Heteroscedastic Nested Error Regression Models with Variance Functions

Shonosuke Sugasawa* and Tatsuya Kubokawa†

Abstract

The article considers a nested error regression model with heteroscedastic variance functions for analyzing clustered data, where the normality for the underlying distributions is not assumed. Classical methods in normal nested error regression models with homogenous variances are extended in the two directions: heterogeneous variance functions for error terms and non-normal distributions for random effects and error terms. Consistent estimators for model parameters are suggested, and second-order approximations of their biases and variances are derived. The mean squared errors of the empirical best linear unbiased predictors are expressed explicitly to second-order. Second-order unbiased estimators of the mean squared errors are provided analytically in closed forms. The proposed model and the resulting procedures are numerically investigated through simulation and empirical studies.

1 Introduction

Linear mixed models and the model-based estimators including empirical Bayes (EB) estimator or empirical best linear unbiased predictor (EBLUP) have been studied quite extensively in the literature from both theoretical and applied points of view. Of these, the small area estimation (SAE) is an important application, and methods for SAE have received much attention in recent years due to growing demand for reliable small area estimates. For a good review and account on this topic, see Ghosh and Rao (1994), Rao (2003), Datta and Ghosh (2012) and Pfeffermann (2014). The linear mixed models used for SAE are the Fay-Herriot model suggested by Fay and Herriot (1979) for area-level data and the nested error regression (NER) models given in Battese, Harter and Fuller (1988) for unit-level data. Especially, the NER model has been used in application of not only SAE but also biological experiments and econometric analysis. Besides the noise, a source of variation is added to explain the correlation among observations within clusters, or subjects, and to allow the analysis to 'borrow strength' from other clusters. The resulting estimators, such as EB or EBLUP, for small-cluster means or subject-specific values provide reliable estimates with higher precisions than direct estimates like sample means.

In the NER model with m small-clusters, let $(y_{i1}, \boldsymbol{x}_{i1}), \ldots, (y_{in_i}, \boldsymbol{x}_{in_i})$ be n_i individual observations from the i-th cluster for $i = 1, \ldots, m$, where $\boldsymbol{x}_{ij}$ is a p-dimensional known vector of covariates. The normal NER model is written as

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i,$$

*Graduate School of Economics, University of Tokyo

†Faculty of Economics, University of Tokyo

where v_i and ε_{ij} denote the random effect and samping error, respectively, and they are mutually independently distributed as $v_i \sim N(0,\tau^2)$ and $\varepsilon_{ij} \sim N(0,\sigma^2)$. The mean of y_{ij} is $\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ for regression coefficients $\boldsymbol{\beta}$, and the variance of y_{ij} is decomposed as

$$\mathrm{Var}(y_{ij}) = E[(y_{ij} - \boldsymbol{x}_i'\boldsymbol{\beta})^2] = \tau^2 + \sigma^2. \tag{1}$$

which is the same for all the clusters. However, Jiang and Nguyen (2012) illustrated that the within-cluster sample variances change dramatically from cluster to cluster for the data given in Battese, et al. (1988). Also, the normality assumptions for random effects and error terms are not always appropriate in practice. Thus, we want to address the issue of relieving these assumptions of normal NER models in the two directions: heterogeneity of variances and non-normality of underlying distributions.

In real application, we often encounter the situation where the sampling variance $\mathrm{Var}(\varepsilon_{ij})$ is affected by the covariate $\boldsymbol{x}_{ij}$. In such case, the variance function is a useful tool for describing its relationship. Variance function estimation has been studied in the literature in the framework of heteroscedastic nonparametric regression. For example, see Cook and Weisberg (1983), Hall and Carroll (1989), Muller and Stadtmuller (1987, 1993) and Ruppert, Wand, Holst and Hossjer (1997). Thus, in this paper, we propose use of the technique to introduce the heteroscedastic variances into NER model without assuming normality of underlying distributions.

The variance structure we consider is

$$\mathrm{Var}(y_{ij}) = \tau^2 + \sigma_{ij}^2, \tag{2}$$

namely, the setup means that the sampling error ε_{ij} has heteroscedastic variance $\mathrm{Var}(\varepsilon_{ij}) = \sigma_{ij}^2$. Then we suggest the variance function model given by $\sigma_{ij}^2 = \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$, where the details are explained in Section 2.

Related to this paper, Jiang and Nguyen (2012) proposed the heteroscedastic nested error regression model with the setup that variance $\mathrm{Var}(y_{ij})$ is proportional to σ_i^2, namely

$$\mathrm{Var}(y_{ij}) = (\lambda + 1)\sigma_i^2. \tag{3}$$

This is equivalent to the assumption that $\mathrm{Var}(v_i) = \lambda\sigma_i^2$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma_i^2$. For setup (3), Jiang and Nguyen (2012) assumed normality for v_i and ε_{ij} and demonstrated the quite interesting result that the maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and λ are consistent for large m, which implies that the resulting empirical Bayes estimator estimates the Bayes estimator consistently. In setup (3), however, there is no consistent estimator for the heteroscedastic variance σ_i^2, and the mean squared error (MSE) of the EB cannot be estimated consistently, since it depends on σ_i^2. To fix the inconsistent estimation of σ_i^2, Maiti, Ren and Sinha (2014) suggested the hierarchical model such that σ_i^2's are random variables and σ_i^{-2} has a gamma distribution. Maiti, et al. (2014) applied this setup to the Fay-Herriot model with statistics for estimating σ_i^2. However, the resulting EB estimator and the MSE can not be expressed in closed forms. The same setup of σ_i^2 was used recently by Kubokawa, Sugasawa, Ghosh and Choudhuri (2014) who derived explicit expressions of the EB estimator and the MSE to second-order. In their simulation study, however, the finite sample properties of estimators of two hyper-parameters in the gamma prior distribution of σ_i^2 are not so well. Although the hierarchical models used in Maiti, et al. (2014) and Kubokawa, et al. (2014) provide consistent estimators for model parameters and predictors, both models assume parametric hierarchical structures based on normal distributions of v_i and ε_{ij}. However, the normality assumption is not always appropriate and another heteroscedastic models are useful for such a situation when the normality assumption does not seem to be correct.

In contrast to the existing results, the proposed model with variance function does not assume normality for either v_i nor ε_{ij}. The advantage of this paper is that the MSE of the EB or EBLUP and its unbiased estimator are derived analytically in closed forms up to second-order without assuming normality for v_i and ε_{ij}. Nonparametric approach to SAE has been studied by Jiang, Lahiri and Wan (2002), Hall and Maiti (2006), Lohr and Rao (2009) and others. Most estimators of the MSE have been given by numerical methods such as Jackknife and bootstrap methods except for Lahiri and Rao (1995), who provided an analytical second-order unbiased estimator of the MSE in the Fay-Heriot model. Hall and Maiti (2006) developed a moment matching bootstrap method for nonparametric estimation of MSE in nested error regression models. The suggested method is actually convenient but it requires bootstrap replication and has computational burden. In this paper, without assuming the normality, we derive not only second-order biases and variances of estimators for the model parameters, but also a closed expression for a second-order unbiased estimator of the MSE in a closed form. Thus our MSE estimator does not require any resampling method and is useful in practical use. Also our MSE estimator can be regarded as a generalization of the robust MSE estimator given in Lahiri and Rao (1995).

The paper is organized as follows: A setup of the proposed HNER model and estimation strategy with asymptotic properties are given in Section 2. In Section 3, we obtain the EBLUP and the second-order approximation of the MSE. Further, we provide the second-order unbiased estimators of MSE by the analytical calculation. In Section 4, we investigate the performance of the proposed procedures through simulation and empirical studies. The technical proofs are given in the Appendix.

2 HNER Models with Variance Functions

2.1 Model settings

Suppose that there are m small clusters, and let $(y_{i1}, \boldsymbol{x}_{i1}), \ldots, (y_{in_i}, \boldsymbol{x}_{in_i})$ be the pairs of n_i observations from the i-th cluster, where $\boldsymbol{x}_{ij}$ is a p-dimensional known vector of covariates. We consider the heteroscedastic nested error regression model

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m, \tag{4}$$

where $\boldsymbol{\beta}$ is a p-dimenstional unknown vector of regression coefficients, and v_i and ε_{ij} are mutually independent random variables with mean zero and variances $\mathrm{Var}(v_i) = \tau^2$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma_{ij}^2$, which are denoted by

$$v_i \sim (0, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \sim (0, \sigma_{ij}^2). \tag{5}$$

It is noted that no specific distributions are assumed for v_i and ε_{ij}. It is assumed that the heteroscedastic variance σ_{ij}^2 of ε_{ij} is given by

$$\sigma_{ij}^2 = \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}), \quad i = 1, \ldots, m, \tag{6}$$

where $\boldsymbol{z}_{ij}$ is a q-dimensional known vector given for each cluster, and $\boldsymbol{\gamma}$ is a q-dimensional unknown vector. The variance function $\sigma^2(\cdot)$ is a known (user specified) function whose range is nonnegative. Some examples of the variance function are given below. The model parameters are $\boldsymbol{\beta}$, τ^2 and $\boldsymbol{\gamma}$, whereas the total number of the model parameters is $p + q + 1$.

Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})'$ and $\boldsymbol{\epsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})'$. Then the model (4) is expressed in a vector form as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + v_i\boldsymbol{1}_{n_i} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, m,$$

where $\mathbf{1}_n$ is an $n \times 1$ vector with all elements equal to one, and the covariance matrix of $\boldsymbol{\epsilon}_i$ is

$$\boldsymbol{\Sigma}_i = \mathrm{Var}(\boldsymbol{y}_i) = \tau^2 \boldsymbol{J}_{n_i} + \boldsymbol{W}_i,$$

for $\boldsymbol{J}_{n_i} = \mathbf{1}_{n_i}\mathbf{1}_{n_i}'$ and $\boldsymbol{W}_i = \mathrm{diag}(\sigma_{i1}^2, \ldots, \sigma_{in_i}^2)$. It is noted that the inverse of $\boldsymbol{\Sigma}_i$ is expressed as

$$\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{W}_i^{-1}\left(\boldsymbol{I}_{n_i} - \frac{\tau^2 \boldsymbol{J}_{n_i}\boldsymbol{W}_i^{-1}}{1 + \tau^2 \sum_{j=1}^{n_i} \sigma_{ij}^{-2}}\right),$$

where $\boldsymbol{W}_i^{-1} = \mathrm{diag}(\sigma_{i1}^{-2}, \ldots, \sigma_{in_i}^{-2})$. Further, let $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_m')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_m')'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \ldots, \boldsymbol{\epsilon}_m')'$ and $\boldsymbol{v} = (v_1\mathbf{1}_{n_1}', \ldots, v_m\mathbf{1}_{n_m}')'$. Then, the matricial form of (4) is written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon}$, where $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma} = \text{block diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m)$.

Now we give some examples of the variance function $\sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$ in (6).

(a) In the case that the dispersion of the sampling error is proportional to the mean, it is reasonable to put $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}^{(s)}$ and $\sigma^2(\boldsymbol{x}_{(s)ij}'\boldsymbol{\gamma}) = (\boldsymbol{x}_{(s)ij}'\boldsymbol{\gamma})^2$ for the sub-vector $\boldsymbol{x}_{(s)ij}'$ of the covariate $\boldsymbol{x}_{ij}$. For identifiability of $\boldsymbol{\gamma}$, we restrict $\gamma_1 > 0$.

(b) Consider the case that m clusters are decomposed into q homogeneous groups $S_1, \ldots, S_q$ with $\{1, \ldots, m\} = S_1 \cup \ldots \cup S_q$. Then, we put

$$\boldsymbol{z}_{ij} = \left(1_{\{i \in S_1\}}, \ldots, 1_{\{i \in S_q\}}\right)',$$

which implies that

$$\sigma_{ij}^2 = \gamma_t^2 \quad \text{for} \quad i \in S_t.$$

Note that $\mathrm{Var}(y_{ij}) = \tau^2 + \gamma_t^2$ for $i \in S_t$. Thus, the models assumes that the m clusters are divided into known q groups with their variance are equal over the same groups. Jiang and Nguyen (2012) used a similar setting and argued that the unbiased estimator of the heteroscedastic variance is consistent when $|S_k| \to \infty, k = 1, \ldots, q$ as $m \to \infty$, where $|S_k|$ denotes the number of elements in S_k.

(c) Log linear functions of variance were treated in Cook and Weisberg (1983) and others. That is, $\log \sigma_{ij}^2$ is a linear function, and σ_{ij}^2 is written as $\sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}) = \exp(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$. Similarly to (a), we put $\boldsymbol{z}_{ij} = \boldsymbol{x}_{(s)ij}$.

For the above two cases (a) and (b), we have $\sigma^2(x) = x^2$, while the case (c) corresponds to $\log\{\sigma^2(x)\} = x$. In simulation and empirical studies in Section 4, we use the log-linear variance model. As given in subsequent section, we show consistency and asymptotic expression of estimators for $\boldsymbol{\gamma}$ as well as $\boldsymbol{\beta}$ and τ^2.

2.2 Estimation

We here provide estimators of the model parameters $\boldsymbol{\beta}$, τ^2 and $\boldsymbol{\gamma}$. When values of $\boldsymbol{\gamma}$ and τ^2 are given, the vector $\boldsymbol{\beta}$ of regression coefficients is estimated by the generalized least squares (GLS) estimator

$$\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\tau^2, \boldsymbol{\gamma}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \left(\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1} \sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i. \tag{7}$$

This is not a feasible form since $\boldsymbol{\gamma}$ and τ^2 are unknown. When estimators $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$ are for τ^2 and $\boldsymbol{\gamma}$, we get the feasible estimator $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\tau}^2, \widehat{\boldsymbol{\gamma}})$ by replacing τ^2 and $\boldsymbol{\gamma}$ in $\widetilde{\boldsymbol{\beta}}$ with their estimators.

