N1 code available in survey	a	N3 code available in survey
b	N1 code n.a. in survey	b	N3 code n.a. in survey
c	no survey available	c	no survey available

Turkey: no data available in NewCronos

For EU15 countries, we have access to micro data for up to 8 ECHP waves; for Poland and Romania, single waves of similar living conditions or consumption surveys are available. For the remaining EU Member State and Candidate countries, only a very limited set of indicators published by Eurostat are available to the project team.[30] These indicators are at the national level, and all concern only monetary poverty.

For all countries (except Turkey) NewCronos tables are available, though individual items of information may be lacking to varying degrees.

In most surveys, the code to identify NUTS1 regions is present. There are a couple of exceptions, namely the Netherlands and Finland, where no such codes are provided in ECHP Users' Data Base (UDB).

By contrast, in most ECHP survey data, no code is available in UDB to identify NUTS2 or lower level areas - the only exceptions being the data for the UK and Portugal. NUTS2 codes are also present for the available data from the Polish and Romanian surveys. Under a separate research agreement, ISTAT has provided to University of Siena the ECHP Production Data Base (PDB) which includes the necessary area identification codes for Italy to NUTS3 level.

NUTS3 of lower level coding is not available in any of the other survey data sets.

Structure of the modelling

In view of the above data situation, the options we have considered are summarised in Table 11.2.

With a few thousand households observed in a survey such as the ECHP, most estimates at the national level are sufficiently accurate (have small sampling error) to be directly reported.

Below the national level, we have used area-level EBLUB composite estimators throughout in countries where available data permitted that, that is, access to area-coded survey data was available. In these countries (identified in Table 11.3), the following applies.

- At NUTS1 level, the available sample sizes are generally smaller and consequently sampling errors are larger. In some cases, the NUTS1 samples are very small, and significant gains in precision are obtained by using composite estimates. However, overall the introduction of modelling and composite estimation adds only marginally to the precision of the direct estimates from the survey at NUTS1 level, especially when data can be cumulated over time as in the case of ECHP.[31]
- The gains from modelling are naturally more significant at NUTS2 level, and substantially more so at NUTS3 level.
- NUTS3 are not always 'small' areas in terms of population size; their smallness in the SAE methodology refers to the smallness of the *samples* available for direct estimation.

---

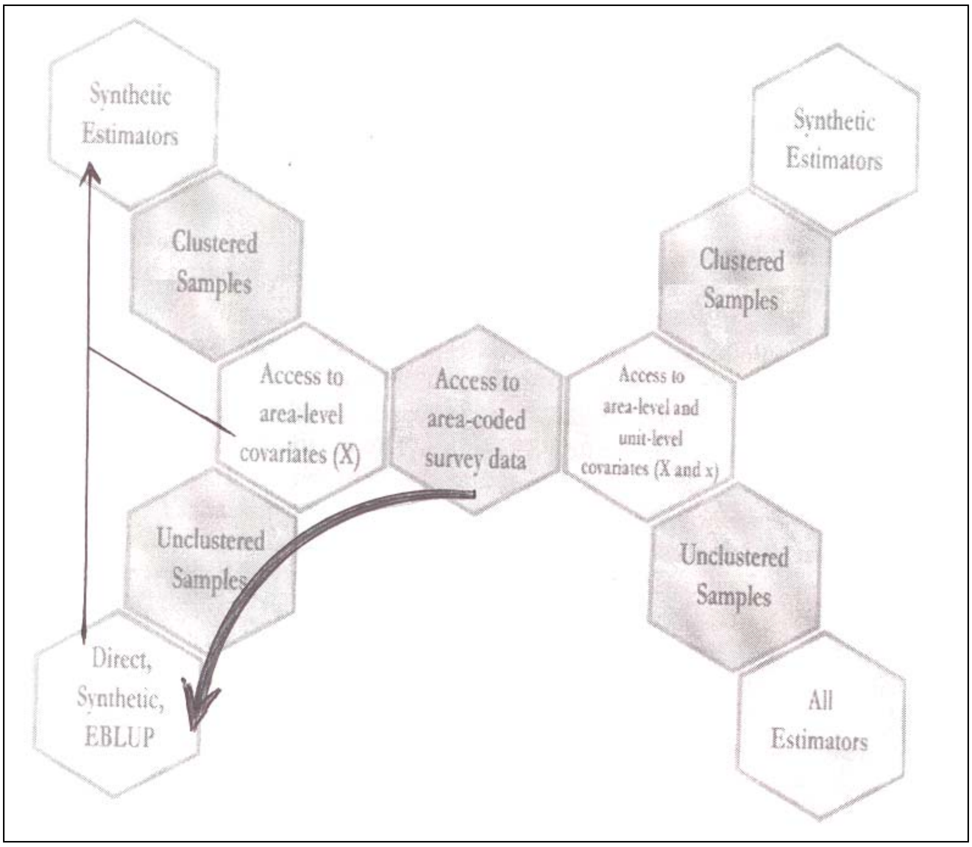
[30] Eurostat (2003), Monetary Poverty in EU Acceding and Candidate Countries, *Statistics in Focus* Theme 3, 21/2003. Also, Eurostat (2004), Monetary Poverty in New Member States and Candidate Countries, *Statistics in Focus* Theme 3, 12/2004. Mostly, these indicators are based on national household budget surveys.

[31] Because of their exceptional heterogeneity, we have divided Italy and Germany into two parts each, for the purpose of country-to-NUTS1 modelling.

Table 11.2 Structure of the modelling

	Data situation	Type of estimator used
	Access to area-coded survey data + Access to area-level covariates +Unclustered samples	Composite (area-level EBLUP)
	Lack of access to area-coded survey data, or access only to country-level survey estimates + Access to area-level covariates	Synthetic (regression-prediction)

The following is a schematic representation of the above in terms of the general structure proposed in Eurarea project report.



Note: By "unclustered sample" is meant a sample where the primary sampling units (PSUs) are confined to be within (or at least coincide with) the areas for which estimates are required. In the present application, the interest is in area units such as NUTS3 regions, and in the ECHP and other samples used, the primary sampling units tended to be much smaller in size than NUTS3 regions. In other words, the sample could be regarded as "unclustered" in the sense intended.

- o With the type of surveys available (HBS, ECHP, or subsequently EU-SILC), the sample sizes are generally likely to be too small to provide useful information for estimation at NUTS4 or NUTS5 level, even after consolidation of the data over a number of years.[32] It is also not possible to go beyond NUTS3 in the type of models developed here using NewCronos tables, since those data are available with up to NUTS3 breakdown at most.

It may be pointed out that, in contrast to the above, Eurarea project simulations considered small area estimates for statistics such as poverty rates going down to NUTS4 and NUTS5 levels. However, this was done on the basis of artificial (simulated) samples selected from census frames onto which information on income had been imputed at the micro level. This provided great flexibility in the size and spread of the samples, something which is not available in real intensive surveys such as ECHP and EU-SILC.

Production of estimates at lower (NUTS4 and NUTS5) levels would require models of a different type, such as those described in Chapter 14. These models are statistically less precise. They involve imputing the required target variables – such as poverty measures – to areas or to individual households in a large data set such as a population census, essentially on the basis of a regression model fitted from a small-scale survey containing common covariates and the required target variables. Such models may of course vary among themselves in the degree of sophistication depending on whether they are *area-level* or *unit-level* models and whether they are *stochastic* or *deterministic*.

The regression-prediction models

In countries where no area-coded survey data are available, we have to resolve to much simpler and cruder modelling. In most cases this situation arose simply because no survey micro data for the country were available to the project team. However, it also arose in cases where the micro data were available but contained no code to identify regions. This applied to a couple of cases in ECHP for NUTS1 identification, but to a majority of the countries in relation to NUTS2 regions, and to all (except Italy) in relation to NUTS3 regions. This is an 'artificial' limitation, and presumably will not be relevant in 'real' applications of our procedures at the national level.

In the absence of area-coded survey data, the procedure we have followed is to use the regression coefficients determined from the corresponding EBLUP model (for the same target variable and the same - NUTS1 or NUTS2 - level), and simply use these coefficients to predict the target variables on the basis of available predictors from NewCronos. This situation is indicated in the second row of Table 11.2, and in the following diagram by the two arrows leading to the box "Synthetic Estimators" in the top left corner.

The results of such modelling depend critically on how good the available predictors are in predicting the target variables. *The illustrative results presented here should be treated with caution, pending the development of better models on the basis of better data.*

---

[32] The potential for cumulating data over time would be generally greater in EU-SILC compared to that in ECHP where the former is based on the 'integrated design' with rotational sampling. Sample rotation provides smaller correlations and hence more efficient cumulation over time (Verma, 2001).

Table 11.3 Regional indicators: small area estimation structure

Domain		Country	NUTS1 : Country	NUTS2 : NUTS1	NUTS3 : NUTS2
9N IT (N)	Italy (N)	survey	**SAE model1****	**SAE model2****	**SAE model3 ****
9S IT (S)	Italy (S)	survey	**SAE model1****		
7 UK	United Kingdom	survey	**SAE model1****	**SAE model2****	#
23 PL	Poland	survey	**SAE model1****	**SAE model2****	#
12 PT	Portugal	survey	**SAE model1****	**SAE model2****	#
27 RO	Romania	survey	-	**SAE model2****	#
1A DE (1)	Germany (1)	survey	**SAE model1****	Regression-Prediction2*	#
1B DE (2)	Germany (2)	survey	**SAE model1****		#
4 BE	Belgium	survey	**SAE model1****	Regression-Prediction2*	#
6 FR	France	survey	**SAE model1****	Regression-Prediction2*	#
10 GR	Greece	survey	**SAE model1****	Regression-Prediction2*	#
11 ES	Spain	survey	**SAE model1****	Regression-Prediction2*	#
13 AT	Austria	survey	**SAE model1****	Regression-Prediction2*	#
3 NL	Netherlands	survey	Regression-Prediction1*	Regression-Prediction2*	#
14 FI	Finland	survey	Regression-Prediction1*	Regression-Prediction2*	#
19 HU	Hungary	published indicators	Regression-Prediction1*	Regression-Prediction2*	#
8 IE	Ireland	survey	-	Regression-Prediction2*	#
15 SE	Sweden	survey	-	Regression-Prediction2*	#
26 BG	Bulgaria	published indicators	-	Regression-Prediction2*	#
17 CZ	Czech Republic	published indicators	-	Regression-Prediction2*	#
25 SK	Slovakia	published indicators	-	Regression-Prediction2*	#
2 DK	Denmark	survey	-	-	#
18 EE	Estonia	published indicators	-	-	#
20 LV	Latvia	published indicators	-	-	#
21 LT	Lithuania	published indicators	-	-	#
22 MT	Malta	published indicators	-	-	#
24 SI	Slovenia	published indicators	-	-	#
5 LU	Luxembourg	survey	-	-	-
16 CY	Cyprus	published indicators	-	-	-
28 TR	Turkey	published indicators	#	#	#

**	Survey + NewCronos
*	NewCronos + Regression coefficients from SAE
-	not applicable
#	not implemented in this illustration
DE(2)=	DE1, DE2, DE3, DE7, DE8, DE9, DEX
IT(S)=	IT8, IT9, ITA, ITB

Note. Because of their exceptional heterogeneity, we have divided Italy and Germany into two parts each for the purpose of country-to-NUTS1 modelling.

11.4 Description of the SAE model used

The ratio approach

Returning to the present application, it should be noted that – compared to the scope of modelling considered in a large project such as Eurarea, for instance – the present application is somewhat simplistic in that it does not attempt to incorporate temporal or spatial autocorrelations. On the other hand, however, *a major positive feature of the present approach is that the modelling strategy is designed to be hierarchical.* We begin with poverty rates and other target variables at the national level, using essentially direct survey estimates without involving any modelling.[33] Modelling at the level of countries is problematic in any case because most pertinent explanatory variables able to distinguish among national patterns are likely to be institutional and historical – variables which are often too complex and almost impossible to quantify. Sometimes countries or national systems are classified into types in an attempt to capture these aspects, such as 'social democratic', 'liberal, 'corporatist', 'residual', and so on[34]. For some purposes categorisation such as the above might be of some use. But generally such schemes are too simplistic to be illuminating. And we suspect that not too infrequently, such "ideal types" are constructed merely to express or promote ideological prejudice.

We can expect 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. Thus 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. In this way the effect of the difficult-to-qualify institutional and historical factors, common to the country and its regions, is abstracted. Similarly, in going from NUTS1 region i to its NUTS2 region j, we express the model variables in the form $R_{ij} = Y_{ij} / Y_i$; and similarly from NUTS2 to NUTS3 in the form $R_{ijk} = Y_{ijk} / Y_{ij}$.

The resulting estimates of the R values can be 'raked' for consistency across levels by ensuring:

$$\Sigma_i W_i.R_i = 1,\ \Sigma_i W_{ij}.R_j = 1,\ \text{etc.},$$

where W_i etc. are the appropriate population weights for the regions, scaled to give $\Sigma_i W_i = 1$.

Occasionally, it may be efficient to specify this type of modelling separately for different parts of a large or exceptionally heterogeneous country, examples being eastern and western parts of Germany, or the northern and southern parts of Italy. The same may apply to metropolitan versus other areas in some countries, such as the UK and France.

The same ideas are extended to the modelling of subpopulations, such as children, old persons, single person households, etc. Consider for instance child poverty rate, say Z_0 at

---

[33] The only exception to using survey estimates directly at the national level is the consolidation we have used to reduce sampling variability by 'benchmarking' the results, as explained earlier, to certain aspects of the pattern averaged over group of countries. See Chapters 4 and 5.

[34] E.g., Berthoud (2004), Patterns of poverty across Europe. The Policy Press.

the national level, Z_i at NUTS1 level, and Z_{ij} at NUTS2 level, with (Y_0, Y_i, Y_{ij}) as the corresponding poverty rates for the total population. Then for NUTS1 regions, we can first model the proportion (Y_i/Y_0) as above, and then using those results model to the ratio:

$$r_i = \frac{(Z_i/Z_0)}{(Y_i/Y_0)} = \frac{(Z_i/Y_i)}{(Z_0/Y_0)}$$

to obtain the required child poverty rate Z_i for region i. Factor r_i indicates how the ratio of child to all-person poverty rate varies across regions. Similarly for NUTS2 regions we can first model the variation in all-person poverty rates (Y_{ij}/Y_i) as before, and then using those results model the ratio:

$$r_{ij} = \frac{(Z_{ij}/Z_i)}{(Y_{ij}/Y_i)} = \frac{(Z_{ij}/Y_{ij})}{(Z_i/Y_i)},$$

and so on.

The target variables

The poverty and deprivation indicators listed in Table 11.4 are taken as the target variables in the SAE model estimated.

Table 11.4 Poverty indicators (Target Variables for SAE models)

1	HCR_C	Head Count Ratio, using country poverty lines (consolidated over computations using 50, 60 and 70% of median equivalised income);
2	HCR_N2	Head Count Ratio, using nuts2 poverty lines (consolidated over computations using 50, 60 and 70% of median equivalised income);
3	LogIncPC	Mean of logarithm of the per capita income;
4	logEqInc	Mean of logarithm of equivalised income;
5	FM_C	Fuzzy monetary poverty rate (scaled to equal HCR_C at EU15 level);
6	FS_C	Fuzzy supplementary (non-monetary) deprivation rate (scaled to equal HCR_C at EU15 level);
7	LAT_C	Latent deprivation rate;
8	MAN_C	Manifest deprivation rate.
9	FSUP-1	Fuzzy supplementary deprivation rate: dimension 1 (basic life-style);
10	FSUP-2	Fuzzy supplementary deprivation rate: dimension 2 (secondary life-style);
11	FSUP-3	Fuzzy supplementary deprivation rate: dimension 3 (housing facilities);
12	FSUP-4	Fuzzy supplementary deprivation rate: dimension 4 (housing deterioration);
13	FSUP-5	Fuzzy supplementary deprivation rate: dimension 5 (environmental problems);

Head count ratios have been computed for two poverty line levels: poverty lines defined with respect to income distribution at the country level, and with respect to income distribution separately within each NUTS2 region. All other poverty or deprivation rates (measures 5-13) have been computed with reference to HCR_C, i.e., using only country-level poverty lines.

Country-specific details on the availability of these variables in EU25 and Candidate countries will be provided in the next chapter (Table 12.2). The main point to note is that for countries other than EU15, Poland and Romania, we have no micro data available and only two of the target variables could be constructed from published data: head count ratio with country poverty line, and median equivalised income.

Models used

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

One such model has been estimated for each target variable at each NUTS level; all countries with area-coded survey data and the particular target variable available are pooled together for the estimation of model parameters at the level concerned.

Such pooling across countries is clearly an over-simplification, and has been introduced here primarily for practical reasons. Nevertheless, the 'ratio approach' described above makes this procedure quite reasonable, we believe. This is because the approach removes the effect of factors common to an area and its components at the next level of the NUTS hierarchy.

Model 3 has been estimated for Italy only, as no NUTS3 codes are available in any other survey.

In countries where no area-coded survey data are available, we have had to resolve to much simpler and cruder regression-prediction models. As noted earlier, this procedure involves using the regression coefficients determined from the corresponding EBLUP model (for the same target variable and the same NUTS level) to predict the target variables on the basis of available predictors.

Model results need to be evaluated with reference to external criteria, as well as internally for consistency. For internal evaluation of the models, the following features should be examined: (a) linearity of the regression; (b) choice of prediction variables; (c) normality of standardised residuals; (d) homogeneity of the variance for standardised residuals; and (e) residual analysis to detect outliers.

On these diagnostic aspects, only preliminary analysis could be done within the framework and resources of the present project. Our aim has been primarily illustrative; some deeper analysis must of course be performed in real life application.

11.5 Methodological note on estimating sampling error for disaggregated direct (survey) estimates

Introduction

In the production of small area estimates (SAE) using a procedure such as EBLUP, a major technical requirement is the production of sampling error estimates for the disaggregated estimates produced directly from the survey. *Direct estimates* refer to the estimates derived from the survey data for the small areas concerned, taking into account the sampling design. *Synthetic estimates* are those derived by fitting an appropriate small area model. A weighted combination of these two types of estimates is then taken to produce the final *composite estimates* (SAEs). The weights in the combination depend on the relative magnitudes of the *design variance* pertaining to the direct estimates, and the *model variance* of the synthetic estimates.

Model variance or error is a measure of the disparity (variability) between the direct survey estimates (assuming those to be based on 100% coverage of population) of the target variables of interest, and the model estimates based on the predictor variables ('regressors'); its primary determinant is how well the model fits the data.

Sampling variance (or its square-root, standard error) is a measure of the variability in the direct estimates as a result of those being based only on a sample of the population. Apart from the design, the primary determinant of the magnitude of the sampling error is sample size; hence this component of error increasingly predominates as we move to small areas and domains.

Sampling error estimates in this context are doubly complex: because the statistics of interest in the study of poverty and deprivation are generally complex, much more so than for instance ordinary proportions, means and ratios; and also because the sample designs on which they are based are complex, involving unequal selection probabilities, stratification, multi-stage selections, aggregation over different samples and times, etc.

Secondly, typically very large number of estimates are required. This may be because of the need to include different types of measures, possibly over different subpopulations, but primarily this arises because of the large number of small domains for which the estimates must be produced.

The third difficulty arises from the fact that the estimates of sampling error are themselves subject to variability, which increases with the degree of disaggregation of results as the sample size is reduced. Results of individual computations – even if computationally possible – cannot be always trusted or directly used; this can apply not only to the estimates of variances but also to the estimated statistics themselves.

Fourthly, samples are not always designed in practice so as to permit rigorous estimation of sampling errors from the sample itself. Approximations are often required in making these estimates.

Finally, there is often a problem of insufficient or incomplete documentation and coding in the micro-data of the structure of the sample so as to permit valid estimates of sampling error taking into account the complex sample structure. See Verma (1993).

Practical approaches and procedures are required to overcome such common difficulties. These involve using approximate procedures and modelling and averaging of individual computations as necessary and appropriate. By 'appropriate' we mean procedures which – while not exact or perfect – nevertheless provide estimates which can be considered sufficiently valid and usable for the *purpose for which they are produced*. In this context it is important to note that the requirement of accuracy of the sampling error estimates for the purpose of SAE is somewhat *less stringent* than, for instance, the situation when such estimates are required for constructing confidence interval and the like for individual statistics produced from the survey. This is because in the context of SAE, the role of sampling error is, in the first instance, only to determine the relative weight of the direct survey estimate in the final composite estimate. Of course error in the final estimates does depend on the sampling error, but increasingly less so as the domain sample size goes down. Approximations in the sampling error estimates can be accepted to the extent the final results from the SAE process are not sensitive to those.

Modelling of sampling errors

A common practical procedure for estimating sampling errors for a set of related statistics is to seek a so-called *generalised variance function* (GVF) which relates the required error of a statistic to some simple and known characteristics of the statistic, such as its value and the sample size. Different functional forms may be required for different types of statistics to produce reasonable approximations of sampling errors; the functional relationships have to be established and validated empirically. There are many well-known examples of the use of such functions in official statistics, for instance US Bureau of the Census (1978). In the specific SAE context, an important example of use of GVFs is National Academy of Sciences (1998), reporting the work of the Panel on Estimates of Poverty for Small Geographic Areas in the United States.

Any GVF implies, implicitly or explicitly, constancy of certain parameters (the population variance, coefficient of variation, the design effect, etc.) determining the magnitude of the sampling error of statistics in the group to which it applies. At least the statistics must be similar, and based on the same or a similar design. *This restriction makes such an approach unsuitable in our multi-country, EU-wide context.* The survey statistics which we must use are based on national samples with different designs and structures, even for standardised surveys such as the ECHP. (With the replacement of ECHP by EU-SILC, this diversity is likely to be substantially greater.) For this reason, a different and more flexible approach is required.

The following describes the procedures we have adopted for calculating standard errors for the poverty and related measures estimated at regional level, going down from the country level to NUTS1, and then to NUTS2, and even to NUTS3 where the necessary information for the purpose is available in the survey. On the basis of experience with analysis of patterns of variation of sampling errors, and taking into account our specific multi-country EU-wide context, the approach we recommend and have used has the following features.

1. The standard error of any statistics is broken down into a number of factors which together account for its magnitude. Each factor represents some aspect(s) of the complexity of the sampling design and the estimation procedure (stratification and clustering, weighting, aggregation over surveys, etc.).

2. There is a considerable body of empirical evidence suggesting that many of these factors act more or less independently of each other, so that the factor effects can be taken as multiplicative (see for instance Verma *et al.* 1980 and 1996). In any case, such a simplifying assumption is usually unavoidable in practice.
3. Each factor depends on parameters corresponding to a number of dimensions, in our specific context from the following set: the statistic or variable concerned (v) and the population (u) over which it is defined; country (c) and its particular domain (i); and for a panel survey such as ECHP, the survey wave (w). Reliable estimates of the factors taking into account all these parameters simultaneously are not possible, or even necessary in practice. On the basis of theoretical and empirical considerations, we simply - and also make the result more robust - by *averaging over dimensions* as appropriate for each parameter. The most obvious and common example is averaging over waves in a panel.
4. Further simplifying assumptions are often required, whether because of lack of sufficient information (such as on aspects of the sampling design), or because the statistic involved is too complex to permit more precise treatment, or simply to make the task manageable.
5. Specifically, we often have to borrow parameters from simpler statistics for use with more complex, less traceable statistics.

These features will be illustrated in the following, stating with the most important and basic statistic – estimated poverty rate or head count ratio (HCR).

Domain sampling error for HCR

For the head count ratio (HCR), we may factorise the standard error estimate (se) into components as follows:

$$se_V = (ser_V).k_V.d_V.s_V.f_V.g_V.r_V$$

Subscript V is general notation for parameters corresponding to various dimensions, such as the statistic or variable concerned (v), the population (u) over which it is defined, country (c) and its particular domain (i), and survey wave (w). Each of the factors are described below in turn.

(1) Simple random sample standard error (ser_V)

The first factor in the equation above stands for standard error which would be obtained in a simple random sample of the same size (n_V), without complexities which the other factors represent. Neglecting minor factors such as the "finite population correction", this factor depends on the sample size in a simple way as follows:

$$ser_V = \left(sd_V / \sqrt{n_V}\right),$$

where sd_V is the standard deviation, a measure of dispersion of the variable in the population, independent of the sample design or size. For a simple proportion p, $sd_V = \sqrt{p.(1-p)}$, which is insensitive to variations in p values over a wide range such as 0.25-0.75, and is well estimated even from samples of small size. The statistic HCR is more complex than a simple proportion, as it is defined in terms of a poverty line which is itself subject to sampling variability. However, empirical results indicate that sd_V defined as above still provides a reasonable approximation for it (Berger and Skinner, 2003; Verma, 2004). In any case, it is reasonable to average the results over waves and even domains within a country so as to obtain more reliable and stable estimates. In other words, for HCR=p and domain sample size n_i, we can approximate its simple random sample standard error as:

$$ser_{HCR} = \sqrt{p.(1-p)/n_i}\ .$$

(2) Effect of sample weights (Kish factor, k_V)

Often variations in sampling rates and hence in the sample weights are determined by reporting requirements and other 'external' considerations largely independent of statistical characteristics of the domains of interest. In this sense the weights may be considered arbitrary or haphazard, the effect of which is to inflate the variance of overall estimates. The important thing is that such unequal weights tend to *affect (inflate) the variance of all estimates for different variables in a rather uniform way*, independently of the structure of the sample except for the weighting itself. Herein lies the practical utility of isolating this effect. It is well approximated by the following simple expression (Kish 1965, p.427, and 1989, p.183):

$$k_i = \sqrt{n_i.\sum w_j^2 \Big/ \left(\sum w_j\right)^2} = \sqrt{1 + cv_i^2(w_j)}$$

where the sum is over the n_i sample cases, and cv_i is the coefficient of variation of individual weights w_j in domain i. Note that the factor has been taken to depend only on domain i. Some variation can be expected to occur over waves because of changes in the panel sample, but these are normally minor and the results can be averaged over waves.

(3) Design factor (d_V)

Design factor (or its square, **design effect**) is a comprehensive summary measure of the effect on sampling error of various complexities in the design. It is the factor by which the actual standard error is different from the error in a simple random sample of the same size. Here this factor represents primarily the effect of stratification and clustering, in so far as the effect of sample weights has already been isolated in terms of (k_i) above. The design effect depends on the structure of the sample as well as the variable being estimated. In the ECHP-UDB data available for the present research, codes for the identification of the sample structure have not been provided generally; consequently, full computation of design effects is not possible at present. However, in Eurostat PAN doc.138 (2000), the information shown in Table 11.5 below is provided on design effects averaged over household income related variables. Note that with the exception of Portugal and Italy, the design effects are quite small, all within the range 1.0-1.2. In Denmark, Luxembourg and the Netherlands, practically simple random samples were used so that d_c=1.0. For Finland, for Sweden (register data), as well as for the survey data from Poland and Romania, we have assumed similar values in the absence of better information at hand. In view of the generally small range within which the design effects vary in the present case, it is sufficient to assume that, within each country, a common design effect value can be used for the set of income poverty and deprivation variables of interest, and that the same value applies across different regions in the country (c).

Table 11.5 Design and Kish factors for income-related variables

	1 DE	2 DK	3 NL	4 BE	5 LU	6 FR	7 UK	8 IE	9 IT	10 GR	11 ES	12 PT	13 AT	14 FI	15 SE	mean
k_i	1,07	1,06	1,10	1,08	1,08	1,03	1,04	1,11	1,13	1,09	1,13	1,36	1,19	(1,10)	(1,10)	1,11
d_c	1,12	1,00	1,00	1,11	1,00	1,12	1,13	1,19	1,86	1,23	1,14	1,67	1,13	(1,00)	(1,00)	1,16
k_i*d_c	1,20	1,06	1,10	1,20	1,08	1,15	1,17	1,32	2,10	1,34	1,29	2,27	1,34	(1,10)	(1,10)	1,29

(..) assumed values; PL & RO: the design effect was assumed as 1,20. The Kish factor in Romania is 1,31.

(4) Subpopulation factor (s_V)

For a subpopulation distributed reasonably uniformly across the population, the sampling error for an estimate over the subpopulation (s) can be related in a simple form to that for an estimate over the total population (c). Examples are HCR for children or old persons, compared to the HCR estimated for the total population. The approximate relationship is

$$s_V = \left(\frac{se_s}{se_c}\right) = \left(\frac{n_c}{n_s}\right)^{1/2} \cdot \left(1 + \left(\frac{n_s}{n_c}\right) \cdot \left(d_c^2 - 1\right)\right)^{1/2}, \; d_c \geq 1,$$

where n_s is the number in the sample from the subpopulation (children, elderly persons, recent school-leavers, etc.), and n_c is number in the sample from the total population. The first factor is the increase in sampling error because of the reduced sample size when we consider only the subpopulation of interest. This is partly balanced by the second factor which gives the reduced design effect. (The design effect is reduced because of reduced cluster size when units belonging to the subpopulation only are considered.)

(5) Reduction due to aggregation over waves (f_V)

Factor f_V reduces the standard error because of consolidation of measures over waves. Of course, we cannot merely add up the sample seizes over waves since ECHP is a panel survey and there is a high positive correlation in the poverty measures among the years, which reduces the gain from cumulation. The correlation can be estimated as follows. Consider two adjacent waves, with proportion poor as p and p', respectively, with the following individual-level overlaps between the two waves:

	Wave w+1		
Wave w	Poor (p'_i=1)	Non-poor (p'_i=0)	total
Poor (p_i=1)	a	b	p=a+b
Non-poor (p_i=0)	c	d	1-p=c+d
total	p'=a+c	1-p'=b+d	1=a+b+c+d

Indicating by p_j and p'_j the {1,0} indicators of poverty of individual j over the two waves, we have, with the sum over all (n) individuals:

$$\text{var}(p_j) = \Sigma(p_j - p)^2 / n = p.(1-p) = v_1, \text{ say;}$$
$$\text{cov}(p_j, p'_j) = \Sigma(p_j - p).(p'_j - p')/n = a - p.p' = c_1, \text{ say.}$$

For data averaged over two adjacent years (and ignoring the difference between p and p'), variance is given by:

$$v_2 = \frac{1}{4}.(v_1 + v_1 + 2.c_1) = \frac{v_1}{2}.\left(1 + \frac{c_1}{v_1}\right).$$

The correlation (c_1/v_1) between two periods is expected to decline as the two become more widely separated. Let (c_i/v_1) be the correlation between two points i waves apart. A simple and reasonable model of the attenuation with increasing i is:

$$(c_i/v_1) = (c_1/v_1)^i.$$

Now in a set of I periods (waves) there are (I-i) pairs exactly i periods apart, i=1 to (I-1). It follows from the above that variance v_I of an average over I periods relates to variance v_1 of the estimate from a single wave as:

$$f_c^2 = \left(\frac{v_I}{v_1}\right) = \frac{1}{I}.\left(1 + 2.\Sigma_{i=1}^{I-1}\left(\frac{I-i}{I}\right).\left(\frac{c_1}{v_1}\right)^i\right), \text{with } \left(\frac{c_1}{v_1}\right) \approx a - p^2,$$

where a is the overall rate of persistent poverty between pairs of adjacent waves (averaged over I-1 pairs), and p is the (cross-sectional) poverty rate averaged over I waves. The ratio of the corresponding standard errors is f_c. Due to averaging over I waves, the effective sample size is increased by $(1/f_c^2)$. We take factor f_c to be country-specific, more or less independent of the particular variable in the set. Table 11.6 shows values of the parameters actually obtained for ECHP data over 8 waves. The last two rows show, respectively, the gain in precision (reduced standard error) over a single wave as a result of cumulation, and the factor by which the effective sample size achieved exceeds the average sample size for a single wave.

Table 11.6 Reduction in standard error resulting from cumulation over waves

	1 DE	2 DK	3 NL	4 BE	5 LU	6 FR	7 UK	8 IE	9 IT	10 GR	11 ES	12 PT	13 AT	14 FI	15 SE
p (average HCR)	0,12	0,11	0,11	0,14	0,12	0,15	0,19	0,19	0,19	0,21	0,19	0,21	0,13	0,10	0,10
a (persistent poverty rate)	0,07	0,06	0,06	0,09	0,09	0,10	0,12	0,13	0,13	0,14	0,12	0,15	0,08	0,06	0,01
I (no of waves)	8	8	8	8	7	8	8	8	8	8	8	8	7	6	5
Gain over single wave f_c	0,59	0,54	0,57	0,59	0,69	0,60	0,60	0,63	0,60	0,60	0,57	0,65	0,61	0,66	0,45
Effective number of waves	2,90	3,39	3,13	2,86	1,86	2,76	2,77	2,49	2,76	2,80	3,02	2,36	2,67	2,29	5,00

Poland and Romania: no cumulation over waves is involved

(6) Reduction from averaging different poverty thresholds (g_V)

One of the important recommendation of this research is that in constructing regional poverty rates and similar statistics from limited sample sizes, some gain in efficiency can be achieved by computing those using different poverty thresholds (such as 50, 60 and 70% of the median income), and then taking an appropriately weighted average of those. It is desirable to take these weights as externally determined constants. The methodology has been explained earlier in Chapter 4.

Consider three poverty line thresholds, with poverty rates p_i, $p_1 < p_2 < p_3$, such the with fixed weights W_i, the final rate is computed as $p = \Sigma_i W_i.p_i$. Its variance is given by:

$$\mathrm{var}(p) = \Sigma_i W_i^2.\mathrm{var}(p_i) + 2.\Sigma_{j<i} W_i W_j.\mathrm{cov}(p_i, p_j).$$

By considering the poverty indicator variables $p_{i,k} = \{0,1\}$ for individuals j in the population, it can be easily seen that the above equation becomes:

$$\mathrm{var}(p) = \Sigma_i W_i^2.p_i.(1-p_i) + 2.\Sigma_{j<i} W_i W_j.p_j.(1-p_i).$$

It is this variance that we compare with the variance of a rate (p_2) computed using a single poverty line such as 60% of the median, as is normally done: $\mathrm{ver}(p_2) = p_2.(1-p_2)$. The ratio:

$$g_V = \left(\mathrm{var}(p)/\mathrm{var}(p_2)\right)^{1/2}$$

gives the required factor by which the standard error is reduced. Table 11.7 gives the actual factors obtained for ECHP data, using appropriately weighted consolidation over three poverty line thresholds, namely 50%, 60% and 70% of the median as explained above. In fact, computations have been performed using different poverty line levels in the sense described earlier, that is by defining the median for populations aggregations to different levels such as NUTS2, NUTS1, Country or EU where possible. The factors are remarkably robust to such changes in the level as seen in the table. (Only country and NUTS2 poverty line results are shown in the table, as they are the most relevant.)

Table 11.7 Reduction in standard error from consolidation over different poverty line thresholds

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	25
Poverty line level	DE	DK	NL	BE	LU	FR	UK	IE	IT	GR	ES	PT	AT	FI	SE	PL
NUTS2	0,93	0,87	0,91	0,90	0,87	0,91	0,94	0,88	0,94	0,96	0,94	0,95	0,89	0,88	0,92	0,95
Country	0,93	0,87	0,91	0,90	0,87	0,91	0,93	0,88	0,95	0,97	0,95	0,96	0,89	0,87	0,92	0,96

This factor equals 1.0 for RO as no consolidation over poverty line thresholds was carried out in this country.

(7) Standard error of ratio of estimates in a hierarchy (r_V)

As noted earlier, it is more efficient to model the small area estimates in a hierarchical manner. In place of estimating the absolute value of any statistic (say e_2), we estimate instead the *ratio* ($r=e_2/e_1$) of the statistic at one level such as NUTS2, to its estimate at the preceding (higher) level such as NUTS1. The objective is to obtain var(r), given var(e_2) obtained as described in the steps above. We have:

$$\mathrm{var}(r) = \mathrm{var}\left(\frac{e_2}{e_1}\right) = \frac{1}{e_1^2}.\left(\mathrm{var}(e_2) + r^2.\mathrm{var}(e_2) - 2.r.\mathrm{cov}(e_1, e_2)\right).$$

The covariance easily evaluated by noting that sample "2" is just a subsample of "1", with the same measurements so that correlation between them is 1.0. It can be shown that with n_2 as the size of the subsample of sample n_1:

$$\mathrm{cov}(e_1, e_2) = [\mathrm{var}(e_1).\mathrm{var}(e_2).(n_2/n_1)]^{1/2} .$$

With the reasonable assumption:

$$\frac{\mathrm{var}(e_2)}{e_2^2} = \frac{\mathrm{var}(e_1)}{e_1^2}.\left(\frac{n_1}{n_2}\right), \text{ that is, } \mathrm{var}(e_2) = r^2.\mathrm{var}(e_1).\left(\frac{n_1}{n_2}\right),$$

we get the simple expression for the required factor (r_V):

$$r_V = \left(\frac{\mathrm{var}(r)}{\mathrm{var}(e_2)}\right)^{1/2} = \left(\frac{1}{e_2}\right).\left(\frac{n_1 - n_2}{n_1}\right)^{1/2},$$

the second factor representing the gain resulting from the fact that sample "2" is simply a subsample of "1".

Standard errors for other statistics

In this application we have considered 13 main poverty measures, listed in Table 11.4 above, for which estimates of standard errors are required at various levels (Country, NUTS1, NUTS2, …). Using the factors for the head count ratios [statistics 1-2] derived above, the corresponding factors for the other statistics have been obtained using the following simplified procedures.

Income related measures [statistics 3-4]

The main differences from the HCR sampling error concern the computation of the standard deviation sd_V, and factor g_V which equals 1.0 since no consolidation over poverty lines is involved. We have assumed all other factors to be the same as those for HCR.

For a variable y such as log-income, standard deviation is computed as:

$$sd_{c,v,w} = \left(\Sigma w_j.(y_j - \bar{y})^2 / \Sigma w_j\right)^{1/2},$$

with $\bar{y} = (\Sigma w_j y_j./\Sigma w_j)$, and w_j as the sample weights. The subscripts have been used in the above to indicate that the expression is specific to country (or region), variable and wave. In order to average values over waves, it is preferable to work with the coefficient of variation $cv_{c,v,w} = sd_{c,v,w}/\bar{y}$, which is scale free and therefore not affected by inflation or unit of measurement. This permits its straightforward averaging over waves:

$$cv_{c,v} = \frac{1}{T}.\sum_{w:1}^{T} cv_{c,v,w} .$$

After that, averaged value of standard deviation can be calculated as:

$$sd_{c,v} = \frac{cv_{c,v}}{T}.\sum_{w:1}^{T} \bar{y}_{c,v,w}$$

Fuzzy measures [statistics 5-13]

We assume that the same structure and parameters as above for sampling error of HCR apply for related fuzzy measures of the degree of poverty and deprivation. Of course, standard deviation is computed with reference to proportion p_v for the variable concerned, which may differ significantly from p for HCR. Fuzzy measures have been computed here with reference to a single poverty threshold (60% of the median income), rather than consolidated over three thresholds as was done in the case of the HCR. Consequently, factor g_c=1. On the other hand, however, we expect fuzzy measures to have smaller variance than conventional HCR based on a dichotomous (yes-no) variable. We have not investigated the magnitude of this effect in the present (ECHP) data, but have simply kept the HCR g_c<1 unchanged to make an allowance for it.

