

MEASURING SUB-NATIONAL INCOME POVERTY BY USING A SMALL AREA MULTIVARIATE APPROACH

BY ENRICO FABRIZI

DISES, Catholic University of the Sacred Heart, Piacenza, Italy

MARIA ROSARIA FERRANTE*

University of Bologna, Italy

AND

SILVIA PACEI

University of Bologna, Italy

The present paper proposes a statistical strategy for the analysis of regional disparities in income poverty. For the EU countries, information on individual income has been collected until now by the European Community Household Panel survey, which only yields reliable estimates for very large regions within countries. In order to obtain reliable estimates for some of the poverty indicators suggested by the Laeken Council at the sub-national level, we suggest the adoption of a multivariate small area estimation approach which enables us to reduce estimate variability. We concentrate on Italy, the country with the lowest degree of regional cohesion within the EU. Results show that disparity cannot be reduced to the so-called “North–South divide,” with the “poor” South separated from the “affluent” North, as both these macro-regions display large internal differences in terms of both poverty level and income inequality. The strategy we propose could also be adopted in order to measure poverty in other European regions, using information produced by the new EU Survey on Income and Living Conditions, which is replacing the European Community Household Panel.

1. INTRODUCTION

Social inclusion has been one of the declared objectives of the European Union (EU) since the Lisbon Summit of March 2000, and is increasingly recognized as a major challenge for the international community. The importance of social inclusion policy has been recently confirmed by the European Council, which has stressed the importance of fighting poverty and of raising the standards of living of all European citizens (European Commission, 2005a).

In order to monitor national progress towards the EU’s common objectives, the Laeken European Council (held in Belgium in 2001) endorsed a set of 18 indicators of the multi-dimensional phenomenon of social inclusion. Almost all of the chosen indicators are supposed to be measured at a national level, although the Council also emphasized the importance of the local dimension.

Note: We would like to thank ISTAT for kindly providing the data used in the present work. Thanks to a special research agreement between Bologna University’s Department of Statistics and ISTAT, we have been able to access the ECHP data base for Italy, together with some other sampling design information which was required in order to obtain the estimates presented herein. Research was partially funded by Miur-PRIN 2003, no. 2003133249.

*Correspondence to: Maria Rosaria Ferrante, Department of Statistics, University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy (maria.ferrante@unibo.it).

An understanding of where social cohesion is most at risk is important in shaping preventative policies, in evaluating their effects and in developing hypotheses regarding the nature of the causal mechanism. Moreover, the problem of monitoring regional poverty is particularly relevant in those countries with strong regional disparities, such as Italy. Within the EU, Italy displays the lowest degree of regional cohesion (European Commission, 2005b) and its marked differences in regional labor markets and economies, together with the concentration of industrial districts in some geographical areas, can give rise to significant differences in income distribution.

The analysis of regional disparities in poverty and income inequality requires reliable information at a detailed territorial level.

In the EU, the measurement of income and poverty has been based until now on the information collected by the European Community Households Panel (ECHP). This annual survey, coordinated by Eurostat (Eurostat, 2002), is designed to provide reliable estimates of poverty indicators only for large areas within countries, known as NUTS1 (NUTS stands for the EU's "Nomenclature of Units for Territorial Statistics"; see <http://europa.eu.int/comm/eurostat/ramon/>). Randomization based unbiased estimators of NUTS2 level population parameters, based on ECHP data, are in most cases characterized by unacceptably large standard errors, and are therefore unreliable, which hinders any regional breakdown of most of the standard EU social inclusion indicators.

Moreover, the ensemble of NUTS2 estimates provides a misleading picture of the distribution of the underlying regional parameters across the country. It can easily be shown that they are overdispersed (Louis, 1984; Heady and Ralphs, 2004); that is, the descriptive variance of these so-called "direct" estimates is larger (far larger if their reliability is low) than the real variance of the actual population of regional parameters.

The aim of this paper is twofold: to propose a strategy designed to produce reliable estimates of poverty related parameters for Italy's NUTS2 regions; and to analyze regional disparities on the basis of the estimates obtained.

In order to provide a multifaceted picture of disparity between Italian regions, in terms of social cohesion, we concentrate on a selection of Laeken indicators. We use data from the last wave of the ECHP survey (conducted in 2001). The common opinion is that Italy's lack of cohesion may be expressed in terms of the North–South divide. The availability of reliable estimates at the NUTS2 level may tell us whether this is the case, or whether there are substantial differences within these macro areas. We are going to focus in particular on the South, as this is the traditional target of developmental and poverty reduction policies.

To obtain reliable estimates of the selected poverty parameters at the NUTS2 level, we propose a multivariate small area estimation (SAE) strategy, elements of which have already been dealt with in the literature, but have never been applied together in the empirical analysis of real data.

The SAE strategy we follow in this paper is model based, that is, we link direct regional estimates of the quantities in question to concomitant auxiliary information about the regions, by means of an explicit model. In particular, we consider a multivariate normal linear mixed model. The application of a similar model to

income distribution related parameters (median income for households of various sizes) may be found in Ghosh *et al.* (1996).

The approach to estimation is Bayesian. We treat the unknown estimands as random variables, and estimate them by means of suitable summaries of their posterior distributions (in practice, their posterior means). These are called “Bayes estimators” (Ghosh, 1992). Although regarded with some suspicion in frequentist-oriented literature, if we focus on small area studies we will see that the theory of Bayes estimators runs parallel to that of frequentist Empirical Best predictors (EBP). Moreover, given the same model, EBP and Bayes estimators almost agree in the case of reasonable choices of prior distributions, and Bayesian measures of uncertainty provide good estimators of repeated sampling Mean Square Error (MSE) of Bayes estimators (see Rao, 2003, ch. 10 for a general introduction to this topic).

Regardless of the approach to estimation, small area estimation based on mixed models yields estimators with reduced MSE at the price of shrinking direct estimates toward the model prediction (this is known as the “synthetic component”). This shrinkage may lead to the misrepresentation of the actual distribution of the ensemble of regional parameters, and in general to the underestimation of the “ensemble” variance.

To overcome this problem, which could jeopardize comparisons, we consider “constrained Bayes” estimators in which the first two moments of the distribution of the small area estimates are adjusted to meet the first two of the underlying population (ensemble) of parameters. Such adjustments are quite well known in the case of univariate models (Rao, 2003, ch. 10.13). An extension of this methodology to multivariate models is briefly discussed, within a frequentist framework, in Ghosh and Maiti (1999); however, to the best of our knowledge, it has never been used in applied estimation problems.

The same data were analyzed in Fabrizi *et al.* (2005), which focused more on the statistical details of estimation such as the estimation of design based variances the comparison between univariate and multivariate models, and the adoption of constrained predictors. In the present paper, we start from the results obtained in Fabrizi *et al.* (2005) and we discuss the statistical problems that usually arise when analyzing regional disparities; we deal with the use of small area estimates within this context, focusing on the reasons for the adopted definitions and estimation strategies, and on the implications thereof. Moreover, we propose an economic analysis of regional disparities, based on the estimation results.

Finally, it should be noted that the ECHP is gradually being replaced by a new survey, the “European Union—Statistics on Income and Living Condition” (EU-SILC). Unfortunately, the EU-SILC data has not been published yet, or at least not with regard to Italy. Nevertheless, due to the similarity between the two surveys, the proposals defined in this paper could also be extended to measure poverty in all European NUTS2 or NUTS3 regions, based on information provided by the new EU-SILC.