Concerning estimation of τ^2, we use the second moment of observations y_{ij}'s. From model (4), it is seen that

$$E\left[(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2\right] = \tau^2 + \sigma^2(\boldsymbol{z}_{ij}'\gamma). \tag{8}$$

Based on the ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_{\text{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, a moment estimator of τ^2 is given by

$$\widehat{\tau}^2 = \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\text{OLS}})^2 - \sigma^2(\boldsymbol{z}_{ij}'\gamma)\right\}, \tag{9}$$

with substituting estimator $\widehat{\boldsymbol{\gamma}}$ into $\boldsymbol{\gamma}$, where $N = \sum_{i=1}^m n_i$.

For estimation of $\boldsymbol{\gamma}$, we consider the within difference in each cluster. Let $\bar{y}_i$ be the sample mean in the i-th cluster, namely $\bar{y}_i = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}$. It is noted that for $\bar{\varepsilon}_i = n_i^{-1}\sum_{j=1}^{n_i}\varepsilon_{ij}$,

$$y_{ij} - \bar{y}_i = (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{ij} - \bar{\varepsilon}_i),$$

which dose not include the term of v_i. Then it is seen that

$$E\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\right] = \left(1 - 2n_i^{-1}\right)\sigma^2(\boldsymbol{z}_{ij}'\gamma) + n_i^{-2}\sum_{h=1}^{n_i}\sigma^2(\boldsymbol{z}_{ih}'\gamma),$$

which motivates us to estimate $\boldsymbol{\gamma}$ by solving the following estimating equation given by

$$\frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\text{OLS}}\right\}^2 - \left(1 - 2n_i^{-1}\right)\sigma^2(\boldsymbol{z}_{ij}'\gamma) - n_i^{-2}\sum_{h=1}^{n_i}\sigma^2(\boldsymbol{z}_{ih}'\gamma)\right]\boldsymbol{z}_{ij} = \boldsymbol{0},$$

which is equivalent to

$$\frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\text{OLS}}\right\}^2\boldsymbol{z}_{ij} - \sigma^2(\boldsymbol{z}_{ij}'\gamma)(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right] = \boldsymbol{0} \tag{10}$$

where $\bar{\boldsymbol{z}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}$. It is noted that, In case of homoscedastic case, namely $\sigma^2(\boldsymbol{z}_{ij}'\gamma) = \delta^2$, the estimator δ^2 and τ^2 reduces to the estimator identical to Prasad-Rao estimator (Prasad and Rao, 1990) up to the constant factor.

Note that the objective function (10) for estimation of $\boldsymbol{\gamma}$ does not depend on $\boldsymbol{\beta}$ and τ^2 and that the estimator of τ^2 depends on $\boldsymbol{\gamma}$. These suggest the following algorithm for calculating the estimates of the model parameters: We first obtain the estimate $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ by solving (10), and then we get the estimate $\widehat{\tau}^2$ from (9) with $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$. Finally we have the GLS estimate $\widehat{\boldsymbol{\beta}}$ with substituting $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\tau}^2$ in (7).

2.3 Large sample properties

In this section, we provide large sample properties of the estimators given in the previous subsection when the number of clusters m goes to infinity, but n_i's are still bounded. To establish asymptotic results, we assume the following conditions under $m \to \infty$.

Assumption (A)

1. There exist $\underline{n}$ and $\overline{n}$ such that $\underline{n} \le n_i \le \overline{n}$ for $i = 1, \ldots, m$. The dimensions p and q are bounded, namely $p, q = O(1)$. The number of clusters with one observation, namely $n_i = 1$, is bounded.

2. The variance function $\sigma^2(\cdot)$ is twice differentiable and its derivatives are denoted by $(\sigma^2)^{(1)}(\cdot)$ and $(\sigma^2)^{(2)}(\cdot)$, respectively.

3. The following matrices converge to non-singular matrices:

$$m^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}', \quad m^{-1}\boldsymbol{X}'\boldsymbol{X}, \quad m^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\sigma^2)^{(a_1)}(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}', \quad m^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{a_2}\boldsymbol{X}$$

   for $a_1 = 1, 2$ and $a_2 = \pm 1$.

4. The forth moments of v_i and ε_{ij} exist, namely $E[v_i^4] < \infty$ and $E[\varepsilon_{ij}^4] < \infty$.

The conditions 1 and 3 are the standard assumptions in small area estimation. The condition 2 is also non-restrictive, and the simple variance function $\sigma^2(x) = x^2$ and $\sigma^2(x) = \exp(x)$ obviously satisfies the assumption. The moment condition 4 is necessary for existence of MSE of the EBLUP, and it is satisfied by many continuous distributions, including normal, shifted gamma, Laplace and t-distribution with degrees of freedom larger than 5.

In what follows, we use the notations

$$\sigma_{ij}^2 \equiv \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}), \quad \sigma_{ij(k)}^2 \equiv (\sigma^2)^{(k)}(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}), \quad k = 1, 2$$

for simplicity. To derive asymptotic approximations of the estimators, we define the following statistics in the i-th cluster:

$$u_{1i} = \frac{m}{N}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2 - \tau^2\right\}, \tag{11}$$

$$\boldsymbol{u}_{2i} = \frac{m}{N}\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right]. \tag{12}$$

Moreover, we define

$$\boldsymbol{T}_1(\boldsymbol{\gamma}) = \sum_{k=1}^{m}\sum_{h=1}^{n_k}\sigma_{kh(1)}^2\boldsymbol{z}_{kh}, \quad \boldsymbol{T}_2(\boldsymbol{\gamma}) = \left(\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kh(1)}^2(\boldsymbol{z}_{kh} - 2n_k^{-1}\boldsymbol{z}_{kh} + n_k^{-1}\bar{\boldsymbol{z}}_k)\boldsymbol{z}_{kh}'\right)^{-1}, \tag{13}$$

noting that $\boldsymbol{T}_1(\boldsymbol{\gamma}) = O(m)$ and $\boldsymbol{T}_2(\boldsymbol{\gamma}) = O(m^{-1})$ under Assumption (A). Then we obtain the asymptotically linear expression of the estimators.

Theorem 1. *Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\gamma}}', \widehat{\tau}^2)'$ be the estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \tau^2)'$. Under Assumption* (A)*, it follows that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(m^{-1/2})$ with the asymptotically linear expression*

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \frac{1}{m}\sum_{i=1}^{m}((\boldsymbol{\psi}_i^{\boldsymbol{\beta}})', (\boldsymbol{\psi}_i^{\boldsymbol{\gamma}})', \psi_i^{\tau})' + o_p(m^{-1/2}),$$

where

$$\boldsymbol{\psi}_i^{\boldsymbol{\gamma}} = N\boldsymbol{T}_2(\boldsymbol{\gamma})\boldsymbol{u}_{2i}, \quad \psi_i^{\tau} = u_{1i} - \boldsymbol{T}_1(\boldsymbol{\gamma})'\boldsymbol{T}_2(\boldsymbol{\gamma})\boldsymbol{u}_{2i}, \quad \boldsymbol{\psi}_i^{\boldsymbol{\beta}} = m\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}).$$

From Theorem 1, it follows that $m^{1/2}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ have an asymptotically normal distribution with mean vector $\mathbf{0}$ and covariance matrix $m\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a $(p+q+1)\times(p+q+1)$ matrix partitioned as

$$m\boldsymbol{\Omega}\equiv\begin{pmatrix} m\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\beta}} & m\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\gamma}} & m\boldsymbol{\Omega}_{\boldsymbol{\beta}\tau} \\ m\boldsymbol{\Omega}'_{\boldsymbol{\beta}\boldsymbol{\gamma}} & m\boldsymbol{\Omega}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} & m\boldsymbol{\Omega}_{\boldsymbol{\gamma}\tau} \\ m\boldsymbol{\Omega}'_{\boldsymbol{\beta}\tau} & m\boldsymbol{\Omega}'_{\boldsymbol{\gamma}\tau} & m\Omega_{\tau\tau}\end{pmatrix}=\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\begin{pmatrix} E[\psi_i^{\boldsymbol{\beta}}\psi_i^{\boldsymbol{\beta}'}] & E[\psi_i^{\boldsymbol{\beta}}\psi_i^{\boldsymbol{\gamma}'}] & E[\psi_i^{\boldsymbol{\beta}}\psi_i^{\tau}] \\ E[\psi_i^{\boldsymbol{\gamma}}\psi_i^{\boldsymbol{\beta}'}] & E[\psi_i^{\boldsymbol{\gamma}}\psi_i^{\boldsymbol{\gamma}'}] & E[\psi_i^{\boldsymbol{\gamma}}\psi_i^{\tau}] \\ E[\psi_i^{\tau}\psi_i^{\boldsymbol{\beta}'}] & E[\psi_i^{\tau}\psi_i^{\boldsymbol{\gamma}'}] & E[\psi_i^{\tau}\psi_i^{\tau}]\end{pmatrix}.$$

It is noticed that $E[u_{1i}(y_{ij}-\boldsymbol{x}'_{ij}\boldsymbol{\beta})]=0$ and $E[u_{2i}(y_{ij}-\boldsymbol{x}'_{ij}\boldsymbol{\beta})]=0$ when y_{ij} are normally distributed. In such a case, it follows $\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\gamma}}=0$ and $\boldsymbol{\Omega}_{\boldsymbol{\beta}\tau}=\mathbf{0}$, namely $\boldsymbol{\beta}$ and $\boldsymbol{\phi}=(\boldsymbol{\gamma}',\tau^2)'$ are asymptotically orthogonal. However, since we do not assume that normality for observations $y'_{ij}s$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ are not necessarily orthogonal.

The asymptotic covariance matrix $m\boldsymbol{\Omega}$ or $\boldsymbol{\Omega}$ can be easily estimated from samples. For example, $m\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\beta}}=\lim_{m\to\infty}m^{-1}\sum_{i=1}^{m}E[\psi_i^{\boldsymbol{\beta}}\psi_i^{\boldsymbol{\beta}'}]$ can be estimated by

$$m\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\beta}\boldsymbol{\beta}}=\frac{1}{m}\sum_{i=1}^{m}\widehat{\psi_i^{\boldsymbol{\beta}}}\widehat{\psi_i^{\boldsymbol{\beta}'}},$$

where $\widehat{\psi_i^{\boldsymbol{\beta}}}$ is obtained by replacing unknown parameters $\boldsymbol{\theta}$ in $\psi_i^{\boldsymbol{\beta}}$ with estimates $\widehat{\boldsymbol{\theta}}$. It is noted that the accuracy of estimation is given by

$$\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\beta}\boldsymbol{\beta}}=\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\beta}}+o_p(m^{-1}),$$

from Theorem 1 and $\boldsymbol{\Omega}=O(m^{-1})$. The estimator $\widehat{\boldsymbol{\Omega}}$ will be used to get the estimators of mean squared errors of predictors in Section 3.

We next provide the asymptotic properties of conditional covariance matrix given in the following corollary where the proof is given in the Appendix.

Corollary 1. *Under Assumption* (A), *for $i=1,\ldots,m$, it follows that*

$$E\left((\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\Big|\boldsymbol{y}_i\right)=\boldsymbol{\Omega}+o_p(m^{-1}).\tag{14}$$

This property is used for estimation and evaluating the mean squared errors of EBLUP discussed in the subsequent section. Moreover, in the evaluation of the mean squared errors of EBLUP and derivation of its estimators, we need to obtain the conditional and unconditional asymptotic bias of estimators $\widehat{\boldsymbol{\theta}}$.

Let $\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i),\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)$ and $\boldsymbol{b}_{\tau}^{(i)}(\boldsymbol{y}_i)$ be the second-order conditional asymptotic bias defined as

$$E[\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|\boldsymbol{y}_i]=\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)+o_p(m^{-1}),\quad E[\widehat{\gamma}-\gamma|\boldsymbol{y}_i]=\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)+o_p(m^{-1}),$$
$$E[\widehat{\tau}^2-\tau^2|\boldsymbol{y}_i]=\boldsymbol{b}_{\tau}^{(i)}(\boldsymbol{y}_i)+o_p(m^{-1}).$$

In the following theorem, we provide the analytical expressions of $\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i),\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)$ and $\boldsymbol{b}_{\tau}^{(i)}(\boldsymbol{y}_i)$. Define $\boldsymbol{b}_{\boldsymbol{\beta}}$, $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and b_τ by

$$\begin{aligned}\boldsymbol{b}_{\boldsymbol{\beta}}=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\Bigg\{&\sum_{s=1}^{q}\sum_{k=1}^{m}\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k\left(\boldsymbol{\Omega}_{\beta^*\gamma_s}-\boldsymbol{\Omega}_{\beta\gamma_s}\right)\\ &+\sum_{k=1}^{m}\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{J}_{n_k}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k(\boldsymbol{\Omega}_{\boldsymbol{\beta}^*\tau}-\boldsymbol{\Omega}_{\boldsymbol{\beta}\tau})\Bigg\}\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{b}_{\boldsymbol{\gamma}} = \boldsymbol{T}_2(\gamma)\Big[& 2\sum_{k=1}^{m} \mathrm{col}\left\{\mathrm{tr}\left(\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\left[\boldsymbol{V}_{\mathrm{OLS}}\boldsymbol{X}_k' - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\right]\right)\right\}_r \\
& - \sum_{k=1}^{m}\sum_{j=1}^{n_k} \boldsymbol{z}_{kj}\sigma_{kj(2)}(\boldsymbol{z}_{kj} - 2n_k^{-1}\boldsymbol{z}_{kj} + n_k^{-1}\bar{\boldsymbol{z}}_k)'\boldsymbol{\Omega}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\boldsymbol{z}_{kj}\Big],
\end{aligned} \tag{15}
$$

and

$$
\begin{aligned}
b_\tau = & -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(1)}^2\boldsymbol{z}_{jk}'\boldsymbol{b}_\gamma - \frac{2}{N}\sum_{k=1}^{m}\mathrm{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\boldsymbol{X}_k\right\} \\
& - \frac{1}{2N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(2)}^2\boldsymbol{z}_{kj}'\boldsymbol{\Omega}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\boldsymbol{z}_{kj} + \frac{1}{N}\sum_{k=1}^{m}\mathrm{tr}\left(\boldsymbol{X}_k'\boldsymbol{X}_k\boldsymbol{V}_{\mathrm{OLS}}\right),
\end{aligned}
$$

where $\boldsymbol{E}_k = \boldsymbol{I}_{n_k} - n_k^{-1}\boldsymbol{J}_{n_k}$, $\boldsymbol{V}_{\mathrm{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\boldsymbol{Z}_{kr} = \mathrm{diag}(z_{k1r}, \ldots, z_{kn_kr})$ for r-th element z_{kjr} of $\boldsymbol{z}_{kj}$, $\boldsymbol{\Omega}_{\boldsymbol{\beta}^*a}$ for $a \in \{\tau, \gamma_1, \ldots, \gamma_q\}$ and $\boldsymbol{W}_{i(s)}$ are defined in the proof of Theorem 2, and $\mathrm{col}\{a_r\}_r$ denotes a q-dimensional vector $(a_1, \ldots, a_q)'$. It is noted that $\boldsymbol{b}_{\boldsymbol{\beta}}, \boldsymbol{b}_{\boldsymbol{\gamma}}, b_\tau$ are of order $O(m^{-1})$. Now we provide the second-order approximation to the conditional asymptotic bias.

