Indicators of Social Exclusion and Poverty in Europe's Regions

Vijay Verma, Gianni Betti, Michela Natilli, Achille Lemmi

Working Paper n. 59, April 2006

Indicators of Social Exclusion and Poverty in Europe's Regions[1]

Verma V., Betti G., Natilli M., Lemmi A.
verma@unisi.it; betti2@unisi.it; natilli@unisi.it; lemmi@unisi.it
Dipartimento di Metodi Quantitativi, Università degli Studi di Siena

Abstract

This paper provides an overview of a methodology and application for the construction of regional (sub-national) indicators of poverty and social exclusion in Europe. Starting from Laeken Indicators defined at the country level, proposals are made for their regional adaptation. A strategy is developed and implemented for the construction of regional indicators. It involves making use of micro-level survey data and region-level data from other sources to produce regional indicators using composite area-level EBLUP methodology. Illustrative results are shown of estimates down to NUTS2 level for most EU regions and to NUTS3 level in the case of Italy.

1. Introduction

Indicators of poverty and social exclusion are an essential tool for monitoring progress in the reduction of these problems. In the EU-wide context, these indicators need to be comparable across countries and time. For this purpose, the European Commission has adopted a common set of indicators, referred to as the Laeken Indicators. A critical review of these indicators provides the stating point of our research (Section 2). Hitherto, most of these indicators have been defined only at the national level. These are not necessarily appropriate or sufficient for regional analysis. In Section 3, we make specific proposals in going from the country list to a regional list of indicators.

The strategy we have adopted for the construction of regional indicators has three fundamental aspects (Section 4): making the best use of available survey data (in this case, mainly ECHP); exploiting to the maximum 'meso' data (here regional tabulations

---

[1] This paper provides an overview of a methodology and application for the construction of regional (sub-national) indicators of poverty and social exclusion in Europe. It is based on research carried out during 2004-2005 under project Regional Indicators to Reflect Social Exclusion and Poverty (VT/2003/43) of the European Commission, Employment and Social Affairs DG. The project was executed by Centro di Ricerca Interdipartimentale sulla Distribuzione del Reddito (C.R.I.DI.RE.) at Università degli Studi di Siena. Achille Lemmi acted as the Project Director, and Vijay Verma as the Research Director for the project. Other main contributors to the research included Gianni Betti, Anna Mulas, Michela Natilli, Laura Neri, and Nicola Salvati.

of NewCronos); and using the two sources in combination to produce the best possible estimates for regions using appropriate small area estimation (SAE) techniques. Implementation of SAE methodology is described in Section 5. Two distinct approaches are involved, depending on the data situation. Where micro-level data containing the necessary identifiers for areas and area-level covariates are available, composite area-level EBLUP estimation methodology has been used. In other countries, simply the regression coefficient estimated from the above are used for prediction.

The main application involves the production of estimates of country-to-NUTS1 and NUTS1-to-NUTS2 ratios. A very small selection of the results on indicators at NUTS2 level is shown in Section 6. Because of the more favourable data availability situation in the case of Italy, it was possible to go one step further, to NUTS3 level estimation (Section 7).

2. Laeken indicators of poverty and social exclusion

Indicators of poverty and social exclusion are an essential tool for monitoring progress in the reduction of these problems. In the EU-wide context, these indicators are most useful when they are comparable across countries, so that the situation in any EU Member State can be evaluated in relation to the situation in other countries. In order to monitor trends, these indicators also need to be comparable over time. In order to meet these objectives, the European Commission has adopted a common set of indicators of poverty and social exclusion, with standardised definitions and procedures for their construction. These are referred to as the Laeken Indicators, deriving their name from the location of the European Council meeting where they were adopted. The set of common indicators is supplemented by country-specific indicators, chosen flexibly according to the requirements and data availability in individual countries. Hitherto, most of the indicators have been defined and constructed only at the national level, except for occasional breakdown for special subpopulations such as children, other groups by age and gender, or different household types.

Our main interest in this paper is to define indicators at the regional (sub-national) level, and describe and illustrate the statistical methodology for their construction.

Some of the country-level indicators can be usefully classified down to the regional level in their existing form; some other may need modification (simplification) before such classification. There are, however, also country-level indicators which are not suitable (meaningful, useful, feasible) for regional breakdown. It is also necessary to consider additional, specifically regional indicators which are not covered in the country-level list. Still, the established set of country-level indicators provides the basis for developing indicators suitable for the regional level. Therefore, we begin by describing in this section the indicators defined for country-level application. Selection, adaptation, and supplementation of these indicators for regional application will be taken up in the next section.

2.1. Background to the Laeken Indicators

Following a series of consultations, the Laeken European Council endorsed a first set of 18 indicators of social exclusion and poverty, organised in a two-level structure of 10 primary indicators – covering the broad fields considered to be the most important elements leading to social exclusion – and 8 secondary indicators – intended to support the lead indicators and describe other dimensions of the problem.

Countries have supplemented these standard primary and secondary indicators by tertiary country-specific indicators, often based on different data sources, and/or using alternative definitions. Such indicators are meant to provide useful complementary information to that of relative poverty risk. The sub-national dimension of poverty and social exclusion is in some instances described through a breakdown by region or type of place (urban/rural) of the common indicators. These indicators are mostly 'performance-related'. In addition, countries also use 'input-related' indicators; 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, or 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. A common example is benefit-dependency indicators.

The EU list of indicators is of course not fixed: the EU Social Protection Committee undertakes, through its Indicators Sub-Group, further development of the common indicators. The methodological principles guiding the selection of indicators are the following: "The portfolio of EU indicators should be balanced across different dimensions and 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 original list of indicators has continued to be extended. However, in the current list of common indicators as a whole, the concept of social exclusion remains 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.

2.2. Laeken indicators of cross-sectional measures of income poverty

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

Primary indicators

- Indicators 1: At-risk-of-poverty rate, broken down by various characteristics, such as (a) age and gender, (b) most frequent activity status, (c) household type, and (d) accommodation tenure status, (e) gender among workers, and (f) work intensity.
- 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

Secondary indicators

- o Indicator 11: Dispersion around the at-risk-of-poverty threshold
- o Indicator 13: At-risk-of-poverty rate before social transfers, by age and gender
- o Indicator 14: Inequality of income distribution Gini coefficient

From time to time, these indicators are refined and other ones added for specific EU publications. Here are some examples.

- o Poverty risk by work intensity of households
- o At-risk-of-poverty rate among workers by gender (derived from Indicator 1b)
- o Relative median income ratio, by age
- o Structure of income, by age
- o Further breakdown of indicators (1d, 2, 11) by age

The above indicators involve the following definitions.

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 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.[2]

At-risk-of poverty rate broken down according to certain characteristics

The 'at-risk-of poverty rate (after social transfers)' broken down by population group defined in terms of certain characteristics 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'. It is important to note that 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)

A set of illustrative values of the 'at-risk-of-poverty threshold' for different types of households. An illustrative value is obtained by multiplying the threshold for the total population, by equivalised size of the household type being considered.

---

[2] For any indicator, the EU average, when required, is calculated as a weighted average of the indicators established for each country, where the weight is proportional to the number of persons living in private households in each country.