The outline of the present paper is as follows. Section 2 provides some background information on estimates of regional poverty indicators. Section 3 contains a description of the features of those Laeken indicators chosen to describe regional poverty and the ECHP data set. Section 4 focuses on the small area

nature of our estimation problem. Section 5 provides a description of the adopted multivariate model and of the Hierarchical Bayes method used in its estimation. Section 6 presents the results of the regional estimates, while Section 7 compares the estimates in order to provide a picture of the regional disparity in poverty within Italy. Finally, Section 8 offers some concluding remarks and some ideas for the further development of the present work.

2. INCOME DISPARITIES AND POVERTY ESTIMATION IN ITALY AND THE EU: THE STATE OF THE ART

The European Commission (2005a) and, in the case of Italy, the Italian Statistical Institute (ISTAT) and the Bank of Italy, have made substantial efforts to fill the gap created by the lack of reliable information on regional poverty. Moreover, methods for small area estimation of per-capita income have been investigated in the EURAREA research project, funded by Eurostat under the EU's Fifth Framework Programme (see <http://www.statistics.gov.uk/eurarea/default.asp>).

ISTAT bases its estimates of poverty on the Consumer Expenditure Survey (Coccia and Masi, 2003). However, it should be said that poverty evaluated using income information is not affected by the choices that households make between saving and consuming. Moreover, using household consumption to build poverty indicators represents a departure from the practice of other member states, whereby income poverty indicators are calculated from income figures. Given that poverty indicators based on consumption generally indicate a lower poverty level, this limits the comparability of information between European regions situated within different countries. Another source of income information available in Italy is the biennial Survey of Income and Wealth carried out by the Bank of Italy. This survey produces reliable estimates at the macro-area level, the macro-areas in question being obtained by grouping together diverse NUTS2 regions. An attempt to compare the distribution between NUTS2 regions has been based on pooling three successive repetitions of the said survey (Cannari and D'Alessio, 2003), a strategy similar to that adopted in a recent report carried out by the European Commission (2005a). This approach enables estimates for more than one year to be obtained, and as such cannot be considered a suitable way of producing annual estimates or of analyzing the temporal dynamics of poverty.

The strategy we propose, focusing as it does on income, in theory makes it possible to compare the NUTS2 regions of all EU member states. In considering various income indicators, it gives a more complete picture of the multidimensional phenomenon of poverty. A multivariate small area model that utilizes the correlation between different indicators in order to improve estimate reliability, avoids the need to consolidate information from different waves.

3. DATA AND INDICATORS

The ECHP is a harmonized, cross-national, longitudinal EU survey which was carried out yearly between 1994 and 2001. It focuses on household income and living conditions. The ECHP's target population consists of all resident

households of EU member countries. As far as Italy is concerned, the first wave's design is a stratified two stage design, in which strata were formed by grouping the PSUs (municipalities) according to geographic administrative region (NUTS2) and demographic size (for more details see Eurostat, 2002). These various features make this dataset a valuable source of statistical information for the social inclusion process in the EU. Besides all the improvements and additions it is going to produce, the EU-SILC survey, which is currently replacing the ECHP, will continue to possess all of the aforesaid features.

For the purposes of this paper, we consider a selection of the indicators suggested in the Laeken Council report, which help provide an informative picture of income inequality and poverty at the level of Italy's administrative regions (NUTS2).

All of the indicators are obtained on the basis of the "personal equivalent total net income," which shall be referred to hereinafter simply as "income" or y .

As far as the definition of income is concerned, the ECHP questionnaire enables us to obtain a good approximation of the concept of net income, that is, income net of income taxes and social insurance deductions at source. This definition may not coincide exactly with the notion of disposable income, which is also net of final settlement payments or refunds on income taxes. Such information is not requested in the ECHP interview.

Total net household income is calculated as the sum of the net incomes of all members of the household. Personal equivalent net income is calculated by dividing total net household income by equivalent household size according to the OECD scale used by Eurostat (which gives a weight of 1.0 to the first adult, 0.5 to the other persons aged 14 or over who are living in the household, and 0.3 to children under the age of 14). Consequently, the same equivalent total net income is assigned to each person in the same household.

The considered indicators are:

- (1) *Per-capita equivalent net income (PCI)*.
- (2) *Regional at-risk-of-poverty threshold (RPT)*. This is defined as 60 percent of the regional median of personal equivalent total net income.
- (3) *At-risk-of-poverty rate, regional threshold (ARPR_r)*. The share of persons with an equivalent total net income of less than 60 percent of the RPT.
- (4) *At-risk-of-poverty rate, national threshold (ARPR_n)*. The share of persons with an equivalent total net income of less than 60 percent of the national median income (that is, the National Poverty Threshold, NPT).
- (5) *Gini coefficient (G)*.

Some of the indicators are useful to describe the distribution of y , while ARPR_n is the most widely used poverty indicator in OECD countries. Although most poverty analyses are based on the definition of a single national threshold, there may also be grounds for a discussion of poverty rates based on regional thresholds. First, in assuming the existence of a national poverty line, it is implicit that all individuals within the country in question face paying the same prices (see Mogstad *et al.*, 2007 for a discussion of this topic). Although ISTAT does not calculate region-specific price indexes for Italy, common sense tells us that price levels differ widely between NUTS2 regions, especially between those of the affluent North and those of the less developed South. Generally speaking,

if we assume the “derived relativist” approach advocated by Sen (1979, 1983), a common national poverty line implies that the same amount of income is needed to reach the same level of “capabilities” throughout the country. Again, this may not be true for Italy where there are substantial regional differences in social structures, lifestyles, social norms and forms of consumption. Moreover, the use of regional poverty lines may also be justified by the subjective poverty argument. If we restrict to those based on income, subjective poverty lines (Goedhart *et al.*, 1977) are based on a respondent’s perception of the minimum income needed to make ends meet. Even if income and household composition are taken into account, these subjective lines usually display substantial variations from one group of individuals to the next. Due to the limited data available, we cannot calculate subjective poverty lines from ECHP figures. Furthermore, if we take the provisional EU-SILC figures for 2004 (at the time of writing, the official figures have yet to be published), there is some evidence that subjective lines vary considerably across regions, thereby substantiating the idea of considering regional thresholds. The EU-SILC questionnaire contains a question about the “lowest monthly income to make ends meet.” We analyzed the resulting data and found that once adjusted for income and for different household compositions (using the method described in Kapteyn *et al.*, 1988), the average minimum income “to make ends meet” is roughly 50 percent higher in the affluent central and northern regions than it is in the poorer southern regions. In view of this fact, the adoption of a regional threshold (i.e. based on regional median income) may also be motivated as an approximation of subjective thresholds calculated on a regional basis.

4. THE SMALL AREA NATURE OF OUR ESTIMATION PROBLEM

The number of individuals from each region interviewed in the 2001 ECHP survey ranges from 201 to 1,837, with a country-wide total of 15,943.

