Statistica Sinica **26** (2016), 465-492
doi:http://dx.doi.org/10.5705/ss.202014.0070

# PREDICTION IN HETEROSCEDASTIC NESTED ERROR REGRESSION MODELS WITH RANDOM DISPERSIONS

Tatsuya Kubokawa[1], Shonosuke Sugasawa[1], Malay Ghosh[2] and Sanjay Chaudhuri[3]

[1] *University of Tokyo,* [2] *University of Florida and* [3] *National University of Singapore*

*Abstract:* The paper considers small-area estimation for a heteroscedastic nested error regression (HNER) model that assumes that the within-area variances are different among areas. Although HNER is useful for analyzing data where the within-area variation changes from area to area, it is difficult to provide good estimates for the error variances because of small sample sizes for small-areas. To address this issue, we suggest a random dispersion HNER model which assumes a prior distribution for the error variances. The resulting Bayes estimates of small area means provide stable shrinkage estimates even for areas with small sample sizes. Next we propose an empirical Bayes approach for estimating the small area means. For measuring uncertainty of the empirical Bayes estimators, we use the conditional and unconditional mean squared errors (MSE) and derive second-order correct approximations. It is interesting to note that the difference between the two MSEs appears in the first-order terms while the difference appears in the second-order terms for classical normal linear mixed models. Second-order unbiased estimators of the two MSEs are given with an application to posted land price data. Also, some simulation results are provided.

*Key words and phrases:* Asymptotic approximation, conditional mean squared error, empirical Bayes, parametric bootstrap, second-order approximation, second-order unbiased estimate, small area estimation.

## 1. Introduction

Small area estimation (SAE) using linear mixed models has been extensively studied in the literature from both theoretical and applied points of view. For a good review and account on this topic, see Ghosh and Rao (1994), Pfeffermann (2002), Rao (2003), and Datta (2009). Of these, the nested error regression (NER) model introduced by Battese, Harter and Fuller (1988) has been used in SAE. Suppose that there are $m$ small areas. For area $i$, $n_i$ individual data $(y_{i1}, \boldsymbol{x}_{i1}), \ldots, (y_{1n_i}, \boldsymbol{x}_{in_i})$ are observed, where $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ is a vector of auxiliary variables with $x_{ij1} = 1$. Then, the NER model is expressed as the mixed effects model

$$y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad i = 1, \ldots, m;\ j = 1, \ldots, n_i, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is an unknown vector of regression coefficients, the $v_i$'s and $\varepsilon_{ij}$'s are mutually independent random errors with $v_i \sim \mathcal{N}(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. However, Jiang and Nguyen (2012) illustrated that the within-area sample variances change dramatically from small-area to small-area for the data given in Battese, Harter and Fuller (1988). This motivated us to extend the NER model to cases with heteroscedastic variances.

As mixed models with heteroscedastic variances, two setups have been treated in the literature: One setup is the assumption that sampling errors are heteroscedastic, $Var(\varepsilon_{ij}) = \sigma_i^2$. In this case, the variance of $y_{ij}$ is decomposed as

$$Var(y_{ij}) = E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2] = \sigma_v^2 + \sigma_i^2, \tag{1.2}$$

for $i = 1, \ldots, m$. The other setup assumes that the variance of $y_{ij}$ is proportional to the heteroscedastic quantity $\sigma_{0,i}^2$, and that the ratio $Var(y_{ij})/\sigma_{0,i}^2$ is homoscedastic and is decomposed as $Var(y_{ij})/\sigma_{0,i}^2 = \kappa_v^2 + \kappa_e^2$, where $\kappa_v^2$ and $\kappa_e^2$ come from $v_i$ and $\varepsilon_{ij}$, respectively. Thus, the ratio of $y_{ij} - \boldsymbol{x}_i^T\boldsymbol{\beta}$ to the heteroscedastic quantity $\sigma_{0,i}$ can be modeled as the homoscedastic nested error structure

$$\frac{y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}}{\sigma_{0,i}} = v_{0,i} + \varepsilon_{0,ij},$$

where $v_{0,i} \sim \mathcal{N}(0, \kappa_v^2)$ and $\varepsilon_{0,ij} \sim \mathcal{N}(0, \kappa_e^2)$. Let $\sigma_i^2 = \kappa_e^2\sigma_{0,i}^2$ and $\lambda = \kappa_v^2/\kappa_e^2$. Then, the variance of $y_{ij}$ is expressed as

$$Var(y_{ij}) = E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2] = (\lambda + 1)\sigma_i^2. \tag{1.3}$$

The setups (1.2) and (1.3) were used by Maiti, Ren, and Sinha (2014) and Jiang and Nguyen (2012), respectively. The sampling errors have the heteroscedastic variance $\sigma_i^2$ in (1.2), while individual observations $y_{ij}$ have variances proportional to $\sigma_i^2$ in (1.3). Both models relax the homogeneity assumption of the usual NER model (1.1).

In mixed models with heteroscedastic variances, it is difficult to provide good estimates for $\sigma_i^2$ because of small samples sizes, $n_i$'s, for small-areas. To fix this difficulty, Maiti, Ren, and Sinha (2014) suggested that $\sigma_i^2$ has an inverse gamma distribution in the model (1.2). It is interesting to point out that the resulting empirical Bayes (EB) estimator of $\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta} + v_i$ for a known $p$-variate vector $\boldsymbol{c}_i$ shrinks both means and variances. Since the EB includes integration with respect to $\sigma_i^2$, however, the EB and its mean squared error (MSE) cannot be expressed in closed forms. Thus one needs heavy numerical computations to provide values of the EB and its MSE.

In this paper, we treat the heteroscedastic nested error regression (HNER) model with the variance structure (1.3) given in Jiang and Nguyen (2012). Assuming an inverse gamma distribution for $\sigma_i^2$, we suggest a random dispersion

HNER (RHNER) model. The resulting Bayes estimator of $\xi_i$ and the conditional variance of $\xi_i$ given data are expressed in closed forms, and the EB estimator of $\xi_i$ shrinks both means and variances. Also, the EB estimator of $\sigma_i^2$ provides stable shrinkage estimates even for $n_i - p = 0$.

For measuring uncertainty of the empirical Bayes estimator $\hat{\xi}_i^{EB}$ of $\xi_i$, we use the conditional and unconditional mean squared errors (MSE)

$$\begin{aligned} cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2|\boldsymbol{y}_i], \\ MSE(\omega; \hat{\xi}_i^{EB}) =& E[(\hat{\xi}_i^{EB} - \xi_i)^2], \end{aligned}$$

where $\omega$ is a vector of unknown parameters. When data of the small area of interest are observed as $\boldsymbol{y}_i$, and one wants to know the prediction error of the EB estimators based on these data, the conditional mean squared error (cMSE) given $\boldsymbol{y}_i$ is used instead of the conventional unconditional MSE. Booth and Hobert (1998) showed that the difference between the cMSE and MSE is quite small and appears in the second-order terms in classical normal linear mixed models. Here, however, we show that the difference appears in the leading or the first-order terms in the RHNER model.

The paper is organized as follows: A setup of the RHNER model and its motivation are given in Section 2. In Section 3, maximum likelihood (ML) estimators are provided for the unknown $\boldsymbol{\beta}$, $\lambda$, and hyper-parameters of the gamma distribution. The consistency of the ML estimators is shown and their asymptotic variances and covariances are derived through calculation of the Fisher information. In Section 4, we provide second-order approximations of the conditional and unconditional MSEs of the EB estimator for $\xi_i$; their second-order unbiased estimators are based on the parametric bootstrap method. In Section 5, we investigate the performance of the proposed procedures through simulation and empirical studies. Concluding remarks are given in Section 6 and the technical proofs are given in the Appendix.

## 2. HNER Models with Random Dispersions

### 2.1. Setup of models and predictors

We begin with the model given in (1.1) and (1.3). For stable estimators of the $\sigma_i^2$'s, we need a sufficient amount of data from each area. Since the $n_i$'s are typically small, $\sigma_i^2$ cannot usually be estimated with reasonable precision. To give more stable estimators for $\sigma_i^2$, we assume a prior distribution for $\sigma_i^2$. Let $\eta_i = 1/\sigma_i^2$. We assume that $\eta_1, \ldots, \eta_m$ are mutually independent and identically distributed with common pdf

$$\pi(\eta_i|\tau_1, \tau_2) \sim \mathcal{G}a(\frac{\tau_1}{2}, \frac{2}{\tau_2}), \tag{2.1}$$

a gamma distribution with mean $\tau_1/\tau_2$ and variance $2\tau_1/\tau_2^2$. Such a parametrization on $\tau_1$ and $\tau_2$ give us simple expression for the Bayes estimator (2.5) of $\sigma_i^2$ and simplifies also the subsequent calculations. Since $E[\sigma_i^2] = E[\eta_i^{-1}] = \tau_2/(\tau_1 - 2)$ and $Var[\sigma_i^2] = Var[\eta_i^{-1}] = 2\tau_2^2/\{(\tau_1 - 2)^2(\tau_1 - 4)\}$, the variance of $\sigma_i^2$ does not exist for $\tau_1 \leq 4$. The HNER model given in (1.1) and (1.3) with the random dispersion (2.1) is called a *Random Heteroscedastic Nested Error Regression* (RHNER) model.

Let $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_m^T)^T$, $\boldsymbol{X}_i^T = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,n_i})$ and $\boldsymbol{X} = (\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_m^T)$. All the unknown parameters are denoted by $\omega = (\boldsymbol{\beta}^T, \lambda, \boldsymbol{\tau})^T$ for $\boldsymbol{\tau} = (\tau_1, \tau_2)$. Then, the RHNER model is given by

$$\begin{aligned} \boldsymbol{y}_i|v_i, \eta_i \sim& \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + v_i\boldsymbol{j}_{n_i}, \eta_i^{-1}\boldsymbol{I}_{n_i}), \\ v_i|\eta_i \sim& \mathcal{N}(0, \lambda\eta_i^{-1}), \\ \eta_i \sim& \mathcal{G}a(\frac{\tau_1}{2}, \frac{2}{\tau_2}). \end{aligned} \tag{2.2}$$

The conditional distribution of $v_i$ given $\boldsymbol{y}_i$ and $\eta_i$ is $\mathcal{N}(\hat{v}_i, \lambda\eta_i^{-1}/(n_i\lambda+1))$, where

$$\hat{v}_i = \hat{v}_i(\boldsymbol{\beta}, \lambda) = \frac{n_i\lambda}{n_i\lambda + 1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}). \tag{2.3}$$

It is noted that $\hat{v}_i = E[v_i|\boldsymbol{y}_i]$ does not depend on $\eta_i$ or $\sigma_i^2$.

In this paper, we consider the problem of predicting the mixed quantity

$$\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta} + v_i, \quad i = 1, \ldots, m,$$

where $\boldsymbol{c}_i$ is a known $p$-variate vector. A typical example of $\boldsymbol{c}_i$ is the population mean of the covariates in the $i$-th area. The conditional expectation of $\xi_i$ given $\boldsymbol{y}_i$ and $\eta_i$ is

$$\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda) = E[\xi_i|\boldsymbol{y}_i, \sigma_i^2] = \boldsymbol{c}_i^T\boldsymbol{\beta} + \hat{v}_i(\boldsymbol{\beta}, \lambda).$$

This is interpreted as the Bayes estimator of $\xi_i$ under squared error loss. Since it does not depend on $\eta_i$, the estimator $\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)$ continues to be the conditional expectation of $\xi_i$ given $\boldsymbol{y}_i$ after integrating out the $\eta_i$, that is the Bayes estimator of $\xi_i$ is the same in the two situations. However, the empirical Bayes estimators, which substitute estimators of $\boldsymbol{\beta}$ and $\lambda$ into $\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)$, are different between the HNER and RHNER models.

In the HNER model, we need to estimate $(m + p + 1)$ parameters $\boldsymbol{\beta}$, $\lambda$, and $\sigma_1^2, \ldots, \sigma_m^2$. As the number of parameters increases as $m$ increases and the $n_i$'s are bounded in small-area estimation, we are faced with the problem of inconsistency and instability of the estimators of $\sigma_i^2$. In this situation, Jiang and Nguyen (2012) established the surprising result that the MLEs of $\boldsymbol{\beta}$ and $\lambda$ are consistent, which lead to the consistency of the EB estimator $\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$.

