CIRJE-F-754

On Measuring Uncertainty of Small Area Estimators with Higher Order Accuracy

Tatsuya Kubokawa
University of Tokyo

July 2010

On Measuring Uncertainty of Small Area Estimators with Higher Order Accuracy

Tatsuya Kubokawa*
University of Tokyo

July 28, 2010

Abstract

The empirical best linear unbiased predictor (EBLUP) or the empirical Bayes estimator (EB) in the linear mixed model is recognized useful for the small area estimation, because it can increase the estimation precision by using the information from the related areas. Two of the measures of uncertainty of EBLUP is the estimation of the mean squared error (MSE) and the confidence interval, which have been studied under the second-order accuracy in the literature. This paper provides the general analytical results for these two measures in the unified framework, namely, we derive the conditions on the general consistent estimators of the variance components to satisfy the third-order accuracy in the MSE estimation and the confidence interval in the general linear mixed normal models. Those conditions are shown to be satisfied by not only the maximum likelihood (ML) and restricted maximum likelihood (REML), but also the other estimators including the Prasad-Rao and Fay-Herriot estimators in specific models.

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

1 Introduction

The linear mixed models (LMM) and the empirical best linear unbiased predictor (EBLUP) or the empirical Bayes estimator (EB) induced from LMM have been studied for a long time in the literature. Especially, they have been recognized in recent years as useful tools in small area estimation. Small area refers to a small geographical area or a group for which little information is obtained from the sample survey, and the direct estimator based only on the data from a given small area is likely to be unreliable because only a few

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

observations are available from the small area. To increase the precision of the estimate, relevant supplementary information such as data from other related small areas is used via suitable linking models. The typical models used for the small area estimation are the Fay-Herriot model and the nested error regression model (NERM), and the model-based estimates including EBLUP or EB are found very useful as 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 (1999, 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 accuracy have been derived based on the Taylor series expansion by Kackar and Harville (1984), Prasad and Rao (1990), Harville and Jeske (1992), Datta and Lahiri (2000), Datta, Rao and Smith (2005) and Das, Jiang and Rao (2004). For some recent results including jackknife and bootstrap methods, see Lahiri and Rao (1995), Butar and Lahiri (2003), Hall and Maiti (2006a), Slud and Maiti (2006) and Chen and Lahiri (2008). Another method for measuring uncertainty of EBLUP is to provide a confidence interval based on EBLUP, and the confidence intervals which satisfy the nominal confidence level with the second-order accuracy have been derived based on the Taylor series expansion by Datta, Ghosh, Smith and Lahiri (2002), Basu, Ghosh and Mukerjee (2003) and Kubokawa (2010). Recently, Hall and Maiti (2006b) and Chatterjee, Lahiri and Li (2008) developed the method based on parametric bootstrap.

In this paper, we treat the problem of predicting the general linear combination of the regression coefficients and the random effects in the general linear mixed model under the normality assumption, and we construct the asymptotically unbiased estimator of MSE of EBLUP and the confidence interval based on EBLUP, both of which guarantee the third-order accuracy in the unified framework. The results obtained in this paper extend the results given in the literature to the following four directions: (1) treating the two problems of the MSE estimation and the confidence interval in the unified setup, (2) the third-order accuracy, (3) the general LMM, and (4) the general consistent estimators of unknown parameters embedded in the covariance matrices.

Concerning the points (1) and (2), the MSE estimation and the confidence intervals have been treated separately in the literature, and the results given in the literature have been derived under the second-order accuracy.

Concerning the point (3), Datta and Lahiri (2000) dealt with a general linear mixed model where the covariance matrices of the random effects and the error terms are assumed to be linear in the unknown parameters, denoted by $\boldsymbol{\theta}$. This assumption is reasonable when the elements of $\boldsymbol{\theta}$ are variance components, but it may be restrictive because the covariance matrices are non-linear functions of $\boldsymbol{\theta}$ when the random effects or error terms have autoregressive structures like $AR(1)$. This difference in the setup of the covariance matrices appears in the bias of the restricted maximum likelihood estimator (REML) of $\boldsymbol{\theta}$, namely, the second-order bias of REML vanishes when the covariance matrices are liner in $\boldsymbol{\theta}$, but it does not vanish without the linearity assumption. Das, *et al*. (2004) handled the general LMM without assuming the linearity of covariance matrices in $\boldsymbol{\theta}$ and derived the general asymptotically unbiased estimator of MSE with the second-order accuracy, where their estimators of $\boldsymbol{\theta}$ are given as solutions of score-like equations which include ML and REML.

In this paper, we consider the general consistent estimators of $\boldsymbol{\theta}$ in the general LMM without assuming that the covariance matrices are linear in $\boldsymbol{\theta}$. Then, we develop unified conditions on the general consistent estimators of $\boldsymbol{\theta}$ under which the derived estimator estimates the MSE of EBLUP asymptotically unbiasedly with the third-order accuracy and the constructed confidence interval based on EBLUP satisfies the nominal confidence level with the third-order accuracy. A feature of this paper is that the Stein identity given by Stein (1981) is used to evaluate the MSE of EBLUP, which enables us to generate the general conditions on estimators of $\boldsymbol{\theta}$.

The paper is organized as follows: The main results on the MSE estimation and confidence intervals are given in Section 2. The conditions and the results for the second-order approximation are described in Subsection 2.2, and those for the third-order approximation are provided in Subsection 2.3. Two simple and instructive examples are given in Subsection 2.4. The second-order and third-order expansions of ML and REML estimators of $\boldsymbol{\theta}$ are studied in in Section 3. The third-order approximations in the MSE estimation and confidence intervals based on ML and REML are applied to some specific models including the Fay-Herriot model, the nested error regression model and a basic area level model proposed by Rao and Yu (1994) for combining the time-series and cross-sectional data. The proofs of the main results are given in Section 4.

Finally, it should be remarked that the validity of the asymptotic expansions will not be discussed here. All the results are based on major terms obtained by Taylor series expansions as used in Datta and Lahiri (2000). To establish the validity in the third-order approximation, we need more appropriate conditions like those given in Das, *et al.* (2004) who gave the rigorous proofs in the second-order approximation.

2 MSE Estimation and Confidence Interval Based on EBLUP

2.1 The model and notations

Consider the general linear mixed model

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

where $\boldsymbol{y}$ is an $N \times 1$ observation vector of the response variable, $\boldsymbol{X}$ 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 $\boldsymbol{v} \sim \mathcal{N}_M(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\theta}))$ and $\boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \boldsymbol{R}(\boldsymbol{\theta}))$, 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

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

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be $p \times 1$ and $M \times 1$ vectors of fixed constants, and suppose that we want to estimate the scalar quantity $\mu = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{v}$. Since the conditional distribution of $\boldsymbol{v}$ given $\boldsymbol{y}$ is given by

$$\boldsymbol{v}|\boldsymbol{y} \sim \mathcal{N}_M\big(\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), (\boldsymbol{G}(\boldsymbol{\theta})^{-1} + \boldsymbol{Z}'\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{Z})^{-1}\big), \tag{2.2}$$

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

$$\begin{aligned}\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) =& E[\mu|\boldsymbol{y}] = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ =& \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),\end{aligned} \tag{2.3}$$

where $\boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{Z}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{b}$. This can be interpreted as the Bayes estimator of μ in the Bayesian context. The generalized least squares estimator of $\boldsymbol{\beta}$ for given $\boldsymbol{\theta}$ is given by

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

which is substituted into $\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta})$ to get the estimator

$$\widehat{\mu}^{EB}(\boldsymbol{\theta}) = \widehat{\mu}^B(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{a}'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})). \tag{2.4}$$

This estimator is the best linear unbiased predictor (BLUP) of μ. When an estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{y})$ 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. We give a higher order approximation to MSE of EBLUP, an asymptotic unbiased estimator of the MSE and a confidence interval based on EBLUP with higher order accuracy.

We here explain the notations used through the paper. Let $\mathcal{C}_{\boldsymbol{\theta}}^{[k]}$ denote a set of k times continuously differentiable functions with respect to $\boldsymbol{\theta}$. As partial derivatives with respect to $\boldsymbol{\theta}$, we use the notations defined by

$$\begin{aligned}\boldsymbol{A}_{(i)}(\boldsymbol{\theta}) =& \partial_i \boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i}, \quad \boldsymbol{A}_{(ij)}(\boldsymbol{\theta}) = \partial_{ij}\boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial^2 \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}, \\ \boldsymbol{A}_{(ijk)}(\boldsymbol{\theta}) =& \partial_{ijk}\boldsymbol{A}(\boldsymbol{\theta}) = \frac{\partial^3 \boldsymbol{A}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad \frac{\partial}{\partial \boldsymbol{\theta}} = (\partial_1, \ldots, \partial_q)', \quad \boldsymbol{\nabla}_y = \frac{\partial}{\partial \boldsymbol{y}},\end{aligned}$$

for matrices or vectors $\boldsymbol{A}(\boldsymbol{\theta})$, where we use the same notations for scalars. For $0 \leq i, j, k \leq q$, let $\lambda_1 \leq \cdots \leq \lambda_N$ be the eigenvalues of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and let those of $\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}_{(ij)}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{(ijk)}(\boldsymbol{\theta})$ be λ_a^i, λ_a^{ij} and λ_a^{ijk} for $a = 1, \ldots, N$ respectively, where $|\lambda_1^i| \leq \cdots \leq |\lambda_N^i|$, $|\lambda_1^{ij}| \leq \cdots \leq |\lambda_N^{ij}|$ and $|\lambda_1^{ijk}| \leq \cdots \leq |\lambda_N^{ijk}|$.

2.2 Second-order approximation

[1] Approximation of MSE. We begin by the second-order approximation to MSE of EBLUP. To this end, we assume the following conditions for large N and $1 \leq i, j, k \leq q$:

(A1) The elements of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{G}(\boldsymbol{\theta})$, $\boldsymbol{R}(\boldsymbol{\theta})$, $\boldsymbol{a}$ and $\boldsymbol{b}$ are uniformly bounded, and p, q and M are bounded. The matrix $\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}$ is positive definite and $\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}/N$ converges to a positive definite matrix;

(A2) (i) $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$, and $\lim_{N\to\infty} \lambda_1 > 0$, $\lim_{N\to\infty} \lambda_N < \infty$, $\lim_{N\to\infty} |\lambda_N^i| < \infty$ and $\lim_{N\to\infty} |\lambda_N^{ij}| < \infty$. (ii) $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$, and $(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}(\boldsymbol{\theta}) = O_p(1)$, $(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(i)}(\boldsymbol{\theta}) = O_p(1)$, $(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(ij)}(\boldsymbol{\theta}) = O_p(1)$ and $\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{s}_{(j)}(\boldsymbol{\theta}) = O(1)$.

(A3) $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{y}) = (\hat{\theta}_1, \ldots, \hat{\theta}_q)'$ 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) $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is expanded as

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**} + O_p(N^{-3/2}), \tag{2.5}$$

where $\widehat{\boldsymbol{\theta}}^* = O_p(N^{-1/2})$ and $\widehat{\boldsymbol{\theta}}^{**} = O_p(N^{-1})$. For $\widehat{\boldsymbol{\theta}}^* = (\hat{\theta}_1^*, \ldots, \hat{\theta}_q^*)'$, it is assumed that $\hat{\theta}_i^*$ satisfies that (i) $E[\hat{\theta}_i^*] = O(N^{-1})$ and (ii) $\boldsymbol{s}_{(j)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\nabla}_y\hat{\theta}_i^* = O_p(N^{-1})$.

The assumption (A1) implies that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{\beta} = O_p(N^{-1/2})$, and (A1) and (A2) (i) mean that $\widehat{\boldsymbol{\beta}}_{(i)}(\boldsymbol{\theta}) = O_p(N^{-1/2})$ and $\widehat{\boldsymbol{\beta}}_{(ij)}(\boldsymbol{\theta}) = O_p(N^{-1/2})$. Also, (A1) and (A2) imply that $\widehat{\mu}^{EB}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$, $\widehat{\mu}_{(i)}^{EB}(\boldsymbol{\theta}) = O_p(1)$ and $\widehat{\mu}_{(ij)}^{EB}(\boldsymbol{\theta}) = O_p(1)$.

Under the above assumptions, we can derive the second-order approximation to MSE. Define $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$ and $g_3^*(\boldsymbol{\theta})$ by

$$\begin{aligned} g_1(\boldsymbol{\theta}) =& \boldsymbol{b}'(\boldsymbol{G}(\boldsymbol{\theta})^{-1} + \boldsymbol{Z}'\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{Z})^{-1}\boldsymbol{b}, \\ g_2(\boldsymbol{\theta}) =& (\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta}))'(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}(\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta})), \\ g_3^*(\boldsymbol{\theta}) =& \mathrm{tr}\Big[\Big(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\Big)\boldsymbol{\Sigma}(\boldsymbol{\theta})\Big(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\Big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\Big], \end{aligned} \tag{2.6}$$

for $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*) = E[(\widehat{\boldsymbol{\theta}}^* - E[\widehat{\boldsymbol{\theta}}^*])(\widehat{\boldsymbol{\theta}}^* - E[\widehat{\boldsymbol{\theta}}^*])']$. It can be seen that $g_1(\boldsymbol{\theta})$ is rewritten as

$$g_1(\boldsymbol{\theta}) = \boldsymbol{b}'\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{b} - \boldsymbol{s}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{s}(\boldsymbol{\theta}). \tag{2.7}$$

Theorem 2.1 *Assume the conditions* (A1)-(A4). *Then the MSE of* $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$ *is approximated as*

$$MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3^*(\boldsymbol{\theta}) + O(N^{-3/2}). \tag{2.8}$$

All the proofs of theorems given in this section will be given in Section 4.

[2] Approximated unbiased estimator of MSE. We next provide an asymptotically unbiased estimator of $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ with second-order accuracy. Define $g_{11}(\boldsymbol{\theta})$ and $g_{12}(\boldsymbol{\theta})$ by

$$\begin{aligned} g_{11}(\boldsymbol{\theta}) =& \Big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big)' E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}], \\ g_{12}(\boldsymbol{\theta}) =& \frac{1}{2}\mathrm{tr}\Big[\boldsymbol{B}(\boldsymbol{\theta})\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\Big], \end{aligned} \tag{2.9}$$

where the (i,j)-th element of $\boldsymbol{B}(\boldsymbol{\theta})$ is given by

$$(\boldsymbol{B}(\boldsymbol{\theta}))_{i,j} = (\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta}))'(\partial_{ij}\boldsymbol{G}(\boldsymbol{\theta}))(\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta})) + \boldsymbol{s}(\boldsymbol{\theta})'(\partial_{ij}\boldsymbol{R}(\boldsymbol{\theta}))\boldsymbol{s}(\boldsymbol{\theta}). \tag{2.10}$$

It is noted that $g_{12}(\boldsymbol{\theta})=0$ when $\boldsymbol{G}$ and $\boldsymbol{R}$ are matrices of linear functions of $\boldsymbol{\theta}$. Define $mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ by

$$mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\widehat{\boldsymbol{\theta}}) + g^{\#}(\widehat{\boldsymbol{\theta}}), \tag{2.11}$$

where

$$g^{\#}(\boldsymbol{\theta}) = g_2(\boldsymbol{\theta}) + 2g_3^*(\boldsymbol{\theta}) - g_{11}(\boldsymbol{\theta}) - g_{12}(\boldsymbol{\theta}). \tag{2.12}$$

Since $E[\widehat{\boldsymbol{\theta}}^*] = O(N^{-1})$ from the condition (A4), it is noted that $g_{11}(\boldsymbol{\theta}) = O(N^{-1})$, so that $g^{\#}(\boldsymbol{\theta}) = O(N^{-1})$. The following theorem shows that $mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ is a second-order unbiased estimator of the MSE of $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$ under the additional assumption:

(B1) 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) $g^{\#}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[1]}$ and $\partial_i g^{\#}(\boldsymbol{\theta}) = O(N^{-1})$.

Theorem 2.2 *Assume the conditions* (A1)-(A4) *and* (B1). *Then,*

$$E[mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] = MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O(N^{-3/2}). \tag{2.13}$$

[3] Corrected confidence intervals. We construct a confidence interval of $\mu = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{b}'\boldsymbol{v}$ which satisfies the nominal confidence level with the second-order accuracy. Let $mse(\widehat{\boldsymbol{\theta}}) = mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\widehat{\boldsymbol{\theta}}) + g^{\#}(\widehat{\boldsymbol{\theta}})$ for $g^{\#}$ given in (2.12). Since $mse(\widehat{\boldsymbol{\theta}})$ is an asymptotically unbiased estimator of the MSE of the empirical Bayes estimator $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$, it is reasonable to consider the confidence interval of the form

$$I^{EB}(\widehat{\boldsymbol{\theta}}):\ \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) \pm z_{\alpha/2}\sqrt{mse(\widehat{\boldsymbol{\theta}})}. \tag{2.14}$$

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 to adjust the significance point $z_{\alpha/2}$ as $z_{\alpha/2}\{1+h(\widehat{\boldsymbol{\theta}})\}$ by using an appropriate correction function $h(\widehat{\boldsymbol{\theta}})$. That is, the corrected confidence interval is described as

$$I^{CEB}(\widehat{\boldsymbol{\theta}}):\ \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) \pm z_{\alpha/2}\left[1+h(\widehat{\boldsymbol{\theta}})\right]\sqrt{mse(\widehat{\boldsymbol{\theta}})}.$$

Here, we define the function $h(\boldsymbol{\theta})$ by

$$h(\boldsymbol{\theta}) = \frac{z_\alpha^2+1}{8g_1(\boldsymbol{\theta})^2}\mathrm{tr}\,[(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)]. \tag{2.15}$$

The following theorem shows that $I^{CEB}(\widehat{\boldsymbol{\theta}})$ satisfies the nominal confidence coefficient up to the second-order under the additional assumption:

(C1) $h(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[1]}$, $\partial_i h(\boldsymbol{\theta}) = O(N^{-1})$ for $1 \leq i \leq q$.

Theorem 2.3 *Assume the conditions* (A1)-(A4), (B1) *and* (C1). *Then,*

$$P[\mu \in I^{CEB}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + O(N^{-3/2}). \tag{2.16}$$

2.3 Third-order approximation

We now show that all the results given in Theorems 2.1, 2.2 and 2.3 hold with third-order accuracy under some additional assumptions. We here assume the following conditions:

(A5) (i) $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[3]}$, and $\lim_{N\to\infty} |\lambda_N^{ijk}| < \infty$. (ii) $\boldsymbol{s}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[3]}$, and $(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{s}_{(ijk)}(\boldsymbol{\theta}) = O_p(1)$ and $\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{s}_{(jk)}(\boldsymbol{\theta}) = O(1)$.

