The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1–32
La revue canadienne de statistique

Robust Estimation of Mean Squared Prediction Error in Small-area Estimation

Ping Wu[1] and Jiming Jiang[2]*

[1]*Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai, China*
[2]*University of California, Davis, USA*

Key words and phrases: Best linear unbiased prediction; jackknife; mean-squared prediction error; moment-matching bootstrap; nested-error regression.

MSC 2010: Primary 62D05; secondary 62F40

Abstract: The nested-error regression model is one of the best-known models in small area estimation. A small area mean is often expressed as a linear combination of fixed effects and realized values of random effects. In such analyses, prediction is made by borrowing strength from other related areas or sources, and mean-squared prediction error (MSPE) is often used as a measure of uncertainty. In this paper, we propose a bias-corrected analytical estimation of MSPE as well as a moment-match jackknife method to estimate the MSPE without specific assumptions about the distributions of the data. Theoretical and empirical studies are carried out to investigate performance of the proposed methods with comparison to existing procedures. *The Canadian Journal of Statistics* xx: 1–32; 20?? © 20?? Statistical Society of Canada

Résumé: Insérer votre résumé ici. We will supply a French abstract for those authors who can't prepare it themselves. *La revue canadienne de statistique* xx: 1–32; 20?? © 20?? Société statistique du Canada

1. INTRODUCTION

Considerable attention has been given in recent years to small area estimation (SAE) due to an increasing demand in applications in various federal and local governments. Statistical models are used to borrow strength from related areas or sources in order to overcome insufficiency of sample size and provide reliable estimates. See, for example, Pfeffermann (2013), and Rao & Molina (2015), for reviews of important recent developments in SAE.

One of the best-known models in SAE is the nested-error regression (NER) model, first introduced by Battese, Harter, & Fuller (1988). Assume that the data $Y_{ij}, i = 1, \dots, m, j = 1, \dots, n_i$ are clustered such that there is independence between clusters, that is, $Y_i = (Y_{ij})_{1\le j\le n_i}, i = 1, \dots, m$, are independent, but correlated within clusters. Specifically, the NER model can be expressed as

$$Y_{ij} = X_{ij}'\beta + b_i + d_{ij}\varepsilon_{ij}, \qquad i = 1, \dots, m,\ j = 1, \dots, n_i, \tag{1}$$

where X_{ij} is a p-vector of known covariates, β is a p-vector of unknown fixed effects, n_i is the numbers of sampled units from the ith area, and d_{ij} is a known scalar. Furthermore, b_i is an area-specific random effects and ε_{ij} are sampling errors. It is assumed that the random effects b_i are i.i.d. with mean zero and variance $\sigma_b^2 > 0$, that the errors ε_{ij} are i.i.d. with mean zero and variance $\sigma_\varepsilon^2 > 0$, and that the random effects and errors are mutually independent.

* *Author to whom correspondence may be addressed.*
E-mail: jimjiang@ucdavis.edu

In the literature, there are some different assumptions for the sampling variance $\mathrm{Var}(\varepsilon_{\mathrm{ij}})$ in model (1). In the simplest case, $d_{ij} = 1$ for all i, j, and this model is called a nested error linear model with equal error variances. Battese, Harter, & Fuller (1988) used this model to predict the areas of corn and soybeans for 12 counties in North Central Iowa. In fact, d_{ij} always depends on the covariate X_{ij}. By choosing d_{ij} to be the square root of some covariate, Rao & Choudhry (1995) studied the population of unincorporated tax filers from the province of Nova Scotia, Canada. The sampling variance $\mathrm{Var}(\varepsilon_{\mathrm{ij}})$ is assumed to be a function of some unknown parameters as in Sugasawa & Kubokawa (2017). In this paper, we study the case with known d_{ij}.

Under model (1), the best linear unbiased predictor (BLUP) of the small area mean, $\theta_i = \bar{X}_{i.}'\beta + b_i$, where $\bar{X}_{i.}'$ is the population mean of the covariates X_{ij} for area i, can be expressed as

$$\tilde{\theta}_i = \bar{X}_{i.}'\tilde{\beta} + \frac{\sigma_b^2}{\sigma_b^2 1_{n_i}' D_i^{-1} 1_{n_i} + \sigma_\varepsilon^2} 1_{n_i}' D_i^{-1}(Y_i - X_i\tilde{\beta}), \tag{2}$$

where $X_i = (X_{i1}', \ldots, X_{in_i}')'$, $D_i = \mathrm{diag}(\mathrm{d}_{\mathrm{i1}}^2, \ldots, \mathrm{d}_{\mathrm{in_i}}^2)$, $\tilde{\beta}$ is the weighted least square estimator defined in Equation (4) below, and where 1_{n_i} denotes the n_i-vector with a value of 1 for each element. Once some consistent estimators, $\hat{\sigma}_b^2$ and $\hat{\sigma}_\varepsilon^2$, are obtained, the corresponding empirical BLUP (EBLUP) or two-stage predictor is given by

$$\hat{\theta}_i = \bar{X}_{i.}'\hat{\beta} + \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 1_{n_i}' D_i^{-1} 1_{n_i} + \hat{\sigma}_\varepsilon^2} 1_{n_i}' D_i^{-1}(Y_i - X_i\hat{\beta}), \tag{3}$$

where $\hat{\beta}$ is $\tilde{\beta}$ with σ_b^2 and σ_ε^2 replaced by $\hat{\sigma}_b^2$ and $\hat{\sigma}_\varepsilon^2$, respectively.

For the EBLUP defined above, it is not necessary to assume normality of the data. However, normality is often needed to derive an estimator of the mean-squared prediction error (MSPE) of $\hat{\theta}_i$, which is widely used as a measure of uncertainty (e.g., Rao & Molina, 2015). Under the normality assumption, Kackar & Harville (1984) and Harville & Jeske (1992) studied various approximations to the MSPE. Prasad & Rao (1990) studied accuracy of a second-order approximation by the Taylor series approximation, or linearization, using the method of moments to estimate the variance components. Datta & Lahiri (2000) studied the Prasad-Rao approach using maximum likelihood, or restricted maximum likelihood, estimators of the variance components. However, they did not give a rigorous proof of the results, which was later given by Das, Jiang, & Rao (2004). In the context of resampling methods, Booth & Hobert (1998) proposed a parametric bootstrap method to estimate the MSPE under generalized linear mixed models (GLMM; e.g., Jiang, 2007). Butar & Lahiri (2003) studied parametric bootstrap under linear mixed models (LMM). Jiang, Lahiri, & Wan (2002) proposed a jackknife estimator of the MSPE under LMM and GLMM. Hall & Maiti (2006a) proposed parametric bootstrap methods under very general settings.

In practice, however, specific parametric distributional assumptions often do not hold. There have been some results in MSPE estimation without parametric distributional assumptions. Under only moment conditions, Lahiri & Rao (1995) demonstrated robustness of the Prasad-Rao estimator of the MSPE under the Fay-Herriot model (e.g., Fay & Herriot, 1979). Hall & Maiti (2006b) studied model (1) and noted that, essentially, only the second and fourth moments of the random effects and errors influence the bias of the MSPE estimator; they further proposed a moment-matching, double-bootstrap procedure to estimate the MSPE.

The main purpose of this paper is to study estimation of the MSPE without specific assumptions about the distribution of the data. Firstly we derive a naive analytical estimator, which has a different term compared to that under the normality assumption. We then correct the bias of the naive estimator to $o(m^{-1})$. Secondly, following Hall & Maiti (2006b), we propose a moment-matching jackknife estimator. More specifically, we first apply the moment-matching bootstrap method to obtain a naive MSPE estimator whose bias is $O(m^{-1})$. Since estimation of the third and fourth moments are notorious for their large standard errors in the case of small samples, we use a jackknife algorithm, instead of double bootstrap, to correct the bias of the bootstrap estimator, once again to $o(m^{-1})$.

The rest of the paper is organized as follows. We consider estimation of the model parameters in Section 2. In Section 3 we apply the bias-correction methods to estimate the MSPE. Section 4 reports results of simulation studies, and Section 5 is data analysis. Proofs of theoretical results are given in Appendix.

2. ESTIMATION OF MODEL PARAMETERS

Write $Y = (Y_1', \ldots, Y_m')'$, $X = (X_1', \ldots, X_m')'$, $Z = \mathrm{diag}(1_{n_1}, \ldots, 1_{n_m})$, and $D = \mathrm{diag}(D_1, \ldots, D_m)$. Then, model (1) can be written in matrix form as

$$Y = X\beta + Zb + D^{1/2}\varepsilon,$$

where $b = (b_1, \ldots, b_m)'$, and $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_m')'$. It follows that

$$\mathrm{Cov(Y) = diag(V_1, \ldots, V_m) \equiv: V}$$

with $V_i = \sigma_b^2 1_{n_i} 1_{n_i}' + \sigma_\varepsilon^2 D_i$. Now the weighted least squares estimator of β in Equation (2) can be expressed as

$$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y. \tag{4}$$

In the remaining part of this section we show how to obtain $\hat{\sigma}_b^2$ and $\hat{\sigma}_\varepsilon^2$, and how to estimate the fourth moments of the random effects and errors; the latter will be used to approximate the MSPE. There are various non-parametric methods of estimating variance components and high moments in the literature. See, for example, Harville (1974), Stukel & Rao (1997), Wu & Zhu (2010), Wu, Stute, & Zhu (2012), and Hall & Yao (2003). Below we propose a new set of estimators of the variance components and of higher moments.

2.1. Estimation of variance components

The (ordinary) least-squares estimation of β is given by $\hat{\beta}_{\text{lse}} = (X'X)^{-1}X'Y$. It is easy to derive the following:

$$\begin{aligned} \mathrm{E}\{(Y - X\hat{\beta}_{\text{lse}})(Y - X\hat{\beta}_{\text{lse}})'\} &= P_{X^\perp}\mathrm{E}\{(Y - X\beta)(Y - X\beta)'\}P_{X^\perp} \\ &= \sigma_b^2 P_{X^\perp} ZZ' P_{X^\perp} + \sigma_\varepsilon^2 P_{X^\perp} D P_{X^\perp}, \end{aligned}$$

where $P_{X^\perp} = I_N - X(X'X)^{-1}X'$ is an orthogonal projection matrix with $N = \sum_{i=1}^m n_i$. Thus, an unbiased estimator of $\psi = (\sigma_b^2, \sigma_\varepsilon^2)'$ is obtained as follows

$$\begin{aligned} \tilde{\psi} &= (W'W)^{-1}\, W'(P_{X^\perp}Y \otimes P_{X^\perp}Y) \\ &= \begin{pmatrix} \mathrm{tr\{(Z'P_{X^\perp}Z)^2\}} & \mathrm{tr(Z'P_{X^\perp}DP_{X^\perp}Z)} \\ \mathrm{tr(Z'P_{X^\perp}DP_{X^\perp}Z)} & \mathrm{tr\{(P_{X^\perp}D)^2\}} \end{pmatrix}^{-1} \begin{pmatrix} Y'P_{X^\perp}ZZ'P_{X^\perp}Y \\ Y'P_{X^\perp}DP_{X^\perp}Y \end{pmatrix}, \end{aligned} \tag{5}$$

where $W = (\mathrm{vec(P_{X^\perp}ZZ'P_{X^\perp}), vec(P_{X^\perp}DP_{X^\perp})})$ with vec being the operator stacking columns of a matrix one underneath the other that results in a column vector, and where $\otimes$

denotes the Kronecker product. Component-wise, from Equation (5), unbiased estimators of σ_ε^2 and σ_b^2 are given, respectively, by

$$\tilde{\sigma}_\varepsilon^2 = \frac{Y' P_{X^\perp} \left[\mathrm{tr}\left\{(\mathrm{Z'P_{X^\perp}Z})^2\right\}\mathrm{D} - \mathrm{tr}(\mathrm{Z'P_{X^\perp}DP_{X^\perp}Z})\mathrm{ZZ'}\right] P_{X^\perp} Y}{\mathrm{tr}\{(\mathrm{Z'P_{X^\perp}Z})^2\}\mathrm{tr}\{(\mathrm{P_{X^\perp}D})^2\} - \mathrm{tr}^2(\mathrm{Z'P_{X^\perp}DP_{X^\perp}Z})},$$

$$\tilde{\sigma}_b^2 = \frac{Y' P_{X^\perp} \left[\mathrm{tr}\{(\mathrm{P_{X^\perp}D})^2\}\mathrm{ZZ'} - \mathrm{tr}(\mathrm{Z'P_{X^\perp}DP_{X^\perp}Z})\mathrm{D}\right] P_{X^\perp} Y}{\mathrm{tr}\{(\mathrm{Z'P_{X^\perp}Z})^2\}\mathrm{tr}\{(\mathrm{P_{X^\perp}D})^2\} - \mathrm{tr}^2(\mathrm{Z'P_{X^\perp}DP_{X^\perp}Z})}.$$

Then a simple positive of ψ is given by $\hat{\psi} = (\hat{\sigma}_b^2, \hat{\sigma}_\varepsilon^2)$ with $\hat{\sigma}_b^2 = \max\{\tilde{\sigma}_b^2, 0\}$ and $\hat{\sigma}_\varepsilon^2 = \max\{\tilde{\sigma}_\varepsilon^2, 0\}$.

2.2. Estimation of the fourth moments μ_{b4} and $\mu_{\varepsilon 4}$

Denote the fourth moments of the random effects b_i and errors ε_{ij} by μ_{b4} and $\mu_{\varepsilon 4}$, respectively. Following Hall & Maiti (2006b), we first obtain a consistent estimator of $\mu_{\varepsilon 4}$. Let $e_{ij} = Y_{ij} - X_{ij}'\beta$ for all i. Then, we have

$$\mathrm{E}(\mathrm{e_{ij}} - \mathrm{e_{ik}})^4 = \mathrm{E}(\mathrm{d_{ij}}\varepsilon_{\mathrm{ij}} - \mathrm{d_{ik}}\varepsilon_{\mathrm{ik}})^4 = \mu_{\varepsilon 4}(\mathrm{d_{ij}^4} + \mathrm{d_{ik}^4}) + 6(\sigma_\varepsilon^2)^2 \mathrm{d_{ij}^2 d_{ik}^2},$$

It follows that $\mu_{\varepsilon 4}$ can be estimated consistently by

$$\hat{\mu}_{\varepsilon 4} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} (\hat{e}_{ij} - \hat{e}_{ik})^4 - 6(\hat{\sigma}_\varepsilon^2)^2 \sum_{i=1}^m \sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} d_{ij}^2 d_{ik}^2}{\sum_{i=1}^m (n_i - 1) \sum_{j=1}^{n_i} d_{ij}^4},$$

where $\hat{e}_{ij}$ is the empirical approximation of e_{ij} by replacing β with $\hat{\beta}_{\mathrm{lse}}$.

Now consider estimation of the fourth moment of the random effects b_i. We first apply a transformation to model (1) as follows:

$$d_{ij}^{-1} Y_{ij} = d_{ij}^{-1} X_i \beta + d_{ij}^{-1} b_i + \varepsilon_{ij}, \qquad i = 1, \dots, m, \quad j = 1, \dots, n_i.$$

Write $\check{e}_{ij} = d_{ij}^{-1} b_i + \varepsilon_{ij}$. After some tedious calculations, we obtain

$$\sum_{j=1}^{n_i} \mathrm{E}(\check{\mathrm{e}}_{\mathrm{ij}}^4) = \mu_{b4} \sum_{j=1}^{n_i} d_{ij}^{-4} + 6\sigma_b^2 \sigma_\varepsilon^2 \sum_{j=1}^{n_i} d_{ij}^{-2} + n_i \mu_{\varepsilon 4}$$

and

$$\mathrm{E}\left(\sum_{\mathrm{j=1}}^{\mathrm{n_i}} \check{\mathrm{e}}_{\mathrm{ij}}^3 \sum_{\mathrm{j=1}}^{\mathrm{n_i}} \check{\mathrm{e}}_{\mathrm{ij}}\right) = \mu_{b4} \sum_{j=1}^{n_i} d_{ij}^{-3} \sum_{j=1}^{n_i} d_{ij}^{-1} + 6\sigma_b^2 \sigma_\varepsilon^2 \left(\sum_{j=1}^{n_i} d_{ij}^{-2} + \sum_{j=1}^{n_i} \sum_{k=j+1}^{n_i} d_{ij}^{-1} d_{ik}^{-1}\right)$$
$$+ n_i \mu_{\varepsilon 4}.$$

The above two moment equations lead to an estimate of μ_{b4} as follows:

$$\hat{\mu}_{b4} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \hat{\check{e}}_{ij}^3 \sum_{j=1}^{n_i} \hat{\check{e}}_{ij} - \sum_{i=1}^m \sum_{j=1}^{n_i} \hat{\check{e}}_{ij}^4}{\sum_{i=1}^m \sum_{j \neq k}^{n_i} d_{ij}^{-3} d_{ik}^{-1}}$$
$$- \frac{3\hat{\sigma}_\varepsilon^2 \hat{\sigma}_b^2 \sum_{i=1}^m \sum_{j \neq k} d_{ij}^{-1} d_{ik}^{-1}}{\sum_{i=1}^m \sum_{j \neq k}^{n_i} d_{ij}^{-3} d_{ik}^{-1}},$$

where $\hat{\check{e}}_{ij}$ is the empirical approximation of $\check{e}_{ij}$ by replacing β with $\hat{\beta}_{\rm lse}$.

