Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 66(1) 2012 105-120

Fast EB Method for Estimating Complex Poverty Indicators in Large Populations

Caterina Ferretti[1] and Isabel Molina[2*]

[1]*University of Florence*

[2]*Department of Statistics, Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe (Madrid), Spain*

Received 01 May 2011; Revised 08 September 2011; Accepted 08 September 2011

SUMMARY

This paper studies small area estimation of computationally complex poverty indicators; more concretely, we study fuzzy monetary and fuzzy supplementary poverty indicators. These two indicators do not need to set a poverty line because they are based on the degree of poverty of each individual relative to the population to which it belongs. Moreover, the latter takes into account the non-monetary and multidimensional nature of poverty. For this, a faster version of the empirical best/ bayes (EB) method of Molina and Rao (2010) is proposed. This new method allows feasible estimation of computationally complex indicators in large populations, and can still reduce considerably the computation time when the original EB method is feasible. In simulations, the proposed fast EB method is compared with the original EB method when estimating the mentioned indicators along with the poverty incidence in small areas. Results show negligible loss of efficiency of the fast EB method as compared to the original one, while allowing estimation of complex indicators that require sorting all population elements. The method is applied to the estimation of poverty indicators in the region of Tuscany, both at province and municipality levels, using data from the Italian Survey on Income and Living Conditions.

Keywords : Empirical best estimator, Fuzzy poverty measures, Small area estimation.

1. INTRODUCTION

Traditional poverty measures are based on a simple dichotomization of each individual of the population as poor or non poor according to a selected welfare variable and a given poverty line. This poverty line typically represents a percentage (say 50%, 60% or 70%) of the median of the equivalised income distribution, see e.g. Foster *et al.* (1984). On the one hand, this approach is unidimensional, that is, it refers only to one proxy of poverty, namely low income or consumption expenditure. On the other, the choice of the poverty line can be regarded as arbitrary. Here we consider small area estimation of traditional poverty measures such as the poverty incidence (PI) or head count ratio, but also other two more appropriate poverty measures. The first one, called fuzzy monetary indicator (FMI), does not need to fix any poverty threshold because it is based on measuring the degree of poverty of each individual in comparison with the other ones in the same population (by ranking all individuals according to the welfare proxy). The second, called fuzzy supplementary indicator (FSI), is based on a ranking obtained after applying a multidimensional approach that takes into account a variety of non-monetary aspects of deprivation.

Traditional model based small area methods can not be routinely applied to estimate these poverty

Corresponding author : Isabel Molina
E-mail address : isabel.molina@uc3m.es

measures, FMI and FSI, because they are non-linear functions of the values of a variable in the population units. Moreover, they are also computationally complex because they require sorting all population elements. Thus, direct application of the empirical best (EB) method of Molina and Rao (2010) can be unfeasible for a large population. Here we introduce a modification of the EB method, called fast EB method, which reduces drastically the computing time, making feasible the estimation of complex non-linear quantities under large populations, but at the same time losing little efficiency.

The original EB method is based on Monte Carlo generation of vectors corresponding to the non-sample part of the population. This requires to identify in the census or in the administrative register from which auxiliary variables come from, the units that were sampled in the survey data. This many times is not possible due to the use of different identification codes in the two data sets. Another advantage of the fast EB method is that it does not need to identify the sample units in the census.

In simulations we compare the results of different small area estimation methods, including the original EB and the fast EB method, for the estimation of the mentioned complex poverty measures using a unit level linear regression model. Finally, the proposed approach is applied to the estimation of poverty indicators in Tuscany provinces.

The organization of the paper is as follows. Section 2 introduces the poverty indicators considered in the paper. Section 3 describes the EB method and introduces the new faster version of this method. Section 4 reports the results of a model based simulation experiment designed to compare the fast EB estimators with the original EB counter parts, the design-based and the ELL estimators of Elbers *et al.,* (2003). Section 5 gives the results of a design-based simulation experiment. Section 6 describes an application to the estimation of the considered poverty indicators in provinces and municipalities of Tuscany. Section 7 gives some concluding remarks. Finally, detailed results of the application for Tuscany municipalities are provided in Appendix.

2. FUZZY MONETARY AND SUPPLEMENTARY INDICATORS

Let $U = \{1,..., N\}$ be a finite population of size N, where E_i is the value of a welfare variable (e.g. equivalised income) for individual i. Let us consider the finite population distribution function of $\{E_1,...,E_N\}$, defined as

$$F_E(x) = \frac{1}{N}\sum_{j=1}^{N} I\,\{E_j \le x\},\ x \in \mathbb{R}, \tag{1}$$

where $I\{E_j \le x\} = 1$ if $E_j \le x$ and 0 otherwise. Let z denote the poverty line for the given population. Note that setting $x = z$ in (1) we obtain the poverty incidence or proportion of individuals under the poverty line, that is, $PI = F_E(z)$.

Consider also the finite population Lorenz curve, given by

$$L_E(x) = \frac{\sum_{j=1}^{N} E_j I\{E_j \le x\}}{\sum_{j=1}^{N} E_j},\ x \in \mathbb{R}.$$

Following the Integrated Fuzzy and Relative (IFR) approach of Betti *et al.* (2006), the Fuzzy Monetary Index (FMI) for individual i is defined as

$$FM_i = \left\{\frac{N}{N-1}(1 - F_E(E_i))\right\}^{\alpha-1} \{1 - L_E(E_i)\}$$

$$= \left\{\frac{1}{N-1}\sum_{j=1}^{N} I\left\{E_j > E_i\right\}\right\}^{\alpha-1}$$

$$\times \left\{\frac{\sum_{j=1}^{N} E_j I\,\{E_j > E_i\}}{\sum_{j=1}^{N} E_j}\right\},\ i \in U.$$

Here, $1 - F_E(E_i)$ is the proportion of individuals that are less poor than individual i and α is a positive constant (see Remark 1 below). This index gives a degree of poverty for individual i and it was proposed by Cheli and Lemmi (1995) as a poverty indicator. Observe that $N(1 - F_E(E_i))/(N - 1)$ is equal to 1 when individual i is the poorest. Moreover, $1 - L_E(E_i)$ is the share of the total welfare of all individuals that are less poor than this individual, indicator that was proposed by Betti and Verma (1999). The average FMI for the population is given by

$$FM = \frac{1}{N}\sum_{i=1}^{N} FM_i \tag{2}$$

Observe that the FMI for individual i depends on the whole population of welfare values, $\{E_1, ..., E_N\}$.

Consider now a score variable S_i for i-th individual defined using the IFR approach, instead of a welfare variable E_i. These scores S_i are obtained by applying a multidimensional approach that takes into account a variety of non-monetary indicators of deprivation. More concretely, the score for i-th individual is an unweighted mean of deprivation scores corresponding to different dimensions, which were obtained as follows. First, from the large set of EU-SILC variables, a set of indicators was selected that were considered substantively meaningful and useful for the measurement of life-style deprivation. All these indicators were considered at household level, even if some of them had been taken from the individual data and then were aggregated at household level. Each item was re-scaled to make it range in the [0,1] interval. Exploratory and confirmatory factor analysis allowed to identify the different dimensions of deprivation. Following exploratory factor analysis, nine dimensions were selected. Some factors were then rearranged in the dimensions in order to create more meaningful groups. Finally, a confirmatory factor analysis was carried out to test the goodness of the hypothesized model. In summary, the following seven dimensions were finally selected: basic life-style, consumer durable goods, housing amenities, financial situation, environmental problems, work/education and health issues.

Thus, the Fuzzy Supplementary Index (FSI) for individual i is defined analogously to the FMI, but now in terms of the obtained scores $\{S_1, ..., S_N\}$, as

$$FS_i = \left\{\frac{N}{N-1}(1 - F_S(S_i))\right\}^{\alpha-1} \{1 - L_S(S_i)\}$$

$$= \left\{\frac{N}{N-1}\sum_{j=1}^{N} I\{S_j > S_i\}\right\}^{\alpha-1}$$

$$\times \left\{\frac{\sum_{j=1}^{N} S_j\, I\{S_j > S_i\}}{\sum_{j=1}^{N} S_j}\right\},\ i \in U.$$

Here, $F_S(x)$ and $L_S(x)$ are respectively the finite population distribution function and Lorenz curve of the score variables $\{S_1, ..., S_N\}$. Similarly, $1 - F_S(S_i)$ is the proportion of individuals who are less deprived than individual i and $1 - L_S(S_i)$ is the share of the total lack of deprivation score assigned to all individuals less deprived than individual i. The average FSI for the population is given by

$$\text{FS} = \frac{1}{N}\sum_{i=1}^{N} FS_i \tag{3}$$

Now consider that the population U is partitioned into D domains or areas $U_1, ..., U_D$ of sizes $N_1, ..., N_D$. Let E_{dj} be the welfare for individual j within domain d. The average fuzzy monetary index for domain d is

$$FM_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj},\ d = 1, ..., D, \tag{4}$$

where FM_{dj} is the FMI for j-th individual from d-th domain.

