CIRJE-F-801

Parametric Bootstrap Methods for Bias Correction in Linear Mixed Models

Tatsuya Kubokawa
University of Tokyo

Bui Nagashima
Graduate School of Economics, University of Tokyo

April 2011

Parametric Bootstrap Methods for Bias Correction in Linear Mixed Models

Tatsuya Kubokawa*and Bui Nagashima†
University of Tokyo

April 21, 2011

Abstract

The empirical best linear unbiased predictor (EBLUP) in the linear mixed model (LMM) is useful for the small area estimation, and the estimation of the mean squared error (MSE) of EBLUP is important as a measure of uncertainty of EBLUP. To obtain a second-order unbiased estimator of the MSE, the second-order bias correction has been derived mainly based on Taylor series expansions. However, this approach is harder to implement in complicated models with more unknown parameters like variance components, since we need to compute asymptotic bias, variance and covariance for estimators of unknown parameters as well as partial derivatives of some quantities. The same difficulty occurs in construction of confidence intervals based on EBLUP with second-order correction and in derivation of second-order bias correction terms in the Akaike Information Criterion (AIC) and the conditional AIC. To avoid such difficulty in derivation of second-order bias correction in these problems, the parametric bootstrap methods are suggested in this paper, and their second-order justifications are established. Finally, performances of the suggested procedures are numerically investigated in comparison with some existing procedures given in the literature.

Key words and phrases: Best linear unbiased predictor, confidence interval, empirical Bayes procedure, Fay-Herriot model, second-order correction, linear mixed model, maximum likelihood estimator, mean squared error, nested error regression model, parametric bootstrap, restricted maximum likelihood estimator, small area estimation.

1 Introduction

The linear mixed models (LMM) and the model-based estimates including empirical best linear unbiased predictor (EBLUP) or the empirical Bayes estimator (EB) have been

*Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

†Graduate School in Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: bui-bui.miu-miu@kde.biglobe.ne.jp

recognized useful in small area estimation. The typical models used for the small area estimation are the Fay-Herriot model and the nested error regression model (NERM), and the usefulness of EBLUP is illustrated by Fay and Herriot (1979) and Battese, Harter and Fuller (1988). For a good review and account on this topic, see Ghosh and Rao (1994), Rao (2003) and Pfeffermann (2002).

When EBLUP is used to estimate a small area mean based on real data, it is important to assess how much EBLUP is reliable. One method for the purpose is to estimate the mean squared error (MSE) of EBLUP, and asymptotically unbiased estimators of the MSE with the second-order bias correction have been derived based on the Taylor series expansion by Prasad and Rao (1990), Datta and Lahiri (2000), Datta, Rao and Smith (2005), Das, Jiang and Rao (2004), Kubokawa (2010b) and others. A drawback of this method is that it is harder to compute the second-order bias, variance and covariance of estimators of more unknown parameters including variance components, and that it is troublesome to derive partial derivatives of some matrices with respect to unknown parameters. To avoid this difficulty, Butar and Lahiri (2003) proposed the parametric bootstrap method, which is easy to implement, since we do not need to compute the second-order bias, variance and partial derivatives. For some recent results including nonparametric methods, see Lahiri and Rao (1995), Hall and Maiti (2006a) and Chen and Lahiri (2008).

We have other problems to be faced with the same difficulty as in the MSE estimation. One is the problem of constructing a confidence interval based on EBLUP such that it satisfies the nominal confidence level with the second-order accuracy. Datta, Ghosh, Smith and Lahiri (2002), Basu, Ghosh and Mukerjee (2003) and Kubokawa (2010a) derived such confidence intervals using the Taylor series expansion. To avoid the difficulty in derivation of second-order moments, Hall and Maiti (2006b) and Chatterjee, Lahiri and Li (2008) proposed the confidence intervals using the parametric bootstrap method.

A similar difficulty occurs in evaluating the bias terms of AIC and conditional AIC. The Akaike Information Criterion (AIC) originated from Akaike (1973, 74) is recognized very useful for selecting models in general situations, and it is also useful for selecting variables in LMM. When unknown parameters in the model are estimated by the maximum likelihood estimator, the penalty term, which is a kind of bias, is known to be $2 \times p$ for dimension p of unknown parameters. When the unknown parameters in LMM are estimated by other estimators, however, Kubokawa (2011) showed that the penalty term includes partial derivatives of the estimator and the covariance matrix. Concerning the conditional AIC, on the other hand, Vaida and Blanchard (2005) and Liang, Wu and Zou (2008) proposed the conditional AIC in LMM, but their derivations were limited to the cases that the parameters in LMM are partly known. Recently, Kubokawa (2011) derived the second-order bias correction term for the conditional AIC, but it is harder to compute in more complicated models since the penalty term consists of the second-order bias, variance and covariance of estimators and partial derivatives of the covariance matrix.

In this paper, we treat the problems mentioned above, and provide useful procedures based on the parametric bootstrap method to avoid the computational difficulties. In Section 2, we suggest the MSE estimator, the confidence interval, AIC and the conditional AIC using the parametric bootstrap. Concerning the MSE estimation, Butar and Lahiri

(2003) estimated the third term of the MSE, denoted by g_3, based on the parametric bootstrap, while in this paper, we consider to estimate the second-order approximation of g_3 using the parametric bootstrap method. A similar approach applies to the confidence interval, and we estimate the second-order correction term based on the parametric bootstrap method. This is different from the parametric bootstrap procedure suggested by Chatterjee, *et al*. (2008) who obtained two end-points of a confidence interval based on a distribution generated by the parametric bootstrap sampling. In Section 3, we give the proofs for the second-order justifications of the procedures given in the paper under several conditions.

In Section 4, we carry out simulation experiments in the Fay-Herriot model to compare the proposed procedures with ones given in the literature. In Section 5, the nested error regression models (NERM) are treated to analyze the posted land price data, and the suggested information criteria are used for selecting regressors. For estimating the averages of land prices in small areas, we give values of EBLUPs, proposed estimates of their MSE and proposed confidence intervals based on the NERM with the selected regressors. These numerical investigations demonstrate that the proposed procedures based on the parametric bootstrap methods work well and are useful.

2 MSE Estimation, Confidence Interval and AIC Based on the Parametric Bootstrap Method

2.1 Linear mixed model and parametric bootstrap method

Consider the following two linear mixed models: One is a model which original data follow, and the other is a model which generates simulated data through the parametric bootstrap method.

[1] Model 1. An $N \times 1$ observation vector $\boldsymbol{y}$ of the response variable has the general linear mixed model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are $N \times p$ and $N \times M$ matrices, respectively, of the explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown vector of the regression coefficients, $\boldsymbol{v}$ is an $M \times 1$ vector of the random effects, and $\boldsymbol{\epsilon}$ is an $N \times 1$ vector of the random errors. Here, $\boldsymbol{v}$ and $\boldsymbol{\epsilon}$ are mutually independently distributed as $\boldsymbol{v} \sim \mathcal{N}_M(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta}))$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{\theta}))$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ is a q-dimensional vector of unknown parameters, and $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\theta})$ and $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\theta})$ are positive definite matrices. Then, $\boldsymbol{y}$ has a marginal distribution $\mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ for

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{R}(\boldsymbol{\theta}) + \boldsymbol{Z}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}'.$$

Throughout the paper, for simplicity, it is assumed that $\boldsymbol{X}$ is of full rank. Also, we often drop $(\boldsymbol{\theta})$ in $\boldsymbol{G}(\boldsymbol{\theta})$, $\boldsymbol{R}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and others for notational convenience.

The unknown parameters in Model 1 are $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. When $\boldsymbol{\theta}$ is known, the regression coefficients vector $\boldsymbol{\beta}$ is estimated by the generalized least squares estimator given by

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}.$$

The parameter $\boldsymbol{\theta}$ consists of variance components and others, and it is estimated by consistent estimator $\widehat{\boldsymbol{\theta}}$ based on $\boldsymbol{y}$ through various methods including maximum likelihood and restricted maximum likelihood methods. Then, $\boldsymbol{\beta}$ is estimated by $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$.

[2] Model 2. An $N \times 1$ random vector $\boldsymbol{y}^*$ given $\boldsymbol{y}$ has the linear mixed model

$$\boldsymbol{y}^* = \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{Z}\boldsymbol{v}^* + \boldsymbol{\epsilon}^*, \tag{2.2}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the same matrices as given in (2.1), and given $\boldsymbol{y}$, $\boldsymbol{v}^*$ and $\boldsymbol{\epsilon}^*$ are conditionally mutually independently distributed as $\boldsymbol{v}^*|\boldsymbol{y} \sim \mathcal{N}_M(\boldsymbol{0}, \boldsymbol{G}(\widehat{\boldsymbol{\theta}}))$ and $\boldsymbol{\epsilon}^*|\boldsymbol{y} \sim \mathcal{N}_N(\boldsymbol{0}, \boldsymbol{R}(\widehat{\boldsymbol{\theta}}))$.

Before stating the main results, it may be instructive to explain the intuitive idea of using the parametric bootstrap method based on Model 2. Let $g(\boldsymbol{\theta})$ be a differentiable function with $g(\boldsymbol{\theta}) = O(1)$. Although $g(\widehat{\boldsymbol{\theta}})$ is an asymptotically unbiased estimator of $g(\boldsymbol{\theta})$, in general, there exists a second-order bias. Then, we need to approximate the expectation $E[g(\widehat{\boldsymbol{\theta}})]$ up to $O(N^{-1})$. It is supposed that the approximation is given by $E[g(\widehat{\boldsymbol{\theta}})] = g(\boldsymbol{\theta}) + b_g(\boldsymbol{\theta}) + O(N^{-3/2})$ where $b_g(\boldsymbol{\theta})$ is a continuously differentiable function with $O(N^{-1})$. Thus, we can get the second-order unbiased estimator $g(\widehat{\boldsymbol{\theta}}) - b_g(\widehat{\boldsymbol{\theta}})$, namely,

$$E[g(\widehat{\boldsymbol{\theta}}) - b_g(\widehat{\boldsymbol{\theta}})] = \{g(\boldsymbol{\theta}) + b_g(\boldsymbol{\theta})\} - b_g(\boldsymbol{\theta}) + O(N^{-3/2}) = g(\boldsymbol{\theta}) + O(N^{-3/2}). \tag{2.3}$$

Since, in general, the function $b_g(\boldsymbol{\theta})$ can be derived using the Taylor series expansion, it is based on partial derivatives with respect to θ_i, $i = 1, \ldots, q$, and moments of estimator $\widehat{\boldsymbol{\theta}}$, and we need to derive partial derivatives and moments for each model and each estimator $\widehat{\boldsymbol{\theta}}$. This means that different calculations are requested for different models and different estimators, so that those calculations must be harder for more complicated models. Instead of this method, Butar and Lahiri (2003) and Chatterjee *et al*. (2008) proposed an alternative approach through the parametric bootstrap method. The function $b_g(\boldsymbol{\theta})$ is estimated by $E_*[g(\widehat{\boldsymbol{\theta}}^*) - g(\widehat{\boldsymbol{\theta}})|\boldsymbol{y}]$, where $E_*[\cdot|\boldsymbol{y}]$ is the expectation with respect to Model 2 given $\boldsymbol{y}$, and the calculation of $\widehat{\boldsymbol{\theta}}^*$ is the same as that of $\widehat{\boldsymbol{\theta}}$ except that $\widehat{\boldsymbol{\theta}}^*$ is calculated based on $\boldsymbol{y}^*$ instead of $\boldsymbol{y}$. Since given $\boldsymbol{y}$, the expectation $E_*[g(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}]$ can be approximated as

$$E_*[g(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}] = g(\widehat{\boldsymbol{\theta}}) + b_g(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}),$$

it is seen from (2.3) that

$$\begin{aligned} E\Big[2g(\widehat{\boldsymbol{\theta}}) - E_*[g(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}]\Big] =& E\Big[g(\widehat{\boldsymbol{\theta}}) - E_*[g(\widehat{\boldsymbol{\theta}}^*) - g(\widehat{\boldsymbol{\theta}})|\boldsymbol{y}]\Big] \\ =& E[g(\widehat{\boldsymbol{\theta}}) - b_g(\widehat{\boldsymbol{\theta}})] + O(N^{-3/2}) \\ =& g(\boldsymbol{\theta}) + O(N^{-3/2}). \end{aligned} \tag{2.4}$$

Thus, we obtain the second-order unbiased estimator $2g(\widehat{\boldsymbol{\theta}}) - E_*[g(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}]$ which is free from differentiations or moments of $\widehat{\boldsymbol{\theta}}$. This idea can be used in this paper to provide an estimator of MSE of EBLUP, a corrected confidence interval based on EBLUP and conditional and unconditional Akaike Information Criteria.

2.2 Estimation of MSE of EBLUP

Based on the parametric bootstrap method, we first derive an estimator of MSE of EBLUP for the general scalar quantity

$$\mu = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{v},$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ be $p \times 1$ and $M \times 1$ vectors of fixed constants. It is noted that the marginal and the conditional distributions of $\boldsymbol{y}$ given $\boldsymbol{v}$ are, respectively,

$$\begin{aligned}\boldsymbol{y} \sim & \mathcal{N}_N(\boldsymbol{X\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta})),\\ \boldsymbol{y}|\boldsymbol{v} \sim & \mathcal{N}_N(\boldsymbol{X\beta} + \boldsymbol{Zv}, \boldsymbol{R}(\boldsymbol{\theta})).\end{aligned}$$

Let $\boldsymbol{\mu}_v = \boldsymbol{GZ}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$ and $\boldsymbol{\Sigma}_v = \boldsymbol{G} - \boldsymbol{GZ}'\boldsymbol{\Sigma}^{-1}\boldsymbol{ZG}$. Since the conditional distribution of $\boldsymbol{v}$ given $\boldsymbol{y}$ is

$$\boldsymbol{v}|\boldsymbol{y} \sim \mathcal{N}_M(\boldsymbol{\mu}_v, \boldsymbol{\Sigma}_v), \tag{2.5}$$

the conditional expectation $E[\mu|\boldsymbol{y}]$ is written as

$$\begin{aligned}\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) =& E[\mu|\boldsymbol{y}] = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\\ =& \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X\beta}),\end{aligned} \tag{2.6}$$

where $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{ZG}(\boldsymbol{\theta})\boldsymbol{b}$. This can be interpreted as the Bayes estimator of μ in the Bayesian context. The generalized least squares estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ is substituted into $\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta})$ to get the estimator

$$\widehat{\mu}^{EB}(\boldsymbol{\theta}) = \widehat{\mu}^B(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{a}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})), \tag{2.7}$$

which is the best linear unbiased predictor (BLUP) of μ. When an estimator $\widehat{\boldsymbol{\theta}}$ is available for $\boldsymbol{\theta}$, we can estimate μ by the empirical (or estimated) best linear unbiased predictor (EBLUP) $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$, which is also called an empirical Bayes estimator in the Bayesian context.

