On small area estimation under a sub-area level model

Mahmoud Torabi[a], J.N.K. Rao[b]

[a] *Department of Community Health Sciences, University of Manitoba, MB, R3E 0W3, Canada*
[b] *School of Mathematics and Statistics, Carleton University, Ottawa, ON, K1S 5B6, Canada*

Abstract

We propose an extension of the well-known Fay and Herriot [6] area level model to sub-area level. Not only this model may be used to estimate small area means by borrowing strength from related areas, but also by borrowing strength from sub-areas to obtain more efficient sub-area estimators. Model-based empirical best linear unbiased prediction (EBLUP) estimators are obtained from the BLUP estimators by replacing the model parameters by suitable estimators, using an iterative method based on weighted residual sum of squares. Second order approximations to the mean squared error (MSE) of the EBLUP estimators are obtained and then used to drive MSE estimators unbiased to second order. Results of simulation studies on the performance of the proposed estimators are also provided.

Keywords: Best linear unbiased prediction; Fay-Herriot model; Linear mixed models; Mean squared error; Variance components

Email address: `torabi@cc.umanitoba.ca; Fax: +1-204-789-3905` (Mahmoud Torabi)

1. Introduction

Sample surveys are generally designed to provide estimates of totals and means of items of interest for large subpopulations (or domains). Such estimates are "direct" in the sense of using only the domain-specific sample data, and the domain sample sizes are large enough to support reliable direct estimates that are "design based". The associated inferences (standard errors, confidence intervals, etc.) are based on the probability distribution induced by the sampling design with the population item values held fixed. Standard text books on sampling (e.g. Cochran [2], Thompson [19], Lohr [15]) provide extensive accounts of design-based direct estimation.

In recent years, demand for reliable estimates for small domains (small areas) has greatly increased worldwide due to their growing use in formulating policies and programmes, allocation of government funds, regional planning, marketing decisions at local level and other uses. Examples of small domain estimation include poverty counts of school-age children at the county level, income for small places, monthly unemployment rates for Census Metropolitan Areas, health-related estimates for local areas and so on (Rao [17], chapter 5). However, due to cost and operational considerations, it is seldom possible to procure a large enough overall sample size to support direct estimates for all domains of interest. We use the term "small area" to denote any domain for which direct estimates of adequate precision cannot be produced due to small domain-specific sample size. It is often necessary to employ "indirect" estimates for small areas that can increase the "effective" domain sample size by "borrowing strength" from related areas through linking models, using census and administrative data and other auxiliary data associated with the small areas. Such small area models may be classified into two broad types: (i) Area-level models that relate small area direct estimates to area-specific covariates; such models are used if unit-level data are not available. (ii) Unit-level models that relate the unit values of a study variable to associated unit-level covariates with known area means and area-specific covariates. A comprehensive account of model-based small area estimation under area-level and unit-level models is given by Rao [17]; see also Jiang & Lahiri [12] and Datta [4] for recent overviews.

In this paper, we study model-based estimators for sub-areas nested within areas. We introduce a sub-area level model that relates a sub-area direct estimator to sub-area specific covariates, sub-area random effect and associated area random effect. Such a model is useful if unit level auxiliary variables are not available. The proposed model is a natural extension of the well-known Fay and Herriot [6] area-level model to sub-area level. The sub-area model is used to estimate small area means by borrowing strength from related areas. In addition, it can borrow strength from sub-areas to obtain more efficient sub-area estimators. Empirical best linear unbiased prediction (EBLUP) estimators of sub-area level and area level means are obtained from the BLUP estimators by the model parameters estimate using an iterative method based on weighted residual sum of squares. We obtain second order approximations to the mean squared error (MSE) of the EBLUP estimators and then use them to derive MSE estimators unbiased to second order. Our approximations to MSE and its estimator assume that the number of sampled areas is large but the number of sampled sub-areas within a sampled area can be small. Our paper extends the results of Datta et al. [5] for the area level model to the sub-area level model.

The paper is organized as follows: In Section 2, we introduce the sub-area model and derive EBLUP estimators of area and sub-area means when the variance components σ_v^2 and σ_u^2, corresponding to areas and sub-areas, are estimated iteratively based on a weighted residual sum of squares method. In Section 3, we derive second order approximations to MSE of the EBLUP estimators. In Section 4, estimation of MSEs, unbiased to second order, is studied. Simulation studies, reported in Section 5, provide results on the performance of the proposed estimators.

2. Empirical best linear unbiased prediction

In the context of linear mixed models, we propose the following linking model for the sub-area means μ_{ij}:

$$\mu_{ij} = x'_{ij}\beta + v_i + u_{ij}, \qquad i = 1, ..., m; j = 1, ..., N_i, \tag{2.1}$$

where j denotes a sub-area within area i, x_{ij} is a $p \times 1$ vector of sub-area level auxiliary variables ($m > p$), β is a $p \times 1$ vector of regression parameters, $v_i \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ are area random effects, and $u_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ are sub-area random effects. We assume that n_i sub-areas are sampled from the N_i sub-areas in the *i-th* area.

On the other hand, the sampling model is given by

$$y_{ij} = \mu_{ij} + e_{ij}, \tag{2.2}$$

where y_{ij} is a direct estimator of μ_{ij} with sampling error e_{ij}, and $e_{ij}|\mu_{ij} \overset{ind}{\sim} N(0, \sigma_{eij}^2)$ with known sampling variances σ_{eij}^2. Assuming no sample selection bias, the sampling model (2.2) combined for the level model with the linking model (2.1) leads to the sub-area

$$y_{ij} = x'_{ij}\beta + v_i + u_{ij} + e_{ij}, \qquad i = 1, ..., m; j = 1, ..., n_i. \tag{2.3}$$

Model (2.3) accounts for the sub-area level effect u_{ij} as well as the area level effect v_i. It enables us to estimate both small area means, μ_i, and sub-area means, μ_{ij}, by borrowing strength from related areas as well as sub-areas, where μ_{ij} is given by (2.1) and $\mu_i = \sum_{j=1}^{N_i} N_{ij}\mu_{ij}/N_{i+} = \bar{X}_i'\beta + v_i + \bar{U}_i$ is the mean of area i. Here, $N_{i+} = \sum_{j=1}^{N_i} N_{ij}, \bar{X}_i = \sum_{j=1}^{N_i} N_{ij}x_{ij}/N_{i+}, \bar{U}_i = \sum_{j=1}^{N_i} N_{ij}u_{ij}/N_{i+}$ and N_{ij} is the number of units in sub-area j of area i.

Fuller and Goyeneche [7] proposed a sub-area model, similar to our model (2.3), in the context of Small Area Income and Poverty Estimation (SAIPE) in the United States. In this application, county is the sub-area nested within a state (area) and direct county estimates obtained from the Current Population Survey (CPS) data. County-level auxiliary variables are ascertained from census and administrative records.

In matrix notation, the model (2.3) can be written as

$$\mathrm{y}_i = X_i\beta + v_i 1_{n_i} + u_i + e_i, \qquad i = 1, ..., m,$$

where $\mathrm{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})'$ is a $n_i \times 1$ vector, X_i is a $n_i \times p$ matrix with rows $x'_{ij}, (j = 1, ..., n_i)$, $u_i = (u_{i1}, ..., u_{in_i})'$ and $e_i = (e_{i1}, ..., e_{in_i})'$. Equivalently, we have

$$\mathrm{y}_i = X_i\beta + Z_i b_i + e_i, \qquad i = 1, ..., m, \tag{2.4}$$

where $Z_i = (1_{n_i}|I_{n_i})$ with 1_{n_i} as the vector of ones and I_{n_i} as the identity matrix with dimension n_i, and $b_i = (v_i, u_i')'$. Model (2.4) is a linear mixed model with a block diagonal covariance structure with blocks $cov(\mathrm{y}_i) = V_i$ with

$$V_i = \sigma_v^2 J_{n_i} + diag(\sigma_u^2 + \sigma_{ei1}^2, ..., \sigma_u^2 + \sigma_{ein_i}^2), \tag{2.5}$$

where $J_{n_i} = 1_{n_i} 1'_{n_i}$.

For a given $\delta = (\sigma_v^2, \sigma_u^2)'$, the BLUP estimator of b_i is given by

$$\hat{b}_i(\delta) = G_i Z_i' V_i^{-1} \{\mathrm{y}_i - X_i \tilde{\beta}(\delta)\}, \tag{2.6}$$

where $\tilde{\beta}(\delta) = (\sum_{i=1}^m X_i' V_i^{-1} X_i)^{-1} (\sum_{i=1}^m X_i' V_i^{-1} \mathrm{y}_i)$ is the weighted least squares (WLS) estimator of β, and $V_i = R_i + Z_i G_i Z_i'$ with $R_i = diag(\sigma_{ei1}^2, ..., \sigma_{ein_i}^2)$ and

$$G_i = \begin{pmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_u^2 I_{n_i} \end{pmatrix}.$$

For simplicity of notation, we let $\tilde{\beta}(\delta) = \tilde{\beta}$. We employ the following lemma 1 to obtain V_i^{-1} from (2.5).

LEMMA 1. *Let A be a $k \times k$ nonsingular matrix and u and v be two $k \times 1$ vectors such that $A + uv'$ is nonsingular. Then we have*

$$(A + uv')^{-1} = A^{-1} - A^{-1} uv' A^{-1} (1 + v' A^{-1} u)^{-1}.$$

We have

$$V_i^{-1} = diag_j \{(\sigma_u^2 + \sigma_{eij}^2)^{-1}\} - \frac{\sigma_u^2 \gamma_i}{\gamma_{i.}} w_i w_i', \tag{2.7}$$

where $\gamma_{ij} = \sigma_u^2 / (\sigma_u^2 + \sigma_{eij}^2)$, $\gamma_{i.} = \sum_{j=1}^{n_i} \gamma_{ij}$, $\gamma_i = \sigma_v^2 / (\sigma_v^2 + \sigma_u^2 / \gamma_{i.})$, $w_{ij} = (\sigma_u^2 + \sigma_{eij}^2)^{-1}$ and $w_i = (w_{i1}, ..., w_{in_i})'$.

Using (2.7) in (2.6), we obtain, after simplification, the BLUP estimators of v_i and u_{ij}, for a given δ, as

$$\tilde{v}_i(\delta) = \gamma_i (\bar{y}_{i\gamma} - \bar{x}_{i\gamma}' \tilde{\beta}) \tag{2.8}$$

and

$$\tilde{u}_{ij} = \gamma_{ij}(y_{ij} - x'_{ij}\tilde{\beta}) - \gamma_i \gamma_{ij}(\bar{y}_{i\gamma} - \bar{x}'_{i\gamma}\tilde{\beta}), \tag{2.9}$$

where $\bar{y}_{i\gamma} = \frac{1}{\gamma_{i.}} \sum_{j=1}^{n_i} \gamma_{ij} y_{ij}, \bar{x}_{i\gamma} = \frac{1}{\gamma_{i.}} \sum_{j=1}^{n_i} \gamma_{ij} x_{ij}$. The corresponding BLUP estimators of μ_{ij} for the sampled sub-areas are given by

$$\tilde{\mu}_{ij} = \tilde{\mu}_{ij}(\delta) = x'_{ij}\tilde{\beta}(\delta) + \tilde{v}_i(\delta) + \tilde{u}_{ij}(\delta), \quad i = 1, ..., m; j = 1, ..., n_i, \tag{2.10}$$

where $\tilde{v}_i(\delta)$ and $\tilde{u}_{ij}(\delta)$ are obtained from (2.8) and (2.9). For the non-sampled sub-areas, we use the pseudo-BLUP estimator

$$\tilde{\mu}^*_{ij} = \tilde{\mu}^*_{ij}(\delta) = x'_{ij}\tilde{\beta}(\delta) + \tilde{v}_i(\delta), \quad i = 1, ..., m; j = n_{i+1}, ..., N_i. \tag{2.11}$$

The BLUP estimator of area mean μ_i is given by

$$\tilde{\mu}^*_i = \tilde{\mu}^*_i(\delta) = \Big\{ \sum_{j=1}^{n_i} N_{ij}\tilde{\mu}_{ij}(\delta) + \sum_{j=n_{i+1}}^{N_i} N_{ij}\tilde{\mu}^*_{ij}(\delta) \Big\} / N_{i+}. \tag{2.12}$$

Note that the sub-area estimators automatically benchmark to the corresponding area estimator.

If N_i is large, then $\bar{U}_i \approx 0$ and we can approximate μ_i by

$$\mu_i \approx \bar{X}'_i\beta + v_i,$$

and the BLUP $\tilde{\mu}^*_i(\delta)$ by

$$\tilde{\mu}_i(\delta) \approx \bar{X}'_i\tilde{\beta} + \tilde{v}_i(\delta), \qquad i = 1, ..., m, \tag{2.13}$$

assuming that n_i is small.