Theorem 2. *Under Assumption* (A)*, we have*

$$
\begin{aligned}
& \boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i) = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) + \boldsymbol{b}_{\boldsymbol{\beta}}, \quad \boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i) = \boldsymbol{T}_2(\gamma)\boldsymbol{u}_{2i} + \boldsymbol{b}_{\boldsymbol{\gamma}} \\
& b_\tau^{(i)}(\boldsymbol{y}_i) = m^{-1}u_{1i} - m^{-1}\boldsymbol{T}_1(\gamma)'\boldsymbol{T}_2(\gamma)\boldsymbol{u}_{2i} + b_\tau,
\end{aligned} \tag{16}
$$

where $\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)$, $\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)$ *and* $b_\tau^{(i)}(\boldsymbol{y}_i)$ *are of order* $O_p(m^{-1})$*, and* u_{1i} *and* u_{2i} *are given in* (11) *and* (12)*, respectively.*

From the above theorem, we immediately obtain the unconditional asymptotic bias of the estimators $\widehat{\boldsymbol{\theta}}$ by taking expectation with respect to $\boldsymbol{y}_i$ given in the following Corollary.

Corollary 2. *Under Assumption* (A)*, it follows that*

$$
E[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = (\boldsymbol{b}_{\boldsymbol{\beta}}', \boldsymbol{b}_{\boldsymbol{\gamma}}', b_\tau)' + o(m^{-1}),
$$

where $\boldsymbol{b}_{\boldsymbol{\beta}}$, $\boldsymbol{b}_{\boldsymbol{\gamma}}$ *and* b_τ *are given in* (15).

3 Prediction with Risk Evaluation

3.1 EBLUP

We now consider the prediction of

$$
\mu_i = \boldsymbol{c}_i'\boldsymbol{\beta} + v_i,
$$

where $\boldsymbol{c}_i$ is a known (user specified) vector and v_i is the random effect in model (4). The typical choice of $\boldsymbol{c}_i$ is $\boldsymbol{c}_i = \bar{\boldsymbol{x}}_i$ which corresponds to the prediction of mean of the i-th cluster. A predictor $\widetilde{\mu}(\boldsymbol{y}_i)$ of μ_i is evaluated in terms of the MSE $E[(\widetilde{\mu}(\boldsymbol{y}_i) - \mu_i)^2]$. In the general forms of $\widetilde{\mu}(\boldsymbol{y}_i)$, the minimizer (best predictor) of the MSE cannot be obtain without a distributional assumption for v_i and ε_{ij}. Thus we focus on the class of linear and unbiased predictors, and the best linear unbiased predictor (BLUP) of μ_i in terms of the MSE is given by

$$
\widetilde{\mu}_i = \boldsymbol{c}_i'\boldsymbol{\beta} + \boldsymbol{1}_{n_i}'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}).
$$

This can be simplified to

$$\widetilde{\mu}_i = \boldsymbol{c}_i'\boldsymbol{\beta} + \sum_{j=1}^{n_i} \lambda_{ij}\left(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}\right),$$

where $\lambda_{ij} = \tau^2\sigma_{ij}^{-2}\eta_i^{-1}$ for $\eta_i = 1 + \tau^2\sum_{h=1}^{n_i}\sigma_{ih}^{-2}$. In case of homogeneous variances, namely $\sigma_{ij}^2 = \delta^2$, it is confirmed that the BLP reduces to $\widetilde{\mu}_i = \boldsymbol{c}_i'\boldsymbol{\beta} + \lambda_i(\bar{y}_i - \bar{\boldsymbol{x}}_i'\boldsymbol{\beta})$ with $\lambda_i = n_i\tau^2(\delta^2 + n_i\tau^2)^{-1}$ as given in Hall and Maiti (2006). The BLUP is not feasible since it depends on unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and τ^2. Plugging the estimators into $\widetilde{\mu}_i$, we get the empirical best linear unbiased predictor (EBLUP)

$$\widehat{\mu}_i = \boldsymbol{c}_i'\widehat{\boldsymbol{\beta}} + \sum_{j=1}^{n_i}\widehat{\lambda}_{ij}\left(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}\right), \qquad \widehat{\lambda}_{ij} = \widehat{\tau}^2\widehat{\sigma}_{ij}^{-2}\widehat{\eta}_i^{-1} \tag{17}$$

for $\widehat{\eta}_i^{-1} = 1 + \widehat{\tau}^2\sum_{h=1}^{n_i}\widehat{\sigma}_{ih}^{-2}$. In the subsequent section, we consider the mean squared errors (MSE) of EBLUP (17) without any distributional assumptions for v_i and ε_{ij}.

3.2 Second-order approximation to MSE

To evaluate uncertainty of EBLUP given by (17), we evaluate the MSE defined as $\mathrm{MSE}_i(\boldsymbol{\phi}) = E\left[(\widehat{\mu}_i - \mu_i)^2\right]$ for $\boldsymbol{\phi} = (\boldsymbol{\gamma}', \tau^2)'$. The MSE is decomposed as

$$\begin{aligned}\mathrm{MSE}_i(\boldsymbol{\phi}) &= E\left[(\widehat{\mu}_i - \widetilde{\mu}_i + \widetilde{\mu}_i - \mu_i)^2\right]\\ &= E\left[(\widetilde{\mu}_i - \mu_i)^2\right] + E\left[(\widehat{\mu}_i - \widetilde{\mu}_i)^2\right] + 2E\left[(\widehat{\mu}_i - \widetilde{\mu}_i)(\widetilde{\mu}_i - \mu_i)\right].\end{aligned}$$

From the expression of $\widetilde{\mu}_i$, we have

$$\widetilde{\mu}_i - \mu_i = \left(\sum_{j=1}^{n_i}\lambda_{ij} - 1\right)v_i + \sum_{j=1}^{n_i}\lambda_{ij}\varepsilon_{ij},$$

which leads to

$$R_{1i}(\boldsymbol{\phi}) \equiv E\left[(\widetilde{\mu}_i - \mu_i)^2\right] = \left(\sum_{j=1}^{n_i}\lambda_{ij} - 1\right)^2\tau^2 + \sum_{j=1}^{n_i}\lambda_{ij}^2\sigma_{ij}^2 = \tau^2\eta_i^{-1}. \tag{18}$$

For the second term, however, we cannot obtain an exact expression, so that we obtain the approximation up to $O(m^{-1})$. Using the Taylor series expansion and Theorem 1, we have

$$\begin{aligned}E\left[(\widehat{\mu}_i - \widetilde{\mu}_i)^2\right] &= E\left[\left\{\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right\}^2\right] + o(m^{-1})\\ &= \mathrm{tr}\left\{E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'E\left((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Big|\boldsymbol{y}_i\right)\right]\right\} + o(m^{-1})\\ &= \mathrm{tr}\left\{E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'\right]\boldsymbol{\Omega}\right\} + o(m^{-1}) \equiv R_{2i}(\boldsymbol{\phi}) + o(m^{-1}),\end{aligned}$$

where we used Corollary 1 and the fact that $\partial\widehat{\mu}_i/\partial\boldsymbol{\theta}$ does not depend on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ except for $\boldsymbol{y}_i$. The straightforward calculation shows that

$$\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\beta}} = \boldsymbol{c}_i - \sum_{j=1}^{n_i}\lambda_{ij}\boldsymbol{x}_{ij}, \quad \frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\gamma}} = \eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}), \quad \frac{\partial\widetilde{\mu}_i}{\partial\tau^2} = \eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta}), \tag{19}$$

where

$$\boldsymbol{\delta}_{ij} = \tau^4 \sum_{h=1}^{n_i} \sigma_{ih}^{-4}\sigma_{ih(1)}^2 \boldsymbol{z}_{ih} - \tau^2\eta_i\sigma_{ij}^{-2}\sigma_{ij(1)}^2\boldsymbol{z}_{ij},$$

so that we have

$$\begin{aligned}R_{2i}(\boldsymbol{\phi}) =&\eta_i^{-4}\tau^2\left(\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}\right)'\boldsymbol{\Omega}_{\gamma\gamma}\left(\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}\right)+\eta_i^{-4}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\boldsymbol{\Omega}_{\gamma\gamma}\boldsymbol{\delta}_{ij}\\&+2\eta_i^{-3}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\boldsymbol{\Omega}_{\gamma\tau}+\eta_i^{-3}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\Omega_{\tau\tau}+\left(\boldsymbol{c}_i-\sum_{j=1}^{n_i}\lambda_{ij}\boldsymbol{x}_{ij}\right)'\boldsymbol{\Omega}_{\boldsymbol{\beta\beta}}\left(\boldsymbol{c}_i-\sum_{j=1}^{n_i}\lambda_{ij}\boldsymbol{x}_{ij}\right),\end{aligned} \tag{20}$$

which is of order $O(m^{-1})$. We next evaluate the cross term $E\left[(\widehat{\mu}_i-\widetilde{\mu}_i)(\widetilde{\mu}_i-\mu_i)\right]$. This term vanishes under the normality assumptions for v_i and ε_{ij}, but in general, it cannot be neglected. As in the case of R_{2i}, we obtain an approximation of $E\left[(\widehat{\mu}_i-\widetilde{\mu}_i)(\widetilde{\mu}_i-\mu_i)\right]$ up to $O(m^{-1})$. In the evaluation, we assume that $E(v_i^3)=E(\varepsilon_{ij}^3)=0$. To this end, we expand $\widehat{\mu}_i-\widetilde{\mu}_i$ as

$$\widehat{\mu}_i-\widetilde{\mu}_i=\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})+\frac{1}{2}\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)+o_p(m^{-1}).$$

It follows that

$$\widetilde{\mu}_i-\mu_i=\left(\sum_{j=1}^{n_i}\lambda_{ij}-1\right)v_i+\sum_{j=1}^{n_i}\lambda_{ij}\varepsilon_{ij}\equiv w_i,$$

and then,

$$E\left[(\widehat{\mu}_i-\widetilde{\mu}_i)(\widetilde{\mu}_i-\mu_i)\right]=E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\right]+\frac{1}{2}E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right]+o(m^{-1}).$$

Using the expression of (19) and Corollary 1, the straightforward calculation (whose details are given in the Appendix) shows that

$$R_{32i}(\boldsymbol{\phi})=E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right]=o(m^{-1}),$$

under the assumption $E(v_i^3)=E(\varepsilon_{ij}^3)=0$. Moreover, from Theorem 2, we obtain

$$E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\right]=R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})+o(m^{-1}),$$

for

$$\begin{aligned}R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=&\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\left(\sum_{k=1}^{m}\sum_{h=1}^{n_k}\sigma_{kh(1)}^2\boldsymbol{z}_{kh}\boldsymbol{z}_{kh}'\right)^{-1}\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\\&+m^{-1}\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\Big\{M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})-\boldsymbol{T}_1(\gamma)'\boldsymbol{T}_2(\boldsymbol{\gamma})\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\Big\},\end{aligned} \tag{21}$$

where

$$
\begin{aligned}
M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= mN^{-1}\tau^2\eta_i^{-1}\Big\{n_i\tau^2(3-\kappa_v)+\sigma_{ij}^2(\kappa_\varepsilon-3)\Big\}\\
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= mN^{-1}\tau^2\eta_i^{-1}n_i^{-2}(n_i-1)^2(\kappa_\varepsilon-3)\sigma_{ij}^2\boldsymbol{z}_{ij},
\end{aligned}
$$

and κ_v, κ_ε is defined as $E(v_i^4)=\kappa_v\tau^4$ and $E(\varepsilon_{ij}^4)=\kappa_\varepsilon\sigma_{ij}^4$, respectively, and $\boldsymbol{\kappa}=(\kappa_v,\kappa_\varepsilon)'$. The derivation of the expression of $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})$ is also given in the Appendix. From the expression (21), it holds that $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=O(m^{-1})$.

Under the normality assumption of v_i and ε_{ij}, we immediately obtain $M_{1ij}=0$ and $\boldsymbol{M}_{2ij}=\boldsymbol{0}$ since $\boldsymbol{\kappa}=(3,3)'$. This leads to $R_{31}=0$, which means that the cross term does not appear in the second-order approximated MSE, that is our result is consistent to the well-known result.

Now, we summarize the result for the second-order approximation of the MSE.