Since the administrative (NUTS2) regions of Italy are planned domains for the ECHP survey, design nearly unbiased, consistent estimates of the parameters defined in the previous section can be obtained by applying standard weighted estimators based on the weights available in publicly released files. In small area jargon, such estimates are called “direct” since they use the values of the target variable only from the area-specific sample units. Evaluating the variance of these direct estimators is in this particular case a complicated task, as the underlying design is complex, with the weights used in their computation incorporating calibration corrections for non-response and attrition. With the exception of PCI, they are non linear, sometimes (ARPR_n, ARPR_r) complicated functions of the data; moreover, in view of the multivariate model we want to estimate, we also need to obtain estimates of covariances between the various direct estimators.

For these reasons, and in keeping with other work in this field (Verma and Betti, 2005), we have opted for a solution based on re-sampling algorithms. In particular we use a bootstrap estimation strategy.

Bootstrap variance estimators have been proposed and analyzed for sampling designs as general as the multi-stage sampling design with stratification of primary units often used in large socio-economic surveys. See Rao (1999) for further

TABLE 1
SUMMARY STATISTICS FOR DIRECT ESTIMATES

	Parameter				
	PCI*	RPT*	ARPRr**	ARPRn**	G**
$\min(\hat{\theta}_{ik})$	15.58	7.74	15.76	2.71	17.99
$\max(\hat{\theta}_{ik})$	27.95	15.51	20.78	43.72	32.92
average[se($\hat{\theta}_{ik}$)]	1.21	0.66	3.48	4.15	2.35
min[se($\hat{\theta}_{ik}$)]	0.32	0.25	1.46	0.92	1.10
max[se($\hat{\theta}_{ik}$)]	2.51	1.27	8.01	9.58	3.43

Notes: *In thousands of euros; **as a percentage.

details. These estimators rely on the assumptions that the number of strata is large and that few primary units (but at least two) are sampled from each stratum, so that the sampling fraction at the first stage is negligible. This latter assumption is not met in our case, as there is a stratum of self-representative municipalities (primary units) that are always included in the sample. This is why Ferrante *et al.* (2004) propose a bootstrap algorithm in which any bootstrap sample is the union of two sub-samples, one taken by re-sampling the population in the non self-representative strata, and the other drawn from the self-representative stratum, where the sampling design is actually single stage. This algorithm has been tested by means of simulation exercises, and found to provide estimates close to those obtained using the linearization method (which we calculated for per capita income).

Table 1 provides some basic summary statistics about direct estimates and their standard errors, calculated using the described bootstrap method ($\hat{\theta}_{ik}$ is the direct estimate of the k -th parameter ($k = 1, \dots, K = 5$) for the i -th small area ($i = 1, \dots, m = 21$)).

The reliability of estimates varies from one indicator to another, but in general standard errors are too large to allow for a meaningful analysis of the distribution of the parameters across the nation. The basic reason is that region-specific sample sizes are not large enough. Focusing on individual parameters, we may observe how the estimates of ARPRr and ARPRn are the least efficient (the CVs are, on average, equal to 0.25 and 0.28 respectively). This is due to the further source of sampling variability represented by the fact that they are proportions based on estimated thresholds. On the other hand, the estimates of PCI and RPT are the most reliable, and may be considered sufficiently efficient for various NUTS2 regions. Estimates of G show an acceptable level of variability for certain regions only.

Moreover, as anticipated in the introduction, the “ensemble” of direct estimates is over-dispersed, that is, more widely spread than the underlying ensemble of actual parameters. This situation justifies the recourse to small area estimation methods.

The estimated coefficients of correlation between pairs of poverty indicators, averaged over the areas, are substantial in nearly every case, ranging from 0.38 to 0.76. This feature suggests the use of the multivariate model we are about to introduce in the next section.

5. A MULTIVARIATE HIERARCHICAL SMALL AREA MODEL

Generally speaking, any subset of a given, finite population defined according to a chosen geographic criterion, is considered a “small area” if the survey sample in the area is too small to obtain reliable estimates for that population. For this reason, special estimators, often explicitly relying on models connecting parameters of interest with auxiliary information, should be used to produce estimates with a more acceptable mean square error. We are now going to illustrate the multivariate hierarchical model we use to obtain the estimates of the poverty parameters for Italy’s administrative regions.

Let $\theta_i = (\theta_{ik})_{1 \leq k \leq K}$ be the K -vector of the population parameters we are interested in, pertaining to area i ; let $\hat{\theta}_i$ be the vector of “direct” estimates of these population quantities. Our small area estimation strategy is based on the following model assumptions:

$$(1) \quad \hat{\theta}_i | \theta_i, \Psi_i \stackrel{ind}{\sim} N_K(\theta_i, \Psi_i)$$

$$(2) \quad \theta_i | \beta, \Sigma_v \stackrel{ind}{\sim} N_K(\mathbf{Z}_i \beta, \Sigma_v)$$

where the $K \times K$ positive definite $\Psi_i = V(\hat{\theta}_i | \theta_i)$ is assumed to be known and equal to the estimate obtained using the bootstrap method illustrated in the previous section. This assumption, although questionable, is very common in small area literature (see Rao 2003, p. 76). \mathbf{Z}_i is a $K \times Kp$ known matrix with the k -th row given by $(\mathbf{0}', \dots, \mathbf{0}', \mathbf{z}'_{ik}, \mathbf{0}', \dots, \mathbf{0}')$, and contains the available auxiliary information about area i , while β is a vector of regression coefficients. Σ_v is an assumed positive definite $K \times K$ prior variance matrix. Note that assumptions (1) and (2) may be reformulated as

$$(3) \quad \hat{\theta}_i = \mathbf{Z}_i \beta + \mathbf{v}_i + \mathbf{e}_i$$

with $\mathbf{v}_i | \Sigma_v \stackrel{ind}{\sim} N_K(0, \Sigma_v)$, $\mathbf{e}_i \stackrel{ind}{\sim} N_K(\mathbf{0}, \Psi_i)$, \mathbf{v}_i independent of \mathbf{e}_i and $\theta_i = \mathbf{Z}_i \beta + \mathbf{v}_i$, a more common representation for linear mixed models. Note that the random effects are introduced in order to represent area-specific characteristics not represented by the covariate used in the model.

Note that in small area estimation, when area-level models such (1)–(2) are adopted there is no real choice between random and fixed effects (as discussed, in the general context, by Greene 2003, pp. 301–3) as a model with a distinct fixed effect for each area would be unidentifiable.

More details about a model close to (1)–(2) may be found in Ghosh *et al.* (1996). In particular note that if β, Σ_v are known, the mixed effect specification leads to a “Bayes estimator” of θ_i , i.e. the posterior mean $E(\theta_i | D)$ with $D = \{\hat{\theta}_i, \Psi_i\}$ of the following form:

$$\hat{\theta}_i^B = \Sigma_v (\Psi_i + \Sigma_v)^{-1} \hat{\theta}_i + \Psi_i (\Psi_i + \Sigma_v)^{-1} \mathbf{Z}_i \beta.$$

This estimator has a nice weighted average structure, in which the “synthetic” component $\mathbf{Z}_i\boldsymbol{\beta}$ receives a weight that is inversely proportional to the precision of direct estimators. As a consequence, $\hat{\boldsymbol{\theta}}_i^B$ is consistent. Assume an asymptotic framework for finite populations similar to that of Isaki and Fuller (1982), and assume the consistency of $\hat{\boldsymbol{\theta}}_i$; as $n_i \rightarrow \infty$, $\boldsymbol{\Psi}_i \rightarrow \mathbf{0}_{K \times K}$, and $\hat{\boldsymbol{\theta}}_i^B \rightarrow \hat{\boldsymbol{\theta}}_i$. The Bayes estimator $\hat{\boldsymbol{\theta}}_i^B$ may also be given a frequentist justification; in fact it may be obtained as the Best Linear Unbiased Predictor of $\boldsymbol{\theta}_i$ (Robinson, 1991).

