REVSTAT – Statistical Journal
Volume 18, Number 5, October 2020, 613–635

THE FAY–HERRIOT MODEL IN SMALL AREA ESTIMATION: EM ALGORITHM AND APPLICATION TO OFFICIAL DATA

Authors: José Luis Ávila-Valdez
– Department of Mathematics, Universidad Popular Autónoma del Estado de Puebla, Mexico
joseluis.avila@upaep.mx

Mauricio Huerta
– School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile
mauricio.huerta.a@gmail.com

Víctor Leiva
– School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Chile
victorleivasanchez@gmail.com www.victorleiva.cl

Marco Riquelme
– Institute of Statistics, Universidad de Valparaíso, Chile
mriquelmealamos@gmail.com

Leonardo Trujillo
– Department of Statistics, Universidad Nacional de Colombia, Colombia
ltrujilloo@unal.edu.co

Received: July 2017 Revised: May 2018 Accepted: July 2018

Abstract:

- Standard methods of variance component estimation used in the Fay-Herriot model for small areas can produce problems of inadmissible values (negative or zero) for these variances. This implies that the empirical best linear unbiased predictor of a small area mean does not take into account the variance of the random effect of the corresponding area, reducing it to a regression estimator. In this paper, we propose an approach based on the expectation-maximization (EM) algorithm to solve the problem of inadmissibility. As stated in the theory of variance component estimation, we confirm through Monte Carlo simulations that the EM algorithm always produces strictly positive variance component estimates. In addition, we compare the performance of the proposed approach with two recently proposed methods in terms of relative bias, mean square error and mean square predictor error. We illustrate our approach with official data related to food security and poverty collected in Mexico, showing their potential applications.

Key-Words:

- *empirical best linear unbiased predictor; food security and poverty; Monte Carlo simulation; R software; random effects; variance components.*

AMS Subject Classification:

- 62D05, 62F99, 62J99.

1. INTRODUCTION

Surveys are intended not only to estimate population target parameters, but also to estimate characteristics for a variety of subpopulations commonly known as domains or areas. An area is considered as small if the sample domain is not sufficiently large to have a direct estimate of the area parameter with adequate precision. Then, the goal of the small area estimation is to produce reliable estimates of subpopulation target parameters for areas with small samples or even where the area is not sampled at all; see Pfeffermann [30] (2013).

Currently, the small area estimation methodology is playing an important role in both public and private sectors. Different government agencies around the world, for example, the Bureau of Labor Statistics and Census Bureau in the United States (US), Ministry of Social Development of Chile, National Administrative Department of Statistics in Colombia, National Council for the Evaluation of Social Development Policy in Mexico and Office of National Statistics in the United Kingdom (UK) are adopting such a methodology. This is due to the need for reliable estimates on parameters of interest in specific areas or domains; see Rao and Molina [35] (2015).

Because of the wide acceptance about small area estimation in recent years, several models have been developed, applied and studied. Pfeffermann [30] (2013), Rao [34] (2003) and Rao and Molina [35] (2015) reviewed the advances in this methodology from its beginnings to the present. Small area estimation methodology can be divided into two parts (Lohr [24, pp. 518-522], 1999): (i) design-based techniques (for example, direct, synthetic and composite estimators) and (ii) model-based techniques (for example, area-level models and unit-level models); see Coelho and Pereira [6] (2011), Pereira and Coelho [29] (2012) and Rueda *et al.* [36] (2018). On the one hand, in design-based techniques, the existence of a model is not recognized. Implicit models are sometimes proposed as an assisting tool, linking a number of small areas according to administrative or census records, which is considered as auxiliary data. Then, even when the model is misspecified, design-based properties can hold; see Lehtonen and Veijanen [22] (2009). On the other hand, model-based techniques rely on explicit super-population models (Datta [7], 2009) and include area-level models, relating each small area characteristic to auxiliary data that are available for each area. Area-level modeling is often described by the popular Fay-Herriot (FH) model (Fay and Herriot [13], 1979), which has been widely used in small area estimation. Li and Lahiri [23] (2010) emphasized that the main reasons for its widespread usage include: (i) its simplicity, (ii) its ability to protect the confidentiality of microdata, and (iii) its ability to produce design-consistent estimators. Other advantages of the FH model are that it takes into account the sampling design (level 1 model) and only requires area auxiliary variables that, in general, are more easily available in practice than unit auxiliary variables. Applications of the FH model have been extensive, mainly in the study of poverty and other related socio-demographic variables. For a reference in the context of big data sources in small area estimation through the FH model, see Marchetti *et al.* [25] (2015). A recent application of the FH model for poverty mapping in Chile can be found in Casas-Cordero *et al.* [2] (2016). Also, model-based techniques include unit-level models relating the unit values of the response variable to auxiliary variables for each individual in the survey; see Coelho and Casimiro [5] (2008). A well-known model proposed by Battese *et al.* [1] (1988) is a particular example of a unit-level model, corresponding to a nested regression model. Area and unit model-based techniques in small area estimation are presented by Jiang and Lahiri [20] (2006), Datta [7] (2009) and Datta and Ghosh [8] (2012), among others. Linear mixed models have played a crucial role in model-based techniques. Note also that these techniques can be based on either Bayesian or frequentist methods. In this paper, we consider a frequentist model-based technique employing the FH model under a non-informative sampling design. For informative sampling, see Pfeffermann and Sverchkov [31] (2007).

A problem detected in small areas, using the FH model, is that the standard methods utilized for variance component estimation may produce a negative or zero value. For more details about these methods, see Fay and Herriot [13] (1979), Prasad and Rao [33] (1990) for moment estimation (PR method), and Datta and Lahiri [9] (2000) for maximum likelihood (ML) and residual or restricted ML (RML) estimation. Note that the empirical best linear unbiased predictor (EBLUP) of a small area mean does not take into account the variance of the random effect for the corresponding area, reducing it to a regression estimator. Li and Lahiri [23] (2010) and Yoshimori and Lahiri [42] (2014) solved this problem adjusting the associated likelihood function.

The expectation-maximization (EM) algorithm is a popular iterative approach to estimate parameters by the ML method in models with incomplete data (unobserved or missing). This algorithm is used in many applications of mixed models, because there the unobserved data occur naturally. A comprehensive account of the EM algorithm is found in Laird and Were [21] (1982), van Dyk [39] (2000) and McLachlan and Krishnan [26] (2008). Some advantages of the EM algorithm are the following: (i) it is more stable than other algorithms, due to its property of monotone convergence (Laird and Were [21], 1982); (ii) it is more robust to starting values than other algorithms (Demidenko [11], 2013); and (iii) it generates positive definite matrix estimates if the starting matrix is positive definite (Thompson and Meyer [38], 1986; Searle *et al.* [37], 2006; Demidenko [11], 2013; El-Leithy *et al.* [12], 2016). An important feature related to (iii), stated by Searle *et al.* [37, pp. 297-298] (2006), is that the iterations will always remain in the parameter space, since the ML estimation is performed for the complete data.

The main objectives of this research are: (i) to review the estimation methods proposed at date on the topic; (ii) to propose an alternative approach for avoiding a negative or zero value in the variance component estimates, using the EM algorithm in both ML (MLEM) and RML (RMLEM) methods; (iii) to evaluate the proposed approach by Monte Carlo (MC) simulations; and (iv) to illustrate potential applications of our approach with official data related to food security and poverty. The proposed approach is compared to the methods presented in Li and Lahiri [23] (2010) and Yoshimori and Lahiri [42] (2014).

The outline of this paper is as follows. Section 2 introduces background to the FH model, the EBLUP of a small area mean, and a measure of its uncertainty. In addition, in this section, some variance component estimation methods are reviewed, highlighting their advantages and shortcomings. In Section 3, we propose an approach based on the EM algorithm to get positive values for the variance component estimates. In Section 4, the results of an MC simulation are presented to assess the performance of the proposed approach, comparing it to two alternative methods recently introduced. In Section 5, we apply our approach to estimate the small area means of monthly per capita expenditure in a food security and poverty study conducted in Mexico; see CIESIN [4] (2005). Conclusions and future research are discussed in Section 6.

2. THE FAY-HERRIOT MODEL

2.1. Formulation

Suppose that there are m small areas labeled as $i = 1, ..., m$. Assuming a $p \times 1$ vector of observed values $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^\top$ for auxiliary variables is available for each area i, Fay and Herriot [13] (1979) proposed their model to improve the direct estimator $\widehat{\theta}_i$ used to compute the true small area mean θ_i, consisting of the following two levels:

- Level 1 (sampling model): $\widehat{\theta}_i|\theta_i \overset{\text{IND}}{\sim} \mathrm{N}(\theta_i, \psi_i)$,
- Level 2 (linking model): $\theta_i \overset{\text{IND}}{\sim} \mathrm{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2), \quad i = 1, ..., m,$

where "IND" stands for "independent", ψ_i is the known variance of the sampling error, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is a vector of unknown regression coefficients to be estimated, and σ^2 is the unknown variance of the area-specific random effect to be estimated. Level 1 accounts for the sampling variability of the survey estimates $\widehat{\theta}_i$ of θ_i, whereas Level 2 links θ_i to the vector of p known area-specific auxiliary variables; see Jiang and Lahiri [20] (2006) and Li and Lahiri [23] (2010). Then, the FH model can be written as

$$\widehat{\theta}_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + b_i + \varepsilon_i, \quad i = 1, ..., m, \tag{2.1}$$

where $b_i \overset{\text{IID}}{\sim} \mathrm{N}(0, \sigma^2)$ are independent and identically distributed (IID) area-specific random effects with unknown σ^2 to be estimated from the data, and $\varepsilon_i \overset{\text{IND}}{\sim} \mathrm{N}(0, \psi_i)$ represent sampling errors with known variances ψ_i. Although in this paper we are considering ψ_i as known, in practical cases when the variances ψ_i are not available, Fay and Herriot [13] (1979) employed generalized variance functions (Wolter [41, Chapter 7], 2007) to estimate them. In addition, it is assumed that b_i and ε_i are independent.

2.2. Estimation of a small area mean

We are interested in estimating or predicting the small area mean $\theta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + b_i$, as well as in obtaining a measurement of uncertainty associated with that prediction. Under the model given in (2.1), the best predictor (BP) of θ_i, which minimizes the mean squared error (MSE), can be expressed by a weighted average of the direct estimator $\widehat{\theta}_i$ and the regression-synthetic estimator $\boldsymbol{x}_i^\top \boldsymbol{\beta}$, being it defined as

$$\widehat{\theta}_i^{\text{BP}} = (1 - B_i)\widehat{\theta}_i + B_i\, \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad i = 1, ..., m, \tag{2.2}$$

with the weight $0 < B_i < 1$ defined as $B_i = \psi_i/(\sigma^2 + \psi_i)$. Note that $(1 - B_i)$ is a function of the variance ratio σ^2/ψ_i and measures the uncertainty when estimating θ_i relative to the total variance $\sigma^2 + \psi_i$; see Rao and Molina [35] (2015). In addition, the parameter σ^2 is a measure of homogeneity of the areas after accounting for the auxiliary variables $\boldsymbol{x}_i$. If σ^2 is known, $\boldsymbol{\beta}$ can be obtained by the standard weighted least squares estimator $\overline{\boldsymbol{\beta}}$; see Mert [27] (2015). Hence, by replacing it in (2.2), one gets the best linear unbiased predictor (BLUP) of θ_i expressed as

$$\widehat{\theta}_i^{\,\mathrm{BLUP}} = (1 - B_i)\widehat{\theta}_i + B_i\, \boldsymbol{x}_i^\top \overline{\boldsymbol{\beta}}, \quad i = 1, ..., m, \tag{2.3}$$

where

$$\overline{\boldsymbol{\beta}} = \frac{\sum_{i=1}^{m} \boldsymbol{x}_i \widehat{\theta}_i/(\sigma^2 + \psi_i)}{\sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^\top/(\sigma^2 + \psi_i)}.$$

The BLUP of θ_i given in (2.3) depends on σ^2, which is unknown in practical applications. Replacing σ^2 in (2.3) with a general estimator, that for now we denote by $\widehat{\sigma}^{\,2}$ (see details in Section 2.3), we obtain the EBLUP of θ_i as

$$\widehat{\theta}_i^{\,\mathrm{EBLUP}} = (1 - \widehat{B}_i)\widehat{\theta}_i + \widehat{B}_i\, \boldsymbol{x}_i^\top \widetilde{\boldsymbol{\beta}}, \tag{2.4}$$

where $\widehat{B}_i$ and $\widetilde{\boldsymbol{\beta}}$ are the estimators of B_i and $\boldsymbol{\beta}$ when σ^2 is replaced with $\widehat{\sigma}^{\,2}$ in (2.2) and (2.3), respectively. Note that the model given in (2.1) can be rewritten in matrix terms as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{I}_m \boldsymbol{b} + \boldsymbol{\varepsilon}, \tag{2.5}$$

where $\boldsymbol{Y} = (Y_1, ..., Y_m)^\top$, with $Y_i = \widehat{\theta}_i$, $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_m)^\top$ is of full rank, $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, $\boldsymbol{\beta}$ is defined as above, $\boldsymbol{b} = (b_1, ..., b_m)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_m)^\top$. In addition, as mentioned in scalar terms, $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$ are independently distributed with $\boldsymbol{b} \sim \mathrm{N}_m(\boldsymbol{0}, \boldsymbol{G})$ and $\boldsymbol{\varepsilon} \sim \mathrm{N}_m(\boldsymbol{0}, \boldsymbol{S})$, for $\boldsymbol{G} = \sigma^2 \boldsymbol{I}_m$ and $\boldsymbol{S} = \mathrm{diag}\{\psi_1, ..., \psi_m\}$. The model defined in (2.5) is a particular case of a more general linear mixed model (Datta *et al.* [10], 2005) with its variance-covariance matrix taking the form $\boldsymbol{V} = \boldsymbol{G} + \boldsymbol{S}$.

