Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 1, 2019 (New Series), pp 11–22

Small Area Estimation Methods for Poverty Mapping: A Selective Review

Isabel Molina[1], **J. N. K. Rao**[2] **and Maria Guadarrama**[3]

[1] *Universidad Carlos III de Madrid, Madrid, Spain*
[2] *Carleton University, Ottawa, Canada*
[3] *Luxembourg Institute of Socio-Economic Research (LISER), Luxembourg*

Received: January 10, 2019; Accepted: January 10, 2019

Abstract

Poverty mapping in small areas is currently having increasing interest, because those maps aid governments and international organizations to design, apply and monitor more effectively regional development polices, directing them to the actual places or population subgroups where they are more urgently needed. After a simulated census method used by the World Bank, several other procedures have been developed that proved to have better properties. We will review several methods that are applied for poverty mapping in small areas, including those based on area level modes and used by the U. S. Census Bureau for estimating poor school age children and methods based on unit level models such as the traditional method used by the World Bank and empirical best (EB) and hierarchical Bayes (HB) methods based on optimality criteria. We will also discuss some variations of the unit level model methods that can used to deal with certain situations such as informative sampling or two-stage sampling. We will discuss pros and cons of these methods from a practical point of view, but based on the theory that is currently known.

Key words: Area level model; Empirical best estimation; Hierarchical bayes estimation; Local poverty indicators; Unit level models.

1 Introduction

Poverty maps are important sources of information on the regional distribution of poverty and are currently used extensively to support regional policy making and for allocating funds to local jurisdictions. Major applications include the poverty and inequality maps produced by the World Bank for many developing countries. In the U.S., the Small Area Income and Poverty Estimates (SAIPE) program (http://www.census.gov/programs-surveys/saipe.html) of the Census Bureau provides annual estimates of poor school-age children for all school districts, counties and states for the allocation of federal funds to local jurisdictions. In Europe, several efforts have been made to create regional databases and associated maps of poverty and social exclusion indicators in order to support regional development policies; see, for example, the joint project "Poverty Mapping in the New Member States of the European Union" between the World Bank and the European Commission and the TIPSE project (The Territorial Dimension of Poverty and Social Exclusion in Europe) commissioned by the European Observation Network for Territorial

Corresponding Author: J.N.K. Rao
E-mail: jrao@math.carleton.ca

Development and Cohesion (ESPON) program. In Mexico, the National Council for the Assessment of the Social Development Policy (CONEVAL) is committed by law to produce regular poverty and inequality estimates at the state level by population subgroups and at the municipality level.

Producing accurate poverty maps at high level of disaggregation is not straightforward because of insufficient sample sizes of official surveys in some of the target regions. As a result, direct estimates, obtained with the region-specific sample data, are unstable in the sense of leading to very large sampling variances for regions with small sample sizes. In the paper, we review the main methods for estimating general non-linear small area parameters, focusing for illustrative purposes on a specific family of poverty indicators, known as FGT family of poverty indicators Foster *et al.*(1984), widely used by World Bank and others. Specifically, we describe direct estimation, empirical best linear unbiased prediction (EBLUP) based on the well-known Fay-Herriot area level model Fay and Herriot (1979), the method of Elbers *et al.* (2003), called the ELL method and used by the World Bank, empirical best or empirical Bayes (EB) method of Molina and Rao (2010), hierarchical Bayes (HB) method of Molina *et al.* (2014), and other variants of the EB method to deal with two-stage sampling or informative sampling of units. For the benefit of a practitioner, we discuss, as objectively as possible, the benefits and drawbacks of each method.

2 Poverty Indicators

Suppose that the population consists of N population units with N_d units belonging to local (or small) area $d\,(d=1,\dots,D)$. We denote the welfare variable of interest for unit i $(i=1,\dots,N_d)$ in area d by E_{di}. The family of FGT poverty indicators for area d may be expressed as

$$F_{\alpha d}=N_d^{-1}\sum\nolimits_{i=1}^{N_d}[(z-E_{di})/z]^{\alpha}I(E_{di}<z)\equiv N_d^{-1}\sum\nolimits_{i=1}^{N_d}F_{\alpha di},\alpha\geq 0, \quad (1)$$

where z is a specified poverty line (or threshold), $I(E_{di}<z)=1$ if E_{di} is below z and $I(E_{di}<z)=0$ otherwise. For $\alpha=0$, we obtain the proportion of individuals "at risk of poverty", that is, the poverty incidence or at-risk-of-poverty rate or Head Count Ratio (HCR). For $\alpha=1$, we get the average of the relative distances to not being "at risk of poverty", called the poverty gap. Poverty incidence measures the frequency of poverty, whereas poverty gap measures intensity of poverty. It is clear from (1) that FGT measure is a separable non-linear function of the welfare variable E_{di}. Poverty line might vary across areas d if such area-specific thresholds are available. A limitation of the FGT measure is that it requires the specification of the poverty line/s.

