Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Comparison of some positive variance estimators for the Fay-Herriot small area model

by Susana Rubin-Bleuer and Yong You

Release date: June 22, 2016

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
p preliminary
r revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
E use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2016
Vol. 42, No. 1, pp. 63-85
Statistics Canada, Catalogue No. 12-001-X

Comparison of some positive variance estimators for the Fay-Herriot small area model

Susana Rubin-Bleuer and Yong You[1]

Abstract

The restricted maximum likelihood (REML) method is generally used to estimate the variance of the random area effect under the Fay-Herriot model (Fay and Herriot 1979) to obtain the empirical best linear unbiased (EBLUP) estimator of a small area mean. When the REML estimate is zero, the weight of the direct sample estimator is zero and the EBLUP becomes a synthetic estimator. This is not often desirable. As a solution to this problem, Li and Lahiri (2011) and Yoshimori and Lahiri (2014) developed adjusted maximum likelihood (ADM) consistent variance estimators which always yield positive variance estimates. Some of the ADM estimators always yield positive estimates but they have a large bias and this affects the estimation of the mean squared error (MSE) of the EBLUP. We propose to use a MIX variance estimator, defined as a combination of the REML and ADM methods. We show that it is unbiased up to the second order and it always yields a positive variance estimate. Furthermore, we propose an MSE estimator under the MIX method and show via a model-based simulation that in many situations, it performs better than other 'Taylor linearization' MSE estimators proposed recently.

Key Words: Variance estimation; Adjusted maximum likelihood; REML; Order of bias; MSE estimation.

1 Introduction

The Fay-Herriot model (Fay and Herriot 1979) is a basic area level model used to estimate small area means, when available direct survey estimates are imprecise due to small sample sizes. In this model, the small area mean is represented by a non-random linear term in the covariates, plus a random area effect. The best linear unbiased prediction (BLUP) estimator of a small area mean, under the Fay-Herriot model, can be obtained by minimizing the mean squared error (MSE) among the class of linear unbiased estimators. The BLUP is a weighted average of the direct survey estimator and the regression-synthetic estimator, with weights depending on the variance of the random area effects, σ_v^2. Usually, this variance has to be estimated from the data under the Fay-Herriot model. The empirical best linear unbiased (EBLUP) estimator of the small area mean is obtained by replacing the variance in the formula of the BLUP with an estimate. There are many well-known methods of variance estimation used in this context but the variance estimator used most often is the restricted maximum likelihood (REML) estimator because it accounts for the loss of degrees of freedom due to estimating the regression coefficient. Furthermore, it is unbiased up to the second order, and it also converges faster in terms of the number of iterations. Despite these important characteristics, occasionally, and particularly when the number of areas, m, is small or moderate, the REML method yields a zero variance estimate. This implies zero weight to the direct survey estimator in the EBLUP formula and hence the EBLUP estimator becomes a regression-synthetic estimator. However, most practitioners are reluctant to use synthetic estimators for small area means, since these ignore the survey based information and are often quite biased. When dealing with real data sets, for which models are never

1. Susana Rubin-Bleuer and Yong You, International Cooperation and Corporate Statistical Methods Division, Statistics Canada. E-mail: susana.rubin-bleuer@canada.ca; yong.you@canada.ca.

perfect, a positive estimate for σ_v^2 reduces the bias of the EBLUP over the synthetic model. Certainly, a positive random effects variance estimate, results in a 'conservative' EBLUP estimator in the sense that it gives a positive weight to the direct survey estimator. Furthermore, it can be viewed as the sum of the regression estimator plus a non-zero term that accounts for part of the 'model bias'. This feature gives rise to a series of variance estimation methods that yield positive estimates.

In this article, we focus on the adjusted likelihood variance estimators developed by Lahiri and Li (2009) and we propose a MIX variance estimator. Our MIX variance estimator is the combination of a REML estimator and any of the adjusted likelihood methods. We also put forward an estimator of the MSE of the EBLUP under the MIX and investigate the theoretical and finite sample properties of both the MIX variance estimator and MSE estimator.

Morris (2006) and Lahiri and Li (2009) proposed adjusted likelihood variance estimators resulting from optimizing the profile and residual likelihood adjusted with a factor $h(\sigma_v^2), \sigma_v^2 > 0.$ Li and Lahiri (2011) proposed two methods of variance estimation (the AM.LL and AR.LL methods, associated with the profile and residual likelihoods respectively) that ensure positive estimates with adjustment factor $h_{\mathrm{LL}}(\sigma_v^2) = \sigma_v^2.$ Yoshimori and Lahiri (2014) proposed two other variance estimators (the AM.YL and AR.YL methods) derived from adjusting the the profile and residual likelihoods with factor

$$h_{\mathrm{YL}}(\sigma_v^2) = \left\{\arctan\left[\sum_{i=1}^{m} \sigma_v^2 \Big/ (\sigma_v^2 + \psi_i)\right]\right\}^{1/m}$$

where ψ_i is the sampling variance for the i^{th} area. It is well known that the LL estimators are biased, especially for small or moderate number of areas (see Lahiri and Pramanik 2011). The YL method that adjusts the profile likelihood also leads to a biased estimator of $\sigma_v^2.$ However the bias of the variance estimator does not affect the MSE of the EBLUP: the second order asymptotic approximation to the MSE shows that the MSE depends on the asymptotic variance and not on the bias of the variance estimator. However, the bias of the variance estimators affects, the Taylor linearization MSE estimators and it can lead to negatively biased MSE estimators. It is desirable then to investigate alternative positive variance estimators.

The method of combining the AM.LL and the REML variance estimators was first mentioned by Yuan (2009) for the Fay-Herriot model. However, Yuan (2009) did not study its properties, empirically or otherwise. Rubin-Bleuer, Yung and Landry (2010, 2011 and 2012) carried out empirical comparisons of a MIX variance estimator under a time series and cross-sectional area level model and Rubin-Bleuer and You (2012) studied the asymptotic and finite sample properties of the MIX variance estimator for the Fay-Herriot model.

Here we formalize the MIX method for the Fay-Herriot model and prove that the MIX variance estimator is unbiased up to the second order. Furthermore, we propose an MSE estimator of the Taylor linearization type. We also examine the empirical performance of the MIX for a small and moderate number of areas. With respect to MSE estimation, Rubin-Bleuer and You (2012) and Molina, Rao and Datta (2015) each proposed a different 'split' MSE estimator under MIX variance estimation. We show that both the Rubin-Bleuer and You (2012) and the Molina et al. (2015) MSE estimators are unbiased up to the second order.

These ‘split’ MSE estimators were assigned a rule for populations that yielded zero estimates under REML variance estimation, and another rule for populations that yielded positive estimates under REML variance estimation. Both papers mentioned above showed that for a small number of areas, these ‘split’ estimators behaved well empirically in terms of average relative bias. However this outcome could be misleading, since the MSE estimators are usually negatively biased for populations where the REML variance estimate is zero, and they are positively biased for populations with positive REML estimates: the bias cancels out on average. In view of this issue we propose another MSE estimator, and we compare it to other MSE estimators when conditioned to populations where the REML estimate is zero.

In Section 2, we introduce the Fay-Herriot model, the EBLUP estimator of the small area mean and a second order approximation of the MSE of the EBLUP under the model. In Section 3, we describe the REML estimator and the *.LL and *.YL variance estimators. In Section 4, we present a general MIX variance estimator and we prove that its bias is of the same order as the bias of the REML estimator. We propose an unbiased (up to the second order) estimator of the MSE under the MIX method. In Section 5, we conduct an empirical study to compare the different variance estimators. Note that we defined the MIX variance estimator as a combination between the REML and any of the adjusted likelihood variance estimators, but the MIX variance estimator we chose for this study is the combination of the REML estimator and the AM.LL variance estimator. We selected this combination because Li and Lahiri (2011) reported that the adjusted profile likelihood performed better than adjusted residual likelihood (AR.LL) and because the adjustment factor in the Yoshimori and Lahiri (2014) variance estimators is too close to zero (in log terms), to improve significantly on the REML method. Finally in Section 6, we present the simulation results, analysis and conclusion.

2 EBLUP and MSE of the EBLUP under the Fay-Herriot model

Let $y_i, i = 1, \ldots, m,$ be the direct survey estimators of the small area means $\theta_i, i = 1, \ldots, m.$ The Fay-Herriot model consists of the following sampling and linking models:

$$\textbf{Sampling model: } y_i = \theta_i + e_i, e_i | \theta_i \overset{\text{i.d.}}{\sim} (0, \psi_i), \quad i = 1, \ldots, m, \tag{2.1}$$

$$\textbf{Linking model: } \theta_i = \mathbf{z}_i' \boldsymbol{\beta} + v_i, v_i \overset{\text{i.i.d.}}{\sim} \left(0, \sigma_v^2\right), \sigma_v^2 > 0, \quad i = 1, \ldots, m, \tag{2.2}$$

where e_i are the sampling errors, independently distributed with mean zero and “known” sampling variances ψ_i, $\mathbf{z}_i \, (p \times 1)$ are known vectors of covariate values; $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed regression coefficients; and v_i are independent and identically distributed random effects with mean zero and model variance σ_v^2. Combining (2.1) and (2.2) we obtain:

$$y_i = \mathbf{z}_i' \boldsymbol{\beta} + v_i + e_i, \quad i = 1, \ldots, m, \tag{2.3}$$

with both model and sampling errors. The $y_i, i = 1, \ldots, m,$ can be viewed as outcomes in the combined design-model space (see Rubin-Bleuer and Schiopu-Kratina 2005).

Under model (2.3), the EBLUP of the small area mean θ_i is given by:

$$\hat{\theta}_i\left(\hat{\sigma}_v^2\right) = \mathbf{z}_i'\hat{\boldsymbol{\beta}}\left(\hat{\sigma}_v^2\right) + \hat{\gamma}_i\left[y_i - \mathbf{z}_i'\hat{\boldsymbol{\beta}}\left(\hat{\sigma}_v^2\right)\right] = \hat{\gamma}_i y_i + (1-\hat{\gamma}_i)\,\mathbf{z}_i'\hat{\boldsymbol{\beta}}\left(\hat{\sigma}_v^2\right), \quad i = 1,\ldots,m, \tag{2.4}$$

where $\hat{\sigma}_v^2$ is a consistent estimator of σ_v^2,

$$\hat{\gamma}_i = \hat{\sigma}_v^2 \Big/ \left(\hat{\sigma}_v^2 + \psi_i\right), \quad \text{and} \quad \hat{\boldsymbol{\beta}}\left(\hat{\sigma}_v^2\right) = \left[\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i' \Big/ \left(\hat{\sigma}_v^2 + \psi_i\right)\right]^{-1} \left[\sum_{i=1}^m \mathbf{z}_i y_i \Big/ \left(\hat{\sigma}_v^2 + \psi_i\right)\right]. \tag{2.5}$$

To calculate the Mean Squared Error (MSE) of the EBLUP, we set the following regularity conditions:

1) The $\boldsymbol{\psi}_i$ are bounded from above and away from zero,
2) The $\mathbf{z}_i, 1 \le i \le m$ are bounded, and
3) $\liminf \boldsymbol{\lambda}_{\min}\left(1/m \sum_i \mathbf{z}_i \cdot \mathbf{z}_i'\right) > 0$ where $\boldsymbol{\lambda}_{\min}(A)$ = minimum eigenvalue of matrix A.

