博士論文

Multivariate Linear Mixed Models with Application to Small Area Estimation

（多変量線形混合モデルと小地域推定への応用）

Tsubasa Ito
伊藤　翼

Abstract

This thesis studies multivariate linear mixed model with application to small area estimation. In the small area estimation, the empirical best linear unbiased predictor (EBLUP) in the linear mixed model is useful because it gives a stable estimate for a mean of a small area. For measuring uncertainty of EBLUP, much of research is focused on second-order unbiased estimation of mean squared prediction errors in the univariate case.

In this thesis, the multivariate Fay-Herriot model and nested-error regression model where the covariance matrix of random effects is fully unknown are considered. When the EBLUP is measured in terms of a mean squared error matrix (MSEM), a second-order approximation of MSEM of the EBLUP and a second-order unbiased estimator of the MSEM are derived analytically in closed forms. Confidence region and confidence interval of the small area mean centered around EBLUP, which are second order correct are also constructed.

Contents

Chapter 1

Introduction

Linear mixed models and model-based predictors in small area estimation have been studied extensively and actively in recent years due to the growing demand for reliable small area estimates. In small area estimation, direct design-based estimates for small area means have large standard errors due to small sample sizes from small areas. In order to improve accuracy, the linear mixed models are considered which consist of fixed effects based on common parametes and random effects depending on areas, and the resulting empirical best linear unbiased predictors (EBLUP) provide more reliable estimates by 'borrowing strength' from neighboring areas. This is because EBLUP shrinks the sample mean of the small area towards a stable quantity costructed by pooling all the data through the fixed effects, and the shrinkage arises from the random eects. Then, EBLUP can be interpreted as the empirical Bayes estimator, which is discussed by Efron and Morris (1975).

The linear mixed models used in small area estimation are the Fay-Herriot model for analyzing area-level data by Fay and Herriot (1978) and the nested error regression (NER) model for analyzing unit-level data by Battese, Harter and Fuller (1988). Various extensions and generalizations of these models and many statistical methods for inference have been studied in the literature.

In the linear mixed models, best linear unbiased predictors (BLUP) are derivated at first, which cannot be used since they contain unknown variance components. Several methods for estimation of these parameters are known. Two major methods are maximum likelihood and restricted maximum likelihood methods based on the marginal density function of the observations. Other methods are moment estimators, which are proposed by Prasad and Rao (1990) and Fay and Herriot (1979). Their properties are studied in the literatures and the most important one is that they converge to the true values as the number of areas increases under the some regularized conditions. EBLUP is obtained by plugging the estimates of variance components in BLUP.

When EBLUP is used for the estimation of small area means, we need to evaluate the uncertainty of EBLUP. Mean squared error (MSE) of EBLUP is usually considered for this, and its estimator is needed in practice. MSE of EBULP is approximated at first with second-order accuracy when the number of areas increases by Taylar's expansion. Then, the second-order unbiased edtimator of MSE is obtained in the same way as the MSE approximation. Another method for assessing the reliability of EBLUP is the construction of confidence intervals of small area means. Confidence intervals constructed by using EBLUP and estimates of its MSE based on the standard statistical theory are known to have the coverage probabilities (CP) which have the second-order bias for the nominal confidence level. Several methods are proposed to correct this. One is to derive the correction term analytically by approximating the distribution of the satistic of interest via the Taylor's expansion. Another is to approximate computationally the distribution function of the statistic by the parametric bootstrap method.

For comprehensive reviews of small area estimation, see Ghosh and Rao (1994), Datta and Ghosh (2012), Pfeffermann (2013) and Rao and Molina (2015).

When multivariate data with correlations are observed from small areas for estimating multi-dimensional characteristics, like poverty and unemployment indicators, Fay (1987) suggested a multivariate extension of the univariate Fay-Herriot model, called a multivariate Fay-Herriot model, to produce reliable estimates of median incomes for four-, three- and five-person families. Fuller and Harter (1987) also considered a multivariate modeling for estimating a finite population mean vector. Datta, Day and Basawa (1999) provided unified theories in empirical linear unbiased prediction or empirical Bayes estimation in general multivariate mixed linear models. Datta, Day and Maiti (1998) suggested a hierarchical Bayesian approach to multivariate small area estimation. Datta, *et al.* (1999) showed the interesting result that the multivariate modeling produces more efficient predictors than the conventional univariate modeling. Porter, Wikle and Holan (2015) used the multivariate Fay-Herriot model for modeling spatial data. Ngaruye, von Rosen and Singull (2016) applied a multivariate mixed linear model to crop yield estimation in Rwanda.

Although Datta, *et al.* (1999) developed the general and unified theories concerning the empirical best linear unbiased predictors (EBLUP) and their uncertainty, it is definitely more helpful and useful to provide concrete forms with closed expressions for EBLUP, the second-order approximation of the mean squared error matrix (MSEM) and the second-order unbiased estimator of the mean squared error matrix. Recently, Benavent and Morales (2016) treated the multivariate Fay-Herriot model with the covariance matrix of random effects depending on unknown parameters. As a structure in the covariance matrix, they considered diagonal, AR(1) and the related structures and employed the residual maximum likelihood (REML) method for estimating the unknown parameters embedded in the covariance matrix. A second-order approximation and estimation of the MESM were also derived. For some examples, however, it is difficult or impossible to assume specific structures without prior knowledge or information on covariance matrices. Then, this thesis studies multivariate linear mixed model with application to small area estimation where the structure of the covariance matrix of random effects is fully unknown.

In this thesis, some problems are considered and new results are obtained. Firstly, the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown is considered. The empirical best linear unbiased predictors are provided, and second-order approximation of their mean squared error matrices and their second-order unbiased estimators of the MSEM are derived with closed expressions. The problem of costructing confidence regions for small area mean vectors is also cinsidered. These are achieved under tha normal assumption. This assumption, however, is often strict in practice, and then we need the robust estimator for the MSEM of EBLUP. This is considered in Chapter 3. In Chapter 4, the multivariate nested-error regression model where the covariance matrix of random effects is fully unknown is considered. The construction of the confidence interval for the liner combination of a small area mean and some vector is also considered.

Chapter 2

Multivariate Fay-Herriot Model

2.1 Motivation

In this chapter, the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown is considered. This situation has been studied by Fay (1987), Fuller and Harter (1987), Datta, *et al*. (1998), and useful in the case that statisticians have little knowledge on structures in correlation. As a specific estimator of the covariance matrix, Prasad-Rao type estimators with closed forms and use the modified versions which are restricted over the space of nonnegative definite matrices is employed. The empirical best linear unbiased predictors are provided based on the Prasad-Rao type estimators, and second-order approximation of their mean squared error matrices and their second-order unbiased estimators of the MSEM are derived with closed expressions. These are multivariate extensions of the results given by Prasad and Rao (1990) and Datta, *et al*. (2005) for the univariate case. It is noted that empirical best linear unbiased predictors for small area means are empirical Bayes estimators and related to the so-called James-Stein estimators. In this sense, the prediction in the multivariate Fay-Herriot model corresponds to the empirical Bayes estimation of a mean matrix of a multivariate normal distribution, which is related to the estimation of a precision matrix from a theoretical aspect as discussed in Efron and Morris (1976). In this framework, several types of estimators are suggested for estimation of the precision matrix, and it may be an interesting query whether those estimators provide improvements in the multivariate small area estimation.

Another topic in this capter is the construction of confidence regions. Confidence regions are more useful for measuring uncertainty of EBLUP, but there is no literature about confidence regions for multivariate small area estimation problems to the best of our knowledge. Naive confidence regions can be constructed easily by using the Bayes estimators of small area means and their MSEM. As is the case in the univariate small area estimation problem, the coverage probability of the naive methods cannot be guaranteed to be greater than or equal to the nominal confidence coefficient $1 - \alpha$. Recently, in the univariate Fay-Herriot model, Diao, Smith, Datta, Maiti and Opsomer (2014) constructed closed-form confidence intervals whose coverage probability is identical to the nominal confidence coefficient up to the second-order for small area means under the normality assumption. Although the approach discussed in this chapter is a multivaliate extension of Diao et al. (2014), there are two difficulties: One is how to construct a confidence region on the multi-dimensional space, and the other is how to construct a positive-definite and consistent estimator of the covariance matrix of random effects. We here consider a confidence region based on the Mahalanobis distance centerd around EBLUP, and use the asymptotic expansion of the characteristic function of this distance to approximate the coverage probability based on the chi-square distributions. We obtain the correction term in a closed form, and provide the confidence region that is second order correct. Concerning the estimation of the covariance matrix, the Prasad-Rao type estimator with non-negative definite modification can be given in a closed form by the moment method. When the covariance matrix is estimated with the zero matrix or a singular matrix close to the zero matrix, however, the correction term becomes instable in the confidence region. This fact is well known in the univariate confidence interval. Thus, a new method for obtaining a positive-definite and sencon-order unbiased estimator of the covariance matrix is suggested. Moreover, the extension of our results to construction of corrected confidence regions for the difference of two small area mean vectors is considered. Another approach to construction of corrected confidence regions is the bootstrap method which needs heavy burden in computation. Because the corrected confidence region suggested here is provided in closed forms, it is easy to implement, which is a merit of our method.

This chapter is organized as follows: Section 2.2 introduces the multivariate Fay-Herriot model and gives the EBLUP and its prediction risk approximation. In section 2.3, the second-order approximation of MSEM of EBLUP and the second-order unbiased estimator of the MSEM are derived. In section 2.4, the proposed confidence region is derived. Section 2.5 gives the Prasad-Rao type estimator of the covariance matrix of the random effects and its positive-definite modification with second-order unbiasedness and consistency. In section 2.6, the extension to the confidence regions for the difference of two small area means is described. The performances of our proposed methods are investigated in Section 2.7. This numerical study illustrates that the proposals have good performances for the low-dimensional case. However, a $k \times k$ covariance matrix has $k(k+1)/2$ parameters, and we need more data so as to maintain the performances of the proposals for higher-dimensional cases.

2.2 Multivariate Fay-Herriot Model and Empirical Best Linear Unbiased Predictor

Suppose that area-level data $(\boldsymbol{y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{y}_m, \boldsymbol{X}_m)$ are observed, where m is the number of small areas, $\boldsymbol{y}_i$ is a k-variate vector of direct survey estimates and $\boldsymbol{X}_i$ is a $k \times s$ matrix of covariates associated with $\boldsymbol{y}_i$ for the i-th area. The multivariate Fay-Herriot model suggested by Fay (1987) is described as

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, m, \tag{2.2.1}$$

where $\boldsymbol{\beta}$ is an s-variate vector of unknown regression coefficients, $\boldsymbol{v}_i$ is a k-variate vector of random effects depending on the i-th area and $\boldsymbol{\varepsilon}_i$ is a k-variate vector of sampling errors. It is assumed that $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$ are mutually independently distributed as

$$\boldsymbol{v}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Psi}) \quad \text{and} \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{D}_i), \tag{2.2.2}$$

where $\boldsymbol{\Psi}$ is a $k \times k$ unknown and nonsingular covariance matrix and $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m$ are $k \times k$ known covariance matrices. This is a multivariate extension of the so-called Fay-Herriot model suggested by Fay and Herriot (1979). Letting $\boldsymbol{\theta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{v}_i$ for $i = 1, \ldots, m$, we can rewrite the model given in (2.2.1) and (2.2.2) as

$$\begin{aligned} \boldsymbol{y}_i \mid \boldsymbol{\theta}_i \sim& \mathcal{N}_k(\boldsymbol{\theta}_i, \boldsymbol{D}_i), \\ \boldsymbol{\theta}_i \sim& \mathcal{N}_k(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi}), \end{aligned} \tag{2.2.3}$$

for $i = 1, \ldots, m$. Thus, the multivariate Fay-Herriot model is interpreted as the Bayes model with the prior distribution of $\boldsymbol{\theta}_i$. It may be convenient to express model (2.2.1) in a matrix form. Let $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_m^\top)^\top$, $\boldsymbol{v} = (\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_m^\top)^\top$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_m^\top)^\top$. Then, model (2.2.1) is expressed as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\varepsilon}, \tag{2.2.4}$$

where $\boldsymbol{v} \sim \mathcal{N}_{km}(\boldsymbol{0}, \boldsymbol{I}_m \otimes \boldsymbol{\Psi})$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_{km}(\boldsymbol{0}, \boldsymbol{D})$ for $\boldsymbol{D} = \text{block diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m)$. Throughout the paper, it is assumed that $\boldsymbol{X}$ is of full rank.

For example, we consider the crop data of Battese, Harter and Fuller (1988), who analyze the data in the nested error regression model. For the i-th county, let y_{i1} and y_{i2} be survey data of average areas of corn and soybean, respectively. Also let x_{i1} and x_{i2} be satellite data of average areas of corn and soybean, respectively. In this case, $\boldsymbol{y}_i$, $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ correspond to

$$\boldsymbol{y}_i = (y_{i1}, y_{i2})^\top, \quad \boldsymbol{X}_i = \begin{pmatrix} 1 & x_{i1} & x_{i2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_{i1} & x_{i2} \end{pmatrix}, \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)^\top$$

for $k = 2$ and $s = 6$. Battese, *et al.* (1988) applied a univariate nested error regression model for each of y_{i1} and y_{i2}, while we can use the multivariate model (2.2.1) for analyzing both data simultaneously.

We want to predict and construct a confidence region of $\boldsymbol{\theta}_a$ for the a-th area. To this end, we begin by deriving the Bayes estimator of $\boldsymbol{\theta}_a$. The posterior distribution of $\boldsymbol{\theta}_i$ given $\boldsymbol{y}_i$ and the marginal distribution of $\boldsymbol{y}_i$ are

$$\begin{aligned} \boldsymbol{\theta}_i \mid \boldsymbol{y}_i \sim& \mathcal{N}_k(\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}), (\boldsymbol{\Psi}^{-1} + \boldsymbol{D}_i^{-1})^{-1}), \\ \boldsymbol{y}_i \sim& \mathcal{N}_k(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Psi} + \boldsymbol{D}_i), \end{aligned} \quad i = 1, \ldots, m, \tag{2.2.5}$$

where

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}) =& \boldsymbol{X}_i^\top\boldsymbol{\beta} + \boldsymbol{\Psi}(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}) \\ =& \boldsymbol{y}_i - \boldsymbol{D}_i(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}), \end{aligned}$$

which is the Bayes estimator of $\boldsymbol{\theta}_i$.

When $\boldsymbol{\Psi}$ is known, the maximum likelihood estimator or generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) =& \{\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{y} \\ =& \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{y}_i. \end{aligned} \tag{2.2.6}$$

Substituting $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ into $\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi})$ yields the empirical Bayes estimator

$$\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) = \boldsymbol{y}_a - \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a - \boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}. \tag{2.2.7}$$

Datta, *et al.* (1999) showed that $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta}_a$. It can be also demonstrated that $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is the Bayes estimator against the uniform prior distribution of $\boldsymbol{\beta}$ as well as the empirical Bayes estimator as shown above, which is called the Bayes empirical Bayes estimator.

Because $\boldsymbol{\Psi}$ is unknown, we need to estimate the covariance matrix $\boldsymbol{\Psi}$. Estimators used in the univariate case are the ANOVA type estimator given by Prasad and Rao (1990), the Fay-Herriot estimator suggested by Fay and Herriot (1979), and the ML and REML methods used in Datta and Lahiri (2000). Corresponding to the univariate case, we consider the general class of estimators $\widehat{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$ which satisfy the following conditions:

(H1) $\widehat{\boldsymbol{\Psi}}$ is an even function of $\boldsymbol{y}$; $\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}) = \widehat{\boldsymbol{\Psi}}(-\boldsymbol{y})$

(H2) $\widehat{\boldsymbol{\Psi}}$ is a translation invariant function ; $\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T}) = \widehat{\boldsymbol{\Psi}}(\boldsymbol{y})$ for any $\boldsymbol{T} \in \mathbb{R}^s$ and all $\boldsymbol{y}$.

The modified Prasad-Rao estimator suggested later in this paper and the ML method satisfy these conditions. We replace $\boldsymbol{\Psi}$ in $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ with the estimator $\widehat{\boldsymbol{\Psi}}$, and the resulting empirical Bayes (EB) estimator is

$$\widehat{\boldsymbol{\theta}}_a^{EB} = \widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}) = \boldsymbol{y}_a - \boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a - \boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})\}. \tag{2.2.8}$$

This is also interpreted as the empirical best linear unbiased predictor (EBLUP) in the context of the linear mixed models.

2.3 Second-order Approximation of Mean Squared Error Matrix

For evaluating the uncertainty of $\widehat{\boldsymbol{\theta}}_a^{EB}$, we prepare three lemmas.

Lemma 2.3.1 *$\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ is independent of $\boldsymbol{P}\boldsymbol{y}$ for $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}$. Also, $\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is a function of $\boldsymbol{P}\boldsymbol{y}$, and independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$.*

The proof of Lemma 2.3.1 is given in the section 2.8. It is noted that $\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a = (\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a) + (\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \widehat{\boldsymbol{\theta}}_a^{EB})$. From Lemma 2.3.1, $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \widehat{\boldsymbol{\theta}}_a^{EB}$ is a function of $\boldsymbol{P}\boldsymbol{y}$ and is independent of $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a$. It is noted that

$$\begin{aligned}
&E[(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a)(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a)^\top] \\
&=E[(\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi}) - \boldsymbol{\theta}_a)(\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi}) - \boldsymbol{\theta}_a)^\top + (\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi}))(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi}))^\top] \\
&=\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) =&(\boldsymbol{\Psi}^{-1} + \boldsymbol{D}_a^{-1})^{-1} = \boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a, \\
\boldsymbol{G}_{2a}(\boldsymbol{\Psi}) =&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\boldsymbol{X}^\top(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a.
\end{aligned} \tag{2.3.1}$$

Because $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a$ is independent of $\boldsymbol{P}\boldsymbol{y}$, it is observed that given $\boldsymbol{P}\boldsymbol{y}$, the conditional distribution of $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a$ is $\mathcal{N}_k(\boldsymbol{0}, \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}))$. This implies the following lemma which will be used for constructing a confidence region in the next section.

Lemma 2.3.2 *Under the conditions* (H1) *and* (H2), *the conditional distribution of* $\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a$ *given* $\boldsymbol{Py}$ *is*

$$\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a | \boldsymbol{Py} \sim \mathcal{N}_k(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}), \boldsymbol{H}_a(\boldsymbol{\Psi})). \tag{2.3.2}$$

for $\boldsymbol{H}_a(\boldsymbol{\Psi}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi})$.

For evaluating uncertainty of the EBLUP asymptotically, we assume the conditions given below for $m \to \infty$:.

(H3) $\widehat{\boldsymbol{\Psi}}$ is $\sqrt{m}$-consistent and second-order unbiased, namely $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = O(m^{-1/2})$ and $E[\widehat{\boldsymbol{\Psi}}] = \boldsymbol{\Psi} + o(m^{-1})$.

(H4) $0 < k < \infty$, $0 < s < \infty$.

(H5) There exist positive constants $\underline{d}$ and $\overline{d}$ such that $\underline{d}$ and $\overline{d}$ do not depend on m and satify $\underline{d}\boldsymbol{I}_k \leq \boldsymbol{D}_i \leq \overline{d}\boldsymbol{I}_k$ for $i = 1, \dots, m$.

(H6) $\boldsymbol{X}^\top \boldsymbol{X}$ is nonsingular and $\boldsymbol{X}^\top \boldsymbol{X}/m$ converges to a positive definite matrix.

Under these conditions, we can obtain the important approximations which will be useful for evaluating the mean squared error (MSE) matrix of $\widehat{\boldsymbol{\theta}}_a^{EB}$ and for constructing corrected confidence region based on $\widehat{\boldsymbol{\theta}}_a^{EB}$.

Lemma 2.3.3 *Under conditions* (H1)-(H6), *the following approximations hold:*

(1) $E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top] = \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O(m^{-3/2})$, *where*

$$\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) = \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} E\Big[(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\Big](\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a, \tag{2.3.3}$$

(2) $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) - \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O(m^{-3/2})$.

The proof of Lemma 2.3.3 is given in the section 2.8. Using Lemma 2.3.2 and Lemma 2.3.3 (1), we can approximate the MSE matrix of $\widehat{\boldsymbol{\theta}}_a^{EB}$ as

$$\begin{aligned}\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) =& \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top] \\ =& \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O(m^{-3/2}).\end{aligned} \tag{2.3.4}$$

Using Lemma 2.3.3 (2), we can obtain the second-order unbiased estimator of $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$, which is given by

$$msem(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) + \boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}}) + 2\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}), \tag{2.3.5}$$

namely, $E[msem(\widehat{\boldsymbol{\theta}}_a^{EB})] = \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) + O(m^{-3/2})$. Lemma 2.3.3 will be also used for deriving corrected confidence region in the next section.

2.4 Confidence Region with Corrected Coverage Probability

We now construct a confidence region of $\boldsymbol{\theta}_a$ based on $\widehat{\boldsymbol{\theta}}_a^{EB}$ with second-order accuracy. When $\boldsymbol{\Psi}$ is known, it follows from Lemma 2.3.2 that the confidence region based on the Mahalanobis distance with $100(1-\alpha)\%$ confidence coefficient is $\{\boldsymbol{\theta}_a \mid (\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})) \leq \chi^2_{k,1-\alpha}\}$

for $\boldsymbol{H}_a(\boldsymbol{\Psi}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi})$, where $\chi^2_{k,1-\alpha}$ is the $100\alpha\%$ upper quantile of the chi-squared distribution with degrees of freedom k. For a matrix $\boldsymbol{A}(\boldsymbol{\Psi})$, $\boldsymbol{A}^{-1}(\boldsymbol{\Psi})$ denotes the inverse matrix of $\boldsymbol{A}(\boldsymbol{\Psi})$. Since $\boldsymbol{\Psi}$ is unknown, we replace $\boldsymbol{\Psi}$ with estimator $\widehat{\boldsymbol{\Psi}}$ to get the naive confidence region

$$CR_0 = \{\boldsymbol{\theta}_a \mid (\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}})) \le \chi^2_{k,1-\alpha}\}. \tag{2.4.1}$$

Under appropriate conditions, it can be shown that the coverage probability tends to the nominal confidence coefficient $1-\alpha$, namely $\lim_{m\to\infty} P(\boldsymbol{\theta}_a \in CR_0) = 1-\alpha$. However, this confidence region has the second-order bias, because $P(\boldsymbol{\theta}_a \in CR_0) = 1-\alpha+O(m^{-1})$. Thus, we want to derive a corrected confidence region CR such that $P(\boldsymbol{\theta}_a \in CR) = 1-\alpha+O(m^{-3/2})$.

Define B_1, B_2 and B_3 by

$$\begin{aligned}
B_1 =& B_1(\boldsymbol{\Psi}) = -\frac{1}{2}\mathrm{tr}\left(E[\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))]\right),\\
B_2 =& B_2(\boldsymbol{\Psi}) = -\frac{1}{8}\left\{E[\mathrm{tr}^2(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))] + 2\mathrm{tr}\left(E[(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))^2]\right)\right\},\\
B_3 =& \mathrm{tr}\,(\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi})),
\end{aligned} \tag{2.4.2}$$

where $\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}) = \boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})$ and $\mathrm{tr}^2(\boldsymbol{A}) = (\mathrm{tr}\,\boldsymbol{A})^2$ for matrix $\boldsymbol{A}$. It can be seen that $B_1 = O(m^{-1})$, $B_2 = O(m^{-1})$ and $B_3 = O(m^{-1})$. Then, we provide the main theorem which will be proved in the section 2.8.

Theorem 2.4.1 *Under the conditions* (H1)-(H6)*, it holds that*

$$\begin{aligned}
&P\{(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a) \le x\}\\
&\qquad = F_k(x) + 2(B_1 - B_2 - B_3)f_{k+2}(x) + 2B_2 f_{k+4}(x) + o(m^{-1}),
\end{aligned} \tag{2.4.3}$$

where $F_k(x)$ and $f_k(x)$ are the cumulative distribution and probability density functions of the chi-squared distribution with the degree of freedom k, respectively.

We can consider the Bartlett-type correction using the asymptotic expansion (2.4.3). For $h = O(m^{-1})$, it is observed that

$$\begin{aligned}
&P\{(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a) \le x(1+h)\}\\
&\qquad = F_k(x) + hxf_k(x) + 2(B_1 - B_2 - B_3)f_{k+2}(x) + 2B_2 f_{k+4}(x) + o(m^{-1}).
\end{aligned}$$

Note that $hxf_k(x) + 2(B_1 - B_2 - B_3)f_{k+2}(x) + 2B_2 f_{k+4}(x)$ is of order $O(m^{-1})$. Thus, the second-order term vanishes if

$$hxf_k(x) = -2(B_1 - B_2 - B_3)f_{k+2}(x) - 2B_2 f_{k+4}(x) = 0. \tag{2.4.4}$$

Since $\Gamma(x+1) = x\Gamma(x)$ for the gamma function $\Gamma(x)$, the solution of the equation (2.4.4) on h is

$$h^*(\boldsymbol{\Psi}) = -2\{(B_1 - B_3 - B_2)/k + B_2 x/k(k+2)\}. \tag{2.4.5}$$

For $h^*(\boldsymbol{\Psi})$ given in (2.4.5), it holds that for any $x > 0$,

$$P\{(1 + h^*(\widehat{\boldsymbol{\Psi}}))^{-1}(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a) \le x\} = F_k(x) + o(m^{-1}).$$

Hence, the corrected confidence region is given by

$$CR = \{\boldsymbol{\theta}_a \mid (\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a^{EB})^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\boldsymbol{\theta}_a - \widehat{\boldsymbol{\theta}}_a^{EB}) \le \{1 + h^*(\widehat{\boldsymbol{\Psi}})\}\chi^2_{k,1-\alpha}\}. \tag{2.4.6}$$

Corollary 2.4.1 *Under conditions* (H1)-(H6), *it holds that*

$$P(\boldsymbol{\theta}_a \in CR) = 1 - \alpha + o(m^{-1}).$$

2.5 Derivation of a Second-order Unbiased and Positive-definite Estimator of the Covariance Matrix of Random Effects

We here provide a new method for deriving a second-order unbiased and positive-definite estimator of $\boldsymbol{\Psi}$. As well known in the univariate case, the Prasad-Rao estimator of the 'between' component of variance takes a negative value with a positive probability, and the nonnegative estimator which truncates it at zero is used. The maximum likelihood (ML) and restricted maximum likelihood (REML) estimators take values of zero with positive probabilities. To fix this drawback, Li and Lahiri (2010) suggested the adusted maximum likelihood method for giving a positive and consistent estimator. As pointed out by Yoshimori and Lahiri (2014), this problem causes instability of the corrected confidence interval. In the multivariate case, since $\boldsymbol{G}_{2a}(\boldsymbol{\Psi}) = O(m^{-1})$, it is seen that $\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi}) = \boldsymbol{G}_{1a}^{-1}(\boldsymbol{\Psi}) + O(m^{-1}) = \boldsymbol{\Psi}^{-1} + \boldsymbol{D}_a^{-1} + O(m^{-1})$. This means that the correction function $h^*(\boldsymbol{\Psi})$ takes a large value when some eigenvalues of estimator $\widehat{\boldsymbol{\Psi}}$ are zero.

To derive a positive-definite and consistet estimator of $\boldsymbol{\Psi}$, let $\boldsymbol{U}$ be a $k \times k$ orthogonal matrix $\boldsymbol{U}$ such that $\widehat{\boldsymbol{\Psi}} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}^\top$ for a diagonal matrix $\boldsymbol{L} = \mathrm{diag}\,(\ell_1, \ldots, \ell_k)$. Then, we consider adjusted estimators of the form

$$\widehat{\boldsymbol{\Psi}}_{(A)} = \frac{1}{2}(\widehat{\boldsymbol{\Psi}} - a\boldsymbol{I}_k + \boldsymbol{U}\boldsymbol{L}_{(A)}\boldsymbol{U}^\top), \tag{2.5.1}$$

where

$$\boldsymbol{L}_{(A)} = \mathrm{diag}\,(\sqrt{(\ell_1 - \hat{a})^2 + \hat{b}_1}, \ldots, \sqrt{(\ell_k - \hat{a})^2 + \hat{b}_k}),$$

for some statistics $\hat{a}$ and $\hat{b}_1, \ldots, \hat{b}_k$.

Proposition 2.5.1 *Assume that* $\hat{a} = O_p(m^{-1})$, $E[\hat{a}^2] = o(m^{-1})$ *and that* $\hat{b}_i$*'s are positive almost surely and* $\hat{b}_i = O_p(m^{-1})$ *for* $i = 1, \ldots, k$. *Let* $\widehat{\boldsymbol{\Psi}}$ *be a consistent estimator of* $\boldsymbol{\Psi}$ *as* $m \to \infty$.

(1) $\widehat{\boldsymbol{\Psi}}_{(A)}$ *given in* (2.5.1) *is positive-definite almost surely, and* $\widehat{\boldsymbol{\Psi}}_{(A)} = \widehat{\boldsymbol{\Psi}} + O_p(m^{-1})$.

(2) *If* $\widehat{\boldsymbol{\Psi}}$ *is second-order unbiased, namely,* $E[\widehat{\boldsymbol{\Psi}}] = \boldsymbol{\Psi} + o(m^{-1})$, *and if* $\hat{b}_i = 4\hat{a}(\ell_i - \hat{a})0$ *is almost surely positive, then* $\widehat{\boldsymbol{\Psi}}_{(A)}$ *is positive definite almost surely and second-order unbiased.*

Proof. It is clear that $\widehat{\boldsymbol{\Psi}}_{(A)}$ is positive definite almost surely. Note that there exists positive λ_i such that ℓ_i converges to λ_i, because $\widehat{\boldsymbol{\Psi}}$ is consistent. Since $\hat{a} = O_p(m^{-1})$ and $E[\hat{a}^2] = o(m^{-1})$, it is seen that $P(\lambda_i - \hat{a} < 0) = o(m^{-1})$. Then, the eigenvalues of $\widehat{\boldsymbol{\Psi}}_{(A)}$ are approximated as

$$\begin{aligned}
\ell_i - \hat{a} + \sqrt{(\ell_i - \hat{a})^2 + \hat{b}_i} =& \ell_i - \hat{a} + \sqrt{(\ell_i - \lambda_i + \lambda_i - \hat{a})^2 + \hat{b}_i} \\
=& \ell_i - \hat{a} + |\lambda_i - \hat{a}|\sqrt{1 + \frac{2(\lambda_i - \hat{a})(\ell_i - \lambda_i) + (\ell_i - \lambda_i)^2 + \hat{b}_i}{(\lambda_i - \hat{a})^2}} \\
=& 2\ell_i - 2\hat{a} + \frac{\hat{b}_i}{2(\lambda_i - \hat{a})} + o_p(m^{-1}).
\end{aligned} \tag{2.5.2}$$

This implies that $\widehat{\boldsymbol{\Psi}}_{(A)} = \widehat{\boldsymbol{\Psi}} + O_p(m^{-1})$, which shows part (1). For part (2), let $\hat{b}_i = 4\hat{a}(\lambda_i - \hat{a})$. Then we can see that the second term is equal to the third term in RHS of (2.5.2), and the second-order bias vanishes. Thus, the part (2) is shown by replacing λ_i with ℓ_i. □

Before constructing the estimator $\widehat{\boldsymbol{\Psi}}_{(A)}$ with specific $\hat{a}$ and $\hat{b}_i$'s, we obtain estimator $\widehat{\boldsymbol{\Psi}}$ which satisfies conditions (H1), (H2) and (H3). When $\boldsymbol{\Psi}$ is a fully unknown covariance matrix, it is hard to derive the ML and REML estimates numerically. Instead, we begin by deriving a Prasad-Rao type estimator based on the moment method. Because $E[(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top] = \boldsymbol{\Psi} + \boldsymbol{D}_i$ for $i = 1, \ldots, m$, we have $\sum_{i=1}^m E[(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top] = m\boldsymbol{\Psi} + \sum_{i=1}^m \boldsymbol{D}_i$. Substituting the ordinary least squares estimator $\widehat{\boldsymbol{\beta}}^{OLS} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ into $\boldsymbol{\beta}$, we get the Prasd-Rao type consistent estimator

$$\widehat{\boldsymbol{\Psi}}_0^{PR} = \frac{1}{m}\sum_{i=1}^m \big\{(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{OLS})(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}^{OLS})^\top - \boldsymbol{D}_i\big\}. \tag{2.5.3}$$

It is noted that this estimator has a second-order bias. In fact, the bias, given by $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0^{PR}}(\boldsymbol{\Psi}) = E[\widehat{\boldsymbol{\Psi}}_0^{PR}] - \boldsymbol{\Psi}$, is

$$\begin{aligned}\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0^{PR}}(\boldsymbol{\Psi}) =& \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi} + \boldsymbol{D}_j)\boldsymbol{X}_j\Big\}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top \\ & - \frac{1}{m}\sum_{i=1}^m(\boldsymbol{\Psi} + \boldsymbol{D}_i)\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top - \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i).\end{aligned} \tag{2.5.4}$$

Substituting $\widehat{\boldsymbol{\Psi}}_0^{PR}$ into $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0^{PR}}(\boldsymbol{\Psi})$ provides the bias-corrected estimator

$$\widehat{\boldsymbol{\Psi}}^{PR} = \widehat{\boldsymbol{\Psi}}_0^{PR} - \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0^{PR}}(\widehat{\boldsymbol{\Psi}}_0^{PR}). \tag{2.5.5}$$

The estimator $\widehat{\boldsymbol{\Psi}}^{PR}$ satisfies conditions (H1), (H2) and (H3). However, it still has a drawback of taking a negative value with a positive probability. For applying the method suggested in Proposition 2.5.1, let

$$\hat{a} = \mathrm{tr}\,(\widehat{\boldsymbol{\Psi}}^{PR})/mk \quad \text{and} \quad \hat{b}_i = \max\{4\hat{a}(\ell_i^{PR} - \hat{a}), 1/m\}, \quad \text{for} \quad i = 1, \ldots, k,$$

where ℓ_i^{PR}'s are eigenvalues of $\widehat{\boldsymbol{\Psi}}^{PR}$. Note that $P\{\hat{a}(\ell_i^{PR} - \hat{a}) < 1/(4m)\} = o(m^{-1})$. Then, we suggest the adjusted estimator

$$\widehat{\boldsymbol{\Psi}}_{(A)}^{PR} = \frac{1}{2}(\widehat{\boldsymbol{\Psi}}^{PR} - \hat{a}\boldsymbol{I}_k + \boldsymbol{U}^{PR}\boldsymbol{L}_{(A)}^{PR}(\boldsymbol{U}^{PR})^\top), \tag{2.5.6}$$

where column vectors of $\boldsymbol{U}^{PR}$ are the eigenvectors of $\widehat{\boldsymbol{\Psi}}^{PR}$ and

$$\boldsymbol{L}_{(A)}^{PR} = \mathrm{diag}\,(\sqrt{(\ell_1^{PR} - \hat{a})^2 + \hat{b}_1}, \ldots, \sqrt{(\ell_k^{PR} - \hat{a})^2 + \hat{b}_k}).$$

It follows from Proposition 2.5.1 that $\widehat{\boldsymbol{\Psi}}_{(A)}^{PR}$ is positive-definite and second-order unbiased.