(A6) $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ can be further expanded as

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**} + \widehat{\boldsymbol{\theta}}^{***} + O_p(N^{-2}), \tag{2.17}$$

where $\widehat{\boldsymbol{\theta}}^* = O_p(N^{-1/2})$, $\widehat{\boldsymbol{\theta}}^{**} = O_p(N^{-1})$ and $\widehat{\boldsymbol{\theta}}^{***} = O_p(N^{-3/2})$. It is assumed that these satisfy the following: (i) $E[\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*] = O(N^{-2})$ and $E[\hat{\theta}_i^*\hat{\theta}_j^{**}] = O(N^{-2})$ and (ii) $\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\nabla}_y\hat{\theta}_j^{**} = O_p(N^{-3/2})$ and $E[\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\{\boldsymbol{\nabla}_y\boldsymbol{\nabla}_y'\hat{\theta}_i^*\}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{s}_{(j)}(\boldsymbol{\theta})\hat{\theta}_j^*] = O(N^{-2})$.

The assumptions (A1), (A2)(i) and (A5)(i) imply that $\widehat{\boldsymbol{\beta}}_{(ijk)}(\boldsymbol{\theta}) = O_p(N^{-1/2})$. Also, (A1), (A2) and (A5) imply that $\widehat{\mu}_{(ijk)}^{EB}(\boldsymbol{\theta}) = O_p(1)$.

Theorem 2.4 *Assume the conditions* (A1)-(A6)*. Then the MSE of* $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})$ *is approximated as*

$$MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3^*(\boldsymbol{\theta}) + O(N^{-2}). \tag{2.18}$$

To give an asymptotically unbiased estimator of $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))$ with the third-order accuracy, assume that

(B2) For $1 \le i,j,k,\ell \le q$, (i) $g_1(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[4]}$ and $\partial_{ijk\ell}g_1(\boldsymbol{\theta}) = O(1)$, (ii) $g^{\#}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$ and $\partial_{ij}g^{\#}(\boldsymbol{\theta}) = O(N^{-1})$ and (iii) $E[\widehat{\boldsymbol{\theta}}^{***}] = O_p(N^{-2})$.

Theorem 2.5 *Assume the conditions* (A1)-(A6)*,* (B1) *and* (B2)*. Then,*

$$E[mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] = MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) + O(N^{-2}). \tag{2.19}$$

Finally, assume that

(C2) $h(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$ and $\partial_{ij}h(\boldsymbol{\theta}) = O(N^{-1})$ for $1 \le i,j \le q$.

Theorem 2.6 *Assume the conditions* (A1)-(A6), (B1), (B2), (C1) *and* (C2)*. Then,*

$$P[\mu \in I^{CEB}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + O(N^{-2}). \tag{2.20}$$

We conclude this subsection with some remarks.

Remark 2.1 When the covariance matrix $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)$ and the bias $E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}]$ are approximated as $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*) = \boldsymbol{C}^* + O(N^{-3/2})$ and $E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}] = \boldsymbol{b}^* + O(N^{-3/2})$, we can replace $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)$ and $E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}]$ with $\boldsymbol{C}^*$ and $\boldsymbol{b}^*$, respectively, in Theorems 2.1 - 2.6.

Remark 2.2 The model treated by Datta and Lahiri (2000) is $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{v}_i + \boldsymbol{\epsilon}_i$ for $i = 1, \ldots, k$, where $\boldsymbol{v}_i \sim \mathcal{N}_{b_i}(\boldsymbol{0}, \boldsymbol{G}_i(\boldsymbol{\theta}))$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{n_i}(\boldsymbol{0}, \boldsymbol{R}_i(\boldsymbol{\theta}))$ and it is assumed that $\boldsymbol{R}_i(\boldsymbol{\theta}) = \sum_{j=0}^q \theta_j \boldsymbol{D}_{ij}\boldsymbol{D}'_{ij}$ and $\boldsymbol{G}_i(\boldsymbol{\theta}) = \sum_{j=0}^q \theta_j \boldsymbol{F}_{ij}\boldsymbol{F}'_{ij}$ for $\theta_0 = 1$ and known matrices $\boldsymbol{D}_{ij}$ and $\boldsymbol{F}_{ij}$. It is also assumed that the elements of $\boldsymbol{D}_{ij}$ and $\boldsymbol{F}_{ij}$ are uniformly bounded. For the other notations, see Datta and Lahiri (2000). They provided the corresponding results to Theorem 2.1 and 2.2. It can be seen that the conditions (A1)-(A3) satisfies conditions (a), (b), (c), (d), (f) of Datta and Lahiri (2000) except the conditions that $\sup_{1\le i\le k} n_i$ is bounded and that $k \to \infty$, which are implicitly assumed in the condition (A4) in this paper.

Remark 2.3 As mentioned below (2.9), the term $g_{12}(\boldsymbol{\theta}) = 2^{-1}\mathrm{tr}\left[\boldsymbol{B}(\boldsymbol{\theta})\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\right]$ does not appear under the condition

(A7) $\boldsymbol{G}(\boldsymbol{\theta})$ and $\boldsymbol{R}(\boldsymbol{\theta})$ are matrices of linear functions of $\boldsymbol{\theta}$,

since $(\boldsymbol{B}(\boldsymbol{\theta}))_{i,j} = (\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta}))'(\partial_{ij}\boldsymbol{G}(\boldsymbol{\theta}))(\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta})) + \boldsymbol{s}(\boldsymbol{\theta})'(\partial_{ij}\boldsymbol{R}(\boldsymbol{\theta}))\boldsymbol{s}(\boldsymbol{\theta})$. The model of Datta and Lahiri (2000) satisfies the condition (A7). Since most models studied in the literature satisfy (A7), the term $g_{12}(\boldsymbol{\theta})$ has not explicitly appeared in the literature except Das, *et al.* (2004), who treated the model with general covariance structures, and the term $g_{12}(\boldsymbol{\theta})$ is implicitly included by $\Delta_2(\sigma)$ given in (4.5) of their paper.. When $\boldsymbol{G}(\boldsymbol{\theta})$ or $\boldsymbol{R}(\boldsymbol{\theta})$ have time-series or longitudinal structures, however, the term $g_{12}(\boldsymbol{\theta})$ cannot be ignored. For this point, see Section 3.5. The models for analyzing time-series and cross-section data have been actively and extensively studied in the literature. Of these, Rao and Yu (1994) and Datta, Kahiri and Maiti (2002) have provided the explicit forms of MSE estimators of EBLUP. Rao and Yu (2002) derived the MSE estimator in the case that the AR(1) coefficient ρ is known, and used the plug-in estimator when ρ is unknown. Datta, *et al.* (2002) treated a random walk model, namely the case of $\rho = 1$. Thus, the term $g_{12}(\boldsymbol{\theta})$ does not appear in these papers, although both handled time-series structures.

Remark 2.4 It is noted that the validity of the asymptotic expansions will not be discussed here. All the results in this paper are based on major terms obtained by Taylor series expansions which is a similar method as used in Datta and Lahiri (2000). The validity of the second-order approximations in MSE and its estimation has been shown by Prasad and Rao (1990) for unbiased estimators of $\boldsymbol{\theta}$ in some specific models, and by Das, *et al.* (2004) for ML and REML in the general LMM. Although this paper provides the third order approximations without the validity, we need more conditions and much more steps for establishing the validity of the third-order approximations.

Remark 2.5 The corrected function $h(\boldsymbol{\theta})$ given in (2.15) includes $g_1(\boldsymbol{\theta})$ in the denominator, and this may cause the instability of the corrected confidence interval $I^{CEB}(\widehat{\boldsymbol{\theta}})$ near $\boldsymbol{\theta} = \boldsymbol{0}$. For example, as given in Example 2.2, we have $g_1(\theta_1, \theta_2) = \theta_1\theta_2/(\theta_1 + n_s\theta_2)$ in the nested error regression model, where θ_1 and θ_2, respectively, are the 'within' and 'between' components of variance, and n_s is a sample size of a small area. When θ_2 is close to zero, the estimator $\hat{\theta}_2$ and $g_1(\hat{\theta}_1, \hat{\theta}_2)$ take values near zero, which leads to the instability of the confidence interval. One method for fixing this problem is to use the truncation of the estimator $\hat{\theta}_2$ as $\hat{\theta}_2^{TR} = \max\{\hat{\theta}_2, N^{-2/3}\}$, which was suggested in Kubokawa (2010), For the practical use of $I^{CEB}(\widehat{\boldsymbol{\theta}})$, we need such a modification of the estimator $\widehat{\boldsymbol{\theta}}$.

2.4 Instructive examples

In this section, the results given in the previous sections are applied to specific models, and the corresponding forms of the MSE estimators and the confidence intervals are derived.

Example 2.1 (Fay-Herriot model) As a simple basic area model, we consider the Fay-Herriot model described by

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i + \varepsilon_i, \quad i = 1, \ldots, k,$$

where k is the number of small areas, $\boldsymbol{x}_i$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients, and v_i's and ε_i's are mutually independently distributed random errors such that $v_i \sim \mathcal{N}(0, \theta)$ and $\varepsilon_i \sim \mathcal{N}(0, d_i)$. 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 $\boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_k)$ and $N = k$. It is assumed that $\sup_{i\geq 1} d_i < \infty$, $\inf_{i\geq 1} d_i > 0$ and that $k \to \infty$.

When we want to estimate $\mu_s = \boldsymbol{x}_s'\boldsymbol{\beta} + v_s$, the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ used in Section 2 correspond to $\boldsymbol{a} = \boldsymbol{x}_s$ and $\boldsymbol{b} = (0, \ldots, 0, 1, 0, \ldots, 0)'$ such that $\boldsymbol{b}'\boldsymbol{v} = v_s$. The EBLUP or empirical Bayes estimator of μ_s is written as

$$\widehat{\mu}_s^{EB}(\hat{\theta}) = \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta}) + \big(\hat{\theta}/(\hat{\theta} + d_s)\big)(y_s - \boldsymbol{x}_s'\widehat{\boldsymbol{\beta}}(\hat{\theta})),$$

and the functions $g_1(\theta)$, $g_2(\theta)$, $g_3^*(\theta)$, $g_{11}(\theta)$ and $h(\theta)$ are expressed as $g_1(\theta) = \theta(\theta + d_s)^{-1}$, $g_2(\theta) = d_s^2(\theta + d_s)^{-2}\boldsymbol{x}_s'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}_s$,

$$\begin{aligned} g_3^*(\theta) =& d_s^2(\theta + d_s)^{-3} Var(\hat{\theta}^*), \\ g_{11}(\theta) =& d_s^2(\theta + d_s)^{-2} E[\hat{\theta}^* + \hat{\theta}^{**}], \\ h(\theta) =& \frac{z_\alpha^2 + 1}{8\theta^2(\theta + d_s)^2} Var(\hat{\theta}^*), \end{aligned} \tag{2.21}$$

and $g_{12}(\theta) = 0$. In this model, the conditions (A2) and (A5) given hold, and the conditions (A4) (ii) and (A6)(i) are rewritten as $\partial\hat{\theta}^*/\partial y_s = O_p(N^{-1})$, $\partial\hat{\theta}^{**}/\partial y_s = O_p(N^{-3/2})$ and $E[(\partial^2\hat{\theta}^*/\partial y_s^2)\hat{\theta}^*] = O(N^{-2})$. Assume the conditions (A1), (A3), (A4) and (A6) and that $E[\hat{\theta}^{***}] = O(N^{-2})$, $Var(\hat{\theta}^*) \in \mathcal{C}_\theta^{[2]}$ and $E[\hat{\theta}^* + \hat{\theta}^{**}] \in \mathcal{C}_\theta^{[2]}$. Then, we can obtain the third-order approximations given in Theorems 2.4, 2.5 and 2.6.

[Prasad-Rao estimator] A simple estimator of θ is the unbiased estimator suggested by Prasad and Rao (1990) given by $\hat{\theta}^U = (k - p)^{-1}(\boldsymbol{y}'\boldsymbol{W}_0\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{W}_0])$ for $\boldsymbol{W}_0 = \boldsymbol{I}_k - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. In this case, $\hat{\theta}^U - \theta = \hat{\theta}^{U*} = (k - p)^{-1}\mathrm{tr}\,[\boldsymbol{W}_0(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]$, and it is easy to see that $E[\hat{\theta}^{U*}] = 0$ and $Var[\hat{\theta}^{U*}] = 2k^{-2}\mathrm{tr}\,\boldsymbol{\Sigma}^2 + O(k^{-2})$ as described in Prasad and Rao (1990). Since all the conditions other than (A1) are satisfied, from Remark 2.1, we get the results in Theorems 2.4, 2.5 and 2.6 under (A1). ■

Example 2.2 (Nested error regression model(NERM)) The model we next handle is the nested error regression model (NERM) given by

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

where k is the number of small areas, $N = \sum_{i=1}^{k} n_i$, $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients, and v_i's and ε_{ij}'s are mutually independently distributed as $v_i \sim \mathcal{N}(0, \sigma_v^2)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Here, σ_v^2 and σ^2 are referred to as, respectively, 'between' and 'within' components of variance, and both are unknown. Let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{i,n_i})'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_k')'$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i,n_i})'$, $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_k')'$ and let $\boldsymbol{\epsilon}$ be similarly defined. Let $\boldsymbol{v} = (v_1, \ldots, v_k)'$ and $\boldsymbol{Z} = \text{block diag}(\boldsymbol{j}_{n_1}, \ldots, \boldsymbol{j}_{n_k})$ for $\boldsymbol{j}_{n_i} = (1, \ldots, 1)' \in \boldsymbol{R}^{n_i}$. Then, the model is expressed in vector notations as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v} + \boldsymbol{\epsilon}$. It is assumed that $\sup_{i \geq 1} n_i < \infty$ and that $k \to \infty$.

We want to estimate the mean $\mu_s = \overline{\boldsymbol{x}}_s'\boldsymbol{\beta} + v_s$ of the s-th small area for $\overline{\boldsymbol{x}}_s = \sum_{j=1}^{n_s} \boldsymbol{x}_{sj}/n_s$. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ used in Section 2 correspond to $\boldsymbol{a} = \overline{\boldsymbol{x}}_s$ and $\boldsymbol{b} = (0, \ldots, 0, 1, 0, \ldots, 0)'$ such that $\boldsymbol{b}'\boldsymbol{v} = v_s$. Also, $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ and $\boldsymbol{\Sigma}$ correspond to $\theta_1 = \sigma^2$, $\theta_2 = \sigma_v^2$ and $\boldsymbol{\Sigma} = \text{block diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k)$ for $\boldsymbol{\Sigma}_i = \theta_1 \boldsymbol{I}_{n_i} + \theta_2 \boldsymbol{J}_{n_i}$, $\boldsymbol{J}_{n_i} = \boldsymbol{j}_{n_i}\boldsymbol{j}_{n_i}'$, $\boldsymbol{I}_{n_i}$ being the $n_i \times n_i$ identity matrix. Since $\boldsymbol{\Sigma}_s^{-1} = \theta_1^{-1}(\boldsymbol{I}_{n_s} - \theta_2/(\theta_1 + n_s\theta_2)\boldsymbol{J}_{n_s})$ and $\boldsymbol{\Sigma}_s^{-1}\boldsymbol{j}_{n_s} = (\theta_1 + n_s\theta_2)^{-1}\boldsymbol{j}_{n_s}$, $\boldsymbol{s}(\boldsymbol{\theta})'$ is expressed as $\boldsymbol{s}(\boldsymbol{\theta})' = (\theta_2/(\theta_1 + n_s\theta_2)(\boldsymbol{0}_{n_1}', \ldots, \boldsymbol{0}_{n_{s-1}}', \boldsymbol{j}_{n_s}', \boldsymbol{0}_{n_{s+1}}', \ldots, \boldsymbol{0}_{n_k}')'$ where $\boldsymbol{0}_{n_j}$ is the n_j-dimensional zero vector such that $\boldsymbol{0}' = (\boldsymbol{0}_{n_1}', \ldots, \boldsymbol{0}_{n_s}', \ldots, \boldsymbol{0}_{n_k}')$. From this expression, $\partial_1 \boldsymbol{s}(\boldsymbol{\theta})$ and $\partial_2 \boldsymbol{s}(\boldsymbol{\theta})$ can be derived. Then, the EBLUP or empirical Bayes estimator of μ_s is written as

$$\widehat{\mu}_s^{EB}(\widehat{\boldsymbol{\theta}}) = \overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) + \big(n_s\hat{\theta}_2/(\hat{\theta}_1 + n_s\hat{\theta}_2)\big)(\overline{y}_s - \overline{\boldsymbol{x}}_s'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})),$$

and the functions $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$, $g_3^*(\boldsymbol{\theta})$, $g_{11}(\boldsymbol{\theta})$ and $h(\boldsymbol{\theta})$ are expressed as $g_1(\boldsymbol{\theta}) = \theta_1\theta_2(\theta_1 + n_s\theta_2)^{-1}$, $g_2(\boldsymbol{\theta}) = \theta_1^2(\theta_1 + n_s\theta_2)^{-2}\overline{\boldsymbol{x}}_s'(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_s$,

$$\begin{aligned} g_3^*(\boldsymbol{\theta}) =& n_s(\theta_1 + n_s\theta_2)^{-3}(-\theta_2, \theta_1)\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)(-\theta_2, \theta_1)', \\ g_{11}(\boldsymbol{\theta}) =& (\theta_1 + n_s\theta_2)^{-2}(n_s\theta_2^2, \theta_1^2)E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}], \\ h(\boldsymbol{\theta}) =& \frac{z_\alpha^2 + 1}{8(\theta_1\theta_2)^2(\theta_1 + n_s\theta_2)^2}(n_s\theta_2^2, \theta_1^2)\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)(n_s\theta_2^2, \theta_1^2)', \end{aligned} \tag{2.22}$$

and $g_{12}(\boldsymbol{\theta}) = 0$. In this model, conditions (A2) and (A5) hold. It is also noted that in (A4) and (A6), the conditions $\boldsymbol{s}_{(j)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\nabla}_y\hat{\theta}_i^* = O_p(N^{-1})$, $\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{\nabla}_y\hat{\theta}_j^{**} = O_p(N^{-3/2})$ and $E[\boldsymbol{s}_{(i)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\{\boldsymbol{\nabla}_y\boldsymbol{\nabla}_y'\hat{\theta}_i^*\}\boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{s}_{(j)}(\boldsymbol{\theta})\hat{\theta}_j^*] = O(N^{-2})$ are rewritten as $\boldsymbol{\nabla}_s\hat{\theta}_i^* = O_p(N^{-1})$, $\boldsymbol{\nabla}_s\hat{\theta}_i^{**} = O_p(N^{-3/2})$ and $E[\{\boldsymbol{\nabla}_s\boldsymbol{\nabla}_s'\hat{\theta}_i^*\}\hat{\theta}_j^*] = O(N^{-2})$, respectively, for $1 \leq i, j \leq 2$ and $\boldsymbol{\nabla}_s = \partial/\partial\boldsymbol{y}_s$. Assume the conditions (A1), (A3), (A4) and (A6) and that $E[\widehat{\boldsymbol{\theta}}^{***}] = O(N^{-2})$, $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$ and $E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}] \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$. Then, we can obtain the third-order approximations given in Theorems 2.4, 2.5 and 2.6.

