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

Survey Methodology

Mean squared prediction error estimators of the empirical best linear unbiased predictor of a small area mean under a semi-parametric Fay-Herriot model

by Shijie Chen, Partha Lahiri and J.N.K. Rao

Release date: June 30, 2025

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 infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. 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.

Survey Methodology, June 2025
Vol. 51, No. 1, pp. 233-250
Statistics Canada, Catalogue No. 12-001-X

Mean squared prediction error estimators of the empirical best linear unbiased predictor of a small area mean under a semi-parametric Fay-Herriot model

Shijie Chen, Partha Lahiri and J.N.K. Rao[1]

Abstract

In this paper, we derive a second-order unbiased (or nearly unbiased) mean squared prediction error (MSPE) estimator of the empirical best linear unbiased predictor (EBLUP) of a small area mean for a semi-parametric extension to the well-known Fay-Herriot model. Specifically, we derive our MSPE estimator essentially assuming certain moment conditions on both the sampling errors and random effects distributions. The normality-based Prasad-Rao MSPE estimator has a surprising robustness property in that it remains second-order unbiased under the non-normality of random effects when a simple Prasad-Rao method-of-moments estimator is used for the variance component and the sampling error distribution is normal. We show that the normality-based MSPE estimator is no longer second-order unbiased when the sampling error distribution has non-zero kurtosis or when the Fay-Herriot moment method is used to estimate the variance component, even when the sampling error distribution is normal. Interestingly, when the simple method-of moments estimator is used for the variance component, our proposed MSPE estimator does not require the estimation of kurtosis of the random effects. Results of a simulation study on the accuracy of the proposed MSPE estimator, under non-normality of both sampling and random effects distributions, are also presented.

Key Words: Linear mixed model; Mean squared prediction errors; Variance components.

1. Introduction

In the context of estimating per-capita income for small places (population less than 1,000) from the 1970 Census of Population and Housing, Fay and Herriot (1979) used a small area regression model and demonstrated that the resulting empirical best linear unbiased predictors (EBLUP) have smaller average error than either the traditional survey estimators or an alternative method using county averages. Let Y_i be a direct estimator of the ith small area mean θ_i and $\mathbf{x}_i = (x_{i1},\ldots,x_{ip})'$ be a $p \times 1$ vector of associated predictor variables, $i = 1,\ldots,m.$ The Fay-Herriot model may be written as $Y_i = \theta_i + e_i$ and $\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i,$ or as a linear mixed model: $Y_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i + e_i,$ where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are independent with $e_i \overset{\text{ind}}{\sim} N(0, D_i)$ and $v_i \overset{\text{ind}}{\sim} N(0,\psi), i = 1,\ldots,m.$ The sampling variances D_i are assumed to be known, but $\boldsymbol{\beta}$ and ψ are to be estimated from the data $(Y_i, \mathbf{x}_i), i = 1,\ldots,m.$ In practice, D_i's are externally estimated using the generalized variance function (GVF) method; see Wolter (1985), Fay and Herriot (1979), Otto and Bell (1995), among others. The Fay-Herriot (FH) model has been used extensively in small area estimation and related problems for a variety of reasons, including its simplicity, its ability to protect confidentiality of microdata and its ability to produce design-consistent estimators [see Rao and Molina (2015), Chapter 7].

1. Shijie Chen, Vice President, Enanta Pharmaceuticals. E-mail: scip6153@gmail.com; Partha Lahiri, Professor, Joint Program in Survey Methodology & Department of Mathematics, University of Maryland, College Park MD 20742, U.S.A. E-mail: plahiri@umd.edu; J.N.K. Rao, Distinguished Research Professor, Carleton University, Ottawa, Ontario, Canada, K1S 5B6. E-mail: jrao@math.carleton.ca.

While the empirical best linear unbiased predictor (EBLUP) of a small area total or mean under the FH model is easy to obtain, an accurate estimation of its mean square prediction error (MSPE) is a challenging problem. A naive MSPE estimator is given by the MSPE of the best linear unbiased predictor (BLUP) with the model variance ψ replaced by a suitable estimator. But, it usually underestimates the true MSPE of EBLUP mainly for two reasons. First, it fails to incorporate the extra variability incurred due to the estimation of ψ and the order of this underestimation is $O(m^{-1})$, for large m. Secondly, the naive MSPE estimator underestimates even the true MSPE of the BLUP, the order of the underestimation being $O(m^{-1})$. Prasad and Rao (1990) demonstrated the importance of accounting for these two sources of underestimation, and using a Taylor linearization method produced a second-order unbiased (or nearly unbiased) MSPE estimator of EBLUP when the variance component is estimated by a simple method-of-moments. The bias of that MSPE estimator is of order $o(m^{-1})$. The derivation of their MSPE estimator involves essentially two main steps. First a second-order correct MSPE expansion is obtained by neglecting all terms of order $o(m^{-1})$. The second step involves the estimation of this second-order correct MSPE approximation such that bias is of lower order, i.e. $o(m^{-1})$. Datta and Lahiri (2000) extended the Prasad-Rao MSPE estimation method to maximum likelihood (ML) and restricted ML (REML) estimators of the model parameters. Datta, Rao and Smith (2005) obtained a nearly unbiased MSPE estimator when ψ is estimated by the Fay and Herriot (1979) method-of-moments. Das, Jiang and Rao (2004) generalized the Taylor method to general linear mixed models and obtained nearly unbiased MSPE estimators.

Note that the derivation of EBLUP, using a moment estimator of ψ, does not require normality assumptions. However, for the estimation of MSPE, Prasad and Rao (1990), Datta and Lahiri (2000), Das et al. (2004), Datta et al. (2005), and others used the normality assumption. One exception is the paper by Lahiri and Rao (1995) who assumed normality of the sampling errors $\{e_i\}$, but replaced the normality of the random effects $\{v_i\}$ by certain moment conditions. They showed that the normality-based Prasad-Rao estimator of the MSPE of EBLUP remains second-order unbiased under this non-normal set-up when the simple method-of-moments estimator of ψ is used. This is indeed a surprising result, demonstrating the robustness of the Prasad-Rao MSPE estimator under unspecified non-normality of the random effects $\{v_i\}$. Does this result hold when the sampling errors $\{e_i\}$ are non-normal or when ψ is estimated by the moment estimator of ψ proposed by Fay and Herriot (1979)? The case of non-normal sampling errors may be useful when the sample sizes are small for several areas. In the latter case, the effect of the central limit theorem may not hold.

In Section 2, we introduce a semi-parametric Fay-Herriot model that assumes certain moment conditions and briefly review the empirical best linear unbiased prediction (EBLUP). In Section 3, we obtain a second-order approximation to the MSPE of the EBLUP without the normality assumption. In Section 4, we propose a second-order (or nearly) unbiased MSPE estimator of EBLUP, using the approximation to the MSPE obtained in Section 3. Finally, some empirical results from a small simulation study are reported in Section 5. Proofs of our results are sketched in the Appendix.

2. A semi-parametric Fay-Herriot model and EBLUP

We introduce a semi-parametric Fay-Herriot model as $Y_i = \theta_i + e_i$ and $\theta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i, i = 1,\ldots,m,$ or as a linear mixed model,

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i + e_i, \quad i = 1,\ldots,m, \tag{2.1}$$

where the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$ are uncorrelated with $e_i \sim [0, D_i, \kappa_{ei}]$ and $v_i \sim [0, \psi, \kappa_v]$, $[\mu, \sigma^2, \kappa]$ representing a probability distribution with mean μ, variance σ^2 and kurtosis κ. We define kurtosis of a distribution as $\kappa = \mu_4 / \sigma^4 - 3$, where σ^2 and μ_4 are the variance and the fourth central moment of the distribution, respectively. We assume that $\boldsymbol{\beta}$, ψ, and κ_v are unknown. In addition to the conventional assumption of the normality-based Fay-Herriot model for the sampling variance, which posits that the sampling variance D_i is known, we also assume that the sampling kurtosis κ_{ei} is known.

Define $\mathbf{X}' = (x_1,\ldots,x_m)$, and $\boldsymbol{\Sigma}(\psi) = \text{diag}\{\psi + D_j;\ j = 1,...,m\}$ and $\boldsymbol{\beta}$ can be estimated by $\hat{\boldsymbol{\beta}}(\psi) = [\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\psi)\mathbf{X}]^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\psi)\mathbf{Y}$, a weighted least squares estimator of $\boldsymbol{\beta}$ for a given ψ. The BLUP of θ_i under the FH model (2.1) is given by:

$$\hat{\theta}_i(Y_i;\psi) = B_i Y_i + (1 - B_i)\,\mathbf{x}_i'\hat{\boldsymbol{\beta}}(\psi),$$

where $B_i = \frac{\psi}{D_i + \psi}, i = 1,\ldots,m.$ An EBLUP of θ_i is then obtained as

$$\hat{\theta}_i(Y_i;\hat{\psi}) = \hat{B}_i Y_i + (1 - \hat{B}_i)\mathbf{x}_i'\hat{\boldsymbol{\beta}}(\hat{\psi}) =: \hat{\theta}_i,$$

where $\hat{B}_i = \frac{\hat{\psi}}{D_i + \hat{\psi}}, i = 1,\ldots,m,$ and $\hat{\psi}$ is a moment estimator of ψ. Note that the BLUP and EBLUP, based on $\hat{\psi}$, do not require normality of $\{e_i\}$ and $\{v_i\}$.