However, there are no consistent estimators of the $\sigma_i^2$. This problem can be fixed if instead the RHNER model is used. In fact, the parameters we need to estimate in the RHNER model are $\boldsymbol{\beta}$, $\lambda$, $\tau_1$, and $\tau_2$, and we can provide their consistent estimators.

### 2.2. A motivation from estimation of dispersions

We give more detailed motivation for considering random dispersions in the HNER model. We first treat the simple case that $\boldsymbol{\beta} = \mathbf{0}$ and $n_1 = \cdots = n_m = n$ in (1.1). Let $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_m^2)^T$ and $\gamma = 1/(1+n\lambda)$. It follows from the equation (4) of Jiang and Nguyen (2012) that the log-likelihood is then

$$L^H(\gamma, \boldsymbol{\sigma}^2) = \frac{1}{2}\sum_{i=1}^{m}\Big[-n\log\sigma_i^2 + \log\gamma - \frac{\{\sum_{j=1}^{n}(y_{ij}-\overline{y}_i)^2 + n\gamma\overline{y}_i^2\}}{\sigma_i^2}\Big] + K,$$

where $K$ is a generic constant. Differentiating $L^H(\gamma, \boldsymbol{\sigma}^2)$ with respect to $\gamma$ and the $\sigma_i^2$'s, we see that the maximum likelihood (ML) estimators, $\hat{\gamma}^H$ and $\hat{\sigma}_{(H)i}^2$, of $\gamma$ and the $\sigma_i^2$'s are solutions of the equations

$$\begin{aligned} \hat{\gamma}^H &= \frac{m}{\sum_{i=1}^{m} n\overline{y}_i^2/\hat{\sigma}_{(H)i}^2}, \\ \hat{\sigma}_{(H)i}^2 &= \frac{1}{n}\Big\{\sum_{j=1}^{n}(y_{ij}-\overline{y}_i)^2 + n\hat{\gamma}^H\overline{y}_i^2\Big\}. \end{aligned} \tag{2.4}$$

Here $\hat{\sigma}_{(H)i}^2$ is not consistent when $m \to \infty$, but $n$ is bounded. Thus, we need to modify $\hat{\sigma}_{(H)i}^2$ when $n$ is small. For example, we look at the empirical Bayes estimator of $\xi_i$. In the simple case we treat here, we have $\xi_i = v_i$, and the EB estimator of $\xi_i$ is

$$\hat{\xi}_i^H = (1-\hat{\gamma}^H)\overline{y}_i = \Big\{1 - \frac{m}{\sum_{i=1}^{m} n\overline{y}_i^2/\hat{\sigma}_{(H)i}^2}\Big\}\overline{y}_i,$$

from (2.4). This is a natural shrinkage estimator, and it is reasonable for large $m$ since $\hat{\gamma}^H$ is consistent. When $m$ is not large, however, there is a concern about the precision of the estimator $\hat{\sigma}_{(H)i}^2$. Since $\sum_{j=1}^{n}(y_{ij}-\overline{y}_i)^2 \le n\hat{\sigma}_{(H)i}^2 \le \sum_{j=1}^{n} y_{ij}^2$, it is seen that

$$\frac{\overline{y}_i^2}{\sum_{j=1}^{n} y_{ij}^2/n} \le \frac{\overline{y}_i^2}{\hat{\sigma}_{(H)i}^2} \le \frac{\overline{y}_i^2}{T_i/n},$$

for $T_i = \sum_{j=1}^{n}(y_{ij}-\overline{y}_i)^2$. When $n$ is small, clearly $\overline{y}_i^2/\hat{\sigma}_{(H)i}^2$ has a large variation, which leads to the instability of the empirical Bayes estimator $\hat{\xi}_i^H$. Although the simple case of equal replications $n$ is considered here, in the survey data we need to handle involve small sample sizes $n_i$'s for some small-areas, and resulting estimators of $\hat{\sigma}^2_{(H)i}$'s lack stability owing to small degrees of freedom.

To overcome this drawback, we need to stabilize $\hat{\sigma}^2_{(H)i}$ via shrinkage. The random dispersion model is helpful for this purpose. Since $T_i\eta_i = T_i/\sigma_i^2 \sim \chi^2_{n-1}$, from the joint distribution of $(T_i, \eta_i)$, the posterior mean of $\sigma_i^2$ given $T_i$ is

$$E[\sigma_i^2|T_i] = \frac{T_i + \tau_2}{n - 1 + \tau_1}. \tag{2.5}$$

When $\hat{\tau}_1$ and $\hat{\tau}_2$ are estimators of $\tau_1$ and $\tau_2$ based on the statistics $T_1, \ldots, T_m$, it is reasonable to estimate $\sigma_i^2$ by

$$\hat{\sigma}^2_{(RH)i} = \frac{T_i + \hat{\tau}_2}{n - 1 + \hat{\tau}_1}.$$

Clearly, $\hat{\sigma}^2_{(RH)i}$ is more stable than the unbiased estimator $T_i/(n-1)$ when $n$ is small. Replacing $\hat{\sigma}^2_{(H)i}$ in $\hat{\xi}_i^H$ with a shrinkage estimator like $\hat{\sigma}^2_{(RH)i}$, one can get the more stabilized empirical Bayes estimator

$$\hat{\xi}_i^{RH} = \Big\{1 - \frac{m}{\sum_{i=1}^m n\overline{y}_i^2/\hat{\sigma}^2_{(RH)i}}\Big\}\overline{y}_i.$$

Another need for a consistent estimator of $\sigma_i^2$ appears in evaluation of uncertainty of the empirical Bayes estimator $\hat{\xi}_i^H$. When the mean squared error is used for measuring the uncertainty, the MSE of $\hat{\xi}_i^H$, denoted by $E[(\hat{\xi}_i^H - \xi_i)^2]$ converges to

$$E[\mathrm{Var}(v_i|\boldsymbol{y}_i)] = \frac{\sigma_i^2\lambda}{(1 + n\lambda)} = \frac{\sigma_i^2(1-\gamma)}{n}$$

for large $m$. In order to estimate the uncertainty of $\hat{\xi}_i^H$, we want to estimate the leading term of the MSE consistently. Since $\hat{\sigma}^2_{(H)i}$ is not consistent, however, we cannot provide any consistent estimator of the leading term in the MSE of $\hat{\xi}_i^H$ in the HNER model. This drawback is overcome in the RHNER model.

## 3. Predictors and Asymptotic Properties of MSE

### 3.1. MLE of parameters and the empirical Bayes estimator

Consider the RHNER model given in (2.2). When $\lambda$ and $\boldsymbol{\beta}$ are known, the best predictor or the Bayes estimator of $\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta} + v_i$ is given by

$$\begin{aligned} \hat{\xi}_i^B &= \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda) = E[\xi_i|\boldsymbol{y}_i] \\ &= \boldsymbol{c}_i^T\boldsymbol{\beta} + (1-\gamma_i)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}), \end{aligned} \tag{3.1}$$

where $\gamma_i = \gamma_i(\lambda) = 1/(n_i\lambda+1)$. In our case since $\lambda$ and $\boldsymbol{\beta}$ are unknown, we need to estimate them from the marginal distributions of the $\boldsymbol{y}_i$. We provide maximum likelihood (ML) estimators for unknown parameters $\omega = (\boldsymbol{\beta}^T, \lambda, \tau_1, \tau_2)^T$.

The marginal likelihood of $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)^T$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^T$ after integrating out the full joint likelihood with respect to $v_i$'s can be expressed as

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{\eta}|\omega) &= \prod_{i=1}^m \Big\{ \frac{\eta_i^{n_i/2}}{(2\pi)^{n_i/2}\sqrt{n_i\lambda+1}} \exp\Big[-\frac{\eta_i}{2}\Big\{\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2 \\ &\quad - \frac{n_i^2\lambda}{n_i\lambda+1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2\Big\}\Big]\pi(\eta_i|\tau_1,\tau_2)\Big\} \\ &= \prod_{i=1}^m \Big\{ \frac{\tau_2^{\tau_1/2}\eta_i^{(n_i+\tau_1)/2-1}2^{-(n_i+\tau_1)/2}}{\pi^{n_i/2}\Gamma(\tau_1/2)\sqrt{n_i\lambda+1}} \exp\Big[-\frac{\eta_i}{2}\{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2\}\Big]\Big\}, \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} Q_i = Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) &= \sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2 - \frac{n_i^2\lambda}{n_i\lambda+1}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 \\ &= \sum_{j=1}^{n_i}\{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}^2 + n_i\gamma_i(\lambda)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2, \end{aligned} \tag{3.3}$$

where $\gamma_i = \gamma_i(\lambda) = 1/(n_i\lambda + 1)$. Integrating out the joint density $f(\boldsymbol{y}, \boldsymbol{\eta}|\omega)$ in (3.2) with respect to $\boldsymbol{\eta}$ yields the marginal likelihood of $\boldsymbol{y}$ given by

$$f(\boldsymbol{y}|\omega) = \prod_{i=1}^m \Big\{ \frac{\tau_2^{\tau_1/2}\Gamma((n_i+\tau_1)/2)}{\pi^{n_i/2}\sqrt{n_i\lambda+1}\Gamma(\tau_1/2)}\{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2\}^{-(n_i+\tau_1)/2}\Big\}. \tag{3.4}$$

Let $L = L(\boldsymbol{\beta}, \lambda, \tau_1, \tau_2) = \log f(\boldsymbol{y}|\omega)$. Then,

$$\begin{aligned} 2L = &-\sum_{i=1}^m n_i \log \pi + m\tau_1 \log \tau_2 + 2\sum_{i=1}^m \log\{\Gamma(\frac{n_i+\tau_1}{2})\} - 2m\log\{\Gamma(\frac{\tau_1}{2})\} \\ &-\sum_{i=1}^m \log(n_i\lambda+1) - \sum_{i=1}^m (n_i+\tau_1)\log(Q_i+\tau_2). \end{aligned}$$

Let $L_{\boldsymbol{\beta}}$, $L_\lambda$, $L_{\tau_1}$ and $L_{\tau_2}$ be the derivatives of $L$ with respect to $\boldsymbol{\beta}$, $\lambda$, $\tau_1$, and $\tau_2$. Then,

$$\begin{aligned} 2L_{\boldsymbol{\beta}} &= -\sum_{i=1}^m \frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial Q_i}{\partial \boldsymbol{\beta}}, \\ 2L_\lambda &= -\sum_{i=1}^m \frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial Q_i}{\partial \lambda} - \sum_{i=1}^m n_i\gamma_i, \\ 2L_{\tau_1} &= \sum_{i=1}^m \log\Big(\frac{\tau_2}{Q_i+\tau_2}\Big) + \sum_{i=1}^m \Big\{\psi(\frac{n_i+\tau_1}{2}) - \psi(\frac{\tau_1}{2})\Big\}, \end{aligned} \tag{3.5}$$

$$2L_{\tau_2} = m\frac{\tau_1}{\tau_2} - \sum_{i=1}^{m} \frac{n_i + \tau_1}{Q_i + \tau_2},$$

where $\psi(a)$ is the digamma function defined by $\psi(a) = \Gamma'(a)/\Gamma(a)$, $\partial Q_i/\partial\lambda = -n_i^2\gamma_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2$ for $\partial\gamma_i/\partial\lambda = -n_i\gamma_i^2$, and

$$\frac{\partial Q_i}{\partial \boldsymbol{\beta}} = -2\sum_{j=1}^{n_i}\{(y_{ij} - \overline{y}_i) - (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i) - 2n_i\gamma_i(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i. \quad (3.6)$$

The MLEs of $\boldsymbol{\beta}$, $\lambda$, $\tau_1$, and $\tau_2$ are solutions of the simultaneous equations $L_{\boldsymbol{\beta}} = 0$, $L_\lambda = 0$, $L_{\tau_1} = 0$ and $L_{\tau_2} = 0$. The MLEs are denoted by $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$, and $\hat{\tau}_2$. The empirical Bayes estimator of $\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta} + v_i$ is provided by

$$\hat{\xi}_i^{EB} = \hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) = \boldsymbol{c}_i^T\widehat{\boldsymbol{\beta}} + (1 - \hat{\gamma}_i)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\widehat{\boldsymbol{\beta}}), \quad (3.7)$$

where $\hat{\gamma}_i = \gamma_i(\hat{\lambda}) = 1/(n_i\hat{\lambda} + 1)$.

