A review of small area estimation methods for poverty mapping

Isabel Molina

Department of Statistics, Universidad Carlos III de Madrid
C/Madrid 126, 28903 Getafe, Madrid
isabel.molinauc3m.es

María Guadarrama

Luxembourg Institute of Socio-Economic Research (LISER), Luxemburg

J.N.K. Rao

School of Mathematics and Statistics, Carleton University, Ottawa

Résumé. L'étude des indicateurs de pauvreté pour les petits domaines est d'intérêt grandissant ce dernier temps. Ces indicateurs aident les gouvernements et les organisations internationales à planifier et gérer plus facilement les mesures à l'échelle régionale en assurant une meilleure réponse vers les populations qui ont vraiment besoin. Plusieurs méthodes ont été proposées ce dernier temps pour étudier ces indicateurs pour les petits domaines. Ces méthodes donnent en général des meilleurs résultats que les méthodes traditionnelles employées par la Banque Mondiale. Dans cette présentation, nous allons présenter plusieurs méthodes pour l'étude d'indicateurs de pauvreté pour les petits domaines, dont celles basées sur les modèles au niveau domaine et utilisées par U.S. Census Bureau ainsi que les méthodes basées sur les modèles au niveau unité utilisées par la Banque Mondiale. Nous allons aussi discuter quelques modifications des procédures de base qui peuvent être utilisées pour traiter le cas d'un sondage non-informatif ou à deux degrés. Nous allons présenter aussi les avantages et inconvénients de ces méthodes d'un point de vue pratique et théorique.

Mots-clés. Bayes hiérarchique, indicateur de pauvreté, meilleur estimateur empirique, modèle niveau unité, modèle niveau domaine, paramètres non-linéaires.

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 policies, directing them to the actual places or population subgroups where they are more urgently needed. After the traditional 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 models and used by the U.S. Census Bureau and methods based on unit level models such as the traditional method used by the Word Bank. We will also discuss some variations of the basic procedures that can be used to deal with certain situations such as informative sampling or two stage sampling. We will discuss the pros and cons of these methods from a practical point of view, but based on the theory that is known for them.

Keywords. Area level model; Non-linear parameters; Empirical best estimator; Hierarchical Bayes; Poverty mapping; Unit level models.

1 Introduction

Poverty maps are an important source of information on the regional distribution of poverty and are currently used to support regional policy making and to allocate funds to local jurisdictions. Good examples are the poverty and inequality maps produced by the World Bank for many countries all over the world. In the U.S., the Small Area Income and Poverty Estimates (SAIPE) program (http://www.census.gov/hhes/www/saipe) of the Census Bureau provides annual estimates of income and poverty statistics for all school districts, counties, and states, for the administration of federal, state and local programs and the allocation of federal funds to local jurisdictions. In Europe, several efforts have been done to create regional databases and associated maps of poverty and social exclusion indicators in order to support regional development policies, see e.g. the joint project èèPoverty 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 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 municipality level.

Obtaining accurate poverty maps at high levels of disaggregation is not straightforward because of insufficient sample size of official surveys in some of the target regions. Direct estimates, obtained with the region-specific sample data, are unstable in the sense of having very large sampling errors for regions with small sample size. Here we review the main methods for the estimation of general non-linear small area parameters, focusing for illustrative purposes on a specific family of poverty indicators. Specifically, we describe direct estimation, the EBLUP based on the Fay-Herriot area level model (Fay and Herriot, 1979), the method of Elbers, Lanjouw and Lanjouw (2003), the empirical Best/Bayes (EB) method of Molina and Rao (2010) together with its variation called Census EB, the hierarchical Bayes (HB) method of Molina, Nandram and Rao (2014), and other variants of the EB method to deal two-stage sampling or informative sampling. We put ourselves in the point of view of a practitioner and discuss, as objectively as possible, the benefits and drawbacks of each method.

2 Poverty indicators

In this paper, for illustration purposes we will focus on the FGT family of poverty indicators introduced by Foster, Greer and Thorbecke (1984), although most of the methods that will be presented can be applied to general poverty or inequality indicators.