Prasad and Rao (1990) proposed the following simple method-of-moments estimator of ψ:

$$\hat{\psi}_{\text{PR}} = \max\left\{0, (m-p)^{-1}\sum_{j=1}^{m}\left\{\left(Y_j - \mathbf{x}_j'\hat{\boldsymbol{\beta}}_{\text{OLS}}\right)^2 - \left(1 - h_{jj}\right) D_j\right\}\right\},$$

where $\hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$ the ordinary least squares estimator of $\boldsymbol{\beta}$, and $h_{jj} = \mathbf{x}_j'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j,$ the leverage for the jth small area $(j = 1,\ldots,m)$. Fay and Herriot (1979) obtained a different moment estimator $\hat{\psi}_{\text{FH}}$, by solving the following equation iteratively for ψ:

$$A(\psi) = \frac{1}{m-p}\mathbf{Y}'\mathbf{Q}(\psi)\mathbf{Y} - 1 = 0, \tag{2.2}$$

where

$$\mathbf{Q}(\psi) = \boldsymbol{\Sigma}^{-1}(\psi) - \boldsymbol{\Sigma}^{-1}(\psi)\mathbf{X}\{\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\psi)\mathbf{X}\}^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\psi),$$

and $\mathbf{Y}'\mathbf{Q}(\psi)\,\mathbf{Y} = \sum_{j=1}^{m}(D_j + \psi)^{-1}\{Y_j - \mathbf{x}_j'\hat{\boldsymbol{\beta}}(\psi)\}^2$ is the weighted residual sum of squares. Pfeffermann and Nathan (1981) proposed a similar moment method in the context of regression analysis of survey data. For the special case $D_i = D\,(i = 1,\ldots,m),$ $\hat{\psi}_{\text{PR}} = \hat{\psi}_{\text{FH}}.$ In this paper, we focus on EBLUP based on $\hat{\psi}_{\text{PR}}$ or $\hat{\psi}_{\text{FH}}$.

The estimators $\hat{\psi}_{\mathrm{PR}}$ and $\hat{\psi}_{\mathrm{FH}}$ are typically consistent estimators for large m under the following regularity conditions:

(r.1) $0 < D_L \le D_j \le D_U < \infty$, $j = 1,\ldots,m$,

(r.2) $\sup_{j\ge 1} h_{jj} = O(\frac{1}{m})$;

i.e., $\hat{\psi} - \psi = O_p(m^{-1/2})$ for $\hat{\psi} = \hat{\psi}_{\mathrm{PR}}$ and $\hat{\psi}_{\mathrm{FH}}$.

Under non-normality and regularity conditions, the bias of $\hat{\psi}_{\mathrm{PR}}$ is of order $o(m^{-1})$. However, unless $D_i = D\ (i = 1,\ldots,m)$, the bias of $\hat{\psi}_{\mathrm{FH}}$ is of order $O(m^{-1})$, even under normality, and it is given by

$$E[\hat{\psi}_{\mathrm{FH}} - \psi] \;=\; b(\hat{\psi}_{\mathrm{FH}}; \psi, \kappa_v) + o(m^{-1}), \tag{2.3}$$

where

$$b(\hat{\psi}_{\mathrm{FH}}; \psi, \kappa_v) \;= b_N(\hat{\psi}_{\mathrm{FH}}; \psi) + \alpha(\hat{\psi}_{\mathrm{FH}}; \psi, \kappa_v)$$

$$b_N(\hat{\psi}_{\mathrm{FH}}; \psi) \;= \frac{2[m\mathrm{tr}(\boldsymbol{\Sigma}^{-2}) - \{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})\}^2]}{[\mathrm{tr}(\boldsymbol{\Sigma}^{-1})]^3}$$

$$\alpha(\hat{\psi}_{\mathrm{FH}}; \psi, \kappa_v) \;= \frac{[\mathrm{tr}(\boldsymbol{\Sigma}^{-2})]^2 - \mathrm{tr}(\boldsymbol{\Sigma}^{-3})\mathrm{tr}(\boldsymbol{\Sigma}^{-1})}{[\mathrm{tr}(\boldsymbol{\Sigma}^{-1})]^3}\psi^2\kappa_v + \frac{\mathrm{tr}(\mathbf{D}^2\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-2})\ \mathrm{tr}(\boldsymbol{\Sigma}^{-2}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1})\ \mathrm{tr}(\mathbf{D}^2\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-3})}{[\mathrm{tr}(\boldsymbol{\Sigma}^{-1})]^3},$$

$$\mathbf{D} = \mathrm{diag}\{D_j;\ j = 1,\ldots,m\},$$

$$\boldsymbol{\Phi} = \mathrm{diag}\{\kappa_{ej};\ j = 1,\ldots,m\}.$$

The proof of (2.3) is given in the Appendix. In the above, $b_N(\hat{\psi}_{\mathrm{FH}}; \psi)$ denotes the bias of $\hat{\psi}_{\mathrm{FH}}$ up to order $O(m^{-1})$ under normality [see Datta et al. (2005)] and $\alpha(\hat{\psi}_{\mathrm{FH}}; \psi, \kappa_v)$ is the additional non-normality effect on the bias.

The variance of $\hat{\psi}$ may be expressed as

$$\mathrm{var}(\hat{\psi}) \;=\; \mathrm{var}_N(\hat{\psi}) \;+\; \eta(\hat{\psi}; \psi, \kappa_v), \tag{2.4}$$

where $\mathrm{var}_N(\hat{\psi})$ denotes the variance of $\hat{\psi}$ up to order $O(m^{-1})$ under normality, and $\eta(\hat{\psi}, \psi, \kappa_v)$ denotes the additional non-normality effect on the variance. From Datta et al. (2005), we have

$$\mathrm{var}_N(\hat{\psi}_{\mathrm{PR}}) \;=\; 2m^{-2}\sum_{j=1}^{m}(\psi + D_j)^2 \;=\; 2m^{-2}\mathrm{tr}(\boldsymbol{\Sigma}^2),$$

$$\mathrm{var}_N(\hat{\psi}_{\mathrm{FH}}) \;=\; 2m\left\{\sum_{j=1}^{m}(\psi + D_j)^{-1}\right\}^{-2} \;=\; 2m\left\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})\right\}^{-2}.$$

It can be shown that

$$\eta(\hat{\psi}_{\mathrm{PR}};\psi,\kappa_v) = m^{-1}\left\{\kappa_v\psi^2 + \frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right\} = m^{-1}\left\{\kappa_v\psi^2 + m^{-1}\mathrm{tr}(\mathbf{D}^2\mathbf{\Phi})\right\}$$

$$\eta(\hat{\psi}_{\mathrm{FH}};\psi,\kappa_v) = \left\{\sum_j(\psi+D_j)^{-1}\right\}^{-2}\sum_{j=1}^{m}\left\{(\psi+D_j)^{-2}[\kappa_v\psi^2+\kappa_{ej}D_j^2]\right\}$$
$$= \left[\mathrm{tr}(\mathbf{\Sigma}^{-1})\right]^{-2}\left\{\mathrm{tr}(\mathbf{\Sigma}^{-2})\kappa_v\psi^2 + \mathrm{tr}(\mathbf{D}^2\mathbf{\Phi}\mathbf{\Sigma}^{-2})\right\}.$$

See Appendix for a proof of the asymptotic variance of $\hat{\psi}$. When both $\{e_i\}$ and $\{v_i\}$ are normal, we have $\mathrm{var}(\hat{\psi}_{\mathrm{PR}}) = \mathrm{var}_N(\hat{\psi}_{\mathrm{PR}})$, $\mathrm{var}(\hat{\psi}_{\mathrm{FH}}) = \mathrm{var}_N(\hat{\psi}_{\mathrm{FH}})$ and $\mathrm{var}_N(\hat{\psi}_{\mathrm{FH}}) \le \mathrm{var}_N(\hat{\psi}_{\mathrm{PR}})$ with equality for the balanced case $D_i = D, (i=1,\ldots,m)$, see Datta et al. (2005). It is interesting to note that the latter result does not extend to the situation with non-zero kurtosis. For the balanced case $D_i = D\,(i=1,\ldots,m)$, we have $b(\hat{\psi}_{\mathrm{PR}};\psi,\kappa_v) = b(\hat{\psi}_{\mathrm{FH}};\psi,\kappa_v) = 0$ and $\mathrm{var}(\hat{\psi}_{\mathrm{PR}}) = \mathrm{var}(\hat{\psi}_{\mathrm{FH}}) = 2m^{-1}(\psi+D)^2 + m^{-1}\{\kappa_v\psi^2 + m^{-1}D^2\mathrm{tr}(\mathbf{\Phi})\}$ simply because in this situation $\hat{\psi}_{\mathrm{PR}} = \hat{\psi}_{\mathrm{FH}}$.