Before calculating some moments given in B_1, B_2 and B_3, we need a closed-form expression of $\boldsymbol{G}_{3a}(\boldsymbol{\Psi})$ given in (2.3.3), which is stated in the following lemma.

Lemma 2.5.1 *By using the Prasad-Rao type estimetor given in* (2.5.5) *or* (2.5.6)*, we can write* $\boldsymbol{G}_{3a}(\boldsymbol{\Psi})$ *in* (2.3.3)*, as*

$$\begin{aligned}\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) =& \frac{1}{m^2}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big[\sum_{i=1}^m(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\ &+\sum_{i=1}^m\{\mathrm{tr}\,[(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}]\}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a.\end{aligned} \tag{2.5.7}$$

Finally, we calculate some moments given in B_1, B_2 and B_3 for the estimator $\widehat{\boldsymbol{\Psi}}_{(A)}^{PR}$. This calculation is used for providing the correction function $h^*(\boldsymbol{\Psi})$.

Lemma 2.5.2 *Assume conditions* (H4)-(H6)*. For* $\widehat{\boldsymbol{\Psi}}_{(A)}^{PR}$ *as in* (2.5.6)*, the values of* B_1 *and* B_2 *in* (2.4.2) *are given by*

$$\begin{aligned}B_1 =& -\frac{1}{2m^2}\sum_{i=1}^m\Big\{\mathrm{tr}\,\Big((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\ &\qquad\times(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big)\\ &\quad+\mathrm{tr}\,\Big((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big)\\ &\qquad\times\mathrm{tr}\,\Big((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big)\Big\}+o(m^{-1}),\\ B_2 =& -\frac{1}{4m^2}\sum_{i=1}^m\Big\{\mathrm{tr}\,\Big(((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))^2\Big)\\ &\quad+\mathrm{tr}\,\Big(((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))^2\Big)\\ &\quad+\mathrm{tr}^2\Big(((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))\Big)\Big\}+o(m^{-1}),\end{aligned}$$

and the value of B_3 *in* (2.4.2) *is* $B_3 = \mathrm{tr}\,(\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi}))$ *for* $\boldsymbol{G}_{3a}(\boldsymbol{\Psi})$ *given in* (2.5.7)*.*

By substituting these values into (2.4.5), we can construct the confidence region in the closed-form. Moreover, by substituting (2.5.7) into (2.3.5), we can obtain an estimator of closed-form approximation of the MSE matrix of $\widehat{\boldsymbol{\theta}}_a^{EB}$ as a by-product.

2.6 Confidence Region for the Difference of Two Small Area Means

In this section, we extend the results in Section 2.4 to the construction of a confidence region for $\boldsymbol{\theta}_a - \boldsymbol{\theta}_b$ for $a \neq b$. This enables us to conduct a statistical test under the null hypothsis $H_0 : \boldsymbol{\theta}_a = \boldsymbol{\theta}_b$. Since the corrected confidence region of $\boldsymbol{\theta}_a - \boldsymbol{\theta}_b$ can be constructed by the same arguments as in Section 2.4, we here provide the sketch of the result.

Let $\boldsymbol{G}_{ab}(\boldsymbol{\Psi}) = \boldsymbol{H}_a(\boldsymbol{\Psi}) + \boldsymbol{H}_b(\boldsymbol{\Psi}) - \boldsymbol{G}_{2ab}(\boldsymbol{\Psi})$, where $\boldsymbol{G}_{2ab}(\boldsymbol{\Psi}) = E[(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)(\widehat{\boldsymbol{\theta}}_b^{EB} - \boldsymbol{\theta}_b)^\top] + E[(\widehat{\boldsymbol{\theta}}_b^{EB} - \boldsymbol{\theta}_b)(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top]$. Then, it can be evaluated as

$$\begin{aligned}\boldsymbol{G}_{2ab}(\boldsymbol{\Psi}) =&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\boldsymbol{X}^\top(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}_b^\top(\boldsymbol{\Psi}+\boldsymbol{D}_b)^{-1}\boldsymbol{D}_b\\ &+\boldsymbol{D}_b(\boldsymbol{\Psi}+\boldsymbol{D}_b)^{-1}\boldsymbol{X}_b\{\boldsymbol{X}^\top(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a.\end{aligned}$$

The asymptotic expansion of the cumulative distribution function is

$$\begin{aligned}&P\{(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)^\top\boldsymbol{G}_{ab}^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)\le x\}\\ &=F_k(x)+2(\tilde{B}_1-\tilde{B}_2-\tilde{B}_3)f_{k+2}(x)+2\tilde{B}_2f_{k+4}(x)+o(m^{-1}),\end{aligned}$$

where $\tilde{B}_1$, $\tilde{B}_2$ and $\tilde{B}_3$ are

$$\begin{aligned}\tilde{B}_1 =&-\frac{1}{2}\mathrm{tr}\Big(E[\boldsymbol{G}_{ab}^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+\boldsymbol{G}_{1b}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1b}(\boldsymbol{\Psi}))\boldsymbol{G}_{ab}^{-2}(\boldsymbol{\Psi})\\ &\times(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+\boldsymbol{G}_{1b}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1b}(\boldsymbol{\Psi}))\boldsymbol{G}_{ab}^{-1/2}(\boldsymbol{\Psi})]\Big),\\ \tilde{B}_2 =&-\frac{1}{8}\Big\{E[\mathrm{tr}^2((\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+\boldsymbol{G}_{1b}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1b}(\boldsymbol{\Psi}))\boldsymbol{G}_{ab}^{-1}(\boldsymbol{\Psi}))]\\ &+2\mathrm{tr}\Big(E[((\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+\boldsymbol{G}_{1b}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1b}(\boldsymbol{\Psi}))\boldsymbol{G}_{ab}^{-1}(\boldsymbol{\Psi}))^2]\Big)\Big\},\\ \tilde{B}_3 =&\mathrm{tr}(\boldsymbol{G}_{ab}^{-1}(\boldsymbol{\Psi})(\boldsymbol{G}_{3a}(\boldsymbol{\Psi})+\boldsymbol{G}_{3b}(\boldsymbol{\Psi})-\boldsymbol{G}_{2ab}(\boldsymbol{\Psi}))),\end{aligned}\tag{2.6.1}$$

Setting $\tilde{h}^* = -2\{(\tilde{B}_1-\tilde{B}_3-\tilde{B}_2)/k+\tilde{B}_2x/k(k+2)\}$, we have

$$P((1+\tilde{h}^*)^{-1}(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)^\top\boldsymbol{G}_{ab}^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)\le x)=F_k(x)+o(m^{-1}),$$

namely, $P(\boldsymbol{\theta}_a-\boldsymbol{\theta}_b\in CR_{ab})=1-\alpha+o(m^{-1})$ for the corrected confidence region

$$CR_{ab}=\{\boldsymbol{\theta}_a-\boldsymbol{\theta}_b|(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)^\top\boldsymbol{G}_{ab}^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_b^{EB}+\boldsymbol{\theta}_b)\le(1+\tilde{h}^*)\chi^2_{k,1-\alpha}\}.$$

When the adjusted Prasd-Rao type estimator $\widehat{\boldsymbol{\Psi}}_{(A)}^{PR}$ given in (2.5.6) is used for estimating $\boldsymbol{\Psi}$, the functions $\tilde{B}_1$ and $\tilde{B}_2$ are calculated as Then, we have

$$\begin{aligned}\tilde{B}_1 =&-\frac{1}{2}\mathrm{tr}(\boldsymbol{V}_{1aa}+\boldsymbol{V}_{1bb}+\boldsymbol{V}_{1ab}+\boldsymbol{V}_{1ba})+o(m^{-1}),\\ \tilde{B}_2 =&-\frac{1}{4m^2}\sum_{i=1}^m\mathrm{tr}\Big(\Big\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{G}_{ab}^{-1}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\ &+(\boldsymbol{\Psi}+\boldsymbol{D}_b)^{-1}\boldsymbol{D}_b\boldsymbol{G}_{ab}^{-1}\boldsymbol{D}_b(\boldsymbol{\Psi}+\boldsymbol{D}_b)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}^2\Big)\\ &-\frac{1}{4}\mathrm{tr}(\boldsymbol{V}_{2aa}+\boldsymbol{V}_{2bb}+\boldsymbol{V}_{2ab}+\boldsymbol{V}_{2ba})+o(m^{-1}),\end{aligned}$$

where for $(c,d)=(a,a),(a,b),(b,a)$ and (b,b),

$$
\begin{aligned}
\boldsymbol{V}_{1cd} =&\boldsymbol{G}_{cd}^{-1/2}(\boldsymbol{\Psi})\boldsymbol{D}_c(\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\\
&\times\Big[\frac{1}{m^2}\sum_{i=1}^m\Big\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\boldsymbol{D}_c\boldsymbol{G}_{cd}^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_d(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\
&\qquad+\mathrm{tr}\,((\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\boldsymbol{D}_c\boldsymbol{G}_{cd}^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_d(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}\Big]\\
&\times\boldsymbol{D}_d(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}\boldsymbol{G}_{cd}^{-1/2}(\boldsymbol{\Psi}),\\
\boldsymbol{V}_{2cd} =&\boldsymbol{G}_{cd}^{-1/2}(\boldsymbol{\Psi})\boldsymbol{D}_c(\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\\
&\times\Big[\frac{1}{m^2}\sum_{i=1}^m\Big\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\boldsymbol{D}_c\boldsymbol{G}_{cd}^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_d(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\
&\qquad+\mathrm{tr}\,((\boldsymbol{\Psi}+\boldsymbol{D}_c)^{-1}\boldsymbol{D}_c\boldsymbol{G}_{cd}^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_d(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}\Big]\\
&\times(\boldsymbol{\Psi}+\boldsymbol{D}_d)^{-1}\boldsymbol{D}_d\boldsymbol{G}_{cd}^{-1/2}(\boldsymbol{\Psi}).
\end{aligned}
$$

Also, $\tilde{B}_3$ is obtained by using the expression in (2.5.7).

2.7 Simulation and Empirical Studies

2.7.1 Finite sample performances

We now investigate finite sample performances of EBLUP in terms of MSEM and the second-order unbiased estimator of MSEM by simulation.

[1] Setup of simulation experiments. We treat the multivariate Fay-Herriot model (2.2.1) for $k=2,3$ and $m=30,60$ without covariates, namely $\boldsymbol{X}_i=\boldsymbol{I}_k$. As a setup of the covariance matrix $\boldsymbol{\Psi}$ of the random effects, we consider

$$
\boldsymbol{\Psi}=\begin{cases}\rho\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top+(1-\rho)\mathrm{diag}(\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top) & \text{for } k=2,\\ \rho\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top+(1-\rho)\mathrm{diag}(\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top) & \text{for } k=3,\end{cases}
$$

where $\boldsymbol{\psi}_2=(\sqrt{1.5},\sqrt{0.5})^\top$, $\boldsymbol{\psi}_3=(\sqrt{1.5},1,\sqrt{0.5})^\top$, and $\mathrm{diag}(\boldsymbol{A})$ denotes the diagonal matrix consisting of diagonal elements of matrix $\boldsymbol{A}$. Here, ρ is the correlation coefficient, and we handle the three cases $\rho=0.25,0.5,0.75$. The cases of negative correlations are omitted, because we observe the same results with those of positive ones.

Concerning the dispersion matrices $\boldsymbol{D}_i$ of sampling errors $\boldsymbol{\varepsilon}_i$, we treat the two $\boldsymbol{D}_i$-patterns: (a) $0.7\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.3\mathbf{I}_k$ and (b) $2.0\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.2\mathbf{I}_k$. In the univariate Fay-Herriot model, these cases are treated by Datta, *et al.* (2005). There are five groups $G_1,\ldots,G_5$ corresponding to these $\boldsymbol{D}_i$-patterns, and there are six and twelve small areas in each group for $m=30$ and 60, respectively, where the sampling covariance matrices $\boldsymbol{D}_i$ are the same for areas within the same group.

[2] Comparison of MSEM. We begin with obtaining the true mean squared error matrices of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB}=\widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}^+)$ by simulation. Let $\{\boldsymbol{y}_i^{(r)}, i=1,\ldots,m\}$ be the simulated data in the r-th replication for $r = 1, \ldots, R$ with $R = 50,000$. Let $\widehat{\mathbf{\Psi}}^{+(r)}$ and $\boldsymbol{\theta}_a^{(r)}$ be the values of $\widehat{\mathbf{\Psi}}^{+}$ and $\boldsymbol{\theta}_a$ in the r-th replication. Then the simulated value of the true mean squared error matrices is calculated by

$$\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = R^{-1}\sum_{i=1}^{R}\big\{\widehat{\boldsymbol{\theta}}_a(\widehat{\mathbf{\Psi}}^{+(r)}) - \boldsymbol{\theta}_a^{(r)}\big\}\big\{\widehat{\boldsymbol{\theta}}_a(\widehat{\mathbf{\Psi}}^{+(r)}) - \boldsymbol{\theta}_a^{(r)}\big\}^{\top}.$$

As an estimator of $\mathbf{\Psi}$, we here use the simple estimator $\widehat{\mathbf{\Psi}}_0^{+}$, because there is little difference between $\widehat{\mathbf{\Psi}}_0^{+}$ and $\widehat{\mathbf{\Psi}}_1^{+}$ in simulated values of MSEM under the setup of $\boldsymbol{X}_i = \mathbf{I}_k$. Simulated values of the mean squared error matrices, averaged over areas within groups G_t, are reported in Tables 2.1, 2.3, and 2.5. To measure relative improvement of EBLUP, we calculate the percentage relative improvement in the average loss (PRIAL) of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$, defined by

$$\mathrm{PRIAL}(\widehat{\boldsymbol{\theta}}_a^{EB}, \boldsymbol{y}_a) = 100 \times \Big[1 - \frac{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\}}{\mathrm{tr}\,\{\mathrm{MSEM}(\boldsymbol{y}_a)\}}\Big].$$

It is also interesting to compare $\widehat{\boldsymbol{\theta}}_a^{EB}$ with the EBLUP $\widehat{\boldsymbol{\theta}}_a^{uEB}$ derived from the univariate Fay-Herriot model. Thus, we calculate the PRIAL given by

$$\mathrm{PRIAL}(\widehat{\boldsymbol{\theta}}_a^{EB}, \widehat{\boldsymbol{\theta}}_a^{uEB}) = 100 \times \Big[1 - \frac{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\}}{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{uEB})\}}\Big],$$

and those values are reported in Tables 2.2, 2.4 and 2.6.

Table 2.1 reports the simulated values of the true MSEM of $\widehat{\boldsymbol{\theta}}_a^{EB}$ for $k = 2$, $\boldsymbol{D}_i$-patterns (a), $m = 30, 60$ and $\rho = 0.25, 0.5, 0.75$. For fixed m, the values of MSEM decrease as the correlation ρ in the random effect becomes large. For fixed ρ, the values of MSEM decrease as m becomes large. Table 2.2 reports the values of PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ under the same setup as in Table 2.1. In all the cases, $\widehat{\boldsymbol{\theta}}_a^{EB}$ improves on $\boldsymbol{y}_a$ largely and the improvement rates are larger for larger ρ. In comparison with $\widehat{\boldsymbol{\theta}}_a^{uEB}$, the univariate EBLUP $\widehat{\boldsymbol{\theta}}_a^{uEB}$ is slightly better than $\widehat{\boldsymbol{\theta}}_a^{EB}$ for $\rho = 0.25$, but the difference is not significant. The values of PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\widehat{\boldsymbol{\theta}}_a^{uEB}$ get larger as ρ increases. In the case of $m = 60$, the improvements of $\widehat{\boldsymbol{\theta}}_a^{EB}$ in light of PRIAL get larger for larger ρ. In the case of $\rho = 0.25$, the improvement of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\widehat{\boldsymbol{\theta}}_a^{uEB}$ is better for $m = 60$ than for $m = 30$. This is because the low accuracy in estimation of the covariance matrix $\mathbf{\Psi}$ has more adverse influence on prediction than the benefit from incorporating the small correlation into the estimation.

The comparison of performances between $\boldsymbol{D}_i$-patterns (a) and (b) is investigated in Tables 2.3 and 2.4. The simulated values of the MSEM of $\widehat{\boldsymbol{\theta}}_a^{EB}$ in $\boldsymbol{D}_i$-patterns (a) and (b) are reported in Table 2.3 for $k = 2$, $m = 30, 60$ and $\rho = 0.5$. As the increment of variance of sampling error in G_1, the MSEM in G_1 becomes larger, and the other groups have slightly larger MSEM except G_5. The values of PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ are given in Table 2.4 for $\boldsymbol{D}_i$-patterns (a) and (b). under the same setup as in Table 2.3. As seen from the table, the improvement of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ in G_1 is larger for $\boldsymbol{D}_i$-pattern (b) because of the large sampling variance. However, $\widehat{\boldsymbol{\theta}}_a^{EB}$

is not better than $\widehat{\boldsymbol{\theta}}^{uEB}$ in G_4 and G_5 for $m = 30$ and in G_5 for $m = 60$ in $\boldsymbol{D}_i$-pattern (b). This implies that incorporating the information of areas with large sampling variances affects more adversely estimation of areas with small sampling variances in the multivariate model than in the univariate model.

Tables 2.5 and 2.6 report the values of MSEM and PRIAL for $k = 3$, $m = 30$, $\rho = 0.5$ and $\boldsymbol{D}_i$-pattern (a). From Table 2.6, it is revealed that PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ are larger for $k = 3$ than for $k = 2$ in the case of $\rho = 0.75$, but smaller in the case of $\rho = 0.25$. When m is fixed as $m = 30$, the accuracy in estimation of the covariance matrix $\boldsymbol{\Psi}$ gets smaller for the larger dimension. This demonstrates that it is not appropriate to treat the multivariate Fay-Herriot model with a large covariance matrix when m is not large.

$m = 30$

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
G_1	$\begin{bmatrix} 49.8 & 3.8 \\ 3.8 & 32.6 \end{bmatrix}$	$\begin{bmatrix} 48.7 & 8.1 \\ 8.1 & 30.1 \end{bmatrix}$	$\begin{bmatrix} 46.5 & 13.8 \\ 13.8 & 25.3 \end{bmatrix}$
G_2	$\begin{bmatrix} 44.7 & 3.1 \\ 3.1 & 30.4 \end{bmatrix}$	$\begin{bmatrix} 43.8 & 6.5 \\ 6.5 & 28.3 \end{bmatrix}$	$\begin{bmatrix} 41.4 & 11.6 \\ 11.6 & 23.7 \end{bmatrix}$
G_3	$\begin{bmatrix} 39.0 & 2.4 \\ 2.4 & 27.9 \end{bmatrix}$	$\begin{bmatrix} 38.0 & 5.3 \\ 5.3 & 26.3 \end{bmatrix}$	$\begin{bmatrix} 36.6 & 9.2 \\ 9.2 & 21.8 \end{bmatrix}$
G_4	$\begin{bmatrix} 33.1 & 1.7 \\ 1.7 & 25.3 \end{bmatrix}$	$\begin{bmatrix} 32.4 & 3.8 \\ 3.8 & 23.6 \end{bmatrix}$	$\begin{bmatrix} 30.6 & 6.8 \\ 6.8 & 19.8 \end{bmatrix}$
G_5	$\begin{bmatrix} 26.1 & 1.1 \\ 1.1 & 21.6 \end{bmatrix}$	$\begin{bmatrix} 25.6 & 2.3 \\ 2.3 & 20.4 \end{bmatrix}$	$\begin{bmatrix} 24.2 & 4.6 \\ 4.6 & 17.4 \end{bmatrix}$

$m = 60$

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
G_1	$\begin{bmatrix} 49.0 & 4.1 \\ 4.1 & 30.7 \end{bmatrix}$	$\begin{bmatrix} 47.4 & 8.2 \\ 8.2 & 28.0 \end{bmatrix}$	$\begin{bmatrix} 45.2 & 14.0 \\ 14.0 & 23.6 \end{bmatrix}$
G_2	$\begin{bmatrix} 43.5 & 3.4 \\ 3.4 & 28.6 \end{bmatrix}$	$\begin{bmatrix} 42.5 & 7.0 \\ 7.0 & 26.5 \end{bmatrix}$	$\begin{bmatrix} 40.3 & 11.7 \\ 11.7 & 22.1 \end{bmatrix}$
G_3	$\begin{bmatrix} 37.9 & 2.6 \\ 2.6 & 26.0 \end{bmatrix}$	$\begin{bmatrix} 37.1 & 5.7 \\ 5.7 & 24.5 \end{bmatrix}$	$\begin{bmatrix} 35.2 & 9.6 \\ 9.6 & 20.4 \end{bmatrix}$
G_4	$\begin{bmatrix} 31.9 & 1.8 \\ 1.8 & 23.4 \end{bmatrix}$	$\begin{bmatrix} 31.4 & 4.1 \\ 4.1 & 21.8 \end{bmatrix}$	$\begin{bmatrix} 29.8 & 7.3 \\ 7.3 & 18.5 \end{bmatrix}$
G_5	$\begin{bmatrix} 25.2 & 1.2 \\ 1.2 & 19.8 \end{bmatrix}$	$\begin{bmatrix} 24.8 & 2.7 \\ 2.7 & 18.7 \end{bmatrix}$	$\begin{bmatrix} 23.8 & 5.1 \\ 5.1 & 16.1 \end{bmatrix}$

Table 2.1: Simulated values of mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ multipled by 100 for $k = 2$, $\boldsymbol{D}_i$-patterns (a)

[3] MSEM approximation and its estimator. We next investigate the performance of the second-order approximation of MSEM of EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB}$ given in (2.3.4) and the second-order unbiased estimator msem$(\widehat{\boldsymbol{\theta}}_a^{EB})$ of MSEM given in (2.3.5). The values of the second-order approximation of MSEM are given in Table 2.7 for $k = 2$, $m = 3$ and $\boldsymbol{D}_i$-pattern (a). Comparing the values in Table 2.7 with the corresponding true values of the MSEM in Table 2.1, we can

		$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$			$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$	
$m = 30$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
G_1	41.2	43.8	48.9	-0.5	3.8	11.6
G_2	37.2	40.1	45.7	0.0	3.5	12.3
G_3	33.0	35.8	41.8	-0.7	3.4	11.8
G_4	27.3	29.8	37.2	-1.9	1.8	11.0
G_5	20.8	23.5	30.4	-2.5	1.1	10.0
		$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$			$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$	
$m = 60$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
G_1	43.2	45.8	51.0	-0.6	4.6	13.9
G_2	39.8	42.4	48.1	0.2	5.1	13.6
G_3	35.6	38.7	44.2	1.3	4.6	14.3
G_4	30.6	33.7	39.8	0.3	3.5	13.2
G_5	24.8	27.5	33.6	0.4	2.9	11.0

Table 2.2: PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ for $k = 2$, $\boldsymbol{D}_i$-patterns (a)

$m = 30$	Pattern (a)	Pattern (b)	$m = 60$	Pattern (a)	Pattern (b)
G_1	$\begin{bmatrix} 48.7 & 8.1 \\ 8.1 & 30.1 \end{bmatrix}$	$\begin{bmatrix} 89.9 & 19.7 \\ 19.7 & 42.9 \end{bmatrix}$	G_1	$\begin{bmatrix} 47.4 & 8.2 \\ 8.2 & 28.0 \end{bmatrix}$	$\begin{bmatrix} 86.8 & 20.1 \\ 20.1 & 40.0 \end{bmatrix}$
G_2	$\begin{bmatrix} 43.8 & 6.5 \\ 6.5 & 28.3 \end{bmatrix}$	$\begin{bmatrix} 44.5 & 6.0 \\ 6.0 & 30.2 \end{bmatrix}$	G_2	$\begin{bmatrix} 42.5 & 7.0 \\ 7.0 & 26.5 \end{bmatrix}$	$\begin{bmatrix} 42.9 & 6.5 \\ 6.5 & 27.8 \end{bmatrix}$
G_3	$\begin{bmatrix} 38.0 & 5.3 \\ 5.3 & 26.3 \end{bmatrix}$	$\begin{bmatrix} 39.3 & 4.7 \\ 4.7 & 28.3 \end{bmatrix}$	G_3	$\begin{bmatrix} 37.1 & 5.7 \\ 5.7 & 24.5 \end{bmatrix}$	$\begin{bmatrix} 37.8 & 5.0 \\ 5.0 & 25.8 \end{bmatrix}$
G_4	$\begin{bmatrix} 32.4 & 3.8 \\ 3.8 & 23.6 \end{bmatrix}$	$\begin{bmatrix} 33.4 & 3.2 \\ 3.2 & 25.9 \end{bmatrix}$	G_4	$\begin{bmatrix} 31.4 & 4.1 \\ 4.1 & 21.8 \end{bmatrix}$	$\begin{bmatrix} 32.0 & 3.6 \\ 3.6 & 23.8 \end{bmatrix}$
G_5	$\begin{bmatrix} 25.6 & 2.3 \\ 2.3 & 20.4 \end{bmatrix}$	$\begin{bmatrix} 19.1 & 0.1 \\ 0.1 & 18.8 \end{bmatrix}$	G_5	$\begin{bmatrix} 24.8 & 2.7 \\ 2.7 & 18.7 \end{bmatrix}$	$\begin{bmatrix} 18.1 & 0.6 \\ 0.6 & 16.4 \end{bmatrix}$

Table 2.3: Simulated values of mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ multipled by 100 for $k = 2$, $\rho = 0.5$

	$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$		$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$	
$m = 30$	Pattern (a)	Pattern (b)	Pattern (a)	Pattern (b)
G_1	43.8	66.4	3.8	2.1
G_2	40.1	37.0	3.5	0.8
G_3	35.8	32.1	3.4	1.0
G_4	29.8	26.2	1.8	-0.2
G_5	23.5	4.2	1.1	-8.5
	$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$		$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$	
$m = 60$	Pattern (a)	Pattern (b)	Pattern (a)	Pattern (b)
G_1	45.8	68.5	4.6	3.1
G_2	42.4	40.7	5.1	3.1
G_3	38.7	36.2	4.6	2.7
G_4	33.7	30.6	3.5	1.3
G_5	27.5	13.9	2.9	-2.7

Table 2.4: PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ for $k = 2$, $m = 30, 60$, $\rho = 0.5$, $\boldsymbol{D}_i$-patterns (a), (b)

	$m = 30$		
	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
G1	$\begin{bmatrix} 50.0 & 3.4 & 3.5 \\ 3.4 & 44.3 & 3.4 \\ 3.5 & 3.4 & 33.3 \end{bmatrix}$	$\begin{bmatrix} 48.0 & 7.0 & 6.4 \\ 7.0 & 41.1 & 6.5 \\ 6.4 & 6.5 & 29.5 \end{bmatrix}$	$\begin{bmatrix} 42.0 & 12.3 & 10.0 \\ 12.3 & 34.8 & 9.4 \\ 10.0 & 9.4 & 23.1 \end{bmatrix}$
G2	$\begin{bmatrix} 45.2 & 2.6 & 2.8 \\ 2.6 & 39.8 & 2.9 \\ 2.8 & 2.9 & 31.2 \end{bmatrix}$	$\begin{bmatrix} 42.8 & 5.8 & 5.4 \\ 5.8 & 37.7 & 5.5 \\ 5.4 & 5.5 & 28.2 \end{bmatrix}$	$\begin{bmatrix} 38.2 & 10.0 & 8.3 \\ 10.0 & 31.8 & 8.0 \\ 8.3 & 8.0 & 21.7 \end{bmatrix}$
G3	$\begin{bmatrix} 40.0 & 2.0 & 1.9 \\ 1.9 & 36.1 & 2.1 \\ 1.9 & 2.1 & 29.0 \end{bmatrix}$	$\begin{bmatrix} 37.5 & 4.1 & 3.9 \\ 4.1 & 33.9 & 4.1 \\ 3.9 & 4.1 & 25.8 \end{bmatrix}$	$\begin{bmatrix} 33.5 & 7.7 & 7.0 \\ 7.7 & 28.8 & 6.4 \\ 7.0 & 6.4 & 20.5 \end{bmatrix}$
G4	$\begin{bmatrix} 33.4 & 1.3 & 1.5 \\ 1.3 & 31.0 & 1.6 \\ 1.5 & 1.6 & 26.0 \end{bmatrix}$	$\begin{bmatrix} 32.1 & 2.7 & 2.9 \\ 2.7 & 29.2 & 3.0 \\ 2.9 & 3.0 & 20.7 \end{bmatrix}$	$\begin{bmatrix} 29.2 & 5.6 & 5.1 \\ 5.6 & 25.2 & 5.1 \\ 5.1 & 5.1 & 18.4 \end{bmatrix}$
G5	$\begin{bmatrix} 26.3 & 0.7 & 0.7 \\ 0.7 & 25.4 & 1.0 \\ 0.7 & 1.0 & 22.7 \end{bmatrix}$	$\begin{bmatrix} 25.8 & 1.6 & 1.5 \\ 1.6 & 24.1 & 1.8 \\ 1.5 & 1.8 & 20.7 \end{bmatrix}$	$\begin{bmatrix} 23.4 & 3.1 & 3.2 \\ 3.1 & 21.0 & 3.4 \\ 3.2 & 3.4 & 16.5 \end{bmatrix}$

Table 2.5: Simulated values of mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ multiplied by 100 for $k = 3$, $m = 30$, $\boldsymbol{D}_i$-patterns (a)

	$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$			$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$		
$k=2$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$
G_1	41.2	43.8	48.9	-0.5	3.8	11.6
G_2	37.2	40.1	45.7	0.0	3.5	12.3
G_3	33.0	35.8	41.8	-0.7	3.4	11.8
G_4	27.3	29.8	37.2	-1.9	1.8	11.0
G_5	20.8	23.5	30.4	-2.5	1.1	10.0
	$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$			$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$		
$k=3$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$
G1	39.5	43.6	52.4	-1.9	5.9	20.3
G2	35.2	40.0	48.9	-2.6	4.8	19.2
G3	30.3	35.1	44.6	-3.9	3.2	18.2
G4	25.2	29.7	39.6	-4.4	1.9	15.7
G5	17.6	21.6	32.0	-5.7	-0.3	12.8

Table 2.6: PRIAL of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$ for $k=2,3$, $m=30$, $\boldsymbol{D}_i$-patterns (a)

see that the second-order approximation can approximate the true MSEM precisely for every G_t and ρ.

Concerning the performance of the second-order unbiased estimator msem$(\widehat{\boldsymbol{\theta}}_a^{EB})$ given in (2.3.3), we compute the simulated values of relative bias of the estimator msem$(\widehat{\boldsymbol{\theta}}_a^{EB})$, averaged over areas within groups G_t. Those values are reported in Table 2.8 for $k=2$, $m=30,60$ and $\boldsymbol{D}_i$-pattern (a). It is revealed from Table 2.8 that the relative bias gets larger for larger ρ. Also, the values of the relative bias are smaller for $m=60$ than for $m=30$, namely, the relative bias gets small as m increases.

	$m=30$		
	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$
G_1	$\begin{bmatrix} 49.8 & 3.7 \\ 3.7 & 32.6 \end{bmatrix}$	$\begin{bmatrix} 48.6 & 7.9 \\ 7.9 & 30.3 \end{bmatrix}$	$\begin{bmatrix} 46.2 & 13.2 \\ 13.2 & 25.9 \end{bmatrix}$
G_2	$\begin{bmatrix} 44.6 & 3.1 \\ 3.1 & 30.4 \end{bmatrix}$	$\begin{bmatrix} 43.6 & 6.6 \\ 6.6 & 28.4 \end{bmatrix}$	$\begin{bmatrix} 41.5 & 11.1 \\ 11.1 & 24.4 \end{bmatrix}$
G_3	$\begin{bmatrix} 38.9 & 2.4 \\ 2.4 & 27.8 \end{bmatrix}$	$\begin{bmatrix} 38.1 & 5.2 \\ 5.2 & 26.1 \end{bmatrix}$	$\begin{bmatrix} 36.3 & 8.9 \\ 8.9 & 22.6 \end{bmatrix}$
G_4	$\begin{bmatrix} 32.6 & 1.7 \\ 1.7 & 24.7 \end{bmatrix}$	$\begin{bmatrix} 32.0 & 3.8 \\ 3.8 & 23.3 \end{bmatrix}$	$\begin{bmatrix} 30.6 & 6.6 \\ 6.6 & 20.5 \end{bmatrix}$
G_5	$\begin{bmatrix} 25.7 & 1.1 \\ 1.1 & 20.7 \end{bmatrix}$	$\begin{bmatrix} 25.3 & 2.4 \\ 2.4 & 20.0 \end{bmatrix}$	$\begin{bmatrix} 24.4 & 4.3 \\ 4.3 & 17.8 \end{bmatrix}$

Table 2.7: Second order approximations of mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ multiplied by 100 for $k=2$, $\boldsymbol{D}_i$-patterns (a)

Pattern (a)

$m=30$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$
G_1	$\begin{bmatrix} -0.3 & -2.6 \\ -2.6 & 1.1 \end{bmatrix}$	$\begin{bmatrix} -0.9 & -4.1 \\ -4.1 & 2.9 \end{bmatrix}$	$\begin{bmatrix} 0.6 & -9.5 \\ -9.5 & 10.1 \end{bmatrix}$
G_2	$\begin{bmatrix} 0.6 & 3.1 \\ 3.1 & 0.9 \end{bmatrix}$	$\begin{bmatrix} 0.3 & -3.5 \\ -3.5 & 2.7 \end{bmatrix}$	$\begin{bmatrix} 1.1 & -10.4 \\ -10.4 & 13.1 \end{bmatrix}$
G_3	$\begin{bmatrix} -0.6 & -5.8 \\ -5.8 & 1.2 \end{bmatrix}$	$\begin{bmatrix} 1.3 & -7.8 \\ -7.8 & 4.6 \end{bmatrix}$	$\begin{bmatrix} 1.2 & -16.7 \\ -16.7 & 13.6 \end{bmatrix}$
G_4	$\begin{bmatrix} -0.4 & -4.6 \\ -4.6 & 2.9 \end{bmatrix}$	$\begin{bmatrix} 0.4 & -10.8 \\ -10.8 & 4.7 \end{bmatrix}$	$\begin{bmatrix} 1.2 & -23.4 \\ -23.4 & 17.8 \end{bmatrix}$
G_5	$\begin{bmatrix} 0.3 & -24.4 \\ -24.4 & 2.2 \end{bmatrix}$	$\begin{bmatrix} 0.6 & -26.1 \\ -26.1 & 7.7 \end{bmatrix}$	$\begin{bmatrix} 3.4 & -42.3 \\ -42.3 & 23.1 \end{bmatrix}$

Pattern (a)

$m=60$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$
G_1	$\begin{bmatrix} -0.1 & -2.3 \\ -2.3 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.2 & -0.2 \\ -0.2 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.4 & -0.7 \\ -0.7 & 1.7 \end{bmatrix}$
G_2	$\begin{bmatrix} 0.7 & -3.4 \\ -3.4 & -0.2 \end{bmatrix}$	$\begin{bmatrix} 0.4 & -0.8 \\ -0.8 & -0.5 \end{bmatrix}$	$\begin{bmatrix} 0.8 & -0.0 \\ -0.0 & 2.1 \end{bmatrix}$
G_3	$\begin{bmatrix} 0.2 & -5.1 \\ -5.1 & -0.2 \end{bmatrix}$	$\begin{bmatrix} 0.1 & -1.5 \\ -1.5 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 0.3 & -1.4 \\ -1.4 & 2.9 \end{bmatrix}$
G_4	$\begin{bmatrix} -0.1 & -5.1 \\ -5.1 & -0.4 \end{bmatrix}$	$\begin{bmatrix} -0.4 & -2.4 \\ -2.4 & 0.3 \end{bmatrix}$	$\begin{bmatrix} 1.4 & -2.9 \\ -2.9 & 3.6 \end{bmatrix}$
G_5	$\begin{bmatrix} 0.2 & -3.3 \\ -3.3 & -0.2 \end{bmatrix}$	$\begin{bmatrix} 0.3 & -5.9 \\ -5.9 & -0.1 \end{bmatrix}$	$\begin{bmatrix} 0.6 & -8.1 \\ -8.1 & 5.1 \end{bmatrix}$

Table 2.8: Simulated values of percentage average relative bias of mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ multiplied by 100 for $k=2$, $m=30,60$, $\boldsymbol{D}_i$-pattern (a)

[4] Confidence region. We next investigate finite sample performances of the proposed confidence regions by simulation in the multivariate Fay-Herriot model (2.2.1) for $k = 2, 3$ and $m = 30$. The design matrix, $\boldsymbol{X}_i$ is a $k \times 2k$ matrix, such that

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & x_{i1} & 0 & 0 \\ 0 & 0 & 1 & x_{i2} \end{pmatrix}, \boldsymbol{X}_i = \begin{pmatrix} 1 & x_{i1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{i2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{i3} \end{pmatrix}$$

for $k = 2, 3$ respectevely, where x_{ij}'s are generated from the uniform distribution on $(-1, 1)$, which are fixed through the simulation runs. As a setup of the covariance matrix $\boldsymbol{\Psi}$ of the random effects, we consider

$$\boldsymbol{\Psi} = \begin{cases} \rho\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top + (1-\rho)\mathrm{diag}(\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top) & \text{for } k = 2, \\ \rho\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top + (1-\rho)\mathrm{diag}(\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top) & \text{for } k = 3, \end{cases}$$

where $\boldsymbol{\psi}_2 = (\sqrt{1.6}, \sqrt{0.8})^\top$, $\boldsymbol{\psi}_3 = (\sqrt{1.6}, \sqrt{1.2}, \sqrt{0.8})^\top$, and $\mathrm{diag}(\boldsymbol{A})$ denotes the diagonal matrix consisting of diagonal elements of matrix $\boldsymbol{A}$. Here, ρ is the correlation coefficient, and we handle the three cases $\rho = 0.2, 0.4, 0.6$. The cases of negative correlations are omitted, because we observe the same results with those of positive ones.