Of course, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_v$ are unknown in most applications. In this case, there are at least two options open. The first and most popular one consists of estimating $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_v$ from the data, usually using Maximum Likelihood or Restricted Maximum Likelihood methods. The estimators obtained in this way are known as “Empirical Bayes” (see Rao, 2003, ch. 6 and 9). The second estimation strategy we adopt here, is carried out noting the fact that $\hat{\boldsymbol{\theta}}_i^B$ is a posterior mean conditional on $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_v$, i.e. $\hat{\boldsymbol{\theta}}_i^B = E(\boldsymbol{\theta}_i | \boldsymbol{\beta}, \boldsymbol{\Sigma}_v, \hat{\boldsymbol{\theta}}_i)$. In this context, the Hierarchical Bayes estimator is obtained simply as:

$$(4) \quad \hat{\boldsymbol{\theta}}_i^{HB} = E(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_i) = E_{\boldsymbol{\beta}, \boldsymbol{\Sigma}_v}(\hat{\boldsymbol{\theta}}_i^B)$$

where the expectation $E_{\boldsymbol{\beta}, \boldsymbol{\Sigma}_v}(\hat{\boldsymbol{\theta}}_i^B)$ implies the specification of a prior distribution for $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_v$. Note that $\hat{\boldsymbol{\theta}}_i^{HB}$ enjoys the same consistency property as $\hat{\boldsymbol{\theta}}_i^B$.

In general $\hat{\boldsymbol{\theta}}_i^{HB}$ will be difficult to obtain in a closed form, but may be easily approximated by means of the output of Markov Chain Monte Carlo algorithms, which will be discussed a little more in detail in the next section.

With regard to the prior distribution on $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_v$, we assume that $p(\boldsymbol{\beta}, \boldsymbol{\Sigma}_v) = p(\boldsymbol{\beta})p(\boldsymbol{\Sigma}_v)$. For the regression parameter, we specify the prior $\boldsymbol{\beta} \sim N(0, a_1 \mathbf{I}_{KP})$, a conjugate prior that warrants posterior property and fast computation. As regards $\boldsymbol{\Sigma}_v = \text{diag}(\sigma_{v,k})$: for these variances on the diagonal of $\boldsymbol{\Sigma}_v$, we assume *a priori* independence and $p(\sigma_{v,k}) \sim \text{Unif}(0, a_2)$, that is, uniform priors for the standard deviations. These priors for the variance components are justified by Gelman (2006) in terms of both posterior properness and the implications for the Markov Chain Monte Carlo algorithms.

The constants a_1, a_2 are large compared with the scale of data, so as to reflect the lack of prior information about model parameters, thus defining “diffuse proper priors,” whose main goal is to have a very mild impact on posterior inferences. This has been checked by means of sensitivity analyses based on different values of a_1, a_2 and different prior distributions.

Although posterior means are the most popular estimates of area-specific parameters, their empirical distribution function misrepresents one of the underlying population parameters (Ghosh, 1992; Heady and Ralphs, 2004), as it is generally under-dispersed, while that of “direct” estimates is in all cases over-dispersed. To obtain a more reliable picture of the “ensemble” of regional parameters, we consider constrained hierarchical Bayes estimators $\hat{\boldsymbol{\theta}}_i^{CHB}$, first introduced by Ghosh and Maiti (1999), albeit not in a hierarchical context. They may be characterized as the most efficient summaries \mathbf{t}_i of the posterior distribution $p(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_i)$ matching the variability of the small area parameters. That is, we are looking for quantities \mathbf{t}_i that minimize $E \left[\sum_{i=1}^m (\boldsymbol{\theta}_i - \mathbf{t}_i)(\boldsymbol{\theta}_i - \mathbf{t}_i)' | D \right]$ and satisfy the constraints:

$$(5) \quad m^{-1} \sum_{i=1}^m E(\bar{\theta}|D) = m^{-1} \sum_{i=1}^m \mathbf{t}_i$$

$$(6) \quad E \left[\sum_{i=1}^m (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})' | D \right] = m^{-1} \sum_{i=1}^m (\mathbf{t}_i - \bar{\mathbf{t}})(\mathbf{t}_i - \bar{\mathbf{t}})'$$

where $\bar{\theta} = m^{-1} \sum_{i=1}^m \theta_i$, $\bar{\mathbf{t}} = m^{-1} \sum_{i=1}^m \mathbf{t}_i$. Constrained hierarchical Bayes estimators are given by

$$(7) \quad \hat{\theta}_i^{CHB} = \hat{\theta}^{CHB} + (\mathbf{H}_1 + \mathbf{H}_2)^{1/2} \mathbf{H}_2^{-1/2} (\theta_i^{HB} - \hat{\theta}^{HB})$$

where $\mathbf{H}_1 = \sum_{i=1}^m V(\theta_i|D) - mV(\bar{\theta}|D)$ and $\mathbf{H}_2 = \sum_{i=1}^m (\hat{\theta}_i^{HB} - \mathbf{t}_i)(\hat{\theta}_i^{HB} - \mathbf{t}_i)'$. The matrix square roots in (7) are clearly defined, since both \mathbf{H}_1 and \mathbf{H}_2 are semi-positive definite.

Uncertainty associated with $\hat{\theta}_i^{CHB}$ may be measured by its Posterior Mean Square Error, defined as $PMSE(\hat{\theta}_i^{CHB}) = V(\theta_i|D) + (\hat{\theta}_i^{CHB} - \hat{\theta}_i^{HB})(\hat{\theta}_i^{CHB} - \hat{\theta}_i^{HB})'$. In Fabrizi *et al.* (2007) properties of constrained estimated are empirically evaluated based on a simulated data set. The results obtained show that $PMSE(\hat{\theta}_i^{CHB})$ provides an approximation of the frequentist Mean Square Error associated to $\hat{\theta}_i^{CHB}$ with properties similar to those of frequentist estimators of the Mean Square Error of Empirical Best predictors as those considered in Prasad and Rao (1990).

In selecting covariates to enter in the \mathbf{Z}_i matrix, given that our main aim is to reduce the variability of direct estimates, we need to limit the choice to the available information about regional populations or to reliable regional estimates. We then consider the following as possible covariates measured by Censuses or administrative sources: the *percentage of employed people working in the manufacturing sector out of the total number of employed* (MAN), the *percentage of people with a secondary school education* (SS), *average household size* (HS), *per-capita GDP* (GDP) obtained from the National Accounts, and *per-capita disposable income* (DI) obtained from tax documents. As far as regards sampling estimates, we consider the *unemployment rate* (U) provided by the Italian Labour Force Survey (Istat, 2003) which is characterized by very low variability (on average the regional CV is 0.003).