Consider a population P of size N that is partitioned into D domains or areas $P_1, \ldots, P_D$, of sizes $N_1, \ldots, N_D$. Let E_{di} be a measure of welfare for individual i ($i = 1, \ldots, N_d$) in area d ($d = 1, \ldots, D$). Let z be the poverty line, that is, the value such that when $E_{di} < z$, individual i from area d is regarded as èèat risk of poverty”. Then, the FGT family of poverty indicators for area d is given by

$$F_{\alpha d} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{z - E_{di}}{z} \right)^{\alpha} I(E_{di} < z), \quad \alpha \geq 0,\, d = 1, \ldots, D, \tag{1}$$

where $I(E_{di} < z) = 1$ if $E_{di} < z$, and $I(E_{di} < z) = 0$ otherwise. For $\alpha = 0$ we obtain the proportion of individuals èèat risk of poverty”, that is, the poverty incidence or at-risk-of-poverty rate. For $\alpha = 1$, we get the average of the relative distances to not being èèat risk of poverty”, called the poverty gap. The poverty incidence measures the frequency of poverty, whereas the poverty gap measures the intensity of poverty.

3 Direct estimators

Let s be a sample drawn from the population P. We denote by $s_d = s \cap P_d$ the subsample from area d of size $n_d < N_d$ and by $r_d = P_d - s_d$ the complement of s_d, of size $N_d - n_d$. The overall sample size is $n = n_1 + \cdots + n_D$.

If we wish to estimate a given characteristic in a domain or area, a direct estimator is that estimator obtained using only the observations from that area, provided that this area has been sampled (i.e., with strictly positive sample size). The FGT poverty indicator (1) of order α for area d can be expressed as a linear parameter as follows

$$F_{\alpha d} = N_d^{-1} \sum_{i=1}^{N_d} F_{\alpha di}, \quad F_{\alpha di} = \left(\frac{z - E_{di}}{z} \right)^{\alpha} I(E_{di} < z), \quad i = 1, \ldots, N_d.$$

Then, the basic direct estimator of $F_{\alpha d}$ is simply given by

$$\hat{F}_{\alpha d}^{\mathrm{DIR}} = N_d^{-1} \sum_{i \in s_d} w_{d,i} F_{\alpha di}, \tag{2}$$

where $w_{d,i}$ is the survey weight of unit i from area d.

4 Fay-Herriot model

The Fay-Herriot (FH) area level model, introduced by Fay and Herriot (1979), links the parameters of interest for all the areas, $F_{\alpha d}$, $d = 1, \ldots, D$, through a linear model as

$$F_{\alpha d} = \mathbf{x}_d' \boldsymbol{\beta} + u_d, \quad d = 1, \ldots, D. \tag{3}$$

Here, $\mathbf{x}_d$ is a p-vector of area level covariates, $\boldsymbol{\beta}$ is the regression parameter common for all areas, and u_d is the area-specific regression error, also called random effect for area d. We assume that area random effects u_d are independent and identically distributed (iid), with unknown variance σ_u^2, that is, $u_d \stackrel{iid}{\sim} (0, \sigma_u^2)$. Note that true values $F_{\alpha d}$ are not observable and therefore model (3) cannot be directly fitted. However, we can make use of a direct estimator $\hat{F}_{\alpha d}^{\text{DIR}}$ of $F_{\alpha d}$. FH model assumes that $\hat{F}_{\alpha d}^{\text{DIR}}$ is design-unbiased, with

$$\hat{F}_{\alpha d}^{\text{DIR}} = F_{\alpha d} + e_d, \quad d = 1, \ldots, D, \tag{4}$$

where e_d is the sampling error for domain d. We assume that sampling errors e_d are independent of random effects u_d and satisfy $e_d \stackrel{ind}{\sim} (0, \psi_d)$, where the sampling variances ψ_d, $d = 1, \ldots, D$, are assumed to be known. Combining (3) and (4), we obtain a linear mixed model

$$\hat{F}_{\alpha d}^{\text{DIR}} = \mathbf{x}_d' \boldsymbol{\beta} + u_d + e_d, \quad d = 1, \ldots, D. \tag{5}$$