Inequality of income distribution: S80/S20, the income quintile share ratio

The ratio of the sum of equivalised disposable income received by 20% of the country's population having the highest equivalised disposable income (top inter-quintile interval) to that received by 20% of the country's population having the lowest equivalised disposable income (lowest inter-quintile interval).

Relative median at-risk-of-poverty gap

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.

Dispersion around the at-risk-of-poverty threshold

The proportion of 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

The percentage of persons whose equivalised disposable income, before receiving social transfers including any old-age and survivors' benefits, is below the national 'at-risk-of-poverty threshold'.

Inequality of income distribution: Gini coefficient

A measure of 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

Work intensity of a household refers to the number of months that all working-age household members have worked during the income reference year, as a proportion of the total number of working months theoretically available to the household.

In-work poverty risk

The proportion at risk of poverty among individuals who are classified as employed or self-employed according to the situation of most frequent activity status.

2.3. Longitudinal indicators of poverty and deprivation

In the longitudinal dimension, indicators may be designed to capture the experience of poverty and deprivation at any time during a certain period, or persistently or continuously over the period. 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. It is defined as the share of persons with an equivalised disposable income below the at-risk-of-poverty threshold in the current year, and also during at least two of the preceding three years. Hence it is based on longitudinal data covering four consecutive years.

Indicators 12 and 15 are among the "secondary" indicators. Indictor 12 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, normally the appropriate inflation rate will be that for the period (t-4) to (t-1). Indicator 15 is merely a variant of standard Laeken indicator 3, except for using a lower poverty threshold.

2.4. Other Laeken indicators

Apart from monetary poverty indicators described above, 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 include the following.

- Indicator 5: Regional cohesion (dispersion of regional employment rates), by gender
- 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 rate, by gender
- Indicator 17: Very long-term unemployment rate, by gender
- Indicator 18: Persons with low educational attainment, by age and gender (also, low reading literacy performance of pupils)

Indicator 5, regional cohesion (dispersion of regional employment rates), refers to the coefficient of variation of regional employment rates. This indicator has some serious statistical shortcomings, and alternatives need to be considered (see Section 3.7).

Indicator 9, life expectancy at birth by gender, is relatively complex. Furthermore, except perhaps in the largest countries, it is not considered among high priority indicators. Statistical work carried out in connection with this paper 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 for small regions, normally from administrative sources directly.

Indicator 10, self defined health status by income level, adopted only tentatively at Laeken, is calculated as the ratio of the proportions in the bottom and the top income quintile groups of the population aged 16 and over who classify themselves as being in a bad or very bad state of health. However, the indicator is still not established with an agreed methodology.

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.

3. Adaptation to the regional level

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:

- o Average measures, i.e. ordinary measures such as totals, means, rates and proportions constructed by aggregating or averaging individual values. (Examples: area unemployment rate; population proportion having a certain characteristic).
- o Distributional measures, such as measures of variation or dispersion among households and persons in the region; such measures depend on the distribution of the whole population.

The patterns of variation and relationship for the two types of measures can differ from each other, and hence require separate statistical models. Average measures are often more easily constructed or are available from alternative sources. Distributional measures tend to be more complex and are less readily available from sources other than complex surveys; at the same time, such measures are more pertinent to 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: the dispersion of regional means is of direct relevance in the identification of geographical disparity.

3.1. Choice of units to serve as 'regions'

The first issue in developing regional indicators concerns the choice of the type of units to serve as 'regions'. For a number of substantive and practical reasons, we consider geographical-administrative regions, specifically NUTS regions (and LAUs) at various level of classification, as the most appropriate choice for EU countries. The reasons for this choice include the following. NUTS regions are the most commonly used units for the formulation and implementation of social policy: the units are well-defined and identifiable, and are already widely accepted and used by different users and producers of statistical information. Despite the fact that NUTS units are not defined in exactly the same way in different countries and can differ greatly in size and homogeneity, 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. The classification covers each country exhaustively, providing a hierarchical set of units for which data can be linked across different levels. A lot of information already exists for this type of units from many different sources. Above all, data availability for the purpose of constructing the required indicators is the major reason for the choice of NUTS regions for the purpose.

This by no means precludes the above being supplemented by other dimensions. For instance, it is 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). Alternatively, 'regions' may refer to disaggregation according to population subgroups, i.e., groups identified by characteristics of individual households and persons: children, elderly persons, national minorities, immigrants, etc. Indeed, the analysis can accommodate different types of units simultaneously. 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.

The average population size for EU countries is somewhat over 5 million for NUTS1, of the order of 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 across countries declines as we go down the hierarchy.

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 paper 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 and beyond, even if the results remain tentative.

3.2. Cross-sectional measures of income poverty

Basic indicators

Henceforth the Laeken indicators have been applied primarily at the national level. It is necessary to adapt them for regional application, taking into account differences in the requirements and the data situations. As a general rule, it is necessary to focus on the more basic among the indicators. This is because the data requirements are substantially increased when the results are 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, especially when the information comes from sample surveys of limited size (say less than 2,000 sample households per region). Broad classification, such as distinguishing children, youth and elderly persons, may be possible, but even that has to be subsidiary to the need for adequate regional breakdown.

For the purpose of regional indicators, the focus has to be primarily on ordinary poverty rates for the total population, possibly with some major breakdowns. 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. Examples are Gini coefficient, relative median at-risk-of-poverty gap, and at-risk-of-poverty rate before social transfers.

On the other hand, poverty rates have 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 may also be useful because they can be decomposed into within and between region components.

Poverty line thresholds

Returning to indicators of the poverty rate, different values are obtained depending on the threshold and level of the chosen poverty line. By 'poverty line threshold' we mean the percentage of the median income defining the poverty line. The country-level Laeken indicators include measures of dispersion of the poverty rate around the thresholds taken as 40, 50, 60 and 70% of the national median income. In view of the reduced sample sizes in moving to the regional level, it is generally desirable to avoid producing too many individual figures each subject to large sampling variability. Instead, we propose to consolidate the poverty rates corresponding to different thresholds by taking an appropriately weighted average of them (see Section 4.1), so as 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 empirical income distribution based on small regional samples.

Poverty line levels

By the "level of poverty line" we mean the population level to 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 (for instance as 60% of the national median income). This applies even when the indicators are aggregated over countries, or are disaggregated to regions within a country. 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 by region and obtain regional poverty rates defined according to the national poverty line in each country.

It is also useful to consider poverty lines at other levels. For instance, we may pool the data across countries to construct a single income distribution (and hence a single poverty line) for the whole EU, and use this to compute poverty rates at the EU level, or for individual countries, or for any level of regions within any country. Specifically relevant for constructing regional indicators is the use of regional poverty lines, i.e. a poverty line defined for each region based only on the income distribution within that region. The numbers of poor persons identified with these lines can then be used to estimate regional poverty rates. They can also be aggregate upwards to give alternative national poverty rates, or disaggregated downwards to produce sub-regional poverty rates – but in all cases based on the regional poverty lines.

Different levels for the poverty line can 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 a certain regional level (such as NUTS2), the use of the regional 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 poverty lines defined at a higher level (such as NUTS1 in this case) 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:

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, while country poverty rates can differ greatly when a EU poverty line is used, the country rates tend to differ much less whether we use a poverty line defined at the national, NUTS1 or NUTS2 level.

