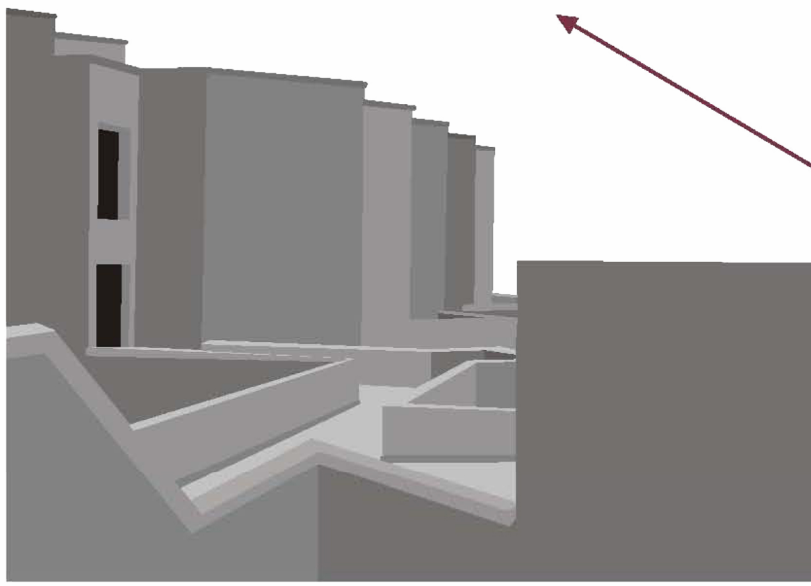

Comunicaciones del CIMAT

Small Area Estimation Based on a Two-Fold Nested Error Lognormal Model

Georges Bucyibaruta and Rogelio Ramos Quiroga

Comunicación del CIMAT
I-18-03/07.09.2018
PE/CIMAT

CIMAT

Small Area Estimation Based on a Two-Fold Nested Error Lognormal Model

Georges Bucyibaruta[1] and Rogelio Ramos Quiroga[2]
[1]georges.bucy@cimat.mx, [2]rramosq@cimat.mx

CIMAT

Centro de Investigación en Matemáticas, A.C.

Abstract

Estimation of small area population means under a two-fold nested error lognormal model is derived. Different authors have worked on the one-fold nested error model where only one aggregated level is modeled. Based on a two-stage sampling design, the aim of this manuscript is to propose a model which represents both types of variation, the variation across domains and the variation across cluster within each domain, under the log-transformation of the variable of interest. The minimum mean error squared predictor (MMSEP) of a small-area mean is obtained explicitly under this model. Replacing the unknown variance components involved in the MMSEP estimator by their estimators obtained from the Fisher scoring algorithm for restricted maximum likelihood, we provided the Emprirical best (EB) predictor. In the framework of Prasad and Rao (1990), we obtained the closed form expressions of the mean squared error (MSE). Closed form expressions, based on bias correction, for the EB predictor and the bootstrap bias-corrected MSE estimator are obtained. Results are supported by a simulation study.
Keywords: Two-fold nested error lognormal model; Empirical Bayes prediction; Mean squared error; Bias-correction; Small area estimation.

Resumen

En este trabajo se estudia una variable de interés, en el contexto de estimación para áreas pequeñas, cuya distribución es asimétrica. El modelo SAE (Small Area Estimation) que consideramos, es un modelo lineal mixto bajo una transformación logarítmica. El modelo mixto con efectos anidados a un nivel ha sido estudiado por diferentes autores; en este modelo, las áreas pequeñas son modeladas asumiendo un solo nivel de agregación. Sin embargo, en muchos estudios es de interés incorporar al modelo niveles adicionales de agregación, para tomar en cuenta variabilidad extra o incorporar características del diseño muestral. En este reporte presentaremos los predictores óptimos de la media poblacional de la variable de interés (BLUP's) y los predictores EBLUP's corregidos, así como expresiones analíticas para los Errores Cuadráticos Medios (MSE)(Prasad and Rao 1990). Además, se propone un método bootstrap paramátrico para corregir el sesgo en la estimación de MSE. La teoría propuesta se evalúa mediante un estudio de simulación.

Table of Contents

List of Figures

List of Tables

1 Introduction

Public and private sectors use survey information, provided by statistical agencies, to support government policies and business decisions, position a product on the market, etc. Typically these surveys are designed to produce information for the target sampling population and for large population subgroups.

Multistage sampling designs are used in many practical cases, when a design involves two different aggregation levels, domain (small area) and sub-domains (sub-small areas or clusters), it is reasonable to assume a twofold nested error model including random effects explaining the heterogeneity at the two levels of aggregation. In this report, we consider the two-fold nested error lognormal regression model for estimating small area means. This model includes small area and sub-small area (cluster) effects to account for the unexplained between-area and between-cluster heterogeneity, respectively.

Following the definition given by Rao (2003), the term small area or small domain refers to a subpopulation for which the domain-specific sample is not large enough to produce direct estimates with reliable precision. This subpopulation can be a small geographical area (county, state, district, etc.), a demographic group within a geographical region (specific sex-age group, etc.) or any subdivision of the population. One possible solution to improve direct estimates is to borrow strength from other related data sets by using the data either from similar areas, or using relevant auxiliary information (covariates) obtained from census or some other administrative records.

Some variables of interest are skewed distributed and there is a need to provide small area estimates for these variables. The problem of highly skewed data is, according to Barnett and Lewis (1994), particularly common in business and social surveys. Usual standard estimation methods, under a linear model, for the characteristic of interest (mean in this case) of a skewed variable can be inappropriate.

Small area methods based on the ideas of non-normal distributions have been considered. Slud and Maiti (2006) proposed an empirical Bayes or best (EB) predictor for a small area mean assuming that the area-level direct estimators have a lognormal distribution. Ghosh and Maiti (2004) discussed a small area unit-level model based on natural exponential quadratic variance function families, where they assumed that the covariates are the same across units in a single small area. Chandra and Chambers (2011) considered a lognormal distribution as a basis for constructing a model-based direct estimator for a small area mean. Its model is a weighted sum of sampled units in which the weights are defined to give the minimum mean squared error linear predictor of the population mean when the parameters of the lognormal distribution were known.

The model considered by Berg and Chandra (2014) is referred to the one-fold nested error model since only one aggregated level, the small area, is modeled. In this study, we propose a two-fold lognormal model which is different from the Berg and Chandra (2014) because we work with a two-fold unit-level data instead of one-fold unit-level data. Furthermore, we consider the case of cluster-specific covariates as in Datta and Ghosh (1991) and Pfeffermann and Barnard (1991).

The rest of this report is organized as follows. In section 2, the general form of a two-fold nested error lognormal model is presented; the minimum meam squared error predictor (MMSEP) of the small-area mean $\bar{Y}_i$ is explicitly obtained, this predictor depends on unknown variance components which are estimated by the restricted maximun likelihood (REML) method in order to obtain the empirical Bayes (EB) predictor. The MSE estimation of the resulting EB predictor, is obtained in

Section 3. The corrected EB predictor to the second-order, is obtained in Section 4. In Section 5, we propose an approximately unbiased estimator of the MSE corrected to the second order based on bootstrap method. Results of a simulation study on the relative bias of the MSE estimator are reported in Section 6. Section 7 contains the concluding remarks.

2 Two-fold nested error lognormal model

2.1 Two-fold nested error linear model

Different authors have worked on the one-fold nested error model where only one aggregated level, the small area, is modeled. However, in many real applications, it may be of interest to incorporate additional aggregated levels in the model to account for extra variability or to reflect the sampling design. In complex survey sampling, the sample is selected in stages to reduce the cost. At the first stage, primary sampling units (PSU) or clusters are selected. Within each PSU, secondary sampling units (SSU) are selected where in some situations are considered as observational sampling units or individuals units. Fuller and Battese (1973) proposed a two-fold model, that can be used to model data from such complex design in order to capture variability from both the PSU and SSU levels, and its transformation where the transformed quantities are the differences between the original observations and multiples of averages of subsets of obsevations. Later, Datta and Ghosh (1991, under a Bayesian framework, and Pfeffermann and Barnard (1991) used the two-fold model for the special case of cluster-specific covariates. Stukel and Rao (1999) extended the results of Datta and Ghosh (1991) and Pfeffermann and Barnard (1991) to general two-fold nested error regression models, considering the unit-level covariates to be available.
For the classical model-based approach, the characteristics of interest, y, and the coavariates, X, are available at the unit level and the linear mixed models (LMM) are used to represent the assumed stochastic relationship between the quantities (Battese et al., 1988). The two-fold nested error linear model is formally defined as (Stukel and Rao, 1999)

$$Y_{ijk} = x_{ijk}^T\beta + v_i + u_{ij} + e_{ijk}; \quad i = 1, \dots, M; j = 1, \dots, M_i; k = 1, \dots, N_{ij}, \tag{1}$$

where the value y_{ijk} is the observed characteristic of interest associated to unit k from cluster j within small area i, the covariates $x_{ijk}^T = (x_{ijk1}, \dots, x_{ijkp})$ are a $1 \times p$ vector of known variables, β is a $p \times 1$ vector of unknown regression parameters, and the area effects v_i, the cluster effects u_{ij} and the residual errors e_{ijk} are assumed to be mutually independent. Furthermore,

$$v_i \sim N(0, \sigma_v^2), \;\; u_{ij} \sim N(0, \sigma_u^2), \;\; e_{ijk} \sim N(0, \sigma_e^2).$$

The parameter of interest is the small area population mean

$$\bar{Y}_i = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}.$$

In the settings of the theory of prediction presented by Henderson (1975), the following theorem gives the form of the best predictor (BP) and shows why it has a minimum mean squared error.

Theorem 1. Under the two-fold nested error linear model (1), the best predictor of the small area mean $\bar{Y}_i$, $i = 1, \dots, M$ is given by $\bar{Y}_i^{BP} = E(\bar{Y}_i | y_s)$ where y_s is the observed characteristic of interest, with

$$\hat{\bar{Y}}_i^{BP}(\theta) = \frac{1}{N_i}\left[\sum_{j,k \in s_i} y_{ijk} + \sum_{j \in s_i} \sum_{k \in \bar{s}_{ij}} \tilde{y}_{ijk}^{*BP} + \sum_{j \in \bar{s}_i} \sum_{k=1}^{N_{ij}} \tilde{y}_{ijk}^{**BP} \right], \tag{2}$$

where $\bar{s}_{ij}$ is the set of nonsampled units in the j^{th} sampled cluster and $\bar{s}_i$ is the set of nonsampled sampled clusters in small area i. Also, the predictors $\tilde{y}^{*BP}_{ijk}$ and $\tilde{y}^{**BP}_{ijk}$, from (2), are defined as follows

$$\begin{aligned} \tilde{y}^{*BP}_{ijk} &= x^T_{ijk}\beta + \tilde{v}^{BP}_i + \tilde{u}^{BP}_{ij} \\ \tilde{y}^{**BP}_{ijk} &= x^T_{ijk}\beta + \tilde{v}^{BP}_i, \end{aligned} \tag{3}$$

where $\tilde{v}^{BP}_i = E(v_i|y_s)$ and $\tilde{u}^{BP}_{ij} = E(u_{ij}|y_s)$.

Proof. Let's consider another estimator $\hat{\bar{Y}}_i$ of $\bar{Y}_i$ function of y_s then

$$\begin{aligned} MSE(\hat{\bar{Y}}_i) =& E(\hat{\bar{Y}}_i - \bar{Y}_i)^2 \\ =& E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i + \bar{Y}^{BP}_i - \bar{Y}_i)^2 \\ =& E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i)^2 + E(\bar{Y}^{BP}_i - \bar{Y}_i)^2 + 2E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i)(\bar{Y}^{BP}_i - \bar{Y}_i) \\ =& MSE(\bar{Y}^{BP}_i) + E(\bar{Y}^{BP}_i - \bar{Y}_i)^2 + 2E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i)(\bar{Y}^{BP}_i - \bar{Y}_i). \end{aligned}$$

Since $E(\bar{Y}^{BP}_i - \bar{Y}_i)^2$ is positive, it suffices to show that $E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i)(\bar{Y}^{BP}_i - \bar{Y}_i) = 0$. Note that $\hat{\bar{Y}}_i - \bar{Y}^{BP}_i$ is a function of y_s say $f(y_s)$, it follows that

$$\begin{aligned} E(\hat{\bar{Y}}_i - \bar{Y}^{BP}_i)(\bar{Y}^{BP}_i - \bar{Y}_i) =& E(f(y_s)(\bar{Y}^{BP}_i - \bar{Y}_i)) \\ =& E(f(y_s)\bar{Y}^{BP}_i) - E(f(y_s)\bar{Y}_i) \\ =& E(f(y_s)\bar{Y}_i) - E(f(y_s)\bar{Y}_i) \\ =& 0. \end{aligned}$$

Which means that $MSE(\hat{\bar{Y}}_i) \geq MSE(\bar{Y}^{BP}_i)$.
The above equality is the direct application of the following conditional expectation property

$$E(f(X)E(Y|X)) = E(f(X)Y).$$

Therefore, the best predictor $\bar{Y}^{BP}_i$ can be written as:

$$\begin{aligned} \bar{Y}^{BP}_i =& E(\bar{Y}_i|y_s) \\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} E[y_{ijk}|y_s] + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} E[y_{ijk}|y_s]\right]. \end{aligned} \tag{4}$$

2.2 Two-fold nested error lognormal model

Suppose that the ith small area contains M_i first-stage units or primary sampling units (or clusters) and that the jth cluster in the ith area contains N_{ij} second-stage units or observational (or simple) sampling units (elements). Let (Y_{ijk}, X_{ij}) be the y-value and x-value for the the kth element in the jth cluster from the ith area $(k = 1, 2, \ldots, N_{ij}; j = 1, 2, \ldots, M_i; i = 1, 2, \ldots, M)$. Under this population structure, we consider a two-stage sampling in each small area, where a sample s_i, of m_i clusters is selected from the ith sampled small area and, if the jth cluster is sampled, then a subsample, s_{ij}, of n_{ij} elements is selected from it. Without loss of generality, the sample values are denoted by (y_{ijk}, x_{ij}), $(k = 1, 2, \ldots, n_{ij}; j = 1, 2, \ldots, m_i; i = 1, 2, \ldots, m)$.
Under the aforementioned population structure, we obtain $\hat{\bar{Y}}^{MMSE}_i$ using the following nested error

two-fold regression model on the logarithm of the variable of interest. The proposed model in its general form for all population units, is given by

$$log(y_{ijk}) \equiv l_{ijk} = x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}, \quad \begin{array}{l} k = 1,2,\ldots,N_{ij} \\ j = 1,2,\ldots,M_i \\ i = 1,2,\ldots,M, \end{array} \tag{5}$$

where the area effects v_i, the cluster effects u_{ij} and the residual errors e_{ijk} are assumed to be mutually independent. Furthermore,

$$v_i \sim N(0,\sigma_v^2), \;\; u_{ij} \sim N(0,\sigma_u^2), \;\; e_{ijk} \sim N(0,\sigma_e^2).$$

The objective is to predict the value of small area population mean:

$$\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} y_{ijk}. \tag{6}$$

2.3 Minimun MSE predictor

Since the variance of a small area estimator based on the direct small area sample is excessively large, there is a need for constructing model based estimators with low mean squared prediction error (MSPE). This section introduces the minimum mean squared predicted error (MMSE), known also as Best/Bayes predictor (BP) of a function of a random vector in a finite population.
We assume that the sample values have the mentioned structure and follow the assumed model (5). Thus the sample model may be written as

$$log(y_{ijk}) \equiv l_{ijk} = x_{ij}^T\beta + v_i + u_{ij} + e_{ijk}, \quad \begin{array}{l} k = 1,2,\ldots,n_{ij} \\ j = 1,2,\ldots,m_i \\ i = 1,2,\ldots,m, \end{array} \tag{7}$$

where, for notational simplicity, the sample clusters, si, are denoted as $j = 1,2,\ldots,m_i$ and sample elements s_{ij} as $k = 1,2,\ldots,n_{ij}$.
Following **theorem.1**, the minimum MSE predictor of the $\bar{Y}_i$ is $E[\bar{Y}_i|(y,x)]$, where $(y,x) = \{y_{ijk}, i \in s, j \in s_i, k \in s_{ij}\} \cup \{x_{ij}, i = 1,\ldots,m; j = 1,\ldots,M_i\}$. Where s is the set of indices of those small areas that are in the sample.