2.1. Estimation of variance components

At this stage, we need to estimate the variance components σ_v^2 and σ_u^2 to obtain the EBLUP estimators of μ_{ij} and μ_i. We first extend the Fay and Herriot [6] iterative moment method of estimating the variance component in the Fay-Herriot area level model to the sub-area level model (2.3). Based on weighted residual sum of squares, we have

$$E\Big\{\sum_{i=1}^{m}(\mathrm{y}_i - X_i\tilde{\beta})'V_i^{-1}(\mathrm{y}_i - X_i\tilde{\beta})\Big\} = n - p.$$

The method of moments is then used to form the following estimating equation:

$$\sum_{i=1}^{m}(\mathrm{y}_i - X_i\tilde{\beta})'V_i^{-1}(\mathrm{y}_i - X_i\tilde{\beta}) = n - p. \tag{2.14}$$

On the other hand, we take the average of model (2.3) over j to get

$$\begin{aligned}\bar{y}_i &= \bar{x}_i'\beta + v_i + \bar{u}_i + \bar{e}_i, \quad i = 1, ..., m,\\ &= \bar{x}_i'\beta + \epsilon_i, \end{aligned}\tag{2.15}$$

where $\bar{y}_i, \bar{x}_i, \bar{u}_i$ and $\bar{e}_i$ are averages over $j = 1, ..., n_i$ and $\epsilon_i \overset{ind}{\sim} N\{0, \tilde{\sigma}_i^2 = n_i^{-1}(n_i\sigma_v^2 + \sigma_u^2 + \sigma_{ei}^2)\}$. We then obtain the WLS estimator of β associated with (2.15) as $\tilde{\beta}^* = (\sum_{i=1}^m \bar{x}_i\bar{x}_i'/\tilde{\sigma}_i^2)^{-1}(\sum_{i=1}^m \bar{x}_i\bar{y}_i/\tilde{\sigma}_i^2)$ and the associated weighted residual sum of squares $\sum_{i=1}^m(\bar{y}_i - \bar{x}_i'\tilde{\beta}^*)^2/\tilde{\sigma}_i^2$ with expectation equal to $m-p$. We then use the method of moments to form the estimating equation

$$\sum_{i=1}^{m}(\bar{y}_i - \bar{x}_i'\tilde{\beta}^*)^2/\tilde{\sigma}_i^2 = m - p. \tag{2.16}$$

The estimators $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ are obtained by solving the two equations (2.14) and (2.16) iteratively. Since $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ may take negative values, we define $\hat{\sigma}_v^2 = \max(0, \tilde{\sigma}_v^2)$ and $\hat{\sigma}_u^2 = \max(0, \tilde{\sigma}_u^2)$.

Substituting $\hat{\delta} = (\hat{\sigma}_v^2, \hat{\sigma}_u^2)'$ for $\delta = (\sigma_v^2, \sigma_u^2)'$ in (2.10) we get the EBLUP estimator of μ_{ij} for sampled sub-areas as $\hat{\mu}_{ij} = \tilde{\mu}_{ij}(\hat{\delta})$. Similarly, for the non-sampled sub-areas, a pseudo-EBLUP estimator of μ_{ij} is obtained from (2.11) as $\hat{\mu}_{ij}^* = \tilde{\mu}_{ij}^*(\hat{\delta})$. The exact EBLUP of μ_i is obtained from (2.12) as $\hat{\mu}_i^* = \tilde{\mu}_i^*(\hat{\delta})$ and its approximation from (2.13) as $\hat{\mu}_i = \tilde{\mu}_i(\hat{\delta})$. Note that the estimators $\hat{\mu}_i, \hat{\mu}_{ij}$ and $\hat{\mu}_{ij}^*$ do not require normality assumption. However, in Section 3 we use normality of v_i, u_{ij} and e_{ij} to derive a second order approximation to the MSE of the estimators.

3. Mean squared error approximation

In this section, we obtain a second order approximation to the MSE of the EBLUP estimators $\hat{\mu}_{ij}, \hat{\mu}_{ij}^*$ and $\hat{\mu}_i$, in the sense that the neglected terms are of order $o(m^{-1})$ for large m. We assume normality of v_i, u_{ij} and e_{ij}.

3.1. Sub-area estimators

We first consider the EBLUP $\hat{\mu}_{ij}$ of sampled sub-area j nested within area i. Under normality of the random effects v_i, u_{ij} and e_{ij}, we can express $\text{MSE}(\hat{\mu}_{ij}) = E(\hat{\mu}_{ij} - \mu_{ij})^2$ as

$$\text{MSE}(\hat{\mu}_{ij}) = \text{MSE}(\tilde{\mu}_{ij}) + E(\hat{\mu}_{ij} - \tilde{\mu}_{ij})^2, \tag{3.1}$$

where $\text{MSE}(\tilde{\mu}_{ij}) = E(\tilde{\mu}_{ij} - \mu_{ij})^2$. Further, an exact expression for $\text{MSE}(\tilde{\mu}_{ij})$ is given by

$$\text{MSE}(\tilde{\mu}_{ij}) = g_{1ij}(\delta) + g_{2ij}(\delta), \tag{3.2}$$

where

$$g_{1ij}(\delta) = (1 - \gamma_{ij})^2\{\sigma_u^2 + (1 - \gamma_i)\sigma_v^2\} + \gamma_{ij}^2\sigma_{eij}^2, \tag{3.3}$$

and

$$g_{2ij}(\delta) = (1-\gamma_{ij})^2(x_{ij}-\gamma_i\bar{x}_{i\gamma})'\Phi(x_{ij}-\gamma_i\bar{x}_{i\gamma}),$$

with $\Phi = var(\tilde{\beta}) = (\sum_{i=1}^m X_i'V_i^{-1}X_i)^{-1}$. It may be noted that (3.2) does not require the normality assumption.

The second term $g_{2ij}(\delta)$ in (3.2), due to estimating β, is of order $O(m^{-1})$, while the first term $g_{1ij}(\delta)$, given by (3.3), is $O(1)$. We can therefore interpret $g_{1ij}(\delta)$ as the MSE when all parameters are known, and $g_{2ij}(\delta)$ as the inflation to MSE due to estimating β. It can be shown that the leading term $g_{1ij}(\delta) \leq \sigma^2_{eij}$, where $\sigma^2_{eij} = E(y_{ij}-\mu_{ij})^2$ is the MSE of the direct estimator y_{ij}.

It remains to evaluate the last term $E(\hat{\mu}_{ij}-\tilde{\mu}_{ij})^2$ in (3.1). Following Das, Jiang, and Rao [3], we propose the following approximation, based on Taylor linearization:

$$E(\hat{\mu}_{ij}-\tilde{\mu}_{ij})^2 \approx E(\frac{d\tilde{\mu}^B_{ij}}{d\sigma^2_v})^2 var(\tilde{\sigma}^2_v) + E(\frac{d\tilde{\mu}^B_{ij}}{d\sigma^2_u})^2 var(\tilde{\sigma}^2_u)$$
$$+2E\{(\frac{d\tilde{\mu}^B_{ij}}{d\sigma^2_v})(\frac{d\tilde{\mu}^B_{ij}}{d\sigma^2_u})\}cov(\tilde{\sigma}^2_v,\tilde{\sigma}^2_u) \equiv g_{3ij}(\delta),$$

where

$$\tilde{\mu}^B_{ij} = x'_{ij}\beta + \gamma_i(1-\gamma_{ij})(\bar{y}_{i\gamma}-\bar{x}'_{i\gamma}\beta) + \gamma_{ij}(y_{ij}-x'_{ij}\beta).$$

After considerable simplification (see Torabi [20]), $g_{3ij}(\delta)$ is obtained as

$$g_{3ij}(\delta) = (1-\gamma_{ij})^2\Bigg\{(\frac{\sigma^2_u}{\gamma_{i.}})^2(\frac{\gamma_i}{\sigma^2_v})^3 var(\tilde{\sigma}^2_v) + \frac{\sigma^2_v}{\gamma_i}\Big[\frac{\gamma_i}{\sigma^2_u\gamma_{i.}}\sum_j \gamma^2_{ij}(1-\frac{\gamma_i\sigma^2_u}{\gamma_{i.}\sigma^2_v}) - \frac{\gamma_i\gamma_{ij}}{\sigma^2_u}\Big]\Big[-\frac{\gamma_i}{\sigma^2_u\gamma_{i.}}$$
$$.\sum_j \gamma^2_{ij}(1+\frac{\gamma_i\sigma^2_u}{\gamma_{i.}\sigma^2_v}) + \frac{\gamma_{ij}}{\sigma^2_u}(2-\gamma_i)\Big] + \sigma_u^{-4}\Big[\gamma^2_{ij}\sigma^2_v + \sigma^2_u\gamma_{ij} + \frac{\gamma^2_i}{\gamma^2_{i.}}\{\sigma^2_u\sum_j\gamma^3_{ij} + \sigma^2_v(\sum_j\gamma^2_{ij})^2\}$$
$$-2\frac{\gamma_i\gamma_{ij}}{\gamma_{i.}}(\sigma^2_v\sum_j\gamma^2_{ij} + \sigma^2_u\gamma_{ij})\Big] var(\tilde{\sigma}^2_u) + 2\frac{\gamma_i}{\gamma_{i.}\sigma^2_v}\Big[\gamma_{ij}(1-\gamma_i) - \frac{\gamma^2_i\sigma^2_u}{\gamma^2_{i.}\sigma^2_v}\sum_j\gamma^2_{ij}\Big] cov(\tilde{\sigma}^2_v,\tilde{\sigma}^2_u)\Bigg\}. \quad (3.4)$$

The term $g_{3ij}(\delta)$ is of the same order as $g_{2ij}(\delta)$. Combining (3.2) and (3.4), we get a second order approximation to MSE($\hat{\mu}_{ij}$) as

$$\text{MSE}(\hat{\mu}_{ij}) \approx g_{1ij}(\delta) + g_{2ij}(\delta) + g_{3ij}(\delta). \tag{3.5}$$

The proof that neglected terms in the approximation (3.5) are of lower order, $o(m^{-1})$, for large m, is omitted, for simplicity (see Torabi [20]). However, $var(\tilde{\sigma}_v^2)$, $var(\tilde{\sigma}_u^2)$ and $cov(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)$ are derived in Appendix A.

The MSE approximation of the non-sampled sub-area pseudo-EBLUP $\hat{\mu}_{ij}^*$ is similarly obtained (details omitted). We have MSE($\hat{\mu}_{ij}^*$) = MSE($\tilde{\mu}_{ij}^*$) + $E(\hat{\mu}_{ij}^* - \tilde{\mu}_{ij}^*)^2$ and

$$\text{MSE}(\tilde{\mu}_{ij}^*) = g_{1ij}^*(\delta) + g_{2ij}^*(\delta), \tag{3.6}$$

where

$$g_{1ij}^*(\delta) = \sigma_u^2\{1 + \gamma_i\gamma_{i.}^{-1}(1 - 2\gamma_i)\}$$

and

$$g_{2ij}^*(\delta) = (x_{ij} - \gamma_i\bar{x}_{i\gamma})'\Phi(x_{ij} - \gamma_i\bar{x}_{i\gamma}).$$

Further,

$$\begin{aligned} E(\hat{\mu}_{ij}^* - \tilde{\mu}_{ij}^*)^2 &\approx g_{3ij}^*(\delta) \\ = \gamma_i^2\Big\{&[\sigma_v^2 + \gamma_{i.}^{-1}(\gamma_{i1}\sigma_u^2 + \gamma_{i1e})]\{(\frac{1-\gamma_i}{\sigma_v^2})^2 var(\tilde{\sigma}_v^2) \\ &+\sigma_u^{-4}(\gamma_i - 1 + \gamma_{i1})^2 var(\tilde{\sigma}_u^2) + [2(1-\gamma_i)(\gamma_i - 1 + \gamma_{i1})\sigma_u^{-2}\sigma_v^{-2}]cov(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)\} \\ &+\sigma_u^{-4}[(\sigma_v^2 + \sigma_u^2)\gamma_{i2} + \gamma_{i2e} - 2\gamma_{i1}(\gamma_i - 1 + \gamma_{i1})(\sigma_v^2 + \sigma_u^2\gamma_{i.}^{-1})]var(\tilde{\sigma}_u^2) \\ &-2\gamma_{i1}(1-\gamma_i)(\sigma_v^2 + \gamma_{i.}^{-1}\sigma_u^2)\sigma_u^{-2}\sigma_v^{-2}cov(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)\Big\}, \end{aligned} \tag{3.7}$$

where $\gamma_{i1} = \gamma_{i.}^{-1}\sum_{j=1}^{n_i}\gamma_{ij}^2, \gamma_{i1e} = \gamma_{i.}^{-1}\sum_{j=1}^{n_i}\gamma_{ij}^2\sigma_{eij}^2, \gamma_{i2} = \gamma_{i.}^{-2}\sum_{j=1}^{n_i}\gamma_{ij}^4$, and $\gamma_{i2e} = \gamma_{i.}^{-2}\sum_{j=1}^{n_i}\gamma_{ij}^4\sigma_{eij}^2$. The term $g_{3ij}^*(\delta)$ is of the same order as $g_{2ij}^*(\delta)$.

Considering (3.6) and (3.7), we get a second order approximation to MSE($\hat{\mu}^*_{ij}$) as

$$\text{MSE}(\hat{\mu}^*_{ij}) \approx g^*_{1ij}(\delta) + g^*_{2ij}(\delta) + g^*_{3ij}(\delta). \tag{3.8}$$

3.2. Area estimators

Similar to Section 3.1, we follow Das, Jiang, and Rao [3] to obtain a second order approximation to MSE of the area EBLUP $\hat{\mu}_i$, assuming that N_i is large. We have

$$\text{MSE}(\hat{\mu}_i) = \text{MSE}(\tilde{\mu}_i) + E(\hat{\mu}_i - \tilde{\mu}_i)^2,$$

where

$$\text{MSE}(\tilde{\mu}_i) = g_{1i}(\delta) + g_{2i}(\delta), \tag{3.9}$$

with

$$g_{1i}(\delta) = \frac{\gamma_i}{\gamma_{i.}}\sigma_u^2,$$
$$g_{2i}(\delta) = (\bar{X}_i - \gamma_i \bar{x}_{i\gamma})'\Phi(\bar{X}_i - \gamma_i \bar{x}_{i\gamma}).$$

The terms $g_{1i}(\delta)$ and $g_{2i}(\delta)$ are of order $O(1)$ and $O(m^{-1})$ respectively.