3. Approximation to MSPE

The MSPE of the EBLUP $\hat{\theta}_i$ is given by $\mathrm{MSPE}(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2$, where the expectation is taken over the marginal distribution of $\mathbf{Y}$ under the semi-parametric Fay-Herriot model (2.1). The MSPE of the BLUP of $\hat{\theta}_i(Y_i,\psi)$ is not affected by non-normality and it is given by

$$\mathrm{MSPE}[\hat{\theta}_i(Y_i,\psi)] = g_{1i}(\psi) + g_{2i}(\psi),$$

where

$$g_{1i}(\psi) = \frac{\psi D_i}{\psi + D_i},$$

$$g_{2i}(\psi) = \frac{D_i^2}{(\psi+D_i)^2}\mathrm{var}[\mathbf{x}_i'\hat{\mathbf{\beta}}(\psi)] = \frac{D_i^2}{(\psi+D_i)^2}\,\mathbf{x}_i'\left[\mathbf{X}'\mathbf{\Sigma}^{-1}(\psi)\,\mathbf{X}\right]^{-1}\mathbf{x}_i.$$

A naive estimator of $\mathrm{MSPE}[\hat{\theta}_i(Y_i,\psi)]$ that does not incorporate variability incurred due to estimation of ψ is given by

$$\mathrm{mspe}_i^{\mathrm{naive}} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}).$$

We are interested in approximating the MSPE of the EBLUP under the proposed semi-parametric model (2.1) that accounts for the estimation of ψ and is second-order accurate, i.e., accurate up to order $O(m^{-1})$.

We decompose the MSPE of the EBLUP $\hat{\theta}_i$ as

$$\begin{aligned}\mathrm{MSPE}[\hat{\theta}_i(Y_i,\hat{\psi})] &= \mathrm{MSPE}[\hat{\theta}_i(Y_i,\psi)] + E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)]^2 \\ &\quad + 2E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi) - \theta_i].\end{aligned} \tag{3.1}$$

where $\hat{\theta}_i(Y_i,\hat{\psi}) = \hat{\theta}_i$ and $\hat{\theta}_i(Y_i,\psi)$ is the BLUP. The cross-product term in (3.1) is zero under normality of $\{v_i\}$ and $\{e_i\}$; see Kackar and Harville (1984), but it is of $O(m^{-1})$ under the proposed semi-parametric FH model (2.1) and hence not negligible. Result 1, given below, provides the following approximations to the last two terms of (3.1). A proof of Result 1 is given in the Appendix.

Result 1: Under the proposed semi-parametric FH model (2.1) and regularity conditions (r.1), (r.2) and (r.3): $\sup_{j\geq 1} E\,|\,v_j\,|^{8+\delta} < \infty,\ 0 < \delta < 1,$ we have

$$(i)\quad E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)]^2 = g_{3i}(\psi,\kappa_v) + o(m^{-1}),$$

$$(ii)\quad E[\hat{\theta}_i(Y_i,\hat{\psi}) - \hat{\theta}_i(Y_i,\psi)][\hat{\theta}_i(Y_i,\psi) - \theta_i] = g_{4i}(\psi,\kappa_v) + o(m^{-1}),$$

where

$$g_{3i}(\psi,\kappa_v) = \frac{D_i^2}{(\psi + D_i)^3}\,\mathrm{var}(\hat{\psi}),$$

$$g_{4i}(\psi,\kappa_v) = \frac{\psi D_i^2}{m(\psi + D_i)^3}\left[D_i\kappa_{ei} - \psi\kappa_v\right] c(\hat{\psi};\psi),$$

$c(\hat{\psi}_{\mathrm{PR}};\psi) = 1$ and $c(\hat{\psi}_{\mathrm{FH}};\psi) = m(\psi + D_i)^{-1}\left\{\sum_j (\psi + D_j)^{-1}\right\}^{-1}$.

Thus, a second-order expansion to MSPE of the EBLUP $\hat{\theta}_i$ is given by

$$\begin{aligned}\mathrm{AMSPE}_i &= g_{1i}(\psi) + g_{2i}(\psi) + g_{3i}(\psi,\kappa_v) + 2g_{4i}(\psi,\kappa_v)\\ &= \frac{\psi D_i}{\psi + D_i} + \frac{D_i^2}{(\psi + D_i)^2}\,\mathrm{var}[\mathbf{x}_i'\hat{\boldsymbol{\beta}}(\psi)] + \frac{D_i^2}{(\psi + D_i)^3}\,\mathrm{var}(\hat{\psi}) + \frac{2\psi D_i^2}{m(\psi + D_i)^3}\left[D_i\kappa_{ei} - \psi\kappa_v\right] c(\hat{\psi};\psi) \quad (3.2)\\ &= \mathrm{AMSPE}_{i,N} + \frac{D_i^2}{(\psi + D_i)^3}\,\eta(\hat{\psi};\psi,\kappa_v) + 2g_{4i}(\psi,\kappa_v),\end{aligned}$$

where $\mathrm{AMSPE}_{i,N}$ is the normality-based MSPE approximation as given in Prasad and Rao (1990) and Datta et al. (2005). The term $g_{3i}(\psi,\kappa_v)$ is the additional uncertainty due to the estimation of the variance component ψ and the term $2g_{4i}(\psi,\kappa_v)$ is needed to adjust for non-zero kurtosis. Under the regularity conditions, $g_{1i}(\psi)$ is the leading term [of order $O(1)$] and the remaining terms are all of order $O(m^{-1})$. Note that non-normality affects both $\mathrm{var}(\hat{\psi})$ and the cross-product term $2E[\hat{\theta}_i(\hat{\psi},\mathbf{Y}) - \hat{\theta}_i(\psi,\mathbf{Y})]\,[\hat{\theta}_i(\psi,\mathbf{Y}) - \theta_i]$. When both $\{e_i\}$ and $\{v_i\}$ are normal, the above approximation reduces to the Prasad and Rao (1990) approximation when $\hat{\psi} = \hat{\psi}_{\mathrm{PR}}$ and the Datta et al. (2005) approximation when $\hat{\psi} = \hat{\psi}_{\mathrm{FH}}$. When the $\{e_i\}$ are normal and $\hat{\psi} = \hat{\psi}_{\mathrm{PR}}$, the MSPE approximation (3.2) reduces to the Lahiri and Rao (1995) approximation.

4. Nearly unbiased estimator of MSPE

The second-order MSPE approximation AMSPE_i, given by (3.2), involves unknown parameters ψ and κ_v. Let $\hat{\kappa}_v$ be a consistent estimator of κ_v. Then $g_{2i}(\hat{\psi})$, $g_{3i}(\hat{\psi},\hat{\kappa}_v)$, and $g_{4i}(\hat{\psi},\hat{\kappa}_v)$ are second-order unbiased (or nearly unbiased) estimators of $g_{2i}(\psi)$, $g_{3i}(\psi,\kappa_v)$, and $g_{4i}(\psi,\kappa_v)$, respectively, since latter functions of ψ and κ_v are already of order $O(m^{-1})$. However, estimation of the the leading term $g_{1i}(\psi)$ in (3.2) needs special attention since it is of the order $O(1)$.

Under regularity conditions (r.1) and (r.2), it can be shown that

$$E[g_{1i}(\hat{\psi})] = g_{1i}(\psi) - g_{3i}(\psi,\kappa_v) + g_{5i}(\psi,\kappa_v) + o(m^{-1}), \tag{4.1}$$

where

$$g_{5i}(\psi,\kappa_v) = \frac{D_i^2}{(\psi+D_i)^2} b(\hat{\psi};\psi,\kappa_v).$$

Using (3.2) and (4.1), a second-order unbiased (or nearly unbiased) MSPE estimator is then given by

$$\text{mspe}_i = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + 2g_{3i}(\hat{\psi},\hat{\kappa}_v) + 2g_{4i}(\hat{\psi},\hat{\kappa}_v) - g_{5i}(\hat{\psi},\hat{\kappa}_v). \tag{4.2}$$

When $\hat{\psi}=\hat{\psi}_{\text{PR}}$, we have

$$\begin{aligned}\text{mspe}_i^{\text{PR}} &= \text{mspe}_{i,N}^{\text{PR}} + \frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei} + \frac{1}{m}\sum_{j=1}^{m}\kappa_{ej}D_j^2\right] \\ &= \text{mspe}_{i,N}^{\text{PR}} + \frac{2D_i^2}{m(\hat{\psi}+D_i)^3}\left[\hat{\psi}D_i\kappa_{ei} + m^{-1}\text{tr}(\mathbf{D}^2\boldsymbol{\Phi})\right],\end{aligned} \tag{4.3}$$

where

$$\text{mspe}_{i,N}^{\text{PR}} = g_{1i}(\hat{\psi}) + g_{2i}(\hat{\psi}) + \frac{2D_i^2}{m^2(\hat{\psi}+D_i)^3}\sum_{j=1}^{m}(\hat{\psi}+D_j)^2$$

is the normality-based MSPE estimator first proposed by Prasad and Rao (1990). It is interesting to note that the MSPE estimator (4.3) does not require the estimation of κ_v although κ_v is involved in the MSPE approximation (3.2). For normal $\{e_i\}$ but unspecified non-normal $\{v_i\}$, $\text{mspe}_i^{\text{PR}} = \text{mspe}_{i,N}^{\text{PR}}$ (Lahiri and Rao, 1995). It is interesting to note that the robustness of the Prasad-Rao MSPE estimator does not extend to the case when $\{e_i\}$ have non-zero kurtosis. When $\kappa_{ej} > 0$, $j=1,\ldots,m$, $\text{mspe}_i^{\text{PR}}$ overestimates whenever $\text{mspe}_{i,N}^{\text{PR}}$ overestimates.