[Prasad-Rao estimator] Prasad and Rao (1990) suggested estimators based on unbiased estimators of $\theta_1 = \sigma^2$ and $\theta_2 = \sigma_v^2$. 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 $\boldsymbol{E}_i = \boldsymbol{I}_{n_i} - n_i^{-1}\boldsymbol{J}_{n_i}$. Then, unbiased estimators of θ_1 and θ_2 suggested by Prasad and Rao (1990) are

$$\hat{\theta}_1^U = S_1/(N - k - p) \quad \text{and} \quad \hat{\theta}_2^U = \{S - (N - p)\hat{\theta}_1^U\}/N^*,$$

where $N_* = N - \text{tr}\,\{(\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{i=1}^{k} n_i^2\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'\}$. In this case, $\hat{\theta}_i^U - \theta_i = \hat{\theta}_i^{U*}$ for $i = 1, 2$, and it is easy to see that $E[\hat{\theta}_1^{U*}] = 0$, $E[\hat{\theta}_2^{U*}] = 0$, $\boldsymbol{\nabla}_i S_1 = 2(\boldsymbol{E}_i\boldsymbol{y}_i - \boldsymbol{E}_i\boldsymbol{X}_i(\boldsymbol{X}'\boldsymbol{E}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}\boldsymbol{y}) =$

$O_p(1)$, $\nabla_i \nabla_i' S_1 = 2(\boldsymbol{E}_i - \boldsymbol{E}_i \boldsymbol{X}_i (\boldsymbol{X}'\boldsymbol{E}\boldsymbol{X})^{-1} \boldsymbol{X}_i' \boldsymbol{E}_i) = O(1)$, $\nabla_i S = 2(\boldsymbol{y}_i - \boldsymbol{X}_i (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y}) = O_p(1)$, $\nabla_i \nabla_i' S = 2(\boldsymbol{I}_{n_i} - \boldsymbol{X}_i (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}_i) = O(1)$, $\boldsymbol{X}'\nabla_y S_1 = \boldsymbol{0}$ and $\boldsymbol{X}'\nabla_y S = \boldsymbol{0}$. Based on these observations, we can check the conditions (A3), (A4) and (A6). From (5.4)-(5.6) of Prasad and Rao (1990), $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{U*})$ can be approximated as

$$\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{U*}) = \frac{2\theta_1^2}{N-k} \begin{pmatrix} 1 & -k/N \\ -k/N & \{k^2 + (N-k)\sum_{i=1}^k (1 + n_i\theta_2/\theta_1)^2\}/N^2 \end{pmatrix} + O(N^{-2}).$$

Thus from Remark 2.1, we can get the corresponding results in Theorems 2.4, 2.5 and 2.6 under (A1). ■

3 ML and REML methods

3.1 Notations and assumptions

In this section, we derive higher order expansions described in (2.5) and (2.17) for the ML and REML estimators of $\boldsymbol{\theta}$, and show that the conditions (A3), (A4) and (A6) are satisfied, and the corresponding results given in Theorems 2.1 - 2.6 are provided.

For notational simplicity, we here omit $(\boldsymbol{\theta})$ in $\boldsymbol{A}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ and others, and use the vector and matrix notations $\mathbf{col}_i(a_i)$ and $\mathbf{mat}_{ij}(b_{ij})$ defined by

$$\mathbf{col}_i(a_i) = \begin{pmatrix} a_1 \\ \vdots \\ a_q \end{pmatrix}, \quad \mathbf{mat}_{ij}(b_{ij}) = \begin{pmatrix} b_{11} & \cdots & b_{1q} \\ \vdots & \ddots & \vdots \\ b_{q1} & \cdots & b_{qq} \end{pmatrix}.$$

We here use the same notations as in Subsection 2.1. Also, for $0 \le i, j, k, \ell, m \le q$, let eigenvalues of $\boldsymbol{\Sigma}_{(ijk\ell)}$ and $\boldsymbol{\Sigma}_{(ijk\ell m)}$ be $\lambda_a^{ijk\ell}$ and $\lambda_a^{ijk\ell m}$ for $a = 1, \ldots, N$ respectively, where $|\lambda_1^{ijk\ell}| \le \cdots \le |\lambda_N^{ijk\ell}|$ and $|\lambda_1^{ijk\ell m}| \le \cdots \le |\lambda_N^{ijk\ell m}|$.

(M1) Let $\boldsymbol{A}_2 = \mathbf{mat}_{ij}(\mathrm{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}])$. Assume that $\boldsymbol{A}_2$ is a $q \times q$ positive definite matrix, and $\boldsymbol{A}_2/N$ converges to a positive definite matrix.

(M2) $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[4]}$, and $\lim_{N\to\infty} |\lambda_N^{ijk\ell}| < \infty$.

(M3) $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[5]}$, and $\lim_{N\to\infty} |\lambda_N^{ijk\ell m}| < \infty$.

Under these conditions with the conditions (A1), (A2)(i) and (A5)(i), we derive the second- and third-order expansions of ML and REML, which are defined as follows:

[1] ML method. The ML estimator $\widehat{\boldsymbol{\theta}}^M = (\hat{\theta}_1^M, \ldots, \hat{\theta}_q^M)'$ of $\boldsymbol{\theta}$ is defined as the solution of the equations $L_i(\widehat{\boldsymbol{\theta}}^M) = 0$ for $i = 1, \ldots, q$, where

$$L_i(\boldsymbol{\theta}) = L_i = \boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}], \tag{3.1}$$

for $\boldsymbol{P}(\boldsymbol{\theta}) = \boldsymbol{P} = \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$. Since $\boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{y} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, the condition (A3) is clearly satisfied and we can put $\boldsymbol{\beta} = \boldsymbol{0}$ without any loss of generality.

[2] REML method. The REML estimator $\widehat{\boldsymbol{\theta}}^R = (\hat{\theta}_1^R, \ldots, \hat{\theta}_q^R)'$ of $\boldsymbol{\theta}$ is defined as the solution of the equations $L_i^R(\widehat{\boldsymbol{\theta}}^R) = 0$ for $i = 1, \ldots, q$, where

$$L_i^R(\boldsymbol{\theta}) = \boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{y} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}]. \tag{3.2}$$

It is clear that the condition (A3) is satisfied, and we can put $\boldsymbol{\beta} = \mathbf{0}$ without any loss of generality.

The consistency of the ML and REML has been studied by Sweeting (1980), Mardia and Marshall (1984) and Cressie and Lahiri (1993). It can be seen that the conditions of Theorem 2 in Mardia and Marshall (1984) are satisfied by (A1), (A2)(i), (A5)(i) and (M1), so that we can see that $\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta} = O_p(N^{-1/2})$ and $\widehat{\boldsymbol{\theta}}^R - \boldsymbol{\theta} = O_p(N^{-1/2})$.

To derive asymptotic expansions of ML and REML, the following equalities are useful:

$$\begin{aligned} E\big[\mathrm{tr}\,[\boldsymbol{C}_1(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[\boldsymbol{C}_2(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big] =& 2\mathrm{tr}\,[\boldsymbol{C}_1\boldsymbol{\Sigma}\boldsymbol{C}_2\boldsymbol{\Sigma}], \\ E\big[\mathrm{tr}\,[\boldsymbol{C}_1(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[\boldsymbol{C}_2(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[\boldsymbol{C}_3(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big] =& 8\mathrm{tr}\,[\boldsymbol{C}_1\boldsymbol{\Sigma}\boldsymbol{C}_2\boldsymbol{\Sigma}\boldsymbol{C}_3\boldsymbol{\Sigma}], \end{aligned} \tag{3.3}$$

where $\boldsymbol{C}_1$, $\boldsymbol{C}_2$ and $\boldsymbol{C}_3$ are $N \times N$ matrices and $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ for $\boldsymbol{\beta} = \mathbf{0}$.

3.2 Expansions of ML and the corresponding results

We first derive the third-order expansion given in (2.17) for the ML estimator of $\boldsymbol{\theta}$ under the conditions (A1), (A2)(i), (A5)(i) and (M1)-(M3) where $\boldsymbol{\beta} = 0$.

[1] Taylor series expansion of ML. From the Taylor series expansion of (3.1), it is observed that

$$\begin{aligned} 0 = L_i(\widehat{\boldsymbol{\theta}}^M) =& L_i(\boldsymbol{\theta}) + \sum_a L_{i(a)}(\hat{\theta}_a^M - \theta_a) + \frac{1}{2}\sum_{a,b} L_{i(ab)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b) \\ &+ \frac{1}{6}\sum_{a,b,c} L_{i(abc)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)(\hat{\theta}_c^M - \theta_c) \\ &+ \frac{1}{24}\sum_{a,b,c,d} L_{i(abcd)}(\widetilde{\boldsymbol{\theta}})(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)(\hat{\theta}_c^M - \theta_c)(\hat{\theta}_d^M - \theta_d), \end{aligned} \tag{3.4}$$

where $\widetilde{\boldsymbol{\theta}}$ is a point satisfying $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \leq \|\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta}\|$ for the Euclidean norm $\|\cdot\|$, and $L_{i(ab)} = \partial_{ab}L_i$, $L_{i(abc)} = \partial_{abc}L_i$ and $L_{i(abcd)} = \partial_{abcd}L_i$. Also, $\sum_{a,b,c,d}$ means summation over $1 \leq a, b, c, d \leq q$, and $\sum_a$, $\sum_{a,b}$ and $\sum_{a,b,c}$ are defined similarly. Since $L_{i(abcd)}(\widetilde{\boldsymbol{\theta}}) = O_p(N)$, the last term is up to $O_p(N^{-1})$. Then, the equality (3.4) is expressed as

$$\begin{aligned} \mathbf{0} =& \mathbf{col}_i(L_i) + \mathbf{mat}_{ia}(L_{i(a)})(\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta}) + \frac{1}{2}\mathbf{col}_i\Big(\sum_{a,b} L_{i(ab)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)\Big) \\ &+ \frac{1}{6}\mathbf{col}_i\Big(\sum_{a,b,c} L_{i(abc)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)(\hat{\theta}_c^M - \theta_c)\Big) + O_p(N^{-1}), \end{aligned}$$

which implies that

$$\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta} =\{\mathbf{mat}_{ia}(-L_{i(a)})\}^{-1}\Big\{\mathbf{col}_i(L_i) + \frac{1}{2}\mathbf{col}_i\big(\sum_{a,b} L_{i(ab)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)\big)$$
$$+ \frac{1}{6}\mathbf{col}_i\big(\sum_{a,b,c} L_{i(abc)}(\hat{\theta}_a^M - \theta_a)(\hat{\theta}_b^M - \theta_b)(\hat{\theta}_c^M - \theta_c)\big)\Big\} + O_p(N^{-2}). \tag{3.5}$$

Thus, we need to evaluate each term in (3.5).

Since $(\boldsymbol{\Sigma}^{-1})_{(i)} = -\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}$, L_i is expressed as

$$\begin{aligned} L_i =& - \boldsymbol{y}'\{(\boldsymbol{\Sigma}^{-1})_{(i)} + \boldsymbol{Q}_i\}\boldsymbol{y} + \mathrm{tr}\,[\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(i)}] \\ =& - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(i)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{y}\boldsymbol{y}'], \end{aligned} \tag{3.6}$$

where $\boldsymbol{Q}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1} - \boldsymbol{P}\boldsymbol{\Sigma}_{(i)}\boldsymbol{P}$. From (3.6), it is observed that

$$L_{i(a)} =\mathrm{tr}\,[\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}^{-1})_{(i)}] - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{Q}_{i(a)}\boldsymbol{y}\boldsymbol{y}'], \tag{3.7}$$
$$L_{i(ab)} =B_{iab} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{Q}_{i(ab)}\boldsymbol{y}\boldsymbol{y}'], \tag{3.8}$$
$$L_{i(abc)} =C_{iabc} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iabc)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{Q}_{i(abc)}\boldsymbol{y}\boldsymbol{y}'], \tag{3.9}$$

where $B_{iab} = \partial_b\{\mathrm{tr}\,[\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}^{-1})_{(i)}]\}+\mathrm{tr}\,[\boldsymbol{\Sigma}_{(b)}(\boldsymbol{\Sigma}^{-1})_{(ia)}]$ and $C_{iabc} = \partial_c\big\{\partial_b\{\mathrm{tr}\,[\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}^{-1})_{(i)}]\}+ \mathrm{tr}\,[\boldsymbol{\Sigma}_{(b)}(\boldsymbol{\Sigma}^{-1})_{(ia)}]\big\} + \mathrm{tr}\,[\boldsymbol{\Sigma}_{(c)}(\boldsymbol{\Sigma}^{-1})_{(iab)}]$. From (3.6), L_i is written as

$$\mathbf{col}_i(L_i) = \boldsymbol{a}_1 - \boldsymbol{a}_0, \tag{3.10}$$

where $q \times 1$ vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_0$ are defined by

$$\boldsymbol{a}_1 = \mathbf{col}_i(-\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(i)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]), \quad \boldsymbol{a}_0 = \mathbf{col}_i(\mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{y}\boldsymbol{y}']). \tag{3.11}$$

It is noted that $\boldsymbol{X}'\boldsymbol{F}\boldsymbol{y} = O_p(N^{1/2})$ and $\mathrm{tr}\,[\boldsymbol{F}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] = O_p(N^{1/2})$ provided $\boldsymbol{F}$ satisfies $\boldsymbol{X}'\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{X} = O(N)$ and $\mathrm{tr}\,[\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{\Sigma}] = O(N)$, respectively, since $E[\{\mathrm{tr}\,[\boldsymbol{F}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\}^2] = \mathrm{tr}\,[\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{\Sigma}]$ by the equality in (3.3). Hence from conditions (A1), (A2)(i) and (M1), it follows that $\boldsymbol{a}_1 = O_p(N^{1/2})$ and $\boldsymbol{a}_0 = O_p(1)$. From (3.7),

$$\mathbf{mat}_{ia}(-L_{i(a)}) = \boldsymbol{A}_2 + \boldsymbol{A}_1 + \boldsymbol{A}_0, \tag{3.12}$$

where $q \times q$ matrices $\boldsymbol{A}_2$, $\boldsymbol{A}_1$ and $\boldsymbol{A}_0$ are defined by

$$\boldsymbol{A}_2 = \mathbf{mat}_{ia}(-\mathrm{tr}\,[\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}^{-1})_{(i)}]), \quad \boldsymbol{A}_1 = \mathbf{mat}_{ia}(\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]), \tag{3.13}$$

and $\boldsymbol{A}_0 = \mathbf{mat}_{ia}(\mathrm{tr}\,[\boldsymbol{Q}_{i(a)}\boldsymbol{y}\boldsymbol{y}'])$. It is noted that

$$(\boldsymbol{\Sigma}^{-1})_{(ij)} = \boldsymbol{\Sigma}^{-1}\{\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)} + \boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)} - \boldsymbol{\Sigma}_{(ij)}\}\boldsymbol{\Sigma}^{-1},$$

and

$$(\mathrm{tr}\,[\boldsymbol{C}\boldsymbol{D}])^2 \le \mathrm{tr}\,[\boldsymbol{C}^2]\mathrm{tr}\,[\boldsymbol{D}^2],$$

for symmetric matrices $\boldsymbol{C}$ and $\boldsymbol{D}$. Then from conditions (A1), (A2)(i) and (M1), it follows that $\boldsymbol{A}_2/N$ converges a positive definite matrix, and $\boldsymbol{A}_1 = O_p(N^{1/2})$ and $\boldsymbol{A}_0 = O_p(1)$, so that the inverse matrix of $\mathbf{mat}_{ia}(-L_{i(a)})$ can be expanded as

$$\{\mathbf{mat}_{ia}(-L_{i(a)})\}^{-1} = \boldsymbol{A}_2^{-1} - \boldsymbol{A}_2^{-1}\boldsymbol{A}_1\boldsymbol{A}_2^{-1} + \boldsymbol{A}_2^{-1}(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1 - \boldsymbol{A}_0)\boldsymbol{A}_2^{-1} + O_p(N^{-5/2}), \quad (3.14)$$

where $\boldsymbol{A}_2^{-1} = O(N^{-1})$, $\boldsymbol{A}_2^{-1}\boldsymbol{A}_1\boldsymbol{A}_2^{-1} = O_p(N^{-3/2})$ and $\boldsymbol{A}_2^{-1}(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1 - \boldsymbol{A}_0)\boldsymbol{A}_2^{-1} = O_p(N^{-2})$. Similarly, from (3.8) and (3.9), $L_{i(ab)}$ and $L_{i(abc)}$ can be evaluated as

$$\begin{aligned} L_{i(ab)} =& B_{iab} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + O_p(1), \\ L_{i(abc)} =& C_{iabc} + O_p(N^{1/2}). \end{aligned} \quad (3.15)$$

Hence from (3.5), $\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta}$ can be approximated as

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta} =& \big\{\boldsymbol{A}_2^{-1} - \boldsymbol{A}_2^{-1}\boldsymbol{A}_1\boldsymbol{A}_2^{-1} + \boldsymbol{A}_2^{-1}(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1 - \boldsymbol{A}_0)\boldsymbol{A}_2^{-1} + O_p(N^{-5/2})\big\} \\ & \times \Big\{\boldsymbol{a}_1 - \boldsymbol{a}_0 + \frac{1}{2}\mathbf{col}_i\big(\sum_{a,b}\big\{B_{iab} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big\}(\hat\theta_a^M - \theta_a)(\hat\theta_b^M - \theta_b)\big) \\ & + \frac{1}{6}\mathbf{col}_i\big(\sum_{a,b,c} C_{iabc}(\hat\theta_a^M - \theta_a)(\hat\theta_b^M - \theta_b)(\hat\theta_c^M - \theta_c)\big)\Big\} + O_p(N^{-2}). \end{aligned} \quad (3.16)$$

[2] First- and second-order terms. From the approximation (3.16), it follows that $\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{M*} + O_p(N^{-1})$, where

$$\widehat{\boldsymbol{\theta}}^{M*} = \boldsymbol{A}_2^{-1}\boldsymbol{a}_1 = \boldsymbol{A}_2^{-1}\mathbf{col}_i(-\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(i)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]). \quad (3.17)$$

Using the approximations (3.15) and (3.17), we can see that

$$\mathbf{col}_i\big(\sum_{a,b} L_{i(ab)}(\hat\theta_a^M - \theta_a)(\hat\theta_b^M - \theta_b)\big) = \boldsymbol{b}_0 + O_p(N^{-1/2}),$$

for

$$\boldsymbol{b}_0 = \mathbf{col}_i\big(\sum_{a,b} B_{iab}\hat\theta_a^{M*}\hat\theta_b^{M*}\big).$$

Hence from (3.16), it is seen that $\widehat{\boldsymbol{\theta}}^M - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{M*} + \widehat{\boldsymbol{\theta}}^{M**} + O_p(N^{-3/2})$, where

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{M**} =& -\boldsymbol{A}_2^{-1}\big\{\boldsymbol{a}_0 - \frac{\boldsymbol{b}_0}{2} + \boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{a}_1\big\} \\ =& \boldsymbol{A}_2^{-1}\Big\{-\mathbf{col}_i\big(\mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{y}\boldsymbol{y}']\big) + \mathbf{col}_i\big(\sum_{a,b} B_{iab}\hat\theta_a^{M*}\hat\theta_b^{M*}\big)/2\Big\} \\ & - \boldsymbol{A}_2^{-1}\mathbf{mat}_{ia}\big(\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big)\widehat{\boldsymbol{\theta}}^{M*}. \end{aligned} \quad (3.18)$$

[3] Third-order term. To evaluate the approximation (3.16) up to $O_p(N^{-2})$, we observe that for (3.15),

$$\begin{aligned} \mathbf{col}_i\big(\sum_{a,b} L_{i(ab)}(\hat\theta_a^M - \theta_a)(\hat\theta_b^M - \theta_b)\big) =& \boldsymbol{b}_0 + \boldsymbol{b}_{-1} + O_p(N^{-1}), \\ \mathbf{col}_i\big(\sum_{a,b,c} L_{i(abc)}(\hat\theta_a^M - \theta_a)(\hat\theta_b^M - \theta_b)(\hat\theta_c^M - \theta_c)\big) =& \boldsymbol{c}_{-1} + O_p(N^{-1}), \end{aligned}$$

for

$$\begin{aligned}\boldsymbol{b}_{-1} =&\mathbf{col}_i\big(\sum_{a,b} B_{iab}(\hat\theta_a^{M*}\hat\theta_b^{M**}+\hat\theta_b^{M*}\hat\theta_a^{M**})\big)\\&-\mathbf{col}_i\big(\sum_{a,b}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}'-\boldsymbol{\Sigma})]\hat\theta_a^{M*}\hat\theta_b^{M*}\big),\\ \boldsymbol{c}_{-1} =&\mathbf{col}_i\big(\sum_{a,b,c} C_{iabc}\hat\theta_a^{M*}\hat\theta_b^{M*}\hat\theta_c^{M*}\big).\end{aligned}$$

Thus, the third-order term is given by

$$\widehat{\boldsymbol{\theta}}^{M***} = \boldsymbol{A}_2^{-1}\big\{\frac{\boldsymbol{b}_{-1}}{2}+\frac{\boldsymbol{c}_{-1}}{6}+\boldsymbol{A}_1\boldsymbol{A}_2^{-1}(\boldsymbol{a}_0-\frac{\boldsymbol{b}_0}{2})+(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1-\boldsymbol{A}_0)\widehat{\boldsymbol{\theta}}^{M*}\big\}. \tag{3.19}$$

[5] Expansion of ML and the corresponding results. From these arguments, under the conditions (A1), (A2)(i), (A5)(i), (M1)-(M3), we obtain the expansion

$$\widehat{\boldsymbol{\theta}}^{M}-\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{M*}+\widehat{\boldsymbol{\theta}}^{M**}+\widehat{\boldsymbol{\theta}}^{M***}+O_p(N^{-2}), \tag{3.20}$$

where it is verified that $\widehat{\boldsymbol{\theta}}^{M*}=O_p(N^{-1/2})$, $\widehat{\boldsymbol{\theta}}^{M**}=O_p(N^{-1})$ and $\widehat{\boldsymbol{\theta}}^{M***}=O_p(N^{-3/2})$.

