Communications for Statistical Applications and Methods
2020, Vol. 27, No. 3, 349–363

https://doi.org/10.29220/CSAM.2020.27.3.349
Print ISSN 2287-7843 / Online ISSN 2383-4757

Mean estimation of small areas using penalized spline mixed-model under informative sampling

Angela N. R. Chytrasari[ab], Sri Haryatmi Kartiko[1,a], Danardono Danardono[a]

[a]Department of Mathematics, FMIPA, Universitas Gadjah Mada, Indonesia;
[b]Department of Mathematic Education, FKIP, Universitas Pancasakti Tegal, Indonesia

Abstract

Penalized spline is a suitable nonparametric approach in estimating mean model in small area. However, application of the approach in informative sampling in a published article is uncommon. We propose a semiparametric mixed-model using penalized spline under informative sampling to estimate mean of small area. The response variable is explained in terms of mean model, informative sample effect, area random effect and unit error. We approach the mean model by penalized spline and utilize a penalized spline function of the inclusion probability to account for the informative sample effect. We determine the best and unbiased estimators for coefficient model and derive the restricted maximum likelihood estimators for the variance components. A simulation study shows a decrease in the average absolute bias produced by the proposed model. A decrease in the root mean square error also occurred except in some quadratic cases. The use of linear and quadratic penalized spline to approach the function of the inclusion probability provides no significant difference distribution of root mean square error, except for few smaller samples.

Keywords: small area estimation, mixed-model, semiparametric, informative sampling, likelihood

1. Introduction

The need of reliable statistical information for sub-populations that are limited by the small size of samples has led to the development of small area estimation methods, see for instance Molina *et al.* (2014), Tzavidis *et al.* (2012), Clement (2014), Burgard *et al.* (2014) and Hwang and Kim (2015). Small area estimation should be based on a model-based approach (Rao, 2003). In model-based approach, the parametric assumption stating the relationship of response variables and auxiliary variables are often limited. Ruppert *et al.* (2003) studied semiparametric regression with an unspecified mean function that assumed to be approximated sufficiently by a penalized spline (p-spline). They obtained an empirically best linear unbiased predictor (EBLUP) using mixed-model formulation. Opsomer *et al*. (2008) extended the results of Ruppert *et al.* (2003) to small area estimation that includes the random area effect and obtained the EBLUP of mean of small area. Meanwhile, Rao *et al.* (2014) applied the approach of Opsomer *et al*. (2008) to develop a robust EBLUP of means small area. Opsomer *et al*. (2008) and Rao *et al.* (2014) involved p-spline in the context of a small area mixed-model. However, they developed their semiparametric model-based approach to estimate small area means under a noninformative sampling assumption in which standard inference procedures can be applied.

[1] Corresponding author: Department of Mathematics, FMIPA, Universitas Gadjah Mada, Jl. Sekip Utara Yogyakarta, Indonesia. E-mail: s_kartiko@yahoo.com

Published 31 May 2020 / journal homepage: http://csam.or.kr

Complex sampling designs are often used to collect sample data. In a complex sampling design, if the sample is informative, the model that is applied to the sample may be different from the model used in the population. Therefore, a sample model in standard inference may produce a heavy bias. Both Pfeffermann and Sverchkov (2007, 2009) and Burgard *et al.* (2014) stated that the informative effect of the sample must be taken into account in the inference process to reduce bias.

Verret *et al.* (2015) in added a function of inclusion probability of unit j in area i, $g(\pi_{ij})$ as a covariate into the unit-level error regression model (Battese *et al.*, 1988), to reduce the informative effect on the prediction of small area mean. The selection of the $g(\pi_{ij})$ function is done by first plotting the residual model without the variable $g(\pi_{ij})$ to $g(\pi_{ij})$ function. The plot graph which tends to be linear determines the $g(\pi_{ij})$ chosen. Their approach to determine the $g(\pi_{ij})$ function is applied to a linear mixed-model. However, in non-linear models, the $g(\pi_{ij})$ function may be difficult to be determined because of the possibility of not obtaining the plot that tends to be linear.

We propose a predictive approach for small area means based on semiparametric mixed-model using p-spline under informative sampling. We take into account the effects of informative sample by adding a p-spline function of the inclusion probability to the model. Model performances were measured using mean square error and absolute bias calculated by bootstrap method. We gave a simulation to evaluate the proposed predictor performance. This article is structured as follows. Section 1 gives the introduction. Section 2 give brief definition of population, sample and informative sampling. We present our model in Section 3 where we define the model, derive estimator parameters; in addition, we show the estimator properties of the coefficients model and assess model performance. We present simulation study in Section 4. Summary of our paper is presented in Section 5.

2. Population, sample and informative sampling

Let U be a population of values x and y. U is partitioned into M clusters denoted by U_i, $i = 1, \ldots, M$ that are seen as small areas. Each cluster contains N_i unit elements. A sample of (x_{ij}, y_{ij}) of n_i, denoted as $S_i = \{(x_{ij}, y_{ij}) | j = 1, \ldots, n_i;\ i = 1, \ldots, M\}$, is taken independently in each area i using informative sampling. Pfeffermann and Sverchkov (2009) stated that informative sampling is a sampling mechanism with the probability of inclusion that depends on the response variable. Referring to Pfeffermann *et al.* (1998), mathematically the informative sampling conditions in the area i can be explained as follows. Let I_{ij} denotes $(N_i \times 1)$ indicator variable such that $I_{ij} = 1$ if $j \in S_i$ and $I_{ij} = 0$ if $j \notin S_i$. Suppose the population unit on area i, y_{ij} $(j = 1, \ldots, N_i;\ i = 1, \ldots, M)$ is an independent realization of a distribution with the probability of density expressed as $f_{U_i}(y_{ij} | \mathbf{x}_{ij})$ dependent on the concomitant $\mathbf{x}_{ij}$ which may include auxiliary variables and design variables. The marginal probability function of the sample y_i in area i can be written using the Bayes theorem:

$$\begin{aligned} f_{S_i}(y_{ij}|\mathbf{x}_{ij}) &= f_{U_i}(y_{ij}|\mathbf{x}_{ij}, j \in S_i) \\ &= f_{U_i}(y_{ij}|\mathbf{x}_{ij}, I_{ij} = 1) \\ &= \frac{P(I_{ij} = 1|\mathbf{x}_{ij}, y_{ij}).f_{U_i}(y_{ij}|\mathbf{x}_{ij})}{P(I_{ij} = 1|\mathbf{x}_{ij})}. \end{aligned} \tag{2.1}$$

If $P(I_{ij} = 1|\mathbf{x}_{ij}, y_{ij}) \neq P(I_{ij} = 1|\mathbf{x}_{ij})$, then the sample probability function is different from the population probability function. On this condition, the sampling design is called informative. Sample drawn under informative sampling is called informative sample.

3. Proposed model

Let the unit value of observation variable, y_{ij}, can always be obtained and covariate variable unit, x_{ij}, is an univariate variable. The inclusion probability of unit j in the area i is denoted as π_{ij} ($j = 1, \dots, N_i$, $i = 1, \dots, M$). Suppose the values of x_{ij} and π_{ij} known for every i and j. The variable y_{ij} has a correlation with x_{ij} which is expressed in an unknown function $g_1(x_{ij})$. The random area effect, v_i, is also considered. The sample model is:

$$y_{ij} = g_1(x_{ij}) + v_i + e_{ij}, \quad j = 1, \dots, n_i;\ i = 1, \dots, M, \tag{3.1}$$

with e_{ij} is unit error. We assume e_{ij} and v_{ij} have a normal distribution with mean zero and variance σ_e^2 and σ_e^2 respectively. Furthermore, we concern the informative sampling effects by adding an unknown probability inclusion functions $g_2(\pi_{ij})$ into the model:

$$y_{ij} = g_1(x_{ij}) + g_2(\pi_{ij}) + v_i + e_{ij}, \quad j = 1, \dots, n_i;\ i = 1, \dots, M. \tag{3.2}$$

We add the function of inclusion probability to the model based on the Pfeffermann and Sverchkov (2009) affirmation that the inclusion probability is a rough summary that allows the design variables. This statement provides an alternative way to overcome the effects of informative sampling in a model-based approach, especially when it is encountered with a very complex or even unknown sampling design.