Theorem 2. Under the assumed model, (5), the expression for the minimum MSE predictor is

$$\begin{aligned} \hat{\bar{Y}}_i^{MMSE} =& E[\bar{Y}_i|(y,x)] \\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} E[y_{ijk}|(y,x)] + \sum_{j\in\bar{s}_i}\sum_{k=1}^{N_{ij}} E[y_{ijk}|(y,x)]\right] \\ =& \frac{1}{N_i}\left[\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} E[\exp\{l_{ijk}\}|(y,x)] + \sum_{j\in\bar{s}_i}\sum_{k=1}^{N_{ij}} E[\exp\{l_{ijk}\}|(y,x)]\right]. \end{aligned} \tag{8}$$

This expression reflects two cases to be discussed, first: The sub-small area within the small area sampled and contains some observations from the sample, which corresponds to the second term to the right (8); second: The sub-small area within the small area sampled, and does not contain any observations from the sample. This corresponds to the third term to the right of (8) .

The model (7) in matrix form for each $j \in s_i$ is as follows:

$$l_{ij} = X_{ij}\beta + v_i 1_{n_{ij}} + u_{ij} 1_{n_{ij}} + e_{ij}, \tag{9}$$

where

$$l_{ij} = \begin{bmatrix} l_{ij1} \\ l_{ij2} \\ \vdots \\ l_{ijn_{ij}} \end{bmatrix}_{n_{ij}\times 1} ; \quad X_{ij} = \begin{bmatrix} x_{ij}^T \\ x_{ij}^T \\ \vdots \\ x_{ij}^T \end{bmatrix}_{n_{ij}\times p} = 1_{n_{ij}} \otimes x_{ij}^T ; \quad e_{ij} = \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijn_{ij}} \end{bmatrix}_{n_{ij}\times 1} .$$

From (9), the variance of the vector of transformed variable, l_{ij}, is given by

$$\begin{aligned} var(l_{ij}) =& \sigma_v^2 1_{n_{ij}} 1_{n_{ij}}^T + \sigma_u^2 1_{n_{ij}} 1_{n_{ij}}^T + \sigma_e^2 I_{n_{ij}} \\ =& \sigma_v^2 J_{n_{ij}} + \sigma_u^2 J_{n_{ij}} + \sigma_e I_{n_{ij}}. \end{aligned}$$

Then by a given i we have the vector, l_i, that combines the expressions represented in (9)

$$l_i = \begin{bmatrix} l_{i1}^T \\ l_{i2}^T \\ \vdots \\ l_{im_i}^T \end{bmatrix} = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im_i} \end{bmatrix} \beta + \begin{bmatrix} 1_{n_{i1}} \\ 1_{n_{i2}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix} v_i + \begin{bmatrix} 1_{n_{i1}} & & & \\ & 1_{n_{i2}} & & \\ & & \ddots & \\ & & & 1_{n_{im_i}} \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{im_i}^T \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{im_i} \end{bmatrix},$$

and its variance is given by

$$V_i = \begin{bmatrix} (\sigma_v^2 + \sigma_u^2) J_{n_{i1}} + \sigma_e^2 I_{n_{i1}} & \sigma_v^2 J_{n_{i1} n_{i2}} & \dots & \sigma_v^2 J_{n_{i1} n_{im_i}} \\ \sigma_v^2 J_{n_{i2} n_{i1}} & (\sigma_v^2 + \sigma_u^2) J_{n_{i2}} + \sigma_e^2 I_{n_{i2}} & \dots & \sigma_v^2 J_{n_{i2} n_{im_i}} \\ \vdots & & & \\ \sigma_v^2 J_{n_{im_i} n_{i1}} & \sigma_v^2 J_{n_{im_i} n_{i2}} & \dots & (\sigma_v^2 + \sigma_v^2) J_{n_{im_i}} + \sigma_e^2 I_{n_{im_i}} \end{bmatrix},$$

where 1_K is a vectors of $1's$ with a length K, I_K is an identity matrix and J_K is a matrix of $1's$ of dimension $K \times K$ respectively.
Now, we have a joint distribution of the expressions represented by (7) for a given small area i.

$$l_i = \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{in} \end{bmatrix} \sim N\Big(\begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im_i} \end{bmatrix} \beta, V_i \Big) \equiv N(X_i\beta, V_i).$$

It follows that

$$\begin{aligned} \bar{l}_i = \begin{bmatrix} \bar{l}_{i1} \\ \bar{l}_{i2} \\ \vdots \\ \bar{l}_{in} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T l_{i1} \\ \frac{1}{n_{i2}} 1_{n_{i2}}^T l_{i2} \\ \vdots \\ \frac{1}{n_{im_i}} 1_{n_{im_i}}^T l_{im_i} \end{bmatrix} = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}}^T & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}}^T \end{bmatrix} \begin{bmatrix} l_{i1} \\ l_{i2} \\ \vdots \\ l_{im_i} \end{bmatrix} \\ =& W_i l_i. \end{aligned}$$

Note that

$$W_i X_i \beta = \begin{bmatrix} x_{i1}^T \\ x_{i2}^T \\ \vdots \\ x_{in}^T \end{bmatrix} \beta = x_i \beta,$$

and

$$\begin{aligned}\bar{V}_i =& W_i V_i W_i^T = \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}}^T & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}}^T & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}}^T \end{bmatrix} V_i \begin{bmatrix} \frac{1}{n_{i1}} 1_{n_{i1}} & & & \\ & \frac{1}{n_{i2}} 1_{n_{i2}} & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} 1_{n_{im_i}} \end{bmatrix} \\ =& \begin{bmatrix} (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{i1}} \sigma_e^2 & \sigma_v^2 & \cdots & \sigma_v^2 \\ \sigma_v^2 & (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{i2}} \sigma_e^2 & \cdots & \sigma_v^2 \\ \vdots & & & \\ \sigma_v^2 & \sigma_v^2 & \cdots & (\sigma_v^2 + \sigma_u^2) + \frac{1}{n_{im_i}} \sigma_e^2 \end{bmatrix}.\end{aligned}$$

Then

$$\bar{l}_i \sim N(x_i \beta, \bar{V}_i),$$

Considering the case where the sub-small area, j, has some observations within the sample, it follows that

$$\begin{aligned} cov(u_{ij}, \bar{l}_i) &= cov(u_{ij}, W_i l_i) \\ &= C_1^{(j)T} W_i^T = W_i C_1^{(j)} = \sigma_u^2 \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m_i \times 1)} = \alpha^{(j)}, \\ cov(v_i, \bar{l}_i) &= cov(v_i, W_i l_i) = \sigma_v^2 W_i cov(l_i, v_i) \\ &= \sigma_v^2 W_i \begin{bmatrix} 1_{n_{i1}} \\ \vdots \\ 1_{n_{ij}} \\ \vdots \\ 1_{n_{im_i}} \end{bmatrix}_{(n_i \times 1)} = \sigma_v^2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{(m_i \times 1)} = \gamma. \end{aligned}$$

Now the joint distribution is

$$\begin{bmatrix} \bar{l}_i \\ u_{ij} \\ v_i \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i \beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & \alpha^{(j)} & \gamma \\ \alpha^{(j)T} & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \Big). \tag{10}$$

From (10), we have

$$E(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i) = \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} (\bar{l}_i - x_i \beta) \equiv \mu_{1j}$$

and

$$var\left(\begin{bmatrix} u_{ij} \\ v_i \end{bmatrix} \mid \bar{l}_i\right) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} \equiv \Sigma_{1j}.$$

Using the above conditional expressions and the moment generating function of the lognormal distribution,it follows

$$E(\exp\{v_i + u_{ij}\} \mid \bar{l}_i) = \exp\{1^T \mu_{1j} + \frac{1}{2} 1^T \Sigma_{1j} 1\}.$$

Now the expression of the second term in (8) is given by

$$\tilde{y}^*_{ijk} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T \beta + 1^T \mu_{1j} + \frac{1}{2} 1^T \Sigma_{1j} 1 + \frac{1}{2}\sigma_e^2\}. \tag{11}$$

Next we consider the second case with a sub-small area, r, in the sampled small area, but does not have any observation in the sample, it follows that

$$E(u_{ir} \mid l_i) = E(u_{ir} \mid \bar{l}_i) = 0.$$

Now the joint distribution is

$$\begin{bmatrix} \bar{l}_i \\ u_{ir} \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 0 & \gamma \\ 0^T & \sigma_u^2 & 0 \\ \gamma^T & 0 & \sigma_v^2 \end{bmatrix} \right). \tag{12}$$

Proceeding the same way as in the first case and using (12) we have

$$E\left(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i\right) = \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} (\bar{l}_i - x_i\beta) \equiv \mu_2$$

and

$$var\left(\begin{bmatrix} u_{ir} \\ v_i \end{bmatrix} \mid \bar{l}_i\right) = \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix} \equiv \Sigma_2.$$

Using the above results the moment generating function of the lognormal distribution, we have

$$E(\exp\{v_i + u_{ir}\} \mid \bar{l}_i) = \exp\{1^T \mu_2 + \frac{1}{2} 1^T \Sigma_2 1\},$$

and the expression of the third term in (8) is given by

$$\tilde{y}^{**}_{ijk} \equiv E(y_{ijk} \mid \bar{l}_i) = \exp\{x_{ij}^T \beta + 1^T \mu_2 + \frac{1}{2} 1^T \Sigma_2 1 + \frac{1}{2}\sigma_e^2\}. \tag{13}$$

Substituting the expressions (11) and (13) in (8), the minimum MSE predictor, under the assumption that $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$ is known, is given by

$$\begin{aligned} \hat{\bar{Y}}_i^{MMSE}(\theta) =& E[\bar{Y}_i | (y, x)] \\ =& \frac{1}{N_i} \left[\sum_{j,k \in s_i} y_{ijk} + \sum_{j \in s_i} \sum_{k \in \bar{s}_{ij}} \tilde{y}^*_{ijk}(\theta) + \sum_{j \in \bar{s}_i} \sum_{k=1}^{N_{ij}} \tilde{y}^{**}_{ijk}(\theta) \right]. \end{aligned} \tag{14}$$

2.4 Empirical Bayes predictor

In practice, θ is not unknown, so it is not possible to calculate (**??**). We replace the true value of θ with its consistent estimator to obtain the Emprical Bayes (EB) predictor. Let $\hat{\theta}^T = (\hat{\beta}, \hat{\sigma})$ be a restricted maximum likelihood (REML) estimator. By substituting the true θ in (**??**) with an estimator, we obtain

$$\hat{\hat{Y}}_i^{EB} = \frac{1}{N_i}\left\{\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} \tilde{y}^*_{ijk}(\hat{\theta}) + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} \tilde{y}^{**}_{ijk}(\hat{\theta})\right\}, \qquad (15)$$

where
$\hat{\hat{Y}}_i^{EB} = \hat{Y}_i^{MMSE}(\hat{\theta})$, $\hat{y}^{*EB}_{ijk} = \tilde{y}^*_{ijk}(\hat{\theta})$, and $\hat{y}^{**EB}_{ijk} = \tilde{y}^{**}_{ijk}(\hat{\theta})$.

2.5 Fisher-scoring algorithm under restricted maximum likelihood

The sample model (7) may be seen as a special case of a general linear mixed model with block diagonal covariance structure, involving fixed and random effects, and a small-area mean can be expressed as a linear combination of fixed effects and realized values of random effects i.e., a model composed by m independent submodels:

$$l_{ij} = col_{1\le k\le n_{ij}}(l_{ijk}),\ l_i = col_{1\le j\le m_i}(l_{ij}), l = col_{1\le j\le m}(l_i),$$

where

$$l_{ij} = 1_{n_{ij}}(x_{ij}^T\beta) + 1_{n_{ij}}v_i + 1_{n_{ij}}u_{ij} + e_{ij},$$

and

$$l_i = diag_{1\le j\le m_i}(1_{n_{ij}} \otimes x_{ij}^T\beta) + 1_{n_i}v_i + diag_{1\le j\le m_i}(1_{n_{ij}})u_i + e_i,$$

with $1_{n_i} = col_{1\le j\le m_i}(1_{n_{ij}})$.
Then, the matrix form of the model is

$$l = X\beta + Z_1 v + Z_2 u + e, \qquad (16)$$

where
$X = X_{n\times p}$, $Z_1 = diag_{1\le i\le m}(1_{n_i})$, $Z_2 = diag_{1\le i\le m}(diag_{1\le j\le m_i}(1_{n_{ij}}))_{n\times d}$,
$n = \sum_{i=1}^m n_i$, $n_i = \sum_{j=1}^{m_i} n_{ij}$, $d = \sum_{i=1}^m m_i$.
The model (16) can be rewritten in the following form

$$l = X\beta + Zw + e, \qquad (17)$$

where, $Z = (Z_1, Z_2)$ and $w = (v^T, u^T)^T$.
The variance, V, of l is given by $V(\sigma) = diag_{1\le i\le m}(V_i(\sigma))$,
where V_i is defined in the previous section and $\sigma = (\sigma_v^2, \sigma_u^2, \sigma_e^2)^T$ is the vector of unkown parameters involved in the covariance structure of the model.
Following Henderson (1975), the best linear unbiased estimator (BLUE) of β in (17) is given by

$$\tilde{\beta}(\sigma) = (X^T V^{-1}(\sigma)X)^{-1}X^T V^{-1}(\sigma)l.$$

Replacing an restricted maximum likelihood (REML) estimator, $\hat{\sigma}$, of σ in previous equation we obtain the so called empirical BLUE (EBLUE) $\hat{\beta} = \tilde{\beta}(\hat{\sigma})$.

