STATISTICS IN TRANSITION new series and *SURVEY METHODOLOGY*
Joint Issue: Small Area Estimation 2014
Vol. 17, No. 1, pp. 41–66

A COMPARISON OF SMALL AREA ESTIMATION METHODS FOR POVERTY MAPPING

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

ABSTRACT

We review main small area estimation methods for the estimation of general non-linear parameters focusing on FGT family of poverty indicators introduced by Foster, Greer and Thorbecke (1984). In particular, we consider direct estimation, the Fay-Herriot area level model (Fay and Herriot, 1979), the method of Elbers, Lanjouw and Lanjouw (2003) used by the World Bank, the empirical Best/Bayes (EB) method of Molina and Rao (2010) and its extension, the Census EB, and finally the hierarchical Bayes proposal of Molina, Nandram and Rao (2014). We put ourselves in the point of view of a practitioner and discuss, as objectively as possible, the benefits and drawbacks of each method, illustrating some of them through simulation studies.

Key words: 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, the joint project "Poverty Mapping in the New Member States of the European Union" between the World Bank and the European Commission was aimed to construct poverty maps for the new members of the EU.

[1]Department of Statistics, Universidad Carlos III de Madrid. Address: C/Madrid 126, 28903 Getafe (Madrid), Spain, Tf: +34 916249859. E-mail: maria.guadarrama@uc3m.es

[2]Department of Statistics, Universidad Carlos III de Madrid. Address: C/Madrid 126, 28903 Getafe (Madrid), Spain, Tf: +34 916249887. E-mail: isabel.molina@uc3m.es

[3]School of Mathematics and Statistics, Carleton University. E-mail: jrao@math.carleton.ca

The TIPSE (The Territorial Dimension of Poverty and Social Exclusion in Europe) project, commissioned by the European Observation Network for Territorial Development and Cohesion (ESPON) program, aims to support policy by creating a regional database and associated maps of poverty and social exclusion indicators. 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. Very unstable poverty estimates might make the seemingly poorer regions in one period appear as the richer in the next period, which can be contradictory. On the other hand, very stable but biased estimates (e.g., too homogeneous across regions) might make identification of the poorer regions difficult.

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 introduced in Section 2. Specifically, in Section 3 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, and hierarchical Bayes (HB) method of Molina, Nandram and Rao (2014). We discuss advantages and disadvantages of each procedure from a practical point of view. In Section 4 we illustrate their performance in simulations under several scenarios, including the cases of informative sampling or the presence of outliers. Finally, in Section 5 we draw some conclusions.

2. Poverty indicators

In this paper, we will focus on the FGT family of poverty indicators introduced by Foster, Greer and Thorbecke (1984). 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 “at 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 “at 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 “at risk of poverty", called the poverty gap. The poverty incidence measures the frequency of poverty, whereas the poverty gap measures the intensity of poverty. We remark that the unit level methods introduced in this paper can be applied to estimate any desired population characteristic that is obtained as a real measurable function of a continuous variable, as long as this variable follows the considered model in each method.

3. Estimators

Estimation of population characteristics is typically based on a sample s 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 $n = n_1 + \cdots + n_D$. The following subsections describe common estimators of poverty indicators obtained from the sample data.

3.1. Direct estimators

Turning now to estimation in a given domain or area d, a direct estimator is an estimator obtained using only the n_d observations from that area, provided that this area has been sampled (i.e., $n_d > 0$). 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 $w_{d,i} = \pi_{d,i}^{-1}$ is the sampling weight of unit i from area d and $\pi_{d,i}$ is the inclusion probability of unit i in the subsample s_d.

Below we list the advantages and disadvantages of direct estimators, such as (2), for small area estimation.

Advantages:

- They are (at least approximately) design-unbiased and design-consistent (as

$n_d \to \infty$). Thus, they perform well under complex sampling designs, including informative sampling, as long as they are calculated using the correct inclusion probabilities.

- They do not require model assumptions; that is, they are completely nonparametric.

Disadvantages:

- They are very inefficient for areas with very small n_d.
- They cannot be calculated for nonsampled areas (i.e., with $n_d = 0$).

3.2. Fay-Herriot model

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

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

where $\mathbf{x}_d$ is a p-vector of area level covariates, β 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, $u_d \overset{iid}{\sim} (0, \sigma_u^2)$. 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 $\hat{F}_{\alpha d}^{\mathrm{DIR}}$ of $F_{\alpha d}$. FH model assumes that $\hat{F}_{\alpha d}^{\mathrm{DIR}}$ is design-unbiased, with

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

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 $e_d \overset{ind}{\sim} (0, \psi_d)$, where the sampling variances ψ_d, $d = 1,\dots,D$, are assumed to be known. Combining (3) and (4), we obtain a linear mixed model

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

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

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

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

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

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 β 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 $\hat{F}_{\alpha d}^{\mathrm{FH}}$ and called hereafter FH estimator.

A second-order correct estimator of MSE ($\hat{F}_{\alpha d}^{\mathrm{FH}}$) 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.

Advantages:

- The BLUP under FH model can be expressed as a weighted combination of the direct and the regression-synthetic estimators, that is,

$$\tilde{F}_{\alpha d}^{\mathrm{FH}} = \gamma_d \hat{F}_{\alpha d}^{\mathrm{DIR}} + (1-\gamma_d)\mathbf{x}_d'\tilde{\beta}, \quad d = 1, \ldots, D. \tag{7}$$

  with weight $\gamma_d = \sigma_u^2/(\sigma_u^2 + \psi_d)$. Then, for an area d in which the direct estimator $\hat{F}_{\alpha d}^{\mathrm{DIR}}$ is inefficient, that is, with a large sampling variance ψ_d compared to the unexplained between-area variability σ_u^2, γ_d becomes small and $\tilde{F}_{\alpha d}^{\mathrm{FH}}$ borrows more strength from the other areas through the regression-synthetic estimator $\mathbf{x}_d'\tilde{\beta}$. On the other hand, for an area d in which the direct estimator $\hat{F}_{\alpha d}^{\mathrm{DIR}}$ is efficient, that is, with small sampling variance ψ_d compared to the unexplained between-area variability σ_u^2, γ_d is large and $\tilde{F}_{\alpha d}^{\mathrm{FH}}$ attaches more weight to the direct estimator. Thus, FH estimator automatically borrows strength for the areas where it is needed.
- If $\gamma_d > 0$ for area d, it makes use of the sampling weights $w_{d,i}$ through the direct estimator $\hat{F}_{\alpha d}^{\mathrm{DIR}}$. Thus, it is design-consistent (as $n_d \to \infty$). As a consequence, it is less affected by informative sampling provided that the direct estimator is calculated using the correct inclusion probabilities.
- Due to the aggregation of data, it is not very much affected by isolated unit level outliers.
- It requires only area level auxiliary information and therefore avoids the confidentiality issues associated with micro-data.