More complex poverty measures have also been proposed, including the fuzzy monetary and supplementary indices based on ranking the individuals with respect to their level of poverty or welfare. An advantage of those measures is that the specification of a poverty line is not required. We refer the reader to Neri *et al.* (2005) for a description of alternative poverty indicators.

3 Direct Estimators of FGT Poverty Indicators

Let s be a probability sample of size n drawn from the finite population P. We denote the subsample of size n_d drawn from area d by s_d ($n_1+\dots+n_D=n$). A direct estimator of the poverty indicator $F_{\alpha d}$ is obtained by using only sample observations $E_{di}, i \in s_d$, from area (or domain) d, provided n_d is strictly positive.

A basic design-based direct estimator of $F_{\alpha d}$ is simply given by

$$\hat{F}_{\alpha d} = N_d^{-1} \sum\nolimits_{i \in s_d} w_{di} F_{\alpha di} \tag{2}$$

provided the domain population size N_d is known, where w_{di} is the survey weight attached to unit i from area d. If the domain population size N_d is unknown, then we replace N_d in (2) by its design-based estimator $\hat{N}_d = \sum\nolimits_{i \in s_d} w_{di}$ (Biggeri *et al.*, 2018). The resulting estimator suffers from ratio bias if the domain sample size n_d is small. Both estimators can lead to large sampling errors if n_d is small because only area-specific sample observations are used. Here we will focus on the direct estimators (2), assuming known N_d, and use them as inputs in an area-level linking model to obtain more efficient model-based estimators of the poverty indicators $F_{\alpha d}$.

4 EBLUP Method under an Area Level Linking Model

Fay and Herriot (1979) introduced an area level model that links the parameters of interest $F_{\alpha d}$ for all the areas $d = 1, ..., D$. For this purpose, we need a p-vector of covariates x_d for all the areas related to the parameters of interest $F_{\alpha d}$. This supplementary data might be obtained from a recent census and/or current administrative records. Under this set up, the linking model is specified through a linear regression model

$$F_{\alpha d} = x_d' \beta + u_d, d = 1, ..., D. \tag{3}$$

Here, β is the p-vector of regression parameters common for all areas and u_d is the area-specific regression error, also called random effect for area d. The random area effects u_d are assumed to be independent and identically distributed (iid) with mean 0 and unknown variance σ_u^2, that is, $u_d \sim_{iid} (0, \sigma_u^2)$. Note that the true values $F_{\alpha d}$ are not observable and therefore model (3) cannot be directly fitted. However, we can make use of the direct estimators $\hat{F}_{\alpha d}$ given by (2) through the sampling model

$$\hat{F}_{\alpha d} = F_{\alpha d} + e_d, d = 1, ..., D \tag{4}$$

where e_d is the sampling error associated with $\hat{F}_{\alpha d}$. We assume that sampling is done independently across all the areas so that the sampling errors e_d are independent given the true values $F_{\alpha d}$. We further assume that the sampling errors are independent of the random area effects and satisfy $e_d \sim_{ind} (0, \psi_d)$ with known sampling variance ψ_d. The above assumptions are somewhat restrictive in practice, and methods for relaxing some of the assumptions are discussed in Rao and Molina (2015). Combining the sampling model (4) with the matched linking model (3) leads to a linear mixed model given by

$$\hat{F}_{\alpha d} = x_d'\beta + u_d + e_d, d = 1, \ldots, D. \tag{5}$$

Standard theory for linear mixed models can be applied to the model (5) to obtain the best linear unbiased predictor (BLUP) of $F_{\alpha d}$; see Rao and Molina (2015), chapter 6. Normality of the random effects and the sampling errors is not needed to obtain the BLUP.