The REML method maximizes the joint probability density of $n-p$ linear independent contrasts $\omega = Al$, where A^T is an $n \times (n-p)$ full column rank matrix satisfying $AA^T = I_{n-p}$ and $BX = 0$. Thus, the probability density function of ω does not depend on β and is given by

$$L(\omega|\sigma) = (2\pi)^{\frac{-(n-p)}{2}} |X^T X|^{1/2} |V(\sigma)|^{-1/2} |X^T V^{-1}(\sigma) X|^{-1/2} \exp\left[-\frac{1}{2} l^T P(\sigma) l\right],$$

where

$$P(\sigma) = V^{-1}(\sigma) - V^{-1}(\sigma) X (X^T V^{-1}(\sigma) X)^{-1} X^T V^{-1}(\sigma).$$

Note that $P(\sigma)$ satisfies $P(\sigma)V(\sigma)P(\sigma) = P(\sigma)$ and $P(\sigma)X = 0_n$.

The REML estimator of σ is the maximizer of $l_{REML}(\sigma) = logL(\omega|\sigma)$. The fact that the REML equations are nonlinear, they do not have closed analytical solutions. We adapt the iterative technique for solving the REML equations. A common variant of Newton-Raphson (NR) algorithm is Fisher-scoring method, which is appears to be slightly more robust to initial values than strict NR (Jennrich and Sampson, 1976), which replaces the inverse of the Hessian matrix by its expected value, which after allowing for a change in sign, turns out to be defined by the inverse of Fisher's information matrix.

Let $S(\sigma) = \partial l_{REML}(\sigma)/\partial\sigma = (S_1(\sigma), \ldots, S_3(\sigma))$ and $F(\sigma) = -E[\partial l_{REML}(\sigma)/\partial\sigma\partial\sigma^T] = (F_{qr}(\sigma))$ be the scores vector and the Fisher information matrix respectively. Using the fact that

$$\frac{\partial P(\sigma)}{\partial \sigma_s} = -P(\sigma) \frac{\partial V(\sigma)}{\partial \sigma_s} P(\sigma), \quad q = 1, 2, 3,$$

the first order partial derivative of $l_{REML}(\sigma)$ with respect to σ_s is given by

$$S_q(\sigma) = -\frac{1}{2} trace\left[P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_q}\right] + \frac{1}{2} l^T P(\sigma) \frac{\partial V(\sigma)}{\sigma_q} P(\sigma) l, \quad q = 1, 2, 3.$$

Then, taking the negative expectation of second order partial derivative of $l_{REML}(\theta)$ with respect to σ_q and σ_r, the element (q, r) of the Fisher information matrix is obtained by

$$F_{qr}(\sigma) = \frac{1}{2} trace\left[P(\sigma)\frac{\partial V(\sigma)}{\partial \sigma_q} P(\sigma) \frac{\partial V(\sigma)}{\partial \sigma_r}\right], \quad q, r = 1, \ldots, 4.$$

Then, assuming σ^s to be the value of the estimator at iteration s, the updating expression of the Fisher-scoring algorithm is

$$\sigma^{s+1} = \sigma^s + [F(\sigma^s)]^{-1} S(\sigma^s).$$

3 MSE estimation for the EB predictor

Since our goal is to use the predictor of the population mean in practice, we need to compute the mean square error of MMSE predictor, as well as its estimator.
The MSE of an EB predictor, or EBLUP under normal assumption, can be written as a sum of two terms (Kackar and Harville, 1984; Prasad and Rao, 1990):

$$MSE(\hat{\bar{Y}}_i^{EB}) = M_{i1}(\theta) + M_{i2}(\theta), \tag{18}$$

where
$M_{1i} = E[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2]$ and $M_{2i} = E[(\hat{\bar{Y}}_i^{MMSE} - \hat{\bar{Y}}_i^{EB})^2]$.

The first term, $M_{i1}(\theta)$, is the variance of the error in the minimum MSE predictor (14), the predictor obtained under the true (unknown) θ. The second term accounts for variability of the predictor due to estimation of the parameters in $\theta = (\beta^T, \sigma^T)^T$.
In the next two subsections, we give a closed form expression for $M_{i1}(\theta)$ and a linear approximation for $M_{i2}(\theta)$ respectively.

3.1 Derivation of the expressions of M_{1i}

In this subsection we derive the expression of the first term of the right-hand side of (18).

$$
\begin{aligned}
M_{1i} &= E[(\bar{Y}_i - \hat{\bar{Y}}_i^{MMSE})^2] = E[(\bar{Y}_i - E[\bar{Y}_i|(y,x)])^2] = E[var(\bar{Y}_i|(y,x))]\\
&= E\left\{\frac{1}{N_i^2}\left[var\left(\sum_{j\in s_i}\sum_{k\in s_{ij}} y_{ijk} + \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk} + \sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right)\right]\right\}\\
&= E\left\{\frac{1}{N_i^2}\left[var\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}|(y,x)\right) + var\left(\sum_{j\in \bar{s}_i}\sum_{k=1}^{N_{ij}} y_{ijk}|(y,x)\right) + 2cov\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}, \sum_{r\in \bar{s}_i}\sum_{p=1}^{N_{ir}} y_{irp}|(y,x)\right)\right]\right\}\\
&= \frac{1}{N_i^2}\Big(E(V_1) + E(V_2) + 2E(C_1)\Big). \qquad (19)
\end{aligned}
$$

Starting by the first term, it follows that

$$
\begin{aligned}
V_1 &= var\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}|(y,x)\right)\\
&= cov\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}, \sum_{q\in s_i}\sum_{p\in \bar{s}_{iq}} y_{iqp}|(y,x)\right)\\
&= \sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in \bar{s}_{ir}} cov(y_{ijk}, y_{irp}|(y,x)). \qquad (20)
\end{aligned}
$$

The conditional covariance is defined by

$$
cov(y_{ijk}, y_{irp}|(y,x)) = E[y_{ijk}y_{iqp}|(y,x)] - E[y_{ijk}|(y,x)]E[y_{iqp}|(y,x)].
$$

Note that

$$
\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ij} \\ u_{iq} \end{bmatrix} \sim N\left(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 2\gamma & \alpha^{(j)} & \alpha^{(q)} \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ \alpha^{(j)T} & 0 & \sigma_u^2 & 0 \\ \alpha^{(q)T} & 0 & 0 & \sigma_u^2 \end{bmatrix}\right). \qquad (21)
$$

From (21), it follows that

$$
E\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i\right) = \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ \alpha^{(r)T} \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{1jq} \qquad (22)
$$

and

$$
var\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{iq} \end{bmatrix} \mid \bar{l}_i\right) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ \alpha^{(q)T} \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & \alpha^{(q)} \end{bmatrix} \equiv \Sigma_{1jq}. \qquad (23)
$$

Taking into acount for different cases that can occcure between j and q, and between k and p, we have

$$\begin{aligned}
&E[y_{ijk}y_{iqp}|(y,x)] = \\
&[\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\}I(k=p)+ \\
&\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\}I(k\neq p)]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\}]I(j\neq q).
\end{aligned} \tag{24}$$

and

$$\begin{aligned}
&E[y_{ijk}|(y,x)]E[y_{iqp}|(x,y)] = \\
&= [\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\}]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j}+\mu_{1q}) + \frac{1}{2}1^T(\Sigma_{1j}+\Sigma_{1q})1 + \sigma_e^2\}]I(j\neq q).
\end{aligned} \tag{25}$$

We have from (24) and (25) that

$$\begin{aligned}
&cov(y_{ijk}, y_{iqp} \mid \bar{l}_i) = \\
&[(\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1 + \sigma_e^2)\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\})I(k=p)+ \\
&(\exp\{2(x_{ij}^T\beta + 1^T\mu_{1j} + 1^T\Sigma_{1j}1) + \sigma_e^2\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_{1j}) + 1^T\Sigma_{1j}1 + \sigma_e^2\})I(k\neq p)]I(j=q)+ \\
&[\exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T\mu_{1jq} + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + 1^T(\mu_{1j}+\mu_{1q})+ \\
&\frac{1}{2}1^T(\Sigma_{1j}+\Sigma_{1q})1 + \sigma_e^2\}]I(j\neq q).
\end{aligned}$$

From the subsection 2.3, it follows that

$$\begin{aligned}
E(1^T\mu_{1j}) =& 0 \\
var(1^T\mu_{1j}) =& 1^T V_{\mu_{1j}} 1,
\end{aligned}$$

where

$$V_{\mu_{1j}} = var(\mu_{1j}) = \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix}.$$

Then we have

$$\begin{aligned}
&E[cov(y_{ijk}, y_{iqp} \mid \bar{l}_i)] = \\
&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T\Sigma_{1j}1 + \sigma_e^2\}(\exp\{1^T\Sigma_{1j}1 + \sigma_e^2\} - 1)I(k=p)+ \\
&(\exp\{1^T\Sigma_{1j}1\} - 1)I(k\neq p)]I(j=q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T V_{\mu_{1jq}} 1 + \frac{1}{2}1^T\Sigma_{1jq}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1+ \\
&\frac{1}{2}1^T(\Sigma_{1j}+\Sigma_{1q})1 + \sigma_e^2\}]I(j\neq q).
\end{aligned}$$

Within the expression above we add and subtract $\exp\{1^T\Sigma_{1j}1\}I(k=p)$

$$
\begin{aligned}
&E[cov(y_{ijk}, y_{iqp} \mid \bar{l}_i)] = \\
&[\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}(\exp\{1^T \Sigma_{1j} 1 + \sigma_e^2\} - \exp\{1^T \Sigma_{1j} 1\}) I(k=p) + \\
&\exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}(\exp\{1^T \Sigma_{1j} 1\} - 1)] I(j=q) \\
&+ [\exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T V_{\mu_{1jq}} 1 + \frac{1}{2}1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}] I(j \neq q) \\
&= v_{11j} I(k=p) I(j=q) + v_{12j} I(j=q) + v_{13jq} I(j \neq q),
\end{aligned}
$$

where

$$
\begin{aligned}
v_{11j} &= \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}(\exp\{1^T \Sigma_{1j} 1 + \sigma_e^2\} - \exp\{1^T \Sigma_{1j} 1\}) \\
v_{12j} &= \exp\{2(x_{ij}^T\beta + 1^T V_{\mu_{1j}} 1) + 1^T \Sigma_{1j} 1 + \sigma_e^2\}(\exp\{1^T \Sigma_{1j} 1\} - 1). \\
v_{13jr} &= \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T V_{\mu_{1jq}} 1 + \frac{1}{2}1^T \Sigma_{1jq} 1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{iq}^T\beta + \frac{1}{2}1^T (V_{\mu_{1j}} + V_{\mu_{1q}}) 1 + \\
&\quad \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_{1q}) 1 + \sigma_e^2\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E(V_1) =& \sum_{j\in s_i} \sum_{k\in \bar{s}_{ij}} \sum_{q\in s_i} \sum_{p\in \bar{s}_{iq}} v_{11j} I(p=k) I(q=j) + \sum_{j\in s_i} \sum_{k\in \bar{s}_{ij}} \sum_{q\in s_i} \sum_{p\in \bar{s}_{iq}} v_{12j} I(q=j) + \\
& \sum_{j\in s_i} \sum_{k\in \bar{s}_{ij}} \sum_{q\in s_i} \sum_{p\in \bar{s}_{iq}} v_{13jq} I(q \neq j) \\
=& \sum_{j\in s_i} \sum_{k\in \bar{s}_{ij}} v_{11j} \left[\sum_{q\in s_i} \sum_{p\in \bar{s}_{iq}} I(p=k) I(q=j) \right] + \sum_{j\in s_i} \sum_{k\in \bar{s}_{ij}} v_{12j} \left[\sum_{q\in s_i} \sum_{p\in \bar{s}_{ir}} I(q=j) \right] + \\
& \sum_{j\in s_i} \sum_{q\in s_i} v_{13jq} I(q \neq j) \left[\sum_{k\in \bar{s}_{ij}} \sum_{p\in \bar{s}_{ir}} 1 \right] \\
=& \sum_{j\in s_i} (N_{ij} - n_{ij}) v_{11j} + \sum_{j\in s_i} (N_{ij} - n_{ij})^2 v_{12j} + \sum_{j\in s_i} \sum_{q\in s_i} (N_{ij} - n_{ij})(N_{iq} - n_{iq}) v_{13jq} I(q \neq j).
\end{aligned}
\tag{26}
$$

Here we calculate the expression of the second term of (19)

$$
\begin{aligned}
V_2 =& var\left(\sum_{g\in \bar{s}_i} \sum_{k=1}^{N_{ig}} y_{igk} | (y,x) \right) \\
=& cov\left(\sum_{g\in \bar{s}_i} \sum_{k=1}^{N_{ig}} y_{igk}, \sum_{r\in \bar{s}_i} \sum_{k=1}^{N_{ir}} y_{irk} | (y,x) \right) \\
=& \sum_{g\in \bar{s}_i} \sum_{k=1}^{N_{ig}} \sum_{r\in \bar{s}_i} \sum_{p=1}^{N_{ir}} cov(y_{ijk}, y_{irp} | (y,x)).
\end{aligned}
\tag{27}
$$

The covariance in (27) is given by

$$
cov(y_{igk}, y_{irp}|(y,x)) = E[y_{igk} y_{irp}|(y,x)] - E[y_{igk}|(y,x)] E[y_{irp}|(y,x)].
$$

Note that

$$\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \sim N\Big(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V}_i & 2\gamma & 0 & 0 \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ 0^T & 0 & \sigma_u^2 & 0 \\ 0^T & 0 & 0 & \sigma_u^2 \end{bmatrix} \Big). \tag{28}$$

From (28), it follows that

$$E\Big(\begin{bmatrix} 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \Big) = \begin{bmatrix} 2\gamma^T \\ 0^T \\ 0^T \end{bmatrix} \bar{V}_i^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_3, \tag{29}$$

and

$$var\Big(\begin{bmatrix} 2v_i \\ u_{ig} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \Big) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ 0^T \\ 0^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 2\gamma & 0 & 0 \end{bmatrix} \equiv \Sigma_3. \tag{30}$$

