On Measuring the Variability of Small Area Estimators in a Multivariate Fay-Herriot Model

Tsubasa Ito* and Tatsuya Kubokawa†

Abstract

This paper is concerned with the small area estimation in the multivariate Fay-Herriot model where covariance matrix of random effects are fully unknown. The covariance matrix is estimated by a Prasad-Rao type consistent estimator, and the empirical best linear unbiased predictor (EBLUP) of a vector of small area characteristics is provided. When the EBLUP is measured in terms of a mean squared error matrix (MSEM), a second-order approximation of MSEM of the EBLUP and a second-order unbiased estimator of the MSEM is derived analytically in closed forms. The performance is investigated through numerical and empirical studies.

Key words and phrases: Empirical Bayes method, empirical best linear unbiased prediction, mean squared error matrix, second-order approximation, small area estimation.

1 Introduction

Mixed effects models and their model-based estimators have been recognized as a useful method in statistical inference. In particular, small area estimation is an important application of mixed effects models. Although direct design-based estimates for small area means have large standard errors because of small sample sizes from small areas, the empirical best linear unbiased predictors (EBLUP) induced from mixed effects models provide reliable estimates by “borrowing strength” from neighboring areas and by using data of auxiliary variables. Such a model-based method for small area estimation has been studied extensively and actively from both theoretical and applied aspects, mostly for handling univariate survey data. For comprehensive reviews of small area estimation, see Ghosh and Rao (1994), Datta and Ghosh (2012), Pfeffermann (2013) and Rao and Molina (2015).

When multivariate data with correlations are observed from small areas for estimating multidimensional characteristics, like poverty and unemployment indicators, Fay (1987) suggested a multivariate extension of the univariate Fay-Herriot model, called a multivariate Fay-Herriot model, to produce reliable estimates of median incomes for four-, three- and five-person families. Fuller and Harter (1987) also considered a multivariate modeling for estimating a finite population mean vector. Datta, Day and Basawa (1999) provided unified theories in empirical linear unbiased prediction or empirical Bayes estimation in general multivariate mixed linear

*Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN.
E-Mail: tsubasa.ito.0710@gmail.com
†Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN.
E-Mail: tatsuya@e.u-tokyo.ac.jp
models. Datta, Day and Maiti (1998) suggested a hierarchical Bayesian approach to multivariate small area estimation. Datta, et al. (1999) showed the interesting result that the multivariate modeling produces more efficient predictors than the conventional univariate modeling. Porter, Wikle and Holan (2015) used the multivariate Fay-Herriot model for modeling spatial data. Ngaruye, von Rosen and Singull (2016) applied a multivariate mixed linear model to crop yield estimation in Rwanda.

Although Datta, et al. (1999) developed the general and unified theories concerning the empirical best linear unbiased predictors (EBLUP) and their uncertainty, it is definitely more helpful and useful to provide concrete forms with closed expressions for EBLUP, the second-order approximation of the mean squared error matrix (MSEM) and the second-order unbiased estimator of the mean squared error matrix. Recently, Benavent and Morales (2016) treated the multivariate Fay-Herriot model with the covariance matrix of random effects depending on unknown parameters. As a structure in the covariance matrix, they considered diagonal, AR(1) and the related structures and employed the residual maximum likelihood (REML) method for estimating the unknown parameters embedded in the covariance matrix. A second-order approximation and estimation of the MSEM were also derived. For some examples, however, we cannot assume specific structures without prior knowledge or information on covariance matrices.

In this paper, we treat the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown. This situation has been studied by Fay (1987), Fuller and Harter (1987), Datta, et al. (1998), and useful in the case that statisticians have little knowledge on structures in correlation. As a specific estimator of the covariance matrix, we employ Prasad-Rao type estimators with closed forms and use the modified versions which are restricted over the space of nonnegative definite matrices. The empirical best linear unbiased predictors are provided based on the Prasad-Rao type estimators, and second-order approximation of their mean squared error matrices and their second-order unbiased estimators of the MSEM are derived with closed expressions. These are multivariate extensions of the results given by Prasad and Rao (1990) and Datta, et al. (2005) for the univariate case.

The paper is organized as follows: Section 2 gives the Prasad-Rao type estimators and their nonnegative-definite modifications for the covariance matrix of the random effects, and shows their consistency. In Sections 3 and 4 the second-order approximation of MSEM of EBLUP and the second-order unbiased estimator of the MSEM are derived in closed forms. The performance of EBLUP and the MSEM estimator are investigated in Section 5. This numerical study illustrates that the proposals have good performances for the low-dimensional case. However, a $k \times k$ covariance matrix has $k(k+1)/2$ parameters, and we need more data so as to maintain the performances of the proposals for higher-dimensional cases.

Finally, it is noted that empirical best linear unbiased predictors for small area means are empirical Bayes estimators and related to the so-called James-Stein estimators. In this sense, the prediction in the multivariate Fay-Herriot model corresponds to the empirical Bayes estimation of a mean matrix of a multivariate normal distribution, which is related to the estimation of a precision matrix from a theoretical aspect as discussed in Efron and Morris (1976). In this framework, several types of estimators are suggested for estimation of the precision matrix, and it may be an interesting query whether those estimators provide improvements in the multivariate small area estimation.
2 Empirical Best Linear Unbiased Prediction

In this paper, we assume that area-level data \((y_1, X_1), \ldots, (y_m, X_m)\) are observed, where \(m\) is the number of small areas, \(y_i\) is a \(k\)-variate vector of direct survey estimates and \(X_i\) is a \(k \times s\) matrix of covariates associated with \(y_i\) for the \(i\)-th area. Then, the multivariate Fay-Herriot model is described as

\[
y_i = X_i \beta + v_i + \varepsilon_i, \quad i = 1, \ldots, m,
\]

where \(\beta\) is an \(s\)-variate vector of unknown regression coefficients, \(v_i\) is a \(k\)-variate vector of random effects depending on the \(i\)-th area and \(\varepsilon_i\) is a \(k\)-variate vector of sampling errors. It is assumed that \(v_i\) and \(\varepsilon_i\) are mutually independently distributed as

\[
v_i \sim N_k(0, \Psi) \quad \text{and} \quad \varepsilon_i \sim N_k(0, D_i),
\]

where \(\Psi\) is a \(k \times k\) unknown and nonsingular covariance matrix and \(D_1, \ldots, D_m\) are \(k \times k\) known covariance matrices. This is a multivariate extension of the so-called Fay-Herriot model suggested by Fay and Herriot (1979).

For example, we consider the crop data of Battese, Harter and Fuller (1988), who analyze the data in the nested error regression model. For the \(i\)-th county, let \(y_{i1}\) and \(y_{i2}\) be survey data of average areas of corn and soybean, respectively. Also let \(x_{i1}\) and \(x_{i2}\) be satellite data of average areas of corn and soybean, respectively. In this case, \(y_i, X_i\) and \(\beta\) correspond to

\[
y_i = (y_{i1}, y_{i2})^\top, \quad X_i = \begin{pmatrix} 1 & x_{i1} & x_{i2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & x_{i1} & x_{i2} \end{pmatrix}, \quad \beta = (\beta_1, \ldots, \beta_6)^\top
\]

for \(k = 2\) and \(s = 6\). Battese, et al. (1988) applied a univariate nested error regression model for each of \(y_{i1}\) and \(y_{i2}\), while we can use the multivariate model \((\mathbf{1})\) for analyzing both data simultaneously.

We now express model \((\mathbf{1})\) in a matrix form. Let \(y = (y_1^\top, \ldots, y_m^\top)^\top, X = (X_1^\top, \ldots, X_m^\top)^\top, v = (v_1^\top, \ldots, v_m^\top)^\top\) and \(\varepsilon = (\varepsilon_1^\top, \ldots, \varepsilon_m^\top)^\top\). Then, model \((\mathbf{1})\) is expressed as

\[
y = X \beta + v + \varepsilon,
\]

where \(v \sim N_{km}(0, I_m \otimes \Psi)\) and \(\varepsilon \sim N_{km}(0, D)\) for \(D = \text{block diag}(D_1, \ldots, D_m)\).

For the \(a\)-th area, we want to predict the quantity \(\theta_a = X_a \beta + v_a\), which is the conditional mean \(E[y_a \mid v_a]\) given \(v_a\). A reasonable estimator can be derived from the conditional expectation \(E[\theta_a \mid y_a] = X_a \beta + E[v_a \mid y_a]\). The conditional distribution of \(v_i\) given \(y_i\) and the marginal distribution of \(y_i\) are

\[
v_i \mid y_i \sim N_k(v_i^*(\beta, \Psi), (\Psi^{-1} + D_i^{-1})^{-1}), \quad y_i \sim N_k(X_i \beta, \Psi + D_i), \quad i = 1, \ldots, m,
\]

where

\[
v_i^*(\beta, \Psi) = \Psi(\Psi + D_i)^{-1}(y_i - X_i \beta) = \{I_k - D_i(\Psi + D_i)^{-1}\}(y_i - X_i \beta).
\]

Thus, we get the estimator

\[
\theta_a^*(\beta, \Psi) = X_a \beta + E[v_a \mid y_a] = X_a \beta + v_a^*(\beta, \Psi)
\]

\[
= y_a - D_a(\Psi + D_a)^{-1}(y_a - X_a \beta),
\]

3
which corresponds to the Bayes estimator of $\theta_a$ in the Bayesian framework.

