Statistica Sinica **27** (2017), 1101-1123
doi:https://doi.org/10.5705/ss.202015.0318

HETEROSCEDASTIC NESTED ERROR REGRESSION MODELS WITH VARIANCE FUNCTIONS

Shonosuke Sugasawa and Tatsuya Kubokawa

The Institute of Statistical Mathematics and University of Tokyo

Abstract: The nested error regression model is a useful tool for analyzing clustered (grouped) data, especially so in small area estimation. The classical nested error regression model assumes normality of random effects and error terms, and homoscedastic variances. These assumptions are often violated in applications and more flexible models are required. This article proposes a nested error regression model with heteroscedastic variances, where the normality for the underlying distributions is not assumed. We propose the structure of heteroscedastic variances by using some specified variance functions and some covariates with unknown parameters. Under this setting, we construct moment-type estimators of model parameters and some asymptotic properties including asymptotic biases and variances are derived. For predicting linear quantities, including random effects, we suggest the empirical best linear unbiased predictors, and the second-order unbiased estimators of mean squared errors are derived in closed form. We investigate the proposed method with simulation and empirical studies.

Key words and phrases: Empirical best linear unbiased predictor, heteroscedastic variance, mean squared error, nested error regression, small area estimation, variance function.

1. Introduction

Linear mixed models and the model-based estimators including empirical Bayes (EB) estimators or empirical best linear unbiased predictors (EBLUP) have been studied quite extensively in the literature. Of them, 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 on this topic, see Ghosh and Rao (1994); Rao and Molina (2015); Datta and Ghosh (2012); Pfeffermann (2014). The linear mixed models used for SAE are the Fay-Herriot models 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 not only SAE but also biological experiments and econometric analysis

In the NER model, a cluster-specific variation is added to explain the correlation among observations within clusters besides the noise, which 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 NER model proposed by Battese, Harter and Fuller (1988) is given by

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

where v_i and ε_{ij} denote random effect and samping error, respectively, and mutually independently with $v_i \sim N(0, \tau^2)$ 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 $\mathrm{Var}(y_{ij}) = \tau^2 + \sigma^2$, the same for all the clusters. Jiang and Nguyen (2012) illustrated that the within-cluster sample variances change dramatically from cluster to cluster for the data given in Battese, Harter and Fuller (1988);they proposed the heteroscedastic nested error regression (HNER) model in which $\mathrm{Var}(y_{ij})$ is proportional to σ_i^2, $\mathrm{Var}(y_{ij}) = (\lambda + 1)\sigma_i^2$. This is equivalent to the assumption that $\mathrm{Var}(v_i) = \lambda\sigma_i^2$ and $\mathrm{Var}(\varepsilon_{ij}) = \sigma_i^2$. Under this setup, Jiang and Nguyen (2012) assumed normality for v_i and ε_{ij}, and showed that the maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and λ are consistent for large m, which implies that the resulting EB estimator is asymptotically equivalent to the Bayes estimator. Thorough simulation studies, Jiang and Nguyen (2012) found that that the EBLUP from HNER model can improve the prediction accuracy over that from NER model when the data is generated from HNER model. However, there is no consistent estimator for the heteroscedastic variance σ_i^2 because of finiteness of n_i, and the mean squared error (MSE) of the EBLUP cannot be estimated consistently since it depends on σ_i^2. To fix the inconsistent estimation of σ_i^2, Kubokawa et al. (2016) proposed the hierarchical model such that the σ_i^2's are random variables and the σ_i^{-2} have a gamma distribution. The same dispersion structure was used in Maiti, Ren and Sinha (2014) who applied this hierarchical structure to the Fay-Herriot model with statistics for estimating σ_i^2. Kubokawa et al. (2016) proposed the ML estimators of model parameters, including the shape and scale parameters in the dispersion distribution of σ_i^2. They also showed the consistency of the model parameters and constructed the second-order unbiased mean squared errors of MSE by using the parametric bootstrap.

While these two HNER models are useful for analyzing unit-level data with heteroscedastic variances, the serious drawback is that both require the normality assumption for random effects and error terms, which are not necessary satisfied in applications. We address the issue of relaxing assumptions of classical normal NER models in two directions: heteroscedasticity of variances and non-normality of underlying distributions.

In data analysis, one often encounters situations in which the sampling variance $\mathrm{Var}(\varepsilon_{ij})$ is affected by the covariate $\boldsymbol{x}_{ij}$. In such a 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, see Cook and Weisberg (1983); Hall and Carroll (1989); Muller and Stadtmuller (1987, 1993) and Ruppert et al. (1997). We propose the use of the technique to introduce heteroscedastic variances into the NER model without assuming normality of underlying distributions. The variance structure we consider is $\mathrm{Var}(y_{ij}) = \tau^2 + \sigma_{ij}^2$, the sampling error ε_{ij} has heteroscedastic variance $\mathrm{Var}(\varepsilon_{ij}) = \sigma_{ij}^2$. We suggest that the variance function model be given by $\sigma_{ij}^2 = \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$. In terms of modeling the heteroscedastic variances with covariates, the generalized linear mixed models (Jiang (2006)) are also useful tools. The small area models using generalized linear mixed models are investigated in Ghosh et al. (1998), but they require strong parametric assumptions compared to the heteroscedastic model without assuming underlying distributions.

We propose flexible and tractable HNER models without assuming normality for either v_i nor ε_{ij}. The advantage of the proposed model 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}. The 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 such numerical methods as the Jackknife and the bootstrap, 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 convenient but brings a computational burden. We derive a closed expression for a second-order unbiased estimator of the MSE using second-order biases and variances of estimators of the model parameters. It 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 is given in Section 2. In Section 3, we obtain the EBLUP and the second-order approximation of the MSE. Further, we provide second-order unbiased estimators of the MSE by calculation. In Section 4, we investigate the performance of the proposed procedures through simulation and empirical studies. Proofs are given in the supplementary materials.