The best linear unbiased predictor (BLUP) of $F_{\alpha d} = \mathbf{x}_d' \boldsymbol{\beta} + u_d$ under model (5) is given by

$$\tilde{F}_{\alpha d}^{\text{FH}} = \mathbf{x}_d' \tilde{\boldsymbol{\beta}} + \tilde{u}_d, \tag{6}$$

where $\tilde{u}_d = \gamma_d (\hat{F}_{\alpha d}^{\text{DIR}} - \mathbf{x}_d' \tilde{\boldsymbol{\beta}})$ is the BLUP of u_d, with $\gamma_d = \sigma_u^2 / (\sigma_u^2 + \psi_d)$ and where $\tilde{\boldsymbol{\beta}}$ is the weighted least squares estimator of $\boldsymbol{\beta}$, given by

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{d=1}^{D} \gamma_d \mathbf{x}_d \mathbf{x}_d' \right)^{-1} \sum_{d=1}^{D} \gamma_d \mathbf{x}_d \hat{F}_{\alpha d}^{\text{DIR}}.$$

In practice, the variance σ_u^2 of the area effects u_d is unknown and needs to be estimated. Common estimation methods are maximum likelihood (ML) and restricted maximum likelihood (REML). REML corrects for the degrees of freedom due to estimating $\boldsymbol{\beta}$ and leads to a less biased estimator of σ_u^2 for finite sample size n. Let $\hat{\sigma}_u^2$ be the resulting estimator. Replacing $\hat{\sigma}_u^2$ for σ_u^2 in (6), we obtain the empirical BLUP (EBLUP) of $F_{\alpha d}$, denoted here as $\hat{F}_{\alpha d}^{\text{FH}}$ and called hereafter FH estimator.

A second-order correct estimator of MSE $(\hat{F}_{\alpha d}^{\text{FH}})$ is given in Rao (2003, Chapter 7), assuming normality of u_d and e_d. Good and bad properties of FH estimator (6) are listed below, including particular properties for poverty mapping.

5 ELL method

The method of Elbers, Lanjouw and Lanjouw (2003), called hereafter ELL method, assumes a unit level linear mixed model for a log-transformation of the variable measuring welfare of individuals, with random effects for the sampling clusters or primary sampling units. For comparability with the rest of the methods presented here, in the following we assume that the sampling clusters are the areas. In this case, the model becomes the nested error model of Battese, Harter and Fuller (1988) for the log-transformation of the welfare variables, that is, $Y_{di} = \log(E_{di})$ is assumed to be linearly related with a p-vector of auxiliary variables $\mathbf{x}_{di}$, which may include unit-specific and area-specific covariates, and includes random area effects u_d as follows

$$Y_{di} = \mathbf{x}_{di}'\boldsymbol{\beta} + u_d + e_{di}, \quad i = 1, \dots, N_d, \, d = 1, \dots, D. \tag{7}$$

Here, $\boldsymbol{\beta}$ is a p-vector of regression coefficients, $u_d \overset{iid}{\sim} (0, \sigma_u^2)$, $e_{di} \overset{ind}{\sim} (0, \sigma_e^2 k_{di}^2)$, where u_d and e_{di} are independent and k_{di} are known constants.

ELL estimator of $F_{\alpha d}$ is given by the marginal expectation $\hat{F}_{\alpha d}^{\text{ELL}} = E[F_{\alpha d}]$ under model (7). This estimator and its MSE are approximated by a bootstrap method. In this bootstrap procedure, random effects u_d^* and model errors e_{di}^* are generated from residuals obtained by fitting model (7) to survey data. Then, a bootstrap census of Y-values is generated as

$$Y_{di}^* = \mathbf{x}_{di}'\hat{\boldsymbol{\beta}} + u_d^* + e_{di}^*, \quad i = 1, \dots, N_d, \, d = 1, \dots, D,$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$. The generation is repeated for $a = 1, \dots, A$, obtaining A censuses. Then, for each bootstrap census a, the FGT poverty indicator for area d is calculated as

$$F_{\alpha d}^{*(a)} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{z - \exp(Y_{di}^{*(a)})}{z} \right)^{\alpha} I(\exp(Y_{di}^{*(a)}) < z).$$