When $\hat{\psi}=\hat{\psi}_{\text{FH}}$, we have

$$\begin{aligned}\text{mspe}_i^{\text{FH}} &= \text{mspe}_i^{\text{DRS}} + \frac{2D_i^2}{(\hat{\psi}_{\text{FH}}+D_i)^3}\eta(\hat{\psi}_{\text{FH}};\hat{\psi}_{\text{FH}},\hat{\kappa}_v) \\ &\quad + 2g_{4i}(\hat{\psi}_{\text{FH}},\hat{\kappa}_v) - \frac{D_i^2}{(\hat{\psi}_{\text{FH}}+D_i)^2}\alpha(\hat{\psi}_{\text{FH}};\hat{\psi}_{\text{FH}},\hat{\kappa}_v)\end{aligned} \tag{4.4}$$

where

$$\text{mspe}_i^{\text{DRS}} = g_{1i}(\hat{\psi}_{\text{FH}}) + g_{2i}(\hat{\psi}_{\text{FH}}) + \frac{2D_i^2}{(\hat{\psi}_{\text{FH}} + D_i)^3}\,\text{evar}_N(\hat{\psi}_{\text{FH}}) - \frac{D_i^2}{(\hat{\psi}_{\text{FH}} + D_i)^2}\,b_N(\hat{\psi}_{\text{FH}}; \hat{\psi}_{\text{FH}})$$

is the normality-based estimator proposed by Datta et al. (2005). In the above $\text{evar}_N(\hat{\psi}_{\text{FH}})$ is obtained from $\text{var}_N(\hat{\psi}_{\text{FH}})$ using $\hat{\psi}_{\text{FH}}$ in place of ψ. When both $\{e_i\}$ and $\{v_i\}$ are normal, $\text{mspe}_i^{\text{FH}}$, given by (4.4), reduces to $\text{mspe}_i^{\text{DRS}}$. However, when $\{e_i\}$ are normal but not $\{v_i\}$, $\text{mspe}_i^{\text{FH}}$ is not identical to $\text{mspe}_i^{\text{DRS}}$ unless $D_i = D\,(i = 1,\ldots,m)$. Hence $\text{mspe}_i^{\text{DRS}}$, unlike $\text{mspe}_i^{\text{PR}}$, is not robust under non-normality of $\{v_i\}$, even when $\{e_i\}$ are normal. It is easy to check that for the balanced case $D_i = D\,(i = 1,\ldots,m)$, $\text{mspe}_i^{\text{FH}} = \text{mspe}_i^{\text{PR}}$.

To obtain a consistent estimator of κ_v, we replace ψ and $\text{var}(\hat{\psi}_{\text{FH}})$ in (2.4) by $\hat{\psi}_{\text{FH}}$ and

$$v_{\text{WJ}} = \sum_{u=1}^{m} w_u (\hat{\psi}_{\text{FH},(-u)} - \hat{\psi}_{\text{FH}})^2,$$

a weighted jackknife estimator of $\text{var}(\hat{\psi}_{\text{FH}})$ considered by Chen and Lahiri (2008), where $w_u = 1 - h_{uu}$ and $\hat{\psi}_{\text{FH},(-u)}$ is the Fay-Herriot estimator of ψ, using all but the u th small area data. The resulting equation is given by:

$$k(\hat{\psi}_{\text{FH}}) + l(\hat{\psi}_{\text{FH}})\,\kappa_v = 0, \tag{4.5}$$

where

$$k(\psi) = 2m + \text{tr}(\mathbf{D}^2\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-2}) - \{\text{tr}(\boldsymbol{\Sigma}^{-1})\}^2 v_{\text{WJ}}$$
$$l(\psi) = \text{tr}(\boldsymbol{\Sigma}^{-2})\,\psi^2.$$

Solving (4.5) for κ_v, we obtain a closed-form estimator of κ_v: $\hat{\kappa}_v = -k(\hat{\psi}_{\text{FH}})/l(\hat{\psi}_{\text{FH}})$ if $\hat{\psi}_{\text{FH}} > 0$ and 0 otherwise. Since this is a smooth function and $\hat{\psi}_{\text{FH}}$ and v_{WJ} are consistent for ψ and $\text{var}(\hat{\psi}_{\text{FH}})$ respectively, we have consistency of the estimator $\hat{\kappa}_v$.

5. Simulation study

Finite-sample accuracy of the proposed robust estimator of the MSPE of the EBLUP is investigated in this section through a Monte Carlo simulation study, for the special case $\mathbf{x}_i'\boldsymbol{\beta} = \mu$ and $\kappa_{ei} = \kappa_e\,(i = 1,\ldots,m)$. Noting that the MSPE is translation invariant (i.e., it remains the same when Y_i is changed to $Y_i - \mu$), 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 still estimate the zero mean. We have considered different values of m, the number of small areas. But, to save space, we report the simulation results only for $m =$ 60 and summarize results for other values of m. Further, we consider nine combinations of $\kappa_v = 0$, 3, 6 and $\kappa_e =$ 0, 3, 6. Note that in our simulation $\kappa = 0$, 3, and 6 correspond to the normal, double exponential, and *shifted* exponential with mean zero distributions, respectively. We set $\psi = 1$. Regarding the known sampling

variances, we consider both a balanced case, i.e. $D_j = D = \psi$ $(j = 1,\ldots,m)$ and an unbalanced case that corresponds to the Type II pattern of Datta et al. (2005). Thus, for the unbalanced case, there are 5 groups $G_t, (t = 1,\ldots,5)$ of equal number of small areas such that within each group sampling variances are the same. Specifically, $D_j = 2.0$ $j \in G_1; D_j = 0.6$ $j \in G_2; D_j = 0.5$ $j \in G_3; D_j = 0.4$ $j \in G_4; D_j = 0.2$ $j \in G_5$.

For both balanced and unbalanced cases, we compared the performance of our proposed robust MSPE estimator of the EBLUP based on the Prasad-Rao simple method-of-moments estimator of ψ with the naive and the normality-based Prasad-Rao (PR) MSPE estimator. In addition, for the unbalanced case, we compared the performance of our proposed robust MSPE estimator of the EBLUP based on the Fay-Herriot method-of-moments estimator with the naive and the normality-based Datta-Rao-Smith (DRS) MSPE estimators. Note that in the balanced case $\hat{\psi}_{\text{PR}} = \hat{\psi}_{\text{FH}}$ and in this case our robust method does not involve estimation of the unknown kurtosis κ_v. For the unbalanced case, estimation of κ_v is not needed for the Prasad-Rao method of estimating ψ, but it is needed for the Fay-Herriot method of estimating ψ.

For each case with specified parameters, we generated $R =$ 10,000 independent set of variates $\{v_i, e_i, i = 1,\ldots,m\}$. Simulated values of the relative bias (RB) and of the relative root mean squared error (RRMSE) of estimators of mean squared prediction error of the EBLUP were then computed for all the m areas as follows:

$$\text{RB}_i = 100\ \frac{E(\text{mspe}_i) - \text{MSPE}_i}{\text{MSPE}_i},$$

$$\text{RRMSE}_i = 100\ \frac{\sqrt{E(\text{mspe}_i - \text{MSPE}_i)^2}}{\text{MSPE}_i}$$

where mspe_i denotes an estimator of MSPE_i, the MSPE of the EBLUP $\hat{\theta}_i(y_i; \hat{A})$ of the true small-area mean θ_i $(i = 1,\ldots,m)$, and the expectation "E" is approximated by the Monte Carlo method. The RB and RRMSE, averaged over areas with the same sampling variances, were then reported in Tables 5.1-5.4.

Table 5.1 reports the percent RB of MSPE estimators for the balanced case. For each MSPE estimator, the absolute RB decreases as the number of small areas, m, increases. For all the cases, the naive estimator leads to underestimation, ranging from about 7% to 23% for $m = 30$ and about 3% to 14% for $m = 60$. Results for the Prasad-Rao MSPE estimator and the proposed MSPE estimator are almost identical and RB is negligible whenever the sampling errors $\{e_i\}$ are normally distributed; this is consistent with previous theory (Lahiri and Rao, 1995). When the $\{e_i\}$ are non-normal, the normality-based Prasad-Rao MSPE estimator leads to underestimation, some times as large as about 8% for $m =$ 60 and about 10% for $m =$ 30. On the other hand, the proposed MSPE estimator corrects for the underestimation in all cases. We note that for $m =$ 30, our proposed robust MSPE estimator could lead to overestimation; but the overestimation decreases dramatically as m increases to 60. In contrast, the underestimation for the Prasad-Rao MSPE estimation decreases very slowly. To highlight this point, consider the case where both $\{e_i\}$ and $\{v_i\}$ follow

the location shifted exponential distribution. In this case, as we increase m from 30 to 60, underestimation for the Prasad-Rao MSPE estimator decreases slowly from about 10% to 8%; in contrast, overestimation for the robust MSPE estimator decreases dramatically from about 15% to 4%.