### 3.2. Asymptotic properties of MLE

To evaluate the mean squared errors of the empirical Bayes estimator $\hat{\xi}_i^{EB}$ asymptotically, we need to derive asymptotic variances and covariances of the MLE when $m$ tends to infinity. To derive asymptotic properties of the MLE, we assume the following.

(A1) The sample sizes $n_i$'s are bounded below and above as $\underline{n} \le n_i \le \overline{n}$ for constants $\underline{n}$ and $\overline{n}$. The elements of $\boldsymbol{X}$ are uniformly bounded, $\boldsymbol{X}^T\boldsymbol{X}$ is positive definite, and $\boldsymbol{X}^T\boldsymbol{X}/m$ converges to a positive definite matrix.

Since asymptotic variances and covariances of MLE's of $\boldsymbol{\beta}$, $\lambda$, $\tau_1$, and $\tau_2$ are derived from the Fisher information matrix, we begin with the derivation of $\boldsymbol{I}_{\boldsymbol{\beta},\boldsymbol{\theta}}$, the Fisher information matrix of $\boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\lambda, \tau_1, \tau_2)^T$. The Fisher information matrix and the inverse of $\boldsymbol{\theta} = (\lambda, \tau_1, \tau_2)^T$ are denoted by

$$\boldsymbol{I}_{\boldsymbol{\theta\theta}} = \begin{pmatrix} I_{\lambda\lambda} & I_{\lambda\tau_1} & I_{\lambda\tau_2} \\ I_{\lambda\tau_1} & I_{\tau_1\tau_1} & I_{\tau_1\tau_2} \\ I_{\lambda\tau_2} & I_{\tau_1\tau_2} & I_{\tau_2\tau_2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{I}_{\boldsymbol{\theta\theta}}^{-1} = \begin{pmatrix} I^{\lambda\lambda} & I^{\lambda\tau_1} & I^{\lambda\tau_2} \\ I^{\lambda\tau_1} & I^{\tau_1\tau_1} & I^{\tau_1\tau_2} \\ I^{\lambda\tau_2} & I^{\tau_1\tau_2} & I^{\tau_2\tau_2} \end{pmatrix}.$$

Then, exact expression of the Fisher information matrix is given in the following theorem. The proof is deferred to the Appendix.

**Theorem 1.** *The Fisher information of* $\boldsymbol{\beta}$ *is given by*

$$\boldsymbol{I}_{\boldsymbol{\beta\beta}} = \frac{\tau_1}{\tau_2}\sum_{i=1}^{m}\frac{n_i + \tau_1}{n_i + \tau_1 + 2}\Big\{\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)^T + n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T\Big\}.$$

*Also,* $\boldsymbol{I}_{\boldsymbol{\beta}\lambda} = \boldsymbol{0}$, $\boldsymbol{I}_{\boldsymbol{\beta}\tau_1} = \boldsymbol{0}$, *and* $\boldsymbol{I}_{\boldsymbol{\beta}\tau_2} = \boldsymbol{0}$. *The elements of* $2\boldsymbol{I}_{\boldsymbol{\theta\theta}}$ *are*

$$2I_{\lambda\lambda} =\sum_{i=1}^{m} \frac{(n_i+\tau_1-1)n_i^2\gamma_i^2}{n_i+\tau_1+2}, \qquad 2I_{\lambda\tau_1} = -\sum_{i=1}^{m} \frac{n_i\gamma_i}{n_i+\tau_1},$$
$$2I_{\lambda\tau_2} =\frac{\tau_1}{\tau_2}\sum_{i=1}^{m} \frac{n_i\gamma_i}{n_i+\tau_1+2}, \qquad 2I_{\tau_1\tau_1} = \frac{1}{2}\sum_{i=1}^{m}\{\psi'(\frac{\tau_1}{2})-\psi'(\frac{n_i+\tau_1}{2})\},$$
$$2I_{\tau_1\tau_2} =-\frac{1}{\tau_2}\sum_{i=1}^{m} \frac{n_i}{n_i+\tau_1}, \qquad 2I_{\tau_2\tau_2} = \frac{\tau_1}{\tau_2^2}\sum_{i=1}^{m} \frac{n_i}{n_i+\tau_1+2}.$$

It follows from Theorem 1 and assumption (A1) that $m^{-1}\boldsymbol{I}_{\boldsymbol{\beta\beta}} = O(1)$ and $m^{-1}\boldsymbol{I}_{\boldsymbol{\theta\theta}} = O(1)$, and the limiting values of these quantities are away from zero. The proof of the following is given in the Appendix.

**Theorem 2.** *Assume* (A1). *Then, for* $\widehat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\tau}_1, \hat{\tau}_2)^T$,

$$\begin{aligned} E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T|\boldsymbol{y}_i] =&(\boldsymbol{I}_{\boldsymbol{\beta\beta}})^{-1}+O_p(m^{-3/2}),\\ E[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T|\boldsymbol{y}_i] =&(\boldsymbol{I}_{\boldsymbol{\theta\theta}})^{-1}+O_p(m^{-3/2}),\\ E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T|\boldsymbol{y}_i] =&O_p(m^{-3/2}). \end{aligned} \tag{3.8}$$

*Thus* $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i = O_p(m^{-1/2})$ *and* $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}|\boldsymbol{y}_i = O_p(m^{-1/2})$. *The conditional biases satisfy* $E[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i] = O(m^{-1})$ *and* $E[\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}|\boldsymbol{y}_i] = O(m^{-1})$.

## 4. Measures of Uncertainty of the Empirical Bayes Estimator

### 4.1. Second-order approximations of the conditional and unconditional MSEs

We derive a second-order approximation of the MSE of the empirical Bayes (EB) estimator and its second-order unbiased estimator. We want to predict $\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta}+v_i$ with EB $\hat{\xi}_i^{EB} = \hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) = \boldsymbol{c}_i^T\widehat{\boldsymbol{\beta}}+\hat{v}_i(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$. For measuring uncertainty of EB, we use the conditional and unconditional mean squared errors (MSE) defined by

$$\begin{aligned} cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =&E[(\hat{\xi}_i^{EB}-\xi_i)^2|\boldsymbol{y}_i],\\ MSE(\omega; \hat{\xi}_i^{EB}) =&E[(\hat{\xi}_i^{EB}-\xi_i)^2]. \end{aligned}$$

The conditional and unconditional MSEs can be decomposed as

$$\begin{aligned} cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) =&E[\{\xi_i-\hat{\xi}_i^B(\boldsymbol{\beta},\lambda)\}^2|\boldsymbol{y}_i]+E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}},\hat{\lambda})-\hat{\xi}_i^B(\boldsymbol{\beta},\lambda)\}^2|\boldsymbol{y}_i]\\ =&g_1^c(\omega|\boldsymbol{y}_i)+g_2^c(\omega|\boldsymbol{y}_i), \quad \text{(say)} \qquad (4.1)\\ MSE(\omega; \hat{\xi}_i^{EB}) =&E[\{\xi_i-\hat{\xi}_i^B(\boldsymbol{\beta},\lambda)\}^2]+E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}},\hat{\lambda})-\hat{\xi}_i^B(\boldsymbol{\beta},\lambda)\}^2]\\ =&g_1(\omega)+g_2(\omega). \quad \text{(say)} \qquad (4.2) \end{aligned}$$

The first term $g_1^c(\omega|\boldsymbol{y}_i)$ is the posterior variance of $\xi_i$ given $\boldsymbol{y}_i$,

$$g_1^c(\omega|\boldsymbol{y}_i) = Var(\xi_i|\boldsymbol{y}_i) = \frac{\lambda}{n_i\lambda+1}E[\eta_i^{-1}|\boldsymbol{y}_i] = \frac{\lambda}{n_i\lambda+1}\frac{Q_i+\tau_2}{n_i+\tau_1-2}, \tag{4.3}$$

where $Q_i$ is given in (3.3). Similarly, $g_1(\omega)$ is given by

$$g_1(\omega) = E[Var(\xi_i|\boldsymbol{y}_i)] = \frac{\lambda}{n_i\lambda+1}E[\eta_i^{-1}] = \frac{\lambda}{n_i\lambda+1}\frac{\tau_2}{\tau_1-2}. \tag{4.4}$$

Noting that $g_1^c(\omega|\boldsymbol{y}_i) = O_p(1)$, $g_2^c(\omega|\boldsymbol{y}_i) = O_p(m^{-1})$, $g_1(\omega) = O(1)$, and $g_2(\omega) = O(m^{-1})$, we can see that the difference between the cMSE and MSE appears in the leading or the first-order terms. This is an interesting fact, because the difference is small and appears in the second-order terms in the classical normal theory mixed models, as demonstrated by Booth and Hobert (1998). They also showed that the difference is significant and appears in the first-order terms for distributions far from normality. This is consistent in that the random dispersion model (2.2) is not a normal distribution, but close to a $t$-distribution.

In the case of the HNER model, $Var(\xi_i|\boldsymbol{y}_i)$ is identical to $E[Var(\xi_i|\boldsymbol{y}_i)]$ since $y_i$ has a normal distribution, and is given by $\sigma_i^2\lambda/(n_i\lambda+1)$. Thus, we cannot estimate the first-order term $\sigma_i^2\lambda/(n_i\lambda+1)$ consistently in the HNER model, since $n_i$ is bounded. However, we can estimate $g_1^c(\omega|\boldsymbol{y}_i)$ and $g_1(\omega)$ consistently in the RHNER model (2.2) since $\lambda$, $\tau_1$ and $\tau_2$ are estimated consistently.

**Theorem 3.** *Under* (A1)*, the conditional MSE of* $\hat{\xi}_i^{EB}$ *is approximated as*

$$\begin{aligned} cMSE(\lambda,\boldsymbol{\tau};\hat{\xi}_i^{EB}|\boldsymbol{y}_i) =& \frac{1-\gamma_i}{n_i}\frac{Q_i+\tau_2}{n_i+\tau_1-2} + \gamma_i^2\boldsymbol{c}_i^T\left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1}\boldsymbol{c}_i \\ &+ n_i^2\gamma_i^4(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 I^{\lambda\lambda} + O_p(m^{-3/2}), \end{aligned} \tag{4.5}$$

*for* $\gamma_i = 1/(n_i\lambda+1)$*, and the unconditional MSE is approximated as*

$$\begin{aligned} MSE(\lambda,\boldsymbol{\tau};\hat{\xi}_i^{EB}) =& \frac{1-\gamma_i}{n_i}\frac{\tau_2}{\tau_1-2} + \gamma_i^2\boldsymbol{c}_i^T\left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1}\boldsymbol{c}_i \\ &+ n_i\gamma_i^3\frac{\tau_2}{\tau_1-2}I^{\lambda\lambda} + O(m^{-3/2}). \end{aligned} \tag{4.6}$$

### 4.2. Second-order unbiased estimators of the conditional and unconditional MSEs

We derive second-order unbiased estimators of the unconditional and conditional MSEs. Since it is hard to derive second-order biases of the MLEs of $\boldsymbol{\beta}$, $\lambda$, $\tau_1$, and $\tau_2$, we cannot provide analytical second-order unbiased estimators of the MSEs. Instead, we use parametric bootstrap methods, which provide second-order unbiased MSE estimators.

We begin by treating the unconditional case. The parametric bootstrap sample in this case is denoted as

$$y_{ij}^* = \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}} + v_i^* + \varepsilon_{ij}^*, \quad i = 1, \ldots, m;\ j = 1, \ldots, n_i, \tag{4.7}$$

where $v_i^*$'s and $\varepsilon_{ij}^*$'s are conditionally mutually independent given $\eta_i^*$'s and

$$\begin{aligned} v_i^* | \eta_i^* &\sim \mathcal{N}(0, \frac{\hat{\lambda}}{\eta_i^*}), \\ \varepsilon_{ij}^* | \eta_i^* &\sim \mathcal{N}(0, \frac{1}{\eta_i^*}), \\ \eta_i^* &\sim \sim \mathcal{G}a(\frac{\hat{\tau}_1}{2}, \frac{2}{\hat{\tau}_2}). \end{aligned} \tag{4.8}$$

The estimator of the unconditional MSE, $MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB})$, is given by

$$mse^*(\hat{\xi}_i^{EB}) = \hat{g}_1^* + \hat{g}_2^*,$$

where

$$\begin{aligned} \hat{g}_1^* &= 2g_1(\hat{\lambda}, \widehat{\boldsymbol{\tau}}) - E_*[g_1(\hat{\lambda}^*, \widehat{\boldsymbol{\tau}}^*)], \\ \hat{g}_2^* &= \hat{\gamma}_i^2 E^*[\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\}^2] + n_i \hat{\gamma}_i^3 \frac{\hat{\tau}_2}{\hat{\tau}_1 - 2} E^*[(\hat{\lambda}^* - \hat{\lambda})^2]. \end{aligned}$$

Then, we can show that $mse^*(\hat{\xi}_i^{EB})$ is a second-order unbiased estimator of $MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB})$. The proof is given in the Appendix.