Theorem 3. *Under Assumption* (A) *and* $E[v_i^3]=E[\varepsilon_{ij}^3]=0$*, the second-order approximation of the MSE is given by*

$$
\mathrm{MSE}_i(\boldsymbol{\phi})=R_{1i}(\boldsymbol{\phi})+R_{2i}(\boldsymbol{\phi})+2R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})+o(m^{-1}),
$$

where $R_{1i}(\boldsymbol{\phi})$*,* $R_{2i}(\boldsymbol{\phi})$ *and* $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})$ *are given in* (18)*,* (20) *and* (21)*, respectively, and* $R_{1i}(\boldsymbol{\phi})=O(1)$*,* $R_{2i}(\boldsymbol{\phi})=O(m^{-1})$ *and* $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=O(m^{-1})$*.*

The approximated MSE given in Theorem 3 depends on unknown parameters. Thus, in the subsequent section, we derive the second-order unbiased estimator of the MSE by the analytical and the matching bootstrap methods.

3.3 Analytical estimator of the MSE

We first derive the analytical second-order unbiased estimator of the MSE. From Theorem 3, $R_{2i}(\boldsymbol{\phi})$ is $O(m^{-1})$, so that it can be estimated by the plug-in estimator $R_{2i}(\widehat{\boldsymbol{\phi}})$ with second-order accuracy, namely $E[R_{2i}(\widehat{\boldsymbol{\phi}})]=R_{2i}(\boldsymbol{\phi})+o(m^{-1})$. For $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})$ with order $O(m^{-1})$, if a consistent estimator $\widehat{\boldsymbol{\kappa}}$ is available for $\boldsymbol{\kappa}$, this term can be estimated by the plug-in estimator with second-order unbiasedness. To this end, we construct a consistent estimator of $\boldsymbol{\kappa}$ using the expression of fourth moment of observations. The straightforward calculation shows that

$$
\begin{aligned}
&E\left[\sum_{j=1}^{n_i}\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\}^4\right]\\
&\quad=\kappa_\varepsilon n_i^{-4}(n_i-1)(n_i-2)(n_i^2-n_i-1)\left(\sum_{j=1}^{n_i}\sigma_{ij}^4\right)+3n_i^{-3}(2n_i-3)\left\{\left(\sum_{j=1}^{n_i}\sigma_{ij}^2\right)^2-\sum_{j=1}^{n_i}\sigma_{ij}^4\right\},
\end{aligned}
$$

whereby we can estimate κ_ε by

$$
\widehat{\kappa}_\varepsilon=\frac{1}{N^*}\sum_{i=1}^m\left[\sum_{j=1}^{n_i}\Big\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}\Big\}^4-3n_i^{-3}(2n_i-3)\left\{\left(\sum_{j=1}^{n_i}\sigma_{ij}^2\right)^2-\sum_{j=1}^{n_i}\sigma_{ij}^4\right\}\right], \tag{22}
$$

where $N^*=n_i^{-4}(n_i-1)(n_i-2)(n_i^2-n_i-1)\sum_{j=1}^{n_i}\sigma_{ij}^4$ and $\widehat{\boldsymbol{\beta}}$ is feasible GLS estimator of $\boldsymbol{\beta}$ given in Section 2. For κ_v, it is observed that

$$
E\left[\left(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta}\right)^4\right]=\tau^4\kappa_v+6\tau^2\sigma_{ij}^2+\kappa_\varepsilon\sigma_{ij}^4,
$$

which leads to the estimator of κ_v given by

$$\widehat{\kappa}_v = \frac{1}{N\widehat{\tau}^4}\sum_{i=1}^m\sum_{j=1}^{n_i}\left\{\left(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}\right)^4 - 6\widehat{\tau}^2\widehat{\sigma}_{ij}^2 - \widehat{\kappa}_\varepsilon\widehat{\sigma}_{ij}^4\right\}. \tag{23}$$

From Theorem 1, it is immediately follows that the estimators given in (22) and (23) are consistent. Using these estimators, we can estimate R_{31i} by $R_{31i}(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\kappa}})$ with second-order accuracy.

Finally, we consider the second-order unbiased estimation of R_{1i}. The situation is different than before since $R_{1i} = O(1)$, which means that the plug-in estimator $R_{1i}(\widehat{\boldsymbol{\phi}})$ has the second-order bias with $O(m^{-1})$. Thus we need to obtain the second-order bias of $R_{1i}(\widehat{\boldsymbol{\phi}})$ and correct them. By the Taylor series expansion, we have

$$R_{1i}(\widehat{\boldsymbol{\phi}}) = R_{1i}(\boldsymbol{\phi}) + \left(\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}\right)(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) + \frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi})'\left(\frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right)(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) + o_p(m^{-1})$$

from Theorem 1. Then, the second-order bias of $R_{1i}(\widehat{\boldsymbol{\phi}})$ is expressed as

$$\begin{aligned}
&E[R_{1i}(\widehat{\boldsymbol{\phi}})] - R_{1i}(\boldsymbol{\phi}) \\
&\quad= \left(\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}\right)E[\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}] + \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right)E\left[(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})'\right]\right\} + o(m^{-1}) \\
&\quad= \left(\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}\right)\boldsymbol{b}_\phi + \frac{1}{2}\mathrm{tr}\left\{\left(\frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right)\boldsymbol{\Omega}_\phi\right\} + o(m^{-1}),
\end{aligned}$$

where $\boldsymbol{\Omega}_{\boldsymbol{\phi}}$ is the sub-matrix of $\boldsymbol{\Omega}$ with respect to $\boldsymbol{\phi}$, and $\boldsymbol{b}_\phi$ is the second-order bias of $\widehat{\boldsymbol{\phi}}$ given in Corollary 2. The straightforward calculation shows that

$$\begin{aligned}
&\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial\tau^2} = \eta_i^{-2}, \quad \frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\gamma}} = -\tau^2\eta_i^{-2}\boldsymbol{\eta}_{i(1)}, \quad \frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\tau^2\partial\tau^2} = 2\tau^{-2}(\eta_i^{-3} - \eta_i^{-2}), \\
&\frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\gamma}\partial\tau^2} = -2\eta_i^{-3}\boldsymbol{\eta}_{i(1)}, \quad \frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} = \tau^2\eta_i^{-3}(2\boldsymbol{\eta}_{i(1)}\boldsymbol{\eta}_{i(1)}' - \eta_i\boldsymbol{\eta}_{i(2)}),
\end{aligned}$$

where

$$\boldsymbol{\eta}_{i(1)} \equiv \frac{\partial\eta_i}{\partial\boldsymbol{\gamma}} = -\tau^2\sum_{j=1}^{n_i}\sigma_{ij}^{-4}\sigma_{ij(1)}^2\boldsymbol{z}_{ij}, \quad \boldsymbol{\eta}_{i(2)} \equiv \frac{\partial^2\eta_i}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} = \tau^2\sum_{j=1}^{n_i}\left(2\sigma_{ij}^{-2}\sigma_{ij(1)}^4 - \sigma_{ij(2)}^2\right)\sigma_{ij}^{-4}\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}'.$$

Therefore, we obtain the expression of the second-order bias given by

$$\begin{aligned}
B_i(\boldsymbol{\phi}) = &-\tau^2\eta_i^{-2}\boldsymbol{\eta}_{i(1)}'\boldsymbol{b}_\gamma + \eta_i^{-2}b_\tau - 2\eta_i^{-3}\boldsymbol{\eta}_{i(1)}'\boldsymbol{\Omega}_{\gamma\tau} + \tau^{-2}(\eta_i^{-3} - \eta_i^{-2})\Omega_{\tau\tau} \\
&+ \tau^2\eta_i^{-3}\left\{\boldsymbol{\eta}_{i(1)}'\boldsymbol{\Omega}_{\gamma\gamma}\boldsymbol{\eta}_{i(1)} - \frac{1}{2}\eta_i\mathrm{tr}\left(\boldsymbol{\eta}_{i(2)}\boldsymbol{\Omega}_{\gamma\gamma}\right)\right\},
\end{aligned} \tag{24}$$

with $B_i = O(m^{-1})$. Noting that B_i can be estimated by $B_i(\widehat{\boldsymbol{\phi}})$ with $E[B_i(\widehat{\boldsymbol{\phi}})] = B_i + o(m^{-1})$ from Theorem 1, we propose the bias corrected estimator of R_{1i} given by

$$\widehat{R_{1i}}(\widehat{\boldsymbol{\phi}})^{bc} = R_{1i}(\widehat{\boldsymbol{\phi}}) - B_i(\widehat{\boldsymbol{\phi}}),$$

which is second-order unbiased estimator of R_{1i}, namely

$$E[\widehat{R_{1i}}(\widehat{\boldsymbol{\phi}})^{bc}] = R_{1i}(\boldsymbol{\phi}) + o(m^{-1}).$$

Now, we summarize the result for the second-order unbiased estimator of MSE in the following theorem.

Theorem 4. *Under Assumption* (A) *and* $E[v_i^3] = E[\varepsilon_{ij}^3] = 0$, *the second-order unbiased estimator of* MSE_i *is given by*

$$\widehat{\mathrm{MSE}}_i = \widehat{R_{1i}}(\widehat{\boldsymbol{\phi}})^{bc} + R_{2i}(\widehat{\boldsymbol{\phi}}) + 2R_{31i}(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\kappa}}),$$

that is, $E\left[\widehat{\mathrm{MSE}}_i\right] = \mathrm{MSE}_i + o(m^{-1})$.

It is remarked that the proposed estimator of MSE does not require any resampling methods such as bootstrap. This means that the analytical estimator can be easily implemented and has less computational burden compared to bootstrap. Moreover, we do not assume normality of v_i and ε_{ij} in the derivation of the MSE estimator as in Lahiri and Rao (1995). Thus the proposed MSE estimator is expected to have a robustness property, which will be investigated in the simulation studies.

4 Simulation and Empirical Studies

4.1 Model based simulation

We first compare the performances of EBLUP obtained from the proposed HNER with variance functions (HNERVF) with the conventional NER and the HNER with random dispersions (HNERRD) proposed in Kubokawa, et al. (2014) in terms of simulated MSE. To this end, we consider the following data generating process:

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m, \\ v_i \sim (0, \tau^2), \qquad \varepsilon_{ij} \sim (0, \exp(\gamma_0 + \gamma_1 z_{ij})). \end{aligned} \tag{25}$$

We take $m = 20$, $n_i = 8$, $\beta_0 = 1, \beta_1 = 0.8$, $\tau = 1.2$. For the values of γ_0 and γ_1, we consider two patterns: $(\gamma_0, \gamma_1) = (1, -0.4), (1, 0)$. Note that $\gamma_1 = -0.4$ indicates that the true model holds the heteroscedasticity in sampling variances while $\gamma_1 = 0$ indicates the true model has homoscedastic variance in which both HNER models are overfitted. We generate x_{ij} and z_{ij} from the uniform distribution on $(0, 2)$ and $(0, 5)$, respectively, which are fixed through the simulation runs. Following Hall and Maiti (2006), we consider five patterns of distributions of v_i and ε_{ij}, that is , M1: v_i and ε_{ij} are both normally distributed, M2: v_i and ε_{ij} are both scaled t-distribution with degrees of freedom 6, M3: v_i and ε_{ij} are both scaled and located χ_5 distribution, M4: v_i are ε_{ij} are scaled and located χ_5 and $-\chi_5$ distribution, respectively, and M5: v_i are ε_{ij} are both logistic distribution. Based on $R = 10,000$ simulation runs, we calculate the MSE of each area defined as

$$\mathrm{MSE}_i = \frac{1}{R} \sum_{r=1}^{R} (\widehat{\mu}_i^{(r)} - \mu_i^{(r)})^2, \tag{26}$$

where $\widehat{\mu}_i^{(r)}$ and $\mu_i^{(r)}$ are obtained values of the EBLUP and the true values of $\mu_i = \beta_0 + \beta_1 \overline{x}_i + v_i$ in the r-th iteration, respectively. For estimation of the variance component in the NER model, we use the Prasad and Rao estimator (Prasad and Rao, 1990). The resulting simulated MSE values for five distribution and two values of γ_1 are given in Figure 1 (in case of $\gamma_1 = -0.4$) and Figure 2 (in case of $\gamma_1 = 0$). From Figure 1, it is observed that the HNERVF provides least values of MSE in all areas. It is a natural result that the HNERRD provides second best prediction in terms of MSE values, but the MSE values are not so different from the NER model. Thus the model specification is appropriate, the EBLUP obtained from HNERVF performs so well compared to the existing models. On the other hand, in Figure 2, the HNERVF provides little larger MSE values than the HNERRD and NER in normal case (M1). It is not surprising result since the parameter γ_1 in the HNERVF is 0 in the true model and the estimation error of γ_1 inflates the MSE values. However, in other cases (M2$\sim$M5), the HNERVF provides the close MSE values to the NER and HNERRD although the true model has homoscedastic variances. Thus we may conclude that the HNERVF has little disadvantages of over-specification in terms of MSE values of the EBLUP.