Table 5.1
Simulated values of percent relative bias (RB) of mean squared prediction error estimators for the balanced case: $\psi = D = 1$ and ψ is estimated by the Prasad-Rao (same as the Fay-Herriot) method-of-moments

		Distribution of v					
		Normal		Double Exponential		Shifted Exponential	
Distribution of e		$m = 30$	$m = 60$	$m = 30$	$m = 60$	$m = 30$	$m = 60$
Normal	Naive	-12.10	-6.64	-6.66	-2.92	-12.29	-6.34
	Prasad-Rao	0.86	-0.11	0.53	0.54	1.61	0.54
	Proposed	0.86	-0.11	0.53	0.54	1.61	0.54
Double Exponential	Naive	-17.90	-10.67	-9.60	-4.96	-17.59	-10.82
	Prasad-Rao	-5.35	-4.22	-2.59	-1.52	-4.21	-4.16
	Proposed	5.76	1.10	1.82	0.42	8.27	1.53
Shifted Exponential	Naive	-22.13	-14.17	-11.91	-6.67	-22.65	-14.18
	Prasad-Rao	-9.86	-7.90	-4.85	-3.25	-9.69	-7.60
	Proposed	12.4	2.61	4.33	0.66	15.05	3.86

Table 5.2 reports the percent relative root mean squared error (RRMSE) of the MSPE estimators for the balanced case. It shows that the proposed MSPE estimator performs the best, in terms of RRMSE, when the sampling errors $\{e_i\}$ are non-normal, while the RRMSE in the normal $\{e_i\}$ case are almost identical for the proposed and the Prasad-Rao estimators. The naive MSPE estimator leads to relatively large RRMSE due to the large squared bias.

Table 5.2
Simulated values of percent relative root mean squared error (RRMSE) of the mean squared prediction error estimators for the balanced case: $\psi = D = 1$ and ψ is estimated by the Prasad-Rao (same as the Fay-Herriot) method-of-moments

		Distribution of v					
		Normal		Double Exponential		Shifted Exponential	
Distribution of e		$m = 30$	$m = 60$	$m = 30$	$m = 60$	$m = 30$	$m = 60$
Normal	Naive	4.29	2.03	2.48	1.05	5.99	3.20
	Prasad-Rao	2.60	1.57	1.62	0.87	3.87	2.61
	Proposed	2.60	1.57	1.62	0.87	3.87	2.61
Double Exponential	Naive	5.95	2.95	3.04	1.34	7.59	4.20
	Prasad-Rao	3.22	2.13	1.83	1.04	4.47	3.21
	Proposed	1.77	1.57	1.14	0.86	2.61	2.39
Shifted Exponential	Naive	7.58	3.88	3.72	1.61	9.30	5.02
	Prasad-Rao	4.04	2.75	2.18	1.21	5.25	3.73
	Proposed	1.61	1.41	0.80	0.80	2.24	1.99

Tables 5.3 and 5.4 report the percent RB and RRMSE for the unbalanced case and $m = 60$ when ψ is estimated by the Prasad-Rao simple method-of-moments and the Fay-Herriot method-of-moments, respectively. As in the balanced case, the naive MSPE estimator underestimates the true MSPE in all situations, and the extent of underestimation depends on the method of estimation of ψ as well as the distribution combination of the sampling errors $\{e_i\}$ and the random effects $\{v_i\}$. It appears that for both balanced and unbalanced cases, non-normality of $\{e_i\}$ causes more underestimation than that of $\{v_i\}$.

Table 5.3 shows that for the unbalanced case the RB for the Prasad-Rao normality-based MSPE estimator is negligible whenever the sampling errors $\{e_i\}$ are normally distributed regardless of the distribution of $\{v_i\}$; this is again consistent with previous theory (Lahiri and Rao, 1995). Our theory in Section 4 suggests that the normality-based Datta-Rao-Smith MSPE estimator, unlike the Prasad-Rao MSPE estimator, is not second-order unbiased even under the normality of $\{e_i\}$. It is, however, interesting to note from Table 5.4 that there is virtually no difference between the Datta-Rao-Smith estimator and the proposed robust estimator (4.4) when $\{e_i\}$ are normal. Note that there are minor differences between these two estimators in terms of RB even when both $\{e_i\}$ and $\{v_i\}$ are normally distributed; but this difference can be attributed to the fact that even in this case these two formulas differ simply because we are using an estimator of the kurtosis of the normal distribution in the robust MSPE formula (4.4).

Table 5.3
Percent Relative Bias (Relative Root Mean Squared Error) of the mean squared prediction error estimators for the unbalanced case and $m = 60$ when ψ is estimated by the Prasad-Rao simple method-of-moments

		Dist. of v								
		Normal			Double Exp.			Shifted Exp.		
Dist. of e	Group	Naive	PR	Prop.	Naive	PR	Prop.	Naive	PR	Prop.
Normal	G1	-4.68 (3.49)	0.19 (3.34)	0.19 (3.34)	-3.11 (2.96)	0.13 (2.83)	0.13 (2.83)	-4.97 (6.00)	-0.03 (5.84)	-0.03 (5.84)
	G2	-5.97 (0.82)	-0.33 (0.51)	-0.33 (0.51)	-2.43 (0.37)	0.15 (0.27)	0.15 (0.27)	-5.41 (1.43)	0.84 (0.96)	0.84 (0.96)
	G3	-5.78 (0.62)	-0.29 (0.34)	-0.29 (0.34)	-2.33 (0.25)	0.05 (0.18)	0.05 (0.18)	-6.54 (1.12)	-0.41 (0.62)	-0.41 (0.62)
	G4	-5.48 (0.43)	-0.24 (0.20)	-0.24 (0.20)	-2.17 (0.15)	-0.06 (0.10)	-0.06 (0.10)	-4.71 (0.74)	1.46 (0.37)	1.46 (0.37)
	G5	-3.97 (0.11)	0.12 (0.03)	0.12 (0.03)	-0.61 (0.03)	0.77 (0.02)	0.77 (0.02)	-4.66 (0.22)	0.77 (0.05)	0.77 (0.05)
Double Exp.	G1	-10.88 (5.13)	-6.31 (4.56)	-0.38 (4.03)	-5.12 (3.42)	-1.95 (3.15)	1.16 (2.88)	-10.61 (7.55)	-5.90 (7.01)	0.24 (6.40)
	G2	-8.50 (1.22)	-2.78 (0.72)	1.68 (0.43)	-2.35 (0.42)	0.26 (0.31)	1.55 (0.26)	-10.35 (2.03)	-4.05 (1.20)	1.32 (0.63)
	G3	-9.03 (0.97)	-3.42 (0.49)	0.79 (0.24)	-3.15 (0.32)	-0.74 (0.22)	0.37 (0.17)	-9.00 (1.53)	-2.54 (0.80)	2.86 (0.38)
	G4	-6.80 (0.61)	-1.27 (0.27)	2.75 (0.14)	-1.94 (0.19)	0.21 (0.13)	1.14 (0.11)	-8.66 (1.10)	-2.16 (0.46)	3.20 (0.20)
	G5	-4.44 (0.16)	0.12 (0.04)	3.39 (0.12)	-1.69 (0.04)	-0.33 (0.03)	0.22 (0.02)	-7.14 (0.35)	-0.75 (0.07)	4.88 (0.63)
Shifted Exp.	G1	-13.28 (6.11)	-8.79 (5.37)	2.89 (4.32)	-7.67 (3.92)	-4.56 (3.47)	1.53 (2.79)	-14.44 (8.85)	-9.94 (8.04)	1.77 (6.57)
	G2	-10.33 (1.46)	-4.56 (0.84)	4.63 (0.34)	-3.77 (0.49)	-1.15 (0.35)	1.45 (0.22)	-11.05 (2.28)	-4.67 (1.36)	6.41 (0.50)
	G3	-10.78 (1.18)	-5.10 (0.60)	3.63 (0.18)	-4.20 (0.36)	-1.78 (0.23)	0.46 (0.13)	-11.66 (1.84)	-5.21 (0.94)	5.80 (0.30)
	G4	-9.49 (0.80)	-3.93 (0.35)	4.42 (0.15)	-4.68 (0.26)	-2.53 (0.16)	-0.67 (0.08)	-9.71 (1.27)	-3.05 (0.52)	8.31 (0.46)
	G5	-6.34 (0.21)	-1.56 (0.05)	5.41 (0.41)	-2.91 (0.06)	-1.51 (0.03)	-0.43 (0.02)	-8.55 (0.42)	-1.78 (0.08)	10.68 (2.63)

The situation, however, is quite different when the sampling errors e_i are not normally distributed. The normality-based Prasad-Rao and Datta-Rao-Smith MSPE estimators both tend to underestimate, although in some non-normal situations both of them perform well. It appears that the underestimation is more prominent in the Prasad-Rao MSPE estimator than in the Datta-Rao-Smith MSPE estimator. As in the

balanced case, the distribution combination of $\{e_i\}$ and $\{v_i\}$ is an important factor in determining the extent of underestimation in both Prasad-Rao and Datta-Rao-Smith MSPE estimators. For example, when both e_i and v_i follow the location shifted exponential distribution, there is about an 8% underestimation for the Datta-Rao-Smith MSPE estimator for G_1; the corresponding underestimation for the Prasad-Rao MSPE estimator being about 10%. We also note that there is usually more underestimation for group G_1, the group with the smallest ratio ψ / D_i than for the other groups.

