Small Area Methodologies for Poverty Estimation: an Application to Italian Data

Giusti, Caterina
Department of Statistics and Mathematics Applied to Economics, University of Pisa
Via C.Ridolfi, 10
56124 Pisa, Italy
E-mail: caterina.giusti@ec.unipi.it

Marchetti, Stefano
Department of Statistics and Mathematics Applied to Economics, University of Pisa
Via C.Ridolfi, 10
56124 Pisa, Italy
E-mail: stefano.marchetti@for.unipi.it

Pratesi, Monica
Department of Statistics and Mathematics Applied to Economics, University of Pisa
Via C.Ridolfi, 10
56124 Pisa, Italy
E-mail: m.pratesi@ec.unipi.it

Tzavidis, Nikos
Social Statistics and Southampton Statistical Science Research Institute, University of Southampton
Southampton
SO17 1BJ, United Kingdom
E-mail: : n.tzavidis@soton.ac.uk

The estimation and dissemination of poverty, inequality and life condition indicators all over the European Union is nowadays one topic of primary interest. Such indicators should assist in monitoring living conditions and in guiding the implementation of policies that aim at improving the living conditions in the EU Member States. This interest is reflected in the major investment that the European Commission has made by funding the SAMPLE (Small Area Methods for Poverty and Living conditions Estimates) project under the auspices of the 7th Framework (www.sample-project.eu). In particular, the knowledge of the cumulative distribution function of the household disposable equivalised income represents an important source of information for the living conditions in a given area. From the cumulative distribution function of the household income many quantities (e.g. the median income, the income quantiles) and monetary poverty indicators (e.g. the Head Count Ratio or at-risk-of-poverty-rate) can be computed. Other quantities, such as the average of the household income, can be highly influenced by outlying values. The estimation of these target parameters at the small area level using data coming from major sample surveys, such as the EU-SILC (European Union Statistics on Income and Living Conditions) survey, can be performed by a variety of methods. Since M-quantile models (Chambers and Tzavidis, 2006 and Tzavidis et al. 2008; 2010) do not impose strong distributional assumptions and are outlier robust, the use of these models for poverty estimation may protect against departures from assumptions of the traditional unit-level nested error regression model for small area estimation. Estimation of the variability of M-quantile poverty estimates can be obtained analytically in the case of the average income (Chambers et al., 2007), or using a nonparametric bootstrap approach for the quantiles of the income and for the HCR indicator. The aim of this work is to estimate some measure of poverty for the Provinces of three

Italian regions, Lombardia, in the North of the Country, Toscana, in Central Italy, and Campania, in Southern Italy by using the M-quantile methods using data from the EU-SILC survey 2008 and from the Population Census of Italy.

Theory

Let $\mathbf{x}_i$ be a known vector of p auxiliary variables for each population unit j in small area i and assume that information for the variable of interest y is available only on the sample. Chambers and Tzavidis (2006) have developed an approach to small area estimation based on the quantiles of the conditional distribution of the variable of study (y) given the covariates (Breckling and Chambers, 1988). The qth M-quantile $Q_q(x;\psi)$ of the conditional distribution of y given x satisfies:

(1) $$Q_q(\mathbf{x}_{ij};\psi) = \mathbf{x}_{ij}^T\beta_\psi(q)$$

where ψ denotes the influence function associated with the M-quantile. For specified q and continuous ψ, an estimate $\hat{\beta}_\psi(q)$ of $\beta_\psi(q)$ is obtained via an iterative weighted least squares algorithm. When (1) holds the bias adjusted M-quantile predictor of m_i, the mean of the variable of interest in area i, is:

(2) $$\hat{m}_i^{MQ/CD} = N_i^{-1}\Big[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i}\mathbf{x}_{ij}^T\hat{\beta}_\psi(\hat{\theta}_i) + \frac{N_i - n_i}{n_i}\sum_{j\in s_i}(y_{ij} - \hat{y_{ij}})\Big]$$

where s_i denotes the n_i sampled units in area i, r_i denotes the remaining $N_i - n_i$ units in the area ($s_i \cup r_i = U_i$, with U_i the population in area i), $\hat{y}_{ij} = \mathbf{x}_{ij}^T\hat{\beta}_\psi(\hat{\theta}_i)$ is a linear combination of the auxiliary variables and $\hat{\theta}_i$ is an estimate of the average value of the M-quantile coefficients of the units in area i (Tzavidis and Chambers, 2007). The MSE of the estimator (2) can be estimated analytically as suggested in Chambers et al. (2007).