2.3. Asymptotic properties

In this section, we state asymptotic properties of the estimators suggested above as m tends to infinity. We need the following conditions.

(C1) There exist positive integers such that $2 < n_l \leq n_i \leq n_u$ for all i.
(C2) There exist positive constants g_1 and g_2 such that

$$g_1 \leq \lambda_{\min}(\frac{1}{N}X'X) \qquad \text{and} \qquad \| \mathrm{X_{ij}} \| \leq \mathrm{g_2} \ \text{ for all i and j},$$

where $\lambda_{\min}(A)$ means the least eigenvalue of some matrix A, $\|\cdot\|$ denotes the Euclidean norm, and where X_{ij} is the transpose of the jth row of X_i.
(C3) There exist positive constants d_u and d_l such that $d_l \leq d_{ij} \leq d_u$ for all i and j.
(C4) Assume $\mathrm{E}|\mathrm{b}_1|^{8+\delta} < \infty$ and $\mathrm{E}|\varepsilon_{11}|^{8+\delta} < \infty$, for any $0 < \delta < 1$.

Condition $(C1)$ is reasonable in small area estimation. Condition $(C2)$ is a standard assumption in linear models. Condition $(C3)$ means that the sampling variances are bounded uniformly in each cluster. Condition $(C4)$ is satisfied by many continuous distributions, including uniform, normal, gamma or lognormal with zero mean, double exponential. Please refer to Lahiri & Rao (1995).

In Appendix A, we show that

$$\lim_{m\to\infty} \frac{1}{m}(W'W) = \lim_{m\to\infty} \frac{1}{m} \begin{pmatrix} \mathrm{tr}\{(\mathrm{Z'Z})^2\} & \mathrm{tr}(\mathrm{Z'DZ}) \\ \mathrm{tr}(\mathrm{Z'DZ}) & \mathrm{tr}(\mathrm{D}^2) \end{pmatrix} \equiv: \Sigma_W. \tag{6}$$

The asymptotic properties of estimation of $\hat{\phi}$ are stated in the following theorems.

Theorem 1. *Suppose that conditions $(C1)$-$(C3)$ are satisfied. If μ_{b4} and $\mu_{\varepsilon 4}$ are finite, we have, as $m \to \infty$,*

$$\sqrt{m}(\hat{\psi} - \psi) \xrightarrow{d} N(0, \Sigma_W^{-1}\Sigma\Sigma_W^{-1}),$$

where $\Sigma = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m (\Sigma_{i1} + \Sigma_{i2})$ with

$$\Sigma_{i1} = \left\{\mu_{b4} - 3(\sigma_b^2)^2\right\} \begin{pmatrix} n_i^4 & n_i^2\mathrm{tr}(\mathrm{D_i}) \\ n_i^2\mathrm{tr}(\mathrm{D_i}) & \mathrm{tr}^2(\mathrm{D_i}) \end{pmatrix} + \left\{\mu_{\varepsilon 4} - 3(\sigma_\varepsilon^2)^2\right\} \begin{pmatrix} \mathrm{tr}(\mathrm{D_i^2}) & \mathrm{tr}(\mathrm{D_i^3}) \\ \mathrm{tr}(\mathrm{D_i^3}) & \mathrm{tr}(\mathrm{D_i^4}) \end{pmatrix}, \tag{7}$$

$$\Sigma_{i2} = 2 \begin{pmatrix} \{n_i^2\sigma_b^2 + \mathrm{tr}(\mathrm{D_i})\sigma_\varepsilon^2\}^2 & \mathrm{tr}\{(\mathrm{n_i}\sigma_\mathrm{b}^2\mathrm{D_i^{1/2}} + \mathrm{D_i^{3/2}}\sigma_\varepsilon^2)^2\} \\ \mathrm{tr}\{(\mathrm{n_i}\sigma_\mathrm{b}^2\mathrm{D_i^{1/2}} + \mathrm{D_i^{3/2}}\sigma_\varepsilon^2)^2\} & \mathrm{tr}\{(\sigma_\mathrm{b}^2\mathrm{D_i} + \mathrm{D_i^2}\sigma_\varepsilon^2)^2 + (\sigma_\mathrm{b}^2)^2\{\mathrm{tr}^2(\mathrm{D_i}) - \mathrm{tr}(\mathrm{D_i^2})\}\} \end{pmatrix} \tag{8}$$

Corollary 1. *Assume that the conditions of Theorem 1 hold. Under Condition (C_4), for any s satisfying $0 < s \leq 2 + \delta'$ with $0 < \delta' < \frac{1}{4}\delta$ and $0 < \delta < 1$,*

$$\begin{aligned} &\mathrm{E}|\tilde{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}}), && P(\tilde{\sigma}_b^2 < 0) = O(m^{-s}), \\ &\mathrm{E}|\hat{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}}), && \mathrm{E}|\hat{\sigma}_\mathrm{b}^2 - \tilde{\sigma}_\mathrm{b}^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}/2}), \\ &\mathrm{E}|\tilde{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}}), && P(\tilde{\sigma}_\varepsilon^2 < 0) = O(m^{-s}), \\ &\mathrm{E}|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}}), && \mathrm{E}|\hat{\sigma}_\varepsilon^2 - \tilde{\sigma}_\varepsilon^2|^{2\mathrm{s}} = \mathrm{O}(\mathrm{m}^{-\mathrm{s}/2}). \end{aligned}$$

We assume existence of the following limits:

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}(n_i-1)\sum_{j=1}^{n_i}d_{ij}^4=c_1, \tag{9}$$

$$\lim_{m\to\infty}\frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=2}^{n_i}d_{ij}^2d_{ik}^2}{\sum_{i=1}^{m}(n_i-1)\sum_{j=1}^{n_i}d_{ij}^4}=c_2, \tag{10}$$

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\sum_{j\neq k}^{n_i}d_{ij}^{-3}d_{ik}^{-1}=c_3, \tag{11}$$

$$\text{and}\quad \lim_{m\to\infty}\frac{\sum_{i=1}^{m}\sum_{j\neq k}d_{ij}^{-1}d_{ik}^{-1}}{\sum_{i=1}^{m}\sum_{j\neq k}^{n_i}d_{ij}^{-3}d_{ik}^{-1}}=c_4. \tag{12}$$

Theorem 2. *Suppose that conditions $(C1)$-$(C4)$ are satisfied. Then, as $m\to\infty$,*

$$\sqrt{m}(\hat{\mu}_{\varepsilon 4}-\mu_{\varepsilon 4})\xrightarrow{d}N(0,\upsilon_\varepsilon), \tag{13}$$

$$\sqrt{m}(\hat{\mu}_{b4}-\mu_{b4})\xrightarrow{d}N(0,\upsilon_b), \tag{14}$$

where $\upsilon_\varepsilon=\lim_{m\to\infty}m^{-1}\sum_{i=1}^{m}\mathrm{E}\lambda_i^2$, $\upsilon_b=\lim_{m\to\infty}m^{-1}\sum_{i=1}^{m}\mathrm{E}\chi_i^2$, and λ_i, χ_i are respectively defined in Equations $(A.4)$ and $(A.5)$.

Remark 1. *If the random effects and sampling errors are all normally distributed, $\mu_{\varepsilon 4}=3(\sigma_\varepsilon^2)^2$ and $\mu_{b4}=3(\sigma_b^2)^2$ hold. It follows that Σ_{i1} in Equation (7) vanishes.*

3. ESTIMATION OF MSPE

With the variance components estimated consistently, one can easily obtain the EBLUP $\hat{\theta}_i$ in Equation (3). In Subsection 3.1 we consider estimation of the MSPE of the EBLUP.

3.1. Analytical estimation of MSPE

For the ith small area, the MSPE is defined as $\mathrm{MSPE}_i=\mathrm{E}(\hat{\theta}_i-\theta_i)^2$, where the expectation is taken under model (1). Under the normality assumption, one can show that

$$\mathrm{MSPE}_i=\mathrm{E}(\hat{\theta}_i-\tilde{\theta}_i)^2+\mathrm{E}(\tilde{\theta}_i-\theta_i)^2,$$

where the cross-term $\mathrm{E}\{(\hat{\theta}_i-\theta_i)(\tilde{\theta}_i-\theta_i)\}=0$. However, under non-normal distributions, the previous equation does not hold. In order to study analytical estimation of MSPE_i, we use the Taylor expansion

$$\begin{aligned}\hat{\theta}_i &= \tilde{\theta}_i+\frac{\partial\tilde{\theta}_i}{\partial\psi'}(\hat{\psi}-\psi)+\frac{1}{2}(\hat{\psi}-\psi)'\frac{\partial\tilde{\theta}_i(\tilde{\psi})}{\partial\psi\partial\psi'}(\hat{\psi}-\psi)\\ &\equiv: \tilde{\theta}_i+R_{i1}+R_{i2},\end{aligned}$$

where $\tilde{\psi}$ lies between ψ and $\hat{\psi}$. Then, we have

$$\mathrm{MSPE_i}=\mathrm{E}(\tilde{\theta}_i-\theta_i+R_{i1})^2+2\mathrm{E}\{(\tilde{\theta}_i-\theta_i+R_{i1})R_{i2}\}+\mathrm{E}(R_{i2}^2) \tag{15}$$

with $\mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}} + \mathrm{R}_{\mathrm{i1}})^2 = \mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}})^2 + 2\mathrm{E}\{\mathrm{R}_{\mathrm{i1}}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}})\} + \mathrm{E}(\mathrm{R}_{\mathrm{i1}}^2)$. Simply, one can derive the expression

$$\mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}})^2 = \mathrm{f}_{\mathrm{i1}}(\psi) + \mathrm{f}_{\mathrm{i2}}(\psi), \tag{16}$$

where $f_{i1}(\psi) = \rho_i \sigma_\varepsilon^2$ with $\rho_i = \sigma_b^2 (T_i \sigma_b^2 + \sigma_\varepsilon^2)^{-1}$ and $T_i = 1_{n_i}' D_i^{-1} 1_{n_i}$, and

$$f_{i2} = (\bar{X}_{i\cdot} - \rho_i X_i' D_i^{-1} 1_{n_i})' (X' V^{-1} X)^{-1} (\bar{X}_{i\cdot} - \rho_i X_i' D_i^{-1} 1_{n_i}).$$

Note that Equation (16) holds as long as the random effects and errors have zero mean and bounded variances.

By Lemma 1 in Appendix B, we have

$$\mathrm{E}\{(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}})\mathrm{R}_{\mathrm{i1}}\} = f_{i3}(\psi^*) + O(m^{-2}), \tag{17}$$

where $\psi^* = (\sigma_b^2, \sigma_\varepsilon^2, \mu_{b4}, \mu_{\varepsilon 4})'$ is a vector, and

$$\begin{aligned} f_{i3}(\psi^*) = \frac{1}{(T_i \sigma_b^2 + \sigma_\varepsilon^2)^3} \mathrm{tr} \Bigg((\mathrm{W}'\mathrm{W})^{-1} \Bigg[\begin{pmatrix} n_i \\ \mathrm{tr}(\mathrm{D_i}) \end{pmatrix} \sigma_{\mathrm{b}}^2 \left\{ \mu_{\varepsilon 4} - 3(\sigma_\varepsilon^2)^2 \right\} \\ - \begin{pmatrix} n_i^2 T_i \\ T_i \mathrm{tr}(\mathrm{D_i}) \end{pmatrix} \sigma_\varepsilon^2 \left\{ \mu_{b4} - 3(\sigma_b^2)^2 \right\} \Bigg] \begin{pmatrix} \sigma_\varepsilon^2 \\ -\sigma_b^2 \end{pmatrix}' \Bigg). \end{aligned} \tag{18}$$

Under the normality assumption, it is easy to see that $f_{i3}(\psi^*) = 0$; otherwise, this term is of the order $O(m^{-1})$, hence can not be neglected.

By Lemma 2 in Appendix B, we have

$$\mathrm{E}(\mathrm{R}_{\mathrm{i1}}^2) = f_{i41}(\psi^*) + f_{i42}(\psi) + O(m^{-2}), \tag{19}$$

where

$$f_{i41}(\psi^*) = \frac{T_i}{(T_i \sigma_b^2 + \sigma_\varepsilon^2)^3} \mathrm{tr} \left\{ (\mathrm{W}'\mathrm{W})^{-1} \sum_{\mathrm{i}=1}^{\mathrm{m}} \Sigma_{\mathrm{i1}} (\mathrm{W}'\mathrm{W})^{-1} \begin{pmatrix} (\sigma_\varepsilon^2)^2 & -\sigma_b^2 \sigma_\varepsilon^2 \\ -\sigma_b^2 \sigma_\varepsilon^2 & (\sigma_b^2)^2 \end{pmatrix} \right\},$$

and f_{i42} is defined as f_{i41} with Σ_{i1} substituted by Σ_{i2}. By the definitions of Σ_{i1} and Σ_{i2} in Equations (7) and (8), respectively, $f_{i41}(\psi^*)$ vanishes under normality, but $f_{i42}(\psi)$ is always of order $O(m^{-1})$.

Based on Equations (15)–(19), we have

$$\mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}} + \mathrm{R}_{\mathrm{i1}})^2 = \mathrm{f}_{\mathrm{i1}}(\psi) + \mathrm{f}_{\mathrm{i2}}(\psi) + 2\mathrm{f}_{\mathrm{i3}}(\psi^*) + \mathrm{f}_{\mathrm{i41}}(\psi^*) + \mathrm{f}_{\mathrm{i42}}(\psi) + \mathrm{O}(\mathrm{m}^{-2}). \tag{20}$$

Especially, under normality, we have

$$\mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}} + \mathrm{R}_{\mathrm{i1}})^2 = f_{i1}(\psi) + f_{i2}(\psi) + f_{i42}(\psi) + O(m^{-2}).$$

But after all, the following theorem states that $\mathrm{MSPE_i} - \mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}} + \mathrm{R}_{\mathrm{i1}})^2 = \mathrm{o}(\mathrm{m}^{-1})$.

Theorem 3. *Assume that the kth moments of the random effects and errors are finite. Under conditions* $(C1)$-$(C3)$*, we have*

$$\mathrm{MSPE_i} = \mathrm{E}(\tilde{\theta}_{\mathrm{i}} - \theta_{\mathrm{i}} + \mathrm{R}_{\mathrm{i1}})^2 + \mathrm{O}(\mathrm{m}^{-1-\varpi}),$$

where $\varpi = (2\delta) \wedge (k(1/2 - \delta) - 1)$ *with* $0 < \delta < 1/2$ *and* $k(1/2 - \delta) > 1$.

Remark 2. *It is easy to see that* $\varpi \in (0, 1)$. *Specifically, we have* $\varpi = 1/2$ *with* $k = 6$ *and* $\delta = 1/4$.

3.2. Bias-corrected analytical MSPE estimator

By Theorem 3 and Equation (20), a naive estimator of MSPE_i is given by

$$\widetilde{\mathrm{MSPE}}_i = f_{i1}(\hat{\psi}) + f_{i2}(\hat{\psi}) + 2f_{i3}(\hat{\psi}^*) + f_{i41}(\hat{\psi}^*) + f_{i42}(\hat{\psi}).$$

However, it is well known that $\widetilde{\mathrm{MSPE}}_i$ is only first-order unbiased, because $f_{i1}(\hat{\psi})$ is not bias-corrected to $o(m^{-1})$. We now explore how to obtain a bias-corrected estimator of MSPE_i.

From Lemma 4 in Appendix C, we see that $f_{i2}(\psi)$, $f_{i3}(\psi^*)$, $f_{i41}(\psi^*)$, and $f_{i42}(\psi)$ can be, respectively, estimated by $f_{i2}(\hat{\psi})$, $f_{i3}(\hat{\psi}^*)$, $f_{i41}(\hat{\psi}^*)$, and $f_{i42}(\hat{\psi})$ with corresponding biases of $o(m^{-1})$. However, $f_{i1}(\hat{\psi})$ as an estimator of $f_{i1}(\psi)$ has bias of $O(m^{-1})$. In fact, by Equation $(A.11)$, we have $\mathrm{Ef_{i1}}(\hat{\psi}) = \mathrm{f_{i1}}(\psi) - \mathrm{f_{i41}}(\psi^*) - \mathrm{f_{i42}}(\psi) + \mathrm{O}(\mathrm{m}^{-3/2})$. It follows that $f_{i1}(\hat{\psi}) + f_{i41}(\hat{\psi}^*) + f_{i42}(\hat{\psi})$ is an estimator $f_{i1}(\psi)$ whose bias is $o(m^{-1})$. Hence a bias-correct estimator of MSPE_i is given by

$$\widehat{\mathrm{MSPE}}_{1i} = f_{i1}(\hat{\psi}) + f_{i2}(\hat{\psi}) + 2f_{i3}(\hat{\psi}^*) + 2f_{i41}(\hat{\psi}^*) + 2f_{i42}(\hat{\psi}). \tag{21}$$

Moreover, by Lemma 4 in Appendix C again, we have

$$\mathrm{E}(\widehat{\mathrm{MSPE}}_{1\mathrm{i}}) = \mathrm{MSPE_i} + \mathrm{O}(\mathrm{m}^{-3/2}).$$

3.3. Bias-corrected resampling MSPE estimation

In Hall & Maiti (2006b), $f_{i3}(\psi^*)$, $f_{i41}(\psi^*)$, and $f_{i42}(\psi)$ were not given in closed-form expressions; thus, the above bias-corrected analytical estimation (Eq. 21) could not be obtained in this article. In order to overcome this difficulty, the authors proposed a double bootstrap procedure and a non-parametric estimator of MSPE_i. The procedure is computationally expensive because it must apply a second bootstrap to correct the bias; in fact, the algorithm runs at a very slow speed. In this section, we follow Hall & Maiti (2006b) but only apply a one-step wild bootstrap to obtain a naive MSPE estimator; we then suggest two methods to correct the bias.