Further, by Taylor linearization, $E(\hat{\mu}_i - \tilde{\mu}_i)^2$ is approximated as

$$E(\hat{\mu}_i - \tilde{\mu}_i)^2 \approx E(\frac{d\tilde{\mu}_i^B}{d\sigma_v^2})^2 var(\tilde{\sigma}_v^2) + E(\frac{d\tilde{\mu}_i^B}{d\sigma_u^2})^2 var(\tilde{\sigma}_u^2)$$

$$+2E\{(\frac{d\tilde{\mu}_i^B}{d\sigma_v^2})(\frac{d\tilde{\mu}_i^B}{d\sigma_u^2})\}cov(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2) \approx g_{3i}(\delta),$$

where

$$\tilde{\mu}_i^B = \bar{X}_i'\beta + \gamma_i(\bar{y}_{i\gamma} - \bar{x}_{i\gamma}'\beta).$$

After some calculation we obtain

$$g_{3i}(\delta) = (\frac{\gamma_i}{\gamma_{i.}})^2\Bigg[\frac{\sigma_u^4\gamma_i}{\sigma_v^6}var(\tilde{\sigma}_v^2)+\sigma_u^{-4}\sum_j\gamma_{ij}^4\Big\{(1-\gamma_i)\sigma_v^2+\sigma_u^2(1-2\frac{\gamma_i}{\gamma_{i.}})+\sigma_{eij}^2\Big\}var(\tilde{\sigma}_u^2)$$

$$+2\sigma_v^{-4}\gamma_i\sum_j\gamma_{ij}^2\Big\{(\gamma_i-1)\sigma_v^2+\frac{\sigma_u^2}{\gamma_{i.}}(\gamma_i-\gamma_{ij})-\frac{\gamma_{ij}}{\gamma_{i.}}\sigma_{eij}^2\Big\}cov(\tilde{\sigma}_v^2,\tilde{\sigma}_u^2)\Bigg], \quad (3.10)$$

where $g_{3i}(\delta)$ is of order $O(m^{-1})$. Combining (3.9) and (3.10), we obtain a second order approximation to MSE($\hat{\mu}_i$) as follows

$$\text{MSE}(\hat{\mu}_i) \approx g_{1i}(\delta) + g_{2i}(\delta) + g_{3i}(\delta). \quad (3.11)$$

The neglected terms in the approximation (3.11) are $o(m^{-1})$ for large m, similar to Section 3.1, but the proof is omitted (see Torabi [20]).

We now turn to the MSE of the exact EBLUP $\hat{\mu}_i^*$ of μ_i. We can express MSE($\hat{\mu}_i^*$) in terms of the approximation MSE($\hat{\mu}_i$) as follows:

$$\text{MSE}(\hat{\mu}_i^*) = \text{MSE}(\hat{\mu}_i) + \sum_{j=n_{i+1}}^{N_i} w_{ij}^2\sigma_u^2 + E\{\sum_{j=1}^{n_i} w_{ij}(\hat{u}_{ij}-u_{ij})\}^2$$

$$+2E\{(\bar{X}_i'\hat{\beta}+\hat{v}_i)-(\bar{X}_i'\beta+v_i)\}\{\sum_{j=1}^{n_i} w_{ij}(\hat{u}_{ij}-u_{ij})\}, \quad (3.12)$$

where MSE($\hat{\mu}_i$) is given by (3.11). Details of the evaluation of the last two terms in (3.12) are available from the authors. Derivation of second-order unbiased estimator of MSE($\hat{\mu}_i^*$) is also available. In this paper, for simplicity, we confine ourselves to the approximation pseudo-EBLUP $\hat{\mu}_i$ of the area mean μ_i.

4. Estimation of mean squared error

Using the second order MSE approximations (3.5), (3.8) and (3.11), we now obtain MSE estimators that are second-order unbiased in the sense that $E[\text{mse}(\hat{\mu}_{ij})] - \text{MSE}(\hat{\mu}_{ij}) = o(m^{-1}), E[\text{mse}(\hat{\mu}^*_{ij})] - \text{MSE}(\hat{\mu}^*_{ij}) = o(m^{-1})$, and $E[\text{mse}(\hat{\mu}_i)] - \text{MSE}(\hat{\mu}_i) = o(m^{-1})$. Since $g_{2ij}(\delta)$ and $g_{3ij}(\delta)$ are of order $O(m^{-1})$, it follows that

$$E[g_{2ij}(\hat{\delta})] = g_{2ij}(\delta) + o(m^{-1}),$$

$$E[g_{3ij}(\hat{\delta})] = g_{3ij}(\delta) + o(m^{-1}).$$

However, $g_{1ij}(\delta)$ is of order $O(1)$, so $g_{1ij}(\hat{\delta})$ is not a second-order correct estimator of $g_{1ij}(\delta)$, since its bias is of order $O(m^{-1})$. Thus, the bias in $g_{1ij}(\delta)$ must be estimated to the correct order. Following Das, Jiang, and Rao [3] and Datta et al. [5], we can write

$$E[g_{1ij}(\hat{\delta}) + (\hat{\delta} - \delta)'\nabla g_{1ij}(\delta)|_{\delta=\hat{\delta}} + g_{3ij}(\hat{\delta})] = g_{1ij}(\delta) + o(m^{-1}).$$

Therefore, combining the above results, a second order unbiased estimator of $\text{MSE}(\hat{\mu}_{ij})$ is given by

$$\text{mse}(\hat{\mu}_{ij}) = g_{1ij}(\hat{\delta}) + g_{2ij}(\hat{\delta}) + 2g_{3ij}(\hat{\delta}) - \frac{dg_{1ij}(\delta)}{d\sigma_v^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_v^2) - \frac{dg_{1ij}(\delta)}{d\sigma_u^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_u^2), \tag{4.1}$$

where $\hat{b}(\hat{\sigma}_v^2)$ and $\hat{b}(\hat{\sigma}_u^2)$ are estimators of biases $b(\hat{\sigma}_v^2) = E(\hat{\sigma}_v^2 - \sigma_v^2)$ and $b(\hat{\sigma}_u^2) = E(\hat{\sigma}_u^2 - \sigma_u^2)$, respectively. Expressions for $\hat{b}(\hat{\sigma}_v^2)$ and $\hat{b}(\hat{\sigma}_u^2)$ are given in Appendix B.

Similarly,

$$\text{mse}(\hat{\mu}^*_{ij}) = g^*_{1ij}(\hat{\delta}) + g^*_{2ij}(\hat{\delta}) + 2g^*_{3ij}(\hat{\delta}) - \frac{dg^*_{1ij}(\delta)}{d\sigma_v^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_v^2) - \frac{dg^*_{1ij}(\delta)}{d\sigma_u^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_u^2), \tag{4.2}$$

and

$$\text{mse}(\hat{\mu}_i) = g_{1i}(\hat{\delta})+g_{2i}(\hat{\delta})+2g_{3i}(\hat{\delta})-\frac{dg_{1i}(\delta)}{d\sigma_v^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_v^2)-\frac{dg_{1i}(\delta)}{d\sigma_u^2}|_{\delta=\hat{\delta}}\hat{b}(\hat{\sigma}_u^2). \quad (4.3)$$

We now find $\frac{dg_{1ij}(\delta)}{d\sigma_v^2}$ and $\frac{dg_{1ij}(\delta)}{d\sigma_u^2}$ in (4.1). Recalling that $g_{1ij}(\delta) = (1-\gamma_{ij})^2[\sigma_u^2 + \sigma_v^2(1-\gamma_i)] + \gamma_{ij}^2\sigma_{eij}^2$, we obtain

$$\frac{dg_{1ij}(\delta)}{d\sigma_v^2} = (1-\gamma_{ij})^2(1-\gamma_i - \frac{\gamma_i^2\sigma_u^2}{\gamma_{i.}\sigma_v^2}),$$

and

$$\frac{dg_{1ij}(\delta)}{d\sigma_u^2} = (1-\gamma_{ij})\Big\{\frac{-2\gamma_{ij}(1-\gamma_{ij})}{\sigma_u^2}[\sigma_u^2+(1-\gamma_i)\sigma_v^2]+(1-\gamma_{ij})(1+\frac{\gamma_i^2\sum_{j=1}^{n_i}\gamma_{ij}^2}{\gamma_{i.}^2})+\frac{2\gamma_{ij}^2\sigma_{eij}^2}{\sigma_u^2}\Big\}.$$

Combining the above results, we obtain $\text{mse}(\hat{\mu}_{ij})$ from (4.1).

THEOREM 1. *A second order unbiased estimator of the MSE*$(\hat{\mu}_{ij})$ *is given by mse*$(\hat{\mu}_{ij})$ *with the property* $E[mse(\hat{\mu}_{ij})] = MSE(\hat{\mu}_{ij}) + o(m^{-1})$.

Proof of Theorem 1 is given in Appendix C.

We now need to find $\frac{dg^*_{1ij}(\delta)}{d\sigma_v^2}$ and $\frac{dg^*_{1ij}(\delta)}{d\sigma_u^2}$ in (4.2). Recalling that $g^*_{1ij}(\delta) = \sigma_u^2[1+\gamma_i\gamma_{i.}^{-1}(1-2\gamma_{ij})]$, we obtain

$$\frac{dg^*_{1ij}(\delta)}{d\sigma_v^2} = \gamma_i\gamma_{i.}^{-1}(1-\gamma_i)(1-2\gamma_{ij})\sigma_u^2\sigma_v^{-2},$$

and

$$\frac{dg^*_{1ij}(\delta)}{d\sigma_u^2} = \gamma_i\gamma_{i.}^{-1}\Big\{(1-2\gamma_{ij})[\gamma_i+\gamma_{i1}-(1+2\gamma_{ij})]-2\gamma_{ij}^2+\gamma_i^{-1}\gamma_{i.}\Big\}.$$

Combining the above results, we obtain $\text{mse}(\hat{\mu}^*_{ij})$ from (4.2).

THEOREM 2. *A second order unbiased estimator of the MSE*$(\hat{\mu}^*_{ij})$ *is given by mse*$(\hat{\mu}^*_{ij})$ *with the property* $E[mse(\hat{\mu}^*_{ij})] = MSE(\hat{\mu}^*_{ij}) + o(m^{-1})$.

Proof of Theorem 2 is similar to Theorem 1 and omitted for simplicity.

We now turn to $mse(\hat{\mu}_i)$. We need to obtain $\frac{dg_{1i}(\delta)}{d\sigma_v^2}$ and $\frac{dg_{1i}(\delta)}{d\sigma_u^2}$ in (4.3). Recalling that $g_{1i}(\delta) = \gamma_i \frac{\sigma_u^2}{\gamma_{i.}}$, we obtain $\frac{dg_{1i}(\delta)}{d\sigma_v^2} = \frac{\gamma_i^2 \sigma_u^4}{\gamma_{i.}^2 \sigma_v^4}$ and $\frac{dg_{1i}(\delta)}{d\sigma_u^2} = (\frac{\gamma_i}{\gamma_{i.}})^2 \sum_{j=1}^{n_i} \gamma_{ij}^2$. Combining the above results, we obtain $\text{mse}(\hat{\mu}_i)$.

THEOREM 3. *A second order unbiased estimator of the MSE of $\hat{\mu}_i$ is given by $mse(\hat{\mu}_i)$ with the property $E[mse(\hat{\mu}_i)] = MSE(\hat{\mu}_i) + o(m^{-1})$.*

Proof of Theorem 3 follows along the lines of Appendix C and hence omitted.

5. Simulation study

5.1. Model-based simulation study

A model-based simulation study was undertaken in order to investigate the performance of the proposed sub-area model. For simplicity, we considered the simple balanced sub-area model with a common mean and error variances $\sigma_{eij}^2 = \sigma_{ei}^2$ for each sub-area. It is given by

$$y_{ij} = \mu + v_i + u_{ij} + e_{ij}, \quad j = 1, ..., \bar{n}; i = 1, ..., m. \tag{5.1}$$

Model (5.1) is a special case of the sub-area level small area model (2.3) with $x_{ij} = 1, \beta = \mu, p = 1, n_i = \bar{n}$.

In our simulation study, we used $\bar{n} = 3$ sample sub-areas in each area, $m = 30$ small areas and consequently $n = 90$ sub-areas in each sample. Without loss of generality, we set $\mu = 0$. However, to account for the estimation of unknown regression parameters that arise in application, we still need to estimate this zero mean. We fixed $\sigma_v^2 = \sigma_u^2 = 300$, while we considered three patterns for σ_{ei}^2, variance of sampling error; (a) $\sigma_{ei}^2 =$

$(260, 280, 300, 320, 340, 360)$; (b) $\sigma_{ei}^2 = (230, 280, 300, 320, 365, 650)$ and (c) $\sigma_{ei}^2 = (100, 280, 300, 380, 750, 2000)$, which are similar to those in Datta et al. [5]. There are six groups $A_1, ..., A_6$ and five small areas in each group with three small sub-areas in each area. The sampling variances σ_{ei}^2 are the same within the same group. Pattern (c) has the largest variability in the σ_{ei}^2-values, while pattern (a) is the least variable and pattern (b) has intermediate variability.