2.3. Estimation of σ^2

Note that the EBLUP given in (2.4) depends on the way how σ^2 is estimated. Different methods have been proposed in the literature to estimate σ^2; see Fay and Herriot [13] (1979) and Prasad and Rao [33] (1990). In those cases when the estimate of σ^2 takes a negative value, Prasad and Rao [33] (1990) suggested to truncate the negative estimate at zero. They also showed that the probability of having a negative estimate goes to zero as $m \to \infty$; see Datta [7] (2009). As an alternative, the ML method has been widely used in small area estimation; see Jiang and Lahiri [20] (2006) and Rao and Molina [35] (2015). It was employed by Datta and Lahiri [9] (2000) in the context of the FH model, in whose case the log-likelihood function takes the form

$$\ell_{\mathrm{ML}}(\sigma^2, \boldsymbol{\beta}; \boldsymbol{Y}) = c - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{2.6}$$

where c is a constant that is independent of σ^2. By differentiating (2.6) with respect to $\boldsymbol{\beta}$ and σ^2, we have

$$\frac{\partial \ell_{\mathrm{ML}}(\sigma^2, \boldsymbol{\beta}; \boldsymbol{Y})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{Y} - \boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{\beta}, \tag{2.7}$$

$$\frac{\partial \ell_{\mathrm{ML}}(\sigma^2, \boldsymbol{\beta}; \boldsymbol{Y})}{\partial \sigma^2} = \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{V}^{-2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) - \frac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}). \tag{2.8}$$

Thus, equating (2.7) and (2.8) to zero, and solving them simultaneously with respect to σ^2 and $\boldsymbol{\beta}$, we obtain the ML estimators of σ^2, denoted by $\widehat{\sigma}^2_{\mathrm{ML}}$, and of $\boldsymbol{\beta}$ given in (2.3). If we replace $\boldsymbol{\beta}$ by $\widetilde{\boldsymbol{\beta}}$ in (2.6), we have the corresponding profile log-likelihood (PML) function expressed as

$$\ell_{\mathrm{PML}}(\sigma^2; \boldsymbol{Y}) = c - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}\boldsymbol{Y}^\top \boldsymbol{P}\boldsymbol{Y}, \tag{2.9}$$

where $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{V}^{-1}$. Equating (2.9) to zero and solving it with respect to σ^2, we have an estimator of σ^2 identical to that obtained through (2.6) by means of the ML method; see Jiang [19] (2007). Consequently, the associated estimate computed with the PML method is not analyzed here. Datta and Lahiri [9] (2000) obtained both the asymptotic variance and bias of $\widehat{\sigma}^2_{\mathrm{ML}}$, given respectively by

$$\begin{aligned} \overline{V}[\widehat{\sigma}^2_{\mathrm{ML}}] &= \frac{2}{\mathrm{tr}(\boldsymbol{V}^{-2})} + o(m^{-1}), \\ \mathrm{Bias}[\widehat{\sigma}^2_{\mathrm{ML}}] &= \frac{\mathrm{tr}(\boldsymbol{P} - \boldsymbol{V}^{-1})}{\mathrm{tr}(\boldsymbol{V}^{-2})} + o(m^{-1}). \end{aligned} \tag{2.10}$$

Note that the ML estimates tend to underestimate the variance components and then the RML estimation is preferred; see Pinheiro and Bates [32] (2004). A feature of the RML method is that, when estimating variance components, it takes into account the degrees of freedom involved in estimating the fixed effects, which is not considered by the ML method; see Searle *et al.* [37] (2006). Several alternative derivations of the RML method have been presented in the literature; see Harville [18] (1977), Jiang [19] (2007) and references therein.

Verbyla [40] (1990) proposed an approach which divides the likelihood function into two independent parts, one related to the fixed effect ($\boldsymbol{Y}_1 = \boldsymbol{L}_1^\top\boldsymbol{Y}$) and the another part related to the residual contrasts $\boldsymbol{Y}_2 = \boldsymbol{L}_2^\top\boldsymbol{Y}$, where $\boldsymbol{L} = [\boldsymbol{L}_1\ \ \boldsymbol{L}_2]$ is a non-singular matrix, with $\boldsymbol{Y}$ given in (2.5) and $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ being $m \times p$ and $m \times (m-p)$ matrices, respectively, both of full column rank, which are chosen to satisfy $\boldsymbol{L}_1^\top\boldsymbol{X} = \boldsymbol{I}_p$ and $\boldsymbol{L}_2^\top\boldsymbol{X} = \boldsymbol{0}$. Therefore, $\boldsymbol{Y}$ is transformed as

$$\boldsymbol{L}^\top\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{L}_1^\top\boldsymbol{Y} \\ \boldsymbol{L}_2^\top\boldsymbol{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} \sim \mathrm{N}_m\left(\begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{L}_1^\top\boldsymbol{V}\boldsymbol{L}_1 & \boldsymbol{L}_1^\top\boldsymbol{V}\boldsymbol{L}_2 \\ \boldsymbol{L}_2^\top\boldsymbol{V}\boldsymbol{L}_1 & \boldsymbol{L}_2^\top\boldsymbol{V}\ \boldsymbol{L}_2 \end{bmatrix} \right).$$

The probability density function (PDF) of $\boldsymbol{L}^\top\boldsymbol{Y}$ can be expressed as the product of the conditional PDF of $\boldsymbol{Y}_1$ given $\boldsymbol{Y}_2$ and the marginal PDF of $\boldsymbol{Y}_2$. Hence, the log-likelihood function of $\boldsymbol{L}^\top\boldsymbol{Y}$ is $\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{L}^\top\boldsymbol{Y}) = \ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{Y}_1|\boldsymbol{Y}_2) + \ell(\sigma^2; \boldsymbol{Y}_2)$. Since $\boldsymbol{Y}_1$ is a $p \times 1$ vector and $\ell(\sigma^2; \boldsymbol{Y}_2)$ is not a function of $\boldsymbol{\beta}$, the fixed effects are estimated from $\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{Y}_1|\boldsymbol{Y}_2)$. Once $\boldsymbol{\beta}$ has been estimated, there is no information left for estimating σ^2 and $\ell(\sigma^2; \boldsymbol{Y}_2)$ is used for estimating σ^2. The function $\ell(\sigma^2; \boldsymbol{Y}_2)$ is known as residual or restricted log-likelihood function, from which the RML estimator is obtained. Then, under the FH model defined in (2.5), it is expressed as

$$\ell_{\mathrm{RML}}(\sigma^2; \boldsymbol{Y}_2) = c - \frac{1}{2}\log(|\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}|) - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}\boldsymbol{Y}^\top\boldsymbol{P}\boldsymbol{Y}. \tag{2.11}$$

Thus, the RML estimator of σ^2, $\widehat{\sigma}^2_{\mathrm{RML}}$ namely, is generated as a solution of the equation

$$\frac{\partial \ell_{\mathrm{RML}}(\sigma^2; \boldsymbol{Y}_2)}{\partial \sigma^2} = 0.$$

Datta and Lahiri [9] (2000) showed that $\overline{V}[\widehat{\sigma}^2_{\text{RML}}]$ is identical to $\overline{V}[\widehat{\sigma}^2_{\text{ML}}]$ given in (2.10), whereas $\widehat{\sigma}^2_{\text{RML}}$ is asymptotically unbiased for σ^2. All above estimators hold the following properties: (i) they are $m^{\frac{1}{2}}$-consistent, that is, $\widehat{\sigma}^2 - \sigma^2 = O(m^{\frac{1}{2}})$; (ii) they are even functions of $\boldsymbol{Y}$ and hence $\widehat{\sigma}^2(-\boldsymbol{Y}) = \widehat{\sigma}^2(\boldsymbol{Y})$; and (iii) they are invariant functions under translation and so $\widehat{\sigma}^2(\boldsymbol{Y}+\boldsymbol{X}\boldsymbol{d}) = \widehat{\sigma}^2(\boldsymbol{Y})$, for any $\boldsymbol{d} \in \mathbb{R}^p$ and for all $\boldsymbol{Y}$; see Datta *et al.* [10] (2005). In contrast to these properties (i)-(iii), the FH, ML, PR and RML methods can provide non-admissible negative or zero values for the estimates of $\widehat{\sigma}^2$, especially when the number of small areas is low; see Li and Lahiri [23] (2010) and Yoshimori and Lahiri [42] (2014). However, as happens in practice with any of these methods, the estimate $\widehat{\sigma}^2 = \max(\widehat{\sigma}^2_{\text{M}}, 0)$ is used, where "M" indicates the FH, PR, ML or RML method. Then, if $\widehat{\sigma}^2 = 0$ (when the Level 2 model is perfect), the EBLUP in (2.4) reduces to the simple regression-synthetic estimator (since $\widehat{B}_i = 1$), which typically has an overshrinking problem. Thus, as mentioned by Li and Lahiri [23] (2010), this situation is unrealistic, because Level 2 model cannot be perfect and $\widehat{\sigma}^2$ should be always greater than zero. To solve the problem of a negative or zero value for the variance component estimate, various methods have been proposed. Li and Lahiri [23] (2010) adjusted the ML (LML) method defining a product of σ^2 and a standard likelihood function, introducing the adjusted log-likelihood function $\ell_{\text{LML}}(\sigma^2;\boldsymbol{Y}) = \ell(\sigma^2;\boldsymbol{Y}) + \log(\sigma^2)$, where $\ell(\sigma^2;\boldsymbol{Y})$ may be chosen from (2.9) or (2.11). Its maximization produces the LML and Li-Lahiri RML (LRML) estimators of σ^2, denoted as $\widehat{\sigma}^2_{\text{LML}}$ and $\widehat{\sigma}^2_{\text{LRML}}$, respectively. Both $\widehat{\sigma}^2_{\text{LML}}$ and $\widehat{\sigma}^2_{\text{LRML}}$ are strictly positive, even for small m. Li and Lahiri [23] (2010) showed that their asymptotic variances are as given in (2.10). In addition, the corresponding biases are expressed as

$$\begin{aligned}
\text{Bias}[\widehat{\sigma}^2_{\text{LML}}] &= \frac{\text{tr}(\boldsymbol{P}-\boldsymbol{V}^{-1}) + 2/\sigma^2}{\text{tr}(\boldsymbol{V}^{-2})} + o(m^{-1}),\\
\text{Bias}[\widehat{\sigma}^2_{\text{LRML}}] &= \frac{2/\sigma^2}{\text{tr}(\boldsymbol{V}^{-2})} + o(m^{-1}).
\end{aligned}$$

Yoshimori and Lahiri [42] (2014) proposed other adjusted ML method, with adjusted likelihood function defined as the product of a function $h(\sigma^2)$ and a standard likelihood function. In this case, the adjusted log-likelihood function is defined as

$$\ell_{\text{YML}}(\sigma^2;\boldsymbol{Y}) = \ell(\sigma^2;\boldsymbol{Y}) + \log(h(\sigma^2)), \tag{2.12}$$

where $\ell(\sigma^2;\boldsymbol{Y})$ expressed in (2.12) can be chosen from (2.9) or (2.11), and $h(\sigma^2) = (\tan^{-1}(\text{tr} - (\boldsymbol{I}_m - \boldsymbol{B})))^{\frac{1}{m}}$, with $\boldsymbol{B} = \text{diag}\{B_1, \dots, B_m\}$ and B_i as defined in (2.2). Thus, the Yoshimori-Lahiri ML (YML) and Yoshimori-Lahiri RML (YRML) estimators of σ^2, denoted by $\widehat{\sigma}^2_{\text{YML}}$ and $\widehat{\sigma}^2_{\text{YRML}}$, respectively, are obtained by maximizing (2.12) with respect to σ^2. Both $\widehat{\sigma}^2_{\text{YML}}$ and $\widehat{\sigma}^2_{\text{YRML}}$ are also strictly positive, even for small m. Yoshimori and Lahiri [42] (2014) showed that their asymptotic variances are identical as in (2.10). In addition, we have that

$$\text{Bias}[\widehat{\sigma}^2_{\text{YML}}] = \frac{\text{tr}(\boldsymbol{P}-\boldsymbol{V}^{-1})}{\text{tr}(\boldsymbol{V}^{-2})} + o(m^{-1})$$

and $\widehat{\sigma}^2_{\text{YRML}}$ is asymptotically unbiased for σ^2.

2.4. Uncertainty of the EBLUP

A measure of uncertainty of the EBLUP of θ_i given in (2.4) is obtained by its mean squared predicted error (MSPE), also known as MSE or predicted mean squared error (Rao and Molina [35, Section 5.2, p. 119], 2015), defined by

$$\text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}] = \text{E}[\widehat{\theta}_i^{\,\text{EBLUP}} - \theta_i]^2, \tag{2.13}$$

which, under certain regularity conditions, can be decomposed as (Datta *et al.* [10], 2005)

$$\text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}] = g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + \text{E}[\widehat{\theta}_i^{\,\text{EBLUP}} - \widehat{\theta}_i^{\,\text{BLUP}}]^2, \tag{2.14}$$

where

$$g_{1i}(\sigma^2) = \frac{\sigma^2\psi_i}{\sigma^2+\psi_i}, \quad g_{2i}(\sigma^2) = \frac{\psi_i^2}{(\sigma^2+\psi_i)^2}\boldsymbol{x}_i^\top\left(\sum_{j=1}^{m}\frac{1}{\sigma^2+\psi_i}\boldsymbol{x}_j\boldsymbol{x}_j^\top\right)^{-1}\boldsymbol{x}_i. \tag{2.15}$$

The term $g_{1i}(\sigma^2)$ is of order $O(1)$, which captures the uncertainty of the BP given in (2.2), whereas the term $g_{2i}(\sigma^2)$ is of order $O(m^{-1})$, capturing the uncertainty due to the estimation of $\boldsymbol{\beta}$. The last term in (2.14) considers the uncertainty due to the estimation of σ^2. Ignoring this term seriously underestimates the MSPE. However, there is no closed-form expression available for it, but an approximation of order $O(m^{-1})$ can be expressed by (Li and Lahiri [23], 2010)

$$\text{E}[\widehat{\theta}_i^{\,\text{EBLUP}} - \widehat{\theta}_i^{\,\text{BLUP}}]^2 = g_{3i}(\sigma^2) + o(m^{-1}),$$

where

$$g_{3i}(\sigma^2) = \frac{\psi_i^2}{(\sigma^2+\psi_i)^3}\overline{V}[\widehat{\sigma}^{\,2}]. \tag{2.16}$$

Therefore, a second-order approximation to $\text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}]$ in (2.13) or (2.14), under certain regularity conditions, is defined as

$$\text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}] = g_{1i}(\sigma^2) + g_{2i}(\sigma^2) + g_{3i}(\sigma^2) + o(m^{-1}). \tag{2.17}$$

It is noteworthy that both terms $g_{1i}(\sigma^2)$ and $g_{2i}(\sigma^2)$ given in (2.17) do not depend on the estimation method for σ^2 or B_i, but σ^2 affects the term $g_{3i}(\sigma^2)$ through $\overline{V}[\widehat{\sigma}^{\,2}]$. For the FH model, Datta *et al.* [10] (2005) and Datta [7] (2009) showed that the term $g_{3i}(\sigma^2)$ is the smallest in the ML and RML methods, but it is the largest in the PR and FH methods.

