CIRJE-F-939

Prediction in Heteroscedastic Nested Error Regression Models with Random Dispersion

Tatsuya Kubokawa
The University of Tokyo

Shonosuke Sugasawa
Graduate School of Economics, The University of Tokyo

Malay Ghosh
University of Florida

Sanjay Chaudhuri
National University of Singapore

August 2014

Prediction in Heteroscedastic Nested Error Regression Models with Random Dispersions

Tatsuya Kubokawa[*], Shonosuke Sugasawa[†], Malay Ghosh[‡], and Sanjay Chaudhuri[§]

University of Tokyo, University of Florida and National University of Singapore

August 29, 2014

Abstract

The paper concerns small-area estimation in the heteroscedastic nested error regression (HNER) model which 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 samples sizes for small-areas. To fix this difficulty, 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 small sample sizes. Next we propose an empirical Bayes procedure 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 their second-order 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 the posted land price data.

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

[*]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. 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

1 Introduction

Linear mixed (LM) models and the model-based estimators including empirical Bayes estimator (EB) or empirical best linear unbiased predictor (EBLUP) have been studied quite extensively 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 as a unit-level model. The NER model with m small-areas assumes that the data $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ are taken from the i-th small-area, $i = 1, \ldots, m$, where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually independent. It is further assumed that y_{ij} is normally distributed with $E[y_{ij}] = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$, $Var(y_{ij}) = \sigma_y^2$ and $Corr(y_{ij}, y_{ik}) = \rho$, $j \neq k$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$, σ_y^2 and ρ are unknown parameters, $\boldsymbol{x}_{ij}$'s are known vectors of covariates, and $Corr(y_{ij}, y_{ik})$ denoted the correlation coefficient of y_{ij} and y_{ik}. The NER model can be expressed as a random effects model with

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

where v_i's and ε_{ij}'s are mutually independent with $v_i \sim \mathcal{N}(0, \lambda\sigma^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Then $Var(v_i)/Var(\varepsilon_{ij}) = \lambda$, and that σ_y^2 and ρ correspond to

$$\sigma_y^2 = (1 + \lambda)\sigma^2 \quad \text{and} \quad \rho = \lambda/(1 + \lambda).$$

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, *et al.* (1988). Figure 1, given in Section 5 in this paper, also indicates variability of the within-area variances. Jiang and Nguyen (2012) assumed that the variance $E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2]$ is proportional to σ_i^2, which depends on the area i. Since $E[(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta})^2] = E[(v_i + \varepsilon_{ij})^2] = Var(v_i) + Var(\varepsilon_{ij})$, this assumption implies that $Var(v_i) + Var(\varepsilon_{ij}) = C\sigma_i^2$ for some constant C. Letting $Var(\varepsilon_{ij}) = \sigma_i^2$, we can see that $\{Var(v_i)/Var(\varepsilon_{ij}) + 1\}Var(\varepsilon_{ij}) = \{Var(v_i)/Var(\varepsilon_{ij}) + 1\}\sigma_i^2 = C\sigma_i^2$, which means that

$$Var(v_i)/Var(\varepsilon_{ij}) = C - 1,$$

namely, $Var(v_i)/Var(\varepsilon_{ij})$ is a constant. Using the same notation as in the NER model, we write $Var(v_i)/Var(\varepsilon_{ij}) = \lambda$. Thus, the heteroscedastic nested error regression (HNER) model suggested by Jiang and Nguyen (2012) is the model given in (1.1) with

$$Var(v_i) = \lambda\sigma_i^2 \quad \text{and} \quad Var(\varepsilon_{ij}) = \sigma_i^2. \tag{1.2}$$

In the HNER model, Jiang and Nguyen (2012) demonstrated that the maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and λ are consistent for large m and that the resulting empirical Bayes (EB) estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ ($\overline{\boldsymbol{x}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}$) estimates the Bayes estimator consistently since the Bayes estimator does not depend on $\sigma_1^2, \ldots, \sigma_m^2$, but on $\boldsymbol{\beta}$ and λ. This is quite interesting, because the number of unknown variances σ_i^2's increases as m tends to infinity. However, the posterior variance of v_i given $(y_{i1}, \ldots, y_{i,n_i})$ is

$$Var(v_i|y_{i1}, \ldots, y_{i,n_i}) = \sigma_i^2 \lambda/(1 + n_i\lambda),$$

which implies that the mean squared error (MSE) of the EB of ξ_i depends on σ_i^2. Then, we need to estimate σ_i^2 for estimating the MSE of the EB of ξ_i. Since the sample sizes n_i is small in the small-area estimation, we cannot provide good estimates for σ_i^2 with reasonable precision.