2. HNER Models with Variance Functions

2.1. Model settings

Suppose 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

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, m, \tag{2.1}$$

where $\boldsymbol{\beta}$ is a p-dimensional 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 $\mathrm{Var}(\varepsilon_{ij}) = \sigma_{ij}^2$, denoted by

$$v_i \sim (0, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \sim (0, \sigma_{ij}^2). \tag{2.2}$$

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

$$\sigma_{ij}^2 = \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}), \quad i = 1, \ldots, m, \tag{2.3}$$

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 are given below. The model parameters are $\boldsymbol{\beta}$, τ^2 and $\boldsymbol{\gamma}$, 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 (2.1) is expressed in a vector form as

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

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 $\boldsymbol{W}_i = \mathrm{diag}(\sigma_{i1}^2, \ldots, \sigma_{in_i}^2)$. 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 $\boldsymbol{v} = (v_1\boldsymbol{1}_{n_1}',\ldots,v_m\boldsymbol{1}_{n_m}')'$. Then (2.1) is written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon}$, where $\mathrm{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma} = \text{block diag}(\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_m)$. Three examples of the variance function in (2.3) are as follows.

(a) In the case that the dispersion of the sampling error is proportional to the mean, it is reasonable to put $\boldsymbol{z}_{ij} = \boldsymbol{x}_{(s)ij}$ and $\sigma^2(\boldsymbol{x}_{(s)ij}'\gamma) = (\boldsymbol{x}_{(s)ij}'\gamma)^2$ for a 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 (a) and (b), we have $\sigma^2(x) = x^2$, while (c) corresponds to $\log\{\sigma^2(x)\} = x$. In simulation and empirical studies in Section 4, we use the log-linear variance model.

2.2. Estimation

We here provide estimators of $\boldsymbol{\beta}$, 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

$$\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\tau^2,\boldsymbol{\gamma}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \left(\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^m \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i. \tag{2.4}$$

As $\boldsymbol{\gamma}$ and τ^2 are unknown, $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$ are used for τ^2 and $\boldsymbol{\gamma}$ to get $\widehat{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\widehat{\tau}^2, \widehat{\boldsymbol{\gamma}})$.

For estimation of τ^2, we use the second moment of the y_{ij}'s. From (2.1), it is seen that

$$E\left[(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2\right] = \tau^2 + \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}). \tag{2.5}$$

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

$$\widehat{\tau}^2 = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\text{OLS}})^2 - \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})\right\}, \tag{2.6}$$

substituting $\widehat{\boldsymbol{\gamma}}$ 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. 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 does not include the term v_i. Then

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

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

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

which is equivalent to

$$\frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\Bigg[\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\text{OLS}}\}^2\boldsymbol{z}_{ij} - \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\Bigg] = 0, \tag{2.7}$$

where $\bar{\boldsymbol{z}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}$. In the homoscedastic case with $\sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}) = \delta^2$, the estimators of δ^2 and τ^2 reduce to the estimators identical to the Prasad-Rao estimators (Prasad and Rao (1990)), up to a constant factor.

The function given as the left side of (2.7) does not depend on $\boldsymbol{\beta}$ and τ^2 and the estimator of τ^2 does not depend on $\boldsymbol{\beta}$ but on $\boldsymbol{\gamma}$. This suggests a simple algorithm for calculating the estimates: obtain $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ by solving (2.7), then get the estimate $\widehat{\tau}^2$ from (2.6) with $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$. Finally we have the GLS estimate $\widehat{\boldsymbol{\beta}}$ by substituting $\widehat{\gamma}$ and $\widehat{\tau}^2$ in (2.4).

2.3. Large sample properties

In this section, we provide large sample properties of our estimators when the number of clusters, m, goes to infinity, but the n_i's are still bounded. We need the following conditions under $m \to \infty$.

(A1) 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. The number of clusters with one observation is bounded.

(A2) The variance function $\sigma^2(\cdot)$ is twice differentiable with derivatives $(\sigma^2)^{(1)}(\cdot)$ and $(\sigma^2)^{(2)}(\cdot)$.

(A3) The following matrices converge to non-singular matrices:

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

for $a_1 = 1, 2$ and $a_2 = -1, 0, 1$.

(A4) $E[|v_i|^{8+c}] < \infty$ and $E[|\varepsilon_{ij}|^{8+c}] < \infty$ for $0 < c < 1$.

(A5) For all i and j, there exist $0 < \underline{c_1}, \overline{c_1} < \infty$ and values $\underline{c_2}, \overline{c_2}$ such that $\underline{c_1} < \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}) < \overline{c_1}$ and $\underline{c_2} < (\sigma^2)^{(k)}(\boldsymbol{z}_{ij}'\boldsymbol{\gamma}) < \overline{c_2}$ with $k = 1, 2$ in the neighborhood of the true values.

Conditions (A1) and (A3) are the standard assumptions in small area estimation. Condition (A2) is non-restrictive, and the typical variance functions $\sigma^2(x) = x^2$ and $\sigma^2(x) = \exp x$ satisfy it. The condition (A4) is used for deriving the second-order approximation of the MSE of the EBLUP discussed in Section 3, and it is satisfied by many continuous distributions, including the normal, shifted gamma, Laplace, and t-distribution with degrees of freedom larger than 9. The three examples given in Section 2.1 satisfy (A5).

In what follows, we write $\sigma_{ij}^2 \equiv \sigma^2(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$, $\sigma_{ij(k)}^2 \equiv (\sigma^2)^{(k)}(\boldsymbol{z}_{ij}'\boldsymbol{\gamma})$, $k = 1, 2$, for simplicity. We use the following notations in the i-th cluster:

$$u_{1i} = \frac{m}{N}\sum_{j=1}^{n_i}\left\{(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})^2 - \sigma_{ij}^2 - \tau^2\right\}, \tag{2.8}$$

$$\boldsymbol{u}_{2i} = \frac{m}{N}\sum_{j=1}^{n_i}\left[\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta}\right\}^2\boldsymbol{z}_{ij} - \sigma_{ij}^2(\boldsymbol{z}_{ij} - 2n_i^{-1}\boldsymbol{z}_{ij} + n_i^{-1}\bar{\boldsymbol{z}}_i)\right], \tag{2.9}$$

with

$$\boldsymbol{T}_1(\boldsymbol{\gamma}) = \sum_{k=1}^{m}\sum_{h=1}^{n_k} \sigma_{kh(1)}^2 \boldsymbol{z}_{kh},$$
$$\boldsymbol{T}_2(\boldsymbol{\gamma}) = \left(\sum_{k=1}^{m}\sum_{j=1}^{n_k} \sigma_{kh(1)}^2 (\boldsymbol{z}_{kh} - 2n_k^{-1}\boldsymbol{z}_{kh} + n_k^{-1}\bar{\boldsymbol{z}}_k)\boldsymbol{z}_{kh}'\right)^{-1}. \tag{2.10}$$

Here $\boldsymbol{T}_1(\boldsymbol{\gamma}) = O(m)$ and $\boldsymbol{T}_2(\boldsymbol{\gamma}) = O(m^{-1})$ under (A1)-(A5).