We assume that function $g_1(x_{ij})$ and $g_2(\pi_{ij})$ in (3.2) are smooth, therefore it can be approached by a p-spline function. The proposed sample model is:

$$y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ij}^k + \sum_{k=1}^{K_1} t_k (x_{ij} - q_k)_+^p + \sum_{k=1}^{s} \delta_k \pi_{ij}^k + \sum_{k=1}^{K_2} r_k (\pi_{ij} - Q_k)_+^s + v_i + e_{ij}, \tag{3.3}$$

for $j = 1, \dots, n_i$; $i = 1, \dots, M$ where p and s are degrees of p-spline function for $g_1(x_{ij})$ and $g_2(\pi_{ij})$ respectively. The coefficients of the parametric part and spline part for $g_1(x_{ij})$ are expressed sequentially as β_k and t_k. Meanwhile, δ_k and r_k are for $g_2(\pi_{ij})$. Also, the followings were defined $(x_{ij} - q_k)_+^p = \max(0, x_{ij} - q_k)^p$ and $(\pi_{ij} - Q_k)_+^s = \max(0, \pi_{ij} - Q_k)^s$. Hereafter, q_k and Q_k are knots in x and π respectively. Model (3.3) is defined for each area in the population. Let $\boldsymbol{x}_{ij} = [1\ \ x_{ij}\ \cdots\ x_{ij}^p]$, $\boldsymbol{\pi}_{ij} = [\pi_{ij}\ \pi_{ij}^2\ \cdots\ \pi_{ij}^s]$, and $\boldsymbol{z}_{1ij} = [(x_{ij} - q_1)_+^p\ \cdots\ (x_{ij} - q_{K_1})_+^p]$, $\boldsymbol{z}_{2ij} = [(\pi_{ij} - Q_1)_+^s\ \cdots\ (\pi_{ij} - Q_{K_2})_+^s]$. By defining the following vectors, $\boldsymbol{\beta} = [\beta_0\ \cdots\ \beta_p]^t$, $\boldsymbol{\delta} = [\delta_0\ \cdots\ \delta_p]^t$, $\boldsymbol{h}_1 = [t_1\ \cdots\ t_{K_1}]^t$, and $\boldsymbol{h}_2 = [r_1\ \cdots\ r_{K_2}]^t$, equation (3.3) can restated as

$$y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{\pi}_{ij}\boldsymbol{\delta} + \boldsymbol{z}_{1ij}\boldsymbol{h}_1 + z_{2ij}\boldsymbol{h}_2 + v_i + e_{ij}; \quad j = 1, \dots, n_i;\ i = 1, \dots, M. \tag{3.4}$$

Furthermore, by denoting $\mathrm{col}\{a_i\}_{i=1}^n$ as column matrix with element $(a_1, \dots, a_n)$ and $\mathrm{diag}\{a_i\}_{i=1}^n$ as diagonal matrix with diagonal element $(a_1, \dots, a_n)$, we define matrices $\boldsymbol{Z}_1 = \mathrm{col}(\{\mathrm{col}(\{\boldsymbol{z}_{1ij}\}_{j=1}^{n_i})\}_{i=1}^M)$, $\boldsymbol{Z}_2 = \mathrm{col}(\{\mathrm{col}(\{\boldsymbol{z}_{2ij}\}_{j=1}^{n_i})\}_{i=1}^M)$, $\boldsymbol{Z}_3 = \mathrm{diag}(\{\mathbf{1}_{n_i}\}_{i=1}^M)$, $\boldsymbol{h}_3 = [v_1\ \cdots\ v_M]^t$, and $\boldsymbol{e} = \mathrm{col}(\{\mathrm{col}(\{e_{ij}\}_{j=1}^{n_i})\}_{i=1}^M)$.

Using all sample data, model (3.4) can be stated briefly in a matrix form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\Pi\delta} + \boldsymbol{Zh} + \mathbf{e}, \tag{3.5}$$

where $\boldsymbol{y} = \mathrm{col}(\{\mathrm{col}(\{y_{1ij}\}_{j=1}^{n_i})\}_{i=1}^M)$, $\boldsymbol{X} = \mathrm{col}(\{\mathrm{col}(\{\boldsymbol{x}_{1ij}\}_{j=1}^{n_i})\}_{i=1}^M)$, and $\boldsymbol{\Pi} = \mathrm{col}(\{\mathrm{col}(\{\boldsymbol{\pi}_{1ij}\}_{j=1}^{n_i})\}_{i=1}^M)$. Meanwhile, $\boldsymbol{h} = \begin{bmatrix}\boldsymbol{h}_1^t & \boldsymbol{h}_2^t & \boldsymbol{h}_3^t\end{bmatrix}^t$ and $\boldsymbol{Z} = \begin{bmatrix}\boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3\end{bmatrix}$ are partition matrices. We assume also that matrix

$\boldsymbol{X}_{n\text{x}(p+1)}$ and $\boldsymbol{\Pi}_{n\text{x}s}$ have full column rank. Penalized spline fitting criteria after divided by $\sigma^2_{\boldsymbol{e}}$, is equal to best unbiased linear prediction criteria in mixed-model if $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are considered as a set of random coefficients with $\text{cov}(\boldsymbol{h}_1) = \sigma^2_{\boldsymbol{h}_1}\boldsymbol{I}_{K_1}$ and $\text{cov}(\boldsymbol{h}_2) = \sigma^2_{\boldsymbol{h}_2}\boldsymbol{I}_{K_2}$. In our case, parameters of smoothing p-spline fulfill $\lambda_1^2 = \sigma^2_{\boldsymbol{e}}/\sigma^2_{\boldsymbol{h}_1}$ and $\lambda_2^2 = \sigma^2_{\boldsymbol{e}}/\sigma^2_{\boldsymbol{h}_2}$. We assume that $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ independently and identically normal distribution around zero with a certain variance. In brief, least square problem in p-spline in this paper is equivalent to best unbiased prediction problem in (3.5). Let $\boldsymbol{G}_k = \sigma^2_{\boldsymbol{h}_k}\boldsymbol{I}_{K_k}$ for $k = 1, 2, 3$ where $K_3 = M$. The model involves functions which are mixture of fixed parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ and linear function for random quantity in $\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3$, where $\boldsymbol{h}_k \overset{iid}{\sim} N(0, \sigma^2_{\boldsymbol{h}_k}\boldsymbol{I}_{K_k})$, for $k = 1, 2, 3$ and $\boldsymbol{e} \overset{iid}{\sim} N(0, \sigma^2_{\boldsymbol{e}}\boldsymbol{I}_n)$.

3.1. Estimator of coefficient vector

Estimators for $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and predictors for $\boldsymbol{h}$ can be obtained simultaneously through Henderson mixed-model equation. By defining $\boldsymbol{Z}_4 = \boldsymbol{I}_n, \boldsymbol{G}_4 = \sigma^2_{\boldsymbol{h}_4}\boldsymbol{I}_{K_4}$, where $\boldsymbol{h}_4 = \boldsymbol{e}$, and $K_4 = n$ ($n = \sum_{i=1}^{M} n_i$), variance of $\boldsymbol{y}$ can be written as

$$\boldsymbol{V} = \sum_{k=1}^{4} \boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t, \tag{3.6}$$

$\boldsymbol{V}$ is a diagonal matrix so $\boldsymbol{V}$ is symmetric. Subsequently, by assuming $(\boldsymbol{y}|\boldsymbol{h}) \sim N(\boldsymbol{X\beta}+\boldsymbol{\Pi\delta}+\boldsymbol{Zh}, \boldsymbol{G}_4)$, the conditional density for y with given $\boldsymbol{h}$ can be written as $f(\boldsymbol{y}|\boldsymbol{h}) \approx \exp\{-(1/2)(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{\Pi\delta} - \boldsymbol{Zh})^t\boldsymbol{G}_4^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{\Pi\delta} - \boldsymbol{Zh})\}$. The joint density for $\boldsymbol{y}$ and $\boldsymbol{h}$ can be obtained using formula $f(\boldsymbol{y}, \boldsymbol{h}) = f(\boldsymbol{y}|\boldsymbol{h})f(\boldsymbol{h}) = f(\boldsymbol{y}|\boldsymbol{h})f(\boldsymbol{h}_1)f(\boldsymbol{h}_2)f(\boldsymbol{h}_3)$. The joint density function $f(\boldsymbol{y}, \boldsymbol{h})$ is then maximized against $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, $\boldsymbol{h}_3$ through its logarithm function to get the following Henderson mixed-model equation,

$$\begin{bmatrix} \boldsymbol{X}^t\boldsymbol{X} & \boldsymbol{X}^t\boldsymbol{\Pi} & \boldsymbol{X}^t\boldsymbol{Z}_1 & \boldsymbol{X}^t\boldsymbol{Z}_2 & \boldsymbol{X}^t\boldsymbol{Z}_3 \\ \boldsymbol{\Pi}^t\boldsymbol{X} & \boldsymbol{\Pi}^t\boldsymbol{\Pi} & \boldsymbol{\Pi}^t\boldsymbol{Z}_1 & \boldsymbol{\Pi}^t\boldsymbol{Z}_2 & \boldsymbol{\Pi}^t\boldsymbol{Z}_3 \\ \boldsymbol{Z}_1{}^t\boldsymbol{X} & \boldsymbol{Z}_1{}^t\boldsymbol{\Pi} & \boldsymbol{Z}_1{}^t\boldsymbol{Z}_1 + \lambda_1^2\boldsymbol{I}_{K_1} & \boldsymbol{Z}_1{}^t\boldsymbol{Z}_2 & \boldsymbol{Z}_1{}^t\boldsymbol{Z}_3 \\ \boldsymbol{Z}_2{}^t\boldsymbol{X} & \boldsymbol{Z}_2{}^t\boldsymbol{\Pi} & \boldsymbol{Z}_2{}^t\boldsymbol{Z}_1 & \boldsymbol{Z}_2{}^t\boldsymbol{Z}_2 + \lambda_2^2\boldsymbol{I}_{K_2} & \boldsymbol{Z}_2{}^t\boldsymbol{Z}_3 \\ \boldsymbol{Z}_3{}^t\boldsymbol{X} & \boldsymbol{Z}_3{}^t\boldsymbol{\Pi} & \boldsymbol{Z}_3{}^t\boldsymbol{Z}_1 & \boldsymbol{Z}_3{}^t\boldsymbol{Z}_2 & \boldsymbol{Z}_3{}^t\boldsymbol{Z}_3 + \lambda_3^2\boldsymbol{I}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \boldsymbol{h}_1 \\ \boldsymbol{h}_2 \\ \boldsymbol{h}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^t\boldsymbol{y} \\ \boldsymbol{\Pi}^t\boldsymbol{y} \\ \boldsymbol{Z}_1{}^t\boldsymbol{y} \\ \boldsymbol{Z}_2{}^t\boldsymbol{y} \\ \boldsymbol{Z}_3{}^t\boldsymbol{y} \end{bmatrix}, \tag{3.7}$$

where $\lambda_3^2 = \sigma^2_{\boldsymbol{e}}/\sigma^2_{\boldsymbol{h}_3}$. Let $\boldsymbol{J}$ is the invers of coefficient matrix on equation (3.7) with $\boldsymbol{J}_{bb}$ is the element in the b^{th} row and b^{th} column in matrix $\boldsymbol{J}$. The solution of (3.7) leads to an estimators for the model coefficient parameters, that is