Chapter 12
Small Area Estimation (SAE): Application and Results

APPLICATION

12.1 Target variables and covariates

It will be recalled that three different types of SAE models have been estimated:

- o SAE Model 1: estimated on the ratio NUTS1/Country;
- o SAE Model 2: estimated on the ratio NUTS2/ NUTS ;
- o SAE Model 3: estimated on the ratio NUTS3/ NUTS2 (for Italy only).

One such model has been estimated for each target variable at each level; all countries with area-coded survey data and the particular target variable available are pooled together for the estimation of model parameters at the level concerned.

Simple regression-prediction models have been used in countries or regions where no area-coded survey data are available. One such model corresponds to each SAE model; it uses the regression coefficients determined from the corresponding SAE model to predict the target variables on the basis of available predictors.

In this chapter are shown results obtained by the estimation of the SAE models and results of the regression models.

Table 12.1 lists the 13 target variables for Models 1 and 2. The variables have been described in the previous chapter, and are grouped into three sets: income poverty related measures; overall deprivation measures; and dimension-specific deprivation measures.

Table 12.1 Target variables for SAE models 1 and 2 (based on Table 11.4)

Income poverty related measures		
1	HCR_C	Head Count Ratio – country poverty line
2	HCR_N2	Head Count Ratio – NUTS2 poverty line
3	LogIncPC	Mean log(per capita income)
4	LogEqInc	Mean log(equivalised income)
5	FM_C	Fuzzy monetary poverty rate
Overall deprivation measures		
6	FS_C	Fuzzy supplementary (non-monetary) deprivation rate
7	LAT_C	Latent deprivation rate
8	MAN_C	Manifest deprivation rate
Dimension-specific deprivation measures		
9	FSUP-1	Deprivation rate: dimension 1 (basic life-style);
10	FSUP-2	Deprivation rate: dimension 2 (secondary life-style);
11	FSUP-3	Deprivation rate: dimension 3 (housing facilities);
12	FSUP-4	Deprivation rate: dimension 4 (housing deterioration);
13	FSUP-5	Deprivation rate: dimension 5 (environmental problems);

As explained in Chapters 5 and 6, measures FM_C and FS_C are linked to HCR_C based on the country poverty lines. The remaining measures (7-13) are derived from these as described in Charter 6.

The availability of these variables in EU25 and Candidate countries is presented in Table 12.2. Sufficient information is not available in the ECHP surveys in Germany, Luxembourg and Sweden to construct deprivation measures in specific dimensions (variables Fuzzy Supplementary 1-5). Only monetary measures could be computed from the survey in Romania. It should also be noted that some of the non-monetary measures for Poland lack comparability with corresponding ECHP measures because of differences in the survey questions used.

Table 12.2 The availability of the target variables in EU25 and Candidate countries

		target variable	1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSUP-1	10 FSUP-2	11 FSUP-3	12 FSUP-4	13 FSUP-5
1	DE	Germany	X	X	X	X	X	X	X	X					
2	DK	Denmark	X	X	X	X	X	X	X	X	X	X	X	X	X
3	NL	Netherlands	X	X	X	X	X	X	X	X	X	X	X	X	X
4	BE	Belgium	X	X	X	X	X	X	X	X	X	X	X	X	X
5	LU	Luxembourg	X	X	X	X	X	X	X	X					
6	FR	France	X	X	X	X	X	X	X	X	X	X	X	X	X
7	UK	United Kingdom	X	X	X	X	X	X	X	X	X	X	X	X	X
8	IE	Ireland	X	X	X	X	X	X	X	X	X	X	X	X	X
9	IT	Italy	X	X	X	X	X	X	X	X	X	X	X	X	X
10	GR	Greece	X	X	X	X	X	X	X	X	X	X	X	X	X
11	ES	Spain	X	X	X	X	X	X	X	X	X	X	X	X	X
12	PT	Portugal	X	X	X	X	X	X	X	X	X	X	X	X	X
13	AT	Austria	X	X	X	X	X	X	X	X	X	X	X	X	X
14	FI	Finland	X	X	X	X	X	X	X	X	X	X	X	X	X
15	SE	Sweden	X	X	X	X	X	X	X	X					
23	PL	Poland	X	X	X	X	X	X	X	X	X	X	X	X	X
27	RO	Romania #	X	X	X	X	X								
16	CY	Cyprus	*		*										
17	CZ	Czech Republic	*		*										
18	EE	Estonia	*		*										
19	HU	Hungary	*		*										
20	LT	Latvia	*		*										
21	LV	Lithuania	*		*										
22	MT	Malta	*		*										
24	SI	Slovenia	*		*										
25	SK	Slovakia	*		*										
26	BG	Bulgaria	*		*										
28	TR	Turkey	*		*										

\# Romania: Consumption instead of income variables
* Only published indicators available on HCR (poverty line 60% of national median), and national median income, mostly based on HBS.

For countries other than EU15, Poland and Romania, we have no micro data available and only two of the target variables could be constructed from published data (in Eurostat publications Statistics in Focus and also recorded in NewCronos): head count ratio with country poverty line, and log equivalised income.[35]

Tables 12.3 list the 8 independent variables (covariates) used in models 1 and 2. These are all obtained from the tabulations provided in NewCronos.

Table 12.4 gives a picture of the availability of these variables in individual countries.

[35] An approximation is involved in the construction of the second variable, log equivalised income. We have found only median income level in the published data. These were converted to mean values using the average mean:median ratio from ECHP surveys in EU15 (=1.15), and then taking the log of the mean so estimated. Hence we do not have mean of logarithm of income (as in other countries, estimated from micro data), but logarithm of mean income.

Table 12.3 Covariates Variables available at Nuts1 and Nuts2 level

1	Disposable income	PPS per capita 2000, net
2	Net/Gross	Ratio (Net/Gross) Income 2000; constructed from the secondary distribution of income account of households: current taxes on income, wealth etc. (uses), and disposable income net (uses)
3	Activity rate	mean of activity rates for 1999 and 2000; from domain Regional Labour Market
4	Unemp rate	Unemployment rate 2000; from Regional Unemployment: LFS adjusted series
5	Long-term unemp	Long-term unemployment rate 2000 (unemployed for 12 months or longer; from Regional Unemployment: LFS adjusted series
6	% in manufacturing	Percentage of workers in manufacturing 2000; constructed from structural business statistics.
7	IMR	Infant mortality rate 2000; constructed from demographic statistics.
8	HH Size	Household size 2000; constructed using the number of households and the total population

Table 12.4 The availability of covariates in EU25 and Candidate countries

			1 Disposable income	2 Net/Gross	3 Activity rate	4 Unemp rate	5 Long term unemp	6 % in manifacturing	7 IMR	8 HH size
1	DE	Germany	X	X	X	X	X	X	X	X
2	DK	Denmark	X	X	X	X	X	X	X	
3	NL	Netherlands	X	X	X	X	X		X	X
4	BE	Belgium	X	X	X	X	X	X	X	X
5	LU	Luxembourg			X	X	X	X	X	X
6	FR	France	X	X	X	X	X	X	X	X
7	UK	United Kingdom	X	X	X	X	X	X	X	X
8	IE	Ireland	X	X	X	X	X	X	X	
9	IT	Italy	X	X	X	X	X	X	X	X
10	GR	Greece	X	X	X	X	X	X	X	X
11	ES	Spain	X	X	X	X	X	X	X	X
12	PT	Portugal	X	X	X	X	X	X	X	X
13	AT	Austria	X	X	X	X	X	X	X	X
14	FI	Finland	X	X	X	X	X	X	X	
15	SE	Sweden	X	X	X	X	X	X	X	
23	PL	Poland	X	X	X	X	X	X	X	
27	RO	Romania	X	X	X	X	X	X	X	X
16	CY	Cyprus	-	-	-	-	-	-	-	-
17	CZ	Czech Republic	X	X	X	X	X		X	X
18	EE	Estonia	-	-	-	-	-	-	-	-
19	HU	Hungary	X	X	X	X	X	X	X	X
20	LT	Latvia	-	-	-	-	-	-	-	-
21	LV	Lithuania	-	-	-	-	-	-	-	-
22	MT	Malta	-	-	-	-	-	-	-	-
24	SI	Slovenia	-	-	-	-	-	-	-	-
25	SK	Slovakia	X	X	X	X	X	X	X	X
26	BG	Bulgaria		X	X	X	X	X	X	X
28	TR	Turkey								

- not applicable (covariates not required as no regions to be modelled in the country)

blank = missing in NewCronos (March 2005)

With regards to the choice of the independent variables (covariates) for building the models, they were selected if the required data were available for most of the countries involved in the estimation models and in the regression models. Substantive considerations were also involved in the selection of the covariates used. We decided to estimate the models considering the full set of covariates available and selected, including some which were statistically non-significant.

12.2 SAE Model 1

This concerns EBLUP models for going from country to NUTS1 level, utilising in combination survey data and the information compiled in NewCronos.

We expect that *the model as implemented here could be improved by fitting it separately for different groups of countries, or introducing the country effect through dummy variables* added to the list of covariates in Table 12.3.

As explained in Section 11.4, we used the 'ratio approach' to improve the precision of the models. Under this approach, the model input consists of

(a) NUTS1-to-Country ratio for the statistic concerned, as directly estimated from the survey

(b) standard error of this ratio estimator

The output from the model consists of

(a)* model estimate of NUTS1-to-Country ratio for the statistic concerned

(b)* mean-squared error of this estimate

Performance measures

Table 12.5 shows some 'performance measures' of SAE Model 1. For each model (i.e., target) variable, three measures are shown:

i) the model parameter gamma (γ). It is the ratio between the model variance and the total variance, and is the share of the weight given to the direct survey estimate in the final composite estimate.

ii) ratio (a)*/(a), i.e., the ratio between the EBLUP estimated value of (a)* and the corresponding direct estimation (a). This is to check the extent to which the modelling changes the input direct estimates.

iii) ratio (b)*/(b), i.e., the ratio between mean-squared error (MSE) of the EBLUP estimate of the NUTS1: Country ratio, and MSE of direct survey estimate of this ratio. This is to check the extent to which the modelling has improved precision of the estimates.

For each of the above, the following summary statistics are given: the mean value over all NUTS1 areas in the model; the coefficient of variation of those values; and the minimum and maximum values.

Overall the results are as expected: the SAE Model1 for NUTS1 level does not provide much gain, as can be seen from the mean ratio of mean-squared errors . This is because the sample sizes for most NUTS1 areas are actually quite large; NUTS1 can hardly be called 'small areas'. The large sample sizes are achieved by cumulation of data over survey waves.

The largest gains in efficiency are for Manifest Deprivation Rate and HCR with NUTS2 poverty line. It is particular noteworthy for HCR_N2, where the MSE is reduced to two-thirds, which implies more than doubling the effective sample size. Since this variable is based on income distributions within each NUTS2 region (even though the modelling being discussed is from country to NUTS1 level of aggregation), it is possible that the sampling errors of the direct estimates are larger. The main reason for the better performance of the model, however, must be a stronger relationship of HCR_N2 with the predictor variables used, compared to the same relationship for HCR_C. This is an important observation because of the substantive importance, as noted earlier, of HCR_N2 as a regional indicator of poverty.

Table 12.5 Performance measures for SAE Model 1

(gamma value, ratio of EBLUP estimates to direct estimates, ratio of EBLUP standard error to direct standard error)

	Gamma				Estimate EBLUP/direct estimate				Mean-squared error (MSE) MSE(EBLUP)/MSE(direct estimate			
	mean	CV	min	max	mean	CV	min	max	mean	CV	min	max
1 HCR_C	0,86	0,15	0,41	0,99	0,99	0,10	0,70	1,49	0,95	0,19	0,35	1,90
2 HCR_N2	0,35	0,47	0,03	0,73	1,00	0,05	0,84	1,14	0,67	0,23	0,23	0,93
3 logEqInc	0,95	0,05	0,71	0,99	1,00	0,00	1,00	1,00	0,98	0,02	0,89	1,00
4 logIncPC	0,95	0,05	0,71	0,99	1,00	0,00	1,00	1,00	0,98	0,02	0,89	1,00
5 FM_C	0,83	0,16	0,35	0,98	0,99	0,05	0,72	1,05	0,92	0,07	0,68	0,99
6 FS_C	0,83	0,16	0,39	0,98	1,00	0,05	0,84	1,28	0,93	0,07	0,70	0,99
7 Latent	0,86	0,14	0,38	0,98	1,00	0,03	0,81	1,11	0,94	0,06	0,70	0,99
8 Manifest	0,66	0,36	0,15	0,96	0,98	0,12	0,60	1,39	0,83	0,18	0,43	0,99
9 Fsup_1	0,93	0,05	0,74	0,99	1,00	0,02	0,96	1,03	0,97	0,02	0,89	1,00
10 Fsup_2	0,86	0,10	0,65	0,98	1,00	0,03	0,89	1,11	0,94	0,04	0,84	0,99
11 Fsup_3	0,70	0,32	0,08	0,98	0,99	0,17	0,36	1,32	0,86	0,16	0,29	1,00
12 Fsup_4	0,88	0,09	0,65	0,98	1,00	0,02	0,94	1,06	0,96	0,04	0,84	0,99
13 Fsup_5	0,88	0,07	0,73	0,98	1,00	0,02	0,96	1,05	0,96	0,03	0,89	0,99

Parameter estimation and significance level

Table 12.6 shows the estimated regression coefficients and the associated significance levels.

In relation to *income poverty related measures* as the target variables, the most significant covariates tend to be disposable income, infant mortality rate and unemployment rate. It is interesting to note that infant mortality rate is significant even if the countries involved in the model are all developed countries. The sign of the coefficient of unemployment rate on the income poverty measure needs to be noticed: it is negative for HCR_C and positive for disposable income per capita. This may be considered somewhat unexpected. However, it is possible that there are confounding effects: for instance in Germany and France there are higher unemployment rates, but at the same time these are richer countries; by contrast, Portugal has a lower unemployment rate but is a poorer country. As noted, perhaps the model could be improved by fitting it separately for different groups of countries, or introducing the country effect through dummy variables. It should also be remembered that the associations being discussed are at the area level, not at the level of individual persons or households.

Concerning *overall deprivation measures*, the Latent and Manifest indicators have more or less the same significant variables: disposable income, activity rate, infant mortality rate and household size. The sign of the parameters seems to be reasonable. With regards to the general fuzzy supplementary indicator (FS_C) the significant variables of the models are more or less the same as those mentioned above.

The models having as dependent variables the measures of *dimension-specific deprivation* are less homogeneous: for each dimension there are different subsets of significant variables. Disposable income is a significant variable in the models relating to the lack of desired possession (FSUP-2), the lack of housing facilities (FSUP-3) and the housing deterioration (FSUP-4). All these dimensions are related to lacking something related to living in the household. Variables activity rate and long-term unemployment are significant only for the model regarding the environmental problem (FSUP-5); this seems reasonable given that the models considers correlations at area level, and it is plausible that the area (NUTS1) activity rate has a significant/positive correlation with the general situation in the area. Infant mortality rate is highly significant for the dimension FSUP-1 and FSUP-5 that concern, respectively, basic non-monetary deprivation and environmental problems. It should be interesting to analyse the association between these two variables more deeply to identify the direction of causality.

Table 12.6 SAE Model 1: Parameter estimation and significance level

	Income poverty related measures	**HCR_C**	**HCR_N2**	**LogEqInc**	**LogIncPC**	**FM_C**
0	Intercept	-0,415	3,082 ***	1,143 ***	1,184 ***	-0,137
1	Disposable income	-1,810 ***	-0,132	0,077 ***	0,090 ***	-1,682 ***
2	Net/Gross	2,557	-2,116 ***	-0,199 ***	-0,213 ***	2,392
3	Activity rate	1,237	0,149	-0,048	-0,066	1,055
4	Unemp rate	-0,351 ***	0,040	0,017 ***	0,021 ***	-0,321 ***
5	Long-term unemp	0,155	0,006	-0,004	-0,007	0,151
6	% in manufacturing	-0,203 **	-0,056	0,008 *	0,009 *	-0,168 *
7	IMR	0,557 ***	0,092	-0,016 **	-0,018 **	0,428 ***
8	HH size	-0,708	-0,066	0,021	-0,001	-0,702

	Overall deprivation measures	**FS_C**	**LAT_C**	**MAN_C**
0	Intercept	-1,191	-0,665	-0,753
1	Disposable income	-0,925 **	-1,061 ***	-2,186 ***
2	Net/Gross	0,876	1,574	1,993
3	Activity rate	3,161 ***	1,872 ***	2,964 ***
4	Unemp rate	0,017	-0,122	-0,276 **
5	Long-term unemp	0,386 *	0,256	0,367
6	% in manufacturing	-0,133	-0,150 **	-0,158
7	IMR	0,452 ***	0,345 ***	0,774 ***
8	HH size	-1,639 ***	-1,038 ***	-1,731 **

	Dimension-specific measures	**FSUP-1**	**FSUP-2**	**FSUP-3**	**FSUP-4**	**FSUP-5**
0	Intercept	4,133	2,013	4,390	3,979	-1,438
1	Disposable income	-0,748	-1,063 **	-2,077 *	-1,719 ***	0,499
2	Net/Gross	-0,369	0,124	-2,025	-1,109	-0,959
3	Activity rate	-1,115	0,730	2,536	1,391	3,028 ***
4	Unemp rate	-0,340 **	-0,142	-0,248	-0,235 *	0,381 ***
5	Long-term unemp	-0,110	0,063	0,116	0,244	0,491 **
6	% in manufacturing	-0,410 ***	-0,191 *	-0,490 **	-0,314 ***	0,079
7	IMR	0,760 ***	0,311 *	0,289	-0,011	0,380 **
8	HH size	-0,795	-0,837	-1,536	-1,209 *	-1,454 ***

(significance levels: *** 99%;**95%;*90%)

12.3 SAE Model 2

This concerns EBLUP models for going from NUTS2 to NUTS1 level, again utilising in combination survey data and the information compiled in NewCronos.

The list of target variables and covariates is the same as that for SAE Model 1 described already. However, there are differences in the extent to which the variables are available and the nature of the countries involved in the same model.

Performance measures

Table 12.7 shows some 'performance measures' of SAE Model 2. For each model (target variable), three measures are shown as in Table 12.5.

The performance of the model in terms of gain in efficiency is obviously better for Model 2 (NUTS2 level) compared to Model 1 (NUTS1 level). This is because here the sample sizes available for direct estimates are smaller. The highest gains, of 20-25%, are for Latent, Manifest and FSUP-1 deprivation measures. Again, as with Model1, the gain for HCR_N2 is almost twice as large as that for HCR_C. This is important in the context of constructing regional indicators. The gain for HCR_C, FSUP-2, FSUP-4 and FSUP-5 is around 10%, while no prediction is possible for FSUP-3 for lack of adequate data. For logarithm of equivalised income and the logarithm of the per capita income, the relative gains are the smallest among the variables.

Table 12.7 Performance measures for the SAE Model 2

(gamma value, ratio of EBLUP estimates to direct estimates, ratio of EBLUP standard error to direct standard error)

		Gamma				Estimate EBLUP/direct estimate				Standard error (SE) SE(EBLUP)/SE(direct estimate)			
		mean	CV	min	max	mean	CV	min	max	mean	CV	min	max
1	HCR_C	0,80	0,22	0,45	0,98	1,01	0,08	0,86	1,34	0,90	0,11	0,71	1,00
2	HCR_N2	0,66	0,38	0,19	0,95	1,01	0,07	0,83	1,30	0,82	0,22	0,47	1,00
3	logEqInc	0,81	0,23	0,44	0,98	1,00	0,00	1,00	1,01	0,94	0,18	0,68	1,35
4	logIncPC	0,85	0,14	0,65	0,99	1,00	0,00	0,99	1,01	0,92	0,12	0,74	1,21
5	FM_C	0,75	0,27	0,40	0,98	1,02	0,14	0,80	1,63	0,88	0,12	0,66	1,02
6	FS_C	0,68	0,32	0,32	0,97	1,02	0,09	0,85	1,45	0,85	0,14	0,63	0,99
7	Latent	0,61	0,36	0,23	0,96	1,01	0,08	0,84	1,41	0,81	0,16	0,50	0,98
8	Manifest	0,55	0,49	0,12	0,97	1,06	0,25	0,71	2,25	0,76	0,24	0,36	1,00
9	Fsup_1	0,60	0,41	0,22	0,97	1,01	0,08	0,86	1,28	0,80	0,18	0,54	1,00
10	Fsup_2	0,73	0,22	0,47	0,97	1,01	0,07	0,87	1,26	0,88	0,09	0,70	0,99
11	Fsup_3												
12	Fsup_4	0,77	0,15	0,51	0,97	1,01	0,05	0,88	1,24	0,90	0,06	0,76	0,99
13	Fsup_5	0,76	0,22	0,49	0,98	1,00	0,05	0,87	1,11	0,89	0,10	0,72	1,01

Parameter estimation and significance level

Table 12.8 shows the estimated regression coefficients and the associated significance levels.

It is worth emphasising that for the *income poverty related measures*, the models include Italy, UK, Portugal, as well as Poland and Romania. The set of countries involved in the model is quite heterogeneous. For this set of measures, the covariates significant in four out of five models (the only exception is HCR_N2) are the infant mortality rate and the percentage of people employed in manufactory. The infant mortality rate, as in Models 1, seems to be strongly related with poverty. Regarding the significant effect of the percentage of people employed in manufactory, it is interesting to observe that this variable was significant also in Model 1, but in Model 2 the level of significance is higher.

The models having as target variables the *non-monetary deprivation measures* do not provide very good performance, in terms of significance of the covariates. No doubt a better set of covariates are required than could be extracted only from the data compiled in NewCronos. It is important to take note of the fact that no covariates are found to be significant for the model regarding the variable FSUP-3 (lacking of housing facilities). Hence for this variable it was not possible to predict any value. For the other models, unemployment rate it always a significant variable.

The models for latent and manifest deprivation measures show a similar behaviour: the same set of covariates are significant, and the sign of the coefficients seem reasonable.

Comparing only the covariates which are significant in both Model 1 and Model 2, we can see that the estimated parameters (regression coefficients) are generally consistent in sign and magnitude – except for the covariate Unemployment Rate for which the results tend to be rather variable.

Table 12.8 SAE Model 2: Parameter estimation and significance level

	Income poverty related measures	HCR_C	HCR_N2	logEqInc	LogIncPC	FM_C
0	Intercept	0,512	-1,201	1,014 ***	0,945 ***	1,192
1	Disposable income	-0,410	0,111	0,017	0,031 **	-0,612 **
2	Net/Gross	0,419	0,915	-0,060	-0,013	0,290
3	Activity rate	0,029	0,863 *	0,037	0,049	-0,045
4	Unemp rate	0,057	0,148 *	0,008	0,014 **	0,029
5	Long-term unemp	0,018	-0,056	-0,005	-0,009	0,067
6	% in manufacturing	-0,292 **	-0,017	0,010 **	0,009 *	-0,271 **
7	IMR	0,484 **	0,161	-0,021 ***	-0,028 ***	0,366 **
8	HH size	0,195	0,070	-0,001	0,000	0,004
	Overall deprivation measures	**FS_C**	**LAT_C**	**MAN_C**		
0	Intercept	2,019	1,340	0,477		
1	Disposable income	-0,642	-0,635 **	-1,011 *		
2	Net/Gross	-1,550	-0,575	-0,519		
3	Activity rate	0,750	0,589	1,535		
4	Unemp rate	0,371 ***	0,228 **	0,281 *		
5	Long-term unemp	-0,001	0,035	0,165		
6	% in manufacturing	-0,173	-0,228 **	-0,365 **		
7	IMR	0,238	0,230 *	0,460 **		
8	HH size	-0,018	0,018	-0,012		
	Dimension-specific measures	**FSUP-1**	**FSUP-2**	**FSUP-3**	**FSUP-4**	**FSUP-5**
0	Intercept	1,212	-0,354	-4,386	-3,666	3,091
1	Disposable income	-0,816 **	-0,121	0,336	0,552	-0,716 *
2	Net/Gross	-1,110	-0,404	3,750	2,289	-1,963
3	Activity rate	1,569 **	1,174	1,172	1,659 *	0,340
4	Unemp rate	0,201 *	0,270 **	0,136	0,390 **	0,293 **
5	Long-term unemp	0,008	0,163	-0,126	-0,075	0,107
6	% in manufacturing	-0,242 **	-0,078	-0,164	-0,166	-0,009
7	IMR	0,180	0,318 *	0,144	0,003	0,041
8	HH size	0,000	0,032	0,177	0,022	-0,205 *

(significance levels: *** 99%;**95%;*90%)

12.4 SAE Model 3

This concerns EBLUP models for going from NUTS2 to NUTS3 level. SAE Models 3 are estimated for Italy (only this database makes possible the access to area-coded survey at NUTS3 level). Given the high level of disaggregation it has been decided to consider only three poverty indicators (consequently three models): the HCR_C, the HCR_N2, logEqInc. The list of the independent variables available is also more limited; it is confined to the relevant covariates tables for which are provided in NewCronos at NUTS3 level.

The available set covariates is very limited indeed. It would be important to find additional and better covariates in real-life replications of SAE Model 3.

Performance measures

Table 12.9 shows some 'performance measures' of SAE Model 3. For each model (target variable), three measures are shown as in Table 12.7.

In this case we really have small areas with very small sample sizes. The average gain in precision is at least 20%, and it is quite consistent across the target variables. It is interesting to note the minimum value of the ratio between the EBLUP standard error and the direct standard error: the minimum values in all the three models is less the 0.10. This means that in some areas the EBLUP estimator provides a gain in efficiency, compared to the direct survey estimates, that is higher than 90%.

Table 12.9 Performance measurement for the SAE Model 3

(gamma value, ratio of EBLUP estimates to direct estimates, ratio of EBLUP standard error to direct standard error)

	Gamma				Estimate EBLUP/direct estimate				Standard error (SE) SE(EBLUP)/SE(direct estimate)			
	mean	CV	min	max	mean	CV	min	max	mean	CV	min	max
HCR_C	0.70	0.41	0.01	1.00	1.05	0.27	0.44	2.53	0.81	0.30	0.10	1.00
H_N2	0.76	0.36	0.00	1.00	1.03	0.20	0.46	2.21	0.85	0.27	0.08	1.00
logEqInc	0.62	0.44	0.00	1.00	1.00	0.01	0.96	1.05	0.77	0.32	0.05	0.98

Parameter estimation and significance level

Table 12.10 shows the estimated regression coefficients and the associated significance levels.

The models for the three target variables show a similar behaviour: for all of them the highest significance level covariate is gross domestic product (GDP), which is obviously related to income level and poverty measures. The model regarding the more purely relative poverty measure HCR_N2 shows that the significant variable is the unemployment rate.

Table 12.10 SAE Model 3: Parameter estimation and significance level

Independent variables	HCR_C	HCR_N2	logEqInc
Intercept	0.758	0.865	0.935 ***
Unemployment rates 2000	0.288	0.317 *	-0.008
Crude death rate 2000	-0.172	-0.253	0.024
Crude birth rate 2000	0.856	1.228 *	-0.027
Gross Domestic Product 2000	-1.708 ***	-1.726 ***	0.085 ***
Population Density 2000	-0.065	-0.042	0.000
Activity Rates 2000	1.081	0.644	-0.009

significance levels: *** 99%;**95%;*90%

SOME RESULTS

The numerical results obtained from the application of the SAE models described above, and of the regression-predictions for areas where that was not possible, are listed fully in the Statistical Annex at the end of this report. Below a specimen of the results are shown with brief remarks.[36]

12.5 Income poverty

Figure 12.1 Head Count Ratio NUTS2 regions (country poverty lines)

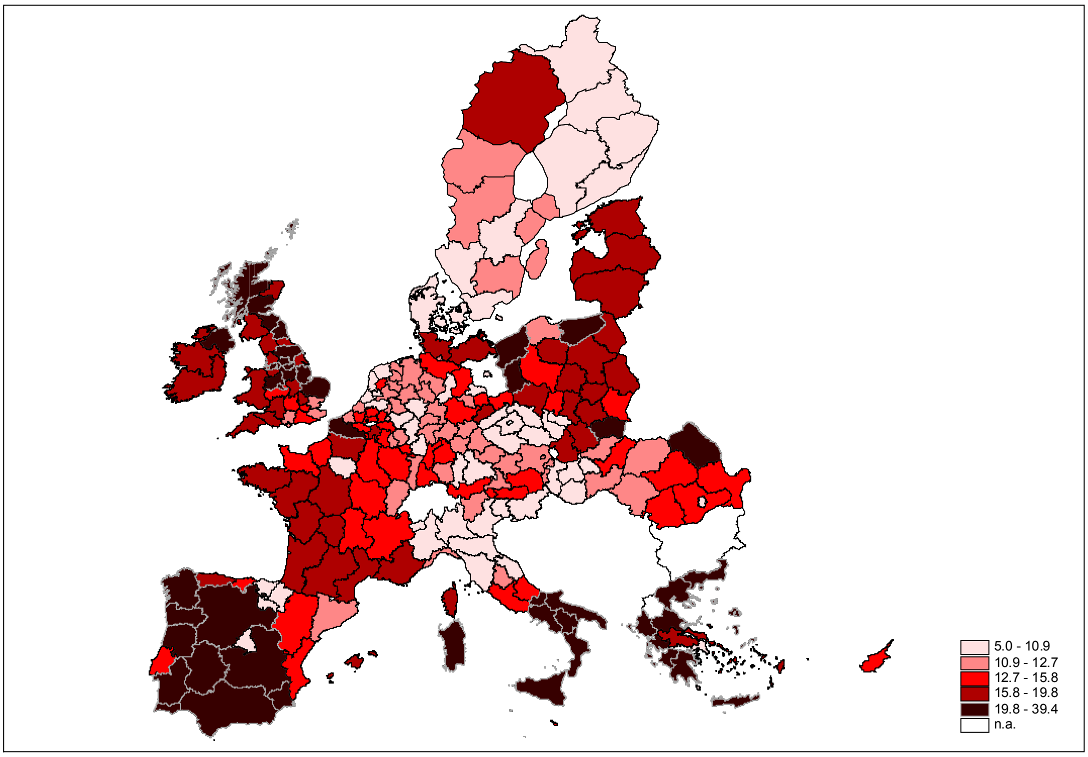


Figure 12.1 shows the concentration of the poorest (bottom quintile) areas to be in Portugal, Spain, Greece and Southern Italy. The highest estimated poverty rate using country poverty lines is in Sicilia (ITA0 with 39.4%), and the next highest in Calabria (IT93 with 39.2%).

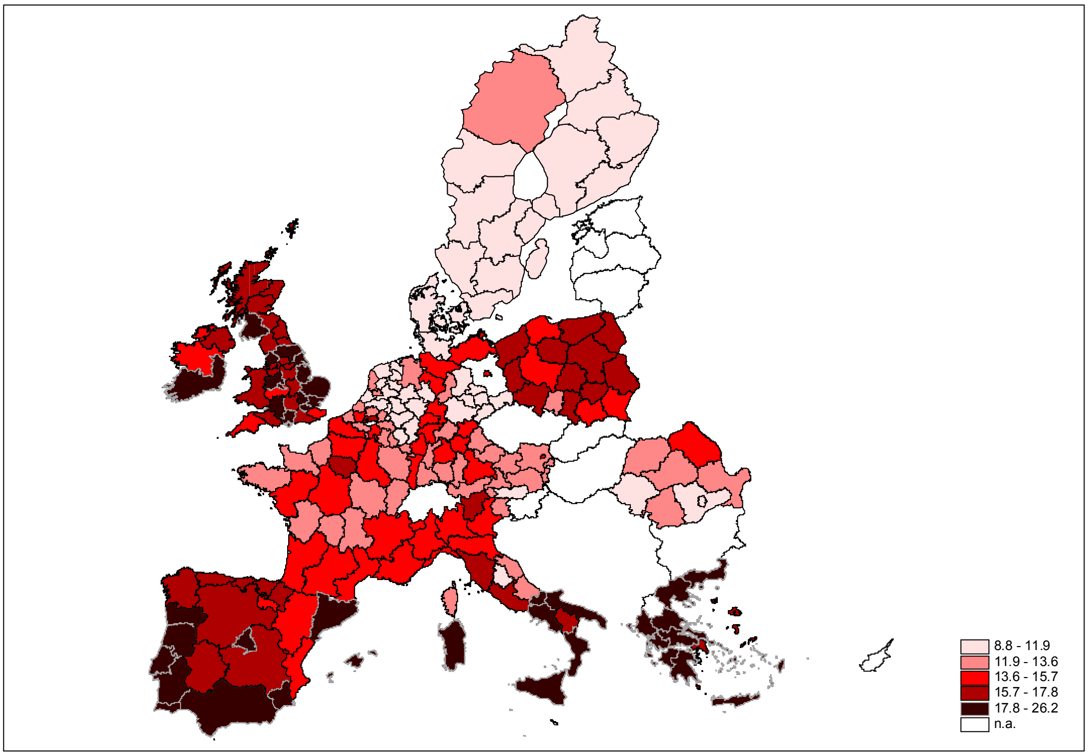
In UK as well, the proportion in poverty is also quite high in many areas outside the south.

In Italy it is interesting to note the striking difference between the south and the north. For this reason it will be very interesting to analyse the situation at NUTS3 level (see Chapter 13).

---

[36] Note: In the following maps results for Brandenburg, Prov. Brabant Wallon, Praha, Ciudad Autónoma de Ceuta and Melilla, Bratislavský and Dytiki Makedonia have not reported because of the large number of missing values in the covariates.

Figure 12.2 Head Count Ratio NUTS2 regions (NUTS2 poverty lines)



Observing the map in Figure 12.2 we can see less homogenous situation in terms of head count ratio with NUTS2 poverty lines, compared with the previous map using country poverty lines. It is the effect of the definition of the poverty line. Defining the poverty line at NUTS2 level makes the poverty measure more truly relative. It is interesting to note that some areas that shows the worst situation in Figure 12.1 do no belong to the last category in Figure 12.2 (for example Basilicata in Italy). On the other hand, some areas that belong to the best class in Figure 12.1, move to the middle bracket - the red one - in Figure 12.2, such as Toscana, Emilia Romagna and Lombardia in Italy. The same applies also to some regions of Spain.

Countries where regional differences in levels of income are small tend to present similar pictures irrespective of whether country or NUTS2 poverty line is used.

It is interesting to compare Figures 12.3 and 12.4 below for the United Kingdom: the former relates to the head count ratio, computed using a poverty line at country level, the latter relates to the same poverty measure computed using a NUTS2 poverty line. The difference between the indicators HCR_C and HCR_N2 can appreciated clearly by observing area Borders-Central-Fife-Lothian-Tayside (UKA1) for instance: it shows a very high poverty rate with reference to the country poverty line, but a low poverty rate with reference to the NUTS2 poverty line. (This should mean that the area is really poor in terms of level of income, but within it, there is less inequality.) The opposite situation is seen in area Greater London (UK55).