Disadvantages:

- The sampling variances ψ_d are assumed to be known, but in practice they are estimated. It is not easy to incorporate the uncertainty due to estimation of the sampling variances in the MSE.
- The number of observations used to fit the FH model is the number of areas D, which is typically much smaller than the number of observations used to fit unit level models, n. Thus, model parameters are estimated with less efficiency and therefore the efficiency gains with respect to direct estimators are expected to be smaller than under unit level models.
- It requires normality of u_d and e_d for MSE estimation. This might not hold for very complex poverty indicators.
- If we want to estimate several indicators depending on a common continuous variable, it requires separate modeling and searching for good covariates for each indicator.
- Once the model is fitted at the area level, small area estimates $\hat{F}_{\alpha d}^{\mathrm{FH}}$ cannot be further disaggregated for subdomains or subareas within the areas unless a new good model is found at that subarea level.

3.3. 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

$$Y_{di} = \mathbf{x}_{di}'\beta + u_d + e_{di}, \quad i = 1,\ldots,N_d, d = 1,\ldots,D. \tag{8}$$

Here, β 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 $\hat{F}_{\alpha d}^{\mathrm{ELL}} = E[F_{\alpha d}]$ under model (8). 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 (8) to survey data. Then, a bootstrap census of Y-values is generated as

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

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$. The generation is repeated for $a = 1,\ldots,A$, 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,

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

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

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

Advantages and disadvantages of ELL method are listed below.

Advantages:

- It is based on unit level data, which are richer than area level data and sample size is much larger (n compared to D).
- ELL method can be applied to estimate general indicators defined as a function of the model response variables Y_{di}.
- They are model-unbiased if the model parameters are known.
- Once the model is fitted, estimates can be obtained at whatever subarea level.

Disadvantages:

- In terms of model MSE, ELL estimates perform poorly and can even perform worse than direct estimators when unexplained between-area variation is significant, see Molina and Rao (2010). In fact, for the estimation of domain means, ELL estimates are basically equal to regression-synthetic estimators, which assume the regression model without further between-area variation.
- They are based on a model assumption. Hence, model checking is crucial.

- They are not design-unbiased and can be seriously biased under informative sampling.
- They can be seriously affected by unit level outliers.
- If cluster effects are included in the model instead of area effects, but area effects are significant, ELL estimates of the model MSE can seriously underestimate the true MSE. Even if area effects are included in the model, ELL estimates of MSE do not track correctly the true MSE for each area.

3.4. 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 (8) with normality of random effects u_d and errors e_{di}. Under that model, the area vectors $\mathbf{Y}_d = (Y_{d1},\ldots,Y_{dN_d})'$ are independent for $d = 1,\ldots,D$ and satisfy $\mathbf{Y}_d \overset{ind}{\sim} N(\mu_d, \mathbf{V}_d)$, where $\mu_d = \mathbf{X}_d\beta$ and $\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{N_d}\mathbf{1}'_{N_d} + \sigma_e^2 \mathbf{A}_d$, for $\mathbf{A}_d = \mathrm{diag}(k_{di}^2; i = 1,\ldots,N_d)$. For an area parameter $\delta_d = h(\mathbf{Y}_d)$, the estimator that minimizes the MSE, called best estimator, is given by

$$\hat{\delta}_d^{\mathrm{B}} = E_{\mathbf{Y}_{dr}}[h(\mathbf{Y}_d)|\mathbf{Y}_{ds};\theta] = \int h(\mathbf{Y}_d) f(\mathbf{Y}_{dr}|\mathbf{Y}_{ds};\theta) d\mathbf{Y}_{dr}, \tag{9}$$

where $f(\mathbf{Y}_{dr}|\mathbf{Y}_{ds};\theta)$ 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 θ is the vector of model parameters. Now replacing θ in (9) by an estimator $\hat{\theta}$, we get the empirical best (EB) estimator, $\hat{\delta}_d^{\mathrm{EB}}$.

Under the nested error model (8), 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 $\mathbf{Y}_{dr}|\mathbf{Y}_{ds} \overset{ind}{\sim} N(\mu_{dr|s}, \mathbf{V}_{dr|s})$, where the conditional mean vector and covariance matrix are given by

$$\mu_{dr|s} = \mathbf{X}_{dr}\beta + \gamma_{dc}(\bar{y}_{dc} - \bar{\mathbf{x}}_{dc}^T\beta)\mathbf{1}_{N_d-n_d}, \tag{10}$$

$$\mathbf{V}_{dr|s} = \sigma_u^2(1-\gamma_d)\mathbf{1}_{N_d-n_d}\mathbf{1}_{N_d-n_d}^T + \sigma_e^2 \mathrm{diag}_{i\in r_d}(k_{di}^2). \tag{11}$$

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

$$\bar{y}_{dc} = \frac{1}{c_{d\cdot}} \sum_{i \in s_d} c_{di} Y_{di}, \quad \bar{\mathbf{x}}_{dc} = \frac{1}{c_{d\cdot}} \sum_{i \in s_d} c_{di} \mathbf{x}_{di}. \tag{12}$$

For complex non-linear parameters $\delta_d = h(\mathbf{Y}_d)$, the expectation given in (9) cannot be calculated analytically. In those cases, the EB estimator $\hat{\delta}_d^{\mathrm{EB}}$ is approximated by Monte Carlo. This requires to simulation of 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 replication 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 (11), corresponds to the covariance matrix of a random vector $\mathbf{Y}_{dr}^{(a)}$ generated from the model

$$\mathbf{Y}_{dr}^{(a)} = \mu_{dr|s} + v_d^{(a)} \mathbf{1}_{N_d - n_d} + \boldsymbol{\varepsilon}_{dr}^{(a)}, \tag{13}$$

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

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

see Molina and Rao (2010). Using model (13), 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 $\varepsilon_{di}^{(a)} \overset{ind}{\sim} N(0, \sigma_e^2 k_{di}^2)$, for $i \in r_d$. Then, we obtain the corresponding out-of-sample values $Y_{di}^{(a)}$, $i \in r_d$, from (13) using as means the corresponding elements of $\mu_{dr|s}$ given by (10). Using the vector $\mathbf{Y}_{dr}^{(a)}$ generated from (13), 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 (13) with $\gamma_{dc} = 0$ and in this case $\mathbf{Y}_d^{(a)} = \mathbf{Y}_{dr}^{(a)}$. The Monte Carlo approximation to the EB estimator (9) of $\delta_d = h(\mathbf{Y}_d)$ is then given by

$$\hat{\delta}_d^{\mathrm{EB}} \approx \frac{1}{A} \sum_{a=1}^{A} h(\mathbf{Y}_d^{(a)}). \tag{14}$$