In this paper, we suggest a random dispersion HNER (RHNER) model which assumes that the priors of $(\sigma_i^2)^{-1}$, $i = 1, \ldots, m$, are mutually independent gamma random variables. The resulting Bayes estimator of σ_i^2 gives stable shrinkage estimates of small area means even for $n_i - p = 0$.

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

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

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 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) demonstrated that the difference between the cMSE and MSE is quite small and appears in the second-order terms in classical normal linear mixed models. In this paper, however, we show that the difference appears in the leading or the first-order terms for 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, The maximum likelihood (ML) estimators are described for estimating the unknown $\boldsymbol{\beta}$, λ 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 for ξ_i and their second-order unbiased estimators 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 (1.2). For stable estimators of σ_i^2's, we need enough data from each area. Since n_i's are typically small, σ_i^2 cannot usually be estimated with reasonable precision. To give more stable estimators for σ_i^2, we assume a prior distribution for σ_i^2. Let $\eta_i = 1/\sigma_i^2$. It is assumed that $\eta_1, \ldots, \eta_m$ are independent and identically distributed with common pdf

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

a gamma distribution with mean τ_1/τ_2 and variance $2\tau_1/\tau_2^2$. The HNER model given in (1.1) and (1.2) 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 $\omega = (\boldsymbol{\beta}, \lambda, \boldsymbol{\tau})$ for $\boldsymbol{\tau} = (\tau_1, \tau_2)$. Then, the RHNER model is given by

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

The conditional distribution of v_i given $\boldsymbol{y}_i$ and η_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 η_i or σ_i^2.

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

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

The conditional expectation of ξ_i given $\boldsymbol{y}_i$ and η_i is

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

This is interpreted as the Bayes estimator of ξ under squared error loss. Since it does not depend on η_i, the estimator $\hat{\xi}_i^B(\boldsymbol{\beta}, \lambda)$ continues to be the conditional expectation of ξ_i given $\boldsymbol{y}_i$ after integrating out the η_i, that is the Bayes estimator of ξ_i is the same in the two situations. However, the empirical Bayes estimators, which substitute estimators of $\boldsymbol{\beta}$ and λ 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 $\boldsymbol{\beta}$, λ and $\sigma_1^2, \ldots, \sigma_m^2$. Noting that the number of parameters increases as m increases and that n_is are bounded in small-area estimation, we are faced with the problem of consistency and instability of the estimators of σ_i^2. In this situation, Jiang and Nguyen (2012) established the surprising result that the MLEs of $\boldsymbol{\beta}$ and λ are consistent, which lead to the consistency of the EB $\hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$. However, there are no consistent estimators of the σ_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 $\boldsymbol{\beta}$, λ, τ_1 and τ_2, and we can provide their consistent estimators.

2.2 A motivation from estimation of dispersions

We give more detailed motivation from the estimation of the dispersion parameters σ_i^2 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

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

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

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

It is interesting to point out that $\hat{\sigma}_i^{2H}$ is an asymptotically unbiased estimator of σ_i^2. For the proof, we can use the fact that $n\overline{y}_i^2$ and $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2$ are mutually independent with $n\overline{y}_i^2/(1 + n\lambda) \sim \sigma_i^2 \chi_1^2$ and $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 \sim \sigma_i^2 \chi_{n-1}^2$. Then, from the consistency of $\hat{\gamma}^H$, it follows that $E[\hat{\sigma}_i^{2H}]$ converges to

$$\frac{1}{n} E\Big[\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 + n\gamma \overline{y}_i^2 \Big] = \frac{\sigma_i^2}{n} E[\chi_{n-1}^2 + \gamma(1 + n\lambda)\chi_1^2] = \sigma_i^2,$$

which shows that $\hat{\sigma}_i^{2H}$ is an asymptotically unbiased estimator of σ_i^2.