The ELL estimator of $F_{\alpha d}$ is then approximated by averaging over the A generated censuses, that is,

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

The MSE of $\hat{F}_{\alpha d}^{\text{ELL}}$ is then estimated as follows

$$\text{mse}(\hat{F}_{\alpha d}^{\text{ELL}}) = \frac{1}{n_d} \sum_{a=1}^{A} (F_{\alpha d}^{*(a)} - \hat{F}_{\alpha d}^{\text{ELL}})^2.$$

6 Empirical Best/Bayes EB method

The empirical Best/Bayes (EB) method of Molina and Rao (2010) assumes that the population variables Y_{di} follow the nested error model (7) with normality of random effects u_d and errors e_{di}. Under that model, the area vectors $\mathbf{Y}_d = (Y_{d1}, \ldots, Y_{d,N_d})'$ are independent for $d = 1, \ldots, D$ and satisfy $\mathbf{Y}_d \stackrel{ind}{\sim} N(\boldsymbol{\mu}_d, \mathbf{V}_d)$, where $\boldsymbol{\mu}_d = \mathbf{X}_d\boldsymbol{\beta}$ and $\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{N_d}\mathbf{1}'_{N_d} + \sigma_e^2 \mathbf{A}_d$, for $\mathbf{A}_d = \text{diag}(k_{di}^2; i = 1, \ldots, N_d)$. For an area parameter $\delta_d = h(\mathbf{Y}_d)$, the estimator that minimizes the MSE, called best estimator, is given by

$$\hat{\delta}_d^{\text{B}} = E_{\mathbf{Y}_{dr}}[h(\mathbf{Y}_d)|\mathbf{Y}_{ds}; \boldsymbol{\theta}] = \int h(\mathbf{Y}_d) f(\mathbf{Y}_{dr}|\mathbf{Y}_{ds}; \boldsymbol{\theta}) d\mathbf{Y}_{dr}, \tag{8}$$

where $f(\mathbf{Y}_{dr}|\mathbf{Y}_{ds}; \boldsymbol{\theta})$ is the conditional distribution of the vector of out-of-sample values $\mathbf{Y}_{dr}$ in domain d given the sampled values $\mathbf{Y}_{ds}$ in that domain and $\boldsymbol{\theta}$ is the vector of model parameters. Now replacing $\boldsymbol{\theta}$ in (8) by an estimator $\hat{\boldsymbol{\theta}}$, we get the empirical best (EB) estimator, $\hat{\delta}_d^{\text{EB}}$.

Under the nested error model (7), the distribution of $\mathbf{Y}_{dr}|\mathbf{Y}_{ds}$ is easy to derive. First, we decompose $\mathbf{X}_d$ and $\mathbf{V}_d$ into sample and out-of-sample elements similarly as we do with $\mathbf{Y}_d$, that is,

$$\mathbf{Y}_d = \begin{pmatrix} \mathbf{Y}_{ds} \\ \mathbf{Y}_{dr} \end{pmatrix}, \quad \mathbf{X}_d = \begin{pmatrix} \mathbf{X}_{ds} \\ \mathbf{X}_{dr} \end{pmatrix}, \quad \mathbf{V}_d = \begin{pmatrix} \mathbf{V}_{ds} & \mathbf{V}_{dsr} \\ \mathbf{V}_{drs} & \mathbf{V}_{dr} \end{pmatrix}.$$