Of the above-mentioned covariates, U generally proves to be the best predictor, as it displays good or satisfactory predictive power for all outcome parameters (with R^2 ranging from 0.36 to 0.88). The inclusion of the other possible covariates in the model produces a negligible improvement in the goodness of fit. Given this, together with the timeliness of the information on U, we decided to consider the latter as a covariate for all indicators. Eventually, we should point out that U is the proxy variable for poverty that is recommended by the Laeken Council. However, even if the sampling variability of U is very low, it could raise the variance of the final estimates. If the HB approach is adopted, this further source of uncertainty can be taken into account simply by modeling \mathbf{Z}_i as random, given that $\mathbf{Z}_i \sim N(\mathbf{X}_i, \hat{V}(\mathbf{X}_i))$, $\hat{V}(\mathbf{X}_i)$ being the randomization variance matrixes calculated according to Istat (2003).

6. MODEL-BASED REGIONAL POVERTY ESTIMATES

We approximate the estimators $\hat{\theta}_i^{HB}$ defined in (4) and used in (7) by means of a Monte Carlo Markov Chain (MCMC) algorithm. For a concise introduction to this numerical method applied to Bayesian inference, see, for instance, Carlin and Louis (2000, ch. 5). Since they are becoming increasingly popular, several software packages implementing MCMC procedures are now available. We have chosen to use the Openbugs software package (Thomas *et al.*, 2006) which is very widely used in applied hierarchical modeling, is downloadable for free on the internet, and is open source.

As algorithms are based on the idea of simulating Markov Chains converging to the posterior distribution in question, special care needs to be taken when assessing the convergence. To this end, we run three parallel chains of $R = 25,000$ runs each, the starting point of which is taken from an over-dispersed distribution, and we monitor convergence by visual inspection of the chains plots. Moreover, the Gelman and Rubin statistic is also computed (Gelman and Rubin, 1992) and the autocorrelation diagrams analyzed. Although all the chains involved in our model converge quickly, as a precaution we conservatively discard the first 5,000 iterations from each chain.

Once convergence is assessed and the pre-convergence (or “burn-in”) parts of the chains discarded, generations from the three parallel chains pertaining to a parameter of interest are pooled together and the posterior expectation approximated by the sample mean. Since these pooled samples from the posterior distributions are large (60,000) and the chains, as it is often the case in multivariate models, exhibit a fairly strong degree of autocorrelation, we decided to reduce the chain to one third of its original size by keeping only 1 out of every 3 values generated. To put it briefly, we use the following approximation of (4):

$$(8) \quad \hat{\theta}_i^{HB,MCMC} = (3S)^{-1} \sum_{s=1}^S \sum_{c=1}^3 (\theta_i(s, c) | \mathbf{y})$$

where $S = 20,000$ is the MCMC sample size and $\theta_i(s, c) | \mathbf{y}$ is the s -th value generated from the c -th chain after discarding the burn in parts of the chains, thinning for autocorrelation and pooling the chains together.

The constrained estimators $\hat{\theta}_i^{CHB,MCMC}$ are obtained by plugging $\hat{\theta}_i^{HB,MCMC}$ into (7).

In keeping with most Bayesian literature, the model’s goodness of fit is assessed by means of Posterior Predictive checks (see Carlin and Louis, 2000, section 2.4). By adopting the discrepancy measure proposed in Datta *et al.* (1999), we evaluated the fit of the model as being adequate, the posterior predictive checks ranging from 0.20 to 0.68.

The resulting model based estimates and their standard errors are reported in Table 2. It should be pointed out that almost all estimates reveal a much smaller standard error than that of their direct counterparts, and one that is generally adequate for the purposes of comparative analysis. A more detailed analysis of results, contained in Fabrizi *et al.* (2005), shows that the reduction in the variance of model based estimators compared to direct estimators (averaged over the regions) ranges from 27.3 percent of RPT to 63.6 percent of ARPR; it is relatively

TABLE 2
MODEL-BASED ESTIMATES AND THEIR STANDARD ERRORS (SE)

Macro-region ^(a)	Region	PCI		RPT		ARPRr		ARPRn		G	
		Estimate*	SE	Estimate*	SE	Estimate**	SE	Estimate**	SE	Estimate**	SE
North	Piemonte	24.9	0.75	13.7	0.55	12.2	1.70	8.9	1.86	24.7	1.48
	Valle D'Aosta	25.4	1.02	14.2	0.42	11.5	2.07	11.6	2.33	26.5	1.86
	Liguria	23.9	0.95	13.2	0.53	12.9	1.81	9.2	2.11	23.2	1.39
	Lombardia	27.1	0.81	14.8	0.44	15.6	1.34	5.7	1.76	24.4	1.22
	Bolzano	24.8	0.99	14.0	0.42	12.9	1.94	9.5	2.29	25.1	1.76
	Trento	22.6	0.68	14.2	0.71	15.1	2.56	14.6	3.80	20.6	2.06
	Veneto	25.0	0.75	13.9	0.28	12.5	1.37	5.8	1.04	26.2	1.84
	Friuli V.G.	26.2	0.79	14.3	0.57	6.6	2.17	4.1	0.95	20.4	1.63
	E. Romagna	26.2	0.79	14.7	0.44	13.0	1.30	5.0	0.85	24.3	1.21
	Toscana	23.9	0.72	13.4	0.54	14.6	2.19	9.3	2.69	24.0	1.68
Centre	Umbria	22.2	0.89	13.1	0.52	9.9	1.97	8.5	1.95	21.3	1.49
	Marche	24.7	0.74	13.7	0.55	11.7	1.99	7.0	1.25	26.1	1.57
	Lazio	21.1	0.63	11.7	0.35	15.0	1.80	17.6	2.29	26.8	1.61
	Abruzzo	21.5	0.43	12.5	0.12	13.3	1.86	12.7	2.03	23.3	1.16
	Molise	18.1	0.72	9.5	0.48	11.1	1.89	21.6	4.54	27.0	1.08
South	Campania	16.1	0.48	8.8	0.26	15.8	1.73	32.6	2.93	26.3	1.05
	Puglia	18.7	0.75	10.1	0.40	16.5	2.31	21.3	3.84	27.9	1.40
	Basilicata	19.6	0.98	10.5	0.52	15.2	2.28	20.7	4.76	27.6	1.65
	Calabria	16.7	1.17	8.6	0.43	15.3	1.68	40.0	4.00	29.3	2.05
	Sicilia	17.5	0.88	8.6	0.34	16.2	1.30	33.5	3.68	31.8	1.59
	Sardegna	18.9	0.76	9.8	0.49	15.1	1.51	31.5	2.83	28.4	1.13

Notes: ^(a)According to the ISTAT classification. *In thousands of euros, **as a percentage.

smaller for PCI (32.8 percent) and ARPRn (39.4 percent), and larger for G (53.2 percent). A model comparison exercise, that leads to the adoption of the model considered in this paper, may also be found in Fabrizi *et al.* (2005).

7. A CROSS-REGIONAL COMPARISON OF INCOME POVERTY

As our interest lies in comparing regions, besides the analysis of regional estimates obtained by following the proposed small area procedure, we also consider their ratios to national estimates (in percentage terms). For each poverty indicator previously described, we can define an index $I_{ik} = (\hat{\theta}_{ik} / \hat{\theta}_{ck}) \times 100$ where $k = \text{PCI, RPT, ARPRr, ARPRn, G}$, and $\hat{\theta}_{ck}$ is the national estimate. Moreover, we also consider the respective differences from the national estimate, that is, $\Delta I_{ik} = I_{ik} - 100$.