A random sample s of size $n \le N$ is drawn from the population. Let s_d be the subsample from domain d, $d = 1, ..., D$. If N_d is known, a design-based estimator of the average FMI for domain d, FM_d, is

$$\widehat{FM}_d^{DB} = \frac{1}{N_d}\sum_{j \in s_d} w_{dj}\, \widehat{FM}_{dj}^{DB},\ d = 1, ..., D, \tag{5}$$

where w_{dj} is the sampling weight for individual j within domain d and

$$\widehat{FM}_{dj}^{DB} = \left\{\frac{\sum_{h=1}^{D}\sum_{i \in s_h} w_{hi}\, I\{E_{hi} > E_{dj}\}}{\sum_{l=1}^{D}\sum_{i \in s_l} w_{hi}}\right\}^{\alpha-1}$$

$$\times \left\{\frac{\sum_{h=1}^{D}\sum_{i \in s_h} w_{hi}\, E_{hi}\ I\{E_{hi} > E_{dj}\}}{\sum_{l=1}^{D}\sum_{i \in s_h} w_{hi} E_{hi}}\right\}. \tag{6}$$

Note that $\widehat{FM}_{dj}^{DB}$ uses the sample data from the whole population and not only from domain d. Similarly, the average FSI for domain d is given by

$$FS_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FS_{dj},\ d = 1,\ldots, D. \quad (7)$$

Finally, a design-based estimator of FS_d would be

$$\widehat{FS}_d^{DB} = \frac{1}{N_d}\sum_{j\in s_d} w_{dj}\, \widehat{FS}_{dj}^{DB}\quad d = 1,\ldots,D. \quad (8)$$

where

$$\widehat{FS}_{dj}^{DB} = \left\{\frac{\sum_{h=1}^{D}\sum_{i\in s_h} w_{hi}\, I\{S_{hi} > S_{dj}\}}{\sum_{h=1}^{D}\sum_{i\in s_h} w_{hi}}\right\}^{\alpha-1} \times \left\{\frac{\sum_{h=1}^{D}\sum_{i\in s_h} w_{hi}\, S_{hi}\, I\{S_{hi} > S_{dj}\}}{\sum_{h=1}^{D}\sum_{i\in s_h} w_{hi} S_{hi}}\right\} \quad (9)$$

Remark 1 : To make FMI comparable with the poverty incidence, Cheli and Betti (1999) chose the parameter α as the value for which FMI is equal to the poverty incidence computed for the official poverty line (say 60% of the median welfare). Increasing the value of α implies giving more weight to the tail of the income distribution corresponding to the poorer people. The mean of $1 - L_E(E_i)$ is equal to 0.5 by definition and the mean of $1 - F_E(E_i)$ is greater than 0.5. Then, the mean of their product is greater than 0.5. In most European countries, the values of the poverty incidence range approximately between 0.08 and 0.30. This means that α is greater than 1 in this case; more concretely, it will be approximately between 9/10 and 3.

Remark 2 : $\widehat{FM}_d^{DB}$ and $\widehat{FS}_d^{DB}$ are design consistent for FM_d and FS_d respectively. They are not exactly design unbiased, but since individual indexes FM_{dj} and FS_{dj} are estimated using the whole sample, their bias is likely to be small.

Remark 3 : In case that the domain size N_d is not known, in (5) and (8) it should be replaced by the unbiased estimator $\hat{N}_d = \sum_{j\in s_d} w_{dj}$. In this case, design based estimators $\widehat{FM}_d^{DB}$ and $\widehat{FS}_d^{DB}$ will have non-negligible ratio bias for the domains with smaller sample sizes.

3. FAST EMPIRICAL BEST PREDICTION

This section describe the EB method of Molina and Rao (2010) to estimate the domain FMI and FSI, FM_d and FS_d respectively, and introduces a modification of this method to make it computationally feasible for large populations and/or computationally complex indicators.

As in Molina and Rao (2010), we consider that there exists a one-to-one transformation $Y_{dj} = T(E_{dj})$ of the welfare variable E_{dj}, which follows a Normal distribution. Concretely, the transformed variable Y_{dj} follows the nested error linear regression model of Battese *et al.* (1988), formulated as

$$\begin{gathered} Y_{dj} = \mathbf{x}_{dj}\beta + u_d + e_{dj}, \\ j = 1,\ldots, N_d,\ d = 1,\ldots, D, \\ u_d \sim \text{iid N}(0,\ \sigma_u^2),\ e_{dj} \sim \text{iid N}(0,\ \sigma_e^2) \end{gathered} \quad (10)$$

where $\mathbf{x}_{dj}$ is a row vector with the values of p explanatory variables, u_d is a random area-specific effect and e_{dj} are residual errors. Let $\mathbf{y}_d = (Y_{d1},\ldots, Y_{dN_D})'$ be vector of responses for domain d and $\mathbf{y} = (\mathbf{y}_1',\ldots, \mathbf{y}_D')'$ be the full population response vector. Then, observe that the individual FMIs can be expressed as

$$FM_{dj} = \left\{\frac{1}{N-1}\sum_{h=1}^{D}\sum_{i=1}^{N_h} I\,\{T^{-1}(Y_{hi}) > T^{-1}(Y_{dj})\}\right\}^{\alpha-1} \times \left\{\frac{\sum_{h=1}^{D}\sum_{i=1}^{N_h} T^{-1}(Y_{hi}) I\,\{T^{-1}(Y_{hi}) > T^{-1}(Y_{dj})\}}{\sum_{h=1}^{D}\sum_{i=1}^{N_h} T^{-1}(Y_{hi})}\right\}.$$

This means that the average FMI for domain d is a non-linear function of the population vector $\mathbf{y}$, that is,

$$FM_d = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj} = h_d(\mathbf{y}),\ d = 1,\ldots, D.$$

Let us split the population vector of model responses $\mathbf{y}$ into sample and non-sample parts, that is, $\mathbf{y} = (\mathbf{y}_s', \mathbf{y}_r')'$, where $\mathbf{y}_s$ is the subvector with sample elements and $\mathbf{y}_r$ contains non-sample elements. Then the BP of FM_d is given by

$$\widehat{FM}_d^B = E_{\mathbf{y}_r}(FM_d \mid \mathbf{y}_s) = E_{\mathbf{y}_r}\{h_d(\mathbf{y}) \mid \mathbf{y}_s\} \quad (11)$$

This expectation can be empirically approximated by Monte Carlo simulation. For this, first fit the nested-error model (10) to sample data $\mathbf{y}_s$, obtaining estimates $\hat{\beta}, \hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ of model parameters β, σ_u^2 and σ_e^2 respectively. Obtain also the EB predictor $\hat{u}_d$ of u_d, given by $E(u_d \mid \mathbf{y}_s)$ with unknown parameters replaced by estimated values. Then, using those estimates, generate a large number L of non-sample vectors $\mathbf{y}_r$ from the estimated conditional distribution $\mathbf{y}_r \mid \mathbf{y}_s$. Let $\mathbf{y}_r^{(1)}$ be the vector generated in l-th generation. We attach this vector to the sample vector $\mathbf{y}_s$, to obtain the full population response vector $\mathbf{y}^{(l)} = (\mathbf{y}_s', (\mathbf{y}_r^{(l)})')'$. Using the elements of $\mathbf{y}^{(l)}$, we calculate the domain parameters of interest $\widehat{FM}_d^{(l)} = h_d(\mathbf{y}^{(l)})$, $d = 1, ..., D$. Then, a Monte Carlo approximation to the EB predictor of FM_d is given by

$$\widehat{FM}_d^{EB} \approx \frac{1}{L}\sum_{l=1}^{L} FM_d^{(l)}, \; d = 1, ..., D. \quad (12)$$