In particular, to estimate the FGT poverty indicator given in (1), Molina and Rao (2010) assumed that $Y_{di} = T(E_{di})$ follow the nested error model (8), where E_{di} are variables measuring welfare and $T(\cdot)$ 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

can be expressed as

$$F_{\alpha d} = \frac{1}{N_d} \sum_{i=1}^{N_d} \left(\frac{z - T^{-1}(Y_{di})}{z} \right)^{\alpha} I(T^{-1}(Y_{di}) < z) = h_{\alpha}(\mathbf{Y}_d), \tag{15}$$

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, $\{\mathbf{x}_{di}; i = 1, \ldots, N_d,\ d = 1, \ldots, D\}$. The EB method requires additionally the identification of 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 (9) by generating $\mathbf{Y}_d^{(a)} = \mu_{d|s} + v_d^{(a)} \mathbf{1}_{N_d - n_d} + \varepsilon_d^{(a)}$, where $\mu_{d|s} = \mathbf{X}_d \beta + \gamma_{dc}(\bar{y}_{dc} - \bar{\mathbf{x}}_{dc}^T \beta) \mathbf{1}_{N_d}$ and $\varepsilon_d^{(a)} \sim N(\mathbf{0}_{N_d}, \sigma_e^2 \mathrm{diag}_{i=1,\ldots,N_d}(k_{di}^2))$. 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.

Good properties and drawbacks of the EB method are listed below.

Advantages:

- It is based on unit level data, which are richer than the area level data and uses much larger sample size to fit the model.
- The EB method can be applied to estimate general indicators defined as functions of the response variables Y_{di}.
- Best estimators are model-unbiased.
- They are optimal in terms of minimizing the model MSE for known values of model parameters.
- EB estimates perform significantly better than ELL estimates when unexplained between-area variation is significant. For out-of-sample areas (with

$n_d = 0$), EB and ELL small area estimates are nearly the same. They are nearly the same for all areas if there is no unexplained between-area variation ($\sigma_u^2 = 0$).

- Once the model is fitted, estimates can be obtained at whatever subarea level.

Disadvantages:

- They are based on a model assumption. Hence, model checking is crucial.
- They are not approximately design-unbiased and can be seriously biased under informative sampling.
- They can be severely affected by unit level outliers.
- Parametric bootstrap estimates of the MSE of EB estimators are computationally intensive.

3.5. 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 (8) in terms of the intraclass correlation coefficient $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$ and considering priors for the model parameters $(\beta, \rho, \sigma_e^2)$ that reflect lack of knowledge. Concretely, the HB model is defined as

$$\text{(i)} \quad Y_{di}|u_d, \beta, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}'_{di}\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,$$

$$\text{(ii)} \quad u_d|\rho, \sigma_e^2 \overset{iid}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma_e^2\right), \quad d = 1, \ldots, D,$$

$$\text{(iii)} \quad \pi(\beta, \rho, \sigma_e^2) \propto \frac{1}{\sigma_e^2}, \quad \varepsilon \leq \rho \leq 1 - \varepsilon, \, \sigma_e^2 > 0, \beta \in \mathscr{R}^p,$$

where $\varepsilon > 0$ 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 $\theta = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ given the sample values $\mathbf{Y}_s$ is given by

$$\pi(\mathbf{u}, \beta, \sigma_e^2, \rho | \mathbf{Y}_s) = \pi_1(\mathbf{u}|\beta, \sigma_e^2, \rho, \mathbf{Y}_s)\pi_2(\beta|\sigma_e^2, \rho, \mathbf{Y}_s)\pi_3(\sigma_e^2|\rho, \mathbf{Y}_s)\pi_4(\rho|\mathbf{Y}_s), \quad (16)$$

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 $\theta = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ can then be generated directly from the posterior distribution in (16), avoiding the use of Markov Chain Monte Carlo (MCMC) methods. Under general conditions, a proper posterior distribution is guaranteed.

Given θ, population variables Y_{di} are all independent, satisfying

$$Y_{di}|\theta \overset{ind}{\sim} N(\mathbf{x}'_{di}\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d, \, d = 1, \ldots, D. \quad (17)$$

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

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

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

$$\hat{\delta}_d^{\mathrm{HB}} = E_{\mathbf{Y}_{dr}}(\delta_d|\mathbf{Y}_s) = \int h(\mathbf{Y}_d) f(\mathbf{Y}_{dr}|\mathbf{Y}_s)d\mathbf{Y}_{dr}. \quad (18)$$

The HB estimator can be approximated by Monte Carlo. For this, we first generate samples from the posterior $\pi(\theta|\mathbf{Y}_s)$. 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 $\beta^{(a)}$ is generated from $\pi_2(\beta|\sigma_e^{2(a)}, \rho^{(a)}, \mathbf{Y}_s)$ and, finally, $\mathbf{u}^{(a)}$ is generated from $\pi_1(\mathbf{u}|\beta^{(a)}, \sigma_e^{2(a)}, \rho^{(a)}, \mathbf{Y}_s)$. This process is repeated a large number A of times to get a random sample $\theta^{(a)}$, $a = 1, \ldots, A$ from $\pi(\theta|\mathbf{Y}_s)$. Now for each generated value $\theta^{(a)}$ from $\pi(\theta|\mathbf{Y}_s)$, we generate the out-of-sample values $\{Y_{di}^{(a)}, i \in r_d\}$ from the distribution defined in (17). 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 (15). 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

$$\hat{F}_{\alpha d}^{\mathrm{HB}} \approx \frac{1}{A}\sum_{a=1}^{A} F_{\alpha d}^{(a)}. \tag{19}$$

Benefits and deficiencies of HB method are listed below.

Advantages:

- It is based on unit level data, which are richer than area level data and uses much larger sample size to fit the model.
- HB method can be applied to estimate general indicators defined as function of the model response variables Y_{di}.
- HB estimators are model-unbiased.
- HB estimators are optimal in terms of minimizing the posterior variance.
- EB and HB methods are expected to give practically the same point estimates, see Molina, Nandram and Rao (2014). Thus, the proposed HB method has good frequentist properties.
- Once the model is fitted, estimates can be obtained at any subarea level.
- The proposed HB approach does not require the use of MCMC methods and therefore avoids the need of monitoring the convergence of Monte Carlo chains.
- Bootstrap methods for MSE estimation are not needed. Therefore, total computational time is considerably lower than in the EB method.
- Calculation of credible intervals or other posterior summaries is straightforward.