Proceeding the same way as in the case of V_1, it follows that

$$\begin{aligned} &E[y_{igk}y_{irp}|(y,x)] = \\ &[\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\} I(k=p) + \\ &\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2\} I(k \neq p)] I(g=r) + \\ &[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\}] I(j \neq r). \end{aligned} \tag{31}$$

and

$$\begin{aligned} &E[y_{igk}|(y,x)]E[y_{irp}|(x,y)] = \\ &[\exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}] I(g=r) + \\ &\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}] I(g \neq r). \end{aligned} \tag{32}$$

we have from (31) and (32) that

$$\begin{aligned} &cov(y_{igk}, y_{irp} \mid \bar{l}_i) = \\ &[(\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2)\} - \exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}) I(k=p) \\ &+(\exp\{2(x_{ig}^T\beta + 1^T\mu_2 + 1^T\Sigma_2 1) + \sigma_e^2\} - \exp\{2(x_{ij}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}) I(k \neq p)] I(g=r) \\ &+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}] I(g \neq r). \end{aligned} \tag{33}$$

Factorizing and adding and subtracting $\exp\{1^T\Sigma_2 1\}$ the expression under the $I(j=r)$ we have

$$\begin{aligned} &cov(y_{igk}, y_{irp} \mid \bar{l}_i) = \\ &\exp\{2(x_{ig}^T\beta + 1^T\mu_2) + 1^T\Sigma_2 1 + \sigma_e^2\}[(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\}) I(k=p) + \\ &(\exp\{1^T\Sigma_2 1\} - 1)] I(g=r) \\ &+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T\mu_3 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T\mu_2 + 1^T\Sigma_2 1 + \sigma_e^2\}] I(j \neq r). \end{aligned}$$

From the subsection 2.3 notice that

$$\begin{aligned} E(1^T\mu_2) =& 0 \\ var(1^T\mu_2) =& 1^T V_{\mu_2} 1, \end{aligned}$$

where

$$V_{\mu_2} = var(\mu_2) = \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}_i^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix}.$$

Then we have

$$\begin{aligned}
&E[cov(y_{igk}, y_{irp} \mid \bar{l}_i)] = \\
&\exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}[(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\})I(k=p) + (\exp\{1^T\Sigma_2 1\} - 1)]I(g=r) \\
&+[\exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_3} 1 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T V_{\mu_2} 1 + 1^T\Sigma_2 1 + \sigma_e^2\}]I(g \neq r) \\
&=v_{21g} I(k=p)I(g=r) + v_{22g} I(g=r) + v_{23gr} I(g \neq r),
\end{aligned}$$

where

$$\begin{aligned}
v_{21g} &= \exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1 + \sigma_e^2\} - \exp\{1^T\Sigma_2 1\}), \\
v_{22g} &= \exp\{2(x_{ig}^T\beta + 1^T V_{\mu_2} 1) + 1^T\Sigma_2 1 + \sigma_e^2\}(\exp\{1^T\Sigma_2 1\} - 1), \\
v_{23gr} &= \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 1^T V_{\mu_3} 1 + \frac{1}{2}1^T\Sigma_3 1 + \sigma_e^2\} - \exp\{x_{ig}^T\beta + x_{ir}^T\beta + 21^T V_{\mu_2} 1 + 1^T\Sigma_2 1 + \sigma_e^2\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
E(V_2) &= \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{21g} I(p=k)I(r=g) + \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{22g} I(r=g) + \sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} v_{23jr} I(r \neq g) \\
&= \sum_{g\in\bar{s}_i} N_{ig} v_{21g} + \sum_{g\in\bar{s}_i} N_{ig}^2 v_{22g} + \sum_{g\in\bar{s}_i}\sum_{r\in\bar{s}_i} N_{ig}N_{ir} v_{23gr} I(r \neq g). \qquad (34)
\end{aligned}$$

The expression of the third term of (19) is given by

$$\begin{aligned}
C_1 =& cov\left(\sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} y_{ijk}, \sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} y_{irp} | (y,x)\right) \\
=& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} cov(y_{ijk}, y_{irp} | (y,x)). \qquad (35)
\end{aligned}$$

The covariance is expressed as follows

$$cov(y_{ijk}, y_{irp}|(y,x)) = E[y_{ijk} y_{irp}|(y,x)] - E[y_{ijk}|(y,x)]E[y_{irp}|(y,x)].$$

Then, note that

$$\begin{bmatrix} \bar{l}_i \\ 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \sim N\Bigg(\begin{bmatrix} x_i\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \bar{V} & 2\gamma & \alpha^{(j)} & 0 \\ 2\gamma^T & 4\sigma_v^2 & 0 & 0 \\ \alpha^{(j)T} & 0 & \sigma_u^2 & 0 \\ 0^T & 0 & 0 & \sigma_u^2 \end{bmatrix} \Bigg). \qquad (36)$$

From (36), it follows that

$$E\Bigg(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \Bigg) = \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1}(\bar{l}_i - x_i\beta) \equiv \mu_{4j}, \qquad (37)$$

and

$$var\left(\begin{bmatrix} 2v_i \\ u_{ij} \\ u_{ir} \end{bmatrix} \mid \bar{l}_i \right) = \begin{bmatrix} 4\sigma_v^2 & 0 & 0 \\ 0 & \sigma_u^2 & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix} - \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & 0 \end{bmatrix} \equiv \Sigma_{4j}. \tag{38}$$

Then by referring $j \in s_i$ and $r \in \bar{s}_i$ it follows

- $E[y_{ijk}y_{irp}|(y,x)]$, from (37) and (38) we have

$$\begin{aligned} &E[y_{ijk}y_{irp}|(y,x)] = \\ &E[\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 2v_i + u_{ij} + u_{ir} + e_{ijk} + e_{irp}\}|(y,x)] \\ =&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T\mu_{4j} + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\}. \end{aligned} \tag{39}$$

- $E[y_{ijk}|(y,x)]E[y_{irp}|(y,x)]$, from (11) and (13)

$$\begin{aligned} &E[y_{ijk}|(y,x)]E[y_{irp}|(y,x)] = \\ &\exp\{x_{ij}^T\beta + 1^T\mu_{1j} + \frac{1}{2}1^T\Sigma_{1j}1 + \frac{1}{2}\sigma_e^2\}\exp\{x_{ir}^T\beta + 1^T\mu_2 + \frac{1}{2}1^T\Sigma_2 1 + \frac{1}{2}\sigma_e^2\} \\ =&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(\mu_{1j} + \mu_2) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}. \end{aligned} \tag{40}$$

Then from (39) and (40) we have

$$\begin{aligned} cov(y_{ijk}, y_{irp}|\bar{l}_i) =& \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T\mu_{4j} + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(\mu_{1j} + \mu_2) + \\ &\frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}. \end{aligned} \tag{41}$$

By (41) it follows

$$\begin{aligned} &E[cov(y_{ijk}, y_{irp}|\bar{l}_i)] \\ =&\exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^TV_{\mu_{4j}}1 + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\} - \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(V_{\mu_{1j}} + V\mu_2)1 + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\} \\ =&c_{1jr} - c_{2jr}, \end{aligned} \tag{42}$$

where

$$\begin{aligned} V_{\mu_{4j}} =& \begin{bmatrix} 2\gamma^T \\ \alpha^{(j)T} \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 2\gamma & \alpha^{(j)} & 0 \end{bmatrix}, \\ c_{1jr} =& \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^TV_{\mu_{4j}}1 + \frac{1}{2}1^T\Sigma_{4j}1 + \sigma_e^2\}, \\ c_{2jr} =& \exp\{x_{ij}^T\beta + x_{ir}^T\beta + 1^T(V_{\mu_{1j}} + V\mu_2) + \frac{1}{2}1^T(\Sigma_{1j} + \Sigma_2)1 + \sigma_e^2\}. \end{aligned}$$

From (42) it follows that

$$\begin{aligned} E(C_1) =& \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}}(c_{1ijr} - c_{2ijr}) \\ =& \sum_{j\in s_i}\sum_{r\in\bar{s}_i}(N_{ij} - n_{ij})N_{ir}(c_{1ijr} - c_{2ijr}). \end{aligned} \tag{43}$$

3.2 Derivation of the expressions of M_{2i}

In this subsection we derive the expression of the second term of the right-hand side of (18).

$$
\begin{aligned}
M_{2i}(\theta) &= E[(\bar{Y}_i^{MMSE}(\theta) - \bar{Y}_i^{MMSE}(\hat{\theta}))] \\
&= \frac{1}{N_i^2} E\left[\left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}^{MMSE}(\theta) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\theta)\right) - \left(\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} y_{ijk}^{MMSE}(\hat{\theta}) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} y_{irk}^{MMSE}(\hat{\theta})\right)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{N_i^2} E\left[\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}} \left(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})\right) + \sum_{r\in\bar{s}_i}\sum_{k=1}^{N_{ir}} \left(y_{irk}^{MMSE}(\theta) - y_{irk}^{MMSE}(\hat{\theta})\right)\right]^2 \\
=&\frac{1}{N_i^2} E\left[\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in \bar{s}_{iq}} \left(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})\right)\left(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta})\right) + \right. \\
& 2\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} \left(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})\right)\left(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})\right) + \\
& \left.\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} \left(y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta})\right)\left(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})\right)\right]. \\
=&\frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{q\in s_i}\sum_{p\in \bar{s}_{iq}} E\left[\left(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})\right)\left(y_{iqp}^{MMSE}(\theta) - y_{iqp}^{MMSE}(\hat{\theta})\right)\right] + \\
& 2\frac{1}{N_i^2}\sum_{j\in s_i}\sum_{k\in \bar{s}_{ij}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[\left(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta})\right)\left(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})\right)\right] + \\
& \frac{1}{N_i^2}\sum_{g\in\bar{s}_i}\sum_{k=1}^{N_{ig}}\sum_{r\in\bar{s}_i}\sum_{p=1}^{N_{ir}} E\left[\left(y_{igk}^{MMSE}(\theta) - y_{igk}^{MMSE}(\hat{\theta})\right)\left(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})\right)\right] \\
=&\frac{1}{N_i^2}\left(H_1 + 2H_2 + H_3\right). \qquad (44)
\end{aligned}
$$

Note that,

$$
\begin{aligned}
&(y_{ijk}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\hat{\theta}))(y_{irp}^{MMSE}(\theta) - y_{irp}^{MMSE}(\hat{\theta})) = \\
&y_{ijk}^{MMSE}(\theta)y_{irp}^{MMSE}(\theta) - y_{ijk}^{MMSE}(\theta)y_{irp}^{MMSE}(\hat{\theta}) - y_{ijk}^{MMSE}(\hat{\theta})y_{irp}^{MMSE}(\theta) + y_{ijk}^{MMSE}(\hat{\theta})y_{irp}^{MMSE}(\hat{\theta}).
\end{aligned}
\qquad (45)
$$

Now we need to find the approximation of $y_{ijk}^{MMSE}(\hat{\theta})$. By (11) we have

$$
\begin{aligned}
\tilde{y}_{ijk}^{*}(\hat{\theta}) =& \exp\{x_{ij}^T\hat{\beta} + 1^T\mu_{1j}(\hat{\theta}) + \frac{1}{2}1^T\Sigma_{1j}(\hat{\theta})1 + \frac{1}{2}\hat{\sigma_e^2}\} \\
=& \exp\{\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})\},
\end{aligned}
\qquad (46)
$$

where

$$
\begin{aligned}
\Delta_1(\hat{\theta}) =& 1^T\mu_{1j}(\hat{\theta}) \\
\Omega_1(\hat{\theta}) =& x_{ij}^T\hat{\beta} + \frac{1}{2}(1^T\Sigma_{1j}(\hat{\theta})1 + \hat{\sigma_e^2}).
\end{aligned}
$$

As Δ_1 and Ω_1 are functions of $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$. When its estimate $\hat{\theta}$ is used, $\Delta_1(\hat{\theta})$ and $\Omega_1(\hat{\theta})$ can be expanded respectively around $\Delta_1(\theta)$ and $\Omega_1(\theta)$, by a Taylor series as

$$\begin{aligned}
\Delta_1(\hat{\theta}) \approx & \Delta_1(\theta) + (\hat{\theta}-\theta)^T \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Delta_1(\theta) + (\hat{\theta}-\theta)^T \Delta_1^*(\theta) \\
\Omega_1(\hat{\theta}) \approx & \Omega_1(\theta) + (\hat{\theta}-\theta)^T \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Omega_1(\theta) + (\hat{\theta}-\theta)^T \Omega_1^*(\theta),
\end{aligned} \tag{47}$$

where the expressions in $\frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are calculated as follows

$$\Delta_1^*(\theta) = \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Delta_1(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix} = 1^T \begin{bmatrix} \frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix}$$

with

$$\begin{aligned}
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\beta}}(\theta) =& -\begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} x_i, \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) =& \Big(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \Big) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) =& \Big(\begin{bmatrix} 1_j^T \\ \\ 0^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \Big) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\
\frac{\partial \mu_{1j}(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) =& -\begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta),
\end{aligned}$$

where

$$1_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(m_i \times 1)} \quad ; \quad J = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & & \\ 1 & \dots & 1 \end{bmatrix}_{(m_i \times m_i)} \quad ; \quad D = \begin{bmatrix} \frac{1}{n_{i1}} & & & \\ & \frac{1}{n_{i2}} & & \\ & & \ddots & \\ & & & \frac{1}{n_{im_i}} \end{bmatrix}_{(m_i \times m_i)},$$

and the expressions in $\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta}$ are obtained as follows

$$\Omega_1^*(\theta) = \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},$$

with

$$
\begin{aligned}
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\beta}}(\theta) =& x_{ij}^T, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_{1j}^{(v)} 1, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_{1j}^{(u)} 1, \\
\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) =& \frac{1}{2} 1^T \frac{\partial \Sigma_{1j}(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) 1 + \frac{1}{2} = \frac{1}{2} 1^T \Sigma_{1j}^{(e)} 1 + \frac{1}{2},
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_{1j}^{(v)} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} + \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix}, \\
\Sigma_{1j}^{(u)} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 1_j^T \\ \\ 0^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} + \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} I \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 1_j & 0 \end{bmatrix}, \\
\Sigma_{1j}^{(e)} &= \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix}.
\end{aligned}
$$