The BLUP of the poverty indicator $F_{\alpha d}$ under model (5) is given by a weighted combination of the direct estimator $\hat{F}_{\alpha d}$ and the synthetic estimator with weights $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi_d)$ and $1 - \gamma_d$ respectively:

$$\tilde{F}_{\alpha d}^{FH} = \gamma_d \hat{F}_{\alpha d} + (1 - \gamma_d) x_d' \tilde{\beta}, \tag{6}$$

where the superscript FH stands for Fay and Herriot (1979) and $\tilde{\beta}$ is the weighted least squares estimator of β given by

$$\tilde{\beta} = (\textstyle\sum_{d=1}^{D} \gamma_d \, x_d x_d')^{-1} (\textstyle\sum_{d=1}^{D} \gamma_d \, x_d \hat{F}_{\alpha d}). \tag{7}$$

In practice, the variance σ_u^2 of the area effects u_d is unknown and needs to be estimated from the data $\{(\hat{F}_{\alpha d}, x_d); d = 1, \ldots, D\}$. Fay and Herriot (1979) proposed a method-of-moments estimator that does not require normality assumption and it performed well in applications. Alternatively, assuming normality, maximum likelihood (ML) method or restricted ML (REML) method may be used. REML corrects for the degrees of freedom due to estimating β and leads to a less biased estimator of σ_u^2 for finite overall sample size n. Moreover, REML estimator of σ_u^2 remains asymptotically consistent under non-normality, under certain regularity conditions (Jiang, 1996).

Let $\hat{\sigma}_u^2$ denote the estimator of σ_u^2 obtained by any one of the above methods. Substituting $\hat{\sigma}_u^2$ for σ_u^2 in (7), we obtain the empirical BLUP (EBLUP) of the poverty measure $F_{\alpha d}$, denoted as $\hat{F}_{\alpha d}^{FH}$ and called hereafter as FH estimator. It follows from (6) that the FH estimator gives more weight to the synthetic estimator $x_d'\hat{\beta}$ as $\hat{\gamma}_d$ decreases or as the sampling variance ψ_d increases. Thus, the FH estimator automatically borrows strength through the synthetic estimator $x_d'\hat{\beta}$ for the areas where the direct estimator is not reliable. Molina and Morales (2009) provide an approximately unbiased estimator of the mean squared error (MSE) of the FH estimator of $F_{\alpha d}$ in the case of REML estimator of σ_u^2, following Datta and Lahiri (2000).

The FH estimator accounts for informative sampling within areas through the use of direct estimators as inputs. It is design-consistent as the area-specific sample size increases. Also, it requires only area-level covariates and thus avoids any confidentiality issues associated with micro (or unit level) data. However, in the context of estimating poverty indicators, available area level covariates may not adequately explain the variation in the poverty indicators across areas. For example, Molina and Morales (2009) applied the FH estimator to data from the 2006 Spanish Survey of Income and Living Conditions (SILC) to estimate poverty incidences and poverty gaps in Spanish provinces. They compared the values of the MSE estimator of the FH estimator to the sampling variance of the direct estimator and found overall gain in precision but the efficiency gains are only modest. To improve the efficiency gains, various extensions of the basic area level model have been proposed, including multivariate FH models (Benavent and Morales 2016) and area level time models that can borrow strength across both areas and time (Esteban *et al.* 2012).

5 ELL Method

As noted earlier, the ELL method is widely used by the World Bank to estimate poverty measures for specified small areas in developing countries. It is based on simulating multiple censuses of the desired welfare variable, calculating the poverty measures for specified small areas from each simulated census and then taking the averages over the censuses as the ELL estimates. To implement the ELL method, we need the following data: (i) unit level auxiliary variables from recent census or other administrative sources and (ii) survey data on the welfare variable of interest and the same census auxiliary variables, obtained from a two-stage sample with clusters as first stage units and households as second stage units. The welfare variable E_{kl} is first transformed to $y_{kl} = \log(E_{kl})$ and a nested error linear regression model is assumed for the sample data $\{(y_{kl}, x_{kl}), l \in s_k; k \in s\}$, where s is the sample of clusters and s_k is the sample of units l in the sample cluster k.