Disadvantages:

- It is based on model assumptions. Hence, model checking is crucial.
- HB estimators are not design-unbiased and can be seriously biased under informative sampling.
- HB estimators can be severely affected by unit level outliers.
- HB method is not directly extendable to more complex models without losing some of the mentioned advantages like avoiding MCMC.

4. Simulation studies

This section illustrates some of the mentioned advantages and drawbacks of the considered poverty mapping methods through simulation studies. Concretely, we will report results of simulations under three different scenarios: (i) Nested error model with simple random sampling. (ii) Nested error model with informative sampling. (iii) Nested error model with outliers.

Simulations were implemented in the statistical software environment R (R development core team 2013) using the package `lme4` (Bates et al. 2014), which fits Gaussian linear and nonlinear mixed-effects models, and the package `sae` (Molina and Marhuenda 2015), which contains functions for small area estimation, including calculation of direct, FH and EB estimates along with their MSE estimates.

4.1. Nested error model with simple random sampling

We consider the same model-based simulation setup as in Molina, Nandram and Rao (2014), where data are generated at the unit level following the nested error model (8). However, here we also include FH estimators derived from the FH area level model with the area means of the auxiliary variables as covariates. In addition, we include ELL and Census-EB estimators. The population is composed of $N = 20,000$ units, distributed in $D = 80$ areas with $N_d = 250$ units in each area. We consider two auxiliary variables X_1 and X_2 with known values for all the population units. Their values are generated as $x_{k,di} \sim \text{Bern}(p_{kd})$, $k = 1,2$, with success probabilities $p_{1d} = 0.3 + 0.5d/D$ and $p_{2d} = 0.2$, $d = 1,\ldots,D$. Response variables Y_{di} are generated from the nested error model (8) and the target variables are $E_{di} = \exp(Y_{di})$. The true values of the regression coefficients are $\beta = (3, 0.03, -0.04)'$. Variances of area effects and errors are taken as $\sigma_u^2 = 0.15^2$ and $\sigma_e^2 = 0.5^2$ respectively. The poverty line is set to $z = 12$, which is approximately 0.6 times the median of $\{E_{di}; i = 1,\ldots,N_d,\ d = 1,\ldots,D\}$ for a population generated as described before, which is the official definition of poverty line used in the EU countries. We draw a sample s_d of size $n_d = 50$, $d = 1,\ldots,D$, using sample random sampling (SRS) without replacement, independently from each area d.

A total of $L = 1,000$ population vectors $\mathbf{Y}^{(\ell)}$, $\ell = 1,\ldots,L$, were generated from the nested error model (8) with the mentioned values of model parameters and auxiliary variables. For each Monte Carlo population $\ell = 1,\ldots,L$, we calculated the true area poverty incidences and poverty gaps. Then, we selected the sample s, which is kept fixed across Monte Carlo replicates. Using the sample data $\{(Y_{di}, x_{1,di}, x_{2,di}); i \in s_d, d = 1,\ldots,D\}$ and the population data on the auxiliary variables, we computed direct estimates $\hat{F}_{\alpha d}^{\text{DIR}}$, FH, ELL, EB, Census-EB and

HB estimates of poverty incidence ($\alpha = 0$) and poverty gap ($\alpha = 1$) for each area $d = 1\ldots, D$. FH, ELL and EB estimates were obtained using REML fitting method.

For the Monte Carlo population ℓ, let $F_{\alpha d}^{(\ell)}$ be the true poverty indicator for area d and $\hat{F}_{\alpha d}^{(\ell)}$ be one of the estimates (direct, FH, ELL, EB, Census-EB or HB). Relative bias (RB) and relative root mean squared error (RRMSE) of an estimator $\hat{F}_{\alpha d}$ are approximated empirically as

$$\mathrm{RB}(\hat{F}_{\alpha d}) = \frac{L^{-1}\sum_{\ell=1}^{L}(\hat{F}_{\alpha d}^{(\ell)} - F_{\alpha d}^{(\ell)})}{L^{-1}\sum_{\ell=1}^{L} F_{\alpha d}^{(\ell)}}, \quad \mathrm{RRMSE}(\hat{F}_{\alpha d}) = \frac{\sqrt{L^{-1}\sum_{\ell=1}^{L}(\hat{F}_{\alpha d}^{(\ell)} - F_{\alpha d}^{(\ell)})^2}}{L^{-1}\sum_{\ell=1}^{L} F_{\alpha d}^{(\ell)}}.$$

For each estimator $\hat{F}_{\alpha d}$, the absolute RB (ARB) and the RRMSE are averaged across areas as

$$\overline{\mathrm{ARB}}_\alpha = D^{-1}\sum_{d=1}^{D}|\mathrm{RB}(\hat{F}_{\alpha d})|, \quad \overline{\mathrm{RRMSE}}_\alpha = D^{-1}\sum_{d=1}^{D}\mathrm{RRMSE}(\hat{F}_{\alpha d}).$$

Figure 1 depicts percent RBs (left) and RRMSEs (right) of the estimators of the domain poverty gaps F_{1d} for each area d. EB and Census-EB estimates are not shown in these plots because they are both practically equal to HB estimates and are plotted separately in Figure 2. We can see in Figure 1 left that direct, ELL and HB estimators are practically unbiased. In contrast, FH estimators display a substantial negative bias. Concerning efficiency, Figure 1 right shows that HB estimators have the smallest RRMSE whereas ELL estimators are the ones with the largest RRMSE. Conclusions for the poverty incidence F_{0d} are very similar.

Table 1 presents averages across areas of absolute RB and RRMSE of all the estimators, for both poverty incidence and poverty gap. We see that, on average, FH estimator presents a large absolute RB (over 6% for poverty incidence and close to 15% for poverty gap), whereas EB, HB and Census-HB estimators have a very small RB ($< 1\%$). The latter estimators also achieve the smallest RRMSEs (slightly over 20% for poverty incidence and over 25% for poverty gap). The largest RRMSE is obtained by ELL estimator (over 58%). Note that both absolute RB and RRMSE increase when estimating the poverty gap, because the poverty gap depends to a greater extent on the extreme of the left tail of the income distribution, which is more difficult to estimate correctly from a (finite) sample.

These results indicate that HB estimators are practically unbiased and clearly the most efficient among the considered estimators when the nested error model holds and the sample is drawn with SRS within each area. The bias of FH estimators is due to the fact that they are attaching most of the weight to the regression-synthetic component, which relies exactly on the model, but here data Y_{di} are generated from

the unit level model (8) and the area means of the covariates $\bar{X}_{k,d} = N_d^{-1} \sum_{i=1}^{N_d} x_{k,di}$ are not linearly related with the poverty indicators $F_{\alpha d}$. Thus, FH model fails due to non-linearity of the poverty indicators $F_{\alpha d}$ in the area level covariates $\bar{X}_{k,d}$, $k = 1, 2$, even if the unit level model holds exactly.