From (46) and (47) we have

$$
\begin{aligned}
\tilde{y}_{ijk}^{*}(\hat{\theta}) =& \exp\{x_{ij}^T \hat{\beta} + 1^T \mu_{1j}(\hat{\theta}) + \frac{1}{2} 1^T \Sigma_{1j}(\hat{\theta}) 1 + \frac{1}{2} \hat{\sigma_e^2}\} \\
\approx& \exp\{\Delta_1(\theta) + (\hat{\theta} - \theta)^T \Delta_1^*(\theta) + \Omega_1(\theta) + (\hat{\theta} - \theta)^T \Omega_1^*(\theta)\}. \qquad (48)
\end{aligned}
$$

Then by (13)

$$
\begin{aligned}
\tilde{y}_{ijk}^{**} =& \exp\{x_{ij}^T \hat{\beta} + 1^T \mu_2(\hat{\theta}) + \frac{1}{2} 1^T \Sigma_2(\hat{\theta}) 1 + \frac{1}{2} \hat{\sigma_e^2}\} \\
=& \exp\{\Delta_2(\hat{\theta}) + \Omega_2(\hat{\theta})\}, \qquad (49)
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_2(\hat{\theta}) =& 1^T \mu_2(\hat{\theta}), \\
\Omega_2(\hat{\theta}) =& x_{ij}^T \hat{\beta} + \frac{1}{2}(1^T \Sigma_2(\hat{\theta}) 1 + \hat{\sigma_e^2}).
\end{aligned}
$$

Taking into account that Δ_2 and Ω_2 are functions of $\theta = (\beta, \sigma_v^2, \sigma_u^2, \sigma_e^2)^T$. When its estimate $\hat{\theta}$ is used, $\Delta_2(\hat{\theta})$ and $\Omega_2(\hat{\theta})$ can be expanded respectively around $\Delta_2(\theta)$ and $\Omega_2(\theta)$, by a Taylor series as

$$
\begin{aligned}
\Delta_2(\hat{\theta}) \approx& \Delta_2(\theta) + (\hat{\theta} - \theta)^T \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Delta_2(\theta) + (\hat{\theta} - \theta)^T \Delta_2^*(\theta), \\
\Omega_2(\hat{\theta}) \approx& \Omega_2(\theta) + (\hat{\theta} - \theta)^T \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\theta}}\Big|_{\hat{\theta}=\theta} = \Omega_2(\theta) + (\hat{\theta} - \theta)^T \Omega_2^*(\theta), \qquad (50)
\end{aligned}
$$

where the expressions in $\left.\frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta}$ are calculated as follows

$$\Delta_2^*(\theta) = \left.\frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Delta_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix} = 1^T \begin{bmatrix} \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},$$

with

$$\begin{aligned} \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) &= -\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1} x_i, \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) &= \left(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \right) \bar{V}^{-1}(\bar{l}_i - x_i\beta), \\ \frac{\partial \mu_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) &= -\begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta), \end{aligned}$$

and the expressions in $\left.\frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta}$ are obtained as follows

$$\Omega_2^*(\theta) = \left.\frac{\partial \Omega_1(\hat{\theta})}{\partial \hat{\theta}}\right|_{\hat{\theta}=\theta} = \begin{bmatrix} \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) \end{bmatrix},$$

with

$$\begin{aligned} \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\beta}}(\theta) &= x_{ij}^T \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) &= \frac{1}{2} 1^T \frac{\partial \Sigma_2(\hat{\theta})}{\partial \hat{\sigma_v^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_2^{(v)} 1, \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) &= \frac{1}{2} 1^T \frac{\partial \Sigma_2(\hat{\theta})}{\partial \hat{\sigma_u^2}}(\theta) 1 = \frac{1}{2} 1^T \Sigma_2^{(u)} 1, \\ \frac{\partial \Omega_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) &= \frac{1}{2} 1^T \frac{\partial \Sigma_2(\hat{\theta})}{\partial \hat{\sigma_e^2}}(\theta) 1 + \frac{1}{2} = \frac{1}{2} 1^T \Sigma_2^{(e)} 1 + \frac{1}{2}, \end{aligned}$$

where

$$\begin{aligned} \Sigma_2^{(v)} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0^T \\ 1^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix} + \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix}, \\ \Sigma_2^{(u)} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} I \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix}, \\ \Sigma_2^{(e)} &= \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1} \begin{bmatrix} 0 & \gamma \end{bmatrix}. \end{aligned}$$

From (49) and (50) we have

$$
\begin{aligned}
\tilde{y}^{**}_{ijk}(\hat{\theta}) =& \exp\{x_{ij}^T\hat{\beta} + 1^T\mu_2(\hat{\theta}) + \frac{1}{2}1^T\Sigma_2(\hat{\theta})1 + \frac{1}{2}\hat{\sigma_e^2}\} \\
\approx& \exp\{\Delta_2(\theta) + (\hat{\theta}-\theta)^T\Delta_2^*(\theta) + \Omega_2(\theta) + (\hat{\theta}-\theta)^T\Omega_2^*(\theta)\}. \qquad (51)
\end{aligned}
$$

In the both above cases, the approximation can be represented as follows

$$
\begin{aligned}
\Delta(\hat{\theta}) + \Omega(\hat{\theta}) \approx& \Delta(\theta) + \Omega(\theta) + \frac{\partial\Delta^T}{\partial\hat{\theta}}(\theta)(\hat{\theta}-\theta) + \frac{\partial\omega^T}{\partial\hat{\theta}}(\theta)(\hat{\theta}-\theta) \\
=& \Delta(\theta) + \Omega(\theta) + \Big(\frac{\partial\Delta^T}{\partial\hat{\beta}}(\beta) + \frac{\partial\Omega^T}{\partial\hat{\beta}}(\beta)\Big)(\hat{\beta}-\beta) + \Big(\frac{\partial\Delta^T}{\partial\hat{\sigma}}(\sigma) + \frac{\partial\Omega^T}{\partial\hat{\sigma}}(\sigma)\Big)(\hat{\sigma}-\sigma), \quad (52)
\end{aligned}
$$

where $\sigma = (\sigma_v^2, \sigma_u^2, \sigma_e^2)^T$.
In continuation we discuss different scenarios for each case

- Case 1:

  1 Expressions in H_1, i.e $j, q \in s_i$ and $j = q$

$$
\begin{aligned}
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta) =& \exp\{2(\Delta_1(\theta) + \Omega_1(\theta))\}, \\
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta}) \approx& \exp\{2(\Delta_1(\theta) + \Omega_1(\theta)) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta) + \Omega_1^*(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma) + \Omega_1^*(\sigma))\}, \\
\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta}) \approx& \exp\{2(\Delta_1(\theta) + \Omega_1(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta) + \Omega_1^*(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma) + \Omega_1^*(\sigma)))\}. \qquad (53)
\end{aligned}
$$

  2 Expressions in H_1, i.e $j, q \in s_i$ and $j \neq q$

$$
\begin{aligned}
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta) =& \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta)\}, \\
\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta}) \approx& \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^{(q)*}(\beta) + \Omega_1^{(q)*}(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_1^{(q)*}(\sigma) + \Omega_1^{(q)*}(\sigma))\}, \\
\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta}) \approx& \exp\{\Delta_1^{(j)}(\theta) + \Omega_1^{(j)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^{(j)*}(\beta) + \Omega_1^{(j)*}(\beta)) + (\hat{\sigma}-\sigma)^T(\Delta_1^{(j)*}(\sigma)) + \\
& \Omega_1^{(j)*}(\sigma)) + \Delta_1^{(q)}(\theta) + \Omega_1^{(q)}(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^{(r)*}(\beta) + \Omega_1^{(r)*}(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_1^{(q)*}(\sigma) + \Omega_1^{(q)*}(\sigma))\}. \qquad (54)
\end{aligned}
$$

- Case 2:

  1 Expressions in H_2, i.e $j \in s_i$ and $r \in \bar{s}_i$

$$
\begin{aligned}
\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta) =& \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\theta) + \Omega_2(\theta)\}, \\
\tilde{y}^*_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{\Delta_1(\theta) + \Omega_1(\theta) + \Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma))\}, \\
\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{\Delta_1(\theta) + \Omega_1(\theta) + (\hat{\beta}-\beta)^T(\Delta_1^*(\beta) + \Omega_1^*(\beta)) + \\
& (\hat{\sigma}-\sigma)^T(\Delta_1^*(\sigma) + \Omega_1^*(\sigma)) + \\
& \Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta}-\beta)^T(\Delta_2^*(\beta) + \Omega_2^*(\beta)) + (\hat{\sigma}-\sigma)^T(\Delta_2^*(\sigma) + \Omega_2^*(\sigma))\}. \\
& \qquad (55)
\end{aligned}
$$

- Case 3:

1 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g = r$

$$
\begin{aligned}
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta) =& \exp\{2(\Delta_2(\theta) + \Omega_2(\theta))\},\\
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{2(\Delta_2(\theta) + \Omega_2(\theta)) + (\hat{\beta} - \beta)^T(\Delta^*_2(\beta) + \Omega^*_2(\beta)) +\\
& (\hat{\sigma} - \sigma)^T(\Delta^*_2(\sigma) + \Omega^*_2(\sigma))\},\\
\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{2(\Delta_2(\theta) + \Omega_2(\theta) + (\hat{\beta} - \beta)^T(\Delta^*_2(\beta) + \Omega^*_2(\beta)) +\\
& (\hat{\sigma} - \sigma)^T(\Delta^*_2(\sigma) + \Omega^*_2(\sigma)))\}.
\end{aligned} \tag{56}
$$

2 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g \neq r$

$$
\begin{aligned}
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta) =& \exp\{\Delta^{(g)}_2(\theta) + \Omega^{(g)}_2(\theta) + \Delta^{(r)}_2(\theta) + \Omega^{(r)}_2(\theta)\},\\
\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{\Delta^{(g)}_2(\theta) + \Omega^{(g)}_2(\theta) + \Delta^{(r)}_2(\theta) + \Omega^{(r)}_2(\theta) + (\hat{\beta} - \beta)^T(\Delta^{(r)*}_2(\beta) + \Omega^{(r)*}_2(\beta)) +\\
& (\hat{\sigma} - \sigma)^T(\Delta^{(r)*}_2(\sigma) + \Omega^{(r)*}_2(\sigma))\},\\
\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta}) \approx& \exp\{\Delta^{(g)}_2(\theta) + \Omega^{(g)}_2(\theta) + (\hat{\beta} - \beta)^T(\Delta^{(g)*}_2(\beta) + \Omega^{(g)*}_2(\beta)) +\\
& (\hat{\sigma} - \sigma)^T(\Delta^{(g)*}_2(\sigma)) + \Omega^{(g)*}_2(\sigma) + \Delta^{(r)}_2(\theta) + \Omega^{(r)}_2(\theta) +\\
& (\hat{\beta} - \beta)^T(\Delta^{(r)*}_2(\beta) + \Omega^{(r)*}_2(\beta)) + (\hat{\sigma} - \sigma)^T(\Delta^{(r)*}_2(\sigma) + \Omega^{(r)*}_2(\sigma))\}.
\end{aligned} \tag{57}
$$

The following step is to calculate the expected value for the above expressions. Let

$$
\begin{aligned}
\delta_{1j} \equiv& \Delta^*_1(\beta) + \Omega^*_1(\beta) = x^T_{ij} - 1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} x_i,\\
\rho_{1j} \equiv& \Delta^*_1(\sigma) + \Omega^*_1(\sigma) = \begin{bmatrix} a_{1j} & b_{1j} & c_{1j} \end{bmatrix}^T,
\end{aligned}
$$

where

$$
\begin{aligned}
a_{1j} =& 1^T \Big(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} J \Big) \bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2} 1^T \Sigma^{(v)}_{1j} 1,\\
b_{1j} =& 1^T \Big(\begin{bmatrix} 1^T_j \\ 0^T \end{bmatrix} - \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \Big) \bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2} 1^T \Sigma^{(u)}_{1j} 1,\\
c_{1j} =& -1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} D \bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2} 1^T \Sigma^{(e)}_{1j} 1 + \frac{1}{2}.
\end{aligned}
$$

Assuming that $\hat{\beta}$ and $\hat{\sigma}$ are unbiased estimators of β and σ respectively (Rao, 2003, chap.6), it follows that

$$
\begin{aligned}
E[\delta^T_{1j}(\hat{\beta} - \beta)] =& 0\\
E[\rho^T_{1j}(\hat{\sigma} - \sigma)] =& 0,
\end{aligned}
$$

and

$$
\begin{aligned}
E[\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})] =& \Omega_1(\theta)\\
\Phi_1(\theta) \equiv& var[\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta})]\\
=& var(\Delta_1(\theta)) + \delta^T_{1j} var(\hat{\beta}) \delta_{1j} + E[\rho_{1j}(\hat{\sigma} - \sigma)]^2\\
=& 1^T V_{\mu_{1j}} 1 + \lambda_{1j}
\end{aligned}
$$

where

$$V_{\mu_{1j}} = 1^T \begin{bmatrix} \alpha^{(j)T} \\ \gamma^T \end{bmatrix} \bar{V}^{-1} \begin{bmatrix} \alpha^{(j)} & \gamma \end{bmatrix} 1$$
$$\lambda_{1j} = \delta_{1j}^T var(\hat{\beta})\delta_{1j} + trace[E(\rho_{1j}\rho_{1j}^T) var(\hat{\sigma})].$$

The last term in λ_{1j} is calculated using $E(w^t u)^2 = trace[E(ww^T)E(u^T u)]$, where w and u are random vectors. $var(\hat{\beta})$ and $var(\hat{\sigma})$ are the asymptotic covariance matrices of the estimators, which are obtained from the inverse of the Fisher Information matrix under the REML procedure. then we have

$$\Delta_1(\hat{\theta}) + \Omega_1(\hat{\theta}) \sim N(\Omega_1(\theta), \Phi_1(\theta)). \tag{58}$$

Now,

- From case 1

  1 Expressions in H_1, i.e $j, q \in s_i$ and $j = q$

  $$\begin{aligned} E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] &= \exp\{2(\Omega_1(\theta) + 1^T V_{\mu_{1j}} 1)\} \\ E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{2\Omega_1(\theta) + \frac{1}{2}\varphi_1(\theta)\} \\ E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{2(\Omega_1(\theta) + \Phi_1(\theta))\}, \end{aligned} \tag{59}$$

  where

  $$\varphi_1(\theta) = 4var(\Delta_1(\theta)) + \lambda_{1j}.$$

  Then

  $$E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] - 2E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] + E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] =$$
  $$\exp\{2(\Omega_1(\theta) + 1^T V_{\mu_{1j}} 1)\} - 2\exp\{2\Omega_1(\theta) + \frac{1}{2}\varphi_1(\theta)\} + \exp\{2(\Omega_1(\theta) + \Phi_1(\theta))\} \equiv h_{1ij}. \tag{60}$$