Although $\hat{\sigma}_i^{2H}$ is asymptotically unbiased, it is clear that $\hat{\sigma}_i^{2H}$ is not consistent when $m \to \infty$, but n is bounded. Thus, we need to modify $\hat{\sigma}_i^{2H}$ when n is small. For example, we look at the empirical Bayes estimator of ξ_i. In the simple case we treat here, we have $\xi_i = v_i$, and the EB of ξ_i is given by

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

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, we have a concern about the precision of the estimator $\hat{\sigma}_i^{2H}$. Since $\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2 \leq n\hat{\sigma}_i^{2H} \leq \sum_{j=1}^n y_{ij}^2$, it is seen that

$$\frac{\overline{y}_i^2}{\sum_{j=1}^n y_{ij}^2 / n} \leq \frac{\overline{y}_i^2}{\hat{\sigma}_i^{2H}} \leq \frac{\overline{y}_i^2}{T_i / n},$$

for $T_i = \sum_{j=1}^n (y_{ij} - \overline{y}_i)^2$. When n is small, clearly $\overline{y}_i^2 / \hat{\sigma}_i^{2H}$ has a large variation, which leads to the instability of the empirical Bayes estimator $\hat{\xi}_i^H$. Although the simple case of the equal replications n is considered here, in the survey data we need to handle the case of small sample sizes n_i's for some small-areas, and the estimators $\hat{\sigma}_i^{2H}$'s cannot have enough degrees of freedom.

To overcome this drawback, we need to stabilize $\hat{\sigma}_i^{2H}$ by shrinking it to a point. The random dispersion model is helpful for the purpose. Assume that $\eta_i = 1/\sigma_i^2$ has a distribution $\mathcal{G}a(\tau_1/2, 2/\tau_2)$. Since $T_i\eta_i = T_i/\sigma_i^2 \sim \chi^2_{n-1}$, from the joint distribution of (T_i, η_i), the posterior mean of σ_i^2 given T_i is

$$E[\sigma_i^2|T_i] = (T_i + \tau_2)/(n - 1 + \tau_1).$$

When $\hat{\tau}_1$ and $\hat{\tau}_2$ are estimators of τ_1 and τ_2 based on the statistics $T_1, \ldots, T_m$, it is reasonable to estimate σ_i^2 by

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

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

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

Another need for a consistent estimator of σ_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

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

for large m. To estimate the uncertainty of $\hat{\xi}_i^H$, we want to estimate the leading term of the MSE consistently. Since $\hat{\sigma}_i^{2H}$ 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

We now return back to the RHNER model given in (2.2). When λ and $\boldsymbol{\beta}$ are known, the best predictor or the Bayes estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ is given by

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

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

The joint 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 joint likelihood with respect to v_i's is expressed as

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

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,

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

where $\psi(a) = \log(\Gamma(a))$. Let $L_{\boldsymbol{\beta}}$, L_λ, L_{τ_1} and L_{τ_2} be the derivatives of L with respect to $\boldsymbol{\beta}$, λ, τ_1 and τ_2. Then,

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

where $\partial Q_i/\partial \lambda = -n_i^2\gamma_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2$ for $\partial \gamma_i/\partial \lambda = -n_i\gamma_i^2$, and

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

The MLEs of $\boldsymbol{\beta}$, λ, τ_1 and τ_2 are the solution of the simultaneous equations $L_{\boldsymbol{\beta}} = 0$, $L_\lambda = 0$, $L_{\tau_1} = 0$ and $L_{\tau_2} = 0$, and the MLEs are denoted by $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$ and $\hat{\tau}_2$. The empirical Bayes estimator of $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ is provided by

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

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 their asymptotic variances and covariances are expressed by the Fisher information matrix, we begin by deriving the Fisher information. Let $\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ be the Fisher information matrix of $\boldsymbol{\beta}$. The Fisher information matrix of $\boldsymbol{\theta} = (\lambda, \tau_1, \tau_2)^T$ and the inverse are denoted by

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

Then, exact expressions of the Fisher information matrices can be derived in the following theorem. The proof is given in the Appendix.

Theorem 3.1 *The Fisher information of $\boldsymbol{\beta}$ is given by*

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

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 $2\boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ are given by

$$2I_{\lambda\lambda} = \sum_{i=1}^{m}\frac{(n_i+\tau_1-1)n_i^2\gamma_i^2}{n_i+\tau_1+2}, \quad 2I_{\lambda\tau_1} = -\sum_{i=1}^{m}\frac{n_i\gamma_i}{n_i+\tau_1},$$

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

$$2I_{\tau_1\tau_2} = -\frac{1}{\tau_2}\sum_{i=1}^{m}\frac{n_i}{n_i+\tau_1}, \quad 2I_{\tau_2\tau_2} = \frac{\tau_1}{\tau_2^2}\sum_{i=1}^{m}\frac{n_i}{n_i+\tau_1+2}.$$

It follows from Theorem 3.1 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 following theorem is essential for approximating the MSE of $\hat{\xi}_i^{EB}$ asymptotically. The proof is given in the Appendix.