By the normality assumption, we have that $\mathbf{Y}_{dr}|\mathbf{Y}_{ds} \stackrel{ind}{\sim} N(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s})$, where the conditional mean vector and covariance matrix are given by

$$\boldsymbol{\mu}_{dr|s} = \mathbf{X}_{dr}\boldsymbol{\beta} + \gamma_{dc}(\bar{y}_{dc} - \bar{\mathbf{x}}_{dc}^T\boldsymbol{\beta})\mathbf{1}_{N_d-n_d}, \tag{9}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}_{N_d-n_d}^T + \sigma_e^2 \text{diag}_{i\in r_d}(k_{di}^2). \tag{10}$$

Here, $\gamma_{dc} = \sigma_u^2(\sigma_u^2 + \sigma_e^2/c_{d\cdot})^{-1}$, for $c_{d\cdot} = \sum_{i\in s_d} c_{di}$ with $c_{di} = k_{di}^{-2}$, and $\bar{y}_{dc}$ and $\bar{\mathbf{x}}_{dc}$ are weighted sample means obtained as

$$\bar{y}_{dc} = \frac{1}{c_{d\cdot}}\sum_{i\in s_d} c_{di}Y_{di}, \quad \bar{\mathbf{x}}_{dc} = \frac{1}{c_{d\cdot}}\sum_{i\in s_d} c_{di}\mathbf{x}_{di}. \tag{11}$$

For complex non-linear parameters $\delta_d = h(\mathbf{Y}_d)$, the expectation given in (8) cannot be calculated analytically. In those cases, the EB estimator $\hat{\delta}_d^{\text{EB}}$ is approximated by Monte Carlo. This requires to simulate multivariate Normal vectors $\mathbf{Y}_{dr}^{(a)}$ of sizes $N_d - n_d$, $d = 1, \ldots, D$, from the (estimated) conditional distribution of $\mathbf{Y}_{dr}|\mathbf{Y}_{ds}$ and then to replicate for $a = 1, \ldots, A$, which may be computationally unfeasible. Simulation of very large multivariate Normal vectors $\mathbf{Y}_{dr}^{(a)}$ can be avoided by noting that the conditional covariance matrix $\mathbf{V}_{dr|s}$, given by (10), corresponds to the covariance matrix of a random vector $\mathbf{Y}_{dr}^{(a)}$ generated from the model

$$\mathbf{Y}_{dr}^{(a)} = \boldsymbol{\mu}_{dr|s} + v_d^{(a)}\mathbf{1}_{N_d-n_d} + \boldsymbol{\epsilon}_{dr}^{(a)}, \tag{12}$$

where $v_d^{(a)}$ and $\boldsymbol{\epsilon}_{dr}^{(a)}$ are independent and satisfy

$$v_d^{(a)} \sim N(0, \sigma_u^2(1-\gamma_d)) \quad \text{and} \quad \boldsymbol{\epsilon}_{dr}^{(a)} \sim N(\mathbf{0}_{N_d-n_d}, \sigma_e^2 \text{diag}_{i\in r_d}(k_{di}^2));$$

see Molina and Rao (2010). Using model (12), instead of generating a multivariate normal vector $\mathbf{Y}_{dr}^{(a)}$ of size $N_d - n_d$, we just need to generate $1 + N_d - n_d$ independent univariate normal variables $v_d^{(a)} \overset{ind}{\sim} N(0, \sigma_u^2(1-\gamma_d))$ and $\epsilon_{di}^{(a)} \overset{ind}{\sim} N(0, \sigma_e^2 k_{di}^2)$, for $i \in r_d$. Then, we obtain the corresponding out-of-sample values $Y_{di}^{(a)}$, $i \in r_d$, from (12) using as means, the corresponding elements of $\boldsymbol{\mu}_{dr|s}$ given by (9). Using the vector $\mathbf{Y}_{dr}^{(a)}$ generated from (12), we construct the census vector $\mathbf{Y}_d^{(a)} = (\mathbf{Y}_{ds}', (\mathbf{Y}_{dr}^{(a)})')'$ and calculate the parameter of interest $\delta_d^{(a)} = h(\mathbf{Y}_d^{(a)})$. For a non-sampled area d (i.e., with $n_d = 0$), we generate $\mathbf{Y}_{dr}^{(a)}$ from (12) with $\gamma_{dc} = 0$ and in this case $\mathbf{Y}_d^{(a)} = \mathbf{Y}_{dr}^{(a)}$. The Monte Carlo approximation to the EB estimator (8) of $\delta_d = h(\mathbf{Y}_d)$ is then given by

$$\hat{\delta}_d^{\text{EB}} \approx \frac{1}{A}\sum_{a=1}^{A} h(\mathbf{Y}_d^{(a)}). \tag{13}$$

