Survey Research Methods (2012)
Vol.6, No.3, pp. 155–163
ISSN 1864-3361
http://www.surveymethods.org

Robust Small Area Estimation and Oversampling in the Estimation of Poverty Indicators

Caterina Giusti, Stefano Marchetti, Monica Pratesi and Nicola Salvati
University of Pisa

There has been rising interest in research on poverty mapping over the last decade, with the European Union proposing a core of statistical indicators on poverty commonly known as Laeken Indicators. They include the incidence and the intensity of poverty for a set of domains (e.g. young people, unemployed people). The EU-SILC (European Union – Statistics on Income and Living Conditions) survey represents the most important source of information to estimate these poverty indicators at national or regional level (NUTS 1-2 level). However, local policy makers also require statistics on poverty and living conditions at lower geographical/domain levels, but estimating poverty indicators directly from EU-SILC for these domains often leads to inaccurate estimates. To overcome this problem there are two main strategies: i. increasing the sample size of EU-SILC so that direct estimates become reliable and ii. resort to small area estimation techniques. In this paper we compare these two alternatives: with the availability of an oversampling of the EU-SILC survey for the province of Pisa, obtained as a side result of the SAMPLE project (Small Area Methods for Poverty and Living Conditions, http://www.sample-project.eu/), we can compute reliable direct estimates that can be compared to small area estimates computed under the M-quantile approach. Results show that the M-quantile small area estimates are comparable in terms of efficiency and precision to direct estimates using oversample data. Moreover, considering the oversample estimates as a benchmark, we show how direct estimates computed without the oversample have larger errors as well as larger estimated mean squared errors than corresponding M-quantile estimates.
Keywords: poverty mapping, oversample, M-quantile models

1 Introduction

Available data to measure poverty and living conditions in Italy come mainly from sample surveys, such as the Survey on Income and Living Conditions (EU-SILC). However, these data can be used to produce direct accurate estimates only at the national or regional level (NUTS 1-2 levels). To obtain estimates referring to smaller unplanned domains, such as provinces and municipalities (LAU 1-2 levels), small area methodologies can be used.

Until very recently the practice of poverty mapping has been dominated by the World Bank method proposed by Elbers et al. (2003). More recently researchers in small area estimation have applied alternative small area estimation methods to poverty mapping. Two such recent methods are the M-quantile approach (Chambers and Tzavidis 2006; Tzavidis et al. 2010; Marchetti et al. 2012) and the Empirical Best Prediction (EBP) approach proposed by Molina and Rao (2010). An alternative to small area methods would be to take an oversample of units (that is, a sample with increased sample size selected using the same design as the original sample design) in the area/s of interest to obtain direct estimates with a reliable level of precision. The drawback is, of course, the high cost of the oversampling procedure, which makes it unfeasible as a standard practice.

The aim of this paper is to employ small area models to estimate some poverty indicators for unplanned domains, using data from the EU-SILC survey 2008 and from the Population Census 2001. The presence of outlier observations is a common feature in income data. With outliers we mean observations that are distinctly separated from the rest of the data. It has been well established that outliers can severely affect the parameter estimates of statistical models, for example random effects models, which can in turn affect the small area estimates produced using these models. For these reasons, we choose to use a robust approach based on the M-quantile estimator to obtain small area estimates. Our small domains are identified by the gender of the head of the household (male or female) and the 10 provinces of Tuscany (LAU 1 level), obtaining 20 domains. We estimate the head count ratio, the poverty gap and the per-capita equivalised mean income using direct and small area estimators for all the 20 domains; for the province of Pisa we compute also direct estimates using the oversample. Thus, our aim in the present paper is also to evaluate the performance of the small area poverty estimates taking the opportunity, for the first time, to use data referring to an EU-SILC 2008 oversampling of households for the province of Pisa, obtained as a side result of the SAMPLE project. This unique opportunity makes it possible for the Province of Pisa to compare the estimates obtained with small area models with the direct estimates using

Contact information: Caterina Giusti, Department of Economics and Management, University of Pisa, e-mail: caterina.giusti@ec.unipi.it

the enlarged sample, giving interesting guidelines for future analysis based on the standard EU-SILC sample. This is particularly relevant in the context of the estimation of monetary poverty indicators, where the data are characterized by the presence of outlier observations that can affect the estimates obtained with traditional small area models.

2 Methods for small area estimation of poverty indicators

Among poverty indicators the so called Laeken indicators are very often used to target poverty and inequalities. They are a core set of statistical indicators on poverty and social exclusion agreed by the European Council in December 2001, in the Brussels suburb of Laeken, Belgium. They include measures of the incidence of poverty, such as the Head Count Ratio (also known as at-risk-of-poverty-rate: HCR) and the intensity of poverty, such as the Poverty Gap (PG). These two poverty indicators are part of the generalized measures of poverty introduced by Foster et al. (1984) (FGT poverty measures hereafter).