The dispersion matrices $\boldsymbol{D}_i$ of sampling errors $\boldsymbol{\varepsilon}_i$ are set in the smae way as before. Then, we treat the two $\boldsymbol{D}_i$-patterns: (a) $0.7\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.3\mathbf{I}_k$ and (b) $2.0\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.2\mathbf{I}_k$. There are five groups $G_1, \ldots, G_5$ corresponding to these $\boldsymbol{D}_i$-patterns, and there are six small areas in each group for $m = 30$, respectively, where the sampling covariance matrices $\boldsymbol{D}_i$ are the same for areas within the same group. Concerning the underlying distributions for $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$, we consider two kinds of distributions, that is, multivariate normal distributions and multivariate normalized chi-squared distributions with degrees of freedom 2, which are denoted by M1 and M2, respectively. The chis-quared distribution is used for investigating robustness of the proposed method against the misspecification of distributions of $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$. The values of coverage probabilities (CP) of the corrected confidence region and the naive confidence region and the values of the Bartlett-type correction term h^* are obtained based on $10,000$ simulation run, where the nominal confidence coefficient is 95%.

The values of CP and the correction term in the case of $k = 2$ are reported in Tables 2.9 and 2.10 for $\boldsymbol{D}_i$-patterns (a) and (b), respectively. From the tables for normal distributions, the corrected method has CP values larger than the nominal confidence coefficient. In contrast, CP values of the naive confidence region are much smaller than the nominal confidence coefficient. For example, CP value for G_1 in Table 2.10 is about 89%. These show that the naive method is not appropriate for a confidence region and the correction by h^* works well. For chi-square distributions, CP values of the corrected method satisfies the nominal confidence coefficient in most cases except few cases where CP values are slightly smaller than, but close to 95%, while the performance of the naive method is worse than that in the normal distributions. Thus, the corrected method remains good and robust for the chi-square distributions. Concerning the Bartlett-type correction, it increases as sampling variances or correlation coefficients ρ increase.

Table 2.11 reports the results for $k = 3$ and $\boldsymbol{D}_i$-pattern (a). Comparing Tables 2.9 and 2.11, we can observe that CP values of the naive confidence region are worse in $k = 3$ than those in $k = 2$. The corrected confidence region satisfies the nominal confidence coefficient for $k = 3$ in most cases except the case of $\rho = 0.2$ in chi-square distributions. Hence, the corrected method works well and is robust still for $k = 3$.

We next investigate the finite sample performance of the corrected confidence region for the difference of two small area means, $\boldsymbol{\theta}_a - \boldsymbol{\theta}_b$ for $k = 2$ and $\boldsymbol{D}_i$-pattern (a), where the corrected method is provided in Section 2.6. In each area group, we consider the difference between the first two small areas means. Table 2.12 reports values of the coverage provabilities (CP) and the Bartlett-type correction term h^* for $\boldsymbol{\theta}_a - \boldsymbol{\theta}_b$. From Table 2.12, it is revealed that the performances are similar to the results in Table 2.9, while values of the Bartlett-type correction term h^* are larger for $\boldsymbol{\theta}_a - \boldsymbol{\theta}_b$.

		Normal			chi-square		
ρ		0.2	0.4	0.6	0.2	0.4	0.6
G_1	CP	0.955	0.968	0.974	0.939	0.945	0.956
		(0.917)	(0.923)	(0.917)	(0.898)	(0.901)	(0.907)
	h^*	0.429	0.492	0.760	0.636	0.697	0.841
G_2	CP	0.962	0.960	0.977	0.941	0.942	0.954
		(0.923)	(0.913)	(0.922)	(0.902)	(0.899)	(0.912)
	h^*	0.534	0.571	0.865	0.598	0.640	0.758
G_3	CP	0.958	0.962	0.978	0.939	0.947	0.953
		(0.921)	(0.921)	(0.922)	(0.901)	(0.906)	(0.912)
	h^*	0.470	0.530	0.843	0.669	0.731	0.849
G_4	CP	0.959	0.965	0.973	0.939	0.944	0.953
		(0.928)	(0.928)	(0.925)	(0.905)	(0.908)	(0.911)
	h^*	0.388	0.441	0.688	0.552	0.610	0.742
G_5	CP	0.954	0.962	0.976	0.951	0.947	0.955
		(0.923)	(0.927)	(0.930)	(0.914)	(0.914)	(0.924)
	h^*	0.441	0.470	0.734	0.480	0.519	0.622

Table 2.9: Coverage probabilities (CP) for nominal 95% confidence regions for $k = 2$ and $\boldsymbol{D}_i$-pattern (a). (the corrected method in the first line and the naive method in parentheses)

2.7.2 Illustrative example

This example, primarily for illustration, uses the multivariate Fay-Herriot model (2.2.1) and data from the 2016 Survey of Family Income and Expenditure in Japan, which is based on two or more person households (excluding agricultural, forestry and fisheries households). The target domains are the 47 Japanese prefectural capitals. The 47 prefectures are divided into 10 regions: Hokkaido, Tohoku, Kanto, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, Kyushu and Okinawa. Each region consists of several prefectures except Hokkaido and Okinawa, which consist of one prefecture.

In this study, as observations $(y_{i1}, y_{i2})^\top$, we use the reported data of the yearly averaged monthly spendings on 'Education' and ' Cultural-amusement' per worker's household, scaled by 1,000 Yen, at each capital city of 47 prefectures. In addition, we use the data in the 2014 National Survey of Family Income and Expenditure. The average spending data in this survey

		Normal			chi-square		
ρ		0.2	0.4	0.6	0.2	0.4	0.6
G_1	CP	0.974	0.980	0.990	0.952	0.957	0.963
		(0.897)	(0.895)	(0.899)	(0.876)	(0.876)	(0.891)
	h^*	1.288	1.645	2.571	1.962	2.145	2.270
G_2	CP	0.969	0.980	0.987	0.953	0.960	0.967
		(0.905)	(0.908)	(0.909)	(0.885)	(0.894)	(0.897)
	h^*	1.493	1.870	2.979	2.207	2.380	2.466
G_3	CP	0.967	0.976	0.984	0.954	0.961	0.969
		(0.912)	(0.908)	(0.908)	(0.894)	(0.898)	(0.901)
	h^*	1.288	1.730	2.876	1.933	2.173	2.332
G_4	CP	0.967	0.974	0.982	0.953	0.958	0.965
		(0.916)	(0.918)	(0.914)	(0.898)	(0.898)	(0.907)
	h^*	1.107	1.433	2.319	1.695	1.854	1.950
G_5	CP	0.966	0.973	0.980	0.954	0.957	0.965
		(0.926)	(0.921)	(0.922)	(0.904)	(0.909)	(0.917)
	h^*	1.169	1.494	2.471	1.696	1.857	1.928

Table 2.10: Coverage probabilities (CP) for nominal 95% confidence regions for $k = 2$ and $\boldsymbol{D}_i$-pattern (b). (the corrected method in the first line and the naive method in parentheses)

		Normal			chi-square		
ρ		0.2	0.4	0.6	0.2	0.4	0.6
G_1	CP	0.964	0.977	0.987	0.941	0.953	0.964
		(0.897)	(0.903)	(0.917)	(0.877)	(0.884)	(0.887)
	h^*	0.527	0.675	0.809	0.816	0.890	1.284
G_2	CP	0.964	0.976	0.989	0.950	0.951	0.965
		(0.897)	(0.897)	(0.920)	(0.883)	(0.878)	(0.886)
	h^*	0.570	0.734	0.882	0.884	0.975	1.440
G_3	CP	0.966	0.975	0.986	0.943	0.955	0.967
		(0.903)	(0.903)	(0.917)	(0.879)	(0.884)	(0.889)
	h^*	0.579	0.744	0.876	0.891	0.981	1.454
G_4	CP	0.965	0.973	0.985	0.940	0.952	0.965
		(0.908)	(0.910)	(0.923)	(0.881)	(0.889)	(0.893)
	h^*	0.488	0.630	0.755	0.772	0.843	1.252
G_5	CP	0.964	0.972	0.983	0.944	0.953	0.968
		(0.916)	(0.912)	(0.922)	(0.893)	(0.898)	(0.904)
	h^*	0.474	0.610	0.727	0.761	0.826	1.247

Table 2.11: Coverage probabilities (CP) for nominal 95% confidence regions for $k = 3$ and $\boldsymbol{D}_i$-pattern (a). (the corrected method in the first line and the naive method in parentheses)

ρ		Normal			chi-square		
		0.2	0.4	0.6	0.2	0.4	0.6
G_1	CP	0.975	0.983	0.990	0.960	0.966	0.973
		(0.912)	(0.919)	(0.928)	(0.895)	(0.911)	(0.911)
	h^*	0.940	0.957	1.364	1.114	1.373	1.969
G_2	CP	0.977	0.986	0.986	0.960	0.967	0.975
		(0.922)	(0.933)	(0.929)	(0.907)	(0.912)	(0.927)
	h^*	0.742	0.770	1.087	0.914	1.106	1.543
G_3	CP	0.971	0.971	0.988	0.959	0.961	0.971
		(0.928)	(0.917)	(0.932)	(0.912)	(0.908)	(0.911)
	h^*	0.715	0.727	1.028	0.870	1.067	1.491
G_4	CP	0.974	0.976	0.983	0.950	0.965	0.978
		(0.936)	(0.931)	(0.924)	(0.904)	(0.912)	(0.913)
	h^*	0.675	0.712	1.035	0.834	1.049	1.514
G_5	CP	0.979	0.977	0.985	0.959	0.961	0.967
		(0.933)	(0.930)	(0.935)	(0.920)	(0.915)	(0.910)
	h^*	0.772	0.752	1.088	0.885	1.112	1.648

Table 2.12: Coverage probabilities (CP) for nominal 95% confidence regions of the difference between two small area means for $k = 2$ and $\boldsymbol{D}_i$-pattern (a). (the corrected method in the first line and the naive method in parentheses)

are more reliable than the Survey of Family Income and Expenditure since the sample sizes are much larger. However, this survey is conducted only once in every five years. As auxiliary variables, we use the data of the average spendings on 'Education' and ' Cultural-amusement', which is denoted by EDU_i and CUL_i, respectively. Then the regressor in the model (2.2.1) is

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & \mathrm{EDU}_i & 0 & 0 \\ 0 & 0 & 1 & \mathrm{CUL}_i \end{pmatrix}.$$

Then we apply the multivariate Fay-Herriot model (2.2.1), where sampling covariance matrices $\boldsymbol{D}_i$ of the i-th region for $i = 1, \ldots, 10$ are calculated based on data of yearly averaged monthly spendings on 'Education' and ' Cultural-amusement' in the past ten years (2006-2015), where $\boldsymbol{D}_i$ is given as the average of the sampling covariance matrices of prefectures within the i-th region. That is, the sampling covariance matrix $\boldsymbol{D}_i$ are the same for prefectures within the same region.

The estimates of the covariance matrix $\boldsymbol{\Psi}$, the correlation coefficient ρ and the regression coefficients are

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 8.99 & 3.19 \\ 3.19 & 10.84 \end{pmatrix}, \hat{\rho} = 0.323 \quad \text{and} \quad \widehat{\boldsymbol{\beta}} = (4.49, 0.82, 12.16, 0.65)^\top.$$

The values of EBLUP and direct estimate of spendings on 'Education' and ' Cultural-amusement' are reported in Table 2.13. We only pick up the three prefectures from three different regions: Tokyo prefecture from the Kanto region, Osaka prefecture from the Kinki region and Fukushima prefecture from the Tohoku region, whose sampling covariance matrices are

$$\begin{pmatrix} 1.1 & 0.3 \\ 0.3 & 3.0 \end{pmatrix}, \begin{pmatrix} 1.1 & -0.2 \\ -0.2 & 3.9 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 4.7 & 3.5 \\ 3.5 & 4.9 \end{pmatrix},$$

respectively. It is seen that as the sampling variances become larger, the direct estimates are more shrunken by the EBLUP in the sense of (direct estimate - EBLUP)/(direct estimate).

Table 2.13: EBLUP and direct estimates

	Tokyo	Osaka	Fukushima
direct estimator (EDU)	32.5	19.0	13.3
EBLUP (EDU)	31.8	19.4	12.6
direct estimator (CUL)	41.8	24.9	29.9
EBLUP (CUL)	40.6	26.1	29.0

The uncertainty of EBLUP is provided by the second-order unbiased estimator of MSEM of EBLUP. Table 2.14 reports the estimates of MSEM averaged over prefectures within each region for 10 regions. We also calculate the percentage relative improvement in the average loss estimate (PRIAL estimate) of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$. Table 2.15 reports the average of those values over each region for spendings on education and cultural-amusement. It is revealed from Table 2.15 that the multivariate EBLUP improves on the direct estimates significantly and that the multivariate EBLUP is slightly better than the univariate EBLUP for most regions except Okinawa, which has a smaller sampling covariance matrix.

Table 2.14: Estimates of the mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$

Hokkaido	Tohoku	Kanto	Hokuriku	Tokai
$\begin{bmatrix} 0.5 & 0.7 \\ 0.7 & 3.8 \end{bmatrix}$	$\begin{bmatrix} 3.2 & 2.2 \\ 2.2 & 3.4 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 2.5 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 0.6 \\ 0.6 & 4.7 \end{bmatrix}$	$\begin{bmatrix} 1.4 & 0.6 \\ 0.6 & 1.8 \end{bmatrix}$
Kinki	Chugoku	Shikoku	Kyushu	Okinawa
$\begin{bmatrix} 1.0 & -0.0 \\ -0.0 & 2.8 \end{bmatrix}$	$\begin{bmatrix} 1.5 & 0.3 \\ 0.3 & 2.6 \end{bmatrix}$	$\begin{bmatrix} 4.2 & 0.9 \\ 0.9 & 3.5 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 0.7 \\ 0.7 & 1.8 \end{bmatrix}$	$\begin{bmatrix} 3.0 & 0.8 \\ 0.8 & 1.7 \end{bmatrix}$

Table 2.16 reports the reduction rates of areas of confidence regions with 95% nominal confidence coefficient for each area group based on EBLUP, which is calculated by $(S(\boldsymbol{y}_a) - S(\widehat{\boldsymbol{\theta}}_a^{EB}))/S(\boldsymbol{y}_a)$, where $S(\boldsymbol{y}_a)$ and $S(\widehat{\boldsymbol{\theta}}_a^{EB}))$ are areas of confidence regions based on direct estimators and confidence regions given in (2.4.6). Table 2.16 also reports the the size of sampling covariance matrices for each area group, which is defined as $\mathrm{tr}\,(\boldsymbol{D}_j^2)/2$ for $j = 1, \ldots, 10$. It is seen that the reduction rate of area of confidence regions based on EBLUP is larger as the size of sampling covariance matrix is large. This is a reliable result because direct estimators for areas with large sampling variances are more shrunk and shows the usefulness of our proposed method.

Table 2.15: PRIAL estimates of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$ and $\widehat{\boldsymbol{\theta}}_a^{uEB}$

$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\boldsymbol{y}_a$

Hokkaido	Tohoku	Kanto	Hokuriku	Tokai	Kinki	Chugoku	Shikoku	Kyushu	Okinawa
82.4	84.7	80.9	84.1	80.5	81.5	81.6	85.7	80.1	81.0

$\widehat{\boldsymbol{\theta}}_a^{EB}$ vs $\widehat{\boldsymbol{\theta}}_a^{uEB}$

Hokkaido	Tohoku	Kanto	Hokuriku	Tokai	Kinki	Chugoku	Shikoku	Kyushu	Okinawa
4.1	7.1	1.9	5.3	1.1	4.5	2.4	3.9	1.3	-33.9

	Hokkaido	Tohoku	Kanto	Hokuriku	Tokai	Kinki	Chugoku	Shikoku	Kyushu	Okinawa
reduction rate (%)	17.3	24.2	13.4	24.6	11.9	16.6	17.1	34.7	9.8	19.4
sampling variance	15.8	35.7	5.2	29.9	3.9	7.4	6.7	37.0	3.2	11.5

Table 2.16: Reduction rate of areas of confidence regions based on EBLUP and direct estimates

2.8 Proofs

2.8.1 Proof of Lemma 2.3.1

The covariance of $\boldsymbol{P}\boldsymbol{y}$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ is

$$\begin{aligned}
&E[\boldsymbol{P}\boldsymbol{y}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta})^\top]\{\boldsymbol{X}^\top(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\}\\
&\quad=E[(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS})(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top](\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\\
&\quad=\Big[(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})-\boldsymbol{X}\{\boldsymbol{X}^\top(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top\Big](\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}\\
&\quad=\boldsymbol{0}.
\end{aligned}$$

This implies that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ is independent of $\boldsymbol{P}\boldsymbol{y}$. Next, we prove that $\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is a function of $\boldsymbol{P}\boldsymbol{y}$. From (H2), $\widehat{\boldsymbol{\Psi}}$ is a function of $\boldsymbol{P}\boldsymbol{y}$. Rewrite $\widehat{\boldsymbol{\theta}}_a^{EB}$ as $\widehat{\boldsymbol{\theta}}_a^{EB}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}),\boldsymbol{y})$, $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ as $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{y})$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ as $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{y})$. Since $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T})=\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{y})+\boldsymbol{T}$, from (2.2.7) and (2.2.8), we have

$$\begin{aligned}
&\widehat{\boldsymbol{\theta}}_a^{EB}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T}),\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T})-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}(\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T}),\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T})\\
=&\widehat{\boldsymbol{\theta}}_a^{EB}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}),\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T})-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}(\boldsymbol{y}),\boldsymbol{y}+\boldsymbol{X}\boldsymbol{T})\\
=&\boldsymbol{y}_a+\boldsymbol{X}_a\boldsymbol{T}-\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y})+\boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a+\boldsymbol{X}_a\boldsymbol{T}-\boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}))-\boldsymbol{X}_a\boldsymbol{T}\}\\
&-\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{T}+\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a+\boldsymbol{X}_a\boldsymbol{T}-\boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{X}_a\boldsymbol{T}\}\\
=&\boldsymbol{y}_a-\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y})+\boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a-\boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}))\}-\boldsymbol{y}_a+\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{y}_a-\boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}\\
=&\widehat{\boldsymbol{\theta}}_a^{EB}(\widehat{\boldsymbol{\Psi}}(\boldsymbol{y}),\boldsymbol{y})-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}(\boldsymbol{y}),\boldsymbol{y}).
\end{aligned}$$

Thus, $\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is translation invariant, which implies that $\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is a function of $\boldsymbol{P}\boldsymbol{y}$. Hence, $\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ is independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$. □

2.8.2 Proof of Lemma 2.3.3

For the proof of part (2), note that

$$(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}. \tag{2.8.1}$$

Then, $G_{1a}(\widehat{\boldsymbol{\Psi}})$ is rewritten as

$$\begin{aligned}\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) =&(\widehat{\boldsymbol{\Psi}}^{-1}+\boldsymbol{D}_a^{-1})^{-1}=\boldsymbol{D}_a-\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\\ =&\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a \qquad (2.8.2)\\ &-\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+O_p(m^{-3/2}),\end{aligned}$$

which implies that $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})]=\boldsymbol{G}_{1a}(\boldsymbol{\Psi})-\boldsymbol{G}_{3a}(\boldsymbol{\Psi})+O(m^{-3/2})$.

For the proof of part (1), it is noted that

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) =&\boldsymbol{D}_a\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}-(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})+\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ &-\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}\}.\end{aligned}$$

Using the equation in (2.8.1), we can see that

$$\begin{aligned}&\boldsymbol{D}_a\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}-(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})+O_p(m^{-1})\end{aligned}$$

and

$$\begin{aligned}&\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}-\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}+O_p(m^{-1}).\end{aligned}$$

Thus, we have

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) =&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &+\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}+O_p(m^{-1})\\ =&I_1+I_2+O_p(m^{-1}). \quad \text{(say)}\end{aligned}$$

For I_2, it is noted that

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\\ &=\Big[\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big]\\ &\quad\times\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\ &\quad+\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m\boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\Big\}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\ &=I_{21}+I_{22}.\end{aligned}$$

We can evaluate I_{21} as

$$
\begin{aligned}
I_{21} =&\Big\{\sum_{j=1}^{m}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^{m}\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\
=&O_p(m^{-1}),
\end{aligned}
$$

because $\sum_{j=1}^{m}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j = O(m)$, $\sum_{i=1}^{m}\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i = O_p(m^{1/2})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}=O_p(m^{-1/2})$. We next estimate I_{22} as

$$
\begin{aligned}
I_{22} =& -\Big\{\sum_{j=1}^{m}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^{m}\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)\boldsymbol{X}_i\Big\}\\
&\times\Big\{\sum_{i=1}^{m}\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})
\end{aligned}
$$

for $\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)=(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}$. It can be seen that $I_{22}=O_p(m^{-1})$ from the same arguments as in I_{21}. Thus, it follows that $I_2=O_p(m^{-1})$. Hence, we have

$$
\begin{aligned}
&E[\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top]\\
&=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E\Big[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\Big]\\
&\quad\times(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+O(m^{-3/2}).
\end{aligned}
$$

Let $\widehat{\boldsymbol{\Psi}}_{(-a)}$ be an estimator of $\boldsymbol{\Psi}$ from the data except the ath area. If we add or remove the data of one area in the estimation of $\boldsymbol{\Psi}$, there is a negligible change in the value of the above expectation since $\widehat{\boldsymbol{\Psi}}-\widehat{\boldsymbol{\Psi}}_{(-a)}=O_p(m^{-1})$. Thus, we have

$$
\begin{aligned}
&E[\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top]\\
&=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E\Big[(\widehat{\boldsymbol{\Psi}}_{(-a)}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}_{(-a)}-\boldsymbol{\Psi})\Big]\\
&\quad\times(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+O(m^{-3/2})\\
&=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E\Big[(\widehat{\boldsymbol{\Psi}}_{(-a)}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}_{(-a)}-\boldsymbol{\Psi})\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+O(m^{-3/2})\\
&=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E\Big[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+O(m^{-3/2}),
\end{aligned}
$$

which is equal to $\boldsymbol{G}_{1a}(\boldsymbol{\Psi})+O(m^{-3/2})$, where the second equation follows from the independence of the data of different areas, and the the third equation follows form the same reason mentioned above. □

2.8.3 Proof of Theorem 2.4.1

Let $\boldsymbol{z}_a=\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a-\widehat{\boldsymbol{\theta}}_a^{EB}+\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))$. From Lemma 2.3.2, the conditional distribution of $\boldsymbol{z}_a$ given $\boldsymbol{P}\boldsymbol{y}$ is $\boldsymbol{z}_a\sim\mathcal{N}_k(\boldsymbol{0},\boldsymbol{I}_k)$, and the mahalanobis distance is approximated as

$$
(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a)^\top\boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB}-\boldsymbol{\theta}_a)
$$

$$
\begin{aligned}
=&\boldsymbol{z}_a^\top \boldsymbol{H}_a^{1/2}(\boldsymbol{\Psi})\boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{1/2}(\boldsymbol{\Psi})\boldsymbol{z}_a + 2(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{1/2}(\boldsymbol{\Psi})\boldsymbol{z}_a\\
&+ (\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))\\
=&\boldsymbol{z}_a^\top \Big[\boldsymbol{I}_k - \boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{H}_a(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{H}_a(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\\
&\qquad + \boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\Big]\boldsymbol{z}_a\\
&+ 2(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{1/2}(\boldsymbol{\Psi})\boldsymbol{z}_a\\
&+ (\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})) + o(m^{-1})\\
=&\boldsymbol{z}_a^\top(\boldsymbol{I}_k - \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))\boldsymbol{z}_a + 2\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}}) + o(m^{-1}),
\end{aligned}
\tag{2.8.3}
$$

where

$$
\begin{aligned}
\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}) =&\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{H}_a(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{H}_a(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\\
&- \boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi}),\\
\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top =&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{1/2}(\boldsymbol{\Psi}),\\
g_{3a}(\widehat{\boldsymbol{\Psi}}) =&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})).
\end{aligned}
$$

From (2.8.3), the characteristic function $\varphi(t) = E[\exp\{it(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)\}]$ is approximated as

$$
\begin{aligned}
\varphi(t) =&E\exp\Big(it\{\boldsymbol{z}_a^\top(\boldsymbol{I}_k - \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))\boldsymbol{z}_a + 2\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}})\}\Big) + o(m^{-1})\\
=&E\Big[e^{it\boldsymbol{z}_a^\top \boldsymbol{z}_a}\Big\{1 + it\{-\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a + 2\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}})\}\\
&- \frac{t^2}{2}\{-\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a + 2\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}})\}^2\Big\}\Big] + o(m^{-1})\\
=&E\Big[e^{it\boldsymbol{z}_a^\top \boldsymbol{z}_a}\Big\{1 + it\{-\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a + 2\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}})\}\\
&- \frac{t^2}{2}\{(\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a)^2 + 4\boldsymbol{z}_a^\top \boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a - 4\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a\}\Big\}\Big] + o(m^{-1}),
\end{aligned}
$$

because $\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}) = O_p(m^{-1/2})$, $\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}}) = O_p(m^{-1/2})$ and $g_{3a}(\widehat{\boldsymbol{\Psi}}) = O_p(m^{-1})$. From the law of iterated expectations and the conditional normality of $\boldsymbol{z}_a$, the above equation reduces to

$$
\begin{aligned}
\varphi(t) =&E\Big[e^{it\boldsymbol{z}_a^\top \boldsymbol{z}_a}\Big\{1 + it\{-\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a + g_{3a}(\widehat{\boldsymbol{\Psi}})\}\\
&\qquad - \frac{t^2}{2}\{(\boldsymbol{z}_a^\top \boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{z}_a)^2 + 4\boldsymbol{z}_a^\top \boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{z}_a\}\Big\}\Big] + o(m^{-1}).
\end{aligned}
$$

For some deterministic matrix $\boldsymbol{A}$ and $\boldsymbol{z} \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{I}_k)$, it holds that

$$
\begin{aligned}
E\Big[e^{it\boldsymbol{z}^\top \boldsymbol{z}}\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z}\Big] =&(2\pi)^{-k/2}\int e^{-\frac{(1-2it)\boldsymbol{z}^\top \boldsymbol{z}}{2}}\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z}d\boldsymbol{z} = (1-2it)^{-k/2-1}\mathrm{tr}\,(\boldsymbol{A}),\\
E\Big[e^{it\boldsymbol{z}^\top \boldsymbol{z}}(\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z})^2\Big] =&(2\pi)^{-k/2}\int e^{-\frac{(1-2it)\boldsymbol{z}^\top \boldsymbol{z}}{2}}(\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z})^2 d\boldsymbol{z} = (1-2it)^{-k/2-2}(\mathrm{tr}^{\,2}(\boldsymbol{A}) + 2\mathrm{tr}\,(\boldsymbol{A}^2)).
\end{aligned}
$$

Using these equalities, from the law of iterated expectations, we have

$$\begin{aligned}\varphi(t) =&(1-2it)^{-k/2}\Big[1+it\Big\{-(1-2it)^{-1}\mathrm{tr}\,(E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})])+E[g_{3a}(\widehat{\boldsymbol{\Psi}})]\Big\}\\&+\frac{(it)^2}{2}\Big\{(1-2it)^{-2}\{E[\mathrm{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))]+2\mathrm{tr}\,(E[\boldsymbol{G}_{12a}^2(\widehat{\boldsymbol{\Psi}})])\}\\&+(1-2it)^{-1}4\mathrm{tr}\,(E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top])\Big\}\Big]+o(m^{-1}).\end{aligned}$$

For notational simplicity, let $C=E[\mathrm{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))]+2\mathrm{tr}\,(E[\boldsymbol{G}_{12a}^2(\widehat{\boldsymbol{\Psi}})])$. Let $s=(1-2it)^{-1}$, or $it=(s-1)/2s$. Then, $(1-2it)^{-k/2}\varphi(t)-1$ can be written as

$$\begin{aligned}&it\Big\{-(1-2it)^{-1}\mathrm{tr}\,(E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})])+E[g_{3a}(\widehat{\boldsymbol{\Psi}})]\Big\}\\&+\frac{(it)^2}{2}\Big\{(1-2it)^{-2}C+(1-2it)^{-1}4E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})]\Big\}\\=&\frac{1}{2s}\Big\{E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})]-E[g_{3a}(\widehat{\boldsymbol{\Psi}})]\Big\}\\&+\Big\{\frac{1}{2}\mathrm{tr}\,(E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})])+\frac{1}{2}E[g_{3a}(\widehat{\boldsymbol{\Psi}})]+\frac{C}{8}-E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})]\Big\}\\&+\Big\{-\frac{1}{2}\mathrm{tr}\,(E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})])-\frac{C}{4}+\frac{1}{2}E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})]\Big\}s+\frac{C}{8}s^2+o(m^{-1}).\end{aligned}\tag{2.8.4}$$

which is a second-order polynomial of s.