Figure 12.3 Head Count Ratio (country poverty lines), United Kingdom, NUTS2 regions

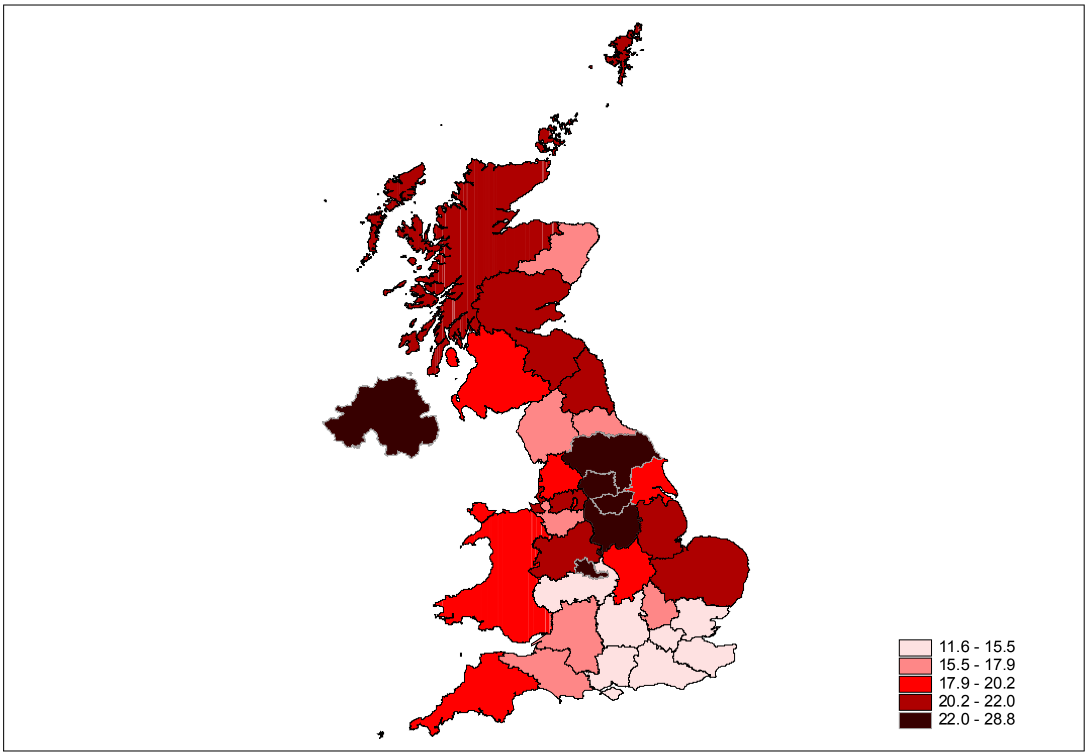
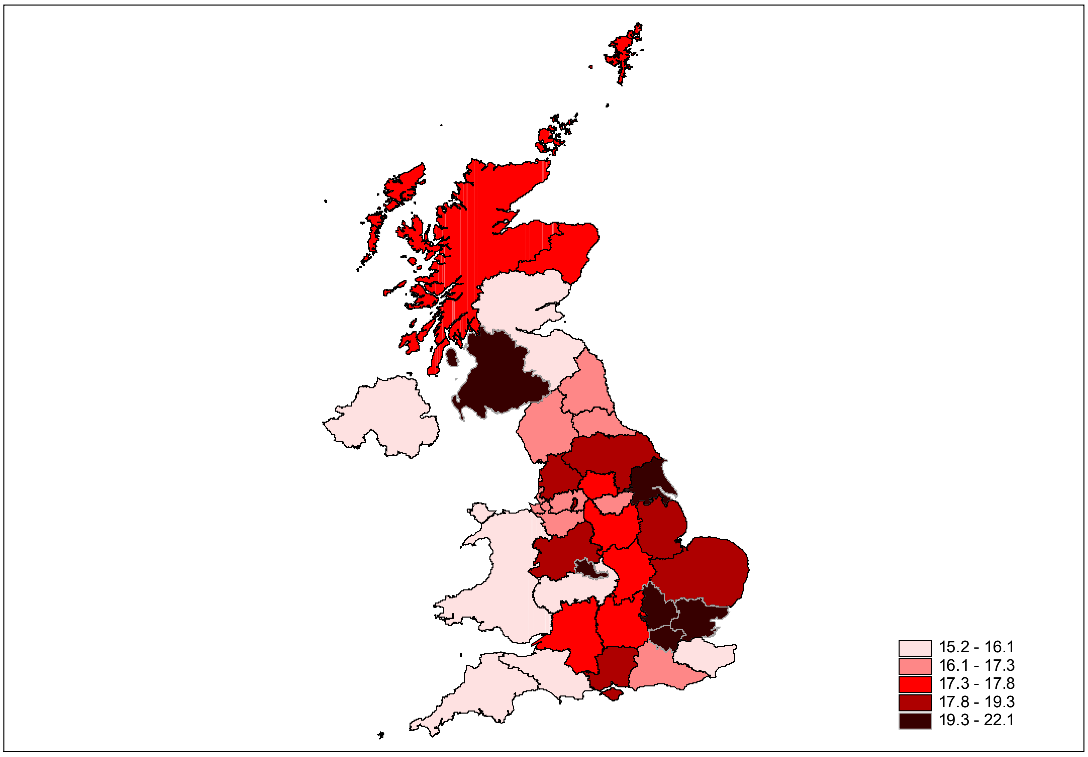


Figure 12.4 Head Count Ratio (NUTS2 poverty lines), United Kingdom NUTS2 regions



12.6 Overall deprivation measures

Figure 12.5 Fuzzy Supplementary: overall deprivation rates (NUTS2 regions)

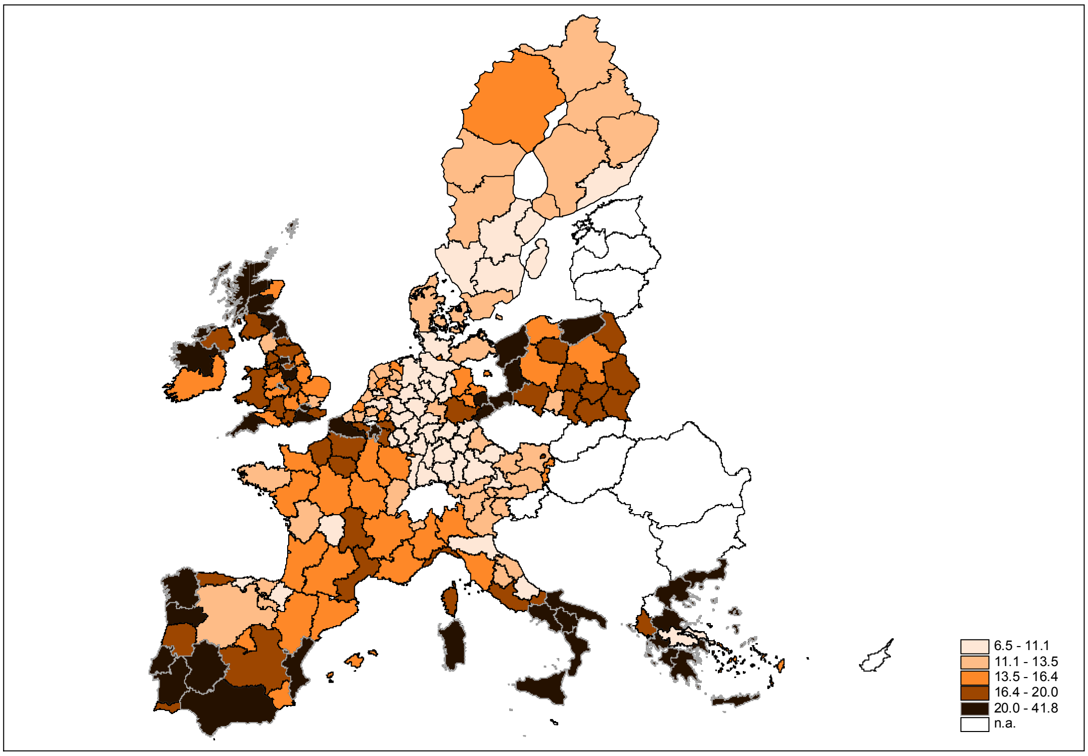


Figure 12.5 shows the variation of the overall non-monetary deprivation rate ('Fuzzy Supplementary') across NUTS2 regions. We observe very high values of deprivation in Portugal, South West Spain and South Italy. A better off country is Germany, showing a very homogenous behaviour among regions.

The following two figures show the same for Latent and Manifest deprivation rates. The Latent rate corresponds to the presence of either form of deprivation – income or non-monetary - at the individual level. The Manifest rate indicates simultaneous incidence of both these forms deprivation. Despite these differences, the overall pattern of results in Figures 12.6 and 12.7 is very similar.

Figure 12.6 Latent deprivation rates (NUTS2 regions)

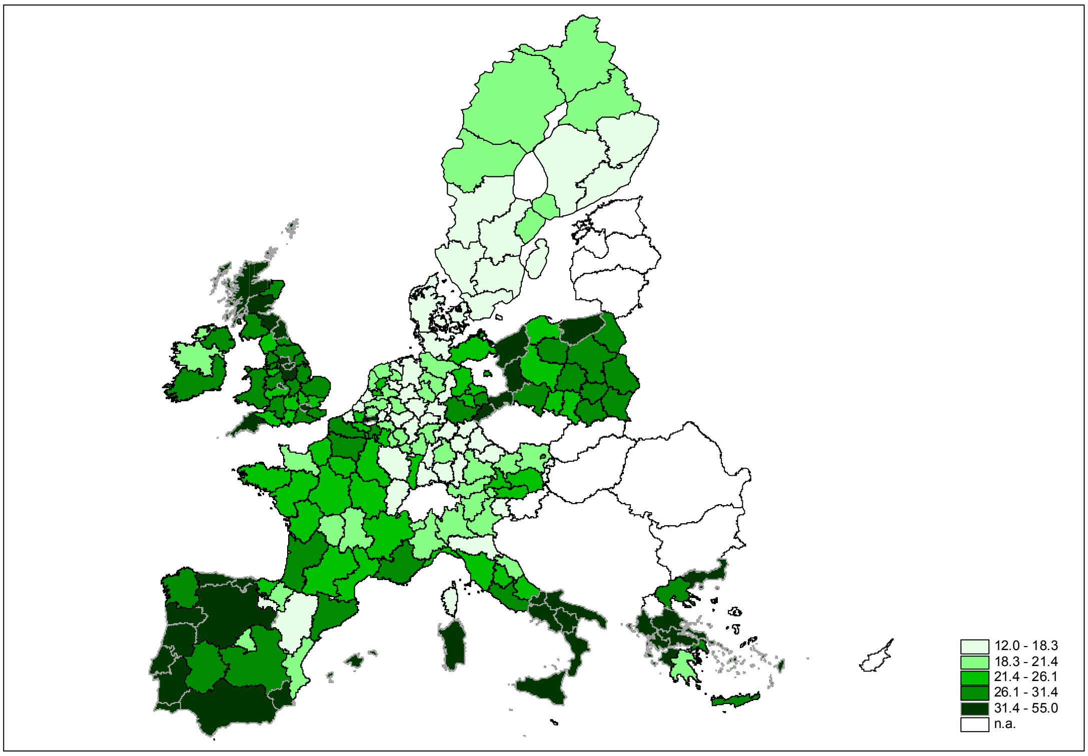
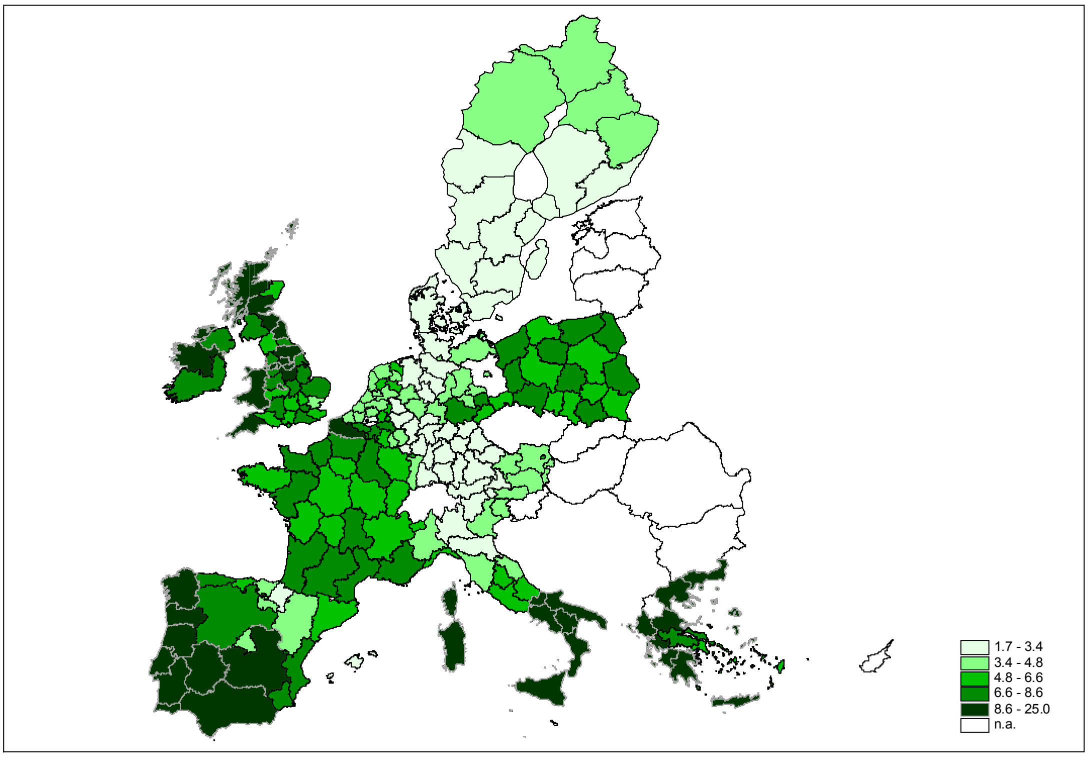


Figure 12.7 Manifest deprivation rates (NUTS2 regions)



12.7 Dimension-specific deprivation measures

Figures 12.8-12.11 display rates of non-monetary deprivation in specific dimensions.

Note that for all the five dimensions, we have no information for Sweden, Luxemburg and Germany.

Figure 12.8 Fuzzy Supplementary, dimension 1: basic life–style deprivation rates (NUTS2 regions)

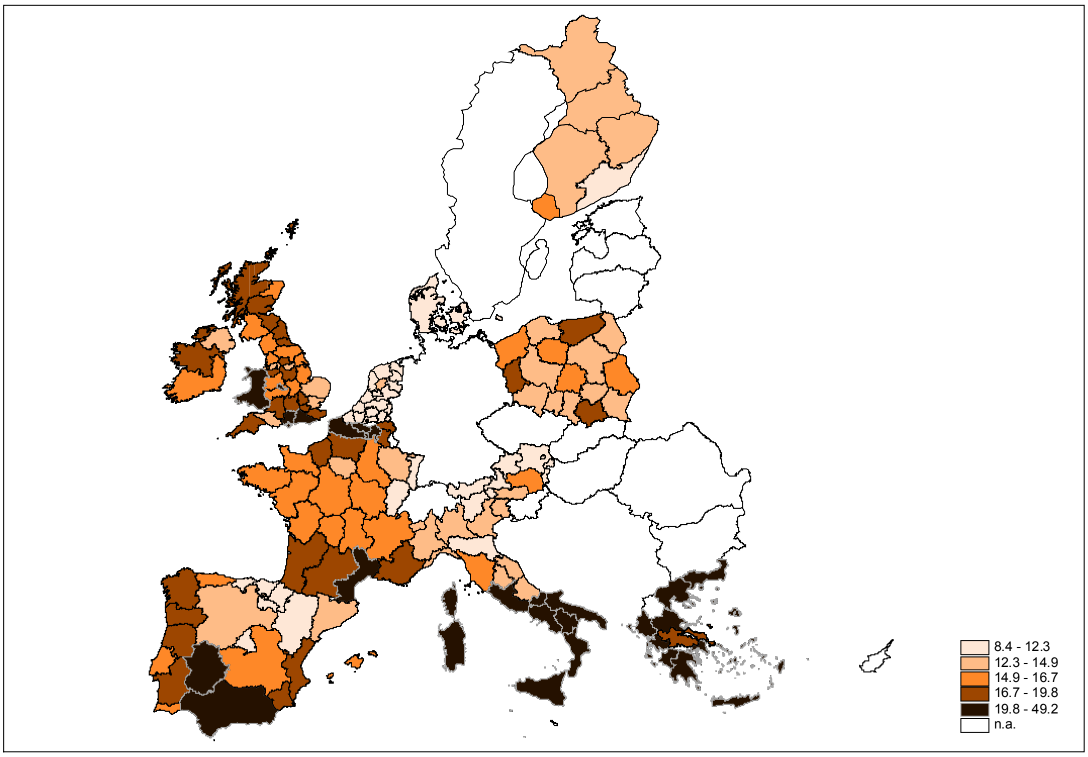
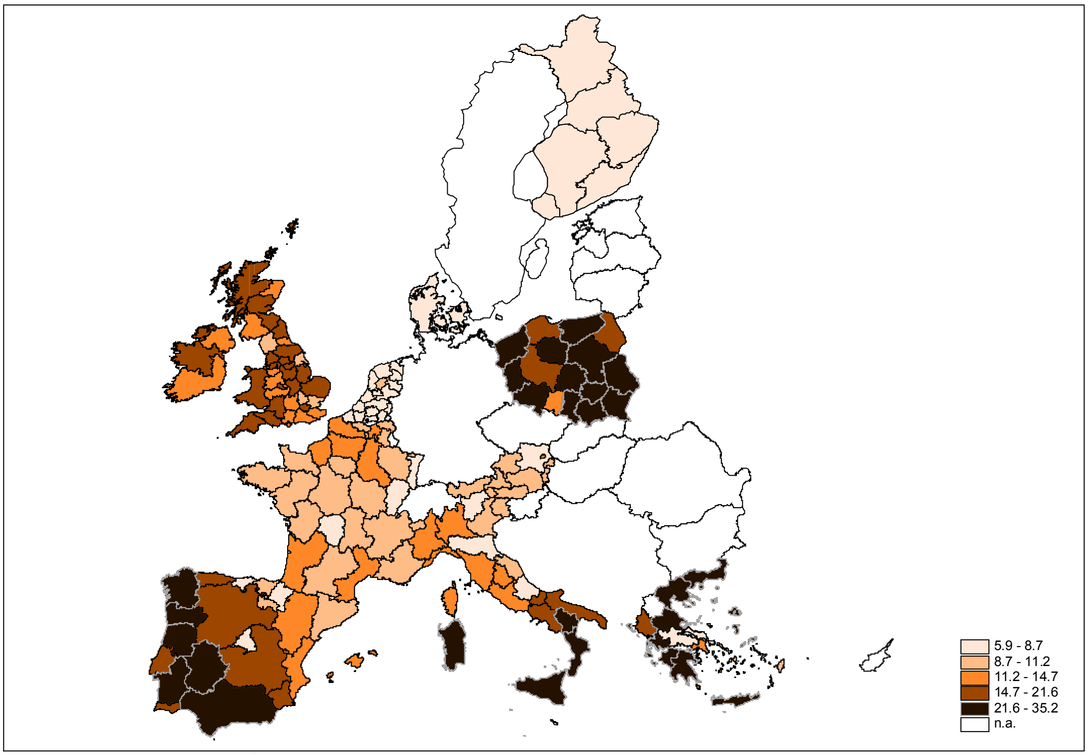


Figure 12.8 concerns the dimension 1 of the supplementary deprivation indicators (Basic Non-monetary Deprivation)

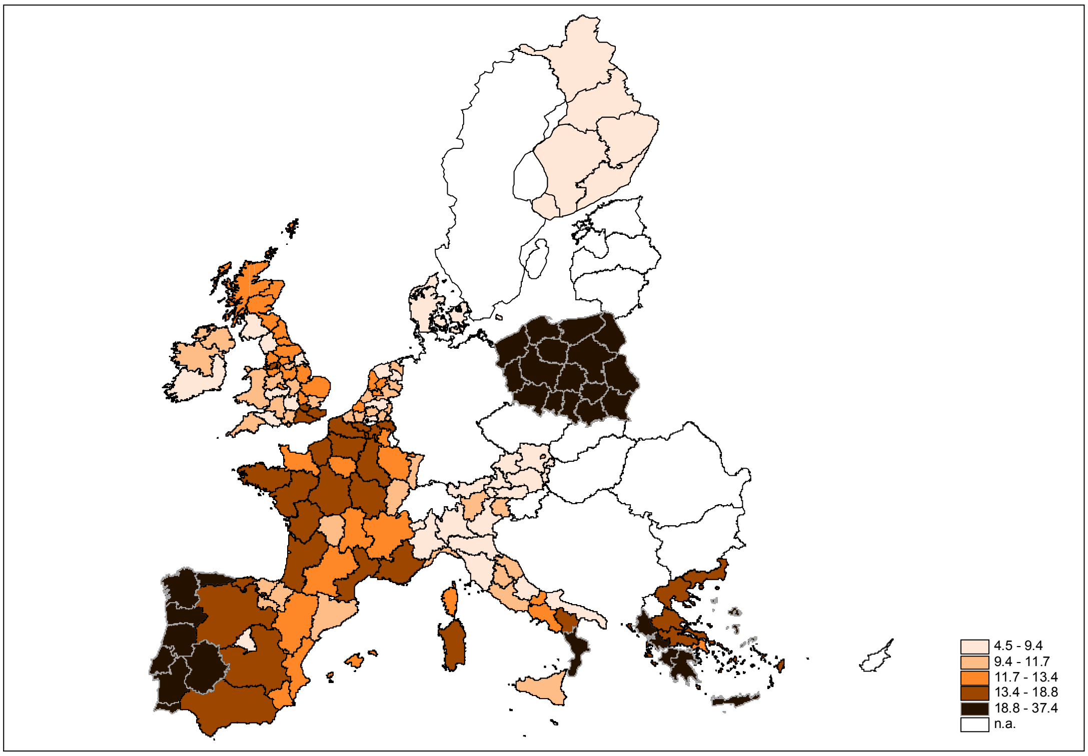
The worst situation is found in South Italy and most of Greece (only areas of these two countries showing a rate of 40% or higher). At the other extreme are areas in Northern Europe, with an interesting exception in Spain (Comunidad Foral de Navarra).

Figure 12.9 Fuzzy Supplementary, dimension 2: secondary life-style deprivation rates (NUTS2 regions)



According to the second dimension of the supplementary indicator FSUP-2 (secondary non monetary deprivation) we observe that southern Italy, some areas of Portugal, South-West Spain, Greece and Poland show the worst situation.

Figure 12.10 Fuzzy Supplementary, dimension 4: housing deterioration (NUTS2 regions)

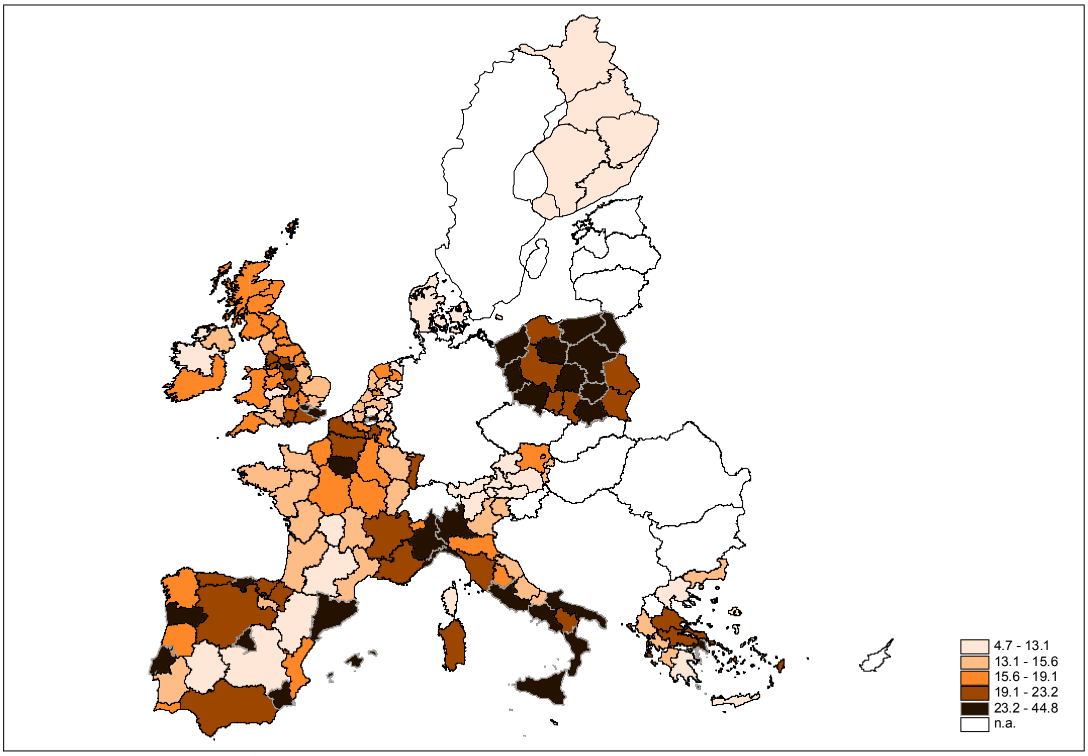


No results are available for FSUP-3 concerning lack of housing facilities because an adequate model could not be developed because of lack of sufficient data.

Concerning more serious housing deterioration (FSUP-4), Figure 12.10 shows a different trend compared to the other supplementary dimensions: very good results for Italy (both North and South), but also for France and UK. The worst figures are in Portugal and North Spain.

The same is indicated for most of Poland, but here it should be noted that the data for Poland on this item are not very comparable with ECHP data from EU15 countries because of differences in the questionnaire. (This item in Poland survey has been actually derived from a single attitudinal question.)

Figure 12.11 Fuzzy Supplementary, dimension 5: Environmental Problems (NUTS2 regions)



Looking at the fifth dimension (Environmental Problems) we can observe how the metropolitan areas (Madrid, Paris, London, Milan) have a very high values for this indicator of deprivation.

Chapter 13
Estimates to NUTS3 level: illustrations from Italy

13.1 Introduction

In this chapter, we present some regional indicators going down to NUTS3 regions in Italy. The estimates have been produced using ECHP data in conjunction with tabulations provided in "Eurostat Free Dissemination Database" (henceforth referred to as NewCronos) using the SAE modelling as described in preceding chapters. This exercise has been possible only because, through a special research agreement with ISTAT, we have been able to access ECHP Production Data Base for Italy, which contains the required NUTS3 identifiers. This highlights the importance of access (with proper confidentiality procedures, of course) to micro data for research aimed at producing useful official statistics.

The regional breakdown in Italy is as follows. The country is divided into 5 macro regions, which constitute the NUTS1 regions, as for instance coded in NewCronos; average population of a macro region is around 11.5 million. There has been a recent change in the NUTS classification as published in "Regional Statistics – Reference Guide 2004". In the case of Italy, a redistribution of NUTS2 regions has been made, and the number of NUTS1 regions has been reduced from earlier 11 to present 5. In the ECHP data, using the earlier classification, these 11 have been identified as NUTS1, which may be considered as the 'main regions' of the country; the average population of these units is around 5 million. It is possible to recode the ECHP regions into the present NUTS1 regions. However, for most of our analysis, we have retained the earlier (smaller) "NUTS1" regions, the main attraction of this choice being that in average size these regions are close to EU15 average NUTS1 size (of around 5.3 million). In any case, all results can be combined to the present NUTS1 classification in proportion to the regions' population size. There are 21 NUTS2 regions, corresponding to Regioni of Italy. NUTS3 regions are Province, there being 103 of them, with an average population of 560 thousand inhabitants.

A detailed and up-to-date description of NUTS is available on the RAMON server
http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html

The table below provides the correspondence between the NUTS levels and the national administrative units (2003) in Italy. At the local level, two levels of Local Administrative Units (LAU) have been defined. The upper LAU level (LAU level 1, formerly NUTS level 4) is defined for most countries, but not for all including Italy. The second LAU level (formerly NUTS level 5) consists of municipalities or equivalent units in the Member States.

	NUTS 1		**NUTS 2**		**NUTS 3**		**LAU 1**		**LAU 2**	
IT	Gruppi di regioni	5	Regioni	21*	Province	103	-		Comuni	8 100

*There are 20 regions in Italy. The results shown below are for these 20 regions. In the standard classification, one region (Trentino Alto Adige) has been divided in two NUTS2 regions.

There are only a very limited number of covariates available from NewCronos at NUTS3 level. This limitation should be born in mind in the interpretation of the results. *Using the methodology developed here, but in conjunction with more and better covariates from alternative sources in Italy, the results can no doubt be improved greatly.*

13.2 Income poverty: NUTS2 level of disaggregation

We begin by presentation of some maps on HCR's computed on the basis of the results of the small area model on ECHP data. We consider three levels of poverty lines in turn:

(1) Country poverty line, defined on the basis of income distribution pooled for the whole country;

(2) NUTS1 poverty line, defined on the basis of income distribution pooled for each NUTS1 region separately;

(3) NUTS2 poverty line, defined on the basis of the income distribution considered separately for each NUTS2 region.

For each of these three poverty line levels, and taking three poverty line thresholds - namely 50, 60 and 70% of the median income at whatever level the income distribution is considered – we compute income poverty rates for each of the 8 ECHP waves. These rates can be produced at any level of aggregation: for the whole country, for each NUTS1, for each NUTS2, each NUTS3 region, etc..

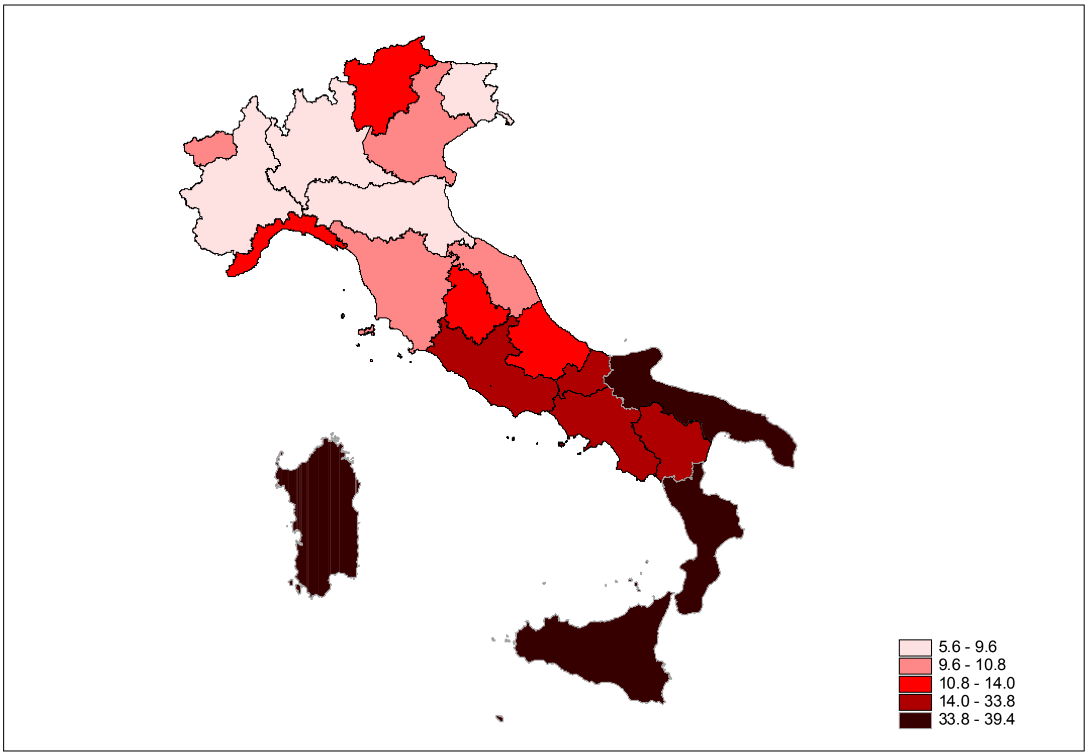
Thereafter, these direct estimates from the survey have been consolidated over 8 ECHP waves, and also over the three poverty line thresholds. As explained in Chapter 5, the consolidation procedure is such that the result can be regarded to correspond quite closely to the standard definition of 60% of the median, averaged over waves.

For poverty line levels (1) and (3) above, the results presented are from the SAE models described in Section 12.4 in the preceding chapter. However, only direct survey estimates are currently available for (2) since we have not applied the model with NUTS1 poverty lines.

We first consider poverty rates at NUTS2 level (Section 13.2). Here NUTS2 is the level of disaggregation of the results. For each NUTS2 region, different poverty rates are produced using different poverty line levels (1)-(3).

Next we will consider poverty rates at NUTS3 level (Section 13.3). Because of the small population and sample sizes, we do not define the poverty line at NUTS3 level (i.e., by considering the income distribution at each NUTS3 separately). Instead, we will focus on two poverty line levels: country and NUTS2. Section 13.4 describes the relationship between the mean level of income and the poverty rate using different poverty line levels, while Section 13.5 considers supplementary (non-monetary) and combined (income and non-monetary) indicators of deprivation.

Figure 13.1 Head Count Ratio, NUTS2 regions (country poverty line). Italy



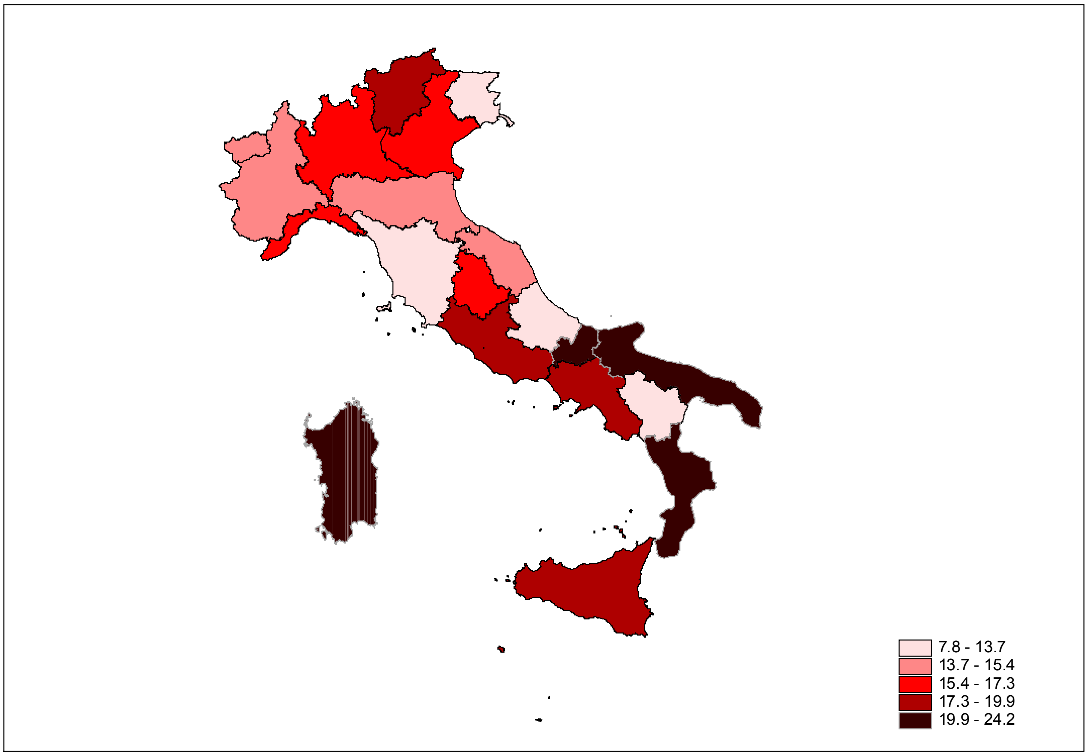
NUTS2 poverty rates using country poverty line

With HCR's computed using the country poverty line, the regional situation is as expected. The regions with the higher head count ratios are in the South Italy. The only interesting exception is Basilicata, with 27% of population in poverty, against 37% and 39% in Puglia and Calabria. The region with the highest poverty rate is Sicilia with 39%.

In North Italy, the regions with a better position (i.e., regions with low HCR's) are, as expected, Friuli Venezia Giulia, Emilia Romagna, Lombardia and Piemonte.

In our view, the result for Trentino Alto Adige appear somewhat unexpected; it should be noted however that in this case the regressors were not always available for good estimation of the model.

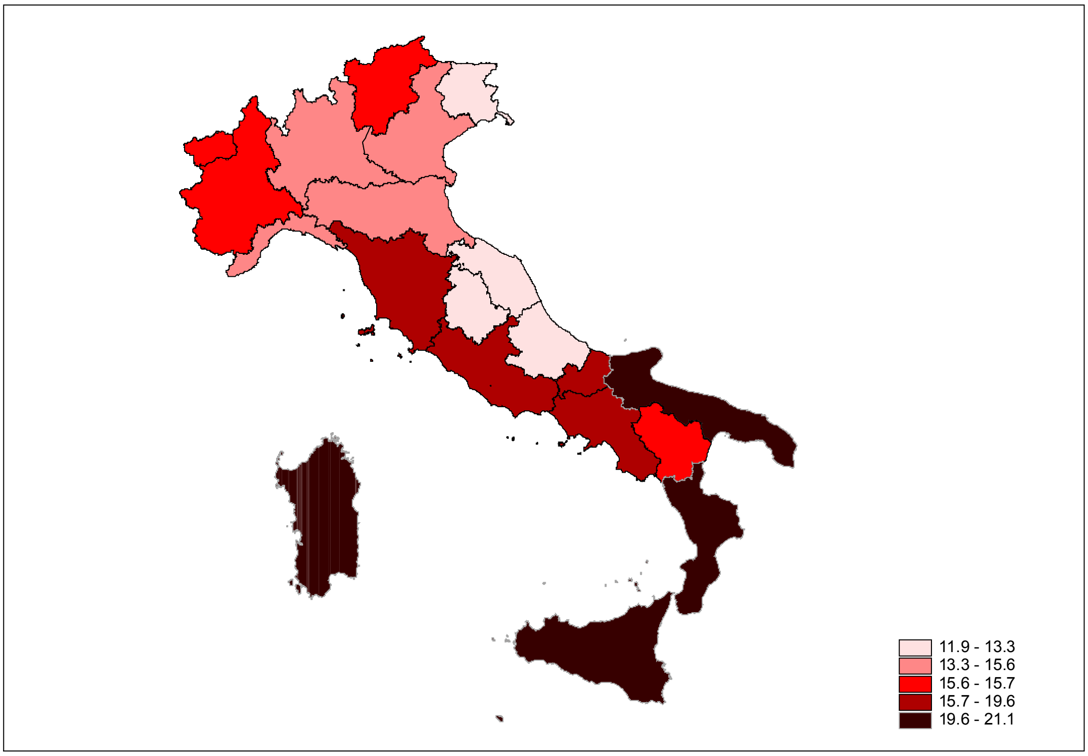
Figure 13.2 Head Count Ratio, NUTS2 regions (NUTS1 poverty lines). Italy



NUTS2 poverty rates using NUTS1 poverty lines

As noted, the results mapped here are based *only on direct survey estimates* from ECHP data. In any case, it is clear that a change in the level at which the poverty line is defined changes the resulting picture. The NUTS1 poverty lines highlight the relative differences within macro-region (NUTS1). For example, it can be seen how the poverty situation in North-East appears to vary geographically: in the same NUTS1 region, we have different performances for each NUTS2 region (Trentino Alto Adige with HCR as 18%; Veneto with 16% and Friuli Venezia Giulia with 8%). A similar pattern can also be seen in the South, where Basilicata has a very low poverty rate compared to Puglia and Calabria.

Figure 13.3 Head Count Ratio, NUTS2 regions (NUTS2 poverty lines). Italy



NUTS2 poverty rates using NUTS2 poverty lines

The logic behind Figure 13.3 is the comparison of inequality *within* NUTS2 regions. Clearly using NUTS2 poverty lines for analysing NUTS2 regions, we have a completely relative measure of HCR. The effect of NUTS2 differences in the levels of income is removed. Also the results are not affected by any cost-of-living differences among NUTS2 regions, since the income distribution for each region is considered separately. Even so, the regions with the higher inequality are in the South, and the ones with the lowest inequality are in the Centre. There is a statistical association between the level of income and the poverty rate, even when the latter is a purely relative measure. This has often been observed at country level in EU, and here we observe the same phenomenon at NUTS2 level in Italy.

An interesting exception in the centre is Toscana, which has generally low poverty rates using country or NUTS1 poverty line, but with quite high inequality (HCR=16%) as measured by NUTS2 poverty line.

The North can be divided in two parts, East and West - the former with less unequal income distribution and the latter with more inequality within NUTS2 regions. It can be seen that in Liguria the poverty rate using a NUTS2 poverty line is lower than that in the other region in the West; this is in contrast to the picture we saw when considering country or NUTS1 poverty lines.