3.3. Indicators 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. The data required for the construction of non-monetary indicators are generally simpler to collect than detailed data on monetary incomes. This makes such indicators more convenient and suitable for regional analysis. One of the important motivations of the work reported in this paper was to incorporate, with increased emphasis, non-monetary dimensions of deprivation to complement regional indicators of income poverty.

An index of non-monetary deprivation which summarises a range of indicators of living conditions should be developed and analysed in its own right. Using the methodology described in Betti and Verma (2002), we have constructed measures of overall non-monetary deprivation (variable FS_C in Table 8, Section 5.3), and similar measures of deprivation in different dimensions of living conditions (variables FSUP_1-5 in the table).

It is also useful to combine monetary and non-monetary measures in order to study the extent to which they overlap. 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. On the same lines as the monetary poverty rate, we can construct non-monetary deprivation rates, and also rates of what we have termed manifest deprivation (representing the presence of both income poverty and non-monetary deprivation simultaneously), and latent deprivation (representing the individual being subject to at least one of the two, income poverty and/or non-monetary deprivation). These are variables MAN_C and LAT_C in Table 8.

3.4. Fuzzy measures

We have also constructed an alternative version of monetary poverty using the fuzzy set approach. In this approach, poverty is seen as a matter of degree determined by the individual's location in the income distribution. For a description of the methodology, see Giorgi and Verma (2002). In Table 8 of Section 5.3, the fuzzy poverty rates are identified as FM_C, as distinct from the conventional poverty rates HCR_C. Numerically, the two rates have been defined to be closely linked. In fact, variables FM_C and FS_C referred to earlier were both constructed using very similar procedures, and their combinations LAT_C and MAN_C have been constructed using standard fuzzy set operations.

3.5. Longitudinal indicators

Longitudinal indicators are less frequently used in social inclusion and other reports than cross-sectional indicators of poverty and exclusion. These indicators are more demanding on the data. In constructing regional indicators, the emphasis has to be shifted away from the study of trends over time and longitudinal measures to essentially cross-sectional measures. 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. Simpler indicators will be more robust and less demanding on the data available.

Indicators over pairs of adjacent years

As to longitudinal indicators, it is preferable to focus on indicators defined over a short time periods. Furthermore, such measures should be aggregated over suitable time periods, so as to illuminate the more stable aspects of the patterns of variation across regions. Simpler indicators will be more robust and less demanding on the data available. Where the available statistical data cover longer time periods, short-duration 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:

- o 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.
- o 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 generally appropriate. In our illustrations, we have taken T as up to 8 years where longitudinal data for the purpose were available. 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 (see Section 2.3) 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).

Non-monetary and combined longitudinal indicators

The longitudinal measures of income poverty can be generalised to multi-dimensional measures of deprivation of the type noted above: any-time, persistent and continuous deprivation, in monetary and non-monetary dimensions, and also in the two dimensions in combination (the above defined latent and manifest forms). The basic cross-sectional rates of monetary and non-monetary deprivation can be combined with each other and then also over time using fuzzy set operations referred to in Section 3.4.

3.6. Other Laeken indicators

Labour force and education related indicators

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

Infant mortality rate

Laeken indicator 9, life expectancy at birth, is likely to be difficult to construct at the regional level, as it would involve the need for regional life-tables. Statistical work carried out in connection with this paper 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 low levels of aggregation, normally from administrative sources directly. This appears to be a more suitable indicator for regional comparisons than life expectancy at birth.

Self-defined health status by income quintile (Laeken indicator 10)

In view of the remarks made above concerning the tentative nature of this indicator even at the country level, time is not yet ripe for such an indicator to be 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' should be considered (Section 2.1).

Just as in the case of income variables, simply comparisons of averages across regions, whether within a country or across EU, can in themselves be regarded as "deprivation indicators" in the sense they indicate regional disparities.

3.7. Indicator of regional cohesion or disparities

Laeken indicator 5, 'regional cohesion (dispersion of regional employment rates)' was proposed in an attempt to measure regional disparities in employment rates. However, alternatives are required to this indicator because of its statistical and substantive shortcomings. 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). While clearly the proposed indicator needs to be improved from a statistical point of view, we do need similar indicators to synthesis the wealth of information contained in the regional breakdown of the common indicators of social inclusion. This applies not only to employment rates, but also to regional disparity in the rates of unemployment, poverty and deprivation etc. The following suggestion are made towards improving the regional disparity or cohesion indicators.

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. 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 a later section.

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, GDP/capita, or indeed to summarise regional dispersion in any of the Laeken indicators which can be reasonably produced at the regional level.

Basically the procedure involves the following steps: (i) the required indicator is computed for each region (such as NUTS2) of the country; (ii) the regions in the country are ranked according to the value of the indicator, say from the smallest to the largest; (iii) population of the regions is cumulated; (iv) 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 (e.g. 20% of the cumulative population) is not automatically obtained in practice when we are dealing with discrete units. Suitable interpolation procedures can be devised to overcome this problem (Verma et al, 2005).

Improving the measure of coefficient of variation

Comparability of the coefficient of variation 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. Below we give a more refined (comparable) measure, replacing cv, which approximately takes into account these sources of variation.[3]

[3] The underlying theory may be found in sampling texts, such as Hansen, Hurwitz and Madow (1953).

$$\rho_i = \left(\frac{B_i}{B}\right)^{+0.4} \cdot \left(cv_i^2 - \left(\frac{se_i}{p_i}\right)^2\right) \cdot \left(\frac{p_i}{(1-p_i)}\right),$$

where pi is a particular measure such as employment rate in country i, se_i is the average sampling error of regional estimates of this measure, and B_i the average region size. B is some arbitrary normalising constant, chosen for instance to make the average values over countries of ρ_i and cv_i^2 to be equal so as to bring out the comparison between the two measures. Simulations indicate that the proposed measure ρ_i tends to be rather stable across countries with regions of quite different sizes. This property is lacking in the coefficient of variation used in the original Laeken indicator of regional cohesion.

3.8. Area-level indicators of deprivation

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. The first issue, commented on in the preceding subsections, is the extent to which new types of indicators have to be added to the 'country-level list' when the focus moves to the regional level.

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. Are there useful indicators which are 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. Indeed, there are many examples of action programmes targeted specifically and in the first instance at areas (regions), rather than on persons and households identified solely on the basis of their personal characteristics and circumstances (see for instance, Glennester et al, 1999).

4. Methodology

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;

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

o 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.

4.1. Survey data

To explain the statistical procedures in concrete terms, it is useful to consider the actual data sets we have utilised. Tables 1 and 2 show the available data. The main results are based on European Community Household Panel (ECHP). Some important characteristics of the data sets available may be noted. For EU-15 countries, we have 8 annual waves of comparable, longitudinal microdata (income reference years 1993-2000), except for some initial waves missing in three countries. Single rounds of similar, but less comparable, microdata are available for two other countries (Poland and Romania).[4] For a number of other (mostly new EU) countries, we have no microdata, but only published aggregate indicators on poverty – see Table 2.

Table 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] Albania is a special case: here we have microdata from the population census and a (rather different) sample survey; this case does not form a part of our main analysis presented in this paper.