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\delta}} \\ \hat{\boldsymbol{h}}_1 \\ \hat{\boldsymbol{h}}_2 \\ \hat{\boldsymbol{h}}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1^t\boldsymbol{y} \\ \boldsymbol{U}_2^t\boldsymbol{y} \\ \boldsymbol{U}_3^t\boldsymbol{y} \\ \boldsymbol{U}_4^t\boldsymbol{y} \\ \boldsymbol{U}_5^t\boldsymbol{y} \end{bmatrix}, \tag{3.8}$$

where, $\boldsymbol{U}_1^t = \boldsymbol{J}_{11}\boldsymbol{X}^t + \boldsymbol{J}_{12}\boldsymbol{\Pi}^t + \boldsymbol{J}_{13}\boldsymbol{Z}_1{}^t + \boldsymbol{J}_{14}\boldsymbol{Z}_2{}^t + \boldsymbol{J}_{15}\boldsymbol{Z}_3{}^t$, $\boldsymbol{U}_2^t = \boldsymbol{J}_{21}\boldsymbol{X}^t + \boldsymbol{J}_{22}\boldsymbol{\Pi}^t + \boldsymbol{J}_{23}\boldsymbol{Z}_1{}^t + \boldsymbol{J}_{24}\boldsymbol{Z}_2{}^t + \boldsymbol{J}_{25}\boldsymbol{Z}_3{}^t$, $\boldsymbol{U}_3^t = \boldsymbol{J}_{31}\boldsymbol{X}^t + \boldsymbol{J}_{32}\boldsymbol{\Pi}^t + \boldsymbol{J}_{33}\boldsymbol{Z}_1{}^t + \boldsymbol{J}_{34}\boldsymbol{Z}_2{}^t + \boldsymbol{J}_{35}\boldsymbol{Z}_3{}^t$, $\boldsymbol{U}_4^t = \boldsymbol{J}_{41}\boldsymbol{X}^t + \boldsymbol{J}_{42}\boldsymbol{\Pi}^t + \boldsymbol{J}_{43}\boldsymbol{Z}_1{}^t + \boldsymbol{J}_{44}\boldsymbol{Z}_2{}^t + \boldsymbol{J}_{45}\boldsymbol{Z}_3{}^t$, and $\boldsymbol{U}_5^t = \boldsymbol{J}_{51}\boldsymbol{X}^t + \boldsymbol{J}_{52}\boldsymbol{\Pi}^t + \boldsymbol{J}_{53}\boldsymbol{Z}_1{}^t + \boldsymbol{J}_{54}\boldsymbol{Z}_2{}^t + \boldsymbol{J}_{55}\boldsymbol{Z}_3{}^t$.

3.2. Properties of estimator of coefficient vector

Properties of (3.8) are derived in line with the ideas of Rao *et al.* (2014). By writing

$$\begin{bmatrix} \boldsymbol{X} & \Pi & \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix}^t \begin{bmatrix} \boldsymbol{X} & \Pi & \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix} \text{ as } \boldsymbol{J}^{-1} - \text{diag}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \lambda_1^2 \boldsymbol{I}_{K_1} & \lambda_2^2 \boldsymbol{I}_{K_2} & \lambda_3^2 \boldsymbol{I}_M \end{bmatrix}$$

thus,

$$\begin{bmatrix} \boldsymbol{U}_1^t \\ \boldsymbol{U}_2^t \\ \boldsymbol{U}_3^t \\ \boldsymbol{U}_4^t \\ \boldsymbol{U}_5 \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \Pi & \boldsymbol{Z}_1 & \boldsymbol{Z}_2 & \boldsymbol{Z}_3 \end{bmatrix} = \boldsymbol{J}\left(\boldsymbol{J}^{-1} - \text{diag}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \lambda_1^2 \boldsymbol{I}_{K_1} & \lambda_2^2 \boldsymbol{I}_{K_2} & \lambda_3^2 \boldsymbol{I}_M \end{bmatrix}\right). \tag{3.9}$$

From (3.8) and (3.9) we may obtain $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{U}_1^t(\boldsymbol{X\beta} + \Pi\boldsymbol{\delta}) = \boldsymbol{\beta}$ and $E(\hat{\boldsymbol{\delta}}) = \boldsymbol{U}_2^t(\boldsymbol{X\beta} + \boldsymbol{\pi\delta}) = \boldsymbol{\delta}$ which show that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ are unbiased, respectively. Likewise, $E(\hat{\boldsymbol{h}}_1 - \boldsymbol{h}_1) = \boldsymbol{U}_3^t(\boldsymbol{X\beta} + \Pi\boldsymbol{\delta}) - \boldsymbol{0} = \boldsymbol{0}$, $E(\hat{\boldsymbol{h}}_2 - \boldsymbol{h}_2) = \boldsymbol{U}_4^t(\boldsymbol{X\beta} + \Pi\boldsymbol{\delta}) - \boldsymbol{0} = \boldsymbol{0}$, $E(\hat{\boldsymbol{h}}_3 - \boldsymbol{h}_3) = \boldsymbol{U}_5^t(\boldsymbol{X\beta} + \Pi\boldsymbol{\delta}) - \boldsymbol{0} = \boldsymbol{0}$ which also show that $\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2$, and $\hat{\boldsymbol{h}}_3$ are unbiased, respectively.

Because $V(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{h}_2, \boldsymbol{h}_3) = \boldsymbol{Z}_1\boldsymbol{G}_1\boldsymbol{Z}_1^t + \boldsymbol{G}_4$ and $E(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{h}_2, \boldsymbol{h}_3) = \boldsymbol{X\beta} + \Pi\boldsymbol{\delta} + \boldsymbol{Z}_2\boldsymbol{h}_2 + \boldsymbol{Z}_3\boldsymbol{h}_3$, and using relation $E(a|b) = E(a) + \text{cov}(a, b).V(b)^{-1}(b - E(b))$, we get

$$E\left(\boldsymbol{h}_1|(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{h}_2, \boldsymbol{h}_3)\right) = \boldsymbol{G}_1\boldsymbol{Z}_1^t\left(\boldsymbol{Z}_1\boldsymbol{G}_1\boldsymbol{Z}_1^t + \boldsymbol{G}_4\right)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta} - \Pi\boldsymbol{\delta} - \boldsymbol{Z}_2\boldsymbol{h}_2 - \boldsymbol{Z}_3\boldsymbol{h}_3).$$

Meanwhile, using the third row of (3.7) we also obtain

$$\hat{\boldsymbol{h}}_1 = \boldsymbol{G}_1\boldsymbol{Z}_1^t\left(\boldsymbol{Z}_1\boldsymbol{G}_1\boldsymbol{Z}_1^t + \boldsymbol{G}_4\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \Pi\hat{\boldsymbol{\delta}} - \boldsymbol{Z}_2\hat{\boldsymbol{h}}_2 - \boldsymbol{Z}_3\hat{\boldsymbol{h}}_3\right).$$

It is proven that $\hat{\boldsymbol{h}}_1 = E(\boldsymbol{h}_1|(\boldsymbol{y}|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{h}}_2, \hat{\boldsymbol{h}}_3))$. Thus $\hat{\boldsymbol{h}}_1$ is the best predictor for $\boldsymbol{h}_1$. The best predictor properties of $\hat{\boldsymbol{h}}_2$ and $\hat{\boldsymbol{h}}_3$ are proven similarly.

3.3. Estimation of variance components

We estimate the variance components $\sigma^2_{\boldsymbol{h}_k}$, for $k = 1, 2, 3, 4$ by using restricted maximum likelihood (REML) method which was first stated by Patterson and Thompson (1971). This estimation is based on linear combination of element $\boldsymbol{y}$, namely $\boldsymbol{K}^t\boldsymbol{y}$, chosen as such that $\boldsymbol{K}^t\boldsymbol{y}$ does not contain fixed effect, that is $\boldsymbol{K}^t\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{K}^t\Pi = \boldsymbol{0}$. We first determine $\boldsymbol{K}^t$, construct REML equation by maximize the likelihood function of $\boldsymbol{K}^t$ to $\sigma^2_{\boldsymbol{h}_k}$, $(k = 1, 2, 3, 4)$, and solving the REML equation to get REML estimator of $\sigma^2_{\boldsymbol{e}}$ and $\sigma^2_{\boldsymbol{h}_k}$.