As m increases, underestimation or overestmation for all the methods decreases since they are all first-order unbiased. However, underestimation for both the naive and the normality-based Prasad-Rao or Datta-Rao-Smith MSPE estimator tend to decrease at a slower rate than for our robust method – thus verifying our theory. For example, consider group G_1 and the case when $\{e_i\}$ and $\{v_i\}$ follow the location shifted exponential distribution. When we increase m from 60 to 100 for group G_1, underestimation for the Prasad-Rao MSPE estimator decreases from about -10% to -7%. We note that for $m = 60$, our robust MSPE estimator suffers from a very minor overestimation of about 2%. When we increase m to 100, overestimation drops to 0.25%. For a small m, the performance of the robust MSPE estimator depends on the estimator of ψ used. For $m = 30$, we have observed that the robust MSPE estimator performs better when the Prasad-Rao method is used in place of the Fay-Herriot method of estimating ψ. This is possibly due to the fact that, unlike the Prasad-Rao method, we need to estimate the kurtosis of $\{v_i\}$ when the Fay-Herriot method is used and the estimation of kurtosis could be problematic for small m.

Tables 5.3 and 5.4 also show that overall our proposed MSPE estimator performs the best, in terms of RRMSE, when the sampling errors are non-normal, while the RRMSE in the normal case are almost identical for the proposed and the Prasad-Rao or DRS MSPE estimators. The naive MSPE estimator leads to relatively large RRMSE due to the large squared bias.

Table 5.4
Percent Relative Bias (Root Mean Squared Error) of the mean squared prediction error estimators for the unbalanced case and $m = 60$ when ψ is estimated by the Fay-Herriot method-of-moments

		Dist. of v								
		Normal			Double Exp.			Shifted Exp.		
Dist. of e	Group	Naive	DRS	Prop.	Naive	DRS	Prop.	Naive	DRS	Prop.
Normal	G1	-3.55 (2.48)	-0.32 (2.51)	0.58 (2.52)	-3.14 (2.73)	-0.49 (2.67)	0.67 (2.70)	-4.80 (4.98)	-1.64 (5.06)	0.69 (5.36)
	G2	-3.92 (0.53)	-0.24 (0.43)	-0.05 (0.42)	-2.45 (0.34)	-0.37 (0.28)	-0.27 (0.27)	-3.82 (1.05)	-0.01 (0.92)	0.49 (0.09)
	G3	-4.26 (0.40)	-0.73 (0.30)	-0.64 (0.30)	-2.39 (0.24)	-0.48 (0.19)	-0.48 (0.19)	-3.83 (0.78)	-0.12 (0.65)	0.10 (0.63)
	G4	-3.24 (0.26)	0.11 (0.19)	0.06 (0.19)	-1.06 (0.13)	0.65 (0.11)	0.57 (0.10)	-2.86 (0.51)	0.72 (0.41)	0.62 (0.04)
	G5	-2.45 (0.06)	-0.01 (0.04)	-0.26 (0.04)	-1.71 (0.03)	-0.64 (0.02)	-0.82 (0.02)	-2.22 (0.13)	0.55 (0.08)	-0.13 (0.09)
Double Exp.	G1	-6.88 (3.15)	-3.77 (3.04)	-0.44 (3.12)	-5.39 (3.06)	-2.79 (2.89)	0.08 (2.81)	-9.36 (5.91)	-6.36 (5.81)	-1.79 (6.06)
	G2	-5.12 (0.65)	-1.48 (0.51)	0.86 (0.45)	-2.34 (0.36)	-0.25 (0.30)	0.70 (0.27)	-5.53 (1.28)	-1.78 (1.08)	1.02 (0.95)
	G3	-4.38 (0.47)	-0.85 (0.36)	1.24 (0.31)	-2.45 (0.25)	-0.54 (0.19)	0.19 (0.17)	-6.11 (0.97)	-2.47 (0.77)	-0.07 (0.63)
	G4	-3.97 (0.31)	-0.64 (0.22)	1.13 (0.18)	-2.60 (0.16)	-0.91 (0.12)	-0.40 (0.10)	-4.54 (0.63)	-0.99 (0.49)	0.99 (0.39)
	G5	-3.63 (0.08)	-1.19 (0.04)	-0.25 (0.03)	-2.53 (0.04)	-1.47 (0.03)	-1.34 (0.02)	-3.28 (0.17)	-0.44 (0.10)	0.53 (0.06)
Shifted Exp.	G1	-9.03 (3.83)	-6.00 (3.64)	-0.44 (3.78)	-7.26 (3.40)	-4.72 (3.13)	-0.25 (2.89)	-10.98 (6.60)	-8.04 (6.44)	-1.40 (6.84)
	G2	-5.44 (0.78)	-1.80 (0.62)	2.69 (0.51)	-4.30 (0.44)	-2.25 (0.34)	-0.47 (0.25)	-8.46 (1.53)	-4.80 (1.24)	0.16 (0.93)
	G3	-6.89 (0.64)	-3.42 (0.46)	0.61 (0.31)	-3.85 (0.31)	-1.96 (0.23)	-0.50 (0.17)	-7.64 (1.13)	-4.02 (0.88)	0.54 (0.61)
	G4	-6.37 (0.44)	-3.08 (0.29)	0.49 (0.17)	-2.85 (0.17)	-1.15 (0.13)	-0.03 (0.10)	-6.44 (0.75)	-2.91 (0.56)	1.19 (0.35)
	G5	-3.78 (0.10)	-1.28 (0.06)	1.03 (0.03)	-1.46 (0.04)	-0.38 (0.03)	0.06 (0.02)	-4.69 (0.20)	-1.81 (0.11)	1.06 (0.04)

6. Discussion

The additional assumption of known sampling kurtosis κ_{ei} warrants further discussion. We will first outline a method for estimating the kurtosis of the sampling distribution of the direct estimator $\overline{y}_{iw} = \sum_{k\in s_i} w_k y_k \Big/ \sum_{k\in s_i} w_k$ of the area mean $\overline{Y}_i$, where s_i is the set of units in area i that are selected in the sample and w_k is the basic design weight of unit k, i.e., reciprocal of the first-order inclusion probability π_k of unit k. The extension to calibration weighting should be possible through linearization but is not attempted here. To achieve this, we treat the sample design as Poisson sampling, which simplifies the evaluation of both the fourth moment and its estimator. In the following, we note that published papers (e.g., Fúquene, Cristancho, Ospina and Morales (2019)) are using a sampling variance estimator, which is identical to that based on Poisson sampling, without realizing this. For smoothing purposes, Poisson sampling assumption is likely to perform well, especially when the sampling fraction within area is not large. Subsequently, we smooth the estimated fourth moment and the second moment to obtain a smoothed estimator of the kurtosis. This smoothed kurtosis estimator is then treated as the true kurtosis, consistent with the approach used in the normality-based Fay-Herriot model for smoothing sampling variances.

Using ratio approximation for the variance, Särndal, Swensson and Wretman (1992) obtained an approximate variance estimator of the direct domain estimator. Denoting π_{kl} the second-order inclusion probability of units k and l, their approximate variance estimator of $\overline{y}_{iw}$ is given by:

$$v(\overline{y}_{iw}) = \hat{N}_{iw}^{-2} \sum_{(k,l)\in s_i} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}} \left(\frac{y_k - \overline{y}_{iw}}{\pi_k} \right) \left(\frac{y_l - \overline{y}_{iw}}{\pi_l} \right),$$

where $\hat{N}_{iw} = \sum_{k\in s_i} w_k$ is an estimator of the population size N_i of area i. Fúquene et al. (2019) proposed the following simplification:

$$v(\overline{y}_{iw}) = \hat{N}_{iw}^{-2} \sum_{k\in s_i} w_k (w_k - 1)\, (y_k - \overline{y}_{iw})^2, \tag{6.1}$$

where $w_k = \pi_k^{-1}$. They treat this variance estimator as the true variance without any smoothing.