**Proposition 1.** *If* (A1) *holds, then* $E[mse^*(\hat{\xi}_i^{EB})] = MSE(\lambda, \boldsymbol{\tau}; \hat{\xi}_i^{EB}) + O(m^{-3/2})$.

We next consider the conditional case. Keeping $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ fixed, a bootstrap sample $\boldsymbol{y}_k^* = (y_{k1}^*, \ldots, y_{kn_k}^*)^T$ is generated from (4.7) for $k \neq i$. As $\boldsymbol{y}_i$ is fixed, we construct the estimators $\widehat{\boldsymbol{\beta}}_{(i)}^*$, $\hat{\lambda}_{(i)}^*$, $\hat{\tau}_{1(i)}^*$, and $\hat{\tau}_{2(i)}^*$ from $\boldsymbol{y}_i$ and the bootstrap sample

$$\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_{i-1}^*, \boldsymbol{y}_i, \boldsymbol{y}_{i+1}^*, \ldots, \boldsymbol{y}_m^* \tag{4.9}$$

with the same technique as used to obtain the estimators $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$, and $\hat{\tau}_2$. Let $E_*[\cdot|\boldsymbol{y}_i]$ be the expectation with regard to the bootstrap sample (4.9). The conditional MSE is given by $cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) = g_1^c(\omega|\boldsymbol{y}_i) + g_2^c(\omega|\boldsymbol{y}_i)$, where $g_1^c(\omega|\boldsymbol{y}_i) = E[\{\xi_i - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i]$ and $g_2^c(\omega|\boldsymbol{y}_i) = E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)\}^2|\boldsymbol{y}_i]$ from (4.1). Since $g_1^c(\omega|\boldsymbol{y}_i) = n_i^{-1}(1 - \gamma_i(\lambda))(Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \tau_2)/(n_i + \tau_1 - 2)$ from (4.3), a second-order unbiased estimator of $g_1^c(\omega|\boldsymbol{y}_i)$ is given by

$$\hat{g}_1^{c*} = 2g_1^c(\widehat{\boldsymbol{\beta}}, \hat{\lambda}, \widehat{\boldsymbol{\tau}}|\boldsymbol{y}_i) - E_*\left[g_1^c(\widehat{\boldsymbol{\beta}}_{(i)}^*, \hat{\lambda}_{(i)}^*, \widehat{\boldsymbol{\tau}}_{(i)}^*|\boldsymbol{y}_i)|\boldsymbol{y}_i\right].$$

Then, it can be verified that $E[\hat{g}_1^{c*}|\boldsymbol{y}_i] = g_1^c(\omega|\boldsymbol{y}_i) + O_p(m^{-3/2})$. Also, $g_2^c(\omega|\boldsymbol{y}_i)$ is estimated via parametric bootstrap as

$$\hat{g}_2^{c*} = E^*\left[\{\hat{\xi}_i^{B*}(\widehat{\boldsymbol{\beta}}_{(i)}^*, \hat{\lambda}_{(i)}^*) - \hat{\xi}_i^{B*}(\widehat{\boldsymbol{\beta}}, \hat{\lambda})\}^2|\boldsymbol{y}_i\right],$$

for $\hat{\xi}_i^{B*}(\boldsymbol{\beta},\lambda) = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + (1-\gamma_i)(\overline{y}_i^* - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})$. Thus,

$$cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i) = \hat{g}_1^{c*} + \hat{g}_2^{c*}, \tag{4.10}$$

which is a second-order unbiased estimator of $cMSE(\omega;\hat{\xi}_i^{EB}|\boldsymbol{y}_i)$. The proof is given in the Appendix.

**Theorem 4.** *If* (A1) *holds, then* $E[cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i)|\boldsymbol{y}_i] = cMSE(\omega;\hat{\xi}_i^{EB}|\boldsymbol{y}_i) + O_p(m^{-3/2})$.

## 5. Simulation and Data Analysis

In this section, we investigate performances of the procedures suggested in the previous sections through simulation and data analysis.

### 5.1. Simulation study

Here we investigate finite sample performances of the ML estimators in the RHNER model and the second-order unbiased estimators for the conditional and unconditional MSEs by Monte Carlo simulation.

The ML estimators $\widehat{\boldsymbol{\beta}}$ and $\hat{\lambda}$ as given in (3.5) based on the RHNER model, as well as the estimators given by Jiang and Nguyen (2012) in HNER, are consistent for large $m$. As discussed in Section 2.2, however, it is expected that the estimators (3.5) still perform well for smaller $m$. Thus, for $m = 10, 20$ and 30, we examined finite sample performances of the estimators (3.5) in RHNER and compared them with the estimators in HNER in light of the mean squared errors (MSE). To this end, we conducted simulation experiments via the simple regression model given by $y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}$ for $j = 1, 2, 3$ and $i = 1, \ldots, m$, where the $x_{ij}$'s were generated from $\mathcal{N}(0, 1)$, and these values were fixed throughout the simulation runs. In this simulation, the true values of $\boldsymbol{\beta}$ and $\lambda$ were $\beta_0 = \beta_1 = \lambda = 1$. For $(\tau_1, \tau_2)$, we treated two cases: (Case I) $(\tau_1, \tau_2) = (8, 4)$ and (Case II) $(\tau_1, \tau_2) = (3, 1/4)$. The values of (mean, variance) of $\eta_i$ were (2, 1) for Case I and (12, 96) for Case II. The variance of $\sigma_i^2$ does not exist for Case II from the note below (2.1). Thus, generated values of $\eta_i$ or $\sigma_i^2$ in Case II are more variable than those in Case I.

We numerically computed values of MSE of the estimators for $(\beta_0, \beta_1, \lambda)$ with

$$\text{MSE} = \frac{1}{K}\sum_{i=1}^{K}(\text{estimate} - \text{true parameter})^2$$

for $K = 1,000$. The square roots of the MSEs of the ML estimator in RHNER are reported in Table 1, where percentages of improvement over the estimators given by Jiang and Nguyen (2012) in HNER are given in the parentheses. It is

Table 1. Square roots of mean squared errors for the ML estimators of $(\beta_0, \beta_1, \lambda, \tau_1, \tau_2)$ in the RHNER model for (Case I) $(\tau_1, \tau_2) = (8, 4)$ and (Case II) $(\tau_1, \tau_2) = (3, 1/4)$. (Values for $\beta_0$ and $\beta_1$ are multiplied by 100. Values in the parentheses denote percentages of improvement over the estimators in the HNER model.)

| Case | Size | $\beta_0$ | $\beta_1$ | $\lambda$ | $\tau_1$ | $\tau_2$ |
|---|---|---|---|---|---|---|
| | $m = 10$ | 29.7 (50.1) | 14.8 (57.6) | 9.3 (18.8) | 69.0 | 43.0 |
| I | $m = 20$ | 21.0 (59.5) | 13.0 (61.7) | 6.4 (21.9) | 55.3 | 33.9 |
| | $m = 30$ | 16.2 (63.6) | 9.3 (63.6) | 5.0 (18.8) | 45.2 | 27.0 |
| | $m = 10$ | 12.4 (48.3) | 8.3 (40.4) | 10.0 (10.1) | 49.5 | 7.0 |
| II | $m = 20$ | 8.8 (56.4) | 5.3 (57.0) | 6.6 (18.9) | 22.1 | 2.8 |
| | $m = 30$ | 7.2 (54.8) | 5.0 (56.8) | 5.3 (21.4) | 13.0 | 1.7 |

observed from Table 1 that the square roots of the MSEs decrease as $m$ increases. This is coincident with the consistency of the estimators in (3.5). Also, the values given in the parentheses illustrate that the estimators in RHNER improve on the estimators in HNER; This seems to be due to the property that the estimates of the variances $\sigma_i^2$ are more stable in RHNER than in HNER. Concerning the difference between Cases I and II, the estimates of $\tau_1$ and $\tau_2$ affect the estimates of $\beta_0$ and $\beta_1$, but do not affect the estimate for $\lambda$ very much. The MSEs of $\tau_1$ and $\tau_2$ seem very large, but the estimators of both conditional and unconditional MSEs give stable estimates as shown in the subsequent simulation studies.

We next investigated finite sample performance for the estimators of conditional and unconditional MSEs as suggested in Section 4.2. For simplicity, we treated the model without covariates given as $y_{ij} = \mu + v_i + \varepsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, m$, for $m = 20$ and 50, where the true values of the unknown parameters were $\mu = 0$, $\lambda = 1$, $\tau_1 = 8$, and $\tau_2 = 4$; the true values of $(\tau_1, \tau_2)$ correspond to Case I in the previous simulation. For the design of $n_i$, we took $n_{1+m(k-1)/5}, \ldots, n_{mk/5} = k$, $k = 1, \ldots, 5$, which means that $m$ areas are divided into five groups and that areas in each group have the same sample size $n_i$.

Concerning the unconditional MSEs, their true values were calculated via simulation with $R = 5,000$ replications as

$$\mathrm{MSE}_i = \frac{\lambda}{n_i\lambda + 1}\frac{\tau_2}{\tau_1 - 2} + \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\xi}_i^{EB(r)} - \hat{\xi}_i^{(B)}\right)^2,$$

where $\hat{\xi}_i^{EB(r)}$ and $\hat{\xi}_i^{B(r)}$ are the EB and Bayes estimators of $\xi_i$ in the $r$-th replication for $r = 1, \ldots, R$. Then, the mean values of the estimator for the MSE and their Percentage Relative Bias (RB) were calculated based on $T = 1,000$ simulation runs with each 100 bootstrap samples, where RB is defined as

$$\mathrm{RB}_i = 100\frac{T^{-1}\sum_{t=1}^{T}\widehat{\mathrm{MSE}}_i^{(t)} - \mathrm{MSE}_i}{\mathrm{MSE}_i},$$

Table 2. Mean and relative bias of the estimators for the unconditional MSE.

| | $m=20$ | | | $m=50$ | | |
|---|---|---|---|---|---|---|
| $n_i$ | MSE | $E[\widehat{\text{MSE}}]$ | RB(%) | MSE | $E[\widehat{\text{MSE}}]$ | RB(%) |
| 1 | 0.377 | 0.346 | -8.12 | 0.350 | 0.337 | -3.71 |
| 2 | 0.249 | 0.228 | -8.73 | 0.231 | 0.224 | -3.27 |
| 3 | 0.183 | 0.169 | -7.82 | 0.173 | 0.167 | -3.04 |
| 4 | 0.145 | 0.134 | -7.07 | 0.137 | 0.133 | -2.86 |
| 5 | 0.119 | 0.111 | -6.33 | 0.114 | 0.111 | -2.74 |

for the MSE estimate $\widehat{\text{MSE}}_t$ in the $t$-th replication for $t = 1, \ldots, T$. For the five groups, Table 2 reports the average values over each group for the MSE estimates and their relative biases. It is observed that the MSE estimates $\widehat{\text{MSE}}$ are close to the true values of MSE, and that their relative biases are small for both $m = 20$ and 50. Although Table 3 (p.599) in Jiang and Nguyen (2012) indicates that the MSE estimates in HNER are not so accurate when $m = 20$, the MSE estimates in RHNER seem appropriate even for $m = 20$. It seems that this comes from stability of estimators of the variances for each small area in RHNER.