Observe that, as pointed out by Molina and Rao (2010), for each population $l = 1, ..., L$, instead of generating a multivariate normal vector of size $N - n$, $\mathbf{y}_r^{(l)}$ we just need to generate univariate non-sample values $Y_{dj}^{(l)}$, for $j \in U_d - s_d$, from the model

$$Y_{dj}^{(l)} = \mathbf{x}_{dj}\hat{\beta} + \hat{u}_d + v_d^{(l)} + \varepsilon_{di}^{(l)},$$
$$v_d^{(l)} \sim \mathrm{N}(0, \hat{\sigma}_u^2(1-\hat{\gamma}_d)), \varepsilon_{dj}^{(l)} \sim \mathrm{N}(0, \hat{\sigma}_e^2) \quad (13)$$

where $\gamma_d = \sigma_u^2(\sigma_u^2 + \sigma_e^2/n_d)^{-1}$ and n_d is the sample size in domain d, for $d = 1, ..., D$. As already mentioned, for large populations and/or complex indicators such as the FMI and the FSI, the EB method can be unfeasible. Take into account that calculation of these indicators requires sorting all population elements, and this needs to be repeated for each $l = 1, ..., L$, which is too time consuming for large N and large L. Here we propose a faster version of the EB method that is based on replacing the true value of the domain average FMI in population l, $FM_d^{(l)}$, by the design-based estimator given in (5). Since the design-based estimator is obtained from a sample drawn from l-th population, this avoids the task of generation of the full population of responses (we need to generate only the responses for the sample elements) and the sorting of all the population elements. More concretely, for each Monte Carlo replication l, we take a sample $s(l) \subseteq U$ using the same sampling scheme and the same sample size allocation as in the original sample s. We take the values of the auxiliary variables corresponding to the units in $s(l)$, that is, we take $\mathbf{x}_{dj}, j \in s_d(l)$, where $s_d(l)$ is the subsample from d-th domain. Then we generate the corresponding responses $Y_{dj}^{(l)}$, $j \in s_d(l)$, for $d = 1, ..., D$, as in (13). Let us denote the vector containing those values as $\mathbf{y}_{s(l)}$. Using $\mathbf{y}_{s(l)}$, calculate the design-based estimator as in (5) and (6), that is, obtain

$$\widehat{FM}_d^{DB}(l) = \frac{1}{N_d}\sum_{j \in s_d(l)} w_{dj}\widehat{FM}_{dj}^{DB}(l), \; d = 1, ..., D, \quad (14)$$

where

$$\widehat{FM}_{dj}^{DB}(l) = \left\{\frac{\sum_{h=1}^{D}\sum_{i \in s_h(l)} w_{hi} I\{E_{hi}^{(l)} > E_{dj}^{(l)}\}}{\sum_{h=1}^{D}\sum_{i \in s_h(l)} w_{hi}}\right\}^{\alpha-1} \times \left\{\frac{\sum_{h=1}^{D}\sum_{i \in s_h(l)} w_{hi}\, E_{hi}^{(l)} I\{E_{hi}^{(l)} > E_{dj}^{(l)}\}}{\sum_{h=1}^{D}\sum_{i \in s_h(l)} w_{hi}\, E_{hi}^{(l)}}\right\}$$

and $E_{dj}^{(l)} = T^{-1}(Y_{dj}^{(l)})$. Finally, the fast EB estimator of FM_d is given by

$$\widehat{FM}_d^{FEB} = \frac{1}{L}\sum_{l=1}^{L}\widehat{FM}_d^{DB}(l), \; d = 1, ..., D.$$

As described in the next section, a model-based simulation study has been carried out to evaluate the performance of the proposed fast EB method to estimate, on the one hand, a traditional poverty measure, namely the domain PI, and the domain FMI and FSI. Results indicate that the fast EB estimator does not lose much efficiency as compared to the standard EB estimator, but it overcomes the computational problems due to large populations or more complex poverty measures.

Remark 4 : Observe that in the fast EB method, the number of univariate normal random variables that needs to be generated is $(n + D)L$ in contrast with the $(N - n + D)L$ variables to be generated in the standard EB method as described in Molina and Rao (2010). This implies a considerable reduction in computation time. As an example, for a population of size $N = 40{,}000{,}000$ with $D = 50$ domains and with sample size $n = 36{,}000$, similarly as in the application with Spanish SILC data in Molina and Rao (2010), the computation time will be reduced by the factor $(n + D)/(N - n + D) \approx 0.0009$.

Remark 5: Given the original sample s and $\mathbf{y}_s$, the random variables $\{FM_d^{(l)}, l = 1, ..., L\}$ are i.i.d. the same as FM_d. Then, by the Khintchine's Strong Law of Large Numbers, when $L \to \infty$, it holds

$$\frac{1}{L}\sum_{l=1}^{L} FM_d^{(l)} \overset{a.s.}{\to} E_{\mathbf{y}_r}(FM_d \mid \mathbf{y}_s) = \widehat{FM}_d^{B}.$$

Similarly, given s and $\mathbf{y}_s$, when $L \to \infty$, the fast EB estimator satisfies

$$\frac{1}{L}\sum_{l=1}^{L} \widehat{FM}_d^{DB}(l) \overset{a.s.}{\to} E_{\mathbf{y}_r, s^*}\left(\widehat{FM}_d^{DB} | \mathbf{y}_s\right),$$

where now the expectation on the right is with respect to the joint distribution of $(\mathbf{y}_r, s^*)$ given s and $\mathbf{y}_s$ because in the fast EB method, $\widehat{FM}_d^{DB}(l)$ is obtained from a second random sample $s^{*(l)}$ drawn from $\mathbf{y}^{(l)}$. Then, expressing the joint distribution in terms of the conditional distribution, we get

$$E_{\mathbf{y}_r, s^*}\left(\widehat{FM}_d^{DB} | \mathbf{y}_s\right) = E_{\mathbf{y}_r}\left\{E_{s^*}\left(\widehat{FM}_d^{DB} \mid \mathbf{y}_s, \mathbf{y}_r\right)\right\}$$

$$\approx E_{\mathbf{y}_r}(FM_d | \mathbf{y}_s) = \widehat{FM}_d^{B},$$

using the approximate design unbiasedness of $\widehat{FM}_d^{DB}$ (see Remark 2 above). Thus, provided that $\widehat{FM}_d^{DB}$ is approximately design unbiased, the fast EB method gives an estimator that is asymptotically equivalent to the Best/Bayes estimator.

4. MODEL BASED SIMULATION EXPERIMENT

The efficiency of fast EB estimators of the domain PIs, in comparison with that of EB estimators, has been studied in a model based simulation experiment. Fast EB estimators of the FMIs have also been compared with design-based and ELL estimators (Elbers *et al.* 2003). For this, we considered a population with $N = 20000$ units, partitioned into $D = 80$ domains with $N_d = 250$ units in each domain d, for $d = 1, ..., D$. The response variables for the population units Y_{dj} were generated from the nested-error model (10) using an intercept and two auxiliary variables, that is, $\mathbf{x}_{dj} = (1, x_{dj1}, x_{dj2})$, where the values of the two auxiliary variables were generated from $x_{dj1} \sim$ Binom (1, 0.2) and $x_{dj2} \sim$ Binom $(1, p_d)$ and, where

$$p_d = 0.3 + 0.5d/\mathrm{D},\ d = 1, . . ., D.$$

We assume that model responses Y_{dj} are the logarithm of the welfare variables E_{dj}. Thus, $E_{dj} = \exp(Y_{dj})$. A set of sample indexes s_d with $n_d = 50$ was drawn independently from each domain d using simple random sampling without replacement (SRSWR). The values of the auxiliary variables for the population units and the sample indexes were kept fixed across Monte Carlo simulations. The intercept and the regression coefficients associated with the two auxiliary variables were taken as $\beta = (3, 0.03, -0.04)'$. The random area effects variance was taken as $\sigma_u^2 = (0.15)^2$ and the error variance as $\sigma_e^2 = (0.5)^2$. The poverty line z was fixed as $z = 12$, which is approximately equal to 0.6 times the median of the welfare variables for a generated population. We generated $I = 1000$ Monte Carlo population vectors $\mathbf{y}^{(i)}$ from the true model. For each Monte Carlo population i, for $i = 1, ..., I$, the following quantities were computed:

1. True domain PIs,

$$PI_d^{(i)} = \frac{1}{N_d}\sum_{j=1}^{N_d} I(E_{dj}^{(i)} < z),\ E_{dj}^{(i)} = \exp(Y_{dj}^{(i)}),\quad d = 1, ..., D,$$

and true domain average FMIs obtained with $\alpha = 2$, that is,

$$FM_d^{(i)} = \frac{1}{N_d}\sum_{j=1}^{N_d} FM_{dj}^{(i)},\ d = 1, ..., D,$$

where

$$FM_{dj}^{(i)} = \left\{ \frac{1}{N-1} \sum_{l=1}^{D} \sum_{k=1}^{N_l} I(E_{lk}^{(i)} > E_{dj}^{(i)}) \right\} \times \left\{ \frac{\sum_{l=1}^{D} \sum_{i=1}^{N_l} E_{lk}^{(i)} I(E_{lk}^{(i)} > E_{dj}^{(i)})}{\sum_{l=1}^{D} \sum_{k=1}^{N_l} E_{lk}^{(i)}} \right\}.$$

2. Design-based estimators of the domain PIs and of the FMIs, using the sample part $\mathbf{y}_s^{(i)}$ of the i-th population vector $\mathbf{y}^{(i)}$.
3. Using the sample $\mathbf{y}_s^{(i)}$, Monte Carlo approximations of the EBPs of the domain PIs were computed the same as in Molina and Rao (2010), with $L = 50$ Monte Carlo replicates.
4. Using the sample $\mathbf{y}_s^{(i)}$, fast EB estimators of the domain PIs and of the average FMIs were calculated, based on $L = 50$ samples drawn from the population using the same sampling scheme as the one used for the original sample s.
5. ELL estimators (Elbers *et al.* 2003) of the domain PIs, based on generation of A =50 censuses using the procedure described in Molina and Rao (2010).
6. Means over Monte Carlo populations i = 1, ..., I of true values, design-based, EB, fast EB and ELL estimators of domain PIs and FMIs. For all these estimators, biases and MSEs over Monte Carlo populations i = 1, ..., I were also computed.