Let a random variable ξ obey a distribution, say $F(\cdot)$ with mean and third moment being zero. Let $\hat{\phi}_b = (\hat{\sigma}_b^2, \hat{\mu}_{b4})'$ and $\hat{\phi}_\varepsilon = (\hat{\sigma}_\varepsilon^2, \hat{\mu}_{\varepsilon 4})'$ be the corresponding estimators suggested in Section 2. For $l = 1, \ldots, k$, generate samples $b^{(l)} = (b_1^{(l)}, \ldots, b_m^{(l)})'$ independently from distribution $F(\hat{\phi}_b)$. Also,

$$\varepsilon^{(l)} = (\varepsilon_{11}^{(l)}, \ldots, \varepsilon_{1n_1}^{(l)}, \ldots, \varepsilon_{m1}^{(l)}, \ldots, \varepsilon_{mn_m}^{(l)})', \quad l = 1, \ldots, k,$$

are sampled independently from distribution $F(\hat{\phi}_\varepsilon)$. Let

$$Y^{(l)} = X\hat{\beta} + Zb^{(l)} + D\varepsilon^{(l)}, \quad l = 1, \ldots, k.$$

Based on the lth bootstrapped data $(X, Y^{(l)})$, we can compute the bootstrapped versions, $\hat{\beta}^{(l)}$, $\hat{\gamma}_b^{2(l)}$, $\hat{\sigma}_\varepsilon^{2(l)}$, and $\hat{\theta}_i^{(l)}$, respectively; and where the bootstrapped small area mean is $\theta_i^{(l)} = \bar{X}_{i\cdot}\hat{\beta} +$

$b_i^{(l)}$. Then, a naive estimator of MSPE$_i$ is given by

$$\widehat{\text{MSPE}}_i^*(\hat{\psi}) = \frac{1}{k}\sum_{l=1}^{k}(\hat{\theta}_i^{(l)} - \theta_i^{(l)})^2, \tag{22}$$

which has bias of order $O(m^{-1})$. In Appendix D, the proof of Theorem 4, defined below, shows that

$$\text{E}(\widehat{\text{MSPE}}_\text{i}^*(\hat{\psi})) = \text{MSPE}_i - f_{i41}(\psi^*) - f_{i42}(\psi) + O(m^{-3/2}).$$

Thus, we obtain the following estimator to correct the bias

$$\widehat{\text{MSPE}}_{2i} = \widehat{\text{MSPE}}_i^*(\hat{\psi}) + f_{i41}(\hat{\psi}^*) + f_{i42}(\hat{\psi}). \tag{23}$$

Alternatively, we may use the jackknife method to bias-correct $\widehat{\text{MSPE}}_i^*(\hat{\psi})$. Note that the bias terms f_{i41} and f_{i42} are introduced by f_{i1}. The latter is a known function of the second moments of the random effects and errors, ψ. Let $\hat{\psi}_{-j}$ be $\hat{\psi}$ with the jth group dropped when computing the estimators. Then, the bias of $\widehat{\text{MSPE}}_i^*(\hat{\psi})$ can also be estimated by

$$\widehat{\text{Bias}} = \frac{m-1}{m}\sum_{j=1}^{m}(f_{i1}(\hat{\psi}_{-j}) - f_{i1}(\hat{\psi})). \tag{24}$$

In view of Equations (22) and (24), a simple bias-corrected estimator of MSPE$_i$ is given by

$$\widehat{\text{MSPE}}_{3i} = \widehat{\text{MSPE}}_i^*(\hat{\psi}) - \widehat{\text{Bias}}_i. \tag{25}$$

In summary, we have the following theorem.

Theorem 4. *Suppose that conditions $(C1)$-$(C4)$ are satisfied. Then, we have*

$$\text{E}(\widehat{\text{MSPE}}_{2\text{i}}) = \text{MSPE}_\text{i} + \text{O}(\text{m}^{-3/2}), \tag{26}$$

$$\text{E}(\widehat{\text{MSPE}}_{3\text{i}}) = \text{MSPE}_\text{i} + \text{O}(\text{m}^{-3/2}). \tag{27}$$

4. SIMULATION STUDY

In this section, we study the finite sample performance of the proposed MSPE estimators, and compare the performance with a number of other measures of uncertainties of EBLUP in the norm and non-normal cases.

We consider $\beta = 1$, $n_i = 3$, $m = 30, 60, 100$, and $d_{ij} = 1$. The X_{ij} are generated from the uniform distribution over $[0.5, 1]$. The average values of relative bias (RB) and the coefficients of variation (CV), over i, are reported, given respectively by,

$$\text{RB}_i = \frac{\text{E}(\widehat{\text{MSPE}}_i) - \text{MSPE}_i}{\text{MSPE}_i}, \qquad \text{CV}_\text{i} = \frac{\text{E}^{1/2}(\widehat{\text{MSPE}}_\text{i} - \text{MSPE}_\text{i})^2}{\text{MSPE}_\text{i}}.$$

All results reported are based on 4000 simulations. The moment-matching distributions needed by the methods defined in Section 3.3 are selected to be the three-point distribution and Student's t-distribution suggested in Hall & Maiti (2006b). The resampling number, needed by the methods suggested in Section 3.3, is also 4000.

In order to compare with Hall & Maiti (2006b), the random effects b_i and errors ε_{ij} are generated from eight different centralized distributions as follows:

M_1: both normal distributions; M_2: both χ^2_5; M_3: both exponential distribution; M_4: χ^2_5 and $-\chi^2_5$; M_5: both Student's t_6; M_6: both logistic distributions; M_7: both $\sqrt{\chi^2_5}$; M_8: both χ^2_{10}.

Here the variances are standardized so that the ratio $\sigma_b^2/\sigma_\varepsilon^2 = 0.5, 1,$ or 2, and $\max\{\sigma_b^2, \sigma_\varepsilon^2\} = 1$, and $\min\{\sigma_b^2, \sigma_\varepsilon^2\} = 0.5$ or 1.

For comparison, the simulating results of Hall & Maiti (2006b), Jiang Lahiri, & Wan (2002) and Prasad & Rao (1990) are given in the rows HM, JLW, and PR, respectively. The proposed Jackknife bias-corrected estimator in Equation (25) with the three-point distribution and Student's t-distribution are denoted by Method_{11} and Method_{12}, respectively; the bias-corrected bootstrap MSPE estimator in Equation(23) with those two distributions are denoted by Method_{21} and Method_{22}, respectively. Also let Method_3 denote the bias-corrected analytical MSPE estimator given in Equation (21). Finally, let Naive_1 and Naive_2 denote the naive MSPE estimator in Equation (22) with the three-point distribution and Student's t-distribution, respectively.

Table 1-Table 4 report the average RB and CV for the case of $m = 60$ and $m = 100$, respectively. Note that the results for HM are copied from Table 1 in Hall & Miti (2006b), for the three-point distribution with equalvariance only. This is because Hall & Miti (2006b) did not state the value of n in Table 2 for the unequal variance cases, that is, $\sigma_b^2/\sigma_\varepsilon^2 = 0.5, 2$; nevertheless, they reported that both RB and CV take higher values for unequal variance components, compared to the equal variance case. From these four tables, it is seen that the naive estimator and Method_3 always take negative RB and hence underestimate the MSPE; but the latter performs better than the former. Moreover, it is seen that all of the methods except Naive_1 and Naive_2 perform more favorably than HM, under all models considered. In the normal-normal case with $m = 60$, PR is more accurate than the other ones; as m increases to 100, JLW performs best; however, the difference is very tiny, compared to our proposed methods. Under all of the other models, it is easily seen that our methods outperform HM, JLW, and PR. Among the proposed methods, it appears that Method_{21} performs the best, and Student's t-distribution seems to perform better. Moreover, the equal variance case tends to have lower CV than the unequal variance cases, and larger values of $\sigma_b^2/\sigma_\varepsilon^2$ tend to improve the RB.

Below, we want to explore the effect of the values of σ_b^2 and σ_ε^2 on the performance of our methods. For $\sigma_\varepsilon^2 = 1$ ($\sigma_b^2 = 1$), σ_b^2 (σ_ε^2) increases from 1 to 4. The resampling distribution is selected to be Student's t-distribution. Table 5 shows the average RB and CV under the norm-norm case M_1. As for the non-normal case, in order to save space, only M_2 is applied. The corresponding simulating results are reported in Table 6.

From Table 5, as for RB, PR still performs best, but the difference is very tiny with our methods. In the terms of CV, the biased-corrected analytical estimator Method_3 is comparable to PR, and they both perform a little better than others. Given $\sigma_\varepsilon^2 = 1$, RB and CV of JLW increase as σ_b^2 increases, but for the other methods, the lager values of σ_b^2 tend to improve the RB, but deteriorate the CV. On the contrary, for fixed $\sigma_b^2 = 1$, all methods performs better as σ_ε^2 decreases. Moreover, the larger value of $\sigma_b^2/\sigma_\varepsilon^2$ tends to improve the RB and CV. Under M_2, Method_3 has the smallest RB, and Method_{22} performs best in terms of CV. Recalling Table 1-Table 4, Method_{22} and Method_3 are very accurate under non-normal case. Moreover, when σ_b^2 (σ_ε^2) decreases, RB and CV becomes better. This is very different with that of the normal case reported in Table 5.

TABLE 1: Comparison of Different Methods of MSPE Estimation: $m = 60$

		$\sigma_b^2/\sigma_\varepsilon^2 = 0.5$		$\sigma_b^2/\sigma_\varepsilon^2 = 1$		$\sigma_b^2/\sigma_\varepsilon^2 = 2$	
Model	Method	RB	CV	RB	CV	RB	CV
M_1	HM			0.091	0.290		
	JLW	0.012	0.147	0.003	0.114	0.002	0.118
	PR	0.006	0.132	0.000	0.108	0.000	0.116
	Naive$_1$	-0.025	0.139	-0.015	0.110	-0.007	0.117
	Naive$_2$	-0.025	0.139	-0.015	0.110	-0.007	0.117
	Method$_{11}$	0.010	0.135	0.003	0.111	0.002	0.119
	Method$_{12}$	0.010	0.134	0.003	0.111	0.002	0.119
	Method$_{21}$	0.008	0.135	0.002	0.110	0.001	0.119
	Method$_{22}$	0.009	0.134	0.002	0.110	0.001	0.119
	Method$_3$	0.004	0.129	-0.002	0.111	0.003	0.116
M_2	HM			0.095	0.331		
	JLW	0.035	0.240	0.024	0.189	0.023	0.182
	PR	0.023	0.178	0.034	0.159	0.018	0.164
	Naive$_1$	-0.037	0.175	-0.021	0.150	-0.005	0.156
	Naive$_2$	-0.033	0.175	-0.019	0.150	-0.003	0.157
	Method$_{11}$	0.006	0.171	0.006	0.154	0.010	0.161
	Method$_{12}$	0.011	0.173	0.008	0.155	0.012	0.162
	Method$_{21}$	0.010	0.173	0.013	0.149	0.008	0.160
	Method$_{22}$	0.014	0.175	0.015	0.150	0.011	0.161
	Method$_3$	0.013	0.186	0.016	0.167	0.014	0.160
M_3	HM			0.108	0.375		
	JLW	0.060	0.383	0.059	0.311	0.042	0.271
	PR	0.035	0.234	0.039	0.209	0.028	0.212
	Naive$_1$	-0.057	0.217	-0.027	0.185	-0.016	0.194
	Naive$_2$	-0.046	0.219	-0.019	0.187	-0.009	0.196
	Method$_{11}$	0.005	0.219	0.013	0.195	0.009	0.203
	Method$_{12}$	0.016	0.225	0.022	0.200	0.015	0.206
	Method$_{21}$	-0.002	0.215	0.007	0.192	0.005	0.201
	Method$_{22}$	0.008	0.220	0.016	0.197	0.012	0.204
	Method$_3$	0.023	0.254	0.020	0.255	0.041	0.219
M_4	HM			0.075	0.317		
	JLW	0.040	0.248	0.029	0.193	0.016	0.178
	PR	0.027	0.180	0.020	0.158	0.010	0.161
	Naive$_1$	-0.033	0.174	-0.017	0.149	-0.013	0.156
	Naive$_2$	-0.029	0.174	-0.015	0.149	-0.011	0.156
	Method$_{11}$	0.015	0.174	0.010	0.154	0.003	0.160
	Method$_{12}$	0.018	0.176	0.013	0.156	0.005	0.160
	Method$_{21}$	0.011	0.172	0.008	0.153	0.001	0.159
	Method$_{22}$	0.014	0.174	0.010	0.154	0.003	0.160
	Method$_3$	0.016	0.159	0.009	0.161	0.014	0.163

5. DATA ANALYSIS

We consider an application of the methods developed in this paper to the Iowa crops data. See Battese, Harter, & Fuller (1988) for a complete description. This data was obtained from the 1978 June Enumerative Survey of the U.S. Department of Agriculture (USDA). Based on this data, some research is aimed at transforming satellite information into good estimates of crop areas at the individual pixel and segment levels. Hanuschak et al. (1979) and Hung & Fuller

TABLE 2: Comparison of methods ... $m = 60$ (Continued)

		$\sigma_b^2/\sigma_\varepsilon^2 = 0.5$		$\sigma_b^2/\sigma_\varepsilon^2 = 1$		$\sigma_b^2/\sigma_\varepsilon^2 = 2$	
Model	Method	RB	CV	RB	CV	RB	CV
M_5	HM			0.106	0.376		
	JLW	0.033	0.253	0.025	0.248	0.014	0.188
	PR	0.021	0.176	0.016	0.172	0.009	0.165
	Naive_1	-0.038	0.171	-0.021	0.151	-0.013	0.156
	Naive_2	-0.033	0.171	-0.018	0.156	-0.011	0.156
	Method_{11}	0.009	0.170	0.006	0.161	0.001	0.161
	Method_{12}	0.013	0.173	0.009	0.168	0.004	0.163
	Method_{21}	0.005	0.168	0.004	0.158	0.000	0.160
	Method_{22}	0.009	0.170	0.007	0.164	0.002	0.162
	Method_3	0.008	0.186	0.010	0.162	0.014	0.166
M_6	HM			0.075	0.317		
	JLW	0.020	0.184	0.015	0.134	0.007	0.144
	PR	0.012	0.154	0.010	0.133	0.004	0.138
	Naive_1	-0.035	0.159	-0.017	0.130	-0.012	0.136
	Naive_2	-0.033	0.159	-0.016	0.131	-0.011	0.136
	Method_{11}	0.007	0.153	0.006	0.133	0.001	0.138
	Method_{12}	0.008	0.154	0.007	0.134	0.001	0.139
	Method_{21}	0.004	0.153	0.005	0.133	0.000	0.138
	Method_{22}	0.006	0.153	0.006	0.133	0.001	0.138
	Method_3	0.009	0.159	0.010	0.138	0.008	0.139
M_7	HM			0.089	0.289		
	JLW	0.021	0.148	0.011	0.118	0.004	0.120
	PR	0.006	0.133	0.003	0.111	0.000	0.117
	Naive_1	-0.033	0.141	-0.014	0.113	-0.008	0.116
	Naive_2	-0.032	0.141	-0.014	0.113	-0.008	0.116
	Method_{11}	0.009	0.135	0.005	0.113	0.001	0.119
	Method_{12}	0.010	0.134	0.006	0.113	0.001	0.119
	Method_{21}	0.006	0.134	0.004	0.113	0.000	0.119
	Method_{22}	0.007	0.134	0.004	0.113	0.000	0.119
	Method_3	0.003	0.132	0.000	0.111	-0.003	0.118
M_8	HM			0.092	0.312		
	JLW	0.030	0.207	0.015	0.150	0.004	0.120
	PR	0.011	0.156	0.005	0.134	0.000	0.117
	Naive_1	-0.029	0.161	-0.018	0.133	-0.009	0.137
	Naive_2	-0.027	0.161	-0.016	0.133	-0.008	0.137
	Method_{11}	0.006	0.154	0.001	0.133	0.001	0.119
	Method_{12}	0.008	0.155	0.003	0.134	0.001	0.119
	Method_{21}	0.003	0.153	-0.001	0.133	0.000	0.119
	Method_{22}	0.005	0.154	0.000	0.133	0.000	0.119
	Method_3	0.006	0.160	0.002	0.136	-0.003	0.118

(1987) concentrated on producing good estimation of total crop areas for both large and small geographical units. Battese, Harter, & Fuller (1988) considered the prediction of areas under corn and soybeans for 12 counties in north-central Iowa.