The nested error linear regression model assumed in the ELL method is given by

$$y_{kl} = x'_{kl}\beta + u_k + v_{kl}, l \in s_k; k \in s, \tag{8}$$

where the cluster effects u_k are assumed to be independent random variables with mean 0 and common variance σ_u^2, and independent of the unit errors v_{kl}, which are also assumed to be independent with mean 0 and possibly unequal variances. ELL method uses ordinary least squares (OLS) to estimate the residuals $\varepsilon_{kl} = u_k + v_{kj}$ as $\hat{\varepsilon}_{kl} = y_{kl} - x'_{kl}\hat{\beta}_{OLS}$ and then split the estimated residuals into components $\hat{u}_k = \hat{\varepsilon}_{k+} = n_k^{-1}\sum_{l \varepsilon s_k} \hat{\varepsilon}_{kl}$ and $\hat{v}_{kl} = \hat{\varepsilon}_{kl} - \varepsilon_{k+}$, where n_k the number of sample units in cluster is k. A simulated census of the welfare variable corresponding to the auxiliary variable x_t in the finite population U is generated as $\{E_t^* = \exp(y_t^*), t \in U\}$ by first randomly drawing values u_t^* and v_t^* with replacement from the fitted sample values $\{\hat{u}_k, \mathrm{k} \in s\}$ and $\{\hat{v}_{kl}, l \in s_k; k \in s\}$ respectively and then taking the simulated value of y_t as

$y_t^* = x_t'\hat{\beta}_{OLS} + u_t^* + v_t^*$. In the original ELL method, the OLS estimator of β is replaced by β^* selected randomly from a normal distribution with mean $\hat{\beta}_{OLS}$ and covariance matrix equal to the estimated covariance matrix of the OLS estimator, but this seems to be not necessary. From the simulated census values $\{y_t^*, t \in U\}$, the poverty measure $F_{\alpha d}$ for any desired small area or domain d may be readily computed and we denote it as $F_{\alpha d}^*$. Note that it is not necessary to specify the domains in advance of generating the simulated census because domain effects are not included in the ELL model (8) unless domains coincide with clusters.

The process of generating a simulated census is repeated a large number of times, A, to obtain A simulated values $\{F_{\alpha d}^{*(a)}; a = 1, ..., A\}$ } of the poverty measure $F_{\alpha d}$ for the domain d. The ELL estimator of $F_{\alpha d}$ is then taken as

$$\hat{F}_{\alpha d}^{ELL} = A^{-1} \sum_{a=1}^{A} F_{\alpha d}^{*(a)} . \tag{9}$$

Note that the ELL method implicitly assumes that the model (8) fitted to the sample data also holds for the population. This assumption is valid under non-informative sampling or absence of sample selection bias.

The ELL estimator of mean squared error (MSE) of $\hat{F}_{\alpha d}^{ELL}$ is simply obtained as

$$\mathrm{mse}(\hat{F}_{\alpha d}^{ELL}) = A^{-1} \sum_{a=1}^{A} (\hat{F}_{\alpha d}^{*(a)} - \hat{F}_{\alpha d}^{ELL})^2 \tag{10}$$

Das and Chambers (2017) noted that the MSE estimator (10) can lead to significant underestimation when the between-area variability is significant and it is not adequately accounted by the covariates used in the model (8) for the sample data. One way to remedy this problem is to include area-level contextual variables in the vector of covariates, if such variables are available both in the sample and in the census. In this case, areas need to be specified in advance of generating simulated censuses. If a different set of area estimates are needed, then different simulated censuses need to be generated by including area level covariates corresponding to those areas.

ELL method can be implemented without linking the sample to the census because a simulated census generates values for all the population units including the sampled units. This is an attractive feature of the method. In the application of the ELL method to developing countries, the number of areas represented in the two-stage sample is typically small compared with the number of areas in the population which means that, for most of the areas in the population, no sample observations are available. In other applications, especially for poverty estimation in European countries, many small areas are represented in the sample and hence methods that take account of the sample data (Section 6 and 7) lead to significantly more efficient estimators for sampled areas than the ELL method. For the non-sampled areas, both ELL and the other methods use regression-type synthetic estimators. In the U.S. SAIPE program, all the counties (small areas) are sampled and the area level FH model is used on direct county estimators.