When $\Psi$ is known, the maximum likelihood estimator or generalized least squares estimator of $\beta$ is

$$
\hat{\beta}(\Psi) = \{X^\top(I_m \otimes \Psi + D)^{-1}X\}^{-1}X^\top(I_m \otimes \Psi + D)^{-1}y
$$

$$
= \left\{ \sum_{i=1}^{m} X_i^\top(\Psi + D_i)^{-1}X_i \right\}^{-1} \sum_{i=1}^{m} X_i^\top(\Psi + D_i)^{-1}y_i. \tag{5}
$$

Substituting $\hat{\beta}(\Psi)$ into $\theta^*(\beta, \Psi)$ yields the estimator

$$
\hat{\theta}_a(\Psi) = y_a - D_a(\Psi + D_a)^{-1}\{y_a - X_a\hat{\beta}(\Psi)\}. \tag{6}
$$

Datta, et al. (1999) showed that $\hat{\theta}_a(\Psi)$ is the best linear unbiased predictor (BLUP) of $\theta_a$. It can be also demonstrated that $\hat{\theta}_a(\Psi)$ is the Bayes estimator against the uniform prior distribution of $\beta$ as well as the empirical Bayes estimator as shown above, which is called the Bayes empirical Bayes estimator.

Concerning estimation of $\Psi$, it is noted that $E[(y_i - X_i\beta)(y_i - X_i\beta)^\top] = \Psi + D_i$ for $i = 1, \ldots, m$, which implies that $\sum_{i=1}^{m} E[(y_i - X_i\beta)(y_i - X_i\beta)^\top] = m\Psi + \sum_{i=1}^{m} D_i$. Substituting the ordinary least squares estimator $\hat{\beta} = (X^\top X)^{-1}X^\top y$ into $\beta$, we get the consistent estimator

$$
\hat{\Psi}_0 = \frac{1}{m} \sum_{i=1}^{m} \{(y_i - X_i\hat{\beta})(y_i - X_i\hat{\beta})^\top - D_i\}. \tag{7}
$$

Taking the expectation of $\hat{\Psi}_0$, we can see that $E[\hat{\Psi}_0] = \Psi + \text{Bias}_{\hat{\Psi}_0}(\Psi)$, where

$$
\text{Bias}_{\hat{\Psi}_0}(\Psi) = \frac{1}{m} \sum_{i=1}^{m} X_i(X^\top X)^{-1}\left\{ \sum_{j=1}^{m} X_j^\top(\Psi + D_j)X_j \right\}(X^\top X)^{-1}X_i^\top
$$

$$
- \frac{1}{m} \sum_{i=1}^{m} (\Psi + D_i)X_i(X^\top X)^{-1}X_i^\top - \frac{1}{m} \sum_{i=1}^{m} X_i(X^\top X)^{-1}X_i^\top(\Psi + D_i). \tag{8}
$$

Substituting $\hat{\Psi}_0$ into $\text{Bias}_{\hat{\Psi}_0}(\Psi)$, we get a bias-corrected given by

$$
\hat{\Psi}_1 = \hat{\Psi}_0 - \text{Bias}_{\hat{\Psi}_0}(\Psi_0). \tag{9}
$$

For notational convenience, we use the same notation $\hat{\Psi}$ for $\hat{\Psi}_0$ and $\hat{\Psi}_1$ without any confusion. It is noted that both estimators are not necessarily nonnegative definite. In this case, there exist a $k \times k$ orthogonal matrix $H$ and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$ such that $\hat{\Psi} = H\Lambda H^\top$. Let $\Lambda^+ = \text{diag}(\max\{0, \lambda_1\}, \ldots, \max\{0, \lambda_k\})$, and let

$$
\hat{\Psi}^+ = H\Lambda^+ H^\top.
$$

Replace $\Psi$ in $\hat{\theta}_a(\Psi)$ with the estimator $\hat{\Psi}^+$, and the resulting estimator is the empirical Bayes (EB) estimator

$$
\hat{\theta}^{EB}_a = \hat{\theta}_a(\hat{\Psi}^+). \tag{10}
$$
To guarantee asymptotic properties of \( \hat{\Psi}^+ \), we assume the following conditions:

(H1) \( 0 < k < \infty \), \( 0 < s < \infty \).

(H2) There exist positive constants \( \underline{d} \) and \( \overline{d} \) such that \( \underline{d} \) and \( \overline{d} \) do not depend on \( m \) and satisfy \( dI_k \leq D_i \leq \overline{d}I_k \) for \( i = 1, \ldots, m \).

(H3) \( X^\top X \) is nonsingular and \( X^\top X/m \) converges to a positive definite matrix.

**Theorem 1** Under conditions (H1)-(H3), the following properties hold for \( \hat{\Psi} = \hat{\Psi}_0 \) and \( \hat{\Psi}_1 \):

1. \( \text{Bias}(\hat{\Psi}_0(\Psi)) = O(m^{-1}) \), which means that \( \hat{\Psi}_0 \) has the second-order bias, while \( \hat{\Psi}_1 \) is a second-order unbiased estimator of \( \Psi \).

2. \( \hat{\Psi} - \Psi = O_p(m^{-1/2}) \) and \( \beta(\hat{\Psi}) - \beta = O_p(m^{-1/2}) \).

3. The nonnegative definite matrix \( \hat{\Psi}^+ \) is consistent for large \( m \), and \( P(\hat{\Psi}^+ \neq \hat{\Psi}) = O(m^{-K}) \) for any \( K \).

**Proof.** We begin with writing \( \hat{\Psi}_0 - \Psi \) as

\[
\hat{\Psi}_0 - \Psi = \frac{1}{m} \sum_{i=1}^{m} \{(y_i - X_i\beta)(y_i - X_i\beta)^\top - (\Psi + D_i)\} + \frac{1}{m} \sum_{i=1}^{m} X_i(\hat{\beta} - \beta)(\hat{\beta} - \beta)^\top X_i^\top
\]

\[
- \frac{1}{m} \sum_{i=1}^{m} (y_i - X_i\beta)(\hat{\beta} - \beta)^\top X_i^\top - \frac{1}{m} \sum_{i=1}^{m} X_i(\hat{\beta} - \beta)(y_i - X_i\beta)^\top,
\]

which yields the bias given in (8). It is easy to check that the bias is of order \( O(m^{-1}) \).

For (2), it is noted that \( \hat{\Psi} - \Psi \) is approximated as

\[
\hat{\Psi} - \Psi = \frac{1}{m} \sum_{i=1}^{m} \{(y_i - X_i\beta)(y_i - X_i\beta)^\top - (\Psi + D_i)\} + O_p(m^{-1})
\]

\[
= \frac{1}{m} \sum_{i=1}^{m} \{u_iu_i^\top - (\Psi + D_i)\} + O_p(m^{-1}), \quad (11)
\]

where \( u_i = y_i - X_i\beta \), having \( N_i(0, \Psi + D_i) \). It is here noted that \( (u_iu_i^\top - (\Psi + D_i))/m \) for \( i = 1, \ldots, m \) are mutually independent and \( E(u_iu_i^\top - (\Psi + D_i))/m = 0 \) for \( i = 1, \ldots, m \). Then the consistency follows because \( \sum_{i=1}^{m} E(u_iu_i^\top - (\Psi + D_i))^2/m^2 = \sum_{i=1}^{m} (2(\Psi + D_i)^2 + \text{tr}(\Psi + D_i)I_k)/m^2 = O(m^{-1}) \) under condition (H2). Using condition (H2) and finiteness of moments of normal random variables, we can show that \( \sqrt{m}(\hat{\Psi} - \Psi) \) converges to a multivariate normal distribution, which implies that \( \hat{\Psi} - \Psi = O_p(m^{-1/2}) \).

We next verify that \( \hat{\beta}(\hat{\Psi}) - \beta = O_p(m^{-1/2}) \). Note that \( \hat{\beta}(\hat{\Psi}) - \beta \) is decomposed as \( \{\hat{\beta}(\hat{\Psi}) - \hat{\beta}(\Psi)\} + \{\hat{\beta}(\Psi) - \beta\} \). For \( \hat{\beta}(\Psi) - \beta \), it is noted that

\[
\hat{\beta}(\Psi) - \beta = \left\{ \sum_{i=1}^{m} X_i^\top(\Psi + D_i)^{-1}X_i \right\}^{-1} \sum_{i=1}^{m} X_i^\top(\Psi + D_i)^{-1}(y_i - Ey_i) \quad (12)
\]

Then, \( \text{Var}(\hat{\beta}(\Psi) - \beta) = \left\{ \sum_{i=1}^{m} X_i^\top(\Psi + D_i)^{-1}X_i \right\}^{-1} = O(1/m) \) and this implies \( \hat{\beta}(\Psi) - \beta = \)
\( O_p(m^{-1/2}) \). We next evaluate \( \hat{\beta}(\hat{\Psi}) - \beta(\Psi) \) as
\[
\hat{\beta}(\hat{\Psi}) - \beta(\Psi) = \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} y_i
- \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} y_i
= \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} y_i
= I_1 + I_2.
\]
First, \( I_1 \) is written as
\[
I_1 = - \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} y_i,
\]
which is of order \( O_p(m^{-1/2}) \), because \( \sum_{i=1}^{m} X_i^T (\hat{\Psi} + D_i)^{-1} X_i = O_p(m) \) and \( \sum_{i=1}^{m} X_i^T (\hat{\Psi} + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} y_i = O_p(m^{1/2}) \). Next, \( I_2 \) is rewritten as
\[
I_2 = - \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} y_i
\times \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} y_i
= \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} X_i
\times \left\{ \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i \right\}^{-1} \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} y_i,
\]
which is of order \( O_p(m^{-1/2}) \), because \( \sum_{i=1}^{m} X_i^T (\hat{\Psi} + D_i)^{-1} X_i = O_p(m) \), \( \sum_{i=1}^{m} X_i^T (\hat{\Psi} + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} X_i = O_p(m^{1/2}) \), \( \sum_{i=1}^{m} X_i^T (\Psi + D_i)^{-1} X_i = O(m) \) and \( \sum_{i=1}^{m} X_i^T (\hat{\Psi} + D_i)^{-1} y_i = O_p(m) \). Thus, we have \( \hat{\beta}(\hat{\Psi}) - \beta(\Psi) = O_p(m^{-1/2}) \), and it is concluded that \( \hat{\beta}(\hat{\Psi}) - \beta = O_p(m^{-1/2}) \).