We are first going to analyze regional disparity on the basis of the calculated indicators, and then consider the five indicators from a multivariate perspective, in order to single out a subdivision of all administrative regions into homogeneous groups in terms of the complete set of poverty indicators in question (see Section 7.2).

In the analysis that follows, no reference will be made to the NUTS1 regions, but to the three Italian macro-regions (North, Centre, and South), a subdivision that is more commonly found in studies of Italy's regional economics. This subdivision is also more relevant from the economic policy viewpoint: all the administrative regions in the South of Italy, with the exception of the Abruzzo (from the mid-1990s onwards) are Objective 1 regions for the purposes of the allocation of EU structural funds (European Commission, 1999).

7.1. Regional Comparisons

If we consider per-capita income (PCI), the regional poverty threshold (RPT), the poverty rate based on the national threshold (*ARPRn*) and, to a lesser extent, the inequality expressed by the Gini coefficient (G), the most striking feature of the estimates set is the large difference between Italy's southern and northern regions, that is, the "North-South divide" that notoriously characterizes the country.

In particular, Table 2 and Figure 1 showing, respectively, level estimates and ΔI_{ik} , reveal the enormous differences in terms of both the proportion of people living below the poverty threshold, and the inequality within regions.

Beside the large differences in mean income, this may also depend on the limited effectiveness of social welfare expenditure in Italy. It has been observed that this spending is less affected by the unemployment rate than it is by the proportion of people above retirement age (which is larger in the northern regions), as the unemployed receive comparatively little compared with pensioners, whose level of pension is quite closely tied to previous income, thus favoring the North of Italy once again (European Commission, 2004, part II).

As expected, the distribution of the proportion of poor people calculated according to region-specific thresholds is much more uniform across the country. Note that the poverty rates based on regional thresholds are sometimes regarded as inequality measures. Therefore the fact that among the seven regions reporting

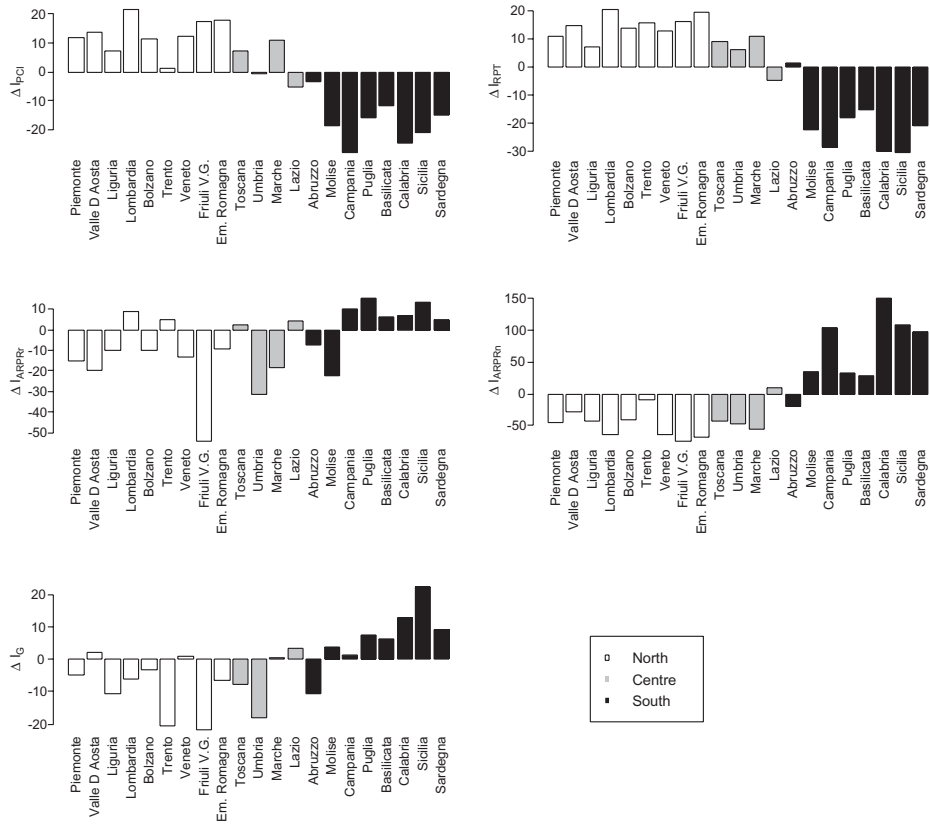


Figure 1. Estimates of Poverty Indicators (difference from the national average)

the highest values of ARPR_r we have the six poorest regions in terms of ARPR_n, may be due to the fact that these regions are the worst off in terms of inequality; this evidence is supported also by the estimates of the Gini index *G*.

Moreover, from ARPR_r estimates we may observe that values above the national average characterize also the Lombardy region, the most populous, highly-industrialized region in the North of the country; although not statistically significant, there is some evidence that this region is threatened by greater problems of social cohesion than the neighboring Emilia Romagna and Veneto regions, which have comparable per-capita levels of income.

Furthermore, the two rates are to some extent consistent. In fact, the seven poorest regions in terms of ARPR_r include the six poorest regions in terms of ARPR_n.

The comparison between *I_{RPT}* and *I_{ARPRr}* reveals certain features of the shape of income distribution for different regions. Figure 2, showing *I_{RPT}* and sorted *I_{ARPRr}*, reveals that the southern regions of Basilicata and Apulia boast higher living standards than the rest of the South: while they share a similar level of exposure to the risk of poverty, their thresholds are higher than those of the other southern regions.

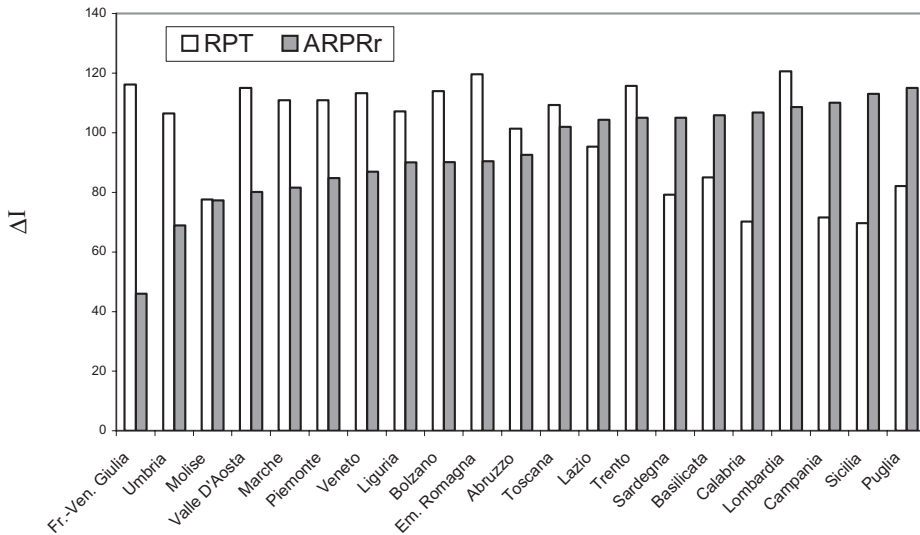


Figure 2. Regional Poverty Thresholds and At-Risk-Of-Poverty Rate Based on Regional Threshold