Following Lahiri and Rao [14] and Datta et al. [5], we considered three different distributions for v_i's and u_{ij}'s, namely normal, double-exponential and location-exponential to evaluate the robustness of second-order unbiasedness of MSE estimators under nonnormality of the random effects v_i and u_{ij} assuming that the sampling errors e_{ij} are normal. For each pattern, we proceeded along the following steps. We generated $B = 5,000$ independent sets of random variables $\{v_i; i = 1, ..., 30\}$ and $\{u_{ij}; j = 1, 2, 3; i = 1, ..., 30\}$ from normal, double-exponential and location-exponential distributions having means zero and specified variances σ_v^2 and σ_u^2, and we also generated random variables $\{e_{ij}; j = 1, 2, 3; i = 1, ..., 30\}$ from normal having mean zero and variance σ_{ei}^2. From those generated datasets the observations $\{y_{ij}; j = 1, 2, 3; i = 1, ..., 30\}$ were obtained using the model $y_{ij} = v_i + u_{ij} + e_{ij}$. By using the generated samples $y_{ij}^{(b)}$, (b=1,...,B=5,000), we calculated $\hat{\sigma}_v^{2(b)}$ and $\hat{\sigma}_u^{2(b)}$ by WLS iteratively as described in section 2. In addition, we estimated the variance components σ_v^2 and σ_u^2 by using the methods of maximum likelihood (ML) and restricted maximum likelihood (REML) under normality. For each generated sample, we calculated

$$\mu_{ij}^{(b)} = v_i^{(b)} + u_{ij}^{(b)}, \quad j = 1, 2, 3; i = 1, ..., 30; b = 1, ..., B,$$

$$\mu_i^{(b)} = v_i^{(b)}, \quad i = 1, ..., 30; b = 1, ..., B.$$

We computed the EBLUP estimates $\hat{\mu}_{ij}^{(b)}$ and $\hat{\mu}_i^{(b)}$ for each generated sample b.

We now turn to the percent relative bias of the second-order correct MSE estimator of sub-area mean over sub-areas and the second-order correct MSE estimator of area mean as

$$\overline{\mathrm{RB}}_i = 100\Big[\bar{n}^{-1}\sum_j \mathrm{RB}_{ij}\Big] \quad (i = 1, ..., m)$$

and

$$\mathrm{RB}_i = 100[B^{-1}\sum_{b=1}^{B} \mathrm{mse}_i^{(b)}/\mathrm{EMSE}_i - 1] \quad (i = 1, ..., m)$$

where

$$\mathrm{RB}_{ij} = B^{-1}\sum_{b=1}^{B} \mathrm{mse}_{ij}^{(b)}/\mathrm{EMSE}_{ij} - 1,$$

$$\mathrm{EMSE}_{ij} = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_{ij}^{(b)} - \mu_{ij}^{(b)})^2,$$

$$\mathrm{EMSE}_i = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_i^{(b)} - \mu_i^{(b)})^2,$$

and mse_{ij} and mse_i are given by (4.1) and (4.3). We then averaged $\overline{\mathrm{RB}}_i$ and RB_i over areas within the same group.

We report the results in Tables 1 and 2. As shown, all three methods, ML, REML and EFH, perform equally well. More specifically, using the EFH method, RB(%) of the second-order correct estimator of MSE_i of sub-area ranges from -3.9% to 1.6% for all three patterns and three distributions, suggesting near unbiasedness. For the EFH method, patterns (a) and (b)

Table 1

Percent average relative bias of sub-area MSE estimator over areas and sub-areas within the same group.

	Pattern (a)			Pattern (b)			Pattern (c)		
	ML	REML	EFH	ML	REML	EFH	ML	REML	EFH
				Normal					
A_1	-0.9	-0.6	-0.5	-0.8	-0.5	-0.5	-0.3	-0.4	-3.9
A_2	0.2	0.6	0.6	0.2	0.6	0.6	0.1	0.5	-0.7
A_3	-0.5	-0.0	0.0	-0.5	-0.0	0.0	-0.6	-0.2	-1.2
A_4	-0.5	-0.1	0.0	-0.5	-0.1	0.0	-0.9	-0.4	-1.1
A_5	0.1	0.6	0.7	0.1	0.6	0.6	-0.9	0.1	-0.3
A_6	-1.0	-0.4	-0.4	-1.5	-0.7	-0.7	-2.7	-0.9	-1.5
				Double-exponential					
A_1	0.3	0.7	0.7	0.5	0.8	0.8	1.0	0.9	-2.5
A_2	1.1	1.5	1.6	1.1	1.5	1.6	0.3	0.6	0.2
A_3	0.1	0.6	0.6	0.1	0.5	0.7	-0.8	-0.3	-0.7
A_4	-0.1	0.3	0.4	-0.2	0.3	0.4	-1.4	-0.9	-0.8
A_5	-1.0	-0.5	-0.4	-1.3	-0.7	-0.4	-3.5	-2.4	-1.7
A_6	-0.9	-0.4	-0.2	-2.0	-1.2	-0.7	-4.4	-2.6	-1.8
				Location-exponential					
A_1	0.7	1.1	1.0	1.1	1.5	1.2	2.0	1.9	-1.8
A_2	0.6	1.0	1.1	0.6	1.1	1.1	-0.5	-0.1	-0.3
A_3	0.1	0.6	0.6	0.2	0.6	0.7	-0.9	-0.5	-0.6
A_4	-0.6	-0.1	0.0	-0.5	-0.1	0.1	-2.1	-1.6	-1.2
A_5	-0.5	0.0	0.2	-0.6	-0.1	0.2	-3.2	-2.2	-1.0
A_6	-1.3	-0.8	-0.6	-2.6	-1.7	-1.2	-5.0	-3.2	-2.1

with least and intermediate variability in σ_{ei}^2 have better performance than pattern (c) which has the largest variability in σ_{ei}^2 for all three distributions. On the other hand, it is clear from Table 2 that mse_i for the EFH method leads to overestimation for all three distributions with RB(%) ranging from 1.3 to 5.4 for pattern (a), 0.3 to 6.1 for pattern (b), and -2.5 to 13.8 for pattern (c). The RB(%) for the ML method range from -1.7 to 2.1 for pattern (a), -4.1 to 2.9 for pattern (b), and -9.0 to 8.6 for pattern (c), and results for REML range from 0.9 to 4.9 for pattern (a), -0.8 to 5.4 for pattern (b), and -4.5 to 11.1 for pattern (c). Also, the variability of RB(%) for pattern (c) is larger than for patterns (a) and (b) for all three methods and three distributions. For pattern (c) with the largest variability in σ_{ei}^2, RB(%) for group A_1 with the smallest σ_{ei}^2 is significantly larger than the RB(%) for the other groups $A_2, ..., A_6$ with larger σ_{ei}^2.

Furthermore, to evaluate the efficiency of sub-area and area EBLUP estimators relative to direct estimators, we computed the percent average relative efficiency (EFF) of sub-area EBLUP estimator over sub-area direct estimator y_{ij} and area EBLUP estimator over area direct estimator $\bar{y}_i = n_i^{-1} \sum_j y_{ij}$ as

$$\text{EFF}_{1i} = 100(\overline{\text{EMSE}}_{i.dir}/\overline{\text{EMSE}}_i)^{1/2} \quad , \quad \text{EFF}_{2i} = 100(\text{EMSE}_{i.dir}/\text{EMSE}_i)^{1/2}$$

respectively, where $\overline{\text{EMSE}}_{i.dir} = \bar{n}^{-1} \sum_j \text{EMSE}_{ij.dir}$, $\text{EMSE}_{ij.dir} = B^{-1} \sum_{b=1}^{B} (y_{ij}^{(b)} - \mu_{ij}^{(b)})^2$, $\text{EMSE}_{i.dir} = B^{-1} \sum_{b=1}^{B} (\bar{y}_i^{(b)} - \mu_i^{(b)})^2$ with $\bar{y}_i^{(b)} = n_i^{-1} \sum_j y_{ij}^{(b)}$. The values of relative efficiency of sub-area EBLUP estimator over sub-area direct estimator averaged over areas within the same group are reported in Table 3. All three methods produced nearly identical results in terms of efficiency of sub-area EBLUP estimator over sub-area direct estimator for

Table 2

Percent average relative bias of area MSE estimator over areas within the same group.

	Pattern (a)			Pattern (b)			Pattern (c)		
	ML	REML	EFH	ML	REML	EFH	ML	REML	EFH
				Normal					
A_1	-0.1	2.4	2.9	0.8	3.3	3.9	4.9	7.2	9.9
A_2	1.2	3.8	4.3	1.5	4.1	4.8	1.1	4.0	5.8
A_3	-1.7	0.9	1.3	-1.5	1.0	1.7	-1.8	1.0	2.5
A_4	1.1	3.7	4.3	1.2	3.9	4.6	-0.1	3.1	4.6
A_5	-0.7	2.0	2.5	-0.8	1.9	2.6	-3.7	-0.1	1.0
A_6	-1.3	1.4	1.9	-3.2	-0.0	0.6	-6.0	-1.5	-1.2
				Double-exponential					
A_1	1.7	4.3	4.8	2.9	5.4	6.0	8.1	10.4	13.4
A_2	2.1	4.9	5.4	2.5	5.3	6.1	1.8	4.8	7.6
A_3	-1.4	1.2	1.8	-1.0	1.6	2.3	-1.4	1.5	3.6
A_4	1.0	3.6	4.2	1.3	4.1	4.8	-0.1	3.0	5.2
A_5	-0.4	2.3	2.9	-0.6	2.1	3.0	-4.9	-1.3	0.6
A_6	-1.2	1.5	2.2	-4.0	-0.8	0.3	-8.0	-3.6	-2.2
				Location-exponential					
A_1	1.3	4.1	4.5	2.8	5.4	5.9	8.6	11.1	13.8
A_2	1.0	3.8	4.3	1.5	4.3	5.0	1.1	4.1	6.7
A_3	0.5	3.1	3.7	1.0	3.7	4.4	0.5	3.4	6.1
A_4	-1.6	1.2	1.8	-1.1	1.7	2.5	-2.8	0.4	2.8
A_5	-0.8	1.9	2.6	-0.8	2.1	3.0	-5.9	-2.1	0.4
A_6	-0.8	2.1	2.8	-4.1	-0.8	0.5	-9.0	-4.5	-2.5

all three distributions and relative efficiency for all three methods and three distributions ranges from 124% to 132% for pattern (a); 121% to 151% for pattern (b) and 110% to 213% for pattern (c). Moreover, EFF_{1i} increases with increasing σ_{ei}^2 for all three patterns and three distributions, and the variability of efficiency for pattern (c) is larger than for patterns (a) and (b) due to large variability in σ_{ei}^2. As a result, for a large σ_{ei}^2, using the sub-area direct estimator leads to significant loss in efficiency. We have similar results on efficiency for area EBLUP estimator over area direct estimator (EFF_{2i}) averaged over areas within the same group, which are reported in Table 4.

If we take $\sigma_v^2 = 0$, our model (2.3) reduces to the Fay-Herriot model. To investigate the loss of efficiency by using the Fay-Herriot model incorrectly, we computed the percent average relative efficiency of sub-area EBLUP estimator over the Fay-Herriot estimator as $\mathrm{EFF}_{3i} = 100(\overline{\mathrm{EMSE}}_{i.FH}/\overline{\mathrm{EMSE}}_i)^{1/2}$ where $\overline{\mathrm{EMSE}}_{i.FH} = \bar{n}^{-1}\sum_j \mathrm{EMSE}_{ij.FH}$, $\mathrm{EMSE}_{ij.FH} = B^{-1}\sum_{b=1}^{B}(\hat{\mu}_{ij.FH}^{(b)} - \mu_{ij}^{(b)})^2$ with $\hat{\mu}_{ij.FH}^{(b)}$ denoting the Fay-Herriot estimator for the b-*th* simulated sample. We then averaged EFF_{3i} over areas within the same group. As shown in Table 5, the sub-area EBLUP estimator is more efficient than the Fay-Herriot estimator for all the three methods and patterns over all the three distributions. Similar to EFF of the sub-area EBLUP estimator over the direct estimator, efficiency of the sub-area EBLUP estimator over the Fay-Herriot estimator is nearly identical for all the three methods ML, REML, and EFH and three distributions normal, double-exponential, and location-exponential. Over all the three distributions, EFF(%) ranges from 104% to 106% for pattern (a), from 104% to 106% for pattern (b), and from

Table 3

Percent average relative efficiency of sub-area EBLUP over sub-area direct estimator averaged over areas within the same group.

	Pattern (a)			Pattern (b)			Pattern (c)		
	ML	REML	EFH	ML	REML	EFH	ML	REML	EFH
				Normal					
A_1	124	124	124	121	122	121	110	110	110
A_2	125	125	125	125	125	125	125	125	125
A_3	127	127	127	127	127	127	127	127	127
A_4	129	129	129	129	129	129	133	133	133
A_5	130	130	130	132	132	132	156	156	155
A_6	131	131	131	149	150	150	212	212	211
				Double-exponential					
A_1	125	125	125	122	122	122	111	111	110
A_2	126	126	126	126	126	126	126	126	126
A_3	128	128	128	128	128	128	128	128	128
A_4	129	129	129	129	129	129	133	133	133
A_5	131	131	131	133	133	133	157	157	157
A_6	132	132	132	150	150	151	213	213	213
				Location-exponential					
A_1	125	125	125	123	123	122	111	111	110
A_2	127	127	127	127	127	127	126	127	126
A_3	128	128	128	128	128	128	128	128	128
A_4	129	129	129	129	129	129	133	133	133
A_5	131	131	131	133	133	133	157	157	157
A_6	132	132	132	151	151	151	212	213	213

102% to 105% for pattern (c). An advantage of our sub-area model is that it provides EBLUP estimator for both area means and sub-area means simultaneously.

An advantage of the proposed sub-area level model is that the associate pseudo-EBLUP, $\hat{\mu}^*_{ij}$, for a non-sampled subarea (i,j) can lead to significant efficiency gains over the corresponding regression synthetic estimator, $x'_{ij}\hat{\beta}_{FH}$, under the Fay-Herriot model $y_{ij} = x'_{ij}\beta + u_{ij} + e_{ij}$ that ignores area effect, where $u_{ij} \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ and $\hat{\beta}_{FH}$ is the WLS estimator of β under this model. The efficiency gains (not reported here) are significantly larger than those for the sampled areas reported in Table 5.