Denoting by y a measure of income for individual/household j, t the poverty line, N_d the number of individuals/households living in area d ($d = 1, \ldots, D$), $I(u \leq k)$ the indicator function (equal to 1 when $u \leq k$ and 0 otherwise) and α a "sensitivity" parameter, the Foster et al. (1984) (FGT) poverty measures for a small area d are defined as:

$$F_{\alpha,d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \Big(\frac{t - y_{jd}}{t}\Big)^{\alpha} \mathrm{I}(y_{jd} \leq t). \qquad (1)$$

The poverty line t is a level of income that defines the state of poverty (units with income below t are considered poor). When $\alpha = 0$, $F_{\alpha,d}$ is the HCR whereas when $\alpha = 1$, $F_{\alpha,d}$ is the PG.

The HCR indicator is a widely used measure of poverty because of its ease of construction and interpretation, since it counts the number of individuals with income below the poverty line. At the same time this indicator also assumes that all poor individuals are in the same situation. For example, the easiest way of reducing the headcount index is by targeting benefits to people just below the poverty line because they are the ones who are cheapest to move across the line. Hence, policies based on the headcount index might be sub-optimal. For this reason we also obtain estimates of the PG indicator. The PG can be interpreted as the average shortfall of poor people. It shows how much would have to be transferred on average to the poors to bring their expenditure up to the poverty line.

In this paper we calculate the HCR, the PG and the mean of the per capita equivalised disposable income in each small area referring to individuals as statistical units. This means that the household equivalised disposable income is calculated as the household total net income divided by equivalised household size according to Eurostat (2007), which gives a weight of 1.0 to the first adult, 0.5 to other persons aged 14 or over who are living in the household and 0.3 to each child aged less than 14. Then, the same household equivalised disposable income is assigned to all members of the same household.

The straightforward approach to calculate FGT poverty indicators referring to the areas of interest is to compute direct estimates, using only the data from the sampled households. Let w_{jd} be the sampling weight (inverse of the probability of inclusion) of household j belonging to area d. Let s_d be the set of n_d sampled observations from area d and c_{jd} be the number of households members of household j in area d.

We compute direct estimators of the FGT poverty indicators adapting the Hájek direct estimator (see Särndal et al. 1992) in the following way

$$F^{dir}_{\alpha,d} = \frac{1}{\sum_{j \in s_d} w_{jd} c_{jd}} \sum_{j \in s_d} w_{jd} \left(\frac{t - y_{jd}}{t}\right)^{\alpha} I(y_{jd} < t) c_{jd}, \qquad d = 1, \ldots, D, \qquad (2)$$

where $\sum_{i \in s_d} w_{jd} c_{jd} = \hat{N}_d$ denotes the estimated population size of small area d, that is the numbers of individuals living in that area. In the same way, the mean of the per-capita equivalised income in each small area can be computed as

$$m^{dir}_d = \frac{1}{\sum_{j \in s_d} w_{jd} c_{jd}} \sum_{j \in s_d} w_{jd} y_{jd} c_{jd}, \qquad d = 1, \ldots, D. \qquad (3)$$

Associated estimates of standard error for estimators (2) and (3) can be computed using Taylor series (linearization) methods for estimators based on complex sample designs (Woodruff 1971; Fuller 1975).

When the sample size in the areas of interest is limited, the standard errors of the direct estimates are too large to be acceptable. For example, direct estimates computed using Italian EU-SILC data have large errors at provincial level or they may not even be computable at municipality level, since many municipalities are not included in the survey sample. Moreover, when a detailed geographical level is combined with a characteristic of interest – for example, estimation of the mean of the household income at provincial level distinguishing for the gender or for the age class of the head of the household – the problem of small sample sizes becomes even more severe.

In these cases small area estimation techniques can be employed. The idea of small area methods is to use statistical models to link the survey variable of interest with covariate information that is also known for out of sample units.

The population mean of the per capita equivalised income of each small area can be written as

$$m_d = N_d^{-1} \Big(\sum_{j \in s_d} y_{jd} c_{jd} + \sum_{j \in r_d} y_{jd} c_{jd} \Big), \qquad (4)$$

where r_d denotes the non-sampled units in area d. Since the y values for the r_d non-sampled units are unknown, they need to be predicted under a given small area model.

In this paper we consider an approach to small area estimation that relaxes the parametric assumptions of traditional random effects for small area estimation (Rao 2003), by employing M-quantile models (Chambers and Tzavidis 2006;

Tzavidis et al. 2010). The authors main motivation for using an M-quantile model was an alternative, less parametric approach for estimating area random effects. The use of M-estimation meant, however, that outlier robust estimation with the M-quantile small area model was automatically achieved.