For (3), let \( \hat{\lambda}_1, \ldots, \hat{\lambda}_k \) be eigenvalues of \( \hat{\Psi} \), and let \( \lambda_1, \ldots, \lambda_k \) be eigenvalues of \( \Psi \). Then, for \( j = 1, \ldots, k \),
\[
P(\hat{\lambda}_j < 0) = P(\hat{\lambda}_j - \lambda_j < -\lambda_j) = P((\hat{\lambda}_j - \lambda_j) > \lambda_j) \leq P(|\sqrt{m}(\hat{\lambda}_j - \lambda_j)| > \sqrt{m}\lambda_j).
\]
Note that \( \lambda_j > 0 \). It follows from the Markov inequality that for any \( K > 0 \),
\[
P(|\sqrt{m}(\hat{\lambda}_j - \lambda_j)| > \sqrt{m}\lambda_j) \leq \frac{E[|\sqrt{m}(\hat{\lambda}_j - \lambda_j)|^{2K}]}{(\sqrt{m}\lambda_j)^{2K}} = O(m^{-K}),
\]
because \( \hat{\lambda}_j - \lambda_j = O_p(m^{-1/2}) \) from \( \hat{\Psi} - \Psi = O_p(m^{-1/2}) \). \( \square \)
3 Second-order Approximation of Mean Squared Error Matrix

Uncertainty of the empirical Bayes estimator $\hat{\theta}_a^{EB}$ in (10) is measured by the mean squared error matrix (MSEM), defined as MSEM($\hat{\theta}_a^{EB}$) = $E[(\hat{\theta}_a^{EB} - \theta_a)(\hat{\theta}_a^{EB} - \theta_a)^\top]$. It is noted that

$$\hat{\theta}_a^{EB} - \theta_a = \{\theta_a^*(\beta, \Psi) - \theta_a\} + \{\hat{\theta}_a(\Psi) - \theta_a^*(\beta, \Psi)\} + \{\hat{\theta}_a(\Psi) - \theta_a(\Psi)\}$$

and that $\hat{\theta}_a(\Psi) - \theta_a^*(\beta, \Psi) = -D_a(\Psi + D_a)^{-1}X_a(\hat{\beta}(\Psi) - \beta)$. The following lemma is useful for evaluating the mean square error matrix.

**Lemma 1** $\hat{\beta}(\Psi)$ is independent of $y - X\hat{\beta}$ or $\hat{\Psi}$. Also, $\hat{\beta}(\Psi)$ is independent of $\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)$.

**Proof.** The covariance of $y - X\hat{\beta}$ and $\hat{\beta}(\Psi)$ is

$$E[(y - X\hat{\beta})(\hat{\beta}(\Psi) - \beta)^\top](I_m \otimes \Psi + D)^{-1}X$$

$$= E[(y - X\hat{\beta})(y - X\hat{\beta})^\top](I_m \otimes \Psi + D)^{-1}X$$

$$= [I_m \otimes \Psi + D - X(X^\top(I_m \otimes \Psi + D)^{-1}X)^{-1}X^\top(I_m \otimes \Psi + D)^{-1}X]$$

$$= 0.$$

This implies that $\hat{\beta}(\Psi)$ is independent of $y - X\hat{\beta}$ or $\hat{\Psi}$. It is also noted that

$$\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi) = -D_a(\Psi + D_a)^{-1}(y_a - X_a\hat{\beta}(\Psi)) + D_a(\Psi + D_a)^{-1}(y_a - X_a\hat{\beta}(\Psi))$$

$$= D_a((\Psi + D_a)^{-1} - (\hat{\Psi} + D_a)^{-1})(y_a - X_a\hat{\beta})$$

$$+ D_a(\hat{\Psi} + D_a)^{-1}X_a\{X^\top(I_m \otimes \hat{\Psi} + D)^{-1}X\}^{-1}X^\top(I_m \otimes \hat{\Psi} + D)^{-1}(y - X\hat{\beta})$$

$$- D_a(\Psi + D_a)^{-1}X_a\{X^\top(I_m \otimes \Psi + D)^{-1}X\}^{-1}X^\top(I_m \otimes \Psi + D)^{-1}(y - X\hat{\beta}),$$

which is a function of $y - X\hat{\beta}$. Hence, $\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)$ is independent of $\hat{\beta}(\Psi)$. \hfill \square

Using Lemma 1 we can decompose the mean squared error matrix as

$$\text{MSEM}(\hat{\theta}_a^{EB}) = E[\{\theta_a^*(\beta, \Psi) - \theta_a\}\{\theta_a^*(\beta, \Psi) - \theta_a\}^\top]$$

$$+ E[\{\hat{\theta}_a(\Psi) - \theta_a^*(\beta, \Psi)\}\{\hat{\theta}_a(\Psi) - \theta_a^*(\beta, \Psi)\}^\top]$$

$$+ E[\{\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)\}\{\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)\}^\top]$$

$$= G_{1a}(\Psi) + G_{2a}(\Psi) + E[\{\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)\}\{\hat{\theta}_a^{EB} - \hat{\theta}_a(\Psi)\}^\top],$$

(16)

where

$$G_{1a}(\Psi) = (\Psi^{-1} + D_a^{-1})^{-1} = \Psi(\Psi + D_a)^{-1}D_a,$$

$$G_{2a}(\Psi) = D_a(\Psi + D_a)^{-1}X_a\{X^\top(I_m \otimes \Psi + D)^{-1}X\}^{-1}X_a^\top(\Psi + D_a)^{-1}D_a.$$

(17)

The third term can be approximated as

$$G_{3a}(\Psi) = \frac{1}{m^2}D_a(\Psi + D_a)^{-1}\left[\sum_{i=1}^m(\Psi + D_i)(\Psi + D_a)^{-1}(\Psi + D_i)\right]$$

$$+ \sum_{i=1}^m\{\text{tr}[((\Psi + D_i)(\Psi + D_a)^{-1})](\Psi + D_i)\}(\Psi + D_a)^{-1}D_a.$$
Theorem 2  The mean squared error matrix of the empirical Bayes estimator $\hat{\theta}_a^{EB}$ is approximated as

$$\text{MSEM}(\hat{\theta}_a^{EB}) = G_{1a}(\Psi) + G_{2a}(\Psi) + G_{3a}(\Psi) + O(m^{-3/2}). \quad (19)$$

Proof. We shall prove that $E[(\hat{\theta}_a^{EB} - \bar{\theta}_a(\Psi))(\hat{\theta}_a^{EB} - \bar{\theta}_a(\Psi))^\top] = G_{3a}(\Psi) + O_p(m^{-3/2})$.

Also from (2) in Theorem 1, it is sufficient to show this approximation for $\hat{\Psi}$ instead of $\hat{\Psi}^+$. It is observed that

$$\hat{\theta}_a^{EB} - \bar{\theta}_a(\Psi) = D_a\{(\Psi + D_a)^{-1} - (\hat{\Psi} + D_a)^{-1}\}(y_a - X_a\beta) + D_a(\hat{\Psi} + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta\} - D_a(\Psi + D_a)^{-1}X_a\{\beta(\Psi) - \beta\}.$$  

Using the equation

$$(\hat{\Psi} + D_i)^{-1} = (\Psi + D_i)^{-1} - (\Psi + D_i)^{-1}(\Psi - \hat{\Psi})(\Psi + D_i)^{-1}, \quad (20)$$

we can see that

$$D_a\{(\Psi + D_a)^{-1} - (\hat{\Psi} + D_a)^{-1}\}(y_a - X_a\beta) = D_a(\Psi + D_a)^{-1}(\Psi - \hat{\Psi})(\Psi + D_a)^{-1}(y_a - X_a\beta) = D_a(\Psi + D_a)^{-1}(\Psi - \hat{\Psi})(\Psi + D_a)^{-1}(y_a - X_a\beta) + O_p(m^{-1})$$

and

$$D_a(\hat{\Psi} + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta\} = D_a(\hat{\Psi} + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta\} - D_a(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)(\hat{\Psi} + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta\} = D_a(\Psi + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta\} + O_p(m^{-1}).$$

Thus, we have

$$\hat{\theta}_a^{EB} - \bar{\theta}_a(\Psi) = D_a(\Psi + D_a)^{-1}(\Psi - \hat{\Psi})(\Psi + D_a)^{-1}(y_a - X_a\beta) + D_a(\Psi + D_a)^{-1}X_a\{\beta(\hat{\Psi}) - \beta(\Psi)\} + O_p(m^{-1}) = I_1 + I_2 + O_p(m^{-1}). \quad (say)$$