5.2. Design-based simulation study

We consider the following two-fold model for a design-based simulation study:

$$y_{ijk} = \mu + v_i + u_{ij} + e_{ijk}, \quad i = 1, ..., m; j = 1, ..., N_i; k = 1, ..., N_{ij},$$

where $v_i \sim N(0, \sigma_v^2), u_{ij} \sim N(0, \sigma_u^2)$ and $e_{ijk} \sim N(0, \sigma_{ei}^2)$. We first generated a fixed finite population $\{y_{ijk}\}$ using the two-fold model with $m = 30$ small areas and $N_i = 18$ sub-areas and $N_{ij} = 300$ elements in each sub-area. Our population area and sub-area means are defined as follows:

$$\bar{Y}_i = \frac{1}{N_i N_{ij}} \sum_{j=1}^{N_i} \sum_{k=1}^{N_{ij}} y_{ijk}, \quad \bar{Y}_{ij} = N_{ij}^{-1} \sum_{k=1}^{N_{ij}} y_{ijk}.$$

We then draw a two-stage simple random sample from each area using $n_i = 10$ and $n_{ij} = 3$. We have $\bar{y}_{ij} = \mu + v_i + u_{ij} + \bar{e}_{ij}$ where $\bar{e}_{ij} \sim N(0, \sigma_{ei}^2/n_{ij})$. We

Table 4

Percent relative efficiency of area EBLUP over area direct estimator averaged over areas within the same group.

	Pattern (a)			Pattern (b)			Pattern (c)		
	ML	REML	EFH	ML	REML	EFH	ML	REML	EFH
					Normal				
A_1	121	122	122	120	120	120	113	114	113
A_2	122	122	122	122	122	122	121	122	121
A_3	123	123	123	122	123	123	122	122	122
A_4	124	124	124	124	124	124	125	126	126
A_5	125	125	125	126	126	126	139	140	140
A_6	125	126	126	136	136	137	179	179	179
					Double-exponential				
A_1	122	123	123	121	121	121	115	116	115
A_2	123	123	123	123	123	123	122	123	122
A_3	124	125	125	124	124	124	124	124	124
A_4	125	125	125	125	125	125	127	127	127
A_5	126	127	127	127	127	127	141	141	142
A_6	127	127	127	137	138	138	180	180	181
					Location-exponential				
A_1	123	123	123	122	122	122	117	117	116
A_2	124	124	124	124	124	124	124	124	123
A_3	126	126	126	126	126	126	125	125	125
A_4	126	126	126	126	126	126	128	128	128
A_5	127	127	128	128	128	128	141	142	142
A_6	128	128	128	138	138	139	180	180	181

Table 5

Percent average relative efficiency of sub-area EBLUP over the Fay-Herriot estimator averaged over areas within the same group.

	Pattern (a)			Pattern (b)			Pattern (c)		
	ML	REML	EFH	ML	REML	EFH	ML	REML	EFH
				Normal					
A_1	104	104	104	104	104	104	102	102	102
A_2	104	104	104	104	104	104	104	104	104
A_3	104	104	104	104	104	104	104	104	104
A_4	105	105	105	105	105	105	105	105	105
A_5	105	105	105	105	105	105	105	105	105
A_6	105	105	105	105	105	106	104	104	104
				Double-exponential					
A_1	105	105	105	105	105	105	103	103	103
A_2	104	105	104	105	105	105	105	105	105
A_3	104	104	104	104	104	104	104	104	104
A_4	106	106	106	106	106	106	104	104	104
A_5	105	105	105	105	105	105	105	105	105
A_6	105	105	105	105	105	105	104	104	104
				Location-exponential					
A_1	105	105	105	105	105	105	104	104	104
A_2	105	105	105	105	105	105	105	105	105
A_3	105	105	105	106	106	106	104	104	104
A_4	105	105	105	106	106	106	105	105	104
A_5	105	105	105	106	106	106	105	105	105
A_6	105	105	105	106	106	106	104	104	104

set the model parameters as in our model-based simulation study (section 5.1). For simplicity, we only considered the pattern (b) with the random effects generated from normal distribution. In particular, we set $\mu = 0, \sigma_v^2 = \sigma_u^2 = 300$, and $\sigma_{ei}^2/n_{ij} = (230, 280, 300, 320, 365, 650)$. We selected $B = 5000$ two-stage samples from the fixed finite population.

We calculate our design-based area EBLUP using (2.13), EBLUP for a sampled sub-area using (2.10), and EBLUP for a non-sampled sub-area using (2.11), with replacing variance components with its estimators.

Similar to Pfefferman and Sverchkov [16], we report the EMSE separately for sampled and non-sampled sub-areas. In the case of sampled sub-areas, we have

$$\text{EMSE}_{ij,s} = \sum_{b=1}^{B} d_{ij}^{(b)} (\hat{\mu}_{ij}^{(b)} - \bar{Y}_{ij})^2 / \sum_{b=1}^{B} d_{ij}^{(b)},$$

where $d_{ij}^{(b)}$ is 1 if the sub-area j in area i selected in the $b-$th sample and it is zero otherwise. In the case of non-sampled sub-areas, we have

$$\text{EMSE}_{ij,ns} = \sum_{b=1}^{B} (1 - d_{ij}^{(b)}) (\hat{\mu}_{ij}^{*(b)} - \bar{Y}_{ij})^2 / \sum_{b=1}^{B} (1 - d_{ij}^{(b)}).$$

The EMSE for area EBLUP is given by

$$\text{EMSE}_i = B^{-1} \sum_{b=1}^{B} (\hat{\mu}_i^{(b)} - \bar{Y}_i)^2.$$

The EMSE values averaged over areas and sampled and non-sampled sub-areas within each group were reported in Figure 1. For the moderate balanced pattern (b) in the case of Normal distribution of the v_i's and u_{ij}'s, the variability of the $\overline{\text{EMSE}}_i$ values over groups $A_1, ..., A_6$ for sampled sub-areas is comparable with the corresponding values in our model-based simulation

(Table 1) and smaller than the corresponding values for the non-sampled sub-areas. Moreover, all three methods, ML, REML and EFH, produced identical results in terms of $\overline{\text{EMSE}}_i$ for both sampled and non-sampled sub-areas. We have similar results for EMSE of area EBLUP (EMSE_i) averaged over areas within same group, which were reported in Figure 2.

We also take the Fay-Herriot sampled sub-area estimator as follows:

$$\hat{\mu}_{ij}^{FH} = x'_{ij}\hat{\beta} + \tilde{u}_{ij}(\hat{\beta}, \hat{\sigma_u^2}), \tag{5.2}$$

where $\tilde{u}_{ij}(\beta, \sigma_u^2) = \gamma_{ij}(y_{ij} - x'_{ij}\tilde{\beta})$, β and σ_u^2 were estimated based on the model (5.2). For the non-sampled sub-area, we used the regression synthetic estimator

$$\hat{\mu}_{ij}^{*FH} = x'_{ij}\hat{\beta}.$$

The values of relative efficiency of sub-area EBLUP estimator over Fay-Herriot sub-area estimator averaged over areas within the same group were reported for both sampled and non-sampled sub-areas in Figure 3. All three methods produced nearly identical results in terms of efficiency of sub-area EBLUP estimator over Fay-Herriot sub-area estimator, and relative efficiency for all three methods ranges from 103% to 124% for sampled sub-areas, and from 99% to 173% for non-sampled sub-areas. We have similar results on efficiency for area EBLUP estimator over area direct estimator averaged over areas within the same group, which were reported in Figure 4.

6. Discussion

Our simulation results indicate that the three methods of estimating model parameters (ML, REML and EFH) perform similarly in terms of MSE

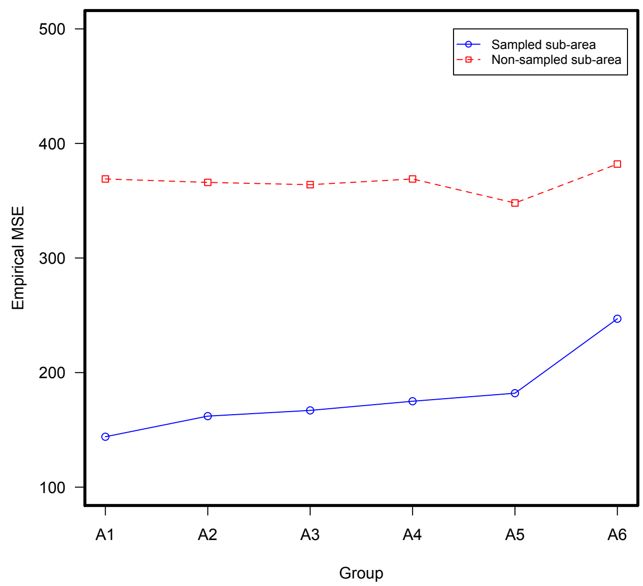

Fig. 1 Empirical mean squared error (EMSE) of sub-area EBLUP over areas and sampled and non-sampled sub-areas within the same group, pattern (b) with normal distribution for random effects, noting that the EMSE values are identical for the methods ML, REML, and EFH for both sampled and non-sampled sub-area.

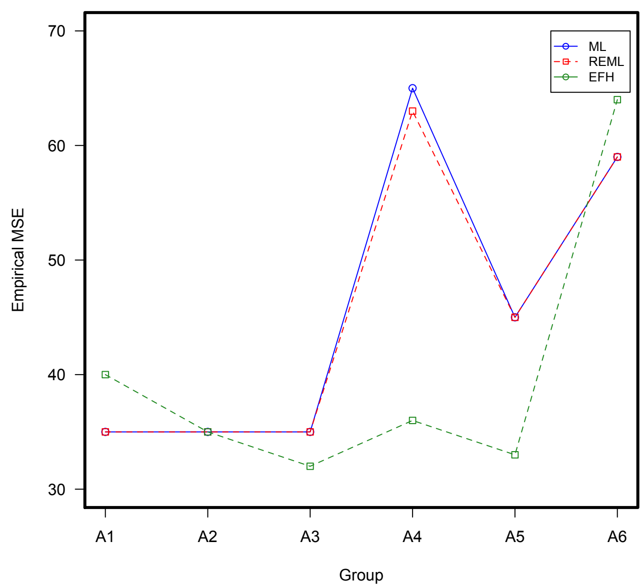

Fig. 2 Empirical mean squared error of area EBLUP over areas within the same group; pattern (b) with normal distribution for random effects.

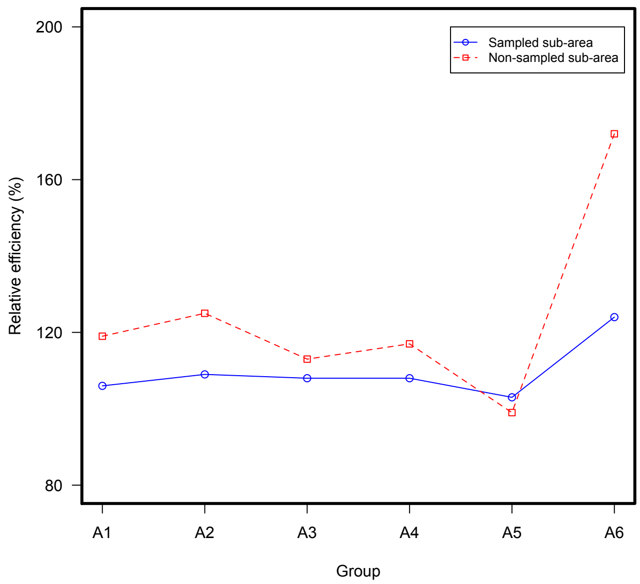

Fig. 3 Percent average relative efficiency (RE) of sub-area EBLUP over the Fay-Herriot estimator averaged over areas and sampled and non-sampled sub-areas within the same group, pattern (b) with normal distribution for random effects, noting that the RE values are identical for the methods ML, REML, and EFH for both sampled and non-sampled sub-area.

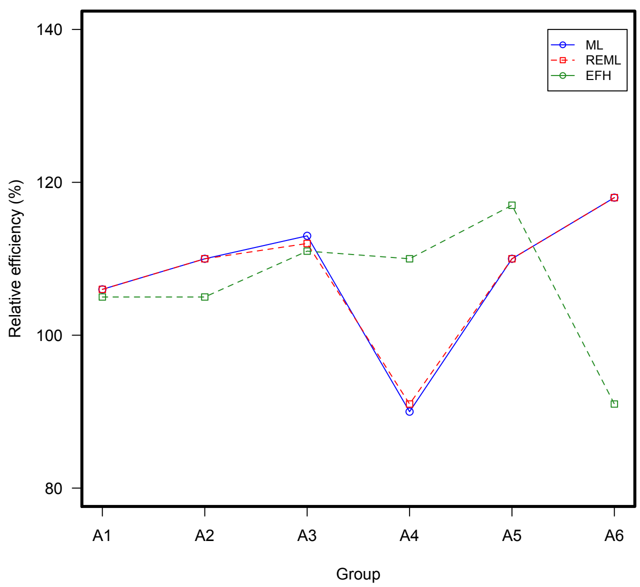

Fig. 4 Percent relative efficiency of area EBLUP over area direct estimator averaged over areas within the same group; pattern (b) with normal distribution for random effects.

of the associated EBLUP estimators of area means and sub-area means and the RB of the associated MSE estimators. Non-normality of the random effects v_i and u_{ij} has not affected MSE and its estimator significantly, even though the MSE estimator is derived under normality.