Theorem 3.2 *Assume condition* (A1)*. Then, for* $\widehat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\tau}_1, \hat{\tau}_2)^T$*, it holds that*

$$
\begin{aligned}
E[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T|\boldsymbol{y}_i] =&(\boldsymbol{I}_{\boldsymbol{\beta\beta}})^{-1} + O_p(m^{-3/2}),\\
E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T|\boldsymbol{y}_i] =&(\boldsymbol{I}_{\boldsymbol{\theta\theta}})^{-1} + O_p(m^{-3/2}),\\
E[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T|\boldsymbol{y}_i] =&O_p(m^{-3/2})0
\end{aligned}
\tag{3.8}
$$

This implies that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{y}_i = O_p(m^{-1/2})$ *and* $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}|\boldsymbol{y}_i = O_p(m^{-1/2})$*. Also, the conditional biases* $E[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{y}_i] = O(m^{-1})$ *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 approximation of the conditional and unconditional MSEs

We shall derive a second-order approximation of the MSE of the empirical Bayes (EB) estimator and its second-order unbiased estimator. Recall that we want to predict $\xi_i = \overline{\boldsymbol{x}}_i^T\boldsymbol{\beta} + v_i$ with EB $\hat{\xi}_i^{EB} = \hat{\xi}_i^B(\widehat{\boldsymbol{\beta}}, \hat{\lambda}) = \overline{\boldsymbol{x}}_i^T\widehat{\boldsymbol{\beta}} + \hat{v}_i(\widehat{\boldsymbol{\beta}}, \hat{\lambda})$. 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

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

and

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

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

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

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. Noting that the random dispersion model (2.2) is not a normal distribution, but close to a t-distribution, we observe that the above fact coincides with their assertion.

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, it should be noted that 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 λ, τ_1 and τ_2 are estimated consistently.

Theorem 4.1 *Under assumption* (A1), *the conditional MSE of* $\hat{\xi}_i^{EB}$ *is approximated as*

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

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

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

Proof. We shall 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

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

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

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

It follows from Theorem 3.2 that $E[\{\overline{\boldsymbol{x}}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2|\boldsymbol{y}_i] = E[\{\overline{\boldsymbol{x}}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}^2] + O_p(m^{-3/2}) = \overline{\boldsymbol{x}}_i^T(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}})^{-1}\overline{\boldsymbol{x}}_i + O_p(m^{-3/2})$, $E[(\hat{\lambda}-\lambda)^2|\boldsymbol{y}_i] = E[(\hat{\lambda}-\lambda)^2] + O_p(m^{-3/2}) = I^{\lambda\lambda} + O_p(m^{-3/2})$ and $E[(\hat{\lambda}-\lambda)\overline{\boldsymbol{x}}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})|\boldsymbol{y}_i] = E[(\hat{\lambda}-\lambda)\overline{\boldsymbol{x}}_i^T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})] + O_p(m^{-3/2}) = O_p(m^{-3/2})$. Thus, one gets

$$g_2^c(\omega|\boldsymbol{y}_i) = \gamma_i^2\overline{\boldsymbol{x}}_i^T\left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1}\overline{\boldsymbol{x}}_i + n_i^2\gamma_i^4(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 I^{\lambda\lambda} + O_p(m^{-3/2}). \tag{4.7}$$

Combining (4.3) and (4.7) yields the approximation given in (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

$$g_2(\omega) = \gamma_i^2\overline{\boldsymbol{x}}_i^T\left(\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\right)^{-1}\overline{\boldsymbol{x}}_i + n_i\gamma_i^3\frac{\tau_2}{\tau_1-2}I^{\lambda\lambda} + O(m^{-3/2}). \tag{4.8}$$

Combining (4.4) and (4.8) gives the expression given in (4.6). Therefore, the proof of Theorem 4.1 is complete. □

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

We now derive second-order unbiased estimators of the unconditional and conditional MSEs. Since it is hard to derive second-order biases of the MLEs of $\boldsymbol{\beta}$, λ, τ_1 and τ_2, we could not provide analytical second-order unbiased estimators of the MSEs. Instead, we use the 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

$$y_{ij}^* = \boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + v_i^* + \varepsilon_{ij}^*, \quad i=1,\ldots,m;\ j=1,\ldots,n_i, \tag{4.9}$$

where v_i^*'s and ε_{ij}^*'s are conditionally mutually independent given η_i^*'s and

$$\begin{aligned}
v_i^*|\eta_i^* \sim& \mathcal{N}(0,\hat{\lambda}/\eta_i^*), \\
\varepsilon_{ij}^*|\eta_i^* \sim& \mathcal{N}(0,1/\eta_i^*), \\
\eta_i^* \sim& \sim \mathcal{G}a(\hat{\tau}_1/2, 2/\hat{\tau}_2).
\end{aligned} \tag{4.10}$$