3.3.1. Determine $\boldsymbol{K}^t$

We determine the matrix $\boldsymbol{K}^t$ by first rewriting fixed effects in (3.5) in partition matrices form, that is $\boldsymbol{y} = [\boldsymbol{X} \;\; \Pi] \begin{bmatrix} \boldsymbol{\beta}^t & \boldsymbol{\delta}^t \end{bmatrix}^t + \sum_{k=1}^3 \boldsymbol{Z}_k\boldsymbol{h}_k + \boldsymbol{e}$. Suppose $\boldsymbol{K}^t$ is as such, thus $[\boldsymbol{K}^t\boldsymbol{X} \;\; \boldsymbol{K}^t\Pi] = \boldsymbol{K}^t[\boldsymbol{X} \;\; \Pi] = [\boldsymbol{0} \;\; \boldsymbol{0}]$. Eligible $\boldsymbol{K}^t$ can be determined as:

$$\boldsymbol{K}^t = \boldsymbol{I} - \begin{bmatrix} \boldsymbol{X} & \Pi \end{bmatrix}\left(\begin{bmatrix} \boldsymbol{X}^t \\ \Pi^t \end{bmatrix}\begin{bmatrix} \boldsymbol{X} & \Pi \end{bmatrix}\right)^{-1}\begin{bmatrix} \boldsymbol{X}^t \\ \Pi^t \end{bmatrix}. \tag{3.10}$$

By developing matrix inverse on (3.10) we have

$$\boldsymbol{K}^t = \boldsymbol{M} - \boldsymbol{M}\boldsymbol{X}\left(\boldsymbol{X}^t\boldsymbol{M}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{M}, \tag{3.11}$$

where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t$. It is easy to prove that $\boldsymbol{M}$ and $\boldsymbol{K}^t$ are idempotent and symmetric matrix. Another form of $\boldsymbol{K}^t$ is $\boldsymbol{K}^t = (\boldsymbol{M} - \boldsymbol{M}\boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{M}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t\boldsymbol{M})$ with $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t$.

3.3.2. REML equation

Furthermore, REML equation will be formed by first establishing likelihood function for $\boldsymbol{K}^t\boldsymbol{y}$. By assuming $\boldsymbol{y} \sim N(\boldsymbol{X\beta} + \boldsymbol{\Pi\delta}, \boldsymbol{V})$ thus $\boldsymbol{K}^t\boldsymbol{y} \sim N(\boldsymbol{0}, \boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K})$, the likelihood function for $\boldsymbol{K}^t\boldsymbol{y}$ can be stated as $L \approx (|\boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K}|)^{-1/2}e^{-(1/2)\boldsymbol{y}^t\boldsymbol{K}(\boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}^t\boldsymbol{y}}$. By maximizing the logarithm function of L on $\sigma^2_{h_k}$, for $k = 1, 2, 3, 4$, we obtain the REML equation in trace matrix:

$$\operatorname{tr}\left(\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma^2_{h_k}}\right) = \boldsymbol{y}^t\boldsymbol{P}\frac{\partial \boldsymbol{V}}{\partial \sigma^2_{h_k}}\boldsymbol{P}\boldsymbol{y}, \tag{3.12}$$

with $\boldsymbol{P} = \boldsymbol{K}(\boldsymbol{K}^t\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}^t$. It is clear that $\boldsymbol{P}$ is a idempotent and symmetric matrix. Considering $\partial\boldsymbol{V}/\partial\sigma^2_{h_k} = \boldsymbol{Z}_k\boldsymbol{Z}^t_k$ for $k = 1, 2, 3, 4$, the REML equation (3.12) can be restated:

$$\operatorname{col}\left\{\operatorname{tr}\left(\boldsymbol{P}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\right)\right\}^4_{k=1} = \operatorname{col}\left\{\boldsymbol{y}^t\boldsymbol{P}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\boldsymbol{P}\boldsymbol{y}\right\}^4_{k=1}$$

that is equivalent to

$$\operatorname{tr}(\boldsymbol{P}) = \boldsymbol{y}^t\boldsymbol{P}^2\boldsymbol{y}, \quad \text{for } k = 4 \tag{3.13}$$

and

$$\operatorname{tr}\left(\boldsymbol{P}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\right) = \boldsymbol{y}^t\boldsymbol{P}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\boldsymbol{P}\boldsymbol{y}, \quad \text{for } k = 1, 2, 3. \tag{3.14}$$

3.3.3. REML estimator of $\sigma^2_{\boldsymbol{e}}$

Estimator for variance of error can be obtained by first multiplying (3.14) with $\sigma^2_{\boldsymbol{h}_k}$ and then adding for $k = 1, 2, 3$. We get

$$\operatorname{tr}\left(\boldsymbol{P}\sum_{k=1}^{3}\sigma^2_{h_k}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\right) = \boldsymbol{y}^t\boldsymbol{P}\sum_{k=1}^{3}\sigma^2_{h_k}\boldsymbol{Z}_k\boldsymbol{Z}^t_k\boldsymbol{P}\boldsymbol{y}. \tag{3.15}$$

Then from (3.15) by considering $\sum_{k=1}^{3}\sigma^2_{\boldsymbol{h}_k}\boldsymbol{Z}_k\boldsymbol{Z}^t_k = \boldsymbol{V} - \sigma^2_{\boldsymbol{h}_4}\boldsymbol{Z}_4\boldsymbol{Z}^t_4$, $\boldsymbol{P}\boldsymbol{V}\boldsymbol{P} = \boldsymbol{P}$ and (3.13), we can write

$$\operatorname{tr}(\boldsymbol{P}\boldsymbol{V}) = \boldsymbol{y}^t\boldsymbol{P}\boldsymbol{y}. \tag{3.16}$$

We used the following theorem stated in Rencher and Schaalje (2008) to develop (3.16).

Theorem 1. *If A is symmetric and idempotent matrix of rank r, then rank(A) = tr(A) = r.*

The left hand side of (3.16) described as follows. $\boldsymbol{V}$ is diagonal matrix with full rank, therefore

$$\begin{aligned}\operatorname{rank}(\boldsymbol{P}\boldsymbol{V}) &= \operatorname{rank}(\boldsymbol{P}) \\ &= \operatorname{rank}\left(\boldsymbol{K}\left(\boldsymbol{K}^t\boldsymbol{K}\right)^{-1}\boldsymbol{K}^t\right) \\ &= \operatorname{rank}\left(\boldsymbol{K}^t\right).\end{aligned} \tag{3.17}$$

Because $\boldsymbol{PVP} = \boldsymbol{P}$ we get $\boldsymbol{PVPV} = \boldsymbol{PV}$ in other word, $\boldsymbol{PV}$ is idempotent. Furthermore,

$$\begin{aligned}(\boldsymbol{PV})^t &= \boldsymbol{V}^t\boldsymbol{P}^t\\ &= \boldsymbol{VP} \quad \text{(because } \boldsymbol{P} \text{ and } \boldsymbol{V} \text{ are symmetric)}\\ &= \boldsymbol{P}^{-1}\boldsymbol{PVP}\\ &= \boldsymbol{I} \quad \text{(because } \boldsymbol{PVP} = \boldsymbol{P}\text{)}.\end{aligned}$$

Because $\boldsymbol{I}$ is symmetric, we can get $(\boldsymbol{PV})^t = \boldsymbol{PV}$, in other word $\boldsymbol{PV}$ is symmetric. By using Theorem (1) and (3.17), we can write the left hand side of (3.16) as:

$$\begin{aligned}\text{tr}(\boldsymbol{PV}) &= \text{rank}(\boldsymbol{PV})\\ &= \text{rank}(\boldsymbol{K}^t). \end{aligned} \tag{3.18}$$

Matrix $\boldsymbol{M}$ is idempotent and symmetric so that $\boldsymbol{MX}(\boldsymbol{X}^t\boldsymbol{MX})^{-1}\boldsymbol{X}^t\boldsymbol{M}$ idempotent and symmetric, therefore by theorem (1), $\text{rank}(\boldsymbol{MX}(\boldsymbol{X}^t\boldsymbol{MX})^{-1}\boldsymbol{X}^t\boldsymbol{M}) = \text{tr}(\boldsymbol{MX}(\boldsymbol{X}^t\boldsymbol{MX})^{-1}\boldsymbol{X}^t\boldsymbol{M})$. Matrix $\boldsymbol{\Pi}_{nxs}$ has full column rank. Using (3.11), equation (3.18) becomes

$$\begin{aligned}\text{tr}(\boldsymbol{PV}) &= \text{rank}(\boldsymbol{M}) - \text{rank}\left(\boldsymbol{MX}\left(\boldsymbol{X}^t\boldsymbol{MX}\right)^{-1}\boldsymbol{X}^t\boldsymbol{M}\right)\\ &= n - s - \text{tr}\left(\boldsymbol{MX}\left(\boldsymbol{X}^t\boldsymbol{MX}\right)^{-1}\boldsymbol{X}^t\boldsymbol{M}\right).\end{aligned} \tag{3.19}$$

We describe the right hand side of (3.16) by first rewriting $\boldsymbol{P}$ in the following theorem.

Theorem 2. *If $\boldsymbol{K}^t\boldsymbol{X} = \boldsymbol{0}$ and $\boldsymbol{K}^t\boldsymbol{\Pi} = \boldsymbol{0}$ and $\boldsymbol{V}$ is a positive definite matrix, then*

$$\boldsymbol{K}\left(\boldsymbol{K}^t\boldsymbol{VK}\right)^{-1}\boldsymbol{K}^t = \boldsymbol{P}, \tag{3.20}$$

with $\boldsymbol{P} \equiv \boldsymbol{M}^* - \boldsymbol{M}^*\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{M}^*\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{M}^*$ and $\boldsymbol{M}^* = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{V}^{-1}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t\boldsymbol{V}^{-1}$.