It is possible to use a jackknife method or a parametric double-bootstrap method along the lines of Jiang, Lahiri and Wan [13] and Hall and Maiti [9] to estimate the MSE of the EBLUP estimators but the methods are computer intensive and the resulting MSE estimators are less stable than the second-order correct MSE estimators and they may take negative values. Those resampling methods are suitable for cases where second-order correct MSE estimators are not tractable or difficult to derive.

A natural alternative to the EBLUP approach, proposed in this paper, is to use a hierarchical Bayes (HB) approach (Rao [17], Chapter 10). The HB approach leads to "exact" inferences, unlike the EBLUP or the empirical Bayes (EB) approach, but it requires the specification of prior distributions on the model parameters. A referee noted that within the HB approach, auxiliary information from misaligned sub-areas can also be included (see e.g., Gotway et al. [8]).

We have studied sub-area level models but unit level can also be used (Rao [17], Chapter 7). A referee suggested a unit level model of the form $y_{ijk} = x'_{ij}\beta_1 + z'_{ijk}\beta_2 + v_i + u_{ij} + e_{ijk}$ that includes both area level covariates x_{ij} and unit level covariates z_{ijk}. Inference under this model requires data (y_{ijk}, z_{ijk}) at the unit level, whereas our sub-area (say county) model requires only sub-area level data (y_{ij}, x_{ij}). Our sub-area level model also accommodates the sampling design within sub-areas through the direct estimators y_{ij} of sub-area means unlike the above unit level model.

Acknowledgements

We would like to thank two referees for constructive comments and suggestions, which led to an improved version of the manuscript. This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix A. Derivation of variances and covariance of $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$

We express the moment equation (2.14) as

$$\mathrm{y}' A_1(\delta)\mathrm{y} = n - p \tag{A. 1}$$

where $\mathrm{y} = col_{1\le i\le m}(\mathrm{y}_i)$ and

$$A_1(\delta) = V^{-1}(\delta) - V^{-1}(\delta)X[X'V^{-1}(\delta)X]^{-1}X'V^{-1}(\delta),$$

with $V(\delta) = diag_{1\le i\le m}(V_i)$ and $X = col_{1\le i\le m}(X_i)$. Similarly, the moment equation (2.16) is expressed as

$$\bar{y}' A_2(\delta)\bar{y} = m - p, \tag{A. 2}$$

where $\bar{y} = col_{1\le i\le m}(\bar{y}_i)$ and

$$A_2(\delta) = \tilde{V}^{-1}(\delta) - \tilde{V}^{-1}(\delta)\bar{x}(\bar{x}'\tilde{V}^{-1}\bar{x})^{-1}\bar{x}'\tilde{V}^{-1}(\delta),$$

with $\tilde{V}(\delta) = diag_{1\le i\le m}(\tilde{\sigma}_i^2)$ and $\bar{x} = col_{1\le i\le m}(\bar{x}_i')$.

Then by a Taylor series expansion, we have from (A. 1) and (A. 2)

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathrm{y}'A_1(\tilde{\delta})\mathrm{y}/(n-p) \\ \bar{\mathrm{y}}'A_2(\tilde{\delta})\bar{\mathrm{y}}/(m-p) \end{pmatrix} = \begin{pmatrix} \mathrm{y}'A_1(\delta)\mathrm{y}/(n-p) \\ \bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}/(m-p) \end{pmatrix}$$

$$+ \begin{pmatrix} \mathrm{y}'A_1^1(\delta)\mathrm{y}/(n-p) & \mathrm{y}'A_1^2(\delta)\mathrm{y}/(n-p) \\ \bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}}/(m-p) & \bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}}/(m-p) \end{pmatrix} \begin{pmatrix} \tilde{\sigma}_v^2 - \sigma_v^2 \\ \tilde{\sigma}_u^2 - \sigma_u^2 \end{pmatrix} + [O(\|\ \tilde{\delta} - \delta\ \|^2)]_{2\times 1}.$$

Hence,

$$\begin{pmatrix} \tilde{\sigma}_v^2 - \sigma_v^2 \\ \tilde{\sigma}_u^2 - \sigma_u^2 \end{pmatrix} = \begin{pmatrix} \mathrm{y}'A_1^1(\delta)\mathrm{y}/(n-p) & \mathrm{y}'A_1^2(\delta)\mathrm{y}/(n-p) \\ \bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}}/(m-p) & \bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}}/(m-p) \end{pmatrix}^{-1} \begin{pmatrix} 1 - \frac{\mathrm{y}'A_1(\delta)\mathrm{y}}{n-p} \\ 1 - \frac{\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}}{m-p} \end{pmatrix}$$

$$+[O(\|\ \tilde{\delta} - \delta\ \|^2)]_{2\times 1}, \tag{A. 3}$$

where $A_i^j(\delta) = \frac{dA_i(\delta)}{d\delta_j}; i, j = 1, 2$.

On the other hand, $(n-p)^{-1}\mathrm{y}'A_1(\delta)\mathrm{y} = (n-p)^{-1}E[\mathrm{y}'A_1(\delta)\mathrm{y}] + O_p(m^{-1/2}) = 1 + O_p(m^{-1/2})$, since $\mathrm{var}[(n-p)^{-1}\mathrm{y}'A_1(\delta)\mathrm{y}] = O(m^{-1})$, (Bishop et al. [1]). Similarly, $(m-p)^{-1}\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}} = (m-p)^{-1}E(\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}) + O_p(m^{-1/2}) = 1 + O_p(m^{-1/2})$, since $var[(m-p)^{-1}\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}] = O(m^{-1})$. That means, $(n-p)^{-1}\mathrm{y}'A_1(\delta)\mathrm{y} - 1$ and $(m-p)^{-1}\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}} - 1$ are of order $O_p(m^{-1/2})$.

Moreover, noting that $V_i = \mathrm{var}(\mathrm{y}_i) = \sigma_v^2 J_{n_i} + \sigma_u^2 I_{n_i} + R_i$ and $V = diag(V_1, V_2, ..., V_m)$, we get $\frac{dV}{d\sigma_v^2} = diag(J_{n_1}, ..., J_{n_m}) \equiv \tilde{J}_n$, where J_{n_i} is a matrix of ones with dimension $n_i \times n_i$ and $\frac{dV}{d\sigma_u^2} = diag(I_{n_1}, ..., I_{n_m}) = I_n$. Hence, we obtain $A_1^1(\delta) \equiv \partial A_1(\delta)/\partial\sigma_v^2 = -A_1\tilde{J}_nA_1$ and $\mathrm{y}'A_1^1(\delta)\mathrm{y} = -\mathrm{y}'A_1(\delta)\tilde{J}_nA_1(\delta)\mathrm{y}$. Therefore,

$$E(\mathrm{y}'A_1^1(\delta)\mathrm{y}) = -tr[\tilde{J}_nE\{(A_1\mathrm{y})(A_1\mathrm{y})'\}],$$

where $E\{(A_1\mathrm{y})(A_1\mathrm{y})'\} = \mathrm{var}(A_1\mathrm{y}) + E(A_1\mathrm{y})E(A_1\mathrm{y})'$ with $\mathrm{var}(A_1\mathrm{y}) = A_1 V A_1 = A_1$ and $E(A_1\mathrm{y}) = E(V^{-1}\mathrm{y} - V^{-1}X\tilde{\beta}) = V^{-1}X\beta - V^{-1}X\beta = 0$. Hence, $E[(A_1\mathrm{y})(A_1\mathrm{y})'] = A_1$.

We may further write $E(\mathrm{y}'A_1^1(\delta)\mathrm{y}) = -tr(\tilde{J}_n A_1) \approx -tr(\tilde{J}_n V^{-1}) = O(m)$ for large m, so that

$$(n-p)^{-1}\mathrm{y}'A_1^1(\delta)\mathrm{y} = -(n-p)^{-1}tr(\tilde{J}_n V^{-1}) + O_p(m^{-1/2}), \qquad \text{(A. 4)}$$

It follows that $(n-p)^{-1}\mathrm{y}'A_1^1(\delta)\mathrm{y} = O_p(1)$, since $\mathrm{var}\{(n-p)^{-1}\mathrm{y}'A_1^1(\delta)\mathrm{y}\} = O(m^{-1})$.

Similarly, $E[\mathrm{y}'A_1^2(\delta)\mathrm{y}] \equiv E(\mathrm{y}'\frac{\partial A_1(\delta)}{\partial \sigma_u^2}\mathrm{y}) \approx -tr(V^{-1}) = O(m)$, so that

$$(n-p)^{-1}\mathrm{y}'A_1^2(\delta)\mathrm{y} = -(n-p)^{-1}tr(V^{-1}) + O_p(m^{-1/2}), \qquad \text{(A. 5)}$$

and, since $\mathrm{var}[(n-p)^{-1}\mathrm{y}'A_1^2(\delta)\mathrm{y}] = O(m^{-1})$, it follows that $(n-p)^{-1}\mathrm{y}'A_1^2(\delta)\mathrm{y} = O_p(1)$.

Moreover, noting that $\tilde{\sigma}_i^2 = \sigma_v^2 + \frac{\sigma_u^2}{n_i} + \frac{\sigma_{e_i}^2}{n_i}$, and $\tilde{V} = diag_{1\le i\le m}(\tilde{\sigma}_i^2)$ such that $\frac{d\tilde{V}}{d\sigma_v^2} = I_m$ and $\frac{d\tilde{V}}{d\sigma_u^2} = diag(n_1^{-1}, ..., n_m^{-1}) \equiv \tilde{N}^{-1}$. As before, we may write $E(\bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}}) \equiv E(\bar{\mathrm{y}}'\frac{\partial A_2(\delta)}{\partial \sigma_v^2}\bar{\mathrm{y}}) \approx -tr(V^{-1}) = O(m)$, so that

$$(m-p)^{-1}\bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}} = -(m-p)^{-1}tr(\tilde{V}^{-1}) + O_p(m^{-1/2}). \qquad \text{(A. 6)}$$

Since $\mathrm{var}[(m-p)^{-1}\bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}}] = O(m^{-1})$, it follows that $(m-p)^{-1}\bar{\mathrm{y}}'A_2^1(\delta)\bar{\mathrm{y}} = O_p(1)$.

Similarly, $E(\bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}}) \equiv E(\bar{\mathrm{y}}'\frac{\partial A_2(\delta)}{\partial \sigma_u^2}\bar{\mathrm{y}}) \approx -tr(\tilde{N}^{-1}\tilde{V}^{-1}) = O(m)$, so that

$$(m-p)^{-1}\bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}} = -(m-p)^{-1}tr(\tilde{N}^{-1}\tilde{V}^{-1}) + O_p(m^{-1/2}). \qquad \text{(A. 7)}$$

Noting that $\mathrm{var}[(m-p)^{-1}\bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}}] = O(m^{-1})$, we get $(m-p)^{-1}\bar{\mathrm{y}}'A_2^2(\delta)\bar{\mathrm{y}} = O_p(1)$.

Hence, it follows from (A. 3) that $\tilde{\sigma}_v^2 - \sigma_v^2 = O_p(m^{-1/2})$ and $\tilde{\sigma}_u^2 - \sigma_u^2 = O_p(m^{-1/2})$. Therefore, by (A. 4)-(A. 7), we may write (A. 3) as

$$\begin{pmatrix} \tilde{\sigma}_v^2 - \sigma_v^2 \\ \tilde{\sigma}_u^2 - \sigma_u^2 \end{pmatrix} = -\begin{pmatrix} tr(\tilde{J}_n V^{-1})/(n-p) & tr(V^{-1})/(n-p) \\ tr(\tilde{V}^{-1})/(m-p) & tr(\tilde{N}^{-1}\tilde{V}^{-1})/(m-p) \end{pmatrix}^{-1}$$

$$\cdot \begin{pmatrix} 1 - \frac{\mathrm{y}'A_1(\delta)\mathrm{y}}{n-p} \\ 1 - \frac{\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}}{m-p} \end{pmatrix} + [O_p(m^{-1})]_{2\times 1}. \tag{A. 8}$$

Then, the asymptotic variance-covariance of $\tilde{\delta}$ is given by

$$\begin{pmatrix} \mathrm{var}(\tilde{\sigma}_v^2) & \mathrm{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2) \\ \mathrm{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2) & \mathrm{var}(\tilde{\sigma}_u^2) \end{pmatrix} = \begin{pmatrix} \frac{tr(\tilde{J}_n V^{-1})}{n-p} & \frac{tr(V^{-1})}{n-p} \\ \frac{tr(\tilde{V}^{-1})}{m-p} & \frac{tr(\tilde{N}^{-1}\tilde{V}^{-1})}{m-p} \end{pmatrix}^{-1}$$

$$\cdot \begin{pmatrix} \mathrm{var}(\frac{\mathrm{y}'A_1(\delta)\mathrm{y}}{n-p}) & \mathrm{cov}(\frac{\mathrm{y}'A_1(\delta)\mathrm{y}}{n-p}, \frac{\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}}{m-p}) \\ \mathrm{cov}(\frac{\mathrm{y}'A_1(\delta)\mathrm{y}}{n-p}, \frac{\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}}{m-p}) & \mathrm{var}(\frac{\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}}{m-p}) \end{pmatrix} \begin{pmatrix} \frac{tr(\tilde{J}_n V^{-1})}{n-p} & \frac{tr(V^{-1})}{n-p} \\ \frac{tr(\tilde{V}^{-1})}{m-p} & \frac{tr(\tilde{N}^{-1}\tilde{V}^{-1})}{m-p} \end{pmatrix}^{-T},$$

where

$$\mathrm{var}[\mathrm{y}'A_1(\delta)\mathrm{y}/(n-p)] = 2(n-p)^{-1}, \tag{A. 9}$$

$$\mathrm{var}[\bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}/(m-p)] = 2(m-p)^{-1}, \tag{A. 10}$$

$$\begin{aligned} \mathrm{cov}[\mathrm{y}'A_1(\delta)\mathrm{y}, \bar{\mathrm{y}}'A_2(\delta)\bar{\mathrm{y}}] &= \mathrm{cov}[\mathrm{y}'A_1(\delta)\mathrm{y}, \mathrm{y}'\tilde{N}_1^{-T}A_2(\delta)\tilde{N}_1^{-1}\mathrm{y}] \\ &= 2tr[A_1(\delta)V\tilde{N}_1^{-T}A_2(\delta)\tilde{N}_1^{-1}V], \end{aligned} \tag{A. 11}$$

and $\tilde{N}_1^{-1}\mathrm{y} = \bar{\mathrm{y}}$ with

$$\tilde{N}_1^{-1} = \begin{pmatrix} n_1^{-1}\mathbf{1}_{n_1} & 0\mathbf{1}_{n_1} & \dots & 0\mathbf{1}_{n_1} \\ 0\mathbf{1}_{n_2} & n_2^{-1}\mathbf{1}_{n_2} & \dots & 0\mathbf{1}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0\mathbf{1}_{n_m} & 0\mathbf{1}_{n_m} & \dots & n_m^{-1}\mathbf{1}_{n_m} \end{pmatrix}'.$$

In fact, to get (A. 9)- (A. 11) we used the following Lemma.