The model is

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + b_i + \varepsilon_{ij}, \qquad i = 1, \ldots, 12,\ j = 1, \ldots, n_i, \tag{28}$$

TABLE 3: Comparison of Different Methods of MSPE Estimation: $m = 100$.

		$\sigma_b^2/\sigma_\varepsilon^2 = 0.5$		$\sigma_b^2/\sigma_\varepsilon^2 = 1$		$\sigma_b^2/\sigma_\varepsilon^2 = 2$	
Model	Method	RB	CV	RB	CV	RB	CV
M_1	HM			0.098	0.280		
	JLW	0.000	0.109	0.001	0.085	-0.000	0.092
	PR	-0.002	0.104	0.002	0.088	-0.001	0.092
	Naive_1	-0.021	0.109	-0.005	0.089	-0.005	0.094
	Naive_2	-0.023	0.114	-0.004	0.090	-0.005	0.094
	Method_{11}	0.000	0.106	0.002	0.088	-0.000	0.095
	Method_{12}	-0.002	0.111	0.002	0.088	-0.000	0.095
	Method_{21}	-0.001	0.106	0.002	0.088	-0.000	0.095
	Method_{22}	-0.003	0.111	0.002	0.088	-0.000	0.095
	Method_3	0.004	0.104	-0.003	0.085	0.000	0.089
M_2	HM			0.067	0.298		
	JLW	0.008	0.131	0.021	0.154	0.009	0.128
	PR	0.005	0.121	0.017	0.135	0.007	0.123
	Naive_1	-0.020	0.133	-0.017	0.118	-0.007	0.122
	Naive_2	-0.016	0.136	-0.014	0.121	-0.006	0.123
	Method_{11}	-0.003	0.121	0.008	0.133	0.002	0.124
	Method_{12}	-0.001	0.122	0.011	0.138	0.003	0.125
	Method_{21}	-0.004	0.120	0.006	0.132	0.001	0.124
	Method_{22}	-0.002	0.122	0.010	0.138	0.002	0.124
	Method_3	0.007	0.144	0.009	0.125	0.012	0.126
M_3	HM			0.079	0.327		
	JLW	0.039	0.238	0.034	0.203	0.026	0.185
	PR	0.029	0.177	0.026	0.162	0.020	0.162
	Naive_1	-0.032	0.164	-0.019	0.150	-0.009	0.154
	Naive_2	-0.026	0.165	-0.014	0.151	-0.005	0.155
	Method_{11}	0.005	0.166	0.006	0.155	0.006	0.159
	Method_{12}	0.012	0.170	0.011	0.157	0.009	0.160
	Method_{21}	0.002	0.165	0.003	0.154	0.004	0.158
	Method_{22}	0.009	0.168	0.008	0.156	0.008	0.159
	Method_3	0.017	0.185	0.015	0.190	0.037	0.171
M_4	HM			0.064	0.312		
	JLW	0.017	0.162	0.010	0.129	0.008	0.128
	PR	0.012	0.140	0.007	0.119	0.006	0.123
	Naive_1	-0.027	0.138	-0.019	0.150	-0.009	0.121
	Naive_2	-0.025	0.138	-0.014	0.151	-0.008	0.122
	Method_{11}	0.002	0.137	-0.001	0.119	0.000	0.124
	Method_{12}	0.004	0.138	0.000	0.119	0.001	0.124
	Method_{21}	-0.000	0.137	-0.002	0.118	-0.001	0.123
	Method_{22}	0.002	0.137	-0.001	0.119	0.000	0.124
	Method_3	0.012	0.125	0.018	0.123	0.016	0.129

where y_{ij} is the number of hectares of corn (or soybeans) in the jth segment of the ith county, x_{ij1} and x_{ij2} are the number of pixels classified as corn and soybeans, respectively, in the jth segment of the ith county, n_i is the number of sample segments in the ith county, and ranges from 1 to 5.

TABLE 4: Comparison of methods ... m=100 (Continued)

		$\sigma_b^2/\sigma_\varepsilon^2 = 0.5$		$\sigma_b^2/\sigma_\varepsilon^2 = 1$		$\sigma_b^2/\sigma_\varepsilon^2 = 2$	
Model	Method	RB	CV	RB	CV	RB	CV
M_5	HM			0.036	0.280		
	JLW	0.015	0.174	0.016	0.153	0.009	0.143
	PR	0.010	0.137	0.012	0.126	0.007	0.131
	Naive$_1$	-0.028	0.134	-0.012	0.120	-0.008	0.125
	Naive$_2$	-0.025	0.140	-0.010	0.122	-0.006	0.127
	Method$_{11}$	0.000	0.133	0.004	0.123	0.002	0.129
	Method$_{12}$	0.003	0.141	0.007	0.127	0.003	0.131
	Method$_{21}$	-0.001	0.132	0.003	0.123	0.001	0.129
	Method$_{22}$	0.002	0.140	0.006	0.126	0.002	0.131
	Method$_3$	0.003	0.141	0.011	0.127	0.009	0.130
M_6	HM			0.097	0.288		
	JLW	0.007	0.130	0.008	0.110	0.004	0.109
	PR	0.004	0.120	0.006	0.106	0.003	0.107
	Naive$_1$	-0.024	0.123	-0.006	0.103	-0.007	0.107
	Naive$_2$	-0.023	0.123	-0.006	0.103	-0.007	0.107
	Method$_{11}$	0.000	0.120	0.002	0.107	0.000	0.109
	Method$_{12}$	0.001	0.121	0.003	0.107	0.001	0.109
	Method$_{21}$	-0.001	0.120	0.002	0.106	-0.000	0.109
	Method$_{22}$	-0.000	0.120	0.002	0.107	0.000	0.109
	Method$_3$	0.006	0.120	0.006	0.104	0.008	0.110
M_7	HM			0.076	0.312		
	JLW	0.011	0.110	0.010	0.090	0.005	0.092
	PR	0.003	0.105	0.006	0.087	0.003	0.092
	Naive$_1$	-0.015	0.107	-0.010	0.089	-0.029	0.094
	Naive$_2$	-0.015	0.107	-0.009	0.089	-0.027	0.094
	Method$_{11}$	0.004	0.106	0.006	0.090	0.003	0.094
	Method$_{12}$	0.004	0.106	0.007	0.090	0.003	0.094
	Method$_{21}$	0.003	0.106	0.006	0.090	0.003	0.094
	Method$_{22}$	0.003	0.106	0.006	0.090	0.003	0.094
	Method$_3$	0.002	0.104	0.004	0.088	0.002	0.092
M_8	HM			0.051	0.279		
	JLW	0.017	0.132	0.010	0.112	0.011	0.111
	PR	0.008	0.121	0.005	0.105	0.008	0.108
	Naive$_1$	-0.019	0.161	-0.011	0.104	-0.004	0.109
	Naive$_2$	-0.019	0.161	-0.011	0.104	-0.004	0.109
	Method$_{11}$	0.003	0.120	0.001	0.106	0.006	0.110
	Method$_{12}$	0.004	0.121	0.002	0.106	0.006	0.110
	Method$_{21}$	0.002	0.120	0.000	0.106	0.005	0.110
	Method$_{22}$	0.003	0.121	0.001	0.106	-0.004	0.110
	Method$_3$	0.005	0.123	0.005	0.107	-0.004	0.110

5.1. Iowa crops data with homogenous sampling errors

In Battese, Harter, & Fuller (1988), the random effects b_i and the errors ε_{ij} are all normally distributed. In addition, the errors are homogeneous. By applying data of corn and soybeans in the $\sum_{i=1}^{m} n_i = 36$ segments of these 12 counties, Battese, Harter, & Fuller (1988) estimated the model parameters and predicted the random effects. In order to obtain the prediction of the mean crop hectares per county defined in Equation (3), this article used the population mean numbers

TABLE 5: RB and CV for the normal-normal case M_1: $m = 30$. (All values are multiplied by 100.)

	Method	$\sigma_b^2 = 4$ RB	CV	$\sigma_b^2 = 3$ RB	CV	$\sigma_b^2 = 2$ RB	CV	$\sigma_b^2 = 1$ RB	CV
$\sigma_\varepsilon^2 = 1$	PR	-0.5	17.2	0.5	16.9	0.5	16.1	0.9	16.0
	JLW	0.0	17.5	1.3	17.6	1.7	17.6	3.1	19.4
	Method$_{11}$	-0.3	17.4	0.8	17.1	1.0	16.3	1.8	16.3
	Method$_{21}$	-0.5	17.4	0.6	17.0	0.6	16.2	1.2	16.0
	Mehtod$_3$	-0.5	17.2	0.6	16.9	0.6	16.1	1.3	15.9
		$\sigma_\varepsilon^2 = 4$		$\sigma_\varepsilon^2 = 3$		$\sigma_\varepsilon^2 = 2$		$\sigma_\varepsilon^2 = 1$	
	Method	RB	CV	RB	CV	RB	CV	RB	CV
$\sigma_b^2 = 1$	PR	4.5	29.4	1.1	23.9	0.1	18.8	0.1	15.9
	JLW	2.5	59.0	2.7	44.0	3.4	28.9	2.3	20.0
	Method$_{11}$	-5.6	55.5	-3.2	39.9	-0.2	23.8	0.8	16.1
	Method$_{21}$	4.2	27.6	0.8	23.1	0.0	18.5	0.2	15.8
	Mehtod$_3$	8.6	26.0	3.5	21.4	1.2	17.6	0.4	15.6

TABLE 6: RB and CV for the χ_5^2-χ_5^2 case M_2: $m = 30$. (All values are multiplied by 100.)

	Method	$\sigma_b^2 = 4$ RB	CV	$\sigma_b^2 = 3$ RB	CV	$\sigma_b^2 = 2$ RB	CV	$\sigma_b^2 = 1$ RB	CV
$\sigma_\varepsilon^2 = 1$	PR	3.26	23.64	3.50	22.87	3.09	22.93	2.49	22.95
	JLW	7.47	40.40	7.77	44.81	7.39	42.35	6.77	43.13
	Method$_{12}$	2.29	22.66	2.48	22.24	2.07	22.10	1.53	22.03
	Method$_{22}$	0.95	21.89	1.35	20.95	0.76	20.94	0.40	21.10
	Mehtod$_3$	1.08	21.67	1.23	21.18	0.68	21.22	0.24	21.26
		$\sigma_\varepsilon^2 = 4$		$\sigma_\varepsilon^2 = 3$		$\sigma_\varepsilon^2 = 2$		$\sigma_\varepsilon^2 = 1$	
	Method	RB	CV	RB	CV	RB	CV	RB	CV
$\sigma_b^2 = 1$	PR	6.03	33.71	5.20	30.34	3.60	26.65	2.44	23.23
	JLW	5.37	70.04	7.68	65.30	8.48	59.22	6.51	40.09
	Method$_{12}$	-6.51	56.44	-2.79	46.08	0.14	32.90	1.34	22.71
	Method$_{22}$	6.71	28.06	3.59	25.55	1.13	23.33	0.25	21.45
	Mehtod$_3$	3.86	30.34	2.18	27.54	0.68	24.41	0.12	21.66

of pixels classed as corn and soybeans per segment in the ith county to replace the sample mean numbers $\sum_{j=1}^{n_i} X_{ij1}/n_i$ and $\sum_{j=1}^{n_i} X_{ij2}/n_i$ of pixels in the n_ith sample segments of the ith county.

In this article, we carry out the SMA estimation for corn and soybean of the 12 small areas and the results are presented in Table 7 and Table 8 respectively. The estimated model parameters for corn are $\hat{\beta} = (51.128, 0.329, -0.135)'$, $\hat{\sigma}_b^2 = 144.397$, $\hat{\sigma}_\varepsilon^2 = 145.233$, $\hat{\mu}_{b4} = 10191610.680$, and $\hat{\mu}_{\varepsilon 4} = 15140.706$. The estimated model parameters for soybean are $\hat{\beta} = (-16.612, 0.0301, 0.494)'$, $\hat{\sigma}_b^2 = 289.680$, $\hat{\sigma}_\varepsilon^2 = 169.623$, $\hat{\mu}_{b4} = 3856.356$, and $\hat{\mu}_{\varepsilon 4} =$

TABLE 7: EBLUP, Measures of Uncertainty of Corn with Homogenous Sampling Errors.

		Predicted hectares		Standard errors					
County	segments	BHF	EBLUP	BHF	PR	JLW	Method_{21}	Method_{22}	Method_{3}
Cerro Gordo	1	122.2	166.2	13.7	10.9	10.1	10.8	10.9	12.3
Hamilton	1	126.3	93.4	12.9	10.7	10.1	10.7	10.7	12.2
Worth	1	106.2	88.4	12.4	10.6	12.9	11.6	11.7	13.0
Humboldt	2	108	155.3	9.7	8.5	9.5	8.4	8.3	9.5
Franklin	3	145	153.9	7.1	7.1	7.6	6.9	6.9	7.7
Pocahontas	3	112.6	99.2	7.2	7.1	7.1	7.0	7.0	7.8
Winnebago	3	112.4	115.9	7.2	7.1	7.2	7.0	6.9	7.8
Wright	3	122.1	143.7	7.3	7.1	6.9	6.9	6.8	7.7
Webster	4	115.8	114.7	6.1	6.2	6.2	6.1	6.1	6.8
Hancock	5	124.3	110.0	5.7	5.6	5.7	5.5	5.4	6.0
Kossuth	5	106.3	113.3	5.5	5.6	6.9	5.5	5.5	6.0
Hardin	5	143.6	118.3	6.1	5.6	5.4	5.6	5.5	6.1

TABLE 8: EBLUP, Measures of Uncertainty of Soybean with Homogenous Sampling Errors.

		Predicted hectares		Standard errors					
County	segments	BHF	EBLUP	BHF	PR	JLW	Method_{21}	Method_{22}	Method_{3}
Cerro Gordo	1	77.8	13.2	15.6	13.4	13.3	13.7	13.4	15.6
Hamilton	1	94.8	102.9	14.8	13.3	12.3	13.8	13.6	15.7
Worth	1	86.9	107.7	14.2	13.2	12.7	13.7	13.5	15.6
Humboldt	2	79.7	41.5	11.1	9.8	10.9	10.3	10.1	11.4
Franklin	3	65.2	56.5	8.1	8.0	8.9	8.1	7.9	8.8
Pocahontas	3	113.8	118.6	8.2	8.0	7.7	8.1	7.9	8.8
Winnebago	3	98.5	85.7	8.3	8.0	8.4	8.2	8.0	8.9
Wright	3	112.8	95.7	8.4	8.0	8.8	8.1	7.8	8.8
Webster	4	109.6	113.5	7.0	6.9	6.8	7.1	6.8	7.5
Hancock	5	101	116.3	6.5	6.2	6.4	6.2	6.0	6.6
Kossuth	5	119.9	114.8	6.3	6.1	7.8	6.1	5.9	6.5
Hardin	5	74.9	102.5	6.9	6.2	6.1	6.2	6.0	6.6

68161.788. Moreover, we compute the EBLUP, the corresponding $\widehat{\text{MSPE}}$, and their square roots as the measures of uncertainty. As for the resampling distribution, the Student's t-distribution is applied.

In Table 7 and Table 8, the predicted hectares and the standard errors of the survey regression predictor, given in the rows BHF, are borrowed from Battese, Harter, & Fuller (1988). This article also reported another two measures of the uncertainty, but stated that the survey regression predictor is biased and most inadequate for the entire data. As for the predicted hectares, it is difficult to compare ours with that of Battese, Harter, & Fuller (1988) because we only apply

TABLE 9: Iowa Crops Data: Within-Area Sample Variances.

area	1	2	3	4	5	6	7	8	9	10
corn	2212.35	2374.98	32.53	1884.08	933.10	2915.89	453.61	245.26	146.85	1354.74
soybean	3120.05	1644.51	269.77	2529.06	109.26	2707.56	553.73	296.16	439.06	717.72

TABLE 10: EBLUP, Measures of Uncertainty of Corn with Heteroscedastic Sampling Errors.

Groups	1	2	3	4	5	6	7	8	9	10
EBLUP	115.2	155.2	155.7	99.3	115.0	143.3	116.0	109.8	112.1	118.1
PR	8.0	9.5	5.4	8.1	5.4	8.0	4.7	4.1	4.1	6.4
JLW	7.3	9.8	5.9	7.6	5.4	7.4	4.5	4.0	4.9	5.9
Method$_{21}$	4.3	4.9	3.6	4.3	3.6	4.4	3.4	3.3	3.3	3.8
Method$_{22}$	4.7	5.2	3.9	4.6	3.9	4.7	3.6	3.5	3.5	4.1
Method$_3$	7.6	9.2	5.1	7.6	5.1	7.6	4.1	3.4	3.4	5.1

the 36 sample segments of pixels. Note that the standard errors of all methods decrease as the number of sample segments increase, and our method Method$_{22}$ has the smallest standard errors in all but the first three counties which have only one sample each. Based on our simulation study in the previous section, there is a reason to believe that our proposed MSPE estimators are more accurate than HM, JLW, and PR when applying to such data.