Proof: It is clear that $\boldsymbol{V}$ is symmetric, nonsingular with full rank. Consequently, symmetric matrix $\boldsymbol{V}^{1/2}$ always exists as such so that $\boldsymbol{V} = (\boldsymbol{V}^{1/2})^2$. We can obtain $(\boldsymbol{V}^{1/2}\boldsymbol{K})^t\boldsymbol{V}^{-1/2}\boldsymbol{X} = \boldsymbol{0}$ and $(\boldsymbol{V}^{1/2}\boldsymbol{K})^t\boldsymbol{V}^{-1/2}\boldsymbol{\Pi} = \boldsymbol{0}$ as a result of $\boldsymbol{K}^t\boldsymbol{X} = \boldsymbol{0}$ dan $\boldsymbol{K}^t\boldsymbol{\Pi} = \boldsymbol{0}$. Applying Searle's idea in our case, $\boldsymbol{K}$ replaced with $\boldsymbol{V}^{1/2}\boldsymbol{K}$, $\boldsymbol{X}$ with $\boldsymbol{V}^{-1/2}\boldsymbol{X}$ and $\boldsymbol{\Pi}$ with $\boldsymbol{V}^{-1/2}\boldsymbol{\Pi}$ in the equation $\boldsymbol{K}(\boldsymbol{K}^t\boldsymbol{K})^{-1}\boldsymbol{K}^t = \boldsymbol{K}^t$ (Searle, 1982). By multiplying $\boldsymbol{V}^{-1/2}$ on the left and right sides, we get

$$\boldsymbol{K}\left(\boldsymbol{K}^t\boldsymbol{VK}\right)^{-1}\boldsymbol{K}^t = \boldsymbol{M}^* - \boldsymbol{M}^*\boldsymbol{X}\left(\boldsymbol{X}^t\boldsymbol{M}^*\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{M}^* \equiv \boldsymbol{P}, \tag{3.21}$$

where $\boldsymbol{M}^* = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{V}^{-1}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t\boldsymbol{V}^{-1}$ and $\boldsymbol{V}^{-1}$ is the matrix inverse of variance $\boldsymbol{y}$. Let $\boldsymbol{D} = \text{diag}(\{\boldsymbol{G}_k\}_{k=1}^3)$, we get

$$\boldsymbol{V}^{-1} = \boldsymbol{G}_4^{-1} - \boldsymbol{G}_4^{-1}\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{G}_4^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1}\right)^{-1}\boldsymbol{Z}'\boldsymbol{G}_4^{-1}. \tag{3.22}$$

□

An alternative for $\boldsymbol{P}$ is $\boldsymbol{P} \equiv \boldsymbol{M}^{**} - \boldsymbol{M}^{**}\boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{M}^{**}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t\boldsymbol{M}^{**}$ where $\boldsymbol{M}^{**} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^t$ $\boldsymbol{V}^{-1}$ using $\boldsymbol{K}^t = (\boldsymbol{M} - \boldsymbol{M}\boldsymbol{\Pi}(\boldsymbol{\Pi}^t\boldsymbol{M}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^t\boldsymbol{M})$. The matrix $\hat{\boldsymbol{h}}$ may be stated as:

$$\hat{\boldsymbol{h}} = \left(\boldsymbol{Z}^t\boldsymbol{G}_4^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1}\right)^{-1}\boldsymbol{Z}^t\boldsymbol{G}_4^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{\Pi}\hat{\boldsymbol{\delta}}\right), \tag{3.23}$$

obtained using third, fourth, and fifth row of (3.7) with $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ in place of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ respectively. Substituting $\hat{\boldsymbol{h}}$ to the first and second row of equation (3.7) and considering (3.22), resulting in the following equation,

$$\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}^t\boldsymbol{V}^{-1}\Pi\hat{\boldsymbol{\delta}} = \boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{3.24}$$

$$\Pi^t\boldsymbol{V}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}} + \Pi^t\boldsymbol{V}^{-1}\Pi\hat{\boldsymbol{\delta}} = \Pi^t\boldsymbol{V}^{-1}\boldsymbol{y}. \tag{3.25}$$

By substituting $\hat{\boldsymbol{\beta}}$ from (3.25) into (3.24), and using $\boldsymbol{P}$ in Theorem 2, we get

$$\boldsymbol{VPy} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \Pi\hat{\boldsymbol{\delta}}\right). \tag{3.26}$$

By using (3.19), (3.26) and relation $\boldsymbol{PVP} = \boldsymbol{P}$, we can write equation (3.16) as:

$$n - s - \operatorname{tr}\left(\boldsymbol{MX}\left(\boldsymbol{X}^t\boldsymbol{MX}\right)^{-1}\boldsymbol{X}^t\boldsymbol{M}\right) = \boldsymbol{y}^t\boldsymbol{PVV}^{-1}\boldsymbol{VPy}. \tag{3.27}$$

Substituting (3.22) in (3.27), the right-hand side of (3.27) becomes $\boldsymbol{G}_4^{-1}\boldsymbol{y}^t(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \Pi\hat{\boldsymbol{\delta}} - \boldsymbol{Z}\hat{\boldsymbol{h}})$ where $\hat{\boldsymbol{h}}$ as showed in (3.23) and $\boldsymbol{G}_4^{-1} = \sigma_{\boldsymbol{e}}^{-2}\boldsymbol{I}_n$. Thus, we can determine the estimator for $\sigma_{\boldsymbol{e}}^2$:

$$\hat{\sigma}_{\boldsymbol{e}}^2 = \frac{\boldsymbol{y}^t\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \Pi\hat{\boldsymbol{\delta}} - \boldsymbol{Z}\hat{\boldsymbol{h}}\right)}{n - s - w}, \tag{3.28}$$

where $w = \operatorname{tr}(\boldsymbol{MX}(\boldsymbol{X}^t\boldsymbol{MX})^{-1}\boldsymbol{X}^t\boldsymbol{M})$.

3.3.4. REML estimator of $\sigma_{\boldsymbol{h}_k}^2$

We determine estimator for $\sigma_{\boldsymbol{h}_k}^2$, where $k = 1, 2, 3$ using approach in Searle *et al.* (2006). Take $\boldsymbol{Z} = \boldsymbol{0}$ in $\boldsymbol{V}$ then substitute it into $\boldsymbol{P}$ in (3.20) to get $\boldsymbol{S} = \boldsymbol{M}^* - \boldsymbol{M}^*\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{M}^*\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{M}^*$ with $\boldsymbol{M}^* = \boldsymbol{G}_4^{-1} - \boldsymbol{G}_4^{-1}\Pi(\Pi^t\boldsymbol{G}_4^{-1}\Pi)^{-1}\Pi^t\boldsymbol{G}_4^{-1}$. Matrix $\boldsymbol{P}$ can be stated in $\boldsymbol{S}$ as $\boldsymbol{P} = \boldsymbol{S} - \boldsymbol{SZ}(\boldsymbol{D}^{-1} + \boldsymbol{Z}'\boldsymbol{SZ})^{-1}\boldsymbol{Z}'\boldsymbol{S}$ and can be rewritten:

$$\boldsymbol{P} = \boldsymbol{S} - \boldsymbol{SZDTZ}^t\boldsymbol{S}, \tag{3.29}$$

with $\boldsymbol{T} = (\boldsymbol{I} + \boldsymbol{Z}^t\boldsymbol{SZD})^{-1}$. Let $\boldsymbol{F}_{kk}$ is defined as $\boldsymbol{D}$ with identity element in σ_k^2 and $\boldsymbol{0}$ in σ_j^2; $j \neq k$. Therefore $\boldsymbol{DF}_{kk} = \sigma_{h_k}^2\boldsymbol{F}_{kk}$ or

$$\boldsymbol{F}_{kk} = \frac{\boldsymbol{DF}_{kk}}{\sigma_{h_k}^2}, \quad k = 1, 2, 3. \tag{3.30}$$

Using (3.29) and (3.30), the left-hand side of (3.14) can be stated:

$$\operatorname{tr}\left(\boldsymbol{PZ}_k\boldsymbol{Z}_k^t\right) = \operatorname{tr}\left(\frac{\boldsymbol{F}_{kk} - \boldsymbol{TF}_{kk}}{\sigma_{h_k}^2}\right). \tag{3.31}$$

We obtain $\boldsymbol{DZ}^t\boldsymbol{V}^{-1} = (\boldsymbol{D}^{-1} + \boldsymbol{Z}^t\boldsymbol{G}_4^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^t\boldsymbol{G}_4^{-1}$. Then by using (3.23) and (3.26) we get

$$\begin{aligned}\boldsymbol{DZ}^t\boldsymbol{Py} &= \boldsymbol{DZ}^t\boldsymbol{V}^{-1}\boldsymbol{VPy}\\ &= \hat{\boldsymbol{h}}\end{aligned} \tag{3.32}$$

Substituting (3.30) and (3.32) on the right-hand side of (3.14), we obtain

$$\boldsymbol{y}^t\boldsymbol{P}\boldsymbol{Z}_i\boldsymbol{Z}_i^t\boldsymbol{P}\boldsymbol{y} = \frac{\hat{\boldsymbol{h}}_k^t\hat{\boldsymbol{h}}_k}{\sigma_{h_k}^4}. \tag{3.33}$$

Equation (3.31) and (3.33) are substituted into (3.14) to obtain estimator for $\sigma_{h_k}^2$, $k = 1, 2, 3$, that is

$$\hat{\sigma}_{h_k}^2 = \frac{\hat{\boldsymbol{h}}_k^t\hat{\boldsymbol{h}}_k}{\operatorname{tr}(\boldsymbol{F}_{kk} - \boldsymbol{T}\boldsymbol{F}_{kk})}; \quad k = 1, 2, 3. \tag{3.34}$$