In particular, to estimate the FGT poverty indicator given in (1), Molina and Rao (2010) assumed that the transformed welfare variables $Y_{di} = T(E_{di})$ follow the nested error model (7), where $T(\cdot)$ is a one-to-one transformation. In terms of the vector of transformed variables, $\mathbf{Y}_d = (Y_{d1}, \ldots, Y_{dN_d})'$, the FGT poverty indicator for domain d can be expressed as

$$F_{\alpha d} = \frac{1}{N_d}\sum_{i=1}^{N_d}\left(\frac{z - T^{-1}(Y_{di})}{z}\right)^{\alpha} I(T^{-1}(Y_{di}) < z) = h_\alpha(\mathbf{Y}_d), \tag{14}$$

and the above EB method can be applied to the area parameter $\delta_d = h_\alpha(\mathbf{Y}_d)$.

In the case of complex parameters such as the FGT poverty indicators, analytic approximations for the MSE are hard to derive. Molina and Rao (2010) obtained a parametric boostrap MSE estimator following the bootstrap method for finite populations of González-Manteiga et al. (2008), see Molina and Rao (2010) for further details.

Note that both ELL and EB methods require a survey data file containing the observations from the target variable and the auxiliary variables, that is, $\{(Y_{di}, \mathbf{x}_{di}); i \in s_d,\ d = 1, \ldots, D\}$, and a census containing the values of the same auxiliary variables for all the units in the population, that is, $\{\mathbf{x}_{di}; i = 1, \ldots, N_d,\ d = 1, \ldots, D\}$. EB method requires additionally to identify the set of out-of-sample units r (or equivalently the sample units s) in the census P. Linking the survey and the census files is not always possible in practice. However, typically the area sample size n_d is really small compared to the population size N_d. Then, we can use the Census-EB estimator proposed by Correa, Molina and Rao (2012), and obtained by generating in each Monte Carlo replicate the full census vector

$\mathbf{Y}_d$ rather than only the vector of out-of-sample observations $\mathbf{Y}_{dr}$. For this, we apply the Monte Carlo approximation (8) by generating $\mathbf{Y}_d^{(a)} = \boldsymbol{\mu}_{d|s} + v_d^{(a)}\mathbf{1}_{N_d-n_d} + \boldsymbol{\epsilon}_d^{(a)}$, where $\boldsymbol{\mu}_{d|s} = \mathbf{X}_d\boldsymbol{\beta} + \gamma_{dc}(\bar{y}_{dc} - \bar{\mathbf{x}}_{dc}^T\boldsymbol{\beta})\mathbf{1}_{N_d}$ and $\boldsymbol{\epsilon}_d^{(a)} \sim N(\mathbf{0}_{N_d}, \sigma_e^2\text{diag}_{i=1,\ldots,N_d}(k_{di}^2))$. If the sampling fraction n_d/N_d is negligible, the Census-EB estimator of $\delta_d = F_{\alpha d}$ is practically the same as the original EB estimator.

Note that the EB method does not take into account the sampling design. Another extension of this EB estimator has been developed by Marhuenda et al. (2017) to deal with two-stage sampling or nested population structures. For unequal probability sampling, specially when the sampling is suspected to be informative, Guadarrama, Molina and Rao (2017) have proposed a variant of the EB estimator that provides protection against informative sampling. Finally, the EB method has been also developed for the specific case of skewed model responses, when even after transformation we do not achieve normality, see Diallo and Rao (2014) for the EB method based on the skew normal distribution and the procedure based on the Generalized Beta distribution of the Second Kind (GB2) by Graf, Marín and Molina (2018).