The microdata contain information on equivalised household disposable income and sample weights, so that diverse measures of poverty and inequality can be estimated. The same applies to non-monetary indicators of deprivation, though here the data are less complete. One critical shortcoming of the data is that regional identifiers are not available at all, or are available only for NUTS1 level in many cases (Table 1, last column).

Table 2 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

Averaging over survey waves

Where the information comes from sample surveys of limited size, a trade-off is required between temporal detail and geographical breakdown. Generally, the different ECHP 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 data from the different waves are highly correlated. Therefore the effective sample size for estimates averaged over waves is increased by much less than proportionally to the number of waves included. Nevertheless, there is a significant increase in the effective sample size due to real variation over 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. Also, averaging of the results over waves helps to smooth out short-term trends and bring out more clearly the underlying structural relationships.[5]

Weighted average over poverty line thresholds

In the standard analysis, as for instance 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 includes a measure of dispersion around the at-risk-of-poverty threshold (computing the percentage of persons, over the total population, with an equivalised disposable income below, respectively, 40%, 50%, 60% and 70% of the national median equivalised disposable income). 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. Irregularities are larger when the distributions are estimated from smaller samples, as normally is the case for disaggregated estimates by region. It is this consideration which is likely to dominate in the context of constructing regional measures.

Some gain in sampling precision can be obtained by computing poverty rates using different thresholds, and then taking their weighted average using some appropriate pre-specified (i.e., constant or external) weights. This is the strategy we have used in the construction of regional indicators, in preference to constructing separate indicators for different thresholds.

4.2. NewCronos

The NewCronos (now termed "Eurostat Free Dissemination Database") provides a valuable data resource for the construction of regional indicators. In itself it is not a

---

[5] 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 below. 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 a simple adjustment to reduce the effect of any such distortion (Verma et al, 2005).

source of original data, but a compilation of information from a diversity of sources presented in the form of very detailed tabulations. NewCronos REGIO domain covers the principal aspects of the economic and social life of the European Union: demography, economic accounts, labour force, health, education, etc., by region. The concepts and definitions used are as close as possible to those used by Eurostat for the production or compilation of statistics at national level. The standard model for compiling regional aggregates at various levels has been as follows: first, data from various national sources are compiled in the National Statistical Offices, and then provided to Eurostat for validation. This data set is then loaded into NewCronos by the thematic unit in question (in our case the REGIO domain of NewCronos).

Table 3 Some 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.

In Table 3 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 the (then) Candidate Countries. Mostly the breakdown is to NUTS2 level. Data available to NUTS3 are highlighted in Table 4.

Table 4 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 we have utilised variables derived from NewCronos for the construction of regional indicators.

Direct deprivation indicators

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.

Predictors

A large number of measures correlated with 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 described in Section 4.3 and 5.

Intermediate output indicators

In addition, NewCronos provides a very large number of measures, giving what has been termed as "intermediate output" indicators. Such indicators express on the one hand the policy effort in favour of those at risk of poverty and social exclusion, and on the other hand the impact of social policies as well as of the economic context. NewCronos is a unique source of such indicators.

We believe that this resource, NewCronos, 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.

- o 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.

- o 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.
- o NewCronos is a dynamic resource, in principle regularly updated as new or improved data become available. Of course, its timeliness, statistical quality and comparability 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.
- o The data base is accessible and convenient to use, and most importantly, this resource is placed in the public domain as Eurostat Free Dissemination Database.

4.3. Small area estimation (SAE)

In this paper we tackily 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. Such a practical approach is much more likely to bring out the variety of situations and problems actually encountered in the course of producing regional indicators of poverty and deprivation.

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 (2004) 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. In any case, the type of data available for poverty analysis at the regional level generally preclude the use of unit (household or person) level models. In our application, direct poverty-related information at the micro (unit) level comes from intensive and small-scale surveys (ECHP). This information can be aggregated to areas such as NUTS regions where the latter contain some sample units & the area identifies are available in the microdata. On the other side, the rich body of correlates of poverty-related characteristics of the areas come from aggregated statistics (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.

In our main application we apply area level random-effect models relating small area direct estimates 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, x_i=(x_i,1, x_i,2, …, x_i,p), are related to the target parameters θ_i (totals, means, proportion, etc.) as follows:

$$\theta_i = x_i \beta + z_i v_i \qquad \text{with i = 1…m}$$

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 σv2. Moreover it is assumed that the direct estimators $\hat{\theta}_i$ are available and design unbiased, in the form

$$\hat{\theta}_i = \theta_i + e_i$$

where e_i are independent sampling errors with zero mean and known variance ψ_i. The Best Linear Unbiased Predictor (BLUP) estimator of θ_i is:

$$\tilde{\theta}_i\left(\sigma_v^2\right) = x_i \hat{\beta} + b_i^T G Z^T V^{-1}\left(\hat{\theta}_i - x_i \hat{\beta}\right)$$

where bit 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:

$$\tilde{\theta}_i\left(\sigma_v^2\right) = \gamma_i \hat{\theta}_i + \left(1 - \gamma_i\right) x_i \hat{\beta}$$

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). An important point to note is that the mean square error of the BLUP estimator depends on the variance parameter σ_v^2, which in practice 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).

Table 6 in Section 5.1 below shows the data situation for the present research in terms of availability of micro-level survey data. In view of this data situation, the options we have considered are summarised in Table 5. Two distinct approaches are involved, depending on the data situation.

(1) EBLUB models

With a few thousand households observed in a sample survey, 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 in countries where available data permitted that, that is, where to area-coded survey data are available. In these countries (identified in Table 7, Section 5.2 below), the following applies.

At NUTS1 level, the available sample sizes are generally smaller and consequently sampling errors are larger than at the country level. 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 to enhance the available sample size, as in the present case.

The gains from modelling are naturally more significant at NUTS2 level, and substantially more so at NUTS3 level. Note that 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.

Table 5 Structure of the modelling

	Data situation	Type of estimator used
(1)	Access to area-coded survey data + Access to area-level covariates +Unclustered samples[#]	Composite (area-level EBLUP)
(2)	Lack of access to area-coded survey data, or access only to country-level survey estimates + Access to area-level covariates	Synthetic (regression-prediction)

[#] By “unclustered sample” is simply meant a sample where the primary sampling units (PSUs) are confined to be within (or at least to coincide with) the areas for which estimates are required. This is generally the case in EU surveys even for NUTS3 regions.

We have been able to extend the SAE models to a limited extent in the case of Italy due to the favourable data available situation (see Section 7).

The available 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. In any case, it is 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.

Production of estimates at lower (NUTS4 and NUTS5) levels would require models of a different type. 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.

(2) 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 the regions. This applied, for instance, to a couple of cases in ECHP for NUTS1 regions, 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.

5. SAE model implementation

5.1. Data availability for regional estimation

Each of the two situations identified in Table 5 types involves estimating several models: for each dependent variable of interest, one model corresponding to a particular level of regions, such as 'Model 1' corresponding to EBLUP model for NUTS1 regions, 'Model 2' corresponding to NUTS2 regions, etc. The important point to note is that in estimating each model, information is pooled over a set of countries. This implies the assumption of similar relationships between the model variables in different countries. This is a strong assumption. Its justification results from the 'ratio approach' we have adopted in the model specification, as explained below.