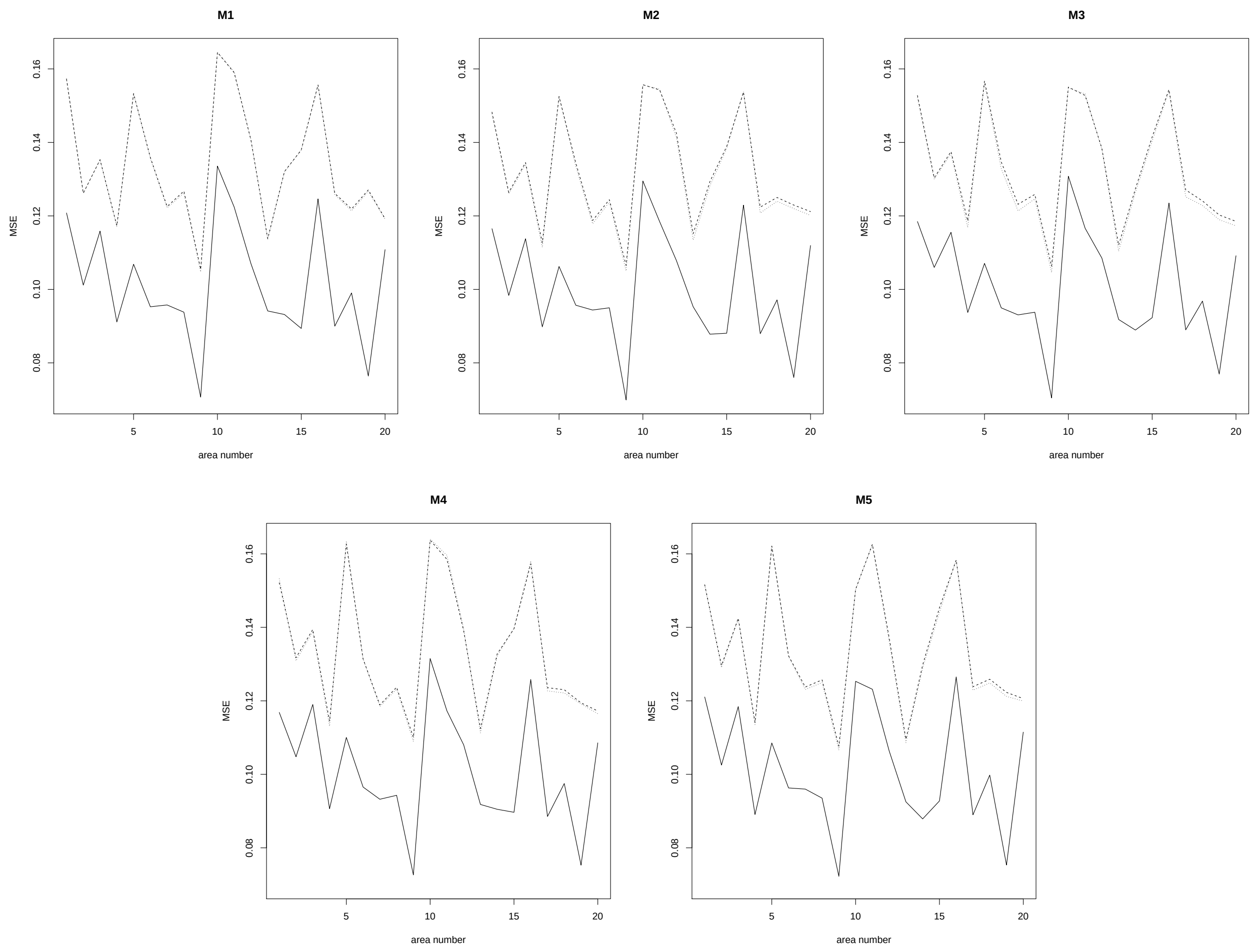

Figure 1: The Simulated Values of MSE in HNERVF (real line), NER (dashed line) and HNERRD (dotted line) in Case of $\gamma_1 = -0.4$ (Heteroscedasticity).

4.2 Finite sample performances of the MSE estimator

We next investigate the finite sample performances of the MSE estimators given in Theorem 4. We use the same data generating process given in (25) and we take $\beta_0 = 1, \beta_1 = 0.8$, $\tau = 1.2$, $\gamma_0 = 1$ and $\gamma_1 = -0.4$. Moreover, we equally divided $m = 20$ areas into four groups ($G = 1, \ldots, 5$), so that each group has five areas and the areas in the same group has the same sample size $n_G = G + 3$. Following the simulation study in the previous subsection, we again consider the five patterns of distributions for v_i and ε_{ij}. The simulated values of the MSE are obtained from (26) based on $R = 10,000$ simulation runs. Then, based on $R = 5,000$ simulation runs, we calculate the relative bias (RB) and coefficient

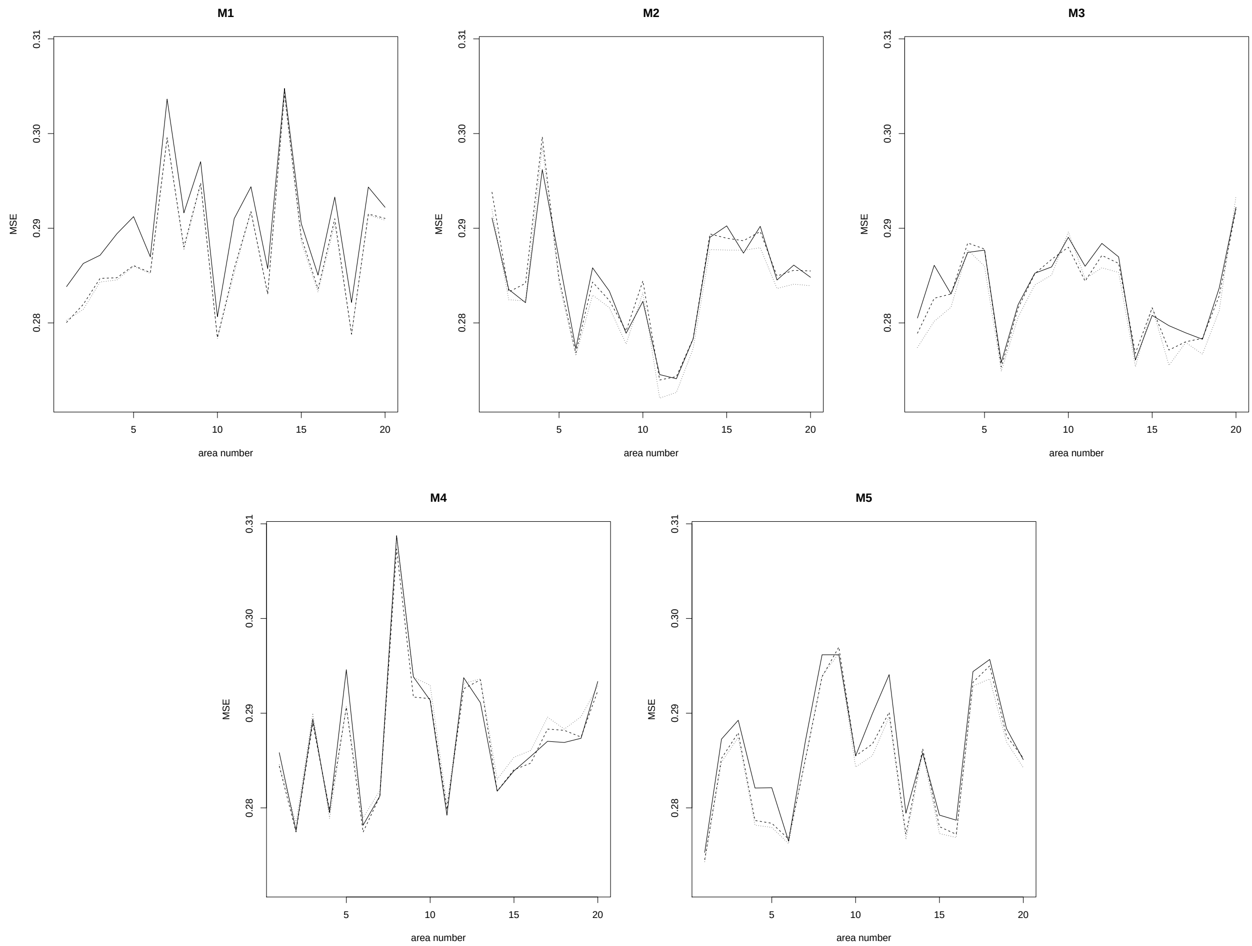

Figure 2: The Simulated Values of MSE in HNERVF (real line), NER (dashed line) and HNERRD (dotted line) in Case of $\gamma_1 = 0$ (Homoscedasticity).

of variation (CV) of MSE estimators given by

$$\mathrm{RB}_i = \frac{1}{R}\sum_{r=1}^{R}\frac{\widehat{\mathrm{MSE}}_i^{(r)} - \mathrm{MSE}_i}{\mathrm{MSE}_i}, \qquad \mathrm{CV}_i^2 = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\widehat{\mathrm{MSE}}_i^{(r)} - \mathrm{MSE}_i}{\mathrm{MSE}_i}\right)^2$$

where $\widehat{\mathrm{MSE}}_i^{(r)}$ is the MSE estimator in the r-th iteration. In Table 1, we report mean and median values of RB_i and CV_i in each group. For comparison, results for the naive MSE estimator, without any bias correction, are reported in Table 1 as well. The naive MSE estimator is the plug-in estimator of the asymptotic MSE (18), namely it is obtained by replacing τ^2 and $\boldsymbol{\gamma}$ in formula (18) by $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$, respectively. In Table 1, the relative bias is small, less than 10% in many cases. When the underlying distributions leave from normality, the MSE estimator still provides small relative bias although it has higher coefficient of variation. The naive MSE estimator is more biased than the analytical MSE estimator in all groups and models, so that the bias correction in MSE estimator is successful.

Table 1: The Mean Values of Percentage Relative Bias (RB) and Coefficient of Variation (CV) of MSE Estimator and Relative Bias of Naive MSE Estimator (RBN) in Each Group.

	G_1			G_2			G_3			G_4		
Model	RB	CV	RBN	RB	CV	RBN	RB	CV	RBN	RB	CV	RBN
M1	-8.72	17.48	-10.67	-7.61	17.52	-9.16	-7.89	19.85	-9.31	-6.52	22.02	-7.83
M2	-12.50	23.60	-13.74	-9.72	23.24	-10.66	-8.39	26.05	-9.43	-4.74	28.37	-5.68
M3	-10.86	22.47	-12.10	-10.58	22.70	-11.48	-7.65	24.66	-8.70	-4.96	26.93	-5.91
M4	-12.51	23.40	-13.57	-10.57	23.03	-11.33	-8.92	25.37	-9.86	-5.65	27.68	-6.52
M5	-11.81	21.24	-13.39	-7.27	20.31	-8.54	-6.34	22.94	-7.58	-4.27	24.98	-5.42

4.3 Illustrative example

We now investigate empirical performances of the suggested model, the empirical Bayes estimator and the second-order unbiased estimator of MSE through analysis of real data. The data used here originates 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, 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.

For $j = 1, \dots, n_i$, y_{ij} denotes the 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. This data set is treated in Kubokawa, et al. (2014), where they pointed out that the heteroscedasticity seem to be appropriate from boxplots of some areas and Bartlett test for testing homoscedastic variance. Figure 3 is the plot of the pairs (D_{ij}, e_{ij}), where e_{ij} is OLS residuals given by $e_{ij} = y_{ij} - (\widehat{\beta}_{0,OLS} + FAR_{ij}\widehat{\beta}_{1,OLS} + T_i\widehat{\beta}_{2,OLS} + D_{ij}\widehat{\beta}_{3,OLS})$. It indicates that the residuals are more variable for small D_{ij} than for large D_{ij}, namely the variances seem functions of D_{ij}. Thus we apply the following HNER model with a variance function given by

$$y_{ij} = \beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i + \varepsilon_{ij}, \tag{27}$$

where $v_i \sim (0, \tau^2)$ and $\varepsilon_{ij} \sim (0, \sigma^2(\gamma_0 + \gamma_1 D_{ij}))$. For the variance function $\sigma^2(\cdot)$, we use $\sigma^2(x) = \exp(x)$ motivated from Figure 3. As a submodel of (27), we also consider the homoscedastic variance model with $\gamma_1 = 0$. Then the estimated values of parameters in these two models are given in the following:

estimates	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\gamma}_0$	$\widehat{\gamma}_1$	$\widehat{\tau}^2$
HNERVF	42.31	2.81	-3.56	-0.66	4.91	-1.82	6.60
NER	33.35	6.58	-3.18	-0.83	3.90	0	8.82

The estimated values of β_2 and β_3, coefficients of T_i and D_{ij}, in both models are negative values which leads to the natural result that the T_i and D_{ij} have negative influence on y_{ij}. The sign of $\widehat{\gamma}_1$ is negative. This corresponds to the variability illustrated in Figure 3. The obtained values of EBLUP given in (17) are given in Table 2 for selected 15 areas. To see the difference of predicted values in terms of the degree of shrinkage, we compute $\text{dif}_i = |\bar{y}_i - \widehat{\mu}_i|$ for each two model and the results are given in Figure 4. It is observed that dif_i in NER model decreases as the area sample size n_i gets large. This is because the sample mean provides better estimates of the true mean as n_i gets larger, so that the sample mean does not need to be shrunk. On the other hand, dif_i in HNERVF is influenced by the estimated heteroscedastic variance $\sigma^2(\widehat{\gamma}_0 + \widehat{\gamma}_1 D_{ij})$ as well as n_i. Thus the plot in Figure 4 shows that the shrinkage degrees in HNERVF has more variability than that in NER. In Table 2 and Figure 3, we also provide the estimates of squared root of MSE (SMSE) given in Theorem 4. It is revealed from Table 2 that the estimates of the SMSE in NER get smaller as n_i gets larger. On the other hands, the SMSE in HNERVF do not have a similar property, because the SMSE in HNERVF is affected by not only n_i but also the heteroscedastic variance as indicated in the MSE formula given in Theorem 3. From Figure 3, we observe that the estimated SMSEs of HNERVF are smaller than that of NER in many areas. Especially, in area 47, 49, 50 and 51, the SMSE values of HNERVF are dramatically small compared to NER. In some other areas, the SMESs of HNERVF is larger than that of NER, but the differences are not so large. These observations and the residual plot in Figure 3 motivate us to utilize the HNERVF in case of heteroscedastic variance explained by some covariates.