The REML estimator for variance components is obtained using (3.28) and (3.34):

$$\hat{\sigma}_{\boldsymbol{e}}^{2(b+1)} = \frac{\boldsymbol{y}^t\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \Pi\hat{\boldsymbol{\delta}} - \boldsymbol{Z}\hat{\boldsymbol{h}}\right)}{n - s - w}, \tag{3.35}$$

$$\hat{\sigma}_{h_1}^{2(b+1)} = \frac{\hat{\boldsymbol{h}}_1^{t(b)}\hat{\boldsymbol{h}}_1^{(b)}}{K_1 - \operatorname{tr}(\boldsymbol{T}_{11})}, \tag{3.36}$$

$$\hat{\sigma}_{h_2}^{2(b+1)} = \frac{\hat{\boldsymbol{h}}_2^{t(b)}\hat{\boldsymbol{h}}_2^{(b)}}{K_2 - \operatorname{tr}(\boldsymbol{T}_{22})}, \tag{3.37}$$

$$\hat{\sigma}_{h_3}^{2(b+1)} = \frac{\hat{\boldsymbol{h}}_3^{t(b)}\hat{\boldsymbol{h}}_3^{(b)}}{m - \operatorname{tr}(\boldsymbol{T}_{33})}, \tag{3.38}$$

where $\boldsymbol{T}_{kk}$ is the element in the k^{th} row and k^{th} column in matrix $\boldsymbol{T}$. Equations (3.35), (3.36), (3.37), and (3.38) are counted with the following iterative process: (i) determine initial value $\boldsymbol{\theta}^{(0)}$ for $\boldsymbol{\theta} = (\sigma_{\boldsymbol{h}_1}^2, \sigma_{\boldsymbol{h}_2}^2, \sigma_{\boldsymbol{h}_3}^2, \sigma_{\boldsymbol{e}}^2)$; (ii) calculate $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{\delta}^{(0)}$, $\boldsymbol{h}_1^{(0)}$, $\boldsymbol{h}_2^{(0)}$, and $\boldsymbol{h}_3^{(0)}$ using $\boldsymbol{\theta}^{(0)}$ in (3.8) and $\boldsymbol{e}^{(0)} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{(0)} + \Pi\boldsymbol{\delta}^{(0)} + \boldsymbol{Z}\boldsymbol{h}^{(0)}$; (iii) calculate right-hand side of (3.35) to (3.38) using the result in (ii) to get updated value of $\boldsymbol{\theta}^{(1)}$; (iv) repeat step (iii) until $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_{\boldsymbol{h}_1}^2, \hat{\sigma}_{\boldsymbol{h}_2}^2, \hat{\sigma}_{\boldsymbol{h}_3}^2, \hat{\sigma}_{\boldsymbol{e}}^2)$ convergent. The obtained value $\hat{\boldsymbol{\theta}}$ is then used in (3.8) to obtain estimators $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}, \hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \hat{\boldsymbol{h}}_3$.

3.4. Predictor of mean and performance model

The prediction for mean of i^{th} small area is calculated:

$$\hat{\mu}_i = \frac{1}{N_i}\left\{\sum_{i \in S_i} y_{ij} + \sum_{i \in \bar{S}_i} \hat{y}_{ij}\right\}, \tag{3.39}$$

where S_i and $\bar{S}_i$ are consecutively state sample unit sets and non-sample unit sets in area-i and $\hat{y}_{ij} = \boldsymbol{x}_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{\pi}_{ij}\hat{\boldsymbol{\delta}} + z_{1ij}\hat{\boldsymbol{h}}_1 + z_{2ij}\hat{\boldsymbol{h}}_2 + \hat{v}_i$ is the predictor of y_{ij} for $j \in \bar{S}_i$ using the proposed model.

We apply the bootstrap procedure stated by Rao *et al.* (2014) to get conditional bootstrap estimation for root mean square error (RMSE) and absolute bias (AB) for $\hat{\mu}_i$. The steps are: (i) generate $\boldsymbol{h}_3^* = [v_1^{*t} \cdots v_M^{*t}]^t$ and $\boldsymbol{e}^*$ with $\boldsymbol{h}_3^* \sim N(\boldsymbol{0}, \hat{\sigma}_{\boldsymbol{h}_3}^2\boldsymbol{I}_M)$ and $\boldsymbol{e}^* \sim N(\boldsymbol{0}, \hat{\sigma}_{\boldsymbol{e}}^2\boldsymbol{I}_n)$; (ii) calculate response $y_{ij}^* = \boldsymbol{x}_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{\pi}_{ij}\hat{\boldsymbol{\delta}} + z_{1ij}\hat{\boldsymbol{h}}_1 + z_{2ij}\hat{\boldsymbol{h}}_2 + v_i^* + e_{ij}^*$ for $j = 1, \ldots, N_i$, $i = 1, \ldots, M$; (iii) calculate bootstrap estimation for $\hat{\boldsymbol{\beta}}^*$, $\hat{\boldsymbol{\delta}}^*$, $\hat{\boldsymbol{h}}_1^*$, $\hat{\boldsymbol{h}}_2^*$, v_i^* by using sample data $(y_{ij}^*, \boldsymbol{x}_{ij}, \boldsymbol{\pi}_{ij})$ for $j \in s_i$ and $i = 1, \ldots, M$; (iv) calculate predictive values for non-sample areas using $\hat{y}_{ij}^* = \boldsymbol{x}_{ij}\hat{\boldsymbol{\beta}}^* + \boldsymbol{\pi}_{ij}\hat{\boldsymbol{\delta}}^* + z_{1ij}\hat{\boldsymbol{h}}_1^* + z_{2ij}\hat{\boldsymbol{h}}_2^* + v_i^*$; for $j \in \bar{s}_i$; (v) calculate empirical bootstrap prediction for $\bar{Y}_i$ with $\tilde{\mu}_i^* = (1/N_i)\{\sum_{j \in s_i} y_{ij}^* + \sum_{j \in \bar{s}_i} \hat{y}_{ij}^*\}$; and bootstrap population mean with $\bar{Y}_i^* = (1/N_i)\sum_{j=1}^{j=N_i} y_{ij}$; (vi) repeat step (i) to (v) B times; (vii) calculate RMSE and AB for bootstrap estimator for $\tilde{\mu}_i$: $\text{RMSE}_{\text{boot}}(\tilde{\mu}_i) = \{(1/B)\sum_{b=1}^{B}\{\tilde{\mu}_i^*(b) - \bar{Y}_i^*(b)\}^2\}^{1/2}$, and $\text{AB}_{\text{boot}}(\tilde{\mu}_i) = (1/B)\sum_{b=1}^{B}|\tilde{\mu}_i^*(b) - \bar{Y}_i^*(b)|$.

4. Simulation study

Let the population be partitioned into 20 small areas of N_i size that are determined randomly in interval $[800, 1000]$, that is $N_i = \{945, 886, 984, 838, 997, 841, 972, 985, 865, 947, 952, 919, 856, 889, 905, 806, 979, 833, 926, 875\}$. We generated the population data y_{ij} for b^{th} simulation ($b = 1, \ldots, 1000$) by the following linear and quadratic model, respectively:

$$\text{linear case: } y_{ij}^{(b)} = 1 + x_{ij} + v_i^{(b)} + e_{ij}^{(b)}; \quad i = 1, \ldots, 20;\ j = 1, \ldots, N_i, \tag{4.1}$$

and

$$\text{quadratic case : } y_{ij}^{(b)} = 1 + x_{ij} + x_{ij}^2 + v_i^{(b)} + e_{ij}^{(b)}; \quad i = 1, \ldots, 20;\ j = 1, \ldots, N_i, \tag{4.2}$$

where $x_{ij} \overset{iid}{\sim} N(1, 2)$, $v_i \overset{iid}{\sim} N(0, 1)$, and $e_{ij} \overset{iid}{\sim} N(0, 1.5)$. The values of x_{ij} was fixed for each run of b^{th} simulation. The sample size n_i vary by 5, 10, and 15 for every area i. The sample data was drawn based on unequal inclusion probability determined by $\pi_{ij} = n(d_{ij}/\sum_{i=1}^{20} d_{ij})$ without replacement. The size variable d_{ij} was Asparouhov's size measure classified in invariant (I) and non invariant (NI) type. The invariant type was independent of the random area effect v_{ij}, in contrast to non-invariant. We set the weight as 1 and stated both in a row:

$$d_{ij} = \left(1 + \exp\left(-\left(\alpha^{-1}e_{ij} + \sqrt{1-\alpha^{-2}}e_i^*\right)\right)\right)^{-1},$$

for invariant, and

$$d_{ij} = \left(1 + \exp\left(-\left(\alpha^{-1}\left(e_{ij} + v_i\right) + \sqrt{1-\alpha^{-2}}\,(e_i^* + v_i^*)\right)\right)\right)^{-1},$$

for non invariant, Asparouhov (2006). Notation α denoted the level of informative effect. Increasing the α's values indicated decreasing informative effect. We observed α for $1, 2, 3$, or ∞, with $\alpha = \infty$ indicated the sampling was non informative. The error unit $e_i^* \overset{iid}{\sim} N(0, 1.5)$ and random effect $v_i^* \overset{iid}{\sim} N(0, 1)$ were generated independently of e_{ij} and v_i. The samples size n_i varies $5, 10$, and 15 in each area.