Using this approach the small area mean estimator is obtained using the Chambers and Dunstan (1996) distribution function estimator (CD hereafter), as shown in Tzavidis et al. (2010). The MQ/CD estimator of the small area mean, using the same notation as in the decomposition (4), is

$$m_d^{MQ/CD} = N_d^{-1}\Big\{\sum_{j\in s_d} y_{jd}c_{jd} + \sum_{j\in r_d} \mathbf{x}_{jd}^T\hat{\beta}_\psi(\hat{\theta}_d)c_{jd} + \frac{N_d - n_d}{n_d}\sum_{j\in s_d}[y_{jd} - \mathbf{x}_{jd}^T\hat{\beta}_\psi(\hat{\theta}_d)]c_{jd}\Big\}, \quad (5)$$

which is based on the linear model $Q_q(\mathbf{x}_{jd},\psi) = \mathbf{x}_{jd}^T\beta_\psi(q_{jd})$ for the M-quantile of order q for the conditional distribution of y given a set of explanatory variables $\mathbf{x}$. Here ψ is the influence function and the estimate $\hat{\beta}_\psi(q_{jd})$ of $\beta_\psi(q_{jd})$ is obtained, for specified q and continuous ψ, via an iterative weighted least squares algorithm. A popular choice for the influence function is the Huber Proposal 2, $\psi(u) = uI(-\delta \le u \le \delta) + \delta\,\mathtt{sgn}(u)I(|u| > \delta)$, where δ is a cut-off constant. In estimator (5), $\hat{\theta}_d$ is an estimate of the average value of the M-quantile coefficients of the units in area d. See Chambers and Tzavidis (2006) and Tzavidis et al. (2010) for further details on the estimation of the M-quantile coefficients at unit level and for the computation of the small area M-quantile coefficients. Estimation of the MSE of estimator (5) can be achieved by using a linearization approach (Chambers et al. 2012) or a bootstrap approach recently proposed by Tzavidis et al. (2010).

The M-quantile approach to small area estimation can be used also to estimate poverty indicators. In this case we use the decomposition

$$F_{\alpha,d} = N_d^{-1}\Big(\sum_{j\in s_d} F_{\alpha,jd} + \sum_{j\in r_d} F_{\alpha,jd}\Big), \quad (6)$$

where $F_{\alpha,jd} = ((t - y_{jd})/t)^\alpha I(y_{jd} \le t)c_{jd}$.

To estimate the out of sample component in this expression we can use the same ideas described above for estimating the small area mean under the M-quantile small area model. Indeed, estimation of some poverty indicators, like the head count ratio, is a special case of cumulative distribution function estimation since we are interested in estimating the number of individuals/households below a threshold. As a result one approach to estimating $F_{\alpha,d}$ is to use a smearing-type estimator (as suggested in Chambers and Dunstan (1996) for estimating the distribution function). In this case, an estimator $\hat{F}_{\alpha,d}^{MQ}$ of $F_{\alpha,d}^{MQ}$ is

$$\hat{F}_{\alpha,d}^{MQ} = N_d^{-1}\left\{\sum_{j\in s_d}\Big(\frac{t - y_{jd}}{t}\Big)^\alpha I(y_{jd} \le t)c_{jd} + \sum_{k\in r_d} n_d^{-1}\sum_{j\in s_d}\Big(\frac{t - (\hat{y}_{kd} + (y_{jd} - \hat{y}_{jd}))}{t}\Big)^\alpha I(\hat{y}_{kd} + (y_{jd} - \hat{y}_{jd}) \le t)c_{jd}\right\} \quad (7)$$

which can be computed using a Monte Carlo procedure described in Marchetti et al. (2012) and Pratesi et al. (2010), similar in spirit to that proposed by Molina and Rao (2010) under the Empirical Best approach. In the previous expression $\hat{y}_{kd} = \mathbf{x}_{kd}^T\hat{\beta}_\psi(\hat{\theta}_d)$ is predicted by linear M-quantile models. The MSE of (7) can be estimated using the bootstrap techniques described in Marchetti et al. (2012).

As already said, the M-quantile method to poverty estimation does not impose strong distributional assumptions and because it is outlier robust, it models the raw income sample values. Previous studies have shown the potentialities of M-quantile estimators to compute poverty measures together with their corresponding variability at the small area level (Tzavidis et al. 2007; Marchetti et al. 2012; Pratesi et al. 2010, 2012). In the present paper we have for the first time the possibility to evaluate the performance of small area M-quantile estimators in an applied perspective, using the results of EU-SILC 2008 oversampling of households for the province of Pisa as benchmarking.

3 Estimation of Poverty Indicators at Provincial Level in Tuscany

In this section we present the results obtained applying small area estimators (5) and (7) to compute the HCR, PG and the mean of the per capita equivalised income in the 10 Provinces of the Tuscany region. To better compare the living conditions in these areas we estimate the indicators considering the gender of the head of the household: that is, we consider also the gender as a further unplanned domain. As a result, we have a total of 20 domains of interest.