LEMMA 2 (Searle [18]): *If $x \sim N(\mu, V)$ where x is a vector, $x'Px$ and $x'Qx$ are two quadratic forms such that P and Q are square matrices, then we have that*

$$\text{cov}(x'Px, x'Qx) = 2tr(PVQV) + 4\mu' PVQ\mu.$$

Hence, combining (A. 9)-(A. 11), the asymptotic variance-covariance of $\tilde{\delta}$ is well-defined. To further simplify, we first note that

$$V_i^{-1} = diag_j[(\sigma_u^2 + \sigma_{eij}^2)^{-1}] - \frac{\sigma_v^2 \sigma_u^2}{\sigma_u^2 + \sigma_v^2 \gamma_{i.}} w_i w_i'.$$

Hence,

$$tr(V^{-1}) = \frac{1}{\sigma_u^2} \sum_{i=1}^{m} \{\gamma_{i.} - \frac{\gamma_i}{\gamma_{i.}} \sum_{j=1}^{n_i} \gamma_{ij}^2\}, \tag{A. 12}$$

$$tr(\tilde{J}_n V^{-1}) = \frac{1}{\sigma_u^2} \sum_{i=1}^{m} \gamma_{i.}(1 - \gamma_i). \tag{A. 13}$$

Furthermore, we have that

$$tr(\tilde{V}^{-1}) = \sum_{i=1}^{m} n_i (n_i \sigma_v^2 + \sigma_u^2 + \sigma_{ei}^2)^{-1}, \tag{A. 14}$$

$$tr(\tilde{N}^{-1} \tilde{V}^{-1}) = \sum_{i=1}^{m} (n_i \sigma_v^2 + \sigma_u^2 + \sigma_{ei}^2)^{-1}. \tag{A. 15}$$

Therefore, we may write the variances and the covariance of $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$ as

$$\text{var}(\tilde{\sigma}_v^2) = 2[tr(\tilde{J}_n V^{-1}) tr(\tilde{N}^{-1}\tilde{V}^{-1}) - tr(\tilde{V}^{-1}) tr(V^{-1})]^{-2} \{n[tr(\tilde{N}^{-1}\tilde{V}^{-1})]^2$$

$$+ m[tr(V^{-1})]^2 - 2tr(V^{-1}) tr(\tilde{N}^{-1}\tilde{V}^{-1}) tr(A_1 V \tilde{N}_1^{-T} A_2 \tilde{N}_1^{-1} V)\} + o(m^{-1}),$$

$$\text{var}(\tilde{\sigma}_u^2) = 2[tr(\tilde{J}_n V^{-1}) tr(\tilde{N}^{-1}\tilde{V}^{-1}) - tr(\tilde{V}^{-1}) tr(V^{-1})]^{-2} \{n[tr(\tilde{V}^{-1})]^2$$

$$+m[tr(\tilde{J}_nV^{-1})]^2-2tr(\tilde{V}^{-1})tr(\tilde{J}_nV^{-1})tr(A_1V\tilde{N}_1^{-T}A_2\tilde{N}_1^{-1}V)\}+o(m^{-1}),$$

$$\mathrm{cov}(\tilde{\sigma}_v^2,\tilde{\sigma}_u^2)=2\Big[tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(\tilde{V}^{-1})tr(V^{-1})\Big]^{-2}\Big\{-ntr(\tilde{V}^{-1})$$

$$.tr(\tilde{N}^{-1}\tilde{V}^{-1})-mtr(V^{-1})tr(\tilde{J}_nV^{-1})+\Big[tr(V^{-1})tr(\tilde{V}^{-1})+tr(\tilde{N}^{-1}\tilde{V}^{-1})tr(\tilde{J}_nV^{-1})\Big]$$

$$.tr(A_1V\tilde{N}_1^{-T}A_2\tilde{N}_1^{-1}V)\Big\}+o(m^{-1}),$$

where $tr(V^{-1}),tr(\tilde{J}_nV^{-1}),tr(\tilde{V}^{-1})$ and $tr(\tilde{N}^{-1}\tilde{V}^{-1})$ are given by (A. 12)-(A. 15) respectively.

Appendix B. Derivation of bias of $\hat{\sigma}_v^2$ and $\hat{\sigma}_u^2$

We now find expressions for bias terms $b(\hat{\sigma}_v^2)$ and $b(\hat{\sigma}_u^2)$. We have $b(\hat{\delta}_i)=E(\hat{\delta}_i-\delta_i)=E(\tilde{\delta}_i-\delta_i)+E(\hat{\delta}_i-\tilde{\delta}_i)$ for $i=1,2$, where $E(\hat{\delta}_i-\tilde{\delta}_i)=-E[\tilde{\delta}_iI(\tilde{\delta}_i\le 0)]\le[E(\tilde{\delta}_i^2)]^{1/2}[Pr(\tilde{\delta}_i\le 0)]^{1/2}=O(m^{-2})$. To get $E(\tilde{\delta}_i-\delta_i)=b(\tilde{\delta}_i),E(\tilde{\sigma}_v^2-\sigma_v^2)$ and $E(\tilde{\sigma}_u^2-\sigma_u^2)$, we follow Datta et al. [5] to derive the bias of $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_u^2$. By (A. 3), letting $A_k(\delta)=A_k,A_k^1(\delta)=A_k^1$ and $A_k^2(\delta)=A_k^2$ for $k=1,2$, we have

$$\frac{1}{n-p}\Big\{\mathrm{y}'A_1\mathrm{y}-(n-p)+[\mathrm{y}'A_1^1\mathrm{y}+tr(\tilde{J}_nA_1)](\tilde{\sigma}_v^2-\sigma_v^2)+[\mathrm{y}'A_1^2\mathrm{y}+tr(A_1)](\tilde{\sigma}_u^2-\sigma_u^2)$$

$$-(\tilde{\sigma}_v^2-\sigma_v^2)tr(\tilde{J}_nA_1)-(\tilde{\sigma}_u^2-\sigma_u^2)tr(A_1)+\frac{1}{2}tr(VA_1^{11})(\tilde{\sigma}_v^2-\sigma_v^2)^2+\frac{1}{2}tr(VA_1^{22})(\tilde{\sigma}_u^2-\sigma_u^2)^2$$

$$+tr(VA_1^{12})(\tilde{\sigma}_v^2-\sigma_v^2)(\tilde{\sigma}_u^2-\sigma_u^2)\Big\}+o_p(m^{-1})=0 \qquad \text{(A. 16)}$$

and

$$\frac{1}{m-p}\Big\{\bar{\mathrm{y}}'A_2\bar{\mathrm{y}}-(m-p)+[\bar{\mathrm{y}}'A_2^1\bar{\mathrm{y}}+tr(A_2)](\tilde{\sigma}_v^2-\sigma_v^2)+[\bar{\mathrm{y}}'A_2^2\bar{\mathrm{y}}+tr(\tilde{N}^{-1}A_2)](\tilde{\sigma}_u^2-\sigma_u^2)$$

$$-(\tilde{\sigma}_v^2-\sigma_v^2)tr(A_2)-(\tilde{\sigma}_u^2-\sigma_u^2)tr(\tilde{N}^{-1}A_2)+\frac{1}{2}tr(\tilde{V}A_2^{11})(\tilde{\sigma}_v^2-\sigma_v^2)^2+\frac{1}{2}tr(\tilde{V}A_2^{22})(\tilde{\sigma}_u^2-\sigma_u^2)^2$$

$$+tr(\tilde{V}A_2^{12})(\tilde{\sigma}_v^2-\sigma_v^2)(\tilde{\sigma}_u^2-\sigma_u^2)\Big\}+o_p(m^{-1})=0, \qquad \text{(A. 17)}$$

where $A_k^{ij}(\delta) = \frac{d^2 A_k(\delta)}{d\delta_i d\delta_j}, (k = 1, 2)$. We then take expectations of (A. 16) and (A. 17) to get

$$\frac{1}{n-p}\Big\{\text{cov}(\text{y}'A_1^1\text{y}, \tilde{\sigma}_v^2) + \text{cov}(\text{y}'A_1^2\text{y}, \tilde{\sigma}_u^2) - tr(\tilde{J}_n A_1)b(\tilde{\sigma}_v^2) - tr(A_1)b(\tilde{\sigma}_u^2)$$

$$+\frac{1}{2}tr(VA_1^{11})\text{var}(\tilde{\sigma}_v^2)+\frac{1}{2}tr(VA_1^{22})\text{var}(\tilde{\sigma}_u^2)+tr(VA_1^{12})\text{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)\Big\}+o(m^{-1}) = 0 \quad \text{(A. 18)}$$

and

$$\frac{1}{m-p}\Big\{\text{cov}(\bar{\text{y}}'A_2^1\bar{\text{y}}, \tilde{\sigma}_v^2) + \text{cov}(\bar{\text{y}}'A_2^2\bar{\text{y}}, \tilde{\sigma}_u^2) - tr(A_2)b(\tilde{\sigma}_v^2) - tr(\tilde{N}^{-1}A_2)b(\tilde{\sigma}_u^2)$$

$$+\frac{1}{2}tr(\tilde{V}A_2^{11})\text{var}(\tilde{\sigma}_v^2)+\frac{1}{2}tr(\tilde{V}A_2^{22})\text{var}(\tilde{\sigma}_u^2)+tr(\tilde{V}A_2^{12})\text{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)\Big\}+o(m^{-1}) = 0. \quad \text{(A. 19)}$$

Therefore, we can find approximations to $b(\tilde{\sigma}_v^2)$ and $b(\tilde{\sigma}_u^2)$ from (A. 18) and (A. 19) as

$$\begin{aligned} b(\tilde{\sigma}_v^2) = \Big[\frac{tr(A_1)tr(A_2)}{tr(\tilde{N}^{-1}A_2)} - tr(\tilde{J}_n A_1)\Big]^{-1}\Big\{ &- \text{cov}(\text{y}'A_1^1\text{y}, \tilde{\sigma}_v^2) - \text{cov}(\text{y}'A_1^2\text{y}, \tilde{\sigma}_u^2) \\ &-tr(A_1\tilde{J}_n A_1 \tilde{J}_n)\text{var}(\tilde{\sigma}_v^2) - tr(A_1A_1)\text{var}(\tilde{\sigma}_u^2) - 2tr(A_1A_1\tilde{J}_n)\text{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2) \\ &+\frac{tr(A_1)}{tr(\tilde{N}^{-1}A_2)}[\text{cov}(\bar{\text{y}}'A_2^1\bar{\text{y}}, \tilde{\sigma}_v^2) + tr(A_2A_2)\text{var}(\tilde{\sigma}_v^2) + \text{cov}(\bar{\text{y}}'A_2^2\bar{\text{y}}, \tilde{\sigma}_u^2) \\ &+tr(A_2\tilde{N}^{-1}A_2\tilde{N}^{-1})\text{var}(\tilde{\sigma}_u^2) + 2tr(A_2A_2\tilde{N}^{-1})\text{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2)]\Big\} \end{aligned} \quad \text{(A. 20)}$$

and

$$\begin{aligned} b(\tilde{\sigma}_u^2) = \Big[\frac{tr(A_1)tr(A_2)}{tr(\tilde{J}_n A_1)} - tr(\tilde{N}^{-1}A_2)\Big]^{-1}\Big\{ &- \text{cov}(\bar{\text{y}}'A_2^1\bar{\text{y}}, \tilde{\sigma}_v^2) - \text{cov}(\bar{\text{y}}'A_2^2\bar{\text{y}}, \tilde{\sigma}_u^2) \\ -tr(A_2A_2)\text{var}(\tilde{\sigma}_v^2) - tr(A_2\tilde{N}^{-1}A_2\tilde{N}^{-1})&\text{var}(\tilde{\sigma}_u^2) - 2tr(A_2A_2\tilde{N}^{-1})\text{cov}(\tilde{\sigma}_v^2, \tilde{\sigma}_u^2) \\ +\frac{tr(A_2)}{tr(\tilde{J}_n A_1)}[\text{cov}(\text{y}'A_1^1\text{y}, \tilde{\sigma}_v^2) + tr(A_1\tilde{J}_n A_1\tilde{J}_n)&\text{var}(\tilde{\sigma}_v^2) + \text{cov}(\text{y}'A_1^2\text{y}, \tilde{\sigma}_u^2) \end{aligned}$$

$$+tr(A_1A_1)\text{var}(\tilde{\sigma}_u^2)+2tr(A_1A_1\tilde{J}_n)\text{cov}(\tilde{\sigma}_v^2,\tilde{\sigma}_u^2)]\Big\}, \qquad \text{(A. 21)}$$

where $\text{cov}(\text{y}'A_1^1\text{y},\tilde{\sigma}_v^2)$, $\text{cov}(\text{y}'A_1^2\text{y},\tilde{\sigma}_u^2)$, $\text{cov}(\bar{\text{y}}'A_2^1\bar{\text{y}},\tilde{\sigma}_v^2)$ and $\text{cov}(\bar{\text{y}}'A_2^2\bar{\text{y}},\tilde{\sigma}_u^2)$ are calculated below.