Although small area averages are widely used in small area applications, relying only on averages may not provide a very informative picture about the distribution of wealth in a small area. In economic applications for example, estimates of average income may not provide an accurate picture of the area wealth due to the high within area inequality. For this reason, it is of interest on poverty analysis to focus on the estimation of poverty indicators such as the Head Count Ratio (HCR) and Poverty Gap (PG) (see Foster et al. 1984). Denoting by t the poverty line, the poverty measures proposed by Foster et al. (1984) for a small area i are defined as

(3) $$F_{\alpha i} = N_i^{-1}\sum_{j\in U_i}\Big(\frac{t - y_{ij}}{t}\Big)^\alpha I(y_{ij} \leq t)$$

Setting $\alpha = 0$ defines the HCR whereas setting $\alpha = 1$ defines the PG. Estimation of these indicators under the M-quantile approach can be obtained by a Monte Carlo procedure:

1. Fit the M-quantile small area model (1) using the sample values and obtain estimates of β_ψ and θ_i;

2. Draw an out of sample vector using $y_{ijh}^* = \mathbf{x}_{ijh}^T\hat{\beta}_\psi(\hat{\theta}_i) + e_{ijh}^*$, where e_{ijh}^* is a vector of size $N_i - n_i$ drawn from the Empirical Distribution Function (EDF) of the estimated M-quantile regression residuals or from a smooth version of this distribution and $\hat{\beta}_\psi$, $\hat{\theta}_i$ are obtained from the previous step;

3. Repeat the process H times. Each time combine the sample data and out of sample data for estimating the target using $\hat{F}_{\alpha i}^{MQ} = N_i^{-1}\Big[\sum_{j\in s_i} I(y_{ij} \leq t) + \sum_{j\in r_i} I(y_{ij}^* \leq t)\Big]$;

4. Average the results over H simulations.

A mean squared error estimate of the M-quantile estimates of the poverty indicators obtained with this procedure can be computed using the non-parametric bootstrap approach described in Tzavidis et al. (2010).

Application

In this section we present the results for the HCR, PG and for the corresponding Root Mean Squared Errors (RMSEs) estimated using the M-quantile models approach for the Provinces of the Tuscany Region. The working M-quantile small area model uses data coming from the EU-SILC survey 2008 for the sampled households in the Region, and data coming from the Population Census 2001 for all the households living in the Region. In the working model the equivalised household income is the outcome variable. The explanatory variables, common in the EU-SILC survey and in the Census micro-data, include the ownership status, the age of the head of the household, the employment status of the head of the household, the gender of the head of the household, the years of education of the head of the household and the household size. To compute the HCR and PG we used as poverty line the 60% of the national median equivalised household income, 9310.74 Euros. Table 1 shows the number of sampled households and of population households in each Province, and the poverty estimates for the Tuscany Region.

Table 1 - Number of sampled households (n), number of population households (N), estimated percentage Head Count Ratio (HCR) and percentage Poverty Gap (PG) with corresponding Root Mean Squared Error (RMSE) referring to the provinces of Tuscany