Note that $\text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}]$ defined in (2.17) depends on σ^2, which is unknown and hence cannot be used to asses the uncertainty of the EBLUP for a certain data set. Then, it is of interest to obtain a second-order unbiased estimator of $\text{MSPE}(\widehat{\theta}_i^{\,\text{EBLUP}})$, denoted as $\widehat{\text{MSPE}}[\widehat{\theta}_i^{\,\text{EBLUP}}]$, which must satisfy

$$\text{E}[\widehat{\text{MSPE}}[\widehat{\theta}_i^{\,\text{EBLUP}})] - \text{MSPE}[\widehat{\theta}_i^{\,\text{EBLUP}}] = o(m^{-1}).$$

Datta and Lahiri [9] (2000) derived a standard second-order unbiased approximation to the MSPE of the EBLUP, which is valid for all estimation methods of σ^2 discussed in this paper, and given by

$$\widehat{\text{MSPE}}[\widehat{\theta}_i^{\,\text{EBLUP}}] = g_{1i}(\widehat{\sigma}^{\,2}) + g_{2i}(\widehat{\sigma}^{\,2}) + 2g_{3i}(\widehat{\sigma}^{\,2}) - \widehat{B}_i^2\widehat{\text{Bias}}[\widehat{\sigma}^{\,2}], \tag{2.18}$$

where $g_{1i}(\widehat{\sigma}^{\,2})$, $g_{2i}(\widehat{\sigma}^{\,2})$ and $g_{3i}(\widehat{\sigma}^{\,2})$ are defined in (2.15) and (2.16), respectively, when σ^2 is replaced by $\widehat{\sigma}^{\,2}$ and $\widehat{\text{Bias}}[\widehat{\sigma}^{\,2}]$ is a second-order unbiased estimator of $\text{Bias}[\widehat{\sigma}^{\,2}]$. It is important to note that a disadvantage of the method proposed by Li and Lahiri [23] (2010) for estimating σ^2 is that it can yield a negative value for the corresponding MSPE; see Yoshimori and Lahiri [42] (2014).

3. EM ALGORITHM IN THE ML ESTIMATION OF σ^2

3.1. The EM algorithm

Let $\boldsymbol{Y}_o$ be the random vector corresponding to the observed data $\boldsymbol{y}_o$, and $\boldsymbol{\theta}$ the parameter of interest corresponding to a $d \times 1$ vector with parameter space Θ. The vector $\boldsymbol{y}_o$ is viewed as being incomplete and is regarded as an observable function of the complete data. The random vector $\boldsymbol{Y}_c = (\boldsymbol{Y}_o^\top, \boldsymbol{U}^\top)^\top$ corresponds to the complete-data vector $\boldsymbol{y}_c = (\boldsymbol{y}_o^\top, \boldsymbol{u}^\top)^\top$, where $\boldsymbol{U}$ is the random vector associated with $\boldsymbol{u}$, the vector of unobserved or missing data. Let $\ell(\boldsymbol{\theta}|\boldsymbol{y}_o)$ be the log-likelihood function for $\boldsymbol{\theta}$ based on observed data. The EM algorithm approaches the problem of solving the incomplete-data likelihood equation $\partial\ell(\boldsymbol{\theta}|\boldsymbol{y}_o)/\partial\boldsymbol{\theta} = 0$ indirectly by proceeding in an iterative form in terms of the complete-data log-likelihood function, $\ell(\boldsymbol{\theta}|\boldsymbol{y}_c)$. As it is unobservable, it is replaced by its conditional expectation given $\boldsymbol{Y}_o = \boldsymbol{y}_o$, using a current estimate of $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}^{(0)}$ be a starting value for $\boldsymbol{\theta}$. Then, on the first iteration, the E-step of the EM algorithm requires the calculation of

$$Q \equiv Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)}) = \mathrm{E}[\ell(\boldsymbol{\theta}|\boldsymbol{Y}_c)|\boldsymbol{Y}_o, \boldsymbol{\theta}^{(0)}], \tag{3.1}$$

whereas its M-step needs the maximization of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$ with respect to $\boldsymbol{\theta}$ over the parameter space $\boldsymbol{\Theta}$. Hence, we choose $\boldsymbol{\theta}^{(1)}$ such that $Q(\boldsymbol{\theta}^{(1)}|\boldsymbol{\theta}^{(0)}) \geq Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)})$, for all $\boldsymbol{\theta} \in \Theta$. The E-step and M-step must be iterated until reaching convergence, for example, when $|\ell(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{Y}_o) - \ell(\boldsymbol{\theta}^{(r)}|\boldsymbol{Y}_o| < 10^{-5}$, where $\widehat{\boldsymbol{\theta}}^{(r+1)}$ is the current ML estimate of $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}^{(r)}$ is its previous estimate; see McLachlan and Krishnan [26, pp. 18-20] (2008). Thus, the $(r+1)$-th iteration of the EM algorithm consists of an E-step followed by an M-step described as:

E-step: Given $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{(r)}$, compute $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)}) = \mathrm{E}[\ell(\boldsymbol{\theta}|\boldsymbol{Y}_c)|\boldsymbol{Y}_o, \widehat{\boldsymbol{\theta}}^{(r)}]$.

M-step: Find $\widehat{\boldsymbol{\theta}}^{(r+1)}$ maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)})$ such that $Q(\boldsymbol{\theta}^{(r+1)}|\widehat{\boldsymbol{\theta}}^{(r)}) \geq Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(r)})$, for all $\boldsymbol{\theta} \in \Theta$.

3.2. The EM algorithm in the ML method for small area estimation

To solve the problem of negative or zero values when estimating the strictly positive variance components mentioned in Section 2.3, we propose to use the EM algorithm. Then, we derive the MLEM and RMLEM approaches. Let $\boldsymbol{Y}_o = \boldsymbol{Y}$, $\boldsymbol{U} = \boldsymbol{b}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^\top$. From (2.5), we have that

$$\boldsymbol{Y}_c = \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{b} \end{pmatrix} \sim \mathrm{N}_{2m}\left(\begin{bmatrix} \boldsymbol{X\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{V} & \sigma^2\boldsymbol{I}_m \\ \sigma^2\boldsymbol{I}_m & \sigma^2\boldsymbol{I}_m \end{bmatrix}\right),$$

with $\boldsymbol{V} = \boldsymbol{G} + \boldsymbol{S}$ given below (2.5). Then, the distribution of $\boldsymbol{b}$ conditional on $\boldsymbol{Y}$ is $\boldsymbol{b}|\boldsymbol{Y} = \boldsymbol{y} \sim \mathrm{N}_m(\sigma^2\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}), \sigma^2(\boldsymbol{I}_m - \sigma^2\boldsymbol{V}^{-1}))$. Thus, the log-likelihood function for $\boldsymbol{\theta}$ based on $\boldsymbol{y}_c$ can be expressed as $\ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}_c) = \ell(\boldsymbol{\beta}, \sigma^2; \boldsymbol{y}|\boldsymbol{b}) + \ell(\sigma^2; \boldsymbol{b})$. Hence, we have that

$$\ell(\beta, \sigma^2; \boldsymbol{y}_c) = c - \frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2}\boldsymbol{\varepsilon}^\top\boldsymbol{S}^{-1}\boldsymbol{\varepsilon} - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\boldsymbol{b}^\top\boldsymbol{b}, \tag{3.2}$$

where $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{b}$ and c is a constant that is independent of σ^2.

Let $Q_1 \equiv Q_1(\boldsymbol{\beta}, \sigma^2|\boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}})$. By eliminating the constant term in (3.2) and according to (3.1), we obtain $Q_1 = \mathrm{E}[\ell(\boldsymbol{\beta}, \sigma^2|\boldsymbol{Y}_c)|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}}]$ as

(3.3)
$$Q_1 = -\frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2}\mathrm{E}[\boldsymbol{\varepsilon}^\top \boldsymbol{S}^{-1}\boldsymbol{\varepsilon}|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}}] - \frac{1}{2}\log(|\sigma^2 \boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\mathrm{E}[\boldsymbol{b}^\top \boldsymbol{b}|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}}].$$

After some algebraic steps, we get

$$\mathrm{E}[\boldsymbol{\varepsilon}^\top \boldsymbol{S}^{-1}\boldsymbol{\varepsilon}|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}}] = \mathrm{tr}(\boldsymbol{S}^{-1}(\widetilde{\boldsymbol{\varepsilon}}_1\, \widetilde{\boldsymbol{\varepsilon}}_1^\top + \mathrm{Var}[\boldsymbol{b}_1])), \tag{3.4}$$

with $\sigma^{2^{(0)}}$ and $\boldsymbol{\beta}^{(0)}$ being starting values for σ^2 and $\boldsymbol{\beta}$, respectively, where $\widetilde{\boldsymbol{\varepsilon}}_1 = \mathrm{E}[\boldsymbol{\varepsilon}|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}\sigma^{2^{(0)}}]$ $= \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \widetilde{\boldsymbol{b}}_1$, and

$$\widetilde{\boldsymbol{b}}_1 = \left(\frac{1}{\sigma^{2^{(0)}}}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)^{-1} \boldsymbol{S}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(0)}), \quad \mathrm{Var}[\boldsymbol{b}_1] = \left(\frac{1}{\sigma^{2^{(0)}}}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)^{-1}.$$

In addition, we have that

$$\mathrm{E}[\boldsymbol{b}^\top \boldsymbol{b}|\boldsymbol{Y}, \boldsymbol{\beta}^{(0)}, \sigma^{2^{(0)}}] = \mathrm{tr}(\widetilde{\boldsymbol{b}}_1\, \widetilde{\boldsymbol{b}}_1^\top + \mathrm{Var}[\boldsymbol{b}_1]), \tag{3.5}$$

so that substituting (3.4) and (3.5) in (3.3), we obtain

$$Q_1 = -\frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2}\widetilde{\boldsymbol{\varepsilon}}_1^\top \boldsymbol{S}^{-1}\widetilde{\boldsymbol{\varepsilon}}_1 - \frac{1}{2}\log(|\sigma^2 \boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{b}}_1^\top \widetilde{\boldsymbol{b}}_1 - \frac{1}{2}\mathrm{tr}\left(\left(\frac{1}{\sigma^2}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)\mathrm{Var}[\boldsymbol{b}_1]\right).$$

Maximizing Q_1 with respect to $\boldsymbol{\beta}$ and σ^2, we get

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{S}^{-1}(\boldsymbol{Y} - \widetilde{\boldsymbol{b}}_1), \quad \widehat{\sigma^2}^{(1)} = \frac{1}{m}\left(\widetilde{\boldsymbol{b}}_1^\top \widetilde{\boldsymbol{b}}_1 + \mathrm{tr}\left(\frac{1}{\sigma^{2^{(0)}}}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)^{-1}\right).$$

Thus, the first main result of this study based on the EM algorithm, for the ML method used in the FH model, is described as follows:

Step 0. Set $r = 0$ and choose starting values $\boldsymbol{\beta}^{(0)}$ and $\sigma^{2^{(0)}}$.

Step 1. For $r \geq 0$, calculate

$$\widetilde{\boldsymbol{b}}_1^{(r+1)} = \left(\frac{1}{\sigma^{2^{(r)}}}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)^{-1} \boldsymbol{S}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{(r)}).$$

Step 2. For $r \geq 0$, compute

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(r+1)} &= (\boldsymbol{X}^\top \boldsymbol{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{S}^{-1}(\boldsymbol{Y} - \widetilde{\boldsymbol{b}}_1^{(r+1)}), \\ \widehat{\sigma^2}^{(r+1)} &= \frac{1}{m}\left(\widetilde{\boldsymbol{b}}_1^{(r+1)\top}\, \widetilde{\boldsymbol{b}}_1^{(r+1)} + \mathrm{tr}\left(\frac{1}{\widehat{\sigma^2}^{(r)}}\boldsymbol{I}_m + \boldsymbol{S}^{-1}\right)^{-1}\right). \end{aligned}$$

Step 3. Iterate Steps 1 and 2 from $r = 1$ until reaching convergence when the difference in absolute value between the iterations $(r+1)$-th and r-th is less than a small preset precision value (for example10^{-5}).

The EM algorithm generates positive definite matrix estimates in Step 2, if the starting matrix is positive definite according to Thompson and Meyer [38] (1986), Searle *et al.* [37] (2006), Demidenko [11] (2013) and El-Leithy *et al.* [12] (2016) in the context of mixed models.

3.3. EM algorithm in the RML method for small area estimation

The joint distribution of $\boldsymbol{b}$ and $\boldsymbol{Y}_2$ (defined in Section 2.3) is given by

$$\begin{pmatrix}\boldsymbol{Y}_2\\ \boldsymbol{b}\end{pmatrix} \sim \mathrm{N}_{2m-p}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \begin{bmatrix}\boldsymbol{L}_2^\top \boldsymbol{V}\boldsymbol{L}_2 & \sigma^2\boldsymbol{L}_2^\top\\ \sigma^2\boldsymbol{L}_2 & \sigma^2\boldsymbol{I}_m\end{bmatrix}\right).$$

From results of the multivariate normal distribution and after some matrix operations, we have that the distribution of $\boldsymbol{b}$ conditional on $\boldsymbol{Y}_2$ is $\boldsymbol{b}|\boldsymbol{Y}_2 = \boldsymbol{y}_2 \sim \mathrm{N}_m(\sigma^2\boldsymbol{P}\boldsymbol{y}_2,\, \sigma^2(\boldsymbol{I}_m - \sigma^2\boldsymbol{P}))$. Since we propose to use the EM algorithm to estimate σ^2 with the RML method, we rewrite the log-likelihood function for $\boldsymbol{Y}_c$ in (3.2) as

$$\ell(\sigma^2;\boldsymbol{Y}_c) = c - \frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2}\boldsymbol{\varepsilon}^\top\boldsymbol{S}^{-1}\boldsymbol{\varepsilon} - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\boldsymbol{b}^\top\boldsymbol{b}. \tag{3.6}$$

Then, we maximize it conditional on $\boldsymbol{Y}_2$.