For $I_2$, it is noted that

$$\hat{\beta}(\hat{\Psi}) - \beta(\Psi) = \left[\left\{\sum_{j=1}^{m} X_j^\top(\hat{\Psi} + D_j)^{-1}X_j\right\}^{-1} - \left\{\sum_{j=1}^{m} X_j^\top(\Psi + D_j)^{-1}X_j\right\}^{-1}\right] \times \sum_{i=1}^{m} X_i^\top(\hat{\Psi} + D_i)^{-1}(y_i - X_i\beta) + \left\{\sum_{j=1}^{m} X_j^\top(\Psi + D_j)^{-1}X_j\right\}^{-1} \sum_{i=1}^{m} X_i^\top\left\{(\hat{\Psi} + D_i)^{-1} - (\Psi + D_i)^{-1}\right\}(y_i - X_i\beta) = I_{21} + I_{22}.$$
We can evaluate $I_{21}$ as

$$I_{21} = \left\{ \sum_{j=1}^{m} X_j^\top (\Psi + D_j)^{-1} X_j \right\}^{-1} \left\{ \sum_{j=1}^{m} X_j^\top (\Psi + D_j)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_j)^{-1} X_j \right\} \{ \hat{\beta}(\hat{\Psi}) - \beta \}
$$

$= O_p(m^{-1})$, because $\sum_{j=1}^{m} X_j^\top (\Psi + D_j)^{-1} X_j = O(m)$, $\sum_{i=1}^{m} X_i^\top (\Psi + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1} X_i = O_p(m^{1/2})$ and $\hat{\beta}(\hat{\Psi}) - \beta = O_p(m^{-1/2})$ from Theorem [1] (2). We next estimate $I_{22}$ as

$$I_{22} = - \left\{ \sum_{j=1}^{m} X_j^\top (\Psi + D_j)^{-1} X_j \right\}^{-1} \left\{ \sum_{j=1}^{m} X_j^\top A(\hat{\Psi}, D_i) X_j \right\}$$

$$\times \left\{ \sum_{i=1}^{m} X_i^\top A(\hat{\Psi}, D_i) X_i \right\}^{-1} \sum_{i=1}^{m} X_i^\top A(\hat{\Psi}, D_i)(y_i - X_i\beta)$$

for $A(\hat{\Psi}, D_i) = (\hat{\Psi} + D_i)^{-1} (\hat{\Psi} - \Psi)(\Psi + D_i)^{-1}$. It can be seen that $I_{22} = O_p(m^{-1})$ from the same arguments as in $I_{21}$. Thus, it follows that $I_{2} = O_p(m^{-1})$. Hence, we have

$$E[\{\hat{\theta}_a - \hat{\omega}_a(\Psi)\} E^\top_a - \hat{\omega}_a(\Psi)]$$

$$= D_a(\Psi + D_a)^{-1} E[(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}(y_a - X_a\beta)(y_a - X_a\beta)^\top (\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)]$$

$$\times (\Psi + D_a)^{-1} D_a + O(m^{-3/2}).$$

It is noted from [3] that $\hat{\Psi} - \Psi$ is approximated as

$$\hat{\Psi} - \Psi = \frac{1}{m} \sum_{i=1}^{m} \{u_i u_i^\top - (\Psi + D_i)\} + O_p(m^{-1}),$$

which is used to evaluate

$$E[(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}(y_a - X_a\beta)(y_a - X_a\beta)^\top (\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)]$$

$$= \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E[\{u_i u_i^\top - (\Psi + D_i)\}(\Psi + D_a)^{-1} u_a u_a^\top (\Psi + D_a)^{-1}\{u_j u_j^\top - (\Psi + D_i)\}]$$

$$+ O(m^{-3/2})$$

$$= \frac{1}{m^2} \sum_{i=1}^{m} E[\{u_i u_i^\top - (\Psi + D_i)\}(\Psi + D_a)^{-1} u_a u_a^\top (\Psi + D_a)^{-1}\{u_i u_i^\top - (\Psi + D_i)\}]$$

$$+ O(m^{-3/2}),$$

since $E[\{u_i u_i^\top - (\Psi + D_i)\}(\Psi + D_a)^{-1} u_a u_a^\top (\Psi + D_a)^{-1}\{u_j u_j^\top - (\Psi + D_i)\}] = 0$ for $i \neq j$. Letting $z_i = (\Psi + D_i)^{-1/2} u_i$, we can see that $z_i \sim N_k(0, I_k)$. Then,

$$\frac{1}{m^2} \sum_{i=1}^{m} E[\{u_i u_i^\top - (\Psi + D_i)\}(\Psi + D_a)^{-1} u_a u_a^\top (\Psi + D_a)^{-1}\{u_i u_i^\top - (\Psi + D_i)\}]$$

$$= \frac{1}{m^2} \sum_{i \neq a} (\Psi + D_i)^{1/2} E[(z_i z_i^\top - I)B z_a z_a^\top B^\top (z_i z_i^\top - I)](\Psi + D_i)^{1/2} + O(m^{-2}),$$

9
for \( B = (\Psi + D_1)^{1/2}(\Psi + D_a)^{-1/2} \). Let \( C = BB^\top = (\Psi + D_1)^{1/2}(\Psi + D_a)^{-1}(\Psi + D_1)^{1/2} \). For \( i \neq a \),

\[
E[(z_i z_i^\top - I)B z_a z_a^\top B^\top (z_i z_i^\top - I)] \\
= E[z_i z_i^\top B z_a z_a^\top B^\top z_i z_i^\top + B z_a z_a^\top B^\top - z_i z_i^\top B z_a z_a^\top B^\top - B z_a z_a^\top B^\top z_i z_i^\top] \\
= E[z_i z_i^\top C z_i z_i^\top - C] = C + (\text{tr} C) I_k,
\]

because \( E[z_i z_i^\top C z_i z_i^\top] = 2C + (\text{tr} C) I_k \). Thus,

\[
\frac{1}{m^2} \sum_{i \neq a} \{(\Psi + D_i)^{1/2}(C + (\text{tr} C) I_k)(\Psi + D_i)^{1/2} \\
= \frac{1}{m^2} \sum_{i = 1}^{m} \{(\Psi + D_i)^{1/2}(C + (\text{tr} C) I_k)(\Psi + D_i)^{1/2} + O(m^{-2}),
\]

which leads to the expression in (13).

\[ \square \]

4 Estimation of Mean Squared Error Matrix

In this section, we obtain a second-order unbiased estimator of the mean squared error matrix of the empirical Bayes estimator \( \hat{\theta}_a^E \) in (12). A naive estimator of \( \text{MSEM}(\hat{\theta}_a^E) \) is the plug-in estimator of (12) given by \( G_{1a}(\hat{\Psi}^+) + G_{2a}(\hat{\Psi}^+) + G_{3a}(\hat{\Psi}^+) \), but this has a second-order bias, because \( E[G_{1a}(\hat{\Psi}^+)] = G_{1a}(\Psi) + O(m^{-1}) \). Thus, we need to correct the second-order bias. Let

\[
G_{4a}(\Psi) = -D_a(\Psi + D_a)^{-1}\text{Bias}_\Phi(\Psi)(\Psi + D_a)^{-1}D_a, \tag{21}
\]

where \( \text{Bias}_\Phi(\hat{\Psi}) \) is the bias of \( \hat{\Psi} \) given by

\[
\text{Bias}_\Phi(\hat{\Psi}) = \begin{cases} \text{Bias}_\Phi(\Psi) & \text{for } \hat{\Psi} = \hat{\Psi}_0, \\ 0 & \text{for } \hat{\Psi} = \hat{\Psi}_1, \end{cases}
\]

where \( \text{Bias}_\Phi(\Psi) \) is given in (5). Define the estimator \( \text{mse}_m(\hat{\theta}_a^E) \) by

\[
\text{mse}_m(\hat{\theta}_a^E) = G_{1a}(\hat{\Psi}^+) + G_{2a}(\hat{\Psi}^+) + 2G_{3a}(\hat{\Psi}^+) + G_{4a}(\hat{\Psi}^+). \tag{22}
\]

**Theorem 3** Under the assumption, \( E[G_{1a}(\hat{\Psi}^+) + G_{3a}(\hat{\Psi}^+) + G_{4a}(\hat{\Psi}^+)] = G_{1a}(\Psi) + O(m^{-3/2}) \), and

\[
E[\text{mse}_m(\hat{\theta}_a^E)] = \text{MSEM}(\hat{\theta}_a^E) + O(m^{-3/2}),
\]

namely, \( \text{mse}_m(\hat{\theta}_a^E) \) is a second-order unbiased estimator of \( \text{MSEM}(\hat{\theta}_a^E) \).