Table 6 shows the data situation for the present research in terms of availability of micro-level survey data.

5.2. The ratio approach and pooling across countries

The SAE approach we have adopted may be considered somewhat simplistic in that it does not attempt to incorporate temporal or spatial autocorrelations. A major positive feature of the approach, however, 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.

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. This makes the pooling of data across different countries for the estimation of a common model more reasonable. 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 $R_{ijk} = Y_{ijk} / Y_{ij}$. This type of modelling is further improved by taking different parts of a large or exceptionally heterogeneous country as separate units, 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. We simply model the ratio of the subpopulation measure to the total population measure.

Table 6 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

Table 7 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.

5.3. Target variables and covariates

As noted, 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/ NUTS ;

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 (see Table 7).

In this and the next Sections, we describe the models and results for Models 1 and 2. Sae Model3 for Italy is described in Section 7.

Corresponding to Models 1 and 2, 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 (for one set of countries), to predict the target variables (for another set of countries) on the basis of available predictors.

Table 8 lists the 13 target variables for Models 1 and 2. The variables were described in Section 3.4, and are grouped into three sets: income poverty related measures; overall deprivation measures; and dimension-specific deprivation measures.

Table 8 Target variables for SAE models 1 and 2

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);

Concerning measures FM_C and FS_C, the availability of these variables in the countries is as follows. Mostly, the variables are available for EU-15 countries. Sufficient information is not available in the ECHP surveys in Germany, Luxembourg and Sweden to construct deprivation measures in specific dimensions (variables F SUP 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.

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.[6]

Table 9 lists the 8 independent variables (covariates) used in models 1 and 2. These are all obtained from the tabulations provided in NewCronos. In most cases, all the required covariate were available in the individual countries.

Table 9 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

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 estimation & 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.

5.4. 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.

As explained in Section 5.2, we used the ‘ratio approach’ to improve the precision of the models. Under this approach, the model input consists of

[6] An approximation is involved in the construction of the second variable, log equivalised income. It is computed from the available median value using the average mean: median ratio=1.15 from ECHP surveys in EU15, and then taking the log of the mean so estimated.

(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 10 shows some performance measures of SAE Model 1. For each model (i.e., each target variable), three measures are shown:

- o 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;
- o ratio (a)*/(a), i.e., the ratio between the EBLUP estimated value (a)* and the corresponding direct estimate (a). This is to check the extent to which the modelling changes the input direct estimates;
- o 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.

Table 10 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

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.

The regression coefficients and the associated significance levels were also estimated in order to evaluate the performance of the models. These may be found in Verma et al (2005).

Overall, the results are as expected: the SAE Model 1 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' in most ECHP survey. Further increase in the effective sample size was achieved by cumulation of data over survey waves.

5.5. 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 above. However, there are differences in the extent to which the variables are available and the nature of the countries involved in the same model. Note that because of the lack of NUTS2 identifiers in the microdata, SAE Model 2 is based only on pooled information from five countries

Table 11 shows some performance measures of SAE Model 2. For each model (target variable), three measures are shown as in Table 10.

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 LAT-C, MAN-c and FSUP-1 deprivation measures. Again, as with Model 1, 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 because of 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 11 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

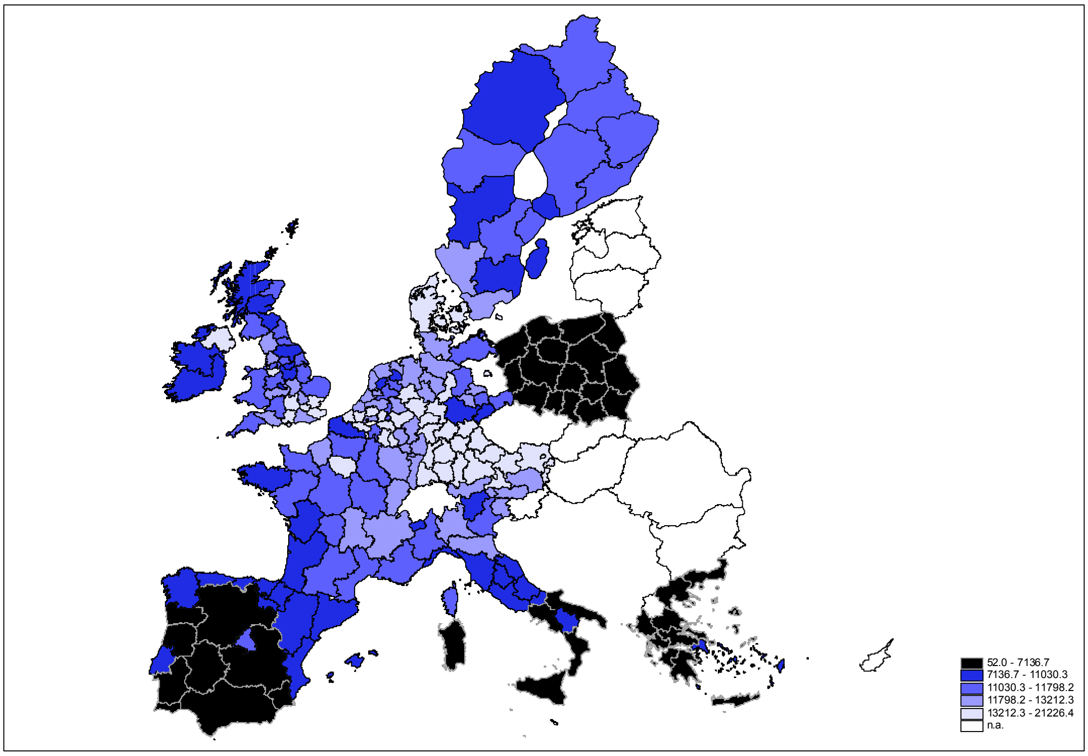
6. Illustrative results

6.1. Poverty rates with country and region-specific poverty lines

In order to understand the effect of using poverty lines defined at different levels, it is useful to begin with regional differences in mean income.

Figure 1 reports the mean net equivalent income at NUTS2 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).