The MSE function of EBLUP $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$ is $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \mu\}^2]$, which can be decomposed as $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3(\boldsymbol{\theta})$ as shown in Prasad and Rao (1990) and Datta and Lahiri (2000), where $g_1(\boldsymbol{\theta}) = E[\{\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) - \mu\}^2]$, $g_2(\boldsymbol{\theta}) = E[\{\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta})\}^2]$ and $g_3(\boldsymbol{\theta}) = E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2]$. The terms $g_1(\boldsymbol{\theta})$ and $g_2(\boldsymbol{\theta})$ can be rewritten as

$$\begin{aligned}g_1(\boldsymbol{\theta}) =& \boldsymbol{b}'(\boldsymbol{G}(\boldsymbol{\theta})^{-1} + \boldsymbol{Z}'\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{Z})^{-1}\boldsymbol{b},\\ g_2(\boldsymbol{\theta}) =& (\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta}))'(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}(\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta})).\end{aligned}$$

Using the argument as in (2.4), we can estimate $g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta})$ by

$$2\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\} - E_*[g_1(\widehat{\boldsymbol{\theta}}^*) + g_2(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}].$$

For $g_3(\boldsymbol{\theta})$, in this paper we use the estimator given by

$$\overline{g}_3{}^*(\widehat{\boldsymbol{\theta}}) = E_*\big[\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}\big|\boldsymbol{y}\big]. \tag{2.8}$$

Thus, we get the estimator

$$mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = 2\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\} - E_*\big[g_1(\widehat{\boldsymbol{\theta}}^*) + g_2(\widehat{\boldsymbol{\theta}}^*)\big|\boldsymbol{y}\big] + \overline{g}_3{}^*(\widehat{\boldsymbol{\theta}}). \tag{2.9}$$

In Section 3.2, we shall show that $E[mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] = MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O(N^{-3/2})$ under some conditions.

2.3 Confidence interval based on EBLUP

We next construct a confidence interval of $\mu = \boldsymbol{a}'\boldsymbol{\beta}+\boldsymbol{b}'\boldsymbol{v}$ based on EBLUP which satisfies the nominal confidence level with the second-order accuracy. Since $mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ is an asymptotically unbiased estimator of the MSE of EBLUP, it is reasonable to consider the confidence interval of the form

$$I^{EB}(\widehat{\boldsymbol{\theta}}):\ \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) \pm z_{\alpha/2}\sqrt{\max\{mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})), 0\}}, \tag{2.10}$$

where $z_{\alpha/2}$ is the $100\times\alpha/2\%$ upper quantile of the standard normal distribution. However, the coverage probability $P[\mu \in I^{EB}(\widehat{\boldsymbol{\theta}})]$ cannot be guaranteed to be greater than or equal to the nominal confidence coefficient $1-\alpha$. To address the problem, we consider the correction function given by

$$h_1^*(\widehat{\boldsymbol{\theta}}) = \frac{1+z_{\alpha/2}^2}{8g_1(\widehat{\boldsymbol{\theta}})^2}E_*\left[\{g_1(\widehat{\boldsymbol{\theta}}^*) - g_1(\widehat{\boldsymbol{\theta}})\}^2 \big| \boldsymbol{y}\right], \tag{2.11}$$

which is an asymptotically unbiased estimator of $h_1(\boldsymbol{\theta})$ given in (3.14). Then, the corrected confidence interval is provided by

$$I_1^{CEB*}(\widehat{\boldsymbol{\theta}}):\ \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) \pm z_{\alpha/2}\left[1+h_1^*(\widehat{\boldsymbol{\theta}})\right]\sqrt{\max\{mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})), 0\}}. \tag{2.12}$$

A drawback of $I_1^{CEB*}(\widehat{\boldsymbol{\theta}})$ is that it cannot give an interval when $mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ takes a negative value. A simulation experiment given in Section 4.2 shows that such a shortcoming occurs in an extreme case. Thus, we suggest the alternative corrected confidence interval

$$I_2^{CEB*}(\widehat{\boldsymbol{\theta}}):\ \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) \pm z_{\alpha/2}\left[1+h_2^*(\widehat{\boldsymbol{\theta}})\right]\sqrt{g_1(\widehat{\boldsymbol{\theta}})+g_2(\widehat{\boldsymbol{\theta}})}, \tag{2.13}$$

where

$$h_2^*(\widehat{\boldsymbol{\theta}}) = h_1^*(\widehat{\boldsymbol{\theta}}) + \frac{g_1(\widehat{\boldsymbol{\theta}}) - E_*[g_1(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}] + \overline{g}_3{}^*(\widehat{\boldsymbol{\theta}})}{2g_1(\widehat{\boldsymbol{\theta}})}. \tag{2.14}$$

In Section 3.2, it can be shown that $P[\mu \in I_i^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1-\alpha+O(N^{-3/2})$ for $i=1,2$ under some conditions.

2.4 AIC and conditional AIC

In this section, we derive the Akaike Information Criterion (AIC) and conditional AIC using the parametric bootstrap method.

[1] AIC_1^* and AIC_2^*. Let us define the *Akaike Information* (AI) by

$$AI(\boldsymbol{\theta}) = -2\iint \{\log f_m(\widetilde{\boldsymbol{y}}|\widehat{\boldsymbol{\beta}}(\boldsymbol{y}), \widehat{\boldsymbol{\theta}}(\boldsymbol{y}))\} f_m(\widetilde{\boldsymbol{y}}|\boldsymbol{\beta},\boldsymbol{\theta}) f_m(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\theta}) \mathrm{d}\widetilde{\boldsymbol{y}}\mathrm{d}\boldsymbol{y}, \tag{2.15}$$

where $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}) = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ are estimators based on $\boldsymbol{y}$, and $f_m(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\theta})$ is a marginal density function of $\boldsymbol{y}$ given by

$$-2\log f_m(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\theta}) = N\log(2\pi) + \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}).$$

Akaike's AIC can be derived as an asymptotically unbiased estimator of $AI(\boldsymbol{\theta})$, namely, $E[AIC] = AI(\boldsymbol{\theta}) + o(1)$ as $N \to \infty$. When AIC is an exact unbiased estimator of $AI(\boldsymbol{\theta})$, it is called the exact AIC, which was suggested by Sugiura (1978), but in general, it is difficult to get in LMM.

Define $\Delta_1^*(\boldsymbol{y})$ and $\Delta_2^*(\boldsymbol{y})$ by

$$\begin{aligned}\Delta_1^*(\boldsymbol{y}) =& -2E_*\big[\boldsymbol{u}^{*\prime}\boldsymbol{P}(\widehat{\boldsymbol{\theta}}^*)\boldsymbol{u}^*\big|\boldsymbol{y}\big],\\ \Delta_2^*(\boldsymbol{y}) =& E_*\big[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{*-1}\boldsymbol{u}^* - \mathrm{tr}\,[\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}^{*-1}]\big|\boldsymbol{y}\big],\end{aligned} \tag{2.16}$$

where $\boldsymbol{u}^* = \boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}}^* = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^*)$ and

$$\boldsymbol{P}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}.$$

Then, we suggest two kinds of AIC given by

$$\begin{aligned}AIC_1^* =& -2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{\theta}}) + 2p - \Delta_2^*(\boldsymbol{y}),\\ AIC_2^* =& -2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}),\widehat{\boldsymbol{\theta}}) - \Delta_1^*(\boldsymbol{y}) - \Delta_2^*(\boldsymbol{y}),\end{aligned} \tag{2.17}$$

In Section 3.3, it can be shown that $E[AIC_i^*] = AI(\boldsymbol{\theta}) + O(N^{-1/2})$ for $i = 1, 2$.

[2] cAIC*. AIC is derived from the marginal (or unconditional) distribution of $\boldsymbol{y}$, and it measures the prediction error of the predictor based on the marginal distribution. This means that AIC is not appropriate for the focus on the prediction of specific areas or specific random effects. For example, EBLUP is used for predicting the random effects associated with a specific area in the context of the small area estimation. Taking this point into account, Vaida and Blanchard (2005) proposed the conditional AIC which measures the prediction error of the predictor incorporating EBLUP based on the conditional distribution given the random effects.

The conditional AIC is derived as an (asymptotically) unbiased estimator of the *conditional Akaike information* (cAI) defined by

$$cAI(\boldsymbol{\theta}) = -2\iiint \log\{f(\widetilde{\boldsymbol{y}}|\widehat{\boldsymbol{v}}(\boldsymbol{y}),\widehat{\boldsymbol{\beta}}(\boldsymbol{y}),\widehat{\boldsymbol{\theta}}(\boldsymbol{y}))\}f(\widetilde{\boldsymbol{y}}|\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{\theta})f(\boldsymbol{y}|\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{\theta})f(\boldsymbol{v}|\boldsymbol{\theta})\mathrm{d}\widetilde{\boldsymbol{y}}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{v},$$

where $\widehat{\boldsymbol{\beta}}(\boldsymbol{y}) = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{v}}(\boldsymbol{y}) = \widehat{\boldsymbol{v}}(\boldsymbol{\theta}) = \boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\{\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})\}$ are estimators based on $\boldsymbol{y}$ and $f(\boldsymbol{v}|\boldsymbol{\theta})$, respectively, are a conditional density function of $\boldsymbol{y}$ given $\boldsymbol{v}$ and a marginal density function of $\boldsymbol{v}$. Note that

$$\begin{aligned}-2\log f(\boldsymbol{y}|\boldsymbol{v},\boldsymbol{\beta},\boldsymbol{\theta}) =& N\log(2\pi) + \log|\boldsymbol{R}(\boldsymbol{\theta})|\\ &+ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v})'\boldsymbol{R}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{v}).\end{aligned} \tag{2.18}$$

Define $\Delta_{c1}^*(\boldsymbol{y})$ and $\Delta_{c2}^*(\boldsymbol{y})$ by

$$\begin{aligned}\Delta_{c1}^*(\boldsymbol{y}) =& -2\Big\{E_*\big[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\boldsymbol{P}(\widehat{\boldsymbol{\theta}}^*)\boldsymbol{u}^* + \mathrm{tr}\,[\widehat{\boldsymbol{R}}^*\widehat{\boldsymbol{\Sigma}}^{*-1}]\,\big|\boldsymbol{y}\big] + N - 2\mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}]\Big\},\\ \Delta_{c2}^*(\boldsymbol{y}) =& E_*\Big[\boldsymbol{u}^{*\prime}(2\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{*-1} - \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{R}}^{*-1}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1})\boldsymbol{u}^*\\ &- \mathrm{tr}\,[\widehat{\boldsymbol{R}}^{*-1}(2\widehat{\boldsymbol{R}} - \widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}})]\,\big|\boldsymbol{y}\Big] + 2N - 2\mathrm{tr}\,[\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}],\end{aligned} \tag{2.19}$$

for $\widehat{\boldsymbol{R}} = \boldsymbol{R}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{R}}^* = \boldsymbol{R}(\widehat{\boldsymbol{\theta}}^*)$. Then, we propose the conditional AIC given by

$$cAIC^* = -2\log f(\boldsymbol{y}|\widehat{\boldsymbol{v}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\theta}}) - \Delta^*_{c1}(\boldsymbol{y}) - \Delta^*_{c2}(\boldsymbol{y}). \tag{2.20}$$

In Section 3.3, it can be shown that $E[cAIC^*] = cAI(\boldsymbol{\theta}) + O(N^{-1/2})$ under some conditions.

3 Proofs of Second-order Approximations

3.1 Notations and common assumptions

In this section, we verify that the procedures proposed in the previous section have second-order approximations. We begin by introducing the notations used here. Let $\mathcal{C}^{[k]}_{\boldsymbol{\theta}}$ denote a set of k times continuously differentiable functions with respect to $\boldsymbol{\theta}$. For partial derivatives with respect to $\boldsymbol{\theta}$, we utilize the notations

$$\boldsymbol{A}_{(i)}(\boldsymbol{\theta}) = \partial_i \boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i}, \quad \boldsymbol{A}_{(ij)}(\boldsymbol{\theta}) = \partial_{ij} \boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial^2 \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j},$$

and

$$\boldsymbol{A}_{(ijk)}(\boldsymbol{\theta}) = \partial_{ijk} \boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial^3 \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k},$$

where $\boldsymbol{A}(\boldsymbol{\theta})$ is a scalar, vector or matrix. For the first and second differential operators with respect to $\boldsymbol{y}$, we use the notations

$$\boldsymbol{\nabla}_{\boldsymbol{y}} = \frac{\partial}{\partial \boldsymbol{y}'}, \quad \boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\nabla}'_{\boldsymbol{y}} = \frac{\partial}{\partial \boldsymbol{y}} \frac{\partial}{\partial \boldsymbol{y}'},$$

namely, the i-th element of $\boldsymbol{\nabla}_{\boldsymbol{y}}$ and the (i,j)-th element of $\boldsymbol{\nabla}_{\boldsymbol{y}} \boldsymbol{\nabla}'_{\boldsymbol{y}}$ are $\partial/\partial y_i$ and $\partial^2/\partial y_i \partial y_j$, respectively.

For $0 \leq i, j, k \leq q$, let $\lambda_1(\boldsymbol{\Sigma}) \leq \cdots \leq \lambda_N(\boldsymbol{\Sigma})$ be the eigenvalues of $\boldsymbol{\Sigma}$ and let those of $\boldsymbol{\Sigma}_{(i)}$, $\boldsymbol{\Sigma}_{(ij)}$ and $\boldsymbol{\Sigma}_{(ijk)}$ be $\lambda^i_a(\boldsymbol{\Sigma})$, $\lambda^{ij}_a(\boldsymbol{\Sigma})$ and $\lambda^{ijk}_a(\boldsymbol{\Sigma})$ for $a = 1, \ldots, N$ respectively, where $|\lambda^i_1(\boldsymbol{\Sigma})| \leq \cdots \leq |\lambda^i_N(\boldsymbol{\Sigma})|$, $|\lambda^{ij}_1(\boldsymbol{\Sigma})| \leq \cdots \leq |\lambda^{ij}_N(\boldsymbol{\Sigma})|$ and $|\lambda^{ijk}_1(\boldsymbol{\Sigma})| \leq \cdots \leq |\lambda^{ijk}_N(\boldsymbol{\Sigma})|$.