5.2. Iowa crops data with heteroscedastic errors

Consider the Iowa crops data again. But here we assume that the errors are heteroscedastic. To explore the within-area variation, the first three counties, which have only 1 sampled segment each, are combined to Iowa form the first small area. Thus there are a total of $m = 10$ small areas. In Table 9, we show the within-area variances. For the corn data, Jiang & Nguyen (2012) suggested two groups: $S_1 = \{1, 2, 4, 6, 10\}$ with variances above 1000 and $S_2 = \{3, 5, 7, 8, 9\}$ with variances below 1000. Similarly the soybean data are also divided into two groups $S_1 = \{1, 2, 4, 6\}$ with variances above 1000 and $S_2 = \{3, 5, 7, 8, 9, 10\}$ with variances below 1000.

Model (28) is still applied. The variances of the errors ε_{ij} are assumed to be $\sigma_\varepsilon^2 d_t^2$ in the group S_t. Here $t = 1, 2$. The d_t^2 are assumed to be $\hat{\sigma}_\varepsilon^2$ based on the homogenous model (28) in S_t. For the corn data, d_1^2 and d_2^2 are found to be 244.17 and 105.62 respectively. For the soybean data, they are 64.09 and 113.13 respectively.

SMA estimation is carried out for the Iowa crops data of the 10 small areas and the results are presented in Table 10 and Table 11 respectively. The estimated model parameters for corn are $\hat{\beta} = (66.260.30 - 0.15)'$, $\hat{\sigma}_b^2 = 1140.63$, $\hat{\sigma}_\varepsilon^2 = 0.74$, $\hat{\mu}_{b4} = 141430.77$, and $\hat{\mu}_{\varepsilon 4} = 4.98$. The estimated model parameters for soybean are $\hat{\beta} = (2.91 - 0.030.48)'$, $\hat{\sigma}_b^2 = 278.90$, $\hat{\sigma}_\varepsilon^2 = 1.80$, $\hat{\mu}_{b4} = 3856.36$, and $\hat{\mu}_{\varepsilon 4} = 49.47$. Moreover, we compute the EBLUP, the corresponding $\widehat{\text{MSPE}}$, and their square roots as the measures of uncertainty. As for the resampling distribution, the Student's t-distribution is applied again.

Comparing Table 10 with Table 7, we can find the results: MSPEs of all methods become smaller; our three methods outperforms PR and JLW; Method$_{21}$ performs best. Comparing Table 11 with Table 8, one draws similar conclusions, but Method$_{22}$ performs best. Thus the NER with heteroscedastic errors is more accurate, and our suggested methods are efficient.

TABLE 11: EBLUP, Measures of Uncertainty of Soybean with Heteroscedastic Sampling Errors.

Groups	1	2	3	4	5	6	7	8	9	10
EBLUP	73.5	39.3	57.1	119.1	85.5	96.0	113.9	116.2	114.1	102.8
PR	6.3	7.7	8.2	6.3	8.3	6.3	7.2	6.4	6.4	6.4
JLW	6.3	8.1	8.8	6.3	8.6	7.1	7.1	6.8	8.1	6.4
Method_{21}	5.4	5.7	5.8	5.3	5.8	5.3	5.5	5.3	5.4	5.3
Method_{22}	5.2	5.4	5.5	5.1	5.5	5.1	5.2	5.1	5.2	5.1
Method_{3}	6.0	7.5	7.6	6.0	7.6	6.0	6.4	5.4	5.4	5.4

6. DISCUSSION

In the context of small-area estimation, normality is often assumed in the literature. In practice, however, this assumption may not hold. Some transformation on the response Y can be applied to overcome this difficulty. However, if the non-normality is in the unobservable random effects, not in the observed data, the transformative method is not practical either. In such cases, Hall & Maiti (2006b) suggested a double moment-matching bootstrap procedure to estimate MSPE. However, the second bootstrap to correct the bias is not very efficient. To overcome this difficulty, we have studied the problem of accuracy measures MSPE of EBLUP in the nested error regression model under moment conditions. We first explored the analytical estimation of MSPE and suggested a bias-corrected analytical estimator $\widehat{\text{MSPE}}_{1i}$ in Equation (21). Secondly we extended the double moment-matching bootstrap method suggested in Hall & Maiti(2006b). One-step moment-matching bootstrap is applied to obtain the the naive estimation $\widehat{\text{MSPE}}_i^*(\hat{\psi})$ in Equation (22), and then two bias-corrected methods $\widehat{\text{MSPE}}_{2i}$ and $\widehat{\text{MSPE}}_{3i}$ are defined in Equations (23) – (24) respectively. A simulation study and data analysis above show that our proposed three MSPE estimators perform powerfully.

Based on our simulation study and data analysis, the estimating efficiency of the fourth moments are not satisfactory, and this has a bad impact on the proposed methods. Part of our future work is to explore new non-parametric estimation of MSPE which does not depend on these high moments. On the other hand, the heteroscedastic parameters d_{ij} of the sampling errors are not easy to be given. We try to study the case of the variance of the errors as a parametric function of some covariates.

ACKNOWLEDGEMENTS

We are very grateful to the editor and two referees for their comments to improve the presentation. The research of Ping Wu is partly supported by the National Natural Science Foundation of China and the 111 project. The research of Jiming Jiang is partially supported by the NSF grant SES-1121794 and NIH grant.

BIBLIOGRAPHY

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An error-Components model for prediction of county crop areas using survey and satellite Data. *Journal of the American Statistical Association*, 83, 28–36.

Booth, J. G.& Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American Statistical Association*, 93, 262–272.

Butar, F. B. & Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small area estimator. Model selection, model diagnostics, empirical Bayes and hierarchical Bayes, Special Issue II (Lincoln, NE, 1999). *Journal of Statistical Planning and Inference*, 112, 63–76.

Das, K., Jiang, J. M., & Rao, N. K. (2004). Mean squared error of empirical predictor. *Annals of Statistics*, 32, 818–840.

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

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

Fuller, W. A. & Battese, G. E. (1973). Transformations of Estimation of Linear Models With Nested-Error Structure. *Journal of the American Statistical Association*, 68, 626–632.

Hall, P. & Maiti, T. (2006a). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society Series B Statistical Methodology*, 68, 221–23,8.

Hall, P. & Maiti, T. (2006b). Non-parametric estimation of mean-squared prediction error in nested-error regression models. *Annals of Statistics*, 34, 1733–1750.

Hall, P. & Yao, Q. (2003). Inference in components of variance models with low replication. *Annals of Statistics*, 31, 414–441.

Hanuschak, G., Sigman, R., Craig, M., Ozga, M., Luebbe, R., Cook, R., Kleweno, D., & Miller, C.(1979). Obtaining timely crop area estimates using ground-gathered and LANDSAT data. *Technical Bulletin 1609*, Washington, DC: Economics, Statistics, and Cooperatives Service, USDA.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* 61, 383-385.

Harville, D. A. & Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87, 724–731.

Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). Random-effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757–765.

Hung, H. M. & Fuller, W. A. (1987). Regression estimation of crop acreages with transformed landsat data as auxiliary variables. *Journal of Business & Economic Statistics*, 5, 475–482.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York.

Jiang, J., Lahiri, P., & Wan, S. (2002). A unified jackknife theory for empirical best prediction with m-estimation. *Annals of Statistics*, 30, 1782–1810.

Jiang, J. & Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *The Canadian Journal of Statistics*, 40, 588–603.

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

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

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistica Sinica*, 28, 40–68.

Prasad, N.G.N. & Rao, J.N.K. (1990). The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association*, 85, 163–171.

Rao, J. N. K. & Molina, I. (2015). *Small Area Estimation*, 2nd ed., Wiley, New York.

Sugasawa, S. & Kubokawa T. (2017). Heteroscedastic nested error regression models with variance functions. *Statista Sinica*, 27, 1101–1123.

Stukel, D. M. & Rao, J. N. K. (1997). Estimation of regression models with nested error structure and unequal error variances under two and three stage cluster sampling. *Statistics & Probability Letters* , 35, 401–407.

Wu, P. & Zhu, L. X. (2010). An orthogonality-based estimation of moments for linear mixed models. *Scandinavian Journal of Statistics* , 37, 253–263.

Wu, P., Stute, W., & Zhu, L. X. (2012). Efficient estimation of moments in linear mixed models. *Bernoulli*, 18, 206–228.

APPENDIX

Appendix A: Proofs of Parameter Estimation

Proof of Equation (5). By matrix properties, it is not difficult to derive that

$$Y'P_{X^\perp}DP_{X^\perp}Y = \mathrm{vec(D)'(P_{X^\perp} \otimes P_{X^\perp})(Y \otimes Y)},$$

$$Y'P_{X^\perp}ZZ'P_{X^\perp}Y = \mathrm{vec(ZZ')'(P_{X^\perp} \otimes P_{X^\perp})(Y \otimes Y)},$$

$$\mathrm{vec(ZZ')'(P_{X^\perp} \otimes P_{X^\perp})vec(ZZ') = tr\left\{(Z'P_{X^\perp}Z)^2\right\}},$$

$$\mathrm{vec(D)'(P_{X^\perp} \otimes P_{X^\perp})vec(D) = tr\left\{(P_{X^\perp}D)^2\right\}},$$

$$\mathrm{vec(ZZ')'(P_{X^\perp} \otimes P_{X^\perp})vec(D) = tr(Z'P_{X^\perp}DP_{X^\perp}Z)}.$$

The proof is finished. ■

Proof of Equation (6). It is easy to derive that

$$\mathrm{tr\{P_{X^\perp}D)^2\} = tr(D^2) - 2tr\{(X'X)^{-1}X'D^2X\} + tr\{(X'X)^{-1}X'DX)^2\}}.$$

By trace inequalities and conditions (C1)-(C3), we have

$$\mathrm{tr}\left\{\mathrm{(X'X)^{-1}X'D^2X}\right\} \le \mathrm{tr}\left\{\mathrm{(X'X)^{-1}}\right\}\sum_{\mathrm{i=1}}^{\mathrm{m}}\mathrm{tr(X_i'D_i^2X_i) = O(1)},$$

$$\mathrm{tr}\left[\left\{\mathrm{(X'X)^{-1}X'DX}\right\}^2\right] \le \mathrm{tr}\left\{\mathrm{(X'X)^{-2}}\right\}\left\{\sum_{\mathrm{i=1}}^{\mathrm{m}}\mathrm{tr(X_i'D_iX_i)}\right\}^2 = \mathrm{O(1)}.$$

Then Equation (6) can be derived. ■

Proof of Theorem 1. Put $e = (e_1', \ldots, e_m')'$ with $e_i = Y_i - X_i\beta$. By conditions $(C1)-(C2)$, we can obtain

$$P_{X^\perp}DP_{X^\perp} - D = \frac{1}{m}O(J_N).$$

Here and below J_n is an $n \times n$ matrix of ones. Then it is not difficult to derive that

$$\frac{1}{\sqrt{m}}\left[e'(P_{X^\perp}DP_{X^\perp} - D)e - \mathrm{E}\left\{\mathrm{e'(P_{X^\perp}DP_{X^\perp} - D)e}\right\}\right] \xrightarrow{p} 0.$$

Similarly we have

$$\frac{1}{\sqrt{m}}\left[e'(P_{X^\perp}ZZ'P_{X^\perp} - ZZ')e - \mathrm{E}\left\{\mathrm{e'(P_{X^\perp}ZZ'P_{X^\perp} - ZZ')e}\right\}\right] \xrightarrow{p} 0.$$

On the other hand, we can obtain

$$e'ZZ'e - \mathrm{E(e'ZZ'e)} = \sum_{\mathrm{i=1}}^{\mathrm{m}}\mathrm{tr\{1_{n_i}1_{n_i}'(e_ie_i' - V_i)\}} \equiv: \sum_{\mathrm{i=1}}^{\mathrm{m}}\zeta_{\mathrm{i}1}, \tag{A.1}$$

$$e'De - \mathrm{E(e'De)} = \sum_{i=1}^{m} \mathrm{tr}\{\mathrm{D_i(e_ie_i' - V_i)}\} \equiv: \sum_{\mathrm{i=1}}^{\mathrm{m}} \zeta_{\mathrm{i2}}. \tag{A.2}$$

Let $\zeta_i = (\zeta_{i1}, \zeta_{i2})'$. It is easy to obtain that $\{\zeta_i\}$ is an independent random vector series with zero mean. After some tedious calculations, we get

$$\begin{aligned}
\mathrm{D}(\zeta_{\mathrm{i1}}) &= n_i^4\{\mu_{b4} - 3(\sigma_b^2)^2\} + \{\mu_{\varepsilon 4} - 3(\sigma_\varepsilon^2)^2\}\mathrm{tr}(\mathrm{D_i^2}) \\
&\quad + 2(\sigma_\varepsilon^2)^2(\mathrm{tr}^2(\mathrm{D_i}) - \mathrm{tr}(\mathrm{D_i^2})) + 4\mathrm{n_i^2}\sigma_\mathrm{b}^2\sigma_\varepsilon^2\mathrm{tr}(\mathrm{D_i}) \\
\mathrm{D}(\zeta_{\mathrm{i2}}) &= \mathrm{tr}^2(\mathrm{D_i})\{\mu_{\mathrm{b4}} - 3(\sigma_\mathrm{b}^2)^2\} + \{\mu_{\varepsilon 4} - 3(\sigma_\varepsilon^2)^2\}\mathrm{tr}(\mathrm{D_i^4}) + 4\sigma_\mathrm{b}^2\sigma_\varepsilon^2\mathrm{tr}(\mathrm{D_i^3}) \\
\mathrm{Cov}(\zeta_{\mathrm{i1}}, \zeta_{\mathrm{i2}}) &= n_i^2\mathrm{tr}(\mathrm{D_i})\{\mu_{\mathrm{b4}} - 3(\sigma_\mathrm{b}^2)^2\} + \{\mu_{\varepsilon 4} - (\sigma_\varepsilon^2)^2\}\mathrm{tr}(\mathrm{D_i^3}) + 4\mathrm{n_i}\sigma_\mathrm{b}^2\sigma_\varepsilon^2\mathrm{tr}(\mathrm{D_i^2}).
\end{aligned}$$

Then it is not difficult to derive that $\mathrm{Cov}(\zeta_\mathrm{i}) = \Sigma_{\mathrm{i1}} + \Sigma_{\mathrm{i2}}$ with Σ_{i1} and Σ_{i2} defined in Equations (7) and (8) respectively.

By the multivariate central limit theorem and Slutsky's theorem, this theorem holds by subtracting $\hat{\psi}$ by $\tilde{\psi}$. Note that $\sqrt{m}(\hat{\psi} - \psi) = \sqrt{m}(\hat{\psi} - \tilde{\psi}) + \sqrt{m}(\tilde{\psi} - \psi)$. We only need to derive, as m tends to ∞,

$$\sqrt{m}(\hat{\psi} - \tilde{\psi}) \xrightarrow{p} 0. \tag{A.3}$$

By the above estimation, it is not difficult to derive that $\mathrm{E}|\tilde{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2|^{4+\delta} = \mathrm{O(m^{-2-\delta/2})}$ and $\mathrm{E}|\tilde{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^{4+\delta} = \mathrm{O(m^{-2-\delta/2})}$. Here $0 < \delta < 1$. Then we have

$$\begin{aligned}
P(\tilde{\sigma}_b^2 < 0) &= P(\tilde{\sigma}_b^2 - \sigma_b^2 < -\sigma_b^2) \le P(|\tilde{\sigma}_b^2 - \sigma_b^2| \ge \sigma_b^2) \\
&\le (\sigma_b^2)^{-4-\delta}\mathrm{E}|\tilde{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2|^{4+\delta} = \mathrm{O(m^{-2-\delta/2})}.
\end{aligned}$$

It follows that $\mathrm{E}(\hat{\sigma}_\mathrm{b}^2 - \tilde{\sigma}_\mathrm{b}^2) = \mathrm{E}(|\tilde{\sigma}_\mathrm{b}^2|1_{\{\tilde{\sigma}_\mathrm{b}^2<0\}}) \le \mathrm{E}^{1/2}|\tilde{\sigma}_\mathrm{b}^2|^2\mathrm{P}^{1/2}(\tilde{\sigma}_\mathrm{b}^2 < 0) = \mathrm{O(m^{-1-\delta/4})}$ and $\mathrm{E}|\hat{\sigma}_\mathrm{b}^2 - \tilde{\sigma}_\mathrm{b}^2|^2 = \mathrm{E}(|\tilde{\sigma}_\mathrm{b}^2|^2 1_{\{\tilde{\sigma}_\mathrm{b}^2<0\}}) \le \mathrm{E}^{1/2}|\tilde{\sigma}_\mathrm{b}^2|^4\mathrm{P}^{1/2}(\tilde{\sigma}_\mathrm{b}^2 < 0) = \mathrm{O(m^{-1-\delta/4})}$, by Markov's inequality. Similarly we can derive that $P(\tilde{\sigma}_\varepsilon^2 < 0) = O(m^{-2-\delta/2})$ and then $\mathrm{E}|\hat{\sigma}_\varepsilon^2 - \tilde{\sigma}_\varepsilon^2|^2 = \mathrm{O(m^{-1-\delta/4})}$. Thus Equation $(A.3)$ holds, and the proof is finished by Slutsky's theorem again. ■

Proof of Corollary 1. The proof is derived similarly to that of Theorem 1. Here we do not derive it in detail. ■

Proof of Theorem 2. By model (1)'s definition and conditions $(C1) - (C3)$, we have $\hat{\beta}_{\mathrm{lse}} - \beta = O_p(m^{-1/2})$. Then it is not difficult to derive that

$$\begin{aligned}
\hat{\mu}_{\varepsilon 4} - \mu_{\varepsilon 4} &= \frac{1}{mc_1}\sum_{i=1}^{m}\sum_{j=1}^{n_i-1}\sum_{k=2}^{n_i}\left\{(d_{ij}\varepsilon_{ij} - d_{ik}\varepsilon_{ik})^4 - (d_{ij}^4 + d_{ik}^4)\mu_{\varepsilon 4} - 6d_{ij}^2d_{ik}^2(\sigma_\varepsilon^2)^2\right\} \qquad (A.4) \\
&\quad - 12c_2\sigma_\varepsilon^2(\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) + o_p(m^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^{m}\lambda_i + o_p(m^{-1/2}),
\end{aligned}$$

where c_1 and c_2 are defined in Equations (9) and (10) respectively, and

$$\lambda_i = c_1^{-1} \sum_{j=1}^{n_i-1} \sum_{k=2}^{n_i} \left\{ (d_{ij}\varepsilon_{ij} - d_{ik}\varepsilon_{ik})^4 - (d_{ij}^4 + d_{ik}^4)\mu_{\varepsilon 4} - 6d_{ij}^2 d_{ik}^2 (\sigma_\varepsilon^2)^2 \right\} - 12c_2\sigma_\varepsilon^2 \zeta_{i1}$$

with ζ_{i1} being defined in Equation $(A.1)$. By conditions $(C1)-(C3)$ and the Lindeberg-Feller central limit theorem, Equation (13) is derived.