Concerning the conditional MSE, we used the same setup as in the simulation of the unconditional MSE except that we took $n_i = 3$ for each small area. Without any loss of generality, it is assumed that values of $y_{1j}$'s in the area 1 are given. As conditioning values for $y_{1j}$'s, we used $\alpha$-quantile points of the marginal distribution of $y_{1j}$, denoted by $y_{1j(\alpha)}$, and select the five quantiles for $\alpha = 0.05$, 0.25, 0.5, 0.75 and 0.95. In the $r$-th iteration, from the sample $\{y_{11(\alpha)}, y_{12(\alpha)}, y_{13(\alpha)}, y_{21}, y_{22}, y_{23}, \ldots, y_{m1}, y_{m2}, y_{m3}\}$, we calculated the values of $\widehat{\xi}_1^{EB(r)}$ and $\widehat{\xi}_1^{B(r)}$. Then, the true values of the conditional MSE of $\widehat{\xi}_1^{EB(r)}$ were numerically calculated as

$$\text{cMSE}_1 = \frac{\lambda}{n_i\lambda + 1} \frac{Q_i + \tau_2}{n_i + \tau_1 - 2} + \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\xi}_1^{EB(r)} - \hat{\xi}_1^{(B)}\right)^2.$$

In the same manner as above, we calculated conditional MSE estimates and their relative biases based on $1,000$ simulation runs with each 100 bootstrap samples. The results from the simulations are reported in Table 3, which shows that values of the conditional MSE are small when the conditioning values are near the median and large when the conditioning values are near the upper or lower quantiles. Also, it is observed that the proposed estimator of the conditional MSE gives appropriate estimates for both $m = 20$ and 50.

Finally, we compared the three models, RHNER, HNER, and NER models in terms of bias and MSE. We considered the data generating process.

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, 5, \quad i = 1, \ldots, 30,$$

Table 3. Mean and relative bias of the estimators for the conditional MSE.

| Areas | $\alpha$ | $y_{1j(\alpha)}$ | $\text{cMSE}_1$ | $E[\widehat{\text{cMSE}_1}]$ | RB(%) |
|---|---|---|---|---|---|
| | 0.05 | -2.30 | 0.269 | 0.244 | -9.4 |
| | 0.25 | -0.72 | 0.129 | 0.121 | -6.3 |
| $m = 20$ | 0.50 | 0.00 | 0.112 | 0.098 | -12.5 |
| | 0.75 | 0.72 | 0.122 | 0.123 | 0.7 |
| | 0.95 | 2.30 | 0.243 | 0.239 | -1.69 |
| | 0.05 | -2.30 | 0.236 | 0.223 | -5.4 |
| | 0.25 | -0.72 | 0.131 | 0.121 | -7.3 |
| $m = 50$ | 0.50 | 0.00 | 0.112 | 0.107 | -3.8 |
| | 0.75 | 0.72 | 0.124 | 0.123 | -0.2 |
| | 0.95 | 2.30 | 0.224 | 0.228 | 2.1 |

where $v_i \sim \mathcal{N}(0, \lambda\sigma_i^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_i^2)$. We set $\beta_0 = \beta_1 = 1,\ \lambda = 1$, generated the $x_{ij}$ from $\mathcal{N}(0, 1)$ and fixed them in the simulation runs. For the setups of $\sigma_i^2$, we considered three patterns, where the second case (2) treats the case that the prior distribution of $\sigma_i^2$ is misspecified.

(1) $\sigma_i^{-2} \sim \mathcal{G}a(\tau_1/2, \tau_2/2)$ with $\tau_1 = 8$ and $\tau_2 = 4$.

(2) $\sigma_i^2 \sim \mathcal{U}(0, 2)$.

(3) $\sigma_i^2 = 1$ corresponding the case of homogeneous variances.

Based on $R = 5,000$ replications in Monte Calro simulation, we calculated the bias and the MSE of EB estimators for each model as

$$\text{bias}_i^{model} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\xi}_i^{model-EB(r)} - \xi_i^{(r)}), \quad \text{MSE}_i^{model} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\xi}_i^{model-EB(r)} - \xi_i^{(r)})^2,$$

where $model \in \{\text{RHNER}, \text{HNER}, \text{NER}\}$ and $\xi_i^{(r)}$ is the true values of $\xi_i$ in the $r$-th iteration. Then we calculated the ratio of absolute bias and MSE given by

$$\text{RB}_i^{\text{RH}} = \frac{|\text{bias}_i^{\text{RHNER}}|}{|\text{bias}_i^{\text{NER}}|}, \qquad \text{RB}_i^{\text{H}} = \frac{|\text{bias}_i^{\text{HNER}}|}{|\text{bias}_i^{\text{NER}}|},$$
$$\text{RMSE}_i^{\text{RH}} = \frac{\text{MSE}_i^{\text{RHNER}}}{\text{MSE}_i^{\text{NER}}}, \qquad \text{RMSE}_i^{\text{H}} = \frac{\text{MSE}_i^{\text{HNER}}}{\text{MSE}_i^{\text{NER}}}.$$

The results for the three designs of $\sigma_i^2$ are given in Figure 1. From Figure 1, the RHNER performs better than NER in the heteroscedastic variance cases (1) and (2) and as good as NER in the case (3) of homoscedastic variance. This means that the RHNER provides more efficient prediction than the NER in the heteroscedastic case and the predictor of the RHNER in the homogeneous case is as efficient as the predictor obtained from NER. Comparing RHNER with

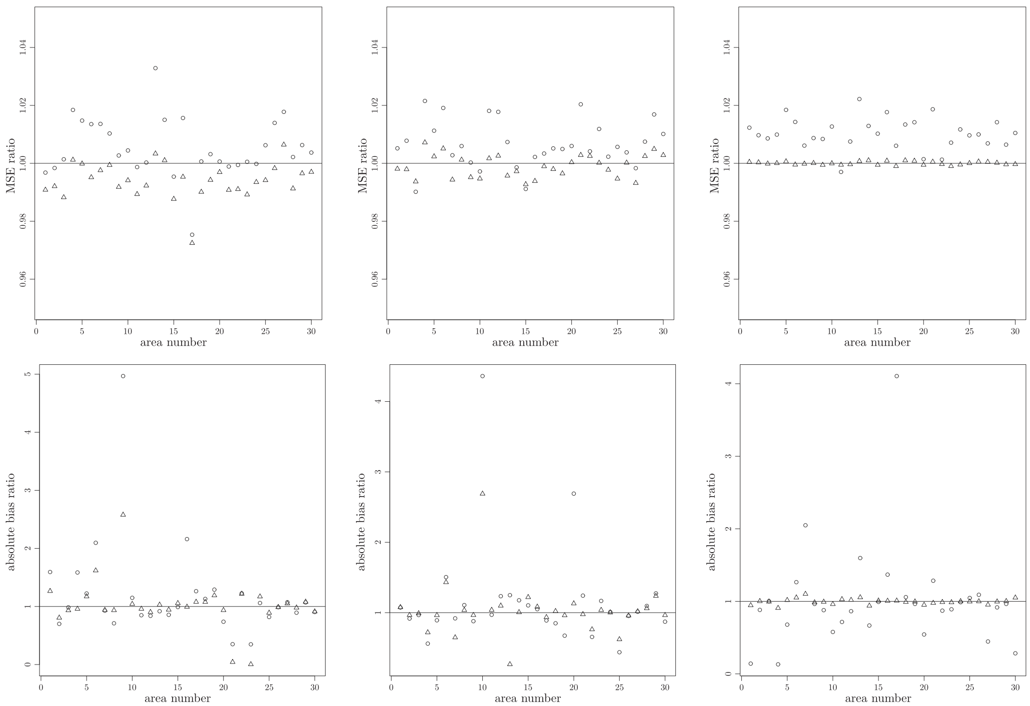


Figure 1. Ratio of the MSE and the Absolute Bias of RHNER (triangle) and HNER (circle) for Three Cases (1), (2) and (3) of $\sigma_i^2$ from Left to Right. (The upper and lower correspond to the MSE and absolute bias, respectively.)

HNER, we see that the RHNER provides better prediction than the HNER in three cases. It may be due to the fact that the RHNER can provide more stable estimates of $(\beta_0, \beta_1, \lambda)$ as evident from Table 1.

### 5.2. Example

This example, primarily for illustration, uses the RHNER model (2.2) and data from the posted land price data along the Keikyu train line in 2001. This train line connects the suburbs in the Kanagawa prefecture to the Tokyo metropolitan area. Those who live in the suburbs in the Kanagawa prefecture take this line to work or study in Tokyo everyday. Thus, it is expected that the land price depends on the distance from Tokyo. The posted land price data are available for 52 stations on the Keikyu train line, and we consider each station as a small area, so $m = 52$. For the $i$-th station, data of $n_i$ land spots are available, where $n_i$ varies around 4 and some areas have only one observation.

To investigate variability within each area, the boxplots are drawn for all areas. For nine selected areas among areas with more than 4 observations, we draw the boxplots in Figure 2, which clearly indicate that the posted land price has the

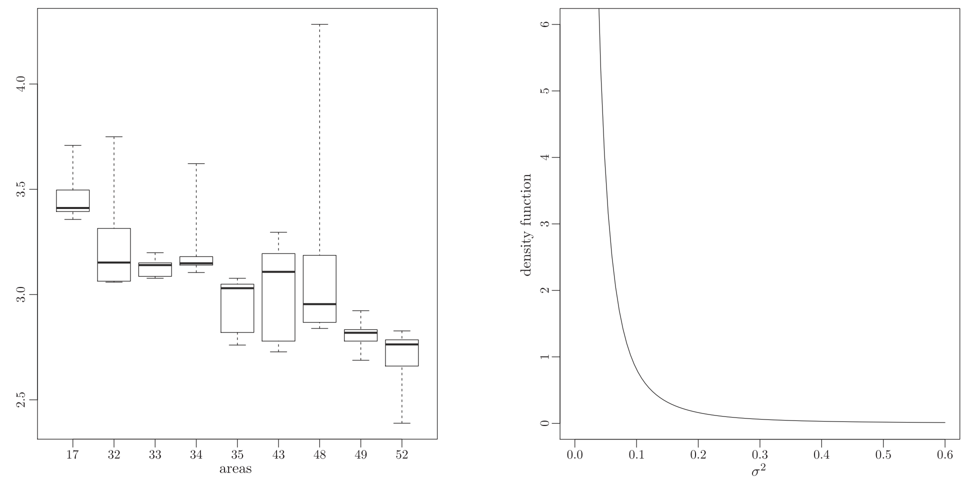


Figure 2. Boxplots of the Posted Land Price Data for Selected Areas (left) and the Estimated Density Function of $\sigma_i^2 = 1/\eta_i$ (right).

large within-area variation and the conventional NER model (which assumes homogeneity of variances) does not seem to be appropriate. Moreover, we apply the Bartlett test for the homogeneity on the variances, namely $H_0 : \sigma_1^2 = \cdots = \sigma_m^2$. Since the Bartlett test need enough sample sizes $n_i$'s, we here investigate the hypothesis $H_0$ for areas with $n_i \geq 6$. The number of areas with $n_i \geq 6$ is twelve. Then, the p-value is 0.011, and the hypothesis $H_0$ is significant for significance level $\alpha = 0.05$. Thus the heterogeneity assumption seems appropriate in this example.

For $j = 1, \ldots, n_i$, let $y_{ij}$ denote the log-transformed value of the posted land price (Yen/10,000) for the unit meter squares of the $j$-th spot, $T_i$ is the time to take from the nearby station $i$ to the Tokyo station around 8:30 in the morning, $D_{ij}$ is the value of geographical distance from the spot $j$ to the station $i$, and $FAR_{ij}$ denotes the floor-area ratio, or ratio of building volume to lot area of the spot $j$. The values of $T_i, D_{ij}$, and $FAR_{ij}$ are transformed by logarithm. Since these data have within-area variability as indicated in Figure 1 (left), we use the RHNER model

$$y_{ij} = \beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i + \varepsilon_{ij}, \tag{5.1}$$

where $v_i$ and $\varepsilon_{ij}$ are mutually independent and distributed as $\mathcal{N}(0, \lambda\sigma_i^2)$ and $\mathcal{N}(0, \sigma_i^2)$, and $\eta_i(= 1/\sigma_i^2)$ is independently distributed as $\mathcal{G}a(\tau_1/2, 2/\tau_2)$.