Throughout the paper, assume the following conditions for large N and $0 \leq i, j, k \leq q$:

(A1) The elements of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{G}(\boldsymbol{\theta})$, $\boldsymbol{R}(\boldsymbol{\theta})$, p, q and M are bounded, and $\boldsymbol{X}'\boldsymbol{X}$ is positive definite and $\boldsymbol{X}'\boldsymbol{X}/N$ converges to a positive definite matrix;

(A2) $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathcal{C}^{[3]}_{\boldsymbol{\theta}}$, $\lim_{N\to\infty} \lambda_1(\boldsymbol{\Sigma}) > 0$, $\lim_{N\to\infty} \lambda_N(\boldsymbol{\Sigma}) < \infty$, $\lim_{N\to\infty} |\lambda^i_N(\boldsymbol{\Sigma})| < \infty$, $\lim_{N\to\infty} |\lambda^{ij}_N(\boldsymbol{\Sigma})| < \infty$ and $\lim_{N\to\infty} |\lambda^{ijk}_N(\boldsymbol{\Sigma})| < \infty$.

(A3) $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ is an estimator of $\boldsymbol{\theta}$ which satisfies that $\widehat{\boldsymbol{\theta}}(-\boldsymbol{y}) = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ and $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{\alpha}) = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ for any p-dimensional vector $\boldsymbol{\alpha}$.

(A4) It is assumed that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is expanded as

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{\dagger} + \widehat{\boldsymbol{\theta}}^{\dagger\dagger} + O_p(N^{-3/2}), \tag{3.1}$$

where $\widehat{\boldsymbol{\theta}}^{\dagger} = O_p(N^{-1/2})$, $\widehat{\boldsymbol{\theta}}^{\dagger\dagger} = O_p(N^{-1})$ and $E[\widehat{\boldsymbol{\theta}}^{\dagger}] = \boldsymbol{0}$.

Example 3.1 As well known, the ML estimator $\widehat{\boldsymbol{\theta}}^M$ and the REML estimator $\widehat{\boldsymbol{\theta}}^R$, respectively, are given as solutions of the following equations:

$$\begin{aligned} &[\text{ML}] \qquad \boldsymbol{y}'\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\theta})\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{y} = \text{tr}\,\left[\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\theta})\right], \\ &[\text{REML}] \qquad \boldsymbol{y}'\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\theta})\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{y} = \text{tr}\,\left[\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\theta})\right], \end{aligned} \tag{3.2}$$

for $i = 1, \ldots, q$, where $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}\left\{\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}\right\}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$. For the details, see Searle, Casella and McCulloch (1992). It is noted that both estimators are location invariant, namely, $\widehat{\boldsymbol{\theta}}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{\alpha}) = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ for any p-dimensional vector $\boldsymbol{\alpha}$. Thus, the condition (A3) is satisfied.

Let $\boldsymbol{A}_2$ be a $q \times q$ matrix such that the (i,j)-th element of $\boldsymbol{A}_2$ is $\text{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}]$. Assume that $\boldsymbol{A}_2$ is positive definite and $\boldsymbol{A}_2/N$ converges to a positive definite matrix. Under this condition, (A1) and (A2), Kubokawa (2010b) showed that $\widehat{\boldsymbol{\theta}}^M$ and $\widehat{\boldsymbol{\theta}}^R$ satisfies the condition (A4).

Other Estimators are also used in some specific models, and Kubokawa (2010b) showed that the Prasad-Rao estimator and the Fay-Herriot estimator satisfy the conditions (A1)-(A4) in the Fay-Herriot model and the nested error regression model.

3.2 MSE estimation and interval estimation

[1] Second-order unbiasedness of the MSE estimator. We first treat the estimation of MSE of the EBLUP $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$, where $\widehat{\mu}^{EB}(\boldsymbol{\theta}) = \widehat{\mu}^{B}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}),\boldsymbol{\theta}) = \boldsymbol{a}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}))$ for $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{Z}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{b}$, and show that the MSE estimator $mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ given in (2.9) has the second-order unbiasedness.

Let $\overline{g}_3(\boldsymbol{\theta})$ and $\overline{g}_4(\boldsymbol{\theta})$ be functions defined by

$$\begin{aligned} \overline{g}_3(\boldsymbol{\theta}) =& \text{tr}\,\Big[\big(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\big)\boldsymbol{\Sigma}(\boldsymbol{\theta})\big(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{\dagger})\Big], \\ \overline{g}_4(\boldsymbol{\theta}) =& \big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)'E[\widehat{\boldsymbol{\theta}}^{\dagger\dagger}] + \frac{1}{2}\text{tr}\,\left[\boldsymbol{B}(\boldsymbol{\theta})\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{\dagger})\right] - \overline{g}_3(\boldsymbol{\theta}), \end{aligned} \tag{3.3}$$

where $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{\dagger}) = E[(\widehat{\boldsymbol{\theta}}^{\dagger})(\widehat{\boldsymbol{\theta}}^{\dagger})']$, and the (i,j)-th element of $\boldsymbol{B}(\boldsymbol{\theta})$ is given by

$$(\boldsymbol{B}(\boldsymbol{\theta}))_{i,j} = (\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta}))'(\partial_{ij}\boldsymbol{G}(\boldsymbol{\theta}))(\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta})) + \boldsymbol{s}(\boldsymbol{\theta})'(\partial_{ij}\boldsymbol{R}(\boldsymbol{\theta}))\boldsymbol{s}(\boldsymbol{\theta}).$$

It is noted that $\boldsymbol{B}(\boldsymbol{\theta}) = \mathbf{0}$ when $\boldsymbol{G}$ and $\boldsymbol{R}$ are matrices of linear functions of $\boldsymbol{\theta}$.

To establish the second-order approximation, we assume the following conditions:

(B1) The elements of $\boldsymbol{a}$ and $\boldsymbol{b}$ are uniformly bounded, and $\boldsymbol{s}(\boldsymbol{\theta})$ satisfies that $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$, $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}(\boldsymbol{\theta}) = O_p(1)$, $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(i)}(\boldsymbol{\theta}) = O_p(1)$, $(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(ij)}(\boldsymbol{\theta}) = O_p(1)$, $\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{s}_{(j)}(\boldsymbol{\theta}) = O(1)$ and $\boldsymbol{s}_{(j)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\nabla}_y\hat{\theta}_i^{\dagger} = O_p(N^{-1})$.

(B2) For $1 \leq i, j, k \leq q$, (i) $g_1(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[3]}$ and $\partial_i g_1(\boldsymbol{\theta}) = O(1)$, $\partial_{ij} g_1(\boldsymbol{\theta}) = O(1)$ and $\partial_{ijk} g_1(\boldsymbol{\theta}) = O(1)$, (ii) $g_2(\boldsymbol{\theta})$, $\overline{g}_3(\boldsymbol{\theta})$ and $\overline{g}_4(\boldsymbol{\theta})$ are continuously differentiable functions satisfying that $\partial_i g_2(\boldsymbol{\theta}) = O(N^{-1})$, $\partial_i \overline{g}_3(\boldsymbol{\theta}) = O(N^{-1})$ and $\partial_i \overline{g}_4(\boldsymbol{\theta}) = O(N^{-1})$.

Theorem 3.1 *Assume the conditions* (A1)-(A4) *and* (B1)-(B2). *Then,* $mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ *is approximated as*

$$mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O_p(N^{-3/2}), \tag{3.4}$$

where

$$mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}}) + \overline{g}_3(\widehat{\boldsymbol{\theta}}) - \overline{g}_4(\widehat{\boldsymbol{\theta}}). \tag{3.5}$$

Thus,

$$E[mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] = MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O(N^{-3/2}). \tag{3.6}$$

Proof. Kubokawa (2010b) proved in his theorems 2.1 and 2.2 that

$$g_3(\boldsymbol{\theta}) = E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2] = \overline{g}_3(\boldsymbol{\theta}) + O(N^{-3/2}), \tag{3.7}$$

$$E[g_1(\widehat{\boldsymbol{\theta}})] = g_1(\boldsymbol{\theta}) + \overline{g}_4(\boldsymbol{\theta}) + O(N^{-3/2}), \tag{3.8}$$

which can be established under the conditions (A1)-(A4), (B1) and (B2)(i). Then from (3.8), it follows that

$$E_*\big[g_1(\widehat{\boldsymbol{\theta}}^*)\big|\boldsymbol{y}\big] = g_1(\widehat{\boldsymbol{\theta}}) + \overline{g}_4(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}). \tag{3.9}$$

Since $g_2(\boldsymbol{\theta}) = O(N^{-1})$, it is seen that under condition (B2)(ii),

$$E_*\big[g_2(\widehat{\boldsymbol{\theta}}^*)\big|\boldsymbol{y}\big] = g_2(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}). \tag{3.10}$$

Hence from (3.9) and (3.10), it is seen that under condition (B2),

$$\begin{aligned} & 2\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\} - E_*\big[g_1(\widehat{\boldsymbol{\theta}}^*) + g_2(\widehat{\boldsymbol{\theta}}^*)\big|\boldsymbol{y}\big] \\ =& g_1(\widehat{\boldsymbol{\theta}}) - \overline{g}_4(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}). \end{aligned} \tag{3.11}$$

Since $\boldsymbol{s}(\boldsymbol{\theta}) = O(1)$, it is noted that

$$\begin{aligned} \boldsymbol{s}(\widehat{\boldsymbol{\theta}}) =& \boldsymbol{s}(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(N^{-1}) \\ =& \boldsymbol{s}(\boldsymbol{\theta}) + \frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\widehat{\boldsymbol{\theta}}^\dagger + O_p(N^{-1}), \end{aligned}$$

which implies that

$$E\big[\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{s}(\boldsymbol{\theta})\}'\boldsymbol{\Sigma}(\boldsymbol{\theta})\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}) - \boldsymbol{s}(\boldsymbol{\theta})\}\big] = \overline{g}_3(\boldsymbol{\theta}) + O(N^{-3/2}). \tag{3.12}$$

Thus, we get

$$E_*\big[\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}\big|\boldsymbol{y}\big] = \overline{g}_3(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}), \tag{3.13}$$

Combining (3.11) and (3.13) gives

$$\begin{aligned} mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) =& 2\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\} - E_*\big[g_1(\widehat{\boldsymbol{\theta}}^*) + g_2(\widehat{\boldsymbol{\theta}}^*)\big|\boldsymbol{y}\big] \\ & + E_*\big[\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\{\boldsymbol{s}(\widehat{\boldsymbol{\theta}}^*) - \boldsymbol{s}(\widehat{\boldsymbol{\theta}})\}\big|\boldsymbol{y}\big] \\ =& \big\{g_1(\widehat{\boldsymbol{\theta}}) - \overline{g}_4(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\big\} + \overline{g}_3(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2}) \\ =& mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O_p(N^{-3/2}). \end{aligned}$$

Finally, it follows from Kubokawa (2010b) that

$$
\begin{aligned}
E[mse^*(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] =& E[mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] + O(N^{-3/2}) \\
=& MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O(N^{-3/2}),
\end{aligned}
$$

which proves (3.6). ■

[2] Second-order corrected confidence interval. We next treat the interval estimation of μ with the corrected confidence intervals $I_1^{CEB*}(\widehat{\boldsymbol{\theta}})$ and $I_2^{CEB*}(\widehat{\boldsymbol{\theta}})$ given in (2.12) and (2.13), and show that the coverage probability can be approximated to the confidence coefficient in the second-order accuracy.

For the purpose, let us define function $h_1(\boldsymbol{\theta})$ by

$$
h_1(\boldsymbol{\theta}) = \frac{z_{\alpha/2}^2 + 1}{8 g_1(\boldsymbol{\theta})^2} \mathrm{tr}\Big[\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)' \mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\Big]. \tag{3.14}
$$

Assume the following condition:

(B3) $h_1(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[1]}$ and $\partial_i h_1(\boldsymbol{\theta}) = O(N^{-1})$ for $1 \le i \le q$.