We note that the assumption regarding the second-order inclusion probabilities, as considered by Fúquene et al. (2019), is satisfied for Poisson sampling. We will first derive the formula (6.1) under Poisson sampling, by letting $\delta_k = 1$ if population unit k in domain i is in the sample and zero otherwise. Note that $\mathbb{E}(\delta_k) = \pi_k$, $\text{Var}(\delta_k) = \pi_k(1-\pi_k)$ and the indicator variables δ_k are independent under Poisson sampling, which implies $\text{Cov}(\delta_k, \delta_l) = 0, k \neq l$. Under the ratio approximation, we apply the formula for estimating the variance of the estimated domain total $\hat{Z}_{iw} = \sum_{k\in s_i} w_k z_k$ of the variable $(y_k - \overline{Y}_i)/N_i = z_k$. The variance of $\hat{Z}_{iw}$ is given by

$$V(\hat{Z}_{iw}) = \mathbb{E}\left[\sum_{k\in U_i} (\delta_k - \pi_k)\, z_k \pi_k^{-1} \right]^2 = \sum_{k\in U_i} \mathbb{E}(\delta_k - \pi_k)^2 z_k^2 \pi_k^{-2} = \sum_{k\in U_i} V(\delta_k)\, z_k^2 \pi_k^{-2},$$

noting that the indicator variables δ_k are independent, where the summation is over U_i, the set of all the population units in the domain i. Now using $V(\delta_k) = \pi_k(1-\pi_k)$, it follows that

$$V(\hat{Z}_{iw}) = \sum_{k\in U_i}(1-\pi_k)\,\pi_k^{-1}z_k^2.$$

Hence, an estimator of variance is given by

$$v(\hat{Z}_{iw}) = \sum_{k\in s_i}(1-\pi_k)\,\pi_k^{-2}z_k^2 = \sum_{k\in s_i}(w_k-1)\,w_k z_k^2.$$

Now replacing z_k by $\hat{z}_k = (y_k - \bar{y}_{iw})/\hat{N}_{iw}$, the above formula reduces to the formula used by Fúquene et al. (2019).

We can use the above approach for variance estimation to obtain the fourth moment of $\bar{y}_{iw}$ and its estimator. We have

$$\begin{aligned}\mu_4(\bar{y}_{iw}) &= \mathbb{E}\left[\sum_{k\in U_i}(\delta_k-\pi_k)\,z_k/\pi_k\right]^4 \\ &= \mathbb{E}\left[\sum_{k\in U_i}(\delta_k-\pi_k)^2 z_k^2/\pi_k^2 + 2\sum_{k<l\in U_i}(\delta_k-\pi_k)(\delta_l-\pi_l)(z_k/\pi_k)(z_l/\pi_l)\right]^2 \qquad (6.2)\\ &= \sum_{k\in U_i}\mathbb{E}(\delta_k-\pi_k)^4 z_k^4/\pi_k^4 + 6\sum_{k<l\in U_i}\mathbb{E}(\delta_k-\pi_k)^2\mathbb{E}(\delta_l-\pi_l)^2(z_k^2/\pi_k^2)(z_l^2/\pi_l^2),\end{aligned}$$

using independence of the indicator variables. We also have

$$\mathbb{E}(\delta_k-\pi_k)^4 = \pi_k(1-\pi_k)[\pi_k^3+(1-\pi_k)^3],\ \mathbb{E}(\delta_k-\pi_k)^2 = \pi_k(1-\pi_k).$$

An estimator for the fourth moment is derived from equation (6.2) by making the following modifications: replace $\sum_{k\in U_i}$ with $\sum_{k\in s_i}$, replace $\sum_{k<l\in U_i}$ with $\sum_{k<l\in s_i}$, multiply the term in the first sum by π_k^{-1}, multiply the term in the second sum by $\pi_k^{-1}\pi_l^{-1}$, and substitute z_k with $\hat{z}_k$. Once we have estimated the variance and fourth central moment of $\bar{y}_{iw}$, we can obtain an estimator for the kurtosis of $\bar{y}_{iw}$.

To stabilize the sampling variance estimates, various smoothing techniques are available in the small area literature; see, for example, Fay and Herriot (1979), Chapter 6 of Rao and Molina (2015), Hawala and Lahiri (2018), Lesage, Beaumont and Bocci (2022), and You and Hidiroglou (2023). Essentially, variance smoothing is done by modeling the sampling variance estimate conditional on a set of relevant auxiliary variables. Often, a log linear model is considered. Beaumont, Bocci, Bosa and Sombo (2024) used random forests to model the estimated sampling variances as a function of area level covariates. This method makes it easier because no specification of the functional form is needed. While the literature on variance smoothing can offer some insights into possible ways to smooth estimated sampling kurtosis, this is a new area of research that needs careful attention.

Smoothed estimate of kurtosis of the sampling error could be helpful in choosing between normality-based Fay-Herriot model and the proposed semi-parametric Fay-Herriot model. However, in the normal case, the proposed MSPE estimator performs similarly to the Prasad-Rao MSPE estimator.

Acknowledgements

S. Chen was supported in part by a Professional Development Award from the RTI International. J.N.K. Rao was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. We thank Ms. Huilin Li for computational support and the Editor and two anonymous referees for their helpful comments and constructive suggestions.

Appendix

The derivation of our MSPE estimator for $\hat{\psi} = \hat{\psi}_{\mathrm{PR}}$ follows along the lines of the proof given in Lahiri and Rao (1995). Therefore, we focus on the proof for $\hat{\psi} = \hat{\psi}_{\mathrm{FH}}$.

Derivation of the asymptotic variance of $\hat{\psi}$:

Expanding $A(\hat{\psi})$ around ψ by Taylor series, we have

$$\frac{1}{m-p}\mathbf{Y}'\mathbf{Q}(\psi)\,\mathbf{Y} - 1 + \frac{1}{m-p}\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\,\mathbf{Y}(\hat{\psi}-\psi) + O(\|\hat{\psi}-\psi\|^2) = 0,$$

where $A(\psi)$ is given by (2.2) and $\mathbf{Q}^{(j)}(\psi)$ is the j th derivative of $\mathbf{Q}(\psi)$ with respect to ψ ($j = 1, 2, \ldots$). Thus,

$$\hat{\psi} - \psi \;=\; -\frac{\frac{1}{m-p}\mathbf{Y}'\mathbf{Q}(\psi)\mathbf{Y} - 1}{\frac{1}{m-p}\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\mathbf{Y}} \;+\; O_p(m^{-1}),$$

noting that $\hat{\psi} - \psi = O_p(m^{-1/2})$.

Define $\mathbf{v} = (v_1, \ldots, v_m)', \mathbf{e} = (e_1, \ldots, e_m)'$, and $\boldsymbol{\varepsilon} = \mathbf{v} + \mathbf{e}$. Consider the transformation: $\tilde{\mathbf{v}} = \boldsymbol{\Sigma}^{-1/2}\mathbf{v}$, $\tilde{\mathbf{e}} = \boldsymbol{\Sigma}^{-1/2}\mathbf{e}$, $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon}$, and $\tilde{\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2}\mathbf{X}$. Then, it is easy to see that $\mathbf{Y}'\mathbf{Q}(\psi)\mathbf{Y} = \tilde{\boldsymbol{\varepsilon}}'\tilde{\mathbf{B}}\tilde{\boldsymbol{\varepsilon}}$, where $\tilde{\mathbf{B}} = \mathbf{I} - \tilde{\mathbf{X}}\{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\}^{-1}\tilde{\mathbf{X}}'$. Since

$$\frac{1}{m-p}\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\,\mathbf{Y} \;=\; -\frac{1}{m-p}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}) \;+\; O_p(m^{-1/2})$$

we obtain

$$\hat{\psi} - \psi \;=\; \frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})}\left[\tilde{\boldsymbol{\varepsilon}}'\tilde{\mathbf{B}}\tilde{\boldsymbol{\varepsilon}} - E(\tilde{\boldsymbol{\varepsilon}}'\tilde{\mathbf{B}}\tilde{\boldsymbol{\varepsilon}})\right] \;+\; O_p(m^{-1}). \tag{A.1}$$

Now use (A.1) and apply part (d) of Lemma C.4 of Lahiri and Rao (1995, page 765) to get the asymptotic variance, $\mathrm{var}(\hat{\psi})$.