13.3 Income poverty: NUTS3 level of disaggregation

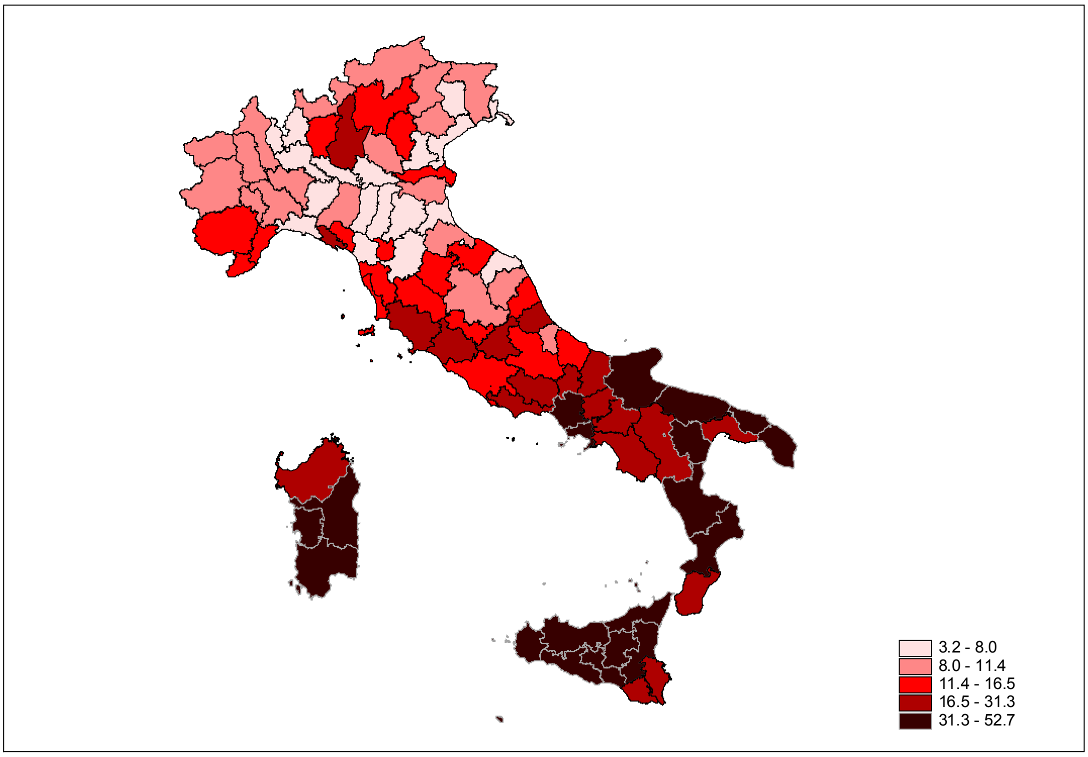
In analysing the poverty situation it is very useful to look at the results disaggregated to NUTS3 level, which corresponds to Province in Italy.

We show results for two poverty line levels:

(1) Country poverty line, defined on the basis of pooled income distribution for the whole country;

(2) NUTS2 poverty line, defined on the basis of the income distribution pooled for NUTS3 regions within each NUTS2.

Figure 13.4 Head Count Ratio, NUTS3 regions (country poverty line). Italy

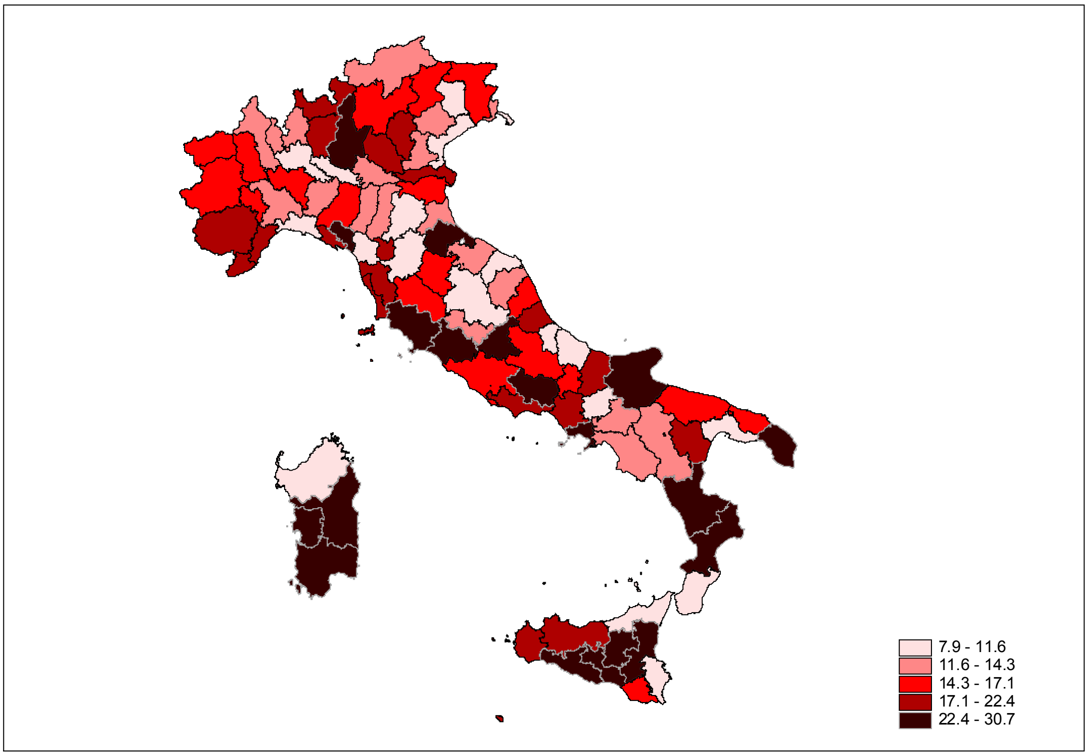


NUTS3 poverty rates using country poverty lines

The results are coherent with the ones shown at NUTS2 level of aggregation, but now we can see which are the provinces with a better (less unequal) or a worse (more unequal) position. The highest head count ratios (above 48%) are for Catania, Enna (in Sicilia), Oristano (in Sardegna) and Foggia (in Puglia).

The provinces with lowest rates are all in North East, in the same region (Friuli Venezia Giulia); the provinces involved being Pordenone, Trieste and Gorizia (all with a ratio lower than 4.5%)

Figure 13.5 Head Count Ratio, NUTS3 regions (NUTS2 poverty lines). Italy



NUTS3 poverty rates using NUTS2 poverty lines

While using the country poverty line provides a more or less absolute measure of regional differences in income, the use of NUTS2 poverty lines gives an approximately relative measure. The latter does depend on regional differences in levels of income, but only among NUTS3 regions within the *same* NUTS2. Differences on a larger geographical scale, such as across NUTS2 or higher-level units, do not matter.

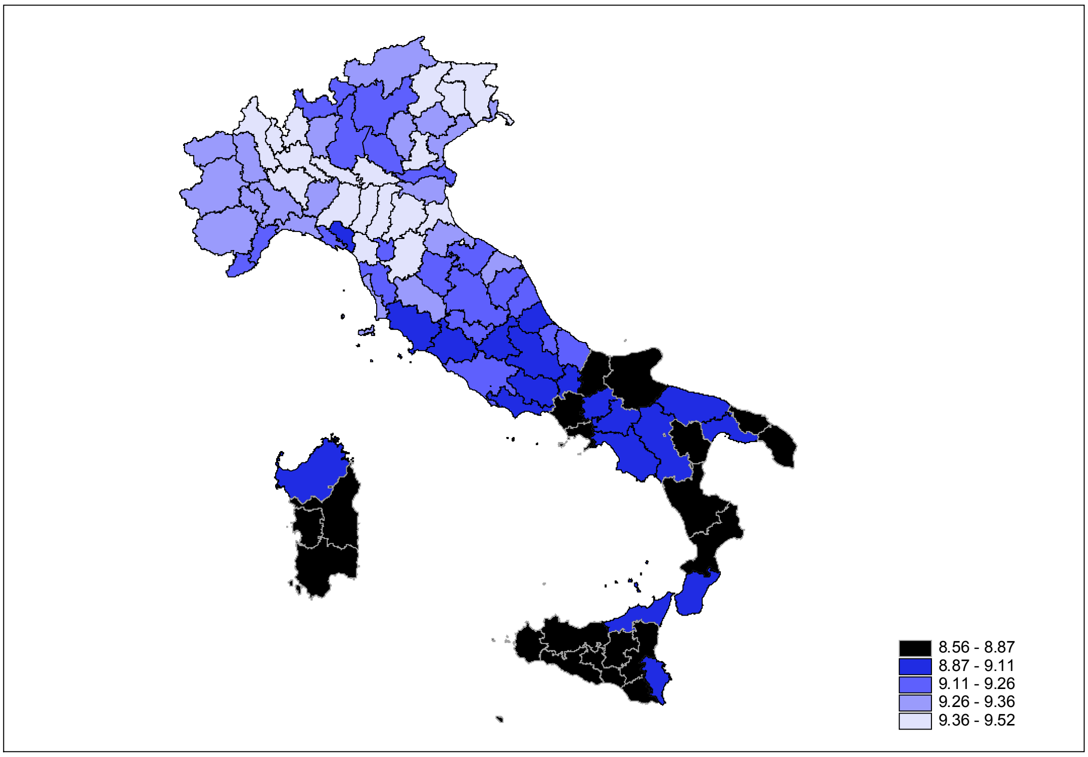
It is also useful to examine how the results for Toscana compare to the results obtained using the recent survey "Indagine sulle Condizioni di Vita in Toscana", which has a much bigger sample size than ECHP for the region concerned. We have found that the results for NUTS3 (Province in Toscana) based on NUTS2 poverty line compare rather well with those (not yet published) of the Tuscanian survey. The main differences are for those provinces which have very small sample sizes in ECHP. This indicates that the model figures are also affected by the small number of regressors available for the model.

An important recommendation: it is possible in principle to incorporate more specific information on provinces, regions and macro regions from diverse sources in Italy, and hence improve the performance of the model. For reasons to do with time and accessibility, we have relied exclusively on ECHP and the tables provided in NewCronos. In practice, all models to low levels such as NUTS3 (or beyond) must be country-specific. It is not necessary or useful to seek standardisation at EU level.

13.4 Level of income and poverty rate

In this section we explore the relationship between the level of income of a region and its poverty rate, in comparison with other regions. The level of disaggregation we consider is NUTS3 regions.

Figure 13.6 Mean net equivalised income (log), NUTS3 regions. Italy



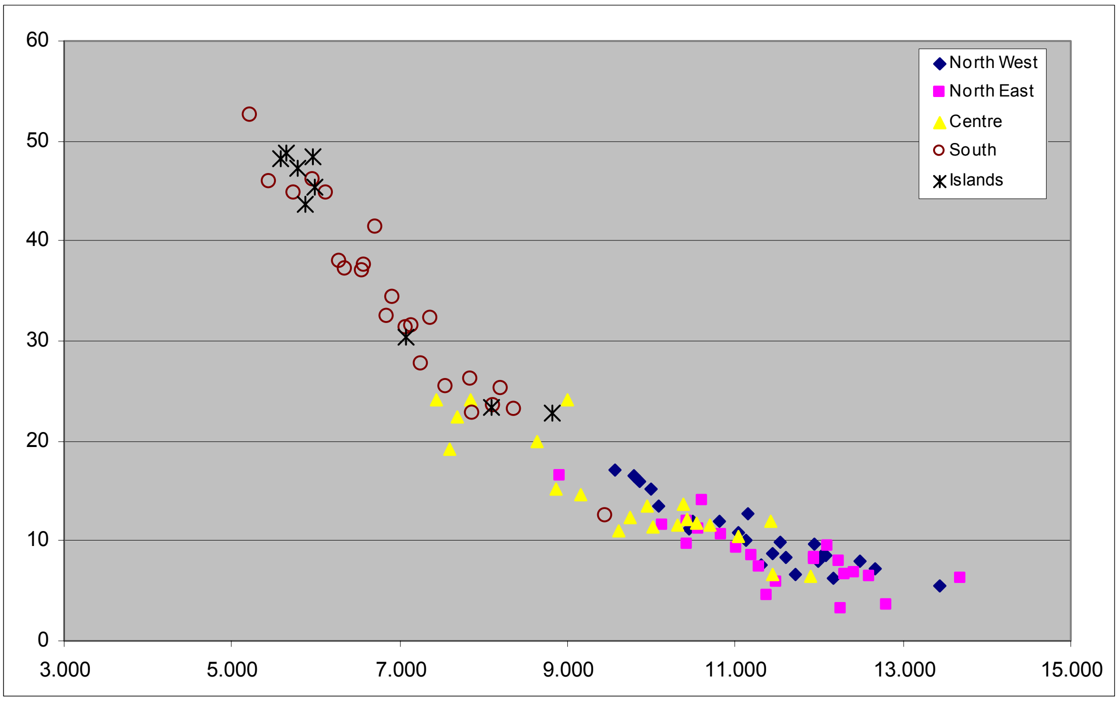
The map above examines variation in the mean level of income across regions.[37] We note how provinces such as Pordenone, Trieste and Gorizia with a very low HCR using country poverty line, also have a high mean income. Trieste is in fact the third province with the highest mean income, together with Bologna and Milano. At the bottom of this ranking we find Catania, Agrigento, Cosenza and Foggia (the latter also has the highest HCR using the country poverty line).

37 The variable actually shown is mean of log equivalised incomes, because it is in that form that the variable is more appropriately modelled. Of course, this transformation leaves the relative position of the regions essentially unchanged.

In order to bring out the relationship between the poverty rate and the income distribution more clearly, we now present two interesting graphs which compare the head count ratio calculated using two poverty lines in turn - (a) a country poverty line, and then (b) a NUTS2 poverty line – against the mean level of income.

The regional variation in income level to poverty rate relationship come out very clearly when we examine the results for provinces separately by macro region. As note, the country is divided into 5 macro region (NUTS1 according to the current classification).

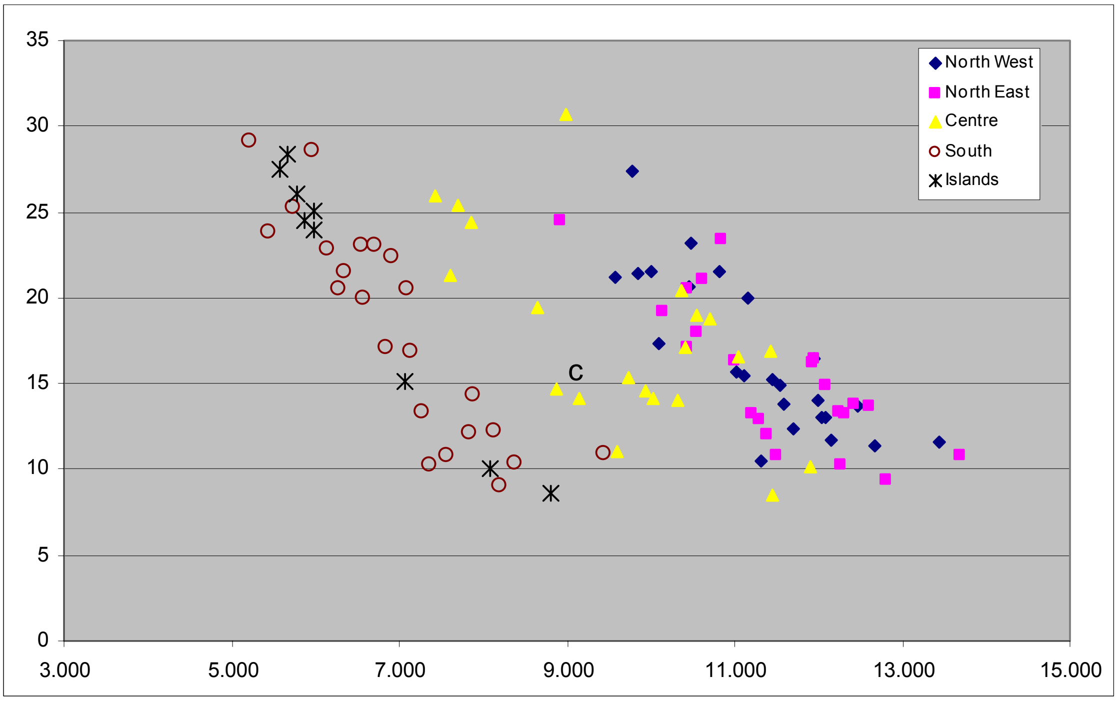
Figure 13.7 Mean equivalised income level versus HCR for NUTS3 regions- country poverty lines. Italy



Country poverty line

There is a clear relationship between the level of income and the poverty rate using country poverty line. Lower income levels tend to go with higher poverty rates, and higher income levels with lower poverty rates. The North-South division is prominently obvious in Figure 13.7. The Centre (Toscana, Umbria, Marche and Lazio) is more mixed and its NUTS2 points cover the whole range in Figure 13.7.

Figure 13.8 Mean equivalised income level versus HCR for NUTS3 regions- NUTS2 poverty line. Italy



NUTS2 poverty line

From Figure 13.8 the strong relationship between income levels and poverty rates in Italy emerges most clearly.

The pattern in Figure 13.8 is remarkable. It displays an almost purely relative measure of income distribution, affected only by NUTS3 variation in levels within NUTS2 regions. Yet the North-South division is still clearly seen.[38] A negative HCR-Income level relationship can be seen within each micro region separately, the series of parallel regression lines moving to the right in Figure 13.8 (higher income levels) from South to North.

In Table 13.1, we present all the figures plotted in the previous graphs; NUTS3 regions are ranked by the mean level of income.

---

[38] Of course, the points in this graph are somewhat more scattered than those in the previous graph, but the latter was based on the country poverty line which increases the strength of the relationship between HCR and income level.

Table 13.1 Mean equivalised income level and poverty rates for NUTS3 regions; Country and NUTS2 poverty lines. Italy

Nuts3	Name	HCR_C	HCR_N2	Income**	Nuts3	Name	HCR_C	HCR_N2	Income**
IT911	Foggia	52,7	29,2	5219	IT327	Rovigo	11,7	19,2	10140
ITA04	Agrigento	47,1	24,1	5324	IT531	Pesaro e urbino	11,5	14,1	10317
IT931	Cosenza	46,0	23,8	5439	IT513	Pistoia	13,6	20,4	10372
ITA07	Catania	48,3	27,5	5587	IT518	Arezzo	12,1	17,1	10419
ITB03	Oristano	48,8	28,4	5656	IT312	Trento	12,0	17,1	10434
IT915	Lecce	44,8	25,3	5731	IT409	Rimini	9,7	20,6	10435
ITB02	Nuoro	47,2	26,0	5777	IT204	Sondrio	11,3	20,7	10447
ITB04	Cagliari	43,8	24,5	5874	IT209	Lodi	12,0	23,2	10490
IT932	Crotone	46,1	28,6	5962	IT517	Pisa	11,7	19,0	10541
ITA06	Enna	48,4	25,0	5975	IT321	Verona	11,2	18,0	10552
ITA05	Caltanissetta	45,5	24,0	5992	IT322	Vicenza	14,1	21,1	10606
IT934	Vibo valentia	44,9	22,8	6134	IT516	Livorno	11,6	18,8	10708
ITA01	Trapani	37,9	20,6	6276	IT206	Bergamo	12,0	21,5	10816
IT801	Caserta	37,2	21,5	6355	IT408	Forlì-cesena	10,6	23,4	10832
IT803	Napoli	37,0	23,1	6551	IT406	Ferrara	9,2	16,4	11007
ITA02	Palermo	37,7	20,0	6566	IT120	Valle d'Aosta	10,8	15,7	11030
IT933	Catanzaro	41,4	23,0	6709	IT515	Prato	10,5	16,6	11041
IT914	Brindisi	32,5	17,1	6850	IT117	Asti	10,1	15,4	11121
IT922	Matera	34,4	22,4	6912	IT116	Cuneo	12,8	20,0	11158
ITA08	Ragusa	30,4	15,2	7075	IT311	Bolzano-Bozen	9,8	14,3	11200
IT722	Campobasso	31,2	20,6	7081	IT324	Treviso	8,5	13,3	11203
IT912	Bari	31,6	16,9	7139	IT401	Piacenza	7,5	12,9	11295
IT805	Salerno	27,8	13,4	7266	IT133	Genova	7,6	10,5	11315
ITA03	Messina	32,3	10,2	7362	IT333	Gorizia	4,6	12,0	11374
IT601	Viterbo	24,0	26,0	7430	IT519	Siena	12,0	16,9	11418
IT913	Taranto	25,5	10,9	7556	IT532	Ancona	6,6	8,5	11446
IT604	Latina	19,1	21,3	7599	IT111	Torino	8,8	15,2	11455
IT605	Frosinone	22,4	25,4	7694	IT325	Venezia	5,8	10,8	11489
IT804	Avellino	26,2	12,2	7847	IT112	Vercelli	10,0	14,9	11541
IT602	Rieti	24,1	24,4	7852	IT118	Alessandria	8,4	13,8	11599
IT721	Isernia	22,8	14,3	7871	IT201	Varese	6,7	12,4	11711
ITB01	Sassari	23,4	10,1	8085	IT514	Firenze	6,4	10,1	11896
IT921	Potenza	23,6	12,2	8125	IT402	Parma	8,4	16,2	11933
IT935	Reggio di calabria	25,2	9,1	8202	IT208	Pavia	9,6	16,4	11941
IT802	Benevento	23,2	10,4	8369	IT332	Udine	8,2	16,4	11950
IT712	Teramo	20,0	19,5	8644	IT202	Como	7,9	14,1	11998
ITA09	Siracusa	22,8	8,6	8809	IT113	Biella	8,5	13,1	12036
IT711	L'aquila	15,1	14,6	8870	IT115	Novara	8,6	13,1	12083
IT511	Massa-carrara	16,5	24,5	8915	IT323	Belluno	9,5	14,9	12095
IT51a	Grosseto	24,2	30,7	8999	IT203	Lecco	6,3	11,8	12159
IT522	Terni	14,6	14,1	9153	IT512	Lucca	4,9	7,9	12242
IT714	Chieti	12,5	10,9	9447	IT326	Padova	8,0	13,3	12248
IT134	La spezia	17,1	21,2	9572	IT331	Pordenone	3,2	10,2	12261
IT521	Perugia	10,9	11,0	9601	IT403	Reggio nell'emilia	6,7	13,3	12304
IT534	Ascoli piceno	12,3	15,3	9746	IT407	Ravenna	6,8	13,8	12428
IT207	Brescia	16,5	27,4	9784	IT20b	Mantova	8,0	13,7	12476
IT132	Savona	15,9	21,4	9859	IT404	Modena	6,4	13,7	12604
IT603	Roma	13,5	14,6	9945	IT20a	Cremona	7,2	11,4	12666
IT114	Verbano-Cusio-Ossola	15,1	21,5	10006	IT334	Trieste	3,6	9,4	12812
IT533	Macerata	11,4	14,1	10029	IT205	Milano	5,5	11,6	13431
IT131	Imperia	13,5	17,4	10087	IT405	Bologna	6,3	10,8	13680
IT713	Pescara	8,9	8,8	10139					

**The income in this table is the geometric (rather than the usual arithmetic) mean of the equivalised income. This is because we have modelled log(income), the antilog of the predicted mean of which gives geometric mean of actual incomes.

As shown in the table, using both the SAE model (and the regression-prediction procedure in a few areas which did not contain any sample point), we have been have to produce indicators on poverty for all the 103 provinces in Italy. The ECHP survey data are not available for all of them, and for some provinces the sample size is too small to produce a reliable figure.

The Statistical Annex at the end of this report gives more detailed information of the full set of results.

13.5 Supplementary and combined deprivation indicators - NUTS2 level of disaggregation

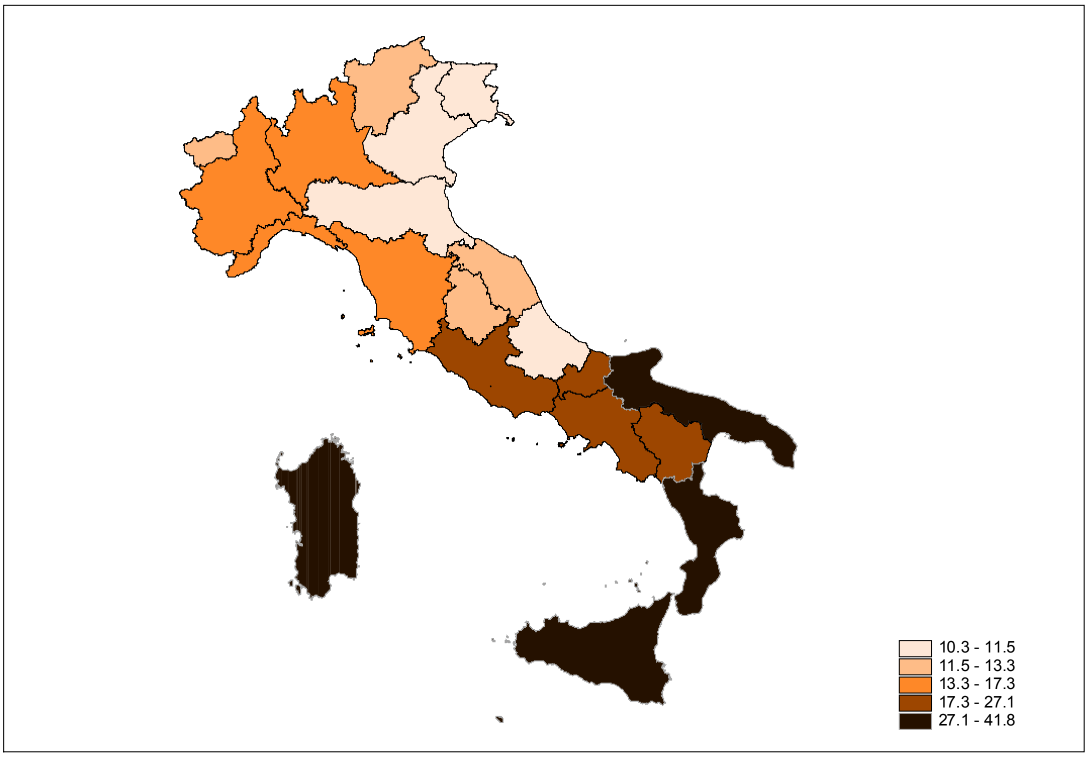
Finally, maps at NUTS2 level are presented for three measures:

(1) "Fuzzy Supplementary", meaning the non-monetary deprivation rate. This has been constructed as described in details in Chapter 6, and the values have been modelled.

(2) Latent deprivation rate, indicating the presence of either of the two forms of deprivation: income poverty and/or non-monetary deprivation.

(3) Manifest deprivation rate, indicating the presence of both of the two forms of deprivation: income poverty and non-monetary deprivation.

Figure 13.9 Fuzzy Supplementary: overall deprivation rates, NUTS2 regions. Italy



Looking at the non monetary indicator of deprivation, Puglia, Calabria, Sicilia and Sardegna are in the group with greatest inequality, as it was the case for monetary indicators. But still, Northern-East Italian regions and Emilia Romagna appear to have relatively less inequality.

The picture is very similar indeed when we examine the pattern according to Latent and Manifest deprivation measures (Figures 13.10 and 13.11).

Figure 13.10 Latent deprivation rates, NUTS2 regions. Italy

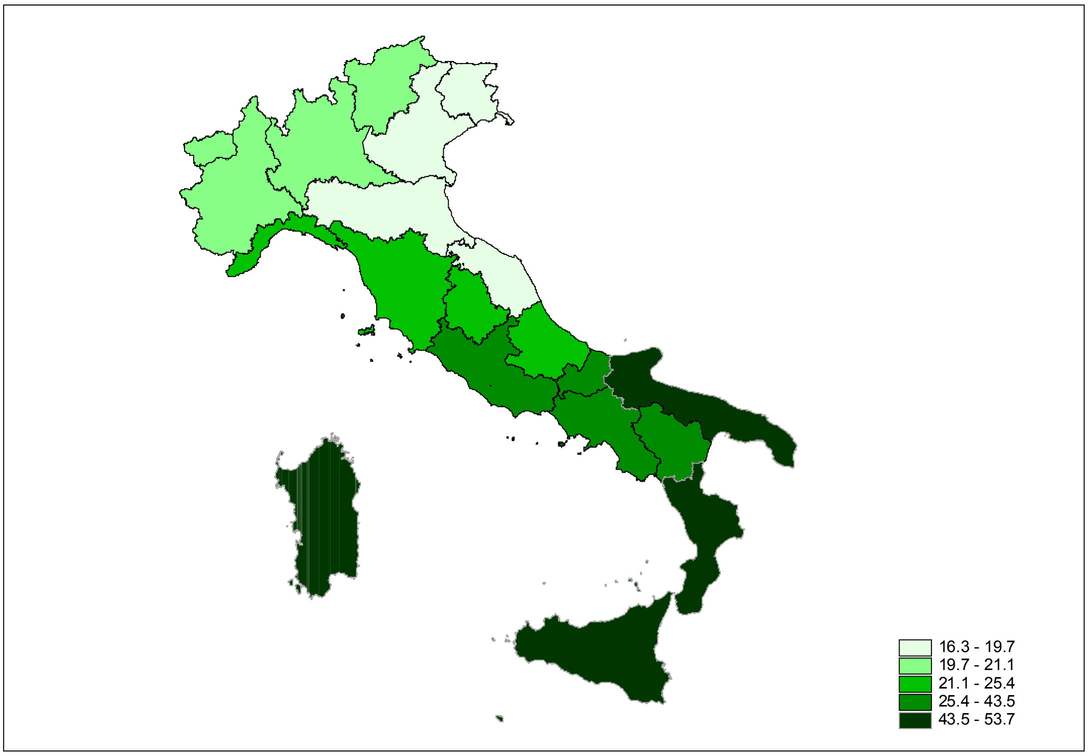
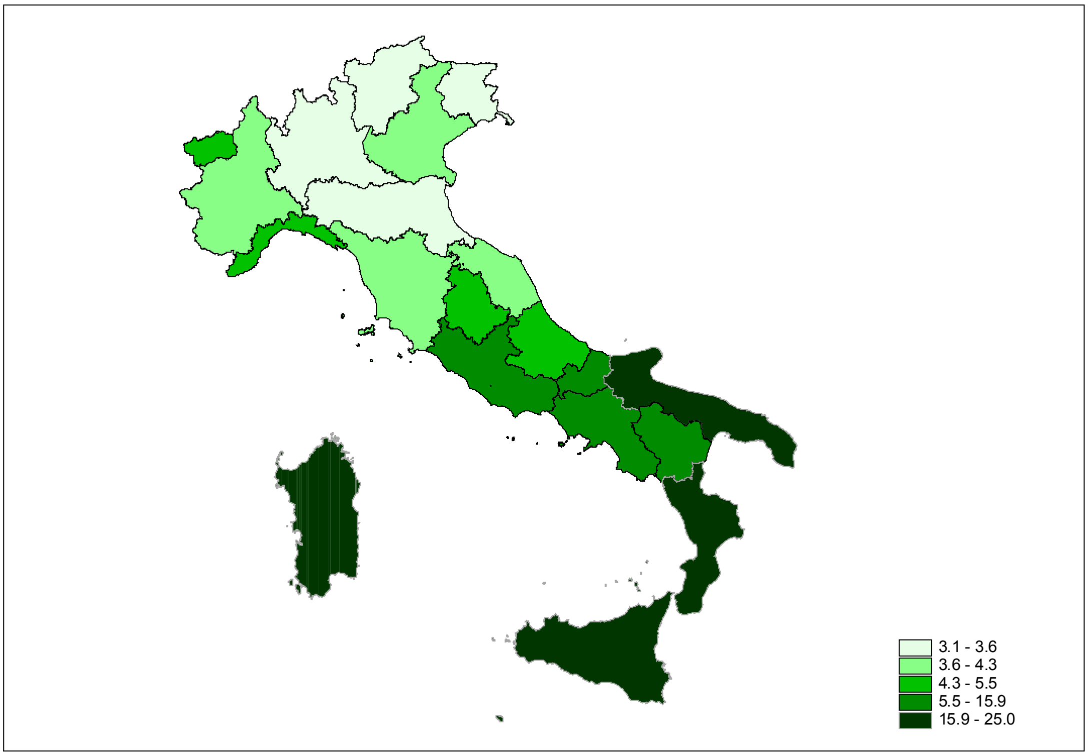


Figure 13.11 Manifest deprivation rates, NUTS2 regions. Italy



Chapter 14
Small Area Estimation (SAE): Going beyond NUTS3

14.1 Introduction

In Chapters 11-13, we have considered SAE methodology going down to NUTS2 or at most to NUTS3 level. With the type of surveys available, such as ECHP, it is generally not possible to go beyond NUTS3. In any case the main source of auxiliary data we have used, namely NewCronos tables, is not available beyond NUTS3 breakdown.

Production of estimates at lower (NUTS4 and NUTS5) levels would require models of a different type. These models are likely to be statistically less precise. Such models may of course vary among themselves in the degree of sophistication depending on whether they are *area-level* or *unit-level* models and whether they are *stochastic* or *deterministic*.

In general terms, these models involve imputing the required dependent variables – such as poverty measures – to areas or to individual households in a large data set such as a population census, essentially on the basis of a regression model fitted from a small-scale survey containing common covariates and the required dependent variables. In this chapter we describe some aspects of methodologies that have been applied in the United Kingdom and Albania using data from a Population Census, in conjunction with an intensive small scale national sample survey. In the application for Albania, the aim was also to perform poverty, deprivation and inequality mapping.[39] Spatial descriptions of the distribution of poverty, deprivation and inequality are of most use to policy-makers and researchers when they are finely disaggregated, i.e. when they represent small geographic units, such as cities, municipalities, districts or other administrative partitions of a country. In order to produce such maps, large data sets are required which include reasonable measures of income or consumption expenditure and which are representative and of sufficient size at low levels of aggregation to yield statistically reliable estimates. Intensive household surveys usually used to calculate distributional measures are rarely of such a sufficient size; whereas census or other large sample surveys large enough to allow disaggregation have little or no information regarding monetary variables.

14.2 A SAE methodology used in the United Kingdom

In the UK deprivation indices are constructed for local government units. Such territorial indicators have been adopted for the targeting government policy geographically. The measurement of deprivation can be performed for very small areas such as wards (LAU2, formerly NUTS5) only using Census data. However, as stated in Payne *et al.* (1996) in relation to 1991 Census, census questions are not specifically designed to measure deprivation, so that any Census-based index will be comprised of variables which are at best proxy measures of deprivation.

[39] We have taken Albania – rather than one of the EU or Candidate countries – for detailed illustration because of the availability to us of micro data not only for an intensive survey containing the required target variables concerning poverty and deprivation but, exceptionally, also for the population census.

Indices of multiple deprivation in local areas

In 2000 the Department of the Environment, Transport and the Regions (DETR) produced the *Index of Local Deprivation* (ID 2000) based on six *domains* of deprivation, plus one more dimension related to child poverty index (Fahmy and Gordon, 2002):

- Poverty rate (*Income*)
- Child poverty rate (*Supplementary*)
- Unemployment rate (*Employment*)
- Percent of people aged 18 and over with no post school qualifications (*Education*)
- Limiting Long Term Illness/Disability rate (*Health*)
- Percent of households with no central heating (*Housing*)
- Percent of households with no car (*Access to services*)

Only indicators contained in last five domains could be constructed on the basis of the 1991 Census. The first two – the most direct measures of income poverty – were derived from the 1990 Breadline Britain Survey and then imputed on to the 1991 Census data (Gordon, 1995; Gordon and Pantazis, 1997).

In 2004 the Office of the Deputy Prime Minister (ODPM) commissioned the Social Disadvantaged Research Centre (SDRC) of the Department of Social Policy and Social Research at the University of Oxford to update the ID2000 for England.

As its predecessor, the Index of Multiple Deprivation 2004 (IMD 2004) is a measure of multiple deprivation at the small area level. The model of multiple deprivation underpinning it is based on the idea of distinct dimensions or domains of deprivation which can be recognised and measured separately. These are experienced by individuals living in an area. People may be counted in one or more of the domains, depending on the number of types of deprivation that they experience. The overall IMD is conceptualised as a weighted area level aggregation of these specific dimensions of deprivation.

This index (ID2004) was is based on seven dimensions as follows:

- Income deprivation
- Employment deprivation
- Health deprivation and disability
- Education, skills and training deprivation
- Barriers to housing and services
- Crime
- Living Environment deprivation

Each of the 32,482 "Super Output Areas" (SOAs) in England has been assigned a score and rank for the Index of Multiple Deprivation; the seven Domain Indices; the sub-domains; and the two supplementary Indices (Income Deprivation Affecting Children and Income Deprivation Affecting Older People).

As for ID2000, the first domain of the ID2004 – concerning direct measurement of income poverty - cannot be directly measured from the Census data set.

Each domain contains a number of *indicators*. The criteria for inclusion of these indicators are that "they should be 'domain specific' and appropriate for the purpose (as direct as possible measures of that form of deprivation); measuring major features of that deprivation (not conditions just experienced by a very small number of people or areas);

up-to-date; capable of being updated on a regular basis; statistically robust; and available for the whole of England at a small area level in a consistent form." (Office of the Deputy Prime Minister, 2004).

The next subsection provides a list of the actual indicators included in each domain.

Weighting of domains for constructing the overall index

The table below sets out the Domain weights which were used to combine the Domains into an Index of Multiple Deprivation.

Domain Weights for the IMD 2004	
Income deprivation	22.5%
Employment deprivation	22.5%
Health deprivation and disability	13.5%
Education, skills and training deprivation	13.5%
Barriers to housing and services	9.3%
Crime	9.3%
Living Environment deprivation	9.3%

Methodology

The following is an example of the methodology for constructing measures of deprivation for local areas. The methodology adopted by Gordon (1995) can be summarised as the estimation of a weighting system of a certain set of variables from the survey, and then the application of this system to the same set of variables in the Census in order to estimate the percentage of poor households (or individuals) in each area (ward or sub-ward).

The construction of the weighting system involved three steps: (a) the selection of the indicators from the survey; (b) identification of a 'poverty line' based on those selected indicators and by means of discriminant analysis (Gordon and Townsend, 1990); and (c) estimation of the weights by means of logistic regression for those indicators present both in the survey and the Census.

(a) The 1990 Breadline Britain Survey questionnaire contained 44 questions designed to cover items related to deprivation. For each item, the respondents were asked whether or not that particular item was a necessity. Only the 32 items considered a necessity by 50% or more of the respondents were included in the analysis.

(b) Gordon and Townsend (1990) assumed that two groups exist, a generally smaller "multiplied deprived" group (the poor) and a larger group which suffers from less deprivation (the non-poor). Discriminant analysis based on those items was implemented to identify the two groups. Assuming a direct relationship between income and deprivation, the income level at which these two groups could be separated (discriminated) was taken as the 'poverty line'. Gordon (1995) applied the same method to the 32 items in the 1990 Breadline Britain Survey in order to find the 'poverty line' corresponding to a "deprivation index score of 3" (i.e. those households/individuals lacking 3 or more of socially perceived necessities).

(c) Finally the weights were estimated using logistic regression between the poor/non-poor dichotomy and a subset of indicator variables present in both the survey and the Census.