Under normality of the sampling errors e_i associated with model (2.3) and the above regularity conditions, a second order approximation to the MSE is given by:

$$\operatorname{MSE}\left\{\hat{\theta}_i\left(\hat{\sigma}_v^2\right)\right\} = g_{1i}\left(\sigma_v^2\right) + g_{2i}\left(\sigma_v^2\right) + g_{3i}\left(\sigma_v^2\right) + o\left(\frac{1}{m}\right), \tag{2.6}$$

with $g_{1i}\left(\sigma_v^2\right) = \gamma_i\psi_i$, $g_{2i}\left(\sigma_v^2\right) = (1-\gamma_i)^2\,\mathbf{z}_i'\left[\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i' \big/ \left(\sigma_v^2 + \psi_i\right)\right]^{-1}\mathbf{z}_i$ and

$$g_{3i}\left(\sigma_v^2\right) = (\psi_i)^2\,\bar{V}\left(\hat{\sigma}_v^2\right) \Big/ \left(\sigma_v^2 + \psi_i\right)^3, \tag{2.7}$$

where $\bar{V}\left(\hat{\sigma}_v^2\right)$ is the asymptotic variance of $\hat{\sigma}_v^2$ (Das, Jiang and Rao 2004).

3 Review of REML and adjusted maximum likelihood methods

3.1 REML method

We consider the combined Fay-Herriot model (2.3) with $\sigma_v^2 > 0$. The REML variance estimator of σ_v^2 is obtained by maximizing the residual likelihood function with respect to σ_v^2 :

$$L_{\text{REML}}\left(\sigma_v^2\right) \propto \left|\left[\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i' \Big/ \left(\sigma_v^2 + \psi_i\right)\right]\right|^{-1/2} \prod_{i=1}^m \left(\sigma_v^2 + \psi_i\right)^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\}$$

where $\mathbf{y} = (y_1,\ldots,y_m)'$, $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{V}^{-1}$, $\mathbf{V} = \operatorname{Var}(\mathbf{y})$, and $\mathbf{Z} = (\mathbf{z}_1,\ldots,\mathbf{z}_m)'$. (Cressie 1992, Datta and Lahiri 2000 and Rao 2003, chapter 6). The REML variance estimator is given by:

$$\hat{\sigma}_{v\text{REML}}^2 = \max\left(\tilde{\sigma}_{v\text{REML}}^2, 0\right), \tag{3.1}$$

where $\tilde{\sigma}_{v\text{REML}}^2$ is the converging value of the REML algorithm. The asymptotic bias and variance of the REML estimator, up to the second order, are respectively given by:

$$\text{Bias}\left(\hat{\sigma}_{v\text{REML}}^2\right) = o\left(\frac{1}{m}\right) \text{ and } V\left(\hat{\sigma}_{v\text{REML}}^2\right) = \frac{2}{\text{tr}\left(\mathbf{V}^{-2}\right)} + o\left(\frac{1}{m}\right). \tag{3.2}$$

A second order unbiased estimator of the MSE of the EBLUP under REML variance estimation is given by (Datta and Lahiri 2000 and Chen and Lahiri 2008, 2011):

$$\text{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}_{v\text{REML}}^2\right)\right\} = \begin{cases} g_{1i}\left(\hat{\sigma}_{v\text{REML}}^2\right) + g_{2i}\left(\hat{\sigma}_{v\text{REML}}^2\right) + 2g_{3i}\left(\hat{\sigma}_{v\text{REML}}^2\right) & \text{if } \hat{\sigma}_{v\text{REML}}^2 > 0 \\ g_{2i}(0) & \text{if } \hat{\sigma}_{v\text{REML}}^2 = 0. \end{cases} \tag{3.3}$$

Remark 3.1. When $\hat{\sigma}_v^2 = 0,$ the EBLUP reduces to the synthetic estimator. However, note that when

$$\hat{\sigma}_v^2 = 0, g_{1i}\left(\hat{\sigma}_v^2\right) = 0, \quad g_{2i}\left(\hat{\sigma}_v^2\right) = \mathbf{z}_i'\left[\sum_{i=1}^m \mathbf{z}_i\mathbf{z}_i' \Big/ \psi_i\right]^{-1}\mathbf{z}_i,$$

and $g_{3i}\left(\hat{\sigma}_v^2\right) = \bar{V}\left(\hat{\sigma}_v^2\right)\big/\psi_i > 0$, i.e., $\text{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}_v^2\right)\right\}$ is not a continuous function of $\hat{\sigma}_v^2$. We will see in the empirical study that when conditioning on $\left\{\hat{\sigma}_v^2 = 0\right\}$, the MSE estimator in (3.3) has significant negative bias, unless the underlying signal to noise ratio $\sigma_v^2\big/\psi_i$ is negligible.

3.2 Adjusted maximum likelihood methods

The adjusted maximum likelihood variance estimators are derived from optimizing either the profile (AM) or the residual (AR) likelihood adjusted with the factor $h\left(\sigma_v^2\right)$. As noted in the introduction, the AM.LL and AR.LL estimators use the adjustment factor $h_{\text{LL}}\left(\sigma_v^2\right) = \sigma_v^2$, and the AM.YL and AR.YL estimators use the adjustment factor

$$h_{\text{YL}}\left(\sigma_v^2\right) = \left\{\text{arc}\tan\left[\sum_{i=1}^m \sigma_v^2 \Big/ \left(\sigma_v^2 + \psi_i\right)\right]\right\}^{1/m}.$$

We denote by $\hat{\sigma}_{v\text{AM.LL}}^2$ and $\hat{\sigma}_{v\text{AM.YL}}^2$ the variance estimators obtained by maximizing the adjusted profile likelihood functions, with respect to σ_v^2 :

$$L_{\text{AM.}*}\left(\sigma_v^2\right) \propto h\left(\sigma_v^2\right)\cdot\prod_{i=1}^m\left(\sigma_v^2 + \psi_i\right)^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\}, \tag{3.4}$$

where $h\left(\sigma_v^2\right) = h_{\text{LL}}\left(\sigma_v^2\right)$ and $h\left(\sigma_v^2\right) = h_{\text{YL}}\left(\sigma_v^2\right)$ for AM.LL and AM.YL respectively. The matrix $\mathbf{P}$ is as in (3.1). The bias of the AM estimators up to the second order (denoted by $\approx$) is:

$$B\left(\hat{\sigma}_{v\text{AM.LL}}^2\right) \approx \frac{\text{tr}\left\{\mathbf{P}-\mathbf{V}^{-1}\right\} + 2\big/\sigma_v^2}{\text{tr}\left(\mathbf{V}^{-2}\right)} = O\left(\frac{1}{m}\right) \quad \text{and} \quad B\left(\hat{\sigma}_{v\text{AM.YL}}^2\right) \approx \frac{\text{tr}\left\{\mathbf{P}-\mathbf{V}^{-1}\right\}}{\text{tr}\left(\mathbf{V}^{-2}\right)} = O\left(\frac{1}{m}\right), \tag{3.5}$$

(Li and Lahiri 2011 and Yoshimori and Lahiri 2014). The AR.LL and AR.YL variance estimators, denoted by $\hat{\sigma}^2_{v\text{AR.LL}}$ and $\hat{\sigma}^2_{v\text{AR.YL}}$, are obtained by maximizing the adjusted residual (AR) likelihood functions with respect to σ^2_v :

$$L_{\text{AR.}*}\left(\sigma_v^2\right) \propto h\left(\sigma_v^2\right) \cdot \left| \sum_{i=1}^{m} \mathbf{z}_i \mathbf{z}_i' \Big/ \left(\sigma_v^2 + \psi_i\right) \right|^{-1/2} \prod_{i=1}^{m} \left(\sigma_v^2 + \psi_i\right)^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}'\mathbf{P}\mathbf{y}\right\} \tag{3.6}$$

where $h\left(\sigma_v^2\right) = h_{\text{LL}}\left(\sigma_v^2\right)$ and $h\left(\sigma_v^2\right) = h_{\text{YL}}\left(\sigma_v^2\right)$ for AR.LL and AR.YL respectively and $\mathbf{P}$ is as in (3.1), The asymptotic bias of the AR estimators are given, respectively by:

$$B\left(\hat{\sigma}^2_{v\text{AR.LL}}\right) \approx \frac{2/\sigma_v^2}{\text{tr}\left(\mathbf{V}^{-2}\right)} = O\left(\frac{1}{m}\right) \text{ and } B\left(\hat{\sigma}^2_{v\text{AR.YL}}\right) = o\left(\frac{1}{m}\right). \tag{3.7}$$

Under the regularity conditions given in Section 2 and $\sigma_v^2 > 0,$ the two LL and the two YL variance estimators exist and are $\sqrt{m}$ – consistent (Li and Lahiri 2011 and Yoshimori and Lahiri 2014). Lahiri and co-authors proposed the following MSE estimators:

$$\text{mse}\left\{\hat{\theta}_i(\cdot)\right\} = g_{1i}(\cdot) + g_{2i}(\cdot) + 2g_{3i}(\cdot) - \psi_i^2 \cdot B(\cdot)\big/(\cdot + \psi_i)^2 \tag{3.8}$$

where the argument in $(\cdot)$ above is either $\hat{\sigma}^2_{v\text{AM.LL}}, \hat{\sigma}^2_{v\text{AR.LL}}$ or $\hat{\sigma}^2_{v\text{AM.YL}}$ under AM.LL, AR.LL and AM.YL variance estimation respectively, and under $\hat{\sigma}^2_{v\text{AR.YL}}$:

$$\text{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}^2_{v\text{AR.YL}}\right)\right\} = g_{1i}\left(\hat{\sigma}^2_{v\text{AR.YL}}\right) + g_{2i}\left(\hat{\sigma}^2_{v\text{AR.YL}}\right) + 2g_{3i}\left(\hat{\sigma}^2_{v\text{AR.YL}}\right). \tag{3.9}$$

Estimators (3.8) and (3.9) are unbiased up to the second order.

Remark 3.2. The sampling errors do not need to be normally distributed for the consistency and asymptotic normality of the LL and YL estimators (see, for example, Rubin-Bleuer et al. 2011).