Bilton *et al.* (2017) relaxed the assumption of linear regression in the ELL model (8) by using classification tree models, which can automatically select predictor variables and readily incorporate interactions between predictor variable by selecting only the important combinations. Molina and Rao (2010) considered a two-stage sampling design with areas as primary sampling units and the sample data is given by $\{(y_{ij}, x_{ij}), j=1,...,n_i; i=1,...,m\}$, where m is the number of sampled areas, n_i is the number of units sampled from sampled area i, $y_{ij} = \log(E_{ij})$ is the transformed welfare variable for the sampled unit j in the sampled area i. Following Molina and Rao (2010) and other papers on unit level models, we denote area by i instead of d and unit within area i by j instead of i used earlier for the area level model (section 4). They used a nested error linear regression model on y_{ij} incorporating random area effects v_i to account for between-area variability not explained by the predictor variables x_{ij}. The model for the sample data is given by

$$y_{ij} = x_{ij}'\beta + v_i + e_{ij}, \; j=1,...,n_i; i=1,...,m, \tag{11}$$

where the random area effects v_i are iid $N(0,\sigma_v^2)$ and independent of the unit errors e_{ij}, which are assumed to be iid $N(0,\sigma_e^2)$ and x_{ij} is a p-vector of covariates including the intercept term. Sampling is assumed to be non-informative at both stages of sampling so that the sample model (11) also holds for the population data $\{(y_{ij}, x_{ij}), j=1,...,N_i; i=1,...,M\}$, where M is the number of areas in the population and N_i is the number of units in the population area i. Under the above set-up, they showed that the EB estimators of the FGT poverty indicators $F_{\alpha i}$ for the sampled areas lead to large gains in efficiency over the corresponding ELL estimators. In fact, the ELL estimators can be less efficient than the corresponding direct estimators $\hat{F}_{\alpha i}$. On the other hand, for the non-sampled areas, the two estimators are comparable in terms of efficiency.

The traditional ELL method applied to (11) is not dependent on the normality assumption, unlike the method studied in Molina and Rao (2010). However, it does not provide a correct prediction of the area-specific effect v_i across the A simulated censuses, because each drawn a selects different values $v_i^{*(a)}$ from the empirical area level residuals $(\hat{v}_1,...,\hat{v}_m)$. As a result, the traditional ELL method uses a combination of estimated random effects from other areas to predict the poverty indicator for a specific area i and leads to loss in efficiency. Diallo and Rao (2018) proposed a modification of the original method to reduce the MSE. This modification retains the estimated area-specific effect $\hat{v}_i$ for area i in constructing the simulated census values $y_{ij}^{*(a)}$ as $y_{ij}^{*(a)} = x_{ij}'\hat{\beta} + \hat{v}_i + e_{ij}^{*(a)}$, where the unit residuals $e_{ij}^{*(a)}$ are drawn from the empirical distribution of the estimated residuals $\hat{e}_{ij}$. Diallo and Rao (2018) conducted a simulation study to demonstrate that the modified ELL method leads to large reduction in MSE related to the traditional ELL method for sampled areas. Note that the modified ELL method also does not require normality assumption.

6 EB Method under Unit Level Model

Molina and Rao (2010) studied EB estimation of FGT poverty indicators written as $F_{\alpha i}=\sum_{j\in U(i)}F_{\alpha ij}=\sum_{j\in U(i)}h_\alpha(y_{ij})$, where $U(i)$ denotes the set of population units in area $i=1,\ldots,M$. The vector of population values y_{ij} in area i is partitioned into the vector of sampled values $y_{ij},\, j\in s(i)$ and the vector of non-sampled values $y_{ij},\, j\in r(i)$, where $s(i)$ and $r(i)$ denote the set of sampled units and the set of non-sampled units, respectively. The nested error regression model (11) is assumed to hold for both the sampled units and the non-sampled units, which is equivalent to assuming non-informative sampling.

The best estimator of $h_\alpha(y_{ij})$ for $j\in r(i)$ is obtained as its expectation with respect to the conditional distribution of y_{ij} given the vector of sampled values in sampled area i. A closed form expression for the best predictor does not exist, but it can be approximated by a Monte Carlo approximation, by generating values $y_{ij}^{(l)}, l=1,\ldots,L$ from the conditional distribution. Under normality of the random area effects v_i and the unit errors e_{ij} in the nested error regression model (11), the desired values $y_{ij}^{(l)}$ can be generated from a univariate normal distribution; see Molina and Rao (2010) for details. The Monte Carlo approximation to the best estimator of $F_{\alpha ij}$ for $j\in r(i)$ is given by

$$\hat{F}_{\alpha ij}^{B}\approx L^{-1}\sum_{l=1}^{L}h_\alpha(y_{ij}^{(l)}),\ j\in r(i)\,. \tag{12}$$