We shall evaluate the moments in (2.8.4). First, $\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})$ can be expanded as

$$\begin{aligned}\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})=&\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{H}_a(\widehat{\boldsymbol{\Psi}})-\boldsymbol{H}_a(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\\&-\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})+o_p(m^{-1}).\end{aligned}\tag{2.8.5}$$

From Lemma 2.3.3, the expectation of the first term in (2.8.5) is $-\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi})\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})+o(m^{-1})$, so that

$$E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})]=-\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi})\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})-E[\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})]+o(m^{-1}),$$

for $\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})=\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}$. Thus,

$$\mathrm{tr}\,(E[\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}})])=-B_3+2B_1+o(m^{-1}),\tag{2.8.6}$$

for B_1 and B_3 defined in (2.4.2). Noting that the first term in (2.8.5) is of order $O(m^{-1/2})$ and the second term is of order $O(m^{-1})$, we can expand $\boldsymbol{G}_{12a}^2(\widehat{\boldsymbol{\Psi}})$ and $\mathrm{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))$ as

$$\begin{aligned}\boldsymbol{G}_{12a}^2(\widehat{\boldsymbol{\Psi}})=&\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})+o_p(m^{-1}),\\\mathrm{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))=&\mathrm{tr}^2(\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1/2}(\boldsymbol{\Psi}))+o_p(m^{-1}),\end{aligned}$$

which lead to $E[\boldsymbol{G}_{12a}^2(\widehat{\boldsymbol{\Psi}})]=E[\{\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})\}^2]+o(m^{-1})$ and $E[\mathrm{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}))]=E[\mathrm{tr}^2(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))]+o(m^{-1})$. Thus,

$$C=E[\mathrm{tr}^2(\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}}))]+2E[\{\boldsymbol{K}_a(\widehat{\boldsymbol{\Psi}})\}^2]+o(m^{-1})=-8B_2+o(m^{-1}),\tag{2.8.7}$$

for B_2 defined in (2.4.2). It can be also observed that

$$\begin{aligned}
\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}}) =&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})\boldsymbol{H}_a(\boldsymbol{\Psi})\boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))\\
=&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})) + o_p(m^{-1}),\\
g_{3a}(\widehat{\boldsymbol{\Psi}}) =&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))\\
=&(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}))^\top \boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})) + o_p(m^{-1}),
\end{aligned}$$

both of which lead to

$$E[\boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})^\top \boldsymbol{g}_{2a}(\widehat{\boldsymbol{\Psi}})] = E[g_{3a}(\widehat{\boldsymbol{\Psi}})] = \mathrm{tr}\,(\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{G}_{3a}(\boldsymbol{\Psi})) + o(m^{-1}) = B_3 + o(m^{-1}). \tag{2.8.8}$$

Combining (2.8.6), (2.8.7) and (2.8.8), we can see that the constant term and the coefficient of s^2 in (2.8.4) are $B_1 - B_3 - B_2$ and $-B_2$ given in (2.4.2), respectively. Thus, the characteristic function of $(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)$ can be written as

$$\varphi(t) = (1-2it)^{-k/2}(1 + B_1 - B_3 - B_2 + (-B_1 + B_3 + 2B_2)s - B_2 s^2) + o(m^{-1}).$$

From the fact that the characteristic function of the chi-squared distribution with degrees of freedom $k+2h$ is given by $(1-2it)^{-k/2-h} = (1-2it)^{-k/2}s^h$, it follows that the asymptotic expansion of the cumulative distribution function of $(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)$ is

$$F_k(x) + (B_1 - B_3 - B_2)F_k(x) + (-B_1 + B_3 + 2B_2)F_{k+2}(x) - B_2 F_{k+4}(x) + o(m^{-1}),$$

where $F_k(x)$ is the cumulative distribution function of the chi-squared distribution with degrees of freedom k. Note that $F_{k+r-2}(x) - F_{k+r}(x) = 2f_{k+r}(x)$, where $f_k(x)$ is the density function of the chi-squared distribution with degrees of freedom k. Then, it is expressed as

$$\begin{aligned}
&P((\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)^\top \boldsymbol{H}_a^{-1}(\widehat{\boldsymbol{\Psi}})(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a) \le x)\\
&\quad =F_k(x) + 2(B_1 - B_3 - B_2)f_{k+2}(x) + 2B_2 f_{k+4}(x) + o(m^{-1}),
\end{aligned}$$

which proves Theorem 2.4.1. □

2.8.4 Proof of Lemma 2.5.1

From Proposition 2.5.1, it is sufficient to show this approximation for $\widehat{\boldsymbol{\Psi}}^{PR}$ instead of $\widehat{\boldsymbol{\Psi}}_{(A)}^{PR}$. It is noted that $\widehat{\boldsymbol{\Psi}}^{PR} - \boldsymbol{\Psi}$ is approximated as

$$\widehat{\boldsymbol{\Psi}}^{PR} - \boldsymbol{\Psi} = \frac{1}{m}\sum_{i=1}^m \{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\} + O_p(m^{-1}), \tag{2.8.9}$$

which is used to evaluate

$$
\begin{aligned}
&E\Big[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\Big]\\
=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m E\Big[\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{u}_a\boldsymbol{u}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{u}_j\boldsymbol{u}_j^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}\Big]\\
&+O(m^{-3/2})\\
=&\frac{1}{m^2}\sum_{i=1}^m E\Big[\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{u}_a\boldsymbol{u}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}\Big]\\
&+O(m^{-3/2}),
\end{aligned}
$$

because $E\Big[\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{u}_a\boldsymbol{u}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{u}_j\boldsymbol{u}_j^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}\Big]=\boldsymbol{0}$ for $i\neq j$. Letting $\boldsymbol{z}_i=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1/2}\boldsymbol{u}_i$, we can see that $\boldsymbol{z}_i\sim\mathcal{N}_k(\boldsymbol{0},\boldsymbol{I}_k)$. Then,

$$
\begin{aligned}
&\frac{1}{m^2}\sum_{i=1}^m E\Big[\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{u}_a\boldsymbol{u}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\{\boldsymbol{u}_i\boldsymbol{u}_i^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}\Big]\\
=&\frac{1}{m^2}\sum_{i\neq a}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}E\Big[(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\Big](\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}+O(m^{-2}),
\end{aligned}
$$

for $\boldsymbol{B}=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1/2}$. Let $\boldsymbol{C}=\boldsymbol{B}\boldsymbol{B}^\top=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}$. For $i\neq a$,

$$
\begin{aligned}
&E[(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})]\\
=&E[\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top\boldsymbol{z}_i\boldsymbol{z}_i^\top+\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top-\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top-\boldsymbol{B}\boldsymbol{z}_a\boldsymbol{z}_a^\top\boldsymbol{B}^\top\boldsymbol{z}_i\boldsymbol{z}_i^\top]\\
=&E[\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{C}\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{C}]=\boldsymbol{C}+(\mathrm{tr}\,\boldsymbol{C})\boldsymbol{I}_k,
\end{aligned}
$$

because $E[\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{C}\boldsymbol{z}_i\boldsymbol{z}_i^\top]=2\boldsymbol{C}+(\mathrm{tr}\,\boldsymbol{C})\boldsymbol{I}_k$. Thus,

$$
\begin{aligned}
&\frac{1}{m^2}\sum_{i\neq a}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\{\boldsymbol{C}+(\mathrm{tr}\,\boldsymbol{C})\boldsymbol{I}_k\}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\\
=&\frac{1}{m^2}\sum_{i=1}^m(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\{\boldsymbol{C}+(\mathrm{tr}\,\boldsymbol{C})\boldsymbol{I}_k\}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}+O(m^{-2}),
\end{aligned}
$$

which leads to the expression in (2.5.7) from Lemma 2.3.3 (1). □

2.8.5 Proof of Lemma 2.5.2

From Proposition 2.5.1, it is sufficient to show this approximation for $\widehat{\boldsymbol{\Psi}}^{PR}$ instead of $\widehat{\boldsymbol{\Psi}}^{PR}_{(A)}$. For some deteministic matrix $\boldsymbol{A}$ and multivariate standard normal variables $\boldsymbol{z}_i$, $i=1,\ldots,k$,

$E[(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})\boldsymbol{A}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})]$ is, from (2.8.9), approximated as

$$
\begin{aligned}
&E[(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})\boldsymbol{A}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})]\\
=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m E\Big[\Big\{(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}\boldsymbol{A}\\
&\times\Big\{(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_j)\Big\}\Big]+O(m^{-3/2})\\
=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}E[(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\boldsymbol{C}_i(\boldsymbol{z}_j\boldsymbol{z}_j^\top-\boldsymbol{I})](\boldsymbol{\Psi}+\boldsymbol{D}_j)^{1/2}+O(m^{-3/2}),
\end{aligned}
$$

for $\boldsymbol{C}_i=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\boldsymbol{A}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}$. For $i\neq j$, $E[(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\boldsymbol{C}_i(\boldsymbol{z}_j\boldsymbol{z}_j^\top-\boldsymbol{I})]=0$, we have

$$
\sum_{i=1}^m\sum_{j=1}^m E[(\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{I})\boldsymbol{C}_i(\boldsymbol{z}_j\boldsymbol{z}_j^\top-\boldsymbol{I})]=\sum_{i=1}^m E[\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{C}_i\boldsymbol{z}_i\boldsymbol{z}_i^\top-\boldsymbol{C}_i].
$$

Because $E[\boldsymbol{z}_i\boldsymbol{z}_i^\top\boldsymbol{C}_i\boldsymbol{z}_i\boldsymbol{z}_i^\top]=2\boldsymbol{C}_i+(\operatorname{tr}\boldsymbol{C}_i)\boldsymbol{I}_k$, it is concluded that

$$
E[(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})\boldsymbol{A}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})]=\frac{1}{m^2}\sum_{i=1}^m((\boldsymbol{\Psi}+\boldsymbol{D}_i)\boldsymbol{A}(\boldsymbol{\Psi}+\boldsymbol{D}_i)+\operatorname{tr}(\boldsymbol{A}(\boldsymbol{\Psi}+\boldsymbol{D}_i))(\boldsymbol{\Psi}+\boldsymbol{D}_i)).
$$

Using this equality, we have

$$
\begin{aligned}
&E[(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})]\\
=&\frac{1}{m^2}\sum_{i=1}^m\Big\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\
&+\operatorname{tr}((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-2}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}+o(m^{-1}),
\end{aligned}
$$

and

$$
\begin{aligned}
&E[(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^{PR})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^{PR})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi}))]\\
=&\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\Big[\frac{1}{m^2}\sum_{i=1}^m\Big\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\
&+\operatorname{tr}((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))(\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+o(m^{-1}),
\end{aligned}
$$

which leads to the first and third expressions in the lemma. From (2.8.2),

$$
\begin{aligned}
E[\operatorname{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}^{PR}))]=&E[\operatorname{tr}^2(\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})(\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^{PR})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi})))]+o(m^{-1})\\
=&E[\operatorname{tr}^2((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi}))]+o(m^{-1}).
\end{aligned}
$$

Letting $\boldsymbol{u}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$, we can see that

$$
\begin{aligned}
&\operatorname{tr}\left((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}^{PR}-\boldsymbol{\Psi})\right)\\
=&\frac{1}{m}\sum_{i=1}^{m}\operatorname{tr}\left((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}(\boldsymbol{u}_i\boldsymbol{u}_i^\top-\boldsymbol{I}_k)(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\right)\\
&+o(m^{-1})\\
=&\frac{1}{m}\sum_{i=1}^{m}\operatorname{tr}\left((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{1/2}\right)\\
&-\frac{1}{m}\sum_{i=1}^{m}\operatorname{tr}\left((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)\right)+o(m^{-1}).
\end{aligned}
$$

Thus we have

$$
E[\operatorname{tr}^2(\boldsymbol{G}_{12a}(\widehat{\boldsymbol{\Psi}}^{PR}))]=\frac{2}{m^2}\sum_{i=1}^{m}\operatorname{tr}\Big(((\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\boldsymbol{H}_a^{-1}(\boldsymbol{\Psi})\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i))^2\Big)+o(m^{-1}),
$$

which leads to the expression in the lemma. □

Chapter 3

Robust Estimation of Mean Squared Error Matrix of Small Area Estimators in a Multivariate Fay-Herriot Model

3.1 Motivation

In this chapter, we treat the multivariate Fay-Herriot model without assuming multivariate normal distributions. We suggest a consistent and nonnegative-definite estimator of the covariance matrix of the random effects, and provide the EBLUP for a vector of small-area characteristics. We derive a second-order approximation of the mean squared error matrix (MSEM) of the EBLUP and obtain the second-order unbiased estimator of the MSEM. This second-order unbiased estimator of the MSEM is robust, because it does not depend on any distributions of the random effects. This is a multivariate extension of the result given in Lahiri and Rao (1995). The difference between the approach discussed here and Lahiri and Rao (1995) is clarified as follows: (1) Lahiri and Rao (1995) assumes that the error terms have univariate normal distributions with known variances, while no distributional assumptions are imposed on the random effects. The present approach does not assume the normality for the error terms, but assumes that the second and fourth moments are known. (2) The present approach handles the multivariate Fay-Herriot model where the covariance matrix of random effects is fully unknown without nomality assumption, while Lahiri and Rao (1995) treated the univariate Fay-Herriot model with unknown variance of random effects.

In Section 3.2, as a specific estimator of the covariance matrix, we employ a Prasad-Rao type estimator with a closed form and use the modified version which is restricted over the space of nonnegative definite matrices. The consistency is also shown. The EBLUP is provided based on the Prasad-Rao type estimator.

In Sections 3.3, we derive the second-order approximation of the mean squared error matrix and the second-order unbiased estimator of the MSEM with a closed form. Similarly to Lahiri and Rao (1995), this MSEM estimator is achieved only under the moment assumptions for random effects, and then, our estimator of MSEM is useful because the normality assumption is very resrtrictive and the specification of the underlying distributions for random effects and sampling errors are difficulut in practice. However, in the multivariate problem, when deriving a second-order approximation of the MSEM of EBLUP and a second-order unbiased estimator of the MSEM, we cannot use the standard technique of approximation via the Taylor's expansion as Lahiri and Rao (1995). Then, the results for the multivariate version are not obvious and we must consider them separately from the univariate problem.

The performances of EBLUP and the MSEM estimator are investigated in Section 3.4 through simulation and empirical studies. The proofs of the main theorems are given in the section 3.5, where the details of tedious calculations are given in the Supplemental Matrial.

3.2 Empirical Best Linear Unbiased Prediction

We assume that area-level data $(\boldsymbol{y}_1, \boldsymbol{X}_1), \ldots, (\boldsymbol{y}_m, \boldsymbol{X}_m)$ are observed, where m is the number of small areas, $\boldsymbol{y}_i$ is a k-variate vector of direct survey estimates and $\boldsymbol{X}_i$ is a $k \times s$ matrix of covariates associated with $\boldsymbol{y}_i$ for the i-th area. Then, we consider the multivariate Fay-Herriot model

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, m, \tag{3.2.1}$$

where $\boldsymbol{\beta}$ is an s-variate vector of unknown regression coefficients, $\boldsymbol{v}_i$ is a k-variate vector of random effects depending on the i-th area, $\boldsymbol{\varepsilon}_i$ is a k-variate vector of sampling errors, and $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$ are mutually independent for $i = 1, \ldots, m$. The distributional assumption of the standard extension of the Fay-Herriot model is that $\boldsymbol{v}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{D}_i)$ for unknown covariance matrix $\boldsymbol{\Psi}$ and known error covariance matrices $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m$. Instead of this assumption, we here assume some moment conditions for

$$\boldsymbol{v}_i = \boldsymbol{\Psi}^{1/2}\boldsymbol{u}_i \quad \text{and} \quad \boldsymbol{\varepsilon}_i = \boldsymbol{D}_i^{1/2}\boldsymbol{e}_i, \tag{3.2.2}$$

namely, $E\boldsymbol{u}_i = E\boldsymbol{e}_i = \boldsymbol{0}$, $\mathrm{Var}(\boldsymbol{u}_{\mathrm{i}}) = \mathrm{Var}(\boldsymbol{e}_{\mathrm{i}}) = \boldsymbol{I}_{\mathrm{k}}$, $Eu_{ij}^4 = \kappa_v$ and $Ee_{ij}^4 = \kappa_\varepsilon$ for $i = 1, \ldots, m$ and $j = 1, \ldots, k$, where $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ik})^\top$ and $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{ik})^\top$. Following the setup of the Fay-Herriot model, we assume that $\boldsymbol{\Psi}$ and κ_v are unknown, but κ_ε and $\boldsymbol{D}_i$'s are known.

Let $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_m^\top)^\top$, $\boldsymbol{v} = (\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_m^\top)^\top$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_m^\top)^\top$. Then, model (3.2.1) is expressed in a matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{\varepsilon}, \tag{3.2.3}$$

where $E[\boldsymbol{v}] = \boldsymbol{0}$, $\mathbf{Cov}\,(\boldsymbol{v}) = \boldsymbol{I}_m \otimes \boldsymbol{\Psi}$, $E[\boldsymbol{\varepsilon}] = \boldsymbol{0}$ and $\mathbf{Cov}\,(\boldsymbol{\varepsilon}) = \boldsymbol{D}$ for $\boldsymbol{D} = \text{block diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m)$.

For the a-th area, we want to predict the quantity $\boldsymbol{\theta}_a = \boldsymbol{X}_a\boldsymbol{\beta} + \boldsymbol{v}_a$. Then, the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta}_a$ can be provided by

$$\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) = \boldsymbol{y}_a - \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\big\{\boldsymbol{y}_a - \boldsymbol{X}_a\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\big\}, \tag{3.2.4}$$

where $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})$ is the generalized least squares estimator

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) =& \{\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{y} \\ =& \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{y}_i.\end{aligned} \tag{3.2.5}$$

Since $\boldsymbol{\Psi}$ is unknown, we need to estimate $\boldsymbol{\Psi}$ consistently. It is noted that $E[(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top] = \boldsymbol{\Psi} + \boldsymbol{D}_i$ for $i = 1, \ldots, m$, which implies that $\sum_{i=1}^m E[(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top] = m\boldsymbol{\Psi} + \sum_{i=1}^m \boldsymbol{D}_i$. Substituting the ordinary least squares estimator $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ into $\boldsymbol{\beta}$, we get the consistent estimator

$$\widehat{\boldsymbol{\Psi}}_0 = \frac{1}{m}\sum_{i=1}^m \left\{(\boldsymbol{y}_i - \boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})(\boldsymbol{y}_i - \boldsymbol{X}_i\widetilde{\boldsymbol{\beta}})^\top - \boldsymbol{D}_i\right\}. \tag{3.2.6}$$

Taking the expectation of $\widehat{\boldsymbol{\Psi}}_0$, we can see that $E[\widehat{\boldsymbol{\Psi}}_0] = \boldsymbol{\Psi} + \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi})$, where

$$\begin{aligned}\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}) =& \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi} + \boldsymbol{D}_j)\boldsymbol{X}_j\Big\}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top \\ &- \frac{1}{m}\sum_{i=1}^m(\boldsymbol{\Psi} + \boldsymbol{D}_i)\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top - \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i).\end{aligned} \tag{3.2.7}$$

Substituting $\widehat{\boldsymbol{\Psi}}_0$ into $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi})$, we get a bias-corrected given by

$$\widehat{\boldsymbol{\Psi}}_1 = \widehat{\boldsymbol{\Psi}}_0 - \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\widehat{\boldsymbol{\Psi}}_0). \tag{3.2.8}$$

For notational convenience, we use the same notation $\widehat{\boldsymbol{\Psi}}$ for $\widehat{\boldsymbol{\Psi}}_0$ and $\widehat{\boldsymbol{\Psi}}_1$ without any confusion. It is noted that both estimators are not necessarily nonnegative definite. In this case, there exist a $k \times k$ orthogonal matrix $\boldsymbol{H}$ and a diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_k)$ such that $\widehat{\boldsymbol{\Psi}} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^\top$. Let $\boldsymbol{\Lambda}^+ = \mathrm{diag}\,(\max\{0, \lambda_1\}, \ldots, \max\{0, \lambda_k\})$, and let

$$\widehat{\boldsymbol{\Psi}}^+ = \boldsymbol{H}\boldsymbol{\Lambda}^+\boldsymbol{H}^\top.$$

Replacing $\boldsymbol{\Psi}$ in $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})$ with the estimator $\widehat{\boldsymbol{\Psi}}^+$, we get the empirical best linear unbiased predictor (EBLUP)

$$\widehat{\boldsymbol{\theta}}_a^{EB} = \widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}^+). \tag{3.2.9}$$

To guarantee asymptotic properties of $\widehat{\boldsymbol{\Psi}}^+$, we assume the following conditions:

(H1) $0 < k < \infty$, $0 < s < \infty$.

(H2) There exist positive constants $\underline{d}$ and $\overline{d}$ such that $\underline{d}$ and $\overline{d}$ do not depend on m and satify $\underline{d}\boldsymbol{I}_k \le \boldsymbol{D}_i \le \overline{d}\boldsymbol{I}_k$ for $i = 1, \ldots, m$.

(H3) $\boldsymbol{X}^\top\boldsymbol{X}$ is nonsingular and $\boldsymbol{X}^\top\boldsymbol{X}/m$ converges to a positive definite matrix.

Then, we obtain the next theorem, which is proved in the section 3.5.

Theorem 3.2.1 *Under conditions* (H1)-(H3)*, the following properties hold for* $\widehat{\boldsymbol{\Psi}} = \widehat{\boldsymbol{\Psi}}_0$ *and* $\widehat{\boldsymbol{\Psi}}_1$*:*

(1) $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}) = O(m^{-1})$*, which means that* $\widehat{\boldsymbol{\Psi}}_0$ *has the second-order bias, while* $\widehat{\boldsymbol{\Psi}}_1$ *is a second-order unbiased estimator of* $\boldsymbol{\Psi}$*.*

(2) $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = O_p(m^{-1/2})$ *and* $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta} = O_p(m^{-1/2})$*.*

(3) *The nonnegative defnite matrix* $\widehat{\boldsymbol{\Psi}}^+$ *is consistent for large* m*, and* $P(\widehat{\boldsymbol{\Psi}}^+ \neq \widehat{\boldsymbol{\Psi}}) = O(m^{-K})$ *for any* K*, provided* $4K$*-th moments of* $\boldsymbol{v}_i$*'s and* $\boldsymbol{\varepsilon}_i$*'s exist.*

3.3 Evaluation of the Mean Squared Error Matrix of EBLUP

Uncertainty of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB}$ in (3.2.9) is measured by the mean squared error matrix (MSEM), defined as $\text{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a\}^\top]$. We begin by deriving the second-order approximation of $\text{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ under the assumptions (H1)-(H3) and the following assumption:

(H4) The eighth moment of $\boldsymbol{u}_i$ and $\boldsymbol{e}_i$ given in (3.2.2) exist, namely $E(\boldsymbol{u}_i\boldsymbol{u}_i^\top)^4 = O(1)$ and $E(\boldsymbol{e}_i\boldsymbol{e}_i^\top)^4 = O(1)$ for $i = 1, \ldots, m$.

Let $\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) = \boldsymbol{y}_a - \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})$. The difference $\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a$ is written as

$$\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a = \{\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) - \boldsymbol{\theta}_a\} + \{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi})\} + \{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}.$$

Thus, the mean squared error matrix is decomposed as

$$\begin{aligned}
&\text{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) \\
&= E[\{\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}\{\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}^\top] + E[\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi})\}^\top] \\
&\quad + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top] \qquad (3.3.1)\\
&\quad + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}^\top] + E[\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top].
\end{aligned}$$

Because $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) = -D_a(\boldsymbol{\Psi} + D_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}$, the first two terms in (3.3.1) can be easily evaluated as $E[\{\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}\{\boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}^\top] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi})$ and $E[\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*(\boldsymbol{\beta}, \boldsymbol{\Psi})\}^\top] = \boldsymbol{G}_{2a}(\boldsymbol{\Psi})$, where

$$\begin{aligned}
\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) &= (\boldsymbol{\Psi}^{-1} + \boldsymbol{D}_a^{-1})^{-1} = \boldsymbol{\Psi}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a, \\
\boldsymbol{G}_{2a}(\boldsymbol{\Psi}) &= \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\boldsymbol{X}^\top(\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1}\boldsymbol{X}\}^{-1}\boldsymbol{X}_a^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a.
\end{aligned} \qquad (3.3.2)$$

For the third term and the last two terms in (3.3.1), define $\boldsymbol{G}_{3a}(\boldsymbol{\Psi})$ and $\widetilde{\boldsymbol{G}}_{3a}(\boldsymbol{\Psi})$ as

$$\begin{aligned}
&\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) \\
&= \frac{1}{m^2}\boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\Big[\sum_{i=1}^m(\boldsymbol{\Psi} + \boldsymbol{D}_i)(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi} + \boldsymbol{D}_i) \\
&\quad + \sum_{i=1}^m\{\text{tr}\,[(\boldsymbol{\Psi} + \boldsymbol{D}_i)(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}]\}(\boldsymbol{\Psi} + \boldsymbol{D}_i) + m(\kappa_v - 3)\boldsymbol{\Psi}^{1/2}\text{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_{\text{a}})^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2} \\
&\quad + \sum_{i=1}^m(\kappa_\varepsilon - 3)\boldsymbol{D}_i^{1/2}\text{diag}(\boldsymbol{D}_{\text{i}}^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_{\text{a}})^{-1}\boldsymbol{D}_{\text{i}}^{1/2})\boldsymbol{D}_{\text{i}}^{1/2}\Big](\boldsymbol{\Psi} + \boldsymbol{D}_{\text{a}})^{-1}\boldsymbol{D}_{\text{a}}, \qquad (3.3.3)
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\boldsymbol{G}}_{4a}(\boldsymbol{\Psi}) = &\frac{1}{m}\boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\Big\{ - (\kappa_v - 3)\boldsymbol{\Psi}^{1/2}\text{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a \\
&+ (\kappa_\varepsilon - 3)\boldsymbol{D}_a^{1/2}\text{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\Big\}. \qquad (3.3.4)
\end{aligned}$$

Then, we can prove the following lemmas, where the proofs are given in the section 3.5.

Lemma 3.3.1 *Under the assumptions* (H1)-(H4)*, it holds that* $E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top] = \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O_p(m^{-3/2})$.

Lemma 3.3.2 *Under the assumptions* (H1)-(H4)*, it holds that* $E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}^\top] = \widetilde{\boldsymbol{G}}_{4a}(\boldsymbol{\Psi}) + O_p(m^{-3/2})$.

Let $\boldsymbol{G}_{4a}(\boldsymbol{\Psi}) = \widetilde{\boldsymbol{G}}_{4a}(\boldsymbol{\Psi}) + \{\widetilde{\boldsymbol{G}}_{4a}(\boldsymbol{\Psi})\}^\top$. Combining (3.3.1), (3.3.2) and Lemmas 3.3.1 and 3.3.2, one gets the following theorem.

Theorem 3.3.1 *Assume* (H1)-(H4)*. Then, the mean squared error matrix of the EBLUP* $\widehat{\boldsymbol{\theta}}_a^{EB}$ *is approximated as*

$$\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{4a}(\boldsymbol{\Psi}) + O(m^{-3/2}). \tag{3.3.5}$$

When $\boldsymbol{v}_i$'s and $\boldsymbol{\varepsilon}_i$'s are normally distributed, the second-order approximation of $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ can be simplified as

$$\begin{aligned}\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + \frac{1}{m^2}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big\{\sum_{i=1}^m \Big[(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1} \\ + \mathrm{tr}\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\}\boldsymbol{I}_k\Big](\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a,\end{aligned}$$

because $\kappa_v = \kappa_\varepsilon = 3$.

We next obtain a second-order unbiased estimator of the mean squared error matrix of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB}$ in (3.2.9). A naive estimator of $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is the plug-in estimator of (3.3.5) given by $\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{4a}(\widehat{\boldsymbol{\Psi}}^+)$, but this has a second-order bias, because $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^+)] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + O(m^{-1})$. Thus, we need to correct the second-order bias. Let

$$\boldsymbol{G}_{5a}(\boldsymbol{\Psi}) = -\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}}(\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a, \tag{3.3.6}$$

where $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}}(\boldsymbol{\Psi})$ is the bias of $\widehat{\boldsymbol{\Psi}}$ given by

$$\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}}(\boldsymbol{\Psi}) = \begin{cases} \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}) & \text{for } \widehat{\boldsymbol{\Psi}} = \widehat{\boldsymbol{\Psi}}_0, \\ \boldsymbol{0} & \text{for } \widehat{\boldsymbol{\Psi}} = \widehat{\boldsymbol{\Psi}}_1, \end{cases}$$

for $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi})$ given in (3.2.7), because $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}) = O(m^{-1})$ and $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_1}(\boldsymbol{\Psi}) = O(m^{-2})$.

In the following theorem, which will be proved in the section 3.5, we obtain the second-order unbiased estimator given by

$$\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}}^+) + 2\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{4a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{5a}(\widehat{\boldsymbol{\Psi}}^+). \tag{3.3.7}$$

Theorem 3.3.2 *Under the assumptions* (H1)-(H4)*, it holds that* $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{5a}(\widehat{\boldsymbol{\Psi}}^+)] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + O(m^{-3/2})$*, and*

$$E[\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})] = \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) + O(m^{-3/2}),$$

namely, $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ *is a second-order unbiased estimator of* $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$.

Although $\boldsymbol{G}_{3a}(\boldsymbol{\Psi})$ and $\boldsymbol{G}_{4a}(\boldsymbol{\Psi})$ depend on the unknown kurtosis κ_v of the distirbution of the random effects, it can be shown that $2\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{4a}(\boldsymbol{\Psi})$ is independent of κ_v. In fact,

$$
\begin{aligned}
&2\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{4a}(\boldsymbol{\Psi}) \\
&\quad= \frac{2}{m^2}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big[\sum_{i=1}^m (\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i) \\
&\qquad + \sum_{i=1}^m \{\mathrm{tr}\,[(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}]\}(\boldsymbol{\Psi}+\boldsymbol{D}_i) \\
&\qquad + \sum_{i=1}^m (\kappa_\varepsilon - 3)\boldsymbol{D}_i^{1/2}\mathrm{diag}(\boldsymbol{D}_{\mathrm{i}}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}^{1/2})\boldsymbol{D}_{\mathrm{i}}^{1/2}\Big](\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{a}} \\
&\qquad + \frac{\kappa_\varepsilon - 3}{m}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2}\mathrm{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi} \\
&\qquad + \frac{\kappa_\varepsilon - 3}{m}\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2}\mathrm{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a,
\end{aligned}
$$

which does not depend on κ_v. Thus, we do not have to estimate κ_v when we provide the estimator $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$. In other words, the estimator $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is the same form as in the case that the random effects have the multivariate normal distribution, which imples that this estimator is robust over distribution of random effects.

Corollary 3.3.1 *The second-order unbiased estimator* $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ *is robust over any distributions of* $\boldsymbol{v}_i$*'s.*

When $\boldsymbol{v}_i$'s and $\boldsymbol{\varepsilon}_i$'s are normally distributed, the second-order unbiased estimator $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is given by replacing $\boldsymbol{\Psi}$ with $\widehat{\boldsymbol{\Psi}}^+$ in the expression

$$
\begin{aligned}
&\boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}) + \frac{1}{m^2}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big\{\sum_{i=1}^m \Big[(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1} \\
&\quad + \mathrm{tr}\,\{(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\}\boldsymbol{I}_k\Big](\boldsymbol{\Psi}+\boldsymbol{D}_i)\Big\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a \\
&\quad - \boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}}(\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a,
\end{aligned}
$$

because $\kappa_v = \kappa_\varepsilon = 3$.

3.4 Simulation and Empirical Studies

3.4.1 Finite sample performances

We now investigate finite sample performances of a second-order unbiased estimator of the mean squared error matrix of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB}$ given in (3.3.7).

We treat the multivariate Fay-Herriot model (3.2.1) for $k = 2$ and $m = 40$. The design matrix, $\boldsymbol{X}_i$ is a $k \times 2k$ matrix, such that

$$
\boldsymbol{X}_i = \begin{pmatrix} 1 & x_{i1} & 0 & 0 \\ 0 & 0 & 1 & x_{i2} \end{pmatrix},
$$

where x_{ij}'s are generated from the uniform distribution on $(-1, 1)$, which are fixed through the simulation runs. As a setup of the covariance matrix $\mathbf{\Psi}$ of the random effects, we consider

$$\mathbf{\Psi} = \rho \boldsymbol{\psi}_2 \boldsymbol{\psi}_2^\top + (1-\rho)\mathrm{diag}(\boldsymbol{\psi}_2 \boldsymbol{\psi}_2^\top)$$

where $\boldsymbol{\psi}_2 = (\sqrt{1.6}, \sqrt{0.8})^\top$ and $\mathrm{diag}(\boldsymbol{A})$ denotes the diagonal matrix consisting of diagonal elements of matrix $\boldsymbol{A}$. Here, ρ is the correlation coefficient, and we handle the three cases $\rho = 0.2, 0.4, 0.6$. The cases of negative correlations are omitted, because we observe the same results with those of positive ones.

Concerning the dispersion matrices $\boldsymbol{D}_i$'s of sampling errors $\boldsymbol{\varepsilon}_i$'s, we treat the two $\boldsymbol{D}_i$-patterns: (a) $0.7\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.3\mathbf{I}_k$ and (b) $2.0\mathbf{I}_k$, $0.6\mathbf{I}_k$, $0.5\mathbf{I}_k$, $0.4\mathbf{I}_k$, $0.2\mathbf{I}_k$. In the univariate Fay-Herriot model, these cases are treated by Datta, *et al*. (2005). There are five groups $G_1, \ldots, G_5$ corresponding to these $\boldsymbol{D}_i$-patterns, and there are eight small areas in each group, where the sampling covariance matrices $\boldsymbol{D}_i$ are the same for areas within the same group.

Concerning the underlying distributions for $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$, we consider three kinds of distributions, that is, multivariate normalized t distributions with degrees of freedom 5, multivariate normalized chi-squared distributions with degrees of freedom 2 and multivariate normalized lognormal distributions $\Lambda(0, 1)$, which are denoted by M1, M2 and M3, respectively.

We begin with obtaining the true mean squared error matrices of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB} = \widehat{\boldsymbol{\theta}}_a(\widehat{\mathbf{\Psi}}^+)$ by simulation. Let $\{\boldsymbol{y}_i^{(r)}, i = 1, \ldots, m\}$ be the simulated data in the r-th replication for $r = 1, \ldots, R$ with $R = 100,000$. Let $\boldsymbol{\theta}_a^{(r)}$, $\widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi})$ and $\widehat{\boldsymbol{\theta}}_a^{EB(r)}$ be the values of $\boldsymbol{\theta}_a$, $\widehat{\boldsymbol{\theta}}_a(\mathbf{\Psi})$ and $\widehat{\boldsymbol{\theta}}_a^{EB}$ in the r-th replication. Then the simulated value of the true mean squared error matrices is calculated by

$$\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\mathbf{\Psi}) + \boldsymbol{G}_{2a}(\mathbf{\Psi}) + R^{-1}\sum_{i=1}^{R}\{\widehat{\boldsymbol{\theta}}_a^{EB(r)} - \widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB(r)} - \widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi})\}^\top$$

$$+ R^{-1}\sum_{i=1}^{R}\{\widehat{\boldsymbol{\theta}}_a^{EB(r)} - \widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi}) - \boldsymbol{\theta}_a^{(r)}\}^\top + R^{-1}\sum_{i=1}^{R}\{\widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi}) - \boldsymbol{\theta}_a^{(r)}\}\{\widehat{\boldsymbol{\theta}}_a^{EB(r)} - \widehat{\boldsymbol{\theta}}_a^{(r)}(\mathbf{\Psi})\}^\top.$$

To measure performances of a second-order unbiased estimator of the MESM of $\widehat{\boldsymbol{\theta}}_a^{EB}$, we obtain the Frobenius risk of a second-order unbiased estimator of the MESM of $\widehat{\boldsymbol{\theta}}_a^{EB}$ with and without the normal assumption for $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_i$ by simulation. A second-order unbiased estimator of the MESM of $\widehat{\boldsymbol{\theta}}_a^{EB}$ with the normal assumption is the same with (3.3.7), but κ_v and κ_ε are eaqual to 0.

Let $\widehat{\boldsymbol{\theta}}_a^{EB(r^*)}$ be the EBLUP for the ath area in the r^*-th replication for $r^* = 1, \ldots, R^*$ with $R^* = 10,000$. The simulated Frobenius risk of a second-order unbiased estimator of the MESM of $\widehat{\boldsymbol{\theta}}_a^{EB}$ is calculated by

$$R^{*-1}\sum_{r^*=1}^{R^*}\mathrm{tr}\left(\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB(r^*)}) - \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\right)\left(\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB(r^*)}) - \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\right)^\top,$$

where $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is a simulated true MESM of $\widehat{\boldsymbol{\theta}}_a^{EB}$.

Tables 3.1 and 3.2 report the simulated Frobenius risks of a second-order unbiased estimator of the MESM for $\boldsymbol{D}_i$-patterns: (a) and (b), respectively. The simulated Frobenius risk of a second-order unbiased estimator of the MESM with a normality assumption is displayed in parenthesis.

It can be seen that for almost all the cases, the frobenius risks of the MSEM estimator without a normality assumption are smaller than those with a normality assumption. For the stable sampling error variances, that is, $\boldsymbol{D}_i$-patterns: (a), there is a littele difference between the simulated Frobenius risk with and without a normality assumption for all ρ patterns and for all area groups when the underlying distribution of random effects and sampling errors are t distributions. On the other hand, when the underlying distribution of random effects and sampling errors have long tails, especially they are log-normal distributions, there are well-marked improvements for the second-order unbiased estimator of the MESM without a normality assumption. This is because kurtosis of the distribution which have long tails tend to be large and there is large difference between the second-order unbiased estimator of the MESM with and without a normality assumption.

When sampling error variances for some area groups are extremely large, that is, for $\boldsymbol{D}_i$-patterns: (b), the results are similar to for that is, $\boldsymbol{D}_i$-patterns: (a). In this case, however, the values of risks of the MESM estimator are larger especially for the area groups with an extremely large and extremely small sampling variance, and there are considerable improvements for the second-order unbiased estimator of the MESM without a normality assumption for all disribution patterns. Then, we can say that a second-order unbiased estimator of the mean squared error matrix of $\widehat{\boldsymbol{\theta}}_a^{EB}$ given in (3.3.7) have robustness for specifying the distribtions of random effects and samplin errors.

	t			chi-square			Log-normal		
ρ	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
G_1	0.115	0.126	0.137	0.171	0.187	0.208	0.789	0.633	0.536
	(0.137)	(0.146)	(0.167)	(0.177)	(0.193)	(0.216)	(1.202)	(1.037)	(0.912)
G_2	0.076	0.083	0.087	0.115	0.126	0.143	0.263	0.281	0.272
	(0.098)	(0.102)	(0.109)	(0.119)	(0.132)	(0.150)	(0.540)	(0.577)	(0.572)
G_3	0.046	0.055	0.051	0.075	0.088	0.081	0.119	0.386	0.119
	(0.064)	(0.078)	(0.065)	(0.080)	(0.093)	(0.087)	(0.346)	(0.695)	(0.317)
G_4	0.018	0.022	0.022	0.037	0.042	0.041	0.025	0.030	0.024
	(0.049)	(0.052)	(0.056)	(0.057)	(0.075)	(0.074)	(0.045)	(0.062)	(0.047)
G_5	0.007	0.006	0.010	0.013	0.014	0.013	0.040	0.033	0.039
	(0.039)	(0.028)	(0.030)	(0.026)	(0.025)	(0.027)	(0.023)	(0.035)	(0.077)

Table 3.1: Frobenius risk of estimator of MSEM with $\boldsymbol{D}_i$-patterns: (a). Values in parenthesis are the risk of estimator of MSEM multiplied by 10 under normality assumption.