The location of k^{th} knots of x was determined by quantile formula:

$$q_k = \left(\frac{k+1}{K_1+2}\right)^{th} \text{ sample quantile of the unique } x_i,$$

for $k = 1, \ldots, K_1$ with $K_1 = \min(1/4 \times \text{ number of unique } x_{ij}, 35)$, as stated by Ruppert *et al.* (2003). The knots Q_k ($k = 1, \ldots, K_2$) for π was obtained analogously. Succinctly, in this simulation we used a linear and quadratic p-spline approach for $g_1(x_{ij})$ and $g_2(\pi_{ij})$ which were used on (3.3) or (3.1). The (3.1) model did not account for informative effect. This model was used as a comparison for the proposed (3.3) model to observe the effect of the addition of function $g_2(\pi_{ij})$ to the model performance. Two cases of the population were constructed using (4.1) and (4.2), where samples were taken with the

Table 1: Models for simulation

Population of y_{ij}	Model approach	Degree of p-spline		Model notation
		$g_1(x_{ij})$	$g_2(\pi_{ij})$	
$y_{ij} = 1 + x_{ij} + v_i + e_{ij}$	$y_{ij} = g_{1ij} + g_{2ij} + v_i + e_{ij}$	1	1	M1
$y_{ij} = 1 + x_{ij} + v_i + e_{ij}$	$y_{ij} = g_{1ij} + g_{2ij} + v_i + e_{ij}$	1	2	M2
$y_{ij} = 1 + x_{ij} + x_{ij}^2 + v_i + e_{ij}$	$y_{ij} = g_{1ij} + g_{2ij} + v_i + e_{ij}$	2	1	N1
$y_{ij} = 1 + x_{ij} + x_{ij}^2 + v_i + e_{ij}$	$y_{ij} = g_{1ij} + g_{2ij} + v_i + e_{ij}$	2	2	N2
$y_{ij} = 1 + x_{ij} + v_i + e_{ij}$	$y_{ij} = g_{1ij} + v_i + e_{ij}$	1	-	R1
$y_{ij} = 1 + x_{ij} + x_{ij}^2 + v_i + e_{ij}$	$y_{ij} = g_{1ij} + v_i + e_{ij}$	2	-	R2

Table 2: The values of $\overline{\text{AB}}$ and $\overline{\text{RMSE}}$ of model M1 and R1 over 20 areas for linear case

Performance measure	α	Size measure	Model M1			Model R1		
			$n_i = 5$	$n_i = 10$	$n_i = 15$	$n_i = 5$	$n_i = 10$	$n_i = 15$
$\overline{\text{AB}}$	1	I	0.1451	0.7779	0.7920	1.0919	1.2184	1.1706
		NI	0.1387	0.5907	0.5884	0.7628	1.0047	0.9332
	2	I	0.1725	0.6744	0.7075	0.6175	1.1499	1.0318
		NI	0.1425	0.7191	0.8323	1.3266	1.1541	1.1335
	3	I	0.1486	0.6370	0.9154	1.3397	1.0592	1.0012
		NI	0.1505	0.7312	0.5961	0.9592	1.0086	1.0429
	∞	I	0.7684	0.6841	0.6858	1.0403	1.2914	1.1369
		NI	0.5741	0.6146	0.7890	1.0649	0.8701	1.0868
$\overline{\text{RMSE}}$	1	I	0.1842	1.2694	1.3022	1.3196	1.4631	1.4065
		NI	0.1837	0.9484	0.9616	0.9268	1.2084	1.1231
	2	I	0.2102	1.0879	1.1480	0.7525	1.3820	1.2395
		NI	0.1806	1.1610	1.3669	1.6024	1.3875	1.3630
	3	I	0.1828	1.0185	1.4941	1.6166	1.2754	1.2031
		NI	0.1800	1.1812	0.9543	1.1578	1.2157	1.2531
	∞	I	1.2126	1.0892	1.1128	1.2583	1.5512	1.3626
		NI	0.8893	0.9638	1.2853	1.2868	1.0508	1.3067

$\overline{\text{AB}}$ = average absolute bias; $\overline{\text{RMSE}}$ = average root mean square error; α = the level of informative effect; I = invariant; NI = non invariant; n_i = sample size; M1 = model (3.3) with linear p-spline for $g_1(x_{ij})$ and $g_2(\pi_{ij})$; R1 = model (3.1), M with linear p-spline for $g_1(x_{ji})$.

inclusion probability determined by the size of invariant (I) and non-invariant (NI). We summarized the models in Table 1.

Values of AB and RMSE were calculated based on $B = 1{,}000$ bootstrap samples. The average absolute bias ($\overline{\text{AB}}$) and average RMSE ($\overline{\text{RMSE}}$) over 20 areas were provided. Table 2 reports the simulation results of $\overline{\text{AB}}$ and $\overline{\text{RMSE}}$ over 20 areas for linear case produced by M1 and R1, for each level of informative effect and sample size taken. As indicated in Table 2, overall, M1 produced a smaller $\overline{\text{AB}}$ than that of the R1. The average bias in the simulated linear case can be reduced by adding the linear p-spline function of the inclusion probability into the semiparametric model. The minimum value of $\overline{\text{AB}}$ for each sample size in a non-invariant case produced by M1 (0.1387 for $n = 5$; 0.5907 for $n = 10$ and 0.5888 for $n = 15$) occurred in a very informative sample ($\alpha = 1$). In the invariant case, as the sample size increased, the minimum $\overline{\text{AB}}$ occurred in a sample with lower level of informative effect (0.145 on $\alpha = 1$, $n = 5$; 0.6370 on $\alpha = 2$, $n = 10$; and 0.6858 on $\alpha = 3$, $n = 15$). The $\overline{\text{RMSE}}$ values generated by M1 were smaller than those generated by R1 except in non-invariant case with $n = 15$, the $\overline{\text{RMSE}}$ value produced by M1 was slightly higher by 0.0039($\alpha = 2$) and 0.291($\alpha = 2$) than that was produced by R1.

Table 3 reports the finding in the quadratic case that also indicates the values of $\overline{\text{AB}}$ and $\overline{\text{RMSE}}$ resulted by N1 and R2 for each level of informative effect and sample size taken. N1 produced $\overline{\text{AB}}$

Table 3: The values of $\overline{\text{AB}}$ and $\overline{\text{RMSE}}$ of model N1 and R2 over 20 areas for quadratic case

Performance measure	α	Size measure	Model N1			Model R2		
			$n_i = 5$	$n_i = 10$	$n_i = 15$	$n_i = 5$	$n_i = 10$	$n_i = 15$
$\overline{\text{AB}}$	1	I	0.1989	0.9478	0.9736	1.0859	1.5856	1.1425
		NI	0.2073	0.7667	0.6582	1.2789	0.9780	1.0293
	2	I	0.1856	0.7285	0.9141	1.4075	1.1320	1,0413
		NI	1.1068	0.6720	0.9196	1.4257	1.1535	1.1089
	3	I	0.1805	0.7114	0.8928	1.3245	1.3317	0.9857
		NI	0.6943	0.2539	0.9659	1.0105	1.0548	1.1109
	∞	I	0.1921	0.6795	0.8530	1.4694	1.2659	1.1131
		NI	0.1841	0.7612	0.9910	1.0421	1.5992	1.0670
$\overline{\text{RMSE}}$	1	I	0.3430	1.4895	1.3653	1.3129	1.8246	1.3791
		NI	0.3538	1.1973	0.9487	1.4388	1.1812	1.2182
	2	I	0.3217	1.2232	1.2859	1.5287	1.3658	1.2490
		NI	1.4027	1.1113	1.2944	1.7009	1.3870	1.3392
	3	I	0.3227	1.1057	1.2595	1.6002	1.5465	1.1878
		NI	0.9370	0.3611	1.3614	1.2112	1.2581	1.3253
	∞	I	0.3354	1.0175	1.1938	1.6825	1.5273	1.3409
		NI	0.3267	1.1591	1.3912	1.2641	1.7663	1.2870

$\overline{\text{AB}}$ = average absolute bias; $\overline{\text{RMSE}}$ = average root mean square error; α = the level of informative effect; I = invariant; NI = non invariant; n_i = sample size; N1 = model (3.3) with quadratic p-spline for $g_1(x_{ij})$ and linear p-spline for $g_2(\pi_{ij})$; $R2$ = model (3.1) with quadratic p-spline for $g_1(x_{ij})$.

values that were smaller than those produced under R2 for each sample size and the level of the informative sample. The average bias in the simulated quadratic case can be reduced by adding the linear p-spline function of the inclusion probability into the semiparametric model. The minimum value of $\overline{\text{AB}}$ occurred in samples that were less informative in invariant type sizes as the sample size increased. In invariant type sizes, as the sample size increased, the minimum value of $\overline{\text{AB}}$ occurred in samples that were less informative (0.1805 for $4\alpha = 3$, $n = 5$) and non informative sample (0.6795 for $n = 10$ and 0.8530 for $n = 15$); conversely, in non-invariant cases (0.1841 for $\alpha = \infty$, $n = 5$; 0.253 for $\alpha = 3$, $n = 10$ and 0.6582 for $\alpha = 1$ and $n = 5$). As sample sizes increased, a minimum $\overline{\text{AB}}$ value was obtained in more informative samples (0.1841 for $\alpha = \infty$, $n = 5$; 0.2539 for $\alpha = 3$, $n = 10$ and 0.1841 for $\alpha = 1$, $n = 15$). N1 also yielded a value of $\overline{\text{RMSE}}$ higher than R2 in five simulation cases in Table 3. A $\overline{\text{RMSE}}$ difference of 0.0161 occurred in a sample size of 10 with non-invariant type measure and $\alpha = 1$. Meanwhile, a $\overline{\text{RMSE}}$ difference of 0.0369 ($\alpha = 2$) and 0.0717 ($\alpha = 3$) in the sample of 15 with invariant measure size, and respectively 0.0361 ($\alpha = 3$) and 0.1042 ($\alpha = \infty$) for non invariant cases.