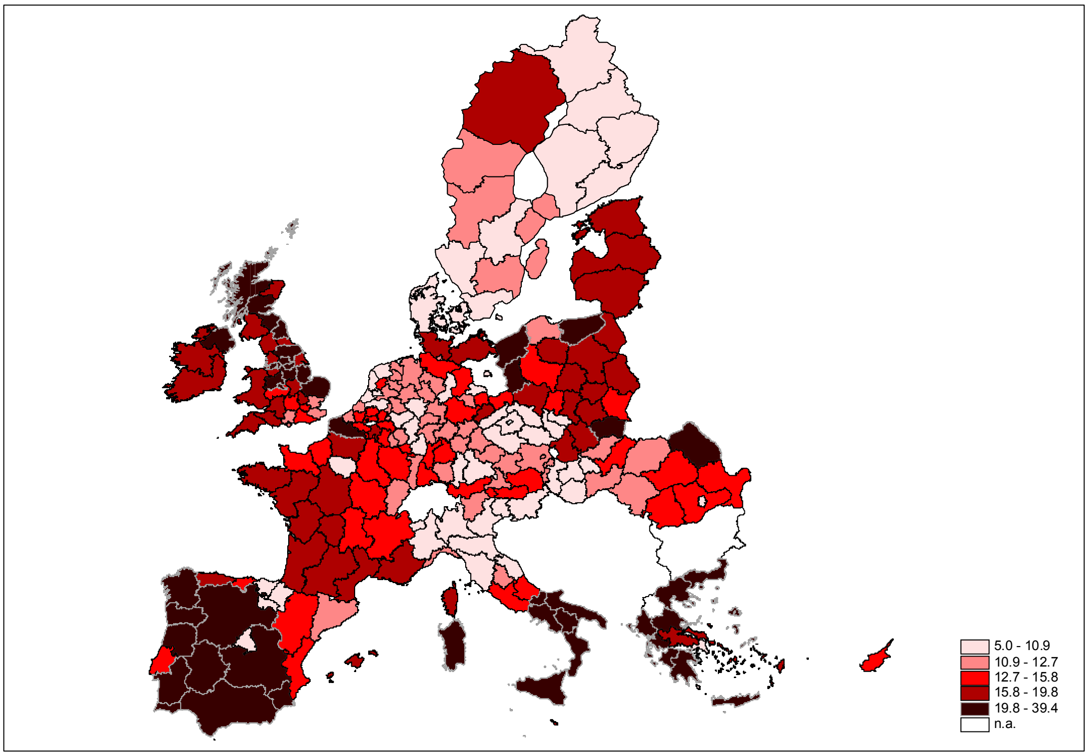
Figure 1 Net equivalent income – NUTS2



Now turning to estimates of regional poverty rates (conventional HCR), it should be noted that the results are based on an extremely large number of computations, performed separately for each survey wave in each country. These are then used to construct consolidated measures of the type we recommend for regional analysis (Section 4.1).

Figure 2 shows the concentration of the areas with the highest poverty rates (bottom quintile) 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 East . In Italy it is interesting to note the striking difference between the South and the North.

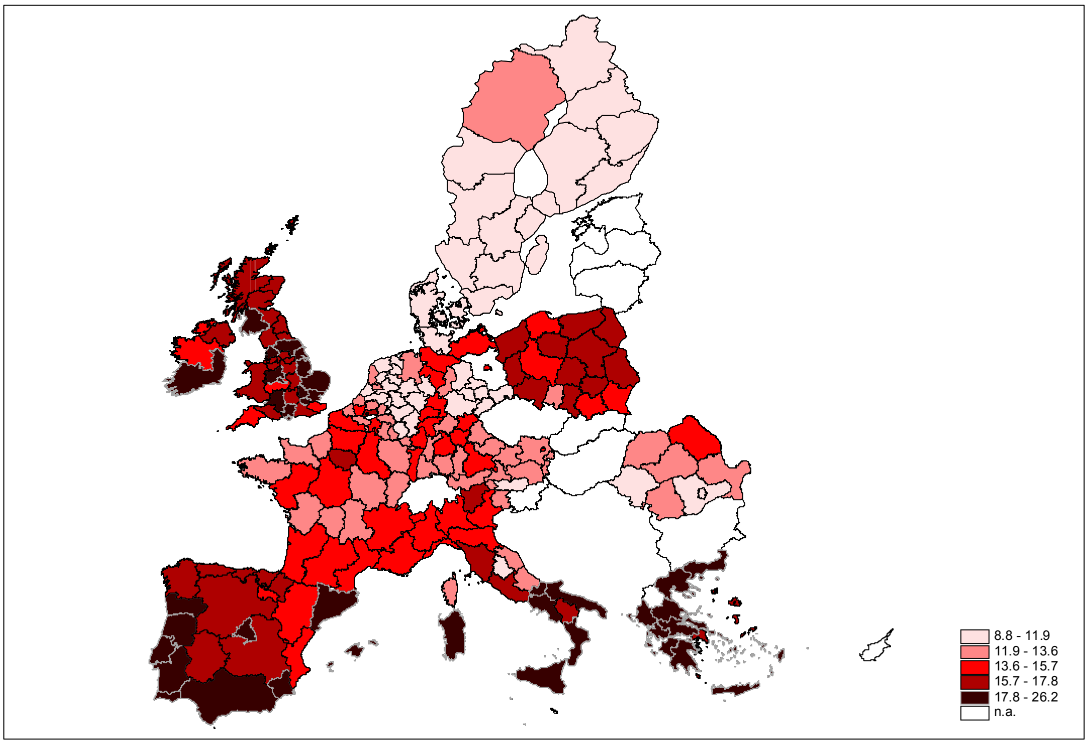
Figure 2 Head Count Ratio NUTS2 regions (country poverty lines)



In Figure 3 with NUTS2 poverty lines, we see a less homogenous situation in terms of head count ratio. 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 2 do no belong to the bottom category in Figure 3 (for example Basilicata in Italy). On the other hand, some areas that belong to the best class in Figure 2, move to the middle bracket in Figure 3, such as Toscana, Emilia Romagna and Lombardia in Italy. The same applies to some regions of Spain.

Countries where regional differences in levels of income are small tend to present similar pictures irrespective of whether the country or NUTS2 poverty lines have been used.

Figure 3 Head Count Ratio NUTS2 regions (NUTS2 poverty lines)



6.2. Non-monetary deprivation

Figure 4 Overall non-monetary deprivation rates (NUTS2 regions)