The outcome variable, used to compute the indicators of interest, is the household equivalised income, which is available for each sampled household from the EU-SILC survey 2008. The explanatory variables are: the marital status of the head of the family (four levels), the working position of the head of the household (working/not working), the education of the head of the household (in years), the mean house surface at area level (in square meters) and the number of households members. The same variables are available for EU-SILC out of sample households from the Population Census 2001. Although the 2008 EU-SILC data were collected six years after the census (2008 EU-SILC data refers to 2007), the 2001–2007 period was one of relatively slow growth and low inflation in Italy, so it is reasonable to assume that there was relatively little change. It is important to underline that EU-SILC and Census data are confidential. These data were

Table 1: Direct estimates of the Head Count Ratio (HCR), Poverty Gap (PG), mean of the per-capita equivalised income (Mean) with corresponding estimated root mean squared errors (in brackets) and number of sampled households (h) by province and gender of the head of the household (HH)

Province	HH gender	h	HCR %	PG %	Mean
Massa Carrara (MS)	Female	34	19.40 (9.31)	4.62 (2.27)	15682.49 (1211.13)
	Male	71	19.82 (5.37)	3.41 (1.20)	15970.05 (881.23)
Lucca (LU)	Female	38	24.24 (6.99)	4.71 (1.72)	15507.76 (1499.11)
	Male	112	8.42 (2.71)	3.74 (1.76)	18882.23 (989.73)
Pistoia (PT)	Female	51	26.27 (9.13)	13.04 (7.15)	16314.72 (2088.05)
	Male	85	7.99 (3.79)	0.88 (0.50)	21273.49 (1456.37)
Firenze (FI)	Female	140	17.31 (3.23)	6.66 (1.55)	17039.98 (821.43)
	Male	275	3.13 (1.00)	0.99 (0.35)	22614.31 (931.65)
Livorno (LI)	Female	31	18.77 (8.07)	4.08 (1.66)	15247.12 (1805.08)
	Male	74	5.36 (2.77)	1.59 (1.09)	19929.71 (1149.26)
Pisa (PI)	Female	44	9.88 (4.28)	4.48 (2.56)	18674.62 (1542.40)
	Male	105	6.62 (2.24)	2.25 (0.91)	21138.21 (1047.76)
Arezzo (AR)	Female	34	21.29 (8.01)	6.02 (3.13)	15324.11 (1204.20)
	Male	109	2.44 (1.14)	0.41 (0.21)	22258.56 (1429.38)
Siena (SI)	Female	29	20.34 (9.93)	6.61 (4.66)	15727.38 (1523.19)
	Male	75	8.83 (3.74)	1.73 (1.34)	22500.07 (1670.64)
Grosseto (GR)	Female	30	12.87 (5.77)	4.32 (2.38)	17860.06 (2172.95)
	Male	35	11.24 (5.43)	3.13 (1.76)	20984.72 (2000.56)
Prato (PO)	Female	37	15.04 (10.04)	1.14 (0.59)	20331.81 (2182.16)
	Male	86	3.57 (1.80)	1.01 (0.58)	20328.12 (1335.89)

provided by ISTAT, the Italian National Institute of Statistics, to the researchers of the SAMPLE project and were analyzed by respecting all confidentiality restrictions.

Before presenting the results of the application, Figures 1 and 2 show some preliminary data and model diagnostics. Figure 1 displays the box-plots of the household equivalised income in each of the 20 domains. These clearly indicate the presence of outlier income observations in some of the domains. Moreover, we can see the gap between the household income distributions when the head of the household is a female compared with when this person is a male: the estimates of the income quantiles are almost always lower in the first case.

We also show some preliminary model diagnostics. Figure 2 shows normal probability plots of level one and level two residuals obtained by fitting a two-level random effects model to the EU-SILC data, considering the households as level one units and the 20 domains as level two units. The two-level random effects model was fitted to data in the raw scale and to log-transformed data, after shifting them to take into account the presence of negative income values. The graphics indicate that the normality hypotheses for the distribution of the residuals of linear mixed models are not met for these data. Computing the Shapiro–Wilk test statistic we reject the hypothesis of normal distribution for level one residuals both in the raw and log scale, while we accept it for the level two residuals. Finally, the Cook's distances indicate that there are outliers in the data even when the log-transformed income is used as the outcome variable.

These preliminary diagnostics confirm that it is appropriate to use M-quantile models in this application. Alternative small area approaches such as the EBP approach of Molina and Rao (2010) and the Robust EBLUP recently proposed by Sinha and Rao (2009) are not considered here. The comparison of these different approaches to that of M-quantile models in the context of estimation of poverty indicators is available in Pratesi et al. (to appear) and in Giusti et al. (in press).