Let $Q_2 \equiv Q_2(\sigma^2|\sigma^{2^{(0)}})$. By eliminating the constant term in (3.6) and according to (3.1), we have that $Q_2 = \mathrm{E}[\ell(\sigma^2|\boldsymbol{Y}_c)|\boldsymbol{Y}_2, \sigma^{2^{(0)}}]$ is such that

$$Q_2 = -\frac{1}{2}\log(|\boldsymbol{S}|) - \frac{1}{2}\mathrm{E}[\boldsymbol{\varepsilon}^\top\boldsymbol{S}^{-1}\boldsymbol{\varepsilon}|\boldsymbol{Y}_2,\sigma^{2^{(0)}}] - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\mathrm{E}[\boldsymbol{b}^\top\boldsymbol{b}|\boldsymbol{Y}_2,\sigma^{2^{(0)}}]. \tag{3.7}$$

After some algebraic steps, we obtain

$$\mathrm{E}[\boldsymbol{\varepsilon}^\top\boldsymbol{S}^{-1}\boldsymbol{\varepsilon}|\boldsymbol{Y}_2,\sigma^{2^{(0)}}] = \mathrm{tr}(\boldsymbol{S}^{-1}(\widetilde{\boldsymbol{\varepsilon}}_2\,\widetilde{\boldsymbol{\varepsilon}}_2^\top + \mathrm{Var}[\boldsymbol{b}_2])), \tag{3.8}$$

where $\mathrm{Var}[\boldsymbol{b}_2] = \sigma^{2^{(0)}}(\boldsymbol{I}_m - \sigma^{2^{(0)}}\boldsymbol{P}^{(0)})$, $\widetilde{\boldsymbol{\varepsilon}}_2 = \mathrm{E}[\boldsymbol{\varepsilon}|\boldsymbol{Y}_2,\sigma^{2^{(0)}}] = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \widetilde{\boldsymbol{b}}_2$, with $\widetilde{\boldsymbol{b}}_2 = \sigma^{2^{(0)}}\boldsymbol{P}^{(0)}\boldsymbol{Y}$ and $\boldsymbol{P}^{(0)}$ being a starting value for $\boldsymbol{P}$. In addition, we have that

$$\mathrm{E}[\boldsymbol{b}^\top\boldsymbol{b}|\boldsymbol{Y}_2,\sigma^{2^{(0)}}] = \mathrm{tr}(\widetilde{\boldsymbol{b}}_2\,\widetilde{\boldsymbol{b}}_2^\top + \mathrm{Var}[\boldsymbol{b}_2]), \tag{3.9}$$

so that substituting (3.8) and (3.9) in (3.7), it conducts to

$$Q_2 = -\frac{1}{2}\log(|\boldsymbol{S}| - \frac{1}{2}\widetilde{\boldsymbol{\varepsilon}}_2^\top\boldsymbol{S}^{-1}\widetilde{\boldsymbol{\varepsilon}}_2 - \frac{1}{2}\log(|\sigma^2\boldsymbol{I}_m|) - \frac{1}{2\sigma^2}\widetilde{\boldsymbol{b}}_2^\top\widetilde{\boldsymbol{b}}_2 - \frac{1}{2}\mathrm{tr}((\sigma^2\boldsymbol{I}_m + \boldsymbol{S}^{-1})\,\mathrm{Var}[\boldsymbol{b}_2]).$$

Maximizing Q_2 with respect to σ^2, we obtain $\sigma^{2^{(1)}} = (1/m)(\widetilde{\boldsymbol{b}}_2^\top\,\widetilde{\boldsymbol{b}}_2 + \mathrm{tr}(\sigma^{2^{(0)}}(\boldsymbol{I}_m - \sigma^{2^{(0)}}\boldsymbol{P}^{(0)})))$. Thus, the second main result of this study based on the EM algorithm, for the RML method in the FH model, is summarized as follows:

Step 0. Set $s = 0$, and choose a starting value $\sigma^{2^{(0)}}$.

Step 1. For $s \geq 0$, calculate $\widetilde{\boldsymbol{b}}_2^{(s+1)} = \widehat{\sigma^2}^{(s)}\boldsymbol{P}^{(s)}\boldsymbol{Y}$.

Step 2. For $s \geq 0$, compute $\widehat{\sigma^2}^{(s+1)} = (1/m)(\widetilde{\boldsymbol{b}}_2^{(s+1)\top}\,\widetilde{\boldsymbol{b}}_2^{(s+1)} + \mathrm{tr}(\widehat{\sigma^2}^{(s)}(\boldsymbol{I}_m - \widehat{\sigma^2}^{(s)}\boldsymbol{P}^{(s)})))$.

Step 3. Iterate Steps 1 and 2 from $r = 1$ until reaching convergence when the difference in absolute value between the iterations $(r+1)$-th and r-th is less a small preset precision value (for example10^{-5}).

4. MONTE CARLO SIMULATION STUDY

4.1. Scenario of the simulation

We present a multi-sample MC simulation study to compare the performance of the approaches proposed in this paper (MLEM and RMLEM) as alternative solutions to the problem of a negative or zero value in the estimate of σ^2. The multi-sample simulation is a common practice in MC procedures whenever we do not have an easy way to estimate measures of dispersion of a statistic, like for the MSE; see Fishman [15] (1973), Figueiredo and Gomes [14] (2004) and Gomes *et al.* [16, 17] (2011, 2016) for details about multi-sample MC simulation. The idea is reasonably simple: in a multi-sample simulation of size $R \times T$, instead of generating a sample of very large size of observed values of a statistic, $N_{\text{sim}} = R \times T$ say, we collect T observations of the statistic on each of the R independent replications of the experiment. The value of T also needs to be large enough to reduce the bias, and eventually provide asymptotic normality. Then, we take the average of the corresponding R estimates as overall estimate of the parameter of interest, where each estimate is computed from T runs. Thus, under very broad conditions, the overall estimator converges to normality as R increases. Moreover, we may estimate the standard error (SE) of this overall estimator, even if R is small. For small values of R, and whenever we may guarantee the asymptotic normality of the estimator for the parameter of interest, we may use the t-student distribution with $R-1$ degrees of freedom to approximate its true distribution. The performance of the approaches proposed in this paper is compared to the LML, LRML, YML and YRML methods, according to their percentage relative bias (PRB) and MSE, as well as the MSPE of the EBLUP estimator. We follow the same scenario used in Yoshimori and Lahiri [42] (2014) to do an effective comparison in relation to that work. Specifically, we consider the FH model defined in (2.1) with a common mean $\mu = \boldsymbol{x}_i^\top \boldsymbol{\beta}$. As the MSE is invariant under translation, we set $\mu = 0$ without loss of generality. However, to account for the uncertainty in the estimation of the common mean that arises in practice, we treat the mean as unknown. We generate $R = 20$ independent MC replications with $T = 500$ runs ($N_{\text{sim}} = 20 \times 500 = 10,000$) of $\{Y_i,\, i = 1, ..., m\}$ using the FH model: $Y_i = b_i + \varepsilon_i$, where b_i and ε_i are independent with $b_i \overset{\text{IID}}{\sim} \mathrm{N}(0, \sigma^2)$ and $\varepsilon_i \overset{\text{IND}}{\sim} \mathrm{N}(0, \psi_i)$. We analyze both balanced (equal sampling variances ψ_i) and unbalanced (unequal sampling variances ψ_i) cases for different values of m. To examine the effect of the number of small areas on the performance of several estimators, we use values of $m \in \{15, 30, 45\}$. In the balanced case, we consider each of the combinations of m and ψ_i, where $\psi_i \in \{0.05, 0.1, 1, 10, 20\}$, fixing $\sigma^2 = 1$. We also examine the effect of σ^2/ψ_i on the performance of the estimators as in Yoshimori and Lahiri [42] (2014). In the unbalanced case, we also fix $\sigma^2 = 1$ and assume the following three patterns of sampling variances as in Yoshimori and Lahiri [42] (2014): (i) Pattern A, $\psi_i \in \{0.1, 0.4, 0.5, 0.6, 4.0\}$, where almost all (but one) of the sampling variances are smaller than σ^2; (ii) Pattern B, $\psi_i \in \{1.5, 2.0, 2.5, 3.0, 3.5\}$, where all sampling variances are slightly greater than σ^2; and (iii) Pattern C, $\psi_i \in \{2, 4, 5, 6, 20\}$, where not only are all sampling variances greater than σ^2, but one is much greater than σ^2, representing a case for extremely small area. Pattern A was also used by Datta and Lahiri [9] (2000) and Datta *et al.* [10] (2005), and Pattern C by Chen and Lahiri [3] (2008) in their simulation studies. In each pattern, we consider five groups (g) of small areas, each with three, six or nine small areas according to $m = 15$, $m = 30$ or

$m = 45$, such that the sampling variances ψ_i are the same within a given group. For example, in Pattern A for $m = 15$, we simulate three small areas for each case with sampling variances $\psi_i = 0.1, 0.4, 0.5, 0.6$ and 4.0. Similarly the other patterns of sampling variances and m were simulated.

4.2. Behavior of $\widehat{\sigma}^2$

The empirical probabilities of obtaining a zero estimate of σ^2 by different methods for balanced and unbalanced cases are reported in Tables 1 (balanced case) and 2 (unbalanced case). In both cases, the MLEM/RMLEM approaches and the LML/LRML/YML/YRML methods produce strictly positive estimates of σ^2. As mentioned in Yoshimori and Lahiri [42] (2014), only the ML and RML methods could yield negative or zero estimate of σ^2. For the balanced case and the ML/RML methods, the probability of getting negative or zero estimate increases as σ^2/ψ_i decreases in both methods, being slightly smaller in the RML method.

Table 1: Percentage of negative or zero estimate of σ^2 for the indicated m, variance ratio and method.

m	σ^2/ψ_i	ML	RML	LML	LRML	YML	YRML	MLEM	RMLEM
15	0.05	57.57	50.50	0	0	0	0	0	0
	0.1	52.28	44.89	0	0	0	0	0	0
	1	8.49	6.49	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0
30	0.05	51.14	46.09	0	0	0	0	0	0
	0.1	43.91	38.97	0	0	0	0	0	0
	1	1.42	1.09	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0
45	0.05	48.18	44.25	0	0	0	0	0	0
	0.1	39.68	36.02	0	0	0	0	0	0
	1	0.26	0.18	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0

Table 2: Percentage of negative or zero estimate of σ^2 for the indicated m, pattern and method.

m	Pattern	ML	RML	LML	LRML	YML	YRML	MLEM	RMLEM
15	A	0.92	0.37	0	0	0	0	0	0
	B	27.33	21.65	0	0	0	0	0	0
	C	35.37	27.04	0	0	0	0	0	0
30	A	0.06	0.03	0	0	0	0	0	0
	B	13.87	11.29	0	0	0	0	0	0
	C	28.21	23.39	0	0	0	0	0	0
45	A	0	0	0	0	0	0	0	0
	B	7.50	6.24	0	0	0	0	0	0
	C	21.40	18.21	0	0	0	0	0	0

As m increases, this probability decreases and is very similar in both methods. In the unbalanced case, Pattern C (having an extreme value in the sampling variance) yields the largest percentages of negative or zero variance component estimates. Similarly to the balanced case, as m increases, this probability decreases, being smaller for the RML method than the ML method. In the remainder of this section, we consider only the performance of those methods mentioned above that produce strictly positive variance, because these methods solve the problem of inadmissibility presented in this paper (we do not further comparisons for the ML and RML methods). An aspect to be evaluated for assessing the performance of different estimators of σ^2 is their bias. We use the PRB of a given estimator of σ^2, $\widehat{\sigma}^2$ say, defined in our simulation study as the sample mean $\widehat{\mathrm{PRB}}[\widehat{\sigma}^2] = (1/R)\sum_{r=1}^{R}(\widehat{\mathrm{PRB}}_r[\widehat{\sigma}^2])$ of the PRB calculated on the $R = 20$ replications, with $\widehat{\mathrm{PRB}}_r[\widehat{\sigma}^2] = (1/T)\sum_{t=1}^{T}((\widehat{\sigma}^{2(t)} - \sigma^2)/\sigma^2) \times 100$, where $\widehat{\sigma}^{2(r)}$ denotes an estimate of σ^2 for the t-th instance in the r-th replication, with an associated SE defined as $((1/(R-1))\sum_{r=1}^{R}(\widehat{\mathrm{PRB}}_r[\widehat{\sigma}^2] - \widehat{\mathrm{PRB}}[\widehat{\sigma}^2])^2)^{1/2}$. The PRBs of estimators for σ^2 are presented in Tables 3 and 4 for balanced and unbalanced cases, respectively.