The estimates of the parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \lambda, \tau_1, \tau_2)$ are

$$\widehat{\beta}_0 = 5.69,\ \widehat{\beta}_1 = 0.11,\ \widehat{\beta}_2 = -0.63,\ \widehat{\beta}_3 = -0.08,\ \hat{\lambda} = 0.22,\ \hat{\tau}_1 = 2.93,\ \hat{\tau}_2 = 0.04.$$

It is interesting to point out that the estimated regression function is a decreasing function of $T_i$ and $D_{ij}$, which means that the land price $y_{ij}$ tends to decrease as the time from Tokyo or distance from nearest station increases. Since $\hat{\tau}_1 = 2.93$ and $\hat{\tau}_2 = 0.04$, the distribution of $\eta_i$ has a large mean about 73 and a heavy tail. Since the estimated value of $\tau_1$ is smaller than 4, from the note below (2.1), the variance of $\sigma_i^2$ does not exist, which corresponds to the observation that the posted land price data have great variability as indicated by the boxplots in Figure 2. Figure 2 (right) draws the estimated density function of $\sigma_i^2 = 1/\eta_i$ where $\eta_i$ has $\Gamma(\hat{\tau}_1/2, 2/\hat{\tau}_2)$, so that the distribution of $\sigma_i^2$ has a small mean, but a heavy tail.

We here estimate the land price of a spot with floor-area ratio 100% and distance 1,000m from the station $i$, namely

$$\xi_i = \beta_0 + FAR_0\beta_1 + T_i\beta_2 + D_0\beta_3 + v_i,$$

for $FAR_0 = \log(100)$ and $D_0 = \log(1,000)$. The predicted values of $\xi_i$ and their conditional and unconditional MSE estimates, based on 1,000 bootstrap samples, are given in Table 4. It is revealed from Table 4 that the estimates of the unconditional MSE get smaller as $n_i$ gets larger. On the other hands, the estimates of the conditional MSE do not have a similar property because the conditional MSE is affected by not only $n_i$, but also the observed values as indicated in Table 3. It is interesting that, in Area 48, the estimated conditional MSE is 1.99, while the estimated unconditional MSE is 0.14. This gives us a warning message on the value 2.98 of the EB based on the data from Area 48. Noting that this area has great variability as shown in Figure 2, it seems that the conditional MSE can capture the variability of areas. Thus, both estimates of the unconditional and conditional MSE are worth reporting. For comparison of RHNER, HNER, and NER, we also report the estimates of MSEs of HNER and NER in Table 4; The estimates of HNER are not credible values because of the inconsistency of $\hat{\sigma}_i^2$. It is observed from Table 4 that the MSEs of RHNER are smaller than that of HNER and NER, which motivate us to utilize RHNER in the case of heteroscedastic variances.

## 6. Concluding Remarks

In the context of small-area estimation, homogeneous linear mixed models like the nested error regression (NER) model have been studied. Jiang and Nguyen (2012) found that the data given in Battese, Harter and Fuller (1988) have heterogeneity, and first suggested the heteroscedastic nested error regression (HNER) model that assumes that the within-area variances are different among areas. Motivated by the inconsistency of the MLE of the dispersion $\sigma_i^2$, we propose the random dispersion heteroscedastic nested error regression (RHNER)

Table 4. Values of EB and Estimates of Unconditional and Conditional MSEs for Selected fifteen Areas of RHNER and Estimates of MSEs of HNER and NER. (Estimates of MSE and cMSE are multiplied by 100).

| | | RHNER | | | HNER | NER |
|---|---|---|---|---|---|---|
| Area | $n_i$ | EB | $\widehat{\text{cMSE}}$ | $\widehat{\text{MSE}}$ | $\widehat{\text{MSE}}$ | $\widehat{\text{MSE}}$ |
| 1 | 1 | 3.92 | 0.63 | 0.35 | 28.2 | 1.03 |
| 5 | 2 | 3.75 | 1.90 | 0.17 | 20.2 | 0.92 |
| 7 | 5 | 3.61 | 0.37 | 0.26 | 0.62 | 0.86 |
| 8 | 3 | 3.62 | 0.26 | 0.22 | 2.42 | 0.76 |
| 13 | 5 | 3.49 | 0.12 | 0.16 | 0.86 | 0.68 |
| 14 | 3 | 3.45 | 0.19 | 0.21 | 1.27 | 0.72 |
| 17 | 7 | 3.36 | 0.15 | 0.13 | 3.96 | 1.06 |
| 25 | 7 | 3.28 | 0.10 | 0.12 | 2.01 | 0.64 |
| 26 | 4 | 3.31 | 0.31 | 0.18 | 4.16 | 0.63 |
| 32 | 6 | 3.12 | 0.30 | 0.14 | 1.67 | 0.71 |
| 33 | 8 | 3.08 | 0.07 | 0.11 | 1.79 | 0.71 |
| 34 | 11 | 3.14 | 0.17 | 0.08 | 3.30 | 1.97 |
| 35 | 7 | 2.92 | 0.36 | 0.12 | 30.7 | 1.10 |
| 48 | 6 | 2.98 | 1.99 | 0.14 | 27.6 | 1.97 |
| 49 | 10 | 2.80 | 0.09 | 0.09 | 28.2 | 0.64 |

model. The consistency of the MLE of the parameters has been shown and their asymptotic variances and covariances have been derived. For measuring uncertainty of the EB estimator, the conditional and unconditional MSE's are approximated up to the second-order, and their second-order unbiased estimators are provided based on the parametric bootstrap method. Although the difference between the cMSE and MSE is quite small and appears in the second-order terms in classical normal linear mixed models, the difference appears in the leading or the first-order terms for the RHNER model.

As for a possible future study, it is interesting to construct a confidence interval of $\xi_i = \boldsymbol{c}_i^T\boldsymbol{\beta} + v_i$. In the RHNER model with random dispersions, it may be computationally harder to get corresponding confidence intervals. To this end, it is noted that

$$\begin{aligned} v_i|(\boldsymbol{y}_i, \eta_i) \sim& \mathcal{N}\Big(\hat{v}_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda), \frac{1-\gamma_i}{n_i\eta_i}\Big), \\ \eta_i|\boldsymbol{y}_i \sim& \mathcal{G}a\Big(\frac{n_i+\tau_1}{2}, \frac{2}{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda)+\tau_2}\Big), \end{aligned} \tag{6.1}$$

where $\hat{v}_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) = (1-\gamma_i)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})$ for $\gamma_i = 1/(n_i\lambda+1)$ and $Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda)$ given in (3.3). Let

$$Z_i = \frac{\sqrt{n_i}}{\sqrt{1-\gamma_i}\sqrt{Q_i+\tau_2}}(v_i - \hat{v}_i).$$

Then the conditional pdf of $Z_i$ given $\boldsymbol{y}_i$, denoted by $f_i(z_i|\tau_1)$, is given as

$$f_i(z|\tau_1) = \frac{\Gamma((n_i+\tau_1+1)/2)}{\sqrt{\pi}\Gamma((n_i+\tau_1)/2)}(1+z^2)^{-(n_i+\tau_1+1)/2}. \tag{6.2}$$

Define $z_{i,\alpha}(\tau_1)$ as the solution of the equation

$$\int_{z_{i,\alpha}(\tau_1)}^{\infty} f_i(z|\tau_1)\mathrm{d}z = \alpha.$$

Hence, $P[\xi_i > U_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda, \boldsymbol{\tau})] = \alpha$, where

$$U_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda, \boldsymbol{\tau}) = \boldsymbol{c}_i^T\boldsymbol{\beta} + \hat{v}_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda) + \frac{\sqrt{1-\gamma_i(\lambda)}\sqrt{Q_i(\boldsymbol{y}_i, \boldsymbol{\beta}, \lambda)+\tau_2}}{\sqrt{n_i}} z_{i,\alpha}(\tau_1).$$

Based on these equalities, we need to show that $P[\xi_i > U_i(\boldsymbol{y}_i, \widehat{\boldsymbol{\beta}}, \hat{\lambda}, \widehat{\boldsymbol{\tau}})] = \alpha + m^{-1}h(\boldsymbol{\beta}, \lambda, \boldsymbol{\tau}) + O(m^{-3/2})$. Since $z_{i,\alpha}(\tau_1)$ depends on unknown $\tau_1$, it is computationally hard to construct a confidence interval with second-order accuracy. This issue will be addressed in the future.

**Acknowledgements**

We would like to thank the Editor, an associate editor and the two reviewers for many valuable comments and helpful suggestions which led to an improvement in the presentation of the paper. Research of the first author was supported in part by Grant-in-Aid for Scientific Research (15H01943 and 26330036) from Japan Society for the Promotion of Science. The second author was supported in part by Grant-in-Aid for Scientific Research (15J10076) from Japan Society for the Promotion of Science (JSPS) The work of the third author was supported by an NSF Grant SES-1026165.

## Appendix

We begin with two lemmas that are used for calculating the Fisher information in Theorem 1.

**Lemma A.1.** Conditional on $\eta_i$, $Q_i$ defined in (3.2) is distributed independently of $(\sum_{j=1}^{n_i}\{(y_{ij}-\overline{y}_i)-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}^2/Q_i, (\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2/Q_i)$.

**Proof.** Assume that $\boldsymbol{\beta}$ and $\lambda$ are fixed. Given the joint pdf in (3.2), conditional on $\eta_i$, $Q_i$ is complete sufficient, while $(\sum_{j=1}^{n_i}\{(y_{ij}-\overline{y}_i)-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}^2/Q_i, (\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2/Q_i)$ is ancillary. We apply Basu's theorem to get Lemma A.1.

**Lemma A.2.** Let $R_i = Q_i/(Q_i+\tau_2)$. After integrating out $\eta_i$, $R_i \sim Beta(n_i/2, \tau_1/2)$.

**Proof.** Here $Q_i|\eta_i \sim \eta_i^{-1}\chi^2_{n_i}$ and $\eta_i \sim \mathcal{G}a(\tau_1/2, 2/\tau_2)$. Integrating out $\eta_i$, the marginal pdf of $Q_i$ is

$$\begin{aligned} f(Q_i) =& \frac{Q_i^{n_i/2-1}\tau_2^{\tau_1/2}}{(Q_i+\tau_2)^{(n_i+\tau_1)/2}} \frac{\Gamma((n_i+\tau_1)/2)}{\Gamma(n_i/2)\Gamma(\tau_1/2)} \\ =& \Big(\frac{Q_i}{Q_i+\tau_2}\Big)^{n_i/2-1}\Big(\frac{\tau_2}{Q_i+\tau_2}\Big)^{\tau_1/2-1} \frac{\tau_2}{(Q_i+\tau_2)^2}\frac{1}{B(n_i/2,\tau_1/2)}. \end{aligned}$$

Then, $R_i$ has pdf $f(R_i) = R_i^{n_i/2-1}(1-R_i)^{\tau_1/2-1}/B(n_i/2,\tau_1/2)$, which proves Lemma A.2.

**Proof of Theorem 1.** It follows as a consequence of Lemma A.2 that

$$E\Big[\frac{1}{Q_i+\tau_2}\Big] =\tau_2^{-1}E[1-R_i] = \frac{\tau_1}{\tau_2}\frac{1}{n_i+\tau_1}, \tag{A.1}$$

$$E\Big[\frac{Q_i}{(Q_i+\tau_2)^2}\Big] =\tau_2^{-1}E[R_i(1-R_i)] = \frac{\tau_1}{\tau_2}\frac{n_i}{(n_i+\tau_1)(n_i+\tau_1+2)}, \tag{A.2}$$

$$E\Big[\frac{Q_i^2}{(Q_i+\tau_2)^2}\Big] =E[R_i^2] = \frac{n_i(n_i+2)}{(n_i+\tau_1)(n_i+\tau_1+2)}. \tag{A.3}$$

We use (A.1), (A.2) and (A.3) repeatedly in the following calculations.