Fig. 1 and 2 show respectively mean values and MSEs of the PI for each area. Observe in Fig.1 that the mean values of the fast EB estimators are very close to those of the EB estimators. Design-based estimators are more unstable across areas, whereas ELL estimators are too stable and do not track the true values. Thus, this figure indicates that the biases of the fast EB estimators are very similar to those of the EB estimators, although the main differences between the estimators come from their MSEs. Observe in Fig. 2 that the MSEs of the EB and fast EB estimators are considerably smaller than those of the other estimators for all areas. Moreover, the MSEs of the fast EB estimators are only slightly larger than those of the EB estimators. These results suggest that the new fast EB estimators can reduce a lot the computational workload, and at the same time loose little efficiency in comparison with the EB estimators.

Analogously, Fig. 3 and Fig. 4 show respectively mean values and MSEs of design-based and fast EB estimators for the domain average FMI. Again, these figures show that the bias of the fast EB estimator is preserved small and similar to that of the design-based estimator, while the MSE is uniformly smaller for all areas.

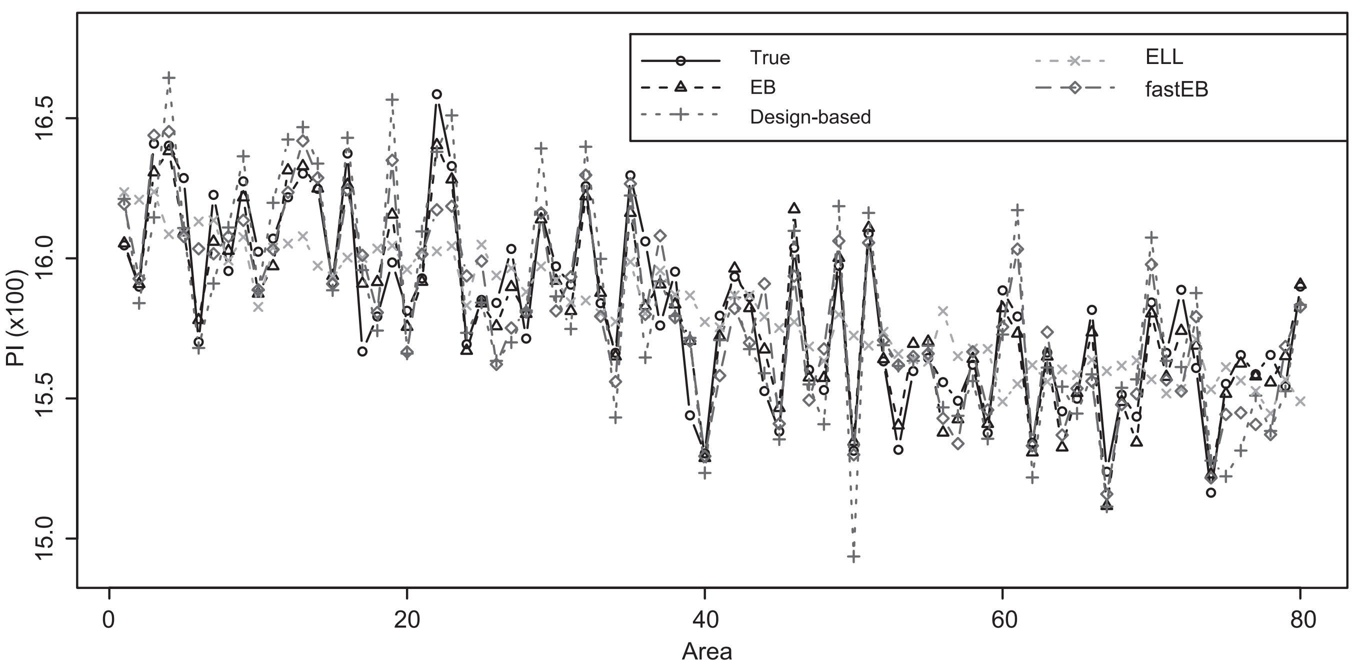

Fig. 1. Mean over simulated populations of true values, EB, design-based, ELL and fast EB estimators of PI (×100) for each area d.

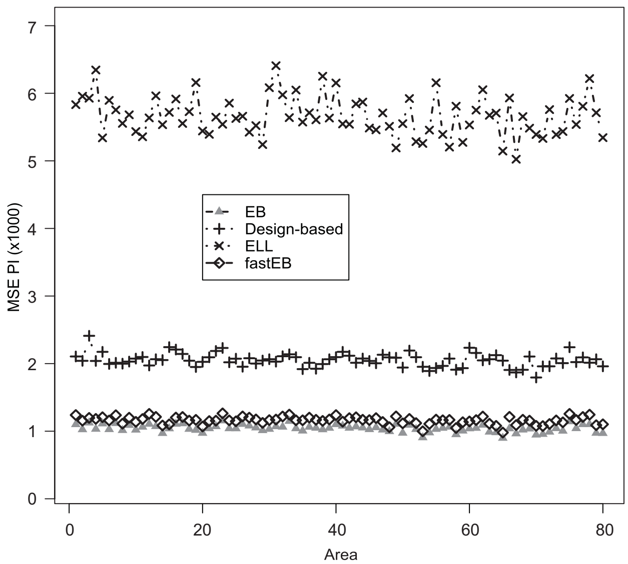

Fig. 2. MSE(×1000) over simulated populations of EB, design-based, ELL and fast EB estimators of PI for each area *d*.

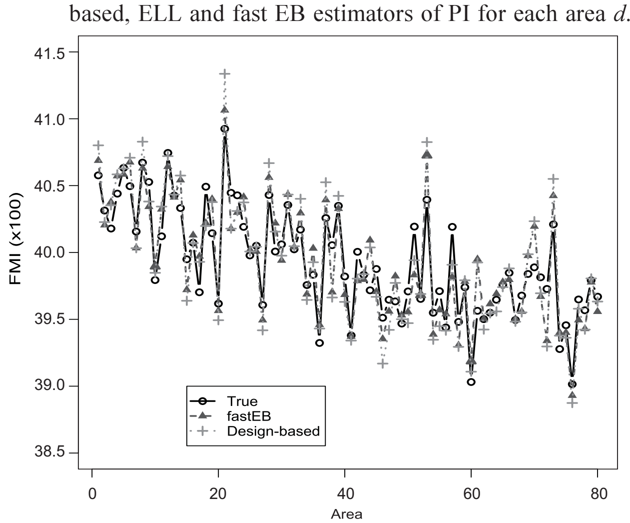

Fig. 3. Mean over simulated populations of true values, fast EB and design-based estimators of average FMI (× 100) for each area *d*.

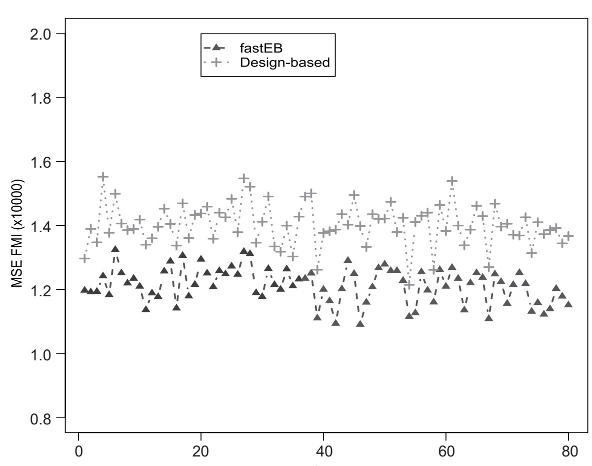

Fig. 4. MSE(×1000) over simulated populations of fast EB and design-based estimators of average FMI for each area *d*.