3.3 Optimization algorithms

Given the data, the REML likelihood function may attain its maximum value at $\sigma_v^2 = 0,$ even when the true underlying value of $\boldsymbol{\sigma}_v^2$ is positive. On the other hand, the LL and YL likelihoods always attain their maximum value at $\sigma_v^2 > 0.$ Yet, the YL residual likelihood is very close to the REML likelihood. Empirical studies show that the scoring algorithm under AR.YL yields $\hat{\sigma}^2_{v\text{AR.YL}} = 0$ in almost as large a percentage as under REML for data sets following a Fay-Herriot model with a small but non-zero true underlying variance. This happens when the scoring algorithm misses the positive maximum value of the AR.YL likelihood and outputs a zero value (see Appendix B for details). To avoid this problem, we use a grid method for optimization (Estevao 2014). In our study, we set the upper boundary of the search interval as $1{,}000 \times \sigma_v^2,$ since we know σ_v^2 a priori. For applications with real data we suggest to obtain an initial estimate $\hat{\sigma}^2_{v\text{AM.LL}}$ by the method of scoring and set $1{,}000 \times \hat{\sigma}^2_{v\text{AM.LL}}$ as the upper boundary. Then keep increasing the boundary until the variance estimate lies within the search interval.

4 The MIX variance estimator

4.1 Variance estimation

The MIX variance estimator is a procedure that first calculates the REML variance estimate and only substitutes it by an adjusted likelihood variance estimate if the REML estimate is negative. The MIX variance estimator is always positive and it is unbiased up to a term of order $o(1/m)$. The MIX variance estimator of σ_v^2 is defined by:

$$\hat{\sigma}_{v\text{MIX}}^2 = \begin{cases} \hat{\sigma}_{v\text{REML}}^2 & \text{if } \hat{\sigma}_{v\text{REML}}^2 > 0 \\ \hat{\sigma}_{v\text{adj}}^2 & \text{if } \hat{\sigma}_{v\text{REML}}^2 = 0, \end{cases} \tag{4.1}$$

where $\hat{\sigma}_{v\text{adj}}^2$ is one of the adjusted likelihood estimators defined in Section 3.

Remark 4.1. The MIX variance estimator automatically carries some of the common properties shared by the REML and the adjusted likelihood variance estimator. For example, it is even and translation invariant. Thus, under normality of the sampling errors, the second order approximation (2.6) of the MSE of the EBLUP is also valid: Theorem 4.1 below shows that the MSE of the EBLUP under the MIX variance estimator inherits the same asymptotic properties as the MSE under the REML variance estimator.

Theorem 4.1. Under regularity conditions 1 through 3 given in Section 2, and the assumption that $\sigma_v^2 > 0$, the MSE of the EBLUP under the MIX variance estimator is equal to the MSE under the REML variance estimator up to the second order. The theorem follows from the fact that the asymptotic variance of $\hat{\sigma}_{v\text{MIX}}^2$ coincides with the asymptotic variance of $\hat{\sigma}_{v\text{REML}}^2$ (see Appendix A for details).

Theorem 4.2. Under the conditions of Theorem 4.1, $\text{Bias}\left(\hat{\sigma}_{v\text{MIX}}^2\right) = o(1/m)$. The proof is given in Appendix A.

4.2 MSE estimation

The fact that the MIX estimator, $\hat{\sigma}_{v\text{MIX}}^2$, is unbiased to the second order, is crucial to show that our proposed MSE estimator is also unbiased up to the second order.

Corollary 4.2. The MSE estimator of the EBLUP under $\hat{\sigma}_{v\text{MIX}}^2$ given by:

$$\text{mse}\left[\hat{\theta}_i\left(\hat{\sigma}_{v\text{MIX}}^2\right)\right] = g_{1i}\left(\hat{\sigma}_{v\text{MIX}}^2\right) + g_{2i}\left(\hat{\sigma}_{v\text{MIX}}^2\right) + 2g_{3i}\left(\hat{\sigma}_{v\text{MIX}}^2\right) \tag{4.2}$$

is second order unbiased. Once given that $\hat{\sigma}_{v\text{MIX}}^2$ is second order unbiased, the result follows along the lines of Datta and Lahiri (2000).

4.3 Alternative MSE estimators

In the following the MIX variance estimator is the combination of REML and AM.LL.

Rubin-Bleuer and You (2012) had suggested another MSE estimator, also unbiased up to the second order: a 'split' MSE estimator of the form:

$$\operatorname{mse}^*\left[\hat{\theta}_i\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)\right]=\begin{cases} g_{1i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)+g_{2i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)+2g_{3i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right) & \text{if } \hat{\sigma}_{v\mathrm{MIX}}^2=\hat{\sigma}_{v\mathrm{REML}}^2, \\ g_{1i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)+g_{2i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right) & \\ \quad +2g_{3i}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)-\left(1-\hat{\gamma}_{i\mathrm{MIX}}\right)^2\cdot \operatorname{Bias}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right) & \text{if } \hat{\sigma}_{v\mathrm{MIX}}^2=\hat{\sigma}_{v\mathrm{AM.LL}}^2. \end{cases} \tag{4.3}$$

Estimator mse^* has a lower average relative bias (ARB) than the MSE estimator given in (4.2). The lower ARB occurs because the MSE estimates overestimate when REML is positive and underestimate when REML is zero. The mse^* estimator is good on average, but for a particular data set the mse^* estimator might take on negative values.

Molina et al. (2015) proposed two different MSE estimators for the EBLUP under the MIX: with PT standing for their proposed preliminary test of hypothesis for zero variance these estimators are:

$$\operatorname{mse}_0\left\{\hat{\theta}_i\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)\right\}=\begin{cases} \operatorname{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}_{v\mathrm{REML}}^2\right)\right\} & \text{if } \hat{\sigma}_{v\mathrm{REML}}^2>0 \\ g_{2i}(0) & \text{if } \hat{\sigma}_{v\mathrm{REML}}^2=0 \end{cases} \tag{4.4}$$

and

$$\operatorname{mse}_{\mathrm{PT}}\left\{\hat{\theta}_i\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right)\right\}=\begin{cases} \operatorname{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}_{v\mathrm{REML}}^2\right)\right\} & \text{if } \hat{\sigma}_{v\mathrm{REML}}^2>0 \text{ and PT rejected} \\ g_{2i}(0) & \text{if } \hat{\sigma}_{v\mathrm{REML}}^2=0 \text{ or PT not rejected.} \end{cases} \tag{4.5}$$

The rationale for mse_0 and $\operatorname{mse}_{\mathrm{PT}}$ is based on the MSE of the BLUP with $\sigma_v^2=0$. Molina et al. (2015) showed in an empirical study that their proposed MSE estimators performed well on average when both σ_v^2 and the number of areas m were small.

Remark 4.2. mse_0 and $\operatorname{mse}_{\mathrm{PT}}$ are also unbiased up to the second order (see Appendix for a brief proof of this property). Our argument against $\operatorname{mse}\left\{\hat{\theta}_i\left(\hat{\sigma}_v^2\right)\right\}$ (in 3.3) is also valid against mse_0 and $\operatorname{mse}_{\mathrm{PT}}$: for a moderate number of areas, the % of populations with $\hat{\sigma}_{v\mathrm{REML}}^2=0$ may be significant even if σ_v^2/ψ_i is not negligible. In this case, the MSE of the EBLUP should account also for the variation due to variance estimation or risk underestimation.

5 Simulation set up and performance measures

5.1 Simulation set-up

We conducted a model-based Monte Carlo simulation, following Rubin-Bleuer and You (2012), to examine the finite sample performance of the various methods. 'Direct' estimates $\left(y_1,\ldots,y_m\right)$ with $m=15, m=45$ and $m=100,$ are generated from the Fay-Herriot model in (2.3) with $\boldsymbol{\beta}'=(5,4,3,2,1)$ and covariates $\mathbf{z}_i'=\left(1,z_{i2},\ldots,z_{ip}\right),$ generated once from normal distributions $z_{ik}\sim k+N(1,1),$ $k=2,\ldots,5,$ $i=1,\ldots,m,$ and held fixed over the repeated populations. The independent normal random area effects v_i are generated with variance $\sigma_v^2 = 1$. Independent sampling errors e_i, are generated with sampling variances $\psi_i \triangleq 50/n_i$, where n_i is the sample size for area $i, i = 1, \ldots, m$. There are five sampling variance groups determined by $n_i = 3, 5, 7, 10$ or 15, with signal to noise ratios $\sigma_v^2/\psi_i = 0.06$, 0.1, 0.14, 0.2 and 0.3, respectively. Thus when $m = 100$ there are 20 areas per signal to noise ratio. We first generated 50,000 sets of direct estimators for each case and computed the EBLUP and the true Monte Carlo MSE of the EBLUP using the REML, AM.LL, MIX, AM.YL and AR.YL variance estimators. We did not study AR.LL due to its poor performance reported by Li and Lahiri (2011). Next we generated 10,000 sets of direct estimators independently of the first 50,000. For each generated set, we computed the five variance estimators. For the MIX variance estimator we looked at three of the four linearization type MSE estimators discussed in Section 4. Since the linearization MSE estimators often do not estimate bias accurately, we also considered the parametric bootstrap MSE (PB MSE) estimator adjusted for bias using Pfeffermann and Glickman's (2004) method and the naïve PB MSE estimator with 500 repetitions each (see Appendix B for the construction of the bootstrap). The Monte Carlo performance measures are defined below.