**Proof.** From (2) in Theorem 1 it is sufficient to show this approximation for \( \hat{\Psi} \) instead of \( \hat{\Psi}^+ \). Using the equation in (20), we can rewrite \( G_{1a}(\hat{\Psi}) \) as

\[
G_{1a}(\hat{\Psi}) = (\hat{\Psi}^{-1} + D_a^{-1})^{-1} = D_a - D_a(\hat{\Psi} + D_a)^{-1}D_a \\
= G_{1a}(\Psi) + D_a(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}D_a - D_a(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}D_a + O_p(m^{-3/2}). \tag{23}
\]
We shall evaluate each term in RHS of the above equality. It is easy to see from (3) that $E[\hat{\Psi} - \Psi] = \text{Bias}(\hat{\Psi})$, which is written as (8). We next evaluate $E[(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)]$, which is, from (3), approximated as

$$
E[(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)]
$$

$$
= \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E \left[ \left\{ (y_i - X_i \beta)(y_j - X_j \beta)^\top - (\Psi + D_i) \right\}(\Psi + D_a)^{-1} \right.
$$

$$
\times \left\{ (y_j - X_j \beta)(y_j - X_j \beta)^\top - (\Psi + D_j) \right\} \right] + O(m^{-3/2})
$$

$$
= \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\Sigma + D_i)^{1/2} E[(z_i z_i^\top - I)C(z_j z_j^\top - I)](\Psi + D_j)^{1/2} + O(m^{-3/2}),
$$

for $C = (\Psi + D_i)^{1/2}(\Psi + D_a)^{-1}(\Psi + D_j)^{1/2}$. For $i \neq j$, $E[(z_i z_i^\top - I)C(z_j z_j^\top - I)] = 0$, we have

$$
\sum_{i=1}^{m} \sum_{j=1}^{m} E[(z_i z_i^\top - I)C(z_j z_j^\top - I)] = \sum_{i=1}^{m} E[z_i z_i^\top C z_i z_i^\top - C].
$$

Because $E[z_i z_i^\top C z_i z_i^\top] = 2C + (\text{tr} C) I_k$, it is concluded that

$$
D_a(\Psi + D_a)^{-1}E[(\hat{\Psi} - \Psi)(\Psi + D_a)^{-1}(\hat{\Psi} - \Psi)](\Psi + D_a)^{-1}D_a = G_{3a}(\Psi) + O(m^{-3/2}).
$$

The above arguments imply that a second-order unbiased estimator of $G_{1a}(\Psi)$ is $G_{1a}(\hat{\Psi}^+) + G_{3a}(\hat{\Psi}^+) + G_{4a}(\hat{\Psi}^+)$. The estimators $G_{2a}(\hat{\Psi}^+)$ and $G_{3a}(\hat{\Psi}^+)$ do not have second-order biases, and the results in Theorem 3 are established. \qed

5 Simulation and Empirical Studies

5.1 Finite sample performances

We now investigate finite sample performances of EBLUP in terms of MSEM and the second-order unbiased estimator of MSEM by simulation.

[1] Setup of simulation experiments. We treat the multivariate Fay-Herriot model (1) for $k = 2, 3$ and $m = 30, 60$ without covariates, namely $X_i = I_k$. As a setup of the covariance matrix $\Psi$ of the random effects, we consider

$$
\Psi = \begin{cases}
\rho \psi_2 \psi_2^\top + (1 - \rho) \text{diag}(\psi_2) & \text{for } k = 2, \\
\rho \psi_3 \psi_3^\top + (1 - \rho) \text{diag}(\psi_3) & \text{for } k = 3,
\end{cases}
$$

where $\psi_2 = (\sqrt{1.5}, \sqrt{0.5})^\top$, $\psi_3 = (\sqrt{1.5}, 1, \sqrt{0.5})^\top$, and $\text{diag}(A)$ denotes the diagonal matrix consisting of diagonal elements of matrix $A$. Here, $\rho$ is the correlation coefficient, and we handle the three cases $\rho = 0.25, 0.5, 0.75$. The cases of negative correlations are omitted, because we observe the same results with those of positive ones.

Concerning the dispersion matrices $D_i$ of sampling errors $\varepsilon_i$, we treat the two $D_i$-patterns: (a) $0.7I_k, 0.6I_k, 0.5I_k, 0.4I_k, 0.3I_k$ and (b) $2.0I_k, 0.6I_k, 0.5I_k, 0.4I_k, 0.2I_k$. In the univariate Fay-Herriot model, these cases are treated by Datta, et al. (2005). There are five groups $G_1, \ldots, G_5$. 

11
corresponding to these $D_i$-patterns, and there are six and twelve small areas in each group for $m = 30$ and $60$, respectively, where the sampling covariance matrices $D_i$ are the same for areas within the same group.

[2] Comparison of MSEM. We begin with obtaining the true mean squared error matrices of the EBLUP $\hat{\theta}_a^{EB} = \hat{\theta}_a(\hat{\Psi}^{+})$ by simulation. Let $\{y_i^{(r)}, i = 1, \ldots, m\}$ be the simulated data in the $r$-th replication for $r = 1, \ldots, R$ with $R = 50,000$. Let $\hat{\Psi}^{+}$ and $\theta_a^{(r)}$ be the values of $\hat{\Psi}^{+}$ and $\theta_a$ in the $r$-th replication. Then the simulated value of the true mean squared error matrices is calculated by

$$\text{MSEM}(\hat{\theta}_a^{EB}) = R^{-1} \sum_{i=1}^{R} \{ \hat{\theta}_a(\hat{\Psi}^{+(r)}) - \theta_a^{(r)} \} \{ \hat{\theta}_a(\hat{\Psi}^{+(r)}) - \theta_a^{(r)} \}^\top.$$

As an estimator of $\Psi$, we here use the simple estimator $\hat{\Psi}_0^{+}$, because there is little difference between $\hat{\Psi}_0^{+}$ and $\hat{\Psi}_1^{+}$ in simulated values of MSEM under the setup of $X_i = I_k$. Simulated values of the mean squared error matrices, averaged over areas within groups $G_i$, are reported in Tables 1, 3, and 5. To measure relative improvement of EBLUP, we calculate the percentage relative improvement in the average loss (PRIAL) of $\hat{\theta}_a^{EB}$ over $y_a$, defined by

$$\text{PRIAL}(\hat{\theta}_a^{EB}, y_a) = 100 \times \left[ 1 - \frac{\text{tr} \{ \text{MSEM}(\hat{\theta}_a^{EB}) \} }{\text{tr} \{ \text{MSEM}(y_a) \} } \right].$$

It is also interesting to compare $\hat{\theta}_a^{EB}$ with the EBLUP $\hat{\theta}_a^{uEB}$ derived from the univariate Fay-Herriot model. Thus, we calculate the PRIAL given by

$$\text{PRIAL}(\hat{\theta}_a^{EB}, \hat{\theta}_a^{uEB}) = 100 \times \left[ 1 - \frac{\text{tr} \{ \text{MSEM}(\hat{\theta}_a^{EB}) \} }{\text{tr} \{ \text{MSEM}(\hat{\theta}_a^{uEB}) \} } \right],$$

and those values are reported in Tables 2, 4, and 6.

Table 1 reports the simulated values of the true MSEM of $\hat{\theta}_a^{EB}$ for $k = 2$, $D_i$-patterns (a), $m = 30, 60$ and $\rho = 0.25, 0.5, 0.75$. For fixed $m$, the values of MSEM decrease as the correlation $\rho$ in the random effect becomes large. For fixed $\rho$, the values of MSEM decrease as $m$ becomes large. Table 2 reports the values of PRIAL of $\hat{\theta}_a^{EB}$ over $y_a$ and $\hat{\theta}_a^{uEB}$ under the same setup as in Table 1. In all the cases, $\hat{\theta}_a^{EB}$ improves on $y_a$ largely and the improvement rates are larger for larger $\rho$. In comparison with $\hat{\theta}_a^{uEB}$, the univariate EBLUP $\hat{\theta}_a^{EB}$ is slightly better than $\hat{\theta}_a^{EB}$ for $\rho = 0.25$, but the difference is not significant. The values of PRIAL of $\hat{\theta}_a^{EB}$ over $\hat{\theta}_a^{uEB}$ get larger as $\rho$ increases. In the case of $m = 60$, the improvements of $\hat{\theta}_a^{EB}$ in light of PRIAL get larger for larger $\rho$. In the case of $\rho = 0.25$, the improvement of $\hat{\theta}_a^{EB}$ over $\hat{\theta}_a^{uEB}$ is better for $m = 60$ than for $m = 30$. This is because the low accuracy in estimation of the covariance matrix $\Psi$ has more adverse influence on prediction than the benefit from incorporating the small correlation into the estimation.

The comparison of performances between $D_i$-patterns (a) and (b) is investigated in Tables 3 and 4. The simulated values of the MSEM of $\hat{\theta}_a^{EB}$ in $D_i$-patterns (a) and (b) are reported in
Table 3 for $k = 2$, $m = 30, 60$ and $\rho = 0.5$. As the increment of variance of sampling error in $G_1$, the MSEM in $G_1$ becomes larger, and the other groups have slightly larger MSEM except $G_5$. The values of PRIAL of $\hat{\theta}_{a}^{EB}$ over $y_a$ and $\hat{\theta}_{a}^{uEB}$ are given in Table 4 for $D_i$-patterns (a) and (b). under the same setup as in Table 3. As seen from the table, the improvement of $\hat{\theta}_{a}^{EB}$ over $y_a$ in $G_1$ is larger for $D_i$-pattern (b) because of the large sampling variance. However, $\hat{\theta}_{a}^{EB}$ is not better than $\hat{\theta}_{a}^{uEB}$ in $G_1$ and $G_5$ for $m = 30$ and in $G_5$ for $m = 60$ in $D_i$-pattern (b). This implies that incorporating the information of areas with large sampling variances affects more adversely estimation of areas with small sampling variances in the multivariate model than in the univariate model.

Tables 5 and 6 report the values of MSEM and PRIAL for $k = 3$, $m = 30, 60$ and $D_i$-pattern (a). From Table 6, it is revealed that PRIAL of $\hat{\theta}_{a}^{EB}$ over $y_a$ and $\hat{\theta}_{a}^{uEB}$ are larger for $k = 3$ than for $k = 2$ in the case of $\rho = 0.75$, but smaller in the case of $\rho = 0.25$. When $m$ is fixed as $m = 30$, the accuracy in estimation of the covariance matrix $\Psi$ gets smaller for the larger dimension. This demonstrates that it is not appropriate to treat the multivariate Fay-Herriot model with a large covariance matrix when $m$ is not large.

