Computational Statistics and Data Analysis 55 (2011) 1736–1747

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Hierarchical Bayes multivariate estimation of poverty rates based on increasing thresholds for small domains

Enrico Fabrizi [a], Maria Rosaria Ferrante [b], Silvia Pacei [b,*], Carlo Trivisano [b]

[a] *DISES, Università Cattolica, Piacenza, Italy*

[b] *Dipartimento di Scienze Statistiche 'P. Fortunati', Università di Bologna, Italy*

ARTICLE INFO

Article history:
Received 28 October 2009
Received in revised form 14 September 2010
Accepted 5 November 2010
Available online 21 November 2010

Keywords:
Fay–Herriot model
Beta distribution
Hierarchical Bayes modeling

ABSTRACT

A model-based small area method for calculating estimates of poverty rates based on different thresholds for subsets of the Italian population is proposed. The subsets are obtained by cross-classifying by household type and administrative region. The suggested estimators satisfy the following coherence properties: (i) within a given area, rates associated with increasing thresholds are monotonically increasing; (ii) interval estimators have lower and upper bounds within the interval (0, 1); (iii) when a large domain-specific sample is available the small area estimate is close to the one obtained using standard design-based methods; (iv) estimates of poverty rates should also be produced for domains for which there is no sample or when no poor households are included in the sample. A hierarchical Bayesian approach to estimation is adopted. Posterior distributions are approximated by means of MCMC computation methods. Empirical analysis is based on data from the 2005 wave of the EU-SILC survey.

1. Introduction

Poverty and social exclusion are unevenly distributed both geographically and across social groups. This is particularly true for Italy, a country characterized by a low degree of regional cohesion (European Commission, 2005) and where the disparities between household types interact with those between the different regions.

We focus on the estimation of three different poverty rates for domains (sub-populations) obtained by cross-classifying the Italian population by household type and Administrative Region. Estimates will be based on data collected by the 'European Union-Statistics on Income and Living Conditions' survey (EU–SILC — 2nd wave, year 2005). The three poverty rates make reference to increasing poverty thresholds that are all defined as fractions of the national median of the equivalized disposable income. As a consequence the three rates are in ascending order and they are respectively aimed at measuring the portion of very poor people, of poor people and of people who are either poor or risk becoming poor (ISTAT, 2007).

The EU–SILC survey is designed to provide reliable estimates of the main parameters of interest for sub-populations that are much bigger than those we target. Moreover our domains are not planned (they are not survey design strata or unions of strata), so no minimum sample size in these domains is guaranteed. The number of units sampled from a large number of the domains we consider is too low. Calibration estimators calculated following the same methodology as that employed by the National Institute of Statistics (ISTAT) for larger domains, are easily shown to be too imprecise. To improve this 'direct estimation' strategy, a small area method is advisable.

* Corresponding address: Dipartimento di Scienze Statitiche 'P. Fortunati', Università di Bologna, via Belle Arti 41, 40126 Bologna, Italy. Tel.: +39 051 2098229; fax: +39 051 232153.
E-mail address: silvia.pacei@unibo.it (S. Pacei).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.11.001

Among the many small area methods proposed in the literature, we refer to model-based small area estimators relying on area-level models (see Rao (2003), section 5.4.1). In our application multivariate models are advisable, since we have three poverty rates for each area and their estimators are of course correlated. Given the nature of the target parameters, we require that small area estimators of the poverty rates satisfy the following coherence properties: (i) within a given area, rates associated with increasing thresholds should also be monotonically increasing; (ii) interval estimators should have lower and upper bounds within the interval (0, 1); (iii) when a large domain-specific sample is available the small area estimate should be reasonably close to the direct estimate; (iv) estimates of poverty rates should also be produced for domains for which we have no sample or when no poor households are included in the sample.