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

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

Proposition 4.1 *Assume the condition* (A1). *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. Since $\boldsymbol{y}_i = (y_{i1},\ldots,y_{in_i})^T$ is fixed, a bootstrap sample $\boldsymbol{y}_k = (y_{k1},\ldots,y_{kn_k})^T$ is generated from (4.9) for $k \neq i$. Noting that $\boldsymbol{y}_i$ is fixed, we construct the estimators $\widehat{\boldsymbol{\beta}}^*_{(i)}$, $\hat{\lambda}^*_{(i)}$, $\hat{\tau}^*_{1(i)}$ and $\hat{\tau}^*_{2(i)}$ from the $\boldsymbol{y}_i$ and the bootstrap sample

$$\boldsymbol{y}_1^*,\ldots,\boldsymbol{y}_{i-1}^*,\boldsymbol{y}_i,\boldsymbol{y}_{i+1}^*,\ldots,\boldsymbol{y}_m^* \tag{4.11}$$

with the same technique as used to obtain the estimator $\widehat{\boldsymbol{\beta}}$, $\hat{\lambda}$, $\hat{\tau}_1$ and $\hat{\tau}_2$. Let $E_*[\cdot|\boldsymbol{y}_i]$ be the expectation with regard to the bootstrap sample (4.11). 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

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

Then, it can be verified that $E[\hat{g}_1^{c*}|\boldsymbol{y}_i] = g_1^c(\omega|\boldsymbol{y}_i) + o_p(m^{-1})$. $g_2^c(\omega|\boldsymbol{y}_i)$, is estimated via parametric bootstrap as

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

Thus,

$$cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i) = \hat{g}_1^{c*} + \hat{g}_2^{c*}. \tag{4.12}$$

Theorem 4.2 *Under the condition* (A1), *the estimator* (4.12) *is a second-order unbiased estimator of cMSE, namely*

$$E[cmse^*(\hat{\xi}_i^{EB}|\boldsymbol{y}_i)|\boldsymbol{y}_i] = cMSE(\omega;\hat{\xi}_i^{EB}|\boldsymbol{y}_i) + o_p(m^{-1}).$$

5 Numerical and Empirical Studies

In this section, we investigate performances of the procedures suggested in the previous sections through the numerical and empirical studies.

5.1 Simulation study

We here investigate finite sample performances of the maximum likelihood (ML) estimators in the RHNER model and the second-order unbiased estimators for the conditional and unconditional MSEs by the Monte Carlo simulation.

The ML estimators $\boldsymbol{\beta}$ and λ 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 examine finite sample performances of the estimators (3.5) in RHNER and compare them with the estimators in HNER in light of the mean squared errors (MSE). To this end, we conduct 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 x_{ij}'s are generated from $\mathcal{N}(0, 1)$, and these values are fixed throughout the simulation runs. In this simulation, the true values of $\boldsymbol{\beta}$ and λ are $\beta_0 = \beta_1 = \lambda = 1$. For (τ_1, τ_2), we treat 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 η_i are (2, 1) for Case I and (12, 96) for Case II. This means that the generated values of $\sigma_i^2 = 1/\eta_i$ in Case I are more variable than those in Case II.

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

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

for $K = 1,000$. Those values for 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 observed from Table 1 that the values 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, which seems to be due to the property that the estimates of the variances σ_i^2 are more stable in RHNER than in HNER. Concerning the difference between Case I and Case II, the variability of η_i or σ_i^2 affects the estimates of β_0 and β_1, but does not affect the estimates for λ very much.

We next investigate finite sample performances for the estimators of conditional and unconditional MSEs suggested in Section 4.2. For simplicity, we here treat 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 are $\mu = 0$, $\lambda = 1$, $\tau_1 = 8$ and $\tau_2 = 4$, namely, the true values of (τ_1, τ_2) correspond to Case II in the previous simulation. For the design of n_i, we consider

$$n_{1+m(k-1)/5}, \ldots, n_{mk/5} = k, \quad k = 1, \ldots, 5,$$

which means that m areas are divided into five groups and that areas in each group have the same sample size n_i.