Theorem 3.2 *Assume the conditions* (A1)-(A4) *and* (B1)-(B3)*. Then,*

$$
P[\mu \in I_1^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + O(N^{-3/2}). \tag{3.15}
$$

Proof. Let us define U and V^* by

$$
\begin{aligned}
U =& \{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{B}(\boldsymbol{\beta}, \boldsymbol{\theta})\}/\sqrt{g_1(\boldsymbol{\theta})}, \\
V^* =& \Big\{[1 + h_1^*(\widehat{\boldsymbol{\theta}})]\sqrt{mse^*(\widehat{\boldsymbol{\theta}})} - \sqrt{g_1(\boldsymbol{\theta})}\Big\}/\sqrt{g_1(\boldsymbol{\theta})},
\end{aligned}
$$

where $\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$. Using the same arguments as in the proof of theorem 2.3 of Kubokawa (2010b), we can verify that

$$
P[\mu \in I_1^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + \phi(z_\alpha) H_1^*(\boldsymbol{\theta}) + O(N^{-3/2}), \tag{3.16}
$$

where

$$
\begin{aligned}
H_1^*(\boldsymbol{\theta}) =& z_{\alpha/2} E\Big[2V^* - \big\{U^2 + z_{\alpha/2}^2 (V^*)^2\big\}\Big] \\
=& z_{\alpha/2} E\Big[\big\{2V^* + (V^*)^2\big\} - U^2 - (1 + z_{\alpha/2}^2)(V^*)^2\Big].
\end{aligned} \tag{3.17}
$$

It is noted that $mse^*(\widehat{\boldsymbol{\theta}}) = mse(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2})$ and $h_1^*(\widehat{\boldsymbol{\theta}}) = h_1(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2})$, where $mse(\widehat{\boldsymbol{\theta}})$ is given in (3.5). Then, it can be seen that

$$
\begin{aligned}
2V^* + (V^*)^2 =& (2 + V^*)V^* = \{1 + h_1^*(\widehat{\boldsymbol{\theta}})\}^2 \frac{mse^*(\widehat{\boldsymbol{\theta}})}{g_1(\boldsymbol{\theta})} - 1 \\
=& \{1 + h_1(\widehat{\boldsymbol{\theta}})\}^2 \frac{mse(\widehat{\boldsymbol{\theta}})}{g_1(\boldsymbol{\theta})} - 1 + O_p(N^{-3/2}) \\
=& 2V + V^2 + O_p(N^{-3/2}),
\end{aligned}
$$

where $V = [1+h_1(\widehat{\boldsymbol{\theta}})]\sqrt{mse(\widehat{\boldsymbol{\theta}})}/\sqrt{g_1(\boldsymbol{\theta})} - 1$. Since $V^* = V + O_p(N^{-1})$, it is verified that $(V^*)^2 = V^2 + O_p(N^{-3/2})$. Hence, from (3.17), it follows that

$$H_1^*(\boldsymbol{\theta}) = H_1(\boldsymbol{\theta}) + O(N^{-3/2}),$$

where $H_1(\boldsymbol{\theta}) = z_\alpha E\big[(2V+V^2) - U^2 - (1+z_{\alpha/2}^2)V^2\big]$. This approximation is used to rewrite (3.16) as

$$P[\mu \in I_1^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + \phi(z_{\alpha/2})H_1(\boldsymbol{\theta}) + O(N^{-3/2}). \tag{3.18}$$

Kubokawa (2010b) showed in the proof of his theorem 2.3 that $H_1(\boldsymbol{\theta}) = O(N^{-3/2})$, which proves the theorem. ■

Theorem 3.3 *Under the same conditions as in Theorem 3.2,*

$$P[\mu \in I_2^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + O(N^{-3/2}). \tag{3.19}$$

Proof. In the proof of Theorem 3.2, we need to replace $mse^*(\widehat{\boldsymbol{\theta}})$, V^* and $h_1^*(\widehat{\boldsymbol{\theta}})$ with $g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})$, W and $h_2^*(\widehat{\boldsymbol{\theta}})$, respectively, where $W = \big\{[1+h_2^*(\widehat{\boldsymbol{\theta}})]\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\}^{1/2} - \sqrt{g_1(\boldsymbol{\theta})}\big\}/\sqrt{g_1(\boldsymbol{\theta})}$. Then,

$$P[\mu \in I_2^{CEB*}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + \phi(z_\alpha)H_2^*(\boldsymbol{\theta}) + O(N^{-3/2}),$$

where

$$H_2^*(\boldsymbol{\theta}) = z_{\alpha/2}E\Big[\big\{2W + W^2\big\} - U^2 - (1+z_{\alpha/2}^2)W^2\Big].$$

Since $2W + W^2 = \{1+h_2^*(\widehat{\boldsymbol{\theta}})\}^2\{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\}/g_1(\boldsymbol{\theta}) - 1 = \{g_1(\widehat{\boldsymbol{\theta}}) + g_2(\widehat{\boldsymbol{\theta}})\}/g_1(\boldsymbol{\theta}) - 1 + 2h_2^*(\widehat{\boldsymbol{\theta}})g_1(\widehat{\boldsymbol{\theta}})/g_1(\boldsymbol{\theta}) + O_p(N^{-3/2})$, it is seen that

$$E[2W + W^2] = \frac{E[g_1(\widehat{\boldsymbol{\theta}})] - g_1(\boldsymbol{\theta})}{g_1(\boldsymbol{\theta})} + \frac{g_2(\boldsymbol{\theta})}{g_1(\boldsymbol{\theta})} + 2E[h(\widehat{\boldsymbol{\theta}})] + O(N^{-3/2}).$$

It follows from the proof of theorem 2.3 of Kubokawa (2010b) that $E[U^2] = \{g_2(\boldsymbol{\theta}) + \overline{g}_3(\boldsymbol{\theta})\}/g_1(\boldsymbol{\theta}) + O(N^{-3/2})$ for $\overline{g}_3(\boldsymbol{\theta})$ given in (3.3) and (3.7). Also it can be shown that $E[(1+z_{\alpha/2}^2)W^2] = 2h_1(\boldsymbol{\theta}) + O(N^{-3/2})$. Combining these approximations gives

$$H_2^*(\boldsymbol{\theta}) = 2z_{\alpha/2}\Big\{h_2^*(\widehat{\boldsymbol{\theta}}) + \frac{E[g_1(\widehat{\boldsymbol{\theta}})] - g_1(\boldsymbol{\theta}) - \overline{g}_3(\boldsymbol{\theta})}{2g_1(\boldsymbol{\theta})} - h_1(\boldsymbol{\theta})\Big\}.$$

It is here noted that

$$E\Big[E_*[g_1(\widehat{\boldsymbol{\theta}}^*)|\boldsymbol{y}] - g_1(\widehat{\boldsymbol{\theta}})\Big] = E[g_1(\widehat{\boldsymbol{\theta}})] - g_1(\boldsymbol{\theta}) + O(N^{-3/2}),$$

which can be verified from (3.8) and (3.9). Also note that $h_1^*(\widehat{\boldsymbol{\theta}}) = h_1(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2})$ and $\overline{g}_3{}^*(\widehat{\boldsymbol{\theta}}) = \overline{g}_3(\widehat{\boldsymbol{\theta}}) + O_p(N^{-3/2})$. This implies that $H_2^*(\boldsymbol{\theta}) = O(N^{-3/2})$, and the proof is complete. ■

3.3 Derivation of $\mathbf{AIC}_1^*$, $\mathbf{AIC}_2^*$ and $\mathbf{cAIC}^*$

We here show that the penalty terms of AIC_1^*, AIC_2^* and cAIC^* given in (2.17) and (2.20) are second-order approximations of unbiased estimators of the corresponding biases.

Concerning AIC_1^* and AIC_2^* given in (2.17), let us define

$$c(\boldsymbol{\theta}) = -\sum_{i=1}^{q} E\big[\mathrm{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}_{\boldsymbol{y}}'\hat{\theta}_i^{\dagger}]\big],$$

for $\widehat{\boldsymbol{\theta}}^{\dagger} = (\hat{\theta}_1^{\dagger}, \ldots, \hat{\theta}_q^{\dagger})'$. Assume the following conditions:

(C1) $\widehat{\boldsymbol{\theta}}^{\dagger}$ and $\widehat{\boldsymbol{\theta}}^{\dagger\dagger}$ satisfy that $E\big[\mathrm{tr}\,[\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}_{\boldsymbol{y}}'\hat{\theta}_i^{\dagger\dagger}]\big] = O(N^{-1})$, $E\big[\mathrm{tr}\,[(\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}_{\boldsymbol{y}}'\hat{\theta}_i^{\dagger})\hat{\theta}_j^{\dagger}]\big] = O(N^{-1})$ and $E\big[\mathrm{tr}\,[(\boldsymbol{\nabla}_{\boldsymbol{y}}\hat{\theta}_i^{\dagger})(\boldsymbol{\nabla}_{\boldsymbol{y}}\hat{\theta}_j^{\dagger})']\big] = O(N^{-1})$ for $\widehat{\boldsymbol{\theta}}^{\dagger\dagger} = (\hat{\theta}_1^{\dagger\dagger}, \ldots, \hat{\theta}_q^{\dagger\dagger})'$.

(C2) $c(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[1]}$ and $\partial_i c(\boldsymbol{\theta}) = O(N^{-1})$ for $1 \le i \le q$.

Theorem 3.4 *Assume the conditions* (A1)-(A4) *and* (C1)-(C2)*. Then,*

$$E\big[-2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\theta}}) - \Delta_1^*(\boldsymbol{y}) - \Delta_2^*(\boldsymbol{y})\big] = AI(\boldsymbol{\theta}) + O(N^{-1/2})$$

for $\Delta_1^*(\boldsymbol{y})$ *and* $\Delta_2^*(\boldsymbol{y})$ *given in* (2.16)*. Also,* $E[\Delta_1^*(\boldsymbol{y})] = -2p + O(N^{-1/2})$ *and* $E[\Delta_2^*(\boldsymbol{y})] = c(\boldsymbol{\theta}) + O(N^{-1/2})$.

Proof. Define $\Delta_1(\boldsymbol{\theta})$ and $\Delta_2(\boldsymbol{\theta})$ by

$$\begin{aligned}\Delta_1(\boldsymbol{\theta}) =& -2E[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{P}(\widehat{\boldsymbol{\theta}})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})],\\ \Delta_2(\boldsymbol{\theta}) =& E\big[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\widehat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \mathrm{tr}\,[\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}^{-1}]\big].\end{aligned}$$

Kubokawa (2011) proved in the proof of his theorem 2.1 that

$$E_{\boldsymbol{y}}\big[-2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\theta}})\big] - AI(\boldsymbol{\theta}) = \Delta_1(\boldsymbol{\theta}) + \Delta_2(\boldsymbol{\theta}),$$

and that $\Delta_1(\boldsymbol{\theta}) = -2p + O(N^{-1/2})$ and $\Delta_2(\boldsymbol{\theta}) = c(\boldsymbol{\theta}) + O(N^{-1/2})$. Thus, it is observed that

$$\begin{aligned}&E_*\Big[(\boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}))'\boldsymbol{P}(\widehat{\boldsymbol{\theta}}^*)(\boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}))\big|\boldsymbol{y}\Big] = p + O_p(N^{-1/2}),\\ &E_*\Big[(\boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}))'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^*)^{-1}(\boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})) - \mathrm{tr}\,[\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^*)^{-1}]\big|\boldsymbol{y}\Big] = c(\widehat{\boldsymbol{\theta}}) + O_p(N^{-1/2}),\end{aligned}$$

Since $c(\boldsymbol{\theta}) = O(1)$, this implies that $E[\Delta_1^*(\boldsymbol{y})] = -2p + O(N^{-1/2})$ and $E[\Delta_2^*(\boldsymbol{y})] = c(\boldsymbol{\theta}) + O(N^{-1/2})$, and we get the theorem. ■

Concerning the $cAIC^*$ given in (2.20), let us define $c_1(\boldsymbol{\theta})$ and $c_2(\boldsymbol{\theta})$ by

$$
\begin{aligned}
c_1(\boldsymbol{\theta}) =& \sum_{i=1}^q \mathrm{tr}\,[(\boldsymbol{R}\boldsymbol{\Sigma}^{-1})_{(i)}]E[\hat{\theta}_i^{\dagger\dagger}] + \frac{1}{2}\sum_{i=1}^q\sum_{j=1}^q \mathrm{tr}\,[(\boldsymbol{R}\boldsymbol{\Sigma}^{-1})_{(ij)}]E[\hat{\theta}_i^\dagger\hat{\theta}_j^\dagger],\\
c_2(\boldsymbol{\theta}) =& 2\sum_{i=1}^q \mathrm{tr}\,\big[\boldsymbol{R}\{(\boldsymbol{\Sigma}^{-1})_{(i)} - (\boldsymbol{R}^{-1})_{(i)}\}\big]E[\hat{\theta}_i^{\dagger\dagger}]\\
&+ \sum_{i=1}^q\sum_{j=1}^q \mathrm{tr}\,\big[\boldsymbol{R}\{(\boldsymbol{\Sigma}^{-1})_{(ij)} - (\boldsymbol{R}^{-1})_{(ij)}\}\big]E[\hat{\theta}_i^\dagger\hat{\theta}_j^\dagger]\\
&+ \sum_{i=1}^q \mathrm{tr}\,\big[\{2\boldsymbol{R}(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{\Sigma} - \boldsymbol{R}(\boldsymbol{R}^{-1})_{(i)}\boldsymbol{R}\}E[\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}_{\boldsymbol{y}}'\hat{\theta}_i^\dagger]\big].
\end{aligned}
$$

Assume the following conditions:

(C3) $\boldsymbol{R}(\boldsymbol{\theta})$ is continuously differentiable three times in $\boldsymbol{\theta}$, and $\lim_{N\to\infty}\lambda_1(\boldsymbol{R}) > 0$, $\lim_{N\to\infty}\lambda_N(\boldsymbol{R}) < \infty$, $\lim_{N\to\infty}|\lambda_N^i(\boldsymbol{R})| < \infty$, $\lim_{N\to\infty}|\lambda_N^{ij}(\boldsymbol{R})| < \infty$ and $\lim_{N\to\infty}|\lambda_N^{ijk}(\boldsymbol{R})| < \infty$, where $\lambda_a(\boldsymbol{R})$, $\lambda_a^i(\boldsymbol{R})$, $\lambda_a^{ij}(\boldsymbol{R})$ and $\lambda_a^{ijk}(\boldsymbol{R})$ are defined similarly to those for $\boldsymbol{\Sigma}$.

(C4) $c_1(\boldsymbol{\theta})$ and $c_2(\boldsymbol{\theta})$ are continuously differentiable functions satisfying $\partial_i c_1(\boldsymbol{\theta}) = O(N^{-1})$ and $\partial_i c_2(\boldsymbol{\theta}) = O(N^{-1})$ for $1 \le i \le q$.