Properties (i) and (ii) are not satisfied by popular multivariate small area models relying on the assumption of normality (see Ghosh et al. (1996) and Datta et al. (1998)). To obtain estimators satisfying these properties, we propose a model assuming that, within each area, the differences between successive poverty rates are conditionally beta distributed and exchangeable. This modelling of differences is justified by the fact that the direct estimators of rates based on increasing thresholds are not independent, while differences between successive poverty rates estimators have a much lower, practically negligible, correlation. This assumption avoids the complicated generalization of the beta to the multivariate case and allows for an unrestricted correlation structure between rate estimators.

We then propose a multivariate logistic-normal model for the expected values of the Beta distributions, thus obtaining a hierarchical two-stage model. The coherence property (iii) requires the inclusion of sampling weights into the estimation process, and this is what we do using the calibration estimators as input for our models; in this respect our approach differs from that of Molina et al. (2007). The estimators in question also meet the coherency requirement (iv). Our proposal represents an alternative to other solutions to the problem of estimating small area proportions recently presented in the literature (Gonzàlez-Manteiga et al., 2002; Liu et al., 2007; Lohr and Rao, 2009; Longford, 2010).

As far as estimation is concerned, we adopt a hierarchical Bayesian approach implemented by means of a MCMC computational method. The results obtained allow us to compare the incidence of poverty by household type in the different Italian administrative regions. The suggested method could be extended to the estimation of other indicators and could be used with data collected by the EU–SILC in other countries.

The paper is organized as follows. In Section 2 we briefly review the EU–SILC survey, while in Section 3, domains and target parameters are defined. Direct estimators and the estimation of their variance are discussed in Section 4. Section 5 introduces the suggested small area models while Section 6 is devoted to model checking and sensitivity analyses. The evaluation of the performance of the associated small area estimators is discussed in Section 7. Conclusions and possible future developments are outlined in the final Section 8.

2. The data

The EU–SILC survey (European Parliament and Council, 2003; Eurostat, 2005) survey has the aim of collecting timely and comparable cross-sectional and longitudinal microdata on income, poverty, social exclusion and living conditions. The sampling design is a rotating panel based on consistent methodology and definitions across most member states of the European Union (EU). The survey is conducted in each country by the relevant national institute of statistics (in Italy, by ISTAT) and coordinated by Eurostat, the Statistical Bureau of the EU. In Italy, the first wave of the EU–SILC survey was launched in 2004. In this paper we analyse data from the 2005 wave. The income reference period is 2004.

Survey units (households) are sampled according to a stratified two-stage sampling design. First-stage units are given by municipalities, stratified according to administrative region and demographic size. Of the municipalities those with at least 30,000 inhabitants are considered self-representative and form a take-all stratum. Secondary sampling units are given by households. The effective sample of the 2005 wave of the survey contains 22,032 households and a total of 56,105 individuals. In Italy, the survey is designed to obtain reliable estimates at the administrative region level. For more details on the EU–SILC survey see also ISTAT (2007). As far as the domains of interest to us are concerned, the number of households ranges from a minimum of 4 to a maximum of 600; 25th, 50th and 75th percentiles are respectively 45, 97 and 165. Note that the domains of interest, that will be described in detail in the next section, are not planned, i.e. they cut across strata, so a minimum sample size in these domains is guaranteed.

3. Definition of the domains and of the target parameters

We identify 180 domains of interest, obtained by cross-classifying the population of the 20 Italian administrative regions by the 9 household types considered in the EU–SILC survey. These types are defined by simultaneously considering the household size, the presence of children and the age of household members. They are defined as follows: 1. one-person households; 2. two adults, no dependent children, both adults under 65 years of age; 3. two adults, no dependent children, at least one adult 65 years or more years old; 4. other households without dependent children; 5. single parent households, one or more dependent children; 6. two adults, one dependent child; 7. two adults, two dependent children; 8. two adults, three or more dependent children; 9. other households with dependent children. Dependent children comprise all persons below the age of 16 and persons aged 16–24 who are living in the household with at least one of their parents and who are economically inactive.