[3] MSEM approximation and its estimator. We next investigate the performance of the second-order approximation of MSEM of EBLUP $\hat{\theta}_{a}^{EB}$ given in Theorem 2 and the second-order unbiased estimator $\text{mse}(\hat{\theta}_{a}^{EB})$ of MSEN given in Theorem 3. The values of the second-order approximation of MSEM are given in Table 7 for $k = 2$, $m = 3$ and $D_i$-pattern (a). Comparing the values in Table 7 with the corresponding true values of the MSEM in Table 1, we can see that the second-order approximation can approximate the true MSEM precisely for every $G_t$ and $\rho$.

Concerning the performance of the second-order unbiased estimator $\text{mse}(\hat{\theta}_{a}^{EB})$ given in (22), we compute the simulated values of relative bias of the estimator $\text{mse}(\hat{\theta}_{a}^{EB})$, averaged over areas within groups $G_t$. Those values are reported in Table 8 for $k = 2$, $m = 30, 60$ and $D_i$-pattern (a). It is revealed from Table 8 that the relative bias gets larger for larger $\rho$. Also, the values of the relative bias are smaller for $m = 60$ than for $m = 30$, namely, the relative bias gets small as $m$ increases.

5.2 Illustrative example

This example, primarily for illustration, uses the multivariate Fay-Herriot model (1) and data from the 2016 Survey of Family Income and Expenditure in Japan, which is based on two or more person households (excluding agricultural, forestry and fisheries households). The target domains are the 47 Japanese prefectural capitals. The 47 prefectures are divided into 10 regions: Hokkaido, Tohoku, Kanto, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, Kyushu and Okinawa. Each region consists of several prefectures except Hokkaido and Okinawa, which consist of one prefecture.

In this study, as observations $(y_{i1}, y_{i2})$', we use the reported data of the yearly averaged monthly spendings on ‘Education’ and ‘Cultural-amusement’ per worker’s household, scaled by 1,000 Yen, at each capital city of 47 prefectures. In addition, we use the data in the 2014 National Survey of Family Income and Expenditure. The average spending data in this survey
Table 1: Simulated values of mean squared error matrices of $\hat{\theta}_a^{EB}$ multiplied by 100 for $k = 2$, $D_t$-patterns (a)

|   | $m = 30$ | $m = 60$ |
|---|---------|---------|
|   | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ |
| $G_1$ | $[49.8, 3.8]$ | $[48.7, 8.1]$ | $[46.5, 13.8]$ | $[49.0, 4.1]$ | $[47.4, 8.2]$ | $[45.2, 14.0]$ |
| $G_2$ | $[3.8, 32.6]$ | $[8.1, 30.1]$ | $[13.8, 25.3]$ | $[4.1, 30.7]$ | $[8.2, 28.0]$ | $[14.0, 23.6]$ |
| $G_3$ | $[44.7, 3.1]$ | $[43.8, 6.5]$ | $[41.4, 11.6]$ | $[43.5, 3.4]$ | $[42.5, 7.0]$ | $[40.3, 11.7]$ |
| $G_4$ | $[3.1, 30.4]$ | $[6.5, 28.3]$ | $[11.6, 23.7]$ | $[4.4, 29.3]$ | $[7.0, 26.5]$ | $[11.7, 22.1]$ |
| $G_5$ | $[49.8, 3.8]$ | $[48.7, 8.1]$ | $[46.5, 13.8]$ | $[49.0, 4.1]$ | $[47.4, 8.2]$ | $[45.2, 14.0]$ |

Table 2: PRIAL of $\hat{\theta}_a^{EB}$ over $y_a$ and $\hat{\theta}_a^{EB}$ for $k = 2$, $D_t$-patterns (a)

|   | $\hat{\theta}_a^{EB}$ vs $y_a$ | $\hat{\theta}_a^{EB}$ vs $\hat{\theta}_a^{EB}$ |
|---|--------------------------------|--------------------------------|
|   | $m = 30$ | $m = 60$ |
| $G_1$ | $[41.2, 43.8]$ | $[48.9, -0.5]$ | $[3.8, 11.6]$ | $[43.2, 45.8]$ | $[51.0, -0.6]$ | $[4.6, 13.9]$ |
| $G_2$ | $[37.2, 40.1]$ | $[45.7, 0.0]$ | $[3.5, 12.3]$ | $[39.8, 42.4]$ | $[48.1, 0.2]$ | $[5.1, 13.6]$ |
| $G_3$ | $[33.0, 35.8]$ | $[41.8, -0.7]$ | $[3.4, 11.8]$ | $[35.6, 38.7]$ | $[44.2, 1.3]$ | $[4.6, 14.3]$ |
| $G_4$ | $[27.3, 29.8]$ | $[37.2, -1.9]$ | $[1.8, 11.0]$ | $[30.6, 33.7]$ | $[39.8, 0.3]$ | $[3.5, 13.2]$ |
| $G_5$ | $[20.8, 23.5]$ | $[30.4, -2.5]$ | $[1.1, 10.0]$ | $[24.8, 27.5]$ | $[33.6, 0.4]$ | $[2.9, 11.0]$ |
Table 3: Simulated values of mean squared error matrices of $\hat{\theta}_a^{EB}$ multiplied by 100 for $k = 2$, $\rho = 0.5$

|      | $m = 30$ |          | $m = 60$ |          |
|------|----------|----------|----------|----------|
|      | Pattern (a) | Pattern (b) | Pattern (a) | Pattern (b) |
| $G_1$ | 48.7 8.1 | 89.9 19.7 | 47.4 8.2 | 86.8 20.1 |
| $G_2$ | 8.1 30.1 | 19.7 42.9 | 8.2 28.0 | 20.1 40.0 |
| $G_3$ | 43.8 6.5 | 44.5 6.0 | 42.5 7.0 | 42.9 6.5 |
| $G_4$ | 6.5 28.3 | 6.0 30.2 | 7.0 26.5 | 6.5 27.8 |
| $G_5$ | 38.0 5.3 | 39.3 4.7 | 37.1 5.7 | 37.8 5.0 |

Table 4: PRIAL of $\hat{\theta}_a^{EB}$ over $y_a$ and $\hat{\theta}_a^{uEB}$ for $k = 2$, $m = 30, 60$, $\rho = 0.5$, $D_i$-patterns (a), (b)

|      | $m = 30$ |          | $m = 60$ |          |
|------|----------|----------|----------|----------|
|      | Pattern (a) | Pattern (b) | Pattern (a) | Pattern (b) |
| $G_1$ | 43.8 66.4 | 3.8 2.1 | 4.2 2.7 | 18.1 0.6 |
| $G_2$ | 40.1 37.0 | 3.5 0.8 | 4.1 21.8 | 3.6 23.8 |
| $G_3$ | 35.8 32.1 | 4.6 2.7 | 3.8 25.9 | 4.1 18.7 |
| $G_4$ | 29.8 26.2 | 1.8 -0.2 | 3.2 25.9 | 2.7 18.7 |
| $G_5$ | 23.5 4.2 | 1.1 -8.5 | 0.1 18.8 | 0.6 16.4 |

|      | $m = 30$ |          | $m = 60$ |          |
|------|----------|----------|----------|----------|
|      | Pattern (a) | Pattern (b) | Pattern (a) | Pattern (b) |
| $G_1$ | 45.8 68.5 | 4.6 3.1 | 4.2 2.7 | 3.5 1.3 |
| $G_2$ | 42.4 40.7 | 5.1 3.1 | 36.2 4.6 | 2.7 1.3 |
| $G_3$ | 38.7 36.2 | 4.6 2.7 | 30.6 4.6 | 2.9 1.3 |
| $G_4$ | 33.7 30.6 | 3.5 1.3 | 30.6 4.6 | 2.9 1.3 |
| $G_5$ | 27.5 13.9 | 2.9 -2.7 | 19.1 0.1 | 0.6 16.4 |
Table 5: Simulated values of mean squared error matrices of $\hat{\theta}^{EB}_{a}$ multiplied by 100 for $k = 3$, $m = 30$, $D_r$-patterns (a)