Theorem 3.5 *Assume the conditions* (A1)-(A4) *and* (C3)-(C4)*. Then,*

$$
E\Big[-2\log f(\boldsymbol{y}|\widehat{\boldsymbol{v}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\theta}}) - \Delta_{c1}^*(\boldsymbol{y}) - \Delta_{c2}^*(\boldsymbol{y})\Big] = cAI(\boldsymbol{\theta}) + O(N^{-1/2}),
$$

for $\Delta_{c1}^*(\boldsymbol{y})$ and $\Delta_{c2}^*(\boldsymbol{y})$ given in (2.19)*. Also, $E[\Delta_{c1}^*(\boldsymbol{y})] = -2\rho(\boldsymbol{\theta}) + O(N^{-1/2})$ and $E[\Delta_{c2}^*(\boldsymbol{y})] = c_2(\boldsymbol{\theta}) + O(N^{-1/2})$, where $\rho(\boldsymbol{\theta})$ is the effective degrees of freedom defined by*

$$
\rho(\boldsymbol{\theta}) = \mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}] + N - \mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}].
$$

Proof. Define $\Delta_{c1}(\boldsymbol{\theta})$ and $\Delta_{c2}(\boldsymbol{\theta})$ by

$$
\begin{aligned}
\Delta_{c1}(\boldsymbol{\theta}) =& -2\Big\{E\big[\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{P}(\widehat{\boldsymbol{\theta}})\boldsymbol{u}\big] + N - \mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}]\Big\},\\
\Delta_{c2}(\boldsymbol{\theta}) =& E\big[\boldsymbol{u}'(2\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{R}\widehat{\boldsymbol{R}}^{-1}\boldsymbol{R}\boldsymbol{\Sigma}^{-1})\boldsymbol{u}\big]\\
&- E\big[\mathrm{tr}\,[\widehat{\boldsymbol{R}}^{-1}(2\boldsymbol{R} - \boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{R})]\big] + 2N - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{R}].
\end{aligned}
$$

Using the arguments as in the proof of the theorem 2.2 in Kubokawa (2011), we can show that

$$
E_{\boldsymbol{y}}\big[-2\log f(\boldsymbol{y}|\widehat{\boldsymbol{v}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \widehat{\boldsymbol{\theta}})\big] - cAI(\boldsymbol{\theta}) = \Delta_{c1}(\boldsymbol{\theta}) + \Delta_{c2}(\boldsymbol{\theta}),
$$

and that

$$
\Delta_{c1}(\boldsymbol{\theta}) = -2\rho(\boldsymbol{\theta}) + O(N^{-1/2}) \quad \text{and} \quad \Delta_{c2}(\boldsymbol{\theta}) = c_2(\boldsymbol{\theta}) + O(N^{-1/2}),
$$

since $\Delta_{c1}(\boldsymbol{\theta})$ and $\Delta_{c2}(\boldsymbol{\theta})$ can be expressed as $\Delta_{c1}(\boldsymbol{\theta}) = -2J_3 - 2N + 2\mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}]$ and $\Delta_{c2}(\boldsymbol{\theta}) = J_1 - J_2 + 2N - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{R}]$ based on the notations J_1, J_2 and J_3 in Kubokawa (2011). This implies that

$$
\Delta_{c2}^*(\boldsymbol{y}) = c_2(\widehat{\boldsymbol{\theta}}) + O_p(N^{-1/2}),
$$

for $\Delta_{c2}^*(\boldsymbol{y})$ given in (2.19). Since $c_2(\boldsymbol{\theta}) = O(1)$, it is observed that

$$E\big[\Delta_{c2}^*(\boldsymbol{y})\big] = E[c_2(\widehat{\boldsymbol{\theta}})] + O(N^{-1/2}) = c_2(\boldsymbol{\theta}) + O(N^{-1/2}) = \Delta_2(\boldsymbol{\theta}) + O(N^{-1/2}). \quad (3.20)$$

For $\Delta_{c1}(\boldsymbol{\theta})$, it is noted that $\Delta_{c1}(\boldsymbol{\theta}) = O(N)$ since $\mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}] = O(N)$. Kubokawa (2010b) showed in the proof of his theorem 2.3 that

$$E\big[\mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}]\big] = \mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}] + c_1(\boldsymbol{\theta}) + O(N^{-1/2}),$$

which leads to $E_*\big[\mathrm{tr}\,[\widehat{\boldsymbol{R}}^*\widehat{\boldsymbol{\Sigma}}^{*-1}]\big|\boldsymbol{y}\big] = \mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}] + c_1(\widehat{\boldsymbol{\theta}}) + O_p(N^{-1/2})$. Then,

$$\begin{aligned} E\Big[2\mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}] - E_*\big[\mathrm{tr}\,[\widehat{\boldsymbol{R}}^*\widehat{\boldsymbol{\Sigma}}^{*-1}]\big|\boldsymbol{y}\big]\Big] =& E[\mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}] - c_1(\widehat{\boldsymbol{\theta}})] + O(N^{-1/2}) \\ =& \mathrm{tr}\,[\boldsymbol{R}\boldsymbol{\Sigma}^{-1}] + O(N^{-1/2}). \end{aligned}$$

Since $E[\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{P}(\widehat{\boldsymbol{\theta}})\boldsymbol{u}] = \mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}] + O(N^{-1/2})$, it is seen that $E_*[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\boldsymbol{P}(\widehat{\boldsymbol{\theta}}^*)\boldsymbol{u}^*] = \mathrm{tr}\,[(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}] + O_p(N^{-1/2})$. Hence,

$$E\Big[E_*\big[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{R}}\boldsymbol{P}(\widehat{\boldsymbol{\theta}}^*)\boldsymbol{u}^* + \mathrm{tr}\,[\widehat{\boldsymbol{R}}^*\widehat{\boldsymbol{\Sigma}}^{*-1}]\,\big|\boldsymbol{y}\big] + N - 2\mathrm{tr}\,[\widehat{\boldsymbol{R}}\widehat{\boldsymbol{\Sigma}}^{-1}]\Big] = \rho(\boldsymbol{\theta}) + O(N^{-1/2}),$$

that is,

$$E[\Delta_{c1}^*(\boldsymbol{y})] = -2\rho(\boldsymbol{\theta}) + O(N^{-1/2}) = \Delta_{c1}(\boldsymbol{\theta}) + O(N^{-1/2}). \quad (3.21)$$

Combining (3.20) and (3.21) proves the theorem. ■

4 Simulation Studies in the Fay-Herriot Model

In this section, we investigate how the proposed procedures perform in comparison with ones given in the literature in the Fay-Herriot model.

4.1 Fay-Herriot model and procedures used for comparison

The basic area level model proposed by Fay and Herriot (1979) is described by

$$y_a = \boldsymbol{x}_a'\boldsymbol{\beta} + v_a + \varepsilon_a, \quad a = 1, \ldots, k, \quad (4.1)$$

where k is the number of small areas, $\boldsymbol{x}_a$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients, and v_a's and ε_a's are mutually independently distributed random errors such that $v_a \sim \mathcal{N}(0, \theta)$ and $\varepsilon_a \sim \mathcal{N}(0, d_a)$ for known d_a's. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)'$, $\boldsymbol{y} = (y_1, \ldots, y_k)'$, and let $\boldsymbol{v}$ and $\boldsymbol{\epsilon}$ be similarly defined. Then, the model is expressed in vector notations as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\epsilon}$ and $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\theta) = \theta\boldsymbol{I}_k + \boldsymbol{D}$ for $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_k)$. In this model, $\boldsymbol{R} = \boldsymbol{D}$ and $\boldsymbol{G} = \theta\boldsymbol{I}_k$.

For estimating θ, the following procedures are known and their biases and variances up to second order are summarized in Kubokawa (2010b).

[ML] The MLE $\hat{\theta}^M$ is given as the solution of the equation $L^M(\hat{\theta}^M) = 0$, where $L^M(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))^2\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{\Sigma}(\theta)^{-1}]$ for $\boldsymbol{P}(\theta) = \boldsymbol{\Sigma}(\theta)^{-1}\boldsymbol{X}\{\boldsymbol{X}'\boldsymbol{\Sigma}(\theta)^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\theta)^{-1}$.

[REML] The REML estimator $\hat{\theta}^R$ is given as the solution of the equation $L^R(\hat{\theta}^R) = 0$, where $L^R(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))^2\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta)]$.

[Fay-Herriot estimator] The Fay-Herriot estimator $\hat{\theta}^{FH}$ is the solution of the equation $L^{FH}(\hat{\theta}^{FH}) = 0$, where $L^{FH}(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))\boldsymbol{y} - (k-p)$.

[Modified Fay-Herriot estimator] The modified Fay-Herriot estimator $\hat{\theta}^m$ is the solution of the equation $L_m(\hat{\theta}^m) = 0$, where $L_m(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))\boldsymbol{y} - (k-p) - m(\theta)$ for $m(\theta) = 2k\mathrm{tr}\,[\boldsymbol{\Sigma}(\theta)^{-2}]/(\mathrm{tr}\,[\boldsymbol{\Sigma}(\theta)^{-1}])^2 - 2$. Although $\hat{\theta}^{FH}$ has a second-order bias, the second-order bias of the modified Fay-Herriot estimator $\hat{\theta}^m$ vanishes by adding the correction term $m(\theta)$.

[Prasad-Rao estimator] The unbiased estimator suggested by Prasad and Rao (1990) is $\hat{\theta}^U = (k-p)^{-1}(\boldsymbol{y}'\boldsymbol{E}_0\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{E}_0])$ for $\boldsymbol{E}_0 = \boldsymbol{I}_k - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Since $\hat{\theta}^U$ takes negative values with a positive probability, they proposed to use the truncated estimator $\hat{\theta}^{TR} = \max\{\hat{\theta}^U, 0\}$.

For estimator $\hat{\theta}$ given above, Model 2 in (2.2) is described as

$$y_a^* = \boldsymbol{x}_a'\widehat{\boldsymbol{\beta}}(\hat{\theta}) + v_a^* + \varepsilon_a^*, \quad a = 1, \ldots, k, \tag{4.2}$$

where v_a^*'s and ε_a^*'s are mutually independently distributed random errors such that $v_a^*|\boldsymbol{y} \sim \mathcal{N}(0, \hat{\theta})$ and $\varepsilon_a^* \sim \mathcal{N}(0, d_a)$ for known d_a's. The estimators $\hat{\theta}^*$ and $\widehat{\boldsymbol{\beta}}^*(\hat{\theta}^*)$ can be obtained from y_a^*, $a = 1, \ldots, k$, by using the same techniques used to obtain $\hat{\theta}$ and $\widehat{\boldsymbol{\beta}}(\hat{\theta})$.

[1] Estimation of MSE of EBLUP. It is supposed that we want to predict $\mu_s = \boldsymbol{x}_s'\boldsymbol{\beta} + v_s$ for some index s among $1, \ldots, k$, namely, the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ used in Section 2.2 correspond to $\boldsymbol{a} = \boldsymbol{x}_s$ and $\boldsymbol{b} = \boldsymbol{j}_s$ where $\boldsymbol{j}_s = (0, \ldots, 0, 1, 0, \ldots, 0)'$, namely, the s-th element is one, and the other elements are zero. EBLUP of μ_s is written as

$$\widehat{\mu}_s^{EB} = \widehat{\mu}_s^{EB}(y_s; \widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) = \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta}) + \{1 - \gamma_s(\hat{\theta})\}(y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})),$$

for $\gamma_s(\theta) = d_s/(\theta + d_s)$, and the functions $\boldsymbol{s}(\theta)$, $g_1(\theta)$ and $g_2(\theta)$ are expressed as

$$\boldsymbol{s}(\theta) = \{1 - \gamma_s(\theta)\}\boldsymbol{j}_s, \quad g_1(\theta) = d_s\{1 - \gamma_s(\theta)\}, \quad g_2(\theta) = \gamma_s(\theta)^2\boldsymbol{x}_s'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_s.$$

Concerning the estimation of the MSE of EBLUP $\widehat{\mu}_s^{EB}(y_s; \widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta})$, we here handle the following four estimators: One is the MSE estimator based on the parametric bootstrap method given by

$$mse^*(\hat{\theta}, \widehat{\mu}_s^{EB}) = 2\{g_1(\hat{\theta}) + g_2(\hat{\theta})\} - E_*\big[g_1(\hat{\theta}^*) + g_2(\hat{\theta}^*)\big|\boldsymbol{y}\big] + \overline{g}_3{}^*(\hat{\theta}), \tag{4.3}$$

which is from (2.9), where $\overline{g}_3{}^*(\hat{\theta}) = E_*\big[\{\gamma_s(\hat{\theta}^*) - \gamma_s(\hat{\theta})\}^2(\hat{\theta} + d_s)|\boldsymbol{y}\big]$. Butar and Lahiri (2003) suggested another MSE estimator based on the parametric bootstrap method given by

$$\begin{aligned} mse^{BL}(\hat{\theta}, \widehat{\mu}_s^{EB}) =& 2\{g_1(\hat{\theta}) + g_2(\hat{\theta})\} - E_*\big[g_1(\hat{\theta}^*) + g_2(\hat{\theta}^*)\big|\boldsymbol{y}\big] \\ &+ E_*\big[\{\widehat{\mu}_s^{EB}(y_s; \widehat{\boldsymbol{\beta}}(\hat{\theta}^*), \hat{\theta}^*) - \widehat{\mu}_s^{EB}(y_s; \widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta})\}^2|\boldsymbol{y}\big], \end{aligned} \tag{4.4}$$

Prasad and Rao (1990) and Datta and Lahiri (2000) suggested the MSE estimator based on the Taylor series expansion given by

$$mse(\hat{\theta}, \widehat{\mu}_s^{EB}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2\{\gamma_s(\hat{\theta})^3/d_s\}V(\hat{\theta}) - \{\gamma_s(\hat{\theta})\}^2 B(\hat{\theta}), \tag{4.5}$$

for $B(\theta) = E[\hat{\theta}^{\dagger\dagger}]$ and $V(\theta) = Var(\hat{\theta}^{\dagger})$. This can be also derived from (3.5). In this model, an exact unbiased estimator of the MSE can be derived by using the Stein identity and Datta, Kubokawa, Molina and Rao (2011) provided the unbiased estimator given by

$$mse^E(\hat{\theta}, \widehat{\mu}_s^{EB}) = d_s - 2d_s \frac{\partial}{\partial y_s}\Big[\gamma_s(\hat{\theta})\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})\}\Big] + \{\gamma_s(\hat{\theta})\}^2\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})\}^2, \tag{4.6}$$

where

$$\begin{aligned}
&\frac{\partial}{\partial y_s}\Big[d_s(\hat{\theta}+d_s)^{-1}\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})\}\Big] \\
=&\frac{d_s}{\hat{\theta}+d_s} - \frac{d_s}{(\hat{\theta}+d_s)^2}\{y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})\}\frac{\partial\hat{\theta}}{\partial y_s} \\
&- \frac{d_s}{\hat{\theta}+d_s}\boldsymbol{x}_s'(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}(\hat{\theta})^{-1}\boldsymbol{X})^{-1}\Big\{\frac{\boldsymbol{x}_s}{\hat{\theta}+d_s} - \sum_j \frac{\boldsymbol{x}_j\{y_j - \boldsymbol{x}_j'\widehat{\boldsymbol{\beta}}(\hat{\theta})\}}{(\hat{\theta}+d_j)^2}\frac{\partial\hat{\theta}}{\partial y_s}\Big\}.
\end{aligned}$$