  2 Expressions in H_1, i.e $j, q \in s_i$ and $j \neq q$

  $$\begin{aligned} E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} \\ E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}} 1 + \Phi_1^{(q)}(\theta))\} \\ E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] &= \exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta) + \frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\}. \end{aligned} \tag{61}$$

  Then

  $$\begin{aligned} &E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\theta)] - E[\tilde{y}^*_{ijk}(\theta)\tilde{y}^*_{iqp}(\hat{\theta})] - E[\tilde{y}^*_{iqp}(\theta)\tilde{y}^*_{ijk}(\hat{\theta})] + E[\tilde{y}^*_{ijk}(\hat{\theta})\tilde{y}^*_{iqp}(\hat{\theta})] = \\ &\exp\{\Omega_1^{(j)}(\theta) + \Omega_1^{(q)}(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}} + V_{\mu_{1q}})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{1j}} 1 + \Phi_1^{(q)}(\theta))\} - \\ &\exp\{\frac{1}{2}(1^T V_{\mu_{1q}} 1 + \Phi_1^{(j)}(\theta))\} + \exp\{\frac{1}{2}(\Phi_1^{(j)}(\theta) + \Phi_1^{(q)}(\theta))\}] \equiv h_{1ijq}. \end{aligned} \tag{62}$$

Then from (60) and (62) it follows that

$$H_1 = \sum_{j \in s_i} (N_{ij} - n_{ij})^2 h_{1ij} + \sum_{j \in s_i} \sum_{q \in s_i} (N_{ij} - n_{ij})(N_{iq} - n_{iq}) h_{1ijq} I(q \neq j). \tag{63}$$

- From case 2

1 Expressions in H_2, i.e $j \in s_i$ and $r \in \bar{s}_i$

$$\begin{aligned} E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta)] =& \exp\{\Omega_1(\theta)+\Omega_2(\theta)+\frac{1}{2}1^T(V_{\mu_{1j}}+V_{\mu_2})1\} \\ E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{\Omega_1(\theta)+\Omega_2(\theta)+\frac{1}{2}(1^TV_{\mu_{1j}}1+\Phi_2(\theta))\} \\ E[\tilde{y}^{*}_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{\Omega_1(\theta)+\Omega_2(\theta)+\frac{1}{2}(\Phi_1(\theta)+\Phi_2(\theta))\}. \end{aligned} \tag{64}$$

Then

$$\begin{aligned} & E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] - E[\tilde{y}^{*}_{ijk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] + E[\tilde{y}^{*}_{ijk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] = \\ & \exp\{\Omega_1(\theta)+\Omega_2(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{1j}}+V_{\mu_2})1\} - \exp\{\frac{1}{2}(1^TV_{\mu_{1j}}1+\Phi_2(\theta))\} - \\ & \exp\{\frac{1}{2}(1^TV_{\mu_2}1+\Phi_1(\theta))\} + \exp\{\frac{1}{2}(\Phi_1(\theta)+\Phi_2(\theta))\}] \equiv h_{2ijr}. \end{aligned} \tag{65}$$

Then from (65) it follows

$$H_2 = \sum_{j\in s_i}\sum_{r\in\bar{s}_i}(N_{ij-n_{ij}})N_{ir}h_{2ijr}. \tag{66}$$

- From case 3

1 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g = r$

$$\begin{aligned} E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] =& \exp\{2(\Omega_2(\theta)+1^TV_{\mu_2}1)\} \\ E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{2\Omega_2(\theta)+\frac{1}{2}\varphi_2(\theta)\} \\ E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =& \exp\{2(\Omega_2(\theta)+\Phi_2(\theta))\}, \end{aligned} \tag{67}$$

where

$$\begin{aligned} \varphi_2(\theta) =& 4var(\Delta_2(\theta))+\lambda_2 \\ \lambda_2 =& \delta_2^T var(\hat{\beta})\delta_2 + trace[E(\rho_2\rho_2^T)var(\hat{\sigma})] \\ \delta_2^T =& \Delta_2^*(\beta)+\Omega_2^*(\beta) \\ \rho_2^T =& \Delta_2^*(\sigma)+\Omega_2^*(\sigma) = \begin{bmatrix} a_2 & b_2 & c_2 \end{bmatrix}^T \\ E[\Delta_2(\hat{\theta})+\Omega_2(\hat{\theta})] =& \Omega_2(\theta) \\ \Phi_2(\theta) \equiv& var[\Delta_2(\hat{\theta})+\Omega_2(\hat{\theta})] \\ =& var(\Delta_2(\hat{\theta})) + \delta_2^T var(\hat{\beta})\delta_2 + E[\rho_2(\hat{\sigma}-\sigma)]^2 \\ =& 1^TV_{\mu_2}1+\lambda_2, \end{aligned}$$

with

$$\begin{aligned} a_2 =& 1^T\Big(\begin{bmatrix} 0^T \\ 1^T \end{bmatrix} - \begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix}\bar{V}^{-1}J\Big)\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(v)}1 \\ b_2 =& -1^T\begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix}\bar{V}^{-1}I\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(u)}1 \\ c_2 =& -1^T\begin{bmatrix} 0^T \\ \gamma^T \end{bmatrix}\bar{V}^{-1}D\bar{V}^{-1}(\bar{l}_i - x_i\beta) + \frac{1}{2}1^T\Sigma_2^{(e)}1 + \frac{1}{2}. \end{aligned}$$

Then

$$E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - 2E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] + E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =$$
$$\exp\{2(\Omega_2(\theta) + 1^T V_{\mu_2} 1)\} - 2\exp\{2\Omega_2(\theta) + \frac{1}{2}\varphi_2(\theta)\} + \exp\{2(\Omega_2(\theta) + \Phi_2(\theta))\} \equiv h_{3ig}. \tag{68}$$

2 Expressions in H_3, i.e $g, r \in \bar{s}_i$ and $g \neq r$

$$\begin{aligned}
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] &= \exp\{\Omega^{(g)}_2(\theta) + \Omega^{(r)}_2(\theta) + \frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\} \\
E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{\Omega^{(g)}_2(\theta) + \Omega^{(r)}_2(\theta) + \frac{1}{2}(1^T V_{\mu_{2g}} 1 + \Phi^{(r)}_2(\theta))\} \\
E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] &= \exp\{\Omega^{(g)}_2(\theta) + \Omega^{(r)}_2(\theta) + \frac{1}{2}(\Phi^{(g)}_2(\theta) + \Phi^{(r)}_2(\theta))\}.
\end{aligned} \tag{69}$$

Then

$$E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\theta)] - E[\tilde{y}^{**}_{igk}(\theta)\tilde{y}^{**}_{irp}(\hat{\theta})] - E[\tilde{y}^{**}_{irp}(\theta)\tilde{y}^{**}_{igk}(\hat{\theta})] + E[\tilde{y}^{**}_{igk}(\hat{\theta})\tilde{y}^{**}_{irp}(\hat{\theta})] =$$
$$\exp\{\Omega^{(g)}_2(\theta) + \Omega^{(r)}_2(\theta)\}[\exp\{\frac{1}{2}1^T(V_{\mu_{2g}} + V_{\mu_{2r}})1\} - \exp\{\frac{1}{2}(1^T V_{\mu_{2g}} 1 + \Phi^{(r)}_2(\theta))\} -$$
$$\exp\{\frac{1}{2}(1^T V_{\mu_{2r}} 1 + \Phi^{(g)}_2(\theta))\} + \exp\{\frac{1}{2}(\Phi^{(g)}_2(\theta) + \Phi^{(r)}_2(\theta))\}] \equiv h_{3igr}. \tag{70}$$

Then from (68) and (70) it follows

$$H_3 = \sum_{g \in \bar{s}_i} N^2_{ig} h_{3ig} + \sum_{g \in \bar{s}_i}\sum_{r \in \bar{s}_i} N_{ig} N_{ir} h_{3igr} I(r \neq g). \tag{71}$$

4 Derivation of the corrected EB predictor

By the fact that the predictor (15) is a nonlinear transformation of estimators of parameters,

$$E[\hat{y}^{EB}_{ijk}] \neq E[y^{MMSE}_{ijk}(\theta)]. \tag{72}$$

The aim of this section is to find approximately unbiased predictor of the non-sampled value of y_{ijk}. Now considering the two cases separately and using the expressions calculated in the subsection 3.2 it follows that:

case 1 : $j \in s_i$

$$E[\tilde{y}^{*}_{ijk}(\theta)] = \exp\{\Omega_1(\theta) + \frac{1}{2}1^T V_{\mu_{1j}} 1\} \tag{73}$$

and

$$E[\tilde{y}^{*}_{ijk}(\hat{\theta})] = \exp\{\Omega_1(\theta) + \frac{1}{2}(1^T V_{\mu_{1j}} 1 + \lambda_{1j})\} \tag{74}$$

Therefore,

$$\frac{E[\tilde{y}^{*}_{ijk}(\hat{\theta})]}{E[\tilde{y}^{*}_{ijk}(\theta)]} \approx \exp\{\frac{1}{2}\lambda_{1j}\}. \tag{75}$$

Now from (75), we define the multiplicative approximately bias-corrected predictor (BCP)

$$\hat{y}^{*EB.BCP}_{ijk} = \hat{y}^{*EB}_{ijk} \exp\{-\frac{1}{2}\hat{\lambda}_{1j}\}, \tag{76}$$

where $\hat{\lambda}_{1j} = \lambda_{1j}(\hat{\theta})$, with $\lambda_{1j}(\theta) = \delta^T_{1j} var(\hat{\beta})\delta_{1j} + trace[E(\rho_{1j}\rho^T_{1j})var(\hat{\sigma})]$.

case 2 : Case $j \in \bar{s}_i$

$$E[\tilde{y}^{**}_{ijk}(\theta)] = \exp\{\Omega_2(\theta) + \frac{1}{2}1^T V_{\mu_2} 1\} \tag{77}$$

and

$$E[\tilde{y}^{**}_{ijk}(\hat{\theta})] = \exp\{\Omega_2(\theta) + \frac{1}{2}(1^T V_{\mu_2} 1 + \lambda_2)\} \tag{78}$$

Therefore,

$$\frac{E[\tilde{y}^{**}_{ijk}(\hat{\theta})]}{E[\tilde{y}^{**}_{ijk}(\theta)]} \approx \exp\{\frac{1}{2}\lambda_2\}. \tag{79}$$

From (79), we define the multiplicative approximately bias-corrected predictor (BCP)

$$\hat{y}^{**EB.BCP}_{ijk} = \hat{y}^{**EB}_{ijk} \exp\{-\frac{1}{2}\hat{\lambda}_2\}, \tag{80}$$

where $\hat{\lambda}_2 = \lambda_2(\hat{\theta})$, with $\lambda_2(\theta) = \delta_2^T var(\hat{\beta})\delta_2 + trace[E(\rho_2\rho_2^T)var(\hat{\sigma})]$.

Now from (76) and (80) the approximately corrected-bias predictor for $\bar{Y}_i^{MMSE}$ is given by:

$$\bar{Y}_i^{EB.BCP} = \frac{1}{N_i}\left\{\sum_{j,k\in s_i} y_{ijk} + \sum_{j\in s_i}\sum_{k\in\bar{s}_{ij}} \hat{y}^{*EB.BCP}_{ijk} + \sum_{j\in\bar{s}_i}\sum_{k=1}^{N_{ij}} \hat{y}^{**EB.BCP}_{ijk}\right\}. \tag{81}$$

.

5 Parametric bootstrap for MSE estimation

The parametric bootstrap that we propose here to estimate the MSE of EB bias-corrected predictors $\bar{Y}_i^{EB.BCP}$, is an extension of the parametric bootstrap method for finite population proposed by González-Manteiga et al., 2008, Molina and Rao, 2009. This parametric procedure is described as below:

1. Fit model (5) to sample data and obtain model parameters estimates $\hat{\beta}$, $\hat{\sigma}_v^2$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_e^2$.
2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, M$.
3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects $u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$.
4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$,$j = 1, \ldots, N_{ij}$.
5. Construct a bootstrap population using the estimated model

$$log(y^*_{ijk}) = l^*_{ijk} = x_{ij}^T\hat{\beta} + v_i^* + u_{ij}^* + e_{ijk}^*, \tag{82}$$

and calculate the small area population mean

$$\bar{Y}_i^* = \frac{1}{N_i}\sum_{j=1}^{M_i}\sum_{k=1}^{N_{ij}} y^*_{ijk}. \tag{83}$$

6. select the elements l^*_{ijk} that correspond to the indices contianed in the sample s, denote l^*_s. Fit the model to l^*_s obtining new model parameters estimates $\hat{\beta}^*$, $\hat{\sigma}^{2*}_v$, $\hat{\sigma}^{2*}_u$, and $\hat{\sigma}^{2*}_e$.

7. Using the bootstrap sample data l^*_s and the known matrix X, apply the EB method with its correction as it was described in sections 2 and 4 respectively, and calculate bootstrap EB prodictors, $\bar{Y}^{EB*}_i$, $i = 1, \ldots, M$.

Note that the bootstrap population model, given the original sample data, preserve properties of the original population model. This can be observed as follows

$$E_*(v^*_i|l) = E_*(u^*_{ij}|l) = E_*(e^*_{ijk}|l) = 0, \;\; var_*(v^*_i|l) = \hat{\sigma}^2_v, \;\; var_*(u^*_{ij}|l) = \hat{\sigma}^2_u \;\; , var_*(e^*_{ijk}|l) = \hat{\sigma}^2_e, \quad (84)$$

where E_* and var_* represent conditional expectation and variance with respect to the distribution defined by the bootstrap model(96) given the sample data l_s.
Thereby, the distribution of the bootstrap population l^* (given sample data l_s) mimics that of the original population l. Then an estimator of $MSE(\bar{Y}^{EB.BCP}_i)$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_*(\bar{Y}^{EB.BCP*}_i) = E_*[(\bar{Y}^{EB.BCP*}_i - \bar{Y}^*_i)^2] \quad (85)$$

In practice, this expression can be approximated through a Monte Carlo simulation, by repeating steps $2-7$ a large number of times, B, and then taking the mean over the the B replicates as follows: Let $\bar{Y}^{*(b)}_i$ and $\bar{Y}^{EB.BCP*(b)}_i$ be the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b, for $b = 1, \ldots, B$. Then, the estimator of the MSE is calculated as

$$MSE(\bar{Y}^{EB.BCP}_i) = \frac{1}{B}\sum_{b=1}^{B}(\bar{Y}^{EB.BCP^{*(b)}}_i - \bar{Y}^{*(b)}_i)^2. \quad (86)$$

5.1 Bias-corrected MSE estimator based on single bootstrap

A naive estimator of MSE is

$$M\hat{S}E_i = M_{1i}(\hat{\theta}) + M_{2i}(\hat{\theta}), \quad (87)$$

where $M_{1i}(\hat{\theta})$ and $M_{2i}(\hat{\theta})$ are the expressions (**??**) and (**??**), respectively, evaluated at the estimator of θ. In general, it is known that $M_{1i}(\hat{\theta})$ is an asymptotically unbiased estimator (Prasad and Rao, 1990). Furthermore, $M_{1i}(\theta)$ is a nonlinear function of θ, the naive estimator $M_{1i}(\hat{\theta})$ in (87) is biased, so that we need to correct the bias. Given the complexity of the expression of $M_{1i}(\theta)$, in subsection 3.1, it is not possible to correct the bias using analytical approach. The alternative solution is to use the bootstrap method. We derive the single bootstrap bias-corrected estimator of $M_{1i}(\theta)$ in two steps (Butar and Lahir, 2003; Rachida, O., 2011; Kubokawa and Nagashima, 2012). At the first step, under the assumption of known parameters the derivation of MSE is presented. At the second step, a parametric bootstrap approach, described at the beginning of this section, is proposed for bias correction and approximation of the uncertainty due to the estimation of θ.