We examine if the order of p-spline approach for $g_2(\pi_{ij})$ in the model will result in different estimators. Graphically, we do this by comparing distribution of RMSE of each p-spline degree, by considering size measure and level of informative sample ($\alpha = 1, 2, 3$). The comparison is done for both population cases. Figure 1 shows a comparison of the distribution of RMSE values generated by M1 and M2 in the case of linear population y_{ij} while Figure 2 indicates for the case of quadratic populations given by N1 and N2.

In the linear population of y_{ij} for an invariant measure case with $n = 5$ and $\alpha = 2$, M1 produced an RMSE distribution with a range and mode that was smaller than that produced by M2. However, the RMSE values did not differ significantly for other sample sizes. This condition is shown in Figure 1(a). In the non-invariant measure case, Figure 1(b) showed the RMSE distribution produced by M1 had a significantly smaller mode and range than M2 with small samples ($n = 5$). However, for $n = 10$ and $n = 15$, functions for $g_2(\pi_{ij})$ did not produce a significantly different RMSE distribution. In the quadratic population, both the invariant and non-invariant cases show the RMSE distribution produced by N1 and N2 did not differ significantly, unless for ($n = 5$) and $\alpha = 2$. We examine if the order of

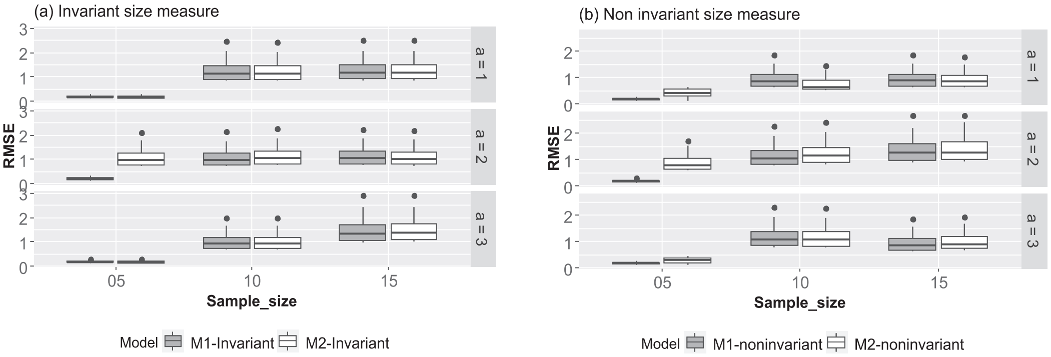

Figure 1: *RMSE distributions produced by model M1 and M2. RMSE = average root mean square error.*

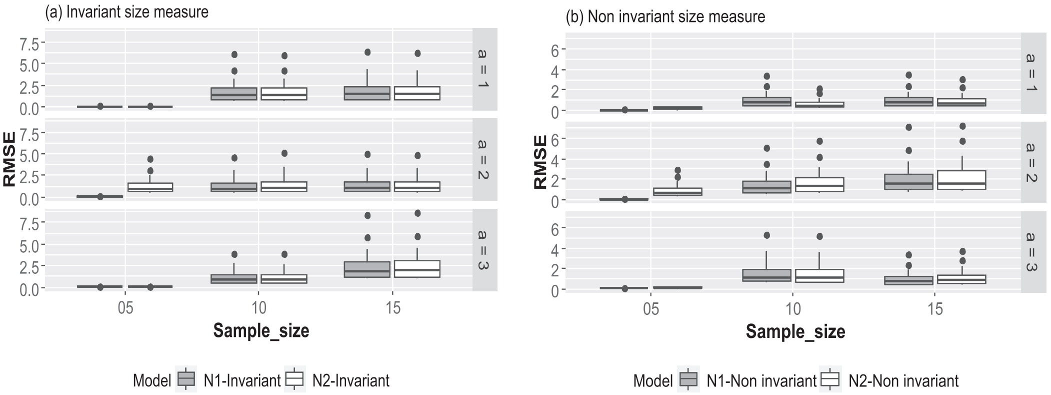

Figure 2: *RMSE distributions produced by model N1 and N2. RMSE = average root mean square error.*

p-spline approach for $g_2(\pi_{ij})$ in the model will result in different estimators. Graphically, we do this by comparing the distribution of RMSE for each p-spline degree, by considering size measure and level of informative sample ($\alpha = 1, 2, 3$). The comparison is done for both population cases.

5. Concluding remarks

Parametric assumptions in statistical models are often restricted in practice; in addition, informative effects of sample must be taken into account in model to reduce bias. We predict the mean of small areas based on a semiparametric mixed-model. We add the inclusion probability function $g(\pi)$ in the model to account for the informative effect. The p-spline applied to approach the function of the covariate variable and the inclusion probability function $g(\pi)$ in the model.

We obtain best unbiased estimators for the model coefficients and REML estimators for the variance components. The simulation results in linear and quadratic population case show that the addition of linear p-spline and quadratic p-spline of inclusion probability into the model can reduce the average absolute bias. By adding linear or quadratic p-spline with variable π we reduce the RMSE average in most cases. We also found that the linear and quadratic p-spline approach for the inclusion probability function in both population case did not provide a significant difference in the RMSE distribution except for the smallest sample sizes with a high degree of informative effect.

The notion of adding the inclusion probability function $g(\pi)$ in a mixed-model has also been studied in Verret *et al.* (2015). They used a nested error regression model and utilized plot of error residuals to determined the form of inclusion probability function $g(\pi)$. In contrast, we use semiparametric mixed-model and utilize p-spline to approach $g(\pi)$. This approach is expected to be an alternative to reduce bias caused by informative sampling. However, like Verret *et al.* (2015), our approach also requires information on inclusion probability for all population units. Limited access to population related to inclusion probability causes constraints in use.

The idea of a p-spline approach as an inclusion probability function in the model can be an alternative approach to reduce bias in small area estimation under informative sampling. In future work, we would like to develop variance components estimators obtained in the recent study as well as obtain a robust prediction to increase the model's performance.

Acknowledgments

Research fund from Directorate of Research and Development, Directorate General of Higher Education, Ministry of Research, Technology and Higher Education, Indonesia.

References

Asparouhov T (2006). General multi-level modeling with sampling weights, *Communication in Statistics, Theory and Methods*, **35**, 439–460.

Battese GE, Harter RM, and Fuller WA (1988). An error-components model for prediction of country crop areas using survey and satellite data, *Journal of the American Statistical Association*, **83**, 28–36.

Burgard JP, Münnich R, and Zimmermann T (2014). The impact of sampling designs on small area estimates for business data, *Journal of Official Statistics*, **30**, 749–771.

Clement EP (2014). Small area estimation with application to disease mapping, *International Journal of Probability and Statistics*, **3**, 15–22.

Hwang J and Kim DH (2015). Bayesian curve-fitting in semiparametric small area models with measurement errors, *Communications for Statistical Applications and Methods*, **22**, 349–359.

Molina I, Nandram B, and Rao JNK (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach, *The Annals of Applied Statistics*, **8**, 852–885.

Opsomer JD, Breidt FJ, Claeskens SG, Kauermann G, and Ranalli MG (2008). Non-parametrik small area estimation using penalized spline regression, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **70**, 265–286.

Patterson HD and Thompson R (1971). Recovery of inter-block information on when block sizes are unequal, *Biometrika*, **58**, 545-554.

Pfeffermann D, Krieger AM and Rinott Y (1998). Parametric distributions of complex survey data under informative probability sampling, *Statistica Sinica*, **8**, 1087–1114.

Pfeffermann D and Sverchkov M (2007). Small-area estimation under informative probability sampling of areas and within the selected areas, *Journal of the American Statistical Association*, **102**, 1427–1439.

Pfeffermann D and Sverchkov M (2009). Inference under informative sampling. In *Sample Survey: Inference and Analysis* (Vol.29B, pp. 455–487), Elsevier, Oxford.

Rao JNK (2003). *Small Area Estimation*, John Wiley & Sons, New Jersey.

Rao JNK, Sinha SK and Dumitrescu L (2014). Robust small area estimation under semi-parametric mixed models, *The Canadian Journal of Statistics*, **42**, 126–141.

Rencher AC and Schaalje GB (2008). *Linear Models in Statistics*, John Wiley & Sons, New Jersey.

Ruppert D, Wand MP, and Carroll RJ (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.

Searle SR (1982). *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.

Searle SR, Casella G, and McCulloch CE (2006). *Variance Components*, John Wiley & Sons, New Jersey.

Tzavidis N, Chambers RL, Salvati N, and Chandra H (2012). Small area estimation in practice an application to agricultural business survey data, *Journal of the Indian Society of Agricultural Statistics*, **66**, 213–228.

Verret F, Rao JNK, and Hidiroglou MA (2015). Model-based small area estimation under informative sampling, *Survey Methodology*, **41**, 333–347.

Received November 16, 2019; Revised February 10, 2020; Accepted March 13, 2020