Begin with the second derivative $L_{\boldsymbol{\beta}\boldsymbol{\beta}}$, which can be written from (3.5) as

$$2L_{\boldsymbol{\beta}\boldsymbol{\beta}} = -\sum_{i=1}^m \frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial^2 Q_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} + \sum_{i=1}^m \frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}. \tag{A.4}$$

But,

$$\frac{\partial^2 Q_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = 2\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T + 2n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T, \tag{A.5}$$

for $\gamma_i = 1/(n_i\lambda+1)$. Now

$$\begin{aligned} \frac{\partial Q_i}{\partial\boldsymbol{\beta}} =& -2\sum_{j=1}^{n_i}\{(y_{ij}-\overline{y}_i)-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i) - 2n_i\gamma_i(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i \\ =& -2\sum_{j=1}^{n_i}(e_{ij}-\overline{e}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i) - 2n_i\gamma_i(v_i+\overline{e}_i)\overline{\boldsymbol{x}}_i, \end{aligned}$$

where $e_{ij} = y_{ij}-\boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ and $\overline{e}_i = n_i^{-1}\sum_{j=1}^{n_i}e_{ij}$. One can also write $Q_i = \sum_{j=1}^{n_i}(e_{ij}-\overline{e}_i)^2 + n_i\gamma_i(v_i+\overline{e}_i)^2$. Once again, conditional on $\eta_i$, $Q_i^{-1}(\partial Q_i/\partial\boldsymbol{\beta})(\partial Q_i/\partial\boldsymbol{\beta}^T)$ is ancillary and is thus independent of $Q_i$. This leads to

$$E\Big[(Q_i+\tau_2)^{-2}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big] =E\Big[\frac{Q_i}{(Q_i+\tau_2)^2}Q_i^{-1}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big]$$

$$=E\Big[\frac{Q_i}{(Q_i+\tau_2)^2}\Big|\eta_i\Big]E\Big[Q_i^{-1}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big].$$

Similarly,

$$E\Big[\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big]=E[Q_i|\eta_i]E\Big[Q_i^{-1}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big],$$

so that

$$E\Big[Q_i^{-1}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big]=E\Big[\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big]/E[Q_i|\eta_i]. \tag{A.6}$$

But, using the fact that $(e_{i1},\ldots,e_{in_i},v_i)$ and $-(e_{i1},\ldots,e_{in_i},v_i)$ have the same distribution and $(e_{i1}-\overline{e}_i,\ldots,e_{in_i}-\overline{e}_i)$ is distributed independently of $(v_i,\overline{e}_i)$ conditional on $\eta_i$, it follows that

$$\begin{aligned}E\Big[\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\boldsymbol{\beta}^T}\Big|\eta_i\Big]=&4E\Big[\sum_{j=1}^{n_i}e_{ij}^2(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T+n_i^2\gamma_i^2(v_i+\overline{e}_i)^2\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T\Big]\\ =&4n_i^{-1}\Big[\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T+n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T\Big].\end{aligned} \tag{A.7}$$

Combining (A.1), (A.2), and (A.4)–(A.7), one gets

$$\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}=E[-L_{\boldsymbol{\beta}\boldsymbol{\beta}}]=\frac{\tau_1}{\tau_2}\sum_{i=1}^m\frac{n_i+\tau_1}{n_i+\tau_1+2}\Big\{\sum_{j=1}^{n_i}(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T+n_i\gamma_i\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i^T\Big\}.$$

Next, observe that

$$\begin{aligned}2L_{\boldsymbol{\beta}\lambda}=&-\sum_{i=1}^m\frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial^2Q_i}{\partial\boldsymbol{\beta}\partial\lambda}+\sum_{i=1}^m\frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\frac{\partial Q_i}{\partial\lambda}\\ =&-2\sum_{i=1}^m n_i^2\gamma_i^2\frac{n_i+\tau_1}{Q_i+\tau_2}(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i\\ &+2\sum_{i=1}^m n_i^2\gamma_i^2\frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\Big\{\sum_{j=1}^{n_i}\big[y_{ij}-\overline{y}_i-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\big](\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)\\ &+n_i\gamma_i(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i\Big\}(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2\\ =&-2\sum_{i=1}^m n_i^2\gamma_i^2\frac{n_i+\tau_1}{Q_i+\tau_2}(v_i+\overline{e}_i)\overline{\boldsymbol{x}}_i\\ &+2\sum_{i=1}^m n_i^2\gamma_i^2\frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\Big\{\sum_{j=1}^{n_i}(e_{ij}-\overline{e}_i)(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)\\ &+n_i\gamma_i(v_i+\overline{e}_i)\overline{\boldsymbol{x}}_i\Big\}(v_i+\overline{e}_i)^2.\end{aligned}$$

Arguing as before, $(e_{i1},\ldots,e_{in_i},v_i)$ and $-(e_{i1},\ldots,e_{in_i},v_i)$ have the same distribution and $(e_{i1}-\overline{e}_i,\ldots,e_{in_i}-\overline{e}_i)$ is distributed independently of $(v_i,\overline{e}_i)$ conditional on $\eta_i$, so $\boldsymbol{I}_{\boldsymbol{\beta}\lambda}=-E[L_{\boldsymbol{\beta}\lambda}]=\mathbf{0}$. Similarly,

$$\begin{aligned}2L_{\boldsymbol{\beta}\tau_1}=&-\sum_{i=1}^m(Q_i+\tau_2)^{-1}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\\ =&2\sum_{i=1}^m\frac{1}{Q_i+\tau_2}\Big\{\sum_{j=1}^{n_i}\big[y_{ij}-\overline{y}_i-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\big](\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)+n_i\gamma_i(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i\Big\},\end{aligned}$$

so that $\boldsymbol{I}_{\boldsymbol{\beta}\tau_1}=-E[L_{\boldsymbol{\beta}\tau_1}]=\mathbf{0}$. Moreover,

$$\begin{aligned}2L_{\boldsymbol{\beta}\tau_2}=&\sum_{i=1}^m(n_i+\tau_1)(Q_i+\tau_2)^{-2}\frac{\partial Q_i}{\partial\boldsymbol{\beta}}\\ =&2\sum_{i=1}^m\frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\Big\{\sum_{j=1}^{n_i}\big[y_{ij}-\overline{y}_i-(\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)^T\boldsymbol{\beta}\big](\boldsymbol{x}_{ij}-\overline{\boldsymbol{x}}_i)+n_i\gamma_i(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\overline{\boldsymbol{x}}_i\Big\},\end{aligned}$$

so that $\boldsymbol{I}_{\boldsymbol{\beta}\tau_2}=-E[L_{\boldsymbol{\beta}\tau_2}]=\mathbf{0}$.

Finally, we evaluate the second derivatives in $L_{\boldsymbol{\theta\theta}}$ for $\boldsymbol{\theta}=(\lambda,\tau_1,\tau_2)^T$. First, we compute

$$\begin{aligned}2L_{\lambda\lambda}=&-\sum_{i=1}^m\frac{n_i+\tau_1}{Q_i+\tau_2}\frac{\partial^2Q_i}{\partial\lambda^2}+\sum_{i=1}^m\frac{n_i+\tau_1}{(Q_i+\tau_2)^2}\Big(\frac{\partial Q_i}{\partial\lambda}\Big)^2+\sum_{i=1}^m n_i^2\gamma_i^2\\ =&-2\sum_{i=1}^m\frac{(n_i+\tau_1)Q_i}{Q_i+\tau_2}n_i^3\gamma_i^3\frac{(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2}{Q_i}\\ &+\sum_{i=1}^m\frac{(n_i+\tau_1)Q_i^2}{(Q_i+\tau_2)^2}n_i^4\gamma_i^4\frac{(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^4}{Q_i^2}+\sum_{i=1}^m n_i^2\gamma_i^2. \qquad (A.8)\end{aligned}$$

Observing that $(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2n_i\gamma_i/Q_i\sim Beta(1/2,n_i/2)$ and is distributed independently of $Q_i$, one gets

$$E\Big[n_i\gamma_i\frac{(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2}{Q_i}\Big]=\frac{1}{n_i}, \qquad (A.9)$$

$$E\Big[(n_i\gamma_i)^2\frac{(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^4}{Q_i^2}\Big]=\frac{3}{n_i(n_i+2)}. \qquad (A.10)$$

Hence, from (A.1), (A.3), and (A.8)–(A.10),

$$\begin{aligned}I_{\lambda\lambda}=&E[-L_{\lambda\lambda}]\\ =&\sum_{i=1}^m E\Big[\frac{(n_i+\tau_1)Q_i}{Q_i+\tau_2}\Big]n_i^2\gamma_i^2n_i^{-1}-\frac{1}{2}\sum_{i=1}^m E\Big[\frac{(n_i+\tau_1)Q_i^2}{(Q_i+\tau_2)^2}\Big]n_i^2\gamma_i^2\frac{3}{n_i(n_i+2)}-\sum_{i=1}^m n_i^2\gamma_i^2\end{aligned}$$

$$=\frac{1}{2}\Big[\sum_{i=1}^m n_i^2\gamma_i^2-3\sum_{i=1}^m \frac{n_i^2\gamma_i^2}{n_i+\tau_1+2}\Big]$$
$$=\frac{1}{2}\sum_{i=1}^m \frac{n_i+\tau_1-1}{n_i+\tau_1+2}n_i^2\gamma_i^2.$$

Next, $2L_{\lambda\tau_1}=-\sum_{i=1}^m(Q_i+\tau_2)^{-1}\partial Q_i/\partial\lambda=\sum_{i=1}^m\{Q_i/(Q_i+\tau_2)\}n_i\gamma_i\{n_i\gamma_i(\overline{y}-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2/Q_i\}$, which yields

$$I_{\lambda\tau_1}=E[-L_{\lambda\tau_1}]=-\frac{1}{2}\sum_{i=1}^m\{\frac{n_i}{n_i+\tau_1}\}n_i\gamma_i n_i^{-1}=-\frac{1}{2}\sum_{i=1}^m\frac{n_i\gamma_i}{(n_i+\tau_1)}.$$

Since $2L_{\lambda\tau_2}=\sum_{i=1}^m(n_i+\tau_1)(Q_i+\tau_2)^{-2}\partial Q_i/\partial\lambda=-\sum_{i=1}^m\{(n_i+\tau_1)Q_i/(Q_i+\tau_2)^2\}n_i\gamma_i\{n_i\gamma_i(\overline{y}-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2/Q_i\}$, one gets

$$I_{\lambda\tau_2}=E[-L_{\lambda\tau_2}]=\frac{1}{2}\frac{\tau_1}{\tau_2}\sum_{i=1}^m\frac{n_i\gamma_i}{n_i+\tau_1+2}.$$

We observe that $2L_{\tau_1\tau_1}=(1/2)\sum_{i=1}^m\{\psi'((n_i+\tau_1)/2)-\psi'(\tau_1/2)\}$, $2L_{\tau_1\tau_2}=m/\tau_2-\sum_{i=1}^m(Q_i+\tau_2)^{-1}$ and $2L_{\tau_2\tau_2}=-m\tau_1/\tau_2^2+\sum_{i=1}^m(n_i+\tau_1)(Q_i+\tau_2)^{-2}$. Then,

$$I_{\tau_1\tau_1}=E[-L_{\tau_1\tau_1}]=\frac{1}{4}\sum_{i=1}^m\Big\{\psi'(\frac{\tau_1}{2})-\psi'(\frac{n_i+\tau_1}{2})\Big\}.$$

Using Lemma A.2, one gets $I_{\tau_1\tau_2}=E[-L_{\tau_1\tau_2}]=-(2\tau_2)^{-1}\sum_{i=1}^m n_i/(n_i+\tau_1)$ and $I_{\tau_2\tau_2}=E[-L\tau_2\tau_2]=\tau_1(2\tau_2^2)^{-1}\sum_{i=1}^m n_i/(n_i+\tau_1+2)$.

**Proof of Theorem 2.** Let $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_{p+3})^T=(\boldsymbol{\beta}^T,\lambda,\boldsymbol{\tau}^T)^T$. The log likelihood of $(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)$, denoted by $\ell(\boldsymbol{\omega})$, is given by $\ell(\boldsymbol{\omega})=\sum_{j=1}^m\ell(\boldsymbol{\omega};\boldsymbol{y}_j)$, where $\ell(\boldsymbol{\omega};\boldsymbol{y}_j)=\log f(\boldsymbol{y}_j|\boldsymbol{\omega})$ is the log likelihood function of $\boldsymbol{y}_j$. Let $\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})=(\partial/\partial\boldsymbol{\omega})\ell(\boldsymbol{\omega})$ and $\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})=(\partial^2/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T)\ell(\boldsymbol{\omega})$. The $(a,b)$-element of $\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})$ is written as $\big(\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\big)_{ab}=\sum_{j=1}^m\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)$, where $\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)=(\partial^2/\partial\omega_a\partial\omega_b)\ell(\boldsymbol{\omega};\boldsymbol{y}_j)$. Since $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m$ are mutually independent, the law of large numbers implies that $-m^{-1}\sum_{j=1}^m\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)$ given $\boldsymbol{y}_i$ converges to the limit of $m^{-1}\sum_{j=1}^m E[-\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)|\boldsymbol{y}_i]$. Let $I_{ab}(\boldsymbol{\omega})=m^{-1}\sum_{j=1}^m E[-\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)]$. Now

$$I_{ab}(\boldsymbol{\omega})-m^{-1}\sum_{j=1}^m E[-\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)|\boldsymbol{y}_i]=\frac{1}{m}\Big\{E[-\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_i)]+\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_i)\Big\},$$

is of order $O_p(m^{-1})$. This shows that given $\boldsymbol{y}_i$, $-m^{-1}\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})|\boldsymbol{y}_i=m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})+O_p(m^{-1/2})$, where $\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})=-E[\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})]$. Unconditionally, $-m^{-1}\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})=m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})+O_p(m^{-1/2})$. Since $\lim_{m\to\infty}m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})$ is positive definite, it follows from Mardia and Marshall (1984) or Sweeting (1980) that $\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega} = O_p(m^{-1/2})$ for the MLE $\widehat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$.