Definition : The single bootstrap bias corrected estimator is defined as

$$M^{BC}_{1i}(\hat{\theta}) = M_{1i}(\hat{\theta}) + b_{1i}(\hat{\theta}), \quad (88)$$

where $b_{1i}(\hat{\theta}) = M_{1i}(\hat{\theta}) - E_{\hat{\theta}}(M_{1i}(\hat{\theta}^*))$.
Below, we present a second stage parametric bootstrap algorithm for bias-correction of the MSE estimator:

1. Fit model (5) to sample data and obtain model parameters estimates $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_v^2, \hat{\sigma}_u^2, \hat{\sigma}_e^2)^T$.
2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, m$.
3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects $u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, m_i$.
4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, m$, $j = 1, \ldots, m_i$,$j = 1, \ldots, n_{ij}$.
5. Construct a bootstrap samples using the estimated model

$$log(y_{ijk}^*) = l_{ijk}^* = x_{ij}^T \hat{\beta} + v_i^* + u_{ij}^* + e_{ijk}^*, \tag{89}$$

and for each bootstrap replicate b, for $b = 1, \ldots, B$ we calculate the bootstrap version $M_{1i}(\hat{\theta})^{(b)}$. Then a Monte Carlo estimate of M_{1i} is given by

$$M_{1i}^{BC}(\hat{\theta}) = 2M_{1i}(\hat{\theta}) - \frac{1}{B} \sum_{1=b}^{B} M_{1i}(\hat{\theta})^{(b)}. \tag{90}$$

Furthermore, the unbiased estimator of the MSE based on the parametric bootstrap is given by

$$m\hat{s}e_i = M_{1i}^{BC}(\hat{\theta}) + M_{2i}(\hat{\theta}). \tag{91}$$

From the described algorithm, we set out the justification behind this approach as it was introduced by Butar and Lahiri (2003) and presented in Kubokawa and Nagashima (2012):
Let $f(\theta)$ be a smooth function. In the spite of the fact that $f(\hat{\theta})$ is an asymptotically unbiased estimator of $f(\theta)$, in general, there exists a second-order bias. Then, we need to approximate the expectation $E[f(\hat{\theta})]$. It is supposed that the approximation is given by

$$E[f(\hat{\theta})] = f(\theta) + b(\theta), \tag{92}$$

where $b(\theta)$ is a smooth function . Then,

$$\begin{aligned} E[f(\hat{\theta}) - b(\hat{\theta})] =& E[f(\hat{\theta})] - E[b(\hat{\theta})] \\ =& \{f(\theta) + b(\theta)\} - b(\theta) \\ =& f(\theta). \end{aligned} \tag{93}$$

Using model (96), it follows that

$$E_{\hat{\theta}}[f(\hat{\theta}^*)|l] = f(\hat{\theta}) + b(\hat{\theta}), \tag{94}$$

where $E_{\hat{\theta}}[.|l]$ is the conditional expectation with respect to the model (96) given l, and the calculation of $\hat{\theta}^*$ is the same as that of $\hat{\theta}$ except that $\hat{\theta}^*$ is calculated based on l^* instead of l. Hence from (93), we have

$$\begin{aligned} E[2f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*)|l]] =& E[f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*) - f(\hat{\theta})|l]] \\ =& E[f(\hat{\theta}) - b(\hat{\theta})] \\ =& f(\theta). \end{aligned} \tag{95}$$

Therefore, $2f(\hat{\theta}) - E_{\hat{\theta}}[f(\hat{\theta}^*)|l]$ is the second-order unbiased estimator of $f(\theta)$.

5.2 Double parametric bootstrap for bias-correction

Following Hall and Maiti (2006) and adopting a double parametric bootstrap to bias-correction, we provide a population double bootstrap bias adjustment to the MSE estimator of EB bias-corrected predictors $\bar{Y}_i^{EB.BCP}$, but in the setting of the parametric bootstrap method for finite population proposed by González-Manteiga et al., 2008, Molina and Rao, 2009. The double parametric procedure is described as below:

1. Fit model (5) to sample data and obtain model parameters estimates $\hat{\beta}$, $\hat{\sigma}_v^2$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_e^2$.
2. Generate bootstrap random area effects as $v_i^* \sim N(0, \hat{\sigma}_v^2)$, $i = 1, \ldots, M$.
3. Generate, independently of random area effects v_i^*, bootstrap random cluster effects $u_{ij}^* \sim N(0, \hat{\sigma}_u^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$.
4. Generate, independently of random area effects v_i^* and random cluster effects u_{ij}^*, bootstrap random errors $e_{ijk}^* \sim N(0, \hat{\sigma}_e^2)$, $i = 1, \ldots, M$, $j = 1, \ldots, M_i$,$j = 1, \ldots, N_{ij}$.
5. Construct a bootstrap population using the estimated model

$$log(y_{ijk}^*) = l_{ijk}^* = x_{ij}^T \hat{\beta} + v_i^* + u_{ij}^* + e_{ijk}^*, \tag{96}$$

and calculate the small area population mean

$$\bar{Y}_i^* = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}^*. \tag{97}$$

6. select the elements l_{ijk}^* that correspond to the indices contianed in the sample s, denote l_s^*. Fit the model to l_s^* obtining new model parameters estimates $\hat{\beta}^*$, $\hat{\sigma}_v^{2*}$, $\hat{\sigma}_u^{2*}$, and $\hat{\sigma}_e^{2*}$.
7. Using the bootstrap sample data l_s^* and the known matrix X, apply the EB method as described in Section 2 and calculate bootstrap EB prodictors, $\bar{Y}_i^{EB*}$, $i = 1, \ldots, M$. Then an estimator of $MSE(\bar{Y}_i^{EB.BCP})$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_*(\bar{Y}_i^{EB.BCP*}) = E_{\hat{\theta}}[(\bar{Y}_i^{EB.BCP*} - \bar{Y}_i^*)^2]. \tag{98}$$

Let's note $\bar{Y}_i^{*(b_1)}$ and $\bar{Y}_i^{EB.BCP*(b_1)}$ as the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b_1, for $b_1 = 1, \ldots, B_1$. Then, the estimator of the MSE is calculated as

$$MSE_*(\bar{Y}_i^{EB.BCP*}) = B_{i1} = \frac{1}{B_1} \sum_{b_1=1}^{B_1} (\bar{Y}_i^{EB.BCP^{*(b_1)}} - \bar{Y}_i^{*(b_1)})^2. \tag{99}$$

8. For each bootstrap replicate $b_1 = 1, \ldots, B_1$, obtain parameters estimates $\hat{\beta}^{*(b_1)}$, $\hat{\sigma}_v^{2*(b_1)}$, $\hat{\sigma}_u^{2*(b_1)}$, and $\hat{\sigma}_e^{2*(b_1)}$, and generate for $b_2 = 1, \ldots, B_2$:

$$\begin{aligned}
v_i^{**} &\sim N(0, \hat{\sigma}_v^{2*(b_1)}), \quad i = 1, \ldots, M \\
u_{ij}^{**} &\sim N(0, \hat{\sigma}_u^{2*(b_1)}), \quad i = 1, \ldots, M, j = 1, \ldots, M_i \\
e_{ijk}^{**} &\sim N(0, \hat{\sigma}_e^{2*(b_1)}), \quad i = 1, \ldots, M, \quad j = 1, \ldots, M_i, \quad j = 1, \ldots, N_{ij}
\end{aligned}$$

9. Constructing a new bootstrap populations using

$$log(y_{ijk}^{**}) = l_{ijk}^{**} = x_{ij}^T \hat{\beta}^{*(b_1)} + v_i^{**} + u_{ij}^{**} + e_{ijk}^{**}, \tag{100}$$

and calculate the small area population mean

$$\bar{Y}_i^{**} = \frac{1}{N_i} \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} y_{ijk}^{**}. \tag{101}$$

10. Select the elements l_{ijk}^{**} that correspond to the indices contianed in the sample s, denote l_s^{**}. Fit the model to l_s^{**} obtining new model parameters estimates $\hat{\beta}^{(b_2)}$, $\hat{\sigma}_v^{2(b_2)}$, $\hat{\sigma}_u^{2(b_2)}$, and $\hat{\sigma}_e^{2(b_2)}$.

11. Using the bootstrap sample data l_s^{**} and the known matrix X, we apply the EB method as described in sections 2 and 4 respectively, and calculate bootstrap EB prodictors, $\bar{Y}_i^{EB**}$, $i = 1, \ldots, M$. Then an estimator of $MSE(\bar{Y}_i^{EB.BCP})$ is the bootstrap MSE of the bootstrap EB.BCP, defined as

$$MSE_{**}(\bar{Y}_i^{EB.BCP**}) = E_{\hat{\theta}^*}[(\bar{Y}_i^{EB.BCP**} - \bar{Y}_i^{**})^2] \tag{102}$$

Noting $\bar{Y}_i^{**(b_2(b_1))}$ and $\bar{Y}_i^{EB.BCP**(b_2(b_1))}$ as the area population mean and its corresponding EB bias-corrected predictor for the bootstrap replicate b_2, for $b_2 = 1, \ldots, B_2$. Then, the estimator of the MSE is calculated as

$$MSE_{**}(\bar{Y}_i^{EB.BCP**}) = B_{i2} = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \frac{1}{B_2} \sum_{b_2=1}^{B_2} (\bar{Y}_i^{EB.BCP^{**(b_2(b_1))}} - \bar{Y}_i^{**(b_2(b_1))})^2. \tag{103}$$

From (99) and (103) we get the population bias-corrected MSE estimator

$$\hat{MSE}(\bar{Y}_i^{EB.BCP}) = 2MSE_*(\bar{Y}_i^{EB.BCP*}) - MSE_{**}(\bar{Y}_i^{EB.BCP**}). \tag{104}$$

5.3 Bias-corrected MSE estimator based on double bootstrap

The population bias-corrected MSE estimator, (104), presented in subsection 5.2 can not be calculated in practical settings since it depends on population quantities, in this subsection we derive a bias-corrected MSE estimator of (87) based on a double bootstrap. As Davison and Hinkley (1997) pointed out, the bootstrap does not provide exact solution, the same as in most statistical methods, regarding the bias correction. However, it is helpful to have available a general technique for making a bias correction to a bootstrap calculation. That technique is the bootstrap itself. The bias-corrected estimator of $M_{1i}(\theta)$ based on double bootstrap is given by (Rachida O., 2011; Chang and Hall, 2015)

$$\hat{m}_{1i}^{bcc} = 3M_{1i}(\hat{\theta}) - 3E_{\hat{\theta}}(M_{1i}(\hat{\theta}^*)|l) + E_{\hat{\theta}^*}(M_{1i}(\hat{\theta}^{**})|l^*), \tag{105}$$

where $E_{\hat{\theta}}[.|l]$ is the conditional expectation with respect to the model (96) given l, and the calculation of $\hat{\theta}^*$ is the same as that of $\hat{\theta}$ except that $\hat{\theta}^*$ is calculated based on l^* instead of l, and $E_{\hat{\theta}^*}[.|l^*]$ is the conditional expectation with respect to the model (100) given l^*, and the calculation of $\hat{\theta}^{**}$ is the same as that of $\hat{\theta}^*$ except that $\hat{\theta}^{**}$ is calculated based on l^{**} instead of l^*.
Applying the bootstrap algorithm of the subsection 5.2, the Monte Carlo approximation to the quantity $\hat{m}_{1i}^{bcc}$ is given by

$$\tilde{m}_{1i}^{bcc} = 3M_{1i}(\hat{\theta}) - \frac{3}{B_1} \sum_{b_1=1}^{B_1} M_{1i}(\hat{\theta}^{*(b_1)}) + \frac{1}{B_1 B_2} \sum_{b_1}^{B_1} \sum_{b_2}^{B_2} M_{1i}(\hat{\theta}^{**(b_2(b_1))}), \tag{106}$$

where $M_{1i}(\hat{\theta})^{*(b_1)}$ is the version of $M_{1i}(\hat{\theta})$ calculated from (96) for each bootstrap replicate b_1, for $b_1 = 1, \ldots, B_1$, and $M_{1i}(\hat{\theta})^{**(b_2(b_1))}$ is the version of $M_{1i}(\hat{\theta})$ obtained from (100) for each b_2, for $b_2 = 1, \ldots, B_2$ for each b_1.
Therefore, from (87) and (106) the bias-corrected MSE estimator based on double bootstrap is given by

$$\hat{mse}_i^{bcc} = \tilde{m}_{1i}^{bcc} + M_{2i}(\hat{\theta}). \qquad (107)$$

6 Simulation study

For the purpose of evaluating the performance of the proposed EB predictors, a simulation experiment is conducted in order to investigate the bias of $MSE(\bar{Y}_i^{EB})$, obtained under a studied model, comparing the derived naive estimator, its proposed bootstrap estimator, and the double bootstrap estimator of the MSE of EB estimators. Note that this experiment will be repited $K = 100$ times. Under the model (5), we generate the response variable for the population units $\log(Y_{ijk})$, similarly to Molina and Rao (2009) but including an indicator of clusters within small area, where the indicator variables mimic the real case where only categorical variables are available. We consider a clustered finite population from which samples are drawn in two stages using simple random sampling at each stage.
In summary, the specifications of the model for the k^{th} simulation, for $k = 1, \ldots, K$, is:

1. We consider a balanced two-fold model, with a population size $N = 120000$ partitioned into $M = 30$ small areas, with small area population size of $N_i = 4000$, $i = 1, \ldots, M$, and each small area is composed of $M_i = 40$ clusters, $i = 1, \ldots, M$. Cluster population sizes are $N_{ij} = 100$, $j = 1, \ldots, M_i$, $i = 1, \ldots, M$.