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* (A1)-(A5),

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

where

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

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

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

One has $E[u_{1i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})] = 0$ and $E[u_{2i}(y_{ij} - \boldsymbol{x}_{ij}'\boldsymbol{\beta})] = 0$ if the y_{ij} are normally distributed and, in such a case, it follows $\boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\gamma}} = 0$ and $\boldsymbol{\Omega}_{\boldsymbol{\beta}\tau} = \boldsymbol{0}$.

The asymptotic covariance matrix $m\boldsymbol{\Omega}$ or $\boldsymbol{\Omega}$ can be easily estimated. 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 $\widehat{\boldsymbol{\theta}}$. One has $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\beta}} + o_p(m^{-1})$, from Theorem 1 and $\boldsymbol{\Omega} = O(m^{-1})$.

The proof of the following result is given in the supplementary materials.

Corollary 1. *Under* (A1)-(A5), *for* $i = 1, \dots, m$,

$$E\left((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \Big| \boldsymbol{y}_i\right) = \boldsymbol{\Omega} + c(\boldsymbol{y}_i) o(m^{-1}), \tag{2.11}$$

where $c(\boldsymbol{y}_i)$ *is a fourth-order function of* $\boldsymbol{y}_i$.

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

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

Define $\boldsymbol{b}_{\boldsymbol{\beta}}$, $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and b_τ by

$$\begin{aligned}
\boldsymbol{b}_{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1} \Bigg\{ \sum_{s=1}^{q} \sum_{k=1}^{m} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{W}_{i(s)} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k \left(\boldsymbol{\Omega}_{\beta^*\gamma_s} - \boldsymbol{\Omega}_{\beta\gamma_s} \right) \\
&\qquad + \sum_{k=1}^{m} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{J}_{n_k} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k (\boldsymbol{\Omega}_{\boldsymbol{\beta}^*\tau} - \boldsymbol{\Omega}_{\boldsymbol{\beta}\tau}) \Bigg\}, \\
\boldsymbol{b}_{\boldsymbol{\gamma}} &= \boldsymbol{T}_2(\boldsymbol{\gamma}) \Bigg[2 \sum_{k=1}^{m} \mathrm{col} \left\{ \mathrm{tr} \left(\boldsymbol{E}_k \boldsymbol{Z}_{kr} \boldsymbol{E}_k \boldsymbol{X}_k \left[\boldsymbol{V}_{\mathrm{OLS}} \boldsymbol{X}_k' - (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k \right] \right) \right\}_r \\
&\qquad - \sum_{k=1}^{m} \sum_{j=1}^{n_k} \boldsymbol{z}_{kj} \sigma_{kj(2)}^2 (\boldsymbol{z}_{kj} - 2 n_k^{-1} \boldsymbol{z}_{kj} + n_k^{-1} \bar{\boldsymbol{z}}_k)' \boldsymbol{\Omega}_{\boldsymbol{\gamma\gamma}} \boldsymbol{z}_{kj} \Bigg], \qquad (2.12) \\
b_\tau &= -\frac{1}{N} \sum_{k=1}^{m} \sum_{j=1}^{n_k} \sigma_{kj(1)}^2 \boldsymbol{z}_{jk}' \boldsymbol{b}_\gamma - \frac{2}{N} \sum_{k=1}^{m} \mathrm{tr} \left\{ (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k \boldsymbol{X}_k \right\} \\
&\qquad - \frac{1}{2N} \sum_{k=1}^{m} \sum_{j=1}^{n_k} \sigma_{kj(2)}^2 \boldsymbol{z}_{kj}' \boldsymbol{\Omega}_{\boldsymbol{\gamma\gamma}} \boldsymbol{z}_{kj} + \frac{1}{N} \sum_{k=1}^{m} \mathrm{tr} \left(\boldsymbol{X}_k' \boldsymbol{X}_k \boldsymbol{V}_{\mathrm{OLS}} \right),
\end{aligned}$$

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

Theorem 2. *Under* (A1)-(A5),

$$\boldsymbol{b}_{\boldsymbol{\beta}}^{(i)}(\boldsymbol{y}_i) = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}_i' \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) + \boldsymbol{b}_{\boldsymbol{\beta}}, \quad \boldsymbol{b}_{\boldsymbol{\gamma}}^{(i)}(\boldsymbol{y}_i) = \boldsymbol{T}_2(\boldsymbol{\gamma}) \boldsymbol{u}_{2i} + \boldsymbol{b}_{\boldsymbol{\gamma}},$$
$$b_\tau^{(i)}(\boldsymbol{y}_i) = m^{-1} u_{1i} - m^{-1} \boldsymbol{T}_1(\boldsymbol{\gamma})' \boldsymbol{T}_2(\boldsymbol{\gamma}) \boldsymbol{u}_{2i} + b_\tau, \tag{2.13}$$

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* (2.8) *and* (2.9), *respectively.*

Corollary 2. *Under* (A1)-(A5), $E[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = (\boldsymbol{b}_{\boldsymbol{\beta}}', \boldsymbol{b}_{\boldsymbol{\gamma}}', b_\tau)' + o(m^{-1})$, *where* $\boldsymbol{b}_{\boldsymbol{\beta}}$, $\boldsymbol{b}_{\boldsymbol{\gamma}}$

and b_τ *are given in* (2.12).

3. Prediction and Risk Evaluation

3.1. Empirical predictor

We 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 (2.1). 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 as

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

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 the case of homogeneous variances, $\sigma_{ij}^2 = \delta^2$, 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). Plugging the estimators into $\widetilde{\mu}_i$, we get the empirical best linear unbiased predictor (EBLUP)

$$\widehat{\mu}_i = \boldsymbol{c}_i'\widehat{\boldsymbol{\beta}} + \sum_{j=1}^{n_i}\widehat{\lambda}_{ij}\left(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}\right), \qquad \widehat{\lambda}_{ij} = \widehat{\tau}^2\widehat{\sigma}_{ij}^{-2}\widehat{\eta}_i^{-1} \tag{3.1}$$

for $\widehat{\eta}_i^{-1} = 1 + \widehat{\tau}^2\sum_{h=1}^{n_i}\widehat{\sigma}_{ih}^{-2}$.