Table 1 reports the results obtained by applying the direct estimators to compute the mean of the per-capita equivalised income, the HCR and the PG in the domains. The poverty line used for the HCR and PG indicators is set to 9310.74 Euros, which corresponds to the 60th percentile of the Italian median equivalised household income, computed using EU-SILC 2008 income values based on the Italian sample of households with weights equal to the cross-sectional EU-SILC household weights. The Table also reports the corresponding estimated Root Mean Squared Errors (RMSEs) and the number of sampled households in each domain. As we can see, the estimates suggest a worst situation of poverty, in terms of incidence, intensity and mean income, in the domains where the head of the household is a female; however, there are also some provinces where the estimates are similar for the two genders, particularly when one takes into account the variability of the estimates.

The results (point and RMSE estimation) from the application of the MQ/CD estimators are presented in Table 2. The estimates confirm the higher prevalence of poverty in the domains characterized by a female head of the household, and this is true in every province. Indeed, the estimated HCR values are always over 20% for the female-headed households, with the only exception being the province of Prato. The corresponding estimated PGs are over or near 10%, indicating that the estimated higher incidence of poverty for the households with a female head are also characterized by a higher intensity of poverty. Finally, the results for per-capita

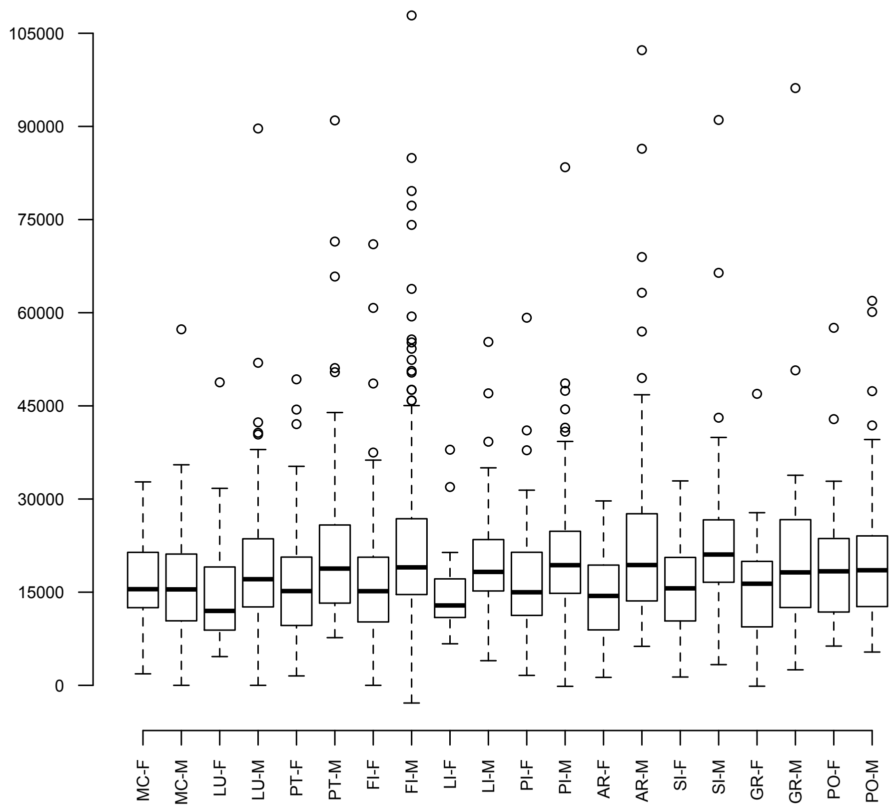

Figure 1. Boxplots of the equivalised household income for Tuscany provinces, by gender of the head of the household

Table 2: MQ/CD estimates of the Head Count Ratio (HCR), Poverty Gap (PG), mean of the per-capita equivalised income (Mean) with corresponding estimated root mean squared errors (in brackets) and number of sampled households (h) by province and gender of the head of the household (HH)

Province	HH gender	h	HCR %	PG %	Mean
Massa Carrara (MS)	Female	34	26.32 (3.91)	11.68 (2.77)	14687.02 (1191.84)
	Male	71	17.49 (3.89)	6.51 (1.87)	14966.98 (846.03)
Lucca (LU)	Female	38	24.42 (3.51)	10.60 (2.33)	15586.29 (1364.45)
	Male	112	12.72 (2.13)	4.39 (0.96)	17742.87 (891.33)
Pistoia (PT)	Female	51	23.72 (3.13)	10.16 (2.01)	15778.08 (1875.19)
	Male	85	9.38 (1.87)	3.03 (0.85)	21364.05 (1522.22)
Firenze (FI)	Female	140	21.34 (1.99)	8.98 (1.25)	15643.43 (773.34)
	Male	275	8.49 (1.27)	2.70 (0.58)	21583.49 (715.13)
Livorno (LI)	Female	31	29.55 (4.10)	13.68 (2.93)	14758.59 (1748.44)
	Male	74	11.90 (2.28)	4.06 (1.08)	18814.98 (1001.09)
Pisa (PI)	Female	44	20.72 (3.13)	8.64 (2.00)	17410.97 (1234.53)
	Male	105	9.02 (1.63)	2.91(0.74)	19962.85 (1051.81)
Arezzo (AR)	Female	34	25.78 (4.11)	11.30 (2.54)	14180.67 (1053.79)
	Male	109	10.08 (1.68)	3.29 (0.75)	21313.21 (1481.55)
Siena (SI)	Female	29	23.91 (3.93)	10.32 (2.47)	13784.54 (1422.86)
	Male	75	9.09 (2.00)	2.92 (0.90)	20312.63 (1206.72)
Grosseto (GR)	Female	30	31.62 (4.19)	15.04 (3.06)	14102.24 (1845.13)
	Male	35	12.23 (3.30)	4.18 (1.60)	19509.73 (1798.83)
Prato (PO)	Female	37	16.03 (3.54)	6.22 (2.07)	19970.95 (2239.82)
	Male	86	10.20 (1.97)	3.32 (0.86)	18876.35 (971.52)