By (A. 8), we may write

$$\tilde{\sigma}_v^2-\sigma_v^2=\frac{tr(\tilde{N}^{-1}\tilde{V}^{-1})[\text{y}'A_1\text{y}-(n-p)]-tr(V^{-1})[\bar{\text{y}}'A_2\bar{\text{y}}-(m-p)]}{tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(V^{-1})tr(\tilde{V}^{-1})}+O_p(m^{-1}).$$

Now using Lemma 2, we get

$$\text{cov}(\text{y}'A_1^1\text{y},\tilde{\sigma}_v^2)=2\Big[tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(V^{-1})tr(\tilde{V}^{-1})\Big]^{-1}$$

$$.\Big[-tr(\tilde{N}^{-1}\tilde{V}^{-1})tr(A_1\tilde{J}_n)-tr(V^{-1})tr(A_1^1V\tilde{N}_1^{-T}A_2\tilde{N}_1^{-1}V)\Big]+o(m^{-1})$$

and

$$\text{cov}(\bar{\text{y}}'A_2^1\bar{\text{y}},\tilde{\sigma}_v^2)=2[tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(V^{-1})tr(\tilde{V}^{-1})]^{-1}\Big[tr(\tilde{N}^{-1}\tilde{V}^{-1})$$

$$.tr(\tilde{N}_1^{-T}A_2^1\tilde{N}_1^{-1}VA_1V)-tr(V^{-1})tr(\tilde{N}_1^{-T}A_2^1\tilde{N}_1^{-1}V\tilde{N}_1^{-T}A_2\tilde{N}_1^{-1}V)\Big]+o(m^{-1}).$$

Similarly, from (A. 8) we have

$$\tilde{\sigma}_u^2-\sigma_u^2=\frac{-tr(\tilde{V}^{-1})[\text{y}'A_1\text{y}-(n-p)]+tr(\tilde{J}_nV^{-1})[\bar{\text{y}}'A_2\bar{\text{y}}-(m-p)]}{tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(V^{-1})tr(\tilde{V}^{-1})}+O_p(m^{-1}).$$

Again using Lemma 2, we have

$$\text{cov}(\text{y}'A_1^2\text{y},\tilde{\sigma}_u^2)=2\Big[tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}V^{*-1})-tr(V^{-1})tr(V^{*-1})\Big]^{-1}$$

$$.\Big[tr(V^{*-1})tr(A_1)+tr(\tilde{J}_nV^{-1})tr(A_1^2V\tilde{N}_1^{-T}A_2\tilde{N}_1^{-1}V)\Big]+o(m^{-1})$$

and

$$\text{cov}(\bar{\text{y}}'A_2^2\bar{\text{y}},\tilde{\sigma}_u^2)=2\Big[tr(\tilde{J}_nV^{-1})tr(\tilde{N}^{-1}\tilde{V}^{-1})-tr(V^{-1})tr(\tilde{V}^{-1})\Big]^{-1}\Big[-tr(\tilde{V}^{-1})$$

$.tr(\tilde{N}_1^{-T} A_2^2 \tilde{N}_1^{-1} V A_1 V) + tr(\tilde{J}_n V^{-1}) tr(\tilde{N}_1^{-T} A_2^2 \tilde{N}_1^{-1} V \tilde{N}_1^{-T} A_2 \tilde{N}_1^{-1} V)\Big] + o(m^{-1}).$

Finally, combining above results, we obtain bias approximations $b(\tilde{\sigma}_v^2)$ and $b(\tilde{\sigma}_u^2)$ from (A. 20) and (A. 21) respectively.

Appendix C. Proof of Theorem 1

We must show that mse$(\hat{\mu}_{ij})$ is correct to $O_p(m^{-1})$. By Taylor expansion,

$$g_{1ij}(\hat{\delta}) = g_{1ij}(\delta) + (\hat{\delta}-\delta)'\nabla g_{1ij}(\delta) + \frac{1}{2}(\hat{\delta}-\delta)'\nabla^2 g_{1ij}(\delta)(\hat{\delta}-\delta) + o_p(m^{-1}).$$

Now taking expectation of the above equation,

$$E[g_{1ij}(\hat{\delta})] = g_{1ij}(\delta) + E[(\hat{\delta}-\delta)'\nabla g_{1ij}(\delta)] + \frac{1}{2}E[(\hat{\delta}-\delta)'\nabla^2 g_{1ij}(\delta)(\hat{\delta}-\delta)] + o(m^{-1}),$$

where

$$E[(\hat{\delta}-\delta)'\nabla g_{1ij}(\delta)] = b(\hat{\sigma}_v^2)\frac{dg_{1ij}}{d\sigma_v^2} + b(\hat{\sigma}_u^2)\frac{dg_{1ij}}{d\sigma_u^2}.$$

In addition, we have

$$E[(\hat{\delta}-\delta)'\nabla^2 g_{1ij}(\delta)(\hat{\delta}-\delta)] = \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E[(\hat{\delta}_k-\delta_k)(\hat{\delta}_l-\delta_l)]$$

$$= \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\hat{\delta}_k-\tilde{\delta}_k+\tilde{\delta}_k-\delta_k)(\hat{\delta}_l-\tilde{\delta}_l+\tilde{\delta}_l-\delta_l)\Big]$$

$$= \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\tilde{\delta}_k-\delta_k)(\tilde{\delta}_l-\delta_l)\Big] + \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\hat{\delta}_k-\tilde{\delta}_k)(\hat{\delta}_l-\tilde{\delta}_l)\Big]$$

$$+ \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\hat{\delta}_k-\tilde{\delta}_k)(\tilde{\delta}_l-\delta_l)\Big] + \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\tilde{\delta}_k-\delta_k)(\hat{\delta}_l-\tilde{\delta}_l)\Big]$$

$$= \sum_{k=1}^{2}\sum_{l=1}^{2}\nabla^2 g_{1ijkl}(\delta)E\Big[(\tilde{\delta}_k-\delta_k)(\tilde{\delta}_l-\delta_l)\Big] + o(m^{-1})$$

$$= tr\Big[\mathrm{var}(\tilde{\delta})\nabla^2 g_{1ij}(\delta)\Big] + o(m^{-1}),$$

since by applying Holder's inequality

$$E[(\hat{\delta}_k - \tilde{\delta}_k)(\hat{\delta}_l - \tilde{\delta}_l)] \le (E|\hat{\delta}_k - \tilde{\delta}_k|^2)^{1/2}.(E|\hat{\delta}_l - \tilde{\delta}_l|^2)^{1/2} = o(m^{-1}),$$

and noting that $\nabla^2 g_{1ijkl}(\delta) = O(1)$, $E|\hat{\delta}_l - \tilde{\delta}_l|^2 = O(m^{-2})$ and $E|\hat{\delta}_k - \tilde{\delta}_k|^2 = O(m^{-2})$ by the following argument:

$$E|\hat{\delta}_k - \tilde{\delta}_k|^2 = E\Big[|\tilde{\delta}_k|^2 I(\tilde{\delta}_k \le 0)\Big]$$

$$\le (E|\tilde{\delta}_k|^4)^{1/2}.\{E[I(\tilde{\delta}_k \le 0)]^2\}^{1/2} = O(m^{-2}),$$

by Holder's inequality, since $\{E[I(\tilde{\delta}_k \le 0)]^2\}^{1/2} = [Pr(\tilde{\delta}_k \le 0)]^{1/2} = O(m^{-2})$ and

$$E|\tilde{\delta}_k - \delta_k + \delta_k|^4 = E\Big[(\tilde{\delta}_k - \delta_k)^2 + \delta_k^2 + 2\delta_k|\tilde{\delta}_k - \delta_k|\Big]^2$$

$$\le 2\Big\{E[(\tilde{\delta}_k - \delta_k)^2 + \delta_k^2]^2 + 4E[\delta_k^2(\tilde{\delta}_k - \delta_k)]^2\Big\}$$

$$= 2\Big[E(\tilde{\delta}_k - \delta_k)^4 + \delta_k^4 + 2\delta_k^2 E(\tilde{\delta}_k - \delta_k)^2 + 4\delta_k^2 E(\tilde{\delta}_k - \delta_k)^2\Big] = O(1).$$

Similarly, $E[(\hat{\delta}_k - \tilde{\delta}_k)(\tilde{\delta}_l - \delta_l)] \le (E|\hat{\delta}_k - \tilde{\delta}_k|^2)^{1/2}.(E|\tilde{\delta}_l - \delta_l|^2)^{1/2} = o(m^{-1})$, by noting that $E|\tilde{\delta}_l - \delta_l|^2 = \mathrm{var}(\tilde{\delta}_l) = o(m^{-1})$ and $E|\hat{\delta}_k - \tilde{\delta}_k|^2 = o(m^{-1})$. By a similar argument, $E[(\tilde{\delta}_k - \delta_k)(\hat{\delta}_l - \tilde{\delta}_l)] = o(m^{-1})$. Combining above results, we have

$$E[g_{1ij}(\hat{\delta})] = g_{1ij}(\delta) + b(\hat{\sigma}_v^2)\frac{dg_{1ij}(\delta)}{d\sigma_v^2} + b(\hat{\sigma}_u^2)\frac{dg_{1ij}(\delta)}{d\sigma_u^2} + \frac{1}{2}tr[\mathrm{var}(\tilde{\delta})\nabla^2 g_{1ij}(\delta)] + o(m^{-1}).$$

It is easy to show that $\frac{1}{2}tr[\mathrm{var}(\tilde{\delta})\nabla^2 g_{1ij}(\delta)] = -g_{3ij}(\delta)$.

Hence,

$$E[g_{1ij}(\hat{\delta})] = g_{1ij}(\delta) + b(\hat{\sigma}_v^2)\frac{dg_{1ij}}{d\sigma_v^2} + b(\hat{\sigma}_u^2)\frac{dg_{1ij}}{d\sigma_u^2} - g_{3ij}(\delta) + o(m^{-1}).$$

Now, since $g_{2ij}(\delta)$ and $g_{3ij}(\delta)$ are of order $O(m^{-1})$, it follows that

$$E[g_{2ij}(\hat{\delta})] = g_{2ij}(\delta) + o(m^{-1}),$$

and

$$E[g_{3ij}(\hat{\delta})] = g_{3ij}(\delta) + o(m^{-1}).$$

Therefore, $g_{2ij}(\hat{\delta})$ and $g_{3ij}(\hat{\delta})$ are the correct estimators of $g_{2ij}(\delta)$ and $g_{3ij}(\delta)$ respectively, each with bias $o(m^{-1})$. Combining above results, we have that

$$E[\text{mse}(\hat{\mu}_{ij})] = \text{MSE}(\hat{\mu}_{ij}) + o(m^{-1}).$$

[1] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis: Theory and Practice, MIT, Cambridge, Mass, 1975.

[2] W.G. Cochran, Sampling Techniques, 3rd edn, Wiley, New York, 1977.

[3] K. Das, J. Jiang, J.N.K. Rao, Mean squared error of empirical predictor. Ann. Statist. 32 (2004) 818–840.

[4] G.S. Datta, Model-based approach to small area estimation. In Sample Surveys: Inference and Analysis (eds D. Pfeffermann & C. R. Rao), Handbook of Statistics 29B 251–288, North-Holland, Amsterdam, 2009.

[5] 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.

[6] R.E. Fay, R.A. Herriot, Estimates of income for small places: an application of James-Stein procedure to census data. J. Am. Statist. Assoc. 74 (1979) 269–277.

[7] W.A. Fuller, J.J. Goyeneche, Estimation of the state variance component.*A draft paper* (1998).

[8] C.A. Gotway, L. J. Young, Combining incompatible spatial data. J. Am. Statist. Assoc. 97 (2002) 632–648.

[9] P. Hall, T. Maiti, On parametric bootstrap methods for small area prediction. J. Roy. Statist. Soc. Ser. B. 68 (2006) 221-238.

[10] C.R. Henderson, Best linear unbiased estimation and prediction under a selection model. Biometrics 31 (1975) 423–447.

[11] J. Jiang, Large Sample Techniques for Statistics, Springer, New York, 2010.

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

[13] J. Jiang, P. Lahiri, S-M. Wan, A unified jackknife theory for empirical best prediction with M-estimation. Ann. Statist. 30 (2002) 1782–1810.

[14] P. Lahiri, J.N.K. Rao, Robust estimation of mean squared error of small area estimators. J. Am. Statist. Assoc. 90 (1995) 758–766.

[15] S.L. Lohr, Sampling: Design and Analysis, Duxbury, Pacific Grove, California, 1999.

[16] D. Pfefferman, M. Sverchkov, Small-area estimation under informative probability sampling of areas and within the selected areas. J. Am. Statist. Assoc. 102 (2007) 1427–1439.

[17] J.N.K. Rao, Small Area Estimation, Wiley, Hoboken, New Jersey, 2003.

[18] S.R. Searle, Linear Models, Wiley, New York, 1971.

[19] M.E. Thompson, Theory of Sample Surveys, Chapman & Hall, London, 1997.

[20] M. Torabi, Some Contributions to Small Area Estimation, Unpublished Ph.D. Thesis, Carleton University, Ottawa, Canada, 2007.