3.2. Second-order approximation to MSE

To evaluate the uncertainty of EBLUP given by (3.1), we evaluate $\mathrm{MSE}_i(\boldsymbol{\phi}) = E\left[(\widehat{\mu}_i - \mu_i)^2\right]$ 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

$$R_{1i}(\boldsymbol{\phi}) \equiv E\left[(\widetilde{\mu}_i - \mu_i)^2\right] = \left(\sum_{j=1}^{n_i} \lambda_{ij} - 1\right)^2 \tau^2 + \sum_{j=1}^{n_i} \lambda_{ij}^2 \sigma_{ij}^2 = \tau^2 \eta_i^{-1}. \tag{3.2}$$

For the second term, using the Taylor series expansion, we have

$$\widehat{\mu}_i - \widetilde{\mu}_i = \left(\frac{\partial \widetilde{\mu}_i}{\partial \boldsymbol{\theta}}\right)' (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left(\frac{\partial^2 \widetilde{\mu}_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right) (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{3.3}$$

where $\boldsymbol{\theta}^*$ is on the line between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$. Calculation shows that

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

$$\frac{\partial \widetilde{\mu}_i}{\partial \tau^2} = \eta_i^{-2} \sum_{j=1}^{n_i} \sigma_{ij}^{-2} (y_{ij} - \boldsymbol{x}_{ij}' \boldsymbol{\beta}), \tag{3.4}$$

where

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

Then each element in $\partial^2 \widetilde{\mu}_i / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ is a linear function of $\boldsymbol{y}_i$. Hence under (A1)-(A5), using similar arguments as in Lahiri and Rao (1995), we can show that

$$E\left[(\widehat{\mu}_i - \widetilde{\mu}_i)^2\right] = R_{2i}(\boldsymbol{\phi}) + o(m^{-1}). \tag{3.5}$$

The detailed proof is given in the supplementary materials. Here

$$\begin{aligned} R_{2i}(\boldsymbol{\phi}) = {} & \eta_i^{-4} \tau^2 \left(\sum_{j=1}^{n_i} \sigma_{ij}^{-2} \boldsymbol{\delta}_{ij}\right)' \boldsymbol{\Omega}_{\gamma\gamma} \left(\sum_{j=1}^{n_i} \sigma_{ij}^{-2} \boldsymbol{\delta}_{ij}\right) + \eta_i^{-4} \sum_{j=1}^{n_i} \sigma_{ij}^{-2} \boldsymbol{\delta}_{ij}' \boldsymbol{\Omega}_{\gamma\gamma} \boldsymbol{\delta}_{ij} \\ & + 2\eta_i^{-3} \sum_{j=1}^{n_i} \sigma_{ij}^{-2} \boldsymbol{\delta}_{ij}' \boldsymbol{\Omega}_{\gamma\tau} + \eta_i^{-3} \sum_{j=1}^{n_i} \sigma_{ij}^{-2} \Omega_{\tau\tau} \\ & + \left(\boldsymbol{c}_i - \sum_{j=1}^{n_i} \lambda_{ij} \boldsymbol{x}_{ij}\right)' \boldsymbol{\Omega}_{\boldsymbol{\beta}\boldsymbol{\beta}} \left(\boldsymbol{c}_i - \sum_{j=1}^{n_i} \lambda_{ij} \boldsymbol{x}_{ij}\right), \end{aligned} \tag{3.6}$$

which is of order $O(m^{-1})$. All the evaluations of residual terms can be similarly, and proofs are omitted in what follows.

The cross term $E\left[(\widehat{\mu}_i - \widetilde{\mu}_i)(\widetilde{\mu}_i - \mu_i)\right]$ vanishes under normality for v_i and ε_{ij} but, in general, it cannot be neglected. Beginning with

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

and using (3.3), we obtain

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

Using (3.4) and Corollary 1, straightforward calculation shows that

$$R_{32i}(\boldsymbol{\phi})\equiv\frac{1}{2}E\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\left(\frac{\partial^2\widetilde{\mu}_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})w_i\right]=o(m^{-1}),$$

under (A1)-(A5). 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

$$R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\boldsymbol{\delta}_{ij}'\left(\sum_{k=1}^{m}\sum_{h=1}^{n_k}\sigma_{kh(1)}^2\boldsymbol{z}_{kh}\boldsymbol{z}_{kh}'\right)^{-1}\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})+m^{-1}\eta_i^{-2}\sum_{j=1}^{n_i}\sigma_{ij}^{-2}\Big\{M_{1ij}(\boldsymbol{\phi},\boldsymbol{\kappa})-\boldsymbol{T}_1(\boldsymbol{\gamma})'\boldsymbol{T}_2(\boldsymbol{\gamma})\boldsymbol{M}_{2ij}(\boldsymbol{\phi},\boldsymbol{\kappa})\Big\},\tag{3.7}$$

where

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

and κ_v, κ_ε are defined as $E(v_i^4)=\kappa_v\tau^4$ and $E(\varepsilon_{ij}^4)=\kappa_\varepsilon\sigma_{ij}^4$, respectively with $\boldsymbol{\kappa}=(\kappa_v,\kappa_\varepsilon)'$. From (3.7), it holds that $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=O(m^{-1})$.

Under normality assumption of v_i and ε_{ij}, we have $M_{1ij}=0$ and $\boldsymbol{M}_{2ij}=\boldsymbol{0}$, since $\boldsymbol{\kappa}=(3,3)'$. This leads to $R_{31}=0$ and our result is consistent with the well-known result.

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

Theorem 3. *Under (A1)-(A5), the second-order approximation of the MSE is*

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

where $R_{1i}(\boldsymbol{\phi})$, $R_{2i}(\boldsymbol{\phi})$ and $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})$ are given in (3.2)*,* (3.6)*, and* (3.7)*, respectively, with $R_{1i}(\boldsymbol{\phi})=O(1)$, $R_{2i}(\boldsymbol{\phi})=O(m^{-1})$ and $R_{31i}(\boldsymbol{\phi},\boldsymbol{\kappa})=O(m^{-1})$.*

The approximated MSE given in Theorem 3 depends on unknown parameters, so we derive its second-order unbiased estimator by the analytical means.