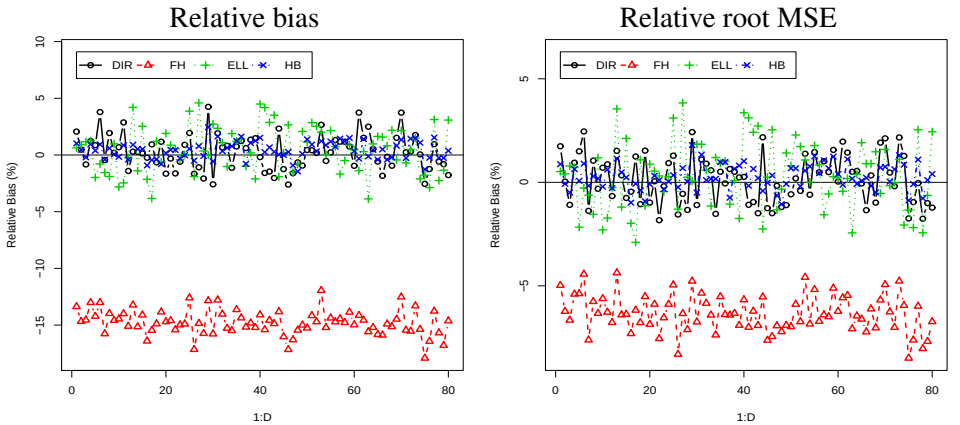

Figure 1. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under the nested error model with simple random sampling.

Table 1. Averages across areas of percent ARB and RRMSE for direct, FH, HB, EB, Census-EB and ELL estimators of poverty incidence F_{0d} and poverty gap F_{1d}, under the nested error model with simple random sampling.

Method	Average ARB (%)		Average RRMSE (%)	
	F_{0d}	F_{1d}	F_{0d}	F_{1d}
Direct	0.99	1.26	28.53	36.33
FH	6.34	14.78	26.26	38.16
HB	0.48	0.65	20.15	25.43
EB	0.51	0.67	20.41	25.73
Census-EB	0.55	0.69	21.15	26.71
ELL	1.31	1.69	47.39	58.63

Figure 2 depicts percent RB (left) and RRMSE (right) of EB and Census-EB estimates of the poverty gap F_{1d} for each area d. Figure 2 left shows the great similarity of EB and Census-EB estimates of F_{1d}, even if sampling fractions in this simulation study are not so small ($n_d/N_d = 1/5$, $d = 1, \ldots, D$). See in Figure 2 right and in Table 1 that the average RRMSE increase of the Census-EB estimator is in this case less than 1%.

Next we study ELL estimator of the MSE of $\hat{F}_{\alpha d}^{\mathrm{ELL}}$. Figure 3 depicts the true MSE of ELL estimators of the poverty gap F_{1d}, labeled “True MSE ELL" and the

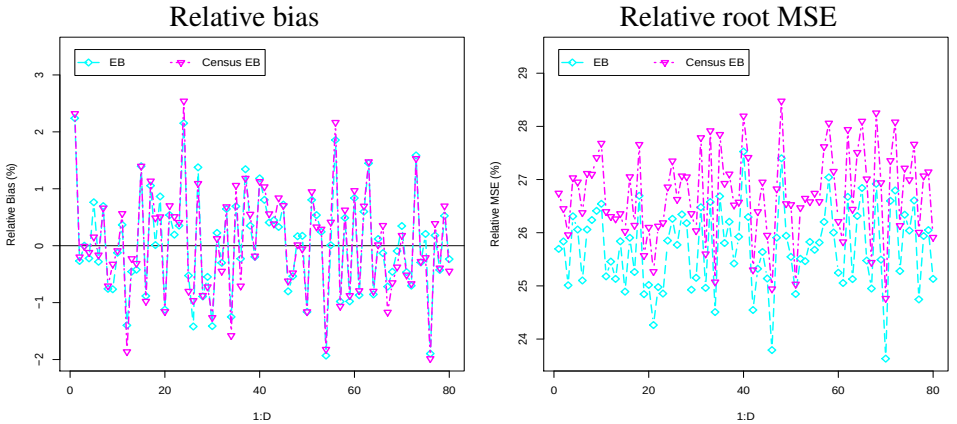

Figure 2. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under the nested error model with simple random sampling.

means across simulations of ELL estimates of the MSE, labeled "MSE ELL", for each area d. This figure shows that ELL estimates of MSE do not really track the true MSEs for each area even if we have considered here random effects for the areas in the model (i.e., sampling clusters equal to areas). In the case that clusters are different from the areas, if we consider the original ELL method that includes only cluster effects but area effects are significant, then ELL estimates might seriously underestimate the MSE.

For EB estimator, the parametric bootstrap procedure proposed by Molina and Rao (2010) approximates the true MSE reasonably well, see Molina and Rao (2010). For HB estimator, posterior variance, approximated by Monte Carlo, is taken as measure of uncertainty.

4.2. Nested error model with informative sampling

We consider the same setup as in the previous simulation study, with the same population sizes, model parameters, auxiliary variables and poverty line. The only difference is that in this simulation study, samples are drawn with informative sampling. When the sampling is informative, the probability of a sample depends on the values of the population vector $\mathbf{Y}$. Thus, under this setup, the simulations need to be performed with respect to the joint distribution of $(\mathbf{Y}, s)$; that is, in each Monte Carlo replicate ℓ, we draw a population vector $\mathbf{Y}^{(\ell)}$ and, given $\mathbf{Y}^{(\ell)}$, we draw a sample $s^{(\ell)}$. A total of $L = 1000$ population vectors $\mathbf{Y}^{(\ell)}$, $\ell = 1, \ldots, L$, are generated from the true nested error model (8). Again, we consider that the target variables

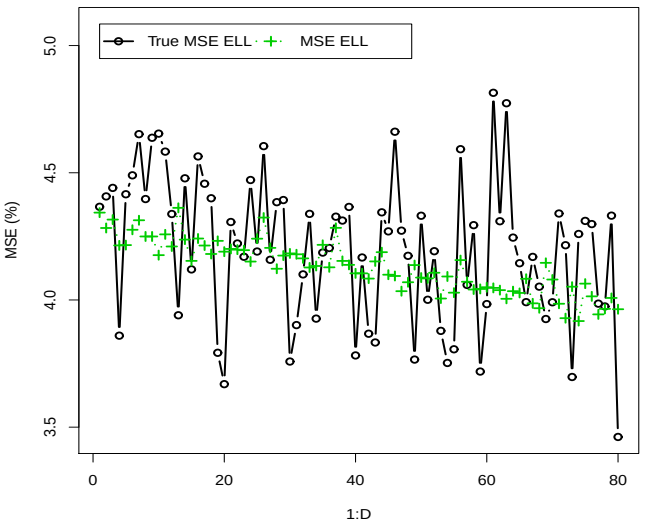

Figure 3. True MSE of ELL estimators of poverty gap F_{1d} and mean across simulations of ELL estimator of the MSE for each area d, under the nested error model with simple random sampling.