1. The MSE of the EBLUP, $\overline{\text{MSE}}_\ell\left(\hat{\theta}_i\right)$, per sampling variance group:

$$\text{MSE}\left(\hat{\theta}_i\right) = \frac{1}{50{,}000}\sum_{r=1}^{50{,}000}\left(\hat{\theta}_i^{(r)} - \theta_i^{(r)}\right)^2, \quad \overline{\text{MSE}}_\ell\left(\hat{\theta}\right) = \frac{5}{m}\sum_{i\in\{j:\psi_j = 50/n_\ell\}}\text{MSE}\left(\hat{\theta}_i\right), \quad \ell = 1,\ldots,5.$$

2. $E\left(\hat{\sigma}_v^2\right) = \sum_{r=1}^{10{,}000}\hat{\sigma}_v^{2(r)}\Big/10{,}000$, $V\left(\hat{\sigma}_v^2\right) = \sum_{r=1}^{10{,}000}\left(\hat{\sigma}_v^{2(r)} - E\left(\hat{\sigma}_v^2\right)\right)^2\Big/10{,}000$, where $\hat{\sigma}_v^{2(r)}$ is the value of $\hat{\sigma}_v^2$ for the r^{th} simulation run $(r = 1,\ldots,10{,}000)$.
3. The Average Relative Bias (ARB) of the MSE per sampling variance group:

$$\text{ARB}_\ell\,(\text{mse}) = \frac{5}{m}\sum_{i\in\{j:\psi_j=50/n_\ell\}}\text{RB}\left(\text{mse}\left(\hat{\theta}_i\right)\right), \quad \ell = 1,\ldots 5,$$

   where $\text{RB}\left(\text{mse}\left(\hat{\theta}_i\right)\right) = \left[\sum_{r=1}^{10{,}000}\text{mse}\left(\hat{\theta}_i^{(r)}\right)\Big/10{,}000 - \text{MSE}\left(\hat{\theta}_i\right)\right]\Big/\text{MSE}\left(\hat{\theta}_i\right)$.
4. The Root Relative MSE of MSE estimators per sampling variance group:

$$\text{RRMSE}_\ell\,(\text{mse}) = \left(\frac{5}{m}\sum_{i\in\{j:\psi_j=50/n_\ell\}}\frac{\sum_{r=1}^{10{,}000}\left(\text{mse}\left(\hat{\theta}_i^{(r)}\right) - \text{MSE}\left(\hat{\theta}_i\right)\right)^2\Big/10{,}000}{\text{MSE}\left(\hat{\theta}_i\right)}\right)^{1/2}.$$

   We also examine the bias of the conditional MSE estimators given that $\left\{\hat{\sigma}_{v\text{REML}}^2 = 0\right\}$ because these are the populations for which the positive estimators were developed.
5. The Average Relative Bias of Conditional MSE estimators:

$$\text{ARB}_{\mathbb{C}} = \frac{5}{m}\sum_{i\in\ell}\text{E}\left[\text{mse}\left(\hat{\theta}_i\right)\middle|\,\hat{\sigma}_{v\text{REML}}^2 = 0\right]\Big/\text{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\middle|\hat{\sigma}_{v\text{REML}}^2 = 0\right] - 1.$$

6 Simulation results and analysis

6.1 Monte Carlo Distribution of the variance estimators

Table 6.1 shows that the REML variance estimator has the lowest bias $(\sigma_v^2 = 1)$ and the highest variance. The lower efficiency of REML may be due to it not being a smooth function of the data caused by its split definition (3.1). The MIX estimator inherits some of this low efficiency. The other variance estimators have lower variability, higher positive bias but the conditional expectation of AM.YL and AR.YL given $\hat{\sigma}_{\text{vREML}}^2 = 0$ is close to zero. The unconditional bias of AM.LL is higher than the unconditional bias of the MIX. By definition of the MIX estimator, the conditional bias of the MIX and AM.LL estimators coincide. The MIX estimator also converges faster than the other estimators. For example, given the probability distribution over the 10,000 variance estimates with $m = 45,$ we calculated the probability of estimates lying within an interval containing $\sigma_v^2 = 1.$ The probability that the MIX estimates lie between 0.6 and 1.4 is 0.47 whereas the probability that AM.YL estimates lie between 0.6 and 1.4 is 0.16. Furthermore, the probability that MIX estimates are smaller than 0.2 is 0.05 whereas the probability that AM.YL estimates are smaller than 0.2 is 0.53.

Table 6.1
Expectation, variance and conditional expectation and variance of $\hat{\sigma}_v^2$

Method	m	$E(\hat{\sigma}_v^2)$	$V(\hat{\sigma}_v^2)$	%REML = 0	$E(\hat{\sigma}_v^2/\text{REML} = 0)$	$V(\hat{\sigma}_v^2/\text{REML} = 0)$
REML	15	1.48	3.38	43%	N/A	N/A
	45	1.21	1.67	29%	N/A	N/A
	100	1.07	0.81	16%	N/A	N/A
AM.LL	15	2.80	1.37	43%	1.80	0.11
	45	1.88	1.01	29%	0.94	0.03
	100	1.49	0.51	16%	0.63	0.01
MIX	15	2.28	1.87	43%	1.80	0.11
	45	1.48	1.31	29%	0.94	0.03
	100	1.17	0.66	16%	0.63	0.01
AR.YL	15	1.66	2.99	43%	0.27	0.01
	45	1.24	1.72	29%	0.06	0.00
	100	1.08	0.80	16%	0.02	0.00
AM.YL	15	0.52	0.84	43%	0.10	0.00
	45	0.65	0.85	29%	0.03	0.00
	100	0.76	0.59	16%	0.01	0.00

6.2 True MSE of the EBLUP, average relative bias and average root relative MSE of the MSE estimators

All variance estimators are consistent and asymptotically normal with variance converging at the same rate. They differ in their bias: REML, AR.YL and MIX have bias of the order of $o(1/m)$ whereas AM.LL and AM.YL have bias of the order $O(1/m)$. The bias inherent in the last three methods impacts the estimation of the MSE of the EBLUP even for a moderate number of areas.

For $m = 100,$ Tables 6.2a and 6.2b show that as σ_v^2/ψ_i increases, the MSE of the EBLUP decreases and this relationship holds irrespective of the number of areas. We observe that the MSE of $\hat{\theta}_i$ under the

REML and the MIX variance estimators are slightly higher than the rest of the MSEs, due to the higher variability inherent in these variance estimators. Table 6.2a presents results for the Taylor linearization MSE estimator and the two parametric MSE estimators under REML, AM.LL, AR.YL and AM.YL variance estimation. Table 6.2b presents results for the following MSE estimators under the MIX variance estimation: RB_Y1 defined in (4.3), RB_Y2 defined in (4.2), M_et_al, defined in (4.5), PB MSE and naïve PB MSE estimators. Among the Taylor MSE estimators, RB_Y1 and M_et_al under MIX exhibit the lowest bias. Among the bootstrap MSE estimators PB under MIX and Naive PB under AR.YL exhibit the lowest bias. Turning to the RRMSE of the MSE estimators, it decreases as σ_v^2 / ψ_i increases. Differences between the RB_Y2 MSE estimator under the MIX and the Taylor MSE estimator under the AM.YL seem small but consistent. While ARB is lower for the RB_Y1, the M_et_al and the Naive MSE estimators under the MIX method than for the RB_Y2 under the MIX, and also lower for the Taylor and the Naive PB under the AR.YL method than for the RB_Y2 under the MIX, the opposite happens in terms of RRMSE. This can be explained in part due to the extreme negative conditional bias exhibited by these MSE estimators (i.e., the RB_Y1 and the M_et_al under the MIX and the Taylor and the Naive PB under the AR.YL method) as shown in Table 6.3. Even for $m = 100$ there is a relatively high proportion (16%) of populations that yield $\hat{\sigma}_{v\text{REML}}^2 = 0$ and in these populations, estimates from most variance methods and most MSE estimators are farthest below the true value. That is, for these MSE estimators, the conditional MSE estimators do not fare well. The PB MSE estimator seems to adjust well for bias, but it is more variable than the Naive PB MSE. When we also include the ARB, the RRMSE and the $\text{ARB}_{\mathbb{C}}$ in the evaluation, the RB_Y2 under the MIX method, followed closely by Naïve PB under the MIX seems to perform the best. This may suggest the superiority of RB_Y2 and Naive under MIX for $m = 100,$ which is a moderate number of areas for this data.

Table 6.2a
MSE, ARB & RRMSE (percentage) of MSE Estimators, $m = 100$

			Taylor MSE estimator		PB estimator		Naïve PB estimator	
Method	σ_v^2 / ψ_i	$\overline{\text{MSE}}$	**ARB**	**RRMSE**	**ARB**	**RRMSE**	**ARB**	**RRMSE**
REML	0.06	135.4	5.1	71.1	-4.4	80.7	1.6	69.9
	0.1	132.1	5.3	64.7	-4.7	74.0	-0.2	63.0
	0.14	119.5	6.0	61.9	-5.5	71.3	-1.8	59.9
	0.2	119.2	6.5	53.6	-5.8	62.4	-3.4	51.7
	0.3	106.6	8.2	46.7	-6.8	55.0	-5.6	44.8
AM.LL	0.06	134.9	6.1	75.4	8.2	66.9	31.3	63.8
	0.1	131.2	6.8	68.1	7.8	59.5	27.5	55.7
	0.14	118.3	8.1	64.6	7.8	55.6	26.5	51.2
	0.2	117.6	8.4	55.4	6.5	46.7	21.6	42.1
	0.3	104.5	10.2	46.7	5.5	38.8	18.2	34.0
AR.YL	0.06	135.4	6.6	69.3	-4.3	80.2	2.1	69.4
	0.1	132.0	7.4	61.9	-4.5	73.4	0.3	62.5
	0.14	119.4	9.0	58.0	-5.3	70.6	-1.2	59.3
	0.2	119.0	10.6	48.2	-5.6	61.8	-2.9	51.1
	0.3	106.4	14.7	38.5	-6.6	54.3	-5.1	44.1
AM.YL	0.06	134.7	10.0	63.2	-12.3	81.0	-19.6	65.9
	0.1	131.3	12.0	56.6	-12.5	75.2	-19.7	61.2
	0.14	118.8	15.0	53.1	-13.7	73.3	-21.4	59.8
	0.2	118.6	18.1	44.8	-13.4	65.2	-20.7	53.5
	0.3	106.4	25.2	38.4	-14.4	58.8	-21.7	48.6

Table 6.2b
MSE, ARB & RRMSE (percentage) of MSE Estimators $m = 100$

	σ_v^2/ψ_i	$\overline{\text{MSE}}$	RB_Y1		RB_Y2		M_et_al		PB estimator		Naïve PB estimator	
			ARB	RRMSE	ARB	RRMSE	ARB	RRMSE	ARB	RRMSE	ARB	RRMSE
MIX	0.06	135.4	2.7	75.7	13.6	63.0	5.2	71.1	-3.0	75.3	8.8	62.4
	0.1	132.1	3.6	68.3	14.9	56.1	5.3	64.7	-3.2	68.3	6.6	55.4
	0.14	119.5	4.9	64.7	16.0	52.4	6.0	61.9	-3.9	65.1	5.3	51.8
	0.2	119.1	6.3	55.2	16.7	43.8	6.5	53.6	-4.4	56.3	2.9	43.7
	0.3	106.5	9.4	46.2	19.9	36.0	8.3	46.7	-5.4	48.6	0.6	36.7

Table 6.3
$\text{MSE}_{\mathbb{C}}$ $\left(\text{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2 \middle| \hat{\sigma}^2_{v\text{REML}} = 0\right]\right)$ and $\text{ARB}_{\mathbb{C}}$ (percentage), $m = 100$