The best estimator (12) depends on the unknown parameters $(\beta,\sigma_v^2,\sigma_e^2)$ of the model. Replacing the model parameters by suitable estimators in (12), such as restricted maximum likelihood (REML) estimators, leads to the empirical best (EB) estimator, denoted by $\hat{F}_{\alpha ij}^{EB}$. The resulting EB estimator of $F_{\alpha i}$ for sampled area i is given by

$$\hat{F}_{\alpha i}^{EB}=N_i^{-1}\Big(\sum_{j\in s(i)}F_{\alpha ij}+\sum_{j\in r(i)}\hat{F}_{\alpha ij}^{EB}\Big)\ , \tag{13}$$

where $F_{\alpha ij}$ for $j\in s(i)$ is computed from the sampled data $y_{ij},\, j\in s(i)$. Molina and Rao (2010) showed in a simulation study that the EB estimator (13) leads to large gains in efficiency over the ELL estimator for sampled areas under the present set up for populations exhibiting significant area effects in the model. For the non-sampled areas, Molina and Rao (2010) proposed a synthetic estimator of the FGT poverty indicator similar to the ELL estimator. They also obtained a proper MSE estimator through parametric bootstrap method by re-estimating the model parameters from each of the L simulated data sets, unlike the ELL estimator of MSE given by (10).

In the case of complex parameters, such as the FGT poverty indicators, analytical approximations to the MSE of the EB estimator are hard to derive. Molina and Rao (2010) proposed a parametric bootstrap-based MSE estimator of the EB estimator, under the normality assumption. It performed well in empirical studies in tracking the true MSE of the EB estimator.

The EB estimator (13) requires the linking of the sample to the population to identify the non-sampled units $r(i)$, unlike the ELL estimator. If the linking is not possible, then we can obtain the EB predictors $\hat{F}_{\alpha ij}^{EB}$ for all $j \in U(i)$ and use the Census EB estimator (Correa *et al.* 2012) of $F_{\alpha i}$ given by

$$\hat{F}_{\alpha i}^{CEB} = N_i^{-1} \sum_{j \in U(i)} \hat{F}_{\alpha ij}^{EB} \quad . \tag{14}$$

The Census EB estimator is less efficient than the EB estimator when the linking of the sample to the population is feasible, but the loss in efficiency is small if the area sampling fraction n_i / N_i is small.

In some applications, we may be interested in producing EB estimates of poverty indicators for both areas and subareas nested within areas. For example, in the Valencia region of Spain provinces are areas and *comarcas* (which are similar to counties) within areas are subareas. Marhuenda *et al.* (2017) proposed a two-fold nested error model to handle this case and obtained EB estimators for both provinces and sampled *comarcas* as well as non-sampled *comarcas*. This model includes random effects to explain the heterogeneity at the two levels of aggregation. Compared to EB estimators based on a one-fold model with only random subarea effects, the EB estimators under the two-fold model lead to considerable gain in efficiency for non-sample subareas by taking advantage of the estimated area effect corresponding to the subareas.

The EB method of Molina and Rao (2010) does not take account of the sampling design by assuming non-informative sampling within areas. Guadarrama *et al.* (2017) developed a variant of the EB estimator, called pseudo-EB estimator, which can provide protection against informative sampling.

The EB method has also been developed for the specific case of skewed response variable, when even after transformation of the welfare variable we do not achieve normality. Diallo and Rao (2018) developed EB estimators and simplified EB estimators under a unit level model with skew normal random effects v_i and skew normal unit errors e_{ij}. They showed that the normality-based EB estimators perform well in terms of efficiency when only the area random effects are skew-normal. On the other hand, if the unit errors are skew normal, considerable loss in efficiency results from the normality-based estimators. Van der Weide and Elbers (2013) studied normal mixture models on the area effects v_i, but assuming normality on the unit errors e_{ij} and their results are in agreement with Diallo and Rao (2018) in the sense that the normality-based EB method of Molina and Rao (2010) is robust provided the unit errors remain normal. Graf *et al.* (2018) developed estimators by modeling the welfare variable using a Generalized Beta distribution of the Second Kind (GB2).