3.4.2 Illustrative example

Data from the 2016 Survey of Family Income and Expenditure in Japan, which is based on two or more person households (excluding agricultural, forestry and fisheries households) can be

	t			chi-square			Log-normal		
ρ	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
G_1	1.397	1.341	1.392	1.796	1.833	1.604	9.375	4.752	4.180
	(1.571)	(1.450)	(1.651)	(1.835)	(1.874)	(1.632)	(12.24)	(6.603)	(5.907)
G_2	0.039	0.042	0.071	0.082	0.092	0.078	0.141	0.180	0.298
	(0.082)	(0.078)	(0.075)	(0.097)	(0.112)	(0.090)	(0.423)	(0.374)	(0.969)
G_3	0.021	0.033	0.087	0.035	0.034	0.041	0.834	0.558	0.845
	(0.037)	(0.033)	(0.037)	(0.046)	(0.043)	(0.046)	(0.082)	(0.142)	(0.077)
G_4	0.043	0.073	0.173	0.014	0.015	0.030	1.529	1.450	1.649
	(0.107)	(0.156)	(0.297)	(0.066)	(0.059)	(0.095)	(1.801)	(1.815)	(1.973)
G_5	0.339	0.534	1.182	0.174	0.279	0.512	10.82	9.075	11.27
	(0.417)	(0.655)	(1.437)	(0.246)	(0.379)	(0.702)	(11.91)	(10.05)	(12.52)

Table 3.2: Frobenius risk of estimator of MSEM with $\boldsymbol{D}_i$-patterns: (b). Values in parenthesis are the risk of estimator of MSEM multiplied by 10 under normality assumption.

obtained and we apply these data to the multivariate Fay-Herriot model (3.2.1) for illustration. The 47 Japanese prefectural capitals are the target domains, and these are divided into 10 regions: Hokkaido, Tohoku, Kanto, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, Kyushu and Okinawa. Each region consists of several prefectures except Hokkaido and Okinawa, which consist of one prefecture.

In this study, the reported data of the yearly averaged monthly spendings on 'Education' and ' Cultural-amusement' per worker's household, scaled by 1,000 Yen, at each capital city of 47 prefectures are observed as $(y_{i1}, y_{i2})^\top$. In addition, we use the data in the 2014 National Survey of Family Income and Expenditure. The average spending data in this survey are more reliable than the Survey of Family Income and Expenditure since the sample sizes are much larger. However, this survey is conducted only once in every five years. As auxiliary variables, we use the data of the average spendings on 'Education' and ' Cultural-amusement', which is denoted by EDU_i and CUL_i, respectively. Then the regressor in the model (3.2.1) is

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & \mathrm{EDU}_i & 0 & 0 \\ 0 & 0 & 1 & \mathrm{CUL}_i \end{pmatrix}.$$

We assume that the sampling covariance matrix $\boldsymbol{D}_i$ of the i-th region are the same for prefectures within the same region. Then, these matrices $\boldsymbol{D}_i$ for $i = 1, \ldots, 10$ are estimated by data of yearly averaged monthly spendings on 'Education' and ' Cultural-amusement' in the past ten years (2006-2015), that is, $\boldsymbol{D}_i$ is given as the average of the sampling covariance matrices of prefectures within the i-th region.

Kurtosis κ_ε of sampling errors is estimated by sample kurtosis after standardizing data of yearly averaged monthly spendings on 'Education' and ' Cultural-amusement' in the past ten years (2006-2015) by sampling covariance matrices $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_{10}$. Then, we obtain $\kappa_\varepsilon = 4.70$. The estimates of the covariance matrix $\boldsymbol{\Psi}$ and the correlation coefficient ρ is

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 8.98 & 3.19 \\ 3.19 & 10.83 \end{pmatrix} \quad \text{and} \quad \hat{\rho} = 0.32.$$

The estimates of the regression coefficients are $\widehat{\boldsymbol{\beta}} = (4.49, 0.82, 12.16, 0.65)^\top$.

The values of direct estimate and EBLUP of spendings on 'Education' and ' Cultural-amusement' are reported in Table 3.3. We pick up one prefecture from each region: Hokakido, Fukushima, Tokyo, Nigata, Aichi, Osaka, Hiroshima, kochi, Fukuoka and Okinawa.

	direct estimator (EDU)	EBLUP (EDU)	direct estimator (CUL)	EBLUP (CUL)
Hokkaido	15.1	15.0	31.5	30.7
Fukushima	13.3	12.7	29.9	29.0
Tokyo	32.5	31.9	41.8	40.6
Nigata	15.4	15.5	28.4	29.7
Aichi	20.2	20.6	30.9	30.6
Osaka	22.8	22.5	29.4	30.1
Hiroshima	19.4	18.7	30.7	30.4
Kochi	27.2	26.3	31.1	30.6
Fukuoka	13.5	13.8	28.0	28.1
Okinawa	13.2	13.3	19.4	20.2

Table 3.3: Direct estimates and EBLUP (1,000yen)

The uncertainty of EBLUP is provided by the second-order unbiased estimator of MSEM of EBLUP. Table 3.4 and 3.5 report the estimates of MSEM averaged over prefectures within each region for 10 regions with and without a normality assumption. It can be seen that the estimates of MSEM with a normality assumption slightly under estimate the risk of EBLUP than those without a normality assumption.

Hokkaido	Tohoku	Kanto	Hokuriku	Tokai
$\begin{bmatrix} 0.51 & 0.73 \\ 0.73 & 3.84 \end{bmatrix}$	$\begin{bmatrix} 3.20 & 2.20 \\ 2.20 & 3.41 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.30 \\ 0.30 & 2.46 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.58 \\ 0.58 & 4.78 \end{bmatrix}$	$\begin{bmatrix} 1.37 & 0.64 \\ 0.64 & 1.78 \end{bmatrix}$
Kinki	Chugoku	Shikoku	Kyushu	Okinawa
$\begin{bmatrix} 0.95 & -0.04 \\ -0.04 & 2.82 \end{bmatrix}$	$\begin{bmatrix} 1.47 & 0.29 \\ 0.29 & 2.57 \end{bmatrix}$	$\begin{bmatrix} 4.23 & 0.88 \\ 0.88 & 3.53 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0.65 \\ 0.65 & 1.74 \end{bmatrix}$	$\begin{bmatrix} 3.02 & 0.78 \\ 0.78 & 1.68 \end{bmatrix}$

Table 3.4: Estimates of the mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ without a normality assumption.

Hokkaido	Tohoku	Kanto	Hokuriku	Tokai
$\begin{bmatrix} 0.51 & 0.71 \\ 0.71 & 3.73 \end{bmatrix}$	$\begin{bmatrix} 3.12 & 2.13 \\ 2.13 & 3.33 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0.30 \\ 0.30 & 2.42 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0.56 \\ 0.56 & 4.60 \end{bmatrix}$	$\begin{bmatrix} 1.36 & 0.63 \\ 0.63 & 1.76 \end{bmatrix}$
Kinki	Chugoku	Shikoku	Kyushu	Okinawa
$\begin{bmatrix} 0.94 & -0.03 \\ -0.03 & 2.76 \end{bmatrix}$	$\begin{bmatrix} 1.46 & 0.29 \\ 0.29 & 2.53 \end{bmatrix}$	$\begin{bmatrix} 4.05 & 0.86 \\ 0.86 & 3.44 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0.64 \\ 0.64 & 1.73 \end{bmatrix}$	$\begin{bmatrix} 2.94 & 0.77 \\ 0.77 & 1.67 \end{bmatrix}$

Table 3.5: Estimates of the mean squared error matrices of $\widehat{\boldsymbol{\theta}}_a^{EB}$ with a normality assumption.

3.5 Proofs

Proof of Theorem 3.2.1

We begin by writing $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi}$ as

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} =& \frac{1}{m}\sum_{i=1}^m \{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\} + \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_i^\top \\ & - \frac{1}{m}\sum_{i=1}^m (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_i^\top - \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top,\end{aligned}$$

which yields the bias given in (3.2.7). It is easy to check that the bias is of order $O(m^{-1})$.

For (2), it is noted that $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}$ is approximated as

$$\begin{aligned}\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} =& \frac{1}{m}\sum_{i=1}^m \{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\} + O_p(m^{-1}) \\ =& \frac{1}{m}\sum_{i=1}^m \{\boldsymbol{u}_i\boldsymbol{u}_i^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\} + O_p(m^{-1}), \end{aligned} \tag{3.5.1}$$

where $\boldsymbol{u}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}$, having $\mathcal{N}_k(\boldsymbol{0}, \boldsymbol{\Psi} + \boldsymbol{D}_i)$. It is here noted that $(\boldsymbol{u}_i\boldsymbol{u}_i^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i))/m$ for $i = 1, \ldots, m$ are mutually independent and $E(\boldsymbol{u}_i\boldsymbol{u}_i^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i))/m = 0$ for $i = 1, \ldots, m$. Then the consistency follows because $\sum_{i=1}^m E(\boldsymbol{u}_i\boldsymbol{u}_i^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i))^2/m^2 = \sum_{i=1}^m (2(\boldsymbol{\Psi} + \boldsymbol{D}_i)^2 + \mathrm{tr}\,(\boldsymbol{\Psi} + \boldsymbol{D}_i)\boldsymbol{I}_k)/m^2 = O(m^{-1})$ under condition (H2). Using condition (H2) and the finiteness of moments of normal random variables, we can show that $\sqrt{m}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})$ converges to a multivariate normal distribution, which implies that $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = O_p(m^{-1/2})$.

We next verify that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta} = O_p(m^{-1/2})$. Note that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta}$ is decomposed as $\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\} + \{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}$. For $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}$, it is noted that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta} = \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i - E\boldsymbol{y}_i). \tag{3.5.2}$$

Then, $\mathrm{Var}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}) = \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1} = O(1/m)$ and this implies $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta} = O_p(m^{-1/2})$. We next evaluate $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta}(\boldsymbol{\Psi})$ as

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta}(\boldsymbol{\Psi}) \\ &= \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_i)^{-1}\boldsymbol{y}_i \\ &- \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\boldsymbol{y}_i \\ &= \Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_i)^{-1} - (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}\Big\}\boldsymbol{y}_i\end{aligned}$$

$$
+\Big[\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}-\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\Big]\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i
$$
$$
= I_1 + I_2. \tag{3.5.3}
$$

First, I_1 is written as

$$
I_1 = -\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i, \tag{3.5.4}
$$

which is of order $O_p(m^{-1/2})$, because $\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i = O_p(m)$ and $\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i = O_p(m^{1/2})$. Next, I_2 is rewritten as

$$
\begin{aligned}
I_2 = &-\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\Big\}\boldsymbol{X}_i\\
&\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i\\
=&\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\\
&\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i,
\end{aligned} \tag{3.5.5}
$$

which is also of order $O_p(m^{-1/2})$, because $\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i = O_p(m)$, $\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i = O_p(m^{1/2})$, $\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i = O(m)$ and $\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{y}_i = O_p(m)$. Thus, we have $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}(\boldsymbol{\Psi}) = O_p(m^{-1/2})$, and it is concluded that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta} = O_p(m^{-1/2})$.

For (3), let $\hat{\lambda}_1,\ldots,\hat{\lambda}_k$ be eigenvalues of $\widehat{\boldsymbol{\Psi}}$, and let $\lambda_1,\ldots,\lambda_k$ be eigenvalues of $\boldsymbol{\Psi}$. Then, for $j=1,\ldots,k$,

$$
P(\hat{\lambda}_j<0)=P(\hat{\lambda}_j-\lambda_j<-\lambda_j)=P(-(\hat{\lambda}_j-\lambda_j)>\lambda_j)\le P(|\sqrt{m}(\hat{\lambda}_j-\lambda_j)|>\sqrt{m}\lambda_j).
$$

Note that $\lambda_j>0$. It follows from the Markov inequality that for any $K>0$,

$$
P(|\sqrt{m}(\hat{\lambda}_j-\lambda_j)|>\sqrt{m}\lambda_j)\le\frac{E[\{|\sqrt{m}(\hat{\lambda}_j-\lambda_j)|\}^{2K}]}{(\sqrt{m}\lambda_j)^{2K}}=O(m^{-K}),
$$

because the $4K$-th moments exist and $\hat{\lambda}_j-\lambda_j=O_p(m^{-1/2})$ from $\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}=O_p(m^{-1/2})$. □

3.5.1 Proof of Lemma 3.3.1

From (2) in Theorem 3.2.1, it is sufficient to show this approximation for $\widehat{\boldsymbol{\Psi}}$ instead of $\widehat{\boldsymbol{\Psi}}^{+}$. It is observed that

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) =&\boldsymbol{D}_a\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}-(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})+\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\
&-\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}\}.
\end{aligned}
$$

Using the equation

$$(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}=(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}, \tag{3.5.6}$$

we can see that

$$\begin{aligned}&\boldsymbol{D}_a\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}-(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})+O_p(m^{-1})\end{aligned}$$

and

$$\begin{aligned}&\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}-\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ &=\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}+O_p(m^{-1}).\end{aligned}$$

Thus, we have

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})=&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\\ &+\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}+O_p(m^{-1})\\ =&I_1+I_2+O_p(m^{-1}).\quad\text{(say)}\end{aligned}$$

For I_2, it is noted that

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\\ &=\Big[\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big]\\ &\quad\times\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\ &\quad+\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m\boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\Big\}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\ &=I_{21}+I_{22}.\end{aligned}$$

We can evaluate I_{21} as

$$\begin{aligned}I_{21}=&\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^m\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\\ =&O_p(m^{-1}),\end{aligned}$$

because $\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j=O(m)$, $\sum_{i=1}^m\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i=O_p(m^{1/2})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}=O_p(m^{-1/2})$ from Theorem 1 (2). We next estimate I_{22} as

$$\begin{aligned}I_{22}=&-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^m\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)\boldsymbol{X}_i\Big\}\\ &\times\Big\{\sum_{i=1}^m\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m\boldsymbol{X}_i^\top\boldsymbol{A}(\widehat{\boldsymbol{\Psi}},\boldsymbol{D}_i)(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\end{aligned}$$

for $\boldsymbol{A}(\widehat{\boldsymbol{\Psi}}, \boldsymbol{D}_i) = (\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1}$. It can be seen that $I_{22} = O_p(m^{-1})$ from the same arguments as in I_{21}. Thus, it follows that $I_2 = O_p(m^{-1})$.

Then, we only need to evaluate $E[I_1 I_1^\top]$ for the approximation of $E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}^\top]$. Let $J_1 = E[(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})]$. It is noted from (3.5.1) that $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}$ is approximated as

$$\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = \frac{1}{m}\sum_{i=1}^{m}\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\} + O_p(m^{-1}),$$

which is used to evaluate

$$\begin{aligned}
J_1 =& \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \\
&\times \{(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}\Big] + O(m^{-3/2}) \\
=& \frac{1}{m^2}\sum_{i=1}^{m} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \\
&\times \{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}\Big] + O(m^{-3/2}) \\
=& \frac{1}{m^2}\sum_{i \neq a} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}\Big] \\
&+ O(m^{-3/2}) \\
=& \frac{1}{m^2}\sum_{i \neq a}\Big\{E\Big[(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top\Big] \\
&- (\boldsymbol{\Psi} + \boldsymbol{D}_i)(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi} + \boldsymbol{D}_i)\Big\} + O(m^{-3/2}).
\end{aligned}$$

The second equality follows since $E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\{(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_j)\}\Big] = \boldsymbol{0}$ for $i \neq j$, and the third equality follows since the term for $i = a$ in the summation is of order $O(m^{-2})$. For $i \neq a$,

$$\begin{aligned}
&E\Big[(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top\Big] \\
=& E\Big[\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top + 2\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top + 2\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top \\
&+ \boldsymbol{\varepsilon}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{\varepsilon}_i^\top + \boldsymbol{v}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{v}_i^\top + \boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top\Big].
\end{aligned}$$

We evaluate $E[\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top]$ and $E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top]$. We can write

$$E[\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top] = \boldsymbol{\Psi}^{1/2}E[\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}^{1/2}\boldsymbol{u}_i\boldsymbol{u}_i^\top]\boldsymbol{\Psi}^{1/2}.$$

The (a, d)-element of $E[\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{B}\boldsymbol{u}_i\boldsymbol{u}_i^\top]$ for some symmetric matrix $\boldsymbol{B}$ is $\sum_{b,c} E[u_{ia}u_{ib}\boldsymbol{B}_{bc}u_{ic}u_{id}]$, where $\boldsymbol{B}_{bc}$ is the (b, c)-element of $\boldsymbol{B}$. Since $\boldsymbol{u}_i \sim (\boldsymbol{0}, \boldsymbol{I}_m)$, we only need to evaluate four cases;

(1) $a=b=c=d$, (2) $a=d, b=c$ and $a\neq b$, (3) $a=b, c=d$ and $a\neq c$, and (4) $a=c, b=d$ and $a\neq b$. For the case (1), we have

$$\sum_{b,c} E[u_{ia}u_{ib}\boldsymbol{B}_{bc}u_{ic}u_{id}] = E[u_{ia}^4\boldsymbol{B}_{aa}] = \kappa_v\boldsymbol{B}_{aa}.$$

For the case (2), we have

$$\sum_{b,c} E[u_{ia}u_{ib}\boldsymbol{B}_{bc}u_{ic}u_{id}] = \sum_{b\neq a} E[u_{ia}^2u_{ib}^2\boldsymbol{B}_{bb}] = \sum_{b}\boldsymbol{B}_{bb} - \boldsymbol{B}aa.$$

These imply that the diagonal elements of $E[\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{B}\boldsymbol{u}_i\boldsymbol{u}_i^\top]$ is those of $(\kappa_v-1)\boldsymbol{B}+\mathrm{tr}\,(\boldsymbol{B})\boldsymbol{I}_m$. For the case (3), we have

$$\sum_{b,c} E[u_{ia}u_{ib}\boldsymbol{B}_{bc}u_{ic}u_{id}] = E[u_{ia}^2u_{id}^2\boldsymbol{B}_{ad}] = \boldsymbol{B}_{ad}.$$

For the case (4), we have

$$\sum_{b,c} E[u_{ia}u_{ib}\boldsymbol{B}_{bc}u_{ic}u_{id}] = E[u_{ia}^2u_{id}^2\boldsymbol{B}_{ad}] = \boldsymbol{B}_{ad}.$$

These imply that the off-diagonal elements of $E[\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{B}\boldsymbol{u}_i\boldsymbol{u}_i^\top]$ is those of $2\boldsymbol{B}$. Hence, we have

$$E[\boldsymbol{u}_i\boldsymbol{u}_i^\top\boldsymbol{B}\boldsymbol{u}_i\boldsymbol{u}_i^\top] = (\kappa_v-1)\mathrm{diag}\boldsymbol{B}+\mathrm{tr}\,(\boldsymbol{B})\boldsymbol{I}+2(\boldsymbol{B}-\mathrm{diag}\boldsymbol{B}) = (\kappa_{\mathrm{v}}-3)\mathrm{diag}\boldsymbol{B}+2\boldsymbol{B}+\mathrm{tr}\,(\boldsymbol{B})\boldsymbol{I}_{\mathrm{m}}.$$

Therefore,

$$\begin{aligned}
&E[\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top]\\
=&(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}+2\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}+\mathrm{tr}\,(\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1})\boldsymbol{\Psi}.
\end{aligned}$$

In the same way, we have

$$\begin{aligned}
&E[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top]\\
=&(\kappa_\varepsilon-3)\boldsymbol{D}_i^{1/2}\mathrm{diag}(\boldsymbol{D}_{\mathrm{i}}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}^{1/2})\boldsymbol{D}_{\mathrm{i}}^{1/2}+2\boldsymbol{D}_{\mathrm{i}}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}+\mathrm{tr}\,(\mathrm{D}_{\mathrm{i}}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1})\boldsymbol{D}_{\mathrm{i}}.
\end{aligned}$$

Hence, for $i\neq a$,

$$\begin{aligned}
&E\Big[(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top\Big]\\
=&E\Big[\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top+2\boldsymbol{v}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top+2\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{v}_i^\top\\
&+\boldsymbol{\varepsilon}_i\boldsymbol{v}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_i\boldsymbol{\varepsilon}_i^\top+\boldsymbol{v}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{v}_i^\top+\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^\top\Big]\\
=&(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}+2\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}+\mathrm{tr}\,(\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1})\boldsymbol{\Psi}\\
&+4\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_i+\mathrm{tr}\,(\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{D}_i+\mathrm{tr}\,(\boldsymbol{D}_i(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{\Psi}\\
&+(\kappa_\varepsilon-3)\boldsymbol{D}_i^{1/2}\mathrm{diag}(\boldsymbol{D}_{\mathrm{i}}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}^{1/2})\boldsymbol{D}_{\mathrm{i}}^{1/2}+2\boldsymbol{D}_{\mathrm{i}}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}+\mathrm{tr}\,(\mathrm{D}_{\mathrm{i}}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1})\boldsymbol{D}_{\mathrm{i}}\\
=&2(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i)+\mathrm{tr}\,((\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})(\boldsymbol{\Psi}+\boldsymbol{D}_i)\\
&+(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}+(\kappa_\varepsilon-3)\boldsymbol{D}_{\mathrm{i}}^{1/2}\mathrm{diag}(\boldsymbol{D}_{\mathrm{i}}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}^{1/2})\boldsymbol{D}_{\mathrm{i}}^{1/2}
\end{aligned}$$

Hence, we have

$$J_1 = \frac{1}{m^2}\sum_{i=1}^{m}\Big[(\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{\Psi}+\boldsymbol{D}_i) + \mathrm{tr}\,((\boldsymbol{\Psi}+\boldsymbol{D}_i)(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})(\boldsymbol{\Psi}+\boldsymbol{D}_i)$$
$$+ (\kappa_v - 3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2} + (\kappa_\varepsilon - 3)\boldsymbol{D}_{\mathrm{i}}^{1/2}\mathrm{diag}(\boldsymbol{D}_{\mathrm{i}}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_{\mathrm{a}})^{-1}\boldsymbol{D}_{\mathrm{i}}^{1/2})\boldsymbol{D}_{\mathrm{i}}^{1/2}\Big]$$
$$+ O(m^{-3/2}),$$

which leads to the expression in (3.3.3) by multipling J_1 by $\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}$ and $(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a$ from left and right sides, respectively. □

3.5.2 Proof of Lemma 3.3.2

We can decompose

$$\begin{aligned}&E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a\}^\top] \\ =&E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*\}^\top] + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top].\end{aligned} \tag{3.5.7}$$

Since $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^* = O(m^{-1/2})$, using the statement in the proof of Lemma 3.3.1, the first term in (3.5.7) can be written as

$$\begin{aligned}&E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*\}^\top] \\ &= - \boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E\Big[(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}^\top\Big]\boldsymbol{X}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a + O(m^{-3/2}).\end{aligned}$$

In the same way, we have

$$\begin{aligned}&E\Big[(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}^\top\Big] \\ =&\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}E\Big[\{(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)^\top\Big] \\ &\times \boldsymbol{X}_j(\boldsymbol{X}(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}^\top)^{-1} + O(m^{-3/2}) \\ =&\frac{1}{m}E\Big[\{(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_a)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top\Big] \\ &\times \boldsymbol{X}_a(\boldsymbol{X}(\boldsymbol{I}_m\otimes\boldsymbol{\Psi}+\boldsymbol{D})^{-1}\boldsymbol{X}^\top)^{-1} + O(m^{-3/2}) \\ =&O(m^{-3/2}),\end{aligned}$$

since the expectation is zero for $i \neq a$ or $j \neq a$. Hence, we have $E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) - \boldsymbol{\theta}_a^*\}^\top] = O(m^{-3/2})$.

Next, we evaluate the second term in (3.5.7). It is noted that

$$\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a = -\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi} = O(1)$$

and

$$\begin{aligned}&\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}) \\ &=\boldsymbol{D}_a\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1} - (\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta}) + \boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta}\} \\ &\quad - \boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}.\end{aligned} \tag{3.5.8}$$

We evaluate the expectation of the product of $\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a$ and the first term in (3.5.8), namely $\boldsymbol{D}_a E[\{(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1} - (\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top]$. Using the equation,

$$
\begin{aligned}
(\boldsymbol{\Psi}+&\boldsymbol{D}_a)^{-1} - (\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\\
=&(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\\
&-(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1} + O(m^{-3/2}),
\end{aligned}
$$

this can be written as

$$
\begin{aligned}
&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big\{E\Big[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big]\\
&\quad + E\Big[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big]\Big\} + O(m^{-3/2}) \qquad (3.5.9)\\
=&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(K_1+K_2) + O(m^{-3/2}). \quad \text{(say)}
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi} =& \frac{1}{m}\sum_{i=1}^m\{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_i)\} + \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top\\
&-\frac{1}{m}\sum_{i=1}^m(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top - \frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top.
\end{aligned}
\qquad (3.5.10)
$$

We can write $K_1 = K_{11} + K_{12} + K_{13} + K_{14}$, where

$$
K_{11} = E\Big[\Big\{\frac{1}{m}\sum_{i=1}^m\{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big],
$$

$$
K_{12} = E\Big[\frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big],
$$

$$
K_{13} = -E\Big[\frac{1}{m}\sum_{i=1}^m(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big],
$$

and

$$
K_{14} = -E\Big[\frac{1}{m}\sum_{i=1}^m \boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{y}_a - \boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top\Big].
$$

K_{11} can be written as

$$
\begin{aligned}
K_{11} =& \frac{1}{m}\sum_{i=1}^m E\Big[\{(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=& \frac{1}{m}E\Big[\{(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top - (\boldsymbol{\Psi}+\boldsymbol{D}_a)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=& \frac{1}{m}E\Big[(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big],
\end{aligned}
$$

since the expectation is zero for $i \neq a$. Using the results from Lemma 3.3.1, the first term in the last eqation is

$$
\begin{aligned}
-&\frac{1}{m}E\Big[(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\boldsymbol{v}_a^\top\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\\
&=-\frac{1}{m}E\Big[\boldsymbol{v}_a\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_a\boldsymbol{v}_a^\top+2\boldsymbol{\varepsilon}_a\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_a\boldsymbol{v}_a^\top+\boldsymbol{v}_a\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_a\boldsymbol{v}_a^\top\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\\
&=-\frac{1}{m}\Big\{(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}+2\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}+\mathrm{tr}\,(\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{\Psi}\\
&\quad+2\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}+\mathrm{tr}\,(\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{\Psi}\Big\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a,
\end{aligned}
$$

and the second term is

$$
\begin{aligned}
&\frac{1}{m}E\Big[(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\boldsymbol{\varepsilon}_a^\top\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\\
&=-\frac{1}{m}E\Big[\boldsymbol{\varepsilon}_a\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_a\boldsymbol{\varepsilon}_a^\top+2\boldsymbol{v}_a\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\varepsilon}_a\boldsymbol{\varepsilon}_a^\top+\boldsymbol{\varepsilon}_a\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{v}_a\boldsymbol{\varepsilon}_a^\top\Big](\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\\
&=-\frac{1}{m}\Big\{(\kappa_\varepsilon-3)\boldsymbol{D}_a^{1/2}\mathrm{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}+2\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\mathrm{tr}\,(\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{D}_a\\
&\quad+2\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\mathrm{tr}\,(\boldsymbol{\Psi}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1})\boldsymbol{D}_a\Big\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
K_{11}=&-\frac{1}{m}(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\\
&+\frac{1}{m}(\kappa_\varepsilon-3)\boldsymbol{D}_a^{1/2}\mathrm{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}.
\end{aligned}
$$

K_{12} can be written as

$$
\begin{aligned}
K_{12}=&\frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m E\Big[\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)(\boldsymbol{v}_k+\boldsymbol{\varepsilon}_k)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&\frac{1}{m}\sum_{i=1}^m E\Big[\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&O(m^{-2}),
\end{aligned}
$$

The third equality follows since for $j \neq a$ or $k \neq a$, the expectation is zero, and the last equality follows since $\boldsymbol{X}^\top\boldsymbol{X} = O(m)$. For the same reason, K_{13} can be written as

$$
\begin{aligned}
K_{13}=&\frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m E\Big[(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&\frac{1}{m}E\Big[(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{X}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)
\end{aligned}
$$

$$
\begin{aligned}
&\times \{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&O(m^{-2}).
\end{aligned}
$$

Similarly, K_{14} is of order $O(m^{-2})$. Hence, we have

$$
\begin{aligned}
&\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E[K_1]\\
&=\frac{1}{m}\boldsymbol{D}_a(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\Big\{-(\kappa_v-3)\boldsymbol{\Psi}^{1/2}\mathrm{diag}(\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}^{1/2})\boldsymbol{\Psi}^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a\\
&\quad+(\kappa_\varepsilon-3)\boldsymbol{D}_a^{1/2}\mathrm{diag}(\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a^{1/2})\boldsymbol{D}_a^{1/2}(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\Big\}+O(m^{-3/2}).
\end{aligned}
$$

Next, we evaluate K_2, which can be written as by (3.5.10)

$$
\begin{aligned}
K_2=&E\Big[\Big\{\frac{1}{m}\sum_{i=1}^m\{(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}+\frac{1}{m}\sum_{i=1}^m\boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top\\
&-\frac{1}{m}\sum_{i=1}^m(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i^\top-\frac{1}{m}\sum_{i=1}^m\boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top\Big\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)\\
&\times\Big\{\frac{1}{m}\sum_{j=1}^m\{(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_j)\}+\frac{1}{m}\sum_{j=1}^m\boldsymbol{X}_j(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_j^\top\\
&-\frac{1}{m}\sum_{j=1}^m(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_j^\top-\frac{1}{m}\sum_{i=1}^m\boldsymbol{X}_j(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})^\top\Big\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top\Big].
\end{aligned}
$$

To evaluate the order of K_2, we need to evaluate the following expectations,

$$
\begin{aligned}
K_{21}=E\Big[&\frac{1}{m^2}\sum_{i=1}^m\{(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)\\
&\times\sum_{j=1}^m\{(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_j)\}(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top\Big],
\end{aligned}
$$

$$
\begin{aligned}
K_{22}=E\Big[&\frac{1}{m^2}\sum_{i=1}^m\{(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)\\
&\times\sum_{j=1}^m\boldsymbol{X}_j(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_j^\top(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top\Big],
\end{aligned}
$$

$$
\begin{aligned}
K_{23}=E\Big[&\frac{1}{m^2}\sum_{i=1}^m\{(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)\\
&\times\sum_{j=1}^m(\boldsymbol{y}_j-\boldsymbol{X}_j\boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_j^\top(\boldsymbol{y}_a-\boldsymbol{X}_a\boldsymbol{\beta})\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top\Big],
\end{aligned}
$$

$$K_{24} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E\Big[\boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a) \boldsymbol{X}_j (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_j^\top \times (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top \Big]$$

$$K_{25} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} E\Big[\boldsymbol{X}_j(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_j^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a) \times (\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_k^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top \Big],$$

and

$$K_{26} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E\Big[\boldsymbol{X}_i(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a) \times (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_j^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top \Big].$$

K_{21} can be written as

$$\begin{aligned} K_{21} =& \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)\{(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)(\boldsymbol{v}_j + \boldsymbol{\varepsilon}_j)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_j)\} \\ & \times (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \Big] \\ =& \frac{1}{m^2} E\Big[\{(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_a)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a)\{(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_j)\} \\ & \times (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \Big] \\ =& O(m^{-2}), \end{aligned}$$

since the expectation is zero for $i \neq a$ or $j \neq a$. K_{22} can be written as

$$\begin{aligned} K_{22} =& \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a) \boldsymbol{X}_j (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_j^\top \\ & \times (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{y}_a - \boldsymbol{X}_a \boldsymbol{\beta}) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \Big] \\ =& \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} \sum_{\ell=1}^{m} E\Big[\{(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_i)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a) \boldsymbol{X}_j (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{v}_k + \boldsymbol{\varepsilon}_k) \\ & \times (\boldsymbol{v}_\ell + \boldsymbol{\varepsilon}_\ell)^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_j^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \Big] \\ =& \frac{1}{m^2} \sum_{j=1}^{m} \sum_{k=1}^{m} E\Big[\{(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a)^\top - (\boldsymbol{\Psi} + \boldsymbol{D}_a)\}(\boldsymbol{\Psi} + \boldsymbol{D}_a) \boldsymbol{X}_j (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{v}_k + \boldsymbol{\varepsilon}_k)(\boldsymbol{v}_k + \boldsymbol{\varepsilon}_k)^\top \\ & \times \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_j^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \Big] \end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{m^2}\sum_{j=1}^m\sum_{k\neq a}E\Big[\{(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_a)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)\boldsymbol{X}_j(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{\Psi}+\boldsymbol{D}_k)\\
&\quad\times\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]+O(m^{-2})\\
=&O(m^{-2}),
\end{aligned}
$$

since the expectation is zero for $i\neq a$ or $k\neq\ell$ and $(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}=O(m^{-1})$. K_{23} can be written as

$$
\begin{aligned}
K_{23}=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m E\Big[\{(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_j^\top\\
&\quad\times(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m E\Big[\{(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_i)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)(\boldsymbol{v}_k+\boldsymbol{\varepsilon}_k)^\top\\
&\quad\times\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&\frac{1}{m^2}E\Big[\{(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_a)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top\\
&\quad\times\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
&\quad+\frac{1}{m^2}\sum_{j\neq a}^m E\Big[\{(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)^\top-(\boldsymbol{\Psi}+\boldsymbol{D}_a)\}(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{\Psi}+\boldsymbol{D}_j)\\
&\quad\times\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big]\\
=&O(m^{-2}),
\end{aligned}
$$

since the expeqtation is zero for $i\neq a$ and $\boldsymbol{X}\boldsymbol{X}^\top=O(m)$. K_{24} can be written as

$$
\begin{aligned}
K_{24}=&\frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{l=1}^m\sum_{p=1}^m\sum_{q=1}^m E\Big[\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_k+\boldsymbol{\varepsilon}_k)(\boldsymbol{v}_l+\boldsymbol{\varepsilon}_l)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_i^\top\\
&\times(\boldsymbol{\Psi}+\boldsymbol{D}_a)\boldsymbol{X}_j(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_p+\boldsymbol{\varepsilon}_p)(\boldsymbol{v}_q+\boldsymbol{\varepsilon}_q)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\\
&\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big].
\end{aligned}
$$

If all k, l, p and q are not equal to a, the expectation is zero, and these terms in the summation vanish. If one of k, l, p and q is equal to a, the expectation is zero unless the others, which are not equal to a are the same, which implies these terms in the summation are of order $O(m^{-3})$ since $\boldsymbol{X}^\top\boldsymbol{X}=O(m)$. If two of k, l, p and q are equal to a, the expectation is zero unless the others, which are not equal to a are the same, which implies these terms in the summation are of order $O(m^{-3})$ since $\boldsymbol{X}^\top\boldsymbol{X}=O(m)$. If three of k, l, p and q are equal to a, the expectation is zero, and these terms in the summation vanish. If all k, l, p and q are equal to a, this term is of order $O(m^{-4})$ since $\boldsymbol{X}^\top\boldsymbol{X}=O(m)$. Then, the above equation is of order $O(m^{-3})$. K_{25} and

K_{26} can be written as

$$K_{25} = \frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{l=1}^m\sum_{p=1}^m\sum_{q=1}^m E\Big[\boldsymbol{X}_j(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_l+\boldsymbol{\varepsilon}_l)(\boldsymbol{v}_p+\boldsymbol{\varepsilon}_p)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_k+\boldsymbol{\varepsilon}_k)$$
$$\times(\boldsymbol{v}_q+\boldsymbol{\varepsilon}_q)^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_k^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big],$$

and

$$K_{26} = \frac{1}{m^2}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m\sum_{l=1}^m E\Big[\boldsymbol{X}_i(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{v}_k+\boldsymbol{\varepsilon}_k)(\boldsymbol{v}_i+\boldsymbol{\varepsilon}_i)^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_j+\boldsymbol{\varepsilon}_j)(\boldsymbol{v}_l+\boldsymbol{\varepsilon}_l)^\top$$
$$\times\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)(\boldsymbol{v}_a+\boldsymbol{\varepsilon}_a)\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\Big].$$

These are of order $O(m^{-3})$ for the similar reason for K_{24}. Hence, K_2 is of order $O(m^{-3/2})$.