Method	σ_v^2/ψ_i	$\overline{\text{MSE}}_{\mathbb{C}}$	Taylor MSE estimator	PB estimator	Naïve PB estimator
REML	0.06	135.6	-76.5	-98.6	-74.8
	0.1	133.0	-74.5	-94.4	-71.8
	0.14	121.5	-78.6	-98.0	-74.9
	0.2	120.4	-73.1	-89.8	-68.6
	0.3	108.0	-73.6	-88.2	-67.3
AM.LL	0.06	135.0	-92.0	-67.6	-26.1
	0.1	132.2	-85.2	-62.0	-24.6
	0.14	120.2	-85.4	-62.3	-25.4
	0.2	118.8	-74.4	-54.1	-22.1
	0.3	105.9	-65.9	-49.7	-20.6
AR.YL	0.06	135.5	-68.6	-96.9	-73.0
	0.1	132.9	-62.4	-92.6	-70.0
	0.14	121.4	-61.1	-96.1	-73.0
	0.2	120.2	-48.9	-87.9	-66.7
	0.3	107.8	-34.5	-86.1	-65.4
AM.YL	0.06	134.9	-45.9	-88.6	-74.7
	0.1	132.1	-39.4	-85.4	-72.2
	0.14	120.4	-36.0	-89.3	-75.7
	0.2	119.6	-23.6	-82.3	-69.7
	0.3	107.6	-6.5	-81.7	-69.3

			RB_Y1	RB_Y2	M_et_al	PB estimator	Naïve PB estimator
MIX	0.06	135.0	-92.0	-22.0	-76.4	-46.0	-27.0
	0.1	132.2	-85.2	-17.7	-74.3	-42.7	-25.9
	0.14	120.2	-85.4	-15.0	-78.3	-43.3	-27.0
	0.2	118.8	-74.4	-7.6	-72.8	-37.6	-23.9
	0.3	105.9	-65.9	1.5	-73.1	-34.6	-22.6

Tables 6.4a and b below display results for $m = 45$ with 9 areas per σ_v^2/ψ_i. The AM.YL yields MSEs smaller than the MIX, with differences in MSEs of at most 2%. As the number of areas decreases, the bias of the variance estimators increase and the MSE estimators are affected by this. Indeed, the ARB of all MSE estimators have increased. In particular, the ARB of the Taylor MSE estimators under YL and LL variance estimation and the ARB of RB_Y2, have increased by 100% over the ARB with 100 areas. In terms of RRMSE, the Taylor MSE under the AM.YL has slightly lower RRMSE than the RB_Y2 under the MIX method for very small σ_v^2/ψ_i. In general, the variability (in RRMSE) of the RB_Y2 is lower than that of the Taylor under LL and YL estimation and than that of the RB_Y1 and the M_et_al. This may be due in part to the underestimation of the MSEs for the populations with zero REML estimates, which, for $m = 45$,

range around 30% of all populations. Table 6.5 illustrates this better: given $\hat{\sigma}^2_{v\text{REML}} = 0$, there is serious underestimation in RB_Y1 and M_et_al.

Table 6.4a
MSE, ARB & RRMSE (percentage) of MSE Estimators, $m = 45$ areas

Method	σ_v^2/ψ_i	$\overline{\text{MSE}}$	Taylor MSE estimator		PB estimator		Naïve PB estimator	
			ARB	RRMSE	ARB	RRMSE	ARB	RRMSE
REML	0.06	171.4	11.8	94.7	-4.7	107.0	6.2	89.2
	0.1	174.1	11.9	83.9	-5.3	93.8	3.0	76.2
	0.14	171.3	12.6	74.5	-5.4	81.9	1.1	65.3
	0.2	166.6	13.9	63.4	-5.8	66.7	-1.2	52.0
	0.3	128.9	20.1	63.0	-7.0	61.4	-3.1	46.7
AM.LL	0.06	171.1	15.5	100.0	16.0	84.9	43.5	83.3
	0.1	173.4	16.8	87.0	14.4	71.1	36.7	68.5
	0.14	170.4	17.7	75.7	12.6	59.7	30.7	56.7
	0.2	165.3	18.2	61.7	9.9	46.2	23.5	43.2
	0.3	127.5	25.6	55.0	10.0	39.7	22.6	36.6
AR.YL	0.06	171.1	17.2	89.9	-3.7	105.0	8.0	87.6
	0.1	173.6	19.6	76.9	-4.3	91.8	4.8	74.6
	0.14	170.8	22.6	65.8	-4.4	79.9	2.7	63.7
	0.2	166.0	27.3	53.7	-4.8	64.8	0.3	50.5
	0.3	128.3	43.8	54.8	-5.7	59.3	-1.3	45.0
AM.YL	0.06	167.5	30.2	78.4	-18.0	97.3	-23.8	73.3
	0.1	169.6	36.5	72.2	-18.0	87.7	-23.6	66.7
	0.14	167.0	42.7	69.3	-17.2	78.0	-22.3	59.7
	0.2	162.8	52.1	70.8	-15.8	65.4	-20.3	50.6
	0.3	126.0	81.3	91.1	-18.0	62.3	-22.9	48.4

Table 6.4b
MSE, ARB & RRMSE (percentage) of MSE Estimators, $m = 45$ areas

	σ_v^2/ψ_i	$\overline{\text{MSE}}$	RB_Y1		RB_Y2		M_et_al		PB estimator		Naïve PB estimator	
			ARB	RRMSE	ARB	RRMSE	ARB	RRMSE	ARB	RRMSE	ARB	RRMSE
MIX	0.06	171.4	9.8	99.4	31.9	84.0	11.8	94.7	3.5	93.8	21.9	78.5
	0.1	174.0	12.1	86.2	33.2	73.1	11.9	83.9	2.6	80.4	17.5	65.1
	0.14	171.2	14.5	74.9	34.4	64.6	12.6	74.5	2.0	68.7	14.0	54.4
	0.2	166.5	17.7	61.7	36.0	55.8	13.9	63.4	0.7	54.5	9.8	41.8
	0.3	128.9	28.8	57.6	48.8	58.2	20.2	63.1	0.3	48.6	8.7	35.9

Taking into account the ARB, the RRMSE and the $\text{ARB}_{\mathbb{C}}$ of the MSE estimators, the Naive PB MSE estimator under the MIX performs the best for larger σ_v^2/ψ_i. Table 6.6 displays performance measures, averaged over the five sampling variance groups, for the three Taylor MSE estimators under the MIX with data from the same model described in 5.1 but with three different values of σ_v^2. The RB_Y2 performs better when $\sigma_v^2 = 1$, but as σ_v^2 becomes smaller, the M_et_al MSE estimator has an advantage, precisely because it was constructed under the premise that σ_v^2 is approximately zero.

Table 6.5
$\mathbf{MSE_C}$ and $\mathbf{ARB_C}$ (percentage). $m = 45$ areas

Method	σ_v^2/ψ_i	$\overline{\mathrm{MSE}}_{\mathrm{C}}$		Taylor MSE estimator		PB estimator	Naïve PB estimator
REML	0.06	170.2		-64.3		-89.7	-60.7
	0.1	173.0		-62.4		-83.7	-57.1
	0.14	170.2		-58.1		-75.5	-51.8
	0.2	165.8		-51.9		-65.1	-44.8
	0.3	131.1		-59.0		-70.5	-49.2
AM.LL	0.06	170.0		-71.5		-49.0	-3.1
	0.1	172.3		-61.5		-42.1	-2.3
	0.14	169.1		-51.1		-35.7	-2.1
	0.2	164.7		-38.3		-28.3	-1.6
	0.3	129.9		-28.8		-29.1	-3.7
AR.YL	0.06	169.9		-48.3		-86.2	-56.7
	0.1	172.6		-38.0		-80.2	-53.2
	0.14	169.7		-25.9		-72.2	-48.2
	0.2	165.3		-7.4		-61.9	-41.5
	0.3	130.5		19.3		-66.8	-45.5
AM.YL	0.06	166.6		-8.2		-73.5	-60.7
	0.1	168.8		3.8		-70.1	-58.1
	0.14	166.1		16.1		-64.1	-53.3
	0.2	162.2		35.9		-56.1	-46.8
	0.3	128.1		72.8		-62.5	-52.5
			RB_Y1	**RB_Y2**	**M_et_al**		
MIX	0.06	170.0	-71.5	6.2	-64.3	-28.1	-4.0
	0.1	172.3	-61.5	13.2	-62.3	-23.8	-3.5
	0.14	169.1	-51.1	18.9	-57.8	-20.0	-3.3
	0.2	164.7	-38.3	26.8	-51.6	-15.7	-2.9
	0.3	129.9	-28.8	40.4	-58.7	-16.7	-5.1

Table 6.6
MSE, ARB, $\mathbf{ARB_C}$ and RRMSE (percentage), 45 areas

			RB_Y1			RB_Y2			M_et_al		
%REML = 0	σ_v^2	$\overline{\mathrm{MSE}}$	**ARB**	**$\mathrm{ARB_C}$**	**RRMSE**	**ARB**	**$\mathrm{ARB_C}$**	**RRMSE**	**ARB**	**$\mathrm{ARB_C}$**	**RRMSE**
29	1	108	16	-50	75	36	21	66	14	-59	75
48	0.2	99	48	-36	101	113	88	114	47	-38	94
51	0.1	91	58	-33	108	137	107	127	58	-32	100

Tables 6.7a and 6.7b below show the outcomes for $m = 15$ areas with 3 areas per σ_v^2/ψ_i. Differences in MSEs per variance estimation method are at most 5%.

There is no monotone relationship between ARB or RRMSE and σ_v^2/ψ_i, which could be an indication that the second order approximation to estimating the MSE is poor under every method of variance estimation. The ARB of all Taylor MSE estimators under the LL and the YL methods of variance estimation are unacceptably high and the same is true for the RRMSE. The RB_Y2 under the MIX does not fare well either. The reason for this last outcome is clear: the high % of zero REML estimates (43%) implies the MIX coincides with AM.LL for the zero REML populations. Thus, the MIX has a positive bias for $m = 15,$ and the RB_Y2 does not account for this bias. The RB_Y1 accounts for the bias in the MIX, but the bias estimator is not very precise for $m = 15.$ The M_et_al MSE estimator almost coincides with the ARB and RRMSE of the Taylor MSE estimator under the REML variance estimation, because by definition they are equal when $\hat{\sigma}^2_{v\text{REML}} = 0.$ The $\text{ARB}_{\mathbb{C}}$ of the three Taylor MSE estimators under the MIX is poor. Taking into account all performance measures, the bootstrap MSE estimators perform better than the Taylor MSE estimators. For $m = 15$ areas with 3 areas per σ^2_v / ψ_i, PB under MIX performs the best, followed by the Naive under AR.YL and AM.YL.