Table 2: The Estimated Results of PLP Data for Selected 15 Areas

		sample	HNERVF		NER	
area	n_i	mean	EBLUP	SMSE	EBLUP	SMSE
1	1	60.70	42.92	3.84	41.49	4.18
10	1	38.20	38.36	3.66	37.92	3.92
7	2	40.10	39.57	3.73	39.12	3.78
19	3	32.30	33.81	3.20	34.23	3.55
15	4	38.50	39.26	3.49	40.78	3.41
41	4	18.20	19.84	2.86	20.73	3.37
12	5	41.46	38.90	3.53	41.15	3.20
51	5	16.54	16.55	1.36	15.55	3.21
46	6	20.57	20.00	1.84	19.96	3.05
52	6	15.00	17.33	1.00	16.83	3.13
25	7	29.74	29.92	2.66	30.99	2.89
50	7	20.30	18.63	1.81	17.25	2.93
33	8	22.86	23.53	2.59	22.15	2.80
49	10	16.64	15.60	1.38	16.36	2.55
34	11	24.94	23.90	2.33	23.41	2.44

5 Concluding Remarks

In the context of small-area estimation, homogeneous nested error regression models have been studied so far in the literature. However, some real data sets show heteroscedasticity in variances as pointed out in Jiang and Nguyen (2012) and Kubokawa, et al. (2014). In such a case, the residuals often indicate that the heteroscedasticity can be explained by some covariates, which motivated us to propose and investigate the heteroscedastic nested error regression model with variance functions (HNERVF). We

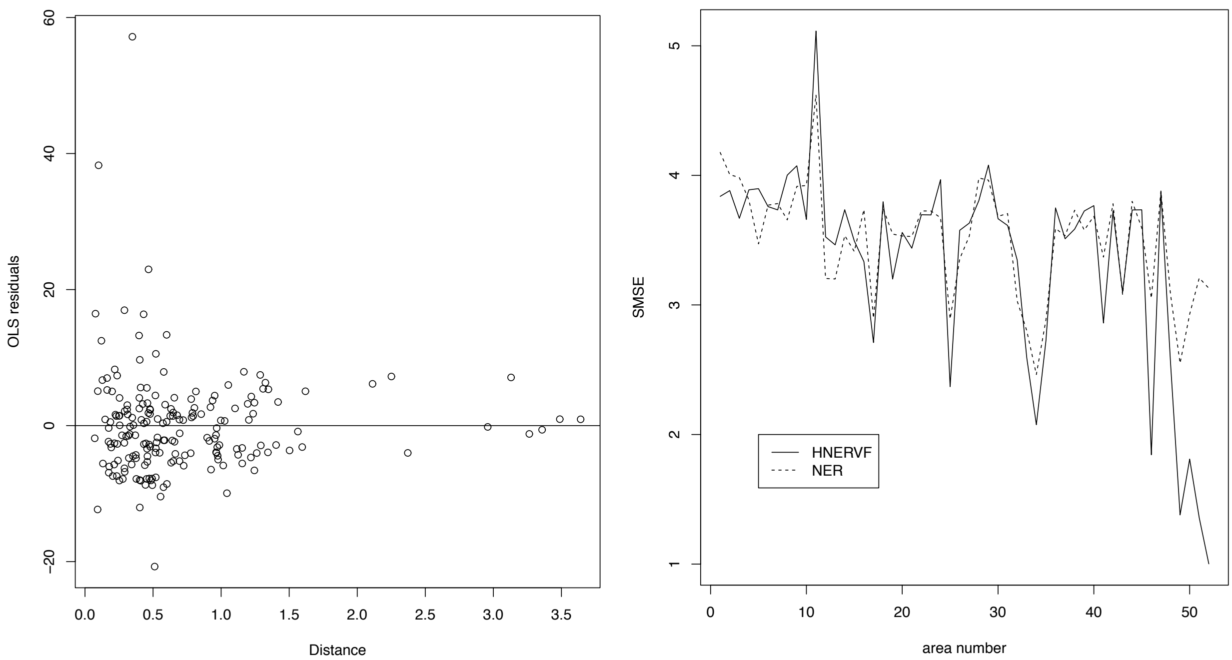

Figure 3: Plot of OLS Residuals Against Distance D_{ij} (Left) and Estimated MSE in HNERVF and NER (Right).

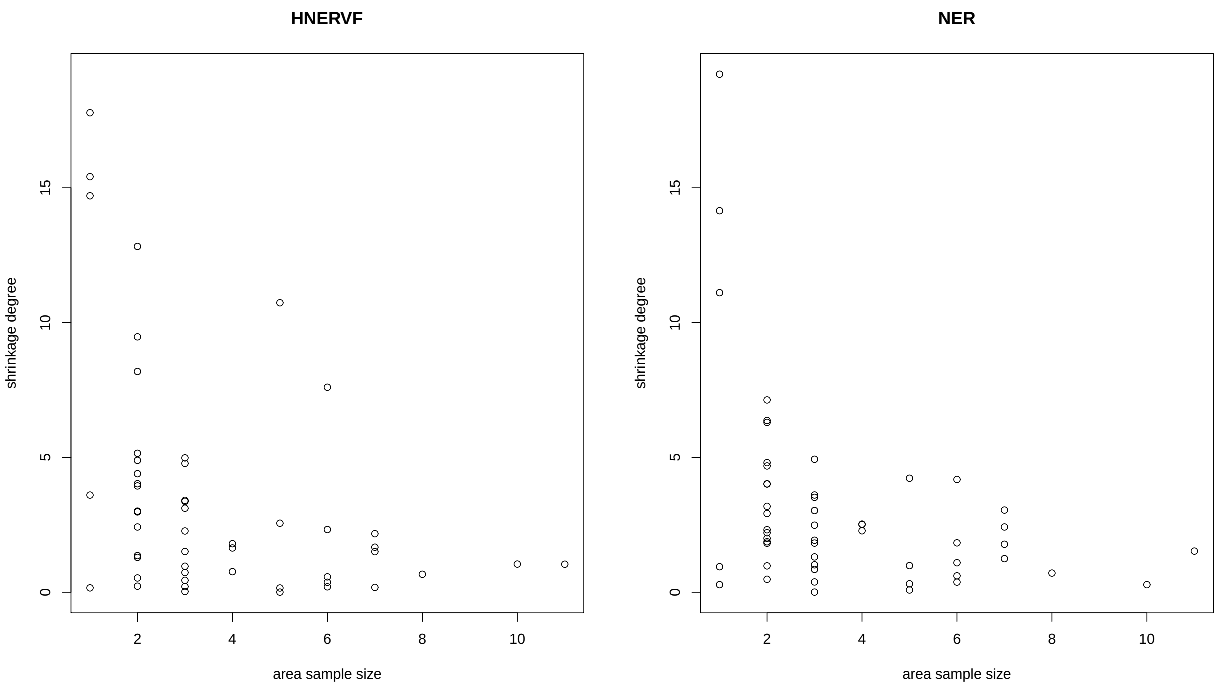

Figure 4: Plots of Shrinkage Degree dif_i Against Area Sample Size n_i in HNERVF and NER

have proposed the estimating method for the model parameters and the asymptotic properties of these estimators have been established without any distributional assumptions for error terms. For measuring uncertainty of the empirical Bayes estimator, the mean squared errors (MSE) have been approximated up to second-order, and their second-order unbiased estimators have been provided in the closed form.

Acknowledgments.
The first author was supported in part by Grant-in-Aid for Scientific Research (15J10076) from Japan Society for the Promotion of Science (JSPS). The second author was supported in part by Grant-in-Aid for Scientific Research (23243039 and 26330036) from Japan Society for the Promotion of Science.

Appendix

A.1. Proof of Theorem 1. Since $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually independent, the consistency of $\widehat{\boldsymbol{\gamma}}$ follows from the standard argument, so that $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\beta}}$ are also consistent. In what follows, we derive the asymptotic expressions of the estimators.

First we consider the asymptotic approximation of $\widehat{\tau}^2 - \tau^2$. From (9), we obtain

$$
\begin{aligned}
\widehat{\tau}^2 - \tau^2 &= \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\rm OLS})^2 - \widehat{\sigma}_{ij}^2\right\} - \tau^2 \\
&= \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2\right\} - \tau^2 - \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\
&\quad - \frac{2}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})\boldsymbol{x}_{ij}'(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) \\
&= \frac{1}{m}\sum_{i=1}^{m}u_{1i} - \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(m^{-1/2}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}), \qquad (28)
\end{aligned}
$$

where $u_{1i} = mN^{-1}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2\right\} - \tau^2$ and we used the fact that $\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta} = O_p(m^{-1/2})$ and $N^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})\boldsymbol{x}_{ij} = O_p(m^{-1/2})$ from the central limit theorem.

For the asymptotic expansion of $\widehat{\boldsymbol{\gamma}}$, remember that the estimator $\widehat{\boldsymbol{\gamma}}$ is given as the solution of the estimating equation

$$
\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\rm OLS}\right\}^2\boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right] = \boldsymbol{0}
$$

Using Taylor expansions, we have

$$
\begin{aligned}
\boldsymbol{0} &= \frac{1}{m}\sum_{i=1}\boldsymbol{u}_{2i} - \frac{2}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}\boldsymbol{z}_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'(\widehat{\boldsymbol{\beta}}_{\rm OLS} - \boldsymbol{\beta}) \\
&\quad - \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(m^{-1/2}),
\end{aligned}
$$

where

$$
\boldsymbol{u}_{2i} = mN^{-1}\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right].
$$

From the central limit theorem, it follows that

$$\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{y_{ij}-\bar{y}_i-(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\}\,\boldsymbol{z}_{ij}(\boldsymbol{x}_{ij}-\bar{\boldsymbol{x}}_i)'=O_p(m^{-1/2}),$$

so that the second terms in the expansion formula is $o_p(m^{-1/2})$. Then we get

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=\frac{N}{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'\right)^{-1}\sum_{i=1}^{m}\boldsymbol{u}_{2i}+o_p(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(m^{-1/2}).$$

Under Assumption (A), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'=O(m).$$

From the independence of $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m$ and the fact $E(\boldsymbol{u}_{2i})=\boldsymbol{0}$, we can use the central limit theorem to show that the leading term in the expansion of $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ is $O_p(m^{-1/2})$. Thus,

$$\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=\frac{N}{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sigma_{ij(1)}^2(\boldsymbol{z}_{ij}-2n_i^{-1}\boldsymbol{z}_{ij}+n_i^{-1}\bar{\boldsymbol{z}}_i)\boldsymbol{z}_{ij}'\right)^{-1}\sum_{i=1}^{m}\boldsymbol{u}_{2i}+o_p(m^{-1/2}).$$

Using the approximation of $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ and $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O_p(m^{-1/2})$, we get the asymptotic expression of $\widehat{\tau}^2-\tau^2$ from (28), which established the result for $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$.

Finally we consider the asymptotic expansion of $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}$. From the expression in (7), it follows that

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}+\sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)'(\widehat{\gamma}_s-\gamma)+\left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)'(\widehat{\tau}^2-\tau^2)+o_p(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(\widehat{\tau}^2-\tau^2).$$

Since

$$\frac{\partial}{\partial\tau^2}\boldsymbol{\Sigma}_i=\boldsymbol{J}_{n_i},\qquad\frac{\partial}{\partial\gamma_s}\boldsymbol{\Sigma}_i=\boldsymbol{W}_{i(s)},\quad s=1,\ldots,q,$$

for $\boldsymbol{W}_{i(s)}=\mathrm{diag}(\sigma_{i1(1)}^2z_{i1s},\ldots,\sigma_{in_i(1)}^2z_{in_is})$, we have

$$\begin{aligned}\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}&=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left(\widetilde{\boldsymbol{\beta}}_\tau^*-\widetilde{\boldsymbol{\beta}}\right),\\ \frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}&=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left(\widetilde{\boldsymbol{\beta}}_{\gamma_s}^*-\widetilde{\boldsymbol{\beta}}\right),\quad s=1\ldots,q,\end{aligned}\tag{29}$$

where

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_\tau^*&=\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{J}_{n_i}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i,\\ \widetilde{\boldsymbol{\beta}}_{\gamma_s}^*&=\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i,\quad s=1,\ldots,q.\end{aligned}$$

Under Assumption (A), we have $\widetilde{\boldsymbol{\beta}}_a^*-\boldsymbol{\beta}=O_p(m^{-1/2})$ for $a\in\{\tau,\gamma_1,\ldots,\gamma_q\}$, whereby $\widetilde{\boldsymbol{\beta}}^*-\widetilde{\boldsymbol{\beta}}=O_p(m^{-1/2})$. Since $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}=O_p(m^{-1/2})$ and $\widehat{\tau}^2-\tau^2=O_p(m^{-1/2})$ as shown above, we get

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})+o_p(m^{-1/2}),$$

which completes the proof.

A2. Proof of Corollary 1. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{p+q+1})' = (\boldsymbol{\beta}', \boldsymbol{\gamma}', \tau^2)'$. Note that $\psi_i^{\theta_k}, k = 1, \ldots, p+q+1$ does not depend on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_m$ and that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually independent. Then,

$$
\begin{aligned}
\frac{1}{m^2} E\left[\left(\sum_{j=1}^m \psi_j^{\theta_k}\right)\left(\sum_{j=1}^m \psi_j^{\theta_l}\right)\Big|\boldsymbol{y}_i\right] &= \frac{1}{m^2}\sum_{j=1, j\neq i}^m E\left[\psi_j^{\theta_k}\psi_j^{\theta_l}\right] + \frac{1}{m^2}\psi_i^{\theta_k}\psi_i^{\theta_l} \\
&= \boldsymbol{\Omega}_{kl} + \frac{1}{m^2}\left\{\psi_i^{\theta_k}\psi_i^{\theta_l} - E\left[\psi_i^{\theta_k}\psi_i^{\theta_l}\right]\right\} = \boldsymbol{\Omega}_{kl} + o_p(m^{-1}),
\end{aligned}
$$

where $\boldsymbol{\Omega}_{kl}$ is the (k,l)-element of $\boldsymbol{\Omega}$ and we used the fact that $E[\psi_j^{\theta_k}|\boldsymbol{y}_i] = E[\psi_j^{\theta_k}] = 0$ for $j \neq i$. Hence, we get the result from the asymptotic approximation of $\widehat{\boldsymbol{\theta}}$ given in Theorem 1.