Using a Taylor series expansion and the above approximation, we can see that

$$\begin{aligned} 0 = \ell_{\boldsymbol{\omega}}(\widehat{\boldsymbol{\omega}}) =& \ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}) + \ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) + O_p(1), \\ =& \ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}) - \boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) + O_p(1). \end{aligned}$$

This implies that

$$\sqrt{m}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) = \left\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^{-1} m^{-1/2}\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}) + O_p(m^{-1/2}). \tag{A.11}$$

Hence, it follows that

$$\begin{aligned} &E[(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})^T|\boldsymbol{y}_i] \\ &\quad = \left\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^{-1}\frac{1}{m^2}E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^T|\boldsymbol{y}_i]\left\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^{-1} + O_p(m^{-3/2}). \end{aligned}$$

Since $E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)] = \boldsymbol{0}$ and $\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}) = \sum_{j=1}^m \ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)$, it is seen that

$$\begin{aligned} &\frac{1}{m^2}E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^T|\boldsymbol{y}_i] \\ &\quad = \frac{1}{m^2}\sum_{j=1, j\neq i}^m E[\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)\right\}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)\right\}^T|\boldsymbol{y}_i] + \frac{1}{m}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}^T \\ &\quad = \frac{1}{m^2}\sum_{j=1}^m E[\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)\right\}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)\right\}^T] \\ &\qquad + \frac{1}{m^2}\Big\{\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}^T - E[\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)\right\}^T]\Big\}, \end{aligned}$$

which implies that

$$\frac{1}{m^2}E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\left\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^T|\boldsymbol{y}_i] = \frac{1}{m^2}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega}) + O_p(m^{-2}).$$

Thus, one gets $E[(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})^T|\boldsymbol{y}_i] = \boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})^{-1} + O_p(m^{-1/2})$, which proves (3.8) in Theorem 2. This implies that conditionally, $\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}|\boldsymbol{y}_i = O_p(m^{-1/2})$.

Concerning the bias of the MLE $\widehat{\boldsymbol{\omega}}$, from (A.11), it follows that

$$E[\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}|\boldsymbol{y}_i] = \left\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\right\}^{-1}m^{-1}E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})|\boldsymbol{y}_i] + O_p(m^{-1}).$$

It is here noted that $E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})|\boldsymbol{y}_i] = \sum_{j=1}^m E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)|\boldsymbol{y}_i] = \sum_{j=1}^m E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_j)] + \{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i) - E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega}; \boldsymbol{y}_i)]\} = 0 + O_p(1)$, so that $E[\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}|\boldsymbol{y}_i] = O_p(m^{-1})$. This proves the second part of Theorem 2, and now the proof of Theorem 2 is complete.

**Proof of Theorem 3.** We evaluate the second terms $g_2^c(\omega|\boldsymbol{y}_i)$ and $g_2(\omega)$. Since $\hat{\xi}_i^{EB} - \hat{\xi}_i^B$ is decomposed as

$$\begin{aligned}\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}},\hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta},\lambda) =& \frac{1}{n_i\hat{\lambda}+1}\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \Big(\frac{n_i\hat{\lambda}}{n_i\hat{\lambda}+1} - \frac{n_i\lambda}{n_i\lambda+1}\Big)(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\\ =& \frac{1}{n_i\hat{\lambda}+1}\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \frac{n_i(\hat{\lambda}-\lambda)}{(n_i\lambda+1)(n_i\hat{\lambda}+1)}(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}),\end{aligned}$$

$g_2^c(\omega|\boldsymbol{y}_i)$ can be expressed as

$$\begin{aligned}g_2^c(\omega|\boldsymbol{y}_i) =& E\Big[\frac{1}{(n_i\hat{\lambda}+1)^2}\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2\Big|\boldsymbol{y}_i\Big]\\ &+ \Big(\frac{n_i}{n_i\lambda+1}\Big)^2 E\Big[\Big(\frac{\hat{\lambda}-\lambda}{n_i\hat{\lambda}+1}\Big)^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2\Big|\boldsymbol{y}_i\Big]\\ &+ 2\frac{n_i}{n_i\lambda+1}E\Big[\frac{\hat{\lambda}-\lambda}{(n_i\hat{\lambda}+1)^2}\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})\Big|\boldsymbol{y}_i\Big]\\ =& \frac{1}{(n_i\lambda+1)^2}E[\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2|\boldsymbol{y}_i] + \frac{n_i^2}{(n_i\lambda+1)^4}E[(\hat{\lambda}-\lambda)^2|\boldsymbol{y}_i](\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2\\ &+ 2\frac{n_i}{(n_i\lambda+1)^3}E[(\hat{\lambda}-\lambda)\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})|\boldsymbol{y}_i](\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta}) + O_p(m^{-3/2}).\end{aligned}$$

It follows from Theorem 2 that $E[\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2|\boldsymbol{y}_i] = E[\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2] + O_p(m^{-3/2}) = \boldsymbol{c}_i^T(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}\boldsymbol{c}_i + O_p(m^{-3/2})$, $E[(\hat{\lambda}-\lambda)^2|\boldsymbol{y}_i] = E[(\hat{\lambda}-\lambda)^2] + O_p(m^{-3/2}) = I^{\lambda\lambda} + O_p(m^{-3/2})$, and $E[(\hat{\lambda}-\lambda)\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})|\boldsymbol{y}_i] = E[(\hat{\lambda}-\lambda)\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})] + O_p(m^{-3/2}) = O_p(m^{-3/2})$. Thus, one gets

$$g_2^c(\omega|\boldsymbol{y}_i) = \gamma_i^2\boldsymbol{c}_i^T\big(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\big)^{-1}\boldsymbol{c}_i + n_i^2\gamma_i^4(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 I^{\lambda\lambda} + O_p(m^{-3/2}). \tag{A.12}$$

Combining (4.3) and (A.12) yields the approximation (4.5). Since

$$E[(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2] = \frac{n_i\lambda+1}{n_i}E[\eta_i^{-1}] = \frac{n_i\lambda+1}{n_i}\frac{\tau_2}{\tau_1-2},$$

one gets

$$g_2(\omega) = \gamma_i^2\boldsymbol{c}_i^T\big(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\big)^{-1}\boldsymbol{c}_i + n_i\gamma_i^3\frac{\tau_2}{\tau_1-2}I^{\lambda\lambda} + O(m^{-3/2}). \tag{A.13}$$

Combining (4.4) and (A.13) gives the expression (4.6). Now the proof of Theorem 3 is complete.

**Proof of Proposition 1.** It follows from (4.4) and the proof of Theorem 3 that $MSE(\omega;\hat{\xi}_i^{EB}) = g_1(\omega) + g_2(\omega) + O(m^{-3/2})$, where $g_2(\omega) = \gamma_i^2 E[\{\boldsymbol{c}_i^T(\widehat{\boldsymbol{\beta}} -$

$\boldsymbol{\beta})\}^2] + n_i\gamma_i^3 E[(\hat{\lambda}-\lambda)^2]$. It is seen from Butar and Lahiri (2003) that $E[\hat{g}_1^*] = g_1(\omega) + O(m^{-3/2})$. Also, it can be shown that $E[\hat{g}_2^*] = g_2(\omega) + O(m^{-3/2})$. This proves Proposition 1.

**Proof of Theorem 4.** It follows from (4.1) and (4.3) that $cMSE(\omega; \hat{\xi}_i^{EB}|\boldsymbol{y}_i) = g_1^c(\omega|\boldsymbol{y}_i) + g_2^c(\omega|\boldsymbol{y}_i) + O_p(m^{-3/2})$, where

$$g_1^c(\omega|\boldsymbol{y}_i) = \frac{\lambda}{n_i\lambda+1}\frac{Q_i+\tau_2}{n_i+\tau_1-2},$$
$$g_2^c(\omega|\boldsymbol{y}_i) = E[\{\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}},\hat{\lambda}) - \hat{\xi}_i^B(\boldsymbol{\beta},\lambda)\}^2|\boldsymbol{y}_i].$$

First, note that $g_1^c(\hat{\omega}|\boldsymbol{y}_i)$ is expanded as $g_1^c(\hat{\omega}|\boldsymbol{y}_i) = g_1^c(\omega|\boldsymbol{y}_i) + G_1(\omega|\boldsymbol{y}_i) + O_p(m^{-3/2})$, where

$$G_1(\hat{\omega},\omega|\boldsymbol{y}_i) = (\hat{\omega}-\omega)^T\frac{\partial g_1^c(\hat{\omega}|\boldsymbol{y}_i)}{\partial\omega} + \frac{1}{2}(\hat{\omega}-\omega)^T\frac{\partial^2 g_1^c(\hat{\omega}|\boldsymbol{y}_i)}{\partial\omega\partial\omega^T}(\hat{\omega}-\omega).$$

Since $E[\hat{\omega}-\omega|\boldsymbol{y}_i] = O_p(m^{-1})$, from Theorem 2, it is seen that $E[G_1(\hat{\omega},\omega|\boldsymbol{y}_i)|\boldsymbol{y}_i] = O_p(m^{-1})$. Hence, $E[g_1^c(\hat{\omega}|\boldsymbol{y}_i)|\boldsymbol{y}_i] = g_1^c(\omega|\boldsymbol{y}_i) + E[G_1(\hat{\omega},\omega|\boldsymbol{y}_i)|\boldsymbol{y}_i] + O_p(m^{-3/2})$. Using the same atguments as in Butar and Lahiri (2003), we can see that $E[\hat{g}_1^{c*}|\boldsymbol{y}_i] = g_1^c(\omega|\boldsymbol{y}_i) + O(m^{-3/2})$. Also, from Theorem 2, it can be shown that $E[\hat{g}_2^{*c}|\boldsymbol{y}_i] = g_2^c(\omega|\boldsymbol{y}_i) + O_p(m^{-3/2})$. This completes the proof of Theorem 4.

## References

Battese, G. E., Harter, R. M. and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83**, 28-36.

Booth, J. S. and Hobert, P. (1998). Standard errors of prediction in generalized linear mixed models. *J. Amer. Statist. Assoc.* **93**, 262-272.

Butar, F. B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plann. Inference* **112**, 63-76.

Datta, G. S. (2009). Model-based approach to small area estimation. In *Handbook of Statistics*, **29B** (Edied by D. Pfeffermann and C. R. Rao), 251-288. North Holland, New York.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statist. Sci.* **9**, 55-93.

Jiang, J. and Nguyen, T. (2012). Small area estimation via heteroscedastic nested-error regression. *Canad. J. Statist.* **40**, 588-603.

Maiti, T., Ren, H. and Sinha, S. (2014). Prediction error of small area predictors shrinking both means and variances. *Scand. J. Statist.* **41**, 775-790.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.

Pfeffermann, D. (2002). Small area estimation: new developments and directions. *Internat. Statist. Inst. Rev.* **70**, 125-143.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, New York.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375-1381.

Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail: tatsuya@e.u-tokyo.ac.jp

Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

E-mail: shonosuke622@gmail.com

Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, Florida 32611, USA.

E-mail: ghoshm@stat.ufl.edu

Department of Statistics and Applied Probability, National University of Singapore, Block S16, Level 7, 6 Science Drive 2, Singapore 117546.

E-mail: stasc@nus.edu.sg

(Received October 2014; accepted July 2015)