5. DESIGN BASED SIMULATION EXPERIMENT

A design based simulation experiment similar to that of Molina and Rao (2010) has been carried out to analyze the design based properties of the fast EB estimator in comparison with the original EB and the design-based estimators. Thus, similarly as in Molina and Rao (2010), only one population was generated in the same way as in Section 4, with the same population and sample sizes, and using the same values of model parameters. Then, in each replication out of $I = 1000$, a new sample was drawn from this fixed population according to SRS without replacement within each area. With each sample, the four types of estimators of poverty measures, namely EB, design-based, ELL and fast EB were calculated. Results on the design bias and design MSE of the estimators for poverty incidence are reported in Fig. 5 and Fig. 6 respectively. As expected, Fig. 5 shows that the empirical bias of the

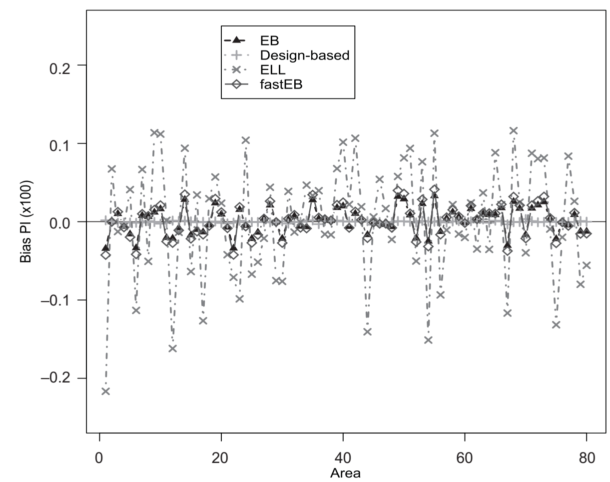

Fig. 5. Design bias(×100) of EB, design-based, ELL and fast EB estimators of PI for each area *d*.

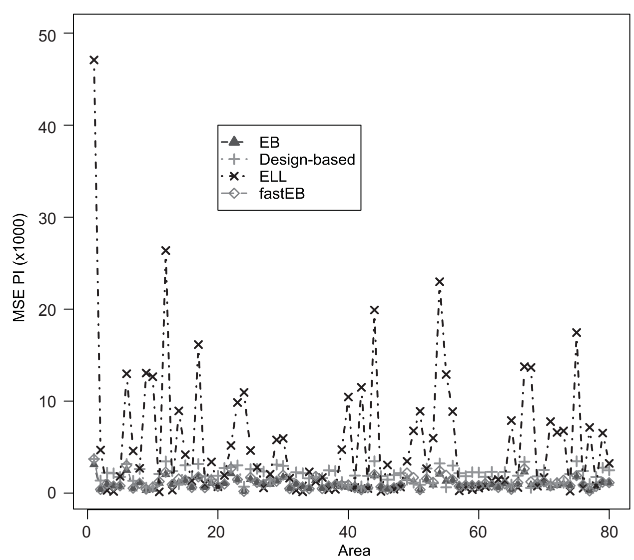

Fig. 6. Design-based MSE (×1000) of EB, design-based, ELL and fast EB estimators of PI for each area *d*.

design-based estimator is practically zero, followed by EB, fast EB and ELL estimators. In terms of MSE, Fig. 6 shows that it remains small for the EB, fast EB and design-based estimators for all areas.

6. APPLICATION TO EU-SILC DATA FROM TUSCANY

The proposed fast EB method was applied to estimate poverty incidences, fuzzy monetary and fuzzy supplementary indicators in Tuscany provinces. For this, data from the 2004 Italian Survey on Income and Living Conditions (SILC) was used.

The regional sample from Tuscany is based on a stratified two stage sampling design: within each province, the primary sampling units (PSUs) are the municipalities, and these are then grouped into strata according to their population size. From these PSUs, households are the secondary sampling units (SSUs), which are then selected by means of systematic sampling. Some of the smaller provinces have very few sampled municipalities and many municipalities are not represented in the sample. For example, in the 2004 survey, only 53 municipalities out of the total of 287 appear in the sample. Then, small area estimation techniques can be required given large errors of direct estimators at province level or the impossibility to compute them at municipalities level. Thus, in this section the fast EB method has been applied to obtain estimates of poverty measures for Tuscany provinces and also for municipalities.

First, we consider as domains of interest the $D = 10$ Tuscany provinces, with sample sizes ranging from 155 (Province of Grosseto) to 1403 (Province of Firenze). The regional sample size is $n = 4{,}426$ individuals. The considered welfare variable is the equivalized annual net income. This variable takes negative values for a group of individuals. To avoid this problem, we followed the recommendation of Eurostat (2006): applying a bottom coding strategy to the lowest values of the distribution. In particular, all values below 15% of the median household income have been set equal to the 15% of the median. This strategy does not affect the poverty line and consequently, neither the design-based estimators (Eurostat 2006, Ciampalini *et al.* 2009 and Neri *et al.* 2009). After that, the equivalized annual net income has been log transformed to obtain an approximately normal distribution. This transformed variable acts as response in the nested-error regression model. As auxiliary variables, we have considered the indicators of 5 quinquennial groups of variable age, the indicator of having Italian nationality, the indicators of 3 levels of the variable education level and the indicators of being unemployed, employed or inactive.

The poverty line for the calculation of the PI is computed as the 60% of the weighted median of the individual equivalised income at Regional level and is equal to 9,372.24 .

Design-based and fast EB estimators of the province PIs, FMIs and FSIs were calculated with $\alpha = 2$. Values of design-based and fast EB estimators of the PI and their associated coefficients of variation (CV) are shown in Table 1 for each Tuscany province. The average over provinces of the estimated PI is 16.4%. The poorer provinces are located mainly in the north-west of Tuscany. According to the fast EB estimators, the province of Massa Carrara has the largest percentage of people under the poverty line (22.4%) followed by Lucca (18.2%) and Pisa (17.8%). On the other hand, the provinces of Arezzo (13.0%) and Firenze (14.4%) contain the less people under the poverty line. The MSEs of fast EB estimators of the PIs are calculated using the parametric bootstrap described in Molina and Rao (2010) with $B = 500$ replicates. The coefficients of variation are obtained as $\mathrm{cv}(\widehat{PI}_d^{\,fast\ EB}) = 100 \times \mathrm{mse}(\widehat{PI}_d^{\,fast\ EB})^{1/2} / \widehat{PI}_d^{\,fast\ EB}$. Table 1 shows that the CVs of fast EB estimators are much smaller than those of design-based estimators and the reduction in CV tends to be greater for domains with smaller sample size. The only exception is Province of Firenze with a large sample size, for which the CV of the fast EB estimator is larger than that of the design-based estimator.

Table 2 reports respectively design-based and fast EB estimates of the province FMIs and FSIs. Fast EB estimators of the FMIs provide a similar picture concerning the monetary poverty in the small areas as the PI. The estimated FMIs are larger than the estimated PIs because in each province there is a concentration of individuals with equivalised income just above the poverty line. The provinces of Arezzo (36.5%) and Firenze (38.0%) remain the richest, whereas the province of Massa Carrara has the largest estimated percent FMI (47.5%) followed by Lucca (42.6%) and Pisa (42.3%).

Table 1. Tuscany province, population size, sample size, design-based and fast EB estimators of PI (×100), CVs of design-based and fast EB estimators (×100).

Province	N_d	n_d	$\widehat{PI}_d^{DB}$	$\widehat{PI}_d^{fast\ EB}$	cv($\widehat{PI}_d^{DB}$)	cv($\widehat{PI}_d^{fast\ EB}$)
Arezzo	304121	416	8.7	13.0	19.09	12.42
Firenze	1119377	1403	13.3	14.4	8.32	10.89
Grosseto	149082	155	12.4	14.7	26.10	11.10
Livorno	290122	339	13.1	14.9	15.61	10.30
Siena	278495	338	11.0	15.6	19.31	10.80
Prato	319320	416	17.0	15.9	14.06	10.77
Pistoia	267076	344	17.4	16.9	13.75	9.87
Pisa	335777	399	16.8	17.8	15.54	8.88
Lucca	265293	315	21.5	18.2	13.30	8.88
Massa Carrara	251471	301	26.0	22.4	13.09	7.30
Average			**15.7**	**16.4**		

Table 2. Tuscany province, sample size, design-based and fast EB estimators of FMI and FSI (×100), latent (LAT_d) and manifest (MAN_d) deprivation (×100), and ratio MAN_d/LAT_d (×100).