A3. Proof of Theorem 2. We begin by deriving the conditional asymptotic bias of $\widehat{\boldsymbol{\gamma}}$. Let $\widetilde{\boldsymbol{\gamma}}$ be the solution of the equation

$$
\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta}) \equiv \frac{1}{N}\sum_{i=1}^m\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2 \boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right] = \boldsymbol{0}
$$

with $\sigma_{ij}^2 = \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$. For notational simplicity, we use $\boldsymbol{F}$ instead of $\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta})$ without any confusion and $F_r, r = 1, \ldots, q$ denotes the r-th component of $\boldsymbol{F}$, namely $\boldsymbol{F} = (F_1, \ldots, F_q)'$. Define the derivatives $\boldsymbol{F}_{(\boldsymbol{a})}$ and $F_{h(\boldsymbol{ab})}$ by

$$
\boldsymbol{F}_{(\boldsymbol{a})} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{a}'}, \qquad F_{r(\boldsymbol{ab})} = \frac{\partial^2 F_r}{\partial \boldsymbol{a}\partial \boldsymbol{b}'}.
$$

It is noted that $F_{h(\boldsymbol{\beta\gamma})} = 0$. Expanding $\boldsymbol{F}(\widehat{\boldsymbol{\gamma}}; \widehat{\boldsymbol{\beta}}_{\text{OLS}}) = \boldsymbol{0}$, we obtain

$$
\boldsymbol{0} = \boldsymbol{F} + \boldsymbol{F}_{(\boldsymbol{\gamma})}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + \boldsymbol{F}_{(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{t}_1 + \frac{1}{2}\boldsymbol{t}_2 + o_p(m^{-1}),
$$

where $\boldsymbol{t}_s = (t_{s1}, \ldots, t_{sq}), s = 1, 2$ for

$$
t_{1r} = (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' F_{r(\boldsymbol{\gamma\gamma})}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}), \qquad t_{2r} = (\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})' F_{r(\boldsymbol{\beta\beta})}(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}).
$$

It is also noted that

$$
\begin{aligned}
\boldsymbol{F}_{(\boldsymbol{\gamma})} &= -\frac{1}{m}\sum_{k=1}^m\sum_{j=1}^{n_k}\sigma_{kj(1)}^2(\boldsymbol{z}_{kj} - 2n_k^{-1}\boldsymbol{z}_{kj} + n_k^{-1}\bar{\boldsymbol{z}}_k)\boldsymbol{z}_{kj}' \\
\boldsymbol{F}_{(\boldsymbol{\beta})} &= -\frac{2}{N}\sum_{k=1}^m\sum_{j=1}^{n_k}\left\{y_{kj} - \bar{y}_k - (\boldsymbol{x}_{kj} - \bar{\boldsymbol{x}}_k)'\boldsymbol{\beta}\right\}\boldsymbol{z}_{ij}(\boldsymbol{x}_{kj} - \bar{\boldsymbol{x}}_k)',
\end{aligned}
$$

so that $\boldsymbol{F}_{(\boldsymbol{\gamma})}$ is non-stochastic. Thus we have

$$
E[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|\boldsymbol{y}_i] = -(\boldsymbol{F}_{(\boldsymbol{\gamma})})^{-1}\left\{E[\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta})|\boldsymbol{y}_i] + E\left[\boldsymbol{F}_{(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] + \frac{1}{2}E[\boldsymbol{t}_1|\boldsymbol{y}_i] + \frac{1}{2}E[\boldsymbol{t}_2|\boldsymbol{y}_i]\right\} + o_p(m^{-1}).
$$

In what follows, we shall evaluate the each term in the parenthesis in the above expression. For the first term, since $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are mutually independent and $E(\boldsymbol{u}_{2i}) = \boldsymbol{0}$, we have

$$
E[\boldsymbol{F}(\boldsymbol{\gamma}; \boldsymbol{\beta})|\boldsymbol{y}_i] = \frac{1}{m}\boldsymbol{u}_{2i}.
$$

For evaluation of the second term, we define $\boldsymbol{Z}_{kr} = \mathrm{diag}(z_{k1r}, \ldots, z_{kn_kr})$, where z_{kjr} denotes the r-th element of $\boldsymbol{z}_{kj}$. Then it follows that

$$
\begin{aligned}
E\left[\boldsymbol{F}_{r(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] &= -\frac{2}{N}\sum_{k=1}^{m} E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] \\
= -\frac{2}{N}\sum_{k=1,k\neq i}^{m} & E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] - \frac{2}{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{E}_i\boldsymbol{Z}_{ir}\boldsymbol{E}_i\boldsymbol{X}_i E\left[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta}\Big|\boldsymbol{y}_i\right].
\end{aligned}
$$

Noting that it holds for $\ell = 1, \ldots, m$ and $k \neq i$

$$
E\left[(\boldsymbol{y}_\ell - \boldsymbol{X}_\ell\boldsymbol{\beta})(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\Big|\boldsymbol{y}_i\right] = 1_{\{\ell=k\}}\boldsymbol{\Sigma}_k, \quad E[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta}|\boldsymbol{y}_i] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}),
$$

we have

$$
\begin{aligned}
E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] &= \sum_{\ell=1}^{m}\mathrm{tr}\left\{\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k' E\left[(\boldsymbol{y}_\ell - \boldsymbol{X}_\ell\boldsymbol{\beta})(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\Big|\boldsymbol{y}_i\right]\right\} \\
&= \mathrm{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\right\},
\end{aligned}
$$

which is $O(m^{-1})$ and

$$
\frac{1}{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k E\left[\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta}\Big|\boldsymbol{y}_i\right] = o_p(m^{-1}).
$$

Thus, we get

$$
E\left[\boldsymbol{F}_{r(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] = -\frac{2}{m}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\mathrm{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\right\} + o_p(m^{-1}), \tag{30}
$$

where the leading term is $O(m^{-1})$. For the third and forth terms, note that

$$
F_{r(\boldsymbol{\gamma\gamma})} = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj} - 2n_k^{-1}\boldsymbol{z}_{kj} + n_k^{-1}\bar{\boldsymbol{z}}_k)\boldsymbol{z}_{kj}' z_{kjr} \qquad F_{r(\boldsymbol{\beta\beta})} = \frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k,
$$

which are non-stochastic. Then for $h = 1, \ldots, q$,

$$
\begin{aligned}
E[t_{1r}|\boldsymbol{y}_i] &= -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k} z_{kjr}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj} - 2n_k^{-1}\boldsymbol{z}_{kj} + n_k^{-1}\bar{\boldsymbol{z}}_k)'\boldsymbol{\Omega}_{\boldsymbol{\gamma\gamma}}\boldsymbol{z}_{kj} + o_p(m^{-1}), \\
E[t_{2r}|\boldsymbol{y}_i] &= \frac{2}{N}\sum_{k=1}^{m}\mathrm{tr}\left(\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\boldsymbol{V}_{\mathrm{OLS}}\right) + o_p(m^{-1}),
\end{aligned}
$$

for $\boldsymbol{V}_{\mathrm{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$, where we used Corollary 1 and

$$
E\left[(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta})'\Big|\boldsymbol{y}_i\right] = \boldsymbol{V}_{\mathrm{OLS}} + o_p(m^{-1}), \tag{31}
$$

which follows from the similar argument in the proof of Corollary 1. Thus we obtain

$$
\begin{aligned}
E[\boldsymbol{t}_1|\boldsymbol{y}_i] &= -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\boldsymbol{z}_{kj}\sigma_{kj(2)}^2(\boldsymbol{z}_{kj} - 2n_k^{-1}\boldsymbol{z}_{kj} + n_k^{-1}\bar{\boldsymbol{z}}_k)'\boldsymbol{\Omega}_{\boldsymbol{\gamma\gamma}}\boldsymbol{z}_{kj} + o_p(m^{-1}), \\
E[\boldsymbol{t}_2|\boldsymbol{y}_i] &= \frac{2}{N}\sum_{k=1}^{m}\left\{\mathrm{tr}\left(\boldsymbol{X}_k'\boldsymbol{E}_k\boldsymbol{Z}_{kr}\boldsymbol{E}_k\boldsymbol{X}_k\boldsymbol{V}_{\mathrm{OLS}}\right)\right\}_r + o_p(m^{-1}),
\end{aligned}
$$

where $\{\boldsymbol{a}_r\}_r$ denotes the q-dimensional vector $(a_1, \dots, a_q)$. Therefore, we establish the result for $\widehat{\boldsymbol{\gamma}}$ in (16).

We next derive the result for $\widehat{\tau}^2$. Let

$$\widetilde{\tau}^2 = \frac{1}{N}\sum_{k=1}^{m}\left\{(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}) - \sum_{j=1}^{n_k}\sigma_{kj}^2\right\}.$$

Using the Taylor series expansion, we have

$$\begin{aligned}\widehat{\tau}^2 = \widetilde{\tau}^2 &+ \frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + \frac{1}{2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right)(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\\ &+ \frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) + \frac{1}{2}(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})'\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta}) + o_p(m^{-1}),\end{aligned}$$

where we used the fact that $\partial^2\widetilde{\tau}^2/\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}' = 0$. The straight calculation shows that

$$\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}} = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(1)}^2\boldsymbol{z}_{kj}, \qquad \frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(2)}^2\boldsymbol{z}_{kj}\boldsymbol{z}_{kj}', \qquad \frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = \frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}_i'\boldsymbol{X}_i,$$

which are non-stochastic. Thus we obtain

$$\begin{aligned}E[\widehat{\tau}^2 - \tau^2|\boldsymbol{y}_i] &= E[\widetilde{\tau}^2 - \tau^2|\boldsymbol{y}_i] + \left(\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}}\right)' E\left[\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}|\boldsymbol{y}_i\right] + \frac{1}{2}\text{tr}\left\{\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right)E\left[(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'|\boldsymbol{y}_i\right]\right\}\\ &+ E\left[\left(\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}}\right)'(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right] + \frac{1}{2}\text{tr}\left\{\left(\frac{\partial^2\widetilde{\tau}^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)E\left[(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})'|\boldsymbol{y}_i\right]\right\} + o_p(m^{-1})\\ &\equiv B_{\tau1}(\boldsymbol{y}_i) + B_{\tau2}(\boldsymbol{y}_i) + B_{\tau3}(\boldsymbol{y}_i) + B_{\tau4}(\boldsymbol{y}_i) + B_{\tau5}(\boldsymbol{y}_i) + o_p(m^{-1}).\end{aligned}$$

From the expression of $\widetilde{\tau}^2$, it holds that

$$\begin{aligned}B_{\tau1}(\boldsymbol{y}_i) &= \frac{1}{N}\sum_{k=1,k\neq i}^{m} n_k\tau^2 + \frac{1}{N}\left\{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) - \sum_{j=1}^{n_i}\sigma_{ij}^2\right\} - \tau^2\\ &= \left(1 - \frac{n_i}{N}\right)\tau^2 + \frac{1}{m}u_{1i} + \frac{n_i}{N}\tau^2 - \tau^2 = \frac{1}{m}u_{1i},\end{aligned}$$

for u_{1i} defined in (11). Also, we immediately have

$$B_{\tau2}(\boldsymbol{y}_i) = -\frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(1)}^2\boldsymbol{z}_{kj}'\boldsymbol{b}_{\gamma}^{(i)}(\boldsymbol{y}_i)$$

For evaluation of $B_{\tau4}(\boldsymbol{y}_i)$, note that

$$\frac{\partial\widetilde{\tau}^2}{\partial\boldsymbol{\beta}} = -\frac{2}{N}\sum_{k=1}^{m}\boldsymbol{X}_k'(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta}).$$

Similarly to (30), we get

$$\begin{aligned}B_{\tau4}(\boldsymbol{y}_i) &= -\frac{2}{N}\sum_{k=1}^{m}E\left[(\boldsymbol{y}_k - \boldsymbol{X}_k\boldsymbol{\beta})'\boldsymbol{X}_k(\widehat{\boldsymbol{\beta}}_{\text{OLS}} - \boldsymbol{\beta})\Big|\boldsymbol{y}_i\right]\\ &= -\frac{2}{N}\sum_{k=1}^{m}\text{tr}\left\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k\boldsymbol{X}_k\right\} + o_p(m^{-1}).\end{aligned}$$

Moreover, Corollary 1 and (31) enable us to obtain the expression of $B_{\tau 3}(\boldsymbol{y}_i)$ and $B_{\tau 5}(\boldsymbol{y}_i)$, whereby we get

$$b_\tau^{(i)}(\boldsymbol{y}_i) = m^{-1}u_{1i} - \frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{n_k}\sigma_{kj(1)}^2 \boldsymbol{z}_{kj}'\left\{\boldsymbol{b}_\gamma^{(i)}(\boldsymbol{y}_i) - \boldsymbol{b}_\gamma\right\} + b_\tau,$$

which completes the proof for $\widehat{\tau}^2$ in (16).