Table 6.7a
MSE, ARB & RRMSE (percentage) of MSE estimators, $m = 15$ areas

			Taylor MSE estimator		PB estimator		Naïve PB estimator	
Method	σ^2_v / ψ_i	$\overline{\text{MSE}}$	**ARB**	**RRMSE**	**ARB**	**RRMSE**	**ARB**	**RRMSE**
REML	0.06	584.8	12.6	87.9	1.2	85.9	6.9	64.5
	0.1	376.7	26.5	106.3	2.3	85.6	9.6	62.8
	0.14	352.5	25.2	90.1	0.7	54.1	4.3	39.3
	0.2	209.4	43.0	123.0	0.4	74.0	6.3	51.1
	0.3	198.7	50.6	124.7	-1.0	46.3	2.6	31.5
AM.LL	0.06	589.3	24.1	89.3	13.7	61.2	24.1	65.8
	0.1	380.7	48.3	107.1	19.4	58.6	32.5	62.9
	0.14	355.7	40.2	88.6	10.0	36.2	16.8	38.1
	0.2	212.5	76.3	117.9	17.8	45.1	28.7	47.3
	0.3	200.7	76.5	105.1	10.7	26.9	17.2	27.6
AR.YL	0.06	583.3	23.8	83.3	3.2	79.5	3.2	61.6
	0.1	375.1	53.3	106.7	5.4	78.6	5.4	59.7
	0.14	351.3	53.3	102.7	2.4	49.4	2.4	37.1
	0.2	207.7	107.3	153.1	4.1	66.2	4.1	47.2
	0.3	197.5	142.0	199.4	1.9	41.1	1.9	28.9
AM.YL	0.06	571.4	41.6	103.5	-8.0	61.2	-9.2	43.3
	0.1	363.3	95.0	161.4	-11.3	62.9	-13.2	44.1
	0.14	342.0	97.2	179.7	-6.7	40.4	-7.8	29.3
	0.2	197.0	198.4	274.6	-14.5	58.2	-16.7	41.7
	0.3	191.4	270.2	362.4	-11.5	38.4	-13.1	28.7

Table 6.7b
MSE, ARB & RRMSE (percentage) of MSE estimators, $m = 15$ areas

			RB_Y1		RB_Y2		M_et_al		PB estimator		Naïve PB estimator	
	σ^2_v / ψ_i	$\overline{\text{MSE}}$	**ARB**	**RRMSE**	**ARB**	**RRMSE**	**ARB**	**RRMSE**	**%ARB**	**%RRMSE**	**%ARB**	**%RRMSE**
MIX	0.06	584.9	21.0	84.7	35.4	93.7	12.6	87.9	10.0	53.8	19.3	62.1
	0.1	377.1	46.0	103.9	68.4	122.6	26.4	106.1	14.8	52.7	26.6	59.9
	0.14	353.0	41.9	91.5	59.4	112.7	25.0	89.9	7.6	33.2	13.7	36.7
	0.2	209.7	83.2	127.8	108.9	155.8	42.8	122.8	14.0	42.5	23.7	46.0
	0.3	198.9	94.8	136.7	117.1	162.2	50.4	124.6	8.7	26.6	14.5	27.7

Summarizing, under the Fay-Herriot model with positive σ_v^2, and among the positive variance estimators under study, the MIX and the AR.YL variance estimators are the only ones with negligible asymptotic bias. The AM.YL and the LL variance estimators have a larger asymptotic bias. On the other hand, our simulation showed that for a moderate number of areas and for populations that yield zero REML estimates, both YL variance estimators were negatively biased, and produced EBLUPs that were close to the synthetic estimator of the mean. In contrast, the MIX, built as the combination of the AM.LL and the REML, was only mildly negatively biased in these populations. Moreover, the unconditional distribution of the MIX approached normality much faster than those of the other variance estimators.

Table 6.8
$\mathrm{MSE}_{\mathbb{C}}$ and $\mathrm{ARB}_{\mathbb{C}}$ m = 15 areas

Method	σ_v^2/ψ_i	$\overline{\mathrm{MSE}}_{\mathbb{C}}$		Taylor MSE estimator		PB estimator	Naïve PB estimator
REML	0.06	594.2		-22.6		-31.7	-16.5
	0.1	381.2		-32.9		-43.2	-22.5
	0.14	345.1		-17.7		-22.7	-10.7
	0.2	212.7		-41.1		-47.3	-25.5
	0.3	197.9		-30.4		-32.7	-17.6
AM.LL	0.06	595.6		-4.1		-5.7	12.1
	0.1	385.7		8.6		-7.0	15.6
	0.14	351.2		18.9		-2.0	10.4
	0.2	216.0		46.4		-5.8	14.4
	0.3	199.5		67.0		-2.9	9.8
AR.YL	0.06	592.2		-0.8		-27.1	-11.0
	0.1	379.7		21.0		-36.5	-14.8
	0.14	344.5		44.0		-18.6	-6.3
	0.2	210.9		98.2		-38.6	-16.4
	0.3	196.6		177.3		-26.1	-11.0
AM.YL	0.06	581.7		30.7		-21.9	-18.0
	0.1	368.6		79.8		-31.5	-25.8
	0.14	333.9		98.3		-15.2	-11.9
	0.2	198.9		198.0		-36.4	-30.0
	0.3	190.0		296.3		-26.2	-21.5
			RB_Y1	**RB_Y2**	**M_et_al**	**PB estimator**	**Naive PB estimator**
MIX	0.06	595.6	-4.1	27.9	-22.9	3.4	17.8
	0.1	385.7	8.6	57.1	-33.7	5.1	22.8
	0.14	351.2	18.9	58.5	-19.1	4.9	14.3
	0.2	216.0	46.4	102.4	-42.0	5.9	20.4
	0.3	199.5	67.0	116.3	-30.9	4.8	13.4

In terms of MSE of the EBLUP, there were considerable gains in precision over the direct estimator, under all methods of variance estimation considered here, even for a small number of areas. The AM.LL and both the AM.YL and the AR.YL variance estimators carried lower variability than the REML and the MIX. It impacted only minimally the MSE of the EBLUP: differences among MSEs for the same signal to noise ratio were small. These differences widened as either the number of areas or the signal to noise ratio decreased. Thus, it may possible that for an extremely low signal to noise ratio, the MSE under MIX would be somewhat larger than under the AM.YL variance estimator.

Under the MIX method of variance estimation, we compared three different Taylor-type MSE estimators and two bootstrap MSE estimators. All three Taylor estimators of the MSE under MIX (RB_Y1, RB_Y2 and M_et_al) are unbiased up to the second order. Also the Taylor-type estimators of the MSE under the LL and the YL are unbiased up to the second order. RB_Y1, AM.LL and AM.YL may yield negative MSE estimates.

The Taylor MSE under the REML method of variance estimation and the M_et_al under the MIX coincide by definition, hence their performance measures have negligible differences (their true MSEs are different, however in our study, for $m = 100,$ the MIX coincided with the REML 84% of the time). For a moderate number of areas, which for this data could be $m = 45$ or $100,$ and for populations that yield zero REML estimates, both the Taylor MSE estimators under the REML and the M_et_al MSE estimators do not account for the variation due to the estimation of σ_v^2 and this is reflected in their very negative $\text{ARB}_{\mathbb{C}},$ which is below -60% for the smaller signal to noise ratios. On the other hand, the RB_Y1 does account for the variation due to the estimation of $\sigma_v^2,$ but its $\text{ARB}_{\mathbb{C}}$ is also very negative: the RB_Y1 is a split MSE estimator that for populations with $\hat{\sigma}_{v\text{REML}}^2 = 0,$ it subtracts a factor of the unconditional bias of the AM.LL, which is always positive, whereas a better formula for a split MSE estimator would be to use an estimator of the conditional bias $E\left(\hat{\sigma}_v^2 / \hat{\sigma}_{v\text{REML}}^2 = 0\right).$ Indeed, even for a moderate number of areas $(m = 100),$ Table 6.1 shows that the unconditional bias of the MIX is 49% whereas the conditional bias of the MIX is -37%.

The PB MSE estimator under the AR.YL and the MIX methods adjusted well for the bias but paid in terms of variance. Among all the MSE estimators it appears that the Naive Bootstrap MSE estimator performed best, and even better under the MIX variance estimation, when taking into account the three measures ARB, $\text{ARB}_{\mathbb{C}}$ and RRMSE together. We found that for a moderate number of areas, the RB_Y2 had the lowest RRMSE among the Taylor estimators under the MIX method. On the other hand, M_et_al is most reliable when the true underlying variance σ_v^2 is very small: in this case M_et_al is effectively the MSE estimator of the synthetic estimator of the small area mean. We do not recommend relying on the second order approximation to the MSE when m is small: the approximation (2.6) to the MSE does not necessarily hold, the performance measures obtained from our study are very unstable and they may vary from data set to data set.

In conclusion, under the hypothesis of $\sigma_v^2 > 0,$ the relative performances of competing positive variance estimators depend on the size of $\sigma_v^2,$ the signal to noise ratio, the number of areas and the objective function. For a moderate number of areas, the MIX variance estimator appeared to perform better than the LL and the YL estimators in this study; under the MIX method, the Naive PB MSE estimator had the lowest $\text{ARB}_{\mathbb{C}}$ and RRMSE combined; the M_et_al MSE estimator under the MIX variance estimator performed marginally better than the RB_Y1 when the underlying σ_v^2 was very small. However, the percentage of REML zeros yielded under the simulation model shows that an outcome of $\hat{\sigma}_{v\text{REML}}^2 = 0$ and/or negative tests of hypothesis do not necessarily mean that σ_v^2 is sufficiently small to rely on M_et_al. In the absence of other information, the Naive PB estimator under the MIX appears to perform better.

Acknowledgements

The authors would like to thank Professor J.N.K. Rao from Carleton University for his useful comments and to Victor Estevao from Statistics Canada for developing the grid maximization especially for this project. We also would like to thank the reviewers for their careful appraisal of our paper and for their suggestions to improve this paper.

Appendix A

Proof of Theorem 4.1

The asymptotic variance of $\hat{\sigma}^2_{v\text{MIX}}$ is given by: $\bar{V}\left(\hat{\sigma}^2_{v\text{MIX}}\right) = \lim_{m\to\infty} E\left(\hat{\sigma}^2_{v\text{MIX}} - \sigma^2_v\right)^2$

We show that $E\left(\hat{\sigma}^2_{v\text{MIX}} - \sigma^2_v\right)^2 \le E\left(\hat{\sigma}^2_{v\text{REML}} - \sigma^2_v\right)^2 + o(1/m)$ as $m \to \infty$.

$$
\begin{aligned}
E\left(\hat{\sigma}^2_{v\text{MIX}} - \sigma^2_v\right)^2 &= \int_{\left\{\hat{\sigma}^2_{v\text{REML}} > 0\right\}} \left(\hat{\sigma}^2_{v\text{REML}} - \sigma^2_v\right)^2 dP + \int_{\left\{\hat{\sigma}^2_{v\text{REML}} = 0\right\}} \left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^2 dP \\
&\le \int_{\Omega} \left(\hat{\sigma}^2_{v\text{REML}} - \sigma^2_v\right)^2 dP + \int_{\left\{\hat{\sigma}^2_{v\text{REML}} = 0\right\}} \left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^2 dP = E\left(\hat{\sigma}^2_{v\text{REML}} - \sigma^2_v\right)^2 \\
&\quad + o\left(\frac{1}{m}\right).
\end{aligned} \tag{A.1}
$$

Indeed, by the Holder and Minkowski inequalities, with any $1 < p < \infty, 1/p + 1/q = 1,$ and setting $X \equiv \left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^2 = \text{O}_p\,(1/m)$ and the indicator $I\left(\hat{\sigma}^2_{v\text{REML}} = 0\right)$ of populations with $\hat{\sigma}^2_{v\text{REML}} = 0,$ we have:

$$
\begin{aligned}
\int_{\left\{\hat{\sigma}^2_{v\text{REML}} < 0\right\}} \left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^2 dP &\le \left(\int_{\Omega} \left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^{2p} dP\right)^{1/p} . \left(P\left\{\hat{\sigma}^2_{v\text{REML}} = 0\right\}\right)^{1/q} \\
&= \left(O\left(\frac{1}{m^p}\right)\right)^{1/p} . \left(o\,(1)\right)^{1/q} = o\left(\frac{1}{m}\right),
\end{aligned} \tag{A.2}
$$

since $\left(\hat{\sigma}^2_{v\text{AM.LL}} - \sigma^2_v\right)^2$ is uniformly bounded and $\hat{\sigma}^2_{v\text{REML}} \xrightarrow{P} \sigma^2_v > 0.$ Note that the AM.LL and REML estimators of σ^2_v are uniformly bounded as a consequence of their almost sure convergence to σ^2_v (see, for example, Yuan and Jennrich 1998).