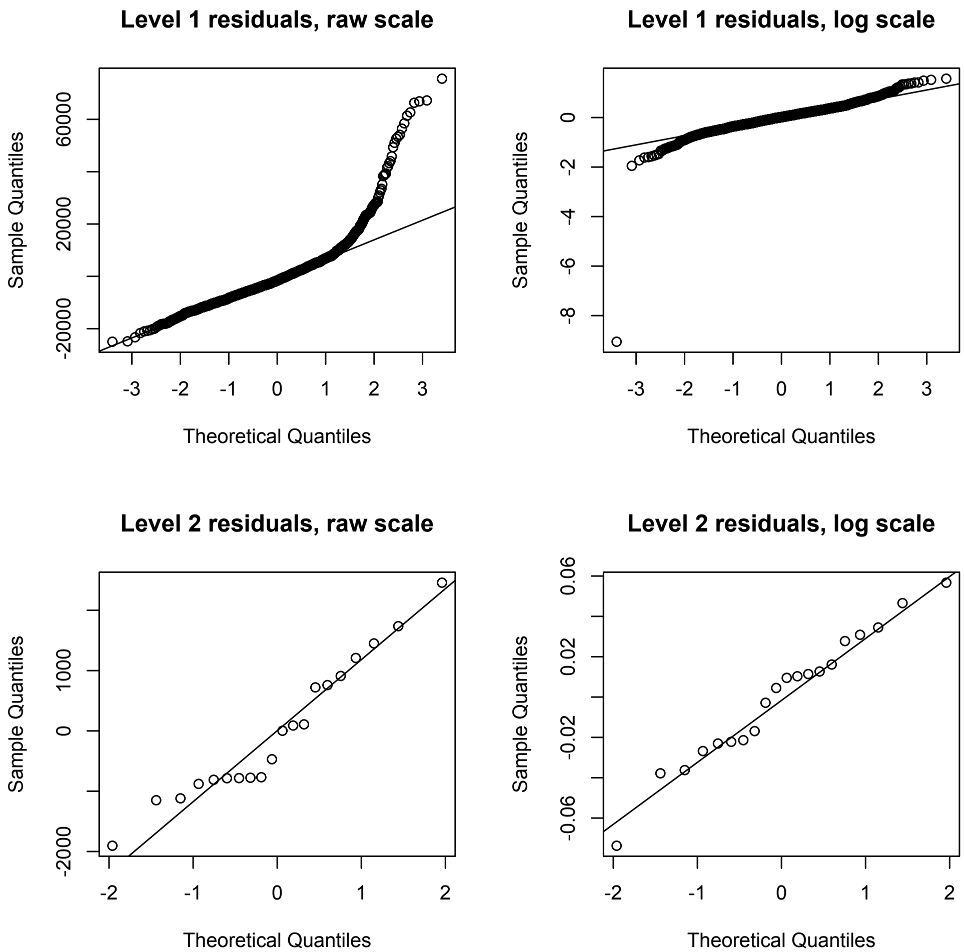

Figure 2. Normal probability plots of level one and level two residuals derived by fitting a two level linear mixed model to EU-SILC data in raw (left) and log (right) scale

Table 3: Direct estimates (without and with oversampling) and MQ/CD estimates of the Head Count Ratio (HCR), Poverty Gap (PG), mean of the per-capita equivalised income (Mean) with corresponding estimated root mean squared errors (in brackets) and number of sampled households (h) in the province of Pisa, by gender of the head of the household (HH)

Estimates	HH gender	h	HCR %	PG %	Mean
Direct estimate	Female	44	9.88 (4.28)	4.48 (2.56)	18674.62 (1542.40)
	Male	105	6.62 (2.24)	2.25 (0.91)	21138.21 (1047.76)
MQ/CD estimates	Female	44	20.72 (3.13)	8.64 (2.00)	17410.97 (1234.53)
	Male	105	9.02 (1.63)	2.91 (0.74)	19962.85 (1051.81)
Direct estimates	Female	193	23.57 (4.92)	6.64 (2.77)	15773.84 (750.19)
(with oversampling)	Male	482	8.21 (1.61)	2.40 (0.60)	20167.39 (944.29)

mean income suggest similar conclusions on the poverty situation in Tuscany, though in this case the estimated (analytic) RMSEs suggest that one should be cautious in making comparisons. Note that the bootstrapped RMSEs of the HCR and PG estimates are often lower than the corresponding direct ones (see Table 1).