3.3. Analytical 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, $E[R_{2i}(\widehat{\boldsymbol{\phi}})] = R_{2i}(\boldsymbol{\phi}) + o(m^{-1})$. 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 fourth moment of observations. Straightforward calculation shows that

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

whereby we can estimate κ_ε by

$$
\begin{aligned}
\widehat{\kappa}_\varepsilon = \frac{1}{N^*}\sum_{i=1}^{m}&\left[\sum_{j=1}^{n_i}\left\{y_{ij} - \bar{y}_i - (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}\right\}^4\right. \\
&\left. -3n_i^{-3}(2n_i-3)\left\{\left(\sum_{j=1}^{n_i}\sigma_{ij}^2\right)^2 - \sum_{j=1}^{n_i}\sigma_{ij}^4\right\}\right],
\end{aligned}
\tag{3.8}
$$

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 the feasible GLS estimator of $\boldsymbol{\beta}$ given in Section 2. For κ_v,

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

which leads to the estimator of κ_v given by

$$
\widehat{\kappa}_v = \frac{1}{N\widehat{\tau}^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{\left(y_{ij} - \boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}}_{\rm OLS}\right)^4 - 6\widehat{\tau}^2\widehat{\sigma}_{ij}^2 - \widehat{\kappa}_\varepsilon\widehat{\sigma}_{ij}^4\right\}. \tag{3.9}
$$

From Theorem 1, the estimators given in (3.8) and (3.9) are consistent. Using them, we can estimate R_{31i} by $R_{31i}(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\kappa}})$ with second-order accuracy.

Consider the second-order unbiased estimation of R_{1i}. Here $R_{1i} = O(1)$, 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 the $R_{1i}(\widehat{\boldsymbol{\phi}})$ and correct them. By a Taylor series expansion,

$$
\begin{aligned}
R_{1i}(\widehat{\boldsymbol{\phi}}) = R_{1i}(\boldsymbol{\phi}) &+ \left(\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right)(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \\
&+ \frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi})'\left(\frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\right)(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) + o_p(\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}\|^2).
\end{aligned}
$$

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

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

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. Straightforward calculation shows that

$$
\frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial \tau^2} = \eta_i^{-2}, \qquad \frac{\partial R_{1i}(\boldsymbol{\phi})}{\partial \boldsymbol{\gamma}} = -\tau^2\eta_i^{-2}\boldsymbol{\eta}_{i(1)}, \qquad \frac{\partial^2 R_{1i}(\boldsymbol{\phi})}{\partial \tau^2\partial \tau^2} = 2\tau^{-2}(\eta_i^{-3} - \eta_i^{-2}),
$$

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

where

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

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

$$
\begin{aligned}
B_i(\boldsymbol{\phi}) = &-\tau^2\eta_i^{-2}\boldsymbol{\eta}_{i(1)}'\boldsymbol{b}_\gamma + \eta_i^{-2}b_\tau - 2\eta_i^{-3}\boldsymbol{\eta}_{i(1)}'\boldsymbol{\Omega}_{\gamma\tau} + \tau^{-2}(\eta_i^{-3} - \eta_i^{-2})\Omega_{\tau\tau} \\
&+ \tau^2\eta_i^{-3}\left\{\boldsymbol{\eta}_{i(1)}'\boldsymbol{\Omega}_{\gamma\gamma}\boldsymbol{\eta}_{i(1)} - \frac{1}{2}\eta_i\mathrm{tr}\left(\boldsymbol{\eta}_{i(2)}\boldsymbol{\Omega}_{\gamma\gamma}\right)\right\},
\end{aligned}
\tag{3.10}
$$

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

Theorem 4. *Under* (A1)-(A5), *the second-order unbiased estimator of* MSE_i *is* $\widehat{\mathrm{MSE}}_i = \widehat{R_{1i}}(\widehat{\boldsymbol{\phi}})^{bc} + R_{2i}(\widehat{\boldsymbol{\phi}}) + 2R_{31i}(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\kappa}})$, *and* $E\left[\widehat{\mathrm{MSE}}_i\right] = \mathrm{MSE}_i + o(m^{-1})$.

The proposed estimator of MSE can be easily implemented and presents less computational burden than the bootstrap. We do not assume normality of v_i and ε_{ij} in the derivation of this estimator, and thus it is expected to have a robustness property.

4. Simulation and Empirical Studies

4.1. Model based simulation

We first compared the performances of EBLUP obtained from the proposed HNER and variance functions (HNERVF) with several existing models in terms of simulated mean squared errors (MSE). We considered the conventional nested error regression (NER) model, heteroscedastic NER model given by Jiang and Nguyen (2012) referred as JN, and the heteroscedastic NER with random dispersions (HNERRD) proposed in Kubokawa et al. (2016). In applying the NER model, we used the unbiased estimator for variance components given in Prasad and Rao (1990) to calculate EBLUP. We also considered log-link gamma mixed (GM) models as competitors from the generalized linear mixed models, as they also allow heteroscedasticity for the variances as the quadratic function of means. We used `glmer` function in `lme4` package in 'R' to apply the GM model.

We set $m = 20$ and $n_i = 8$ in all cases, and we computed the simulated MSE in 10 scenarios denoted by S1, ..., S10. The simulated MSE for some area-specific parameter μ_i was

$$\text{MSE}_i = \frac{1}{R}\sum_{r=1}^{R}(\widehat{\mu}_i^{(r)} - \mu_i^{(r)})^2, \tag{4.1}$$

where $R = 5,000$ was the number of simulation runs, $\widehat{\mu}_i^{(r)}$ the predicted value from some models and $\mu_i^{(r)}$ the true values in the r-th iteration. In all scenarios, we generated covariates x_{ij}'s from the uniform distribution on $(0,1)$, and they were fixed in simulation runs. From S1 to S3, we considered the heteroscedastic model with area-level heteroscedastic variances given by

$$\text{S1} \sim \text{S3}: \ y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad v_i \sim (0,\tau^2), \quad \varepsilon_{ij} \sim (0,\sigma_i^2), \quad \mu_i = \beta_0 + v_i,$$

where $\sigma_i^2 = \exp(0.8 - z_i)$ and $(\beta_0, \beta_1, \tau) = (1, 0.5, 1.2)$. We generated z_i's from the uniform on $(-1,1)$, and they were fixed in simulation runs. The scenarios S1, S2 and S3 had both v_i and ε_{ij} are normal, t with 6 degrees of freedom, and chi-squared with 5 degrees of freedom, respectively, where the t- and chi-squared distributions were scaled and located to meet the specified means and variances. For S4, we took the homoscedastic model