Province	n_d	$\widehat{FM}_d^{DB}$	$\widehat{FM}_d^{fast\ EB}$	$\widehat{FS}_d^{DB}$	$\widehat{FS}_d^{fast\ EB}$	LAT_d	MAN_d	MAN_d/LAT_d
Arezzo	416	35.4	36.5	26.2	29.7	49.4	16.7	33.8
Firenze	1403	37.6	38.0	37.1	36.8	54.2	20.6	38.0
Grosseto	155	39.0	38.3	15.8	20.3	46.0	12.7	27.6
Livorno	339	37.9	39.2	37.0	37.2	55.2	21.2	38.4
Siena	338	38.1	39.6	30.6	32.1	52.6	19.1	36.3
Prato	416	40.2	40.4	33.2	34.7	54.5	20.6	37.7
Pistoia	344	42.4	41.4	41.1	38.0	57.0	22.5	39.4
Pisa	399	43.3	42.3	35.3	34.8	55.8	21.2	38.1
Lucca	315	42.4	42.6	39.8	36.0	56.6	22.0	38.8
Massa Carrara	301	49.6	47.5	33.0	31.6	57.8	21.3	36.8
Average		**40.6**	**40.6**	**32.9**	**33.1**			

Concerning the FSIs, in this case the score variables S_i, constructed as explained in Section 2, are acting as welfare variables. As response variable in the nested-error regression model, we have taken the cloglog transformation of the score. We used the same auxiliary variables employed for the estimation of the PIs and the FMIs. Thus, concerning non-monetary deprivation as measured by the FSIs, the ranking of some of the provinces is completely different from that displayed by the monetary poverty measures, see Table 2. For example, the province of Massa Carrara has a large value of FMI but a small value of FSI, and the opposite holds with Firenze and Livorno provinces.

The propensity to income poverty as measured by FM_{dj} for individual j within domain d and the overall non-monetary deprivation propensity as given by FS_{dj} may be combined to construct composite measures which indicate the extent to which the two aspects, namely income poverty and non-monetary deprivation, overlap for the same individual. For individual j in domain d, the Manifest deprivation is defined as $MAN_{dj} = \min(FM_{dj}, FS_{dj})$, and it represents the propensity to both income poverty and non-monetary deprivation simultaneously. Similarly, we define the Latent deprivation for individual i as $LAT_{dj} = \max(FM_{dj}, FS_{dj})$, which indicates the propensity of individual j to suffer from at least one of the two, income poverty and/or non-monetary deprivation. Then the Manifest (Latent) deprivation index for domain d, MAN_d (LAT_d) is obtained as a weighted mean of the Manifest (Latent) deprivation of individuals in that domain. A useful indicator in this context is the rate between the Manifest deprivation index and the Latent deprivation index, which is included between 0 and 1. When there is no overlap (i.e., when the subpopulation subject to income poverty is entirely different from the subpopulation subject to non-monetary deprivation), the Manifest deprivation rate and hence the above mentioned ratio equals 0. When there is complete overlap, i.e., when each individual is subject to exactly the same degree of income poverty and of non-monetary deprivation, the Manifest and Latent deprivation indexes are the same and hence the above mentioned ratio equals 1.

Note from Table 2 that there are some differences between provinces. The overlap, as measured by the ratio MAN_d/LAT_d, takes the smallest value of 27.6% for Grosseto and the largest of 39.4% for Pistoia. In general, the MAN_d/LAT_d ratio is smaller for areas with lower levels of deprivation, and larger for areas with higher levels. A large value of this ratio means that different types of deprivation overlap and this means that in areas where levels of relative deprivation are already high, deprivation in the income and non-monetary domains is more likely to afflict the same individuals in the population. On the other hand, low values imply the absence of such overlap at micro level.

Fig. 7 and Fig. 8 show respectively the cartograms of the estimated province PIs, FMIs, FSIs and manifest/latent deprivation ratio in Tuscany provinces, constructed using the fast EB estimators.

In the following we show the results obtained for the 53 Tuscany municipalities, with sample sizes ranging from 23 (Municipality of Sovicille) to 408 (Municipality of Firenze). Table A.1 lists the values of design-based estimators and fast EB estimators of the PI, together with CVs of design-based estimators, for each Tuscany municipality. The small difference in the averages of the two estimators might indicate a small bias of the fast EB estimators, probably due to the non-exact normality of the log-income. However, observe that the coeffcients of variation of design-based estimators for most municipalities are considerably large, which means that small area estimation methods are necessary. The fast EB estimators in Table A.1 show large differences among municipalities, from the lowest value of 4.6% for Siena to the highest of 45.9% for Massa with an average over Municipalities of 15.8%. In 13 municipalities, the rate exceeds 20% whereas in 10 of them it is below 12%. The poorer municipalities concentrate mainly in the north-west of Tuscany and

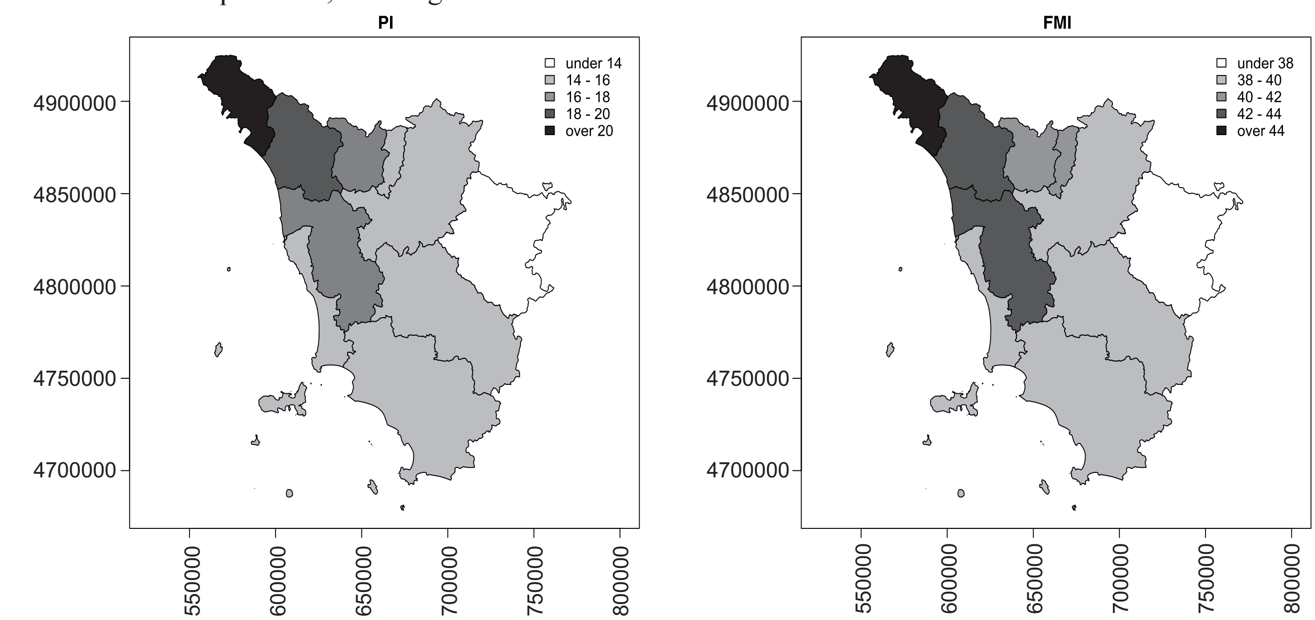

Fig. 7. Cartograms of estimated percent poverty incidence and fuzzy monetary indicators in Tuscany provinces.