[2] Corrected confidence interval. We here treat the following four confidence intervals based on EBLUP: The confidence intervals (2.12) and (2.13) with the correction terms using the parametric bootstrap method written by

$$I_1^{CEB*}(\hat{\theta}):\ \widehat{\mu}^{EB}(\hat{\theta}) \pm z_{\alpha/2}\Big[1 + h_1^*(\hat{\theta})\Big]\sqrt{\max\{mse^*(\hat{\theta}, \widehat{\mu}^{EB}), 0\}}, \tag{4.7}$$

$$I_2^{CEB*}(\hat{\theta}):\ \widehat{\mu}^{EB}(\hat{\theta}) \pm z_{\alpha/2}\Big[1 + h_2^*(\hat{\theta})\Big]\sqrt{g_1(\hat{\theta}) + g_2(\hat{\theta})}, \tag{4.8}$$

where $mse^*(\hat{\theta}, \widehat{\mu}^{EB})$ is given in (4.3) and

$$\begin{aligned}
h_1^*(\hat{\theta}) =& \frac{1 + z_{\alpha/2}^2}{8g_1(\hat{\theta})^2}E_*\Big[\{g_1(\hat{\theta}^*) - g_1(\hat{\theta})\}^2\big|\boldsymbol{y}\Big], \\
h_2^*(\hat{\theta}) =& h_1^*(\hat{\theta}) + \frac{g_1(\hat{\theta}) - E_*[g_1(\hat{\theta}^*)|\boldsymbol{y}] + \overline{g}_3{}^*(\hat{\theta})}{2g_1(\hat{\theta})}.
\end{aligned}$$

As a confidence interval based on the Taylor series expansion, we treat the confidence interval with the correction term, given by

$$I^{CEB}(\hat{\theta}):\ \widehat{\mu}^{EB}(\hat{\theta}) \pm z_{\alpha/2}\Big[1 + h(\hat{\theta})\Big]\sqrt{mse(\hat{\theta}, \widehat{\mu}^{EB})}, \tag{4.9}$$

where $mse(\hat{\theta}, \widehat{\mu}^{EB})$ is given in (4.5) and

$$h(\theta) = \frac{z_{\alpha/2}^2 + 1}{8\theta^2(\theta + d_s)^2}d_s^2 Var(\hat{\theta}^{\dagger}).$$

The confidence interval proposed by Chatterjee, *et al.* (2008) is different from ours. As seen from (2.5), the conditional distribution of μ_s given $\boldsymbol{y}$ is

$$\mu_s|\boldsymbol{y} \sim \mathcal{N}(\widehat{\mu}_s(y_s, \theta, \boldsymbol{\beta}), \sigma_s^2(\theta)),$$

where $\mu_s = \boldsymbol{x}_s'\boldsymbol{\beta} + v_s$, $\widehat{\mu}_s(y_s, \theta, \boldsymbol{\beta}) = \boldsymbol{x}_s'\boldsymbol{\beta} + \theta(\theta + d_s)^{-1}(y_s - \boldsymbol{x}_s'\boldsymbol{\beta})$ and $\sigma_s^2(\theta) = d_s\theta(\theta + d_s)^{-1}$. Thus, $\sigma_s(\theta)^{-1}\{\mu_s - \widehat{\mu}_s(y_s, \theta, \boldsymbol{\beta})\} \sim \mathcal{N}(0, 1)$. Although this suggests to construct a confidence interval from the distribution of $\sigma_s(\hat{\theta})^{-1}\{\mu_s - \widehat{\mu}_s(y_s, \hat{\theta}, \widehat{\boldsymbol{\beta}}(\hat{\theta}))\}$, the distribution is not normal. Define $\mathcal{L}(z)$ and $\mathcal{L}^*(z)$ by

$$\begin{aligned}
\mathcal{L}(z) =& P\left[\sigma_s(\hat{\theta})^{-1}\{\mu_s - \widehat{\mu}_s(y_s, \hat{\theta}, \widehat{\boldsymbol{\beta}}(\hat{\theta}))\} \le z\right], \\
\mathcal{L}^*(z) =& P\left[\sigma_s(\hat{\theta}^*)^{-1}\{\mu_s^* - \widehat{\mu}_s(y_s^*, \hat{\theta}^*, \widehat{\boldsymbol{\beta}}^*(\hat{\theta}^*))\} \le z\right],
\end{aligned}$$

where $\mu_s^* = \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta}) + v_s^*$. Chatterjee, *et al.* (2008) proved that $\mathcal{L}(z)$ can be approximated by $\mathcal{L}^*(z)$ with the second-order accuracy, and proposed the confidence interval of μ_s given by

$$I^{CLL} = \left[\widehat{\mu}_s(y_s, \hat{\theta}, \widehat{\boldsymbol{\beta}}(\hat{\theta})) - q_1\sigma_s(\hat{\theta}),\ \widehat{\mu}_s(y_s, \hat{\theta}, \widehat{\boldsymbol{\beta}}(\hat{\theta})) - q_2\sigma_s(\hat{\theta})\right], \tag{4.10}$$

where q_1 and q_2 satisfies that

$$\mathcal{L}^*(q_2) - \mathcal{L}^*(q_1) = 1 - \alpha.$$

[3] AIC. Two kinds of Akaike Information Criteria based on the parametric bootstrap method are given by

$$\begin{aligned}
AIC_1^* =& -2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) + 2p - \Delta_2^*(\boldsymbol{y}), \\
AIC_2^* =& -2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) - \Delta_1^*(\boldsymbol{y}) - \Delta_2^*(\boldsymbol{y}),
\end{aligned} \tag{4.11}$$

where for $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\boldsymbol{\Sigma}}^* = \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^*)$,

$$\begin{aligned}
\Delta_1^*(\boldsymbol{y}) =& -2E_*\left[\boldsymbol{u}^{*\prime}\boldsymbol{P}(\hat{\theta}^*)\boldsymbol{u}^*\big|\boldsymbol{y}\right], \\
\Delta_2^*(\boldsymbol{y}) =& E_*\left[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{*-1}\boldsymbol{u}^* - \mathrm{tr}\,[\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}^{*-1}]\big|\boldsymbol{y}\right],
\end{aligned}$$

for $\boldsymbol{u}^* = \boldsymbol{y}^* - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\hat{\theta})$. On the other hand, AIC based on the Taylor expansion is described as

$$AIC = -2\log f_m(\boldsymbol{y}|\widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) - \Delta(\hat{\theta}), \tag{4.12}$$

where $\Delta(\theta) = -2p - E\big[\mathrm{tr}\,[\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}_{\boldsymbol{y}}'\hat{\theta}^\dagger]\big]$. This was derived by Kubokawa (2011).

[4] Conditional AIC. The conditional AIC based on the parametric bootstrap method is given by

$$cAIC^* = -2\log f(\boldsymbol{y}|\widehat{\boldsymbol{v}}(\hat{\theta}), \widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) - \Delta_{c1}^*(\boldsymbol{y}) - \Delta_{c2}^*(\boldsymbol{y}), \tag{4.13}$$

where

$$\Delta^*_{c1}(\boldsymbol{y}) = -2\Big\{E_*\big[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}\boldsymbol{P}(\hat{\theta}^*)\boldsymbol{u}^* + \mathrm{tr}\,[\boldsymbol{D}\widehat{\boldsymbol{\Sigma}}^{*-1}]\big|\boldsymbol{y}\big] + k - 2\mathrm{tr}\,[\boldsymbol{D}\widehat{\boldsymbol{\Sigma}}^{-1}]\Big\},$$

$$\Delta^*_{c2}(\boldsymbol{y}) = 2E_*\big[\boldsymbol{u}^{*\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{D}\widehat{\boldsymbol{\Sigma}}^{*-1}\boldsymbol{u}^*\big|\boldsymbol{y}\big] - 2\mathrm{tr}\,[\boldsymbol{D}\widehat{\boldsymbol{\Sigma}}^{-1}].$$

Kubokawa (2011) derived the conditional AIC based on the Taylor series expansion given as

$$cAIC = -2\log f(\boldsymbol{y}|\widehat{\boldsymbol{v}}(\hat{\theta}), \widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta}) - \Delta_c(\hat{\theta}), \tag{4.14}$$

where $\Delta_c(\theta) = -2\rho(\theta) - 2\mathrm{tr}\,[\boldsymbol{D}\boldsymbol{\Sigma}^{-1}E[\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}'_{\boldsymbol{y}}\hat{\theta}^{\dagger}]]$.

4.2 Simulation results

We now investigate the performances of the proposed procedures by simulation and compare them with some existing procedures given in the literature. For the purpose, we adopt part of the simulation framework of Datta, *et al.* (2005) for our study. We consider the Fay-Herriot model (4.1) with $k = 15$, $\theta = 1$ and two d_i-patterns: (a) 0.7, 0.6, 0.5, 0.4, 0.3; (b) 4.0, 0.6, 0.5, 0.4, 0.1, which correspond to patterns (a) and (c) of Datta, *et al.* (2005). Pattern (a) is less variable in d_i-values, while pattern (b) has larger variability. There are five groups $G_1, \ldots, G_5$ and three small areas in each group. The sampling variances d_i are the same for area within the same group. For the sake of computational simplicity, we mainly employ the Prasad-Rao estimator $\hat{\theta} = \hat{\theta}^{TR}$ with the truncation as $\hat{\theta}^{TR} = \max(\hat{\theta}^U, k^{-1/2})$. As shown in Kubokawa (2010b, 11), we observe that $B(\theta)$ and $V(\theta)$ in (4.5) are $B(\theta) = 0$ and $V(\theta) = 2k^{-2}\mathrm{tr}\,\boldsymbol{\Sigma}^2$, and that $\boldsymbol{\nabla}_{\boldsymbol{y}}\hat{\theta} = 2\boldsymbol{E}_0\boldsymbol{y}I(\hat{\theta} > 1/\sqrt{k})$ for the indicator function $I(\cdot)$. Since $\boldsymbol{\nabla}_{\boldsymbol{y}}\boldsymbol{\nabla}'_{\boldsymbol{y}}\hat{\theta} = 2\boldsymbol{E}_0/(k-p)I(\hat{\theta} > 1/\sqrt{k})$, it is seen that $\Delta(\theta) = -2(p+1)$ and $\Delta_c(\theta) = -2\rho(\theta) - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{D}]/k$. For pattern (b), we also investigate the performances for the Fay-Herriot estimator with the truncation at k^{-1}. For the truncated Fay-Herriot estimator, $B(\theta) = 2\{k\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2\}/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3$, $V(\theta) = 2k/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2$, $\boldsymbol{\nabla}_{\boldsymbol{y}}\hat{\theta} = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{\Sigma}^{-2}\boldsymbol{y}I(\hat{\theta} > k^{-1})$, $\Delta(\theta) = -2(p+1)$ and $\Delta_c(\theta) = -2\rho(\theta) - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}\boldsymbol{D}]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]$.

[Simulation experiment I] Let us consider the case that $\boldsymbol{x}'_a\boldsymbol{\beta} = 0$ for simplicity as handled in Chatterjee, *et al.* (2008). Then, $\mu_s = v_s$, $\widehat{\mu}^{EB}_s = \{1 - \gamma_s(\hat{\theta})\}y_s$ and $MSE(\theta, \widehat{\mu}_s(\hat{\theta})) = d_s - d_s\gamma_s(\theta) + E[\{\gamma_s(\hat{\theta}) - \gamma_s(\theta)\}^2 y_s^2]$ for $\gamma_s(\theta) = d_s/(\theta + d_s)$. We prepare the true values of $MSE(\theta, \widehat{\mu}_s(\hat{\theta}))$ in advance, which can be computed based on 100,000 simulated data. The relative bias and the risk functions of MSE estimator mse_s are given by

$$B_s(\theta, mse_s) = 100 \times E\Big[mse_s - MSE(\theta, \widehat{\mu}_s(\hat{\theta}))\Big]/MSE(\theta, \widehat{\mu}_s(\hat{\theta})),$$

$$R_s(\theta, mse_s) = 100 \times E\Big[\big\{mse_s - MSE(\theta, \widehat{\mu}_s(\hat{\theta}))\big\}^2\Big]/\{MSE(\theta, \widehat{\mu}_s(\hat{\theta}))\}^2.$$

For confidence interval CI_s of $\mu_s = v_s$, the coverage probability and the length of CI_s are given by

$$CP_s(\theta, CI_s) = P\Big[\mu_s \in CI_s\Big],$$

$$EL_s(\theta, CI_s) = E\Big[\text{Length of } CI_s\Big].$$

These values are computed as average values based on 10,000 simulation runs where the size of the bootstrap sample is 1,000. Further, those values are averaged over areas within groups G_i, $i = 1, \dots, 5$.

Concerning the MSE estimation, we handle the four estimators mse, mse^*, mse^{BL} and mse^E given in (4.5), (4.3), (4.4) and (4.6), which are referred as TLap, PBap, PBbl and ExactU, respectively, since TLap is based on the Taylor approximation, PBap uses the parametric bootstrap procedure based on the approximation, PBbl is the parametric bootstrap method of Butar and Lahiri (2003) and ExactU is an exact unbiased estimator. The values of their relative biases $10^2 \times B_s(\theta, mse_s)$ and risks $10^2 \times R_s(\theta, mse_s)$ by simulation are reported in Table 1, where the Prasad-Rao estimator is treated in columns (A) and (B) and the Fay-Herriot estimator is treated in column (C).

Since the ExactU is an unbiased estimator, it is clear that the values of the bias of ExactU are quite small, but it has very large risks. This means that estimating the MSE unbiasedly does not necessarily lead to improvement of the risk, but rather yields large variability in general. Investigating the relative biases and risks of TLap, PBap and PBbl in details, we can see that TLap is a positive bias, but PBap and PBbl have negative biases in the column (A) for appropriately balanced d_i's, and TLap has a slightly smaller risk than others. Although the relative biases and the risks of TLap, PBap and PBbl are small in (A), their values are large in the extreme case (B) with quite large d_i in group G_1. Especially, TLap have larger relative biases and risks except group G_1. This may be caused by the estimator of θ. We thus investigate the case (C) where the Fay-Herriot estimator is used for θ. As indicated in (C), the biases and risks of TLap, PBap and PBbl are reasonably small for the Fay-Herriot estimator. In comparison of PBap and PBbl, it is seen that PBbl has a slightly smaller bias, while PBap has a slightly smaller risk, but their difference is little.