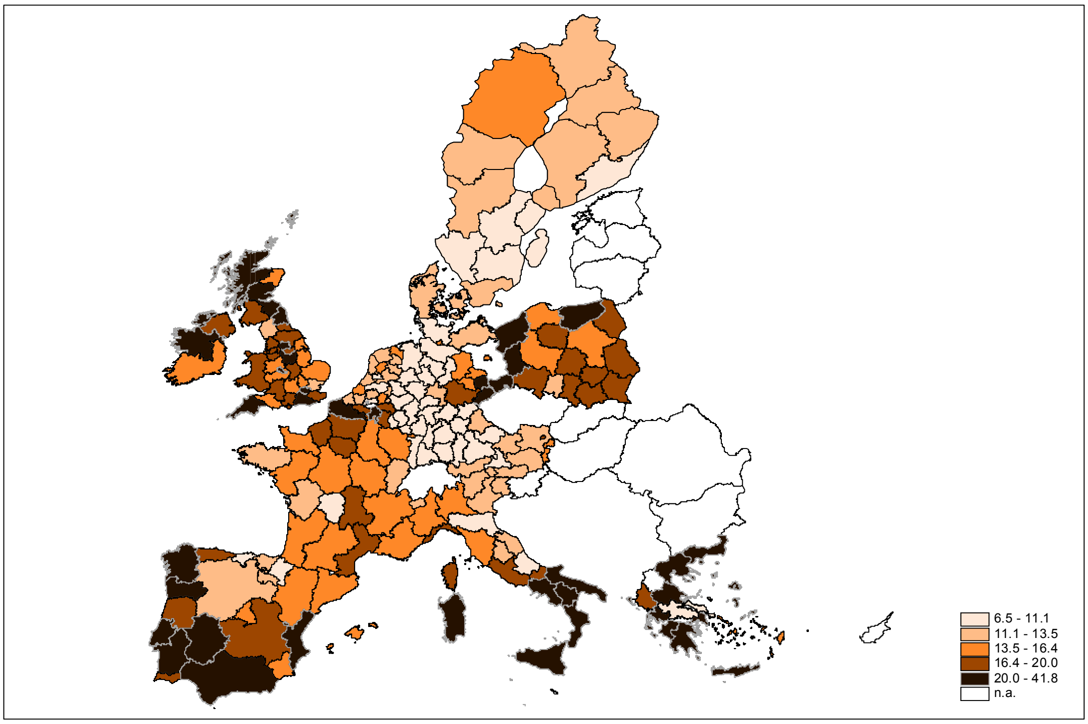


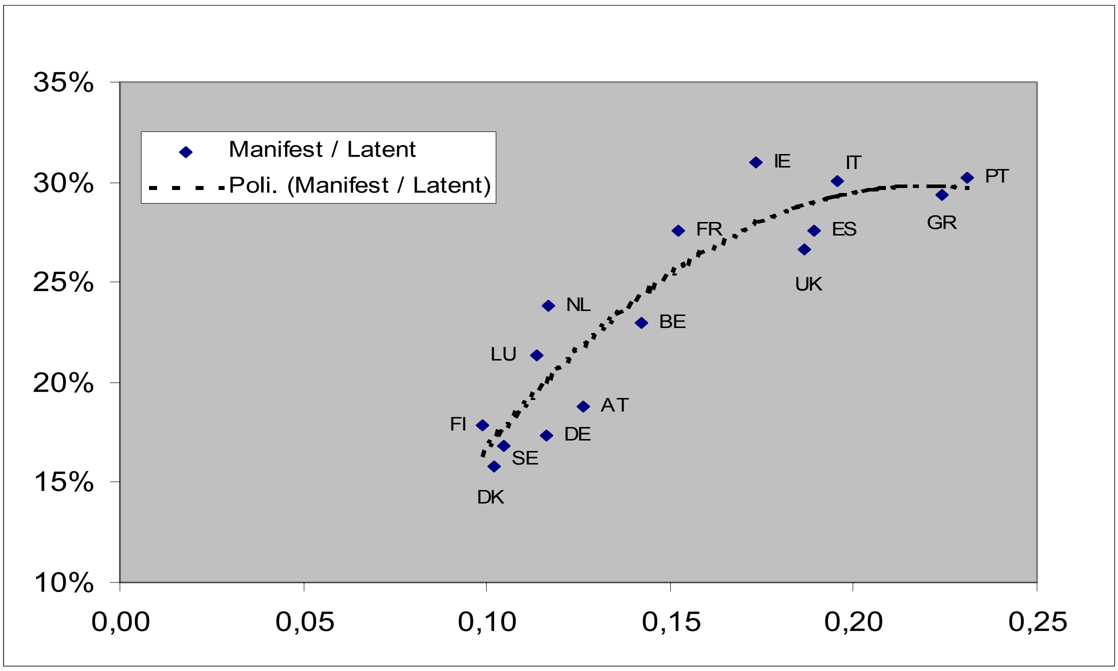
Figure 4 shows the variation of the overall non-monetary deprivation rate across NUTS2 regions. We observe very high values of deprivation in Portugal, South West Spain and South Italy. A better off country is Germany, showing a very homogenous behaviour among regions.

The objective of Figure 5 is to display the degree of overlap, at the level of individual persons, between monetary and non-monetary forms of deprivation. The figure shows the manifest deprivation index as a 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.

By definition, 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 is 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. In richer countries different dimensions have less overlap, and hence deprivation appears to be more multidimensional. In poorer countries, there is more overlap, making deprivation more intensive for the individuals involved.

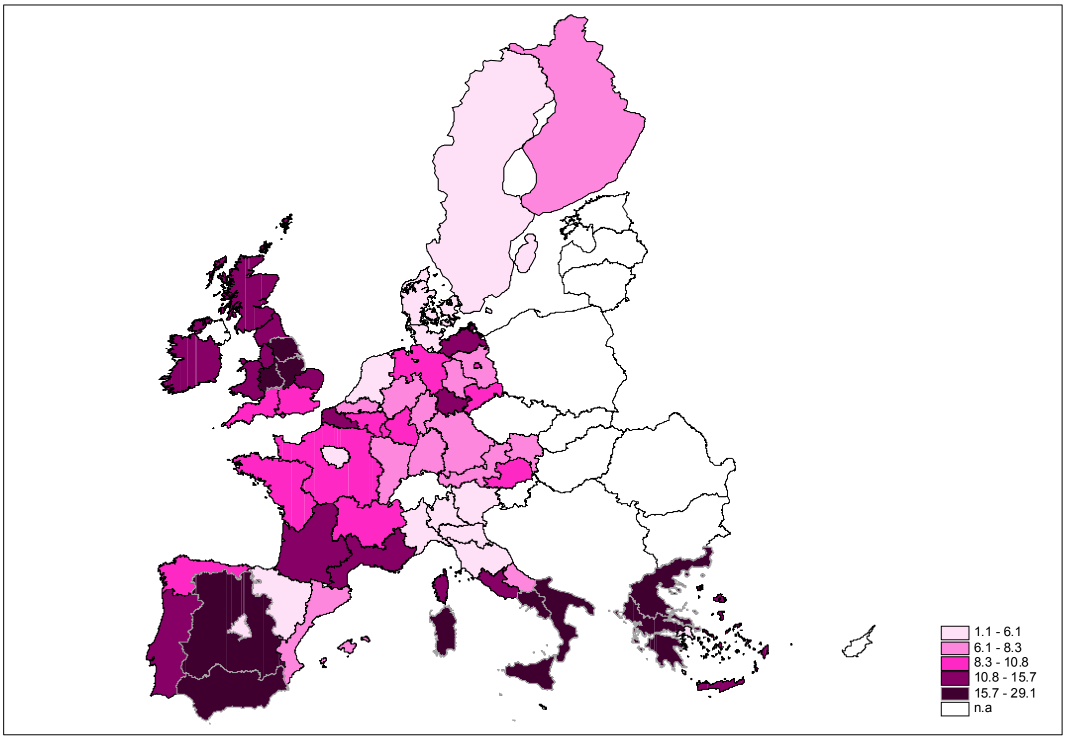
Figure 5 Manifest deprivation rate as a percentage of latent deprivation rate, against a measure of the level of poverty or deprivation in the country. EU15



6.3. Longitudinal poverty

Figure 6 shows the rates of persistent poverty measured over pairs of adjacent years over an 8-year period at NUTS1 level for the EU15 regions – the rate being the proportion who are poor at both years in the pair of years. For comparison, we also estimated the cross-sectional and any-time rates (not shown here). A remarkably high level of consistency was found in the patterns across these three types of measures. In other word, the relative position of EU15 regions is very similar whichever measure is considered.

Figure 6 Percentages in persistent poverty over two adjacent years. NUTS1 regions



7. SAE for Italian NUTS3 regions

SAE Model 3 concerns EBLUP models for going from NUTS2 to NUTS3 (province) level. As noted in Section 5, this was estimated for a limited set of variables; and only for Italy because of lack of data. The results are briefly described in this concluding section.