7 Hierarchical Bayes (HB) method

Computation of EB (and Census-EB) estimates supplemented with their MSE estimates is very intensive and might be unfeasible for very large populations or for very complex indicators. Note that to approximate the EB estimate by Monte Carlo, we need to construct a large number A of censuses $\mathbf{Y}^{(a)}$, where each one might be of huge size. Moreover, to obtain the parametric bootstrap MSE estimator, the Monte Carlo approximation needs to be repeated for each bootstrap replicate. Seeking for a computationally more efficient approach, Molina, Nandram and Rao (2014) developed the alternative hierarchical Bayes (HB) method for estimation of 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 play the role of MSEs, and any other useful posterior summary can be easily obtained.

The HB method is based on reparameterizing the nested error model (7) in terms of the intraclass correlation coefficient $\rho = \sigma_u^2/(\sigma_u^2+\sigma_e^2)$ and considering priors for the model parameters $(\boldsymbol{\beta}, \rho, \sigma_e^2)$ that reflect lack of knowledge. Concretely, the HB model is defined as

$$
\begin{aligned}
&\text{(i)} \quad && Y_{di}|u_d, \boldsymbol{\beta}, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}_{di}'\boldsymbol{\beta} + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,\\
&\text{(ii)} \quad && u_d|\rho, \sigma_e^2 \overset{iid}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma_e^2\right), \quad d = 1, \ldots, D,\\
&\text{(iii)} \quad && \pi(\boldsymbol{\beta}, \rho, \sigma_e^2) \propto \frac{1}{\sigma_e^2}, \quad \epsilon \le \rho \le 1-\epsilon,\ \sigma_e^2 > 0, \boldsymbol{\beta} \in \mathcal{R}^p,
\end{aligned}
$$

where $\epsilon > 0$ is chosen very small to reflect lack of knowledge. See the application carried out by Molina, Nandram and Rao (2014), where inference was not sensitive to a small change of ϵ.

The posterior distribution can be obtained in terms of posterior conditionals using the chain rule of probability as follows. First, note that under the HB approach, the random effects $\mathbf{u} = (u_1, \ldots, u_D)'$ are regarded as additional parameters. Then, the joint posterior pdf of the vector of parameters $\boldsymbol{\theta} = (\mathbf{u}', \boldsymbol{\beta}', \sigma_e^2, \rho)'$ given the sample values $\mathbf{Y}_s$ is given by

$$\pi(\mathbf{u}, \boldsymbol{\beta}, \sigma_e^2, \rho|\mathbf{Y}_s) = \pi_1(\mathbf{u}|\boldsymbol{\beta}, \sigma_e^2, \rho, \mathbf{Y}_s)\pi_2(\boldsymbol{\beta}|\sigma_e^2, \rho, \mathbf{Y}_s)\pi_3(\sigma_e^2|\rho, \mathbf{Y}_s)\pi_4(\rho|\mathbf{Y}_s), \tag{15}$$

where the conditional pdfs $\pi_1, \ldots, \pi_3$ have known forms, but not π_4. However, since ρ is in a closed interval from $(0, 1)$, we can generate values from π_4 using a grid method, for more details see Molina, Nandram and Rao (2014). Samples from $\boldsymbol{\theta} = (\mathbf{u}', \boldsymbol{\beta}', \sigma_e^2, \rho)'$ can then be generated directly from the posterior distribution in (15), avoiding the use of Markov Chain Monte Carlo (MCMC) methods. Under general conditions, a proper posterior distribution is guaranteed.