We finally derive the result for $\widehat{\boldsymbol{\beta}}$. By the Taylor series expansion,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} + \sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\gamma}_s - \gamma) + \left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\tau}^2 - \tau^2) + o_p(m^{-1}),$$

since

$$\left(\frac{\partial\widetilde{\boldsymbol{\beta}}}{\partial\boldsymbol{\phi}}\right)'(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})'\left(\frac{\partial\widetilde{\boldsymbol{\beta}}}{\partial\boldsymbol{\phi}}\right) = o_p(m^{-1}),$$

from $\partial\widetilde{\boldsymbol{\beta}}/\partial\boldsymbol{\phi} = O_p(m^{-1/2})$ as shown in the proof of Theorem 1. From (29), we have

$$\begin{aligned}
&\sum_{s=1}^{q}\left(\frac{\partial}{\partial\gamma_s}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\gamma}_s - \gamma_s)\\
&\quad = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\sum_{s=1}^{q}\left(\sum_{k=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{W}_{i(s)}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)\left\{\left(\widetilde{\boldsymbol{\beta}}_{\gamma_s}^* - \boldsymbol{\beta}\right)(\widehat{\gamma}_s - \gamma_s) - (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\gamma}_s - \gamma_s)\right\},
\end{aligned}$$

and

$$\left(\frac{\partial}{\partial\tau^2}\widetilde{\boldsymbol{\beta}}\right)(\widehat{\tau}^2 - \tau^2) = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\left(\sum_{k=1}^{m}\boldsymbol{X}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{J}_{n_k}\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k\right)\left\{(\widetilde{\boldsymbol{\beta}}_\tau^* - \boldsymbol{\beta})(\widehat{\tau}^2 - \tau^2) - (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\tau}^2 - \tau^2)\right\}.$$

Let $\boldsymbol{\Omega}_{\beta^*\gamma_s} = E[(\widetilde{\boldsymbol{\beta}}_{\gamma_s}^* - \boldsymbol{\beta})(\widehat{\gamma}_s - \gamma_s)]$ and $\boldsymbol{\Omega}_{\beta^*\tau} = E[(\widetilde{\boldsymbol{\beta}}_\tau^* - \boldsymbol{\beta})(\widehat{\tau} - \tau)]$. Then it can be shown that

$$E[(\widetilde{\boldsymbol{\beta}}_\tau^* - \boldsymbol{\beta})(\widehat{\tau} - \tau)|\boldsymbol{y}_i] = \boldsymbol{\Omega}_{\beta^*\gamma_s} + o_p(m^{-1}), \quad E[(\widetilde{\boldsymbol{\beta}}_{\gamma_s}^* - \boldsymbol{\beta})(\widehat{\gamma}_s - \gamma_s)|\boldsymbol{y}_i] = \boldsymbol{\Omega}_{\beta^*\tau} + o_p(m^{-1}),$$

which can be proved by the same arguments as in Corollary 1. Thus from Corollary 1 and the fact that

$$E\left[\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{y}_i\right] = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}),$$

we obtain the result for $\widehat{\boldsymbol{\beta}}$ in (16).

A4. Derivation of $R_{31i}(\boldsymbol{\phi}, \boldsymbol{\kappa})$. Since $\boldsymbol{y}_i$ given $v_i, \boldsymbol{\epsilon}_i$ is non-stochastic, we have

$$\begin{aligned}
E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})w_i\right] &= E\left[E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})w_i\middle|v_i, \boldsymbol{\epsilon}_i\right]\right]\\
&= E\left[E(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}|\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right]\\
&= E\left[\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\beta}}\right)w_i\right] + E\left[\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\gamma}}\right)w_i\right] + E\left[b_\tau^{(i)}(\boldsymbol{y}_i)\left(\frac{\partial\widetilde{\mu}_i}{\partial\tau}\right)w_i\right] + o(m^{-1})\\
&\equiv R_{31i}(\boldsymbol{\phi}) + o(m^{-1}).
\end{aligned}$$

It is noted that $E(w_i)=0$ and

$$E\left[(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right]=E\left[(v_i+\varepsilon_{ij})w_i\right]=\left(\sum_{j=1}^{n_i}\lambda_{ij}-1\right)\tau^2+\sum_{j=1}^{n_i}\lambda_{ij}\sigma_{ij}^2=0. \tag{32}$$

Using the expression (16) and (19), it follows that

$$\begin{aligned}
E\left[\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\beta}}\right)w_i\right]&=\left(\boldsymbol{c}_i-\sum_{j=1}^{n_i}\lambda_{ij}\boldsymbol{x}_{ij}\right)'\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}E\big[(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})w_i\big]=0\\
E\left[\boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i)'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\gamma}}\right)w_i\right]&=\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\left(\sum_{k=1}^{m}\sum_{h=1}^{n_k}\sigma_{kh(1)}^2\boldsymbol{z}_{kh}\boldsymbol{z}_{kh}'\right)^{-1}\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\\
E\left[b_{\tau}^{(i)}(\boldsymbol{y}_i)\left(\frac{\partial\widetilde{\mu}_i}{\partial\tau}\right)w_i\right]&=m^{-1}\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\Big\{M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})-\boldsymbol{T}_1(\boldsymbol{\gamma})'\boldsymbol{T}_2(\boldsymbol{\gamma})\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\Big\},
\end{aligned}$$

where

$$\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})=E\left[\boldsymbol{u}_{2i}(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right],\quad M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})=E\left[u_{1i}(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right].$$

To evaluate M_{1ij} and $\boldsymbol{M}_{2ij}$, we first prove the following result for fixed $j,k,\ell\in\{1,\ldots,n_i\}$.

$$\begin{aligned}
E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\big]=\tau^2\eta_i^{-1}\Big[&\tau^2(3-\kappa_v)+\kappa_\varepsilon\sigma_{ij}^2 1_{\{j=k=\ell\}}+\sigma_{ij}^2(1_{\{j=k\neq\ell\}}-1_{\{j=k\}})\\
&+\sigma_{ij}^2(1_{\{j=\ell\neq k\}}-1_{\{j=\ell\}})+\sigma_{ik}^2(1_{\{k=\ell\neq j\}}-1_{\{k=\ell\}})\Big].
\end{aligned} \tag{33}$$

To show (33), we note that the left side can be rewritten as

$$-\eta_i^{-1}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})v_i\right]+\sum_{h=1}^{n_i}\lambda_{ih}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})\varepsilon_{ih}\right] \tag{34}$$

from the definition of w_i. Using the fact that $\varepsilon_{i1},\ldots,\varepsilon_{in_i}$ and v_i are independent, the first term in (34) is calculated as

$$E\left[v_i^4+(\varepsilon_{ij}\varepsilon_{ik}+\varepsilon_{ij}\varepsilon_{i\ell}+\varepsilon_{ik}\varepsilon_{i\ell})v_i^2\right]=\kappa_v\tau^4+\tau^2\left(\sigma_{ij}^2 1_{\{j=k\}}+\sigma_{ij}^2 1_{\{j=\ell\}}+\sigma_{ik}^2 1_{\{k=\ell\}}\right).$$

Moreover, we have

$$\begin{aligned}
&E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})\varepsilon_{ih}\right]=E\left[\varepsilon_{ih}(\varepsilon_{ij}+\varepsilon_{i\ell}+\varepsilon_{ik})v_i^2+\varepsilon_{ij}\varepsilon_{ik}\varepsilon_{i\ell}\varepsilon_{ih}\right]\\
&\quad=\tau^2\sigma_{ih}^2\left(1_{\{h=j\}}+1_{\{h=k\}}+1_{\{h=\ell\}}\right)+\kappa_\varepsilon\sigma_{ih}^4 1_{\{j=k=\ell=h\}}+\sigma_{ih}^2\left(\sigma_{ij}^2 1_{\{j=k\neq\ell=h\}}+\sigma_{ij}^2 1_{\{j=\ell\neq k=h\}}+\sigma_{ik}^2 1_{\{j=h\neq k=\ell\}}\right),
\end{aligned}$$

whereby the second term in (34) can be calculated as

$$\tau^2\eta_i^{-1}\left[3\tau^2+\kappa_\varepsilon\sigma_{ij}^2 1_{\{j=k=\ell\}}+\sigma_{ij}^2 1_{\{j=k\neq\ell\}}+\sigma_{ij}^2 1_{\{j=\ell\neq k\}}+\sigma_{ik}^2 1_{\{k=\ell\neq j\}}\right],$$

where we used the expression $\lambda_{ih}=\tau^2\eta_i^{-1}\sigma_{ih}^{-2}$. Then we established the result (33). From (33), we immediately have

$$\begin{aligned}
\sum_{\ell=1}^{n_i}E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\big]&=\tau^2\eta_i^{-1}\left[n_i\tau^2(3-\kappa_v)+\sigma_{ij}^2(\kappa_\varepsilon-3)1_{\{j=k\}}\right]\\
&=E\big[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})^2w_i\big].
\end{aligned}$$

Now, we return to the evaluation of M_{1ij} and $\boldsymbol{M}_{2ij}$. It follows that

$$
\begin{aligned}
M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= \frac{m}{N}\sum_{h=1}^{n_i} E\left[(y_{ih}-\boldsymbol{x}_{ih}'\boldsymbol{\beta})^2(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})w_i\right] \\
&= mN^{-1}\eta_i^{-1}\tau^2\Big\{n_i\tau^2(3-\kappa_v)+\sigma_{ij}^2(\kappa_\varepsilon-3)\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= \frac{m}{N}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}E\left[\{v_i+\varepsilon_{ih}-(v_i+\bar{\varepsilon}_i)\}^2(v_i+\varepsilon_{ij})w_i\right] \\
&= \frac{m}{N}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}\Big\{E\left[(v_i+\varepsilon_{ih})^2(v_i+\varepsilon_{ij})w_i\right]-2n_i^{-1}\sum_{k=1}^{n_i}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{ih})w_i\right] \\
&\quad + n_i^{-2}\sum_{k=1}^{n_i}\sum_{\ell=1}^{n_i}E\left[(v_i+\varepsilon_{ij})(v_i+\varepsilon_{ik})(v_i+\varepsilon_{i\ell})w_i\right]\Big\}.
\end{aligned}
$$

Using the identity given in (33), we have

$$
\begin{aligned}
\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa}) &= mN^{-1}\tau^2\eta_i^{-1}\sum_{h=1}^{n_i}\boldsymbol{z}_{ih}\Big\{\sigma_{ij}^2(\kappa_\varepsilon-3)(1_{\{j=h\}}-2n_i^{-1}1_{\{j=h\}}+n_i^{-2})\Big\} \\
&= mN^{-1}\tau^2\eta_i^{-1}n_i^{-2}(n_i-1)^2(\kappa_\varepsilon-3)\sigma_{ij}^2\boldsymbol{z}_{ij},
\end{aligned}
$$

which completes the result in (21).

A5. Evaluation of $R_{32i}(\boldsymbol{\phi})$. Since $\boldsymbol{y}_i$ given $v_i,\boldsymbol{\epsilon}_i$ is non-stochastic, we have

$$
\begin{aligned}
R_{32i}(\boldsymbol{\phi}) &= \frac{1}{2}E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right] \\
&= \frac{1}{2}E\left[E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\Big|v_i,\boldsymbol{\epsilon}_i\right]\right] \\
&= \frac{1}{2}E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'|\boldsymbol{y}_i\right]\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)w_i\right] = \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Omega}E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'w_i\right]\right\}+o(m^{-1}),
\end{aligned}
$$

where we used Corollary 1 in the last equation. Note that

$$
E\left[(y_{ij}-\boldsymbol{x}_{ij}'\boldsymbol{\beta})^2w_i\right] = -\eta_i^{-1}E\left[(v_i+\varepsilon_{ij})^2v_i\right]+\sum_{h=1}^{n_i}E\left[(v_i+\varepsilon_{ij})^2\varepsilon_{ih}\right]=0
$$

since $E[v_i^3]=0$ and $E[\varepsilon_{ij}^3]=0$. Using the expression (19) of $\partial\widetilde{\mu}_i/\partial\boldsymbol{\theta}$ with the above moment results, we obtain

$$
E\left[\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\widetilde{\mu}_i}{\partial\boldsymbol{\theta}}\right)'w_i\right]=0,
$$

which leads to $R_{32i}(\boldsymbol{\phi})=o_p(m^{-1})$.

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] Cook, R.D. and Weisberg, S. (1983). Diagnostics for Heteroscedasticity in Regression. *Biometrika*, **76**, 1-10.

[3] Datta, G. and Ghosh, M. (2012). Small area shrinkage estimation. *Statist. Science*, **27**, 95-114.

[4] 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.

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

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

[7] Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: The effect of estimating the mean. *J. R. Statist. Soc.*, **B 51**, 3-14.

[8] Hall, P. and Maiti, T. (2006). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Ann. Statist.*. **34**, 1733-1750.

[9] Jiang, J., Lahiri, P. and Wan, S.M. (2002). A unified Jackknife theory for emporical best prediction with M-estimation. *Ann. Statist.*, **30**, 1782-1810.

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

[11] Kubokawa, T., Sugasawa, S., Ghosh, M. and Chaudhuri, S. (2014). Prediction in heteroscedastic nested error regression models with random dispersions. *Statist. Sinica.* to appear.

[12] Lahiri, P. and Rao, J.N.K. (1995). Robust estimation of mean squared error of small area estimators. *J. Amer. Statist. Assoc.*, **90**, 758-766.

[13] Lohr, S.L. and Rao, J.N.K. (2009). Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika*, ,

[14] 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.

[15] Muller, H.-G. and Stadtmuller, U. (1987). Estimation of heteroskedasticity in regression analysis. *Ann. Statist.*, **15**, 610-625.

[16] Muller, H.-G. and Stadtmuller, U. (1993). On variance function estimation with quadratic forms. *J. Stat. Plan. Inf.*, **35**, 213-231.

[17] Pfeffermann, D. (2014). New important developments in small area estimation. *Statist. Science*, **28**, 40-68.

[18] 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.

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

[20] Ruppert, D., Wand, M.P, Holst, U. and Hossjer, O. (1997). Local polynomial variance-function estimation. *Technometrics*, **39**, 262-273.