7 HB Method under Unit Level Model

Computation of EB (or Census-EB) estimates supplemented with their bootstrap MSE estimates is very intensive and might not be feasible for very large populations or for very complex poverty indicators. Note that to approximate the EB estimate by Monte Carlo, we need to construct a large number A of simulated censuses, where each one might be of a huge size. Moreover, to obtain the parametric bootstrap MSE estimates, the Monte Carlo approximation needs to be repeated for each bootstrap replicate. Seeking for a computationally more efficient approach, Molina *et al.* (2014) developed a hierarchical Bayes (HB) method for estimating complex non-linear parameters. This approach does not require the use of bootstrap for MSE estimation because it provides samples from the posterior distribution, from which posterior variances (playing the role of MSEs) and any other useful posterior summary, such as credible intervals, can be easily obtained.

The HB method is based on reparametrizing the unit level nested error regression model (11) in terms of the intra-class correlation coefficient $\rho = \sigma_v^2 / (\sigma_v^2 + \sigma_e^2)$ and considering non-informative prior distributions for the resulting model parameters (β, ρ, σ_e^2) that reflect lack of knowledge. In particular, the joint prior distribution is taken as

$$\pi(\beta, \rho, \sigma_e^2) \propto \sigma_e^{-2}, \varepsilon \le \rho \le 1-\varepsilon, \sigma_e^2 > 0, \beta \in R^p, \qquad (15)$$

where $\varepsilon > 0$ is chosen very small to reflect lack of knowledge. Molina *et al.* (2014) showed that posterior inferences are not sensitive to a small change of ε.

Under the HB approach, the vector of random effects $v = (v_1, ..., v_m)'$ is regarded as an additional model parameter. Then, the joint posterior of the model parameters v, β, σ_e^2 and ρ given all the sample data, y_s, may be expressed as the product of (a) conditional posterior of v given β, σ_e^2 and ρ, (b) conditional posterior of β given σ_e^2 and ρ, (c) conditional posterior of σ_e^2 given ρ and (d) the posterior of ρ. The conditional posteriors (a), (b) and (c) have known explicit forms, but not the conditional posterior (d). Samples can be generated directly from the joint posterior distribution avoiding the use of Markov Chain Monte Carlo (MCMC) methods.

The HB estimator of the poverty measure $F_{\alpha i}$ is given by its conditional expectation with respect to the joint posterior predictive distribution of the unobserved y_{ij} for $j \in r(i)$ given y_s. It can be approximated by Monte Carlo as follows: First generate a value $\rho^{(a)}$ from the posterior (d) using a grid method, then a value $\sigma_e^{2(a)}$ from the conditional posterior (c) by letting $\rho = \rho^{(a)}$, next a value $\beta^{(a)}$ from the conditional posterior (b) by letting $\sigma_e^2 = \sigma_e^{2(a)}$ and $\rho = \rho^{(a)}$, and finally a value $v^{(a)}$ from the conditional posterior (a) by letting $\beta = \beta^{(a)}, \sigma_e^2 = \sigma_e^{2(a)}$ and $\rho = \rho^{(a)}$. Now noting that the population value $\{y_{ij}, j \in U(i), i = 1, ..., M\}$ given the model parameters v, β, σ_e^2 and ρ are independent and distributed as $N(x_{ij}'\beta + v_i, \sigma_e^2)$, we generate out of sample values $\{y_{ij}^{(a)}, j \in r(i)\}$ by letting $v_i = v_i^{(a)}, \beta = \beta^{(a)}$ and $\sigma_e^2 = \sigma_e^{2(a)}$ in the conditional distribution of $y_{ij}, j \in r(i)$. We also have the available sample data $\{y_{ij}, j \in s(i)\}$. Putting the sample data and the generated out of sample values, we construct the full population vector $\{y_{ij}, j \in s(i); y_{ij}^{(a)}, j \in r(i)\}$ for each area i. Using the constructed population data, we compute the parameter of interest $F_{\alpha i}$ as

$$F_{\alpha i}^{(a)} = N_i^{-1}\{\sum_{j\in s(i)} h_\alpha(y_{ij}) + \sum_{j\in r(i)} h_\alpha(y_{ij}^{(a)})\}. \tag{16}$$

We repeat the above process a large number of times, A and obtain the values $F_{\alpha i}^{(1)},...,F_{\alpha i}^{(A)}$ from the posterior distribution of $F_{\alpha i}$. The HB estimator is the posterior mean and approximated by

$$\hat{F}_{\alpha i}^{HB} \simeq A^{-1}\sum_{a=1}^{A} F_{\alpha i}^{(a)} . \tag{17}$$

Posterior variance, computed from the generated values $F_{\alpha i}^{(1)},...,F_{\alpha i}^{(A)}$, plays the role of MSE in the Bayesian set up. Other measures, such as credible intervals, can also be computed from the generated values.