The main outcome of the using logistic regression methodology is to produce "relative weights" (W_i) for a set of covariates *i* from the survey. Each covariate is a dichotomy: presence or absence of certain attribute. Examples are whether or not a person is: unemployed, suffers from long-term illness, belongs to a specified social class, has a car, is in owner-occupied accommodation, is a lone-parent. Now assuming that in a small area (k) the census provides information on the percentage (C_{ki}) of the population having the particular attribute *i*, the proportion in poverty (P_k) in that area is estimated as:

$$P_k = \Sigma_i \left(W_i . C_{ki} \right)$$

This methodology, although quite straightforward and practical, and extensively used in the United Kingdom, has two main drawbacks: (i) it does not consider some small-area specific covariates in the discriminant analysis and in the simple logistic regression model; (ii) it does not provide a measure of statistical accuracy of the results, in particular an evaluation of the standard errors of the final estimates of deprivation at local area level. We will describe an alternative methodology in the next section.

Set of indicators included in each domain

The following description is taken from Office of the Deputy Prime Minister (2004).

Income Deprivation Domain

The purpose of this Domain is to capture the proportion of the population experiencing income deprivation in an area.

- o Adults and children in Income Support households (2001).
- o Adults and children in Income Based Job Seekers Allowance households (2001).
- o Adults and children in Working Families Tax Credit households whose equivalised income (excluding housing benefits) is below 60% of median before housing costs (2001).
- o Adults and children in Disabled Person's Tax Credit households whose equivalised income (excluding housing benefits) is below 60% of median before housing costs (2001).
- o National Asylum Support Service supported asylum seekers in England in receipt of subsistence only and accommodation support (2002).

In addition, an Income Deprivation Affecting Children Index and an Income Deprivation Affecting Older People Index were created.

Employment Deprivation Domain

This domain measures employment deprivation conceptualised as involuntary exclusion of the working age population from the world of work.

- o Unemployment claimant count (JUVOS) of women aged 18-59 and men aged 18-64 averaged over 4 quarters (2001).
- o Incapacity Benefit claimants women aged 18-59 and men aged 18-64 (2001).
- o Severe Disablement Allowance claimants women aged 18-59 and men aged 18-64 (2001).

- o Participants in New Deal for the 18-24s who are not included in the claimant count (2001).
- o Participants in New Deal for 25+ who are not included in the claimant count (2001).
- o Participants in New Deal for Lone Parents aged 18 and over (2001).

Health Deprivation and Disability Domain

This domain identifies areas with relatively high rates of people who die prematurely or whose quality of life is impaired by poor health or who are disabled, across the whole population.

- o Years of Potential Life Lost (1997-2001).
- o Comparative Illness and Disability Ratio (2001).
- o Measures of emergency admissions to hospital (1999-2002).
- o Adults under 60 suffering from mood or anxiety disorders (1997-2002).

Education, Skills and Training Deprivation Domain

This Domain captures the extent of deprivation in terms of education, skills and training in a local area. The indicators fall into two sub domains: one relating to education deprivation for children/young people in the area and one relating to lack of skills and qualifications among the working age adult population.

<u>Sub Domain: Children/young people</u>

- o Average points score of children at Key Stage 2 (2002).
- o Average points score of children at Key Stage 3 (2002).
- o Average points score of children at Key Stage 4 (2002).
- o Proportion of young people not staying on in school or school level education above 16 (2001).
- o Proportion of those aged under 21 not entering Higher Education (1999-2002).
- o Secondary school absence rate (2001-2002).

<u>Sub Domain: Skills</u>

- o Proportions of working age adults (aged 25-54) in the area with no or low qualifications (2001).

Barriers to Housing and Services Domain

The purpose of this Domain is to measure barriers to housing and key local services. The indicators fall into two sub-domains: 'geographical barriers' and 'wider barriers' which also includes issues relating to access to housing, such as affordability.

<u>Sub Domain: Wider Barriers</u>

- o Household overcrowding (2001).
- o LA level percentage of households for whom a decision on their application for assistance under the homeless provisions of housing legislation has been made, assigned to SOAs (2002).
- o Difficulty of Access to owner-occupation (2002).

<u>Sub Domain: Geographical Barriers</u>

- o Road distance to GP premises (2003).
- o Road distance to a supermarket or convenience store (2002).

- o Road distance to a primary school (2001-2002).
- o Road distance to a Post Office (2003).

Crime Domain

This Domain measures the incidence of recorded crime for four major crime themes, representing the occurrence of personal and material victimisation at a small area level.

- o Burglary (4 recorded crime offence types, April 2002-March 2003).
- o Theft (5 recorded crime offence types, April 2002-March 2003, constrained to CDRP level).
- o Criminal damage (10 recorded crime offence types, April 2002-March 2003).
- o Violence (14 recorded crime offence types, April 2002-March 2003).

The Living Environment Deprivation Domain

This Domain focuses on deprivation with respect to the characteristics of the living environment. It comprises two sub-domains: the 'indoors' living environment which measures the quality of housing and the 'outdoors' living environment which contains two measures about air quality and road traffic accidents.

<u>Sub-Domain: The 'indoors' living environment</u>

- o Social and private housing in poor condition (2001).
- o Houses without central heating (2001).

<u>Sub-Domain: The 'outdoors' living environment</u>

- o Air quality (2001).
- o Road traffic accidents involving injury to pedestrians and cyclists (2000-2002).

A glimpse of the results

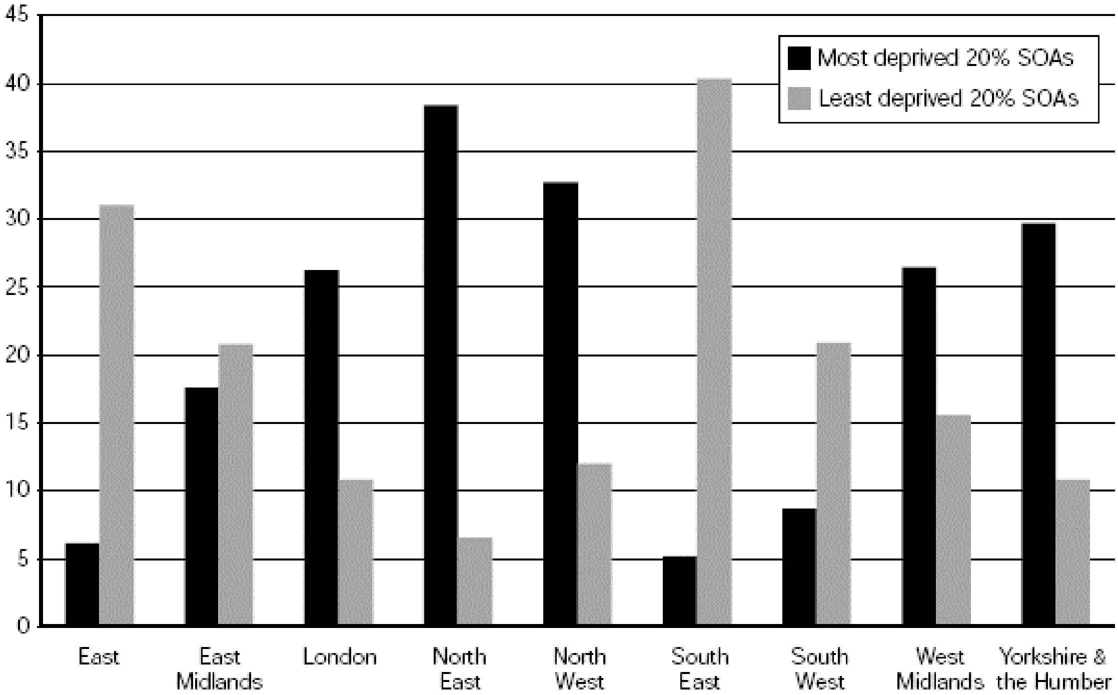
Indicators are developed for each of the 32,482 "Super Output Areas" (SOAs) in England, which are local areas with average population of under 2,000. These areas can be ranked according to the overall deprivation score. The relative position of larger regions can then be indicated by the percentage of local areas in the region which fall in the bottom 20% of all local areas in the country according to the overall deprivation score. (This is similar to the form of presentation employed in Section 9.6 above to identify disadvantaged groups.) Some results are shown in the table below.

Table 14.1 Number of SOAs in the most deprived 20% of SOAs in England on the IMD 2004, by Government Office Region

East	6.2
East Midlands	17.6
London	26.4
North East	38.1
North West	32.8
South East (excluding London)	5.1
South West	8.6
West Midlands	26.3
Yorkshire and the Humber	29.6
Total (England)	**20.0**

The chart below looks at both ends of the distribution: the percentages any region contains of the bottom and of the top 20% of the areas according to the overall deprivation index..

Figure 14.1 Percentage of SOAs in the most and least deprived 20% of SOAs in England on the IMD 2004 by Region.



A few highlights: SOA Level

England's most deprived 20% of SOAs have the following characteristics on average:

- o Just under a third of people are income deprived.
- o One in five of women aged 18-59 and men aged 18-64 are employment deprived.
- o Just under half of children live in families that are income deprived.
- o Just under a third of older people are income deprived.
- o The Region which has the greatest percentage of its SOAs that fall in England's most deprived 20% is the North East (38.1%), followed by the North West (32.8%). The North West has the greatest number of SOAs that fall in England's most deprived 20% (1461), followed by London with 1260.

Local Authority District Level

- o 80 districts fell into the 'most deprived 50' on one or more of the six district level summaries.
- o In the ID 2000, 81 districts fell into this category. Of the 81 districts in the 'most deprived 50' in the ID 2000, 75 remain within this category in the ID 2004.

14.3 SAE methodology used in Albania

We now describe and illustrate an alternative approach aimed at overcoming some drawbacks of the previous approach. The methodology combines census and survey information to produce finely disaggregated maps which describe the spatial distribution of poverty and inequality in the country.[40] The procedure is more demanding than the previous one in terms of the data required: access is required to micro data in the census, though no micro-level matching between survey and census data is involved.

The basic idea is to estimate a linear regression model with local (small area) variance components using information from the smaller and richer sample data - in the case of Albania the Living Standard Measurement Study, LSMS (World Bank, 2002) – in conjunction with aggregate information from a census – for Albania the 2001 Population and Housing Census – possibly supplemented by other sources. Disposable household income or consumption forms the dependent variable in the regression model. The model implies that *the estimated distribution of this variable can be used to generate the distribution for any sub-population in the census conditional to the sub-population's observed characteristics*. From the estimated distribution of the monetary variable in the census data set or in any of its sub-populations, an estimate can be made of a set of poverty measures, such as HCR, the Sen and the Foster-Green-Thorbecke indices, a set of inequality measures such as the Gini coefficient and general entropy measures, etc. To assess the precision of the estimates, standard errors of the poverty and inequality measures need to be computed using an appropriate procedure such as "bootstrapping".

Four important aspects of this methodology should be noted. Firstly, information from the Census is required at micro (household and individual) level; however micro-level linkage between Census and survey data is not required. Secondly, the vector of covariates utilised in the regression model implies that those variables have to be present in both sources. Thirdly and most importantly, the common variables in the sources must be sufficiently comparable; comparability requires the use of common concepts, definitions and measurement procedures. Moreover, especially in countries with rapid changes in living conditions – such as some of the New Member States in EU - it is important that reference periods for the data sets are as close as possible to each other.

This methodology is composed of the following stages: (a) comparison of each of the variable distributions from the Census with the corresponding weighted distribution from the survey; (b) the estimation of stratum-specific linear regression models with variance components on the basis of the survey, and then imputing expenditures in the Census data set; (c) simulation on expenditure, poverty/inequality indicators and relative standard error; (d) the estimation of geographically disaggregated measures of poverty and inequality.

(a) Are the census and the survey comparable?

The two sources of data need to be analysed in order to identify the common concepts and to construct the common variables to be compared. Also, where the survey variables contain missing values, it is normally necessary to impute them in order to avoid loss of

---

[40] The methodology was proposed by Elbers, Lanjouw and Lanjouw (2002, 2003), and adapted for application in Albania by Betti, Ballini and Neri (2003). The research was carried out under a World Bank project.

statistical units (and therefore of "degrees of freedom") in the estimation of the linear regression model with variance components. The imputation procedure used in this particular application was the same that used by Eurostat for ECHP. It is based on the "sequential regression multivariate imputation" (SRMI, Raghunathan *et al.*, 2001) approach as implemented in the imputation software IVE-ware.

Some details of the specific application in Albania are as follows. The original Census and survey variables were transformed to obtain 38 comparable variables. For each variable distributions from the Census were compared with the corresponding weighted distribution from the survey using a chi-square test. An important decision in such analyses concerns the choice of the variables to be included in the regression model as explanatory variables. According to the chi-square test, only 9 of 38 survey distributions fitted the census counterpart closely. There is a trade-off between the use of many explanatory variables (not highly comparable in the two sources) and the use of fewer better-fitting explanatory variables (but then loosing a part of the explanation of the variability in the dependent variable in the model). One way to reduce this problem is to make the variables "coarser" by reducing the number of categories in each, with the objective of obtaining closer distributions between the two sources. In this way all the 38 variables could be used in this case.

(b) Estimation of stratum-specific linear regression models with variance components for imputing expenditures

This stage consisted of developing an accurate empirical model of a logarithmic transformation of the household per-capita total consumption expenditure. Geographical differences in the level of prices were taken into account.

In the model the covariates are assumed to be defined in exactly the same way as in the smaller sample data and in the census. Denoting by $\ln y_{c,h}$ the logarithm consumption expenditure of household h in cluster c, a linear approximation to the conditional distribution of $\ln y_{c,h}$ is considered:

$$\ln y_{c,h} = E\left[\ln y_{c,h} \mid x_{c,h}^T\right] + u_{c,h} = x_{c,h}^T \beta + u_{c,h} \qquad (14.1)$$

where $x_{c,h}^T$ is the vector of covariates and $u_{c,h}$ is the error component.

Previous experience with survey analysis suggests that the proper model to be specified has a complex error structure, in order to allow for a within-cluster correlation in the disturbances as well as heteroschedasticity. To allow for a within cluster correlation in disturbances, the error component is specified as follows:

$$u_{c,h} = \eta_c + \varepsilon_{c,h} \qquad (14.2)$$

where η and ε are independent of each other and not correlated to the matrix of explanatory variables. Since residual location effects can greatly reduce the precision of poverty/inequality measure estimates, it is important to introduce some explanatory variables in the set of covariates which explain the variation in consumption expenditure due to location. For this reason the means of each covariate was introduced into the model; the means were calculated over all the census households in the 450 census enumeration areas which corresponded to the 450 primary sampling units selected in the survey.

Preliminary analyses on the Albanian LSMS suggested that the expenditure behaviour is locally different. In order to avoid forcing the parameter estimates to be the same for the whole country it was decided to estimate separate regression models for each of the following areas:

- o Coastal area (Stratum 1) – rural;
- o Coastal area (Stratum 1) – urban;
- o Central area (Stratum 2) – rural;
- o Central area (Stratum 2) –urban;
- o Mountain area (Stratum 3);
- o Tirana (Stratum 4).

The final results of this first stage are the GLS estimates of the selected model estimated on the LSMS data. In order to reach these final results a number of preliminary steps have to be performed.

In the Albanian example, firstly we verified whether or not weighting helps in the prediction model. The Hausman test implemented here (Deaton, 1997) led to the conclusion that weights need to be used only for Stratum 3. The initial estimate of β in equation (14.1) were obtained from OLS. The proportion of deviance explained by the model ranged between 0.56 and 0.64. With consistent estimate of β, the residuals from the regression were used as estimates of the overall disturbances $\hat{u}_{c,h}$.

The residual is decomposed into uncorrelated household and location components as follows: $\hat{u}_{c,h} = \hat{\eta}_c + e_{c,h}$. The estimated location components ($\hat{\eta}_c$) are the within-cluster means of the overall residual. The household component estimates ($e_{c,h}$) are the overall residual net of location components; these values are used to estimate the variance of $\varepsilon_{c,h}$.

To allow for heteroschedasticity in the household component, a model is chosen which best explains its variation. The covariates of this model can be the usual regressors as well as their squares or interactions between variables. The variance of $\varepsilon_{c,h}$ can be estimated with a logistic model conditional on a set of regressors; the variance of σ_η^2 can be estimated non-parametrically, allowing for heteroschedasticity in $\varepsilon_{c,h}$ (Appendix 2, Elbers, Lanjouw and Lanjouw, 2002). The two variance components are combined in order to calculate the estimated variance covariance matrix ($\hat{\Sigma}$) of the overall residual of the original model. Once $\hat{\Sigma}$ is calculated the original model can be estimated by GLS.

For each model in Albania (one for each stratum or sub-stratum), the significance of the cluster effect was tested by the Lagrange multiplier test for random effects, a test that Var (η_c) = 0; such a hypothesis was rejected at 5% level.

The estimated share of the location component with respect to the total residual variance is represented by $\rho = \sigma_\eta^2 / \sigma_u^2$. In our case, the ρ values ranged between about 4% to 21%, urban areas (Coastal area, Central area and Tirana) showed the lowest effect of the local component, the other rural (Coastal area and Central area) and Mountain areas showed a much more significant local component effect. The idea of estimating different models for each stratum or sub-stratum seems to be appropriate both in terms of local effect and in terms of covariates; in fact different subsets of covariates are significant for each model.

According to the results of the GLS regression, the significant parameter for each stratum/sub-stratum turned out to be the possession of a car. The other significant parameters in almost all the strata/sub-strata were the household size (as logarithmic transformation), the level of education, and the number of nonworking persons in the household. With regards to the possession of durable goods the most important factor was a refrigerator, followed by a TV and heater. Considering the EA mean variables it could be observed that the variables relating to the migration before 1990 and having a separate kitchen were significant in three of the six strata.

The results from this first step consist of a set of estimated GLS parameters for the regression coefficient $\hat{\beta}$, the associated variance covariance matrix and the disturbances at the cluster and the household level. As for the disturbances, attention is focused on their distribution. The results of the tests on normality (Shapiro-Wilk, Kolmogorov-Smirnov and Cramer-von Mises) let us reject the hypothesis of normality in almost all cases; only for the household residuals of the Central urban area did we fail to reject the null hypothesis.

(c) Simulation on expenditure, poverty/inequality indicators and relative standard error

The parameter estimates obtained from the previous step were applied to the census data so as to simulate the expenditure for each household in the census. A set of 100 simulation was conducted. For each simulation a set of the first stage parameters was drawn from their corresponding distribution simulated at the first stage: the beta coefficients, $\tilde{\beta}$, were drawn from a multivariate normal distribution with mean $\hat{\beta}$ (the coefficients of the GLS estimation) and variance covariance matrix equal to the one associated with $\hat{\beta}$. Relating to the simulation of the residual terms $\hat{\eta}_c$ and $e_{c,h}$, assumption of any specific distributional form was avoided by drawing directly from the estimated residuals: for each cluster the residual drawn is $\tilde{\eta}_c$ and for each household $\tilde{\varepsilon}_{c,h}$.

The simulated values are based on both the predicted logarithm of expenditure $x'_{c,h}\tilde{\beta}$, and on the disturbance terms $\tilde{\eta}_c$ and $\tilde{\varepsilon}_{c,h}$ using a bootstrap procedure:

$$\hat{y}_{c,h} = \exp\left(x^T_{c,h}\tilde{\beta} + \tilde{\eta}_c + \tilde{\varepsilon}_{c,h}\right). \tag{14.3}$$

The full set of simulated $\hat{y}_{c,h}$ values was used to calculate the expected value of each of the poverty measures considered. For each of the simulated consumption expenditure distributions a set of poverty and inequality measures was calculated, as was their mean and standard deviation over all the 100 simulations.

(d) construction of poverty and Inequality measures

From the estimated distribution of the monetary variable in the census data set or in any of its sub-populations, an estimate was made of a set of poverty measures based on the Foster-Greer-Thorbecke indices (for α =0,1,2), the Sen index and an absolute poverty line calculated using the information contained in the rich sample survey. In addition, a set of inequality measures based on the Gini coefficient, the Gini coefficient of the poor, and two general entropy (GE) measures (with parameter c=0,1) were estimated. Bootstrap standard errors of the welfare estimates were computed to assess the precision of the estimates.

The procedure for estimating the poverty and inequality measures was applied for the whole of Albania and disaggregated at seven levels:

- rural – urban level;
- the four strata used in sampling the LSMS;
- the six strata for which the linear regression models have been estimated;
- the 12 Prefectures;
- the 36 Districts;
- the 374 Communes/Municipalities;
- the 11 Mini-municipalities into which the city of Tirana is divided.

For any given location, the means constitute the point estimates, while the standard deviations are the bootstrap standard errors of these estimates.

Table 14.2 reports poverty and inequality measures and their bootstrap errors for the whole of Albania, and disaggregated at rural – urban level, by four strata (Regions) and by rural/urban type for the Coastal and Central regions (Stratum 1 and 2). The disaggregation into four strata is very useful for comparing these results to those obtained by LSMS and reported in Table 14.3 (Source: The World Bank, 2003). The census-based predictions are very consistent with those from LSMS: with the sole exception of the Head Count ratio in Stratum 1, in none of the four strata we are able to reject the null hypothesis that the estimate based on census is equal to the LSMS survey mean at 95% confidence level; in the case of the Head count ratio in the Stratum 1, we can reject the null hypothesis at 95% level, but not at 99% level.

According to both sources, Stratum 4 (Region of Tirana) is better off in terms of per capita consumption and percentage of individuals below the poverty line (head count), while in the Stratum 3 (Mountain area) there seems to be the highest proportion of poor individuals.

The last panel in Table 14.2 shows the disaggregation of Central and Coastal strata by rural/urban areas. According to the four poverty indices considered (Head count, FGT(1), FGT(2), Sen) the Mountain region is still the worst off, while in rural areas both in the Coastal and Central strata more than one third of the population is poor. However the region of Tirana shows, according to the Gini coefficient index and the two General Entropy indices used, higher inequality in the distribution of per capita consumption.

Table 14.2 Poverty and inequality indices (%), by urban – rural and strata. Albania

	Head count	FGT(1)	FGT(2)	Gini	Gini-poor	Sen	GE(0)	GE(1)	Con*
ALBANIA	28.60	6.96	2.48	29.54	12.38	5.29	14.28	15.05	7,569.67
	(1.28)	(0.44)	(0.19)	(0.52)	(0.27)	(0.40)	(0.53)	(0.77)	(120.21)
RURAL	36.26	9.06	3.27	27.72	12.57	7.45	12.65	13.74	6,586.25
	(2.18)	(0.73)	(0.32)	(0.72)	(0.33)	(0.76)	(0.74)	(1.35)	(190.99)
URBAN	18.09	4.08	1.40	28.94	11.74	2.78	13.78	14.02	8,919.82
	(0.86)	(0.27)	(0.12)	(0.54)	(0.33)	(0.22)	(0.53)	(0.57)	(170.92)
STRATUM 1	26.64	6.48	2.32	31.57	12.40	4.83	16.36	17.68	8,148.48
	(1.94)	(0.65)	(0.28)	(1.15)	(0.40)	(0.58)	(1.24)	(1.88)	(249.18)
STRATUM 2	29.49	7.00	2.43	27.35	11.94	5.36	12.16	12.43	7,177.76
	(2.32)	(0.76)	(0.33)	(0.54)	(0.46)	(0.71)	(0.51)	(0.52)	(222.95)
STRATUM 3	40.85	10.98	4.20	27.40	13.56	9.43	12.25	12.41	6,181.78
	(1.60)	(0.63)	(0.31)	(0.55)	(0.34)	(0.67)	(0.51)	(0.52)	(120.69)
STRATUM 4	18.01	4.11	1.42	29.35	11.88	2.80	14.18	14.54	8,981.39
	(1.09)	(0.38)	(0.17)	(0.63)	(0.52)	(0.30)	(0.65)	(0.70)	(140.85)
Stratum 1 urban	15.63	3.80	1.40	30.25	12.81	2.54	15.34	15.19	9,935.96
	(1.84)	(0.61)	(0.28)	(0.94)	(0.78)	(0.47)	(1.04)	(1.00)	(467.70)
Stratum 1 rural	34.84	8.47	3.00	28.87	12.23	6.89	13.97	16.54	6,816.09
	(3.22)	(1.09)	(0.46)	(1.74)	(0.48)	(1.09)	(1.94)	(3.87)	(283.67)
Stratum 2 urban	19.48	4.08	1.29	26.34	10.59	2.79	11.28	11.42	8,168.94
	(1.56)	(0.48)	(0.20)	(0.64)	(0.50)	(0.39)	(0.56)	(0.58)	(163.15)
Stratum 2 rural	34.41	8.43	2.99	27.02	12.25	6.80	11.87	12.23	6,689.88
	(3.45)	(1.13)	(0.48)	(0.86)	(0.53)	(1.14)	(0.80)	(0.85)	(327.31)

*Con stands for percapita consumption expenditure; standard errors are in parenthesis.

Table 14.3 Head Count Ratio and Per-capita Consumption: comparison between LSMS and Census*. Albania

	Head count LSMS	Head count Census	Consumption LSMS	Consumption Census
ALBANIA	25.39	28.60	7,800.82	7,569.67
	(1.32)	(1.28)	(117.68)	(120.21)
STRATUM 1	20.60	26.64	8,419.25	8,148.48
	(2.22)	(1.94)	(218.07)	(249.18)
STRATUM 2	25.57	29.49	7,496.12	7,177.76
	(2.32)	(2.32)	(193.63)	(222.95)
STRATUM 3	44.54	40.85	6,168.34	6,181.78
	(2.51)	(1.60)	(149.86)	(120.69)
STRATUM 4	17.82	18.01	9,042.59	8,981.39
	(2.06)	(1.09)	(304.96)	(140.85)

* Standard errors are in parentheses. Those for LSMS are estimated according to Levinson (2001).

Poverty and inequality measures disaggregated at Prefecture, District and Commune/Municipality levels are not reported in this report for sake of space. Anyway it is possible to report that both poverty and inequality are very spatially heterogeneous among Prefectures. In the Prefecture of Vlore there is the highest per capita consumption and the lowest percentage of poor people (18,3%), whereas according to the Gini coefficient consumption is very concentrated (33,5%). On the other hand, the Prefecture of Diber seems to be the worst off with only 6125 lek per month of per capita consumption, and the highest percentage of poor individuals (42,8%). Table 14.4 reports the measure disaggregated at Mini-municipality (for the City of Tirana) level.

Table 14.4 Poverty and inequality indices by Mini-Municipality of Tirana City (%)

Mini municipality	Head count	FGT(1)	FGT(2)	Gini	Gini-poor	Sen	GE(0)	GE(1)	Con
1	16.92	3.81	1.30	28.33	11.68	2.55	13.24	13.46	8,963.78
	(1.26)	(0.43)	(0.19)	(0.73)	(0.66)	(0.33)	(0.72)	(0.78)	(176.52)
2	15.76	3.48	1.17	29.66	11.40	2.29	14.44	14.76	9,510.04
	(1.21)	(0.38)	(0.17)	(0.70)	(0.59)	(0.29)	(0.71)	(0.73)	(178.42)
3	14.89	3.18	1.06	27.71	11.19	2.09	12.63	12.94	9,133.99
	(1.17)	(0.35)	(0.15)	(0.83)	(0.59)	(0.27)	(0.79)	(0.93)	(201.79)
4	20.50	4.70	1.64	27.37	11.95	3.31	12.35	12.43	8,206.32
	(1.32)	(0.45)	(0.21)	(0.58)	(0.61)	(0.37)	(0.57)	(0.55)	(158.31)
5	11.48	2.42	0.80	29.68	11.12	1.53	14.46	14.83	10,424.63
	(0.90)	(0.25)	(0.11)	(0.73)	(0.57)	(0.18)	(0.72)	(0.79)	(198.43)
6	24.76	5.72	1.98	28.52	11.84	4.20	13.44	14.41	7,806.92
	(2.14)	(0.67)	(0.28)	(1.70)	(0.55)	(0.59)	(1.77)	(2.72)	(328.50)
7	15.71	3.59	1.26	28.97	12.07	2.40	13.87	14.10	9,308.23
	(1.10)	(0.35)	(0.16)	(0.61)	(0.63)	(0.27)	(0.62)	(0.63)	(192.80)
8	15.74	3.45	1.17	28.25	11.48	2.29	13.13	13.35	9,137.55
	(1.21)	(0.37)	(0.17)	(0.61)	(0.66)	(0.29)	(0.60)	(0.62)	(202.50)
9	16.01	3.71	1.32	29.59	12.30	2.50	14.48	14.68	9,423.56
	(1.18)	(0.40)	(0.19)	(0.68)	(0.73)	(0.31)	(0.71)	(0.72)	(186.16)
10	9.06	1.67	0.49	27.96	9.51	1.00	12.72	13.12	10,541.28
	(1.02)	(0.26)	(0.09)	(0.73)	(0.59)	(0.17)	(0.66)	(0.75)	(297.74)
11	30.60	7.57	2.74	27.86	12.61	5.91	12.71	12.96	7,115.61
	(2.22)	(0.85)	(0.40)	(0.79)	(0.63)	(0.79)	(0.75)	(0.75)	(175.10)

* Standard errors are in parentheses.

14.4 SAE in Albania: poverty and inequality mapping

The poverty and inequality maps are reported in the Figures 14.2-14.7. We note that poverty is spatially heterogeneous among Municipalities within the same District, but not very much among Districts within the Prefecture to which they belong.

For instance the Prefecture of Berat shows an headcount ratio of 26,4%, whereas the three Districts within the Prefecture range from 22,9% (Skaprar) to 27,8% (Berat); the Prefecture of Gijrokaster shows an headcount ratio of 19,4%, whereas the three Districts within the Prefecture range from 18,3% (Permet) to 22,0% (Telepene). On the other hand, the

District of Kukes shows an headcount of 40,6%, whereas the 15 Municipalities within the District range from 21,4% to 79,5%.

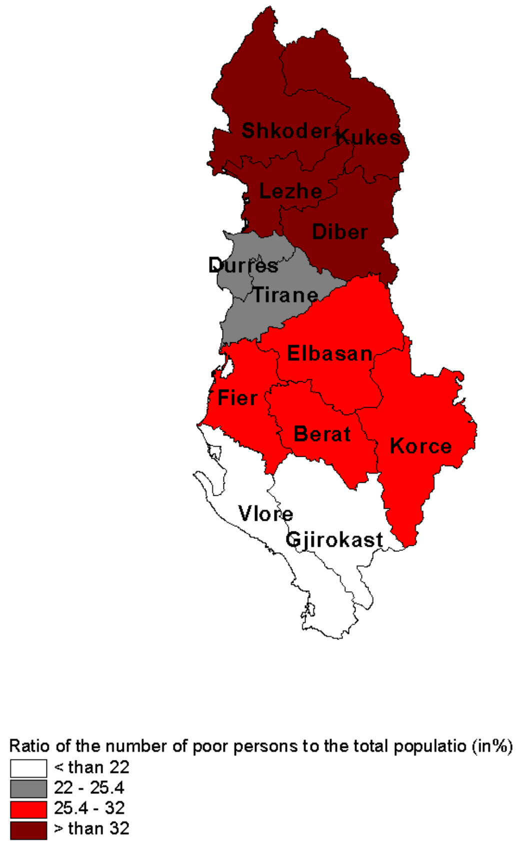
Two thirds of Prefectures have both headcount and per-capita consumption estimates that are significantly different from the corresponding values at national level. In contrast with this, less than 20% of Districts have the same estimates that are significantly different from the Prefecture to which they belong. Finally, at Municipality level, more than 40% of Municipalities have headcount and per-capita consumption measures that are significantly different from the District to which they belong.

In conclusion, in this chapter we have presented a methodology for estimating various measures of poverty and inequality for small administrative units.

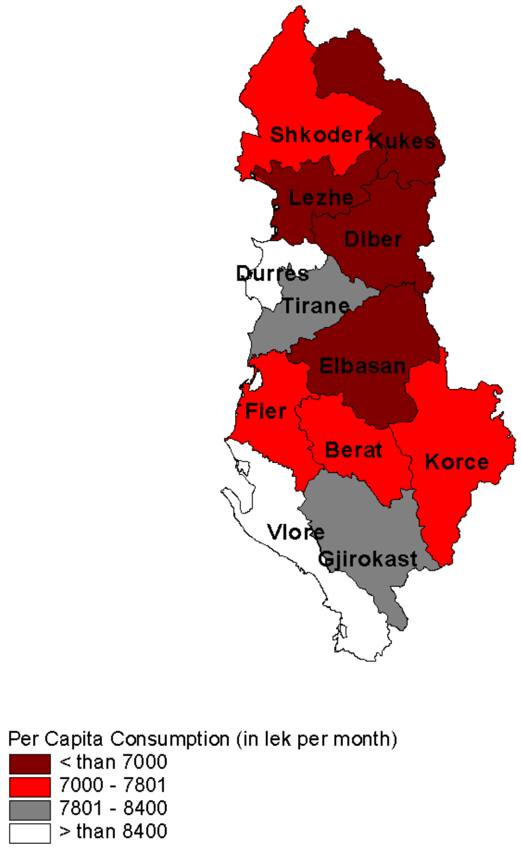
The main findings of research like the present one are potentially very useful for policy-makers. We find, for instance, that in Albania there is considerable heterogeneity of poverty rates across administrative units. The particular spatial pattern of this heterogeneity has important policy implications for poverty alleviation programmes: at the highest level we observe a large spatial heterogeneity among Prefectures; this spatial heterogeneity is much less pronounced among Districts within the same Prefecture; however, it is pronounced again at the lowest level among Municipalities within the same District. What this means for the practitioner and the policymaker is that it is important to disaggregate down to the Commune level when analysing issues and planning interventions, as this will add substantially in terms of precision of the targeting of resources when compared to stopping at the District level.

The quality of the modelling has benefited from the fact that the LSMS survey in Albania followed the Population Census *closely in time as well as in the basic concepts and definitions used*. We would strongly recommend such co-ordination between data sources for the production of useful small area estimates, especially in the situation of countries subject to rapid changes in living conditions, such as some of the New Member States and Candidate countries.

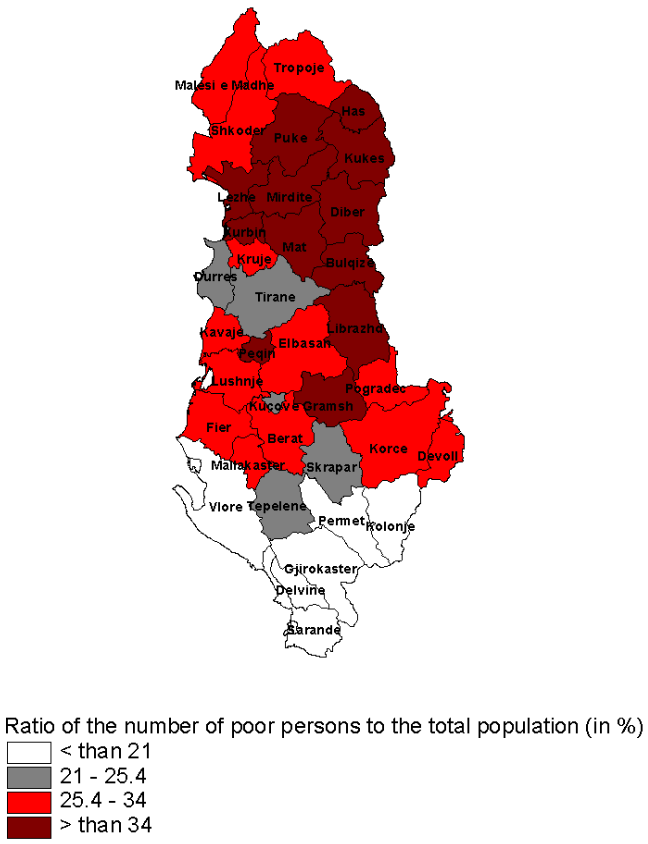
Figures 14.2 Head Count Ratio by Prefectures. Albania



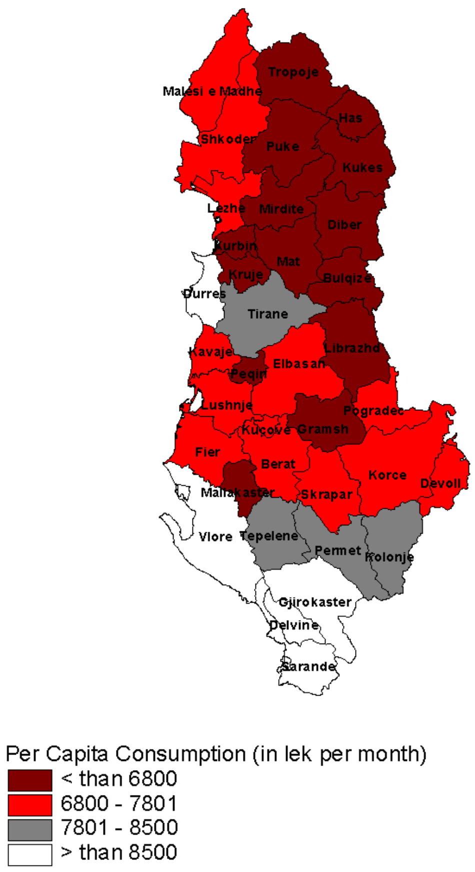
Figures 14.3 Per Capita Consumption by Prefectures. Albania



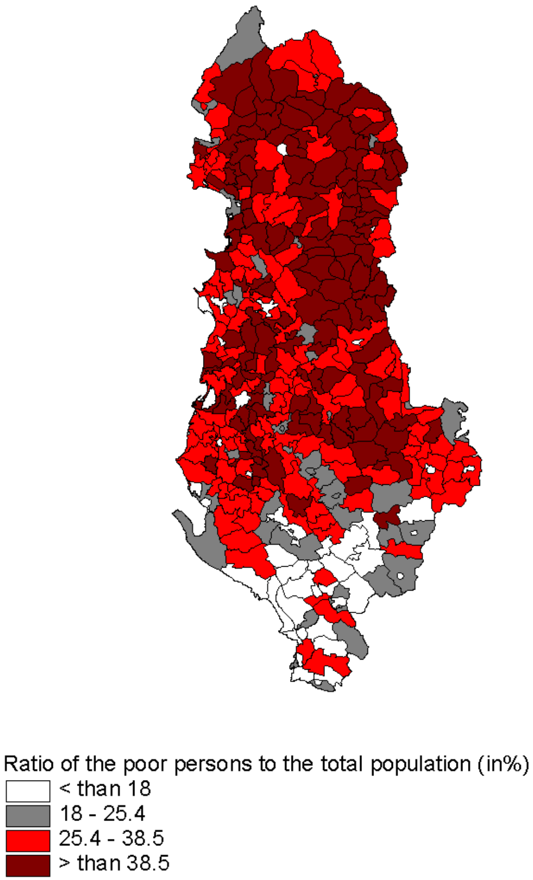
Figures 14.4 Head Count Ratio by District. Albania



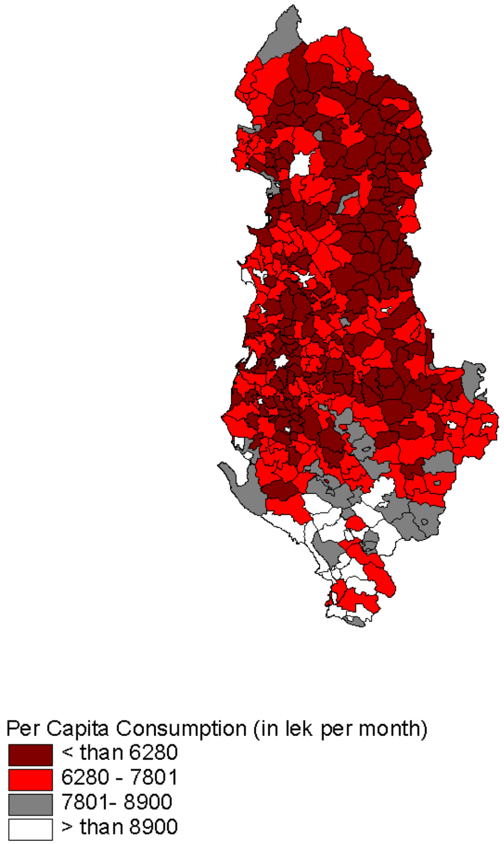
Figures 14.5 Per Capita Consumption by District. Albania



Figures 14.6 Head Count Ratio by Municipality. Albania



Figures 14.7 Per Capita Consumption by Municipality. Albania



Chapter 15
STATISTICAL ANNEX

This STATISTICAL ANNEX provides values of the regional indicators on poverty and deprivation and their standard error where applicable. These represent the final estimates which have been generated using small area estimation (SAE) methodology described in this report. It uses micro-level survey values in conjunction with supplementary information - in the present application from NewCronos tables - where both are available, to generate composite estimates. An attempt has been made to cover up to NUTS2 level all EU Member States (EU25), as well as Candidate countries. For Italy NUTS3 level estimates are also provided for poverty rates and income level.