|       | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ |
|-------|---------------|---------------|---------------|
|       | $\begin{bmatrix} 50.0 & 3.4 & 3.5 \\ 3.5 & 3.4 & 33.3 \end{bmatrix}$ | $\begin{bmatrix} 48.0 & 7.0 & 6.4 \\ 6.4 & 6.5 & 29.5 \end{bmatrix}$ | $\begin{bmatrix} 42.0 & 12.3 & 10.0 \\ 10.0 & 9.4 & 23.1 \end{bmatrix}$ |
| G1    | $\begin{bmatrix} 45.2 & 2.6 & 2.8 \\ 2.8 & 2.9 & 31.2 \end{bmatrix}$ | $\begin{bmatrix} 42.8 & 5.8 & 5.4 \\ 5.4 & 5.5 & 28.2 \end{bmatrix}$ | $\begin{bmatrix} 38.2 & 10.0 & 8.3 \\ 8.3 & 8.0 & 21.7 \end{bmatrix}$ |
| G2    | $\begin{bmatrix} 19.0 & 6.1 & 2.1 \\ 1.9 & 2.1 & 29.0 \end{bmatrix}$ | $\begin{bmatrix} 32.1 & 2.7 & 2.9 \\ 3.9 & 4.1 & 25.8 \end{bmatrix}$ | $\begin{bmatrix} 29.2 & 5.6 & 5.1 \\ 7.0 & 6.4 & 20.5 \end{bmatrix}$ |
| G3    | $\begin{bmatrix} 33.4 & 1.3 & 1.5 \\ 1.5 & 1.6 & 26.0 \end{bmatrix}$ | $\begin{bmatrix} 2.7 & 29.2 & 3.0 \\ 2.9 & 3.9 & 20.7 \end{bmatrix}$ | $\begin{bmatrix} 5.1 & 5.1 & 18.4 \\ 5.6 & 25.2 & 5.1 \end{bmatrix}$ |
| G4    | $\begin{bmatrix} 1.3 & 31.0 & 1.6 \\ 1.3 & 0.7 & 26.0 \end{bmatrix}$ | $\begin{bmatrix} 2.7 & 29.2 & 3.0 \\ 2.9 & 3.9 & 20.7 \end{bmatrix}$ | $\begin{bmatrix} 5.1 & 5.1 & 18.4 \\ 5.6 & 25.2 & 5.1 \end{bmatrix}$ |
| G5    | $\begin{bmatrix} 0.7 & 25.4 & 1.0 \\ 0.7 & 1.0 & 22.7 \end{bmatrix}$ | $\begin{bmatrix} 1.6 & 24.1 & 1.8 \\ 1.5 & 1.8 & 20.7 \end{bmatrix}$ | $\begin{bmatrix} 3.1 & 21.0 & 3.4 \\ 3.2 & 3.4 & 16.5 \end{bmatrix}$ |

Table 6: PRIAL of $\hat{\theta}^{EB}_{a}$ over $y_a$ and $\hat{\theta}^{vEB}_{a}$ for $k = 2, 3$, $m = 30$, $D_r$-patterns (a)

|       | $\hat{\theta}^{EB}_{a}$ vs $y_a$ | $\hat{\theta}^{EB}_{a}$ vs $\hat{\theta}^{vEB}_{a}$ | $\hat{\theta}^{EB}_{a}$ vs $\hat{\theta}^{vEB}_{a}$ |
|-------|-----------------------------|-----------------------------|-----------------------------|
| $k = 2$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ |
|       | $\begin{bmatrix} 41.2 & 43.8 \\ 37.2 & 40.1 \end{bmatrix}$ | $\begin{bmatrix} 48.9 & -0.5 \\ 45.7 & 0.0 \end{bmatrix}$ | $\begin{bmatrix} 3.8 & 11.6 \\ 3.5 & 12.3 \end{bmatrix}$ | $\begin{bmatrix} 11.6 & 3.8 \\ 12.3 & 3.5 \end{bmatrix}$ | $\begin{bmatrix} 11.6 & 3.8 \\ 12.3 & 3.5 \end{bmatrix}$ | $\begin{bmatrix} 11.6 & 3.8 \\ 12.3 & 3.5 \end{bmatrix}$ |
| $G_1$  | $\begin{bmatrix} 33.0 & 35.8 \\ 27.3 & 29.8 \end{bmatrix}$ | $\begin{bmatrix} 41.8 & -0.7 \\ 37.2 & -1.9 \end{bmatrix}$ | $\begin{bmatrix} 3.4 & 11.8 \\ 1.8 & 11.0 \end{bmatrix}$ | $\begin{bmatrix} 11.8 & 3.4 \\ 11.0 & 1.8 \end{bmatrix}$ | $\begin{bmatrix} 11.8 & 3.4 \\ 11.0 & 1.8 \end{bmatrix}$ | $\begin{bmatrix} 11.8 & 3.4 \\ 11.0 & 1.8 \end{bmatrix}$ |
| $G_2$  | $\begin{bmatrix} 20.8 & 23.5 \\ 29.8 & 30.4 \end{bmatrix}$ | $\begin{bmatrix} 23.5 & -2.5 \\ 30.4 & -2.5 \end{bmatrix}$ | $\begin{bmatrix} 1.1 & 10.0 \\ 1.1 & 10.0 \end{bmatrix}$ | $\begin{bmatrix} 10.0 & 1.1 \\ 10.0 & 1.1 \end{bmatrix}$ | $\begin{bmatrix} 10.0 & 1.1 \\ 10.0 & 1.1 \end{bmatrix}$ | $\begin{bmatrix} 10.0 & 1.1 \\ 10.0 & 1.1 \end{bmatrix}$ |
| $k = 3$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ |
|       | $\begin{bmatrix} 39.5 & 43.6 \\ 35.2 & 40.0 \end{bmatrix}$ | $\begin{bmatrix} 52.4 & -1.9 \\ 48.9 & -2.6 \end{bmatrix}$ | $\begin{bmatrix} 5.9 & 20.3 \\ 4.8 & 19.2 \end{bmatrix}$ | $\begin{bmatrix} 20.3 & 5.9 \\ 19.2 & 4.8 \end{bmatrix}$ | $\begin{bmatrix} 20.3 & 5.9 \\ 19.2 & 4.8 \end{bmatrix}$ | $\begin{bmatrix} 20.3 & 5.9 \\ 19.2 & 4.8 \end{bmatrix}$ |
| $G_1$  | $\begin{bmatrix} 30.3 & 35.1 \\ 25.2 & 29.7 \end{bmatrix}$ | $\begin{bmatrix} 44.6 & -3.9 \\ 39.6 & -4.4 \end{bmatrix}$ | $\begin{bmatrix} 3.2 & 18.2 \\ 1.9 & 15.7 \end{bmatrix}$ | $\begin{bmatrix} 18.2 & 3.2 \\ 15.7 & 1.9 \end{bmatrix}$ | $\begin{bmatrix} 18.2 & 3.2 \\ 15.7 & 1.9 \end{bmatrix}$ | $\begin{bmatrix} 18.2 & 3.2 \\ 15.7 & 1.9 \end{bmatrix}$ |
| $G_2$  | $\begin{bmatrix} 17.6 & 21.6 \\ 25.2 & 29.7 \end{bmatrix}$ | $\begin{bmatrix} 32.0 & -5.7 \\ 39.6 & -4.4 \end{bmatrix}$ | $\begin{bmatrix} -0.3 & 12.8 \\ -0.3 & 12.8 \end{bmatrix}$ | $\begin{bmatrix} 12.8 & -0.3 \\ 12.8 & -0.3 \end{bmatrix}$ | $\begin{bmatrix} 12.8 & -0.3 \\ 12.8 & -0.3 \end{bmatrix}$ | $\begin{bmatrix} 12.8 & -0.3 \\ 12.8 & -0.3 \end{bmatrix}$ |
Table 7: Second order approximations of mean squared error matrices of $\hat{\theta}_a$ multiplied by 100 for $k = 2$, $D_i$-patterns

|       | $m = 30$        |           |           |
|-------|----------------|-----------|-----------|
|       | $\rho = 0.25$  | $\rho = 0.5$ | $\rho = 0.75$ |
| $G_1$ | 49.8 3.7       | 48.6 7.9  | 46.2 13.2 |
| $G_2$ | 3.7 32.6       | 7.9 30.3  | 13.2 25.9 |
| $G_3$ | 44.6 3.1       | 43.6 6.6  | 41.5 11.1 |
| $G_4$ | 3.1 30.4       | 6.6 28.4  | 11.1 24.4 |
| $G_5$ | 2.4 27.8       | 5.2 26.1  | 8.9 22.6  |

Table 8: Simulated values of percentage average relative bias of mean squared error matrices of $\hat{\theta}_a$ multiplied by 100 for $k = 2$, $m = 30, 60$, $D_i$-pattern (a)

|               | $m = 30$        |           |           |
|---------------|----------------|-----------|-----------|
|               | $\rho = 0.25$  | $\rho = 0.5$ | $\rho = 0.75$ |
| $G_1$         | -0.3 -2.6      | -0.9 -4.1 | 0.6 -9.5  |
| $G_2$         | -2.6 1.1       | -4.1 2.9  | -9.5 10.1 |
| $G_3$         | 0.6 3.1        | 0.3 -3.5  | 1.1 -10.4 |
| $G_4$         | -3.5 2.7       | 1.3 -7.8  | -10.4 13.1|
| $G_5$         | -5.8 1.2       | -7.8 4.6  | -16.7 13.6|
|               | -0.6 -5.8      | 1.3 -7.8  | 1.2 -16.7 |
|               | -5.8 1.2       | -7.8 4.6  | -16.7 13.6|
|               | -0.4 -4.6      | 0.4 -10.8 | 1.2 -23.4 |
|               | -4.6 2.9       | -10.8 4.7 | -23.4 17.8|
| $G_5$         | 0.3 -24.4      | 0.6 -26.1 | 3.4 -42.3 |
|               | -24.4 2.2      | -26.1 7.7 | -42.3 23.1|

|               | $m = 60$        |           |           |
|---------------|----------------|-----------|-----------|
|               | $\rho = 0.25$  | $\rho = 0.5$ | $\rho = 0.75$ |
| $G_1$         | -0.1 -2.3      | 0.2 -0.2  | 0.4 -0.7  |
| $G_2$         | -2.3 -0.5      | -0.2 -0.5 | -0.7 1.7  |
| $G_3$         | 0.7 -3.4       | 0.4 -0.8  | 0.8 -0.0  |
| $G_4$         | -3.4 -0.2      | -0.8 -0.5 | -0.0 2.1  |
| $G_5$         | 0.2 -5.1       | 0.1 -1.5  | 0.3 -1.4  |
|               | -5.1 -0.2      | -1.5 0.1  | -1.4 2.9  |
|               | -0.1 -5.1      | -0.4 -2.4 | 1.4 -2.9  |
|               | -5.1 -0.4      | -2.4 0.3  | -2.9 3.6  |
| $G_5$         | 0.2 -3.3       | 0.3 -5.9  | 0.6 -8.1  |
|               | -3.3 -0.2      | -5.9 -0.1 | -8.1 5.1  |
are more reliable than the Survey of Family Income and Expenditure since the sample sizes are much larger. However, this survey is conducted only once in every five years. As auxiliary variables, we use the data of the average spendings on ‘Education’ and ‘Cultural-amusement’, which is denoted by EDU\_i and CUL\_i, respectively. Then the regressor in the model (1) is
\[ X_i = \begin{pmatrix} 1 & \text{EDU\_i} & 0 & 0 \\ 0 & 0 & 1 & \text{CUL\_i} \end{pmatrix}. \]