Similarly we have

$$\begin{aligned} \hat{\mu}_{b4} - \mu_{b4} &= \frac{1}{nc_3} \sum_{i=1}^{m} \sum_{j \neq k} \Big\{ (d_{ij}^{-1} b_i + \varepsilon_{ij})^3 (d_{ij}^{-1} b_i + \varepsilon_{ij}) - d_{ij}^{-3} d_{ik}^{-1} \mu_{b4} \\ &\quad -3 d_{ij}^{-1} d_{ik}^{-1} \sigma_b^2 \sigma_\varepsilon^2 \Big\} - 3c_4 \sigma_b^2 (\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2) - 3c_4 \sigma_\varepsilon^2 (\hat{\sigma}_b^2 - \sigma_b^2) + o_p(N^{-1/2}) \qquad \text{(A.5)} \\ &= \frac{1}{n} \sum_{i=1}^{m} \chi_i + o_p(m^{-1/2}), \end{aligned}$$

where c_3 and c_4 are defined in Equations (11) and (12) respectively, and $\chi_i = c_3^{-1} \sum_{j \neq k} \left[(d_{ij}^{-1} b_i + \varepsilon_{ij})^3 (d_{ij}^{-1} b_i + \varepsilon_{ij}) - d_{ij}^{-3} d_{ik}^{-1} \mu_{b4} - 3d_{ij}^{-1} d_{ik}^{-1} \sigma_b^2 \sigma_\varepsilon^2 \right] - 3c_4 \sigma_b^2 \zeta_i - 3c_4 \sigma_\varepsilon^2 \zeta_{i2}$ with ζ_{i1} and ζ_{i2} being defined in Equations $(A.1)$ and $(A.2)$ respectively. By conditions $(C1)-(C3)$ and the Lindeberg-Feller central limit theorem, the proof of Equation (14) is completed. ■

Appendix B: Derivation of MSPE$_i$

By Corollary 1, if $\hat{\sigma}_b^2$ and $\hat{\sigma}_\varepsilon^2$ are replaced by $\tilde{\sigma}_b^2$ and $\tilde{\sigma}_\varepsilon^2$ in the remaining proof below, the difference could be absorbed into the remainders, and the leading terms still hold. For simplicity, we shall assume that this has been done, and use $\tilde{\sigma}_b^2$, $\tilde{\sigma}_\varepsilon^2$, and $\tilde{\psi}$ instead of $\hat{\sigma}_b^2$, $\hat{\sigma}_\varepsilon^2$, and $\hat{\psi}$.

Lemma 1. *Under the conditions* $(C1)-(C3)$*, Equation* (17) *holds.*

Proof of Lemma 1 . Define

$$\frac{\partial \tilde{\theta}_i}{\partial \sigma_b^2} = l_{1i}' e \qquad \text{and} \qquad \frac{\partial \tilde{\theta}_\mathrm{i}}{\partial \sigma_\varepsilon^2} = \mathrm{l}_{2\mathrm{i}}' \mathrm{e}$$

with $l_{ji} = (l_{ji1}', \ldots, l_{jin}')'$. Here $l_{1ii}' = \partial\rho_i/\partial\sigma_b^2 1_{n_i}' D_i^{-1} + J_{1ii}'$, $l_{2ii}' = \partial\rho_i/\partial\sigma_\varepsilon^2 1_{n_i}' D_i^{-1} + J_{2ii}'$, and $l_{jik} = J_{jik}$ for $i \neq k$, where $\partial\rho_i/\partial\sigma_b^2 = \sigma_\varepsilon^2/(T_i\sigma_b^2 + \sigma_\varepsilon^2)^2$, $\partial\rho_i/\partial\sigma_\varepsilon^2 = -\sigma_b^2/(T_i\sigma_b^2 + \sigma_\varepsilon^2)^2$,

$$\begin{aligned} J_{1ik}' &= -\frac{\partial\rho_i}{\partial\sigma_b^2} 1_{n_i}' D_i^{-1} X_i (X'V^{-1}X)^{-1} X_k' V_k^{-1} + (\bar{X}_{i.}' - \rho_i 1_{n_i}' D_i^{-1} X_i)(X'V^{-1}X)^{-1} \\ &\quad \times \left\{ X'V^{-1}ZZ'V^{-1}X(X'V^{-1}X)^{-1} X_k' V_k^{-1} - X_k' V_k^{-1} 1_{m_k} 1_{m_k}' V_k^{-1} \right\}, \end{aligned}$$

and J_{2ik} is defined similarly with $\frac{\partial\rho_i}{\partial\sigma_\varepsilon^2}$ and D_k instead of $\frac{\partial\rho_i}{\partial\sigma_b^2}$ and $1_{m_k} 1_{m_k}'$ respectively. By conditions (C1)-(C3) given above, it is not difficult to derive that $J_{jik} = O(m^{-1}) 1_{m_k}$ for $j =$

$1, 2$, $k = 1, \ldots, m$. Then we have

$$\begin{aligned}\frac{\partial\tilde{\theta}_i}{\partial\psi} &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}1'_{n_i}D_i^{-1}e_i(\sigma_\varepsilon^2, -\sigma_b^2)' + O(\frac{1}{m}1_2)1'_N e \\ &\equiv: I_1 + I_2.\end{aligned}$$

Put $II_1 = (\bar{X}'_{i.} - \rho_i 1'_{n_i}D_i^{-1}X_i)(\tilde{\beta}-\beta)$ and $II_2 = \rho_i 1'_{n_i}D_i^{-1}e_i - b_i$. It follows that $\tilde{\theta}_i - \theta_i = II_1 + II_2$. By Equation (5) and the proof of Theorem 1, we have

$$\begin{aligned}\hat{\psi} - \psi &= (W'W)^{-1}\sum_{i=1}^{m}\zeta_i + (W'W)^{-1}O(\frac{1}{m}1_2)\mathrm{tr}\{\mathrm{J_N(ee'-V)}\} \\ &\equiv: III_1 + III_2.\end{aligned} \tag{A.6}$$

Firstly we deal with $\mathrm{E}((\tilde{\theta}_\mathrm{i} - \theta_\mathrm{i})\mathrm{R_{i1}})$. After some tedious calculations, we have

$$\begin{aligned}\mathrm{E(II_1I_1'III_1)} &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}\mathrm{Etr}\Big\{\mathrm{(W'W)^{-1}}\sum_{\mathrm{j=1}}^{\mathrm{m}}\zeta_\mathrm{j}(\mathrm{\bar{X}'_{i.}} - \rho_\mathrm{i}1'_{\mathrm{n_i}}\mathrm{D_i^{-1}X_i})(\tilde{\beta}-\beta) \\ &\quad \times 1'_{n_i}D_i^{-1}e_i(\sigma_\varepsilon^2, -\sigma_b^2)\Big\} \\ &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}\mathrm{tr}\Big[\mathrm{(W'W)^{-1}E}\Big\{\zeta_\mathrm{i}(\mathrm{\bar{X}'_{i.}} - \rho_\mathrm{i}1'_{\mathrm{n_i}}\mathrm{D_i^{-1}X_i})\mathrm{(X'V^{-1}X)^{-1}} \\ &\quad \times X_i'V_i^{-1}e_i1'_{n_i}D_i^{-1}e_i\Big\}(\sigma_\varepsilon^2, -\sigma_b^2)\Big] \\ &\equiv: O(m^{-2}),\end{aligned}$$

$$\begin{aligned}\mathrm{E(II_2I_1'III_1)} &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}\mathrm{Etr}\left\{\mathrm{(W'W)^{-1}}\sum_{\mathrm{j=1}}^{\mathrm{m}}\zeta_\mathrm{j}(\rho_\mathrm{i}1'_{\mathrm{n_i}}\mathrm{D_i^{-1}e_i} - \mathrm{b_i})1'_{\mathrm{n_i}}\mathrm{D_i^{-1}e_i}\right. \\ &\quad \times(\sigma_\varepsilon^2, -\sigma_b^2)\Big\} \\ &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}\mathrm{tr}\Big[\mathrm{(W'W)^{-1}E}\Big\{\zeta_\mathrm{i}(\rho_\mathrm{i}1'_{\mathrm{n_i}}\mathrm{D_i^{-1}e_i} - \mathrm{b_i})1'_{\mathrm{n_i}}\mathrm{D_i^{-1}e_i}\Big\} \\ &\quad \times(\sigma_\varepsilon^2, -\sigma_b^2)\Big] \\ &= f_{i3}(\psi^*),\end{aligned}$$

where $f_{i3}(\cdot)$ is defined in Equation (18) . Similarly we can derive $\mathrm{E(II_1(I_1'III_2 + I_2'III_1 + I_2'III_2)) = o(m^{-2})}$, $\mathrm{E(II_2(I_1'III_2 + I_2'III_1)) = o(m^{-2})}$, and $\mathrm{E(II_2I_2'III_2) = o(m^{-2})}$. It follows that $\mathrm{E}((\tilde{\theta}_\mathrm{i} - \theta_\mathrm{i})\mathrm{R_{i1}}) = \mathrm{f_{i3}}(\psi^*) + \mathrm{O(m^{-2})}$. ■

Lemma 2. *Under the conditions* $(C1)-(C3)$, *Equation* (19) *holds.*

Proof of Lemma 2 . Similar to the above proof, we can derive $\mathrm{ER}_{i1}^2 = \mathrm{E}(\mathrm{I}_1'\mathrm{II}_1)^2 + \mathrm{O}(\mathrm{m}^{-2})$ with

$$\begin{aligned}
\mathrm{E}(\mathrm{I}_1'\mathrm{II}_1)^2 &= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^4}\mathrm{Etr}\left\{(\mathrm{W}'\mathrm{W})^{-1}\sum_{\mathrm{j}=1}^{\mathrm{m}}\zeta_\mathrm{j}\zeta_\mathrm{j}'(\mathrm{W}'\mathrm{W})^{-1}\right.\\
&\quad\left.\times(1_{n_i}'D_i^{-1}e_i)^2\begin{pmatrix}(\sigma_\varepsilon^2)^2 & -\sigma_b^2\sigma_\varepsilon^2\\ -\sigma_b^2\sigma_\varepsilon^2 & (\sigma_b^2)^2\end{pmatrix}\right\}\\
&= \frac{1}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^4}\mathrm{tr}\left[(\mathrm{W}'\mathrm{W})^{-1}\left\{\sum_{\mathrm{j}\neq\mathrm{i}}\Sigma_\mathrm{j}\mathrm{E}(1_{\mathrm{n_i}}'\mathrm{D}_\mathrm{i}^{-1}\mathrm{e_i})^2\right.\right.\\
&\quad\left.\left.+\mathrm{E}(\zeta_\mathrm{i}(1_{\mathrm{n_i}}'\mathrm{D}_\mathrm{i}^{-1}\mathrm{e_i})^2)\right\}(W'W)^{-1}\begin{pmatrix}(\sigma_\varepsilon^2)^2 & -\sigma_b^2\sigma_\varepsilon^2\\ -\sigma_b^2\sigma_\varepsilon^2 & (\sigma_b^2)^2\end{pmatrix}\right]\\
&= \frac{T_i}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^3}\mathrm{tr}\left\{(\mathrm{W}'\mathrm{W})^{-1}\sum_{\mathrm{j}=1}^{\mathrm{m}}\Sigma_\mathrm{j}(\mathrm{W}'\mathrm{W})^{-1}\right.\\
&\quad\left.\times\begin{pmatrix}(\sigma_\varepsilon^2)^2 & -\sigma_b^2\sigma_\varepsilon^2\\ -\sigma_b^2\sigma_\varepsilon^2 & (\sigma_b^2)^2\end{pmatrix}\right\}+O(m^{-2})\\
&= f_{i41}(\psi^*)+f_{i42}(\psi)+O(m^{-2}),
\end{aligned}$$

where the last equation holds by $\Sigma_j = \Sigma_{j1} + \Sigma_{j2}$ with Σ_{j1} and Σ_{j2} defined in Equations (7) and (8). ■

In order to prove Theorem 3, we need the following lemma in Jiang, Lahiri, & Wan (2002).

Lemma 3. *Assume ξ_m, ξ_{m1}, and ξ_{m2} are random variables, and $\mathcal{B}_m$ is a set, such that*
i) $\mathrm{E}|\xi_\mathrm{m}^2|1_{\mathcal{B}_\mathrm{m}^\mathrm{c}} \le \mathrm{c}_1\mathrm{m}^{-\mathrm{a}_1}$;
ii) $\mathrm{E}|\xi_{\mathrm{m}1}^2|1_{\mathcal{B}_\mathrm{m}^\mathrm{c}} \le \mathrm{c}_2\mathrm{m}^{-\mathrm{a}_2}$, and $\mathrm{E}\xi_{\mathrm{m}1}^2 \le \mathrm{c}_4$; and
iii) $\xi_m = \xi_{m1} + \xi_{m2}$ on $\mathcal{B}_m$ with $|\xi_{m2}| \le m^{-a_3}u_m$ and $\mathrm{Eu}_\mathrm{m}^2 \le \mathrm{c}_3$, where the a's and c's are positive constants. Then, for any $0 < \varpi < a_1 \wedge a_2 \wedge a_3$, we have

$$\mathrm{E}\xi_\mathrm{m}^2 = \mathrm{E}\xi_{\mathrm{m}1}^2 + \mathrm{O}(\mathrm{m}^\varpi).$$

Proof of Theorem 3 . Define $X_i = (X_{i1}, \ldots, X_{ip})'$. It is easy to see that $V_i^{-1} = \sigma_\varepsilon^{-2}D_i^{-1} - \sigma_\varepsilon^{-2}\rho_i D_i^{-1}1_{n_i}1_{n_i}'D_i^{-1}$. By some calculations, $\tilde{\theta}_i$ can be rewritten as

$$\begin{aligned}
\tilde{\theta}_i - \bar{X}_{i.}'\beta &= \rho_i 1_{n_i}'D_i^{-1}e_i \qquad\qquad (\mathrm{A}.7)\\
&\quad+\sum_{j=1}^{m}\sum_{k=1}^{m_j}(\bar{X}_{i.} - \rho_i X_i'D_i^{-1}1_{n_i})'(\sigma_\varepsilon^2 X'V^{-1}X)^{-1}X_{jk}e_{jk}/d_{jk}^2\\
&\quad-\sum_{j=1}^{m}\rho_j(\bar{X}_{i.} - \rho_i X_i'D_i^{-1}1_{n_i})'(\sigma_\varepsilon^2 X'V^{-1}X)^{-1}X_j'D_j^{-1}1_{m_j}1_{m_j}'D_j^{-1}e_j\\
&\equiv: \sum_{j=1}^{K}\lambda_j(\psi)W_j(e).
\end{aligned}$$

Note that $K = n + 1 + \sum_{i=1}^{m} n_i = O(m)$ and the following terms are bounded for some $s > 2$ and $t > 0$:

$$\text{i)} \max_{1\le k\le K} E|W_k(e)|^s,\ \text{ii)} \max_{1\le k\le K} \sup_{\psi} |\lambda_j(\psi)|,\ \text{iii)} \sum_{k=1}^{K} \left|\frac{\partial \lambda_k}{\partial \psi}\right|,\ \text{iv)} \sum_{k=1}^{K} \sup_{|\psi-\psi_0|\le t} \left\|\frac{\partial^2 \lambda_k}{\partial \psi \partial \psi'}\right\|.$$

Here and below ψ_0 denotes the true value of ψ. Define $\mathcal{B} = \{|\hat{\psi} - \psi| < n^{-\delta}, \psi \in \Theta\}$, where $0 < \delta < 1/2$. By the Taylor expansion, on $\mathcal{B}$, we have

$$\begin{aligned} \hat{\theta}_i &= \tilde{\theta}_i + \frac{\partial \tilde{\theta}_i}{\partial \psi'}(\hat{\psi} - \psi) + \frac{1}{2}(\hat{\psi} - \psi)' \frac{\partial \tilde{\theta}_i(\tilde{\psi})}{\partial \psi \partial \psi'}(\hat{\psi} - \psi) \\ &\equiv: \tilde{\theta}_i + R_{i1} + R_{i2}, \end{aligned} \tag{A.8}$$

where $\tilde{\psi}$ lies between ψ and $\hat{\psi}$, and for large m, $|R_{i2}| \le u_i|\hat{\psi} - \psi|^2 \le n^{-2\delta}u_i$ with

$$u_i = \frac{1}{2}\sum_{i=1}^{K} \sup_{|\psi-\psi_0|<n^{-1/2}} \left\|\frac{\partial^2 \lambda_k}{\partial \psi \partial \psi'}\right\| |W_k(e)|.$$

It is easy to show that $E(u_i^2)$ is bounded.