Given $\boldsymbol{\theta}$, population variables Y_{di} are all independent, satisfying

$$Y_{di}|\boldsymbol{\theta} \overset{ind}{\sim} N(\mathbf{x}_{di}'\boldsymbol{\beta} + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,\, d = 1, \ldots, D. \tag{16}$$

The posterior predictive density of $\mathbf{Y}_{dr}$ is then given by

$$f(\mathbf{Y}_{dr}|\mathbf{Y}_s) = \int \prod_{i \in r_d} f(Y_{di}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{Y}_s)d\boldsymbol{\theta}.$$

Finally, the HB estimator of a domain parameter $\delta_d = h(\mathbf{Y}_d)$ is given by

$$\hat{\delta}_d^{\mathrm{HB}} = E_{\mathbf{Y}_{dr}}(\delta_d|\mathbf{Y}_s) = \int h(\mathbf{Y}_d) f(\mathbf{Y}_{dr}|\mathbf{Y}_s)d\mathbf{Y}_{dr}. \tag{17}$$

The HB estimator can be approximated by Monte Carlo. For this, we first generate samples from the posterior $\pi(\boldsymbol{\theta}|\mathbf{Y}_s)$. We generate a value $\rho^{(a)}$ from $\pi_4(\rho|\mathbf{Y}_s)$ using a grid method; then, a value $\sigma_e^{2(a)}$ is generated from $\pi_3(\sigma_e^2|\rho^{(a)}, \mathbf{Y}_s)$; next $\boldsymbol{\beta}^{(a)}$ is generated from $\pi_2(\boldsymbol{\beta}|\sigma_e^{2(a)}, \rho^{(a)}, \mathbf{Y}_s)$ and, finally, $\mathbf{u}^{(a)}$ is generated from $\pi_1(\mathbf{u}|\boldsymbol{\beta}^{(a)}, \sigma_e^{2(a)}, \rho^{(a)}, \mathbf{Y}_s)$. This process is repeated a large number A of times to get a random sample $\boldsymbol{\theta}^{(a)}$, $a = 1, \ldots, A$ from $\pi(\boldsymbol{\theta}|\mathbf{Y}_s)$. Now for each generated value $\boldsymbol{\theta}^{(a)}$ from $\pi(\boldsymbol{\theta}|\mathbf{Y}_s)$, we generate the out-of-sample values $\{Y_{di}^{(a)}, i \in r_d\}$ from the distribution defined in (16). Thus, for each area d, we have generated an out-of-sample vector $\mathbf{Y}_{dr}^{(a)} = \{Y_{di}^{(a)}, i \in r_d\}$, and we have also the available sample data $\mathbf{Y}_{ds}$. Putting them together, we construct the full population vector $\mathbf{Y}_d^{(a)} = (\mathbf{Y}_{ds}', (\mathbf{Y}_{dr}^{(a)})')'$. Now using $\mathbf{Y}_d^{(a)}$, we compute the area parameter $\delta_d^{(a)} = h(\mathbf{Y}_d^{(a)})$. In the particular case of estimating an FGT poverty indicator, we have $\delta_d = F_{\alpha d} = h_\alpha(\mathbf{Y}_d)$ given in (14). Then, in Monte Carlo replicate a, we calculate $F_{\alpha d}^{(a)} = h_\alpha(\mathbf{Y}_d^{(a)})$. Finally, the HB estimator is approximated as

$$\hat{F}_{\alpha d}^{\mathrm{HB}} \approx \frac{1}{A} \sum_{a=1}^{A} F_{\alpha d}^{(a)}. \tag{18}$$

Bibliography

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, *Journal of American Statistical Association*, 83, pp. 28-36.

Correa, L., Molina, I. and Rao, J.N.K., (2012). Comparison of methods for estimation of poverty indicators in small areas. Unpublished report.

Diallo, M. and Rao, J.N.K., (2014). Small Area Estimation of Complex Parameters Under Unit-level Models with Skew-Normal Errors. Proceedings of the Survey Research Section, American Statistical Association.

Elbers, C., Lanjouw, J. O. and Lanjouw, P., (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, 71(1), pp. 355-364.

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

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

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L., (2008). Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78, pp. 443-462.

Graf, M., Marín, J.M. and Molina, I., (2018). Estimation of poverty indicators in small areas under skewed distributions, Unpublished manuscript.

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, pp. 2040.

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 A*, 180 (4), pp. 11111136.

Molina, I. and Rao, J.N.K., (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38, pp. 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. *The Annals of Applied Statistics*, 8(2), pp. 852-885.