$$\text{S4}: \ y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad v_i \sim N(0,\tau^2), \quad \varepsilon_{ij} \sim N(0,\sigma^2), \quad \mu_i = \beta_0 + v_i,$$

with $(\beta_0, \beta_1, \tau, \sigma) = (1, 0.5, 1.2, 1.5)$. In S5 and S6, we used the heteroscedastic model with unit-level heteroscedastic variances,

$$\mathrm{S5,S6:}\ \ y_{ij}=\beta_0+\beta_1x_{ij}+v_i+\varepsilon_{ij},\ \ v_i\sim N(0,\tau^2),\ \ \varepsilon_{ij}\sim N(0,\sigma_{ij}^2),\ \ \mu_i=\beta_0+v_i,$$

where $\sigma_{ij}^2=\exp(0.8-z_{ij})$ in S5 and $\sigma_{ij}^2\sim\Gamma(5,5/\exp(0.8-z_{ij}))$ in S6. For S7 and S8, we considered the mixed model

$$\mathrm{S7,S8:}\ \ y_{ij}=\exp(\beta_0+\beta_1x_{ij}+v_i)\varepsilon_{ij},\ \ \mu_i=\exp(\beta_0+v_i),$$

with $v_i\sim N(0,\tau^2)$, $\varepsilon_{ij}\sim\Gamma(3,3)$ and $(\beta_0,\beta_1,\tau)=(0.5,1,0.3)$ in S7, and $v_i\sim t_6(0,\tau^2)$, $\varepsilon_{ij}\sim SLN(1,\sigma^2)$, and $(\beta_0,\beta_1,\tau,\sigma)=(1.2,0.6,0.4,0.4)$ in S8. Here $t_6(a,b)$ denotes the t-distribution with 6 degrees of freedom with mean a and variance b and $SLN(a,b)$ denotes the scaled log-normal distribution with mean a and variance b. Hence, $S7$ corresponds to the gamma mixed model with log-link function and $S8$ corresponds to its misspecified version. Finally, S9 to S10 are the mixed models

$$\mathrm{S9:}\ \ y_{ij}=(\beta_0+\beta_1x_{ij}+v_i)^2\varepsilon_{ij},\ \ v_i\sim N(0,\tau^2),\ \ \varepsilon_{ij}\sim SLN(1,\sigma^2),\ \ \mu_i=(\beta_0+v_i)^2$$

with $(\beta_0,\beta_1,\tau,\sigma)=(1,0.6,1.5,0.5)$, and

$$\mathrm{S10:}\ \ y_{ij}=\{\exp(\beta_0+\beta_1x_{ij})+v_i\}\varepsilon_{ij},\ \ v_i\sim N(0,\tau^2),\ \ \varepsilon_{ij}\sim SLN(1,\sigma^2),$$
$$\mu_i=\exp(\beta_0)+v_i,$$

with $(\beta_0,\beta_1,\tau,\sigma)=(1,0.3,1.2,0.5)$. Both S9 and S10 are heteroscedastic models in the sense that $\mathrm{Var}(y_{ij})$ depends on x_{ij}.

Under these scenarios, we computed the simulated MSE values of predictors from five methods (HNERVF, HNERRD, NER, JN and GM) in each area. Since one can apply GM only to the data with positive y_{ij}'s, the MSE values of GM model were calculated from S7 to S10. In Table 1, we show the mean, max, and min values of MSE over all areas for each model and scenario. In S1 to S3, HNERVF performs better than the other models, and NER model performs worst since the true model is heteroscedastic. In S4, NER model performs best among four models since it is the true model and other HNER models are overfitted. Here the inefficiency of the prediction of JN is more serious than that of HNERVF and HNERRD. As in S5 and S6, the heteroscedastic variances were unit-level, the amount of improvement of HNERVF over other models was greater. The scenario S7 was a GM model, so that it is reasonable that MSE of GM was smallest among five models. The scenario S8 is not a GM model, but it is close to GM model in that it works well compared to the other models. However, once GM is seriously misspecified as in S9 and S10, GM does not work well because of its parametric assumptions. From S8 to S10, all models were misspecified, but

Table 1. Simulated values of MSE for various scenarios and models.

	Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
mean	HNERVF	0.368	0.370	0.371	0.311	0.280	0.293	0.269	0.619	0.198	0.376
	HNERRD	0.383	0.383	0.387	0.310	0.341	0.379	0.285	0.641	0.259	0.369
	NER	0.398	0.405	0.410	0.307	0.342	0.384	0.375	0.726	0.220	0.384
	JN	0.386	0.392	0.396	0.324	0.357	0.392	0.292	0.684	0.318	0.385
	GM	—	—	—	—	—	—	0.130	0.451	0.231	0.396
max	HNERVF	0.598	0.633	0.569	0.340	0.354	0.469	0.342	1.511	0.299	0.435
	HNERRD	0.630	0.634	0.603	0.342	0.424	0.523	0.405	1.603	0.415	0.419
	NER	0.642	0.639	0.596	0.339	0.423	0.526	0.518	1.992	0.336	0.439
	JN	0.634	0.643	0.618	0.372	0.445	0.545	0.426	1.834	0.532	0.441
	GM	—	—	—	—	—	—	0.149	0.970	0.372	0.473
min	HNERVF	0.138	0.145	0.150	0.272	0.202	0.196	0.205	0.398	0.142	0.297
	HNERRD	0.156	0.157	0.166	0.272	0.254	0.255	0.219	0.408	0.142	0.302
	NER	0.173	0.177	0.202	0.269	0.256	0.256	0.286	0.442	0.152	0.305
	JN	0.157	0.160	0.166	0.288	0.273	0.256	0.220	0.414	0.168	0.314
	GM	—	—	—	—	—	—	0.104	0.335	0.168	0.309

the HNERVF model worked well compared to other models. It is best when it is the true model, but even if HNERVF is misspecified, it works reasonably well owing to its flexible structure.