Table 3: PRB and its corresponding SE (in parentheses) of estimators of σ^2 for the indicated m, variance ratio and method.

m	σ^2/ψ_i	LML		LRML		YML		YRML		MLEM		RMLEM	
15	0.05	1094.36	(24.35)	1272.41	(27.71)	224.55	(18.36)	309.57	(21.42)	171.28	(19.42)	262.68	(22.69)
	0.1	541.57	(13.50)	636.53	(15.31)	94.98	(10.38)	143.26	(12.07)	69.39	(10.79)	121.47	(12.60)
	1	44.33	(2.96)	62.77	(3.25)	-11.38	(2.68)	1.38	(2.90)	-12.14	(2.71)	0.81	(2.93)
	10	10.61	(2.06)	20.65	(2.23)	-6.94	(1.79)	0.42	(1.92)	-6.95	(1.79)	0.42	(1.92)
	20	9.34	(1.64)	18.86	(1.78)	-6.61	(1.42)	0.42	(1.53)	-6.61	(1.42)	0.42	(1.53)
30	0.05	672.31	(13.29)	737.93	(14.26)	144.30	(11.8)	187.01	(12.62)	124.84	(12.02)	175.96	(12.76)
	0.1	322.88	(9.24)	358.37	(9.88)	53.59	(9.12)	78.31	(9.93)	44.51	(9.43)	73.99	(10.29)
	1	21.76	(2.17)	29.46	(2.26)	-6.19	(2.16)	0.44	(2.24)	-6.24	(2.16)	0.51	(2.24)
	10	4.89	(1.22)	9.14	(1.27)	-3.51	(1.14)	0.16	(1.18)	-3.51	(1.14)	0.17	(1.18)
	20	4.19	(1.26)	8.23	(1.31)	-3.44	(1.18)	0.06	(1.22)	-3.44	(1.18)	0.06	(1.22)
45	0.05	515.15	(17.7)	554.32	(18.67)	111.86	(16.27)	140.66	(17.42)	101.51	(16.38)	139.52	(17.67)
	0.1	243.74	(6.9)	265.20	(7.21)	39.55	(7.13)	56.53	(7.48)	34.93	(7.31)	57.02	(7.65)
	1	13.55	(1.98)	18.40	(2.04)	-4.94	(1.99)	-0.51	(2.03)	-4.95	(1.99)	-0.39	(2.04)
	10	3.08	(0.83)	5.78	(0.85)	-2.43	(0.79)	0.01	(0.81)	- 2.43	(0.79)	0.02	(0.81)
	20	2.70	(0.67)	5.26	(0.68)	-2.32	(0.64)	0.01	(0.65)	-2.32	(0.64)	0.01	(0.65)

Table 4: PRB and its corresponding SE (in parentheses) of estimators of σ^2 for the indicated m, pattern and method.

m	Pattern	LML		LRML		YML		YRML		MLEM		RMLEM	
15	A	26.01	(3.07)	42.66	(3.46)	-9.20	(2.43)	1.99	(2.7)	-9.28	(2.42)	1.97	(2.7)
	B	114.71	(5.43)	146.20	(6.02)	-3.76	(4.90)	15.63	(5.46)	-8.00	(5.09)	12.30	(5.65)
	C	258.95	(7.35)	323.53	(8.43)	24.72	(5.83)	58.94	(6.81)	14.73	(5.86)	48.72	(7.19)
30	A	10.19	(1.33)	16.72	(1.39)	-5.91	(1.24)	-0.51	(1.28)	-5.91	(1.24)	-0.47	(1.28)
	B	58.85	(4.26)	71.31	(4.49)	-6.07	(4.24)	4.24	(4.46)	-7.03	(4.30)	4.02	(4.53)
	C	123.79	(6.31)	145.68	(6.79)	0.97	(5.56)	16.90	(6.04)	-2.10	(5.63)	13.38	(6.20)
45	A	6.47	(1.43)	10.52	(1.47)	-4.00	(1.36)	-0.42	(1.39)	-4.00	(1.36)	-0.38	(1.39)
	B	39.56	(2.73)	47.34	(2.81)	-6.07	(2.87)	1.05	(2.93)	-6.40	(2.87)	1.38	(2.93)
	C	87.02	(4.51)	100.23	(4.73)	-1.16	(4.49)	9.85	(4.71)	-2.68	(4.52)	7.83	(4.95)

In the balanced case, when $\sigma^2/\psi_i < 1$, all methods widely overestimate σ^2, with the best performance in those methods based on the ML method than those based on the RML method. In this case, the performance of the MLEM approach is better having the smallest PRBs. When $\sigma^2/\psi_i \geq 1$, the RMLEM approach and the YRML method are the best options being them very similar. Also, the MLEM approach and the YML method always underestimate σ^2. For the unbalanced case, the performance of the RMLEM approach and the YRML method are very similar and have the smallest PRBs in the following cases: (i) for Pattern A and all m (with the PRBs being always smallest for the proposed RMLEM approach) and (ii) for Pattern B, when $m = 30$ and $m = 45$. The performance of the MLEM approach is better for Pattern C when $m = 15$ than in the other cases. For all remaining situations (Pattern B with $m = 15$ and Pattern C with $m = 30, 45$), the smallest PRBs correspond to the YML method. In both balanced and unbalanced cases, as m increases, the performance of all estimators improves. We define the empirical percentage MSE of an estimator $\widehat{\sigma}^2$ of σ^2 based in our simulation study as (Yoshimori and Lahiri [42], 2014) $\widehat{\mathrm{MSE}}_r[\widehat{\sigma}^2] = (1/T)\sum_{t=1}^{T}\big(\widehat{\sigma}^{2^{(t)}} - \sigma^2\big)^2 \times 100$, with $1 \leq r \leq R$, for the t-th instance in the r-th replication, where the overall MSE is then the sample mean $\widehat{\mathrm{MSE}}[\widehat{\sigma}^2] = (1/R)\sum_{t=1}^{T}\widehat{\mathrm{MSE}}_r[\widehat{\sigma}^2]$, with an associated SE given by $((1/(R-1))\sum_{r=1}^{R}(\widehat{\mathrm{MSE}}_r[\widehat{\sigma}^{\,2}] - \widehat{\mathrm{MSE}}[\widehat{\sigma}^{\,2}])]^2)^{1/2}$. The empirical percentage MSEs of different estimators of σ^2 are shown in Tables 5 (balanced case) and 6 (unbalanced case). In Table 5, the performance of the MLEM approach and the YML method are better than the other ones, with a performance much better when σ^2/ψ_i is small, for all m. In other cases, when σ^2/ψ_i is large, all methods have a similar performance, particularly for $m = 45$, but the performance of the MLEM approach and the YML method are still slightly better than other methods. In the unbalanced case, again the MLEM approach and the YML method have a better performance than the other methods, followed by the RMLEM approach and the YRML method, for all patterns and values of m.

Table 5: Percentage MSE and its corresponding SE (in parentheses) of estimators of σ^2 for the indicated m, variance ratio and method.

m	σ^2/ψ_i	LML		LRML		YML		YRML		MLEM		RMLEM	
15	0.05	15530.4	(815.1)	20780.8	(1043.4)	2522.0	(354.8)	3726.0	(460.9)	2507.6	(352.9)	3722.4	(459.3)
	0.1	4031.4	(223.7)	5457.1	(287.1)	770.5	(98.7)	1114.5	(127.4)	789.7	(98.5)	1133.4	(127.0)
	1	74.5	(6.1)	105.5	(8.0)	46.2	(2.9)	52.9	(3.7)	47.4	(2.8)	53.9	(3.7)
	10	21.1	(1.6)	27.7	(2.1)	15.5	(1.0)	17.2	(1.2)	15.5	(1.0)	17.2	(1.2)
	20	20.0	(1.7)	26.0/	(2.2)	14.8	(1.1)	16.5	(1.4)	14.8	(1.1)	16.5	(1.4)
30	0.05	5940.1	(268.4)	7084.0	(308.8)	1332.7	(147.8)	1711.6	(174.0)	1348.0	(147.9)	1746.4	(174.0)
	0.1	1489.0	(90.5)	1795.1	(104.6)	414.3	(48.1)	517.9	(57.2)	426.2	(47.9)	530.9	(56.9)
	1	31.3	(2.0)	37.5	(2.4)	26.1	(1.6)	27.6	(1.7)	26.1	(1.6)	27.6	(1.7)
	10	9.3	(0.9)	10.6	(1.0)	8.1	(0.7)	8.5	(0.7)	8.1	(0.7)	8.5	(0.7)
	20	8.3	(0.7)	9.4	(0.8)	7.2	(0.5)	7.5	(0.6)	7.2	(0.5)	7.5	(0.6)
45	0.05	3545.6	(250.6)	4063.1	(279.6)	935.9	(138.9)	1133.6	(158.9)	947.8	(138.7)	1163.8	(158.5)
	0.1	888.1	(58.1)	1027.1	(64.3)	307.1	(36.6)	361.1	(41.3)	314.0	(36.6)	368.6	(41.3)
	1	19.2	(1.6)	21.7	(1.8)	17.6	(1.3)	18.1	(1.4)	17.6	(1.3)	18.1	(1.4)
	10	5.9	(0.3)	6.5	(0.4)	5.4	(0.3)	5.6	(0.3)	5.4	(0.3)	5.6	(0.3)
	20	5.2	(0.4)	5.7	(0.4)	4.7	(0.3)	4.9	(0.3)	4.7	(0.3)	4.9	(0.3)

Note that, although the results of the MSE in the estimation of the variance component in our simulation study are similar under the YML and MLEM methods, we observe that the estimation by means of the EM algorithm is slightly more accurate (with smaller SEs) than the estimation under the YML and YRML methods, such as occurs when comparing the YRML and RMLEM methods.

Table 6: Percentage MSE and its corresponding SE (in parentheses) of estimators of σ^2 for the indicated m, pattern and method.

m	Pattern	LML		LRML		YML		YRML		MLEM		RMLEM	
15	A	46.6	(5.8)	66.5	(8.1)	29.4	(2.7)	33.5	(3.7)	29.5	(2.7)	33.5	(3.7)
3	B	261.9	(22.0)	375.9	(28.9)	98.3	(8.6)	124.7	(11.8)	104.8	(8.4)	130.0	(11.5)
	C	990.7	(58.4)	1466.0	(78.5)	207.4	(26.6)	303.4	(34.3)	219.8	(26.8)	315.8	(34.4)
30	A	18.3	(1.7)	21.5	(2.0)	15.1	(1.2)	15.9	(1.4)	15.1	(1.2)	15.9	(1.4)
	B	94.2	(8.9)	117.2	(10.4)	59.9	(3.8)	66.3	(4.8)	61.5	(3.7)	67.2	(4.8)
	C	282.7	(28.5)	361.3	(34.1)	112.5	(13.8)	134.5	(16.9)	117.3	(13.8)	140.6	(16.9)
45	A	11.0	(0.7)	12.3	(0.8)	9.7	(0.5)	10.1	(0.6)	9.7	(0.5)	10.1	(0.6)
	B	54.0	(5.9)	63.6	(6.6)	42.4	(4.0)	44.8	(4.5)	42.9	(4.0)	44.9	(4.6)
	C	162.1	(13.9)	195.5	(15.8)	86.7	(7.9)	97.1	(9.2)	89.1	(7.9)	101.1	(9.2)

4.3. Estimator of the MSPE

We study the performance of the estimators of the MSPE of the EBLUP using different estimators of σ^2. Let $\theta_i^{(t)}$ and $\widehat{\theta}_i^{\,\mathrm{EBLUP}(t)}$ be the empirical mean and EBLUP corresponding to several estimators of σ^2 of the area i for the t-th simulation in the r-th replication, respectively, for $i = 1, ..., m$, $t = 1, ..., T$ and $r = 1, ..., R$, with $T = 500$, $R = 20$ and $N_{\mathrm{sim}} = 10,000$, as mentioned. The empirical MSPE of the EBLUP estimator in the area i for the r-th replication is given by $\widehat{\mathrm{MSPE}}_i^{(r)} = (1/T)\sum_{t=1}^{T}\big(\widehat{\theta}_i^{\,\mathrm{EBLUP}(t)} - \widehat{\theta}_i^{(t)}\big)^2$. Then, the overall MSPE is $\widehat{\mathrm{MSPE}}_i = (1/R)\sum_{r=1}^{R}\widehat{\mathrm{MSPE}}_i^{(r)}$, with an associated SE given by $(\sum_{r=1}^{R}(\widehat{\mathrm{MSPE}}_i^{\,r} - \widehat{\mathrm{MSPE}}_i)^2/(R-1))^{1/2}$. Tables 7 (balanced case) and 8 (unbalanced case) report the empirical MSPEs of EBLUPs for different estimation methods of σ^2. In the balanced case, when σ^2/ψ_i is small, both the MLEM approach and the YML method present a performance slightly better than the RMLEM approach and the YRML method, and much better than the LML and LRML methods. This behavior is more pronounced when m is small. If $\sigma^2/\psi_i \geq 1$, all methods provide similar results, and for a fixed σ^2/ψ_i, the performance does not change substantially as m increases. For the unbalanced case, all methods have a similar performance.

Table 7: Empirical MSPE and its corresponding SE (in parentheses) of EBLUPs of $\widehat{\theta}_i$ for the indicated m, ratio and method.

m	σ^2/ψ_i	LML		LRML		YML		YRML		MLEM		RMLEM	
15	0.05	4.75	(0.15)	5.22	(0.16)	2.86	(0.11)	3.08	(0.12)	2.86	(0.11)	3.08	(0.12)
	0.1	2.68	(0.08)	2.90	(0.08)	1.85	(0.06)	1.96	(0.07)	1.87	(0.06)	1.97	(0.07)
	1	0.58	(0.01)	0.59	(0.01)	0.60	(0.01)	0.59	(0.01)	0.60	(0.01)	0.59	(0.01)
	10	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)
	20	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)
30	0.05	2.95	(0.06)	3.13	(0.07)	1.99	(0.05)	2.08	(0.05)	2.00	(0.05)	2.09	(0.05)
	0.1	1.81	(0.05)	1.90	(0.06)	1.43	(0.04)	1.47	(0.04)	1.44	(0.04)	1.48	(0.04)
	1	0.54	(0.01)	0.55	(0.01)	0.55	(0.01)	0.55	(0.01)	0.55	(0.01)	0.55	(0.01)
	10	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)
	20	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)
45	0.05	2.31	(0.07)	2.40	(0.08)	1.68	(0.06)	1.73	(0.06)	1.69	(0.06)	1.74	(0.06)
	0.1	1.53	(0.03)	1.57	(0.03)	1.29	(0.02)	1.31	(0.02)	1.29	(0.02)	1.32	(0.02)
	1	0.53	(0.01)	0.53	(0.01)	0.54	(0.01)	0.53	(0.01)	0.54	(0.01)	0.53	(0.01)
	10	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)
	20	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)	0.05	(<0.01)

Table 8: Empirical MSPE and its corresponding SE (in parentheses) of EBLUP of $\widehat{\theta}_i$ for indicated m, pattern group and method.