Derivation of the second-order approximation to the bias of $\hat{\psi}$

Retaining the next term in the Taylor series expansion of $A(\hat{\psi})$, we obtain

$$\frac{\mathbf{Y}'\mathbf{Q}(\psi)\,\mathbf{Y}}{m-p} - 1 + \frac{\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\,\mathbf{Y}}{m-p}(\hat{\psi}-\psi) + \frac{1}{2}\frac{\mathbf{Y}'\mathbf{Q}^{(2)}(\psi)\,\mathbf{Y}}{m-p}(\hat{\psi}-\psi)^2 + O(\|\hat{\psi}-\psi\|^2) = 0,$$

which is equivalent to

$$\begin{aligned}&\frac{\mathbf{Y}'\mathbf{Q}(\psi)\,\mathbf{Y}}{m-p} - 1 + \frac{\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\,\mathbf{Y} - \mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(1)}(\psi)]}{m-p}(\hat{\psi}-\psi) + \frac{\mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(1)}(\psi)]}{m-p}(\hat{\psi}-\psi)\\ &+ \frac{1}{2}\frac{\mathbf{Y}'\mathbf{Q}^{(2)}(\psi)\,\mathbf{Y} - \mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(2)}(\psi)]}{m-p}(\hat{\psi}-\psi)^2 + \frac{1}{2}\frac{\mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(2)}(\psi)]}{m-p}(\hat{\psi}-\psi)^2 \qquad \text{(A.2)}\\ &= O(\|\hat{\psi}-\psi\|^2).\end{aligned}$$

We now take expectation on both sides of (A.2). To this end, using (A.1), Lemma C.4 of Lahiri and Rao (1995) and after considerable algebra, we obtain

$$E\left\{\mathbf{Y}'\mathbf{Q}^{(1)}(\psi)\,\mathbf{Y} - \mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(1)}(\psi)]\right\}(\hat{\psi}-\psi) = -2 - \frac{\psi^2\kappa_v\,\mathrm{tr}(\boldsymbol{\Sigma}^{-3}) + \mathrm{tr}(\mathbf{D}^2\Phi\boldsymbol{\Sigma}^{-3})}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})} + o(1), \qquad \text{(A.3)}$$

$$E\left\{\mathbf{Y}'\mathbf{Q}^{(2)}(\psi)\,\mathbf{Y} - \mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(2)}(\psi)]\right\}(\hat{\psi}-\psi)^2 = o(1). \qquad \text{(A.4)}$$

To obtain $b(\hat{\psi};\psi,\kappa_v)$, the second-order approximation to the bias of $\hat{\psi}$, we take the expectation in (A.2), and then use (A.3), (A.4), the expression for $\mathrm{var}(\hat{\psi})$, and the following facts:

$$\begin{aligned}E[\mathbf{Y}'\mathbf{Q}(\psi)\,\mathbf{Y}] &= m-p\\ \mathrm{tr}[\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(1)}(\psi)] &= -\mathrm{tr}[\mathbf{Q}(\psi)] = -\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\psi)] + O(1);\\ \mathrm{tr}[\mathbf{Q}(\psi)\boldsymbol{\Sigma}(\psi)\mathbf{Q}^{(1)}(\psi)\boldsymbol{\Sigma}(\psi)] &= -\mathrm{tr}[\mathbf{Q}(\psi)] = -\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\psi)] + O(1);\\ \mathrm{tr}[\boldsymbol{\Sigma}\mathbf{Q}^{(2)}(\psi)] &= 2\mathrm{tr}[\mathbf{Q}^2(\psi)] = 2\mathrm{tr}[\boldsymbol{\Sigma}^{-2}(\psi)] + O(1);\\ \mathbf{X}'\mathbf{Q}(\psi) &= 0.\end{aligned}$$

Proof of Result 1

Expanding $\hat{\theta}_i(\hat{\psi},\mathbf{Y})$ around ψ, we obtain

$$\hat{\theta}_i(\hat{\psi},\mathbf{Y}) - \hat{\theta}_i(\psi,\mathbf{Y}) = (\hat{\psi}-\psi)\,\hat{\theta}_i^{(1)}(\psi,\mathbf{Y}) + O_p(m^{-1}), \qquad \text{(A.5)}$$

where $\hat{\theta}_i^{(1)}(\psi,\mathbf{Y})$ is the first derivative of $\hat{\theta}_i(\psi,\mathbf{Y})$ with respect to ψ. Write

$$\hat{\theta}_i^{(1)}(\psi,\mathbf{Y}) = \tilde{\mathbf{l}}_i'\tilde{\boldsymbol{\varepsilon}}, \qquad \text{(A.6)}$$

where $\tilde{\mathbf{l}}_i' = (\tilde{l}_{i1},\ldots,\tilde{l}_{im})$ with $\tilde{l}_{ii} = \frac{D_i}{(\psi+D_i)^{3/2}} + O(m^{-1})$ and $\tilde{l}_{ij} = O(m^{-1})$, $j \neq i$. Using (A.1) and (A.6) in (A.5), we have

$$E[\hat{\theta}_i(\hat{\psi},\mathbf{Y})-\hat{\theta}_i(\psi,\mathbf{Y})]^2 \;=\; \frac{1}{(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}))^2}\,E\left\{\left(\tilde{\boldsymbol{\varepsilon}}'\tilde{\mathbf{B}}\tilde{\boldsymbol{\varepsilon}}-E[\tilde{\boldsymbol{\varepsilon}}'\tilde{\mathbf{B}}\tilde{\boldsymbol{\varepsilon}}]^2\right)(\tilde{\mathbf{l}}_i'\tilde{\boldsymbol{\varepsilon}})^2\right\} \;+\; o(m^{-1}). \tag{A.7}$$

Application of Lemma C.4 of Lahiri and Rao (1995) to (A.7) and algebra yields part (i) of Result 1.

To prove part (ii) of Result 1, write

$$\hat{\theta}_i(\psi,\mathbf{Y})-\theta_i=\tilde{\mathbf{k}}_i'\tilde{\mathbf{v}}+\tilde{\mathbf{m}}_i'\tilde{\mathbf{e}}, \tag{A.8}$$

where $\tilde{\mathbf{k}}_i'=(\tilde{k}_{i1},\ldots,\tilde{k}_{im}),\ \tilde{\mathbf{m}}_i'=(\tilde{m}_{i1},\ldots,\tilde{m}_{im}),$ with $\tilde{k}_{ii}=-\frac{D_i}{(\psi+D_i)^{1/2}}+O(m^{-1}),\ \tilde{m}_{ii}=\frac{\psi}{(\psi+D_i)^{1/2}}+O(m^{-1})$ and $\tilde{k}_{ij}=\tilde{m}_{ij}=O(m^{-1}),\ j\neq i.$ Using (A.1), (A.3), (A.8) and the regularity conditions (r.1), (r.2) and (r.3), we get

$$\begin{aligned}
&E[\hat{\theta}_i(\hat{\psi},\mathbf{Y})-\hat{\theta}_i(\psi,\mathbf{Y})][\hat{\theta}_i(\psi,\mathbf{Y})-\theta_i]\\
=\;&\frac{1}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1})}E\Big\{[\tilde{\mathbf{v}}'\tilde{\mathbf{B}}\tilde{\mathbf{v}}-E(\tilde{\mathbf{v}}'\tilde{\mathbf{B}}\tilde{\mathbf{v}}]\,(\tilde{\mathbf{l}}_i'\tilde{\mathbf{v}})\,(\tilde{\mathbf{k}}_i'\tilde{\mathbf{v}})+[\tilde{\mathbf{e}}'\tilde{\mathbf{B}}\tilde{\mathbf{e}}-E(\tilde{\mathbf{e}}'\tilde{\mathbf{B}}\tilde{\mathbf{e}}]\,(\tilde{\mathbf{l}}_i'\tilde{\mathbf{e}})\,(\tilde{\mathbf{m}}_i'\tilde{\mathbf{e}})\\
&+\;2(\tilde{\mathbf{v}}'\tilde{\mathbf{B}}\tilde{\mathbf{e}})\,(\tilde{\mathbf{l}}_i'\tilde{\mathbf{v}})\,(\tilde{\mathbf{m}}_i'\tilde{\mathbf{e}})+2(\tilde{\mathbf{v}}'\tilde{\mathbf{B}}\tilde{\mathbf{e}})\,(\tilde{\mathbf{l}}_i'\tilde{\mathbf{e}})\,(\tilde{\mathbf{k}}_i'\tilde{\mathbf{v}})\Big\}\;+\;o(m^{-1}).
\end{aligned} \tag{A.9}$$

Part (ii) of Result 1 follows from (A.9) after an application of Lemma C.4 of Lahiri and Rao (1995) and algebra.

References

Beaumont, J.-F., Bocci, C., Bosa, K. and Sombo, S. (2024). The use of random forests in small area estimation. Presented in the *Seventh International Conference on Establishment Statistics* (ICES VII) 2024.

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.

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.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.

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(1), 183-196.

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

Fúquene, J., Cristancho, C., Ospina, M. and Morales, D. (2019). Prevalence of international migration: An alternative for small area estimation. *arXiv, 1905.00353*.

Hawala, S., and Lahiri, P. (2018). Variance modeling for domains. *Statistics and Applications (New Series)*, 16(1), 399-409.

Kackar, R., and Harville, D. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79, 853-862.

Lahiri, P., and Rao, J.N.K. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association*, 90(430), 758-766.

Lesage, E., Beaumont, J.-F. and Bocci, C. (2022). Two local diagnostics to evaluate the effciency of the empirical best predictor under the Fay-Herriot model. *Survey Methodology*, 47(2), 279-297. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021002/article/00001-eng.pdf.

Otto, M.C., and Bell, W.R. (1995). Sampling error modelling of poverty and income statistics for states. *Proceedings of the Section on Government Statistics*, American Statistical Association, 160-165.

Pfeffermann, D., and Nathan, G. (1981). Regression analysis of data from a cluster sample. *Journal of the American Statistical Association*, 76, 681-689.

Prasad, 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(409), 163-171.

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

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag Inc.

You, Y., and Hidiroglou, M. (2023). Application of sampling variance smoothing methods for small area proportion estimation. *Journal of Official Statistics*, 39, 571-590.