4.2. Finite sample performances of the MSE estimator

We investigated the finite sample performances of the MSE estimators given in Theorem 4. To this end, consider the data generating process

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_i + \varepsilon_{ij}, \quad v_i \sim (0, \tau^2), \quad \varepsilon_{ij} \sim (0, \exp(\gamma_0 + \gamma_1 z_{ij}))$$

with $\beta_0 = 1$, $\beta_1 = 0.8$, $\tau = 1.2$, $\gamma_0 = 1$ and $\gamma_1 = -0.4$. We divided $m = 20$ areas into 5 groups ($G = 1, \ldots, 5$), so that each group had 4 areas and the areas in the same group had the same sample size $n_G = G + 3$. Following Hall and Maiti (2006), we considered five patterns of distributions of v_i and ε_{ij}: M1: v_i and ε_{ij} both normally distributed; M2: v_i and ε_{ij} both scaled t-distribution with degrees of freedom 6; M3: v_i and ε_{ij} both scaled and located χ_5 distribution; M4: v_i and ε_{ij} scaled and located χ_5 and $-\chi_5$ distributions, respectively, and M5: v_i and ε_{ij} both logistic distributions. The simulated values of the MSE were obtained from (4.1) based on $R = 10,000$ simulation runs. Based on $R = 5,000$ simulation runs, we calculate the relative bias (RB) and coefficient of variation (CV) of MSE estimators given by

Table 2. 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.

Group	Measure	M1	M2	M3	M4	M5
	RB	−8.72	−12.50	−10.86	−11.51	−11.81
G_1	CV	17.48	23.60	23.47	23.40	21.24
	RBN	−12.67	−13.74	−13.10	−13.57	−13.39
	RB	−7.61	−9.72	−10.58	−10.57	−7.27
G_2	CV	17.52	23.24	22.70	23.03	20.31
	RBN	−10.16	−12.66	−11.48	−11.33	−10.54
	RB	−7.89	−8.39	−7.65	−8.92	−6.34
G_3	CV	19.85	26.05	24.66	25.37	22.94
	RBN	−9.31	−9.43	−8.70	−9.86	−7.58
	RB	−6.52	−4.74	−4.96	−5.65	−4.27
G_4	CV	22.02	28.37	26.93	27.68	24.98
	RBN	−10.83	−7.68	−7.98	−6.52	−6.42

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

where $\widehat{\mathrm{MSE}}_i^{(r)}$ is the MSE estimator in the r-th iteration. In Table 2, 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 2 as RBN. The naive MSE estimator is the plug-in estimator of the asymptotic MSE (3.2), obtained by replacing τ^2 and $\boldsymbol{\gamma}$ in (3.2) by $\widehat{\tau}^2$ and $\widehat{\boldsymbol{\gamma}}$, respectively. In Table 2, the relative bias is small, less than 10% in many cases. When the underlying distribution is not normal, 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.

4.3. Data application

We applyed the HNERVF model together with HNERRD, NER, JN, and GM models considered in the simulation study in Section 4.1 to the data that originates from the posted land price (PLP) 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 Kanagawa prefecture take this line to work or study in Tokyo, so that it is expected that the land price depends on the distance from Tokyo. The PLP data are available for 52 stations on the Keikyu train line, and we considered each station as a small area, namely,

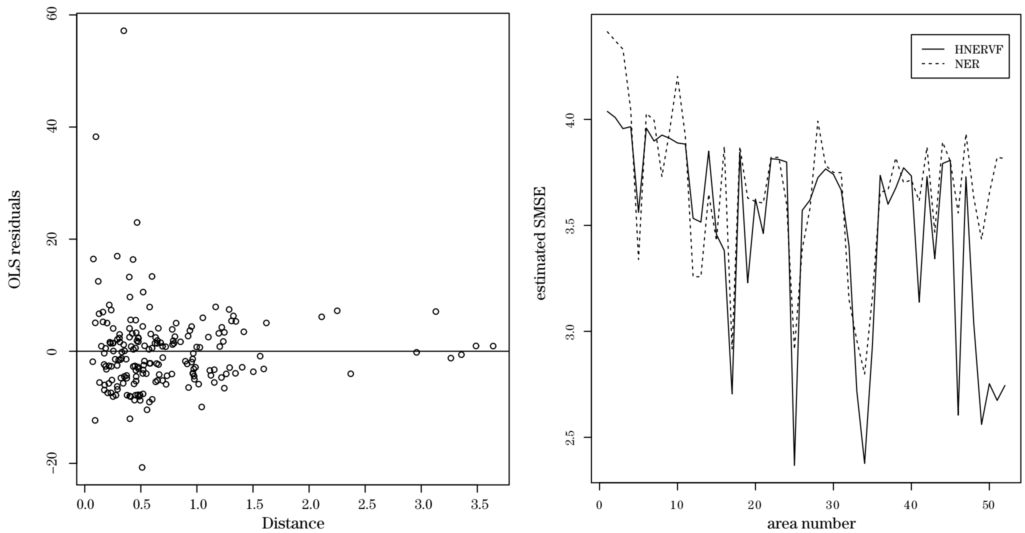

Figure 1. Plot of OLS residuals against distance D_{ij} (Left) and estimated root of MSE (RMSE) in HNERVF and NER models (Right).

$m = 52$. For the i-th station, data of n_i land spots are available, where n_i varies around four and some areas have only one observation.

For $j = 1, \ldots, n_i$, y_{ij} denotes the scaled value of the PLP (Yen/10,000) for the unit meter squares of the j-th spot, T_i is the time to take from the nearby station i to the Tokyo station around 8:30 in the morning, D_{ij} is the value of geographical distance from the spot j to the station i, and FAR_{ij} denotes the floor-area ratio, or ratio of building volume to lot area of the spot j. The three covariates FAR_{ij}, T_i, and D_{ij} are also scaled by 100,10 and 1,000, respectively. This data set is treated in Kubokawa et al. (2016), where they pointed out that the heteroscedasticity seem to be appropriate from boxplots of some areas and the Bartlett test for testing homoscedastic variance. They used the PLP data with log-transformed observations, namely $\log y_{ij}$, but we used y_{ij} in this study since the results are easier to interpret than the results from $\log y_{ij}$. In the left panel of Figure 1, we show the plot of the pairs (D_{ij}, e_{ij}), where the e_{ij} are the OLS residuals

$$e_{ij} = y_{ij} - (\widehat{\beta}_{0,OLS} + FAR_{ij}\widehat{\beta}_{1,OLS} + T_i\widehat{\beta}_{2,OLS} + D_{ij}\widehat{\beta}_{3,OLS}).$$