m	Pattern	g	LML		LRML		YML		YRML		MLEM		RMLEM	
15	A	1	0.92	(0.03)	0.94	(0.03)	0.91	(0.03)	0.92	(0.03)	0.92	(0.03)	0.92	(0.03)
		2	0.41	(0.01)	0.41	(0.01)	0.42	(0.01)	0.42	(0.01)	0.43	(0.01)	0.42	(0.01)
		3	0.36	(0.01)	0.36	(0.01)	0.37	(0.01)	0.37	(0.01)	0.37	(0.01)	0.37	(0.01)
		4	0.31	(0.01)	0.31	(0.01)	0.32	(0.01)	0.32	(0.01)	0.32	(0.01)	0.32	(0.01)
		5	0.09	(<0.01)	0.09	(<0.01)	0.10	(<0.01)	0.10	(<0.01)	0.10	(<0.01)	0.10	(<0.01)
	B	1	1.10	(0.04)	1.15	(0.04)	1.02	(0.04)	1.03	(0.04)	1.03	(0.04)	1.04	(0.05)
		2	1.07	(0.04)	1.11	(0.04)	1.00	(0.03)	1.01	(0.03)	1.02	(0.03)	1.02	(0.03)
		3	1.01	(0.04)	1.05	(0.04)	0.95	(0.04)	0.96	(0.04)	0.96	(0.04)	0.97	(0.04)
		4	0.93	(0.03)	0.96	(0.04)	0.88	(0.03)	0.89	(0.03)	0.90	(0.03)	0.90	(0.03)
		5	0.80	(0.04)	0.82	(0.04)	0.77	(0.04)	0.77	(0.04)	0.79	(0.04)	0.79	(0.04)
	C	1	1.70	(0.06)	1.86	(0.07)	1.39	(0.04)	1.43	(0.04)	1.40	(0.04)	1.44	(0.04)
		2	1.68	(0.08)	1.84	(0.09)	1.32	(0.05)	1.37	(0.06)	1.34	(0.05)	1.39	(0.06)
		3	1.61	(0.10)	1.75	(0.10)	1.27	(0.07)	1.32	(0.08)	1.29	(0.07)	1.34	(0.08)
		4	1.50	(0.07)	1.61	(0.07)	1.21	(0.06)	1.25	(0.06)	1.23	(0.06)	1.27	(0.07)
		5	1.12	(0.04)	1.18	(0.04)	0.96	(0.04)	0.98	(0.04)	0.99	(0.04)	1.01	(0.04)
30	A	1	0.86	(0.02)	0.86	(0.03)	0.86	(0.03)	0.86	(0.03)	0.86	(0.03)	0.86	(0.03)
		2	0.40	(0.01)	0.40	(0.01)	0.40	(0.01)	0.40	(0.01)	0.40	(0.01)	0.40	(0.01)
		3	0.35	(0.01)	0.35	(0.01)	0.35	(0.01)	0.35	(0.01)	0.35	(0.01)	0.35	(0.01)
		4	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)
		5	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)
	B	1	0.93	(0.03)	0.94	(0.03)	0.92	(0.03)	0.92	(0.03)	0.92	(0.03)	0.92	(0.03)
		2	0.90	(0.02)	0.91	(0.02)	0.89	(0.02)	0.89	(0.02)	0.89	(0.02)	0.89	(0.02)
		3	0.85	(0.02)	0.86	(0.02)	0.85	(0.02)	0.85	(0.02)	0.85	(0.02)	0.85	(0.02)
		4	0.80	(0.02)	0.81	(0.02)	0.80	(0.02)	0.80	(0.02)	0.81	(0.02)	0.80	(0.02)
		5	0.70	(0.01)	0.71	(0.01)	0.71	(0.01)	0.71	(0.01)	0.72	(0.01)	0.71	(0.01)
	C	1	1.24	(0.04)	1.27	(0.05)	1.17	(0.04)	1.18	(0.04)	1.18	(0.04)	1.18	(0.04)
		2	1.20	(0.03)	1.24	(0.03)	1.11	(0.03)	1.12	(0.03)	1.11	(0.03)	1.13	(0.03)
		3	1.16	(0.03)	1.20	(0.03)	1.08	(0.03)	1.09	(0.03)	1.09	(0.03)	1.10	(0.04)
		4	1.12	(0.03)	1.16	(0.04)	1.03	(0.03)	1.04	(0.03)	1.04	(0.03)	1.05	(0.03)
		5	0.91	(0.02)	0.93	(0.02)	0.87	(0.02)	0.87	(0.02)	0.88	(0.02)	0.89	(0.02)
45	A	1	0.83	(0.02)	0.83	(0.02)	0.83	(0.02)	0.83	(0.02)	0.83	(0.02)	0.83	(0.02)
		2	0.39	(0.01)	0.39	(0.01)	0.39	(0.01)	0.39	(0.01)	0.39	(0.01)	0.39	(0.01)
		3	0.34	(0.01)	0.34	(0.01)	0.34	(0.01)	0.34	(0.01)	0.34	(0.01)	0.34	(0.01)
		4	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)	0.30	(0.01)
		5	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)	0.09	(<0.01)
	B	1	0.88	(0.02)	0.88	(0.02)	0.88	(0.02)	0.88	(0.02)	0.88	(0.02)	0.88	(0.02)
		2	0.85	(0.02)	0.86	(0.02)	0.86	(0.02)	0.86	(0.02)	0.86	(0.02)	0.86	(0.02)
		3	0.80	(0.02)	0.81	(0.02)	0.81	(0.02)	0.81	(0.02)	0.81	(0.02)	0.81	(0.02)
		4	0.75	(0.01)	0.75	(0.01)	0.76	(0.02)	0.76	(0.02)	0.76	(0.02)	0.76	(0.01)
		5	0.67	(0.02)	0.68	(0.02)	0.69	(0.02)	0.69	(0.02)	0.69	(0.02)	0.69	(0.02)
	C	1	1.13	(0.04)	1.14	(0.04)	1.10	(0.03)	1.11	(0.03)	1.10	(0.03)	1.11	(0.03)
		2	1.07	(0.03)	1.09	(0.03)	1.03	(0.03)	1.03	(0.03)	1.03	(0.03)	1.04	(0.03)
		3	1.05	(0.03)	1.07	(0.03)	1.01	(0.03)	1.02	(0.03)	1.02	(0.03)	1.02	(0.03)
		4	1.01	(0.03)	1.03	(0.03)	0.97	(0.03)	0.98	(0.03)	0.98	(0.03)	0.99	(0.03)
		5	0.83	(0.02)	0.84	(0.02)	0.82	(0.02)	0.82	(0.02)	0.83	(0.02)	0.83	(0.02)

5. APPLICATION

We illustrate the results of this study with real-world data taken from http://dx.doi.org/10.7927/H4FF3Q9B; see CIESIN [4] (2005). The small areas in our application correspond to the 32 states of Mexico. Here, m_i is the number of municipalities in the state i that were used for direct estimation within each state, which ranges from 5 to 578 municipalities.

We are interested in modeling the "average monthly per capita food expenditure for rural household in 2000 (AMPCFERH)". The direct estimator $\widehat{\theta}_i$ is available. We use three auxiliary variables: (i) the size of the illiterate population aged 15 years and above (X_1); (ii) the percentage of the population living in rural areas (X_2); and (iii) the fraction of rural households below the food poverty line (X_3). Our small area model is given by

$$\widehat{\theta}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + b_i + \varepsilon_i, \quad i = 1, ..., 32,$$

where $b_i \overset{\text{IID}}{\sim} \text{N}(0, \sigma^2)$ are area-specific random effects with unknown σ^2; $\varepsilon_i \overset{\text{IND}}{\sim} \text{N}(0, \psi_i)$ represent the sampling errors in the area i with known variance ψ_i, which was calculated from the variance of the AMPCFERH within each municipality. Note that $\psi_i \neq \psi_j$, for $i \neq j$, and hence we are in the unbalanced case. We estimate σ^2 with the LML, LRML, YML, YRML, MLEM and RMLEM methods obtaining the following values: $\widehat{\sigma}^2_{\text{LML}} = 105147.20$, $\widehat{\sigma}^2_{\text{LRML}} = 123306.20$, $\widehat{\sigma}^2_{\text{YML}} = 93504.80$, $\widehat{\sigma}^2_{\text{YLML}} = 108737.40$, $\widehat{\sigma}^2_{\text{MLEM}} = 93503.32$ and $\widehat{\sigma}^2_{\text{RMLEM}} = 108735.6$. The results of the EBLUP of $\widehat{\theta}_i$ generated under different estimators of σ^2 are shown in Table 9.

Table 9: EBLUP of $\widehat{\theta}_i$ with estimates of σ^2 for the indicated Mexican state, m_i and method.

State	m_i	$\widehat{\theta}_i$	LML	LRML	YML	YRML	MLEM	RMLEM
Aguascalientes (Ags)	11	939.35	918.37 ± 21.27	921.38 ± 18.11	915.63 ± 24.81	918.88 ± 21.21	883.70 ± 164.85	921.38 ± 18.11
Baja California (BC)	5	1169.42	1126.60 ± 36.11	1132.35 ± 31.93	1121.75 ± 39.89	1127.78 ± 35.52	1061.41 ± 327.17	1132.35 ± 31.93
Baja California Sur (BCS)	5	1356.74	995.54 ± 182.87	1035.31 ± 166.05	963.68 ± 200.21	1003.73 ± 183.40	912.80 ± 420.93	1035.31 ± 166.05
Campeche (Camp)	11	460.49	479.69 ± 18.28	477.29 ± 15.77	481.72 ± 20.75	479.23 ± 17.97	504.41 ± 148.49	477.29 ± 15.77
Chiapas (Chis)	119	325.82	340.69 ± 15.38	338.74 ± 13.19	342.42 ± 17.79	340.35 ± 15.30	357.94 ± 103.04	338.74 ± 13.19
Chihuahua (Chih)	67	962.03	871.35 ± 94.37	879.37 ± 86.44	864.92 ± 101.19	872.73 ± 93.51	844.74 ± 166.78	879.37 ± 86.44
Coahuila (Coah)	38	879.43	856.88 ± 41.54	860.06 ± 35.87	853.96 ± 47.65	857.35 ± 41.47	834.94 ± 133.54	860.06 ± 35.87
Colima (Col)	10	803.55	789.86 ± 20.16	791.98 ± 17.31	787.95 ± 23.02	790.23 ± 19.89	752.55 ± 189.81	791.98 ± 17.31
Distrito Federal (DF)	7	1526.98	409.99 ± 335.33	459.56 ± 342.75	374.67 ± 332.40	418.91 ± 339.88	355.75 ± 364.64	459.56 ± 342.75
Durango (Dgo)	39	754.39	788.84 ± 47.27	784.37 ± 42.01	792.78 ± 52.38	788.10 ± 46.84	812.29 ± 139.89	784.37 ± 42.01
Guanajuato (Gto)	46	597.01	593.20 ± 14.86	593.98 ± 12.64	592.41 ± 17.36	593.32 ± 14.77	576.37 ± 92.82	593.98 ± 12.64
Guerrero (Gro)	76	605.11	550.19 ± 60.47	557.22 ± 53.54	544.14 ± 66.68	551.49 ± 59.49	495.19 ± 302.22	557.22 ± 53.54
Hidalgo (Hgo)	85	641.63	630.10 ± 43.78	631.72 ± 38.25	628.68 ± 48.94	630.41 ± 43.06	611.97 ± 127.91	631.72 ± 38.25
Jalisco (Jal)	124	808.46	657.37 ± 112.62	664.11 ± 110.76	652.36 ± 114.25	658.54 ± 112.67	639.18 ± 131.28	664.11 ± 110.76
México (Méx)	123	1036.92	622.15 ± 167.89	628.90 ± 167.92	617.30 ± 168.26	623.35 ± 168.21	597.08 ± 180.24	628.90 ± 167.92
Michoacán (Mich)	113	577.81	581.91 ± 17.52	581.69 ± 15.00	582.02 ± 20.23	581.85 ± 17.33	577.63 ± 55.60	581.69 ± 15.00
Morelos (Mor)	33	926.05	787.61 ± 107.11	804.31 ± 96.10	773.48 ± 118.51	790.61 ± 106.99	719.25 ± 310.25	804.31 ± 96.10
Nayarit (Nay)	20	677.62	668.68 ± 34.96	670.25 ± 30.90	667.21 ± 38.93	668.92 ± 34.66	656.15 ± 78.34	670.25 ± 30.90
Nuevo León (NL)	50	1232.20	859.37 ± 132.73	883.75 ± 131.00	841.01 ± 135.47	863.66 ± 134.46	814.11 ± 203.67	883.75 ± 131.00
Oaxaca (Oax)	578	412.72	576.55 ± 100.21	559.66 ± 92.17	590.51 ± 109.23	573.70 ± 101.17	618.84 ± 199.55	559.66 ± 92.17
Puebla (Pue)	223	423.22	462.18 ± 25.78	457.26 ± 22.49	466.51 ± 29.59	461.34 ± 25.91	491.48 ± 139.54	457.26 ± 22.49
Querétaro (Qro)	18	641.84	641.50 ± 18.55	641.64 ± 16.00	641.37 ± 21.14	641.53 ± 18.32	634.05 ± 73.27	641.64 ± 16.00
Quintana Roo (QR)	8	632.76	589.84 ± 73.18	596.10 ± 64.57	584.20 ± 81.90	590.85 ± 72.60	550.42 ± 205.72	596.10 ± 64.57
San Luis Potosí (SLP)	58	504.47	536.93 ± 24.67	532.63 ± 21.44	540.76 ± 28.47	536.20 ± 24.78	567.95 ± 146.40	532.63 ± 21.44
Sinaloa (Sin)	18	927.84	826.22 ± 66.88	838.54 ± 59.69	815.67 ± 74.98	828.40 ± 67.41	765.16 ± 278.01	838.54 ± 59.69
Sonora (Son)	72	989.80	954.36 ± 51.44	958.16 ± 45.92	951.17 ± 56.45	955.01 ± 50.76	937.69 ± 123.42	958.16 ± 45.92
Tabasco (Tab)	17	541.45	554.30 ± 10.05	552.53 ± 8.67	555.89 ± 11.56	553.98 ± 10.00	586.06 ± 161.98	552.53 ± 8.67
Tamaulipas (Tamps)	41	793.39	798.74 ± 33.53	798.14 ± 29.30	799.22 ± 37.98	798.61 ± 33.23	800.82 ± 84.05	798.14 ± 29.30
Tlaxcala (Tlax)	51	777.61	742.60 ± 66.61	747.53 ± 58.27	738.16 ± 75.27	743.37 ± 66.16	715.98 ± 157.75	747.53 ± 58.27
Veracruz (Ver)	216	515.10	515.72 ± 59.07	516.32 ± 52.16	515.10 ± 65.23	515.83 ± 58.07	501.88 ± 126.33	516.32 ± 52.16
Yucatán (Yuc)	100	344.22	369.75 ± 23.10	366.72 ± 20.06	372.30 ± 25.99	369.16 ± 22.72	392.23 ± 136.02	366.72 ± 20.06
Zacatecas (Zac)	57	842.43	836.09 ± 15.82	836.93 ± 13.70	835.33 ± 18.11	836.23 ± 15.62	826.72 ± 73.82	836.93 ± 13.70

Following to Figueiredo and Gomes [14] (2004), we have generated 10000 bootstrap samples of size 32 (the same as the number of states) to calculate the SEs. With these SEs, we build the corresponding bootstrap confidence intervals ($\text{BCI}_{95\%}$) for each EBLUP, using a confidence level of 95%. For the EBLUP, these intervals are obtained as

$$\text{BCI}_{95\%}(\widehat{\theta}_i^{\text{EBLUP}}) = \left[\overline{\widehat{\theta}}_i^{\text{EBLUP}(B)} \pm z_{1-\alpha/2}\text{SD}\left(\widehat{\theta}_i^{\text{EBLUP}(B)}\right)\right], \tag{5.1}$$

where $\overline{\widehat{\theta}}_i^{\text{EBLUP}(B)}$ and $\text{SD}(\widehat{\theta}_i^{\text{EBLUP}(B)})$ are the mean and standard deviation bootstrap, respectively, and $z_{1-\alpha/2}$ is the $(1-\alpha/2) \times 100$-th percentile of the standard normal distribution. The results for BCIs based on (5.1) with the Mexican data are resented in Table 9. Figure 1 shows the estimated AMPCFERH, where the states are colored according to a classification into quartiles for the values of the estimates by the RMLEM approach. Note that, in general, the states with smaller AMPCFERH are mainly concentrated in the southwest part of the country, except for Quintana Roo, while the states with larger AMPCFERH are located at the north part. A measure of uncertainty of the EBLUP of $\widehat{\theta}_i$ is given in Table 10.