are $E_{di} = \exp(Y_{di})$. The sample $s^{(\ell)}$ is drawn by Poisson sampling, with inclusion probabilities $\pi_{d,i}$ depending on a random variable Z_{di} that is correlated with the unexplained part of Y_{di}, that is, the model errors e_{di}. Thus, for each population unit i from area d, we generate a Bernoulli random value $Q_{di} \sim \text{Bern}(\pi_{d,i})$, with $\pi_{d,i} = b^{-1}\exp(-aZ_{di})$, where $a > 0$, $b > 0$ and $Z_{di} \sim \text{Gamma}(\tau_{di}, \theta_{di})$. To choose the values of τ_{di} and θ_{di}, we consider two cases: low and high level of informativeness. In the first case, we take $\tau_{di} = 15 + 0.5e_{di}$ and $\theta_{di} = 0.75 + 0.025e_{di}$, which yield random values Z_{di} with a 20% correlation with the model errors e_{di}. In the second case, we take $\tau_{di} = 22.5 + 7.5e_{di}$ and $\theta_{di} = 1.125 + 0.375e_{di}$, yielding Z_{di} with a 80% correlation with e_{di}, which represents a high level of informativeness. Note that under informative sampling, the sample size is random because each unit in the population comes to the sample depending on its random value Q_{di}. To make this simulation study comparable with the one in previous section, we wish to have a similar average area sample size as before. This is achieved approximately by considering $a = 0.05$ and $b = 2.5$ when the informativeness level is low and taking $a = 0.02$ and $b = 4$ when the informative level is high. With the sample $s^{(\ell)}$ from each population, we compute the five estimators, namely direct, FH, EB, ELL and HB estimators. We excluded here Census-EB estimators because of their similarity with EB estimators.

Figure 4 plots RBs (left) and RRMSEs (right) of the estimators of the poverty gap F_{1d} when the informativeness level is low. Again, EB estimator is excluded because it provides nearly the same results as HB. For low level of informativeness, Figure 4 left shows that the negative bias of the EBLUP based on the FH model, observed in the simulation with SRS, still persists, while the rest of the estimators are almost unbiased. HB estimator still presents the smallest relative MSE among the considered estimators, and ELL estimator performs the worst in terms of relative MSE among the considered estimators. For the poverty incidence F_{0d}, conclusions are similar. These conclusions are confirmed by the averages across areas shown in Table 2 for both poverty incidence and poverty gap. On average, the direct estimator has the smallest absolute RB (about 0.7% for poverty incidence and 0.9% for poverty gap), followed by EB and HB estimators with a bias below 1.4% for both poverty incidence and gap, the smallest RRMSE is for EB estimator (less than 21% for poverty incidence and than 26% for poverty gap) and the largest for ELL estimator (over 47% for poverty incidence and over 58% for poverty gap).

Figure 5 plots RB (left) and RRMSE (right) of the estimators of the poverty gap F_{1d} when the level of informativeness is large. In this case, Figure 5 left shows a negative bias for the FH estimator and a large positive bias of HB and ELL estimators. Looking at Figure 5 right, we can see that now direct and FH estimates, which are calculated using the true inclusion probabilities, present the smallest RRMSE among the considered estimators. Again, conclusions are similar for the poverty incidence F_{0d}. Table 3 lists the averages across areas of ARB and RRMSE for all the considered estimators of the poverty incidence and poverty gap. In this case, the direct estimator has the smallest average ARB (about 0.6% for poverty gap), whereas the average RRMSE of ELL estimator is the largest (99.6%).

To summarize, EB and HB methods are not greatly affected under low level of informativeness, measured in terms of correlation among the design variable used in the inclusion probabilities and the response variable. When the degree of informativeness is high, these two methods are certainly affected because they do not take into account the sampling design. The effect of informative sampling on FH estimator seems to be smaller, and its negative bias is again due to a non-linearity problem of FH model because data actually follows the nested error linear regression model for log income at the unit level. We are currently developing suitable methods to handle informative sampling in the case of unit level models.

4.3. Nested error model with outliers

In this section, we carry out a simulation study under exactly the same conditions as in Section 4.1, but generating the model errors e_{di} from a mixture of normal

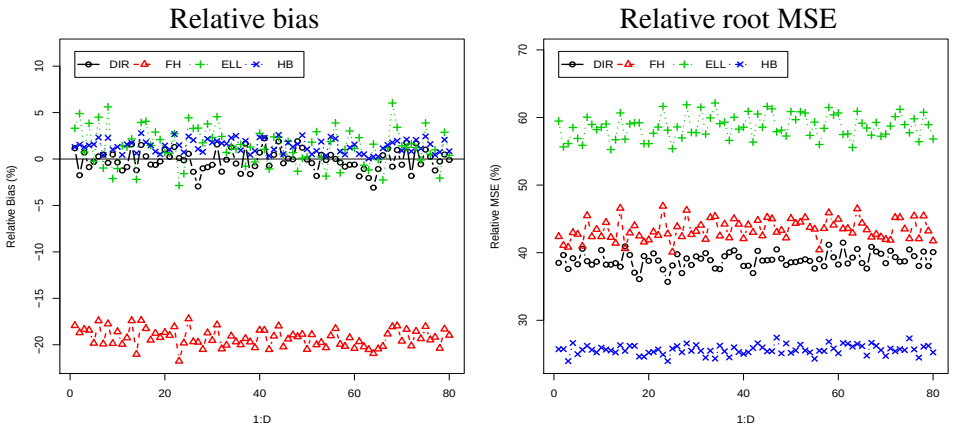

Figure 4. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under low informativeness.

Table 2. Averages across areas of percent ARB and RRMSE for direct, FH, HB, EB and ELL estimators of incidence F_{0d} and poverty gap F_{1d}, under low informativeness.