Concerning the second-order expansion, using the same arguments as in above, we can verify that

$$\widehat{\boldsymbol{\theta}}^{M}-\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{M*}+\widehat{\boldsymbol{\theta}}^{M**}+O_p(N^{-3/2}), \tag{3.21}$$

under the weaker conditions (A1), (A2)(i), (A5)(i), (M1)-(M2).

Proposition 3.1 (i) *Assume the conditions* (A1), (A2), (A5), (M1)-(M3). *Then, the conditions* (A4), (A6), (B1), (B2), (C1) *and* (C2) *are satisfied for* $\widehat{\boldsymbol{\theta}}^{M}$, *and the third-order expansion* (3.20) *is obtained. Especially, it is observed that* $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{M*})=2\boldsymbol{A}_2^{-1}$, $E[\widehat{\boldsymbol{\theta}}^{M*}]=\mathbf{0}$ *and*

$$\begin{aligned}E[\widehat{\boldsymbol{\theta}}^{M**}] =&\boldsymbol{A}_2^{-1}\mathbf{col}_i(\mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}])\\&+\boldsymbol{A}_2^{-1}\mathbf{col}_i(\mathrm{tr}\,[\boldsymbol{A}_2^{-1}\mathbf{mat}_{a,b}(\mathrm{tr}\,[\boldsymbol{\Sigma}_{(ab)}(\boldsymbol{\Sigma}^{-1})_{(i)}])]).\end{aligned} \tag{3.22}$$

It is noted that $E[\widehat{\boldsymbol{\theta}}^{M**}]=\boldsymbol{A}_2^{-1}\mathbf{col}_i(\mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}])$ *when* $\boldsymbol{\Sigma}$ *or* $\boldsymbol{G}$ *and* $\boldsymbol{R}$ *are matrices of linear functions of* $\boldsymbol{\theta}$. *Also,* $E[\hat\theta_i^{M***}]=O(N^{-2})$, $E[\hat\theta_i^{M*}\hat\theta_j^{M**}]=O(N^{-2})$ *and* $E[\hat\theta_i^{M*}\hat\theta_j^{M*}\hat\theta_k^{M*}]=O(N^{-2})$ *for* $1\le i,j,k\le p$.

(ii) *Assume the conditions* (A1), (A2), (A5)(i), (M1)-(M2). *Then, the conditions* (A4), (B1) *and* (C1) *are satisfied, and the second-order expansion* (3.21) *is obtained.*

The proof is given in Section 4. From Proposition 3.1, the assumptions in Theorems 2.4, 2.5 and 2.6 are satisfied by the conditions (A1), (A2), (A5), (M1)-(M3), and we get

$$\begin{aligned}MSE(\boldsymbol{\theta},\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}^{M})) =&g_1(\boldsymbol{\theta})+g_2(\boldsymbol{\theta})+g_3^*(\boldsymbol{\theta})+O(N^{-2}),\\ E[mse(\widehat{\boldsymbol{\theta}}^{M},\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}^{M}))] =&MSE(\boldsymbol{\theta},\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}^{M}))+O(N^{-2}),\\ P[\mu\in I^{CEB}(\widehat{\boldsymbol{\theta}}^{M})] =&1-\alpha+O(N^{-2}),\end{aligned}$$

where $g_1(\boldsymbol{\theta})$, $g_2(\boldsymbol{\theta})$ and $g_3^*(\boldsymbol{\theta})$ are given in (2.6), and $mse(\widehat{\boldsymbol{\theta}}^M, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}^M))$ and $I^{CEB}(\widehat{\boldsymbol{\theta}}^M)$ are defined around Theorems 2.2 and 2.3. Also from Proposition 3.1, the assumptions in Theorems 2.1, 2.2 and 2.3 are satisfied by the conditions (A1), (A2), (A5)(i), (M1)-(M2), and we get the corresponding results with the second-order approximation.

3.3 Expansion of REML and corresponding results

Concerning the expansions of REML defined in (3.2), we can use the same arguments as in the above expansions of ML.

It is noted that L_i^R given in (3.2) is rewritten as $L_i^R = -\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(i)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{Q}_i(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]$ or

$$\mathbf{col}_i(L_i^R) = \boldsymbol{a}_1 - \boldsymbol{a}_0^*,$$

where $\boldsymbol{a}_0^* = \mathbf{col}_i(\mathrm{tr}\,[\boldsymbol{Q}_i(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})])$, and $\boldsymbol{a}_1$ is given in (3.11). Since the term $L_{i(a)}^R$ is given by

$$L_{i(a)}^R = \mathrm{tr}\,[\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}^{-1})_{(i)}] - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + \mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{\Sigma}_{(a)}] - \mathrm{tr}\,[\boldsymbol{Q}_{i(a)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})],$$

the matricial expression $\mathbf{mat}_{ia}(-L_{ia}^R)$ can be written as

$$\mathbf{mat}_{ia}(-L_{i(a)}^R) = \boldsymbol{A}_2 + \boldsymbol{A}_1 + \boldsymbol{A}_0^*,$$

where $\boldsymbol{A}_0^* = \mathbf{mat}_{ia}(-\mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{\Sigma}_{(a)}] + \mathrm{tr}\,[\boldsymbol{Q}_{i(a)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})])$, and $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are given in (3.12). Similarly, $L_{i(ab)}^R$ and $L_{i(abc)}^R$ can be evaluated as $L_{i(ab)}^R = B_{iab} - \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + O_p(1)$ and $L_{i(abc)}^R = C_{iabc} + O_p(N^{1/2})$. From the same arguments as given in the previous subsection, we can approximate $\widehat{\boldsymbol{\theta}}^R$ as

$$\widehat{\boldsymbol{\theta}}^R - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{R*} + \widehat{\boldsymbol{\theta}}^{R**} + \widehat{\boldsymbol{\theta}}^{R***} + O_p(N^{-2}), \tag{3.23}$$

under the same conditions as given in Proposition 3.1. Here, $\widehat{\boldsymbol{\theta}}^{R*}$, $\widehat{\boldsymbol{\theta}}^{R**}$ and $\widehat{\boldsymbol{\theta}}^{R***}$ are the similar forms to $\widehat{\boldsymbol{\theta}}^{M*}$, $\widehat{\boldsymbol{\theta}}^{M**}$ and $\widehat{\boldsymbol{\theta}}^{M***}$, respectively, where $\boldsymbol{a}_0$ and $\boldsymbol{A}_0$ are replaced with $\boldsymbol{a}_0^*$ and $\boldsymbol{A}_0^*$. Especially, $\widehat{\boldsymbol{\theta}}^{R*}$ and $\widehat{\boldsymbol{\theta}}^{R**}$ are given by

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{R*} =& \widehat{\boldsymbol{\theta}}^{M*} = \boldsymbol{A}_2^{-1}\boldsymbol{a}_1,\\
\widehat{\boldsymbol{\theta}}^{R**} =& -\boldsymbol{A}_2^{-1}\{\boldsymbol{a}_0^* - \boldsymbol{b}_0/2 + \boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{a}_1\}\\
=& \boldsymbol{A}_2^{-1}\Big\{-\mathbf{col}_i\big(\mathrm{tr}\,[\boldsymbol{Q}_i(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big) + \mathbf{col}_i\big(\sum_{a,b} B_{iab}\hat{\theta}_a^{M*}\hat{\theta}_b^{M*}\big)/2\Big\}\\
& - \boldsymbol{A}_2^{-1}\mathbf{mat}_{ia}\big(\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big)\widehat{\boldsymbol{\theta}}^{M*}.
\end{aligned} \tag{3.24}$$

Then, we get the following proposition, whose proof is omitted, since it can be verified based on the same arguments as in the proof of Proposition 3.1.

Proposition 3.2 (i) *Assume the conditions* (A1), (A2), (A5), (M1)-(M3). *Then, the conditions* (A4), (A6), (B1), (B2), (C1) *and* (C2) *are satisfied. Thus, the third-order* *expansion* (3.23) *is obtained, and the corresponding results to Theorems 2.4, 2.5 and 2.6 hold. Especially, it is observed that* $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{R*}) = \mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{M*}) = 2\boldsymbol{A}_2^{-1}$, $E[\widehat{\boldsymbol{\theta}}^{R*}] = \mathbf{0}$ *and*

$$E[\widehat{\boldsymbol{\theta}}^{R**}] = \boldsymbol{A}_2^{-1}\mathbf{col}_i\big(\mathrm{tr}\,[\boldsymbol{A}_2^{-1}\mathbf{mat}_{a,b}\big(\mathrm{tr}\,[\boldsymbol{\Sigma}_{(ab)}(\boldsymbol{\Sigma}^{-1})_{(i)}]\big)]\big), \tag{3.25}$$

where $E[\widehat{\boldsymbol{\theta}}^{R**}] = \mathbf{0}$ *when* $\boldsymbol{\Sigma}$ *are matrices of linear functions of* $\boldsymbol{\theta}$*. Also,* $E[\hat{\theta}_i^{R***}] = O(N^{-2})$, $E[\hat{\theta}_i^{R*}\hat{\theta}_j^{R**}] = O(N^{-2})$ *and* $E[\hat{\theta}_i^{R*}\hat{\theta}_j^{R*}\hat{\theta}_k^{R*}] = O(N^{-2})$ *for* $1 \le i, j, k \le p$.

(ii) *Assume the conditions* (A1), (A2), (A5)(i), (M1)-(M2)*. Then, the conditions* (A4), (B1) *and* (C1) *are satisfied, and the second-order expansion and the corresponding results to Theorems 2.1, 2.2 and 2.3 are obtained, where* $\boldsymbol{a}_0$ *is replaced with* $\boldsymbol{a}_0^*$.

3.4 ML and REML in specific models

Example 3.1 (Fay-Herriot model and modified Fay-Herriot estimator) Consider the model treated in Example 2.1.

[ML estimator] The ML estimator $\hat{\theta}_M$ 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 $\boldsymbol{P}(\theta)$ defined in (3.1). The conditions (A2)-(A6), (B1)-(B2), (C1)-(C2) and (M1)-(M2) can be seen to be satisfied, and we get Theorems 2.4, 2.5 and 2.6, where $a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\boldsymbol{yy}' - \boldsymbol{\Sigma})]$, $A_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]$, $\hat{\theta}^{M*} = a_1/A_2$, and $E[\hat{\theta}^{M*}] = 0$, $Var[\hat{\theta}^{M*}] = 2/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]$, and $E[\hat{\theta}^{M**}] = -\mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-2}\boldsymbol{X}/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]$.

[REML estimator] From (3.2), on the other hand, the REML estimator is given as the solution of the equation $L^R(\hat{\theta}^R) = 0$, where

$$L^R(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))^2\boldsymbol{y} - \mathrm{tr}\,[(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))].$$

From the arguments around (3.23), it can be seen that $\hat{\theta}^{R*} = a_1/A_2$, and $E[\hat{\theta}^{R*}] = 0$, $Var[\hat{\theta}^{R*}] = 2/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]$, and $E[\hat{\theta}^{R**}] = 0$. Since all the conditions are satisfied, we get Theorems 2.4, 2.5 and 2.6 under (A1).

[Modified Fay-Herriot estimator] The estimator suggested by Fay and Herriot (1978) is given as the solution of the equation $L^{FH}(\hat{\theta}^{FH}) = 0$, where

$$L^{FH}(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))\boldsymbol{y} - (k - p)$$

for $\boldsymbol{P}(\theta)$ defined in (3.1). Here, we consider the general estimator $\hat{\theta}_m$ given as the solution of the equation $L_m(\theta) = 0$, where

$$L_m(\theta) = \boldsymbol{y}'(\boldsymbol{\Sigma}(\theta)^{-1} - \boldsymbol{P}(\theta))\boldsymbol{y} - (k - p) - m(\theta), \tag{3.26}$$

and $m(\theta)$ is a function of θ with order $m(\theta) = O(1)$. To derive the expression corresponding to (3.20) for $\hat{\theta}_m$, the same arguments as in Subsection 3.2 are used. Especially, the terms corresponding to (3.10), (3.12) and (3.15) are expressed as follows: Since $L_m(\theta) = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\boldsymbol{yy}' - \boldsymbol{\Sigma})] - m(\theta) - \mathrm{tr}\,[\boldsymbol{P}(\boldsymbol{yy}' - \boldsymbol{\Sigma})]$, we can put

$$a_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\boldsymbol{yy}' - \boldsymbol{\Sigma})], \quad a_0 = \mathrm{tr}\,[\boldsymbol{P}(\boldsymbol{yy}' - \boldsymbol{\Sigma})] + m(\theta).$$

Let $f^{(n)}(\theta) = \partial^n f(\theta)/\partial\theta^n$ for a function $f(\theta)$. Since $-L_m^{(1)}(\theta) = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}] + \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + \mathrm{tr}\,[\boldsymbol{P}^{(1)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \mathrm{tr}\,[\boldsymbol{P}] + m^{(1)}(\theta)$ and $\mathrm{tr}\,[\boldsymbol{P}^{(1)}\boldsymbol{\Sigma}] = -\mathrm{tr}\,[\boldsymbol{P}]$, we can put

$$A_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}], \quad A_1 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})], \quad A_0 = \mathrm{tr}\,[\boldsymbol{P}^{(1)}\boldsymbol{y}\boldsymbol{y}'] + m^{(1)}(\theta).$$

Then,

$$\hat{\theta}_m^* = a_1/A_2 = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]/\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}].$$

Since $L_m^{(2)}(\theta) = 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] + 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + O_p(1)$ and $L_m^{(3)}(\theta) = -6\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}] + O_p(k^{1/2})$, we can put

$$\begin{aligned} b_0 =& 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}](\hat{\theta}_m^*)^2, \quad c_{-1} = -6\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}](\hat{\theta}_m^*)^3, \\ b_{-1} =& 4\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]\hat{\theta}_m^*\hat{\theta}_m^{**} + 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})](\hat{\theta}_m^*)^2, \end{aligned}$$

where $\hat{\theta}_m^{**}$ is defined by

$$\begin{aligned} \hat{\theta}_m^{**} =& (-a_0 + b_0/2)/A_2 - (A_1/A_2)\hat{\theta}_m^* \\ =& -\frac{m(\theta)}{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]} + \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]}{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]}(\hat{\theta}_m^*)^2 - \frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]}{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]}\hat{\theta}_m^*. \end{aligned}$$

Also, $\hat{\theta}_m^{***}$ can be derived as

$$\hat{\theta}_m^{***} = A_2^{-1}\big\{b_{-1}/2 + c_{-1}/6 + A_1A_2^{-1}(a_0 - b_0/2) + (A_1^2A_2^{-1} - A_0)\hat{\theta}_m^*\big\}.$$

Hence, $E[\hat{\theta}_m^*] = 0$, $Var[\hat{\theta}_m^*] = 2k/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2$ and

$$E[\hat{\theta}_m^{**}] = 2\frac{k\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2}{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3} - \frac{m(\theta)}{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}]}, \tag{3.27}$$

which were derived by Datta, *et al.* (2005) in the case of $m(\theta) = 0$. That is, $E[\hat{\theta}^{FH*}] = 0$, $Var[\hat{\theta}^{FH*}] = 2k/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2$ and $E[\hat{\theta}^{FH**}] = 2\{k\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}] - (\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2\}/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^3$. From (3.27), we can vanish the second-order term in the bias of $\hat{\theta}_m$ by putting

$$m(\theta) = 2k\frac{\mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}]}{(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2} - 2. \tag{3.28}$$

Then, $E[\hat{\theta}_m^{**}] = 0$ while $Var[\hat{\theta}_m^*] = Var[\hat{\theta}^{FH*}] = 2k/(\mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}])^2$. It can be verified that all the conditions are satisfied, and we get the results in Theorems 2.4, 2.5 and 2.6 under (A1). ■

Example 3.2 (NERM) Consider the model treated in Example 2.2.