Next, we evaluate the expectation of the product of $\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a$ and the second term in (3.5.8), namely $\boldsymbol{D}_aE[(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]$. We can write

$$\begin{aligned}
&E[(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]\\
&=E[(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]+E[(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]\\
&=(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_aE[\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]+(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_aE[\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]\\
&\quad-(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}E[(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{X}_a\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top]+O(m^{-3/2}),
\end{aligned} \tag{3.5.11}$$

since $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})=O(m^{-1})$, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})-\boldsymbol{\beta}=O(m^{-1/2})$ and $\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}=O(m^{-1/2})$. Recall that

$$\begin{aligned}
&\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\\
&=\Big[\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big]\\
&\quad\times\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&\quad+\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m\boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\Big\}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta}).
\end{aligned} \tag{3.5.12}$$

The first term of (3.5.12) can be approximated as

$$\begin{aligned}
&\Big[\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big]\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&=-\Big\{\sum_{j=1}^m\boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^m\boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\\
&\quad\times\Big\{\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m\boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})
\end{aligned}$$

$$
\begin{aligned}
=&\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\\
&\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\\
&\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})+O(m^{-3/2}),
\end{aligned}
$$

and the second term of (3.5.12) can be approximated as

$$
\begin{aligned}
&\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top\Big\{(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}-(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\Big\}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&=-\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\widehat{\boldsymbol{\Psi}}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&=-\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&\quad+\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\\
&\quad\times\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})+O(m^{-3/2}).
\end{aligned}
$$

Then, the expectation in the first term of (3.5.11) is

$$
\begin{aligned}
&E\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi})\}\{\boldsymbol{\theta}_a^*-\boldsymbol{\theta}_a\}^\top\\
&=E\Big[\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\\
&\quad\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}\\
&\quad\times\Big\{\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}\boldsymbol{X}_i\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\Big]\\
&\quad\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}\\
&\quad-E\Big[\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\\
&\quad+\Big\{\sum_{j=1}^m \boldsymbol{X}_j^\top(\boldsymbol{\Psi}+\boldsymbol{D}_j)^{-1}\boldsymbol{X}_j\Big\}^{-1}\\
&\quad\times\sum_{i=1}^m \boldsymbol{X}_i^\top(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})(\boldsymbol{\Psi}+\boldsymbol{D}_i)^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\beta})\Big]\\
&\quad\times\{-\boldsymbol{v}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{D}_a+\boldsymbol{\varepsilon}_a^\top(\boldsymbol{\Psi}+\boldsymbol{D}_a)^{-1}\boldsymbol{\Psi}\}.
\end{aligned}
\tag{3.5.13}
$$

Let $\boldsymbol{S} = \{\sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} \boldsymbol{X}_i\}^{-1}$. The first term of (3.5.13) is

$$\begin{aligned}
&\boldsymbol{S} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m E\Big[\boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} \boldsymbol{X}_i \boldsymbol{S} \boldsymbol{X}_j^\top (\boldsymbol{\Psi} + \boldsymbol{D}_j)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) \\
&\quad \times (\boldsymbol{\Psi} + \boldsymbol{D}_j)^{-1} \boldsymbol{X}_j \boldsymbol{S} \boldsymbol{X}_k^\top (\boldsymbol{\Psi} + \boldsymbol{D}_k)^{-1} (\boldsymbol{y}_k - \boldsymbol{X}_k \boldsymbol{\beta})\Big] \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\},
\end{aligned}$$

and second term of (3.5.13) is

$$\begin{aligned}
&-E\boldsymbol{S} \sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i) \\
&\quad + \boldsymbol{S} \sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i)\Big] \\
&\quad \times \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\}.
\end{aligned}$$

Since $\boldsymbol{S}$ is of order $O(m^{-1})$, these are of order $O(m^{-3/2})$ for the same reason for the second term of (3.5.9). Since $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta} = (\boldsymbol{X}^\top (\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1} \boldsymbol{X})^{-1} \sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})$, the expectation in the second term of (3.5.11) is

$$\begin{aligned}
&E\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top \\
&=(\boldsymbol{X}^\top (\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1} \boldsymbol{X})^{-1} E \sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \\
&=(\boldsymbol{X}^\top (\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} E(\boldsymbol{v}_a + \boldsymbol{\varepsilon}_a) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \\
&=\boldsymbol{0},
\end{aligned}$$

whcich implies that the expectation of the product of $\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a$ and the third term in (3.5.8) is zero, and the expectation in the third term of (3.5.11) is

$$\begin{aligned}
&E(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{X}_a \{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}) - \boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top \\
&=E(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{X}_a (\boldsymbol{X}^\top (\boldsymbol{I}_m \otimes \boldsymbol{\Psi} + \boldsymbol{D})^{-1} \boldsymbol{X})^{-1} \\
&\quad \times \sum_{i=1}^m \boldsymbol{X}_i^\top (\boldsymbol{\Psi} + \boldsymbol{D}_i)^{-1} (\boldsymbol{v}_i + \boldsymbol{\varepsilon}_i) \{-\boldsymbol{v}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + \boldsymbol{\varepsilon}_a^\top (\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{\Psi}\} \\
&=O(m^{-2}),
\end{aligned}$$

for the similar reason for K_1 in (3.5.9). Hence, we have $E\boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_a)^{-1} \boldsymbol{X}_a \{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}) - \boldsymbol{\beta}\}\{\boldsymbol{\theta}_a^* - \boldsymbol{\theta}_a\}^\top = O(m^{-3/2})$. □

3.5.3 Proof of Theorem 3.3.2

From (2) in Theorem 3.2.1, it is sufficient to show this approximation for $\widehat{\boldsymbol{\Psi}}$ instead of $\widehat{\boldsymbol{\Psi}}^+$. Using the equation in (3.5.6), we can rewrite $G_{1a}(\widehat{\boldsymbol{\Psi}})$ as

$$\begin{aligned}
\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}) &= (\widehat{\boldsymbol{\Psi}}^{-1} + \boldsymbol{D}_a^{-1})^{-1} = \boldsymbol{D}_a - \boldsymbol{D}_a(\widehat{\boldsymbol{\Psi}} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a \\
&= \boldsymbol{G}_{1a}(\boldsymbol{\Psi}) + \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a \\
&\quad - \boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1} \boldsymbol{D}_a + O_p(m^{-3/2}).
\end{aligned} \tag{3.5.14}$$

We shall evaluate each term in RHS of the above equality. It is easy to see from (3.5.1) that $E[\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}] = \mathrm{Bias}(\widehat{\boldsymbol{\Psi}})$, which is written as (3.2.7). From Lemma 3.3.1, the expectation of the second term is

$$-E\boldsymbol{D}_a(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})(\boldsymbol{\Psi} + \boldsymbol{D}_a)^{-1}\boldsymbol{D}_a = -\boldsymbol{G}_{3a}(\boldsymbol{\Psi}) + O(m^{-3/2}).$$

The above arguments imply that a second-order unbiased estimator of $\boldsymbol{G}_{1a}(\boldsymbol{\Psi})$ is $\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}^+) + \boldsymbol{G}_{5a}(\widehat{\boldsymbol{\Psi}}^+)$. The estimators $\boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}}^+)$, $\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}^+)$ and $\boldsymbol{G}_{4a}(\widehat{\boldsymbol{\Psi}}^+)$ do not have second-order biases, and the results in Theorem 3.3.2 are established. $\square$

Chapter 4

Multivariate Nested-Error Regression Models

4.1 Motivation

In this chapter, we consider multivariate nested error regression models (MNER) with fixed effests based on a vector of regression coefficients $\boldsymbol{\beta}$ and vectors of random effects $\boldsymbol{v}_i$ and sampling errors $\boldsymbol{\varepsilon}_{ij}$ for the j-th unit in the i-th area. When $\boldsymbol{\theta}_a$, defined by $\boldsymbol{\theta}_a = \boldsymbol{c}_a^\top \boldsymbol{\beta} + \boldsymbol{v}_a$, is a characteristic of interest for the a-th area and constant $\boldsymbol{c}_a$, the Bayes estimator of $\boldsymbol{\theta}_a$ in the Bayesian context is

$$\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \boldsymbol{c}_a^\top \boldsymbol{\beta} + \boldsymbol{B}_a(\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top \boldsymbol{\beta}),$$

for $\boldsymbol{B}_a = \boldsymbol{\Psi}(\boldsymbol{\Psi} + n_a^{-1}\boldsymbol{\Sigma})^{-1}$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ are covariance matrices of $\boldsymbol{\varepsilon}_{ij}$ and $\boldsymbol{v}_i$, respectively, n_a is a size of a sample from the a-th area, and $\overline{\boldsymbol{y}}_a$ and $\overline{\boldsymbol{X}}_a$ are sample means of response variables and the associated explanatory variables in the a-th area. When components of $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_{ij}$ are mutually independent, namely $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are diagonal matrices, it is enough to treat the estimation of each component of $\boldsymbol{\theta}_a$ separately. When components of $\boldsymbol{v}_i$ or $\boldsymbol{\varepsilon}_{ij}$ are correlated each other, however, it could be better to consider the estimation of $\boldsymbol{\theta}_a$ simultaneously. For example, the survey and satellite data of Battese, *et al.* (1988) consist of two crop areas under corn and soybean, and it should be reasonable that the two crop areas are correlated each other.

The multivariate small area estimation has not been studied so much, while most results in small area estimation have been provided in the univariate cases. Fay (1987) proposed a multivariate Fay-Herriot model for analyzing multivariate area-level data. Porter, Wikle and Holan (2015) and Benavent and Morales (2016) suggested multivariate spatial Fay-Herriot models with covariance matrices in which spatial dependence is embedded. Concerning the multivariate nested error regression (MNER) models, Fuller and Harter (1987) obtained the empirical Bayes estimator or the EBLUP and the analytical results for its uncertainty, and Datta, Day and Maiti (1998) developed the fully Bayesian approach. Datta, Day and Basawa (1999) also provided general theoretical results for the multivariate empirical Bayes estimators, but did not give concrete expressions in the fully unknown case of covariance matrices.

The MNER model has the two components of covariance: 'between' component $\boldsymbol{\Psi}$ and 'within' component $\boldsymbol{\Sigma}$. We here use an exact unbiased estimator $\widehat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma}$, and for $\boldsymbol{\Psi}$, we suggest a nonnegative definite and consistent estimator $\widehat{\boldsymbol{\Psi}}$ which is a second-order unbiased estimator of

$\boldsymbol{\Psi}$. For the other estimation methods, see Calvin and Dykstra (1991a, b). Substituting $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Sigma}}$ into $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ in the Bayes estimator $\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})$ and estimating $\boldsymbol{\beta}$ by the generalized least squares estimator $\widehat{\boldsymbol{\beta}}$, one gets the empirical Bayes estimator or EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB} = \widetilde{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})$. We derive analytically a second-order approximation of the MSE matrix of the EBLUP and provide a closed form expression of a second-order unbiased estimator, denoted by $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$, of the MSE matrix of the EBLUP. These results are extensions of the univariate case. It is noted that similar results were given by Fuller and Harter (1987) who considered to estimate $\boldsymbol{B}_a$ nearly unbiasedly, which is slightly different from the approach of this paper.

Another topic addressed in this chapter is the confidence interval problem. As pointed out in Diao, Smith, Datta, Maiti and Opsomer (2014), one difficulty with traditional confidence intervals is that the coverage probabilities do not have second-order accuracy. It is also numerically confirmed that the coverage probabilities are smaller than the nominal confidence coefficient. Diao et al. (2014) suggested the construction of accurate confidence interval based on the EBLUP and the estimator of MSE of EBLUP so that the coverage provability is correct up to second order. For other studies on the confidence interval problem, see Datta, Ghosh, Smith, Lahiri (2002), Basu, Ghosh and Mukerjee (2003), Chatterjee, Lahiri and Li (2008), Kubokawa (2010), Sugasawa and Kubokawa (2015) and Yosimori and Lahiri (2014). In this paper, we consider the confidence interval for the liner combination $\boldsymbol{\ell}^\top \boldsymbol{\theta}_a$ for $\boldsymbol{\ell} \in \mathbb{R}^k$. The naive confidence interval is given by $\boldsymbol{\ell}^\top \widehat{\boldsymbol{\theta}}_a^{EB} \pm z_{\alpha/2} \times \sqrt{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}$ where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. Because this confidence interval does not have second-order accuracy, using similar aguments as in Diao et al. (2014), we construct the closed-form confidence interval whose coverage probability is identical to the nominal confidence coefficient $1-\alpha$ up to second order.

This chapter is organized as follows: In Section 4.2, we probide an exact unbiased estimator $\widehat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma}$ and a nonnegative definit, consistent and second-order unbiased estimator $\widehat{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$. Substituting these estimators into the Bayes estimator yields the empirical Bayes estimator or EBLUP $\widetilde{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})$. In Section 4.3, we derive a second-order approximation of the MSE matrix of the EBLUP and a second-order unbiased estimator of the MSE matrix analytically. Section 4.4 presents the confidence interval with second-order accuracy.. The numerical investigation and emprical studies are given in Section 4.5.

4.2 Empirical Best Linear Unbiased Prediction

In this paper, we assume that data $(\boldsymbol{y}_{ij}, \boldsymbol{X}_{ij})$ for $i = 1 \ldots, m$ and $j = 1, \ldots, n_i$ are observed, where m is the number of small areas, n_i is the number of the subjects in an i-th area such that $\sum_{i=1}^m n_i = N$, $\boldsymbol{y}_{ij}$ is a k-variate vector of direct survey estimates and $\boldsymbol{X}_{ij}$ is a $s \times k$ matrix of covariates associated with $\boldsymbol{y}_{ij}$ for the j-th subject in the i-th area. Then, we assume the multivariate nested-error regression model described as

$$\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i, \tag{4.2.1}$$

where $\boldsymbol{\beta}$ is an s-variate vector of unknown regression coefficients, $\boldsymbol{v}_i$ is a k-variate vector of random effects depending on the i-th area and $\boldsymbol{\varepsilon}_{ij}$ is a k-variate vector of sampling errors. It is assumed that $\boldsymbol{v}_i$ and $\boldsymbol{\varepsilon}_{ij}$ are mutually independently distributed as

$$\boldsymbol{v}_i \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{\Psi}) \quad \text{and} \quad \boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}_k(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are $k \times k$ unknown and nonsingular covariance matrices.

We now express model (4.2.1) in a matrix form. Let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, \ldots, \boldsymbol{y}_{in_i}^\top)^\top$, $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$, $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i})^\top$, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_m^\top)^\top$, $\boldsymbol{\varepsilon}_i = (\boldsymbol{\varepsilon}_{i1}^\top, \ldots, \boldsymbol{\varepsilon}_{in_i}^\top)^\top$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_m^\top)^\top$. Then, model (4.2.1) is expressed as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{1}_{n_i} \otimes \boldsymbol{v}_i + \boldsymbol{\varepsilon}_i, \tag{4.2.2}$$

where $\boldsymbol{1}_{n_i} \otimes \boldsymbol{v}_i \sim \mathcal{N}_{kn_i}(\boldsymbol{0}, \boldsymbol{J}_{n_i} \otimes \boldsymbol{\Psi})$ and $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{kn_i}(\boldsymbol{0}, \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma})$ for $\boldsymbol{J}_{n_i} = \boldsymbol{1}_{n_i}\boldsymbol{1}_{n_i}^\top$.

For the a-th area, we want to predict the quantity $\boldsymbol{\theta}_a = \boldsymbol{c}_a^\top\boldsymbol{\beta} + \boldsymbol{v}_a$, which is the conditional mean $E[\boldsymbol{y}_a \mid \boldsymbol{v}_a]$ given $\boldsymbol{v}_a$ when

$$\boldsymbol{c}_a = \overline{\boldsymbol{X}}_a = n_a^{-1}\sum_{j=1}^{n_a}\boldsymbol{X}_{aj}.$$

A reasonable estimator can be derived from the conditional expectation $E[\boldsymbol{\theta}_a \mid \boldsymbol{y}_a] = \boldsymbol{c}_a^\top\boldsymbol{\beta} + E[\boldsymbol{v}_a \mid \boldsymbol{y}_a]$. The conditional distribution of $\boldsymbol{v}_i$ given $\boldsymbol{y}_i$ and the marginal distribution of $\boldsymbol{y}_i$ are

$$\begin{aligned} \boldsymbol{v}_i \mid \boldsymbol{y}_i \sim& \mathcal{N}_k(\tilde{\boldsymbol{v}}_i(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}), (\boldsymbol{\Psi}^{-1} + n_i\boldsymbol{\Sigma}^{-1})^{-1}), \\ \boldsymbol{y}_i \sim& \mathcal{N}_{kn_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{J}_{n_i} \otimes \boldsymbol{\Psi} + \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}), \end{aligned} \quad i = 1, \ldots, m, \tag{4.2.3}$$

where

$$\tilde{\boldsymbol{v}}_i(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \boldsymbol{\Psi}(\boldsymbol{\Psi} + n_i^{-1}\boldsymbol{\Sigma})^{-1}(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{X}}_i^\top\boldsymbol{\beta}), \tag{4.2.4}$$

where $\overline{\boldsymbol{y}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}$. Thus, we get the estimator

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& \boldsymbol{c}_a^\top\boldsymbol{\beta} + E[\boldsymbol{v}_a \mid \boldsymbol{y}_a] = \boldsymbol{c}_a^\top\boldsymbol{\beta} + \tilde{\boldsymbol{v}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) \\ =& \boldsymbol{c}_a^\top\boldsymbol{\beta} + \boldsymbol{\Psi}(\boldsymbol{\Psi} + n_a^{-1}\boldsymbol{\Sigma})^{-1}(\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta}), \end{aligned} \tag{4.2.5}$$

which corresponds to the Bayes estimator of $\boldsymbol{\theta}_a$ in the Bayesian framework.

When $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are known, the maximum likelihood estimator or generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) = (\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y},$$

where $\boldsymbol{D} = \text{block diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m)$ and $\boldsymbol{D}_i = \boldsymbol{J}_{n_i} \otimes \boldsymbol{\Psi} + \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}$ for $i = 1, \ldots, m$. Substituting $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ into $\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})$ yields the estimator

$$\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \boldsymbol{c}_a^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + \boldsymbol{\Psi}(\boldsymbol{\Psi} + n_a^{-1}\boldsymbol{\Sigma})^{-1}(\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})). \tag{4.2.6}$$

It can be easily verified that this estimator is the best linear unbiased predictor (BLUP) of $\boldsymbol{\theta}_a$.

We provide consistent estimators of the covariance components $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$. Concerning estimation of $\boldsymbol{\Sigma}$, it is noted that $E[\{\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i - (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)^\top\boldsymbol{\beta}\}\{\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i - (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)^\top\boldsymbol{\beta}\}^\top] = (1 - n_i^{-1})\boldsymbol{\Sigma}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, which implies that $\sum_{i=1}^m\sum_{j=1}^{n_i}E[\{\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i - (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)^\top\boldsymbol{\beta}\}\{\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i - (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_i)^\top\boldsymbol{\beta}\}^\top] = (N - m)\boldsymbol{\Sigma}$. Let $\tilde{\boldsymbol{y}}_i = ((\boldsymbol{y}_{i1} - \overline{\boldsymbol{y}}_i)^\top, \ldots, (\boldsymbol{y}_{in_i} - \overline{\boldsymbol{y}}_i)^\top)^\top$, $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{y}}_1^\top, \ldots, \tilde{\boldsymbol{y}}_m^\top)^\top$,

$\widetilde{\boldsymbol{X}}_i = (\boldsymbol{X}_{i1} - \overline{\boldsymbol{X}}_i, \ldots, \boldsymbol{X}_{in_i} - \overline{\boldsymbol{X}}_i)^\top$ and $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{X}}_1^\top, \ldots, \widetilde{\boldsymbol{X}}_m^\top)^\top$. Substituting the statistic $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}})^{-1} \widetilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}$ into $\boldsymbol{\beta}$, we get an unbised estimator of the form

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{N - m - s_0} (\tilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}})^\top, \tag{4.2.7}$$

where s_0 is the rank of $\widetilde{\boldsymbol{X}}$. For estimation of $\boldsymbol{\Psi}$, it is noted that $E[(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top \boldsymbol{\beta})(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top \boldsymbol{\beta})^\top] = \boldsymbol{\Psi} + \boldsymbol{\Sigma}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, which implies that $\sum_{i=1}^m \sum_{j=1}^{n_i} E[(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top] = N(\boldsymbol{\Psi} + \boldsymbol{\Sigma})$. Substituting the ordinary least squares estimator $\widehat{\boldsymbol{\beta}}^{OLS} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ and $\widehat{\boldsymbol{\Sigma}}$ into $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we get the consistent estimator

$$\widehat{\boldsymbol{\Psi}}_0 = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top \widehat{\boldsymbol{\beta}}^{OLS})(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top \widehat{\boldsymbol{\beta}}^{OLS})^\top - \widehat{\boldsymbol{\Sigma}}. \tag{4.2.8}$$

Taking the expectation of $\widehat{\boldsymbol{\Psi}}_0$, we can see that $E[\widehat{\boldsymbol{\Psi}}_0] = \boldsymbol{\Psi} + \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi})$, where

$$\begin{aligned} \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} \boldsymbol{X}_{ij}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X} \boldsymbol{D}^{-1} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_{ij} \\ & - \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} (\boldsymbol{\Sigma} \boldsymbol{X}_{ij}^\top - n_i \boldsymbol{\Psi} \overline{\boldsymbol{X}}_i^\top)(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_{ij} \\ & - \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} \boldsymbol{X}_{ij}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} (\boldsymbol{X}_{ij} \boldsymbol{\Sigma} - n_i \overline{\boldsymbol{X}}_i \boldsymbol{\Psi}), \end{aligned} \tag{4.2.9}$$

where $\boldsymbol{D} = \text{block diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_m)$ for $\boldsymbol{D}_i = \boldsymbol{J}_{n_i} \otimes \boldsymbol{\Psi} + \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma}$, $i = 1, \ldots, m$. Let $\widehat{\boldsymbol{\Psi}}_1 = \widehat{\boldsymbol{\Psi}}_0 - \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\widehat{\boldsymbol{\Psi}}_0, \widehat{\boldsymbol{\Sigma}})$. Then, $\widehat{\boldsymbol{\Psi}}_1$ is a second-order unbiased estimator of $\boldsymbol{\Psi}$. Because $\widehat{\boldsymbol{\Psi}}_1$ takes a negative value, we modify it as

$$\widehat{\boldsymbol{\Psi}} = \boldsymbol{H} \mathrm{diag}\{\max(\lambda_1, 0), \ldots, \max(\lambda_k, 0)\} \boldsymbol{H}^\top, \tag{4.2.10}$$

where $\boldsymbol{H}$ is an orthogonal matrix such that $\widehat{\boldsymbol{\Psi}}_1 = \boldsymbol{H} \mathrm{diag}(\lambda_1, \ldots, \lambda_k) \boldsymbol{H}^\top$.

The consistency of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Psi}}$ can be shown under the assumptions:

(A1) The number of areas m tends to infinity, and k, s and n_i's are bounded with respect to m.

(A2) $\boldsymbol{X}^\top \boldsymbol{X}$ is nonsingular and $\boldsymbol{X}^\top \boldsymbol{X}/m$ converges to a positive definite matrix.

Theorem 4.2.1 *Assume conditions* (A1) *and* (A2)*. Then, the following asymptotic properties hold for* $\widehat{\boldsymbol{\Sigma}}$ *and* $\widehat{\boldsymbol{\Psi}}$*:*

(1) $\widehat{\boldsymbol{\Psi}}$ *is a second-order unbiased estimator of* $\boldsymbol{\Psi}$*, while* $\widehat{\boldsymbol{\Sigma}}$ *is an unbiased estimator of* $\boldsymbol{\Sigma}$*.*

(2) $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = O_p(m^{-1/2})$*,* $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi} = O_p(m^{-1/2})$ *and* $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\beta} = O_p(m^{-1/2})$*.*

(3) *For any* $\delta > 0$*,* $P(\widehat{\boldsymbol{\Psi}} \neq \widehat{\boldsymbol{\Psi}}_1) = O(m^{-\delta})$*.*

The proof is given in the section 4.6 Since $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Psi}}$ are consistent, we can substitute them into (4.2.6) to get the empirical best linear unbiased predictor (EBLUP)

$$\widehat{\boldsymbol{\theta}}_a^{EB} = \boldsymbol{c}_a^\top \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) + \widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\Psi}} + n_a^{-1} \widehat{\boldsymbol{\Sigma}})^{-1} (\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})). \tag{4.2.11}$$

4.3 Evaluation of Uncertainty of EBLUP

The EBLUP suggested in (4.2.11) is expected to have a small estimation error, and it is important to measure how much the estimation error is. In this section, we derive a second-order approximation of the mean squared error matrix (MSEM) of the EBLUP and provide a second-order unbiased estimator of the MSEM. The MSEM of the EBLUP is $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a\}^\top]$. It is noted that

$$\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a = \{\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \boldsymbol{\theta}_a\} + \{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})\} + \{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\},$$

where $\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})$ and $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ are given in (4.2.5) and (4.2.6). The following lemma which will proved in the section 4.6 is useful for evaluating the mean square error matrix.

Lemma 4.3.1 *$\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ is independent of $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}$ and $\tilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}}$, which implies that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ is independent of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Psi}}$. Also, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ is independent of $\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$.*

Noting that $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \{\boldsymbol{c}_a^\top - \boldsymbol{\Psi}(\boldsymbol{\Psi} + n_a^{-1}\boldsymbol{\Sigma})^{-1}\overline{\boldsymbol{X}}_a^\top\}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \boldsymbol{\beta}\}$, from Lemma 4.3.1, we can decompose the MSEM as

$$\begin{aligned}\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) =& E[\{\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \boldsymbol{\theta}_a\}\{\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \boldsymbol{\theta}_a\}^\top] \\ &+ E[\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\Sigma})\}^\top] \\ &+ E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\}^\top] \\ =& \boldsymbol{G}_{1a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\}^\top],\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{G}_{1a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& (\boldsymbol{\Psi}^{-1} + n_a\boldsymbol{\Sigma}^{-1})^{-1} = n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}, \\ \boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& (\boldsymbol{c}_a^\top - \boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top)(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{c}_a - \overline{\boldsymbol{X}}_a\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}),\end{aligned} \tag{4.3.1}$$

for $\boldsymbol{\Lambda}_a = \boldsymbol{\Psi} + n_a^{-1}\boldsymbol{\Sigma}$. In the following theorem which will be proved in the section 4.6, we approximate the third term as

$$\begin{aligned}\boldsymbol{G}_{3a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& \frac{n_a^{-2}}{N^2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m n_i^2\Big\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i + \mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\Big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} \\ &+ \frac{n_a^{-2}}{N^2(N-m)}(N\boldsymbol{\Psi} + m\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\Big\{\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + \mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma})\boldsymbol{\Sigma}\Big\}\boldsymbol{\Lambda}_a^{-1}(N\boldsymbol{\Psi} + m\boldsymbol{\Sigma}).\end{aligned} \tag{4.3.2}$$

Theorem 4.3.1 *The mean squared error matrix of the empirical Bayes estimator $\widehat{\boldsymbol{\theta}}_a^{EB}$ is approximated as*

$$\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + \boldsymbol{G}_{3a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + O(m^{-3/2}). \tag{4.3.3}$$

We next provide a second-order unbiased estimator of the mean squared error matrix of the EBLUP. A naive estimator of $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is the plug-in estimator of (4.3.3) given by $\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})+\boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})+\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})$, but this has a second-order bias, because $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma})+O(m^{-1})$. Correcting this second-order bias, we can derive the second-order unbiased estimator

$$\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB}) = \boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}}) + \boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}}) + 2\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}}). \tag{4.3.4}$$

Theorem 4.3.2 *Under the coditions* (A1) *and* (A2)*, it holds that* $E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})+\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})] = \boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) + O(m^{-3/2})$ *and*

$$E[\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})] = \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) + O(m^{-3/2}),$$

namely, $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ *is a second-order unbiased estimator of* $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$*.*

4.4 Confidence Interval for Linear Combination of EBLUP with Corrected Coverage Probability

In this section, we consider the confidence interval of the liner combination $\boldsymbol{\ell}^\top\boldsymbol{\theta}_a$ for $\boldsymbol{\ell}\in\mathbb{R}^k$ for the a-th area in the MNER.

We begin by estimating the linear combination $\boldsymbol{\ell}^\top\boldsymbol{\theta}_a = \boldsymbol{\ell}^\top(\boldsymbol{c}_a^\top\boldsymbol{\beta}+\boldsymbol{v}_a)$, which is the conditional mean $E[\boldsymbol{\ell}^\top\boldsymbol{y}_a \mid \boldsymbol{v}_a]$ given $\boldsymbol{v}_a$. A reasonable estimator is provided by the conditional expectation $E[\boldsymbol{\ell}^\top\boldsymbol{\theta}_a \mid \boldsymbol{y}_a] = \boldsymbol{\ell}^\top\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma})$, where $\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is given by (4.2.5). By replacing $\boldsymbol{\beta}$ with the generalized least estimator $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) = (\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}$, the BLUP of $\boldsymbol{\ell}\boldsymbol{\theta}_a$ is provided by $\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})$, where $\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is given in (4.2.6). Substituting (4.2.7) and (4.2.10) into the BLUP yields the EBLUP $\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB}$ for $\widehat{\boldsymbol{\theta}}_a^{EB}$ given in (4.2.11). The mean squared error is $E[(\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\ell}^\top\boldsymbol{\theta}_a)^2] = \boldsymbol{\ell}^\top\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}$ and and its second-order unbiased estimator is $\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}$, namely

$$E[\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}] = E[(\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\ell}^\top\boldsymbol{\theta}_a)^2] + o(m^{-1}),$$

where $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$ and $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ are given in (4.3.3) and (4.3.4).

We now construct the confidence interval. The naive confidence interval is given by

$$I^{NCI}\ :\ \boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} \pm z_{\alpha/2}\times\sqrt{\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}, \tag{4.4.1}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. However, this confidence interval does not have the second-order accuracy, namely $P(\boldsymbol{\ell}^\top\boldsymbol{\theta}_a \in I^{NCI}) = 1-\alpha+O(m^{-1})$. To derive a confidence interval with second-order accuracy, we need to evaluate the second moment of $\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}$. Let

$$\begin{aligned}
V(\widehat{\boldsymbol{\theta}}_a^{EB}) =& \frac{n_a^{-4}}{N^2}\sum_{i=1}^m n_i^2\Big\{(\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i\boldsymbol{\ell})^2 + \boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\times\boldsymbol{\ell}^\top\boldsymbol{\Lambda}_i\boldsymbol{\ell}\Big\} \\
&+ \frac{2n_a^{-4}m^2}{N^2(N-m)}(\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell})^2 + \frac{2n_a^{-2}}{N-m}(\boldsymbol{\ell}^\top\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\boldsymbol{\ell})^2 \\
&- \frac{2n_a^{-3}m}{N(N-m)}\Big\{\boldsymbol{\ell}^\top\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\boldsymbol{\ell}\times\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell} + (\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\boldsymbol{\ell})^2\Big\}.
\end{aligned} \tag{4.4.2}$$

Lemma 4.4.1 *Under the coditions* (A1) *and* (A2), *it holds that*

$$E\Big[\Big\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\Big\}^2\Big] = V(\widehat{\boldsymbol{\theta}}_a^{EB}) + o(m^{-1}),$$

and for $c \geq 3$,

$$E\Big[\Big\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\Big\}^c\Big] = o(m^{-1}).$$

Theorem 4.4.1 *Under the coditions* (A1) *and* (A2), *it holds that for any* z,

$$\begin{aligned}&P\Big(\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} - \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}z \leq \boldsymbol{\ell}^\top\boldsymbol{\theta}_a \leq \boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} + \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}z\Big)\\ =&2\Phi(z) - 1 - \frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}(z^3+z)\phi(z) + o(m^{-1}),\end{aligned}$$

where $\Phi(\cdot)$ *and* $\phi(\cdot)$ *are the distribution and density functions of the standard normal distribution.*

Solving the equation

$$2\Phi(z) - 1 - \frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}(z^3+z)\phi(z) = 1-\alpha,$$

we get the solution given by

$$z^* = z_{\alpha/2} + (z_{\alpha/2}^3 + z_{\alpha/2})V(\widehat{\boldsymbol{\theta}}_a^{EB})/8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2,$$

which provides the improved confidence interval

$$I^{ICI}\ :\ \boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} \pm \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}z^*. \tag{4.4.3}$$

Then from Theorem 4.4.1, it follows that $P(\boldsymbol{\ell}^\top\boldsymbol{\theta}_a \in I^{ICI}) = 1-\alpha+o(m^{-1})$.

4.5 Simulation and Empirical Studies

4.5.1 Finite sample performances

We now investigate finite sample performances of EBLUP in terms of MSEM and the second-order unbiased estimator of MSEM by simulation.