Thus, we suggest from Table 1 that the estimators TLap, PBap and PBbl with the Prasad-Rao estimator are good in pattern (a), but for pattern (b), the use of the Fay-Herriot estimator in TLap, PBap and PBbl is recommendable. It is also clear that the exact unbiased estimator ExactU is not useful.

Concerning the interval estimation, we handle the four confidence intervals I^{CEB}, I_1^{CEB*}, I_2^{CEB*} and I^{CLL} given in (4.9), (4.7), (4.8) and (4.10), which are referred as TLap, PB_1, PB_2 and PBcll, respectively. The values of their coverage probabilities $10^2 \times CP$ and expected lengths $10^2 \times EL$ by simulation are reported in Table 2.

In (A) for pattern (a) in Table 2, three confidence intervals PB_1, PB_2 and PBcll have slightly smaller coverage probabilities than 95% nominal coefficient, but their differences are not significant. TLap satisfies the nominal confidence level, but it has a tendency to take slightly larger coverage probability than 95%. In the case (B) for pattern (b), however, the parametric bootstrap procedures PB_1, PB_2 and PBcll are not good since their coverage probabilities are much smaller than the nominal confidence coefficient. As defined in (4.2), the conditional distribution of y_a^* given $\boldsymbol{y}$ is $\mathcal{N}(\boldsymbol{x}_a'\widehat{\boldsymbol{\beta}}(\hat{\theta}), \hat{\theta} + d_a)$, and the variability of y^* strongly depends on the estimate $\hat{\theta}$. This implies that we need to use an estimator of θ with higher precision. Thus, we employ the Fay-Herriot estimator instead of the Prasad-Rao estimator. The resulting CP and EL are given in the column (C).

Table 1: Values of relative biases and risks of the four MSE estimators for $\theta = 1$

	d_i	MSE	**bias** $B(\theta, mse)$ TLap	PBap	PBbl	ExactU	**risk** $R(\theta, mse)$ TLap	PBap	PBbl	ExactU
			(A) pattern (a) for the Prasad-Rao estimator							
G_1	0.7	0.438	1.98	−2.50	−2.51	0.04	3.4	7.4	8.3	128.6
G_2	0.6	0.398	2.21	−2.24	−2.33	−0.08	2.6	6.5	7.4	100.9
G_3	0.5	0.354	2.41	−1.97	−1.98	0.00	1.8	5.5	6.3	81.8
G_4	0.4	0.303	2.60	−1.67	−1.69	−0.37	1.0	4.4	5.2	60.0
G_5	0.3	0.244	2.78	−1.31	−1.18	−0.26	0.4	3.3	4.0	41.4
			(B) pattern (b) for the Prasad-Rao estimator							
G_1	4.0	0.909	8.73	−11.48	−10.72	0.97	28	45	52	2261
G_2	0.6	0.425	43.66	−9.06	−8.81	−0.18	25	15	17	159
G_3	0.5	0.378	47.95	−8.93	−8.53	−0.12	35	13	15	136
G_4	0.4	0.325	52.99	−8.69	−8.19	−0.61	51	11	12	109
G_5	0.1	0.100	60.01	−5.48	−4.17	0.21	105	2	3	27
			(C) pattern (b) for the Fay-Herriot estimator							
G_1	4.0	0.854	−1.59	−2.48	−2.08	1.55	22	23	25	2744
G_2	0.6	0.403	−0.40	0.21	0.09	0.01	4	6	8	111
G_3	0.5	0.358	−0.20	0.88	0.82	−0.23	3	5	7	86
G_4	0.4	0.306	0.05	1.77	1.75	−0.51	2	4	6	64
G_5	0.1	0.094	2.14	6.79	8.09	−0.05	0	1	4	11

As indicated, coverage probabilities of all the confidence intervals are improved in the sense that they are close to 95%. Especially, PB_2 and PBcll have superior behaviors. The difference between PB_1 and PB_2 appears in the coverage probability at G_1 in the case (C). In fact, mse^* sometimes takes negative values in this case, and the resulting confidence interval PB_1 cannot construct an interval, which yields 94.6% coverage probability at G_1, slightly smaller than the nominal confidence coefficient. PB_2 and PBcll are free from such a drawback. Comparing PB_2 and PBcll in case (C), we see that PB_2 has slightly smaller EL than PBcll, but the difference is quite small.

Thus, we suggest from Table 2 that the confidence intervals TLap, PBap and PB-cll with the Prasad-Rao estimator are not bad in pattern (a), but for pattern (b), the confidence intervals PB_2 and PBcll with the Fay-Herriot estimator are good.

[**Simulation experiment II**] We next investigate the performances of AIC and conditional AIC derived in the previous sections through simulation and compare them in terms of the frequencies of selecting the true model.

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ be generated from $\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_x)$ where $\boldsymbol{\Sigma}_x = (1-\rho_x)\boldsymbol{I}_p + \rho_x \boldsymbol{J}_p$ for $\rho_x = 0.1$, where $\boldsymbol{J}_p = \boldsymbol{j}_p \boldsymbol{j}_p'$ for $\boldsymbol{j}_p = (1, \ldots, 1)'$, a p-vector of ones. Suppose that the true model is given by

$$(p^*) \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{v} + \boldsymbol{\epsilon},$$

Table 2: Values of coverage probability and expected length of the four confidence intervals with nominal confidence coefficient 95% for $\theta = 1$

		cov. prob. $10^2 \times CP$				**expected length** EL			
	d_i	TLap	PB_1	PB_2	PBcll	TLap	PB_1	PB_2	PBcll
	(A)	**pattern (a) for the Prasad-Rao estimator**							
G_1	0.7	96.2	94.0	94.0	93.8	2.78	2.62	2.63	2.62
G_2	0.6	96.4	94.2	94.3	93.9	2.64	2.50	2.50	2.49
G_3	0.5	96.2	94.2	94.2	94.0	2.47	2.35	2.35	2.34
G_4	0.4	96.0	94.3	94.4	94.0	2.27	2.17	2.17	2.16
G_5	0.3	96.2	94.7	94.7	94.4	2.02	1.94	1.94	1.93
	(B)	**pattern (b) for the Prasad-Rao estimator**							
G_1	4.0	99.6	95.7	97.9	89.3	7.90	4.20	4.52	3.95
G_2	0.6	98.8	92.9	93.1	91.0	5.44	2.56	2.57	2.51
G_3	0.5	98.7	92.8	93.0	91.3	5.00	2.39	2.40	2.36
G_4	0.4	98.5	92.6	92.7	91.5	4.44	2.20	2.20	2.17
G_5	0.1	97.4	93.8	93.8	93.5	1.73	1.20	1.21	1.19
	(C)	**pattern (b) for the Fay-Herriot estimator**							
G_1	4.0	96.2	94.6	95.9	95.3	4.05	3.94	4.00	4.15
G_2	0.6	96.2	95.2	95.3	95.2	2.66	2.60	2.61	2.67
G_3	0.5	96.1	95.2	95.3	95.3	2.50	2.45	2.45	2.50
G_4	0.4	96.0	95.2	95.3	95.2	2.30	2.26	2.27	2.30
G_5	0.1	95.5	95.7	95.7	95.2	1.25	1.26	1.26	1.23

Table 3: Frequencies selected by the five criteria AIC, AIC_1^*, AIC_2^*, $cAIC$ and $cAIC^*$ for $k = 15$, $\theta = 1$ and the Prasad-Rao estimator: the dimension of a full model is $p = 7$ and the true model is (3)

	pattern (a)					**pattern (b)**				
(m)	AIC	AIC_1^*	AIC_2^*	$cAIC$	$cAIC^*$	AIC	AIC_1^*	AIC_2^*	$cAIC$	$cAIC^*$
(1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.5
(2)	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	1.6	3.5
(3)	62.7	60.2	62.2	73.1	67.4	58.9	56.4	57.9	64.0	57.0
(4)	13.7	13.9	12.7	12.4	11.8	13.1	13.9	12.2	13.6	13.3
(5)	8.5	9.0	8.5	6.2	6.7	9.6	9.9	8.7	8.7	8.5
(6)	7.7	8.3	8.0	4.5	5.8	8.9	9.5	9.4	6.6	7.5
(7)	7.4	8.4	8.4	3.5	6.1	9.3	10.1	11.5	5.2	8.4
(3) + (4)	76.4	74.1	74.9	85.1	79.2	72.0	70.3	70.1	77.6	70.3

where $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_{p^*}, 0, \ldots, 0)'$ for $1 \le p^* \le p$, and $\boldsymbol{v}$ and $\boldsymbol{\epsilon}$ are mutually independent random variables having $\boldsymbol{v} \sim \mathcal{N}_k(\mathbf{0}, \theta \boldsymbol{I})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{D})$. Also, β_ℓ for $1 \le \ell \le p^*$ is generated as a random variable distributed as $\beta_\ell = 2(-1)^{\ell+1}\{1 + U(0,1)\}$ for a uniform random variable $U(0,1)$ on the interval $(0,1)$. We here handle the case that $\theta = 1$, $p = 7$, $p^* = 3$ and $k = 15$. Let (m) be the set $\{1, \ldots, m\}$, and we write the model using the first m regressor variables $\beta_1, \ldots, \beta_m$ by M_m or simply (m). Then, the full model is (7) and the true model is (p^*). As candidate models, we consider the nested subsets $(1), \ldots, (7)$ of $\{1, \ldots, 7\}$, namely,

$$(m) \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\beta}^{(m)} = (\beta_1, \ldots, \beta_m, 0, \ldots, 0)'$. Corresponding to the model (m), we generete random variables from the parametric bootstrap model $\boldsymbol{y}^* = \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{(m)} + \boldsymbol{Z}\boldsymbol{v}^* + \boldsymbol{\epsilon}^*$, where $\boldsymbol{v}^*$ and $\boldsymbol{\epsilon}^*$ are defined below (4.2).

We compare the five selection criteria in the sense of frequency of selecting the true model (3). The criteria we examine are the existing procedures AIC and $cAIC$ based on the Taylor series expansion, and the proposed ones AIC_1^*, AIC_2^* and $cAIC^*$ based on the parametric bootstrap method. In the simulation experiments, we use the Prasad-Rao estimator for θ. For each criterion and each candidate model (m), the number of selecting the model (m) is counted for 10,000 data set. We thus obtain the frequencies of the model (m) selected by the criteria by dividing the number by 10,000. These frequencies are reported in Table 3, where the last column denotes the sum of two frequencies of (3) and (4). Although the model (4) is not true, it includes the true model (3), so that it may be not bad to look at the sum of the two frequencies. From Table 3, it is seen that the parametric bootstrap procedures AIC_1^*, AIC_2^* and $cAIC^*$ are slightly worse than the existing criteria AIC and $cAIC$ for patterns (a) and (b). For pattern (a), conditional information criteria $cAIC$ and $cAIC^*$ have higher frequencies than the unconditional ones AIC and AIC_i^*, $i = 1, 2$. It is also seen that the differences in AIC, AIC_1^* and AIC_2^* are not significant. The frequencies of all the criteria are smaller in pattern (b) than in pattern (a).

The pattern (b) is the extreme case, and it may be better to use the Fay-Herriot estimator instead of the Prasad-Rao estimator. Table 4 reports the frequencies of the five criteria when the Fay-Herriot estimator is used for θ. From this table, we can see that the parametric bootstrap procedures AIC_1^*, AIC_2^* and $cAIC^*$ are much improved in light of frequency, and they are slightly better than AIC and $cAIC$.

Through these experiments, $cAIC$ and $cAIC^*$ with the Prasad-Rao estimator are good for pattern (a) and that the use of the Fay-Herriot estimator for θ is recommendable for pattern (b). However, it seems that there are not significant differences among the frequencies of the five information criteria, and we need to examine more abour their behaviors in other situations and models as a future study.

Table 4: Frequencies selected by the five criteria AIC, AIC_1^*, AIC_2^*, $cAIC$ and $cAIC^*$ for $k = 15$, $\theta = 1$ and the Fay-Herriot estimator: the dimension of a full model is $p = 7$ and the true model is (3)

	pattern (b)				
(m)	AIC	AIC_1^*	AIC_2^*	$cAIC$	$cAIC^*$
(1)	0.0	0.0	0.0	0.7	1.5
(2)	0.0	0.0	0.0	1.5	3.8
(3)	59.1	66.4	66.4	63.1	65.5
(4)	12.8	10.3	10.3	13.8	11.9
(5)	9.6	7.5	7.5	8.3	6.0
(6)	8.6	7.0	7.0	6.2	5.1
(7)	9.7	8.6	8.6	6.0	6.0
(3) + (4)	71.9	76.7	76.7	76.9	77.4

5 An Empirical Study Based on the Nested Error Regression Model

In this section, we investigate whether the proposed procedures work practically through the posted land price data. To analyze the data, we use the nested error regression model, which has been extensively employed in the literature as a unit level model since Battese, *et al*. (1988).

5.1 Nested error regression model

The nested error regression model (NERM) is described as

$$y_{ab} = \boldsymbol{x}'_{ab}\boldsymbol{\beta} + v_a + \varepsilon_{ab}, \quad a = 1, \ldots, k, \ b = 1, \ldots, n_a, \tag{5.1}$$

where k is the number of small areas, $N = \sum_{a=1}^k n_a$, $\boldsymbol{x}_{ab}$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients, and v_a's and ε_{ab}'s are mutually independently distributed as $v_a \sim \mathcal{N}(0, \sigma_v^2)$ and $\varepsilon_{ab} \sim \mathcal{N}(0, \sigma^2)$ for unknown σ_v^2 and σ^2. Let $\boldsymbol{X}_a = (\boldsymbol{x}_{a1}, \ldots, \boldsymbol{x}_{a,n_a})'$, $\boldsymbol{X} = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_k)'$, $\boldsymbol{y}_a = (y_{a1}, \ldots, y_{a,n_a})'$, $\boldsymbol{y} = (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_k)'$ and let $\boldsymbol{\epsilon}$ be similarly defined. Let $\boldsymbol{v} = (v_1, \ldots, v_k)'$ and $\boldsymbol{Z} = \text{block diag}(\boldsymbol{j}_1, \ldots, \boldsymbol{j}_k)$ for $\boldsymbol{j}_a = (1, \ldots, 1)' \in \boldsymbol{R}^{n_a}$. Then, the model is expressed in vector notations as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{\epsilon}$.