[ML estimator] The ML estimators $\widehat{\boldsymbol{\theta}}^M = (\hat{\theta}_1^M, \hat{\theta}_2^M)'$ of $(\theta_1, \theta_2)'$ are given as the solutions of the equations $L_1(\widehat{\boldsymbol{\theta}}^M) = 0$ and $L_2(\widehat{\boldsymbol{\theta}}^M) = 0$, where $L_1(\boldsymbol{\theta})$ and $L_2(\boldsymbol{\theta})$ given in (3.1) can be written as

$$\begin{aligned} L_1(\boldsymbol{\theta}) =& \frac{1}{\theta_1^2}\sum_{i=1}^k \|\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \frac{n_i\theta_2}{\theta_1 + n_i\theta_2}(\overline{y}_i - \overline{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}))\boldsymbol{j}_i\|^2 - \sum_{i=1}^k \frac{n_i}{\theta_1}(1 - \frac{\theta_2}{\theta_1 + n_i\theta_2}), \\ L_2(\boldsymbol{\theta}) =& \sum_{i=1}^k \frac{n_i^2}{(\theta_1 + n_i\theta_2)^2}\{\overline{y}_i - \overline{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})\}^2 - \sum_{i=1}^k \frac{n_i}{\theta_1 + n_i\theta_2}, \end{aligned}$$

since $\boldsymbol{\Sigma}_{(1)} = \boldsymbol{I}$ and $\boldsymbol{\Sigma}_{(2)} = \text{block diag}(\boldsymbol{J}_{n_1}, \ldots, \boldsymbol{J}_{n_k})$. In this model, the conditions (A2)-(A6), (B1)-(B2), (C1)-(C2) and (M1)-(M3) are satisfied, so that the results stated in Proposition 3.1 are established under (A1). Note that $\boldsymbol{A}_2$ and $\boldsymbol{a}_1$ given in (3.13) and (3.10) can be written as

$$
\begin{aligned}
\boldsymbol{A}_2 =& \mathbf{mat}_{ij}(\text{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}]) \\
=& \begin{pmatrix} (N-k)\theta_1^{-2} + \sum_i (\theta_1 + n_i\theta_2)^{-2} & \sum_i n_i(\theta_1 + n_i\theta_2)^{-2} \\ \sum_i n_i(\theta_1 + n_i\theta_2)^{-2} & \sum_i n_i^2(\theta_1 + n_i\theta_2)^{-2} \end{pmatrix}, \\
\boldsymbol{a}_1 =& \begin{pmatrix} \sum_i \text{tr}\,[\boldsymbol{\Sigma}_i^{-2}(\boldsymbol{y}_i\boldsymbol{y}_i' - \boldsymbol{\Sigma}_i)] \\ \sum_i \boldsymbol{j}_i'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i\boldsymbol{y}_i' - \boldsymbol{\Sigma}_i)\boldsymbol{\Sigma}_i^{-1}\boldsymbol{j}_i \end{pmatrix}.
\end{aligned}
$$

Then, $\widehat{\boldsymbol{\theta}}^{M*} = \boldsymbol{A}_2^{-1}\boldsymbol{a}_1$, and $E[\widehat{\boldsymbol{\theta}}^{M*}] = \mathbf{0}$ and

$$
\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{M*}) = \frac{2\theta_1^2}{d(\psi)} \begin{pmatrix} \sum_{i=1}^k n_i^2\gamma_i^2 & -\sum_{i=1}^k n_i\gamma_i^2 \\ -\sum_{i=1}^k n_i\gamma_i^2 & (N-k+\sum_{i=1}^k \gamma_i^2) \end{pmatrix},
$$

where $d(\psi) = (N-k+\sum_{i=1}^k \gamma_i^2)\sum_{i=1}^k n_i^2\gamma_i^2 - (\sum_{i=1}^k n_i\gamma_i^2)^2$ and $\gamma_i = (1+n_i\psi)^{-1}$ for $\psi = \theta_2/\theta_1$. It is easy to see that $g_3^*(\boldsymbol{\theta})$ can be given by $g_3^*(\boldsymbol{\theta}) = 2Nn_s\theta_1\gamma_s^3/d(\psi)$. Also from (3.22), $E[\widehat{\boldsymbol{\theta}}^{M**}]$ can be written as

$$
E[\widehat{\boldsymbol{\theta}}^{M**}] = \frac{\theta_1}{d(\psi)} \begin{pmatrix} -p\sum_{i=1}^k n_i^2\gamma_i^2 + (\sum_{i=1}^k n_i\gamma_i)c(\psi) \\ p\sum_{i=1}^k n_i\gamma_i^2 - (N-k+\sum_{i=1}^k \gamma_i)c(\psi) \end{pmatrix},
$$

where $c(\psi) = \text{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\sum_{i=1}^k n_i^2\gamma_i^2\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i']$. These were obtained by Datta and Lahiri (2000).

[REML estimator] The REML estimators $\widehat{\boldsymbol{\theta}}^R = (\hat{\theta}_1^R, \hat{\theta}_2^R)'$ of $(\theta_1, \theta_2)'$ are given as the solutions of the equations $L_1^R(\widehat{\boldsymbol{\theta}}^R) = 0$ and $L_2^R(\widehat{\boldsymbol{\theta}}^R) = 0$, where $L_1^R(\boldsymbol{\theta})$ and $L_2^R(\boldsymbol{\theta})$ given in (3.2) can be written as

$$
\begin{aligned}
L_1^R(\boldsymbol{\theta}) =& L_1(\boldsymbol{\theta}) + \text{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-2}\boldsymbol{X}], \\
L_2^R(\boldsymbol{\theta}) =& L_2(\boldsymbol{\theta}) + \text{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\text{block diag}(\boldsymbol{J}_{n_1}, \ldots, \boldsymbol{J}_{n_k})\boldsymbol{\Sigma}^{-1}\boldsymbol{X}].
\end{aligned}
$$

Then, it can be seen that $\widehat{\boldsymbol{\theta}}^{R*} = \boldsymbol{A}_2^{-1}\boldsymbol{a}_1 = \widehat{\boldsymbol{\theta}}^{M*}$, and $E[\widehat{\boldsymbol{\theta}}^{R*}] = \mathbf{0}$, $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{R*}) = \mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{M*})$ and $E[\widehat{\boldsymbol{\theta}}^{R**}] = O(N^{-2})$ as shown in Datta and Lahiri (2000). Hence, we can get the corresponding results in Proposition 3.1 under (A1). ■

3.5 A basic area level model for combining time-series and cross-sectional data

Finally we consider a basic area level model proposed by Rao and Yu (1994) for combining the time-series and cross-sectional data. This is an extension of the Fay-Herriot model and is described by

$$
y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_i + u_{it} + \varepsilon_{it}, \quad i = 1, \ldots, k, \; t = 1, \ldots, T, \tag{3.29}
$$

where k is the number of small areas, t is a time index, $N = kT$, $\boldsymbol{x}_{it}$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ unknown common vector of regression coefficients. Here, v_i's, u_{it}'s and ε_{it}'s are mutually independent random variables such that $v_i \sim \mathcal{N}(0, \sigma_v^2)$, $\varepsilon_{it} \sim \mathcal{N}(\mathbf{0}, d_{it})$ and

$$u_{it} = \rho u_{i,t-1} + e_{it}, \qquad |\rho| < 1,$$

where $e_{it} \sim \mathcal{N}(0, \sigma^2)$, and σ_v^2, σ^2 and ρ are unknown parameters. Let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{i,T})'$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{i,T})'$, and let $\boldsymbol{u}_i$ and $\boldsymbol{\epsilon}_i$ be similarly defined. Then, the model is expressed in vector notations as

$$\begin{aligned} \boldsymbol{y}_i =& \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_T v_i + \boldsymbol{u}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, k, \\ =& \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_0\boldsymbol{v}_i^* + \boldsymbol{\epsilon}_i, \end{aligned}$$

for $\boldsymbol{j}_T = (1, \ldots, 1)' \in \boldsymbol{R}^T$, $\boldsymbol{Z}_0 = (\boldsymbol{j}_T, \boldsymbol{I}_T)$ and $\boldsymbol{v}_i^* = (v_i, \boldsymbol{u}_i')'$, and it is seen that $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Psi}(\rho))$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}_i)$, where

$$\boldsymbol{\Psi}(\rho) = \frac{1}{1-\rho^2}\mathbf{mat}_{i,j}(\rho^{|i-j|}) \quad \text{and} \quad \boldsymbol{D}_i = \text{diag}\,(d_{i1}, \ldots, d_{iT}).$$

Let $\theta_1 = \sigma_v^2$, $\theta_2 = \sigma^2$, $\theta_3 = \rho$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$. Thus,

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i(\boldsymbol{\theta})),$$

where for $\boldsymbol{J}_T = \boldsymbol{j}_T\boldsymbol{j}_T'$ and $\boldsymbol{G}_0(\boldsymbol{\theta}) = \text{block diag}(\theta_1, \theta_2\boldsymbol{\Psi}(\theta_3))$, we have

$$\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_0\boldsymbol{G}_0(\boldsymbol{\theta})\boldsymbol{Z}_0' + \boldsymbol{D}_i = \theta_1\boldsymbol{J}_T + \theta_2\boldsymbol{\Psi}(\theta_3) + \boldsymbol{D}_i.$$

Letting $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_k')'$, $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_k')'$ and letting $\boldsymbol{v}^*$ and $\boldsymbol{\epsilon}$ be defined similarly, we can express the model as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v}^* + \boldsymbol{\epsilon}$, where $\boldsymbol{Z} = \boldsymbol{I}_T \otimes \boldsymbol{Z}_0$, $\mathbf{Cov}\,(\boldsymbol{v}^*) = \boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{I}_k \otimes \boldsymbol{G}_0(\boldsymbol{\theta})$ and $\mathbf{Cov}\,(\boldsymbol{\epsilon}) = \boldsymbol{R} = \text{block diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_k)$. Then, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ for $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{Z}\boldsymbol{G}(\boldsymbol{\theta})\boldsymbol{Z}' + \boldsymbol{R} = \text{block diag}(\boldsymbol{\Sigma}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{\Sigma}_k(\boldsymbol{\theta}))$. It is assumed that T is bounded, $\sup_{i\geq 1, t\geq 1} d_{it} < \infty$, $\inf_{i\geq 1, t\geq 1} d_{it} > 0$ and that $k \to \infty$.

We consider to predict the currect mean of the s-th small area $\mu_{sT} = \boldsymbol{x}_{sT}'\boldsymbol{\beta} + v_s + u_{sT}$. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ used in Section 2 correspond to $\boldsymbol{a} = \boldsymbol{x}_{sT}$ and $\boldsymbol{b} = (\mathbf{0}_{T+1}', \ldots, \mathbf{0}_{T+1}, \boldsymbol{b}_s', \mathbf{0}_{T+1}', \ldots, \mathbf{0}_{T+1}')'$ for $\mathbf{0}_{T+1} = (0, \ldots, 0)' \in \boldsymbol{R}^{T+1}$ and $\boldsymbol{b}_s = (1, 0, \ldots, 0, 1)' \in \boldsymbol{R}^{T+1}$ such that $\boldsymbol{b}'\boldsymbol{v}^* = \boldsymbol{b}_s'\boldsymbol{v}_s^* = v_s + u_{sT}$. Then, the EBLUP of μ_{sT} is

$$\begin{aligned} \widehat{\mu}_{sT}^{EB} =& \boldsymbol{x}_{sT}'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{b}_s'\boldsymbol{G}_0(\widehat{\boldsymbol{\theta}})\boldsymbol{Z}_0'\boldsymbol{\Sigma}_s^{-1}(\widehat{\boldsymbol{\theta}})\left\{\boldsymbol{y}_s - \boldsymbol{X}_s\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})\right\} \\ =& \boldsymbol{x}_{sT}'\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}) + \boldsymbol{s}_s'(\widehat{\boldsymbol{\theta}})\left\{\boldsymbol{y}_s - \boldsymbol{X}_s\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})\right\}, \end{aligned}$$

where $\boldsymbol{s}_s'(\boldsymbol{\theta})$ is expressed as

$$\begin{aligned} \boldsymbol{s}_s'(\boldsymbol{\theta}) =& \boldsymbol{b}_s'\boldsymbol{G}_0(\boldsymbol{\theta})\boldsymbol{Z}_0'\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}) = \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\,\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}) \\ =& \{\theta_1\boldsymbol{j}_T' + \theta_2(\theta_3^{T-1}, \ldots, \theta_3, 1)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}), \end{aligned}$$

for $\boldsymbol{c}_T = (0, \ldots, 0, 1)' \in \boldsymbol{R}^T$.

From (2.9), the function $g_1(\boldsymbol{\theta})$ is written as

$$g_1(\boldsymbol{\theta}) = \theta_1 + \theta_2 \boldsymbol{c}_T' \boldsymbol{\Psi}(\theta_3)\boldsymbol{c}_T - \boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{\Sigma}_s(\boldsymbol{\theta})\boldsymbol{s}_s(\boldsymbol{\theta}),$$

and $g_2(\boldsymbol{\theta})$ and $g_3^*(\boldsymbol{\theta})$ are expressed as

$$\begin{aligned} g_2(\boldsymbol{\theta}) =&(\boldsymbol{x}_{sT}' - \boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{X}_s)(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}(\boldsymbol{x}_{sT} - \boldsymbol{X}_s'\boldsymbol{s}_s(\boldsymbol{\theta})),\\ g_3^*(\boldsymbol{\theta}) =&\mathrm{tr}\,\Big[\big(\frac{\partial \boldsymbol{s}_s(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\big)\boldsymbol{\Sigma}_s(\boldsymbol{\theta})\big(\frac{\partial \boldsymbol{s}_s(\boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\Big]. \end{aligned}$$

It can be verified that the expressions of $g_1(\boldsymbol{\theta})$ and $g_2(\boldsymbol{\theta})$ are identical to those given by Rao and Yu (1994) since $\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\boldsymbol{c}_T = (1-\theta_3^2)^{-1}$. Since $\boldsymbol{\Psi}(\theta_3) = (1-\theta_3^2)^{-1}\mathbf{mat}_{i,j}(\theta_3^{|i-j|})$, it can be seen that

$$\begin{aligned} \boldsymbol{\Psi}_{(3)}(\theta_3) = \frac{\partial \boldsymbol{\Psi}(\theta_3)}{\partial \theta_3} =&\frac{2\theta_3}{1-\theta_3^2}\boldsymbol{\Psi}(\theta_3) + \frac{1}{1-\theta_3^2}\mathbf{mat}_{i,j}\big(|i-j|\theta_3^{|i-j|-1}\big),\\ \boldsymbol{\Psi}_{(33)}(\theta_3) = \frac{\partial^2 \boldsymbol{\Psi}(\theta_3)}{\partial \theta_3 \partial \theta_3} =&\frac{2(1+3\theta_3^2)}{(1-\theta_3^2)^2}\boldsymbol{\Psi}(\theta_3) + \frac{4\theta_3}{(1-\theta_3^2)^2}\mathbf{mat}_{i,j}\big(|i-j|\theta_3^{|i-j|-1}\big)\\ &+ \frac{1}{1-\theta_3^2}\mathbf{mat}_{i,j}\big(|i-j|(|i-j|-1)\theta_3^{|i-j|-2}\big). \end{aligned}$$

Then, $\partial \boldsymbol{s}_s(\boldsymbol{\theta})'/\partial \boldsymbol{\theta}$ in $g_3^*(\boldsymbol{\theta})$ can be derived by using the derivatives

$$\begin{aligned} \frac{\partial}{\partial \theta_1}\boldsymbol{s}_s(\boldsymbol{\theta})' = \boldsymbol{s}_{s(1)}(\boldsymbol{\theta})' =&\big(\boldsymbol{j}_T' - \boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{J}_T\big)\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}),\\ \frac{\partial}{\partial \theta_2}\boldsymbol{s}_s(\boldsymbol{\theta})' = \boldsymbol{s}_{s(2)}(\boldsymbol{\theta})' =&\frac{1}{\theta_2}\boldsymbol{s}_s(\boldsymbol{\theta})'(\theta_1\boldsymbol{J}_T + \boldsymbol{D}_s)\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}),\\ \frac{\partial}{\partial \theta_3}\boldsymbol{s}_s(\boldsymbol{\theta})' = \boldsymbol{s}_{s(3)}(\boldsymbol{\theta})' =&\theta_2\big(\boldsymbol{c}_T' - \boldsymbol{s}_s(\boldsymbol{\theta})'\big)\boldsymbol{\Psi}_{(3)}(\theta_3)\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta}). \end{aligned}$$

Also, $g_{11}(\boldsymbol{\theta})$ and $h(\boldsymbol{\theta})$ in (2.9) and (2.15) are

$$\begin{aligned} g_{11}(\boldsymbol{\theta}) =&\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)'E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}],\\ h(\boldsymbol{\theta}) =&\frac{z_\alpha^2+1}{8g_1(\boldsymbol{\theta})^2}\mathrm{tr}\,[\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)], \end{aligned}$$

where the derivatives of $g_1(\boldsymbol{\theta})$ are written as

$$\begin{aligned} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} =&1 - 2\boldsymbol{s}_{s(1)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}_s(\boldsymbol{\theta})^{-1}\boldsymbol{s}_s(\boldsymbol{\theta}) + \boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{J}_T\boldsymbol{s}_s(\boldsymbol{\theta}),\\ \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} =&\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\boldsymbol{c}_T - 2\boldsymbol{s}_{s(2)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}_s(\boldsymbol{\theta})\boldsymbol{s}_s(\boldsymbol{\theta}) + \boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{\Psi}(\theta_3)\boldsymbol{s}_s(\boldsymbol{\theta}),\\ \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_3} =&\theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}_{(3)}(\theta_3)\boldsymbol{c}_T - 2\boldsymbol{s}_{s(3)}(\boldsymbol{\theta})'\boldsymbol{\Sigma}_s(\boldsymbol{\theta})\boldsymbol{s}_s(\boldsymbol{\theta}) + \theta_2\boldsymbol{s}_s(\boldsymbol{\theta})'\boldsymbol{\Psi}_{(3)}(\theta_3)\boldsymbol{s}_s(\boldsymbol{\theta}). \end{aligned}$$

For the function $g_{12}(\boldsymbol{\theta})$ in (2.9), it is given by

$$g_{12}(\boldsymbol{\theta}) = \frac{1}{2}\mathrm{tr}\,\big[\boldsymbol{B}(\boldsymbol{\theta})\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\big],$$

where the (i,j)-th element of $\boldsymbol{B}(\boldsymbol{\theta})$ is given by $(\boldsymbol{b}_s - \boldsymbol{Z}_0'\boldsymbol{s}_s(\boldsymbol{\theta}))'(\partial_{ij}\boldsymbol{G}_0(\boldsymbol{\theta}))(\boldsymbol{b}_s - \boldsymbol{Z}_0'\boldsymbol{s}_s(\boldsymbol{\theta}))$. It is here noted that $\partial_{11}\boldsymbol{G}_0(\boldsymbol{\theta}) = \partial_{12}\boldsymbol{G}_0(\boldsymbol{\theta}) = \partial_{13}\boldsymbol{G}_0(\boldsymbol{\theta}) = \partial_{22}\boldsymbol{G}_0(\boldsymbol{\theta}) = 0$, $\partial_{23}\boldsymbol{G}_0(\boldsymbol{\theta}) = \text{block diag}(0, \boldsymbol{\Psi}_{(3)}(\theta_3))$ and $\partial_{33}\boldsymbol{G}_0(\boldsymbol{\theta}) = \text{block diag}(0, \theta_2\boldsymbol{\Psi}_{(33)}(\theta_3))$. Also note that $\boldsymbol{b}_s - \boldsymbol{Z}_0'\boldsymbol{s}_s(\boldsymbol{\theta}) = \big(1 - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\boldsymbol{j}_T, \boldsymbol{c}_T' - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\big)$. Thus, it is observed that $B_{11} = B_{12} = B_{13} = B_{22} = 0$ and

$$
\begin{aligned}
B_{23} =& \big(\boldsymbol{c}_T' - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\big)\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\}\boldsymbol{\Psi}_{(3)}(\theta_3)\big(\boldsymbol{c}_T' - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\big)',\\
B_{33} =& \big(\boldsymbol{c}_T' - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\big)\theta_2\boldsymbol{\Psi}_{(33)}(\theta_3)\big(\boldsymbol{c}_T' - \{\theta_1\boldsymbol{j}_T' + \theta_2\boldsymbol{c}_T'\boldsymbol{\Psi}(\theta_3)\}\boldsymbol{\Sigma}_s^{-1}(\boldsymbol{\theta})\big)'.
\end{aligned}
$$

In this model, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is not linear in θ_3, so that the function $g_{12}(\boldsymbol{\theta})$ cannot be ignored. Hence, we can compute the requested functions provided $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)$ and $E[\widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**}]$ can be derived for estimator $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$.