[1] Setup of simulation experiments. We treat the multivariate Nested-Error model,

$$\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij},\quad i=1,\ldots,m,\quad j=1,\ldots,n_i.$$

We take $m=40$, $k=2,3$ and $\boldsymbol{\beta}=(0.8,-0.5,-0.3,0.6)^\top$ for $k=2$ and $\boldsymbol{\beta}=(0.8,-0.5,-0.3,0.6,0.4,-0.2)^\top$ for $k=3$. Moreover, we equally divided areas into four groups ($G=1,\ldots,4$), so that each group has ten areas and the areas in the same group has the same sample size $n_G = 3G - 2$. The design matrix, $\boldsymbol{X}_{ij}$ is $2k \times k$ matrix, such that

$$\boldsymbol{X}_{ij} = \begin{pmatrix} 1 & x_{i1} & 0 & 0 \\ 0 & 0 & 1 & x_{i2} \end{pmatrix}^\top, \boldsymbol{X}_{ij} = \begin{pmatrix} 1 & x_{i1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{i2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{i3} \end{pmatrix}^\top$$

for $k = 2, 3$ respectevely. We generate x_{ij} from uniform distribution on $(-1, 1)$, which are fixed through the simulation runs. As a setup of the covariance matrix $\boldsymbol{\Psi}$ of the random effects, we consider

$$\boldsymbol{\Psi} = \begin{cases} \rho\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top + (1-\rho)\mathrm{diag}(\boldsymbol{\psi}_2\boldsymbol{\psi}_2^\top) & \text{for } k = 2, \\ \rho\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top + (1-\rho)\mathrm{diag}(\boldsymbol{\psi}_3\boldsymbol{\psi}_3^\top) & \text{for } k = 3, \end{cases}$$

where $\boldsymbol{\psi}_2 = (\sqrt{1.5}, \sqrt{0.5})^\top$, $\boldsymbol{\psi}_3 = (\sqrt{1.5}, 1, \sqrt{0.5})^\top$, and $\mathrm{diag}(\boldsymbol{A})$ denotes the diagonal matrix consisting of diagonal elements of matrix $\boldsymbol{A}$. Here, ρ is the correlation coefficient, and we handle the three cases $\rho = 0.25, 0.5, 0.75$. The cases of negative correlations are omitted, because we observe the same results with those of positive ones. Concerning the dispersion matrices $\boldsymbol{\Sigma}$ of sampling errors $\boldsymbol{\varepsilon}_i$, we set $\boldsymbol{\Sigma} = \boldsymbol{I}_k$. We consider three patterns of distribution of $\boldsymbol{v}_i$, that is, M1: $\boldsymbol{v}_i$ is normally distributed, M2: $\boldsymbol{v}_i$ follows multivariate t distribution with degrees of freedom 5 and M3: $\boldsymbol{v}_i$ follows multivariate chi-squared distribution with degrees of freedom 2. The distribution of ε_i is normal.

[2] Comparison of MSEM. We begin with obtaining the true mean squared error matrices of the EBLUP $\widehat{\boldsymbol{\theta}}_a^{EB} = \widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})$ by simulation. Let $\{\boldsymbol{y}_i^{(r)}, i = 1, \ldots, m\}$ be the simulated data in the r-th replication for $r = 1, \ldots, R$ with $R = 50,000$. Let $\widehat{\boldsymbol{\Psi}}^{(r)}$, $\widehat{\boldsymbol{\Sigma}}^{(r)}$ and $\boldsymbol{\theta}_a^{(r)}$ be the values of $\widehat{\boldsymbol{\Psi}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\theta}_a = \overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta} + \boldsymbol{v}_a$ in the r-th replication. Then the simulated value of the true mean squared error matrices is calculated by

$$\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = R^{-1}\sum_{i=1}^{R}\big\{\widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}^{(r)}, \widehat{\boldsymbol{\Sigma}}^{(r)}) - \boldsymbol{\theta}_a^{(r)}\big\}\big\{\widehat{\boldsymbol{\theta}}_a(\widehat{\boldsymbol{\Psi}}^{(r)}, \widehat{\boldsymbol{\Psi}}^{(r)}) - \boldsymbol{\theta}_a^{(r)}\big\}^\top. \tag{4.5.1}$$

To measure relative improvement of EBLUP, we calculate the percentage relative improvement in the average loss (PRIAL) of $\widehat{\boldsymbol{\theta}}_a^{EB}$ over $\boldsymbol{y}_a$, defined by

$$\mathrm{PRIAL}(\widehat{\boldsymbol{\theta}}_a^{EB}, \boldsymbol{y}_a) = 100 \times \Big[1 - \frac{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\}}{\mathrm{tr}\,\{\mathrm{MSEM}(\boldsymbol{y}_a)\}}\Big].$$

It is also interesting to compare $\widehat{\boldsymbol{\theta}}_a^{EB}$ with the EBLUP $\widehat{\boldsymbol{\theta}}_a^{uEB}$ derived from the univariate Nestd-Error model. Thus, we calculate the PRIAL given by

$$\mathrm{PRIAL}(\widehat{\boldsymbol{\theta}}_a^{EB}, \widehat{\boldsymbol{\theta}}_a^{uEB}) = 100 \times \Big[1 - \frac{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\}}{\mathrm{tr}\,\{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{uEB})\}}\Big],$$

and those values are reported in Figure 4.1 and 4.2.

Figure 4.1 reports the PRIAL for $k = 2$ and three patterns of distribution of $\boldsymbol{v}_i$; M1, M2 and M3. We can see that the performances of $\widehat{\boldsymbol{\theta}}_a^{EB}$ are stable regardless of the distribution of $\boldsymbol{v}_i$. In all the cases, $\widehat{\boldsymbol{\theta}}_a^{EB}$ improves on $\boldsymbol{y}_a$ largely and the improvement rates are larger for larger ρ; for normal case (M1). On the other hand, $\widehat{\boldsymbol{\theta}}_a^{EB}$ improves on $\widehat{\boldsymbol{\theta}}_a^{uEB}$ for large ρ, but the univariate EBLUP $\widehat{\boldsymbol{\theta}}_a^{uEB}$ is slightly better than $\widehat{\boldsymbol{\theta}}_a^{EB}$ for $\rho = 0.25$ for some areas, but the difference is not significant. This is because the low accuracy in estimation of the covariance matrix $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ has more adverse influence on prediction than the benefit from incorporating the small correlation into the estimation. Moreover, the PRIAL is larger for the groups with small sample size. This is reasonable because the benefit given by incorporating the information from neibouring areas is large for such groups.

Figure 4.2 reports the PRIAL for $k = 3$ and a pattern of distribution of $\boldsymbol{v}_i$; M1. The results are almost the same with the case for $k = 2$. The PRIAL is larger for $k = 3$ than for $k = 2$ in the case of $\rho = 0.75$, but smaller in the case of $\rho = 0.25$. This is because when m is fixed as $m = 40$, the accuracy in estimation of the covariance matrices gets smaller for the larger dimension.

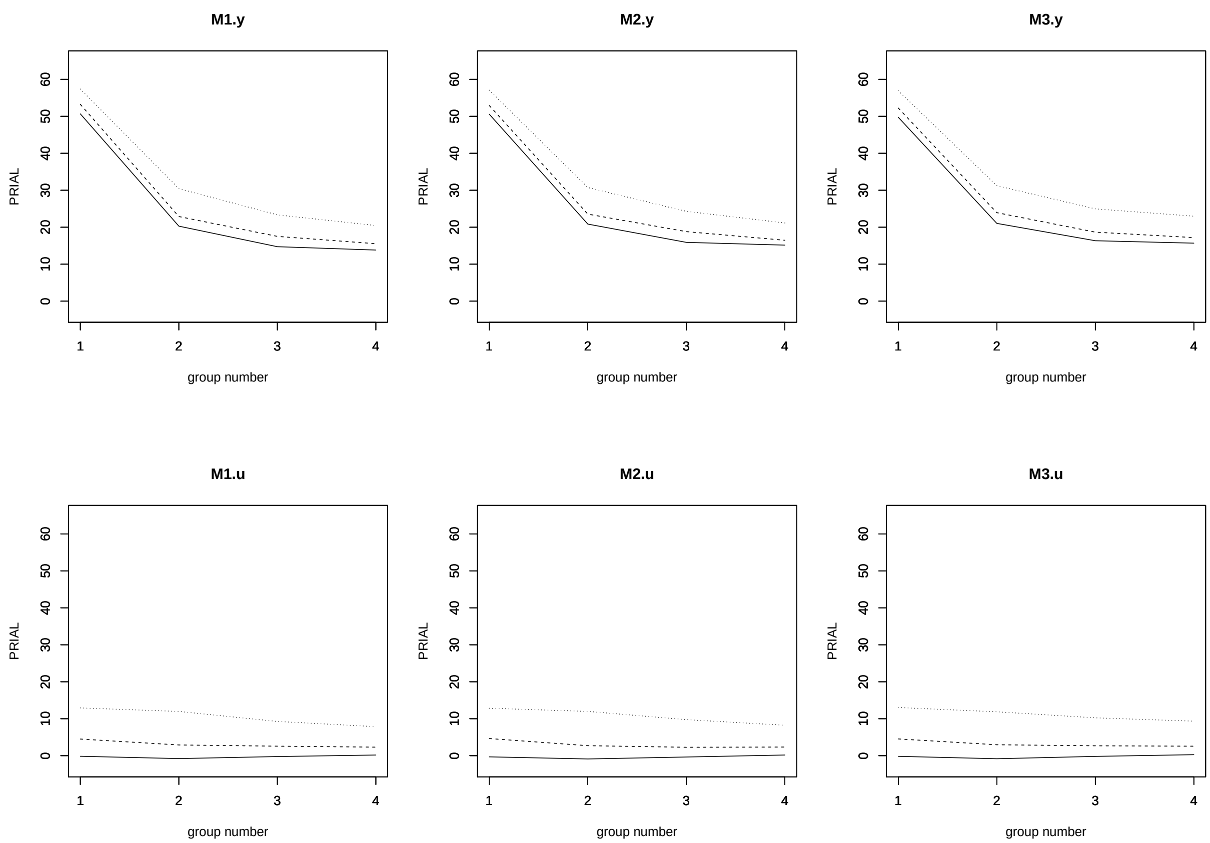

Figure 4.1: PRIAL for $\rho = 0.25$ (real line), $\rho = 0.5$ (dashed line) and $\rho = 0.75$ (dotted line) in case of $k = 2$.

[3] Finite sample performances of the MSEM estimator. We next investigate the performance of the second-order unbiased estimator $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ of MSEM given in Theorem

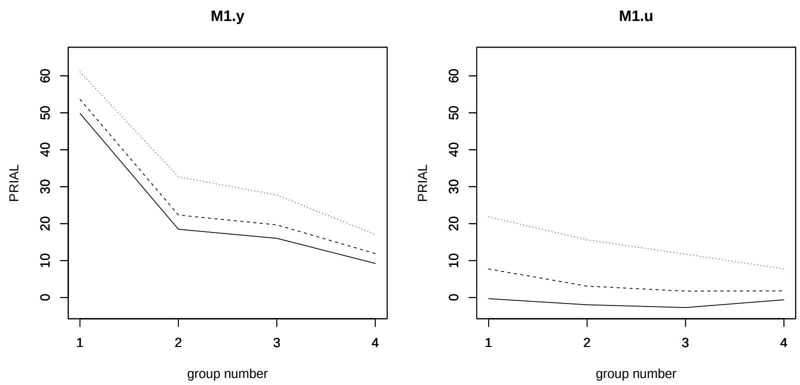

Figure 4.2: PRIAL for $\rho = 0.25$ (real line), $\rho = 0.5$ (dashed line) and $\rho = 0.75$ (dotted line) in case of $k = 3$.

4.3.2. We use the same data generating process as mentioned above and we take only $k = 2$. We consider the normal case (M1) as a pattern of distributions for $\boldsymbol{v}_i$. The simulated values of the MSEM are obtained from (4.5.1) based on $R = 50,000$ simulation runs. Then, based on $R = 5,000$ simulation runs, we calculate the relative bias (RB) of MSEM estimators given by

$$\mathrm{RB}_a = \frac{1}{R}\sum_{r=1}^{R} \frac{\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB(r)}) - \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})}{\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})}$$

where $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB(r)})$ is the MSEM estimator in the r-th replication. In Table 4.1, we report mean values of RB_a in each group. For comparison, results for the naive MSEM estimator, without any bias correction, are reported in Table 4.1 as well. The naive MSEM estimator is the plug-in estimator of the asymptotic MSEM (4.3.3). The relative bias is small for the diagonal elements, less than 10% in almost the cases, whereas considerably large for off-diagonal elements. The naive MSEM estimator is more biased than the analytical MSEM estimator for diagonal elements in all cases, so that the bias correction in MSEM estimator is successful. On the other hand, the analytical MSEM estimator is more biased slightely than the naive MSEM estimator for off-diagonal elements in some cases.

[4] **Finite sample performances of the confidence interval**. We investigate the performance of the improved confidence interval given in (4.4.3). Table 4.2 reports values of coverage probabilities (CP) and average length (AL) for $1 - \alpha = 95\%$ confidence coefficient, where the setup of the simulation experiment is the same as above, namely the three patterns of distributions of $\boldsymbol{v}_i$, M1, M2 and M3 and the three cases of $\rho = 0.25, 0.5, 0.75$ are treated. Table 4.2 also reports values of CP and AL in parentheses for the naive confidence interval (4.4.1).

For all patterns of distributions of $\boldsymbol{v}_i$ and correlation coefficients, values of CP are close to the nominal level of 0.95 and are higher than those for the naive method, especially for areas with small sample size. This is coincident with Diao et al. (2014), which considered the confidence interval estimator under the Fay-Herriot model. Values of CP for areas with large sample sizes

	$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$	
	RB	NRB	RB	NRB	RB	NRB
G_1	$\begin{bmatrix} -3.5 & 0.8 \\ 0.8 & -8.1 \end{bmatrix}$	$\begin{bmatrix} -6.5 & -0.6 \\ -0.6 & -12.2 \end{bmatrix}$	$\begin{bmatrix} -4.6 & -1.8 \\ -1.8 & -7.7 \end{bmatrix}$	$\begin{bmatrix} -7.7 & -3.2 \\ -3.2 & -11.8 \end{bmatrix}$	$\begin{bmatrix} -3.5 & -2.7 \\ -2.7 & -7.6 \end{bmatrix}$	$\begin{bmatrix} -6.3 & -4.3 \\ -4.3 & -11.7 \end{bmatrix}$
G_2	$\begin{bmatrix} -1.9 & 26.7 \\ 26.7 & -4.1 \end{bmatrix}$	$\begin{bmatrix} -4.0 & 36.1 \\ 36.1 & -8.8 \end{bmatrix}$	$\begin{bmatrix} -0.5 & 0.9 \\ 0.9 & -3.1 \end{bmatrix}$	$\begin{bmatrix} -2.8 & 8.1 \\ 8.1 & -8.5 \end{bmatrix}$	$\begin{bmatrix} -0.6 & 3.7 \\ 3.7 & -4.7 \end{bmatrix}$	$\begin{bmatrix} -3.3 & 9.8 \\ 9.8 & -11.4 \end{bmatrix}$
G_3	$\begin{bmatrix} -0.7 & 54.8 \\ 54.8 & -2.7 \end{bmatrix}$	$\begin{bmatrix} -2.3 & 77.1 \\ 77.1 & -6.9 \end{bmatrix}$	$\begin{bmatrix} -0.4 & 19.3 \\ 19.3 & -2.1 \end{bmatrix}$	$\begin{bmatrix} -2.3 & 37.5 \\ 37.5 & -7.2 \end{bmatrix}$	$\begin{bmatrix} -1.7 & 9.9 \\ 9.9 & -2.9 \end{bmatrix}$	$\begin{bmatrix} -4.3 & 26.9 \\ 26.9 & -10.6 \end{bmatrix}$
G_4	$\begin{bmatrix} -0.2 & 30.2 \\ 30.2 & -0.5 \end{bmatrix}$	$\begin{bmatrix} -1.5 & 56.5 \\ 56.5 & -4.2 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 5.6 \\ 5.6 & -0.3 \end{bmatrix}$	$\begin{bmatrix} -1.1 & 29.9 \\ 29.9 & -5.0 \end{bmatrix}$	$\begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.6 \end{bmatrix}$	$\begin{bmatrix} -2.3 & 28.1 \\ 28.1 & -7.8 \end{bmatrix}$

Table 4.1: The Mean Values of Percentage Relative Bias in Each Group (RB) and Relative Bias of Naive MSE Estimator (RBN).

are slightly higher than those for the naive method, but the differences are negligibly small. Values of AL are also larger than those for the naive method for areas with small sample size, and the diferrence is negligible for areas with large sample size.

		Normal			t			chi-square		
ρ		0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
G_1	CP	0.949	0.950	0.957	0.946	0.945	0.950	0.940	0.940	0.943
		(0.936)	(0.933)	(0.928)	(0.935)	(0.929)	(0.923)	(0.927)	(0.924)	(0.918)
	AL	3.700	3.330	2.768	3.655	3.279	2.733	3.615	3.239	2.705
		(3.501)	(3.108)	(2.512)	(3.459)	(3.073)	(2.472)	(3.412)	(3.023)	(2.434)
G_2	CP	0.947	0.943	0.941	0.945	0.941	0.937	0.941	0.940	0.935
		(0.945)	(0.940)	(0.937)	(0.943)	(0.940)	(0.932)	(0.939)	(0.938)	(0.931)
	AL	2.340	2.239	1.978	2.357	2.218	1.948	2.339	2.197	1.924
		(2.364)	(2.226)	(1.950)	(2.343)	(2.204)	(1.925)	(2.321)	(2.177)	(1.898)
G_3	CP	0.947	0.947	0.946	0.947	0.947	0.947	0.948	0.946	0.945
		(0.946)	(0.946)	(0.944)	(0.947)	(0.946)	(0.945)	(0.947)	(0.944)	(0.943)
	AL	1.920	1.845	1.706	1.905	1.833	1.687	1.896	1.822	1.673
		(1.909)	(1.839)	(1.690)	(1.897)	(1.826)	(1.676)	(1.886)	(1.811)	(1.661)
G_4	CP	0.949	0.949	0.951	0.951	0.949	0.951	0.950	0.951	0.952
		(0.947)	(0.948)	(0.949)	(0.950)	(0.948)	(0.950)	(0.949)	(0.950)	(0.950)
	AL	1.653	1.607	1.531	1.643	1.600	1.520	1.639	1.595	1.514
		(1.644)	(1.602)	(1.520)	(1.636)	(1.594)	(1.512)	(1.630)	(1.586)	(1.505)

Table 4.2: Coverage probabilities (CP) and coverage length (AL) for nominal 95% confidence intervals.

4.5.2 Illustrative example

This example, primarily for illustration, uses the multivariate Nested-Error regression model (4.2.1) and data from the posted land price data along the Keikyu train line from 1998 to 2001. This train line connects the suburbs in the Kanagawa prefecture to the Tokyo metropolitan area. Those who live in the suburbs in the Kanagawa prefecture take this line to work or study in Tokyo everyday. Thus, it is expected that the land price depends on the distance from Tokyo.

The posted land price data are available for 53 stations on the Keikyu train line, and we consider each station as a small area, namely, $m = 53$.

For the i-th station, data of n_i land spots are available, where n_i varies around 4 and some areas have only one observation. For $i = 1, \ldots, m$, observations $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, y_{ij3})^\top$ denotes the difference between the value of the posted land price (Yen/1,000) for the unit meter squares of the j-th spot from 1998 to 2001, where y_{ij1} is the a difference between 1998 and 1999, y_{ij2} is the a difference between 1999 and 2000 and y_{ij3} is the a difference between 2000 and 2001. As auxiliary variables, we use the data (T_i, D_{ij}, FAR_{ij}). T_i is the time to take from the nearby station i to the Tokyo station around 8:30 in the morning, D_{ij} is the value of geographical distance from the spot j to the station i and FAR_{ij} denotes the floor-area ratio, or ratio of building volume to lot area of the spot j. Then the regressor in the model (4.2.1) is

$$\boldsymbol{X}_{ij} = \begin{pmatrix} 1 & FAR_{ij} & T_i & D_{ij} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & FAR_{ij} & T_i & D_{ij} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & FAR_{ij} & T_i & D_{ij} \end{pmatrix}^\top.$$

The estimates of the covariance matrix $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 43.4 & 27.3 & 28.4 \\ 27.3 & 33.4 & 20.4 \\ 28.4 & 20.4 & 28.5 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 169.2 & 127.3 & 101.0 \\ 127.3 & 113.5 & 85.3 \\ 101.0 & 85.3 & 77.6 \end{pmatrix}.$$

Thus, the estimated correlation coefficient of random effects $\boldsymbol{\rho} = (\rho_{12}, \rho_{13}, \rho_{23})^\top$ is $(0.72, 0.81, 0.66)$, where ρ_{ab} is the correlation coefficient of v_a and v_b. The estimates of the regression coefficients are $\widehat{\boldsymbol{\beta}} = (-4.28, 16.67, -1.79, -0.13, 6.32, 13.16, -2.34, -0.02, -4.22, 11.16, -0.33, -0.06)^\top$.

All the estimated values of regression coefficients of T_i and D_{ij} are negative values which leads to the natural result that the T_i and D_{ij} have negative influence on y_{ij}, whose magnitude are almost unchanged for three years. On the other hand, the magnitude of the influence of FAR_{ij} on y_{ij} decreases during the same time. The obtained values of EBLUP for a difference between the posted land price data in 2000 and 2001 given in (4.2.11) are give in Table 4.3 for selected 15 areas. To see the difference of predicted values of MNER and NER, Figure 4.3 reports the difference between the degree of shrinkage, which is caluculated by $|\mathrm{dif}(\widehat{\boldsymbol{\theta}}^{EB}) - \mathrm{dif}(\widehat{\boldsymbol{\theta}}^{uEB})|$ where $\mathrm{dif}(\boldsymbol{\theta}) = |\overline{\boldsymbol{y}}_{ij3} - \boldsymbol{\theta}_{ij3}|$. It can be seen that the difference gets smaller as an area sample size n_i gets larger. This is because the smaple mean is reliable when n_i is large, so that the sample mean does not be shrunk and the the degree of shrinkage of MNER and NER have almost no difference. In Table 4.3, we also provide the estimats of squared root of MSE (SMSE) given in (4.3.4). It is revealed from Table 4.3 that SMSE of MNER is smaller than that of NER when n_i is small. On the other hand, SMSE of MNER is larger than that of NER when n_i is large, particularly larger than 5. This is because the low accuracy in estimation of the covariance matrix $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ has more adverse influence on prediction than the benefit from incorporating the small correlation into the estimation. Table 4.4 reports lower bounds (LB) and upper bounds (UB) of the 95% confidence interval estimator of the difference between the value of the posted land price from 1998 to 2001, that is $\boldsymbol{\ell}^\top \boldsymbol{\theta}_a$ where $\boldsymbol{\ell} = (1, 1, 1)^\top$, for selected 15 areas.

4.6 Proofs

In this section, we use the notations $\boldsymbol{\Lambda}_i = \boldsymbol{\Psi} + n_i^{-1}\boldsymbol{\Sigma}$ and $\widehat{\boldsymbol{\Lambda}}_i = \widehat{\boldsymbol{\Psi}} + n_i^{-1}\widehat{\boldsymbol{\Sigma}}$ for $i = 1, \ldots, m$.

		sample	MNER		NER	
area	n_i	mean	EBLUP	SMSE	EBLUP	SMSE
16	1	3.0	2.94	4.93	9.66	5.03
17	1	23.0	29.70	4.85	27.32	4.96
31	1	75.0	38.93	4.74	41.33	4.85
21	2	19.5	19.12	4.29	24.76	4.35
22	2	9.0	11.37	4.29	13.80	4.35
13	3	14.0	16.35	3.87	15.46	3.89
34	3	37.66	36.83	3.87	35.06	3.89
12	4	27.75	28.50	3.58	28.02	3.57
35	4	25.0	25.81	3.56	24.05	3.56
9	5	9.2	9.05	3.38	9.38	3.35
7	6	21.33	18.65	3.15	19.32	3.11
26	7	17.14	18.34	2.95	18.02	2.92
40	8	13.63	11.48	2.81	11.40	2.77
52	10	6.0	6.33	2.58	6.52	2.54
41	11	17.0	14.50	2.47	14.71	2.43

Table 4.3: The estimated results for PLP Data for selected 15 areas.

area	n_i	sample mean	EBLUP	LB	UB
16	1	73.0	48.99	14.97	83.02
17	1	74.0	107.75	74.24	141.24
31	1	261.0	136.79	103.93	169.65
21	2	132.5	109.78	80.78	138.78
22	2	64.0	63.46	34.48	92.44
13	3	50.0	58.86	32.61	85.11
34	3	113.66	120.87	94.63	147.11
12	4	103.5	104.60	80.20	129.00
35	4	71.75	80.66	56.43	104.89
9	5	23.2	21.91	-1.28	45.11
7	6	67.0	56.34	34.68	77.99
26	7	65.86	69.62	49.25	89.99
40	8	35.13	28.40	8.99	47.82
52	10	14.9	15.38	-2.55	33.31
41	11	47.55	38.91	21.75	56.06

Table 4.4: 95% confidence interval estimator for PLP Data for selected 15 areas.

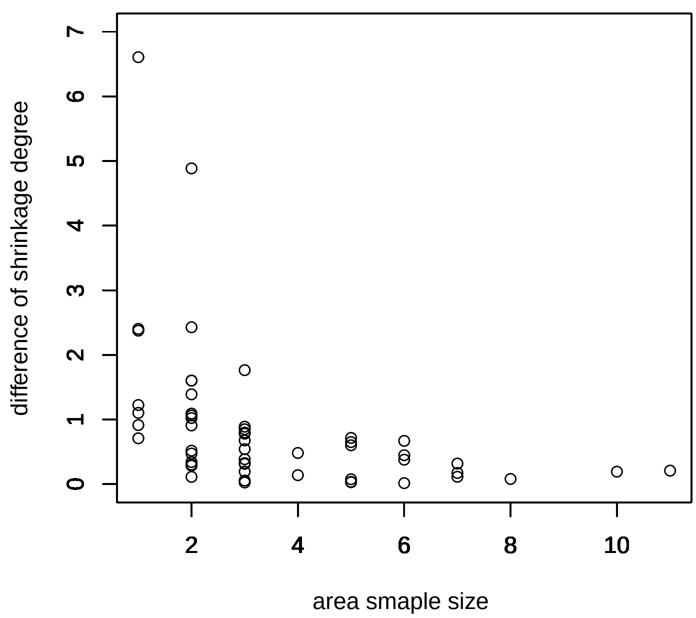

Figure 4.3: Plots of the difference of shrinkage degree against area sample size in MNER and NER.

4.6.1 Proof of Theorem 4.2.1

We first show the results (1) and (2) for the estimators $\widehat{\boldsymbol{\Psi}}_1$ and $\widehat{\boldsymbol{\Sigma}}$. Clearly, $E[\widehat{\boldsymbol{\Sigma}}] = \boldsymbol{\Sigma}$. To show that $E[\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi}] = O(m^{-3/2})$, we begin by writing $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi}$ as

$$
\begin{aligned}
\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} =& \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\{(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta})(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta})^\top - (\boldsymbol{\Psi} + \boldsymbol{\Sigma})\} \\
&+ \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}^\top(\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta})^\top\boldsymbol{X}_{ij} - \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta})^\top\boldsymbol{X}_{ij} \\
&- \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\boldsymbol{X}_{ij}^\top(\widehat{\boldsymbol{\beta}}^{OLS} - \boldsymbol{\beta})(\boldsymbol{y}_{ij} - \boldsymbol{X}_{ij}\boldsymbol{\beta}) - (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}),
\end{aligned}
$$

which yields the bias given in (4.2.9). Since $\widehat{\boldsymbol{\Sigma}}$ is unbiased, the bias $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ of $\widehat{\boldsymbol{\Psi}}_0$ is of order $O(m^{-1})$. Using the results that $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} = O_p(m^{-1/2})$ and $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = O_p(m^{-1/2})$, which will be shown below, we can see that $\widehat{\boldsymbol{\Psi}}_1 = \widehat{\boldsymbol{\Psi}}_0 - \mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\widehat{\boldsymbol{\Psi}}_0, \widehat{\boldsymbol{\Sigma}})$ is a second-oder unbiased estimator of $\boldsymbol{\Psi}$.

For (2), it is noted that $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$ is approximated as

$$
\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = \frac{1}{N-m}\sum_{i=1}^{m}\Big\{\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (n_i - 1)\boldsymbol{\Sigma}\Big\} + O_p(m^{-1}). \qquad (4.6.1)
$$

It is here noted that $\{\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (n_i - 1)\boldsymbol{\Sigma}\}/(N-m)$ for $i = 1, \ldots, m$ are mutually independent and $E\{\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (n_i - 1)\boldsymbol{\Sigma}\}/(N-m) = 0$ for $i = 1, \ldots, m$. Then we can show that $\sqrt{m}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})$ converges to a multivariate normal distributionthe because of the finiteness of moments of normal random variables, which implies that $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} = O_p(m^{-1/2})$.

Concerning $\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi} = O_p(m^{-1/2})$, from the fact that $\mathrm{Bias}_{\widehat{\boldsymbol{\Psi}}_0}(\widehat{\boldsymbol{\Psi}}_0, \widehat{\boldsymbol{\Sigma}}) = O_p(m^{-1})$, it is sufficient to show that $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} = O_p(m^{-1/2})$. Then $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi}$ is approximated as

$$
\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} = \frac{1}{N}\sum_{i=1}^{m}\Big\{\sum_{j=1}^{n_i}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})^\top - n_i(\boldsymbol{\Psi} + \boldsymbol{\Sigma})\Big\} - (\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}) + O_p(m^{-1}). \qquad (4.6.2)
$$

It is here noted that $\{\sum_{j=1}^{n_i}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})^\top - n_i(\boldsymbol{\Psi} + \boldsymbol{\Sigma})\}/N$ for $i = 1, \ldots, m$ are mutually independent and $E\{\sum_{j=1}^{n_i}(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})^\top - n_i(\boldsymbol{\Psi} + \boldsymbol{\Sigma})\}/N = 0$ for $i = 1, \ldots, m$. Then we can show that $\sqrt{m}(\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi})$ converges to a multivariate normal distribution because of the finiteness of moments of normal random variables, which implies that $\widehat{\boldsymbol{\Psi}}_0 - \boldsymbol{\Psi} = O_p(m^{-1/2})$.

We next prove (3) from the fact that $\sqrt{m}(\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi}) = O_p(1)$. The difference between $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Psi}}_1$ is in the case that $\widehat{\boldsymbol{\Psi}}_1$ is not nonegative definite. Thus, we evaluare the probability $P(\boldsymbol{a}^\top\widehat{\boldsymbol{\Psi}}_1\boldsymbol{a} < 0)$ for some $\boldsymbol{a} \in \mathbb{R}^k$. It is noted that the event $\boldsymbol{a}^\top\widehat{\boldsymbol{\Psi}}_1\boldsymbol{a} < 0$ is equivalent to $-\sqrt{m}\boldsymbol{a}^\top(\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi})\boldsymbol{a} > \boldsymbol{a}^\top\boldsymbol{\Psi}\boldsymbol{a}$. Using the Markov inequality, we observe that for any $\delta > 0$,

$$
\begin{aligned}
P(\boldsymbol{a}^\top\widehat{\boldsymbol{\Psi}}_1\boldsymbol{a} < 0) =& P(-\sqrt{m}\boldsymbol{a}^\top(\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi})\boldsymbol{a} > \sqrt{m}\boldsymbol{a}^\top\boldsymbol{\Psi}\boldsymbol{a}) \\
\le& P(|\sqrt{m}\boldsymbol{a}^\top(\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi})\boldsymbol{a}| > \sqrt{m}\boldsymbol{a}^\top\boldsymbol{\Psi}\boldsymbol{a}) \\
\le& E\Big[\Big(\frac{|\sqrt{m}\boldsymbol{a}^\top(\widehat{\boldsymbol{\Psi}}_1 - \boldsymbol{\Psi})\boldsymbol{a}|}{\sqrt{m}\boldsymbol{a}^\top\boldsymbol{\Psi}\boldsymbol{a}}\Big)^{2\delta}\Big] = O(m^{-\delta}),
\end{aligned}
$$

which proves (3) of Theorem 4.2.1.

Using the result (3) of Theorem 4.2.1, we can show that $E[\widehat{\boldsymbol{\Psi}}]-\boldsymbol{\Psi}=O(m^{-3/2})$ and $\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}=O_p(m^{-1/2})$.

Finally we verify that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}=O_p(m^{-1/2})$. Note that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}$ is decomposed as $\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}+\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta}\}$. For $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta}$, it is noted that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta}=(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}).$$

Then, $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta})=(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}=O(1/m)$ and this implies $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta}=O_p(m^{-1/2})$. We next evaluate $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ as

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})\\&=(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{y}-(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}\\&=(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\widehat{\boldsymbol{D}}^{-1}-\boldsymbol{D}^{-1})\boldsymbol{y}+\{(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}-(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}\\&=I_1+I_2,\end{aligned}\tag{4.6.3}$$

where $\widehat{\boldsymbol{D}}$ is obtained by replacing $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ in $\boldsymbol{D}$ with $\widehat{\boldsymbol{\Sigma}}_0$ and $\widehat{\boldsymbol{\Psi}}_0$ respectively. First, I_1 is written as

$$I_1=-(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}(\widehat{\boldsymbol{D}}-\boldsymbol{D})\boldsymbol{D}^{-1}\boldsymbol{y},\tag{4.6.4}$$

which is of order $O_p(m^{-1/2})$, because $(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})=O_p(m)$ and $\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}(\widehat{\boldsymbol{D}}-\boldsymbol{D})\boldsymbol{D}^{-1}\boldsymbol{y}=O_p(m^{1/2})$. Next, I_2 is rewritten as

$$\begin{aligned}I_2=&-(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\widehat{\boldsymbol{D}}^{-1}-\boldsymbol{D}^{-1})\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}\\=&(\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}(\widehat{\boldsymbol{D}}-\boldsymbol{D})\boldsymbol{D}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}\end{aligned}\tag{4.6.5}$$

which is of order $O_p(m^{-1/2})$, because $\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}\boldsymbol{X}=O_p(m)$, $\boldsymbol{X}^\top\widehat{\boldsymbol{D}}^{-1}(\widehat{\boldsymbol{D}}-\boldsymbol{D})\boldsymbol{D}^{-1}\boldsymbol{X}=O_p(m^{1/2})$, $\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{y}=O_p(m)$. Thus, we have $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})=O_p(m^{-1/2})$, and it is concluded that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}=O_p(m^{-1/2})$. □

4.6.2 Proof of Lemma 4.3.1

The covariance of $\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is

$$\begin{aligned}&E[(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS})(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta})^\top]\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X}\\&=E[\{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})-\boldsymbol{X}(\widehat{\boldsymbol{\beta}}^{OLS}-\boldsymbol{\beta})\}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top]\boldsymbol{D}^{-1}\boldsymbol{X}\\&=(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)E[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top]\boldsymbol{D}^{-1}\boldsymbol{X}\\&=(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{D}\boldsymbol{D}^{-1}\boldsymbol{X}=\boldsymbol{0},\end{aligned}$$

which implies that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is independent of $\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}$. We next note that $\tilde{\boldsymbol{y}}=\boldsymbol{Q}\boldsymbol{y}$ and $\widetilde{\boldsymbol{X}}=\boldsymbol{Q}\boldsymbol{X}$ where $\boldsymbol{Q}=\text{block diag}(\boldsymbol{P}_1\otimes\boldsymbol{I}_k,\ldots,\boldsymbol{P}_m\otimes\boldsymbol{I}_k)$ for $\boldsymbol{P}_i=\boldsymbol{I}_{n_i}-n_i^{-1}\boldsymbol{J}_{n_i}$. Then,

$$
\begin{aligned}
&E[(\tilde{\boldsymbol{y}}-\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta})^\top]\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X}\\
=&E[\{(\tilde{\boldsymbol{y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta})-\widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}^\top\widetilde{\boldsymbol{X}})^-\widetilde{\boldsymbol{X}}^\top(\tilde{\boldsymbol{y}}-\widetilde{\boldsymbol{X}}\boldsymbol{\beta})^\top\}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top]\boldsymbol{D}^{-1}\boldsymbol{X}\\
=&\{\boldsymbol{Q}-\boldsymbol{Q}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{X})^-\boldsymbol{X}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\}E[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top]\boldsymbol{D}^{-1}\boldsymbol{X}\\
=&\boldsymbol{Q}\boldsymbol{X}-\boldsymbol{Q}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{X})^-\boldsymbol{X}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{X},
\end{aligned}
$$

which is equal to zero from the property of the generalized inverse. Thus, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is independent of $\tilde{\boldsymbol{y}}-\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{\beta}}$, so that $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is independent of $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Psi}}$.