15.1 Statistics produced

(1) We begin with direct estimates from surveys at the country level for the target variables listed in Table A.1 below. Table A.2 shows which of these are available by country, and provides the directly estimated values.

(2) NUTS1 estimates of the (NUTS1/Country) ratio are provided for each target variable meeting the following conditions:

1. The variable is available at the country level.
2. The country is divided into more than one NUTS1 regions.
3. Code is available in the survey micro data for the identification of NUTS 1 regions.
4. There is some sample data in the survey for the region concerned.
5. The necessary covariates are available in NewCronos (or an alternative source).
6. It is possible to produce statistically valid model estimates for the variable.
7. And that it has been possible to produce such estimates in the time and resources available.

Standard error for each estimated (NUTS1/Country) ratio are also produced by the model. Generally these are smaller than the corresponding standard errors for the direct (survey) estimates. Actual values of NUTS1 estimates and their standard errors can be obtained as explained below in Section 15.2.

Where conditions 3 and/or 4 are not met, but the other condition are satisfied, we have used the regression coefficients estimated above from the SAE procedure to produce more approximate regression-prediction estimates. This applies only to a small minority of NUTS1 regions.

(3) NUTS2 estimates of the (NUTS2/NUTS1) ratio are provided for each target variable meeting conditions similar to the above.

Again, where conditions 3 and/or 4 are not met, but the other conditions are satisfied, we have used the regression coefficients estimated above from the corresponding SAE procedure to produce more approximate regression-prediction estimates. Unfortunately this applies to a large proportion of NUTS2 regions, primarily because of the non-availability of regional codes in the survey data from ECHP.

Standard error for each estimated (NUTS2/NUTS1) ratio are also produced by the model. Generally these are substantially smaller than the corresponding standard errors for the direct (survey) estimates. Actual values of NUTS2 estimates and their standard errors can be obtained as explained below in Section 15.2.

(4) NUTS3 estimates of the (NUTS3/NUTS2) ratio for a small subset of the most important target variables have been produced for NUTS3 regions in Italy, and also the corresponding standard errors. This has been possible because of the provision by ISTAT of the necessary ECHP-PDB data under a special research co-operation agreement with University of Siena. The available covariates from NewCronos at NUTS3 level are also very limited, so that the results of the modelling should be interpreted with caution.

(5) Going below NUTS3 regions is not possible on the basis of the adopted SAE methodology. For this a different type of modelling – which is generally less precise - is required. We have provided some illustrations from the UK and Albania.

15.2 Using the statistical tables

Let y_c be the value of the estimate of a target variable y for country c as given in Table A.2, and y_{ci} the required estimate for it for NUTS1 region i of the country. Table A.3 provides the composite estimate of the ratio $(y_{ci}/y_c) = r_{ci}$, say, and Table A.4 provides the associated standard error

$$se(y_{ci}/y_c) = se(r_{ci}).$$

For consistency with given country-level estimates, the regional estimates of the ratio r_{ci} have been 'raked'. The procedure is straightforward as follows. Let W_{ci} be the weights taken as *proportional to the regional population* given in each table, and s_{ci} the 'unraked' estimates of the ratio as produced by the model or regression-prediction. The corrected or raked estimate for the ratio is

$$r_{ci} = f_c.s_{ci} \quad \text{with} \quad f_c = \Sigma_i W_{ci} / \Sigma_i (W_{ci}.s_{ci}).$$

In any case, the above correction is small in most cases; however, it is prudent to have included it especially for the regression-prediction estimates where inconsistencies can arise in practice.

(1) The required actual statistics for the region is estimated by

$$y_{ci} = y_c.r_{ci} = y_c.f_c.s_{ci}$$

where y is the value of the country-level estimate for any variable (1-13) as shown in Table A.2, and r_{ci} are the corresponding (NUTS1/Country) ratios shown in Table A.3. Note that the 'raking' ensures that $\Sigma_i (W_{ci}.y_{ci}) / \Sigma_i W_{ci} = y_c$, as required.

(2) Standard error of the actual statistic can be approximated from the given tables as follows, following the procedure described in Chapter 11:

$$se(y_{ci}) = y_c.\left(\frac{n_c}{n_c - n_{ci}}\right)^{1/2}.se(r_{ci}),$$

where n_{ci} is the sample size for region i in country c, and n_c is its total for the country. (This formula ignores the effect of raking on the resulting sampling error.) Standard errors for the ratio, $se(r_{ci})$ and the corresponding sample sizes (in terms of total number of households summed up over the available waves of the panel survey) are given in the tables below.

(3) The procedure of producing actual values of the estimates for NUTS2 regions from the (NUTS2/NUTS1) ratios is identical to that described above: we can simply replace the country values by the estimated NUTS1 values, and the NUTS1 quantities in the above by the corresponding NUTS2 values. The same applies to the procedure for obtaining approximate standard error for the statistic.

For instance, for NUTS2 region j in NUTS1 region i of country c,

$$y_{cij} = y_{ci}.r_{cij}; \quad se(y_{cij}) = y_{ci}.\left(\frac{n_{ci}}{n_{ci} - n_{cij}}\right)^{1/2}.se(r_{cij}); \quad n_{ci} = \Sigma_j n_{cij}.$$

Here r_{cij} is the estimated (NUTS2/NUTS1) ratio for each statistics (1-13) as shown in Table A.5, and $se(r_{cij})$ is its estimated standard error shown in Table A.6. The corresponding sample sizes n_{cij} are also given in these tables.

15.3 Tables

Table A.1 List of variables in the annex tables (A.2-A.7)

1	HCR_c	Head Count Ratio using country poverty lines
2	HCR_n2	Head Count Ratio using nuts2 poverty lines
3	logEqInc	Logarithm of the equivalised income
4	LogIncPC	Logarithm of the per capita income
5	FM_c	Fuzzy Monetary (defined at country level)
6	FS_c	Fuzzy Supplementary (defined at country level)
7	LAT_c	Latent deprivation (defined at country level)
8	MAN_c	Manifest deprivation (defined at country level)
9	Fsup1	Fuzzy supplementary dimension 1 (basic life-style deprivation)
10	Fsup2	Fuzzy supplementary dimension 2 (secondary life-style deprivation)
11	Fsup3	Fuzzy supplementary dimension 3 (housing amenities)
12	Fsup4	Fuzzy supplementary dimension 4 (housing deterioration)
13	Fsup5	Fuzzy supplementary dimension 5 (environmental problems)

Table A.2 Target variable values - country level

	sample size*	Pop. mil.	Variable 1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
13 AT	20.647	8.012	12,5	12,6	9,51	9,07	11,8	13,4	21,2	4,0	12,0	10,0	3,8	8,7	13,3
4 BE	23.251	10.251	14,1	14,1	9,50	9,07	13,5	15,0	23,1	5,3	14,4	8,8	3,4	12,1	17,1
1 DE_1	22.057	35.271	12,4	10,8	9,44	9,05	12,7	12,9	21,4	4,2					
1 DE_2	25.387	45.855	12,5	13,8	9,51	9,11	12,1	9,7	18,9	2,9					
2 DK	24.070	5.337	10,3	10,3	9,55	9,15	9,2	11,2	17,6	2,8	9,9	8,3	1,2	8,1	11,5
11 ES	46.089	40.263	18,9	17,6	9,01	8,52	19,5	18,4	29,7	8,2	16,8	15,4	2,0	15,2	22,9
14 FI	25.275	5.176	10,1	10,0	9,33	8,93	8,9	10,9	16,7	3,0	12,5	7,0	2,3	4,7	12,6
6 FR	48.481	58.894	15,1	14,6	9,40	8,97	14,7	15,8	23,9	6,6	16,2	10,9	3,2	13,8	18,7
10 GR	35.921	10.918	21,5	20,2	8,88	8,43	22,2	22,6	34,6	10,2	33,7	20,2	6,1	16,5	21,8
8 IE	24.092	3.787	18,0	17,0	9,23	8,72	16,8	17,8	26,4	8,2	17,1	14,5	2,5	9,3	16,4
9 IT_n	32.728	38.510	10,4	15,4	9,31	8,88	11,8	14,0	21,6	4,2	14,2	11,2	2,0	8,8	20,6
9 IT_s	18.957	19.252	36,8	19,8	8,79	8,30	34,7	29,5	46,0	18,1	34,8	22,5	3,0	12,5	29,1
5 LU	20.361	436	12,1	12,1	9,96	9,53	11,6	11,1	18,7	4,0					
3 NL	44.811	15.926	11,2	11,1	9,40	8,99	11,4	12,0	18,9	4,5	10,6	7,1	0,9	10,6	13,6
12 PT	37.713	10.226	21,3	20,8	8,81	8,32	22,3	23,9	35,5	10,7	18,4	23,7	14,4	27,9	26,6
15 SE	30.735	8.872	10,9	11,1	9,34	8,97	10,0	10,9	17,9	3,0					
7 UK	38.682	58643,2	18,6	18,1	9,39	8,97	18,3	19,0	29,4	7,8	18,0	15,3	1,3	11,9	19,0
16 CY		694	15,0		4,30										
17 CZ		10.273	8,0		4,04										
18 EE		1.370	18,0		3,75										
19 HU		10.024	10,0		3,95										
20 LT		3.500	17,0		3,74										
21 LV		2.373	16,0		3,72										
22 MT		390	15,0		4,23										
23 PL	9.357	38.646	16,8	16,1	3,98	3,37	16,8	16,8	27,1	6,5	14,8	22,2		30,0	23,3
24 SI		1.990	10,0		4,18										
25 SK		5.401	21,0		3,94										
26 RO	31.547	22.442	14,1	12,3	3,95		14,1								
27 BG		8.171	16,0		3,55										
min	9.357	390	8,0	10,0	3,5	3,4	8,9	9,7	16,7	2,8	9,9	7,0	0,9	4,7	11,5
mean	29.482	16.583	15,3	14,6	7,1	8,6	15,4	15,8	24,9	6,4	17,4	14,1	3,5	13,6	19,0
max	48.481	58.894	36,8	20,8	10,0	9,5	34,7	29,5	46,0	18,1	34,8	23,7	14,4	30,0	29,1

Note to table A.2: Germany and Italy have been divided into two parts each

DE(2)= DE1, DE2, DE3, DE7, DE8, DE9, DEX

IT(S)= IT8, IT9, ITA, ITB

* sample size: number of households, aggregated over all available waves

All are direct estimates from ECHP for EU15, or from similar national surveys (PL and RO). For other countries, previously published figures have been reproduced.

Table A.3 Composite estimates of Ratio (NUTS1/Country)

Country	NUTS1	sample size*	Pop. mil.	Variable 1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
	SAE Model															
AT	AT1	9.183	3.365	0,94	1,07	1,005	1,007	0,93	1,14	1,03	1,12	0,90	1,01	1,47	1,04	1,33
AT	AT2	4.853	1.743	1,12	0,92	0,993	0,992	1,13	0,97	1,04	1,04	1,26	1,01	0,99	0,97	0,77
AT	AT3	6.611	2.904	1,00	0,97	0,999	0,997	1,00	0,86	0,95	0,84	0,96	0,98	0,47	0,97	0,76
BE	BE1	3.023	962	0,92	1,08	1,007	1,011	0,90	1,37	1,14	1,20	1,19	1,35	1,42	1,15	1,63
BE	BE2	9.738	5.946	0,96	0,99	1,001	1,000	0,97	0,84	0,92	0,83	0,80	0,86	0,89	0,86	0,80
BE	BE3	10.490	3.343	1,09	0,99	0,997	0,997	1,08	1,18	1,11	1,24	1,29	1,15	1,07	1,21	1,17
DE_1	DE4	6.010	2.601	0,90	1,00	0,993	0,992	1,06	1,33	1,20	1,15					
DE_1	DE5	6.065	662	1,23	0,89	0,999	1,010	0,88	1,00	0,99	0,68					
DE_1	DE6	2.093	1.710	1,02	1,05	1,007	1,020	0,87	1,13	0,99	1,03					
DE_1	DEA	2.164	18.000	0,89	1,00	1,006	1,006	0,90	0,71	0,81	0,80					
DE_1	DED	396	4.443	1,16	0,97	0,988	0,986	1,21	1,72	1,45	1,50					
DE_1	DEE	586	2.633	1,01	1,01	0,993	0,994	1,02	1,21	1,13	1,07					
DE_1	DEF	3.277	2.782	1,41	1,03	1,003	0,999	0,98	0,65	0,84	0,77					
DE_1	DEG	1.466	2.440	1,11	1,01	0,984	0,981	1,44	1,52	1,42	1,67					
DE_2	DE1	3.824	10.500	1,07	0,97	1,001	0,999	0,95	0,95	0,97	0,85					
DE_2	DE2	9.197	12.188	0,88	0,97	1,005	1,005	0,88	0,92	0,91	0,89					
DE_2	DE3	3.946	3.384	0,85	1,12	1,001	1,008	1,09	1,64	1,33	1,39					
DE_2	DE7	2.385	6.058	0,98	1,08	1,005	1,007	0,98	1,03	0,99	1,02					
DE_2	DE8	1.043	1.783	1,54	1,01	0,983	0,982	1,33	1,31	1,25	1,46					
DE_2	DE9	2.396	7.911	0,98	1,00	0,995	0,995	1,10	0,85	0,98	1,07					
DE_2	DEX	2.596	4.030	1,13	0,97	0,991	0,988	1,11	0,94	1,02	1,05					
ES	ES1	6.614	4.286	1,00	0,95	0,992	0,989	1,04	1,11	1,07	1,10	0,99	1,22	1,30	1,70	0,81
ES	ES2	7.013	4.089	0,58	0,93	1,015	1,018	0,64	0,65	0,70	0,47	0,58	0,65	0,87	0,76	0,81
ES	ES3	4.120	5.230	0,43	1,12	1,037	1,039	0,48	0,82	0,69	0,50	0,57	0,49	0,43	0,58	1,34
ES	ES4	7.348	5.255	1,47	0,96	0,979	0,978	1,40	0,95	1,16	1,25	1,02	1,25	1,30	1,30	0,65
ES	ES5	10.070	11.129	0,68	0,98	1,014	1,017	0,73	0,95	0,86	0,76	0,88	0,80	0,88	0,79	1,16
ES	ES6	8.109	8.568	1,53	1,03	0,975	0,971	1,45	1,15	1,26	1,45	1,36	1,37	0,98	0,96	0,92
ES	ES7	2.815	1.707	1,74	1,06	0,971	0,967	1,60	1,86	1,62	2,07	2,22	1,48	2,29	1,72	1,32
FR	FR1	7.629	11.002	0,65	1,15	1,024	1,025	0,65	1,10	0,91	0,78	0,84	1,01	0,83	0,96	1,43
FR	FR2	9.017	10.493	1,00	0,95	0,996	0,996	1,02	1,01	1,01	1,02	1,03	1,00	1,05	1,05	0,94
FR	FR3	3.125	4.010	1,43	1,03	0,984	0,980	1,38	1,38	1,30	1,54	1,36	1,20	1,69	1,28	1,12
FR	FR4	4.482	5.189	0,81	0,94	1,001	0,999	0,86	0,79	0,83	0,84	0,73	0,86	0,91	0,88	0,89
FR	FR5	7.301	7.835	1,11	0,93	0,992	0,992	1,11	0,88	1,00	1,00	1,00	0,98	0,85	1,04	0,75
FR	FR6	5.676	6.220	1,21	0,99	0,992	0,994	1,19	0,92	1,06	1,04	1,13	0,99	1,31	0,93	0,68
FR	FR7	5.516	7.008	0,99	0,97	0,998	0,997	0,98	0,95	0,95	1,00	0,93	0,98	0,89	0,93	1,02
FR	FR8	5.735	7.137	1,14	1,00	0,995	0,996	1,13	1,03	1,07	1,09	1,16	1,03	0,87	0,99	0,98

* sample size: number of households, aggregated over all available waves

(table continued)

Notes to Table A.3.

1. The table gives estimates of ratio (NUTS1/Country) for each statistic (1-13). Actual NUTS1 value of the statistic can be obtained by multiplying this ratio with the corresponding country-level estimate from Table A.2.

2. For countries for which we have survey micro-data, and NUTS1 regions are identified in the data, composite SAE estimates have been produced. However, if there is no NUTS1 code in the data (e.g., NL and FI in ECHP), or no survey data are available to us (as HU), simple regression (with coefficients taken from the corresponding SAE model) has been used to predict the required values.

Table A.3 (cont)

				Variable												
Country	NUTS1	sample	Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13
		size	mil.	HCR_c	HCR_n2	logEqInc	logIncPC	FM_c	FS_c	LAT_c	MAN_c	FSup1	FSup2	FSup3	FSup4	FSup5
GR	GR1	12.215	3.512	1,34	1,07	0,985	0,984	1,29	0,97	1,10	1,23	1,05	1,15	1,32	0,87	0,81
GR	GR2	9.164	2.425	1,33	1,07	0,984	0,982	1,28	0,98	1,09	1,25	1,12	1,22	1,34	1,32	0,63
GR	GR3	9.847	3.885	0,48	0,88	1,025	1,027	0,56	1,05	0,87	0,62	0,87	0,73	0,31	0,76	1,51
GR	GR4	4.695	1.095	1,01	1,05	0,996	0,995	1,01	0,95	0,97	1,04	1,03	1,00	1,66	1,56	0,60
IT_n	IT1	5.176	6.033	0,96	1,00	1,002	1,006	0,98	1,08	1,03	1,07	0,92	1,04	1,06	1,05	1,15
IT_n	IT2	5.804	9.094	0,81	1,01	1,010	1,011	0,80	1,03	0,95	0,80	0,89	1,04	0,84	0,87	1,17
IT_n	IT3	6.494	6.653	0,87	0,98	1,003	1,001	0,92	0,83	0,87	0,85	0,89	0,94	1,16	1,05	0,66
IT_n	IT4	2.999	3.995	0,73	0,96	1,011	1,015	0,75	0,76	0,76	0,75	0,84	0,73	0,78	0,74	0,89
IT_n	IT5	5.742	5.845	1,03	0,95	0,996	0,996	1,06	0,96	0,99	1,06	1,08	1,03	0,98	1,07	0,85
IT_n	IT6	3.465	5.283	1,51	1,11	0,981	0,976	1,42	1,36	1,37	1,43	1,42	1,17	1,24	1,24	1,32
IT_n	IT7	3.048	1.608	1,63	0,93	0,976	0,969	1,57	0,84	1,15	1,45	1,11	0,87	0,84	0,95	0,66
IT_s	IT8	4.665	5.782	0,92	0,99	1,004	1,003	0,93	0,92	0,95	0,88	0,81	0,96	0,88	1,04	1,04
IT_s	IT9	6.929	6.738	1,00	0,98	1,000	1,001	1,00	1,09	1,03	1,09	1,11	1,03	1,32	0,97	1,05
IT_s	ITA	4.215	5.082	1,07	1,01	0,997	0,997	1,06	0,97	1,01	1,02	1,03	0,99	0,62	0,94	0,97
IT_s	ITB	3.148	1.650	1,06	1,07	0,996	0,995	1,05	1,01	1,04	1,01	1,11	1,06	1,32	1,15	0,78
PT	PT1	28.373	9.749	0,97	1,00	1,001	1,002	0,97	1,01	0,99	0,99	0,96	0,99	1,00	1,00	1,03
PT	PT2	4.758	237	1,68	0,99	0,970	0,963	1,54	0,75	1,13	1,13	1,87	1,23	0,81	1,10	0,30
PT	PT3	4.582	240	1,60	0,99	0,970	0,965	1,48	0,81	1,11	1,24	1,76	1,30	1,11	0,77	0,43
UK	UK1	2.426	3.011	1,01	0,91	0,997	0,998	1,03	0,94	0,99	0,97	0,97	0,99	0,52	1,00	0,91
UK	UK2	3.717	4.950	1,27	0,99	0,990	0,989	1,22	0,97	1,07	1,21	0,93	1,02	0,97	0,91	1,06
UK	UK3	3.281	4.157	1,19	0,98	0,991	0,990	1,16	1,04	1,06	1,22	0,93	1,06	0,71	0,96	1,04
UK	UK4	1.708	2.174	1,14	0,99	0,993	0,993	1,13	0,82	1,00	0,88	0,82	1,13	1,10	1,05	0,70
UK	UK5	10.963	18.288	0,77	1,06	1,012	1,013	0,79	1,05	0,95	0,82	1,08	0,96	1,16	1,12	1,12
UK	UK6	3.443	4.909	0,95	0,94	0,999	1,000	0,97	0,92	0,94	0,93	0,99	1,06	1,19	0,81	0,78
UK	UK7	3.390	5.260	1,23	1,01	0,990	0,986	1,19	0,97	1,06	1,16	1,00	0,94	0,57	0,88	1,00
UK	UK8	4.043	6.249	1,04	1,00	1,000	1,001	1,02	1,03	1,01	1,10	0,96	1,06	1,06	1,09	1,05
UK	UK9	2.090	2.900	1,06	0,89	0,994	0,993	1,09	1,03	1,04	1,13	1,17	1,00	0,76	0,90	0,92
UK	UKA	3.621	5.063	1,08	0,98	0,994	0,994	1,08	1,00	1,04	1,06	0,92	0,98	1,26	0,90	0,85
PL	PL1	1.115	7.717	1,01	1,00	0,999	1,000	1,02	1,02	1,02	1,04	1,02	1,01		0,98	1,00
PL	PL2	1.178	8.084	1,00	0,99	1,000	0,999	1,01	1,01	1,02	1,03	1,03	1,01		1,01	1,01
PL	PL3	2.341	6.907	1,00	1,00	1,000	1,000	1,00	1,00	1,00	1,00	0,99	0,99		0,99	1,03
PL	PL4	1.799	6.115	1,05	1,02	0,999	0,999	1,06	1,06	1,05	1,07	1,04	1,05		1,05	1,01
PL	PL5	1.123	4.062	0,96	0,97	1,001	1,002	0,97	0,95	0,97	0,95	0,99	0,90		0,94	0,97
PL	PL6	1.801	5.762	1,03	1,01	0,999	0,999	1,03	1,02	1,03	1,03	1,02	1,02		1,01	1,01
	Regression-prediction															
FI	FI1		5.151	1,00	1,00	1,000	1,000	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
FI	FI2		26	1,11	0,88	0,992	0,990	1,09	1,06	1,10	1,04	1,20	0,97	0,89	0,98	1,01
FR	FR9		1.699													
NL	NL1		1.664	1,02	0,93	0,997	0,998	1,02	1,08	1,05	1,05	0,93	0,99	0,79	0,85	1,16
NL	NL2		3.332	1,11	0,96	0,993	0,991	1,10	0,97	1,02	1,07	0,99	1,01	0,97	0,99	0,86
NL	NL3		7.422	0,95	1,06	1,004	1,005	0,96	1,05	1,01	0,98	1,07	1,03	1,17	1,11	1,08
NL	NL4		3.508	0,99	0,96	0,999	0,998	0,99	0,89	0,94	0,95	0,91	0,93	0,76	0,85	0,90
HU	HU1		2.838	0,70		1,021										
HU	HU2		3.063	0,94		0,998										
HU	HU3		4.124	1,25		0,987										
UK	UKB		1.683	1,22	0,87	1,014	1,007	1,22	0,87	1,06	1,01	0,82	0,94	0,67	0,96	0,75

3. Only countries which have been divided into 2 or more NUTS1 regions appear in the table. Hence not included here are: DK, LU, IE, SE from EU15; all except PO and HU from NMS10; and BG and RO each of which comprises only one NUTS1 region.

4. In the case of DE and IT, the country has been divided into two parts, and for the application of (1), each part is treated separately. Thus for example, for region DE1, we have (NUTS1/Country) ratio for HCR_C as 1.04, and for the 'country' (meaning DE_2), HCR_C is 12.5. This gives actual HCR_C for region DE1 as 1.04*12.5=13.0.

Table A.4 Percentage standard error of composite estimates of Ratio (NUTS1/Country)

				Variable												
Country	NUTS1	sample size*	Pop. mil.	1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
	SAE Model			Percentage standard error												
AT	AT1	9.183	3.365	3,7	3,4	0,07	0,08	3,8	4,1	2,6	10,1	3,7	4,6	14,8	5,3	4,8
AT	AT2	4.853	1.743	4,6	3,6	0,09	0,11	4,8	4,1	3,1	9,2	5,2	5,1	9,3	5,5	3,9
AT	AT3	6.611	2.904	4,0	3,3	0,08	0,09	4,2	3,4	2,6	7,5	4,1	4,7	6,0	5,1	3,2
BE	BE1	3.023	962	5,2	4,4	0,14	0,15	5,3	5,4	3,9	9,7	5,4	8,1	14,7	6,0	5,2
BE	BE2	9.738	5.946	4,1	3,5	0,09	0,09	4,3	3,4	2,6	7,4	3,4	5,8	11,8	4,2	3,0
BE	BE3	10.490	3.343	3,3	2,7	0,06	0,07	3,4	3,5	2,3	8,3	4,0	5,7	10,0	4,2	3,1
DE_2	DE1	3.824	10.500	7,9	4,8	0,20	0,22	9,5	11,0	7,0	17,8					
DE_2	DE2	9.197	12.188	16,6	4,6	0,18	0,20	8,1	9,6	6,0	16,3					
DE_2	DE3	3.946	3.384	14,1	5,6	0,22	0,25	9,9	13,6	8,0	21,7					
DE_1	DE4	6.010	2.601	8,6	5,6	0,18	0,20	7,6	7,9	5,6	13,9					
DE_1	DE5	6.065	662	8,0	6,4	0,49	0,55	15,1	14,7	11,2	21,3					
DE_1	DE6	2.093	1.710	7,7	5,3	0,41	0,46	13,0	12,7	9,6	18,4					
DE_2	DE7	2.385	6.058	11,0	5,0	0,20	0,22	8,8	10,5	6,6	18,5					
DE_2	DE8	1.043	1.783	9,7	5,3	0,25	0,28	10,7	11,4	7,8	20,3					
DE_2	DE9	2.396	7.911	9,7	4,5	0,19	0,21	9,5	9,0	6,4	17,2					
DE_1	DEA	2.164	18.000	8,1	4,9	0,19	0,20	8,2	6,3	4,9	13,9					
DE_1	DED	396	4.443	10,2	4,9	0,14	0,16	7,6	10,0	6,0	17,5					
DE_1	DEE	586	2.633	8,9	5,3	0,18	0,20	7,5	7,4	5,4	13,0					
DE_1	DEF	3.277	2.782	11,0	5,1	0,28	0,31	10,3	9,2	7,2	14,9					
DE_1	DEG	1.466	2.440	9,5	5,5	0,19	0,21	9,0	8,7	6,2	17,7					
DE_2	DEX	2.596	4.030	10,0	4,5	0,21	0,24	9,9	10,3	7,0	17,9					
ES	ES1	6.614	4.286	4,7	3,8	0,16	0,18	4,7	5,2	3,4	9,0	5,1	6,4	27,7	8,9	3,5
ES	ES2	7.013	4.089	3,3	3,6	0,15	0,16	3,3	3,4	2,5	5,1	3,5	3,8	19,7	4,2	3,2
ES	ES3	4.120	5.230	4,3	4,8	0,21	0,23	4,3	5,3	3,5	7,2	4,9	4,9	14,7	5,2	6,2
ES	ES4	7.348	5.255	6,4	3,8	0,15	0,16	5,9	4,4	3,6	9,9	5,1	6,4	27,3	6,7	3,0
ES	ES5	10.070	11.129	3,6	4,0	0,16	0,18	3,7	5,0	3,1	7,2	5,2	4,8	21,3	4,8	5,2
ES	ES6	8.109	8.568	7,4	4,0	0,15	0,16	6,8	5,7	4,2	12,5	7,4	7,8	20,5	5,5	3,9
ES	ES7	2.815	1.707	6,0	4,4	0,19	0,21	5,6	6,5	4,1	12,2	8,0	6,4	40,2	7,0	4,6
FR	FR1	7.629	11.002	4,7	5,3	0,16	0,18	4,7	7,0	4,3	9,2	5,6	8,8	15,7	7,1	7,8
FR	FR2	9.017	10.493	6,3	4,1	0,14	0,15	6,4	6,2	4,4	11,3	6,3	7,9	20,5	7,3	5,0
FR	FR3	3.125	4.010	8,2	4,8	0,16	0,17	8,0	7,6	5,3	15,4	7,6	8,8	35,8	8,1	5,7
FR	FR4	4.482	5.189	5,0	4,1	0,15	0,16	5,2	4,8	3,6	8,4	4,6	6,4	15,2	5,8	4,6
FR	FR5	7.301	7.835	6,4	4,0	0,14	0,15	6,4	5,0	4,0	10,0	5,6	7,1	14,3	6,5	3,9
FR	FR6	5.676	6.220	7,0	4,2	0,14	0,16	6,9	5,4	4,4	10,5	6,4	7,3	22,3	6,0	3,9
FR	FR7	5.516	7.008	6,2	4,2	0,15	0,16	6,1	5,8	4,2	10,9	5,7	7,6	16,3	6,3	5,4
FR	FR8	5.735	7.137	7,1	4,4	0,15	0,16	7,0	6,1	4,6	11,9	6,9	8,0	15,8	6,7	5,2

* sample size: number of households, aggregated over all available waves

(table continued)

Notes to Table A.4

1. The table gives estimates of standard error of the ratio (NUTS1/Country) for each statistic (1-13). Approximate value of standard error of an actual NUTS1 statistic can be obtained by using information on sample sizes, as explained in Section 15.2.

2. Standard errors are given only for regions where composite SAE estimates have been produced (see Note 2 to Table A.3). No standard errors are obtained where simple regression-prediction has been applied. Only countries which have been divided into 2 or more NUTS1 regions appear in the table.

3. Re DE and IT, please see Note 4 to Table A.3.

Table A.4 (cont)

Country	NUTS1	sample size	Pop. mil.	Variable 1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
	SAE Model			**Percentage standard error**												
GR	GR1	12.215	3.512	3,3	2,6	0,09	0,09	3,1	2,3	1,8	5,4	1,8	3,0	8,9	2,7	2,1
GR	GR2	9.164	2.425	3,2	2,7	0,10	0,10	3,0	2,5	1,9	5,3	2,0	3,1	8,4	3,9	2,2
GR	GR3	9.847	3.885	2,2	2,7	0,11	0,12	2,3	3,1	1,9	3,9	2,0	2,6	4,6	3,1	4,4
GR	GR4	4.695	1.095	3,3	3,2	0,14	0,15	3,2	3,2	2,3	5,3	2,4	3,4	9,4	4,6	3,2
IT_n	IT1	5.176	6.033	10,2	4,6	0,21	0,24	9,1	8,6	6,1	16,2	8,3	9,7	28,3	11,3	6,7
IT_n	IT2	5.804	9.094	9,4	4,8	0,21	0,22	8,3	8,8	5,9	15,5	8,3	10,6	27,1	10,3	7,0
IT_n	IT3	6.494	6.653	8,7	4,7	0,18	0,20	8,0	6,6	4,9	14,7	7,2	8,3	33,6	10,6	4,4
IT_n	IT4	2.999	3.995	11,1	5,4	0,26	0,28	9,9	8,6	6,4	17,2	9,4	9,6	25,9	10,8	6,6
IT_n	IT5	5.742	5.845	9,6	4,3	0,19	0,20	8,6	7,3	5,3	15,8	8,3	8,8	26,2	10,6	5,0
IT_n	IT6	3.465	5.283	15,1	4,9	0,23	0,25	12,7	10,4	7,7	20,0	11,6	10,7	29,1	12,8	7,6
IT_n	IT7	3.048	1.608	11,7	4,5	0,23	0,26	10,2	7,8	6,1	16,3	8,7	8,8	21,3	9,8	5,8
IT_s	IT8	4.665	5.782	2,7	4,0	0,18	0,20	2,9	3,3	2,4	4,3	2,7	4,2	16,5	7,3	3,8
IT_s	IT9	6.929	6.738	2,5	3,6	0,15	0,16	2,6	3,4	2,2	4,6	3,0	3,9	24,0	6,1	3,4
IT_s	ITA	4.215	5.082	3,0	4,1	0,19	0,21	3,2	3,5	2,5	4,8	3,2	4,3	12,0	6,6	3,6
IT_s	ITB	3.148	1.650	2,8	3,9	0,20	0,23	2,9	3,4	2,5	4,2	3,1	4,1	17,7	6,5	3,6
PT	PT1	28.373	9.749	3,3	3,0	0,10	0,11	3,1	3,1	2,1	6,0	3,7	3,1	4,9	2,7	2,9
PT	PT2	4.758	237	6,1	4,9	0,28	0,31	5,8	5,7	4,3	8,9	6,7	5,7	8,0	5,2	5,5
PT	PT3	4.582	240	5,2	4,5	0,24	0,26	5,0	4,9	3,6	7,6	5,7	4,8	6,7	4,4	4,6
UK	UK1	2.426	3.011	6,0	4,2	0,19	0,21	6,0	5,7	4,2	9,9	6,2	6,8	20,1	8,1	5,6
UK	UK2	3.717	4.950	6,8	4,3	0,17	0,19	6,5	5,4	4,1	11,2	5,6	6,6	26,1	7,4	5,8
UK	UK3	3.281	4.157	6,4	4,1	0,17	0,19	6,3	5,7	4,1	11,6	5,7	6,7	20,9	7,5	5,7
UK	UK4	1.708	2.174	6,8	4,4	0,21	0,24	6,7	6,0	4,7	10,1	6,4	7,7	27,5	8,7	5,8
UK	UK5	10.963	18.288	5,1	4,6	0,18	0,19	5,3	6,8	4,3	10,3	7,6	7,3	33,6	10,5	7,1
UK	UK6	3.443	4.909	5,5	4,0	0,17	0,19	5,5	5,3	3,9	9,5	5,9	6,8	31,1	6,8	4,9
UK	UK7	3.390	5.260	6,9	4,4	0,17	0,19	6,7	5,6	4,3	11,6	6,1	6,4	19,6	7,4	5,7
UK	UK8	4.043	6.249	5,9	4,2	0,17	0,19	5,8	5,8	4,0	10,7	5,8	6,9	29,0	8,6	5,8
UK	UK9	2.090	2.900	6,4	4,5	0,20	0,22	6,4	6,2	4,5	11,2	7,0	7,0	24,2	8,0	5,9
UK	UKA	3.621	5.063	6,0	3,9	0,17	0,19	6,0	5,5	4,1	10,2	5,6	6,3	36,5	7,2	4,9
PL	PL1	1.115	7.717	9,8	6,7	0,39	0,36	9,5	8,9	6,4	11,5	8,3	8,8		7,9	8,3
PL	PL2	1.178	8.084	9,7	6,6	0,40	0,36	9,4	9,1	6,5	11,7	8,6	8,8		8,0	8,4
PL	PL3	2.341	6.907	6,8	4,6	0,27	0,25	6,5	6,3	4,5	8,3	5,9	6,1		5,5	5,9
PL	PL4	1.799	6.115	7,8	5,4	0,31	0,29	7,6	7,1	5,1	9,1	6,7	7,0		6,4	6,7
PL	PL5	1.123	4.062	10,0	7,0	0,40	0,36	9,6	9,1	6,6	12,1	8,6	8,9		8,0	8,5
PL	PL6	1.801	5.762	7,7	5,2	0,30	0,29	7,5	7,0	5,0	8,9	6,4	6,9		6,3	6,6

Regression-prediction

not applicable

Table A.5 Composite estimates of Ratio (NUTS2/NUTS1)