[REML estimator] We use the REML estimator $\widehat{\boldsymbol{\theta}}^R = (\hat{\theta}_1^R, \hat{\theta}_2^R, \hat{\theta}_3^R)'$ defined in (3.2), which is the solution of the equations $L_i^R(\widehat{\boldsymbol{\theta}}^R) = 0$ for $i = 1, 2, 3$, where

$$
L_i^R(\boldsymbol{\theta}) = \boldsymbol{y}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{y} - \text{tr}\,[(\boldsymbol{\Sigma}^{-1} - \boldsymbol{P})\boldsymbol{\Sigma}_{(i)}].
$$

Here,

$$
\boldsymbol{\Sigma}_{(1)} = \boldsymbol{I}_k \otimes \boldsymbol{J}_T,\ \boldsymbol{\Sigma}_{(2)} = \boldsymbol{I}_k \otimes \boldsymbol{\Psi}(\theta_3),\ \boldsymbol{\Sigma}_{(3)} = \boldsymbol{I}_k \otimes \theta_2\boldsymbol{\Psi}_{(3)}(\theta_3),
$$

since $\boldsymbol{\Sigma}_{\ell(1)}(\boldsymbol{\theta}) = \boldsymbol{J}_T$, $\boldsymbol{\Sigma}_{\ell(2)}(\boldsymbol{\theta}) = \boldsymbol{\Psi}(\theta_3)$ and $\boldsymbol{\Sigma}_{\ell(3)}(\boldsymbol{\theta}) = \theta_2\boldsymbol{\Psi}_{(3)}(\theta_3)$ for $\boldsymbol{\Sigma}_\ell(\boldsymbol{\theta}) = \theta_1\boldsymbol{J}_T + \theta_2\boldsymbol{\Psi}(\theta_3) + \boldsymbol{D}_\ell$. From Proposition 3.2, it follows that $\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^{R*}) = 2\boldsymbol{A}_2^{-1}$ and $E[\widehat{\boldsymbol{\theta}}^{R*}] = \mathbf{0}$, where $\boldsymbol{A}_2 = \mathbf{mat}_{i,j}(\text{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}])$ can be computed by using

$$
\text{tr}\,[\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma}^{-1}] = \sum_{\ell=1}^{k} \text{tr}\,[\boldsymbol{\Sigma}_{\ell(i)}\boldsymbol{\Sigma}_\ell^{-1}\boldsymbol{\Sigma}_{\ell(j)}\boldsymbol{\Sigma}_\ell^{-1}].
$$

Finally, we need to compute the expectation $E[\widehat{\boldsymbol{\theta}}^{R**}]$ given by

$$
E[\widehat{\boldsymbol{\theta}}^{R**}] = -\boldsymbol{A}_2^{-1}\mathbf{col}_i(\text{tr}\,[\boldsymbol{A}_2^{-1}\mathbf{mat}_{a,b}(\text{tr}\,[\boldsymbol{\Sigma}_{(ab)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}])]).
$$

Since $\boldsymbol{\Sigma}_{\ell(11)} = \boldsymbol{\Sigma}_{\ell(12)} = \boldsymbol{\Sigma}_{\ell(13)} = \boldsymbol{\Sigma}_{\ell(22)} = 0$, $\boldsymbol{\Sigma}_{\ell(23)} = \boldsymbol{\Psi}_{(3)}(\theta_3)$ and $\boldsymbol{\Sigma}_{\ell(33)} = \theta_2\boldsymbol{\Psi}_{(33)}(\theta_3)$, it can be seen that

$$
\begin{aligned}
&\text{tr}\,[\boldsymbol{A}_2^{-1}\mathbf{mat}_{a,b}(\text{tr}\,[\boldsymbol{\Sigma}_{(ab)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}^{-1}])]\\
&\quad = 2A_2^{23}\sum_{\ell=1}^{k}\text{tr}\,[\boldsymbol{\Psi}_{(3)}(\theta_3)\boldsymbol{\Sigma}_\ell^{-1}\boldsymbol{\Sigma}_{\ell(i)}\boldsymbol{\Sigma}_\ell^{-1}] + A_2^{33}\sum_{\ell=1}^{k}\text{tr}\,[\theta_2\boldsymbol{\Psi}_{(33)}(\theta_3)\boldsymbol{\Sigma}_\ell^{-1}\boldsymbol{\Sigma}_{\ell(i)}\boldsymbol{\Sigma}_\ell^{-1}],
\end{aligned}
$$

where A_2^{ij} denotes the (i,j)-th element of $\boldsymbol{A}_2^{-1}$.

The conditions (A2), (A5), (M2) and (M3) can be seen to be satisfied. Assuming the conditions (A1), (M1), $\sup_{i\ge1,t\ge1} d_{it} < \infty$, $\inf_{i\ge1,t\ge1} d_{it} > 0$ and that T is bounded and $k \to \infty$, we can see that the results of Proposition 3.2 hold, namely, we obtain the results given in Theorems 2.4-2.6.

4 Proofs

Proof of Theorems 2.1 and 2.4. We begin by proving Theorem 2.4, namely, the third-order approximation given in (2.18) under the conditions (A1)-(A6).

Following Prasad and Rao (1990) and Datta and Lahiri (2000), the MSE of $\widehat{\mu}^{EB}(\widehat{\psi})$ can be written as $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\psi})) = E[\{\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) - \mu\}^2] + E[\{\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta})\}^2] + E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2]$, and the first two terms are expressed as

$$\begin{aligned} E[\{\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}) - \mu\}^2] =& \boldsymbol{b}'(\boldsymbol{G}(\boldsymbol{\theta})^{-1} + \boldsymbol{Z}'\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{Z})^{-1}\boldsymbol{b} = g_1(\boldsymbol{\theta}), \\ E[\{\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta})\}^2] =& (\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta}))'(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}(\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s}(\boldsymbol{\theta})) = g_2(\boldsymbol{\theta}). \end{aligned}$$

Let $g_3(\boldsymbol{\theta}) = E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2]$. From the Taylor series expansion, it follows that

$$\begin{aligned} \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) =& \widehat{\mu}^{EB}(\boldsymbol{\theta}) + \sum_i \widehat{\mu}^{EB}_{(i)}(\boldsymbol{\theta})(\hat{\theta}_i - \theta_i) + \frac{1}{2}\sum_{i,j} \widehat{\mu}^{EB}_{(ij)}(\boldsymbol{\theta})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \\ & + \frac{1}{6}\sum_{i,j,k} \widehat{\mu}^{EB}_{(ijk)}(\widetilde{\boldsymbol{\theta}})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k), \end{aligned} \tag{4.1}$$

where $\widetilde{\boldsymbol{\theta}}$ is a point satisfying $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$ for the Euclidean norm $\|\cdot\|$, and $\widehat{\mu}^{EB}_{(i)}(\boldsymbol{\theta}) = \partial\widehat{\mu}^{EB}(\boldsymbol{\theta})/\partial\theta_i$ and $\widehat{\mu}^{EB}_{(ij)}(\boldsymbol{\theta})$ and $\widehat{\mu}^{EB}_{(ijk)}(\widetilde{\boldsymbol{\theta}})$ are defined similarly. Also $\sum_{i,j,k}$ means summation over $1 \leq i,j,k \leq q$, and $\sum_i$ and $\sum_{i,j}$ are defined similarly.

For notational simplicity, hereafter we omit $(\boldsymbol{\theta})$ in $\widehat{\mu}^{EB}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta})$, $\widehat{\boldsymbol{\beta}}_{(i)}(\boldsymbol{\theta})$, $\boldsymbol{s}(\boldsymbol{\theta})$ and others. Since $\widehat{\mu}^{EB}_{(ijk)}(\widetilde{\boldsymbol{\theta}}) = O_p(1)$ from (A1), (A2) and (A5), note that $\sum_{i,j,k} \widehat{\mu}^{EB}_{(ijk)}(\widetilde{\boldsymbol{\theta}})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k) = O_p(N^{-3/2})$. Then, $g_3(\boldsymbol{\theta})$ can be estimated as

$$\begin{aligned} g_3(\boldsymbol{\theta}) =& E\Big[\sum_{i,j} \widehat{\mu}^{EB}_{(i)}\widehat{\mu}^{EB}_{(j)}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) + \sum_{i,j,k} \widehat{\mu}^{EB}_{(ij)}\widehat{\mu}^{EB}_{(k)}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)\Big] \\ & + O(N^{-2}). \end{aligned} \tag{4.2}$$

It is observed that

$$\begin{aligned} \widehat{\mu}^{EB}_{(i)} =& (\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{s}'_{(i)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{s}'_{(i)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \\ \widehat{\mu}^{EB}_{(ij)} =& (\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(ij)} - \boldsymbol{s}'_{(ij)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \boldsymbol{s}'_{(ij)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ & - \boldsymbol{s}'_{(i)}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{(i)}, \end{aligned}$$

and that $\widehat{\boldsymbol{\beta}}_{(i)} = O_p(N^{-1/2})$ and $\widehat{\boldsymbol{\beta}}_{(ij)} = O_p(N^{-1/2})$ from (A1) and (A2). These facts are used to evaluate $g_3(\boldsymbol{\theta})$ as

$$\begin{aligned} g_3(\boldsymbol{\theta}) =& E\Big[\sum_{i,j} \boldsymbol{s}'_{(i)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{s}'_{(j)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \\ & + 2\sum_{i,j} \boldsymbol{s}'_{(i)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\big\{(\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big\}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \\ & + \sum_{i,j,k} \boldsymbol{s}'_{(k)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{s}'_{(ij)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)\Big] + O(N^{-2}). \end{aligned}$$

Since $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is expanded as $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^* + \widehat{\boldsymbol{\theta}}^{**} + \widehat{\boldsymbol{\theta}}^{***} + O_p(N^{-2})$ from (A6), $g_3(\boldsymbol{\theta})$ can be further approximated as

$$
\begin{aligned}
g_3(\boldsymbol{\theta}) =& \sum_{i,j} E[\boldsymbol{s}'_{(i)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{s}'_{(j)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\hat{\theta}_i^*\hat{\theta}_j^* + 2\hat{\theta}_i^*\hat{\theta}_j^{**})] \\
&+ 2\sum_{i,j} E[\boldsymbol{s}'_{(i)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\big\{(\boldsymbol{a} - \boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big\}\hat{\theta}_i^*\hat{\theta}_j^*] \\
&+ \sum_{i,j,k} E[\boldsymbol{s}'_{(k)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\boldsymbol{s}'_{(ij)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*] + O(N^{-2}) \\
=& I_1 + 2I_2 + I_3 + O(N^{-2}). \qquad \text{(say)}
\end{aligned}
\tag{4.3}
$$

To estimate the first term I_1, we use the following Stein identity given by Stein (1973) for $\boldsymbol{y} \sim \mathcal{N}_N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$:

$$
E[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{g}(\boldsymbol{y})] = E[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma}\boldsymbol{g}(\boldsymbol{y})\}], \tag{4.4}
$$

where $\boldsymbol{g}(\boldsymbol{y}) = (g_1(\boldsymbol{y}), \ldots, g_N(\boldsymbol{y}))'$ is an absolutely continuous function and $\boldsymbol{\nabla}_y$ is the differential operator defined by $\boldsymbol{\nabla}_y = \partial/\partial\boldsymbol{y}$. For example, let $\boldsymbol{A}$ be an $N \times N$ matrix independent of $\boldsymbol{y}$, and let $a(\boldsymbol{y})$ be a scalar function which is twice-differentiable with respect to $\boldsymbol{y}$. Then the Stein identity is used to get that

$$
\begin{aligned}
E[\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}\, a(\boldsymbol{y})] =& E[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{u}\, a(\boldsymbol{y})\}] \\
=& \mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{A}]E[a(\boldsymbol{y})] + E[\boldsymbol{u}'\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{\nabla}_y a(\boldsymbol{y})],
\end{aligned}
$$

for $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. Applying the Stein identity to the second term gives that

$$
E[\boldsymbol{u}'\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{\nabla}_y a(\boldsymbol{y})] = E[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma}\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{\nabla}_y a(\boldsymbol{y})\}] = E[\mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{\nabla}_y\boldsymbol{\nabla}'_y a(\boldsymbol{y})]],
$$

which yields the useful equality

$$
E[\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}a(\boldsymbol{y})] = \mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{A}]E[a(\boldsymbol{y})] + \mathrm{tr}\,[\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}E[\boldsymbol{\nabla}\boldsymbol{\nabla}'a(\boldsymbol{y})]]. \tag{4.5}
$$

Using the Stein identity, we can see that

$$
\begin{aligned}
I_1 =& \sum_{i,j} E\big[\boldsymbol{\nabla}'_y\big\{\boldsymbol{\Sigma}\boldsymbol{s}_{(i)}\boldsymbol{s}'_{(j)}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})(\hat{\theta}_i^*\hat{\theta}_j^* + 2\hat{\theta}_i^*\hat{\theta}_j^{**})\big\}\big] \\
=& \sum_{i,j} \mathrm{tr}\,\big\{\boldsymbol{\Sigma}\boldsymbol{s}_{(i)}\boldsymbol{s}'_{(j)}\big\}E\big[\hat{\theta}_i^*\hat{\theta}_j^* + 2\hat{\theta}_i^*\hat{\theta}_j^{**}\big] + \sum_{i,j} E\big[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(i)}\boldsymbol{s}'_{(j)}\boldsymbol{\Sigma}\boldsymbol{\nabla}_y(\hat{\theta}_i^*\hat{\theta}_j^*)\big] \\
&+ 2\sum_{i,j} E\big[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{s}_{(i)}\boldsymbol{s}'_{(j)}\boldsymbol{\Sigma}\boldsymbol{\nabla}_y(\hat{\theta}_i^*\hat{\theta}_j^{**})\big] \\
=& I_{11} + I_{12} + 2I_{13}. \qquad \text{(say)}
\end{aligned}
$$

Since $E[\hat{\theta}_i^*] = O(N^{-1})$ and $E[\hat{\theta}_i^*\hat{\theta}_j^{**}] = O(N^{-2})$ from (A4) and (A6), it can be seen that

$$
I_{11} = \mathrm{tr}\,\Big[\big(\frac{\partial\boldsymbol{s}'}{\partial\boldsymbol{\theta}}\big)\boldsymbol{\Sigma}\big(\frac{\partial\boldsymbol{s}'}{\partial\boldsymbol{\theta}}\big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)\Big] + O(N^{-2}) \equiv g_3^*(\boldsymbol{\theta}) + O(N^{-2}).
$$

For I_{12}, the Stein identity is applied again to rewrite it as

$$I_{12} = \sum_{i,j} E[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma s}_{(j)}\boldsymbol{s}'_{(i)}\boldsymbol{\Sigma\nabla}_y(\hat{\theta}_i^*\hat{\theta}_j^*)] = \sum_{i,j} \mathrm{tr}\left[\boldsymbol{\Sigma s}_{(j)}\boldsymbol{s}'_{(i)}\boldsymbol{\Sigma}E[\boldsymbol{\nabla}_y\boldsymbol{\nabla}'_y(\hat{\theta}_i^*\hat{\theta}_j^*)]\right]$$
$$=2\sum_{i,j} \boldsymbol{s}'_{(i)}\boldsymbol{\Sigma}E[(\boldsymbol{\nabla}_y\boldsymbol{\nabla}'_y\hat{\theta}_i^*)\hat{\theta}_j^* + (\boldsymbol{\nabla}_y\hat{\theta}_i^*)(\boldsymbol{\nabla}'_y\hat{\theta}_j^*)]\boldsymbol{\Sigma s}_{(j)},$$

which is of order $O(N^{-2})$ as seen from the condition (A4) and (A6) (ii). Also,

$$I_{13} = \sum_{i,j} E\big[(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{s}_{(i)}\boldsymbol{s}'_{(j)}\boldsymbol{\Sigma}\{(\boldsymbol{\nabla}_y\hat{\theta}_i^*)\hat{\theta}_j^{**} + (\boldsymbol{\nabla}_y\hat{\theta}_j^{**})\hat{\theta}_i^*\big],$$

which is of order $O(N^{-2})$ from (A4) and (A6)(ii).

The similar arguments can be used to evaluate the other terms. For I_2, it is observed that

$$I_2 = \sum_{i,j} E\big[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma s}_{(i)}\big[(\boldsymbol{a}-\boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big]\hat{\theta}_i^*\hat{\theta}_j^*\}\big]$$
$$= \sum_{i,j} E\big[\boldsymbol{s}'_{(i)}\boldsymbol{\Sigma\nabla}_y\{\big[(\boldsymbol{a}-\boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\big]\hat{\theta}_i^*\hat{\theta}_j^*\}\big]$$
$$= \sum_{i,j} \boldsymbol{s}'_{(i)}\boldsymbol{\Sigma}\big\{\partial_j\{\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\}(\boldsymbol{a}-\boldsymbol{X}'\boldsymbol{s}) - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{s}_{(j)}\big\}E[\hat{\theta}_i^*\hat{\theta}_j^*]$$
$$+ \sum_{i,j} E\big[\boldsymbol{s}'_{(i)}\boldsymbol{\Sigma}\{(\boldsymbol{\nabla}_y\hat{\theta}_i^*)\hat{\theta}_j^* + (\boldsymbol{\nabla}_y\hat{\theta}_j^*)\hat{\theta}_i^*\}\{(\boldsymbol{a}-\boldsymbol{X}'\boldsymbol{s})'\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{s}'_{(j)}\boldsymbol{X}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\}\big],$$

which is of order $O(N^{-2})$ as seen from (A1), (A2) and (A4), since $\widehat{\boldsymbol{\beta}}_{(j)} = O_p(N^{-1/2})$ and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(N^{-1/2})$.