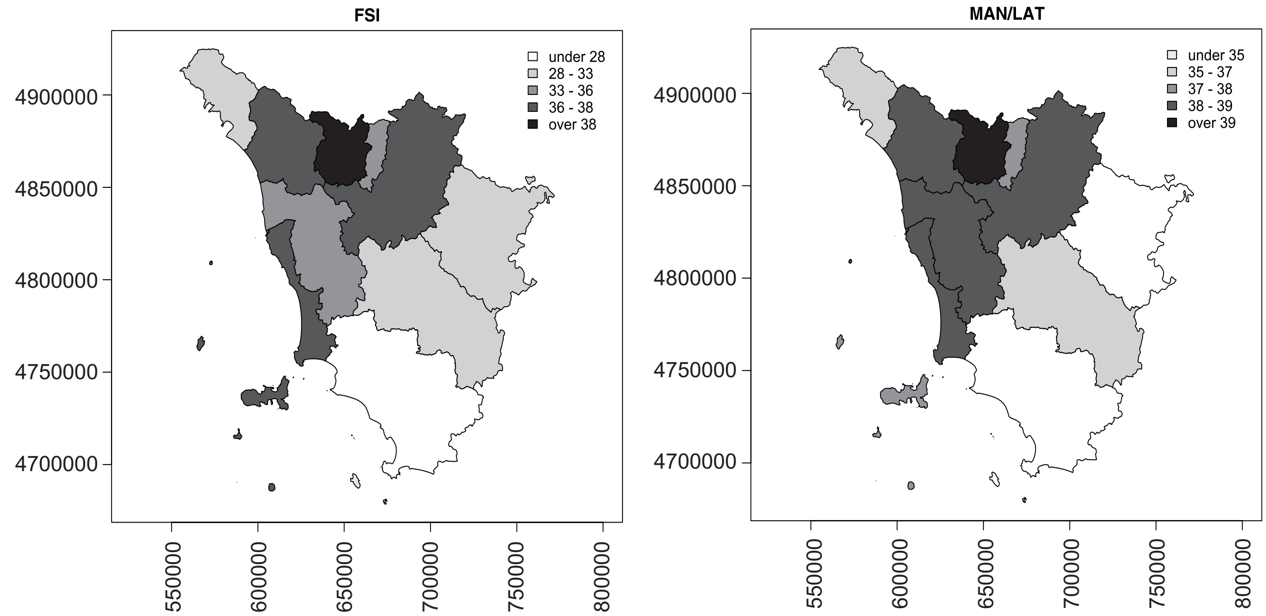

Fig. 8. Cartograms of estimated percent fuzzy supplementary indicators and manifest/latent deprivation ratio in Tuscany provinces.

in particular in the Province of Massa. The municipality of Massa has the largest percentage of poor individuals (45.9%), followed by Certaldo (29.0%) and San Godenzo (27.6%). On the other hand, the municipality of Siena (4.6%), Montemurlo (6.1%) and Pescia (7.8%) appear to be the richest.

Table A.2 lists design based and fast EB estimators of FMI and FSI, together with the poverty indicators obtained as a combination of these two. Fast EB estimators of FMI give a similar picture about the monetary poverty in the municipalities as the PI and the differences between municipalities are similar according to both indicators. Estimated values of FMI are larger than those of the PI since there is a certain concentration in each municipality of individuals with equivalised income just above the poverty line. The municipality of Siena (22.1%) and Montemurlo (25.3%) remain the richest, whereas the municipality of Massa has the largest percentage of poor individuals (67.5%), followed by Certaldo (53.8%) and San Godenzo (52.9%).

Concerning now the non-monetary deprivation as measured by FSI, observe in Table A.2 that the ranking of municipalities is quite similar to that obtained from monetary deprivation at the extreme of the distribution. On the other hand, few municipalities get a completely opposite ranking as compared to the one based on monetary poverty. For example, see that Fivizzano, Isola del Giglio and San Godenzo have large values of estimated FMI together with small values of estimated FSI, and the opposite holds for Firenze and Pian di Sco. The non-monetary dimension is combined with the monetary dimension in order to obtain measures of manifest (MAN_d) and latent (LAT_d) deprivation, which correspond respectively to the intersection and union of the fuzzy sets. The ratio MAN_d/LAT_d, which measures the overlap between the monetary and non-monetary aspects of deprivation, was also obtained for Tuscany municipalities. Table A.2 shows significant differences among municipalities. The ratio MAN_d/LAT_d takes the lowest value of 25.8% for Santa Fiora and Montemurlo and the largest of 49.3% for Massa. The figures displayed in the table suggest that, in this application, the ratio MAN_d/LAT_d tends to be smaller for areas with lower levels of deprivation (of both types) and larger for areas with higher levels of deprivation.

7. CONCLUDING REMARKS

A computationally faster version of the EB method of Molina and Rao (2010) has been proposed. This method makes feasible the estimation of complex indicators, such as the fuzzy poverty measures considered here, in large populations. It can also increase considerably the computational speed when estimating traditional indicators such as the poverty incidence in large populations. Moreover, it avoids the need for identifying the survey units in the census or administrative register from where auxiliary variables are obtained.

Model and design based simulations suggest that the fast EB estimators perform similarly as their original EB counterparts, losing little efficiency. This fast version of the EB method has allowed to obtain model based estimators of monetary and non-monetary poverty indicators, namely FMI and FSI, for the region of Tuscany, at province and municipality level.

The fast EB method proposed in this paper, as the original EB method, requires normality of some transformation of the variable of interest. In the application with SILC data from Tuscany, the normal approximation on the left tail of the distribution is a little rough and this seems to cause a small bias in the EB estimator. Relaxation of this assumption requires further research.

As in the original EB method, in practice we recommend to use a number of Monte Carlo replicates L larger than 50. The computational speed of the fast EB method allows to use larger values of L than in the EB method. A general purpose value for any Monte Carlo approximation is $L = 1000$, but our simulation experience indicates that $L = 200$ gives estimates that are reasonably stable.

Mean square errors of fast EB estimators of fuzzy poverty measures can be obtained by a parametric bootstrap procedure; however, even using the fast EB method, this might be still computationally intensive if the population is too large.

Concerning the parameter α involved in the fuzzy measures, as explained before, in our analysis we have considered $\alpha = 2$. Determining the value of α to make the FMI and FSI equal to the PI in a computationally feasible way deserves further study.

The sampling design used to draw the sample $s^{(l)}$ from U in l-th Monte Carlo replication of the fast EB method does not need to be the same as that one used for the original sample s. When the sampling design used to draw s is not completely known, $s^{(l)}$ can be drawn using simple random sampling without replacement within each domain d, keeping the same allocation of domain sample sizes $\{n_d;\ d = 1, ..., D\}$. Furthermore, allocation of domain sample sizes in the synthetic samples could be optimized to improve the efficiency of the fast EB estimators. This issue deserves also further research.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish grants SEJ2007-64500 and MTM2009-09473 from the Ministerio de Educación y Ciencia in Spain.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller,W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Betti, G., Cheli, B., Lemmi, A., Verma, V. (2006). Multidimensional and longitudinal poverty: An integrated fuzzy approach. In: Lemmi, A., Betti, G. (eds.) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, 111-137, Springer, New York.

Betti, G., Ferretti, C., Gagliardi, F., Lemmi, A., Verma, V. (2009). Proposal for new multidimensional and fuzzy measures of poverty and inequality at national and regional level. Working Paper 83/09, Department of Quantitative Methods, University of Siena.

Betti, G. and Verma V. (1999). Measuring the degree of poverty in a dynamic and comparative context: A multi-dimensional approach using fuzzy set theory. *Proceedings, ICCS-VI*, 11, pp. 289-301, Lahore, Pakistan, August 27-31, 1999.

Cheli, B. and Betti, G. (1999). Totally fuzzy and relative measures of Poverty Dynamics in an Italian Pseudo Panel. 1985-1994. *Metron*, **57(1-2)**, 83-104.

Cheli, B. and Lemmi, A. (1995). A totally fuzzy and relative approach to the multi-dimensional analysis of poverty. *Economic Notes*, **24**, 115-134.

Ciampalini, G., Betti, G. and Verma, V. (2009). Comparability in self-employment income. Working Paper 82/09, Department of Quantitative Methods, University of Siena.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003). Micro-level estimation of poverty and in equality. *Econometrica*, **71**, 455-364.

Eurostat (2006). Treatment of negative income: Empirical assessment of the impact of methods used. Report N. ISR I.04, Project EU-SILC (Community statistics on income and living conditions) 2005/S 116-114302 – Lot 1 (Methodological studies to estimate the impact on comparability of the national methods used).

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

Molina, I. and Rao, J.N.K. (2010). Small area estimation of poverty indicators. *Canad. J. Statist.*, **38**, 369-385.

Neri, L., Gagliardi, F., Ciampalini, G., Verma, V. and Betti, G. (2009). Outliers at upper end of income distribution. Working Paper 86/09, Department of Quantitative Methods, University of Siena.

APPENDIX: RESULTS FOR MUNICIPALITIES

Table A.1. Population size, sample size, design-based and fast EB estimators of PI (×100), and CVs of design-based estimators (×100) for 54 Tuscany Municipalities.