A more effective representation of the MQ/CD computed poverty estimates is in Figures 3, 4 and 5. In each figure, the map on the left refers to the domain where the head of the household is a female, while that on the right refers to the families where the head is a male. In each map the provinces are grouped into four different classes of colors, determined by the quartiles of the represented indicator, with a darker color corresponds to a better figure, namely a higher mean household income and a lower HCR and PG. The representation of the results by means of poverty mapping suggest

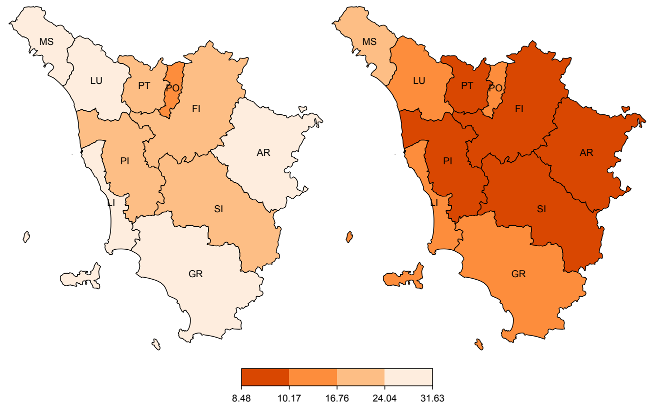

Figure 3. MQ/CD estimated HCR of the per-capita equivalised income for Tuscany provinces, by gender of the head of the household: female (left) and male (right)

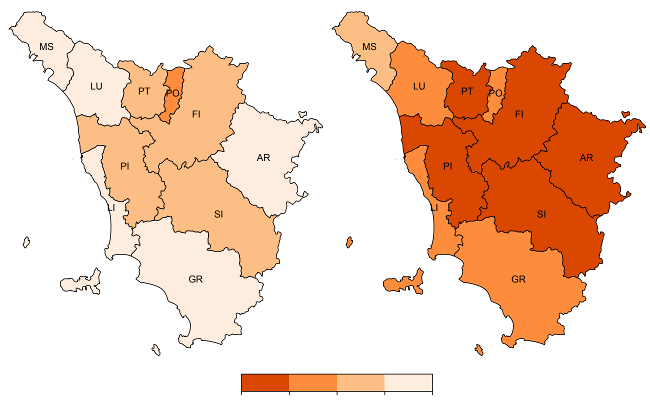

Figure 4. MQ/CD estimated PG of the per-capita equivalised income for Tuscany provinces, by gender of the head of the household: female (left) and male (right)

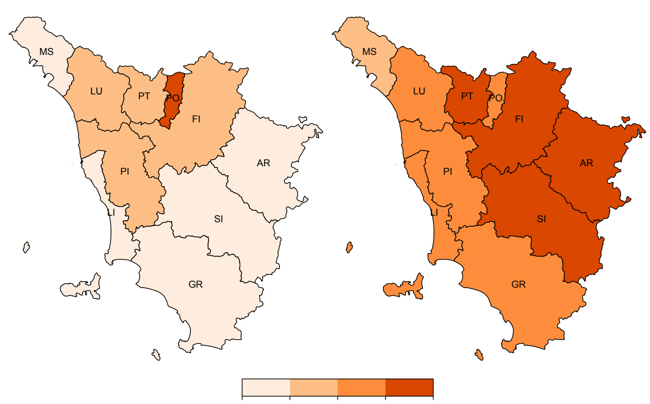

Figure 5. MQ/CD estimated mean of the per-capita equivalised income for Tuscany provinces, by gender of the head of the household: female (left) and male (right

again a big gap in the living conditions depending on the gender of the head of the household. From the maps it is also possible to see that the level of poverty is worse in the provinces of Massa Carrara and Lucca, in the North of the region, and the provinces of Grosseto and Livorno, in the southern/coastal areas.

Finally, Table 3 reports for the Province of Pisa the estimates of the three variables of interest obtained with the direct and MQ/CD estimators (results already presented in the previous Tables) together with the direct estimates (benchmarking values) computed using the EU-SILC 2008 oversample in this Province. With the oversampling the original sample of 149 households has been enlarged to 675 households.