			Variable												
NUTS2	sample	Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13
	size*	mil.	HCR_c	HCR_n2	logEqInc	logIncPC	FM_c	FS_c	LAT_c	MAN_c	FSup1	FSup2	FSup3	FSup4	FSup5
SAE Model															
IT11	2.601	4.289	0,96	1,02	1,002	1,003	0,94	0,95	0,95	0,93	0,95	1,00		0,99	0,99
IT12	1.183	121	1,08	1,02	0,998	0,998	1,05	0,87	0,92	1,10	1,09	0,81		0,97	0,68
IT13	1.392	1.623	1,11	0,96	0,994	0,993	1,14	1,14	1,14	1,17	1,12	1,02		1,02	1,06
IT31	2.208	940	1,20	1,04	0,995	0,994	1,09	1,13	1,06	0,92	0,95	0,79		1,04	0,89
IT32	3.276	4.526	1,06	1,03	0,999	0,998	1,04	0,97	1,00	1,02	1,01	1,03		0,98	1,01
IT33	1.010	1.187	0,62	0,86	1,008	1,013	0,79	1,00	0,94	0,99	1,00	1,06		1,05	1,07
IT51	2.758	3.542	1,00	1,10	1,004	1,005	0,96	1,05	1,01	0,97	1,03	1,04		0,95	1,11
IT52	1.416	838	1,11	0,81	0,988	0,983	1,20	0,99	1,07	1,22	0,91	0,99		1,08	0,90
IT53	1.568	1.465	0,95	0,87	0,998	0,998	1,00	0,89	0,93	0,95	0,97	0,91		1,08	0,79
IT71	1.550	1.280	0,82	0,92	1,005	1,005	0,87	0,87	0,88	0,82	0,88	0,87		0,88	0,98
IT72	1.499	328	1,70	1,31	0,979	0,980	1,52	1,49	1,47	1,71	1,47	1,52		1,45	1,08
IT91	3.562	4.086	1,01	1,02	0,999	0,998	1,01	0,88	0,96	0,91	0,91	0,93		0,61	0,99
IT92	1.632	606	0,74	0,81	1,018	1,019	0,77	0,83	0,84	0,69	0,72	1,04		1,38	0,71
IT93	1.734	2.047	1,07	1,02	0,998	0,999	1,05	1,30	1,14	1,27	1,26	1,13		1,67	1,10
PT11	6.882	3.633	0,97	0,99	0,997	0,995	0,98	1,08	1,02	1,05	1,07	1,24		1,06	0,94
PT12	8.135	2.319	1,28	1,01	0,984	0,984	1,23	0,77	0,95	1,01	0,95	1,03		0,88	0,63
PT13	4.473	2.653	0,66	1,00	1,025	1,027	0,71	1,13	0,99	0,87	0,94	0,64		1,03	1,58
PT14	3.972	765	1,18	0,93	0,989	0,991	1,17	0,96	1,04	1,16	1,07	1,08		1,05	0,57
PT15	4.911	380	1,56	1,15	0,973	0,974	1,43	0,77	1,08	1,04	0,92	0,81		0,88	0,63
UK11	785	1.135	0,88	1,02	1,002	1,000	0,89	0,96	0,93	0,93	0,98	0,89		1,07	0,99
UK12	623	488	0,88	0,99	1,003	0,999	0,90	0,73	0,84	0,71	0,93	0,68		0,52	0,84
UK13	1.018	1.387	1,14	0,99	0,997	1,000	1,13	1,13	1,12	1,16	1,04	1,20		1,11	1,07
UK21	688	868	0,82	1,10	1,009	1,007	0,83	0,80	0,85	0,75	0,90	0,72		0,80	0,91
UK22	605	745	1,01	1,01	0,997	0,999	0,99	0,94	0,96	0,94	0,97	1,03		1,18	0,81
UK23	1.117	1.267	1,08	0,94	0,993	0,994	1,08	1,00	1,03	1,10	0,89	0,96		0,87	1,22
UK24	1.307	2.070	1,02	0,99	1,002	1,001	1,03	1,10	1,06	1,07	1,12	1,13		1,10	0,97
UK31	2.006	1.971	1,13	1,00	0,994	0,992	1,12	1,09	1,08	1,15	1,04	1,04		1,04	1,04
UK32	868	1.546	0,83	0,98	1,008	1,006	0,86	0,98	0,95	0,87	0,96	0,98		0,93	1,05
UK33	407	641	0,99	1,05	1,001	1,008	0,97	0,77	0,90	0,86	0,96	0,93		1,04	0,76
UK51	782	1.593	1,21	1,14	0,995	0,993	1,15	0,76	0,91	0,92	0,89	0,75		0,93	0,83
UK52	1.892	2.089	0,91	0,90	1,000	0,998	0,94	0,71	0,83	0,81	0,95	0,77		0,55	0,76
UK53	1.645	2.548	1,04	0,86	0,995	0,994	1,05	1,11	1,02	1,06	1,18	0,97		1,27	0,98
UK54	977	1.609	0,82	1,00	1,007	1,008	0,85	0,68	0,81	0,69	0,77	0,70		0,75	0,68
UK55	3.703	7.104	1,08	1,10	1,000	1,002	1,05	1,21	1,14	1,19	1,03	1,27		1,15	1,17
UK56	1.109	1.771	0,86	0,98	1,004	1,004	0,89	0,89	0,92	0,81	1,07	0,84		0,79	0,95
UK57	855	1.575	0,81	0,79	0,999	0,998	0,90	0,98	0,95	0,94	0,92	0,86		1,05	1,13
UK61	1.752	2.156	1,02	1,05	0,999	0,997	1,01	1,03	1,01	1,05	1,11	0,97		1,00	0,94
UK62	519	1.568	1,11	0,92	0,994	0,998	1,13	1,20	1,15	1,23	1,05	1,23		1,13	1,13
UK63	1.172	1.185	0,92	0,95	1,004	1,007	0,93	0,87	0,93	0,80	0,77	0,96		0,95	1,05
UK71	646	1.216	0,58	0,83	1,016	1,018	0,65	0,74	0,74	0,61	0,91	0,87		0,73	0,69
UK72	1.368	1.484	0,90	1,01	1,004	1,004	0,92	0,88	0,92	0,85	0,88	0,84		0,94	0,91
UK73	1.376	2.560	1,26	1,07	0,990	0,989	1,21	1,20	1,17	1,27	1,11	1,16		1,16	1,20
UK81	913	982	0,83	0,96	1,010	1,011	0,86	0,84	0,85	0,84	1,00	0,78		0,98	0,94
UK82	1.535	2.487	1,05	1,07	1,000	1,002	1,03	1,07	1,06	1,03	1,10	1,10		1,03	0,98
UK83	751	1.413	0,95	0,99	1,002	1,001	0,95	0,98	0,95	0,98	0,97	1,05		0,92	1,00
UK84	844	1.368	1,09	0,92	0,991	0,987	1,10	1,00	1,06	1,09	0,84	0,93		1,04	1,07
UKA1	1.943	1.901	1,10	0,90	0,996	0,997	1,11	1,18	1,10	1,20	1,08	1,16		1,17	1,00
UKA2	1.279	2.288	0,94	1,08	1,003	1,001	0,94	0,89	0,95	0,88	0,94	0,90		0,81	1,00
UKA4	399	505	0,87	0,98	1,004	1,007	0,87	0,82	0,86	0,77	0,99	0,84		1,20	1,01

* sample size: number of households, aggregated over all available waves

(cont)

Table A.5 (cont)

			Variable												
NUTS2	sample	Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13
	size	mil.	HCR_c	HCR_n2	logEqInc	logIncPC	FM_c	FS_c	LAT_c	MAN_c	FSup1	FSup2	FSup3	FSup4	FSup5
SAE Model (cont)															
PL11	525	2.648	1,06	1,02	0,993	0,999	1,08	1,08	1,08	1,21	1,09	1,06		0,91	1,01
PL12	590	5.069	0,97	0,99	1,004	1,001	0,96	0,96	0,96	0,89	0,95	0,97		1,04	1,00
PL21	615	3.227	1,00	0,96	0,999	0,993	1,05	1,06	1,08	1,16	1,16	1,05		1,06	1,04
PL22	563	4.858	1,00	1,02	1,001	1,004	0,97	0,96	0,95	0,89	0,89	0,97		0,96	0,97
PL31	535	2.233	1,04	1,03	0,997	1,004	1,05	1,01	1,03	1,03	1,08	0,98		1,08	0,90
PL32	600	2.128	0,93	0,94	1,001	0,992	0,96	1,01	0,97	0,98	0,97	1,06		0,93	0,97
PL33	621	1.324	1,04	1,03	1,001	1,001	0,99	0,98	1,00	0,96	0,93	0,99		0,96	1,10
PL34	585	1.222	1,00	1,01	1,002	1,005	1,00	0,99	1,01	1,01	0,98	0,95		1,02	1,12
PL41	643	3.358	0,87	0,95	1,002	1,001	0,85	0,86	0,88	0,84	0,92	0,86		0,87	0,95
PL42	544	1.733	1,17	1,08	0,997	1,002	1,19	1,15	1,15	1,20	1,08	1,20		1,19	1,07
PL43	612	1.024	1,15	1,03	0,997	0,994	1,17	1,19	1,13	1,19	1,13	1,11		1,10	1,03
PL51	609	2.975	1,05	1,05	0,999	0,997	1,04	1,07	1,05	1,07	1,01	1,14		1,08	1,05
PL52	514	1.087	0,85	0,87	1,004	1,009	0,89	0,81	0,87	0,81	0,97	0,63		0,78	0,88
PL61	615	2.101	1,13	1,01	0,995	0,990	1,15	1,04	1,07	1,10	1,03	1,09		1,08	1,07
PL62	592	1.466	1,22	1,08	0,994	0,993	1,28	1,22	1,23	1,27	1,19	1,20		1,05	1,10
PL63	594	2.195	0,73	0,93	1,008	1,014	0,68	0,81	0,78	0,72	0,84	0,78		0,89	0,86
RO01	5.127	3.826	1,52	1,18	0,991	0,989	1,45								
RO02	3.953	2.936	1,04	1,09	1,001	1,001	1,03								
RO03	4.965	3.469	1,05	0,93	0,997	0,996	1,07								
RO04	3.737	2.401	0,92	1,00	1,001	1,001	0,94								
RO05	3.035	2.041	0,81	0,92	1,002	1,003	0,85								
RO06	4.029	2.847	0,89	1,02	1,003	1,002	0,89								
RO07	3.859	2.643	1,02	0,98	0,998	0,998	1,02								
RO08	2.842	2.279	0,36	0,76	1,015	1,019	0,41								
Regression model															
BE21		1.645	1,00	0,97	1,000	0,998	1,00	1,07	0,90	1,10	1,02	1,01		1,03	0,86
BE22		793	0,94	0,92	1,006	1,003	0,97	1,15	0,81	1,21	1,07	1,05		1,02	0,72
BE23		1.363	1,06	1,02	0,998	0,997	1,06	1,08	1,02	1,07	1,07	1,13		1,09	0,99
BE24		1.017	1,05	1,16	0,992	1,003	1,00	0,66	1,53	0,52	0,86	0,86		0,90	1,71
BE25		1.129	0,92	0,93	1,004	1,001	0,94	1,01	0,78	1,06	0,97	0,92		0,93	0,78
BE31		351													
BE32		1.280	1,00	0,95	1,002	1,000	1,01	1,13	0,88	1,17	1,07	1,06		1,04	0,80
BE33		1.020	0,94	0,95	1,003	1,000	0,96	1,05	0,83	1,11	0,99	0,97		1,01	0,78
BE34		248	0,88	0,90	1,007	1,007	0,93	0,96	0,89	0,99	0,98	0,95		0,85	0,88
BE35		445	1,28	1,08	0,990	0,988	1,23	1,18	1,08	1,04	1,19	1,34		1,17	1,12

(table continued)

Notes to Table A.5

1. The table gives estimates of ratio (NUTS2/NUTS1) for each statistic (1-13). Actual NUTS2 value of the statistic can be obtained by multiplying this ratio with the (NUTS1/Country) ratio from Table A.3, and then with the corresponding country-level estimate from Table A.2.

2. For countries for which we have survey micro-data, and NUTS2 regions are identified in the data, composite SAE estimates have been produced. However, if there is no NUTS2 code in the data (as for all ECHP surveys except IT, UK and PT), or no survey data are available to us (as in the case of all NMS10 and Candidate countries except PL and RO), simple regression (with coefficients taken from the corresponding SAE model) has been used to predict the required values.

3. Only NUTS1 regions which have been divided into 2 or more NUTS1 regions appear in the table.

Table A.5 (cont)

NUTS2	sample size	Pop. mil.	Variable 1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
Regression model															
DE11		3.926	0,97	1,06	0,999	1,001	0,96	0,87	1,14	0,86					
DE12		2.680	1,04	0,94	1,000	0,997	1,05	1,14	0,87	1,15					
DE13		2.132	1,05	0,98	1,001	0,999	1,04	1,12	0,93	1,10					
DE14		1.762	0,94	0,98	1,002	1,003	0,96	0,94	0,99	0,96					
DE21		4.244	0,96	1,06	0,997	1,003	0,95	0,73	1,24	0,69					
DE22		1.228	1,04	0,93	1,002	0,997	1,05	1,19	0,75	1,20					
DE23		1.126	1,12	0,95	0,999	0,992	1,12	1,29	0,76	1,31					
DE24		600	1,12	1,04	0,998	0,994	1,08	1,22	0,91	1,19					
DE25		1.764	1,05	1,05	1,000	0,999	1,03	1,14	1,04	1,13					
DE26		1.396	1,01	0,92	1,002	0,998	1,03	1,15	0,79	1,20					
DE27		1.830	0,90	0,93	1,004	1,004	0,94	0,99	0,91	1,06					
DE71		3.728	1,01	1,04	0,998	1,000	1,00	0,92	1,10	0,89					
DE72		1.063	0,94	0,94	1,005	1,004	0,95	1,06	0,87	1,10					
DE73		1.268	1,03	0,94	1,001	0,997	1,04	1,19	0,80	1,25					
DE91		1.669	0,80	0,93	1,009	1,007	0,84	1,00	0,82	1,14					
DE92		2.157	0,96	1,02	1,002	1,004	0,97	0,97	1,10	0,98					
DE93		1.666	1,20	1,07	0,992	0,993	1,16	1,05	1,15	0,92					
DE94		2.420	1,04	0,99	0,998	0,996	1,03	0,99	0,93	0,98					
DEa1		5.258	1,02	1,02	0,998	1,000	1,01	0,98	1,07	0,96					
DEa2		4.270	0,98	0,98	1,001	1,000	0,99	0,99	0,97	1,01					
DEa3		2.610	1,04	0,94	0,999	0,997	1,05	1,04	0,87	1,04					
DEa4		2.052	1,02	1,08	0,998	1,001	0,99	0,89	1,17	0,83					
DEa5		3.809	0,96	0,98	1,003	1,002	0,97	1,07	0,94	1,12					
DEd1		1.631	1,11	1,04	0,995	0,994	1,08	1,02	1,01	0,96					
DEd2		1.719	0,96	1,00	1,001	1,003	0,97	0,97	1,02	0,98					
DEd3		1.094	0,90	0,94	1,006	1,005	0,94	1,02	0,95	1,09					
DEe1		548	0,85	0,97	1,006	1,007	0,88	0,94	0,97	1,01					
DEe2		870	0,95	0,98	1,002	1,003	0,96	0,98	1,00	1,01					
DEe3		1.215	1,10	1,03	0,996	0,995	1,08	1,04	1,01	0,99					
ES11		2.691	1,06	1,01	1,001	0,998	1,06	1,15	0,93	1,13	1,10	1,17		1,09	0,89
ES12		1.063	0,97	0,97	0,997	1,000	0,98	0,88	1,03	0,90	0,92	0,84		0,90	1,07
ES13		532	0,75	1,03	1,002	1,013	0,76	0,49	1,28	0,54	0,65	0,46		0,72	1,43
ES21		2.073	0,89	1,06	1,006	1,011	0,90	0,94	1,25	0,96	0,95	0,92		1,01	1,19
ES22		549	0,92	1,00	1,001	1,002	0,91	0,81	0,94	0,81	0,86	0,80		0,85	1,06
ES23		270	0,91	0,88	1,002	0,998	0,95	0,96	0,73	1,07	0,94	0,89		0,92	0,72
ES24		1.197	1,25	0,93	0,989	0,980	1,22	1,20	0,65	1,15	1,16	1,25		1,07	0,71
ES41		2.462	0,96	1,05	0,996	1,000	0,95	0,74	1,18	0,74	0,85	0,81		0,92	1,29
ES42		1.735	0,90	0,94	1,003	1,001	0,92	0,94	0,80	1,01	0,93	0,88		0,92	0,83
ES43		1.059	1,25	0,99	1,005	0,998	1,24	1,70	0,90	1,59	1,47	1,65		1,32	0,60
ES51		6.251	0,95	1,03	0,999	1,001	0,95	0,85	1,10	0,86	0,90	0,89		0,97	1,16
ES52		4.042	1,00	0,91	1,006	1,001	1,03	1,28	0,77	1,31	1,15	1,15		1,05	0,62
ES53		836	1,33	1,19	0,981	0,987	1,23	0,81	1,41	0,55	1,01	1,10		0,98	1,68
ES61		7.273	1,01	0,99	1,001	1,001	1,01	1,05	1,01	1,05	1,03	1,04		1,01	0,97
ES62		1.159	0,96	1,07	0,995	0,995	0,92	0,71	0,96	0,72	0,79	0,76		0,91	1,18
ES63		71													
ES64		65													
GR11		606	1,04	0,99	1,000	0,996	1,03	1,07	0,86	1,05	1,04	1,08		1,03	0,88
GR12		1.871	0,93	0,90	1,006	1,001	0,97	1,14	0,74	1,21	1,05	1,03		1,00	0,65
GR13		294													
GR14		741	1,04	1,02	0,998	0,999	1,03	0,99	1,11	0,98	1,02	1,05		1,03	1,11
GR21		337	0,91	0,92	1,002	1,005	0,96	0,89	1,06	0,93	0,95	0,87		0,86	1,03
GR22		209	1,04	1,09	1,002	1,012	1,02	0,91	1,46	0,72	1,02	1,01		0,85	1,49
GR23		722	1,21	1,00	0,993	0,989	1,19	1,16	0,96	1,08	1,16	1,28		1,11	0,97
GR24		560	0,63	1,20	1,005	1,016	0,59	0,42	1,35	0,54	0,49	0,25		0,86	1,53
GR25		598	1,13	0,82	1,002	0,990	1,18	1,45	0,52	1,47	1,30	1,44		1,13	0,35
GR41		205	1,41	0,78	0,974	0,958	1,40	1,22	0,36	1,25	1,21	1,30		1,11	0,43
GR42		297	0,73	1,17	1,012	1,031	0,70	0,65	1,62	0,58	0,74	0,49		0,72	1,68
GR43		592	0,99	0,99	1,003	0,999	1,01	1,10	0,91	1,12	1,06	1,15		1,10	0,86

(cont)

Table A.5 (cont)

			Variable												
NUTS2	sample	Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13
	size	mil.	HCR_c	HCR_n2	logEqInc	logIncPC	FM_c	FS_c	LAT_c	MAN_c	FSup1	FSup2	FSup3	FSup4	FSup5
Regression model															
AT11		276	1,08	0,89	0,996	0,992	1,08	1,04	0,72	1,03	1,05	1,03		0,90	0,81
AT12		1.537	1,00	0,97	0,996	0,995	0,98	0,84	0,87	0,85	0,91	0,86		0,89	1,01
AT13		1.551	0,99	1,04	1,004	1,006	1,00	1,15	1,18	1,14	1,09	1,13		1,13	1,02
AT21		560	0,97	0,95	1,000	1,000	0,99	0,95	0,98	0,99	0,98	0,96		0,96	0,97
AT22		1.183	1,02	1,02	1,000	1,000	1,00	1,02	1,01	1,01	1,01	1,02		1,02	1,01
AT31		1.372	0,97	0,99	1,003	1,002	0,97	1,07	0,94	1,09	1,02	1,00		1,02	0,89
AT32		514	0,94	0,99	1,000	1,003	0,96	0,86	1,12	0,87	0,93	0,94		0,94	1,14
AT33		670	1,12	1,02	0,993	0,993	1,09	0,97	1,03	0,91	1,02	1,05		1,00	1,12
AT34		349													
FI13		684	1,01	0,95	1,002	1,003	1,03	1,13	1,00	1,14	1,09	1,03		0,96	0,93
FI18		2.521	0,97	0,99	0,999	0,999	0,96	0,88	0,96	0,89	0,91	0,90		0,96	1,00
FI19		1.317	1,03	1,02	1,000	1,000	1,03	1,06	1,02	1,05	1,05	1,08		1,04	1,01
FI1a		628	1,07	1,06	1,002	1,002	1,07	1,21	1,11	1,18	1,15	1,22		1,13	1,04
FR21		1.343	1,01	1,02	1,001	1,001	1,01	1,02	1,00	1,04	1,00	1,02		1,03	0,95
FR22		1.865	1,07	1,08	0,999	1,000	1,03	1,07	1,11	1,01	1,05	1,08		1,08	1,12
FR23		1.790	0,95	0,98	1,005	1,005	0,98	1,11	1,04	1,15	1,07	1,12		1,09	0,91
FR24		2.454	1,05	1,03	0,996	0,996	1,03	0,95	1,02	0,91	0,98	1,00		1,00	1,10
FR25		1.429	0,94	0,88	1,000	0,998	0,97	0,91	0,79	0,97	0,94	0,85		0,84	0,84
FR26		1.612	0,95	0,97	1,000	1,001	0,96	0,93	0,99	0,95	0,95	0,90		0,94	1,00
FR41		2.314	1,05	0,96	1,000	0,997	1,06	1,15	0,91	1,16	1,10	1,12		1,07	0,84
FR42		1.754	1,00	1,09	0,999	1,003	0,97	0,83	1,20	0,77	0,91	0,92		0,95	1,31
FR43		1.121	0,90	0,94	1,003	1,002	0,93	0,95	0,88	1,03	0,94	0,88		0,94	0,86
FR51		3.255	0,95	1,05	1,005	1,006	0,95	1,04	1,09	1,04	1,01	1,02		1,06	1,04
FR52		2.929	1,04	0,96	0,997	0,996	1,04	0,97	0,93	0,96	1,00	0,98		0,95	0,97
FR53		1.651	1,03	0,97	0,997	0,996	1,03	0,97	0,95	0,99	0,99	0,99		0,98	0,97
FR61		2.933	1,08	1,04	0,997	0,998	1,06	1,03	1,08	0,98	1,04	1,07		1,05	1,10
FR62		2.576	0,94	0,97	1,004	1,003	0,96	1,03	0,95	1,06	1,01	1,00		0,99	0,91
FR63		711	0,88	0,93	0,999	0,998	0,89	0,77	0,83	0,86	0,81	0,71		0,84	0,93
FR71		5.698	1,00	1,02	1,000	1,001	0,99	0,98	1,04	0,96	0,99	1,00		1,00	1,06
FR72		1.311	1,00	0,89	1,000	0,997	1,02	1,10	0,80	1,17	1,05	1,00		0,99	0,73
FR81		2.325	1,06	0,97	0,999	0,996	1,06	1,15	0,86	1,15	1,09	1,12		1,06	0,81
FR82		4.552	0,96	1,03	1,001	1,003	0,96	0,91	1,09	0,91	0,94	0,94		0,98	1,12
FR83		261	1,13	0,82	0,995	0,992	1,14	1,22	0,72	1,24	1,14	1,02		0,92	0,63
FR91		428	0,97	0,95	1,003	1,001	0,99	1,08	0,89	1,11	1,03	1,03		1,02	0,81
FR92		385	1,01	1,05	0,996	0,999	1,02	0,90	1,14	0,89	0,96	0,97		0,97	1,21
FR93		164	1,01	0,97	1,003	1,001	0,99	1,07	0,89	1,06	1,03	1,02		1,03	0,85
FR94		722	1,01	1,01	1,000	1,000	1,00	0,99	1,02	0,98	1,00	1,00		1,00	1,03
IE01		999	0,94	0,87	1,007	1,002	0,99	1,25	0,72	1,34	1,11	1,09		1,04	0,55
IE02		2.788	1,02	1,05	0,997	0,999	1,00	0,91	1,10	0,88	0,96	0,97		0,99	1,16
NL11		565	0,98	0,95	1,004	1,001	1,00	1,15	0,89	1,19	1,08	1,09		1,07	0,75
NL12		628	0,94	1,00	1,001	1,003	0,96	0,93	1,05	0,95	0,95	0,94		0,97	1,07
NL13		472	1,10	1,06	0,993	0,994	1,05	0,92	1,06	0,84	0,97	0,97		0,96	1,21
NL21		1.082	0,93	0,96	1,003	1,002	0,94	1,00	0,87	1,05	0,97	0,91		0,96	0,86
NL22		1.927	1,00	1,00	0,999	0,999	1,01	0,96	1,02	0,96	0,98	0,98		0,99	1,04
NL23		323	1,21	1,12	0,997	0,999	1,17	1,21	1,29	1,05	1,22	1,38		1,19	1,24
NL31		1.113	1,02	1,07	0,996	0,998	1,01	0,87	1,13	0,83	0,94	0,98		1,00	1,22
NL32		2.527	1,00	1,01	0,999	0,999	1,00	0,99	1,02	1,00	0,99	1,00		1,02	1,01
NL33		3.409	0,98	0,96	1,002	1,002	0,99	1,03	0,95	1,05	1,02	1,00		0,98	0,92
NL34		373	1,09	1,05	0,998	0,996	1,04	1,11	0,94	1,07	1,05	1,01		1,04	0,97
NL41		2.366	0,94	1,01	1,002	1,004	0,94	0,91	1,06	0,92	0,94	0,91		0,95	1,08
NL42		1.142	1,13	0,98	0,996	0,992	1,12	1,18	0,88	1,16	1,13	1,20		1,11	0,84

(cont)

Table A.5 (cont)

			Variable												
NUTS2	sample	Pop.	1	2	3	4	5	6	7	8	9	10	11	12	13
	size	mil.	HCR_c	HCR_n2	logEqInc	logIncPC	FM_c	FS_c	LAT_c	MAN_c	FSup1	FSup2	FSup3	FSup4	FSup5
Regression model															
SE01		1.813	1,00	1,04	1,000	1,002	0,96	0,86	1,14	0,78					
SE02		1.491	0,99	1,01	1,002	1,002	0,98	1,00	0,99	0,98					
SE04		1.277	0,91	0,95	1,004	1,006	0,95	1,02	0,99	1,09					
SE06		835	1,10	0,99	0,995	0,992	1,11	1,12	0,92	1,14					
SE07		378	1,05	0,98	0,999	1,000	1,06	1,13	1,04	1,13					
SE08		513	1,47	1,12	0,982	0,977	1,41	1,45	1,07	1,30					
SE09		798	1,07	1,01	0,993	0,987	1,07	0,99	0,83	1,03					
SE0a		1.766	0,84	0,96	1,006	1,008	0,88	0,91	0,96	0,98					
UKA3		370													
CZ01		1.184													
CZ02		1.113	0,90		1,000										
CZ03		1.178	0,97		0,997										
CZ04		1.132	1,17		1,000										
CZ05		1.489	0,94		1,000										
CZ06		1.658	0,88		1,003										
CZ07		1.240	1,00		1,002										
CZ08		1.280	1,19		0,998										
HU21		1.107	0,92		1,006										
HU22		983	0,98		0,998										
HU23		973	1,11		0,995										
HU31		1.266	1,00		1,002										
HU32		1.520	1,06		1,000										
HU33		1.338	0,93		0,998										
SK01		617													
SK02		1.876	0,79		1,006										
SK03		1.356	0,89		1,006										
SK04		1.552	1,35		0,987										
BG01		582	1,17		0,992										
BG02		1.220	0,84		1,004										
BG03		1.339	1,17		0,997										
BG04		2.143	0,82		1,008										
BG05		2.065	1,05		0,996										
BG06		822	1,19		0,994										

In the prediction of values for Hungary there is an inconsistency in the regressors.
In NewCronos database, the infant mortality rate in each nuts1 is higher than every single rate in each correspondent nuts2.
For this reason all the ratios used in the model are smaller than 1, so the values predicted can be affected by this error

Table A.6 Percentage standard error of composite estimates of Ratio (NUTS2/NUTS1)

			Variable												
NUTS1	sample size*	Pop. mil.	1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
SAE Model			Percentage standard error												
IT11	2.601	4.289	7,3	5,0	0,28	0,28	10,2	7,9	5,5	12,8	8,4	9,4		10,2	6,1
IT12	1.183	121	7,3	5,0	0,28	0,29	10,4	8,2	5,7	14,4	9,2	9,7		10,4	6,0
IT13	1.392	1.623	6,5	4,5	0,28	0,29	9,5	7,3	5,1	12,9	8,2	8,7		9,3	5,4
IT31	2.208	940	7,0	4,8	0,19	0,29	9,8	8,8	8,2	15,5	8,2	9,0		8,6	8,5
IT32	3.276	4.526	5,1	3,5	0,19	0,19	7,3	6,6	6,3	12,4	6,2	6,7		6,3	6,3
IT33	1.010	1.187	10,4	6,7	0,35	0,36	12,6	11,1	9,1	15,2	9,7	12,0		12,1	11,5
IT51	2.758	3.542	4,5	3,6	0,19	0,19	6,7	6,1	4,8	10,8	5,5	6,5		6,6	5,2
IT52	1.416	838	7,3	5,5	0,28	0,28	10,1	9,1	7,1	14,2	8,1	9,9		10,2	8,2
IT53	1.568	1.465	6,8	5,1	0,28	0,28	9,5	8,5	6,7	13,3	7,6	9,3		9,5	7,7
IT71	1.550	1.280	3,4	4,2	0,19	0,28	5,6	8,1	3,7	9,6	6,4	9,0		9,2	8,0
IT72	1.499	328	3,5	4,4	0,28	0,29	5,8	8,6	3,9	10,3	6,9	9,5		9,7	8,4
IT91	3.562	4.086	3,6	3,2	0,19	0,19	2,1	2,2	2,1	2,6	1,9	2,7		4,3	2,5
IT92	1.632	606	6,4	5,4	0,28	0,28	3,7	4,0	3,7	4,7	3,4	4,9		7,6	4,5
IT93	1.734	2.047	6,2	5,3	0,28	0,28	3,6	3,8	3,6	4,5	3,3	4,7		7,3	4,4
UK11	785	1.135	6,8	5,4	0,28	0,28	6,7	6,9	4,1	10,0	6,7	7,5		8,4	7,1
UK12	623	488	8,0	6,3	0,28	0,29	7,7	8,0	4,8	11,5	7,7	8,7		9,9	8,3
UK13	1.018	1.387	5,7	4,6	0,19	0,19	5,5	5,8	3,4	8,5	5,7	6,3		7,1	5,9
UK21	688	868	6,4	5,7	0,28	0,29	6,5	7,7	5,4	9,4	7,7	8,3		10,4	7,1
UK22	605	745	6,8	5,9	0,28	0,36	6,9	8,1	5,8	9,8	8,0	8,8		11,0	7,5
UK23	1.117	1.267	4,8	4,4	0,19	0,19	4,8	5,8	4,2	7,2	6,1	6,3		8,0	5,3
UK24	1.307	2.070	4,2	4,0	0,19	0,19	4,4	5,3	3,8	6,4	5,5	5,7		7,3	4,8
UK31	2.006	1.971	2,8	2,5	0,10	0,10	2,8	3,0	3,1	3,9	3,6	3,4		4,4	3,1
UK32	868	1.546	5,6	4,9	0,19	0,29	5,6	6,1	5,9	7,7	6,7	6,8		8,7	6,0
UK33	407	641	8,5	6,9	0,36	0,37	8,5	9,0	8,2	11,2	9,4	10,1		12,6	9,1
UK51	782	1.593	9,8	5,0	0,28	0,28	9,2	7,0	8,0	11,9	6,9	8,6		8,6	6,6
UK52	1.892	2.089	6,4	3,3	0,19	0,20	6,2	4,6	6,2	8,9	4,6	5,7		5,6	4,2
UK53	1.645	2.548	6,9	3,5	0,19	0,19	6,6	4,9	6,4	9,3	5,0	6,1		6,0	4,5
UK54	977	1.609	8,9	4,5	0,28	0,28	8,5	6,4	7,8	11,5	6,3	7,9		7,9	5,9
UK55	3.703	7.104	4,3	2,2	0,10	0,10	4,1	3,1	4,4	6,3	3,2	3,7		3,7	2,8
UK56	1.109	1.771	8,4	4,3	0,19	0,28	8,0	6,1	7,5	11,0	6,0	7,4		7,4	5,6
UK57	855	1.575	9,4	4,8	0,28	0,28	9,0	6,8	8,0	12,0	6,7	8,3		8,3	6,3
UK61	1.752	2.156	4,3	3,2	0,10	0,20	4,2	4,3	3,1	6,5	4,1	4,2		6,4	5,0
UK62	519	498	9,6	6,6	0,36	0,37	9,2	9,1	6,6	12,9	8,5	9,2		13,0	10,5
UK63	1.172	1.185	6,0	4,3	0,19	0,19	5,7	5,8	4,2	8,6	5,6	5,8		8,6	6,8
UK71	646	1.216	6,5	5,5	0,28	0,29	6,5	7,6	5,0	9,4	7,4	8,8		10,9	7,6
UK72	1.368	1.484	3,9	3,5	0,19	0,19	4,0	4,8	3,2	6,0	4,7	5,6		7,0	4,7
UK73	1.376	2.560	3,9	3,5	0,19	0,20	4,0	4,8	3,1	6,0	4,7	5,6		6,9	4,7
UK81	913	982	6,6	4,7	0,28	0,29	6,5	6,3	6,2	8,9	6,7	7,0		7,8	6,1
UK82	1.535	2.487	4,7	3,3	0,19	0,19	4,7	4,5	4,6	6,4	4,9	5,0		5,6	4,3
UK83	751	1.413	7,3	5,1	0,28	0,28	7,3	6,9	6,7	9,7	7,2	7,6		8,6	6,7
UK84	844	1.368	6,9	4,9	0,28	0,29	6,9	6,6	6,4	9,2	6,9	7,3		8,1	6,4
UKA1	1.943	1.901	3,7	3,1	0,10	0,19	3,6	3,9	3,8	5,7	4,3	4,5		5,6	4,5
UKA2	1.279	2.288	5,3	4,3	0,19	0,20	5,3	5,4	5,2	7,9	6,0	6,3		7,8	6,4
UKA4	399	505	10,1	7,4	0,36	0,44	9,8	9,8	8,5	13,2	9,7	11,2		13,6	11,3

* sample size: number of households, aggregated over all available waves

(table continued)

Notes to Table A.6

1. The table gives estimates of standard error of the ratio (NUTS2/NUTS1) for each statistic (1-13). Approximate value of standard error of an actual NUTS2 statistic can be obtained by using information on sample sizes, as explained in Section 15.2.

2. Standard errors are given only for regions where composite SAE estimates have been produced (see Note 2 to Table A.3). No standard errors are obtained where simple regression-prediction has been applied. Only countries or NUTS1 regions which have been divided into 2 or more NUTS2 regions appear in the table.