8 Concluding Remarks

In this review paper, we presented a brief account of model based methods for small area estimation of complex poverty measures, in particular focusing on the ELL, EB and HB methods. Alternative approaches to estimation of small area poverty measures, based on the M-quantile method of Chambers and Tzavidis (2006), are reviewed in Tzavidis *et al.* (2008). The book edited by Pratesi (2016) contains a collection of articles on small area estimation of poverty measures and provides a “comprehensive guide to implementing SAE methods for poverty studies and poverty mapping”.

References

Benavent, R. and Morales, D. (2016). Multivariate Fay-Herriot models for small area estimation. *Computation Statistics and Data Analysis*, **94**, 372-390.

Biggeri, L., Giusti, C., Marchetti, S. and Pratesi, M. (2018). Poverty indicators at local level: Definitions, comparisons in real terms and small area estimation methods. *Statistics and Applications*, **16**, 351-364.

Bilton, P., Jones, G., Ganesh, S. and Haslett, S. (2017). Classification trees for poverty mapping. *Computational Statistics and Data Analysis*, **115**, 53-66.

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

Correa, L., Molina, I. and Rao, J.N.K. (2012). *Comparison of Methods for Estimation of Poverty Indicators in Small Areas*. Unpublished Report.

Das, S. and Chambers, R.L. (2017). Robust mean-squared error estimation for poverty estimates based on the method of Elbers, Lanjouw and Lanjouw. *Journal of the Royal Statistical Society*, **A180**, 1137-1161.

Datta, G. and Lahiri, P. (2000). A unified theory of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, **10**, 613-627.

Diallo, M.S. and Rao, J.N.K. (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors. *Scandinavian Journal of Statistics*, **45**, 1092-1116.

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

Esteban, M. D., Morales, D., Perez, A. and Santamaria, L. (2012). Two area-level time models for estimating small area poverty indictors. *Journal of the Indian Society of Agricultural Statistics*, **66**, 75-89.

Fay, R. E. and Herriot, R. (1979). Estimates of income for small places: An application of the James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269-277.

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

Graf, M., Martin, J. M. and Molina, I. (2018). *Estimation of Poverty Indicators in Small Areas Under Skewed Distributions*. Unpublished Report.

Guadarrama, M., Molina, I. and Rao, J. N. K. (2018). Small area estimation of general parameters under complex sampling designs. *Computational Statistics and Data Analysis*, **121**, 20-40.

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *Annals of Statistics*, **24**, 255-286.

Marhuenda, Y., Molina, I., Morales, D. and Rao, J.N.K. (2017). Poverty mapping in small areas under a twofold nested error regression model. *Journal of the Royal Statistical Society*, **A180**, 1111-1136.

Molina, I. and Morales, D. (2009). Small area estimation of poverty indicators. *Boletin de Estadistica e Investigacion Operativa*, **25**, 318-325.

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

Molina, I., Nandram, B. and Rao, J.N.K. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *Annals of Applied Statistics*, **8**, 852-885.

Neri, L., Ballini, P. and Betti, G. (2005). Poverty and inequality in transition countries. *Statistics in Transition*, **7**, 135-157.

Pratesi, M. (2016). Analysis of poverty data by small area estimation. Hoboken: Wiley.

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation* (Second Edition). Hoboken, New Jersey: Wiley.

Tzavidis, N., Salvati, N., Pratesi, M. and Chambers, R. (2008). M-quantile models with application to poverty mapping. *Statistical Methods and Applications*, **17**, 393-411.

Van der Weide, R. and Elbers, C. (2013). *Estimation of Normal Mixtures in a Nested Error Model with an Application to Small Area Estimation of Welfare*. Presented at the SAE Conference 2013, Bangkok, Thailand.