Method	Average ARB (%)		Average RRMSE (%)	
	F_{0d}	F_{1d}	F_{0d}	F_{1d}
Direct	0.74	0.91	71.69	38.92
FH	10.47	19.26	30.33	43.38
HB	1.10	1.38	20.29	35.63
EB	1.04	1.25	20.48	25.86
ELL	1.63	1.98	47.39	58.65

distributions with different variances in order to create outliers. Concretely, in this simulation study, we generate model errors as $e_{di} \sim (1-\varepsilon)N(0,\sigma_e^2)+\varepsilon N(0,R\sigma_e^2)$, where ε is generated as $\varepsilon \sim \text{Bern}(p)$. We consider two fractions of outliers, $p=0.1$ and $p=0.5$, and two values for the factor R in the variance of outliers, namely $R=10$ and $R=100$. Using the above mechanism to generate model errors, a total of $L=1000$ population vectors $\mathbf{Y}^{(\ell)}, i=1,\ldots,L$, were generated from the nested error model (8). Then, we calculated true area poverty incidences and gaps. Note that the outliers considered in this simulation study are not recording errors in the sample data. They are actually representative outliers appearing in the population. Thus, they are actual realizations of the distribution with heavier tails obtained from the normal mixture, and true values of poverty indicators actually include the generated outliers in the population. The sample is drawn by SRS within each area as in Section 4.1, keeping the sample units s fixed across simulations. With each Monte Carlo sample, direct, FH, EB, ELL and HB estimators were computed.

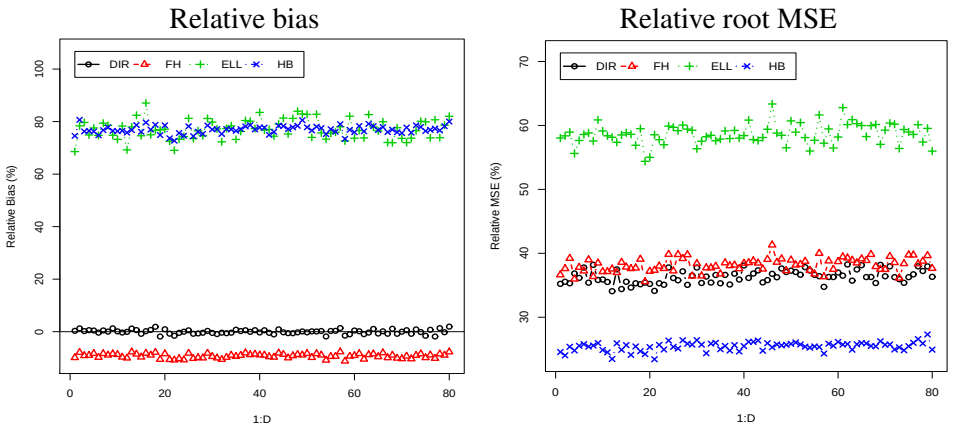

Figure 5. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under high informativeness.

Table 3. Averages across areas of percent ARB and RRMSE for direct, FH, HB, EB and ELL estimators of incidence F_{0d} and poverty gap F_{1d}, under large informativeness.

Method	Average ARB (%)		Average RRMSE (%)	
	F_{0d}	F_{1d}	F_{0d}	F_{1d}
Direct	0.59	0.65	23.62	25.69
FH	6.94	9.21	23.83	29.40
HB	61.64	76.95	66.05	84.95
EB	61.60	73.68	66.08	84.89
ELL	61.69	76.98	72.94	97.29

We report here results for the cases of less frequent mild outliers ($p = 0.1$ and $R = 10$), and of more frequent and extreme outliers ($p = 0.5$ and $R = 100$). For the first case, results for the poverty gap are plotted in Figure 6. Again, EB is excluded in the plots because it provides similar results as HB. Figure 6 left and right show that direct estimators are not practically affected by the outliers, which is expected because this estimator does not rely on any model assumption. Similarly, FH estimator is less affected by outliers because the observed negative bias is again due to non-linearity problems. HB and ELL estimators show a moderate bias, but still HB estimator achieves the lowest error in terms of RRMSE. Averages across areas of ARB and RRMSE for all estimators of poverty incidence and poverty gap are shown in Table 4. We can see that the bias of EB and HB estimators is small (around 4% for poverty incidence and 5% for poverty gap), and the RRMSE has increased only about 0.5% with respect to the case of no outliers (see Table 2) and it is still acceptable (around 21% for poverty incidence and 26% for poverty gap).

For the case of more frequent and extreme outliers ($p = 0.5$ and $R = 100$), Figure 7 left shows that in this case HB, and to a greater extent ELL estimators; present a very large positive bias, see also Table 7 reporting averages across areas. Note that the RRMSE of ELL estimator reaches 226.63% for the poverty gap. In this simulation study, FH estimates perform better than in the previous simulation studies, and this could be due to the fact that, since FH model is less correct when outliers are present, the FH estimator is attaching more weight to the direct estimator, which is practically unbiased. EB, HB and ELL estimators are severely biased when data contains frequent extreme outliers, performing even worse than under high level of informative sampling, but are not too much affected under rare and not so extreme outliers. These methods are based on model assumptions and are not robust to strong model misspecification when the true error distribution has very heavy tails as in the mixture model considered here with $p = 0.5$ and $R = 100$. We are exploring estimation methods for complex parameters that are robust to outliers. Note that previous work on robust estimation, e.g. Sinha and Rao (2009), focused on estimating area means only.

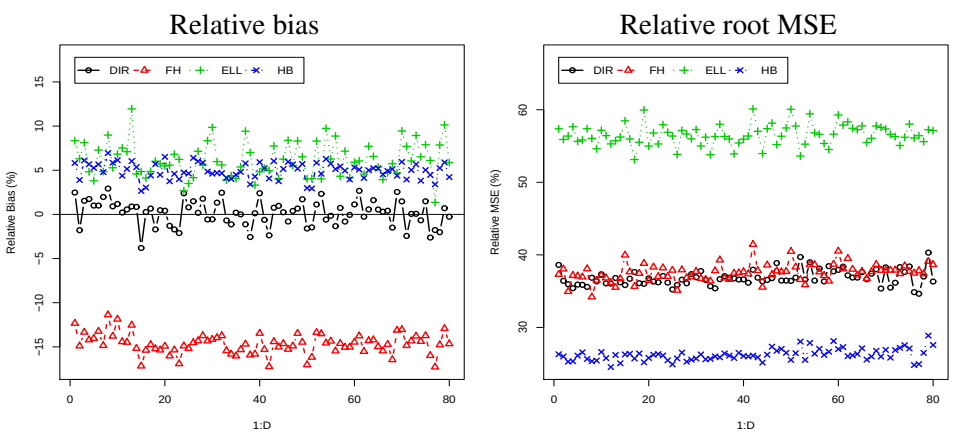

Figure 6. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under nested error model with outliers ($p = 0.01$ and $R = 10$).