Another clear feature of the estimates is the substantial differences within all three macro regions, in terms of all the indicators in question. This is particularly significant in the South, which although being the least developed part of the country, is also the focal point for all national economic policy and EU social welfare action. If we focus on the South, we will note that the Abruzzo is clearly different from all the other regions (more affluent, with lower poverty rates and less inequality). This would seem to support the exclusion of Abruzzo from the Objective 1 intervention region during the 1990s, which at that time was based exclusively on Italy's National Accounts. Moreover, Basilicata and Apulia are relatively better off compared with the rest of the South (higher PCI, RPT, lower ARPRn and G); this is in keeping with the two regions' growth rates—well above the national average—for the second half of the 1990s. Moreover, the high poverty rate, based on the regional threshold, characterizing the Apulia region (and to a slightly lesser extent Basilicata) may be a sign of an unbalance in the relatively rapid economic growth the two regions have recently experienced

In Figure 3, I_{ARPRn} is plotted against I_G : the polarization of the ensemble of Italian regions is clear, with most of the Northern and Central administrative regions lying in the lower left quadrant, while the Southern regions (with the exception of the Abruzzo) are situated in the upper right quadrant. Of the latter, the plight of Sardinia, Campania, Sicily and Calabria is the worst of all. These regions constitute the most economically "lagging behind" part of the country, as is confirmed by the other indicators considered in this paper, and are characterized by a series of well-known structural weakness: high unemployment, low participation in the labor market, a large share of the workforce employed in agriculture, low productivity in the private sector, and the persistent presence of organized crime (Beutel, 2002). It is therefore clear that EU and national economic policies need to focus more closely on this subset of regions rather than on the entire South

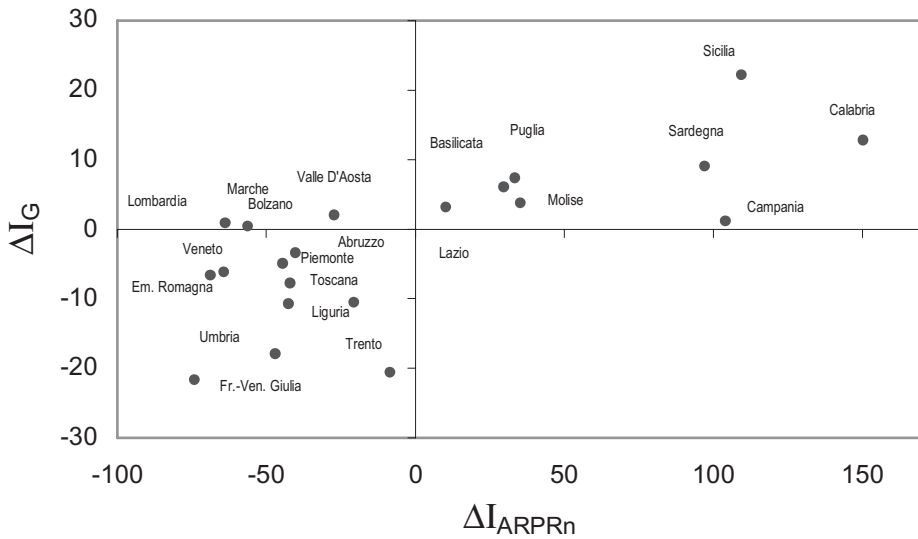


Figure 3. Scatterplots of ARPRn and G (percent differences from the national average)

of Italy. Finally, certain Northern regions boasting a low poverty rate, are also characterized by a non-trivial level of G.

7.2. Towards a Clustering of Regions According to Poverty Indicators

In order to clarify the so-called North–South divide in terms of the indicators in question, we are now going to show how Italy’s administrative regions can be consistently clustered into two distinct groups. This we do by utilizing Principal Component Analysis (PCA) as a pre-processing step towards the clustering of regions, as recommended in Everitt (1993). The reason for this is the high correlation between the indicators (which range, in absolute values, from 0.48 to 0.96). Moreover, PCA attempts to reduce the dimension of the reference space, and to allow a clear graphical representation to be formulated. We consider PCA based on the correlation matrix.

We then perform a Cluster Analysis of the extracted Principal Components (PCs), in order to subdivide the regions into groups. We use the SAS software package for both analyses. The usual criteria employed when selecting the number of PCs led us to choose a solution summarizing the data using just two variables, and thus accounting for 90 percent of the original variance (77 percent and 13 percent respectively). The reduction in the number of dimensions aids data visualization, as it allows most of the information to be visualized in two dimensions (Figure 4). The meaning of these components can be gleaned from their respective coefficients, shown in Table 3.

The first component is a weighted average of PCI, RPT, ARPRn and G. Those regions with positive, high values along this component are characterized by high levels of income poverty and inequality. The second component represents only the ARPRr indicator, as it is positively correlated to it. In order to segment

TABLE 3
EIGENVECTORS OF THE FIRST TWO PRINCIPAL COMPONENTS

	PCI	RPT	ARPRr	ARPRn	G
Component 1	-0.47	-0.48	0.35	0.49	0.42
Component 2	0.34	0.32	0.83	-0.17	0.25

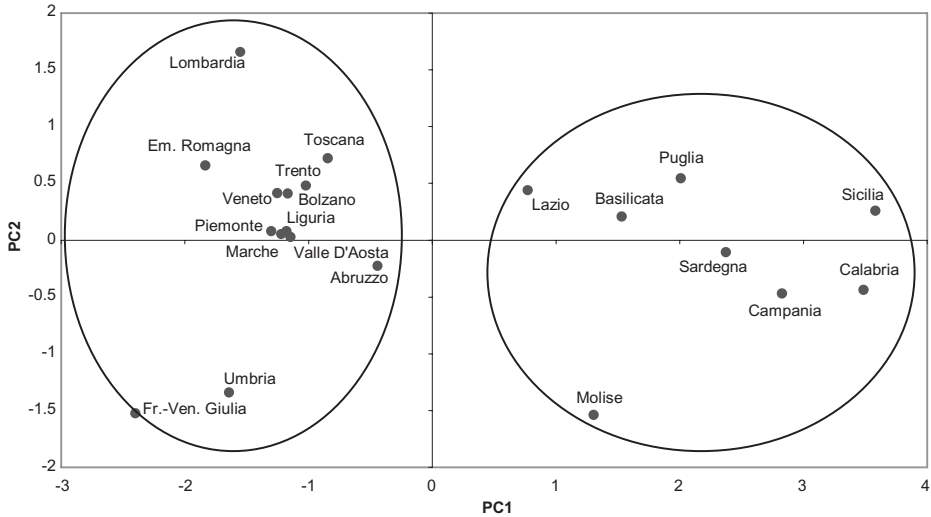


Figure 4. Scatterplot Referred to the First Two Principal Components

all the regions, we adopted a hierarchical (Ward) clustering technique, followed by a non-hierarchical (K-means) clustering technique. This enabled us to identify two groups (with a considerable value of R^2 —equal to 0.72—indicating a good degree of separation between the groups). Figure 4 plots all regions in relation to the first and second PC, and the identified clusters are represented by ellipses.