The figure indicates that the residuals are more variable for small D_{ij} than for large D_{ij}, and the variances are exponentially decreasing with respect to D_{ij}. Thus we applied the HNERVF model with the exponential variance function

$$y_{ij} = \beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i + \varepsilon_{ij}, \tag{4.2}$$

where $v_i \sim (0, \tau^2)$ and $\varepsilon_{ij} \sim (0, \exp(\gamma_0 + \gamma_1 D_{ij}))$. To compare the results, we also applied HNERRD, NER, JN, and GM to the PLP data with the same covariates. In applying the NER model, we regarded it as the submodel of HNERVF by putting $\gamma_1 = 0$ and used the same estimating method with HNERVF. The estimated regression coefficients from the five models are given in Table 3. As the conditional expectation of the GM model is $\exp(\beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i)$, while that of other models has the linear form $\beta_0 + FAR_{ij}\beta_1 + T_i\beta_2 + D_{ij}\beta_3 + v_i$, the scale of the estimated coefficients of GM is different from those of other models. However, the signs of estimated coefficients are the same over all models. The resulting signs are intuitively natural since the PLP is expected to be decreasing as the distance between the spot and the nearest station gets large or the nearest station gets distant from Tokyo station. Moreover, in the HNERVF model, the estimated value of γ_1 is $\widehat{\gamma}_1 = -1.82$, which is consistent with the observation from the left panel of Figure 1. Using Theorem 1, the asymptotic standard error of $\widehat{\gamma}_1$ is 0.492, so that γ_1 seems significant.

We considered estimating the and price of a spot with floor-area ratio 100% and distance from $1,000$m from station i, namely $\mu_i = \beta_0 + \beta_1 + \beta_2 T_i + \beta_3 + v_i$ under the HNERVF, HNERRD, NER, and JN models, and $\mu_i = \exp(\beta_0 + \beta_1 + \beta_2 T_i + \beta_3 + v_i)$ under the GM model. In the figure given in the supplementary materials, we provide the predicted values of μ_i of each model. From the figure, all five models provide relatively similar predicted values, and the predicted values tend to decrease with respect to the area index. This comes from the effect of T_i, since T_i increase as the area index increases.

We calculated the mean squared errors (MSE) of predictors. In the JN model, the consistent estimator of MSE cannot be obtained without any knowledge of grouping of areas (stations), as shown in Jiang and Nguyen (2012). For the GM model, the second-order unbiased estimator of MSE is hard to obtain. Thus, we considered the MSE estimator of the HNERVF, HNERRD and NER models. We used the analytical estimator given in Theorem 4 for HNERVF and NER, and the parametric bootstrap MSE estimator developed in Kubokawa et al. (2016) for HNERRD with 1,000 bootstrap replication. We found that the estimated MSE of the HNERRD model is greater than 700 for all areas, while the estimated MSE of the HNERVF and NER models were smaller than 20. The estimated value of shape parameter in dispersion (gamma) distribution in HNERRD was close to 2, which may inflate the MSE values. The estimated values of the root of the MSE

Table 3. The estimated regression coefficients in each model.

Model	β_0	β_1	β_2	β_3
HNERVF	42.31	2.81	−3.56	−0.661
HNERRD	37.72	3.88	−3.24	−0.960
NER	33.35	6.58	−3.18	−0.832
JN	37.01	3.41	−2.59	−3.19
GM	3.63	0.168	−0.122	−0.039

(RMSE) of the HNERVF and NER models are given in the right panel of Figure 1. The estimated RMSE of HNERVF is smaller than that of NER in many areas. In particular, this is true in 37 areas among 52 areas. Especially, in the latter areas, the amount of improvement is relatively large.

5. Concluding Remarks

In the context of small-area estimation, homogeneous nested error regression models have been extensively studied in the literature. However, some data sets show heteroscedasticity in variances as pointed out in Jiang and Nguyen (2012). To extend the traditional homogeneous nested error regression models, Jiang and Nguyen (2012) and Kubokawa et al. (2016) have proposed heteroscedastic nested error regression models. The drawback of these is the normality assumption required for the response values. To overcome the problem, we have proposed the structure of unit-level heteroscedastic variances modeled by some covariates and unknown parameters, and suggested heteroscedastic nested error regression models without assuming specific underlying distributions. In terms of the variance modeling with covariates, the generalized linear mixed models are also popular tools, but they require somewhat strong parametric assumptions. Therefore, the HNERVF model has clear benefits in applications. Conversely, a drawback of HNERVF is probably the structure of heteroscedastic variances specified by some covariates and unknown parameters, while the heteroscedastic models of Jiang and Nguyen (2012) and Kubokawa et al. (2016) do not requires such a specific structure. However, the heteroscedastic variances can be often modeled by some covariates as in the data application given in Section 4.3.

Supplementary Materials

In the supplementary material, we provide the proofs of Theorem 1, 2, Corollary 1, equation (3.5), the derivation of $R_{31i}(\boldsymbol{\phi}, \boldsymbol{\kappa})$, evaluation of $R_{32i}(\boldsymbol{\phi})$, and the figure showing predicted values in data analysis.

Acknowledgment

We appreciate the valuable comments and suggestions from an Associate editor and the reviewer, which led to the improved version of the paper. 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 (15H01943 and 26330036) from Japan Society for the Promotion of Science.

References

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

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

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

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.

Ghosh, M., Natarajan, K., Stroud, T. W. F. and Carlin, B. P. (1998). Generalized linear models for small area estimation. *J. Amer. Statist. Assoc.* **93**, 273-282.

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

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.

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

Jiang, J. (2006). *Linear and Generalized Linear Mixed Models and Their Applications.* Springer.

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

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

Kubokawa, T., Sugasawa, S., Ghosh, M. and Chaudhuri, S. (2016). Prediction in heteroscedastic nested error regression models with random dispersions. *Statist. Sinica* **26**, 465-492.

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.

Lohr, S. L. and Rao, J. N. K. (2009). Jackknife estimation of mean squared error of small area predictors in nonlinear mixed models. *Biometrika* **96**, 457-468.

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.

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

Muller, H.-G. and Stadtmuller, U. (1993). On variance function estimation with quadratic forms.

J. Stat. Plan. Inf. **35**, 213-231.

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

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.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation, 2nd Edition*. Wiley.

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

Risk Analysis Research Center, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa-shi, Tokyo 190-8562, Japan.

E-mail: sugasawa@ism.ac.jp

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

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

(Received September 2015; accepted May 2016)