Finally, I_3 can be evaluated as

$$I_3 = \sum_{i,j,k} E[\boldsymbol{\nabla}'_y\{\boldsymbol{\Sigma s}_{(k)}\boldsymbol{s}'_{(ij)}(\boldsymbol{y}-\boldsymbol{X\beta})\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*\}]$$
$$= \sum_{i,j,k} E[\boldsymbol{s}'_{(k)}\boldsymbol{\Sigma\nabla}_y\{(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{s}_{(ij)}\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*\}]$$
$$= \sum_{i,j,k} \boldsymbol{s}'_{(k)}\boldsymbol{\Sigma s}_{(ij)}E[\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*]$$
$$+ \sum_{i,j,k} E[\boldsymbol{s}'_{(k)}\boldsymbol{\Sigma}\{(\boldsymbol{\nabla}_y\hat{\theta}_i^*)\hat{\theta}_j^*\hat{\theta}_k^* + (\boldsymbol{\nabla}_y\hat{\theta}_j^*)\hat{\theta}_k^*\hat{\theta}_i^* + (\boldsymbol{\nabla}_y\hat{\theta}_k^*)\hat{\theta}_i^*\hat{\theta}_j^*\}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{s}_{(ij)}], \quad (4.6)$$

which is of order $O(N^{-2})$ as seen from (A2) and (A4). Hence it is concluded that $MSE(\boldsymbol{\theta}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}})) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3^*(\boldsymbol{\theta}) + O(N^{-2})$, and the proof of Theorem 2.4 is complete.

For the proof of Theorem 2.1, the same arguments as given above are used. Especially,

$g_3(\boldsymbol{\theta})$ can be evaluated as

$$\begin{aligned}
g_3(\boldsymbol{\theta}) =& \sum_{i,j} E\Big[\widehat{\mu}_{(i)}^{EB}\widehat{\mu}_{(j)}^{EB}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)\big] + O(N^{-3/2}) \\
=& E\Big[\sum_{i,j} \boldsymbol{s}_{(i)}'(\boldsymbol{y} - \boldsymbol{X\beta})\boldsymbol{s}_{(j)}'(\boldsymbol{y} - \boldsymbol{X\beta})(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) + O(N^{-3/2}) \\
=& \sum_{i,j} E[\boldsymbol{s}_{(i)}'(\boldsymbol{y} - \boldsymbol{X\beta})\boldsymbol{s}_{(j)}'(\boldsymbol{y} - \boldsymbol{X\beta})(\hat{\theta}_i^*\hat{\theta}_j^*)] + O(N^{-3/2}).
\end{aligned}$$

Then the Stein identity can be applied to the above expectation and it is rewritten as

$$\begin{aligned}
g_3(\boldsymbol{\theta}) =& \sum_{i,j} E\big[\boldsymbol{\nabla}_y'\{\boldsymbol{\Sigma s}_{(i)}\boldsymbol{s}_{(j)}'(\boldsymbol{y} - \boldsymbol{X\beta})(\hat{\theta}_i^*\hat{\theta}_j^*)\}\big] + O(N^{-3/2}) \\
=& \sum_{i,j} \mathrm{tr}\,\{\boldsymbol{\Sigma s}_{(i)}\boldsymbol{s}_{(j)}'\}E\big[\hat{\theta}_i^*\hat{\theta}_j^*\big] + \sum_{i,j} E\big[(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{s}_{(i)}\boldsymbol{s}_{(j)}'\boldsymbol{\Sigma\nabla}_y(\hat{\theta}_i^*\hat{\theta}_j^*)\big] + O(N^{-3/2}) \\
=& g_3^*(\boldsymbol{\theta}) + O(N^{-3/2}),
\end{aligned}$$

which proves Theorem 2.1. ■

Proof of Theorems 2.2 and 2.5. We shall prove Theorem 2.5. It is noted that $g_2(\widehat{\boldsymbol{\theta}})$ and $g_3^*(\widehat{\boldsymbol{\theta}})$ are of order $O_p(N^{-1})$, while $g_1(\widehat{\boldsymbol{\theta}}) = O_p(1)$. Since $g_1(\widehat{\boldsymbol{\theta}})$ is not a second-order unbiased estimator of $g_1(\boldsymbol{\theta})$, we need to approximate the expectation $E[g_1(\widehat{\boldsymbol{\theta}})]$. From the Taylor expansion of $g_1(\widehat{\boldsymbol{\theta}})$ around $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, it follows that

$$\begin{aligned}
E[g_1(\widehat{\boldsymbol{\theta}})] =& g_1(\boldsymbol{\theta}) + \sum_i \{\partial_i g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i - \theta_i] + \frac{1}{2}\sum_{i,j}\{\partial_{ij} g_1(\boldsymbol{\theta})\}E[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)] \\
&+ \frac{1}{6}\sum_{i,j,k}\{\partial_{ijk} g_1(\boldsymbol{\theta})\}E[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)] \\
&+ \frac{1}{24}\sum_{i,j,k,\ell}\{\partial_{ijk\ell} g_1(\widetilde{\boldsymbol{\theta}})\}E[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)(\hat{\theta}_\ell - \theta_\ell)],
\end{aligned}$$

where $\widetilde{\boldsymbol{\theta}}$ is a point satisfying $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$. Since $g_1(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[4]}$ and $\partial_{ijk\ell} g_1(\boldsymbol{\theta}) = O(1)$, it can be further approximated as

$$\begin{aligned}
E[g_1(\widehat{\boldsymbol{\theta}})] =& g_1(\boldsymbol{\theta}) + \sum_i \{\partial_i g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^* + \hat{\theta}_i^{**} + \hat{\theta}_i^{***}] \\
&+ \frac{1}{2}\sum_{i,j}\{\partial_{ij} g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^*\hat{\theta}_j^* + \hat{\theta}_i^*\hat{\theta}_j^{**} + \hat{\theta}_j^*\hat{\theta}_i^{**}] \\
&+ \frac{1}{6}\sum_{i,j,k}\{\partial_{ijk} g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*] + O(N^{-2}). \qquad (4.7)
\end{aligned}$$

Let $g_{11}(\boldsymbol{\theta}) = \sum_i \{\partial_i g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^* + \hat{\theta}_i^{**}]$. Note that $\partial_i g_1(\boldsymbol{\theta}) = O(1)$, $\partial_{ij} g_1(\boldsymbol{\theta}) = O(1)$ and $\partial_{ijk} g_1(\boldsymbol{\theta}) = O(1)$. Since $E[\hat{\theta}_i^*\hat{\theta}_j^{**}] = O(N^{-2})$ and $E[\hat{\theta}_i^{***}] = O(N^{-2})$ from (A6) and (B2), it follows that

$$E[g_1(\widehat{\boldsymbol{\theta}})] = g_1(\boldsymbol{\theta}) + g_{11}(\boldsymbol{\theta}) + \frac{1}{2}\sum_{i,j}\{\partial_{ij}g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^*\hat{\theta}_j^*] + O(N^{-2}). \tag{4.8}$$

To evaluate the third term in the r.h.s. of (4.8), we express $g_1(\boldsymbol{\theta})$ as $g_1(\boldsymbol{\theta}) = \boldsymbol{b}'\boldsymbol{G}\boldsymbol{b} - (\boldsymbol{\Sigma s})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma s})$. Then,

$$\partial_j g_1(\boldsymbol{\theta}) = \boldsymbol{b}'\{\partial_j\boldsymbol{G}\}\boldsymbol{b} - \{\partial_j\boldsymbol{\Sigma s}\}'\boldsymbol{s} - \boldsymbol{s}'\{\partial_j\boldsymbol{\Sigma s}\} + \boldsymbol{s}'\{\partial_j\boldsymbol{\Sigma}\}\boldsymbol{s},$$

which leads to

$$\begin{aligned}\partial_{ij}g_1(\boldsymbol{\theta}) =& \boldsymbol{b}'\{\partial_{ij}\boldsymbol{G}\}\boldsymbol{b} - \{\partial_{ij}\boldsymbol{\Sigma s}\}'\boldsymbol{s} - \{\partial_j\boldsymbol{\Sigma s}\}'\{\partial_i\boldsymbol{s}\} - \boldsymbol{s}'\{\partial_{ij}\boldsymbol{\Sigma s}\} - \{\partial_i\boldsymbol{s}\}'\{\partial_j\boldsymbol{\Sigma s}\}\\ &+ \{\partial_i\boldsymbol{s}\}'\{\partial_j\boldsymbol{\Sigma}\}\boldsymbol{s} + \boldsymbol{s}'\{\partial_{ij}\boldsymbol{\Sigma}\}\boldsymbol{s} + \boldsymbol{s}'\{\partial_j\boldsymbol{\Sigma}\}\{\partial_i\boldsymbol{s}\}\\ =& -\left[\{\partial_i\boldsymbol{s}\}'\boldsymbol{\Sigma}\{\partial_j\boldsymbol{s}\} + \{\partial_j\boldsymbol{s}\}'\boldsymbol{\Sigma}\{\partial_i\boldsymbol{s}\}\right]\\ &+ \left[\boldsymbol{b}'\{\partial_{ij}\boldsymbol{G}\}\boldsymbol{b} + \boldsymbol{s}'\{\partial_{ij}\boldsymbol{\Sigma}\}\boldsymbol{s} - \boldsymbol{b}'\{\partial_{ij}\boldsymbol{G}\}'\boldsymbol{Z}'\boldsymbol{s} - \boldsymbol{s}'\boldsymbol{Z}\{\partial_{ij}\boldsymbol{G}\}\boldsymbol{b}\right],\end{aligned}$$

since $\boldsymbol{\Sigma s} = \boldsymbol{ZGb}$. Note that $\boldsymbol{b}'\{\partial_{ij}\boldsymbol{G}\}\boldsymbol{b} + \boldsymbol{s}'\{\partial_{ij}\boldsymbol{\Sigma}\}\boldsymbol{s} - \boldsymbol{b}'\{\partial_{ij}\boldsymbol{G}\}'\boldsymbol{Z}'\boldsymbol{s} - \boldsymbol{s}'\boldsymbol{Z}\{\partial_{ij}\boldsymbol{G}\}\boldsymbol{b} = (\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta}))'(\partial_{ij}\boldsymbol{G}(\boldsymbol{\theta}))(\boldsymbol{b} - \boldsymbol{Z}'\boldsymbol{s}(\boldsymbol{\theta})) + \boldsymbol{s}(\boldsymbol{\theta})'(\partial_{ij}\boldsymbol{R}(\boldsymbol{\theta}))\boldsymbol{s}(\boldsymbol{\theta})$ which is equal to $(\boldsymbol{B}(\boldsymbol{\theta}))_{i,j}$ for $\boldsymbol{B}(\boldsymbol{\theta})$ defined in (2.10). Thus, the third term can be expressed as

$$\frac{1}{2}\sum_{i,j}\{\partial_{ij}g_1(\boldsymbol{\theta})\}E[\hat{\theta}_i^*\hat{\theta}_j^*] = -g_3^*(\boldsymbol{\theta}) + g_{12}(\boldsymbol{\theta}), \tag{4.9}$$

where $g_{12}(\boldsymbol{\theta})$ is defined in (2.9). Hence from (4.8), it can be seen that

$$E[g_1(\widehat{\boldsymbol{\theta}})] = g_1(\boldsymbol{\theta}) + g_{11}(\boldsymbol{\theta}) - g_3^*(\boldsymbol{\theta}) + g_{12}(\boldsymbol{\theta}) + O(N^{-2}) \tag{4.10}$$

Now we can evaluate the expectation of $g^{\#}(\widehat{\boldsymbol{\theta}})$ as

$$\begin{aligned}E[g^{\#}(\widehat{\boldsymbol{\theta}})] =& g^{\#}(\boldsymbol{\theta}) + \sum_i\{\partial_i g^{\#}(\boldsymbol{\theta})\}E[(\hat{\theta}_i - \theta_i)]\\ &+ \frac{1}{2}\sum_{i,j}E[\{\partial_{ij}g^{\#}(\widetilde{\boldsymbol{\theta}})\}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)],\end{aligned} \tag{4.11}$$

where $\widetilde{\boldsymbol{\theta}}$ is a point satisfying $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \le \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$. Since $g^{\#}(\boldsymbol{\theta}) \in \mathcal{C}_{\boldsymbol{\theta}}^{[2]}$, $\partial_i g^{\#}(\boldsymbol{\theta}) = O(N^{-1})$, $\partial_{ij}g^{\#}(\widetilde{\boldsymbol{\theta}}) = O_p(N^{-1})$ and $E[\hat{\theta}_i - \theta_i] = O(N^{-1})$, it can be seen that $E[g^{\#}(\widehat{\boldsymbol{\theta}})] = g^{\#}(\boldsymbol{\theta}) + O(N^{-2})$, so that from (4.10)

$$E[mse(\widehat{\boldsymbol{\theta}}, \widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}))] = E[g_1(\widehat{\boldsymbol{\theta}})] + g^{\#}(\boldsymbol{\theta}) + O(N^{-2}) = g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta}) + g_3^*(\boldsymbol{\theta}) + O(N^{-2}),$$

and the proof of Theorem 2.5 is complete. Theorem 2.2 can be similarly proved. ■

Proof of Theorems 2.3 and 2.6. From (2.2), the conditional distribution of μ given $\boldsymbol{y}$ is given by

$$\mu|\boldsymbol{y} \sim \mathcal{N}(\widehat{\mu}^B(\boldsymbol{\beta}, \boldsymbol{\theta}), g_1(\boldsymbol{\theta})),$$

where $\widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta}) = \boldsymbol{a}'\boldsymbol{\beta} + \boldsymbol{s}(\boldsymbol{\theta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, and $g_1(\boldsymbol{\theta}) = \boldsymbol{b}'(\boldsymbol{G}(\boldsymbol{\theta})^{-1} + \boldsymbol{Z}'\boldsymbol{R}(\boldsymbol{\theta})^{-1}\boldsymbol{Z})^{-1}\boldsymbol{b}$ as given in (2.6). Then, the coverage probability of $I^{CEB}(\widehat{\boldsymbol{\theta}})$ is written as

$$
\begin{aligned}
P[\mu \in I^{CEB}(\widehat{\boldsymbol{\theta}})] =& P[-z + G(-z) < \frac{\mu - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta})}{\sqrt{g_1(\boldsymbol{\theta})}} < z + G(z)] \\
=& E[\Phi(z + G(z)) - \Phi(-z + G(-z))], \qquad (4.12)
\end{aligned}
$$

where $G(z) = U + zV$ for

$$
\begin{aligned}
U =& \{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta})\}/\sqrt{g_1(\boldsymbol{\theta})}, \\
V =& \{[1 + h(\widehat{\boldsymbol{\theta}})]\sqrt{mse(\widehat{\boldsymbol{\theta}})} - \sqrt{g_1(\boldsymbol{\theta})}\}/\sqrt{g_1(\boldsymbol{\theta})}.
\end{aligned} \qquad (4.13)
$$

It is noted that $G(z) = O_p(N^{-1/2})$ as seen below. Then, $\Phi(z + G(z))$ is evaluated as

$$
\begin{aligned}
\Phi(z + G(z)) =& \Phi(z) + G(z)\phi(z) + \frac{G^2(z)}{2}\phi'(z) + \frac{G^3(z)}{6}\phi''(z) \\
& - \frac{1}{6}\int_z^{z+G(z)} (z + G(z) - x)^3\phi'''(x)\mathrm{d}x \\
=& \Phi(z) + \left\{G(z) - \frac{z}{2}G^2(z) + \frac{z^2-1}{6}G^3(z)\right\}\phi(z) \\
& - \frac{1}{6}\int_z^{z+G(z)} (z + G(z) - x)^3(3 - x^2)x\phi(x)\mathrm{d}x.
\end{aligned}
$$

It can be verified that $\int_z^{z+G(z)} (z + G(z) - x)^3(3 - x^2)x\phi(x)\mathrm{d}x = O_p(N^{-2})$. From (4.12), it follows that

$$
P[\mu \in I^{CEB}(\widehat{\boldsymbol{\theta}})] = 1 - \alpha + \phi(z)H(\boldsymbol{\theta}) + O(N^{-2}),
$$

where

$$
H(\boldsymbol{\theta}) = E\big[G(z) - G(-z) - \frac{z}{2}\{G(z)^2 + G(-z)^2\} + \frac{z^2-1}{6}\{G(z)^3 - G(-z)^3\}\big],
$$

which can be rewritten as

$$
H(\boldsymbol{\theta}) = zE\big[2V - (U^2 + z^2V^2) + \frac{z^2-1}{3}(3U^2V + z^2V^3)\big]. \qquad (4.14)
$$

We thus need to show that $H(\boldsymbol{\theta}) = O(N^{-2})$. To this end, we shall verify that $E[2V - (U^2 + z^2V^2)] = O(N^{-2})$ and $E[3U^2V + z^2V^3] = O(N^{-2})$.

It is noted that U is rewritten as $U = \{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}/\sqrt{g_1(\boldsymbol{\theta})} + \{\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta})\}/\sqrt{g_1(\boldsymbol{\theta})}$. Since $\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta}) = (\boldsymbol{a}' - \boldsymbol{s}'\boldsymbol{X})(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) - \boldsymbol{\beta})$, from Kackar and Harville (1984), it follows that $\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta})$ is independent of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})$. Using the result given in Theorem 2.4, we can evaluate $E[U^2]$ and $E[U^2t(\widehat{\boldsymbol{\theta}})]$ for function $t(\cdot)$ as

$$
\begin{aligned}
E[U^2] =& E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2]/g_1(\boldsymbol{\theta}) + E[\{\widehat{\mu}^{EB}(\boldsymbol{\theta}) - \widehat{\mu}^B(\boldsymbol{\beta},\boldsymbol{\theta})\}^2]/g_1(\boldsymbol{\theta}) \\
=& \{g_3^*(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta})\}/g_1(\boldsymbol{\theta}) + O(N^{-2}), \\
E[U^2t(\widehat{\boldsymbol{\theta}})] =& E[\{\widehat{\mu}^{EB}(\widehat{\boldsymbol{\theta}}) - \widehat{\mu}^{EB}(\boldsymbol{\theta})\}^2t(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta})E[t(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}). \qquad (4.15)
\end{aligned}
$$

Also, note that

$$2V + V^2 = (2+V)V = \{1 + h(\widehat{\boldsymbol{\theta}})\}^2 mse(\widehat{\boldsymbol{\theta}})/g_1(\boldsymbol{\theta}) - 1$$
$$= \{g_1(\widehat{\boldsymbol{\theta}}) - g_1(\boldsymbol{\theta})\}/g_1(\boldsymbol{\theta}) + O_p(N^{-1}). \tag{4.16}$$

Since $h(\boldsymbol{\theta}) = O(N^{-1})$ and $g^{\#}(\boldsymbol{\theta}) = O(N^{-1})$, it is seen that

$$E[2V + V^2] = E[mse(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}) - 1 + 2E[h(\widehat{\boldsymbol{\theta}})g_1(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}) + O(N^{-2})$$
$$= \{g_3^*(\boldsymbol{\theta}) + g_2(\boldsymbol{\theta})\}/g_1(\boldsymbol{\theta}) + 2E[h(\widehat{\boldsymbol{\theta}})g_1(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}) + O(N^{-2}).$$

These observations are used to show that

$$E[2V - (U^2 + z^2V^2)] = E[(2V + V^2) - U^2 - (1+z^2)V^2]$$
$$= 2E[h(\widehat{\boldsymbol{\theta}})g_1(\widehat{\boldsymbol{\theta}})]/g_1(\boldsymbol{\theta}) - (1+z^2)E[V^2] + O(N^{-2})$$
$$= 2h(\boldsymbol{\theta}) - (1+z^2)E[V^2] + O(N^{-2}), \tag{4.17}$$

where in the third equality we used the same arguments as in (4.11) for evaluating $E[h(\widehat{\boldsymbol{\theta}})g_1(\widehat{\boldsymbol{\theta}})]$ under (C1) and (C2).