To estimate σ^2 and σ_v^2, we use the Prasad-Rao estimators, which are given as follows: Let $S = \boldsymbol{y}'(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$ and $S_1 = \boldsymbol{y}'(\boldsymbol{E} - \boldsymbol{E}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{E}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E})\boldsymbol{y}$ where $\boldsymbol{E} = \text{block diag}(\boldsymbol{E}_1, \ldots, \boldsymbol{E}_k)$ for $\boldsymbol{E}_a = \boldsymbol{I}_a - n_a^{-1}\boldsymbol{J}_a$. Then, unbiased estimators of σ^2 and σ_v^2 are given by

$$\hat{\sigma}^{2U} = S_1/(N - k - p) \quad \text{and} \quad \hat{\sigma}_v^{2U} = \{S - (N - p)\hat{\sigma}^{2U}\}/N^*,$$

where $N_* = N - \mathrm{tr}\,\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{a=1}^{k} n_a^2 \overline{\boldsymbol{x}}_a \overline{\boldsymbol{x}}_a'\}$. Since $\hat{\sigma}_v^{2U}$ takes a negative value with a positive probability, it may be reasonable to use the truncated estimator $\hat{\sigma}_v^{2TR} = \max\{\hat{\sigma}_v^{2U}, 0\}$.

For estimator $\hat{\sigma}^{2U}$ and $\hat{\sigma}_v^{2TR}$ given above, Model 2 in (2.2) is described as

$$y_{ab}^* = \boldsymbol{x}_{ab}'\widehat{\boldsymbol{\beta}}(\hat{\sigma}^{2U}, \hat{\sigma}_v^{2TR}) + v_a^* + \varepsilon_{ab}^*, \quad a = 1, \ldots, k,\ b = 1, \ldots, n_a, \tag{5.2}$$

where v_a^*'s and ε_{ab}^*'s are mutually independently distributed as $v_a^* \sim \mathcal{N}(0, \hat{\sigma}_v^{2TR})$ and $\varepsilon_{ab}^* \sim \mathcal{N}(0, \hat{\sigma}^{2U})$. The estimators $\hat{\sigma}^{2U*}$, $\hat{\sigma}_v^{2TR*}$ and $\widehat{\boldsymbol{\beta}}^*(\hat{\sigma}^{2U*}, \hat{\sigma}_v^{2TR*})$ can be obtained from y_{ab}^*'s by using the same techniques used to obtain $\hat{\sigma}^{2U}$, $\hat{\sigma}_v^{2TR}$ and $\widehat{\boldsymbol{\beta}}(\hat{\sigma}^{2U}, \hat{\sigma}_v^{2TR})$.

When we want to estimate the mean $\mu_s = \overline{\boldsymbol{x}}_s'\boldsymbol{\beta} + v_s$ of the s-th small area, the EBLUP is $\widehat{\mu}_s^{EB} = \overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}(\hat{\sigma}^{2U}, \hat{\sigma}_v^{2TR}) + \big(n_s\hat{\sigma}_v^{2TR}/(\hat{\sigma}^{2U} + n_s\hat{\sigma}_v^{2TR})\big)(\overline{y}_s - \overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}(\hat{\sigma}^{2U}, \hat{\sigma}_v^{2TR}))$ for $\overline{\boldsymbol{x}}_s = \sum_{j=1}^{n_s} \boldsymbol{x}_{sj}/n_s$. Then, the functions $g_1(\sigma^2, \sigma_v^2)$ and $g_2(\sigma^2, \sigma_v^2)$ are expressed as $g_1(\sigma^2, \sigma_v^2) = \sigma^2\sigma_v^2(\sigma^2 + n_s\sigma_v^2)^{-1}$ and $g_2(\sigma^2, \sigma_v^2) = (\sigma^2)^2(\sigma^2 + n_s\sigma_v^2)^{-2}\overline{\boldsymbol{x}}_s'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_s$. Then, the estimator of MSE of EBLUP and the confidence intervals based on the parametric bootstrap method are given in (2.9), (2.12) and (2.13), where in this model,

$$\overline{g}_3^* = n_s E_*\big[\{\hat{\sigma}_v^{2TR*}(\hat{\sigma}^{2U*} + n_s\hat{\sigma}_v^{2TR*})^{-1} - \hat{\sigma}_v^{2TR}(\hat{\sigma}^{2U} + n_s\hat{\sigma}_v^{2TR})^{-1}\}^2(\hat{\sigma}^{2U*} + n_s\hat{\sigma}_v^{2TR})\big|\boldsymbol{y}\big].$$

Also, AIC and conditional AIC are given in (2.17) and (2.20).

5.2 Posted land price data

We now apply the proposed procedures to the posted land price data along the Keikyu train line which connects the suburbs in Kanagawa prefecture to the Tokyo metropolitan area. Since those who live in the suburbs take this line to work or study in Tokyo on weekdays, it may be expected that the land price depends on the distance from Tokyo.

A data set of the posted land price data in 2001 and their covariates are available for 48 stations on the Keikyu train line, and we consider each station as a small area, namely, $k = 48$. For the a-th station, there are data of n_a land spots, where the average of n_a's is 3.73. For $b = 1, \ldots, n_a$, we use five kinds of observations y_{ab}, TRN_a, DST_{ab}, $FOOT_{ab}$ and FAR_{ab}, where y_{ab} denotes the value of the posted land price (Yen in the hundred of thousands) per m^2 of the b-th spot, TRN_a is the time to take by train from the station a to the Tokyo station around 8:30 in the morning, DST_{ab} is the geographical distance from the spot b to the nearby station a, $FOOT_{ab}$ is the time to take on foot from the spot b to the nearby station a and FAR_{ab} denotes the floor-area ratio of the spot b. As regressor variables, we consider nine variables FAR_{ab}, TRN_a, TRN_a^2, DST_{ab}, DST_{ab}^2, $FOOT_{ab}$, $FOOT_{ab}^2$, $TRN_a \times DST_{ab}$ and $TRN_a \times FOOT_{ab}$, which are denoted by x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 and x_9. Also, the constant term is denoted by x_0. Then we can treat nested error regression model (5.1) with the regressor variables x_0-x_9 as a full model.

Table 5 reports values of AIC_1^* and $cAIC^*$ given in (2.17) and (2.20) for several candidate models, where the regressor variables which minimizes the information criteria are added to the model based on the forward stepwise selection. Among these candidate models, AIC_1^* selects both $\{x_0, x_1, x_2, x_3\}$ and $\{x_0, x_1, x_2, x_3, x_4\}$. Although the minimum of $cAIC^*$ is attained at the model with $\{x_0, x_1, x_2, x_3\}$, we can select the

Table 5: AIC and conditional AIC for selecting regressor variables in the posted land price data

m	x_i	AIC_1^*	$cAIC^*$	$\hat{\sigma}^{2U}$	$\hat{\sigma}_v^{2TR}$
1	x_0	566	512	0.796	1.188
2	x_0, x_1	461	415	0.449	0.481
3	x_0, x_1, x_2	407	396	0.452	0.131
4	x_0, x_1, x_2, x_3	395	389	0.456	0.079
5	x_0, x_1, x_2, x_3, x_4	395	390	0.459	0.071
6	x_0, x_1, x_2, x_3, x_4, x_9	398	393	0.462	0.070
7	x_0, x_1, x_2, x_3, x_4, x_8, x_9	465	425	0.466	0.371
10	all x_i's	489	451	0.470	0.416

variables $\{x_0, x_1, x_2, x_3, x_4\}$ since their difference is not significant. Since it may be better to explain the model including more regressor variables, we suggest the model with $\{x_0, x_1, x_2, x_3, x_4\}$, namely,

$$y_{ab} = \beta_0 + FAR_a\beta_1 + TRN_a\beta_2 + (TRN_a)^2\beta_3 + FOOT_a\beta_4 + v_a + \varepsilon_{ab},$$

for $a = 1, \ldots, k$ and $b = 1, \ldots, n_a$. The parameters are estimated by $\hat{\sigma}^2 = 0.45936$, $\hat{\sigma}_v^2 = 0.07154$ and $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4) = (5.1288, 6.3937 \times 10^{-3}, -0.1076, 7.0710 \times 10^{-4}, -8.1562 \times 10^{-5})$. This result demonstrates that the land prices are not only decreasing as a quadratic function of TRN_a, time to take from the nearby station to Tokyo station, but also decreasing in $FOOT_{ab}$, time to take from the land sopt to the nearby station.

We now estimate the average land price per m^2 around the s-th station, namely, $\mu_s = \overline{\boldsymbol{x}}_s'\boldsymbol{\beta} + v_s$ for $s = 1, \ldots, 48$ where $\overline{\boldsymbol{x}}_s$ is the mean of the regressor variables salected above. For some selected stations, the column of 'prediction' in Table 6 reports the values of n_s, $\overline{y}_s$, $\overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}$ and EBLUP $\widehat{\mu}_s^{EB}$, where all the values are given in Yen in the thousands. Also, the table reports the values of $\sqrt{mse_s^*}$, which is the square of estimates of MSE given in (2.9), and $\hat{\sigma}/\sqrt{n_s}$, which is the square of estimates of the conditional variance of $\overline{y}_s$ given v_s. The lower and upper end-points of the confidence interval I_2^{CEB*}, given in (2.13), with 95% confidence coefficient and the length of the interval are reported in the column of 'confidence interval' in Table 6.

From Table 6, it is revealed that for smaller n_s, the EBLUP $\widehat{\mu}_s^{EB}$ shrinks $\overline{y}_s$ much more toward $\overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}$, which results in smaller values $\sqrt{mse_s^*}$ than $\hat{\sigma}/\sqrt{n_s}$. For example, $\sqrt{mse_s^*}$ is about half of $\hat{\sigma}/\sqrt{n_s}$ for $n_s = 1$, but $\sqrt{mse_s^*}$ is equal to $\hat{\sigma}/\sqrt{n_s}$ for $n_s = 12$. Also, it is seen that the confidence intervals give shorter intervals for larger n_s. Thus, these observations show that the MSE estimator and the confidence interval based on the parametric bootstrap method work well in this example.

Table 6: Values of EBLUP, MSE estimates and confidence intervals for the average land price (Yen in the thousands)

		prediction					confidence interval		
s	n_s	$\overline{y}_s$	$\boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}$	$\widehat{\mu}_s^{EB}$	$\sqrt{mse_s^*}$	$\hat{\sigma}/\sqrt{n_s}$	lower	upper	length
4	1	569	458	473	31	67	400	546	146
8	2	401	426	420	28	47	356	484	128
12	2	316	358	348	27	47	287	409	122
16	2	285	317	309	27	47	249	369	121
20	1	270	300	296	28	67	231	361	130
24	3	312	350	338	26	39	280	395	114
28	5	246	235	240	24	30	187	292	105
32	12	247	177	223	19	19	181	265	84
36	3	187	209	202	26	39	145	260	114
40	3	285	270	275	26	39	216	333	117
44	5	185	188	187	24	30	134	240	106
48	4	159	203	186	28	33	120	252	132

6 Concluding Remarks

In this paper, we have suggested the procedures based on the parametric bootstrap methods in the estimation of the MSE of EBLUP, the confidence interval based on EBLUP, and variable selection problems based on AIC and conditional AIC. These procedures are not only easy to implement practically, but also justified theoretically to have second-order approximations. Also their performances have been investigated through simulation experiments and the empirical study, and it has been shown that the proposed procedures work well and are useful. Since existing procedures derived based on the Taylor series expansions are harder to compute in models with more parameters, the results obtained in this paper show that we can replace the existing procedures with the proposed ones. As shown in Kubokawa (2011), especially, the conditional AIC includes complicated terms in the penalty part, so that the parametric bootstrap method suggested in this paper is useful as an alternative procedure.

Acknowledgments. This research was supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 from Japan Society for the Promotion of Science.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, (B.N. Petrov and Csaki, F, eds.), 267-281, Akademia Kiado, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. System identification and time-series analysis. *IEEE Trans. Autom. Contr.*, **AC-19**, 716-723.

[3] Basu, R., Ghosh, J.K., and Mukerjee, R. (2003). Empirical Bayes prediction intervals in a normal regression model: higher order asymptotics. *Statist. Prob. Letters*, **63**, 197-203.

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

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

[6] Chatterjee, S., Lahiri, P., and Li, H. (2008). Parametric bootstrap approximation to the distribution of EBLUP and related prediction intervals in linear mixed models. *Ann. Statist.*, **36**, 1221-1245.

[7] Chen, S. and Lahiri, P. (2008). On mean squared prediction error estimation in small area estimation problems. *Commun. Statist.-Theory Methods*, **37**, 1792-1798.

[8] Das, K., Jiang, J. and Rao, J.N.K. (2004). Mean squared error of empirical predictor. *Ann. Statist.*, **32**, 818-840.

[9] Datta, G.S., Ghosh, M., Smith, D.D. and Lahiri, P. (2002). On an asymptotic theory of conditional and unconditional coverage probabilities of empirical Bayes confidence Intervals. *Scandinavian J. Statist.*, **29**, 139-152.

[10] Datta, G.S., Kubokawa, T., Molina, I. and Rao, J.N.K. (2011). Estimation of mean squared error of model-based small area estimators. *Test*, to appear.

[11] Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica*, **10**, 613-627.

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

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

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

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

[16] Hall, P. and Maiti, T. (2006b). On parametric bootstrap methods for small area prediction. *J. Royal Statist. Soc.*, **68**, 221-238.

[17] Kubokawa, T. (2010a). Correction of empirical Bayes confidence intervals in nested error regression models. *J. Korean Statist. Soc.*, **39**, 221-236.

[18] Kubokawa, T. (2010b) On Measuring Uncertainty of Small Area Estimators with Higher Order Accuracy Discussion Paper Series, CIRJE-F-754.

[19] Kubokawa, T. (2011). Conditional and unconditional methods for selecting variables in linear mixed model. *J. Multivariate Analysis*, **102**, 641-660.

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

[21] Liang, H., Wu, H.L., and Zou, G. (2008). A note on conditional AIC for mixed-effects models. *Biometrika*, **95**, 773-778.

[22] Pfeffermann, D. (2002). Small area estimation - new developments and directions. *Int. Statist. Rev.*, **70**, 125-143.

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

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

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

[26] Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun. Statist. - Theory Methods*, **1**, 13-26.

[27] Vaida, F., and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, **92**, 351-370.