Table 1: Mean Squared Errors of the Maximum Likelihood Estimators of $(\beta_0, \beta_1, \lambda)$ in the RHNER Model for (Case I) $(\tau_1, \tau_2) = (8, 4)$ and (Case II) $(\tau_1, \tau_2) = (3, 1/4)$. (Values on β_0 and β_1 are multiplied by 100. Values in the parentheses denote percentages of improvement over the estimators in the HNER model.)

Case	Size	β_0	β_1	λ
	$m = 10$	8.85 (50.1)	2.22 (57.6)	0.87 (18.8)
I	$m = 20$	4.42 (59.5)	1.71 (61.7)	0.41 (21.9)
	$m = 30$	2.63 (63.6)	0.88 (63.6)	0.26 (18.8)
	$m = 10$	1.54 (48.3)	0.70 (40.4)	1.01 (10.1)
II	$m = 20$	0.79 (56.4)	0.29 (57.0)	0.44 (18.9)
	$m = 30$	0.53 (54.8)	0.26 (56.8)	0.29 (21.4)

Concerning the unconditional MSEs, their true values are calculated via simulation with $R = 5,000$ replications by calculating the quantity

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

where $\hat{\xi}_i^{EB(r)}$ and $\hat{\xi}_i^{B(r)}$ are the empirical Bayes and Bayes estimator of ξ_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) are calculated based on $T = 1,000$ simulation runs with each 100 bootstrap samples, where RB is defined as

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

for the MSE estimate $\widehat{\mathrm{MSE}}_t$ 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 of $\widehat{\mathrm{MSE}}$ are close to the true values of MSE, and their relative bias 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 the estimators of variances in each small areas in RHNER.

Concerning the conditional MSE, we use the same setup as in the simulation on the unconditional MSE except for $n_i = 3$ for sample sizes in all areas. Without any loss of generality, it is assumed that values of y_{1j} in the area 1 are given. As conditioning values for y_{1j}, we use α-quantile points of the marginal distribution of y_{1j}, denoted by $y_{1j(\alpha)}$, and

Table 2: Mean and Relative Bias of the Estimators for the Unconditional MSE

	$m = 20$			$m = 50$		
n_i	MSE	$\widehat{\text{MSE}}$	RB	MSE	$\widehat{\text{MSE}}$	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

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 calculate the values of $\widehat{\xi}_1^{EB(r)}$ and $\widehat{\xi}_1^{B(r)}$. Then, the true values of the conditional MSE of $\widehat{\xi}_1^{EB(r)}$ are numerically calculated as

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

As the same manner as stated above, we calculate conditional MSE estimates and their relative biases based in $1,000$ simulation runs with each 100 bootstrap samples. The results by simulation 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 when both $m = 20$ and 50.

5.2 Application to PLP data in Japan

We now investigate empirical performances of the suggested model, the empirical Bayes estimator and the second-order unbiased estimators of the conditional and unconditional MSEs through analysis of real data. The data treated here is 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, namely, $m = 52$. 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 in 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

Table 3: Mean and Relative Bias of the Estimators for the Conditional MSE

Areas	α	$y_{1j(\alpha)}$	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

1, which clearly indicates that the posted land price has the large within-area variation and the conventional NER model (which assumes homogeneity of variance) does not seem to be appropriate.

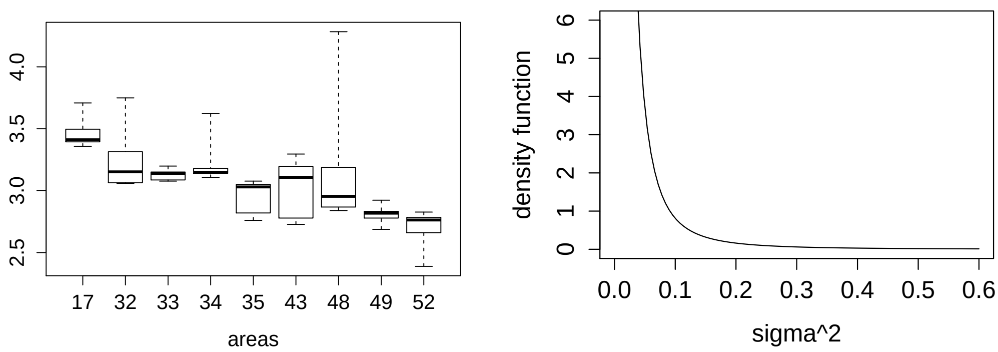

Figure 1: Boxplots of the Posted Land Price Data for Selected Areas (left) and the Estimated Density Function of $\sigma_i^2 = 1/\eta_i$ (right).