We now estimate the term $E[V^2]$. Since

$$V\sqrt{g_1(\boldsymbol{\theta})} = \{\sqrt{g_1(\widehat{\boldsymbol{\theta}})} - \sqrt{g_1(\boldsymbol{\theta})}\} + f(\widehat{\boldsymbol{\theta}}) \tag{4.18}$$

for $f(\boldsymbol{\theta}) = (1 + h(\boldsymbol{\theta}))\sqrt{mse(\boldsymbol{\theta})} - \sqrt{g_1(\boldsymbol{\theta})}\}$, we write $g_1(\boldsymbol{\theta})E[V^2]$ as

$$g_1(\boldsymbol{\theta})E[V^2] = E[\{\sqrt{g_1(\widehat{\boldsymbol{\theta}})} - \sqrt{g_1(\boldsymbol{\theta})}\}^2] + E[\{f(\widehat{\boldsymbol{\theta}})\}^2] + 2E[\{\sqrt{g_1(\widehat{\boldsymbol{\theta}})} - \sqrt{g_1(\boldsymbol{\theta})}\}f(\widehat{\boldsymbol{\theta}})]$$
$$= I_1 + I_2 + 2I_3. \qquad \text{(say)}$$

Noting that

$$\sqrt{g_1(\widehat{\boldsymbol{\theta}})} = \sqrt{g_1(\boldsymbol{\theta})} + \sum_i \{\partial_i \sqrt{g_1(\boldsymbol{\theta})}\}(\hat\theta_i - \theta_i) + \frac{1}{2}\sum_{i,j}\{\partial_{ij}\sqrt{g_1(\boldsymbol{\theta})}\}(\hat\theta_i - \theta_i)(\hat\theta_j - \theta_j)$$
$$+ \frac{1}{6}\sum_{i,j,k}\{\partial_{ijk}\sqrt{g_1(\widetilde{\boldsymbol{\theta}})}\}(\hat\theta_i - \theta_i)(\hat\theta_j - \theta_j)(\hat\theta_k - \theta_k), \tag{4.19}$$

we can see that

$$I_1 = E\Big[\sum_{i,j}\{\partial_i\sqrt{g_1(\boldsymbol{\theta})}\}\{\partial_j\sqrt{g_1(\boldsymbol{\theta})}\}(\hat\theta_i - \theta_i)(\hat\theta_j - \theta_j)$$
$$+ \sum_{i,j,k}\{\partial_k\sqrt{g_1(\boldsymbol{\theta})}\}\{\partial_{ij}\sqrt{g_1(\boldsymbol{\theta})}\}(\hat\theta_i - \theta_i)(\hat\theta_j - \theta_j)(\hat\theta_k - \theta_k)\Big] + O(N^{-2})$$
$$= E\Big[\sum_{i,j}\{\partial_i\sqrt{g_1(\boldsymbol{\theta})}\}\{\partial_j\sqrt{g_1(\boldsymbol{\theta})}\}(\hat\theta_i^*\hat\theta_j^* + 2\hat\theta_i^*\hat\theta_j^{**}) + \sum_{i,j,k}\{\partial_k\sqrt{g_1(\boldsymbol{\theta})}\}\{\partial_{ij}\sqrt{g_1(\boldsymbol{\theta})}\}\hat\theta_i^*\hat\theta_j^*\hat\theta_k^*\Big]$$
$$+ O(N^{-2}),$$

which can be approximated as $I_1 = \sum_{i,j}\{\partial_i\sqrt{g_1(\boldsymbol{\theta})}\}\{\partial_j\sqrt{g_1(\boldsymbol{\theta})}\}E[\hat{\theta}_i^*\hat{\theta}_j^*] + O(N^{-2})$. Since $f(\widehat{\boldsymbol{\theta}})$ can be expressed as

$$f(\widehat{\boldsymbol{\theta}}) = \frac{g^{\#}(\widehat{\boldsymbol{\theta}})}{\sqrt{mse(\widehat{\boldsymbol{\theta}})} + \sqrt{g_1(\widehat{\boldsymbol{\theta}})}} + h(\widehat{\boldsymbol{\theta}})\sqrt{mse(\widehat{\boldsymbol{\theta}})},$$

$g^{\#}(\widehat{\boldsymbol{\theta}}) = O_p(N^{-1})$ and $h(\widehat{\boldsymbol{\theta}}) = O_p(N^{-1})$, it is easy to see that $f(\widehat{\boldsymbol{\theta}}) = O_p(N^{-1})$, and $I_2 = E[\{f(\widehat{\boldsymbol{\theta}})\}^2] = O(N^{-2})$. For I_3, noting that $f(\widehat{\boldsymbol{\theta}}) = f(\boldsymbol{\theta}) + \sum_i(\partial_i f(\widetilde{\boldsymbol{\theta}}))(\hat{\theta}_i - \theta_i)$, from (4.19), it follows that $I_3 = f(\boldsymbol{\theta})\sum_i(\partial_i\sqrt{g_1(\boldsymbol{\theta})})E[\hat{\theta}_i - \theta_i] + O(N^{-2}) = O(N^{-2})$ since $f(\boldsymbol{\theta}) = O(N^{-1})$. Hence from (4.17), we get that

$$E[2V - (U^2 + z^2V^2)] = 2h(\boldsymbol{\theta}) - \frac{1+z^2}{4g_1(\boldsymbol{\theta})^2}\mathrm{tr}\,[\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)\big(\frac{\partial g_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big)'\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*)] + O(N^{-2}),$$

which has order $O(N^{-2})$ from the definition of $h(\boldsymbol{\theta})$ given in (2.15).

Finally, we need to show that $E[3U^2V + z^2V^3] = O(N^{-2})$ or $(3/2)E[U^2(2V + V^2)] - (3/2)E[U^2V^2] + z^2E[V^3] = O(N^{-2})$. From (4.1), (4.2), (4.15) and (4.16), it can be seen that

$$\begin{aligned} E[U^2(2V + V^2)] =& E\Big[\sum_{i,j,k}\widehat{\mu}_{(i)}^{EB}\widehat{\mu}_{(j)}^{EB}\{\partial_k g_1(\boldsymbol{\theta})\}(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)(\hat{\theta}_k - \theta_k)\Big]/g_1(\boldsymbol{\theta}) + O(N^{-2}) \\ =& \sum_{i,j,k}E[\boldsymbol{s}_{(i)}'(\boldsymbol{y} - \boldsymbol{X\beta})\boldsymbol{s}_{(j)}'(\boldsymbol{y} - \boldsymbol{X\beta})\{\partial_k g_1(\boldsymbol{\theta})\}\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*]/g_1(\boldsymbol{\theta}) + O(N^{-2}), \end{aligned}$$

which can be shown to be $O(N^{-2})$ similarly to (4.6). From (4.18) and (4.19), it can be verified that $E[U^2V^2] = O(N^{-2})$ and $E[V^3] = O(N^{-2})$. Therefore, the third-order approximation given in (2.20) is proved.

For the proof of Theorem 2.3, from (4.14), it is noted that

$$H(\boldsymbol{\theta}) = zE\big[2V - (U^2 + z^2V^2)\big] + O(N^{-3/2}),$$

so that we need to show that $E[2V - (U^2 + z^2V^2)] = O(N^{-3/2})$. This can be shown by using the same arguments as used above. ■

Proof of Proposition 3.1. We shall prove part (i) of Proposition 3.1, and the proof of part (ii) is omitted. In this proof, we omit the index M in $\widehat{\boldsymbol{\theta}}^M$, $\widehat{\boldsymbol{\theta}}^{M*}$ and others for the sake of simplicity. Since $\hat{\theta}_i^*$ is written as $\hat{\theta}_i^* = -\sum_a A_2^{ia}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}(\boldsymbol{yy}' - \boldsymbol{\Sigma})]$ where A_2^{ij} is the (i,j)-th element of the inverse matrix $\boldsymbol{A}_2^{-1} = (A_2^{ij})$, it is easy to see that $E[\hat{\theta}_i^*] = 0$ and

$$\begin{aligned} E[\hat{\theta}_i^*\hat{\theta}_j^*] =& \sum_{a,b}A_2^{ia}A_2^{jb}E[\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}(\boldsymbol{yy}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{yy}' - \boldsymbol{\Sigma})]] \\ =& 2\sum_{a,b}A_2^{ia}A_2^{jb}(A_2)_{ab} = 2A_2^{ij}, \end{aligned}$$

where $\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}] = -\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{\Sigma}_{(b)}] = (A_2)_{ab}$. Thus,

$$\mathbf{Cov}\,(\widehat{\boldsymbol{\theta}}^*) = \mathbf{mat}_{ij}(E[\hat{\theta}_i^*\hat{\theta}_j^*]) = 2\boldsymbol{A}_2^{-1}. \tag{4.20}$$

Similarly,

$$\begin{aligned} E[\hat{\theta}_i^*\hat{\theta}_j^*\hat{\theta}_k^*] =& \sum_{a,b,c} A_2^{ia}A_2^{jb}A_2^{kc}E\big[\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] \\ & \times \mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(c)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big] \\ =& 8\sum_{a,b,c} A_2^{ia}A_2^{jb}A_2^{kc}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(c)}\boldsymbol{\Sigma}] = O(N^{-2}). \end{aligned}$$

For $E[\hat{\theta}_i^{**}]$, it is noted that

$$\begin{aligned} \hat{\theta}_i^{**} =& \sum_j A_2^{ij}\Big\{-\mathrm{tr}\,[\boldsymbol{Q}_j\boldsymbol{\Sigma}] - \mathrm{tr}\,[\boldsymbol{Q}_j(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] + \sum_{a,b} B_{jab}\hat{\theta}_a^*\hat{\theta}_b^*/2 \\ & + \sum_{a,b}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]A_2^{ab}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\Big\}, \end{aligned} \tag{4.21}$$

so that from (4.20),

$$E[\hat{\theta}_i^{**}] = \sum_j A_2^{ij}\big\{-\mathrm{tr}\,[\boldsymbol{Q}_j\boldsymbol{\Sigma}] + \sum_{a,b} A_2^{ab}B_{jab} + 2\sum_{a,b} A_2^{ab}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}]\big\}.$$

Since $\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}] = -\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}\boldsymbol{\Sigma}_{(b)}]$, it is observed that

$$\sum_{a,b} A_2^{ab}B_{jab} + 2\sum_{a,b} A_2^{ab}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}] = \sum_{a,b} A_2^{ab}\mathrm{tr}\,[\boldsymbol{\Sigma}_{(ab)}(\boldsymbol{\Sigma}^{-1})_{(j)}],$$

so that

$$E[\hat{\theta}_i^{**}] = \sum_j A_2^{ij}\big\{-\mathrm{tr}\,[\boldsymbol{Q}_j\boldsymbol{\Sigma}] + \sum_{a,b} A_2^{ab}\mathrm{tr}\,[\boldsymbol{\Sigma}_{(ab)}(\boldsymbol{\Sigma}^{-1})_{(j)}]\big\},$$

which can be also expressed as (3.22), since $\mathrm{tr}\,[\boldsymbol{Q}_i\boldsymbol{\Sigma}] = -\mathrm{tr}\,[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1})_{(i)}\boldsymbol{X}]$.

For $E[\hat{\theta}_i^{**}\hat{\theta}_k^*]$, from (4.21), $\hat{\theta}_i^{**}\hat{\theta}_k^*$ is written as

$$\begin{aligned} \hat{\theta}_i^{**}\hat{\theta}_k^* =& \sum_j A_2^{ij}\big\{-\mathrm{tr}\,[\boldsymbol{Q}_j\boldsymbol{\Sigma}]\hat{\theta}_k^* - \mathrm{tr}\,[\boldsymbol{Q}_j(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\hat{\theta}_k^* + \sum_{a,b} B_{jab}\hat{\theta}_a^*\hat{\theta}_b^*\hat{\theta}_k^*/2 \\ & - \sum_{a,b,c} A_2^{ab}A_2^{kc}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(c)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\big\}. \end{aligned}$$

It is noted that $E[\mathrm{tr}\,[\boldsymbol{Q}_j(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\hat{\theta}_k^*] = -2\sum_a A_2^{ka}\mathrm{tr}\,[\boldsymbol{Q}_j\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{\Sigma}] = O(N^{-1})$, $B_{jab} = O(N)$, $E[\hat{\theta}_a^*\hat{\theta}_b^*\hat{\theta}_k^*] = O(N^{-2})$ and

$$\begin{aligned} & E[A_2^{ab}A_2^{kc}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]tr[(\boldsymbol{\Sigma}^{-1})_{(c)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]] \\ & \quad = 8A_2^{ab}A_2^{kc}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ja)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(c)}\boldsymbol{\Sigma}] = O(N^{-1}), \end{aligned}$$

so that

$$E[\hat{\theta}_i^{**}\hat{\theta}_k^*] = O(N^{-2}).$$

We shall show that $E[\hat{\theta}_i^{***}] = O(N^{-2})$ for $\hat{\theta}_i^{***}$ given in (3.19). Note that $B_{iab}E[\hat{\theta}_a^*\hat{\theta}_b^{**}] = O(N^{-1})$, $E[\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(iab)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\hat{\theta}_c^*\hat{\theta}_d^*] = O(N^{-1})$ and $C_{iabc}E[\hat{\theta}_a^*\hat{\theta}_b^*\hat{\theta}_c^*] = O(N^{-1})$. The i-th element of $E[\boldsymbol{A}_1\boldsymbol{A}_2^{-1}(\boldsymbol{a}_0 - \boldsymbol{b}_0/2)]$ is

$$\sum_{a,b} E[\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]A_2^{ab}\Big\{\mathrm{tr}\,[\boldsymbol{Q}_b\boldsymbol{\Sigma}] + \mathrm{tr}\,[\boldsymbol{Q}_b(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})] - \sum_{c,d} B_{bcd}\hat{\theta}_c^*\hat{\theta}_d^*\Big\}],$$

which can be shown to be of order $O(N^{-1})$. Similarly, the i-th element of $E[(\boldsymbol{A}_1\boldsymbol{A}_2^{-1}\boldsymbol{A}_1 - \boldsymbol{A}_0)\widehat{\boldsymbol{\theta}}^*]$ is

$$\begin{aligned}&\sum_{c,d}E[\Big\{\sum_{a,b}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]A_2^{ab}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(bc)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\\ &\qquad - \mathrm{tr}\,[\boldsymbol{Q}_{i(c)}\boldsymbol{\Sigma}] - \mathrm{tr}\,[\boldsymbol{Q}_{i(c)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\Big\}A_2^{cd}\mathrm{tr}\,[(\boldsymbol{\Sigma}_{(d)}^{-1}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]],\end{aligned}$$

which is of order $O(N^{-1})$. Thus, it is concluded that

$$E[\hat{\theta}_i^{***}] = O(N^{-2}).$$

Finally, we check the conditions (A4) (ii) and (A6)(ii). Since $\boldsymbol{\nabla}_y\hat{\theta}_j^* = -2\sum_a A_2^{ja}(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{y}$, it follows that $\boldsymbol{s}_{(i)}'\boldsymbol{\Sigma}\boldsymbol{\nabla}_y\hat{\theta}_j^* = -2\sum_a A_2^{ja}\boldsymbol{s}_{(i)}'\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{y}$, which is of order $O_p(N^{-1})$ from the conditions (A2) and (M1). Since $\boldsymbol{\nabla}_y\boldsymbol{\nabla}_y'\hat{\theta}_j^* = -2\sum_a A_2^{ja}(\boldsymbol{\Sigma}^{-1})_{(a)}$, it is observed that

$$\begin{aligned}&E[\boldsymbol{s}_{(i)}'\boldsymbol{\Sigma}\{\boldsymbol{\nabla}_y\boldsymbol{\nabla}_y'\hat{\theta}_j^*\}\boldsymbol{\Sigma}\boldsymbol{s}_{(i)}\hat{\theta}_k^*]\\ &\quad = -2\sum_a A_2^{ja}\boldsymbol{s}_{(i)}'\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{-1})_{(a)}\boldsymbol{\Sigma}\boldsymbol{s}_{(i)}E[\hat{\theta}_k^*] = 0.\end{aligned}$$

From (4.21), it can be seen that

$$\begin{aligned}\boldsymbol{\nabla}_y\hat{\theta}_j^{**} =& -2\sum_i A_2^{ji}\boldsymbol{Q}_i\boldsymbol{y} - \sum_{i,a,b,c} A_2^{ji}B_{iab}\{A_2^{ac}(\boldsymbol{\Sigma}^{-1})_{(c)}\boldsymbol{y}\hat{\theta}_b^* + A_2^{bc}(\boldsymbol{\Sigma}^{-1})_{(c)}\boldsymbol{y}\hat{\theta}_a^*\}\\ &+ 2\sum_{i,a,b} A_2^{ji}A_2^{ab}\{(\boldsymbol{\Sigma}^{-1})_{(ia)}\boldsymbol{y}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(b)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\\ &\quad + (\boldsymbol{\Sigma}^{-1})_{(b)}\boldsymbol{y}\mathrm{tr}\,[(\boldsymbol{\Sigma}^{-1})_{(ia)}(\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{\Sigma})]\},\end{aligned}$$

which shows that $\boldsymbol{s}_{(i)}'\boldsymbol{\Sigma}\boldsymbol{\nabla}_y\hat{\theta}_j^{**} = O_p(N^{-3/2})$ from the conditions (A2) and (M1). Thus, the conditions (A4)(ii) and (A6)(ii) are satisfied. Therefore, the proposition is proved. ■

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] 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.

[2] 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.

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

[4] 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.

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

[6] Cressie, N. and Lahiri, S.N. (1993). The asymptotic distribution of REML estimators. *J. Multivariate Analysis*, **45**. 217-233.

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

[8] 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.

[9] 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.

[10] Datta, G., Lahiri, P. and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *J. Statist. Plan. Inf.*, **102**, 83-97.

[11] 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.

[12] 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.

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

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

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

[16] Harville, D.A. and Jeske, D.R. (1992). Mean squared error of estimation or prediction under a general linear model. *J. Amer. Statist. Assoc.*, **87**, 724-731.

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

[18] Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.*, **79**, 853-862.

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

[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] Mardia K.V. and Marshall, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, **71**, 135-146.

[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. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, **25**, 175-186.

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

[26] Rao, J.N.K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian J. Statist.*, **22**, 511-528.

[27] Slud, E.V. and Maiti, T. (2006). Mean-squared error estimation in transformed Fay-Herriot models. *J.*, **68**, 239-257.

[28] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.

[29] Sweeting, T.J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.*, **8**, 1375-1381.