Table A.6 (cont)

NUTS1	sample size	Pop. mil.	Variable 1 HCR_c	2 HCR_n2	3 logEqInc	4 logIncPC	5 FM_c	6 FS_c	7 LAT_c	8 MAN_c	9 FSup1	10 FSup2	11 FSup3	12 FSup4	13 FSup5
SAE Model			**Percentage standard error**												
PT11	6.882	3.633	4,2	3,6	0,20	0,20	3,9	3,6	7,8	6,0	4,5	3,7		3,4	3,4
PT12	8.135	2.319	3,6	3,1	0,19	0,20	3,5	3,2	6,2	5,2	3,9	3,3		3,0	3,0
PT13	4.473	2.653	5,3	4,3	0,19	0,28	5,1	4,7	7,6	7,2	5,5	4,8		4,3	4,3
PT14	3.972	765	5,7	4,5	0,28	0,28	5,3	4,9	7,9	7,7	5,8	5,2		4,7	4,5
PT15	4.911	380	5,1	4,3	0,20	0,20	4,9	4,5	8,9	7,3	5,4	4,6		4,1	4,2
PL11	525	2.648	13,5	9,0	0,56	0,50	13,0	12,0	8,6	15,0	11,1	11,9		11,0	11,4
PL12	590	5.069	14,1	9,9	0,55	0,52	13,7	13,3	9,5	17,4	12,5	12,8		11,5	12,1
PL21	615	3.227	13,4	8,9	0,56	0,50	13,0	12,0	8,5	14,9	11,0	11,9		11,0	11,4
PL22	563	4.858	14,1	9,7	0,58	0,51	13,7	13,8	9,8	18,4	13,3	13,1		11,7	12,4
PL31	535	2.233	13,5	9,2	0,53	0,50	13,0	12,6	9,2	17,0	12,0	12,2		11,0	11,8
PL32	600	2.128	13,8	9,3	0,57	0,50	13,3	12,7	9,2	16,7	11,9	12,3		11,3	11,9
PL33	621	1.324	13,3	9,0	0,52	0,50	12,8	12,1	8,7	15,6	11,3	11,9		10,8	11,4
PL34	585	1.222	13,5	9,3	0,53	0,50	13,1	12,8	9,2	17,2	12,1	12,3		11,1	11,9
PL41	643	3.358	13,3	9,0	0,52	0,50	12,8	12,0	8,5	15,1	11,2	11,8		10,8	11,3
PL42	544	1.733	13,9	9,5	0,57	0,51	13,4	12,7	9,1	16,3	11,9	12,5		11,3	11,9
PL43	612	1.024	13,6	9,4	0,53	0,50	13,1	12,3	8,9	16,0	11,5	12,1		11,0	11,6
PL51	609	2.975	13,4	9,1	0,52	0,50	12,9	12,1	8,6	15,2	11,2	11,9		10,9	11,4
PL52	514	1.087	14,9	10,7	0,60	0,52	14,5	13,9	10,2	19,2	13,3	13,3		11,9	12,8
PL61	615	2.101	13,2	8,9	0,52	0,50	12,8	11,8	8,4	14,7	10,8	11,7		10,8	11,2
PL62	592	1.466	13,7	9,5	0,53	0,50	13,2	12,5	9,2	16,6	11,8	12,2		11,0	11,7
PL63	594	2.195	13,3	8,9	0,52	0,50	12,8	11,9	8,5	14,8	10,9	11,8		10,8	11,3
RO01	5.127	3.826	5,5	4,0	0,10	0,10	5,5								
RO02	3.953	2.936	5,5	4,5	0,10	0,10	5,4								
RO03	4.965	3.469	4,8	3,7	0,10	0,10	4,7								
RO04	3.737	2.401	5,2	4,4	0,10	0,10	5,2								
RO05	3.035	2.041	5,4	4,6	0,10	0,10	5,3								
RO06	4.029	2.847	4,7	3,9	0,10	0,10	4,6								
RO07	3.859	2.643	5,3	4,3	0,10	0,10	5,3								
RO08	2.842	2.279	4,1	4,5	0,10	0,10	4,1								
Regression-prediction															
not applicable															

Table A.7 Composite estimates of Ratio (NUTS3/NUTS2), and percentage standard error of this estimated ratio. Italy

				INPUT			OUTPUT			%SE(ratio)		
Nuts2	Nuts3	sample	Pop.	input ratio (NUTS3/NUTS2)			raked	raked	raked	HCR_C	HCR_N2	logEqInc
		size	mil.	HCR_C	HCR_N2	logEqInc	HCR_C	HCR_N2	logEqInc	HCR_C	HCR_N2	logEqInc
SAE Model 3												
IT11	IT111	1120	2214,6	0,80	0,96	1,004	0,92	0,97	1,000	9,22	6,45	0,34
IT11	IT113	13	189,4	0,00	0,00	0,992	0,88	0,84	1,005	30,87	29,01	0,95
IT11	IT114	357	160,7	1,52	1,41	0,989	1,57	1,38	0,986	16,92	12,50	0,60
IT11	IT115	387	344,3	0,69	0,77	1,016	0,89	0,83	1,006	18,11	13,64	0,63
IT11	IT116	287	558,2	1,35	1,40	0,998	1,33	1,28	0,997	21,28	16,44	0,73
IT11	IT117	9	210,5	0,00	0,00	0,991	1,05	0,99	0,997	31,55	29,57	0,97
IT11	IT118	428	430,4	0,92	0,86	1,000	0,88	0,88	1,002	17,71	13,01	0,63
IT13	IT131	1	216,4	0,00	0,00	0,944	1,21	1,18	0,994	30,01	28,66	0,91
IT13	IT132	209	279,7	1,55	1,57	0,986	1,43	1,45	0,992	15,44	11,51	0,60
IT13	IT133	706	905,5	0,60	0,58	1,008	0,68	0,71	1,007	8,26	5,89	0,35
IT13	IT134	476	221,8	1,58	1,25	0,985	1,54	1,44	0,989	11,05	7,91	0,46
IT20	IT201	362	818,4	0,57	0,68	1,001	0,79	0,80	0,997	17,20	10,48	0,55
IT20	IT202	753	541	0,91	0,88	1,002	0,94	0,90	0,999	12,37	7,12	0,40
IT20	IT203	271	310,6	0,47	0,63	1,010	0,74	0,75	1,001	21,02	13,40	0,67
IT20	IT205	2394	3765,8	0,66	0,78	1,010	0,65	0,75	1,011	6,49	3,59	0,21
IT20	IT206	983	969,8	1,43	1,52	0,988	1,42	1,38	0,988	11,26	6,43	0,36
IT20	IT207	530	1105,6	2,23	1,93	0,972	1,96	1,76	0,978	15,34	9,16	0,49
IT20	IT208	300	498,4	1,06	1,05	1,005	1,14	1,05	0,999	20,26	12,61	0,65
IT20	IT209	4	196,5	3,12	3,18	0,963	1,41	1,49	0,985	31,24	29,00	0,95
IT20	IT20a	207	335	0,79	0,71	1,014	0,85	0,73	1,005	20,87	13,46	0,66
IT31	IT311	1293	463,9	1,18	0,98	0,997	0,90	0,91	1,004	6,77	4,74	0,19
IT31	IT312	915	475,8	1,53	1,24	0,990	1,10	1,09	0,996	10,55	7,51	0,29
IT32	IT321	715	825,5	1,26	1,30	0,991	1,17	1,16	0,994	11,79	8,36	0,43
IT32	IT322	355	791,1	1,86	1,45	0,986	1,46	1,37	0,994	17,02	12,58	0,60
IT32	IT323	143	211,1	1,32	1,06	0,998	0,99	0,96	1,008	25,15	19,96	0,84
IT32	IT324	255	788,8	0,81	0,82	1,002	0,89	0,86	1,000	20,59	15,94	0,70
IT32	IT325	635	814,9	0,54	0,64	1,002	0,61	0,70	1,003	13,80	9,90	0,50
IT32	IT326	875	851,5	0,90	0,99	1,011	0,83	0,86	1,010	11,33	8,01	0,41
IT32	IT327	298	243,4	1,29	1,41	0,987	1,21	1,24	0,989	19,86	14,72	0,70
IT33	IT331	372	281,6	0,25	0,66	1,014	0,57	0,79	1,001	12,79	8,62	0,42
IT33	IT332	315	519,6	0,95	1,00	1,011	1,46	1,27	0,998	15,34	10,62	0,50
IT33	IT333	184	138,6	0,42	0,76	1,003	0,81	0,92	0,993	18,94	13,72	0,61
IT33	IT334	139	247,1	0,59	0,55	1,009	0,63	0,73	1,006	25,90	18,33	0,85
IT40	IT401	187	266,5	0,74	0,61	0,992	0,99	0,87	0,992	24,12	17,54	0,70
IT40	IT402	436	398,5	1,30	1,24	0,993	1,11	1,10	0,998	18,86	12,33	0,54
IT40	IT403	470	452,6	0,81	0,88	1,003	0,89	0,90	1,001	18,01	11,45	0,51
IT40	IT404	768	629,2	0,85	0,96	1,003	0,85	0,93	1,004	14,55	8,91	0,41
IT40	IT405	260	919,5	0,94	0,71	1,011	0,83	0,73	1,013	22,86	16,10	0,66
IT40	IT406	5	348,2	0,00	0,00	1,019	1,22	1,11	0,989	33,18	30,66	1,02
IT40	IT407	462	351,4	0,91	0,93	1,004	0,90	0,93	1,002	17,71	11,37	0,50
IT40	IT408	403	355,5	1,68	1,71	0,986	1,40	1,59	0,988	19,70	12,98	0,56
IT40	IT409	8	273,4	0,00	0,00	1,022	1,28	1,39	0,984	34,04	31,70	1,04
IT51	IT511	4	199,5	0,00	1,79	1,004	1,54	1,53	0,978	33,76	31,63	1,05
IT51	IT512	249	375,4	0,26	0,41	1,025	0,46	0,49	1,012	16,15	13,46	0,62
IT51	IT513	661	270	1,24	1,52	0,998	1,27	1,27	0,994	10,74	8,73	0,44
IT51	IT514	930	955,2	0,56	0,70	1,012	0,60	0,63	1,009	7,89	6,35	0,33
IT51	IT515	9	229,2	0,69	0,49	1,034	0,98	1,03	1,001	33,26	31,19	1,04
IT51	IT516	234	334,1	0,97	1,41	1,002	1,09	1,17	0,998	17,76	14,92	0,67
IT51	IT517	244	387	1,24	1,37	0,997	1,09	1,18	0,996	17,63	14,86	0,66
IT51	IT518	13	322,7	0,00	0,00	1,028	1,13	1,07	0,995	29,03	27,42	0,90
IT51	IT519	15	253,4	2,33	3,07	0,980	1,12	1,05	1,005	30,06	28,27	0,93
IT51	IT51a	399	215,5	2,29	2,35	0,978	2,26	1,91	0,979	14,08	11,55	0,56

(cont)

Table A.7 (cont)

				INPUT			OUTPUT			OUTPUT		
	Nuts3	sample	Pop.	input ratio (NUTS3/NUTS2)			raked			%SE(ratio)		
		size	mil.	**HCR_C**	**HCR_N2**	**logEqInc**	**HCR_C**	**HCR_N2**	**logEqInc**	**HCR_C**	**HCR_N2**	**logEqInc**
	SAE Model 3											
IT52	IT521	658	615	0,86	0,59	0,991	0,92	0,93	1,001	8,51	6,59	0,32
IT52	IT522	758	223	1,26	0,81	0,985	1,23	1,19	0,996	7,77	5,99	0,29
IT53	IT531	838	346	1,00	0,95	0,996	1,13	1,11	0,999	6,72	5,11	0,27
IT53	IT532	496	445,3	0,61	0,55	1,007	0,65	0,67	1,010	11,70	9,07	0,45
IT53	IT533	234	303,5	1,24	1,06	0,992	1,11	1,11	0,996	17,18	13,79	0,63
IT60	IT601	432	293	1,39	1,42	0,977	1,53	1,51	0,976	8,10	8,67	0,57
IT60	IT603	2102	3833,3	0,76	0,74	1,012	0,86	0,85	1,008	2,29	2,47	0,18
IT60	IT604	305	511,8	1,17	1,17	0,983	1,22	1,24	0,979	8,02	8,60	0,56
IT60	IT605	626	494,2	1,26	1,30	0,985	1,43	1,48	0,980	5,58	6,01	0,42
IT71	IT711	1	303,7	0,00	0,00	1,021	1,08	1,10	0,995	34,40	32,74	1,06
IT71	IT712	665	291,5	1,11	1,26	0,997	1,43	1,47	0,992	5,93	7,22	0,32
IT71	IT713	217	294,7	0,45	0,38	1,021	0,64	0,66	1,010	12,61	14,74	0,60
IT71	IT714	667	390,3	0,70	0,73	1,007	0,89	0,82	1,002	4,98	6,08	0,27
IT72	IT722	1499	236,1	1,00	1,00	1,000	1,08	1,09	0,997	0,00	0,00	0,00
IT80	IT801	900	856,3	1,17	1,23	0,987	1,10	1,10	0,992	9,34	8,10	0,36
IT80	IT802	756	293,1	0,70	0,55	1,022	0,69	0,53	1,023	10,59	9,18	0,41
IT80	IT803	1486	3099,6	1,17	1,33	0,991	1,10	1,18	0,995	7,12	6,13	0,28
IT80	IT804	479	440,3	0,90	0,72	1,010	0,78	0,62	1,016	13,16	11,48	0,50
IT80	IT805	1044	1092,2	0,88	0,81	1,002	0,82	0,69	1,007	8,47	7,33	0,33
IT91	IT911	883	693,2	1,50	1,64	0,967	1,42	1,47	0,975	9,01	8,07	0,45
IT91	IT912	923	1578,3	0,86	0,90	1,009	0,85	0,85	1,011	8,70	7,81	0,43
IT91	IT913	1107	587,4	0,70	0,53	1,017	0,69	0,55	1,017	7,43	6,65	0,38
IT91	IT914	288	411,3	0,90	0,91	1,005	0,88	0,86	1,006	14,35	12,98	0,64
IT91	IT915	361	815,8	1,22	1,41	0,976	1,21	1,28	0,986	14,10	12,75	0,63
IT92	IT921	1139	399,5	0,67	0,62	1,024	0,87	0,78	1,006	4,57	4,00	0,20
IT92	IT922	493	206	1,03	1,18	1,002	1,26	1,43	0,988	11,31	9,99	0,46
IT93	IT931	787	744,1	1,36	1,30	0,971	1,17	1,20	0,980	7,87	6,88	0,35
IT93	IT932	157	173,7	1,39	1,67	0,984	1,18	1,44	0,991	19,72	17,57	0,77
IT93	IT933	216	381,7	1,31	1,55	0,994	1,06	1,16	1,004	16,71	14,90	0,65
IT93	IT935	574	571,3	0,72	0,48	1,028	0,64	0,46	1,027	9,35	8,18	0,41
ITA0	ITA01	1409	433,5	0,97	1,00	0,996	0,96	1,03	0,998	5,86	5,73	0,27
ITA0	ITA02	806	1235,9	0,95	0,91	1,003	0,96	1,00	1,003	8,45	8,25	0,38
ITA0	ITA03	198	675,5	0,97	0,32	1,020	0,82	0,51	1,016	17,97	17,44	0,71
ITA0	ITA04	335	467,9	1,22	1,07	0,974	1,20	1,21	0,979	12,64	12,31	0,54
ITA0	ITA07	944	1101,1	1,24	1,40	0,981	1,23	1,38	0,984	7,96	7,77	0,36
ITA0	ITA09	523	402,6	0,64	0,42	1,041	0,58	0,43	1,036	10,90	10,63	0,48
ITB0	ITB01	1146	459,2	0,63	0,48	1,027	0,60	0,48	1,027	6,88	6,02	0,31
ITB0	ITB02	253	268,7	1,41	1,44	0,973	1,21	1,23	0,989	16,25	14,49	0,65
ITB0	ITB03	283	156,9	1,33	1,59	0,973	1,25	1,35	0,986	13,75	12,19	0,57
ITB0	ITB04	1466	765,2	1,17	1,21	0,988	1,12	1,16	0,991	5,10	4,45	0,23
	Regression Model 3											
IT11	IT112	0	180,6				1,04	0,95	1,001			
IT20	IT204	0	177,5				1,33	1,33	0,985			
IT20	IT20b	0	375,1				0,94	0,88	1,003			
IT53	IT534	0	370,3				1,20	1,21	0,993			
IT60	IT602	0	150,9				1,53	1,42	0,982			
IT72	IT721	0	91,5				0,79	0,76	1,009			
IT93	IT934	0	176,1				1,14	1,15	0,994			
ITA0	ITA05	0	282,4				1,16	1,20	0,992			
ITA0	ITA06	0	181				1,23	1,25	0,992			
ITA0	ITA08	0	302,4				0,77	0,76	1,011			
	No division in Nuts2											
	IT120	1183	120,5									

Table A.8 Italian regions (up to NUTS3 level)

	Name	population (2000)
IT1	**Nord-Ovest**	**6032,5**
IT11	Piemonte	4288,6
IT111	Torino	2214,6
IT112	Vercelli	180,6
IT113	Biella	189,4
IT114	Verbano-cusio-ossola	160,7
IT115	Novara	344,3
IT116	Cuneo	558,2
IT117	Asti	210,5
IT118	Alessandria	430,4
IT12	Valle daosta	120,5
IT13	Liguria	1623,4
IT131	Imperia	216,4
IT132	Savona	279,7
IT133	Genova	905,5
IT134	La spezia	221,8
IT2	**Lombardia**	**9093,6**
IT20	Lombardia	9093,6
IT201	Varese	818,4
IT202	Como	541,0
IT203	Lecco	310,6
IT204	Sondrio	177,5
IT205	Milano	3765,8
IT206	Bergamo	969,8
IT207	Brescia	1105,6
IT208	Pavia	498,4
IT209	Lodi	196,5
IT20a	Cremona	335,0
IT20b	Mantova	375,1
IT3	**Nord-Est**	**6652,9**
IT31	Trentino-Alto adige	939,7
IT311	Bolzano-Bozen	463,9
IT312	Trento	475,8
IT32	Veneto	4526,3
IT321	Verona	825,5
IT322	Vicenza	791,1
IT323	Belluno	211,1
IT324	Treviso	788,8
IT325	Venezia	814,9
IT326	Padova	851,5
IT327	Rovigo	243,4
IT33	Friuli-venezia giulia	1186,9
IT331	Pordenone	281,6
IT332	Udine	519,6
IT333	Gorizia	138,6
IT334	Trieste	247,1
IT4	**Emilia Romagna**	**3994,9**
IT40	Emilia Romagna	3994,9
IT401	Piacenza	266,5
IT402	Parma	398,5
IT403	Reggio nell'emilia	452,6
IT404	Modena	629,2
IT405	Bologna	919,5
IT406	Ferrara	348,2
IT407	Ravenna	351,4
IT408	Forlì-cesena	355,5
IT409	Rimini	273,4
IT5	**Centro**	**5845,1**
IT51	Toscana	3542,0
IT511	Massa-carrara	199,5
IT512	Lucca	375,4
IT513	Pistoia	270,0
IT514	Firenze	955,2
IT515	Prato	229,2
IT516	Livorno	334,1
IT517	Pisa	387,0
IT518	Arezzo	322,7
IT519	Siena	253,4
IT51a	Grosseto	215,5
IT52	Umbria	838,0
IT521	Perugia	615,0
IT522	Terni	223,0
IT53	Marche	1465,1
IT531	Pesaro e urbino	346,0
IT532	Ancona	445,3
IT533	Macerata	303,5
IT534	Ascoli piceno	370,3
IT6	**Lazio**	**5283,2**
IT60	Lazio	5283,2
IT601	Viterbo	293,0
IT602	Rieti	150,9
IT603	Roma	3833,3
IT604	Latina	511,8
IT605	Frosinone	494,2
IT7	**Abruzzo-Molise**	**1607,8**
IT71	Abruzzo	1280,2
IT711	L'aquila	303,7
IT712	Teramo	291,5
IT713	Pescara	294,7
IT714	Chieti	390,3
IT72	Molise	327,6
IT721	Isernia	91,5
IT722	Campobasso	236,1
IT8	**Campania**	**5781,6**
IT80	Campania	5781,6
IT801	Caserta	856,3
IT802	Benevento	293,1
IT803	Napoli	3099,6
IT804	Avellino	440,3
IT805	Salerno	1092,2
IT9	**Sud**	**6738,3**
IT91	Puglia	4085,9
IT911	Foggia	693,2
IT912	Bari	1578,3
IT913	Taranto	587,4
IT914	Brindisi	411,3
IT915	Lecce	815,8
IT92	Basilicata	605,5
IT921	Potenza	399,5
IT922	Matera	206,0
IT93	Calabria	2046,9
IT931	Cosenza	744,1
IT932	Crotone	173,7
IT933	Catanzaro	381,7
IT934	Vibo valentia	176,1
IT935	Reggio di calabria	571,3
ITA	**Sicilia**	**5082,2**
ITA0	Sicilia	5082,2
ITA01	Trapani	433,5
ITA02	Palermo	1235,9
ITA03	Messina	675,5
ITA04	Agrigento	467,9
ITA05	Caltanissetta	282,4
ITA06	Enna	181,0
ITA07	Catania	1101,1
ITA08	Ragusa	302,4
ITA09	Siracusa	402,6
ITB	**Sardegna**	**1650,0**
ITB0	Sardegna	1650,0
ITB01	Sassari	459,2
ITB02	Nuoro	268,7
ITB03	Oristano	156,9
ITB04	Cagliari	765,2

Table A.9 European regions (up to NUTS2 level)

EU25	European Union (25 countries)
EU15	European Union (15 countries)
NMS10	New Member States
BE	Belgium
BE1	Région de Bruxelles-Capitale/Brussels Hoofdstedeli
BE10	Région de Bruxelles-Capitale/Brussels Hoofdstedel
BE2	Vlaams Gewest
BE21	Prov. Antwerpen
BE22	Prov. Limburg (B)
BE23	Prov. Oost-Vlaanderen
BE24	Prov. Vlaams Brabant
BE25	Prov. West-Vlaanderen
BE3	Région Wallonne
BE31	Prov. Brabant Wallon
BE32	Prov. Hainaut
BE33	Prov. Liège
BE34	Prov. Luxembourg (B)
BE35	Prov. Namur
CZ	Czech Republic
CZ0	Czech Republic
CZ01	Praha
CZ02	Strední Cechy
CZ03	Jihozápad
CZ04	Severozápad
CZ05	Severovýchod
CZ06	Jihovýchod
CZ07	Strední Morava
CZ08	Moravskoslezko
DK	Denmark
DK0	Denmark
DE	Germany (including ex-GDR from 1991)
DE1	Baden-Württemberg
DE11	Stuttgart
DE12	Karlsruhe
DE13	Freiburg
DE14	Tübingen
DE2	Bayern
DE21	Oberbayern
DE22	Niederbayern
DE23	Oberpfalz
DE24	Oberfranken
DE25	Mittelfranken
DE26	Unterfranken
DE27	Schwaben
DE3	Berlin
DE30	Berlin
DE4	Brandenburg
DE41	Brandenburg - Nordost
DE42	Brandenburg - Südwest
DE5	Bremen
DE50	Bremen
DE6	Hamburg
DE60	Hamburg
DE7	Hessen
DE71	Darmstadt
DE72	Gießen
DE73	Kassel
DE8	Mecklenburg-Vorpommern
DE80	Mecklenburg-Vorpommern
DE9	Niedersachsen
DE91	Braunschweig
DE92	Hannover
DE93	Lüneburg
DE94	Weser-Ems
DEA	Nordrhein-Westfalen
DEa1	Düsseldorf
DEa2	Köln
DEa3	Münster
DEa4	Detmold
DEa5	Arnsberg
DEX	Rheinland-Pfalz and Saarland
DEXa	Rheinland-Pfalz
DEXb	Saarland
DEc0	Saarland
DED	Sachsen
DEd1	Chemnitz
DEd2	Dresden
DEd3	Leipzig
DEE	Sachsen-Anhalt
DEe1	Dessau
DEe2	Halle
DEe3	Magdeburg
DEF	Schleswig-Holstein
DEf0	Schleswig-Holstein
DEG	Thüringen
DEg0	Thüringen
EE	Estonia
EE0	Estonia
EE00	Estonia
GR	Greece
GR1	Voreia Ellada
GR11	Anatoliki Makedonia, Thraki
GR12	Kentriki Makedonia
GR13	Dytiki Makedonia
GR14	Thessalia
GR2	Kentriki Ellada
GR21	Ipeiros
GR22	Ionia Nisia
GR23	Dytiki Ellada
GR24	Sterea Ellada
GR25	Peloponnisos
GR3	Attiki
GR30	Attiki
GR4	Nisia Aigaiou, Kriti
GR41	Voreio Aigaio
GR42	Notio Aigaio
GR43	Kriti
ES	Spain
ES1	Noroeste
ES11	Galicia
ES12	Principado de Asturias
ES13	Cantabria
ES2	Noreste
ES21	Pais Vasco
ES22	Comunidad Foral de Navarra
ES23	La Rioja
ES24	Aragón
ES3	Comunidad de Madrid
ES30	Comunidad de Madrid
ES4	Centro (ES)
ES41	Castilla y León
ES42	Castilla-la Mancha
ES43	Extremadura
ES5	Este
ES51	Cataluña
ES52	Comunidad Valenciana
ES53	Illes Balears
ES6	Sur

(cont)

Table A.9 (cont)

ES61	Andalucia	IT60	Lazio
ES62	Región de Murcia	IT7	Abruzzo-Molise
ES63	Ciudad Autónoma de Ceuta (ES)	IT71	Abruzzo
ES64	Ciudad Autónoma de Melilla (ES)	IT72	Molise
ES7	Canarias (ES)	IT8	Campania
ES70	Canarias (ES)	IT80	Campania
FR	France	IT9	Sud
FR1	Île de France	IT91	Puglia
FR10	Île de France	IT92	Basilicata
FR2	Bassin Parisien	IT93	Calabria
FR21	Champagne-Ardenne	ITA	Sicilia
FR22	Picardie	ITA0	Sicilia
FR23	Haute-Normandie	ITB	Sardegna
FR24	Centre	ITB0	Sardegna
FR25	Basse-Normandie	CY	Cyprus
FR26	Bourgogne	CY0	Cyprus
FR3	Nord - Pas-de-Calais	CY00	Cyprus
FR30	Nord - Pas-de-Calais	LV	Latvia
FR4	Est	LV0	Latvia
FR41	Lorraine	LV00	Latvia
FR42	Alsace	LT	Lithuania
FR43	Franche-Comté	LT0	Lithuania
FR5	Ouest	LT00	Lithuania
FR51	Pays de la Loire	LU	Luxembourg (Grand-Duché)
FR52	Bretagne	LU0	Luxembourg (Grand-Duché)
FR53	Poitou-Charentes	LU00	Luxembourg (Grand-Duché)
FR6	Sud-Ouest	HU	Hungary
FR61	Aquitaine	HU1	Közép-Magyarország
FR62	Midi-Pyrénées	HU10	Közép-Magyarország
FR63	Limousin	HU2	Dunántúl
FR7	Centre-Est	HU21	Közép-Dunántúl
FR71	Rhône-Alpes	HU22	Nyugat-Dunántúl
FR72	Auvergne	HU23	Dél-Dunántúl
FR8	Méditerranée	HU3	Alföld és Észak
FR81	Languedoc-Roussillon	HU31	Észak-Magyarország
FR82	Provence-Alpes-Côte d'Azur	HU32	Észak-Alföld
FR83	Corse	HU33	Dél-Alföld
FR9	French overseas departments (FR)	MT	Malta
FR91	Guadeloupe (FR)	MT0	Malta
FR92	Martinique (FR)	MT00	Malta
FR93	Guyane (FR)	NL	Netherlands
FR94	Reunion (FR)	NL1	Noord-Nederland
IE	Ireland	NL11	Groningen
IE0	Ireland	NL12	Friesland
IE01	Border, Midlands and Western	NL13	Drenthe
IE02	Southern and Eastern	NL2	Oost-Nederland
IT	Italy	NL21	Overijssel
IT1	Nord Ovest	NL22	Gelderland
IT11	Piemonte	NL23	Flevoland
IT12	Valle daosta	NL3	West-Nederland
IT13	Liguria	NL31	Utrecht
IT2	Lombardia	NL32	Noord-Holland
IT20	Lombardia	NL33	Zuid-Holland
IT3	Nord Est	NL34	Zeeland
IT31	Trentino-Altoadige	NL4	Zuid-Nederland
IT32	Veneto	NL41	Noord-Brabant
IT33	Friuli-venezia giulia	NL42	Limburg (NL)
IT4	Emilia-Romagna	AT	Austria
IT40	Emilia-Romagna	AT1	Ostösterreich
IT5	Centro (I)	AT11	Burgenland
IT51	Toscana	AT12	Niederösterreich
IT52	Umbria	AT13	Wien
IT53	Marche	AT2	Südösterreich
IT6	Lazio	AT21	Kärnten

Table A.9 (cont)

AT22	Steiermark
AT3	Westösterreich
AT31	Oberösterreich
AT32	Salzburg
AT33	Tirol
PL	Poland
PL1	Centralny
PL11	Lódzkie
PL12	Mazowieckie
PL2	Poludniowy
PL21	Malopolskie
PL22	Slaskie
PL3	Wschodni
PL31	Lubelskie
PL31	Chelmsko-zamojski
PL32	Podkarpackie
PL33	Swietokrzyskie
PL34	Podlaskie
PL4	Pólnocno-Zachodni
PL41	Wielkopolskie
PL42	Zachodniopomorskie
PL43	Lubuskie
PL5	Poludniowo-Zachodni
PL51	Dolnoslaskie
PL52	Opolskie
PL6	Pólnocny
PL61	Kujawsko-Pomorskie
PL62	Warminsko-Mazurskie
PL63	Pomorskie
PT	Portugal
PT1	Continente (PT)
PT11	Norte
PT12	Centro (PT)
PT13	Lisboa
PT14	Alentejo
PT15	Algarve
PT2	Região Autónoma dos Açores (PT)
PT20	Região Autónoma dos Açores (PT)
PT3	Região Autónoma da Madeira (PT)
PT30	Região Autónoma da Madeira (PT)
SI	Slovenia
SI0	Slovenia
SI00	Slovenia
SK	Slovak Republic
SK0	Slovak Republic
SK01	Bratislavský
SK02	Západné Slovensko
SK03	Stredné Slovensko
SK04	Východné Slovensko
FI	Finland
FI1	Manner-Suomi
FI13	Itä-Suomi
FI18	Etelä-Suomi
FI19	Länsi-Suomi
FI1a	Pohjois-Suomi
FI2	Åland
FI20	Åland
SE	Sweden
SE0	Sverige
SE01	Stockholm
SE02	Östra Mellansverige
SE04	Sydsverige
SE06	Norra Mellansverige
SE07	Mellersta Norrland
SE08	Övre Norrland
SE09	Småland med öarna
SE0a	Västsverige
UK	United Kingdom
UK1	North
UK11	Tees Valley and Durham
UK13	Northumberland, Tyne and Wear
UK12	Cumbria
UK8	North West (UK)
UK81	Cheshire
UK82	Greater Manchester
UK83	Lancashire
UK84	Merseyside
UK2	Yorkshire and The Humber
UK21	East Riding and North Lincolnshire
UK22	North Yorkshire
UK23	South Yorkshire
UK24	West Yorkshire
UK3	East Midlands
UK31	Derbyshire and Nottinghamshire
UK32	Leicestershire, Rutland and Northants
UK33	Lincolnshire
UK7	West Midlands
UK71	Herefordshire, Worcestershire and Warks
UK72	Shropshire and Staffordshire
UK73	West Midlands
UK4	East Anglia
UK5	South East
UK51	Bedfordshire, Hertfordshire
UK54	Essex
UK55	London
UK52	Berkshire, Bucks and Oxfordshire
UK53	Surrey, East and West Sussex
UK56	Hampshire and Isle of Wight
UK57	Kent
UK6	South West
UK61	Gloucestershire, Wiltshire and North Somerset
UK63	Dorset and Somerset
UK62	Cornwall and Devon
UK9	Wales
UKA	Scotland
UKA4	North Eastern Scotland
UKA1	Eastern Scotland
UKA2	South Western Scotland
UKA3	Highlands and Islands
UKB	Northern Ireland
UKB0	Northern Ireland
BG	Bulgaria
BG01	Severozapaden
BG02	Severen Tsentralen
BG03	Severoiztochen
BG04	Yugozapaden
BG05	Yuzhen Tsentralen
BG06	Yugoiztochen
RO	Romania
RO01	Nord-Est
RO02	Sud-Est
RO03	Sud
RO04	Sud-Vest
RO05	Vest
RO06	Nord-Vest
RO07	Centru
RO08	Bucuresti

References

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

Atkinson A., Rainwater L., Smeeding T. (1995), *Income Distribution in OECD Countries*, Paris: OECD.

Battese G.E., Harter R.M., Fuller W.A. (1988), An error-component model for prediction of county crop areas using survey and satellite data, *Journal of the American Statistical Association*, 83, 401, pp. 28-36.

Berger Y., Skinner C. (2003), Variance estimation for a low income proportion, *Applied Statistics*, 52 (4), pp 457-468.

Berthoud R. (2004), *Patterns of poverty across Europe*. The Policy Press.

Betti, G., Ballini, F., Neri, L. (2003), Poverty and Inequality Mapping in Albania, *Final Report to the World Bank*, July 2003.

Betti G., Verma V. (1999), Measuring the degree of poverty in a dynamic and comparative context: a multi-dimensional approach using fuzzy set theory, *Proceedings, ICCS-VI*, Vol. 11, pp. 289-301, Lahore, Pakistan.

Betti G., Verma V. (2002), Non-monetary or Lifestyle Deprivation, in Giorgi L. and Verma V. (eds.), European Social Statistics: Income, Poverty and Social Exclusion: 2nd Report. Luxembourg: Office for Official Publications of the European Communities, pp. 76-92.

Carlquist, T. (2004), Rural/urban classification of NUTS3 regions according to 2 concepts. Unpublished memorandum, Eurostat unit E4, Sector for Regional Statistics, 2 December 2004.

Cerioli A., Zani S. (1990), A Fuzzy Approach to the Measurement of Poverty, in Dagum C. and Zenga M. (eds.), *Income and Wealth Distribution, Inequality and Poverty, Studies in Contemporary Economics*. Berlin: Springer Verlag, pp. 272-284.

Chakravarty S. R., D'Ambrosio C. (2002), *The Measurement of Social Exclusion*, mimeo.

Cheli B., Lemmi A. (1995), A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty, *Economic Notes*, 24, pp. 115-134.

Deaton A. (1997), *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. John Hopkins Press and The World Bank, Washington, D.C.

Elbers C., Lanjouw J.O., Lanjouw P. (2002), Micro-level Estimation of Welfare, *Working Paper n. 2911, The World Bank*, Washington, D.C..

Elbers C., Lanjouw J.O., Lanjouw P. (2003), Micro-level Estimation of Poverty and Inequality. *Econometrica*, 71, pp. 355-364.

Eurarea (2004), Enhancing small area estimation techniques to meet European needs. Project IST-2000-26290. Final report. http://www.statistics.gov.uk/eurarea/

European Commission (2000), The Urban Audit: Towards the Benchmarking of Quality of Life in 58 European Cities. Luxembourg: Office for Official Publications of the European Communities.

European Commission (2003), *Joint Report on Social Inclusion*. COM(2003) 773 final.

European Commission (2004), *Report on Social Inclusion 2004. An Analysis of the National Action Plans on Social Inclusion (2004-2006) submitted by the 10 new Member States.*

Eurostat (2000), Sampling Errors: Proportions, means and other ratios estimated from the household questionnaire, *DOC. PAN 138/2000* (unpublished)

Eurostat (2003), Monetary Poverty in EU Acceding and Candidate Countries, *Statistics in Focus*, Theme 3- 21/2003.

Eurostat (2003a), General indicators on transition from work to school. Youth transitions from education to working life in Europe Part I, *Statistics in Focus*, Theme 3 - 4/2003.

Eurostat (2003b), School leavers in Europe and the labour market effects of job mismatches. Youth transitions from education to working life in Europe Part II, *Statistics in Focus*, Theme 3 - 5/2003.

Eurostat (2003c), Young people's social origin, educational attainment and labour market outcomes in Europe. Youth transitions from education to working life in Europe Part III. *Statistics in Focus*, Theme 3 - 6/2003.

Eurostat (2004), Monetary Poverty in New Member States and Candidate Countries, *Statistics in Focus*, Theme 3-12/2004.

Eurostat (2004a), European Regional Statistics: Reference Guide.

Eurostat (2004b), Regions: Statistical Yearbook 2004.

Eurostat (n.d.), Mapping of national education programmes to ISCED 97 for school/academic year 1999/2000.

Fahmy E., Gordon D. (2002), *Mapping deprivation in the South West, University of Bristol*: Townsend Centre for International Poverty Research.

Fay R.E., Herriot R.A. (1979), Estimates of income for small places: an application of James-Stein procedure to census data, *Journal of the American Statistical Association*, 74, pp. 269-277.

Gallego F. (n.d.), Mapping rural/urban areas from population density grids. Institute for Environment and Sustainabilty, JRC, Ispra, Italy.

Giorgi L, Verma V. (2002), European Social Statistics: Income, Poverty and Social Exclusion: 2nd Report. Luxembourg: Office for Official Publications of the European Communities.

Glennerster, H., Lupton, R., Norden, P., Power, A. (1999), Poverty, social exclusion and neighbourhood: studying the area bases of social exclusion. CASEpaper/22. Centre for Analysis of Social Exclusion, London School of Economics

Gordon D. (1995), Census based deprivation indices: their weighting and validation, *Journal of Epidemiology and Community Health*, 49, pp. 39-44.

Gordon D., Townsend P. (1990), Measuring the Poverty Line, *Radical Statistics*, 47, pp. 5-12.

Gordon D., Pantazis C. (Eds.) (1997), *Breadline Britain in the 1990s*, Aldershot, Ashgate.

Gordon D. (1995), Census based deprivation indices: their weighting and validation. *Journal of Epidemiology and Community Health*, 49 (Suppl. 2), S39-44.

Gosh M., Rao J.N.K (1994), Small Area Estimation: An Appraisal (with discussion), *Statistical Science*, 9 (1), 55-93.

Handerson C.R. (1950), Estimation of Genetic Parameters, *Annals of Mathematical Statistics*, 21, pp. 309-310.

Hansen M., Hurwitz W., Madow W. (1953), *Sample Survey Methods and Theory*. Wiley.

Iannelli C., Soro-Bonmatì A. (2001), The Transition from School to Work in Southern Europe: the Cases of Italy and Spain. *Working Paper (7), TSER project: Comparative Analysis of Transitions from Education to Work in Europe.*

Kackar R.N., Harville D.A. (1984), Approximation for standard errors of estimators for fixed and random effects in mixed models, *Journal of the American Statistical Association*, 79, pp. 853-862.

Kangas O., Ritakallio V.M. (2004), Relative to what? Cross-national picture of European poverty measured by regional, national and European standards. *Working Paper no. 384*. Luxembourg Income Study.

Kendall M. (1970), *Rank Correlation Methods* (4th edition). London: Griffin.

Kish L. (1965), *Survey Sampling. New York*. Wiley.

Kish L. (1989), Sampling Methods for Agricultural Surveys. *Statistical Development Series*, No. 3. Rome: FAO

Levinson R. (2001), Sample design for the 2002 Living Standards Measurement Survey (LSMS), *Final Report to the World Bank*, November 2001.

Maasoumi E., Nickelsburg G. (1988), Multivariate measures of well-being and an analysis of inequality in the Michigan data, *Journal of Business & Economic Statistics*, 6, pp. 326-334.

Micklewright J. (2001), in CASE Report of the workshop held on 19 July 2000 organised by Department of Social Security and Centre for Analysis of Social Exclusion, LSE.

National Research Council (1998), *Small-Area Estimates of School-Age Children in Poverty*. Washington DC: National Academy Press (ISBN 0-309-6045-1).

Nolan B., Whelan C.T. (1996), *Resources, deprivation and poverty*. Oxford, Clarendon Press.

Office of the Deputy Prime Minister (n.d.), *Urban and Rural Definitions: A Users' Guide*. London.

Office of the Deputy Prime Minister (2004), *The English Indices Of Deprivation 2004: Summary (revised)*. London

OECD (1994), *Creating Rural Indicators*. Paris

OECD (1998), *Integrating Distressed Urban Areas*. Paris

Payne S., Henson B., Gordon D., Forrest R. (1996), *Poverty and deprivation in West Corwall in the 1990s. A report to the West Cornwall Initiative*. Bristol, Statistical Monitoring Unit.

Raghunathan T.E., Lepkowski J., Van Voewyk J., Solenberger P. (2001), A Multivariate Technique for Imputing Missing Values Using a Sequence of Regression Models, *Survey Methodology*, 27, pp. 85-95.

Ram R. (1982), Composite indices of physical quality of life, basic needs fulfilment and income. A principal component representation, *Journal of Development Economics*, 11, pp. 227-248.

Rao J.N.K (2003), *Small Area Estimation*. Wiley, London.

Rodrigues F.C. (1999), Income Distribution and Poverty in Portugal [1994/95], A Comparison between the European Community Household Panel and the Household Budget Survey, CISEP, ISEG/Universidade Técnica de Lisboa.

Stewart K (2002), Measuring well-being and exclusion in Europe's regions. *CASE paper no. 53*. Centre for Analysis of Social Exclusion, London School of Economics.

Stewart K (2003), Monitoring social exclusion in Europe's regions. *Journal of European Social Policy*, vol 13(4), pp. 335-356.

United States (1978), An error profile: Employment as measured by the Current Population Survey. *Statistical Policy Working Paper*, 3. Washington DC: OFSPS.

Verma V. (1992), *Sampling Errors in Household Surveys: A Technical Study*. New York: United Nations Department for Economic and Social Information and Analysis. (Statistical Division INT-92-P80-15E).

Verma, V. (2001), *EU-SILC Sampling Guidelines*. Report prepared for Eurostat.

Verma V. (2004), Sampling errors and design effects for poverty measures and other complex statistics. Invited Lecture, VII Convegno Internazionale, Metodi Quantitativi per le Scienze Applicate, University of Siena.

Verma V., Scott C., O'Muircheartaigh C. (1980), Sample designs and sampling errors for the World Fertility Survey. *Journal of the Royal Statistical Society*, series A vol. 143(4).

Verma V., Thanh Lê (1996), An Analysis of Sampling Errors for the Demographic and Health Surveys. *International Statistical Review*, 64(3), pp. 265-294.

World Bank (2002), Basic Information Document, Living Standard Measurement Study, Albania, Development Research Group.

World Bank (2003), Construction of the consumption aggregate and estimation of the poverty line, LSMS 2002 - Albania.