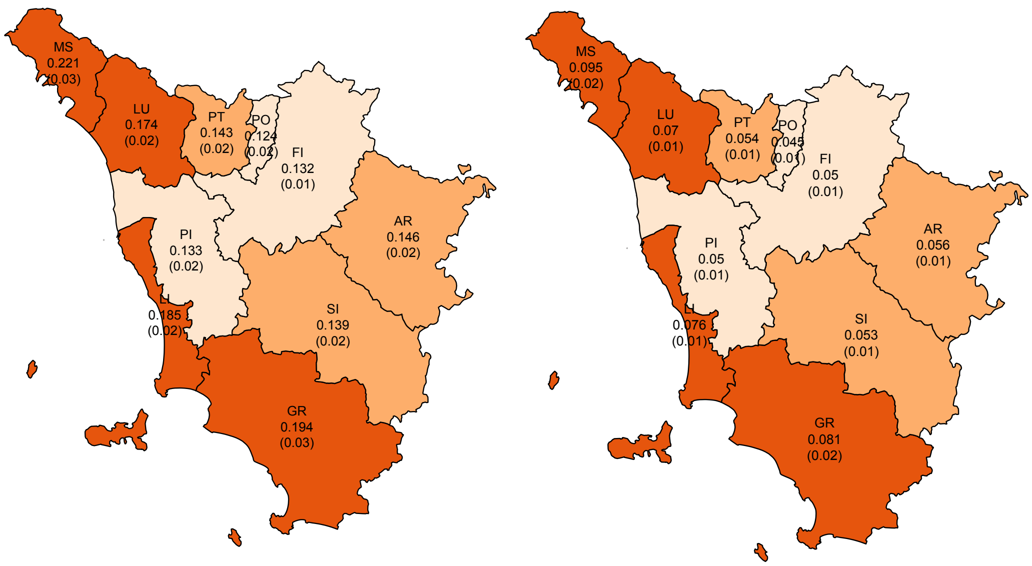
Province	n	N	HCR	HCR RMSE	PG	PG RMSE
Massa-Carrara	105	80810	22.1	2.73	9.52	1.56
Lucca	150	146117	17.39	1.92	7	1.05
Pistoia	136	104466	14.25	1.65	5.43	0.9
Firenze	415	376255	13.18	1.08	4.98	0.6
Livorno	105	133729	18.48	2.1	7.62	1.21
Pisa	149	150259	13.27	1.51	5.05	0.82
Arezzo	143	123880	14.63	1.59	5.6	0.86
Siena	104	101399	13.87	1.74	5.27	0.93
Grosseto	65	87720	19.43	2.55	8.12	1.52
Prato	123	83617	12.37	1.69	4.53	0.87

As we can see from Table 1 and Figure1, the higher estimated HCR and PG are observed for the province of Massa-Carrara: this area is thus the most critical both in terms of incidence and of intensity of poverty. At the opposite the lowest estimates are those referring to the province of Prato.

Conclusions and extensions

The areas we considered here are the Provinces of one Central Italian Region, Tuscany. In future developments of our analyses we will extend the application to other two Regions, Lombardia, in Northern Italy, and Campania, in Southern Italy. The choice of these three regions, out of the 20 existing regions in Italy, is motived by the geographical differences characterizing the Italian territory. In particular, the aim is to investigated the so-called north-south divide characterizing the Italian territory, since each of the three regions can be considered as representative of the corresponding geographical area of Italy (Northern, Central and Southern/Insular Italy). Estimation of the quartiles of the equivalised household income in the same Provinces will be performed as well, to get a more complete picture of poverty in the areas of interest.

Figure 1 - Estimated HCR (left) and PG (right) (Root Mean Squared Errors) for Toscana Provinces. M-quantile model. Poverty line=9310.74 Euros.



REFERENCES (RÉFERENCES)

Chambers R., Tzavidis N. (2006). M-quantile models for small area estimation, Biometrika, 93, 255-268.

Chambers, R., Chandra, H. and Tzavidis, N. (2007). On Robust Mean Squared Error Estimation for Linear Predictors for Domains. CCSR Working paper 2007-10. Cathie Marsh Centre for Census and Survey Research, University of Manchester.

Elbers C., Lanjouw J.O., Lanjouw P. (2003). Micro-level estimation of poverty and inequality. Econometrica, 71, 355-364.

Foster J., Greer J., Thorbecke E. (1984). A class of decomposable poverty measures. Econometrica, 52, 761-766.

Giusti C., Pratesi M., Tzavidis N., Salvati N. (2011) Recent advances in small area methodologies for poverty indicators. Proceedings of the NTTS Conference 2011.

Molina I., Rao J.N.K. (2010). Small Area Estimation of Poverty Indicators. The Canadian Journal of Statistics, 37, 117.

Rao, J.N.K. (2003). Small Area Estimation. New York: Wiley.

Tzavidis N., Salvati N., Pratesi M., Chambers R. (2008). M-quantile Models with Application to Poverty Mapping. Statistical Methods & Applications, 17, 393-411.

Tzavidis N., Marchetti S., Chambers R. (2010). Robust estimation of small area means and quantiles. Australian and New Zeland Journal of Statistics, 52(2), 167-186.