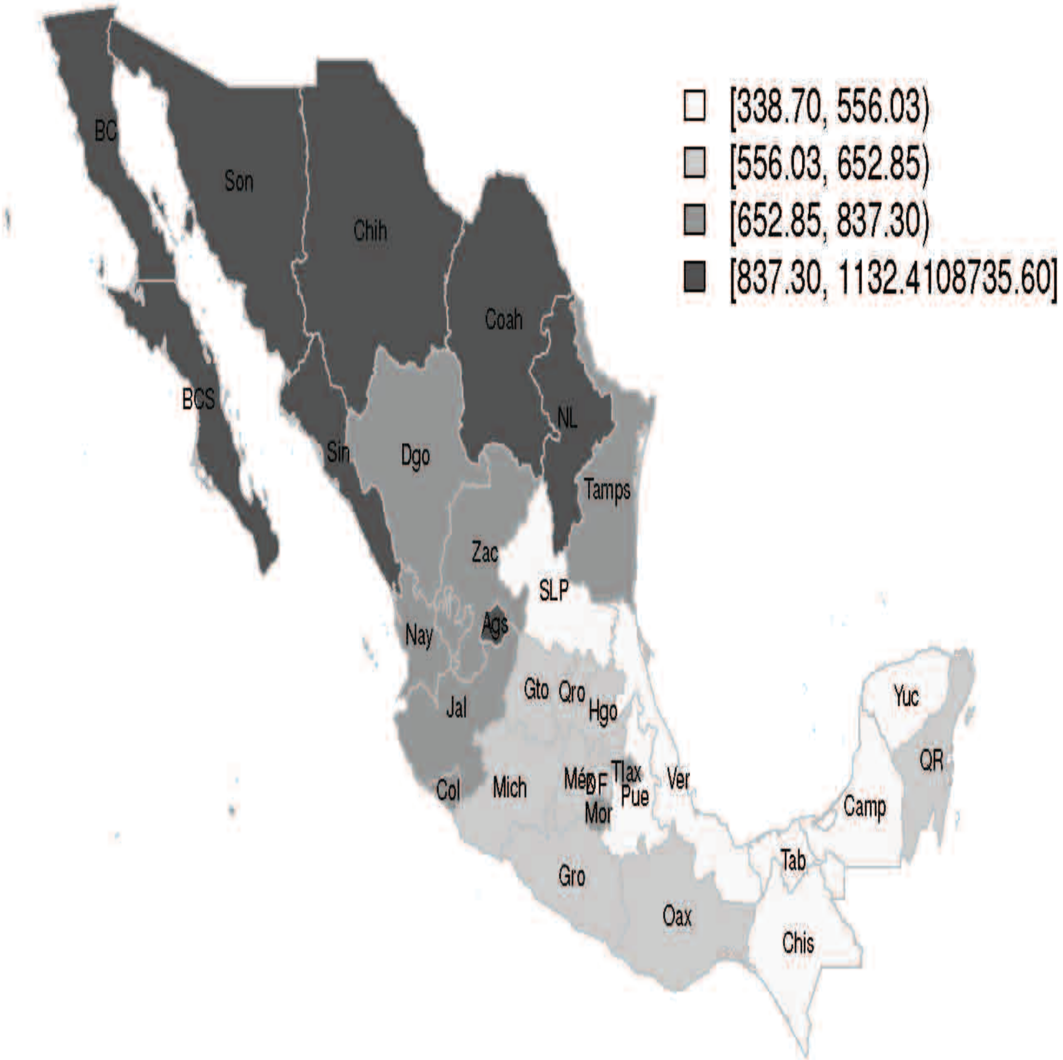

Figure 1: Estimated average monthly per capita food expenditure for rural household in 2000 with the RMLEM approach for Mexican data.

We compute the MSPE and its corresponding $\text{BCI}_{95\%}$, similarly as for the EBLUP given in (5.1), using (2.18) for each of the estimation methods. We highlight two aspects: (i) the values obtained by the MLEM and RMLEM approaches are similar to the corresponding values of the YML and RYML methods, and moreover, its variability presents the same results; and (ii) note that the LML and LRML approaches show lower MSPE indicators, but, as mentioned, these methods can provide a negative value for the MSPE, making their results underestimated and unreliable. As in Molina *et al.* [28] (2014), we calculate the coefficients of variation (CVs), and its correspondig $\text{BCI}_{95\%}$ in terms of the MSPE estimates as $\text{CV}[\widehat{\theta}_i^{\text{EBLUP}}] = ((\text{MSPE}(\widehat{\theta}_i^{\text{EBLUP}}))^{1/2}/\widehat{\theta}_i^{\text{EBLUP}}) \times 100$ for each bootstrap sample, in order to analyze the gain in efficiency of the $\widehat{\theta}_i^{\text{EBLUP}}$ in comparison with direct estimates. Table 11 displays these CVs, from which the two following aspects can be mentioned: (i) there is a clear overall gain in precision when the EBLUP of $\widehat{\theta}_i$ is obtained with the YML/YRML/MLEM/RMLEM methods, if σ^2 is estimated, since in almost all cases the CVs are less than the CVs of the direct estimator; and (ii) in general, this gain in precision has a greater effect in the RMLEM method, and moreover, the variability obtained by the bootstrapp resampling shows that it also has less variability in comparison to the rest of the methods.

Table 10: MSPE of EBLUP of $\widehat{\theta}_i$ with estimates of σ^2 for the indicated Mexican state, m_i and method.

State	m_i	$\widehat{\theta}_i$	LML	LRML	YML	YRML	MLEM	RMLEM
Aguascalientes	11	–	7062.57 ± 88.31	7072.20 ± 84.91	7065.37 ± 87.95	7063.48 ± 87.76	7065.36 ± 87.95	7063.47 ± 87.76
Baja California	5	–	8533.92 ± 113.74	8537.96 ± 110.40	8545.70 ± 112.33	8533.08 ± 113.05	8545.69 ± 112.33	8533.07 ± 113.05
Baja California Sur	5	–	38291.19 ± 3898.23	38779.60 ± 3658.90	38101.79 ± 4049.52	38373.13 ± 3851.71	38101.59 ± 4049.52	38372.91 ± 3851.71
Campeche	11	–	6002.27 ± 58.62	6005.50 ± 56.77	6007.47 ± 57.86	6002.09 ± 58.32	6007.46 ± 57.86	6002.09 ± 58.32
Chiapas	119	–	4727.16 ± 36.06	4728.90 ± 34.86	4730.72 ± 35.61	4726.98 ± 35.83	4730.72 ± 35.61	4726.98 ± 35.83
Chihuahua	67	–	43346.23 ± 3745.72	43585.34 ± 3567.53	43361.26 ± 3843.81	43372.80 ± 3706.31	43361.08 ± 3843.81	43372.60 ± 3706.31
Coahuila	38	–	13307.16 ± 322.52	13340.25 ± 309.35	13314.31 ± 323.26	13310.50 ± 320.07	13314.29 ± 323.26	13310.48 ± 320.07
Colima	10	–	4821.23 ± 42.43	4824.96 ± 40.71	4823.34 ± 42.17	4821.47 ± 42.13	4823.34 ± 42.17	4821.47 ± 42.13
Distrito Federal	7	–	111400.65 ± 32054.55	113227.77 ± 30823.28	110394.68 ± 33003.71	111738.73 ± 31720.72	110393.75 ± 33003.71	111737.70 ± 31720.72
Durango	39	–	17327.26 ± 542.56	17353.87 ± 522.28	17358.39 ± 543.24	17326.94 ± 538.59	17358.36 ± 543.24	17326.91 ± 538.59
Guanajuato	46	–	6224.48 ± 74.76	6232.15 ± 71.48	6226.63 ± 74.66	6225.22 ± 74.12	6226.63 ± 74.66	6225.22 ± 74.12
Guerrero	76	–	18966.35 ± 589.66	18872.63 ± 556.05	19098.20 ± 606.77	18939.14 ± 579.86	19098.18 ± 606.77	18939.12 ± 579.86
Hidalgo	85	–	15972.62 ± 491.00	16015.04 ± 469.76	15984.87 ± 494.07	15976.51 ± 486.51	15984.84 ± 494.07	15976.48 ± 486.51
Jalisco	124	–	95196.34 ± 29228.59	97750.38 ± 28174.99	93689.46 ± 30058.90	95683.09 ± 28945.16	93688.45 ± 30058.90	95681.96 ± 28945.16
México	123	–	171437.83 ± 54229.90	174617.81 ± 52689.00	169339.63 ± 55318.79	172075.09 ± 53758.50	169338.20 ± 55318.79	172073.49 ± 53758.50
Michoacán	113	–	9280.13 ± 175.17	9298.77 ± 167.71	9282.71 ± 175.54	9282.20 ± 173.85	9282.70 ± 175.54	9282.19 ± 173.85
Morelos	33	–	26572.95 ± 1626.86	26759.52 ± 1546.85	26534.16 ± 1661.29	26600.49 ± 1611.46	26534.08 ± 1661.29	26600.39 ± 1611.46
Nayarit	20	–	22314.46 ± 1123.78	22447.65 ± 1070.15	22291.05 ± 1142.82	22333.72 ± 1113.91	22290.98 ± 1142.82	22333.66 ± 1113.91
Nuevo León	50	–	68062.42 ± 12681.30	69275.95 ± 12132.42	67486.95 ± 13078.86	68276.16 ± 12554.36	67486.42 ± 13078.86	68275.58 ± 12554.36
Oaxaca	578	–	35112.56 ± 2740.44	35374.18 ± 2611.48	35068.08 ± 2802.07	35149.58 ± 2712.87	35067.93 ± 2802.07	35149.43 ± 2712.87
Puebla	223	–	10943.82 ± 240.24	10967.14 ± 230.28	10948.83 ± 240.76	10946.20 ± 238.56	10948.81 ± 240.76	10946.18 ± 238.56
Querétaro	18	–	8204.86 ± 134.72	8217.40 ± 129.18	8208.76 ± 134.56	8206.02 ± 133.81	8208.75 ± 134.56	8206.01 ± 133.81
Quintana Roo	8	–	19958.04 ± 797.45	20018.77 ± 765.16	19976.57 ± 803.66	19963.41 ± 791.20	19976.53 ± 803.66	19963.36 ± 791.20
San Luis Potosí	58	–	9360.25 ± 174.23	9377.92 ± 166.87	9363.88 ± 174.46	9362.08 ± 173.01	9363.87 ± 174.46	9362.07 ± 173.01
Sinaloa	18	–	21443.10 ± 972.04	21545.23 ± 928.59	21438.25 ± 984.28	21456.33 ± 963.25	21438.19 ± 984.28	21456.27 ± 963.25
Sonora	72	–	23983.14 ± 1109.54	24042.02 ± 1065.55	24024.91 ± 1119.54	23985.42 ± 1100.99	24024.85 ± 1119.54	23985.35 ± 1100.99
Tabasco	17	–	3501.87 ± 22.68	3503.64 ± 21.93	3503.21 ± 22.36	3501.95 ± 22.57	3503.21 ± 22.36	3501.95 ± 22.57
Tamaulipas	41	–	16459.47 ± 547.96	16507.93 ± 525.90	16469.43 ± 550.71	16464.38 ± 544.25	16469.40 ± 550.71	16464.34 ± 544.25
Tlaxcala	51	–	20426.38 ± 867.03	20518.08 ± 828.69	20423.87 ± 876.94	20438.07 ± 859.53	20423.82 ± 876.94	20438.02 ± 859.53
Veracruz	216	–	22062.93 ± 993.54	22146.37 ± 950.78	22075.62 ± 1005.02	22071.75 ± 984.50	22075.56 ± 1005.02	22071.69 ± 984.50
Yucatán	100	–	7949.13 ± 112.31	7949.75 ± 109.16	7961.83 ± 110.59	7947.76 ± 111.89	7961.82 ± 110.59	7947.75 ± 111.89
Zacatecas	57	–	8189.27 ± 131.47	8198.87 ± 126.70	8195.48 ± 130.75	8189.79 ± 130.78	8195.47 ± 130.75	8189.78 ± 130.78

Table 11: CVs of direct estimator and EBLUP of $\widehat{\theta}_i$ with estimates of σ^2 for the indicated Mexican state, m_i and method.