Let $\xi_i = \sqrt{m}(\hat{\theta}_i - \theta_i)$, $\xi_{i1} = \sqrt{m}(\tilde{\theta}_i + R_{i1} - \theta_i)$, and $\xi_{i2} = \sqrt{m}R_{i2}$. By Lemma 3, we only need to derive i) and ii).

Firstly we deal with $E|\xi_i^2|1_{\mathcal{B}^c}$. By Theorem 1, for any $r > 0$, we have $|\hat{\sigma}_b^2 - \sigma_b^2|^r = O(m^{-r/2})$ and $E|\hat{\gamma}_\epsilon^2 - \sigma_\varepsilon^2|^r = O(m^{-r/2})$. Then it follows from Markov's inequality that $P(\mathcal{B}^c) = O(m^{-k(1/2-\delta)})$ with $k > 0$. By the Cauchy-Schwarz inequality, we have

$$E(|\xi_i^2|1_{\mathcal{B}^c}) \le E^{1/2}|\xi_i|^4 P^{1/2}(\mathcal{B}^c).$$

On the other hand, it is easy to see that

$$|\rho_i| \le T_i^{-1} \quad \text{and} \quad (X'V^{-1}X)^{-1} \le \sigma_\varepsilon^2 d_l^2 (X'X)^{-1}$$

with d_l defined in condition $(C3)$. By Equation $(A.7)$, we have

$$\begin{aligned} |\tilde{\theta}_i - \theta_i|^4 \le c_1 \Big[& b_i^4 + |1_{n_i}' D_i^{-1} e_i|^4 \\ & + \frac{1}{m}\sum_{j=1}^{m}\sum_{k=1}^{m_j} \left\{(\bar{X}_{i.} - \rho_i X_i' D_i^{-1} 1_{n_i})'(\sigma_\varepsilon^2 n^{-1} X'V^{-1}X)^{-1} X_{jk} e_{jk}\right\}^4 \\ & + \frac{1}{m}\sum_{j=1}^{m} \left\{((\bar{X}_{i.} - \rho_i X_i' D_i^{-1} 1_{n_i})'(\sigma_\varepsilon^2 n^{-1} X'V^{-1}X)^{-1} X_j' D_j^{-1} 1_{m_j} 1_{m_j}' D_j^{-1} e_j\right\}^4 \Big] \\ \le c_2 \Big[& b_i^4 + T_i^{-1}|1_{n_i}' D_i^{-1} e_i|^4 \\ & + \frac{1}{m}\sum_{j=1}^{m}\sum_{k=1}^{m_j} (|\bar{X}_{i.}|^4 + |X_i' D_i^{-1} 1_{n_i}|^4) \mathrm{tr}^4\{n(X'X)^{-1}\} |X_{jk}|^4 e_{jk}^4 \\ & + \frac{1}{m}\sum_{j=1}^{m} (|\bar{X}_{i.}|^4 + |X_i' D_i^{-1} 1_{n_i}|^4) \mathrm{tr}^4(n(X'X)^{-1}) |X_j' D_j^{-1} 1_{m_j}|^4 |1_{m_j}' D_j^{-1} e_j|^4 \Big]. \end{aligned}$$

The above inequality tells us that $\mathrm{E}|\tilde{\theta}_{\mathrm{i}}-\theta_{\mathrm{i}}|^4=\mathrm{O}(1)$ and $\mathrm{E}|\hat{\theta}_{\mathrm{i}}-\theta_{\mathrm{i}}|^4=\mathrm{O}(1)$. Hence we have $\mathrm{E}|\xi_{\mathrm{i}}|^4=\mathrm{O}(\mathrm{m}^2)$. It follows that

$$\mathrm{E}(|\xi_{\mathrm{i}}^2|1_{\mathcal{B}^{\mathrm{c}}})=\mathrm{O}(\mathrm{m}^{-(\mathrm{k}(1/2-\delta)-1)}). \tag{A.9}$$

Now we move to deal with $\mathrm{E}|\xi_{\mathrm{i1}}^2|1_{\mathcal{B}^{\mathrm{c}}}$. By Equations (16), (17), and (19) , we have $\mathrm{E}\xi_{\mathrm{i1}}^2=\mathrm{O}(1)$. By the Cauchy-Schwarz inequality again, we have

$$\mathrm{E}(|\xi_{\mathrm{i1}}^2|1_{\mathcal{B}^{\mathrm{c}}})\le \mathrm{E}^{1/2}|\xi_{\mathrm{i1}}|^4\mathrm{P}^{1/2}(\mathcal{B}^{\mathrm{c}})=\mathrm{O}(\mathrm{m}^{-(\mathrm{k}(1/2-\delta)-1)}). \tag{A.10}$$

By Equations $(A.8)-(A.10)$, and Lemma 3, the proof is finished. ■

Appendix C: Derivation of $\widehat{\mathrm{MSPE}}_{1i}$

Lemma 4. *Under the conditions of Theorem 3, we have*

$$\mathrm{Ef}_{\mathrm{i1}}(\hat{\psi})=f_{i1}(\psi)-f_{i41}(\psi^*)-f_{i42}(\psi)+O(m^{-3/2}), \tag{A.11}$$

$$\mathrm{Ef}_{\mathrm{i2}}(\hat{\psi})=f_{i2}(\psi)+O(m^{-3/2}), \tag{A.12}$$

$$\mathrm{Ef}_{\mathrm{i3}}(\hat{\psi}^*)=f_{i3}(\psi^*)+O(m^{-3/2}), \tag{A.13}$$

$$\mathrm{Ef}_{\mathrm{i41}}(\hat{\psi}^*)=f_{i41}(\psi^*)+O(m^{-3/2}), \tag{A.14}$$

$$\mathrm{Ef}_{\mathrm{i42}}(\hat{\psi})=f_{i42}(\psi)+O(m^{-3/2}). \tag{A.15}$$

Proof of Lemma 4 . Firstly we deal with $(A.11)$. By the elementary expansion, we have

$$\begin{aligned}\frac{1}{T_i\hat{\sigma}_b^2+\hat{\sigma}_\varepsilon^2} &= \frac{1}{T_i\sigma_b^2+\sigma_\varepsilon^2}-\frac{T_i(\hat{\sigma}_b^2-\sigma_b^2)+\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2}\\ &\quad+\frac{\{T_i(\hat{\sigma}_b^2-\sigma_b^2)+\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\}^2}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^3}-\frac{\{T_i(\hat{\sigma}_b^2-\sigma_b^2)+\hat{\sigma}_\varepsilon^2-\sigma_\varepsilon^2\}^3}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^3(T_i\hat{\sigma}_b^2+\hat{\sigma}_\varepsilon^2)}\\ &\equiv: K_1+K_2+K_3+K_4.\end{aligned} \tag{A.16}$$

It follows that

$$\begin{aligned}f_{i1}(\hat{\psi}) &= \hat{\sigma}_b^2-\frac{T_i(\hat{\sigma}_b^2)^2}{T_i\hat{\sigma}_b^2+\hat{\sigma}_\varepsilon^2}\\ &= \hat{\sigma}_b^2-T_i(\hat{\sigma}_b^2)^2\sum_{j=1}^4 K_j.\end{aligned} \tag{A.17}$$

Note that

$$(\hat{\sigma}_b^2)^2 \sum_{j=1}^{3} K_j = \frac{(\sigma_b^2)^2}{T_i\sigma_b^2 + \sigma_\varepsilon^2} + \frac{2\sigma_b^2(\hat{\sigma}_b^2 - \sigma_b^2)}{T_i\sigma_b^2 + \sigma_\varepsilon^2} - \frac{(\sigma_b^2)^2\left\{T_i(\hat{\sigma}_b^2 - \sigma_b^2) + \hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right\}}{(T_i\sigma_b^2 + \sigma_\varepsilon^2)^2} \tag{A.18}$$

$$+\frac{1}{(T_i\sigma_b^2 + \sigma_\varepsilon^2)^3}(\hat{\psi} - \psi)' \begin{pmatrix} (\sigma_\varepsilon^2)^2 & -\sigma_b^2\sigma_\varepsilon^2 \\ -\sigma_b^2\sigma_\varepsilon^2 & (\sigma_b^2)^2 \end{pmatrix} (\hat{\psi} - \psi)$$

$$+\frac{\left\{(\hat{\sigma}_b^2 - \sigma_b^2)^2 + 2\sigma_b^2(\hat{\sigma}_b^2 - \sigma_b^2)\right\}\left\{T_i(\hat{\sigma}_b^2 - \sigma_b^2) + \hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2\right\}^2}{(T_i\sigma_b^2 + \sigma_\varepsilon^2)^3}.$$

By Equations $(A.16)$, $(A.18)$, (19), and Theorem 1,

$$\mathrm{E}\left\{\hat{\sigma}_\mathrm{b}^2 - \mathrm{T_i}(\hat{\sigma}_\mathrm{b}^2)^2 \sum_{\mathrm{j}=1}^{3} \mathrm{K_j}\right\} = f_{i1}(\psi) - f_{i41}(\psi) - f_{i42}(\psi^*) + O(m^{-3/2}). \tag{A.19}$$

On the other hand, by the Cauchy-Schwarz inequality and the C_r inequality

$$\begin{aligned} \mathrm{E}|(\hat{\sigma}_\mathrm{b}^2)^2\mathrm{K}_4| &\le \frac{\mathrm{E}\left\{|\hat{\sigma}_\mathrm{b}^2||\mathrm{T_i}(\hat{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2) + \hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^3\right\}}{T_i(T_i\sigma_b^2 + \sigma_\varepsilon^2)^3} \\ &\le c\mathrm{E}^{1/2}|\hat{\sigma}_\mathrm{b}^2|^2(\mathrm{T_i^6}\mathrm{E}^{1/2}|\hat{\sigma}_\mathrm{b}^2 - \sigma_\mathrm{b}^2|^6 + \mathrm{E}^{1/2}|\hat{\sigma}_\varepsilon^2 - \sigma_\varepsilon^2|^6) \\ &= O(m^{-3/2}) \end{aligned} \tag{A.20}$$

with c is some constant. By Equations $(A.17)$, $(A.19)$, and $(A.20)$, Equation $(A.11)$ is resluted.

Now we move to prove Equation $(A.12)$. By the elementary expansion again, we have

$$\begin{aligned} &(X'\hat{V}^{-1}X)^{-1} - (X'V^{-1}X)^{-1} \\ &= (X'V^{-1}X)^{-1}X'V^{-1}(\hat{V} - V)\hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1} \\ &\equiv: G. \end{aligned} \tag{A.21}$$

By the definition of $f_{i2}(\cdot)$ in Equation (16) and $(A.21)$, we have

$$\begin{aligned} f_{i2}(\hat{\psi}) - f_{i2}(\psi) \;=\; & (\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i})' G (\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i}) \\ & -2(\hat{\rho}_i - \rho_i) 1_{n_i}' D_i^{-1} X_i (X'V^{-1}X)^{-1} (\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i}) \\ & -2(\hat{\rho}_i - \rho_i) 1_{n_i}' D_i^{-1} X_i G (\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i}) \\ & +(\hat{\rho}_i - \rho_i)^2 1_{n_i}' D_i^{-1} X_i (X'V^{-1}X)^{-1} X_i' D_i^{-1} 1_{n_i} \\ & +(\hat{\rho}_i - \rho_i)^2 1_{n_i}' D_i^{-1} X_i G X_i' D_i^{-1} 1_{n_i} \\ \equiv: & \sum_{j=1}^{5} L_j. \end{aligned}$$

Next we will derive $\mathrm{EL_j} = \mathrm{O(m^{-3/2})}$ for $j = 1, 2, 3$, and $\mathrm{EL_j} = \mathrm{O(m^{-2})}$ for $j = 4, 5$. Hence Equation $(A.12)$ is produced.

Firstly we deal with L_1. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
|L_1| &\le (\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i})'(\bar{X}_{i\cdot}' - \rho_i X_i' D_i^{-1} 1_{n_i}) \mathrm{tr}^{1/2}(\mathrm{G'G}) \\
&\le c_1 \mathrm{tr}^{1/2}\left\{(\mathrm{X'V^{-1}X})^{-2}\mathrm{X'V^{-2}X}\right\} \mathrm{tr}^{1/2}\left\{(\mathrm{X'\hat{V}^{-1}X})^{-2}\right\} \\
&\quad \times \mathrm{tr}^{1/2}(\mathrm{X'\hat{V}^{-2}X})\mathrm{tr}^{1/2}\left\{(\mathrm{\hat{V}-V})^2\right\} \\
&= c_2 O(m^{-1})(|\hat\sigma_b^2 - \sigma_b^2| + |\hat\sigma_\varepsilon^2 - \sigma_\varepsilon^2|),
\end{aligned} \tag{A.22}
$$

because $\mathrm{tr}\left\{(\mathrm{X'V^{-1}X})^{-2}\mathrm{X'V^{-2}X}\right\} = \mathrm{O(m^{-1})}$, $\mathrm{tr}\left\{(\mathrm{\hat{V}-V})^2\right\} = \mathrm{O(m)}(|\hat\sigma_{\mathrm{b}}^2 - \sigma_{\mathrm{b}}^2| + |\hat\sigma_\varepsilon^2 - \sigma_\varepsilon^2|)$, $\mathrm{tr}\left\{(\mathrm{X'\hat{V}^{-1}X})^{-2}\right\} = |\hat\sigma_\varepsilon^2|^2\mathrm{O(m^{-2})}$, and $\mathrm{tr(X'\hat{V}^{-2}X)} = |\hat\sigma_\varepsilon^2|^{-2}\mathrm{O(m)}$, where the last two equalities are derived by

$$
\begin{aligned}
&\sigma_\varepsilon^{-2}(D^{-1} - \mathrm{diag}\{\mathrm{D_1^{-2}/T_1}, \ldots, \mathrm{D_n^{-2}/T_m}\}) \\
&\le V^{-1} = \gamma_\varepsilon^{-2}(D^{-1} - \mathrm{diag}\{\rho_1\mathrm{D_1^{-2}}, \ldots, \rho_{\mathrm{n}}\mathrm{D_m^{-2}}\}) \le \sigma_\varepsilon^{-2}\mathrm{D^{-1}}.
\end{aligned}
$$

Here c_1 and c_2 are some constants. It follows from $(A.22)$ that $\mathrm{E|L_1| = O(m^{-3/2})}$.

Now we deal with L_2. By the elementary expansion, we have

$$
\hat\rho_i - \rho_i = \frac{\hat\sigma_b^2 - \sigma_b^2}{T_i\sigma_b^2 + \sigma_\varepsilon^2} - \frac{\left\{T_i(\hat\sigma_b^2 - \sigma_b^2) + \hat\sigma_\varepsilon^2 - \sigma_\varepsilon^2\right\}\hat\sigma_b^2}{(T_i\sigma_b^2 + \sigma_\varepsilon^2)^2(T_i\hat\sigma_b^2 + \hat\sigma_\varepsilon^2)}. \tag{A.23}
$$

Then it is not difficult to derive that $\mathrm{E|\hat\rho_i - \rho_i|^k = O(m^{-k/2})}$ for $k \ge 1$. Hence $\mathrm{E|L_2| = O(m^{-3/2})}$.

Next we move to L_3. By the Cauchy-Schwarz inequality,

$$
\mathrm{E|L_3|} \le c\mathrm{E^{1/2}|\hat\rho_i - \rho_i|^2E^{1/2}\{tr(GG')\} = O(m^{-1/2})E^{1/2}\{tr(GG')\}}, \tag{A.24}
$$

where c is some constant. Similar to the proof of $\mathrm{E|L_1| = O(m^{-3/2})}$ in the above, one can obtain $\mathrm{Etr(GG') = O(m^{-2})}$. Hence $\mathrm{E|L_3| = O(m^{-3/2})}$.