For $j = 1, \ldots, n_i$, y_{ij} denotes the value which is transformed by logarithm from 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. These 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 ε_{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 $\Gamma(\tau_1/2, 2/\tau_2)$.

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 η_i has a large mean about 73 and a heavy tail. Since the estimated value of τ_1 is smaller than 4, the variance of η_i or σ_i^2 does not exist, which agrees the observation that the posted land price data has great variability as indicated by the boxplots in Figure 1. Figure 1 (right) draws the estimated density function of $\sigma_i^2 = 1/\eta_i$ where η_i has $\Gamma(\hat{\tau}_1/2, 2/\hat{\tau}_2)$, so that the distribution of σ_i^2 has a small mean, but a heavy tail.

The predicted values of $\overline{\boldsymbol{x}}_i'\boldsymbol{\beta} + v_i$ and their conditional and unconditional MSE estimates, which can be obtained 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 to point out that, in area 48, the estimated conditional MSE is relatively large while the estimated unconditional MSE is not large. Noting that this area has great variability as shown in Figure 1, it seems that the conditional MSE can capture the variability of areas.

6 Concluding Remarks

In the context of small-area estimation, homogeneous linear mixed models like the nested error regression (NER) model have been studied so far in the literature. 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 which assumes that the within-area variances are different among areas. Motivated from the inconsistency of the MLE of the dispersion σ_i^2, this paper suggests the random dispersion heteroscedastic nested error regression (RHNER) 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 empirical Bayes estimator, the conditional and unconditional mean squared errors (MSE) have been approximated up to second-order, and their second-order unbiased estimators have been provided based on the parametric bootstrap method. Although the difference between the cMSE and MSE is quite small and

Table 4: Values of EBLUP and Estimates of Unconditional and Conditional MSES for Selected fifteen Areas. (Estimates of MSE and cMSE are multiplied by 100)

Area	n_i	EBLUP	$\widehat{\text{MSE}}$	$\widehat{\text{cMSE}}$
1	1	4.02	5.19	0.58
4	2	3.91	4.34	0.49
5	5	3.96	3.13	2.31
8	3	3.86	3.83	0.33
17	7	3.50	2.66	1.04
25	7	3.39	2.65	1.37
26	4	3.45	3.42	1.88
32	6	3.22	2.86	2.68
33	8	3.12	2.48	1.90
34	11	3.16	2.09	1.10
35	7	2.99	2.65	3.58
43	6	3.02	2.86	3.73
48	6	3.07	2.86	5.11
49	10	2.82	2.21	2.69
52	6	2.76	2.87	6.55

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 one of future studies, it is interesting to construct a confidence interval of $\xi_i = \overline{\boldsymbol{x}}_i^T \boldsymbol{\beta} + v_i$. 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

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

Thus, the conditional distribution of Z_i given $\boldsymbol{y}_i$ can be expressed as $Z_i|\boldsymbol{y}_i \sim f_i(z_i|\tau_1)$, where the density function of Z_i is given by

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

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

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

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 τ_1, it is computationally hard to construct a confidence interval with second-order accuracy. This issue will be addressed in a future work.

Another interesting topic is that one can consider a different type of a random dispersion model given by

$$\begin{aligned} \boldsymbol{y}_i|v_i, \eta_i \sim& \mathcal{N}_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_{n_i}v_i, \eta_i^{-1}\boldsymbol{I}_{n_i}), \\ v_i \sim& \mathcal{N}(0, \sigma_v^2), \\ \eta_i \sim& \mathcal{G}a(\tau_1/2, 2/\tau_2). \end{aligned} \tag{6.3}$$

In this model, the random effect v_i is distributed independent of η_i or σ_i^2, namely, the variance of $y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - v_i$ changes from area to area, while the variance of $y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$ varies from area to area in the RHNER model (2.2). Since the integration in the marginal likelihood (3.4) with respect to η_i cannot give a closed form in the model (6.3), it may be harder to analyze this model, but it is worth trying it in the future.

A Appendix

A.1 Proof of Theorem 3.1

We begin by providing the following two lemmas which will be used for calculating the Fisher information.

Lemma A.1 *Conditional on η_i, Q_i defined below* (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. Let us assume that $\boldsymbol{\beta}$ and λ are fixed. Given the joint pdf in (3.2), conditional on η_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. Now we apply Basu's theorem to get Lemma A.1. □

Lemma A.2 *Let $R_i = Q_i/(Q_i + \tau_2)$. Marginally, that is, after integrating out η_i, $R_i \sim Beta(n_i/2, \tau_1/2)$.*

Proof. Note that $Q_i|\eta_i \sim \eta_i^{-1}\chi^2_{n_i}$ and $\eta_i \sim \mathcal{G}a(\tau_1/2, 2/\tau_2)$. Now integrating out η_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. □

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 will be using (A.1), (A.2) and (A.3) repeatedly in the following calculations.