The MQ/CD estimates are closer to the direct estimates obtained using the oversampling than the direct estimates based on the original sample, suggesting that direct estimates (without oversampling) are affected by large errors. In particular, this is evident for the Head Count Ratio and for the mean of the per-capita equivalised income when the head of the household is a female. This is not a surprise because the sample size in the Province of Pisa for female-headed households is 44. In contrast, the errors associated with the MQ/CD estimator appear smaller. For example, if we consider the direct estimator values (with oversampling) of HCR as the true values, then the relative error of the MQ/CD estimates is -12% for females and 9.9% for males, whereas the relative bias of the direct estimates (without oversampling) is -58.1% for females and -19.4% for males. Moreover the estimated RMSEs of the MQ/CD estimators tend to the estimated RMSEs of the benchmarking values, whereas the direct estimates (without oversampling) are more variable.

4 Conclusions

In this paper we illustrate the potential of employing M-quantile models for estimating the incidence of poverty and inequality at small area levels and for estimating small area household per-capita equivalised income. In doing so, we use the methodology of Chambers and Tzavidis (2006) and Tzavidis et al. (2007). Unlike traditional random effects models, M-quantile models do not depend on strong distributional assumption and automatically provide outlier robust inference. In this paper we evaluate for the first time the performance M-quantile models for estimating small area poverty indicators taking the opportunity to use data from a EU-SILC 2008 oversampling of households for the province of Pisa, obtained as a side result of the SAMPLE project. Our results provide evidence for the good performance of the estimates based on M-quantile models compared with the direct estimates in terms of bias and variability. Also the analytical and bootstrap MSE estimators track the MSE estimates of the direct estimator obtained by the oversampling especially for HCR and PG.

Further research is necessary in order to understand the behavior of M-quantile approach for estimating income distribution functions and the other small area poverty indicators. Indeed, the indicators computed in the present paper can be considered as a starting point for more in depth studies of poverty and living conditions. In fact, poverty analyses can been done using also non-monetary indicators in order to give a more complete picture of poverty and deprivation (Cheli and Lemmi 1995). As poverty is a question of graduation, the set of indicators is generally enlarged with other indicators of belonging to vulnerable groups from which it can be likely to move towards the status of poverty (for details, see SAMPLE project results at www.sample-project.eu).

Acknowledgements

The authors would like to acknowledge the valuable comments and suggestions of the Associate Editor, the Editor and three referees. These led to a considerable improvement in the paper. This work is financially supported by the European Project SAMPLE Small Area Methods for Poverty and Living Condition Estimates, funded by the European Commissions 7th Framework Programme.

References

Chambers, R. L., Chandra, H., & Tzavidis, N. (2012). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methodology*. (in press)

Chambers, R. L., & Dunstan, R. (1996). Estimating distribution functions from survey data. *Biometrika*, *73*, 597-604.

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

Cheli, B., & Lemmi, A. (1995). A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. *Economic Notes*, *24*, 115-134.

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

Eurostat. (2007). *Description of SILC user database variables. Version 2007.1 from 01-03-2009.* Eurostat.

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

Fuller, W. A. (1975). Regression Analysis for Sample Survey. *Sankhya*, *37*(3), 117-132. (Series C)

Giusti, C., Tzavidis, N., Pratesi, M., & Salvati, N. (in press). *Resistance to outliers of M-quantile and robust random effects small area models.* S3RI Methodology Working Papers M11/04, Southampton Statistical Sciences Research Institute, University of Southampton, UK.

Marchetti, S., Tzavidis, N., & Pratesi, M. (2012). Non-parametric bootstrap mean squared error Estimation for M-quantile Estimators of Small Area Averages, Quantiles and Poverty Indicators. *Computional Statistics & Data Analysis*, *56*(10), 2889-2902.

Molina, I., & Rao, J. N. K. (2010). Small Area Estimation of Poverty Indicators. *Canadian Journal of Statistics*, *38*, 369-385.

Pratesi, M., Giusti, C., & Marchetti, S. (to appear). Small Area Estimation of Poverty Indicators. In C. Davino & L. Fabbris (Eds.), *Survey data collection and integration.* Heidelberg: Springer.

Pratesi, M., Giusti, C., Marchetti, S., Salvati, N., Tzavidis, N., Molina, I., et al. (2010). *SAMPLE DELIVERABLES 12 and 16. Final small area estimation developments and simulations results.* Available at http://www.sampleproject.eu/it/the-project/deliverables-docs.html.

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

Särndal, C. E., Swenson, B., & Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Heidelberg: Springer-Verlag.

Sinha, S. K., & Rao, J. N. K. (2009). Robust Small Area Estimation. *Canadian Journal of Statistics*, *37*, 381-399.

Tzavidis, N., Marchetti, S., & Chambers, R. (2010). Robust prediction of small area means and distributions. *Australian and New Zealand Journal of Statistics*, *52*, 167-186.

Tzavidis, N., Salvati, N., Pratesi, M., & Chambers, R. (2007). M-quantile models with application to poverty mapping. *Statistical Methods and Applications*, *17*, 393-411.

Woodruff, R. S. (1971). A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association*, *66*(334), 411-414.