Then we apply the multivariate Fay-Herriot model (1), where sampling covariance matrices \( D_i \) of the \( i \)-th region for \( i = 1, \ldots, 10 \) are calculated based on data of yearly averaged monthly spendings on ‘Education’ and ‘Cultural-amusement’ in the past ten years (2006-2015), where \( D_i \) is given as the average of the sampling covariance matrices of prefectures within the \( i \)-th region. That is, the sampling covariance matrix \( D_i \) are the same for prefectures within the same region.

The estimates of the covariance matrix \( \Psi \) and the correlation coefficient \( \rho \) is
\[ \hat{\Psi} = \begin{pmatrix} 8.5 & 3.0 \\ 3.0 & 10.2 \end{pmatrix} \quad \text{and} \quad \hat{\rho} = 0.32. \]

The estimates of the regression coefficients and the \( p \)-values for testing \( H_0 : \beta_k = 0 \) for \( k = 1, \ldots, 4 \) are given in Table 9. All the estimates are significant.

| variables | Constant(EDU) | EDU | Constant(CUL) | CUL |
|-----------|---------------|-----|---------------|-----|
| \( \beta \) | 4.47          | 0.82| 12.12         | 0.65|
| \( p \)-value | 0.007         | 0.000| 0.002         | 0.000|

The values of EBLUP and direct estimate of spendings on ‘Education’ and ‘Cultural-amusement’ are reported in Table 10. We only pick up the three prefectures from three different regions: Tokyo prefecture from the Kanto region, Osaka prefecture from the Kinki region and Fukushima prefecture from the Tohoku region, whose sampling covariance matrices are
\[ \begin{pmatrix} 1.1 & 0.3 \\ 0.3 & 3.0 \end{pmatrix}, \quad \begin{pmatrix} 1.1 & -0.2 \\ -0.2 & 3.9 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 4.7 & 3.5 \\ 3.5 & 4.9 \end{pmatrix}, \]
respectively. It is seen that as the sampling variances become larger, the direct estimates are more shrunken by the EBLUP in the sense of \((\text{direct estimate} - \text{EBLUP})/\text{(direct estimate)}\).

The uncertainty of EBLUP is provided by the second-order unbiased estimator of MSEM of EBLUP. Table 11 reports the estimates of MSEM averaged over prefectures within each region for 10 regions. We also calculate the percentage relative improvement in the average loss estimate (PRIAL estimate) of \( \hat{\theta}_a^{EB} \) over \( y_a \) and \( \hat{\theta}_a^{uEB} \). Table 12 reports the average of those values over each region for spendings on education and cultural-amusement. It is revealed from Table 12 that the multivariate EBLUP improves on the direct estimates significantly and that the multivariate EBLUP is slightly better than the univariate EBLUP for most regions except Okinawa, which has a smaller sampling covariance matrix.
Table 10: EBLUP and direct estimates

|                | Tokyo    | Osaka   | Fukushima |
|----------------|----------|---------|-----------|
| direct estimator (EDU) | 32.5     | 19.0    | 13.3      |
| EBLUP (EDU)     | 31.8     | 19.4    | 12.6      |
| direct estimator (CUL) | 41.8     | 24.9    | 29.9      |
| EBLUP (CUL)     | 40.6     | 26.1    | 29.0      |

Table 11: Estimates of the mean squared error matrices of $\hat{\theta}_a^{EB}

\[
\begin{bmatrix}
0.5 & 0.7 \\
0.7 & 3.8 \\
\end{bmatrix},
\begin{bmatrix}
3.2 & 2.2 \\
2.2 & 3.4 \\
\end{bmatrix},
\begin{bmatrix}
1.0 & 0.3 \\
0.3 & 2.5 \\
\end{bmatrix},
\begin{bmatrix}
1.0 & 0.6 \\
0.6 & 4.7 \\
\end{bmatrix},
\begin{bmatrix}
1.4 & 0.6 \\
0.6 & 1.8 \\
\end{bmatrix},
\begin{bmatrix}
0.5 & 0.3 \\
0.3 & 2.6 \\
\end{bmatrix},
\begin{bmatrix}
0.9 & 3.5 \\
0.9 & 3.5 \\
\end{bmatrix},
\begin{bmatrix}
0.7 & 1.8 \\
0.7 & 1.8 \\
\end{bmatrix},
\begin{bmatrix}
0.8 & 1.7 \\
0.8 & 1.7 \\
\end{bmatrix},
\begin{bmatrix}
1.0 & -0.0 \\
-0.0 & 2.8 \\
\end{bmatrix},
\begin{bmatrix}
1.0 & 0.3 \\
1.5 & 0.3 \\
\end{bmatrix},
\begin{bmatrix}
4.2 & 0.9 \\
0.9 & 3.5 \\
\end{bmatrix},
\begin{bmatrix}
1.0 & 0.7 \\
1.0 & 0.7 \\
\end{bmatrix},
\begin{bmatrix}
3.0 & 0.8 \\
3.0 & 0.8 \\
\end{bmatrix}
\end{bmatrix}

Table 12: PRIAL estimates of $\hat{\theta}_a^{EB}$ over $y_a$ and $\hat{\theta}_a^{uEB}$

|                | Hokkaido | Tohoku | Kanto | Hokuriku | Tokai | Kinki | Chugoku | Shikoku | Kyushu | Okinawa |
|----------------|----------|--------|-------|----------|-------|-------|---------|---------|--------|---------|
| $\hat{\theta}_a^{EB}$ vs $y_a$ | 82.4 | 84.7 | 80.9 | 84.1 | 80.5 | 81.5 | 81.6 | 85.7 | 80.1 | 81.0 |
| $\hat{\theta}_a^{EB}$ vs $\hat{\theta}_a^{uEB}$ | 4.1 | 7.1 | 1.9 | 5.3 | 1.1 | 4.5 | 2.4 | 3.9 | 1.3 | -33.9 |
Acknowledgments

Research of the second author was supported in part by Grant-in-Aid for Scientific Research (15H01943 and 26330036) from Japan Society for the Promotion of Science.

References

[1] G.E. Battese, R.M. Harter, W.A. Fuller, An error component model for prediction of mean crop areas using survey and satellite data., J. Am. Statist. Assoc. 95 (1988) 28–36.

[2] R. Benavent, D. Morales, Multivariate Fay-Herriot models for small area estimation, Comp. Statist. Data Anal. 94 (2016) 372–390.

[3] G.S. Datta, B. Day, I.V. Basawa, Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation, J. Statist. Plan. Inf. 75 (1999) 269–279.

[4] G.S. Datta, B. Day, T. Maiti, Multivariate Bayesian small area estimation: An application to survey and satellite data, Sankhya 60 (1998) 344–362.

[5] G.S. Datta, M. Ghosh, Small area shrinkage estimation, Statist. Science 27 (2012) 95–114.

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

[7] B. Efron, C. Morris, Multivariate empirical Bayes estimation of covariance matrices, Ann. Statist. 4 (1976) 22–32.

[8] R. Fay, Application of multivariate regression to small domain estimation, Small Area Statistics (R. Platek, J.N.K. Rao, C.E. Sarndal, M.P. Singh, eds) (1987) 91–102.

[9] R. Fay, R. Herriot, Estimators of income for small area places: an application of James–Stein procedures to census, J. Amer. Statist. Assoc. 74 (1979) 341–353.

[10] W.A. Fuller, R.M. Harter, The multivariate components of variance model for small area estimation, Small Area Statistics (R. Platek, J.N.K. Rao, C.E. Sarndal, M.P. Singh, eds) (1987) 103–137.

[11] M. Ghosh, J.N.K. Rao, Small area estimation: An appraisal, Statist. Science 9 (1994) 55–93.

[12] I. Ngaruye, D. von Rosen, M. Singull, Crop yield estimation at district level for agricultural seasons 2014 in Rwanda, African J. Applied Statist. 3 (2016) 69-90.

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

[14] A.T. Porter, C.K. Wikle, S.H. Holan, Small area estimation via multivariate Fay-Herriot models with latent spatial dependence, Australian and New Zealand J. Statist. 57 (2015) 15–29.

[15] N.G.N. Prasad, J.N.K. Rao, The estimation of mean squared errors of small area estimators, J. Am. Statist. Assoc. 85 (1990) 163–171.
[16] J.N.K. Rao, I. Molina, Small Area Estimation, 2nd Edition (2015) Wiley.