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

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

But,

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

for $\gamma_i = 1/(n_i\lambda+1)$. It is noted that

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

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 η_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

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

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 $(e_{i1},\dots,e_{in_i},v_i)$ and $-(e_{i1},\dots,e_{in_i},v_i)$ have the same distribution and $(e_{i1}-\bar{e}_i,\dots,e_{in_i}-\bar{e}_i)$ is distributed independently of $(v_i,\bar{e}_i)$ conditional on η_i, it follows that

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

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

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

Next, observe that

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

Arguing as belore, $(e_{i1},\dots,e_{in_i},v_i)$ and $-(e_{i1},\dots,e_{in_i},v_i)$ have the same distribution and $(e_{i1}-\bar{e}_i,\dots,e_{in_i}-\bar{e}_i)$ is distributed independently of $(v_i,\bar{e}_i)$ conditional on η_i, it follows that $\boldsymbol{I}_{\boldsymbol{\beta}\lambda} = -E[L_{\boldsymbol{\beta}\lambda}] = \mathbf{0}$. Similarly,

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

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

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

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

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

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

It is here observed that $(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2 n_i\gamma_i/Q_i \sim Beta(1/2, n_i/2)$ and is distributed independently of Q_i. Further,

$$
\begin{aligned}
E[n_i\gamma_i(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^2/Q_i] =&1/n_i, \qquad (A.9)\\
E[(n_i\gamma_i)^2(\overline{y}_i-\overline{\boldsymbol{x}}_i^T\boldsymbol{\beta})^4/Q_i^2] =&3/\{n_i(n_i+2)\}. \qquad (A.10)
\end{aligned}
$$

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

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

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 that

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

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

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

Finally, it is observed 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,

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

Also, 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 $I_{\tau_2\tau_2} = E[-L\tau_2\tau_2] = \tau_1(2\tau_2^2)^{-1}\sum_{i=1}^m n_i/(n_i+\tau_1+2)$.

A.2 Proof of Theorem 3.2

Let $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_{p+3})^T = (\boldsymbol{\beta}^T, \lambda, \boldsymbol{\tau}^T)^T$. The log likelihood of $(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)$ is denoted by $\ell(\boldsymbol{\omega})$, which is also expressed as

$$\ell(\boldsymbol{\omega}) = \sum_{j=1}^m \ell(\boldsymbol{\omega};\boldsymbol{y}_j),$$

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 $\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega}) = (\partial^2/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T)\ell(\boldsymbol{\omega})$. Then, 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 $I_{ab}(\boldsymbol{\omega}) = m^{-1}\sum_{j=1}^m E[-\ell_{ab}(\boldsymbol{\omega};\boldsymbol{y}_j)]$. It is noted that

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

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 $\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega}) = -E[\ell_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})]$. Also, it holds that 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 the 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

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

Hence, it follows that

$$E[(\widehat{\boldsymbol{\omega}}-\boldsymbol{\omega})(\widehat{\boldsymbol{\omega}}-\boldsymbol{\omega})^T|\boldsymbol{y}_i] = \big\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\big\}^{-1}\frac{1}{m^2}E[\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\big\{\ell_{\boldsymbol{\omega}}(\boldsymbol{\omega})\big\}^T|\boldsymbol{y}_i]\big\{m^{-1}\boldsymbol{I}_{\boldsymbol{\omega}\boldsymbol{\omega}}(\boldsymbol{\omega})\big\}^{-1} + O_p(m^{-3/2}).$$

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

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

which implies that

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

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 shows (3.8) in Theorem 3.2. 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

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

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 shows the second part of Theorem 3.2. Therefore, the proof is complete.

Acknowledgments.

Research of the first author was supported in part by Grant-in-Aid for Scientific Research (23243039 and 26330036) from Japan Society for the Promotion of Science.

References

[1] 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.

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

[3] Butar, F.B. and Lahiri, P. (2003). On measures of uncertainty of empirical Bayes small-area estimators. *J. Statist. Plan. Inf.*, **112**, 63-76.

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

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

[6] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

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

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

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

[10] Pfeffermann, D. (2002). Small area estimation: new developments and directions. *International Statistical Institute Review*, **70**, 125-143.

[11] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.

[12] Rao, J.N.K. (2003). *Small Area Estimation*. Wiley.

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