2. Two dummy variables are used as covariates plus intercept. The population values of these indicators for the units are generated from Bernoulli distributions $Ber(p_{hij})$,$h = 1, 2$, with probabilities of success $p_{1ij} = 0.3 + 0.5i/M + 0.1j/M_i$ and $p_{2ij} = 0.2$. The covariates are held fixed across the simulated populations.

3. The fixed effects are $\beta = (6, 0.03, -4)^T$.

4. The small area effects, cluster effects and individual errors are independent; with $v_i \sim N(0, \sigma_v^2)$, $u_{ij} \sim N(0, \sigma_u^2)$ and $e_{ijk} \sim N(0, \sigma_e^2)$, where $\sigma = (\sigma_v^2, \sigma_u^2, \sigma_e^2)$ is such that $\sigma_v^2 = 0.05$, $\sigma_u^2 = 0.2$, $\sigma_e^2 = 0.025$. To imitate different situations that can be existe in real cases, simulation experiments are repeated for various combinations of variance components: small area (Domain) variability, σ_v^2, and cluster (subdomain) variability, σ_u^2.

5. Within each small area i, a sample of $m_i = 5$ clusters is selected using simple random sampling (SRS), and a simple random sample of size $n_{ij} = 10$ is drawn from each sampled cluster. The small area sample sizes are equal $n_i = 50$.

We generate a bootstrap population as it is described at the beginning of section 5. We draw a sample from each Bootstrap population and we fit the model and we compute the MSE estimator (107) and double parametric bootstrap MSE (104).

6.1 Simulation experiments

This experiment of simulation is motivated by the fact that practical usage of EB predictors requires, of course, estimates of variance components. It consists of currying out several runs of the simulation study, keeping constant the sample sizes, the population sizes and the number of levels and sublevels

of the random factors, and varying the values of σ_v^2 and σ_u^2.
Sixteen tests of the experiment are carried out, for the sixteen possible combinations of the values $\sigma_e^2 = 0.025$, $\sigma_u^2 = \{0.05, 0.1, 0.15, 0.2\}$, and $\sigma_v^2 = \{0.05, 0.1, 0.15, 0.2\}$, according to the table 1

r	1	2	3	4	5	6	7	8
$\sigma_u^{2,(r)}$	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1
$\sigma_v^{2,(r)}$	0.05	0.1	0.15	0.2	0.05	0.1	0.15	0.2
r	9	10	11	12	13	14	15	16
$\sigma_u^{2,(r)}$	0.15	0.15	0.15	0.15	0.2	0.2	0.2	0.2
$\sigma_v^{2,(r)}$	0.05	0.1	0.15	0.2	0.05	0.1	0.15	0.2

Table 1: Combinations of σ_u^2 and σ_v^2 for a simulation experiment

6.2 Simulation results

The following plots represent the average of the square roots of the four versions of MSE across the domains with respect to the variance components. In (Figure 1(a)), we show the behavior of those MSEs when the domain and cluster variaances increase simultaneously, while the three remaining plots show the behavior of MSEs when we fix the cluster variance , σ_u^2, and varying the domain variability, σ_v^2.

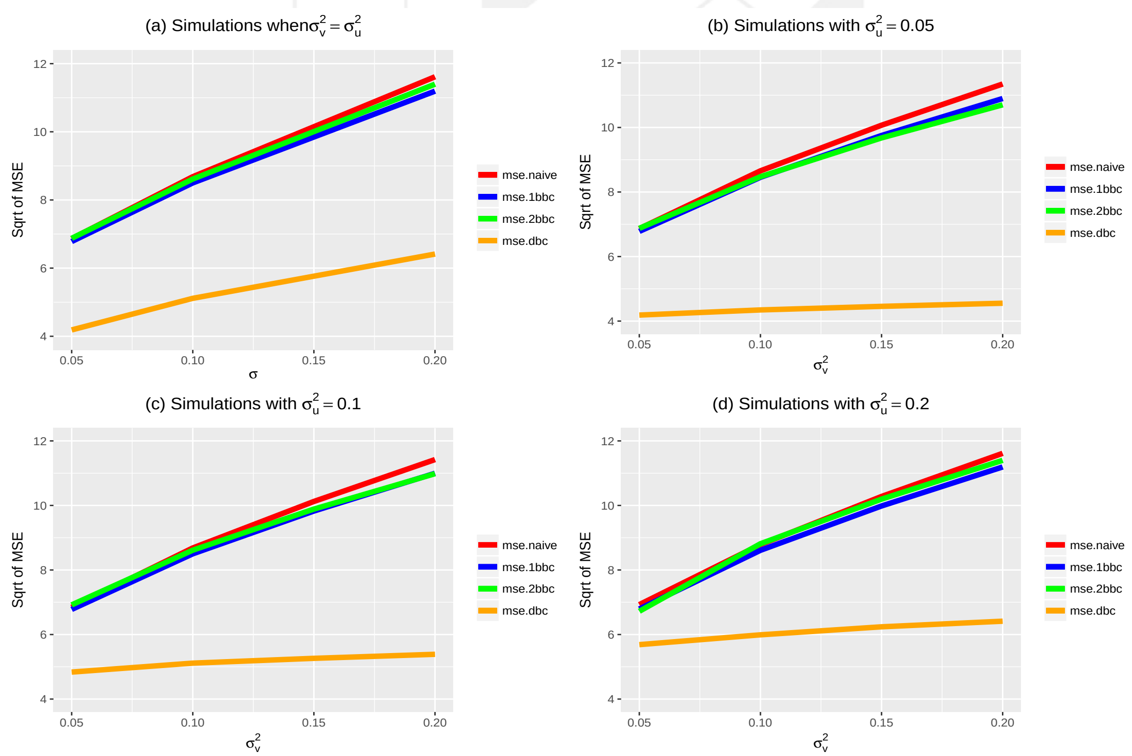

Figure 1: Average sqrt MSE across the domains with respect to the change in variance components.

The simulation experiments were repeated 100 times and the following average estimates were obtained:

$$\hat{\beta} = (5.99, 0.04, -3.99), \quad \hat{\sigma} = (0.0499, 0.199, 0.025).$$

In terms of prediction, after 100 Monte Carlo simulations, the following plot compare the average by domain of population values to their predicted values.

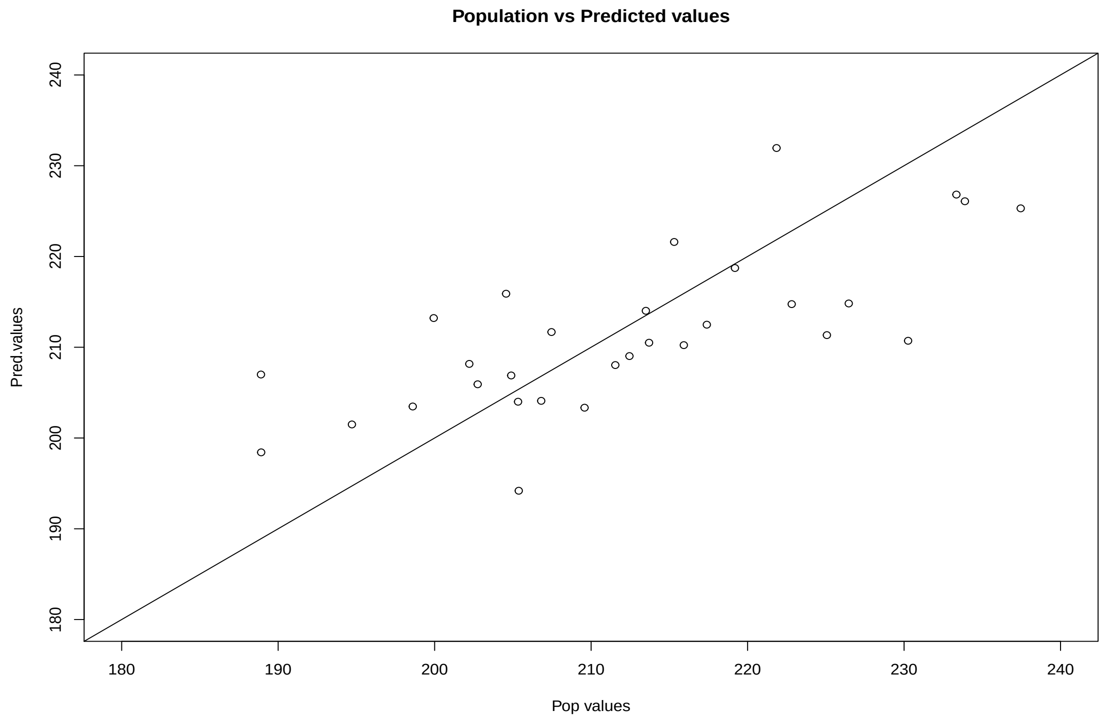

Figure 2: Averages of population and predicted values obtained after 100 simulations, with $\sigma_v^2 = 0.05$ and $\sigma_u^2 = 0.2$.

From the above simulations, we see that the estimates, as well as predictors are close to the true values and thus, the proposed estimators support the theoretical results.

7 Concluding remarks and future research

The minimum mean squared error predictor under the proposed model for small area estimation was developed. To obtain the empirical Bayes predictors of pupulation means for small areas, the Scoring-Fisher algorithm based on restricted maximum likelihood to estimate the variance components was used. Following Prasad and Rao (1990), the estimation theory of MSE for the EB predictor, was adapted to the model under study and the closed form expressions of MSE were obtained. Furthermore, we proposed the bias-corrected estimator of MSE under a parametric bootstrap, as well as a double bootstrap method. We studied the prediction capacity of our model under simulation experiments. The simulation studies established clearly the positive performance of using the proposed model in terms of prediction.

In this study, the assessment of the estimated MSE was based on variance components analysis and was centered on three versions of MSE: naive estimator, bias-corrected estimators (based on simple and double bootstrap) and double bootstrap expressions. Examining the results obtained during the simulation experiments and presented by means of plots (Figure 1), it shows that the MSE along the domains increases in magnitude when the values of σ_u^2 and σ_v^2 increase or decrease simultaneously (Figure 1(a)). In addition, the bias is moderately reduced in magnitude as the values of σ_u^2 and σ_v^2 increased and decreased respectively (Figure 1(b),(c),(d)); that is, the larger

the variance between clusters (sub-domains) and the smaller the variance between domains are, the corrected MSE of the EB predictor becomes closer to the one obtained under the ideal double bootstrap MSE. In summary, as the cluster variability σ_u^2 increases compared with the domain variability, σ_v^2, the corrected MSE estimators and bootstrap MSE versions are getting closer.
For the sake of illustration of our methodology, the simulation experiments were performed only for the balanced case, that is, when the number of samples was the same for each cluster. For further analysis, the experiment can be extended to the unbalanced case. This work confined the attention to the framework of mixed models with homogeneous random area-specific effects. However, in real life, this assumption may not always be justified. The assessment of the performance of the proposed models including spatial dependent random area effects, as well as a development of prediction intervals theoretically appropriate for lognormal data would be interesting avenues for future research. We also wish to have some specific real data set and apply the results of this work.

Acknowledgement

This research was partially supported by the grant from the project 268361 of *Fondos Sectoriales CONACYT-INEGI*, Mexico. The first author gratefully acknowledges the financial support from the scholarship of CONACYT in collabolation with CIMAT, Mexico.

References

[1] Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

[2] Berg, E., Chandra, H. (2014). Small area prediction for unit-level lognormal model. *Computational Statistics and Data Analysis*, 78, 159-175.

[3] Butar, F. B., Lahiri, P. (2003). On measures of uncertainty of empirical bayes small-area estimators, *Journal of Statistical Planning and Inference*, 112, 1, 6376.

[4] Chandra, H., Chambers, R. (2011). Small area estimation under transformation to linearity. *Survey Mothodology*, 37, 39-51.

[5] Chang, J., Hall, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation, *Biometrica*, 102, 1, 203-214.

[6] Datta, G. S., Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *Annals of Statistics*, 19, 1748-1770.

[7] Davison, A. C., Hinkley, D. V., (1997). *Bootstrap methods and their application*, Cambridge University Press, United Kingdom.

[8] Fuller, W. A., Battese, G. E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American Statistical Association*, 68, 626-32.

[9] Ghosh, M., Maiti, T. (2004). Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, 91, 95-112.

[10] Hall P., Maiti T., (2006). On parametric bootstrap methods for small area prediction, *textJ.R. Statist. Soc. B*, 68, 2, 221-238.

[11] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72, 320-340.

[12] Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 75, 423-447.

[13] Kackar, R. N., Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models. *Journal of the American Statistical Association*, 79, 853-862.

[14] Kubokawa,T., Nagashima, B., (2012). Parametric bootstrap methods for bias correction in linear mixed models, *Journal of Multivariate Analysis*, 106, 116.

[15] Molina, I., Rao, J. N. K. (2009). Small area estimation of poverty indicators. *Canadian Journal of Statistics*,38, 369-385.

[16] Pfeffermann, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, 28(1), 40-68.

[17] Pfeffermann, D., Barnard, C.H. (1991). Some new estimators for small area meanswith application to the assessment of farmland values. *Journal of Business and Economic Statistics*, 73-84.

[18] Prasad, N. G. N., Rao, J. N. K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, 85, 409, 163-171.

[19] Rachida O. (2011). Computationally efficient approximation for the double bootstrap mean bias correction. *Economics Bulletin*, 31, 3, 2388-2403.

[20] Rao, J. N. K. (2003). *Small Area Estimation*. New Jersey, John Wiley and Sons, Inc.

[21] Saei, A., Chambers, R. (2003). Small area estimates: A review of methods based on the application of mixed models. Technical Report Southampton Statistical Sciences Research Institute, Methodology Working Paper, M03/16.

[22] Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. New York, Springer-Verlag.

[23] Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. John Wiley and Sons, New York.

[24] Slud, E. V., Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *Journal of the Royal Statistical Society: Sec.B*,68,239-257.

[25] Sugasawa, S., Kubokawa, T., (2017). Transforming response values in small area prediction, *Journal of Computational Statistics and Data Analysis*, 114, 47-60.

[26] Stukel, D. M., Rao, J. N. K. (1999). On Small area estimation under two-fold nested error regression models. *Journal of Statistical Planning and Inference*, 78, 131-147.

[27] Valliant, R., Dortman, A. H., and Royall, R. M. (2000). *Finite Population Sampling and Inference*. John Wiley and Sons, New York.