Proof of Theorem 4.2

We denote by $\hat{\boldsymbol{\sigma}}^2_{v\text{ML}}$ the maximum likelihood variance estimator.

We show first that $\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} - \hat{\boldsymbol{\sigma}}^2_{v\text{ML}} = O_p\,(1/m).$ Let $G_*\left(\boldsymbol{\sigma}^2_v\right) = \partial \log\left(L_*\right) / \partial \boldsymbol{\sigma}^2_v = 0$ be the estimating equation that yields the variance estimator *. Equation (3.4) implies:

$$
G_{\text{AM.LL}}\left(\boldsymbol{\sigma}^2_v\right) - G_{\text{ML}}\left(\boldsymbol{\sigma}^2_v\right) = \partial \log \boldsymbol{\sigma}^2_v / \partial\, \boldsymbol{\sigma}^2_v = \frac{1}{m\boldsymbol{\sigma}^2_v} = O\left(\frac{1}{m}\right). \tag{A.3}
$$

With $G'_{\text{ML}}\,(\cdot) \triangleq \left(\partial G_{\text{ML}} / \partial \boldsymbol{\sigma}^2_v\right)(\cdot)$ and $G''_{\text{ML}}\,(\cdot) \triangleq \left(\partial G'_{\text{ML}} / \partial \boldsymbol{\sigma}^2_v\right)(\cdot),$ equation (A.3) implies:

$$G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right) - G'_{\mathrm{AM.LL}}\left(\boldsymbol{\sigma}_v^2\right) = O\left(\frac{1}{m}\right). \tag{A.4}$$

Now, using equation (A.4), the $\sqrt{m}$ – consistency of the ML and AM.LL estimators of $\boldsymbol{\sigma}_v^2$, the two-term Taylor expansion of $G_{\mathrm{ML}}(\cdot)$ and $G_{\mathrm{AM.LL}}(\cdot)$ at $\boldsymbol{\sigma}_v^2$ and $G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right) = O(1)$ as $m \to \infty$, the left- hand side in (A.3) is equal to:

$$\begin{aligned}
&= G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right)\left(\hat{\boldsymbol{\sigma}}_{v\mathrm{ML}}^2 - \boldsymbol{\sigma}_v^2\right) - G'_{\mathrm{AM.LL}}\left(\boldsymbol{\sigma}_v^2\right)\left(\hat{\boldsymbol{\sigma}}_{v\mathrm{AM.LL}}^2 - \boldsymbol{\sigma}_v^2\right) + O_p\left(\frac{1}{m}\right)\\
&= G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right)\left(\hat{\boldsymbol{\sigma}}_{v\mathrm{ML}}^2 - \hat{\boldsymbol{\sigma}}_{v\mathrm{AM.LL}}^2\right) + \left(G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right) - G'_{\mathrm{AM.LL}}\left(\boldsymbol{\sigma}_v^2\right)\right)\left(\hat{\boldsymbol{\sigma}}_{v\mathrm{AM.LL}}^2 - \boldsymbol{\sigma}_v^2\right) + O_p\left(\frac{1}{m}\right)\\
&= G'_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right)\left(\hat{\boldsymbol{\sigma}}_{v\mathrm{ML}}^2 - \hat{\boldsymbol{\sigma}}_{v\mathrm{AM.LL}}^2\right) + O_p\left(\frac{1}{m^{3/2}}\right) + O_p\left(\frac{1}{m}\right).
\end{aligned}$$

The last equality above implies

$$\hat{\boldsymbol{\sigma}}_{v\mathrm{AM.LL}}^2 - \hat{\boldsymbol{\sigma}}_{v\mathrm{ML}}^2 = O_p\left(\frac{1}{m}\right) \text{ as } m \to \infty. \tag{A.5}$$

Similarly, we establish a relationship between $G_{\mathrm{REML}}\left(\boldsymbol{\sigma}_v^2\right)$ and $G_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right)$: given that $\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{V}^{-1}\right) = O(1)$ follows from conditions 1 through 3 in Section 3 and equation (3.1), we have:

$$G_{\mathrm{REML}}\left(\boldsymbol{\sigma}_v^2\right) - G_{\mathrm{ML}}\left(\boldsymbol{\sigma}_v^2\right) = \frac{1}{m}\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{V}^{-1}\right) = O\left(\frac{1}{m}\right) \text{ as } m \to \infty, \tag{A.6}$$

Equation (A.6) and the same argument as with the AM.LL estimator, imply:

$$\hat{\boldsymbol{\sigma}}_{v\mathrm{REML}}^2 - \hat{\boldsymbol{\sigma}}_{v\mathrm{ML}}^2 = O_p\left(\frac{1}{m}\right) \text{ as } m \to \infty. \tag{A.7}$$

Equations (A.5) and (A.7) combined, yield:

$$\left(\hat{\sigma}_{v\mathrm{REML}}^2 - \hat{\sigma}_{v\mathrm{AM.LL}}^2\right) = O_P\left(\frac{1}{m}\right). \tag{A.8}$$

Now we express the bias of the MIX estimator by:

$$B_{\mathrm{MIX}}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right) = \int_{\left\{\hat{\sigma}_{v\mathrm{REML}}^2 > 0\right\}}\left(\hat{\sigma}_{v\mathrm{REML}}^2 - \sigma_v^2\right)dP + \int_{\left\{\hat{\sigma}_{v\mathrm{REML}}^2 = 0\right\}}\left(\hat{\sigma}_{v\mathrm{AM.LL}}^2 - \sigma_v^2\right)dP.$$

We add and subtract $\int_{\left\{\hat{\sigma}_{v\mathrm{REML}}^2 = 0\right\}}\left(\hat{\sigma}_{v\mathrm{REML}}^2 - \sigma_v^2\right)dP$ from the right-hand side of the equation above to obtain:

$$\begin{aligned}
B_{\mathrm{MIX}}\left(\hat{\sigma}_{v\mathrm{MIX}}^2\right) &= \int_{\Omega}\left(\hat{\sigma}_{v\mathrm{REML}}^2 - \sigma_v^2\right)dP + \int_{\left\{\hat{\sigma}_{v\mathrm{REML}}^2 = 0\right\}}\left(\hat{\sigma}_{v\mathrm{AM.LL}}^2 - \hat{\sigma}_{v\mathrm{REML}}^2\right)dP\\
&= \mathrm{Bias}\left(\hat{\sigma}_{v\mathrm{REML}}^2\right) + \int_{\left\{\hat{\sigma}_{v\mathrm{REML}}^2 = 0\right\}}\left(\hat{\sigma}_{v\mathrm{AM.LL}}^2 - \hat{\sigma}_{v\mathrm{REML}}^2\right)dP.
\end{aligned} \tag{A.9}$$

Now, since $\hat{\sigma}^2_{v\text{AM.LL}} - \hat{\sigma}^2_{v\text{REML}}$ is uniformly bounded, we apply the Holder and Minkowski inequality with $p = q = 2$ and equation (A.8) to the last term in (A.9) to obtain:

$$\begin{aligned} B_{\text{MIX}}\left(\hat{\sigma}^2_{v\text{MIX}}\right) &= \text{Bias}\left(\hat{\sigma}^2_{v\text{REML}}\right) + \left(\int_{\Omega}\left(\hat{\sigma}^2_{v\text{AM.LL}} - \hat{\sigma}^2_{v\text{REML}}\right)^2 dP\right)^{1/2} \cdot P\left\{\hat{\sigma}^2_{v\text{REML}} = 0\right\}^{1/2} \\ &= \text{Bias}\left(\hat{\sigma}^2_{v\text{REML}}\right) + O\left(\frac{1}{m}\right) \cdot o(1) = \text{Bias}\left(\hat{\sigma}^2_{v\text{REML}}\right) + o\left(\frac{1}{m}\right). \end{aligned} \tag{A.10}$$

Proof of Remark 4.2: mse_0 is unbiased up to the second order

$$\begin{aligned} E(\text{mse}_0) - \text{MSE}\left(\hat{\boldsymbol{\theta}}_i\right) &= \int_{\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} > 0\right\}} (g_{1i} + g_{2i} + 2g_{3i})\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP + \int_{\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0\right\}} g_{2i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP - \text{MSE} \\ &= \left[\int_{\Omega}(g_{1i} + g_{2i} + 2g_{3i})\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP - \text{MSE}\right] \\ &\quad + \int_{\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0\right\}} g_{2i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP - \int_{\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0\right\}} (g_{1i} + g_{2i} + 2g_{3i})\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP \\ &= \left[\text{o}\left(\frac{1}{m}\right)\right] - \int_{\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0\right\}} 2g_{3i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) dP, \end{aligned} \tag{A.11}$$

since $g_{1i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) = g_{1i}(0) = 0$ in $\left\{\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0\right\}$ and $g_{2i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right)$ cancels out in (A.11). But

$$g_{3i}\left(\hat{\boldsymbol{\sigma}}^2_{v\text{REML}}\right) = g_{3i}(0) = \frac{\bar{V}(0)}{\boldsymbol{\psi}_i} = O_p\left(\frac{1}{m}\right)$$

and is uniformly bounded under the regularity conditions given in Section 2, hence the last term in (A.11) is also an $\text{o}(1/m)$, which renders mse_0 unbiased up to the second order.