State	m_i	$\widehat{\theta}_i$	LML	LRML	YML	YRML	MLEM	RMLEM
Aguascalientes	11	9.21	9.15 ± 0.21	9.13 ± 0.18	9.18 ± 0.25	9.15 ± 0.21	9.51 ± 0.24	9.12 ± 0.18
Baja California	5	8.14	8.20 ± 0.27	8.16 ± 0.24	8.24 ± 0.30	8.19 ± 0.27	8.71 ± 0.29	8.16 ± 0.24
Baja California Sur	5	17.58	19.66 ± 4.56	19.02 ± 3.76	20.26 ± 5.56	19.52 ± 4.52	21.38 ± 4.45	18.92 ± 3.77
Campeche	11	17.20	16.15 ± 0.60	16.24 ± 0.53	16.09 ± 0.68	16.17 ± 0.59	15.37 ± 0.62	16.23 ± 0.53
Chiapas	119	21.47	20.18 ± 0.24	20.30 ± 0.78	20.09 ± 1.01	20.20 ± 0.88	19.22 ± 0.90	20.30 ± 0.78
Chihuahua	67	25.74	23.89 ± 3.06	23.74 ± 2.74	24.08 ± 3.34	23.86 ± 3.02	24.65 ± 2.77	23.68 ± 2.76
Coahuila	38	13.85	13.46 ± 0.68	13.43 ± 0.58	13.51 ± 0.80	13.46 ± 0.68	13.82 ± 0.72	13.41 ± 0.59
Colima	10	8.81	8.79 ± 0.23	8.77 ± 0.20	8.81 ± 0.26	8.79 ± 0.23	9.23 ± 0.25	8.77 ± 0.20
Distrito Federal	7	39.79	81.41 ± 80.28	73.22 ± 72.47	88.68 ± 83.59	79.80 ± 77.55	93.40 ± 79.40	72.74 ± 70.28
Durango	39	18.60	16.69 ± 1.01	16.79 ± 0.91	16.62 ± 1.11	16.70 ± 1.00	16.22 ± 0.98	16.78 ± 0.91
Guanajuato	46	13.55	13.30 ± 0.35	13.29 ± 0.30	13.32 ± 0.41	13.30 ± 0.35	13.69 ± 0.39	13.28 ± 0.30
Guerrero	76	23.69	25.03 ± 3.17	24.65 ± 2.71	25.40 ± 3.60	24.95 ± 3.09	27.91 ± 2.97	24.70 ± 2.72
Hidalgo	85	21.00	20.06 ± 1.55	20.03 ± 1.35	20.11 ± 1.75	20.05 ± 1.53	20.66 ± 1.51	20.01 ± 1.35
Jalisco	124	77.60	46.94 ± 11.61	47.08 ± 11.20	46.92 ± 11.97	46.97 ± 11.56	47.89 ± 9.14	46.58 ± 11.29
México	123	130.27	66.55 ± 24.00	66.44 ± 23.43	66.66 ± 24.51	66.55 ± 23.95	68.92 ± 17.91	65.96 ± 23.48
Michoacán	113	17.33	16.55 ± 0.53	16.58 ± 0.46	16.55 ± 0.60	16.56 ± 0.52	16.68 ± 0.56	16.56 ± 0.46
Morelos	33	19.88	20.70 ± 3.34	20.34 ± 2.84	21.06 ± 3.94	20.63 ± 3.33	22.65 ± 3.23	20.28 ± 2.85
Nayarit	20	24.39	22.34 ± 1.35	22.35 ± 1.20	22.38 ± 1.49	22.34 ± 1.34	22.75 ± 1.30	22.30 ± 1.21
Nuevo León	50	30.82	30.36 ± 5.98	29.78 ± 5.61	30.89 ± 6.36	30.25 ± 5.99	31.91 ± 5.06	29.57 ± 5.65
Oaxaca	578	53.03	32.50 ± 5.71	33.61 ± 5.56	31.71 ± 5.90	32.68 ± 5.77	30.26 ± 4.71	33.50 ± 5.57
Puebla	223	25.85	22.63 ± 1.28	22.90 ± 1.14	22.43 ± 1.42	22.68 ± 1.28	21.29 ± 1.24	22.88 ± 1.14
Querétaro	18	14.59	14.12 ± 0.43	14.13 ± 0.37	14.13 ± 0.48	14.12 ± 0.42	14.29 ± 0.45	14.12 ± 0.37
Quintana Roo	8	24.17	23.95 ± 3.47	23.74 ± 2.95	24.19 ± 4.10	23.91 ± 3.46	25.68 ± 3.35	23.70 ± 2.96
San Luis Potosí	58	19.93	18.02 ± 0.81	18.18 ± 0.71	17.89 ± 0.92	18.05 ± 0.81	17.04 ± 0.82	18.17 ± 0.72
Sinaloa	18	17.31	17.72 ± 1.66	17.50 ± 1.44	17.95 ± 1.91	17.68 ± 1.65	19.14 ± 1.63	17.47 ± 1.45
Sonora	72	17.13	16.23 ± 0.96	16.18 ± 0.86	16.30 ± 1.06	16.22 ± 0.95	16.53 ± 0.94	16.16 ± 0.87
Tabasco	17	11.08	10.68 ± 0.19	10.71 ± 0.17	10.65 ± 0.22	10.68 ± 0.19	10.10 ± 0.21	10.71 ± 0.17
Tamaulipas	41	17.29	16.06 ± 0.70	16.10 ± 0.62	16.06 ± 0.78	16.07 ± 0.69	16.03 ± 0.70	16.08 ± 0.62
Tlaxcala	51	20.06	19.25 ± 1.96	19.16 ± 1.68	19.36 ± 2.27	19.23 ± 1.95	19.96 ± 1.93	19.12 ± 1.69
Veracruz	216	31.55	28.80 ± 3.66	28.82 ± 3.21	28.84 ± 4.08	28.80 ± 3.59	29.60 ± 3.33	28.77 ± 3.22
Yucatán	100	26.60	24.11 ± 1.46	24.31 ± 1.29	23.97 ± 1.62	24.15 ± 1.44	22.75 ± 1.40	24.31 ± 1.29
Zacatecas	57	11.09	10.82 ± 0.22	10.82 ± 0.19	10.84 ± 0.25	10.82 ± 0.22	10.95 ± 0.24	10.81 ± 0.20

6. CONCLUSIONS AND FUTURE RESEARCH

One of the advantages of using a methodology based on small area estimation is that through auxiliary data we can improve direct estimates of a parameter of interest in small areas. Standard methods of variance component estimation used in the FH model for small areas produce a negative or zero estimate for these variances, with severe implications. In such a context, we proposed alternative approaches to those available in the literature, based on the EM algorithm, for estimating the variance of the random effects in the FH model, when estimating small area means. We showed through a simulation study that the EM algorithm is a good alternative to compute the ML estimate of the variance components, ensuring its strictly positive value. We compared the performance of our approaches with two recently proposed methods by means of statistical indicators. In general, the MLEM and RMLEM approaches performed well and similarly to the YML and YRML methods proposed by Yoshimori and Lahiri [42] (2014), but better than the LML and LRML methods proposed by Li and Lahiri [23] (2010). The proposed approaches have the advantage of working directly with the likelihood function without having to adjust it. A shortcoming of the LML and LRML methods in comparison to the approaches proposed here is that they can yield a negative value for the MSPE. Also, although the results of the MSE in the estimation of the variance component are similar under the YML and MLEM methods, note that the estimation with the EM algorithm is slightly more accurate in terms of SEs than the estimation with the YML and YRML methods, such as occurs when comparing the YRML and RMLEM methods. In an application from the real-world, we confirmed that small area estimation through the FH model helped to improve the direct estimates of the average monthly per capita food expenditure for Mexican rural households in 2000 according to three auxiliary variables. A possible future study can be conducted to compare the YML and YRML methods to their analogous based on the EM algorithm.

ACKNOWLEDGMENTS

The authors thank the editors and reviewers for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This research was supported partially by fellowship “Becas-Conicyt” (J.L. Ávila-Valdez) and grant “Fondecyt 1160868” (V. Leiva), both from the Chilean Council for Scientific and Technology Research.

REFERENCES

[1] Battese, G.E.; Harter, R.M. and Fuller, W.A. (1988). An error components model for prediction of county crop area using survey and satellite data, *Journal of the American Statistical Association*, **83**, 28–36.

[2] [Casas-Cordero et al., 2016]casasep:16 Casas-Cordero, C.; Encina, J. and Lahiri, P. (2016). *Poverty mapping for the Chilean comunas*. In “Analysis of Poverty Data by Small Area Estimation” (M. Pratesi, Ed.), volume 20, pages 379–404, Wiley, Chichester, UK.

[3] Chen, S. and Lahiri, P. (2008). On mean squared prediction error estimation in small area estimation problems, *Communications in Statistics: Theory and Methods*, **37**, 1792–1798.

[4] CIESIN (2005). *Poverty mapping project: poverty and food security case studies*, Technical Report http://dx.doi.org/10.7927/H4FF3Q9B (accessed at 26-Nov-2015), Center for International Earth Science Information Network (CIESIN), Columbia University and Centro Internacional de Tropical Agriculture, New York, US.

[5] Coelho, P.S. and Casimiro, F. (2008). Post Enumeration Survey of the 2001 Portuguese population and housing censuses, *REVSTAT – Statistical Journal*, **6**, 231–252.

[6] Coelho, P.S. and Pereira, L.N. (2011). A spatial unit level model for small area estimation, *REVSTAT – Statistical Journal*, **9**, 155–180.

[7] Datta, G.S. (2009). *Model-based approach of small area estimation.* In "Handbook of Statistics. Sample Surveys: Inference and Analysis" (D. Pfeffermann and C.R. Rao, Eds.), volume 29B, pages 251–288, Elsevier, Oxford, UK.

[8] Datta, G.S. and Ghosh, M. (2012). Small area shrinkage estimation, *Statistical Science*, **27**, 95–114.

[9] Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems, *Statistica Sinica*, **10**, 613–627.

[10] Datta, G.S.; Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183–196.

[11] Demidenko, E. (2013). *Mixed Models: Theory and Applications with R*, Wiley, New Jersey, US.

[12] El-Leithy, H.A.; Wahed, Z.A.A. and Abdallah, M.S. (2016). On non-negative estimation of variance components in mixed linear models, *Journal of Advanced Research*, **7**, 59–68.

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

[14] Figueiredo, F. and Gomes, M.I. (2004). The total median in statistical quality control, *Applied Stochastic Model in Business and Industry*, **20**, 339–353.

[15] Fishman, G.S. (1973). *Concepts and Methods in Discrete Event Digital Simulation*, Wiley, New York, US.

[16] Gomes, M.I.; Henriques-Rodrigues, L. and Miranda, M. (2011). Reduced-bias location-invariant extreme value index estimation: a simulation study, *Communications in Statistics: Simulation and Computation*, **40**, 424–447.

[17] Gomes, M.I.; Henriques-Rodrigues, L. and Miranda, M. (2016). Mean-of-order-p location-invariant extreme value index estimation, *REVSTAT – Statistical Journal*, **14**, 273–296.

[18] Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association*, **72**, 320–340.

[19] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, California, US.

[20] Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation, *TEST*, **15**, 1–96.

[21] Laird, N.M. and Were, J.H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963–974.

[22] Lehtonen, R. and Veijanen, A. (2009). *Design-based methods of estimation for domains and small areas*, In "Handbook of Statistics. Sample Surveys: Inference and Analysis" (D. Pefeffermann and C.R. Rao, Eds.), volume 29B, pages 219–249, Elsevier, Oxford, UK.

[23] Li, H. and Lahiri, P. (2010). An adjusted maximum likelihood method for solving small area estimation problems, *Journal of Multivariate Analysis*, **101**, 882–892.

[24] Lohr, S.L. (1999). *Sampling: Design and Analysis*, Duxbury Press, Pacific Grove, CA, US.

[25] Marchetti, S.; Giusti, C.; Pratesi, M.; Salvati, N.; Giannotti, F.; Pedreschi, D.; Rinzivillo, R.; Pappalardo, L. and Gabrielli, L. (2015). Small area model-based estimators using big data sources, *Journal of Official Statistics*, **31**, 263–281.

[26] McLachlan, G.J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, Wiley, New Jersey, US.

[27] Mert, Y. (2015). Generalized least squares and weighted least squares estimation methods for distributional parameters, *REVSTAT – Statistical Journal*, **13**, 263–282.

[28] 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**, 852–885.

[29] Pereira, L.N. and Coelho, P.S. (2012). Small area estimation using a spatio-temporal linear mixed model, *REVSTAT – Statistical Journal*, **10**, 285–308.

[30] Pfeffermann, D. (2013). New important developments in small area estimation, *Statistical Science*, **28**, 40–68.

[31] Pfeffermann, D. and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas, *Journal of the American Statistical Association*, **102**, 1427–1439.

[32] Pinheiro, J.C. and Bates, D.M. (2004). *Mixed-Effects Models in S and S-Plus*, Springer, New York, US.

[33] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators, *Journal of the American Statistical Association*, **85**, 163–171.

[34] Rao, J.N.K. (2003). *Small Area Estimation*, Wiley, New Yersey, US.

[35] Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*, Wiley, New Yersey, US.

[36] Rueda, M.M.; Arcos, A.; Molina, D. and Trujillo, M. (2018). Model-assisted and model-calibrated estimation for class frequencies with ordinal outcomes, *REVSTAT – Statistical Journal*, **16**, 323–348.

[37] Searle, S.R.; Casella, G. and McCulloch, C.E. (2006). *Variance Components*, Wiley, New York, US.

[38] Thompson, R. and Meyer, K. (1986). Estimation of variance components: what is missing in the EM algorithm, *Journal of Statistical Computation and Simulation*, **24**, 215–230.

[39] van Dyk, D.A. (2000). Fitting mixed-effects models using efficient EM-type algorithms, *Journal of Computational and Graphical Statistics*, **9**, 78–98.

[40] Verbyla, A.P. (1990). A conditional derivation of residual maximum likelihood, *Australian Journal of Statistics*, **32**, 227–230.

[41] Wolter, K.M. (2007). *Introduction to Variance Estimation*, Springer-Verlag, New York, US.

[42] Yoshimori, M. and Lahiri, P. (2014). A new adjusted maximum likelihood method for the Fay-Herriot small area model, *Journal of Multivariate Analysis*, **124**, 281–294.