5. Conclusions

This paper reviews popular poverty mapping procedures focusing on practical aspects. Simulation studies compare these methods under three interesting scenarios that show the good properties when assumptions hold and also the worse performance when some assumptions are not satisfied. These simulation studies illustrate

Table 4. Averages across areas of percent ARB and RRMSE for direct, FH, HB, EB and ELL estimators of incidence F_{0d} and poverty gap F_{1d}, under under nested error model with outliers ($p = 0.01$ and $R = 10$).

Method	Average ARB (%)		Average RRMSE (%)	
	F_{0d}	F_{1d}	F_{0d}	F_{1d}
Direct	0.92	1.18	28.54	36.82
FH	6.16	14.67	26.10	37.55
HB	3.95	4.95	20.81	26.22
EB	3.88	4.79	20.99	26.42
ELL	4.93	6.14	46.65	56.52

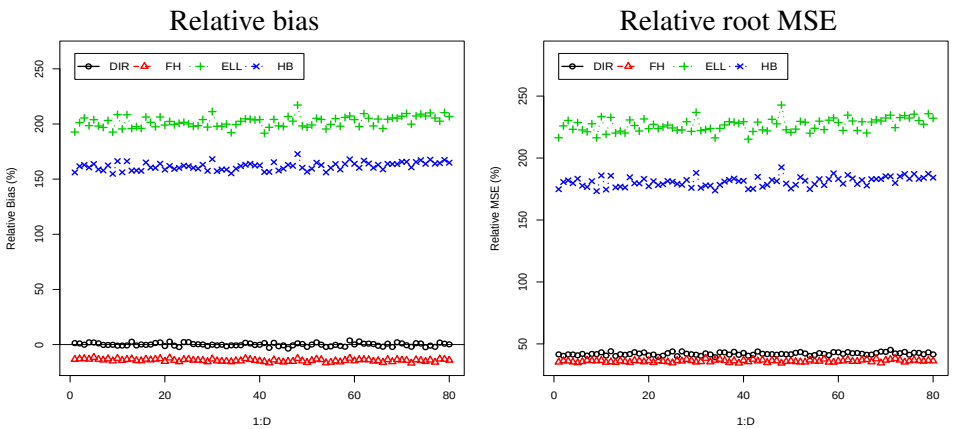

Figure 7. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under nested error model with outliers ($p = 0.05$ and $R = 100$)

that: (i) Even if aggregation protects against model failures in FH area level model, the linearity assumption of the model fails when data follows a unit level model but target parameters are nonlinear functions of the model responses. However, FH estimates are less affected by informative sampling and by symmetric representative unit level outliers. (ii) EB and HB methods perform practically the same, and are the best among the considered estimators when the nested error model with normality holds and sampling is noninformative. They are not very much affected by mildly informative sampling and small proportion of mild outliers, but might be severely affected by highly informative sampling or severe outliers in large proportions. (iii) Census-EB estimators of poverty indicators are practically the same as EB estimators and avoid linking the survey and census data files. (iv) ELL method under a nested error model with random area effects performs the worst in all scenarios

Table 5. Averages across areas of percent ARB and RRMSE for direct, FH, HB, EB and ELL estimators of incidence F_{0d} and poverty gap F_{1d}, under nested error model with outliers ($p = 0.05$ and $R = 100$).

Method	Average ARB (%)		Average RRMSE (%)	
	F_{0d}	F_{1d}	F_{0d}	F_{1d}
Direct	0.96	1.20	29.68	41.99
FH	5.66	14.33	26.65	36.10
HB	74.13	161.73	86.87	180.88
EB	74.11	161.59	86.95	180.81
ELL	92.64	201.97	111.32	226.63

because it does not account for unexplained between-area variation.

Several relaxations of the normality assumption in the EB method have been recently studied. Diallo and Rao (2004) derived EB estimators of poverty indicators assuming the family of skew normal (SN) distributions for the random effects and/or the errors, which includes the normal distribution as a particular case. Their results indicate that the EB method based on normality is robust to deviations from normality of u_d provided e_{di} remains normal. On the other hand, under SN errors e_{di}, normality-based EB estimators can induce significant bias and may not perform well compared to SN-based EB estimators. Van der Weide and Elbers (2014) studied normal mixture models on the area effects u_d and the errors e_{di}. Their results are in agreement with Diallo and Rao (2014) in the sense that the normality-based EB method is robust provided e_{di} remains normal. Graf, Marín and Molina (2015) have also extended the EB method to the generalized Beta distribution of the second kind (GB2), which models income data adequately. They have also shown that using the EB method based on the GB2 distribution leads to clear efficiency gains when the distribution of log income deviates from normality, whereas it does not lose efficiency when log incomes follow the nested error model with normality.

Acknowledgements

Isabel Molina was supported by grants ref. MTM2009-09473, MTM2012-37077-C02-01 and SEJ2007-64500 and J.N.K. Rao's research by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

BATES, D., MAECHLER M., BOLKER, B., WALKER, S., (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7.

BATTESE, G.E., HARTER, R.M., 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, 28–36.

CORREA, L., MOLINA, I., RAO, J.N.K., (2012). Comparison of methods for estimation of poverty indicators in small areas. Unpublished report.

DIALLO, M., 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., LANJOUW, P., (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, 71(1), 355–364.

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

FOSTER, J., GREER, J., THORBECKE, E., (1984). A class of decomposable poverty measures. *Econometrica*, 52, 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, 443–462.

GRAFF, M., MARÍN, J.M., MOLINA, I., (2015). Estimation of poverty indicators in small areas under skewed distributions, Unpublished manuscript.

MOLINA, I., MORALES, D., (2009). Small area estimation of poverty indicators. *Boletín de Estadística e Investigación Operativa*, 25, 318–325.

MOLINA, I., MARHUENDA, Y., (2015), Sae: An R Package for Small Area Estimation, *R Journal*, in print.

MOLINA, I., RAO, J.N.K., (2010). Small area estimation of poverty indicators. *The 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. *The Annals of Applied Statistics*, 8(2), 852–885.

PFEFFERMANN, D., (2013). New important developments on small area estimation. *Statistical Science*, 28, 40–68.

RAO, J.N.K., (2003). *Small Area Estimation*. Hoboken, NJ: Wiley.

RAO, J.N.K., MOLINA, I., (2015). *Small Area Estimation, Second Edition*. Hoboken, NJ: Wiley, in print.

SINHA, S., RAO, J.N.K., (2009). Robust small area estimation. *The Canadian Journal of Statistics*, 37, 381–399.

VAN der WEIDE, R., ELBERS, C. (2013). Estimation of normal mixtures in a nested error model with an application to small area estimation of welfare. Speech presented at the SAE Conference 2013, Bangkok, Thailand.