The results obtained confirm the idea that the central macro-region does not possess any particular economic significance, since the North–South divide is clearly present in this macro-region as well: in fact, the central region appears split into two sub-groups, with Tuscany, Abruzzo, Marche and Umbria part of the larger group containing the northern regions (shown on the left-hand side of the graph), while Lazio and Molise are part of the group of poorer regions (together with the rest of the South).

8. CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER STUDY

In this paper, we have managed to obtain reliable regional estimates for various Laeken poverty indicators, using a multivariate small area model. These estimates allow us to analyze regional disparities in terms of income poverty levels, with reference to the NUTS2 Italian regions.

The results show that the classification of Italian regions by means of administrative criteria into two homogeneous groups—the richer northern regions and the poorer southern regions—can be misleading. What actually emerges from our study is a much more complex picture of poverty. In fact, some of Italy's southern regions appear more similar to their northern counterparts in terms of poverty, while certain northern regions, although characterized by a low poverty rate, nevertheless display a noticeable degree of inequality.

Some aspects of this study merit further investigation: firstly, our analysis could be broadened to include some other Laeken indicators of poverty (e.g. the Income Quintile Share Ratio, the At-Risk-of-Poverty Gap) or of inequality. Moreover, other interesting breakdowns of estimates could be considered besides the traditional regional ones. For example, estimates could be broken down in terms of household type, occupational status and age (in order to highlight child poverty). Finally, our proposed approach could also be adopted in order to measure poverty in other European regions, using the information produced by the new EU Survey on Income and Living Conditions, which is replacing the European Community Household Panel. Furthermore, given that the EU-SILC is a rotating panel survey, the repetition of observations of the same units could be used to further improve the reliability of estimates, and to monitor the time dynamics of poverty.

REFERENCES

- Beutel, Joerg, *The Economic Impact of Objective 1 Intervention for the Period 2000–2006*, Final Report to the Directorate-General for Regional Policies of the European Commission, Basel, August 2002.
- Cannari, Luigi and Giovanni D'Alessio, "Income and Wealth Distribution in the Italian Regions," *Temi di Discussione del Servizio Studi*, Banca D'Italia, 482, 2003 [in Italian].
- Carlin, Bradley P. and Thomas A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall, New York, 2000.
- Coccia, Giuliana and Alessandra Masi, "Analyzing Regional Poverty Indicators," *Rivista di Statistica Ufficiale, Istat*, 2, Franco Angeli, 2003 [in Italian].
- Datta, Gauri S., Partha Lahiri, Tapabrata Maiti, and Kim L. Lu, "Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S.," *Journal of the American Statistical Association*, 94, 1074–82, 1999.
- European Commission, Decision 1999/502/EC of 1 July 1999, *Official Journal of the European Union*, L194 of 07/27/1999, 1999.
- , *A New Partnership for Cohesion: Convergence, Competitiveness, Cooperation*. Third report on economic and social cohesion, Office for the Official Publications of the European Communities, Luxembourg, 2004.
- , *Regional Indicators to Reflect Social Exclusion and Poverty*, Report prepared for Employment and Social Affairs DG, with contribution of Gianni Betti, Achille Lemmi (Project Director), Anna Mulas, Michela Natilli, Laura Neri, Nicola Salvati and Vijay Verma (Research Director), 2005a.
- , *Third Progress Report on Cohesion—Towards a New Partnership for Growth, Jobs and Cohesion*, Brussels, 2005b.
- Eurostat, *European Social Statistics—Income, Poverty and Social Exclusion*, 2nd report, 2002.
- Everitt, Brian, *Cluster Analysis*, 3rd edition, Edward Arnold, 1993.
- Fabrizi, Enrico, Maria R. Ferrante, and Silvia Pacei, "Estimation of Poverty Indicators at the Sub-National Level Using Multivariate Small Area Models," *Statistics in Transition*, 7, 587–608, 2005.
- , "A Comparison of Adjusted Bayes Estimators of Use in Small Area Estimation," University of Bergamo, Technical Report, no. 1, 2007.
- Ferrante, Maria R., Silvia Pacei, and Enrico Fabrizi, "Estimation of Household Income Frequency Distribution in Small Areas," presented as poster at the IMS/ASA SRMS Joint Mini Meeting on Current Trends in Survey Sampling and Official Statistics, Calcutta, India, January 1–3, 2004.

- Gelman, Andrew, "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 1, 513–33, 2006.
- Gelman, Andrew and Donald B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–72, 1992.
- Ghosh, Malai, "Constrained Bayes Estimation with Application," *Journal of the American Statistical Association*, 87, 533–40, 1992.
- Ghosh, Malai and Tapabrata Maiti, "Adjusted Bayes Estimators with Applications to Small Area Estimation," *Sankhyā*, Series B, 61, 71–90, 1999.
- Ghosh, Malai, Narinder Nangia, and Dalho Kim, "Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach," *Journal of the American Statistical Association*, 91, 1423–31, 1996.
- Goedhart, Theo, Victor Halberstadt, Arie Kapteyn, and Bernard van Praag, "The Poverty Line: Concept and Measurement," *Journal of Human Resources*, 12(4), 503–20, 1977.
- Greene, William H., *Econometric Analysis*, Prentice Hall, 2003.
- Heady, Patrick and Martin Ralphs, "Some Findings of the Eurarea Project—and their Implications for Statistical Policy," *Statistics in Transition*, 6(5), 641–54, 2004.
- Isaki, Cary T. and Wayne A. Fuller, "Survey Design Under the Regression Superpopulation Model," *Journal of the American Statistical Association*, 77, 89–96, 1982.
- Istat, Forze di lavoro—Media 2002, *Annuari*, 2003.
- Kapteyn A., P. Kooreman, and R. Willems, "Some Methodological Issues in the Implementation of Subjective Poverty Definitions," *The Journal of Human Resources*, 23(2), 222–42, 1988.
- Louis, Thomas A., "Estimating a Population of Parameter Values Using Bayes and Empirical Bayes Methods," *Journal of the American Statistical Association*, 79, 393–8, 1984.
- Mogstad, Magne, Audun Langørgen, and Rolf Aaberge, "Region-Specific Versus Country-Specific Poverty Lines in Analysis of Poverty," *Journal of Economic Inequality*, 5, 115–22, 2007.
- Prasad, N. G. Narasimha and Jon N. K. Rao, "The Estimation of Mean Squared Error of Small Area Estimators," *Journal of the American Statistical Association*, 85, 163–71, 1990.
- Rao, Jon N. K., "Some Current Trends in Sample Survey Theory and Methods," *Sankhyā*, Series B, 61, 1–57, 1999.
- , *Small Area Estimation*, John Wiley & Sons, New Jersey, 2003.
- Robinson, G. K., "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–32, 1991.
- Sen, Amartya, "Issues in the Measurement of Poverty," *Scandinavian Journal of Economics*, 81, 295–307, 1979.
- , "Poor, Relatively Speaking," *Oxford Economic Papers*, 35(2), 153–69, 1983.
- Thomas, Andrew, Bob O'Hara, Uwe Ligges, and Sibylle Sturz, "Making BUGS Open," *R News*, 6, 12–17, 2006.
- Verma, Vijay and Gianni Betti, "Sampling Errors for Measures of Inequality and Poverty," *Classification and Data Analysis 2005, Book of Short Papers*, CLADAG, 175–8, 2005.