Similar to the above arguments, we obtain $\mathrm{E|L_j| = O(m^{-2})}$ for $j = 4, 5$.

Finally, Equations $(A.13) - (A.15)$ can be derived by the similar arguments. The proof is finished. ■

Appendix D: Derivation of $\widehat{\mathrm{MSPE}}_{2i}$ and $\widehat{\mathrm{MSPE}}_{3i}$

Lemma 5. *Under the conditions (C1) and (C2), we have*

$$
(W_{-j}'W_{-j})^{-1} = (W'W)^{-1} + (W'W)^{-1}\Delta_j(W'W)^{-1} + O(m^{-3})J_2, \tag{A.25}
$$

where $\Delta_j = \begin{pmatrix} \Delta_j^{11} & \Delta_j^{12} \\ \Delta_j^{12} & \Delta_j^{22} \end{pmatrix}$ *with* Δ_j^{11}, Δ_j^{12} *and* Δ_j^{22} *defined in Equations* $(A.26)$-$(A.28)$ *below respectively. Moreover,* $\sum_{j=1}^m \Delta_j = W'W$.

Proof of Lemma 5 . Let (X_{-j}, Z_{-j}, Y_{-j}) be defined as (X, Z, Y) by dropping out the jth group. And then W_{-j} and $P_{X^{\perp}_{-j}}$ are obtained. After some tedious calculations, we have

$$\begin{aligned} \mathrm{tr}\left\{(\mathrm{Z}'_{-j}\mathrm{P}_{\mathrm{X}^{\perp}_{-j}}\mathrm{Z}_{-j})^2\right\} &= \mathrm{tr}\left\{(\mathrm{Z}'\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{Z})^2\right\} - \mathrm{tr}\left\{(\mathrm{Z}_j'\mathrm{Z}_j)^2\right\} + 2\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'(\mathrm{Z}_j\mathrm{Z}_j')^2\mathrm{X}_j\right\} \\ &\quad -\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Z}\mathrm{Z}'\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'\mathrm{Z}_j\mathrm{Z}_j'\mathrm{X}_j\right\} + \mathrm{O}(m^{-1}) \\ &\equiv: \mathrm{tr}\left\{(\mathrm{Z}'\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{Z})^2\right\} - \Delta_j^{11} + \mathrm{O}(m^{-1}) \end{aligned} \quad (A.26)$$

$$\begin{aligned} \mathrm{tr}(\mathrm{Z}'_{-j}\mathrm{P}_{\mathrm{X}^{\perp}_{-j}}\mathrm{D}_{-j}\mathrm{P}_{\mathrm{X}^{\perp}_{-j}}\mathrm{Z}_{-j}) &= \mathrm{tr}(\mathrm{Z}'\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{D}\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{Z}) - \mathrm{tr}(\mathrm{Z}_j'\mathrm{D}_j\mathrm{Z}_j) \\ &\quad +\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'(\mathrm{D}_j\mathrm{Z}_j\mathrm{Z}_j' + \mathrm{Z}_j\mathrm{Z}_j'\mathrm{D}_j)\mathrm{X}_j\right\} \\ &\quad -\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'\mathrm{D}_j\mathrm{X}_j(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Z}\mathrm{Z}'\mathrm{X}\right\} + \mathrm{O}(m^{-1}) \\ &\equiv: \mathrm{tr}(\mathrm{Z}'\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{D}\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{Z}) - \Delta_j^{12} + \mathrm{O}(m^{-1}), \end{aligned} \quad (A.27)$$

$$\begin{aligned} \mathrm{tr}((\mathrm{P}_{\mathrm{X}^{\perp}_{-j}}\mathrm{D}_{-j})^2) &= \mathrm{tr}\left\{(\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{D})^2\right\} - \mathrm{tr}(\mathrm{D}_j^2) + 2\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'\mathrm{D}_j^2\mathrm{X}_j\right\} \\ &\quad -\mathrm{tr}\left\{(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}_j'\mathrm{D}_j\mathrm{X}_j(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{D}\mathrm{X}\right\} + \mathrm{O}(m^{-1}) \\ &\equiv: \mathrm{tr}\left\{(\mathrm{P}_{\mathrm{X}^{\perp}}\mathrm{D})^2\right\} - \Delta_j^{22} + \mathrm{O}(m^{-1}), \end{aligned} \quad (A.28)$$

noting that $(X'_{-j}X_{-j})^{-1} = (X'X)^{-1} + (X'X)^{-1}X_j'(I_{m_j} - X_j(X'X)^{-1}X_j')^{-1}X_j'(X'X)^{-1}$. It follows from Equations $(A.26) - (A.28)$ that

$$W'_{-j}W_{-j} = W'W - \Delta_j + O(\frac{1}{m})J_2. \quad (A.29)$$

On the other hand,

$$\begin{aligned} (W'_{-j}W_{-j})^{-1} &= (W'W)^{-1} + (W'W)^{-1}(W'W - W'_{-j}W_{-j})(W'W)^{-1} \\ &\quad +[(W'W)^{-1}(W'W - W'_{-j}W_{-j})]^2(W'_{-j}W_{-j})^{-1} \\ &= (W'W)^{-1} + (W'W)^{-1}(W'W - W'_{-j}W_{-j})(W'W)^{-1} + O(\frac{1}{m^3})J_2 \end{aligned} \quad (A.30)$$

by $m(W'W)^{-1} = O(1)J_2$ and $m(W'_{-j}W_{-j})^{-1} = O(1)J_2$. By Equations $(A.29)$ and $(A.30)$, Equation $(A.25)$ is derived. ■

Lemma 6. *Under the conditions of Theorem 1,*

$$\hat{\psi}_{-j} - \hat{\psi} = -(W'W)^{-1}\zeta_j + (W'W)^{-1}\Delta_j(W'W)^{-1}\zeta + r_{j1} + r_{j2},$$

where $\hat{\psi}_{-j}$ is defined as $\hat{\psi}$ with the jth group dropped out, $\zeta_j = (\zeta_{j1}, \zeta_{j2})'$ with ζ_{j1} and ζ_{j2} defined in Equations $(A.1)$ and $(A.2)$ respectively, $\zeta = \sum_{j=1}^m \zeta_j$, and r_{j1} and r_{j2} are defined in Equation $(A.33)$ below. Moreover

$$\sum_{j=1}^m (\hat{\psi}_{-j} - \hat{\psi}) = O(\frac{1}{m^2})J_2\zeta + O(\frac{1}{m^3})1_2\eta$$

with $\eta = \mathrm{tr}(\mathrm{J_N}(\mathrm{ee}' - \mathrm{V}))$.

Proof of Lemma 6. Define $\eta_{-j} = \eta - \eta_j$ with $\eta_j = \mathrm{tr}(\mathrm{e_j e_j'} - \mathrm{V_j}) + \sum_{\mathrm{k}\neq\mathrm{j}} \mathrm{e_j' e_k}$. It is easy to derive that $\eta = \sum_{j=1}^m \eta_j$. By Equation $(A.6)$, we have

$$\hat{\psi} - \psi = (W'W)^{-1}\zeta + O(\frac{1}{m})(W'W)^{-1}1_2\eta. \tag{A.31}$$

Similarly we have

$$\hat{\psi}_{-j} - \psi = (W_{-j}'W_{-j})^{-1}\zeta - (W_{-j}'W_{-j})^{-1}\zeta_j + O(\frac{1}{m})(W_{-j}'W_{-j})^{-1}1_2\eta_{-j}. \tag{A.32}$$

It follows from Equations $(A.25)$, $(A.31)$, and $(A.32)$ that

$$\begin{aligned} \hat{\psi}_{-j} - \hat{\psi} &= -(W'W)^{-1}\zeta_j + (W'W)^{-1}\Delta_j(W'W)^{-1}\zeta \\ &\quad + O(\frac{1}{n^2})J_2\zeta_j + O(\frac{1}{m^3})J_2\zeta + O(\frac{1}{m})\left\{-(W'W)^{-1} + O(\frac{1}{m^2})\right\}1_2\eta_{-j} \\ &\quad + O(\frac{1}{m})\left\{(W'W)^{-1}\Delta_j(W'W)^{-1} + O(\frac{1}{m^3})J_2\right\}1_2\eta \\ &\equiv: -(W'W)^{-1}\zeta_j + (W'W)^{-1}\Delta_j(W'W)^{-1}\sum_{k=1}^m \zeta_k + r_{j1} + r_{j2}, \end{aligned} \tag{A.33}$$

where $r_{j1} = O(\frac{1}{m^2})J_2\zeta_j + O(\frac{1}{m^3})J_2\zeta$ and r_{j2} equals the sum of the last terms on the right-hand side of Equation $(A.33)$. By Lemma 5, the proof is finished. ■

Proof of Theorem 4 . Similar to the proof of Theorem 3, we have

$$E(\widehat{\mathrm{MSPE}}_{\mathrm{i}}^{*}(\hat{\psi})|Y) = f_{i1}(\hat{\psi}) + f_{i2}(\hat{\psi}) + f_{i3}(\hat{\psi}^*) + f_{i41}(\hat{\psi}^*) + f_{i42}(\hat{\psi}) + O(m^{-3/2}).$$

By Lemma 4, Equation (23) holds, and then Equation (26) is proved.

Moving to Equation (27). Note that we just need to prove

$$\mathrm{E}(\widehat{\mathrm{Bias}}) = -\mathrm{f_{i41}}(\psi^*) - \mathrm{f_{i42}}(\psi) + \mathrm{O}(\mathrm{m}^{-3/2}). \tag{A.34}$$

By the Taylor expansion, we have

$$\begin{aligned} f_{i1}(\hat{\psi}_{-j}) - f_{i1}(\hat{\psi}) &= \frac{\partial f_{i1}(\hat{\psi})}{\partial \psi'}(\hat{\psi}_{-j} - \hat{\psi}) + \frac{1}{2}(\hat{\psi}_{-j} - \hat{\psi})'\frac{\partial^2 f_{i1}(\breve{\psi})}{\partial\psi\partial\psi'}(\hat{\psi}_{-j} - \hat{\psi}) \\ &= \frac{\partial f_{i1}(\hat{\psi})}{\partial \psi'}(\hat{\psi}_{-j} - \hat{\psi}) + \frac{1}{2}(\hat{\psi}_{-j} - \hat{\psi})'\frac{\partial^2 f_{i1}(\psi)}{\partial\psi\partial\psi'}(\hat{\psi}_{-j} - \hat{\psi}) \\ &\quad + \frac{1}{2}(\hat{\psi}_{-j} - \hat{\psi})'\left\{\frac{\partial^2 f_{i1}(\breve{\psi})}{\partial\psi\partial\psi'} - \frac{\partial^2 f_{i1}(\psi)}{\partial\psi\partial\psi'}\right\}(\hat{\psi}_{-j} - \hat{\psi}) \\ &= L_{j1} + L_{j2} + L_{j3}, \end{aligned}$$

where $\breve{\psi}$ lies between $\hat{\psi}$ and $\hat{\psi}_{-j}$, and $\tilde{\psi}$ lies between $\hat{\psi}$ and ψ. It is sufficient to prove that

$$\mathrm{E}\left(\frac{\mathrm{m}-1}{\mathrm{m}}\sum_{\mathrm{j}=1}^{\mathrm{m}}\mathrm{L}_{\mathrm{j1}}\right)=\mathrm{O}(\mathrm{m}^{-3/2}), \tag{A.35}$$

$$\mathrm{E}\left(\frac{\mathrm{m}-1}{\mathrm{m}}\sum_{\mathrm{j}=1}^{\mathrm{m}}\mathrm{L}_{\mathrm{j2}}\right)=-\mathrm{f}_{41}(\psi^*)-\mathrm{f}_{\mathrm{i}42}(\psi)+\mathrm{O}(\mathrm{m}^{-2}), \tag{A.36}$$

$$\mathrm{E}\left(\frac{\mathrm{m}-1}{\mathrm{m}}\sum_{\mathrm{j}=1}^{\mathrm{m}}\mathrm{L}_{\mathrm{j3}}\right)=\mathrm{O}(\mathrm{m}^{-2}). \tag{A.37}$$

Now we derive Equation $(A.35)$. By Lemma 6 and the Cauchy-Schwarz inequality, we have

$$\mathrm{E}\frac{\mathrm{m}-1}{\mathrm{m}}|\sum_{\mathrm{j}=1}^{\mathrm{m}}\mathrm{L}_{\mathrm{j1}}|\leq O(1)\mathrm{E}^{1/2}\frac{\partial \mathrm{f}_{\mathrm{i1}}(\hat{\psi})}{\partial\psi}\frac{\partial \mathrm{f}_{\mathrm{i1}}(\hat{\psi})}{\partial\psi'}$$
$$\times\mathrm{E}^{1/2}\left\{\sum_{\mathrm{j}=1}^{\mathrm{m}}(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})'\sum_{\mathrm{j}=1}^{\mathrm{m}}(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})\right\}.$$

By the definition of f_{i1} in Equation (16) , we have $\frac{\partial f_{i1}(\hat{\psi})}{\partial\psi}=(\frac{(\sigma_\varepsilon^2)^2}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2},\frac{T_i(\sigma_b^2)^2}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^2})'$, and then $\mathrm{E}(\frac{\partial \mathrm{f}_{\mathrm{i1}}(\hat{\psi})}{\partial\psi'}\frac{\partial \mathrm{f}_{\mathrm{i1}}(\hat{\psi})}{\partial\psi})=\mathrm{O}(1)$. By Lemma 5.6 and the proof of Theorem 1,

$$\mathrm{E}\left\{\sum_{\mathrm{j}=1}^{\mathrm{m}}(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})'\sum_{\mathrm{j}=1}^{\mathrm{m}}(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})\right\}$$
$$=\mathrm{E}\left\{(\mathrm{O}(\frac{1}{\mathrm{m}^2})\mathrm{J}_2\zeta+\mathrm{O}(\frac{1}{\mathrm{m}^3})\mathrm{J}_2\eta)'(\mathrm{O}(\frac{1}{\mathrm{m}^2})\mathrm{J}_2\zeta+\mathrm{O}(\frac{1}{\mathrm{m}^3})\mathrm{J}_2\eta)\right\}$$
$$=O(m^{-3})$$

Hence Equation $(A.35)$ holds.

Now moving on to Equation $(A.36)$. Define $r_{j3}=(W'W)^{-1}\Delta_j(W'W)^{-1}\zeta$. By Lemma 6, we have

$$\sum_{j=1}^{m}\mathrm{E}\left\{(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})(\hat{\psi}_{-\mathrm{j}}-\hat{\psi})'\right\}=(W'W)^{-1}\sum_{j=1}^{m}\mathrm{E}(\zeta_\mathrm{j}\zeta_\mathrm{j}')(\mathrm{W'W})^{-1} \tag{A.38}$$
$$+\sum_{j=1}^{m}\mathrm{E}\left\{(\mathrm{r_{j1}}+\mathrm{r_{j2}}+\mathrm{r_{j3}})(\mathrm{r_{j1}}+\mathrm{r_{j2}}+\mathrm{r_{j3}})'\right\}$$
$$-(W'W)^{-1}\sum_{j=1}^{m}\mathrm{E}\left\{\zeta_\mathrm{j}(\mathrm{r_{j1}}+\mathrm{r_{j2}}+\mathrm{r_{j3}})'\right\}$$
$$=(W'W)^{-1}\sum_{j=1}^{m}\Sigma_j(W'W)^{-1}+O(m^{-2})J_2,$$

noting that

$$(W'W)^{-1}\sum_{j=1}^{m}\mathrm{E}(\zeta_j\zeta_j')(W'W)^{-1} = (W'W)^{-1}\sum_{j=1}^{m}\Sigma_j(W'W)^{-1}$$

$$\sum_{j=1}^{m}\mathrm{E}\left\{(r_{j1}+r_{j2}+r_{j3})(r_{j1}+r_{j2}+r_{j3})'\right\} = O(m^{-2})J_2,$$

$$(W'W)^{-1}\sum_{j=1}^{m}\mathrm{E}\left\{\zeta_j(r_{j1}+r_{j2}+r_{j3})'\right\} = O(m^{-2})J_2.$$

By the definition of f_{i1} in Equation (16) , we have $\frac{\partial^2 f_{i1}(\hat{\psi})}{\partial\psi\partial\psi'} = \frac{-2T_i}{(T_i\sigma_b^2+\sigma_\varepsilon^2)^3}\begin{pmatrix}(\sigma_\varepsilon^2)^2 & -\sigma_\varepsilon^2\sigma_b^2\\ -\sigma_\varepsilon^2\sigma_b^2 & (\sigma_b^2)^2\end{pmatrix}$. By the definitions of f_{i41} and f_{i42} in Equation (19), and Equations $(A.35)$ and $(A.38)$, Equation $(A.36)$ is produced.

Finally let us deal with Equation $(A.37)$. Similar to the proof of Equations $(A.11)$ and $(A.36)$, Equation $(A.37)$ can be derived. The proof is finished. ■

Received 9 April 2019
Accepted 8 January 2020