Municipality	N_d	n_d	$\widehat{PI}_d^{DB}$	$\widehat{PI}_d^{fast\ EB}$	cv($\widehat{PI}_d^{DB}$)
Siena	66988	75	0.0	4.6	–
Montemurlo	53258	75	0.0	6.1	–
Pescia	43181	56	7.5	7.8	50.0
Pian di Sco	44789	71	0.0	8.3	–
Firenze	362591	408	13.2	8.7	16.0
Bagno a Ripoli	76415	94	6.1	8.9	41.0
Bibbiena	48765	67	7.2	10.1	50.2
Grosseto	71211	89	6.6	10.5	42.5
Arezzo	111864	145	9.4	11.0	29.9
Reggello	48522	65	3.7	11.7	70.8
Vaiano	50055	69	7.6	12.1	41.4
Podenzana	41063	69	11.0	12.6	42.7
Livorno	148969	171	12.5	12.8	22.3
Santa Fiora	41672	32	9.6	12.9	58.6
Capannori	94622	120	12.2	12.9	30.9
Calcinaia	33621	43	0.0	12.9	–
Campi Bisenzio	57764	74	3.8	13.0	58.7
Cecina	83711	100	16.5	13.4	26.2
Carmignano	35869	56	17.7	13.7	35.8
Lastra a Signa	50455	73	9.1	13.8	43.4
Pontassieve	49999	65	12.3	14.2	44.7
Pisa122490	122	22.9	14.4	25.0	
Castel Fiorentino	58114	74	14.6	14.9	38.1
San Giuliano Terme	76690	100	6.0	15.1	40.9
Licciana Nardi	51761	63	23.3	15.2	29.1
Pistoia	98096	116	16.1	15.4	24.6
Pelago	42455	62	11.7	15.8	39.7
Empoli Castiglion	77761	109	11.2	16.2	28.5
Fiorentino	46579	72	8.6	16.3	40.6
Fucecchio	60021	78	12.9	16.9	38.5
Scandicci	43404	45	19.9	17.2	33.9
Figline Valdarno	42683	65	15.4	17.4	34.8
Bucine	52124	61	16.3	17.5	37.3
Rapolano Terme	49621	61	7.9	17.5	42.8
Lucca	90046	103	23.1	18.8	22.5
Campiglia Marittima	57442	68	9.5	18.8	39.5
Asciano	43708	54	7.0	18.9	57.9
Incisa in val d'Arno	56247	75	17.5	19.0	29.5
Torrita di Siena	39173	58	10.0	19.1	49.9
Montepulciano	55948	67	25.5	19.1	31.0
Quarrata	48249	65	21.5	19.5	27.7
Pieve a Nievole	28590	54	40.1	20.2	28.2
Carrara	53770	66	16.7	20.7	34.0
Pomarance	58223	85	22.5	22.0	26.6
Prato	180138	216	24.6	23.3	16.0
Isola del Giglio	36199	34	27.3	24.1	37.9
Buggiano	48960	53	11.7	24.2	37.9
Lari	44753	49	23.5	24.7	32.4
Viareggio	80625	92	30.5	25.0	19.1
Fivizzano	53400	51	24.6	25.7	29.0
Sovicelle	23057	23	23.4	27.3	43.4
San Godenzo	43566	52	29.3	27.6	27.9
Certaldo	49380	64	27.9	29.0	24.2
Massa	51477	52	51.7	45.9	21.1
Average			**14.8**	**15.8**	

Table A.2. Sample size,design-based and fast EB estimators of FMI and FSI (×100), latent (LAT_d) and manifest (MAN_d) deprivation (×100), and ratio MAN_d/LAT_d for 54 Tuscany Municipalities (×100).

Municipality	n_d	$\widehat{FM}_d^{DB}$	$\widehat{FM}_d^{fast\ EB}$	$\widehat{FS}_d^{DB}$	$\widehat{FS}_d^{fast\ EB}$	LAT_d	MAN_d	MAN_d/LAT_d
Siena	75	12.9	22.1	23.3	31.9	42.3	11.8	27.8
Montemurlo	75	20.0	25.3	16.2	21.0	36.8	9.5	25.8
Pescia	56	28.8	29.4	28.0	26.5	43.2	12.7	29.3
Pian di Sco	71	25.7	29.9	28.5	39.1	51.2	17.8	34.8
Firenze	408	29.5	30.4	44.3	43.8	54.9	19.2	35.0
Bagno a Ripoli	94	27.6	31.4	24.9	26.3	44.5	13.2	29.7
Bibbiena	67	27.8	32.8	28.6	30.4	47.8	15.4	32.3
Grosseto	89	33.4	33.1	19.1	19.8	41.7	11.2	26.9
Arezzo	145	35.1	34.4	20.1	22.7	44.4	12.7	28.6
Reggello	65	29.6	35.3	20.8	23.2	45.4	13.2	29.0
Vaiano	69	30.9	35.7	30.2	34.1	51.6	18.2	35.4
Podenzana	69	35.6	36.8	13.4	19.9	44.8	11.9	26.5
Campi Bisenzio	74	32.0	36.8	35.3	37.4	54.3	19.9	36.7
Capannori	120	34.6	36.9	32.5	28.6	49.4	16.1	32.7
Santa Fiora	32	38.5	37.2	8.8	18.5	44.4	11.4	25.7
Livorno	171	35.4	37.3	44.2	41.3	56.8	21.8	38.4
Cecina	100	39.7	37.7	33.3	34.5	53.2	19.0	35.8
Carmignano	56	33.0	37.8	53.3	52.2	63.9	26.1	40.9
Calcinaia	43	30.2	37.9	31.1	33.4	52.6	18.6	35.4
Pontassieve	65	40.9	38.5	27.1	28.9	50.6	16.9	33.3
Lastra a Signa	73	34.7	38.6	33.0	37.2	55.2	20.6	37.3
Pisa	122	42.6	38.8	42.5	36.1	54.7	20.2	36.9
San Giuliano Terme	100	38.8	39.8	28.8	28.7	51.0	17.5	34.2
Licciana Nardi	63	42.4	39.9	24.8	26.5	50.3	16.1	32.0
Castel Fiorentino	74	42.6	40.2	43.3	39.0	56.8	22.3	39.4
Pistoia	116	39.8	40.2	29.6	28.3	51.4	17.0	33.0
Pelago	62	40.9	40.3	47.0	45.8	60.8	25.3	41.6
Empoli	109	46.8	41.7	34.9	32.9	54.7	19.9	36.3
Castiglion Fiorentino	72	43.9	42.1	34.8	33.4	55.2	20.3	36.7
Scandicci	45	40.9	42.1	41.5	37.6	57.4	22.3	38.8

Municipality	n_d	$\widehat{FM}_d^{DB}$	$\widehat{FM}_d^{fast\ EB}$	$\widehat{FS}_d^{DB}$	$\widehat{FS}_d^{fast\ EB}$	LAT_d	MAN_d	MAN_d/LAT_d
Fucecchio	78	45.0	42.2	30.2	34.6	55.9	20.9	37.4
Figline Valdarno	65	47.0	42.6	30.7	32.1	55.2	19.4	35.2
Rapolano Terme	61	38.6	43.0	16.5	20.6	50.0	13.6	27.3
Bucine	61	43.8	43.2	24.6	30.5	54.4	19.3	35.4
Asciano	54	45.8	43.5	46.5	41.4	60.3	24.6	40.8
Lucca	103	43.8	44.0	34.2	33.2	56.4	20.8	36.8
Quarrata	65	45.0	44.1	53.9	50.8	65.6	29.3	44.6
Campiglia Marittima	68	41.9	44.2	34.3	34.1	56.7	21.6	38.1
Incisa in val d'Arno	75	45.0	44.5	39.1	39.5	60.0	24.1	40.1
Montepulciano	67	48.6	44.7	30.8	31.9	56.2	20.4	36.3
Torrita di Siena	58	46.9	45.0	36.8	36.3	58.6	22.6	38.6
Pieve a Nievole	54	50.7	45.4	50.8	41.1	61.1	25.4	41.6
Carrara	66	49.0	46.8	48.6	43.2	62.9	27.1	43.0
Pomarance	85	49.4	47.9	34.5	35.5	60.1	23.3	38.7
Prato	216	50.1	49.3	35.7	34.8	60.6	23.5	38.8
Buggiano	53	52.1	49.5	51.9	51.4	68.4	32.5	47.6
Isola del Giglio	34	50.5	50.0	18.7	23.1	56.5	16.6	29.5
Lari	49	54.8	50.1	35.5	35.9	61.5	24.5	39.8
Fivizzano	51	48.0	50.5	21.7	26.6	58.4	18.8	32.1
Viareggio	92	50.2	51.0	50.6	49.4	68.5	32.0	46.7
Sovicelle	23	55.7	51.8	33.0	31.2	61.2	21.8	35.6
San Godenzo	52	58.9	52.9	25.1	27.6	60.2	20.3	33.7
Certaldo	64	55.7	53.8	35.6	37.2	64.4	26.5	41.1
Massa	52	70.2	67.5	56.2	46.1	76.1	37.5	49.3
Average		**39.7**	**39.9**	**34.3**	**34.5**			