Appendix B

B.1 Comparison between REML and AR.YL using the scoring algorithm

The scoring algorithm could sometimes yield zero estimates for the likelihood of the AR.YL. Indeed, for data sets simulated under the model given in Section 5, with $m = 45$ and $\sigma^2_v = 1$, the REML and AR.YL scoring algorithms yielded 28% and 26% zeros respectively. Figures B.1 to B.3 illustrate the why: the likelihoods correspond to a single population generated under the model with $\boldsymbol{\sigma}^2_v = 1$ for which $\hat{\boldsymbol{\sigma}}^2_{v\text{REML}} = 0$.

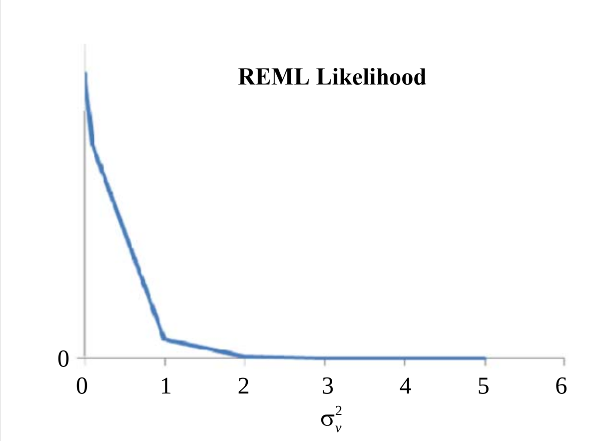

Figure B.1 $L = L_{\text{REML}}\left(\sigma_v^2 \middle| y_1, \ldots, y_{45}\right)$.

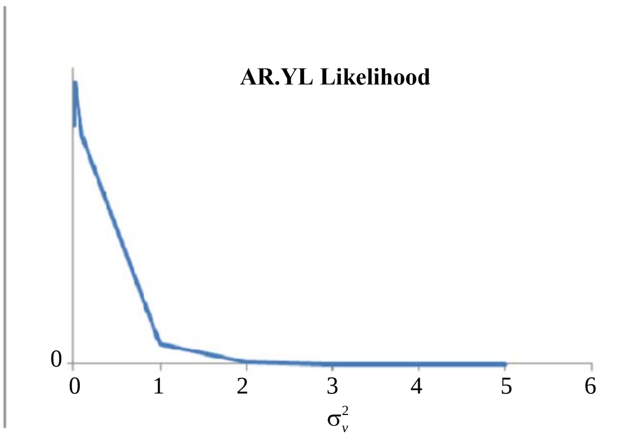

Figure B.2 $L = L_{\text{AR.YL}}\left(\sigma_v^2 \middle| y_1, \ldots, y_{45}\right)$.

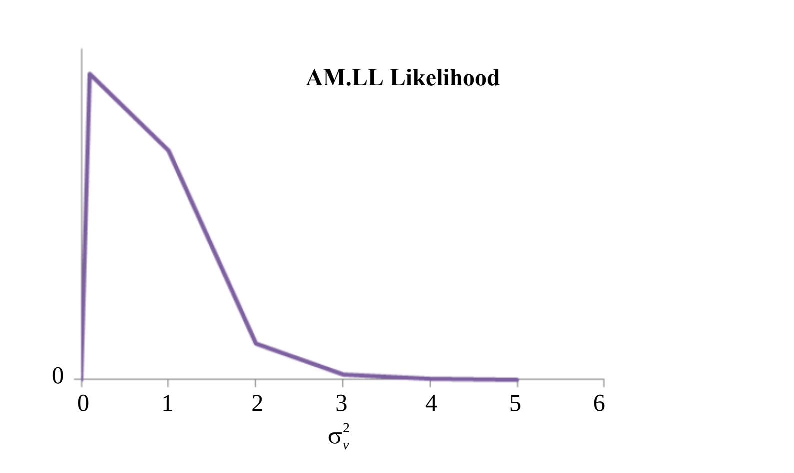

Figure B.3 $L = L_{\text{AM.LL}}\left(\sigma_v^2 \middle| y_1, \ldots, y_{45}\right)$.

Figure B.2 shows that the maximum value of the AR.YL likelihood is very near the border. The scoring algorithm may often miss the maximum and yield a zero value. Figure B.3 shows that the AM.LL likelihood has a maximum value that differentiates better from the border.

B.2 Treatment of zeros in the parametric bootstrap

For each estimate $\hat{\sigma}_v^2 = \hat{\sigma}_v^2\left(\mathbf{y}^{(r)}\right), r = 1, \ldots 10K,$ and each method of variance estimation:

i. Generate a large number B of random area effects $v_i^{(b)} \overset{\text{i.i.d.}}{\sim} N\left(0, \hat{\sigma}_v^2\right), b = 1, \ldots, B,$ and generate, independently of $v_i^{(b)}$, sampling errors $e_i^{(b)} \overset{\text{i.i.d.}}{\sim} N\left(0, \psi_i\right)$, $i = 1, \ldots, m, b = 1, \ldots, B.$ Generate bootstrap data $y_i^{(b)} = \theta_i^{(b)} + e_i^{(b)}, \theta_i^{(b)} = \mathbf{x}_i'\hat{\beta} + v_i^{(b)}, i = 1, \ldots, m.$ If $\hat{\sigma}^2_{v\mathrm{REML}}\left(y^{(r)}\right) = 0,$ then generate $\left(y_i^{(b)}, \theta_i^{(b)}\right),\ b = 1, \ldots, B,$ from the synthetic model (see also Rao and Molina 2015).

ii. Fit the model to the bootstrap data and obtain $\hat{\sigma}_v^{2(b)}$; for the MIX estimator calculate $\hat{\sigma}^{2(b)}_{v\mathrm{MIX}} = \hat{\sigma}^{2(b)}_{v\mathrm{REML}}$ if $\hat{\sigma}_v^{2(b)}$ is positive and $\hat{\sigma}^{2(b)}_{v\mathrm{MIX}} = \hat{\sigma}^{2(b)}_{v\mathrm{AM}}$ otherwise.

iii. Now obtain $\hat{\beta}^{(b)}$, the corresponding EBLUP $\hat{\theta}_i^{(b)}$, the bootstrap components $g_{1i}^{(b)} = g_{1i}\left(\hat{\sigma}_v^2\left(y^{(b)}\right)\right),\ g_{2i}^{(b)} = g_{2i}\left(\hat{\sigma}_v^2\left(y^{(b)}\right)\right)$ and $\bar{g}_{ji}^{\mathrm{PB}} = B^{-1}\sum_b g_{ji}^{(b)},\ j = 1, 2.$

iv. The Naive MSE bootstrap estimator is $\mathrm{mse}_{\mathrm{naive}} = B^{-1}\sum_{b=1}^{B}\left(\hat{\theta}_i^{(b)} - \theta_i^{(b)}\right)^2.$

v. The PB MSE estimator (which is adjusted for bias (Pfeffermann and Glickman 2004) is: $\mathrm{mse}_{\mathrm{PB}}\left(\hat{\theta}_i\right) = g_{1i}\left(\hat{\sigma}_v^2\right) + g_{2i}\left(\hat{\sigma}_v^2\right) - \bar{g}_{1i}^{\mathrm{PB}} - \bar{g}_{2i}^{\mathrm{PB}} + \mathrm{mse}_{\mathrm{naive}}.$

vi. To calculate $\mathrm{ARB}_{\mathbb{C}}$, average $\left(\mathrm{mse}_{\mathrm{PB}}^{(r)}\left(\hat{\theta}_i\right) - \mathrm{MSE}\left(\hat{\theta}_i\right)\right)\big/\mathrm{MSE}\left(\hat{\theta}_i\right)$ over the populations with $(r)\big/\hat{\sigma}^2_{v\mathrm{REML}}\left(\mathbf{y}^{(r)}\right) = 0$ and do similarly with $\mathrm{ARB}_{\mathbb{C}}$ of $\mathrm{mse}_{\mathrm{naive}}$.

References

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.

Chen, S., and Lahiri, P. (2011). On the estimation of Mean Squared Prediction Error in small area estimation. *Calcutta Statistical Association Bulletin*, 63, (Special 7th Triennial Proceedings Volume), Nos. 249-252.

Cressie, N. (1992). REML estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, 18, 1, 75-94.

Das, K., Jiang, J. and Rao, J.N.K. (2004). Mean squared error of empirical predictor. *The Annals of Statistics*, 32, 2, 818-840.

Datta, G., 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.

Estevao, V. (2014). Grid optimization algorithm for maximum likelihood. Internal report, Statistical Research and Innovation Division (SRID), Statistics Canada.

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

Lahiri, P., and Li, H. (2009). Generalized maximum likelihood method in linear mixed models with an application in small area estimation. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, available at http//www.fcsm.gov/events/papers2009.html.

Lahiri, P., and Pramanik, S. (2011). Discussion of “Estimating random effects via adjustment for density maximization” by C. Morris and R. Tang. *Statistical Science*, 26, 2, 291-295.

Li, H., and Lahiri, P. (2011). An adjusted maximum likelihood method for solving small area estimation problems. *Journal of Multivariate Analysis,* 101, 882-892.

Molina, I., Rao, J.N.K. and Datta, G.S. (2015). Small area estimation under a Fay-Herriot model with preliminary testing for the presence of random effects. *Survey Methodology*, 41, 1, 1-19.

Morris, C.N. (2006). Mixed model prediction and small area estimation (with discussions). *Test*, 15, 72-76.

Pfeffermann, D., and Glickman, H. (2004). Mean squarred error approximation in small area estimation by use of parametric and non-parametric bootstrap. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, Alexandria, VA. 4167-78.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation, second edition*. New York: John Wiley & Sons, Inc.

Rubin-Bleuer, S., and Schiopu-Kratina, I. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics,* 33, 6, 2789-2810.

Rubin-Bleuer, S., and You, Y. (2012). A positive variance estimator for the Fay-Herriot small area model. SRID-2012-009E, Statistical Research and Innovation Division (SRID), Statistics Canada.

Rubin-Bleuer, S., Yung, W. and Landry, S. (2010). Adjusted maximum likelihood method for a small area model accounting for time and area effects. SRID-2010-006E, Statistical Research and Innovation Division (SRID), Statistics Canada.

Rubin-Bleuer, S., Yung, W. and Landry, S. (2011). Adjusted maximum likelihood method for a small area model accounting for time and area effects. Long abstract, *Small Area Estimation,* (SAE 20122) in Trier, Germany, International Statistical Institute Satellite Conference.

Rubin-Bleuer, S., Yung, W. and Landry, S. (2012). Variance Component Estimation through the Adjusted Maximum Likelihood Approach. Presentation at the Conference in Honour of the 75th birthday of J.N.K. Rao Carleton University, May 2012, Ottawa.

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.

Yuan, P. (2009). Comparison of SAE methods of variance estimation. Internal document, Statistical Research and Innovation Division (SRID), Statistics Canada.

Yuan, K.H., and Jennrich, R. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis,* 65, 2, 245-260.