7.1. The model

Given the high level of disaggregation, it was 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 very limited; it is confined to the relevant covariates tables for which are provided in NewCronos at NUTS3 level. Unlike the higher level model of Section 6, no pooling over countries is involved.

Table 12 shows some performance measures of SAE Model 3. For each model (target variable), three measures are shown as in Table 10 earlier. 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 that the minimum value of the ratio between the EBLUP standard error and the direct standard error: 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 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

7.2. Regional disparity and the effect of the chosen poverty line level

Figures 7 shows the estimated poverty rates for NUTS3 regions (provinces) of Italy, computed on the basis of the country poverty line; Figure 8 shows the same rates using NUTS2 poverty lines. The difference between the two arises from the large regional disparities in income levels in Italy. By using regional poverty lines, we remove the effect of regional differences in median income level.

Also removed is the affect of any cost-of-living differences among regions, since the income distribution for each region is considered separately. The last mentioned differences of course distort the regional comparison when a common country poverty line is used.

Even so, there is a statistical association between the level of income and the poverty rate, even when the latter is a relative measure within each region [7]. The regions with the highest inequality are in the South, and the ones with the lowest inequality are in the Centre. This phenomenon has been often observed at country level in EU, and here we observe the same at the regional level in Italy.

[7] Even though NUTS3 is the level of disaggregation of the results, we have still used poverty lines at NUTS2 level because of sample size considerations. Hence differences between average income levels and purchasing power parities among NUTS 3 regions in the same NUTS 2 still affect the purely relative nature of the poverty measures displayed in Fig.8.

Figure 7 Head Count Ratio, NUTS3 regions (country poverty line). Italy

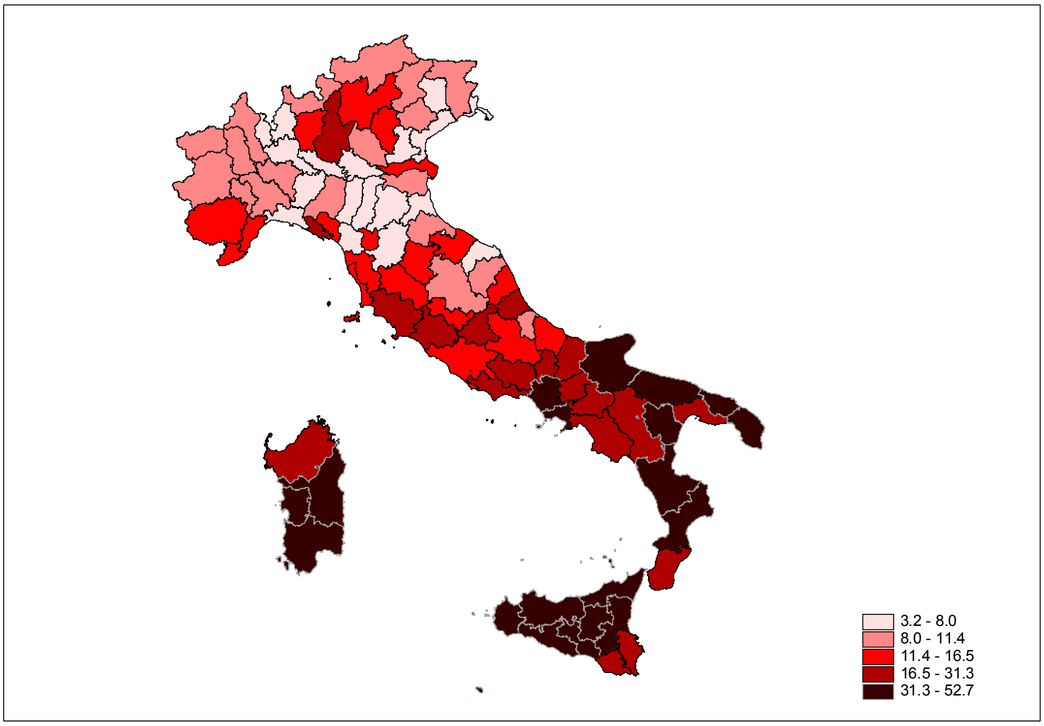
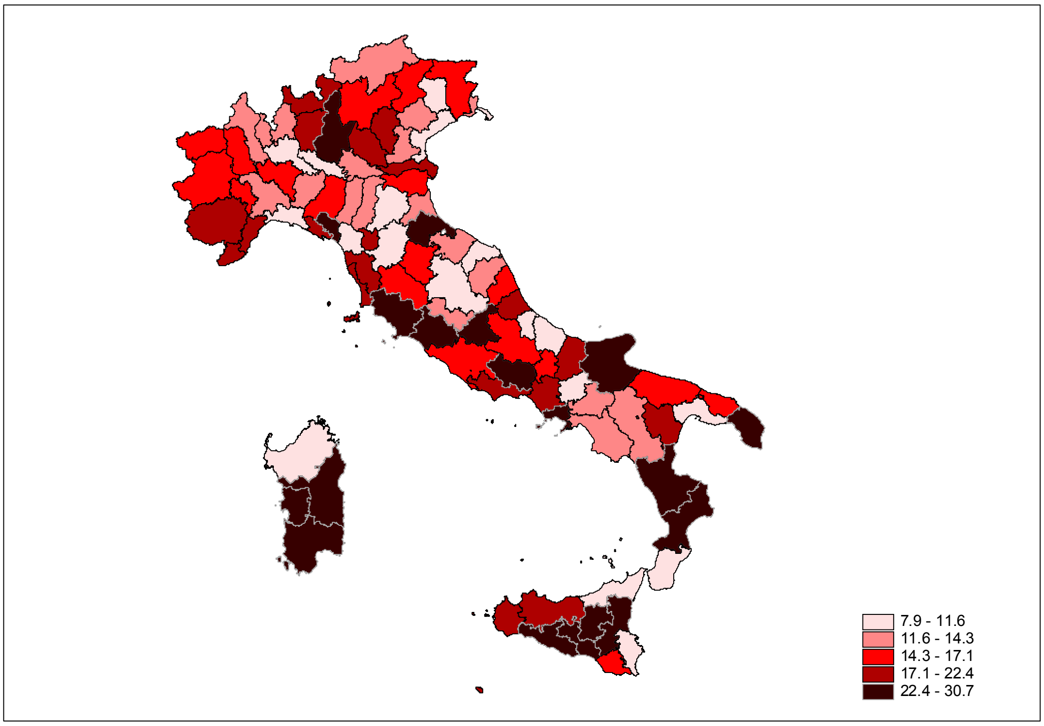


Figure 8 Head Count Ratio, NUTS3 regions (NUTS2 poverty lines). Italy



References

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

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**, pp. 28-36.

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.

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 (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 (2003), Monetary Poverty in EU Acceding and Candidate Countries, *Statistics in Focus*, Theme 3- 21/2003.

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.

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.

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

Gosh M., Rao J.N.K (1994), Small Area Estimation: An Appraisal (with discussion), *Statistical Science*, **9 (1)**, pp. 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.

Rao J.N.K (2003), *Small Area Estimation*. Wiley, London.

Verma V., Betti G., Lemmi A., Mulas A., Natilli M., Neri L., Salvati N. (2005). Regional indicators to reflect social exclusion and poverty. Final report. Project VT/2003/45, European Commission, Employment and Social Affairs D.G.