It is also noted that

$$
\begin{aligned}
&\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\\
&=\boldsymbol{c}_a^\top\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}+\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}\{\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})\}-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\{\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}\\
&=\boldsymbol{c}_a^\top\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}+(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1})(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\widehat{\boldsymbol{\beta}}^{OLS})\\
&\quad-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}\overline{\boldsymbol{X}}_a^\top\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\widehat{\boldsymbol{\beta}}^{OLS}\}+\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\widehat{\boldsymbol{\beta}}^{OLS}\},
\end{aligned}
$$

which is a function of $\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS}$ and $\widehat{\boldsymbol{\Sigma}}$, because $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma}))-\widehat{\boldsymbol{\beta}}^{OLS}=(\boldsymbol{X}^\top\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS})-(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{OLS})$. Hence, $\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})$ is independent of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})$. □

4.6.3 Proof of Theorem 4.3.1

We shall prove that $E[\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}^\top]=\boldsymbol{G}_{3a}(\boldsymbol{\Psi},\boldsymbol{\Sigma})+O_p(m^{-3/2})$. It is observed that

$$
\begin{aligned}
&\widehat{\boldsymbol{\theta}}_a^{EB}-\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\\
&=(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1})(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta})+(\boldsymbol{c}_a^\top-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}\overline{\boldsymbol{X}}_a^\top)\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}\\
&\quad-(\boldsymbol{c}_a^\top-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top)\{\widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi},\boldsymbol{\Sigma})-\boldsymbol{\beta}\}.
\end{aligned}
$$

We can see that

$$
\begin{aligned}
&(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1})(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta})\\
=&\big\{\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\Lambda}}_a^{-1}-\boldsymbol{\Lambda}_a^{-1})+(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\big\}(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta})\\
=&\big\{(\boldsymbol{I}_k-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1})(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\big\}\boldsymbol{\Lambda}_a^{-1}(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta})\\
=&\big\{n_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\big\}\boldsymbol{\Lambda}_a^{-1}(\overline{\boldsymbol{y}}_a-\overline{\boldsymbol{X}}_a^\top\boldsymbol{\beta})+O_p(m^{-1})
\end{aligned}
$$

and

$$
\begin{aligned}
&(\boldsymbol{c}_a^\top-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}\overline{\boldsymbol{X}}_a^\top)\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}\\
=&(\boldsymbol{c}_a^\top-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top)\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}-(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1})\overline{\boldsymbol{X}}_a^\top\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}\\
=&(\boldsymbol{c}_a^\top-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top)\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}\\
&\quad+\big\{n_a^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Lambda}}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})-\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\Lambda}}_a^{-1}n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\big\}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}\\
=&\big\{\boldsymbol{c}_a^\top-\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\overline{\boldsymbol{X}}_a^\top\big\}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{\beta}\}+O_p(m^{-1}).
\end{aligned}
$$

Thus, we have

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& \big\{ n_a^{-1} \boldsymbol{\Sigma} \boldsymbol{\Lambda}_a^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) - \boldsymbol{\Psi} \boldsymbol{\Lambda}_a^{-1} n_a^{-1} (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \big\} \boldsymbol{\Lambda}_a^{-1} (\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top \boldsymbol{\beta}) \\
& + (\boldsymbol{c}_a^\top - \boldsymbol{\Psi} \boldsymbol{\Lambda}_a^{-1} \overline{\boldsymbol{X}}_a^\top) \{ \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\beta} \} + O_p(m^{-1}) \\
=& I_1 + I_2 + O_p(m^{-1}). \quad \text{(say)}
\end{aligned}$$

For I_2, it is noted that

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) =& \big\{ (\boldsymbol{X}^\top \widehat{\boldsymbol{D}}^{-1} \boldsymbol{X})^{-1} - (\boldsymbol{X}^\top \boldsymbol{D}^{-1} \boldsymbol{X})^{-1} \big\} \boldsymbol{X}^\top \widehat{\boldsymbol{D}}^{-1} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
& + (\boldsymbol{X}^\top \boldsymbol{D}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\widehat{\boldsymbol{D}}^{-1} - \boldsymbol{D}^{-1})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
=& I_{21} + I_{22}, \quad \text{(say)}.
\end{aligned}$$

We can evaluate I_{21} as

$$I_{21} = (\boldsymbol{X}^\top \boldsymbol{D}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\boldsymbol{D}}^{-1} (\widehat{\boldsymbol{D}} - \boldsymbol{D}) \boldsymbol{D}^{-1} \boldsymbol{X} \{ \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \boldsymbol{\Sigma}) - \boldsymbol{\beta} \} = O_p(m^{-1}),$$

because $\boldsymbol{X}^\top \boldsymbol{D}^{-1} \boldsymbol{X} = O(m)$, $\boldsymbol{X}^\top \widehat{\boldsymbol{D}}^{-1} (\widehat{\boldsymbol{D}} - \boldsymbol{D}) \boldsymbol{D}^{-1} \boldsymbol{X} = O_p(m^{1/2})$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) - \boldsymbol{\beta} = O_p(m^{-1/2})$ from Theorem 4.2.1 (2). We next estimate I_{22} as

$$\begin{aligned}
I_{22} =& - (\boldsymbol{X}^\top \boldsymbol{D}^{-1} \boldsymbol{X})^{-1} \Big\{ \sum_{i=1}^m \boldsymbol{X}_i^\top \boldsymbol{A}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) \boldsymbol{X}_i \Big\} \\
& \times \Big\{ \sum_{i=1}^m \boldsymbol{X}_i^\top \boldsymbol{A}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) \boldsymbol{X}_i \Big\}^{-1} \sum_{i=1}^m \boldsymbol{X}_i^\top \boldsymbol{A}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}),
\end{aligned}$$

where

$$\boldsymbol{A}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}}) = (\boldsymbol{J}_{n_i} \otimes \widehat{\boldsymbol{\Psi}} + \boldsymbol{I}_{n_i} \otimes \widehat{\boldsymbol{\Sigma}})^{-1} \big\{ \boldsymbol{J}_{n_i} \otimes (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) + \boldsymbol{I}_{n_i} \otimes (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \big\} (\boldsymbol{J}_{n_i} \otimes \boldsymbol{\Psi} + \boldsymbol{I}_{n_i} \otimes \boldsymbol{\Sigma})^{-1}.$$

It can be seen that $I_{22} = O_p(m^{-1})$ from the same arguments as in I_{21}. Thus, it follows that $I_2 = O_p(m^{-1})$.

From equations (4.6.1) and (4.6.2),

$$\begin{aligned}
& \widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) \\
&= \Big\{ n_a^{-1} \boldsymbol{\Sigma} \boldsymbol{\Lambda}_a^{-1} (\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) - \boldsymbol{\Psi} \boldsymbol{\Lambda}_a^{-1} n_a^{-1} (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \Big\} \boldsymbol{\Lambda}_a^{-1} (\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top \boldsymbol{\beta}) + O_p(m^{-1}) \qquad (4.6.6) \\
&= \Big[\frac{n_a^{-1}}{N} \boldsymbol{\Sigma} \boldsymbol{\Lambda}_a^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \big\{ (\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})(\boldsymbol{v}_i + \boldsymbol{\varepsilon}_{ij})^\top - (\boldsymbol{\Psi} + \boldsymbol{\Sigma}) \big\} \\
&\quad - \frac{n_a^{-1}}{N - m} (\boldsymbol{\Psi} + \boldsymbol{\Sigma}) \boldsymbol{\Lambda}_a^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \big\{ (\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (1 - n_i^{-1}) \boldsymbol{\Sigma} \big\} \Big] \\
&\quad \times \boldsymbol{\Lambda}_a^{-1} (\overline{\boldsymbol{y}}_a - \overline{\boldsymbol{X}}_a^\top \boldsymbol{\beta}) + O_p(m^{-1}) \\
&= \frac{n_a^{-1}}{N} \boldsymbol{\Sigma} \boldsymbol{\Lambda}_a^{-1} \sum_{i=1}^m n_i \big\{ (\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)^\top - \boldsymbol{\Lambda}_i \big\} \boldsymbol{\Lambda}_a^{-1} (\boldsymbol{v}_a + \bar{\boldsymbol{\varepsilon}}_a)
\end{aligned}$$

$$
\begin{aligned}
&- \frac{n_a^{-1}}{N(N-m)}(N\boldsymbol{\Psi}+m\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}\big\{(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)^\top-(1-n_i^{-1})\boldsymbol{\Sigma}\big\}\\
&\quad\times \boldsymbol{\Lambda}_a^{-1}(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)+O_p(m^{-1})\\
=&\boldsymbol{A}_1(\boldsymbol{y})-\boldsymbol{A}_2(\boldsymbol{y})+O_p(m^{-1}),\quad \text{(say)}. \qquad (4.6.7)
\end{aligned}
$$

Note that $E[\boldsymbol{A}_1(\boldsymbol{y})\{\boldsymbol{A}_2(\boldsymbol{y})\}^\top]=\boldsymbol{0}$ because $\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i$ is independent of $\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i$. Hence, we need to evaluate $E[\boldsymbol{A}_1(\boldsymbol{y})\{\boldsymbol{A}_1(\boldsymbol{y})\}^\top]$ and $E[\boldsymbol{A}_2(\boldsymbol{y})\{\boldsymbol{A}_2(\boldsymbol{y})\}^\top]$.

Concerning $E[\boldsymbol{A}_1(\boldsymbol{y})\{\boldsymbol{A}_1(\boldsymbol{y})\}^\top]$, it can be seen that

$$
\begin{aligned}
&E\Big[\sum_{i=1}^m n_i\big\{(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)^\top-\boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)^\top\boldsymbol{\Lambda}_a^{-1}\sum_{j=1}^m n_j\big\{(\boldsymbol{v}_j+\bar{\boldsymbol{\varepsilon}}_j)(\boldsymbol{v}_j+\bar{\boldsymbol{\varepsilon}}_j)^\top-\boldsymbol{\Lambda}_j\big\}\Big]\\
&=\sum_{i=1}^m n_i^2E\Big[\big\{(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)^\top-\boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)^\top\boldsymbol{\Lambda}_a^{-1}\big\{(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)^\top-\boldsymbol{\Lambda}_i\big\}\Big]\\
&=\sum_{i\neq a} n_i^2E\Big[\big\{(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)^\top-\boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}\big\{(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i+\bar{\boldsymbol{\varepsilon}}_i)^\top-\boldsymbol{\Lambda}_i\big\}\Big]+O(1)\\
&=\sum_{i\neq a} n_i^2\big\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i+\mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\big\}+O(1)=\sum_{i=1}^m n_i^2\big\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i+\mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\big\}+O(1),
\end{aligned}
$$

so that we have

$$
E[\boldsymbol{A}_1(\boldsymbol{y})\{\boldsymbol{A}_1(\boldsymbol{y})\}^\top]=\frac{n_a^{-2}}{N^2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m n_i^2\big\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i+\mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}+O(m^{-2}). \quad (4.6.8)
$$

Concerning the evaluation of $E[\boldsymbol{A}_2(\boldsymbol{y})\{\boldsymbol{A}_2(\boldsymbol{y})\}^\top]$, let $\boldsymbol{W}=\sum_{i=1}^m\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)^\top$ for simplicity. Then, $\boldsymbol{W}$ has the Wishart distribution $\mathcal{W}_k(N-m,\boldsymbol{\Sigma})$. Because $\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a$ is independent of $\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i$, it follows that

$$
\begin{aligned}
&E\Big[\sum_{i=1}^m\sum_{j=1}^{n_i}\big\{(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij}-\bar{\boldsymbol{\varepsilon}}_i)^\top-(1-n_i^{-1})\boldsymbol{\Sigma}\big\}\boldsymbol{\Lambda}_a^{-1}(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)^\top\boldsymbol{\Lambda}_a^{-1}\\
&\quad\times\sum_{k=1}^m\sum_{\ell=1}^{n_k}\big\{(\boldsymbol{\varepsilon}_{k\ell}-\bar{\boldsymbol{\varepsilon}}_k)(\boldsymbol{\varepsilon}_{k\ell}-\bar{\boldsymbol{\varepsilon}}_k)^\top-(1-n_k^{-1})\boldsymbol{\Sigma}\big\}\Big]\\
=&E\big[\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\boldsymbol{\Lambda}_a^{-1}(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)(\boldsymbol{v}_a+\bar{\boldsymbol{\varepsilon}}_a)^\top\boldsymbol{\Lambda}_a^{-1}\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\big]\\
=&E\big[\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\boldsymbol{\Lambda}_a^{-1}\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\big].
\end{aligned}
$$

From the properties of the Wishart distribution, it is noted that $E[\boldsymbol{W}]=(N-m)\boldsymbol{\Sigma}$ and $E[\boldsymbol{W}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{W}]=(N-m)(N-m+1)\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}+(N-m)\mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma})\boldsymbol{\Sigma}$. Thus,

$$
\begin{aligned}
E[\boldsymbol{A}_2(\boldsymbol{y})\{\boldsymbol{A}_2(\boldsymbol{y})\}^\top]=&\frac{n_a^{-2}}{N^2(N-m)^2}(N\boldsymbol{\Psi}+m\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}E\big[\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\boldsymbol{\Lambda}_a^{-1}\{\boldsymbol{W}-(N-m)\boldsymbol{\Sigma}\}\big]\\
&\times\boldsymbol{\Lambda}_a^{-1}(N\boldsymbol{\Psi}+m\boldsymbol{\Sigma})\\
=&\frac{n_a^{-2}}{N^2(N-m)}(N\boldsymbol{\Psi}+m\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\Big\{\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}+\mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma})\boldsymbol{\Sigma}\Big\}\boldsymbol{\Lambda}_a^{-1}(N\boldsymbol{\Psi}+m\boldsymbol{\Sigma}).
\end{aligned}
\qquad (4.6.9)
$$

Combining (4.6.7), (4.6.8) and (4.6.9) gives the expression in (4.3.2). □

4.6.4 Proof of Theorem 4.3.2

From (2) in Theorem 4.2.1, it is sufficient to show this approximation for $\widehat{\boldsymbol{\Psi}}$. We can rewrite $G_{1a}(\widehat{\boldsymbol{\Psi}})$ as

$$\begin{aligned}\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}}) =& n_a^{-1}\widehat{\boldsymbol{\Psi}}(\widehat{\boldsymbol{\Psi}}+n_a^{-1}\widehat{\boldsymbol{\Sigma}})^{-1}\widehat{\boldsymbol{\Sigma}}\\ =& \boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) + n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\\ &- n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} - n_a^{-2}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\\ &+ n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi} + n_a^{-2}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\\ &+ O_p(m^{-3/2}),\end{aligned}$$

which implies that

$$\begin{aligned}&\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})]\\ &=E\Big[n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + n_a^{-2}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\\ &\quad - n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi} - n_a^{-2}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\Big]\\ &\quad + O(m^{-3/2}),\end{aligned}$$

because $\widehat{\boldsymbol{\Psi}}$ is second-order unbiased and $\widehat{\boldsymbol{\Sigma}}$ is unbiased. On the other hand, from (4.6.6), it follows that

$$\begin{aligned}&E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}^\top]\\ &=E\Big[\big\{n_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}) - \boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\big\}\boldsymbol{\Lambda}_a^{-1}(\bar{\boldsymbol{y}}_a - \bar{\boldsymbol{X}}_a^\top\boldsymbol{\beta})\\ &\quad\times(\bar{\boldsymbol{y}}_a - \bar{\boldsymbol{X}}_a^\top\boldsymbol{\beta})^\top\boldsymbol{\Lambda}_a^{-1}\big\{(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}n_a^{-1}\boldsymbol{\Sigma} - n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\big\}\Big] + O(m^{-3/2})\\ &=E\Big[\big\{n_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi}) - \boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\big\}\boldsymbol{\Lambda}_a^{-1}\big\{(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}n_a^{-1}\boldsymbol{\Sigma} - n_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\big\}\Big]\\ &\quad + O(m^{-3/2}).\end{aligned}$$

Thus, we have

$$\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})] = E[\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}\{\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\}^\top] + O(m^{-3/2}),$$

which yields $\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})] = \boldsymbol{G}_{3a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) + O(m^{-3/2})$. Since $\boldsymbol{G}_{3a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) = O(m^{-1})$, one gets

$$\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) = E[\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}}) + \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})] + O(m^{-3/2}),$$

and Theorem 4.3.2 is established. □

4.6.5 Proof of Theorem 4.4.1

The proof is done along the line given in Diao et al. (2014). Let $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{I}_k - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. From Lemma 4.3.1, $\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}))$ is a function of $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$ and is independent of $\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \boldsymbol{\theta}_a)$ given $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$. It is noted that $E[\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \boldsymbol{\theta}_a)(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \boldsymbol{\theta}_a)^\top\boldsymbol{\ell}] = \boldsymbol{\ell}^\top E[(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma}))(\widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma}))^\top + (\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \boldsymbol{\theta}_a)(\widetilde{\boldsymbol{\theta}}_a(\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\Sigma}) - \boldsymbol{\theta}_a)^\top]\boldsymbol{\ell} = \boldsymbol{\ell}^\top(\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma}) +$

$\boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))\boldsymbol{\ell}$, where $\boldsymbol{G}_{1a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ and $\boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ are given in (4.3.1). Then the conditional distribution of $\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)$ given $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$ is

$$\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)|\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} \sim \mathcal{N}_k(\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})), \boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}), \tag{4.6.10}$$

where $\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) = \boldsymbol{G}_{1a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma}) + \boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$. This implies that

$$\begin{aligned}
&P\Big(\frac{\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a)}{\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}} \le z\Big) \\
=&E\Big[P\Big(\frac{\boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \boldsymbol{\theta}_a) - \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))}{\{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}} \\
&\qquad \le \frac{\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z - \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))}{\{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}}|\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}\Big)\Big] \\
=&E\Big[\Phi\Big(\frac{\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z - \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))}{\{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}}\Big)\Big].
\end{aligned}$$

Thus, it is observed that

$$\begin{aligned}
&P\Big(\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} - \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z \le \boldsymbol{\ell}^\top\boldsymbol{\theta}_a \le \boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB} + \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z\Big) \\
=&E\Big[\Phi\Big(\frac{\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z - \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))}{\{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}}\Big) \\
&\quad - \Phi\Big(\frac{-\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z - \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma}))}{\{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}}\Big)\Big] \\
=&E\Big[\Phi(r_{1a} - r_{2a}) - \Phi(-r_{1a} - r_{2a})\Big] = E\Big[\Phi(r_{1a} + r_{2a}) + \Phi(r_{1a} - r_{2a})\Big] - 1,
\end{aligned}$$

where

$$\begin{aligned}
r_{1a} =& \{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} z / \{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}, \\
r_{2a} =& \boldsymbol{\ell}^\top(\widehat{\boldsymbol{\theta}}_a^{EB} - \widehat{\boldsymbol{\theta}}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})) / \{\boldsymbol{\ell}^\top\boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}\}^{1/2}.
\end{aligned}$$

By the Taylor series expansion, for $r_{1a}^* \in (r_{1a}, r_{1a} + r_{2a})$ and $r_{1a}^{**} \in (r_{1a}, r_{1a} - r_{2a})$, we have

$$\Phi(r_{1a} + r_{2a}) + \Phi(r_{1a} - r_{2a}) = 2\Phi(r_{1a}) + r_{2a}^2\phi^{(1)}(r_{1a}) + \frac{1}{24}r_{2a}^4(\phi^{(3)}(r_{1a}^*) + \phi^{(3)}(r_{1a}^**)), \tag{4.6.11}$$

where $\phi^{(1)}(\cdot)$ and $\phi^{(3)}(\cdot)$ are the first and third derivatives of the standard normal density $\phi(\cdot)$. The Taylor series expansion is also used to get

$$\begin{aligned}
\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} =& \{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} + \boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2} \\
=& \{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}\Big(1 + \frac{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{2\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}} \\
&- \frac{(\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^2}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2} + \cdots\Big).
\end{aligned} \tag{4.6.12}$$

We evaluate the expectation of the first term $2\Phi(r_{1a})$ in (4.6.11). By the Taylor series expansion, for $z^* \in (z, r_{1a})$, it is seen that

$$\Phi(r_{1a}) - \Phi(z) = (r_{1a} - z)\phi(z) + \frac{(r_{1a} - z)^2}{2}\phi^{(1)}(z) + \frac{(r_{1a} - z)^3}{6}\phi^{(2)}(z) + \frac{(r_{1a} - z)^4}{24}\phi^{(1)}(z^*). \tag{4.6.13}$$

From (4.6.12), we can evaluate $E[r_{1a}]$ as

$$\begin{aligned}E[r_{1a}] =&\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2} z\Big[1 + \frac{E[\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}]}{2\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}\\
&- \frac{E[(\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^2]}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2} + \frac{E[(\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^3]}{16\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^3}\\
&+ \frac{1}{16}\frac{5}{8}E\Big[\int_{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}^{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}} \{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{-1/2} x^{-7/2} (\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^3\Big]\Big].\end{aligned} \tag{4.6.14}$$

Since $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB}) = O(1)$ and $\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})$ is a second order unbised estimator of $\mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})$, the second term in the bracket of (4.6.14) is of order $o(m^{-1})$. From Lemma 4.4.1, the moments of higher than three are of order $o(m^{-1})$, and we have $E[\{\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2] = V(\widehat{\boldsymbol{\theta}}_a^{EB}) + o(m^{-1})$. Then, using

$$\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2} = 1 + \frac{1}{2}\frac{\boldsymbol{\ell}^\top \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}} + o(m^{-1}),$$

we have

$$\begin{aligned}E[r_{1a}] - z =&\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2} z\Big[1 - \frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big] - z + o(m^{-1})\\
=&\Big\{\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2} - 1 - \Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2} \frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big\}z\\
&+ o(m^{-1})\\
=&\Big\{\frac{1}{2}\frac{\boldsymbol{\ell}^\top \boldsymbol{G}_{3a}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell}} - \frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big\}z + o(m^{-1}),\end{aligned} \tag{4.6.15}$$

because $\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\theta}}_a^{EB})$ and $V(\widehat{\boldsymbol{\theta}}_a^{EB})$ are of order $O(m^{-1})$.

Since $E[r_{1a}^2] = E[\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} z^2/\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell} = z^2 \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}/\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi}, \boldsymbol{\Sigma})\boldsymbol{\ell} +$

$o(m^{-1})$ and $E[(r_{1a}-z)^2]=E[r_{1a}^2]-2zE[r_{1a}-z]-z^2$, it is observed that

$$
\begin{aligned}
&E[(r_{1a}-z)^2]\\
=&z^2\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{\ell}}-2z\Big(\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2}z\Big[1-\frac{E[(\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}-\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^2]}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big]\\
&-z\Big)-z^2+o(m^{-1})\\
=&\Big\{\Big(\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2}-1\Big)^2+\Big(\frac{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{\ell}}\Big)^{1/2}\frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big\}z^2+o(m^{-1})\\
=&\frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}z^2+o(m^{-1}).
\end{aligned}
\tag{4.6.16}
$$

This implies that $r_{1a}-z=O_p(m^{-1/2})$. Then, the expectation of the third and forth terms of (4.6.13) is of order $O_(m^{-3/2})$.

We evaluate the expectation of the second term in (4.6.11). Since $E[r_{2a}^2]=\boldsymbol{\ell}^\top\boldsymbol{G}_{3a}\boldsymbol{\ell}/\boldsymbol{\ell}^\top\boldsymbol{H}_a\boldsymbol{\ell}$ and $E[r_{1a}]-z$ are of order $O(m^{-1})$, we have

$$
E[r_{2a}^2\phi^{(1)}(r_{1a})]=E[r_{2a}^2\phi^{(1)}(z)]+o(m^{-1})=\frac{\boldsymbol{\ell}^\top\boldsymbol{G}_{3a}\boldsymbol{\ell}}{\boldsymbol{\ell}^\top\boldsymbol{H}_a\boldsymbol{\ell}}\phi^{(1)}(z)+o(m^{-1}). \tag{4.6.17}
$$

Since $r_{2a}^4=O(m^{-2})$ and $E[r_{1a}^*]-z=O(1)$, the expectation of the third term in (4.6.11) is of order $O(m^{-2})$.

Combining (4.6.11), (4.6.13), (4.6.15), (4.6.16) and (4.6.17) gives

$$
\begin{aligned}
&P\Big(\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB}-\{\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}z\le\boldsymbol{\ell}^\top\boldsymbol{\theta}_a\le\boldsymbol{\ell}^\top\widehat{\boldsymbol{\theta}}_a^{EB}+\{\boldsymbol{\ell}^\top\mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^{1/2}z\Big)\\
=&2\Phi(z)+2\Big\{\frac{1}{2}\frac{\boldsymbol{\ell}^\top\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}}{\boldsymbol{\ell}^\top \boldsymbol{H}_a(\boldsymbol{\Psi},\boldsymbol{\Sigma})\boldsymbol{\ell}}-\frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{8\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}\Big\}z\phi(z)\\
&+\frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}z^2\phi^{(1)}(z)+\frac{\boldsymbol{\ell}^\top\boldsymbol{G}_{3a}\boldsymbol{\ell}}{\boldsymbol{\ell}^\top\boldsymbol{H}_a\boldsymbol{\ell}}\phi^{(1)}(z)-1+o(m^{-1})\\
=&2\Phi(z)-1-\frac{V(\widehat{\boldsymbol{\theta}}_a^{EB})}{4\{\boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}\}^2}(z^3+z)\phi(z)+o(m^{-1}),
\end{aligned}
$$

which establishes Theorem 4.4.1. □

4.6.6 Proof of Lemma 4.4.1

It is noted that

$$
\boldsymbol{G}_{1a}(\widehat{\boldsymbol{\Psi}},\widehat{\boldsymbol{\Sigma}})-\boldsymbol{G}_{1a}(\boldsymbol{\Psi},\boldsymbol{\Sigma})=n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}+n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}+O_p(m^{-1}),
$$

and $\boldsymbol{G}_{2a}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})$, $\boldsymbol{G}_{2a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$, $\boldsymbol{G}_{3a}(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ and $\boldsymbol{G}_{3a}(\widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\Sigma}})$ are of order $O(m^{-1})$. Thus,

$$\begin{aligned}
&E[(\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell})^2] \\
=&\boldsymbol{\ell}^\top E(n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi})\boldsymbol{\ell}\boldsymbol{\ell}^\top \\
&\times (n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi})\boldsymbol{\ell} + o(m^{-1}),
\end{aligned}$$

and the moment of $\boldsymbol{\ell}^\top \mathrm{msem}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell} - \boldsymbol{\ell}^\top \mathrm{MSEM}(\widehat{\boldsymbol{\theta}}_a^{EB})\boldsymbol{\ell}$ of order higher than three is of order $o(m^{-1})$.

From equations (4.6.1) and (4.6.2), it follows that

$$\begin{aligned}
&n_a^{-2}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} + n_a^{-1}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}) \\
&\quad = \frac{n_a^{-2}}{N}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m n_i\big\{(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)^\top - \boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} \\
&\qquad - \Big[\frac{n_a^{-2}m}{N(N-m)}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}\big\{(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (1 - n_i^{-1})\boldsymbol{\Sigma}\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma} \\
&\qquad\quad - \frac{n_a^{-1}}{N-m}\boldsymbol{\Psi}\boldsymbol{\Lambda}_a^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}\big\{(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top - (1 - n_i^{-1})\boldsymbol{\Sigma}\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Psi}\Big] + O_p(m^{-1}) \\
&\quad = \boldsymbol{B}_1(\boldsymbol{y}) - \boldsymbol{B}_2(\boldsymbol{y}) + O_p(m^{-1}), \quad \text{(say)}.
\end{aligned}$$

Since $\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i$ is independent of $\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i$, it is seen that $E[\boldsymbol{B}_1(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_2(\boldsymbol{y})\}^\top] = \boldsymbol{0}$. Thus, we shall evaluate $E[\boldsymbol{B}_1(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_1(\boldsymbol{y})\}^\top]$ and $E[\boldsymbol{B}_2(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_2(\boldsymbol{y})\}^\top]$. For the proofs, we can use the same arguments as in (4.6.7), (4.6.8) and (4.6.9).

Concerning $E[\boldsymbol{B}_1(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_1(\boldsymbol{y})\}^\top]$, it is observed that

$$\begin{aligned}
&E\Big[\sum_{i=1}^m n_i\big\{(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)^\top - \boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\sum_{j=1}^m n_j\big\{(\boldsymbol{v}_j + \bar{\boldsymbol{\varepsilon}}_j)(\boldsymbol{v}_j + \bar{\boldsymbol{\varepsilon}}_j)^\top - \boldsymbol{\Lambda}_j\big\}\Big] \\
&= \sum_{i=1}^m n_i^2 E\Big[\big\{(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)^\top - \boldsymbol{\Lambda}_i\big\}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\big\{(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{v}_i + \bar{\boldsymbol{\varepsilon}}_i)^\top - \boldsymbol{\Lambda}_i\big\}\Big] \\
&= \sum_{i=1}^m n_i^2\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i + \mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\},
\end{aligned}$$

so that we have

$$E[\boldsymbol{B}_1(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_1(\boldsymbol{y})\}^\top] = \frac{n_a^{-4}}{N^2}\sum_{i=1}^m n_i^2\{\boldsymbol{\Lambda}_i\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i + \mathrm{tr}\,(\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\boldsymbol{\Sigma}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{\Lambda}_i)\boldsymbol{\Lambda}_i\}. \tag{4.6.18}$$

Concerning $E[\boldsymbol{B}_2(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_2(\boldsymbol{y})\}^\top]$, recall that $\boldsymbol{W} = \sum_{i=1}^m\sum_{j=1}^{n_i}(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)(\boldsymbol{\varepsilon}_{ij} - \bar{\boldsymbol{\varepsilon}}_i)^\top$ has the

Wishart distribution $\mathcal{W}_k(N-m, \mathbf{\Sigma})$. Then, we have

$$
\begin{aligned}
&E[\boldsymbol{B}_2(\boldsymbol{y})\boldsymbol{\ell}\boldsymbol{\ell}^\top\{\boldsymbol{B}_2(\boldsymbol{y})\}^\top] \\
=&E\Big\{\frac{n_a^{-2}m}{N(N-m)}\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\big\{\boldsymbol{W}-(N-m)\mathbf{\Sigma}\big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}-\frac{n_a^{-1}}{N-m}\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\big\{\boldsymbol{W}-(N-m)\mathbf{\Sigma}\big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\Big\} \\
&\times\boldsymbol{\ell}\boldsymbol{\ell}^\top\Big\{\frac{n_a^{-2}m}{N(N-m)}\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\big\{\boldsymbol{W}-(N-m)\mathbf{\Sigma}\big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}-\frac{n_a^{-1}}{N-m}\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\big\{\boldsymbol{W}-(N-m)\mathbf{\Sigma}\big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\Big\} \\
=&\frac{n_a^{-4}m^2}{N^2(N-m)}\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\Big\{\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}+\operatorname{tr}(\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma})\mathbf{\Sigma}\Big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma} \\
&+\frac{n_a^{-2}}{N-m}\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\Big\{\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}+\operatorname{tr}(\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma})\mathbf{\Sigma}\Big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi} \\
&-\frac{n_a^{-3}m}{N(N-m)}\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\Big\{\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}+\operatorname{tr}(\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma})\mathbf{\Sigma}\Big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi} \\
&-\frac{n_a^{-3}m}{N(N-m)}\mathbf{\Psi}\mathbf{\Lambda}_a^{-1}\Big\{\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}+\operatorname{tr}(\mathbf{\Lambda}_a^{-1}\mathbf{\Psi}\boldsymbol{\ell}\boldsymbol{\ell}^\top\mathbf{\Sigma}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma})\mathbf{\Sigma}\Big\}\mathbf{\Lambda}_a^{-1}\mathbf{\Sigma}.
\end{aligned}
\tag{4.6.19}
$$

Multipling $\boldsymbol{\ell}$ by (4.6.18) and (4.6.19) from both sides, we get the expression given in (4.4.2). □

Acknowledgment

I thank my supervisor Prof. Tatsuya Kubokawa for all of his help and lots of encouragement to me during the master's and doctral courses. I was financially supported by JSPS. Lastly, I thank my father, mother, sisiter, grandparents and their families for their support.

Bibliography

[1] R. Basu, J.K. Ghosh, R. Mukerjee, Empirical Bayes prediction intervals in a normal regression model: Higher order asymptotics, Statist. Probab. Letters 63 (2003) 197–203.

[2] G.E. Battese, R.M. Harter, W.A. Fuller, An error component model for prediction of mean crop areas using survey and satellite data., J. Am. Statist. Assoc. 95 (1988) 28–36.

[3] R. Benavent, D. Morales, Multivariate Fay-Herriot models for small area estimation, Comp. Statist. Data Anal. 94 (2016) 372–390.

[4] J.A. Calvin, R.L. Dykstra, Maximum likelihood estimation of a set of covariance matrices under Löwner order restrictions with applications to balanced multivariate variance components models, Ann. Statist. 19 (1991) 850-869.

[5] J.A. Calvin, R.L. Dykstra, Least squares estimation of covariance matrices in balanced multivariate variance components models, J. Am. Statist. Assoc. 86 (1991) 388-395.

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

[7] G.S. Datta, B. Day, I.V. Basawa, Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation, J. Statist. Plan. Inf. 75 (1999) 269–279.

[8] G.S. Datta, B. Day, T. Maiti, Multivariate Bayesian small area estimation: An application to survey and satelite data, Sankhya 60 (1998) 344–362.

[9] G.S. Datta, M. Ghosh, Small area shrinkage estimation, Statist. Science 27 (2012) 95–114.

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

[11] L. Diao, D.D. Smith, G.S. Datta, T. Maiti, J. Opsomer, Accurate confidence interval estimation of small area parameters under the Fay-Herriot model, Scand. J. Statist. 41 (2014) 497–515.

[12] B. Efron, C. Morris, Data analysis using Stein's estimator and its generalizations, J. Amer. Statist. Assoc. 70 (1975), 311–319.

[13] B. Efron, C. Morris, Multivariate empirical Bayes estimation of covariance matrices, Ann. Statist. 4 (1976) 22–32.

[14] R. Fay, Application of multivariate regression to small domain estimation, Small Area Statistics (R. Platek, J.N.K. Rao, C.E. Sarndal, M.P. Singh, eds) (1987) 91–102.

[15] R. Fay, R. Herriot, Estimators of income for small area places: an application of James–Stein procedures to census, J. Amer. Statist. Assoc. 74 (1979) 341–353.

[16] W.A. Fuller, R.M. Harter, The multivariate components of variance model for small area estimation, Small Area Statistics (R. Platek, J.N.K. Rao, C.E. Sarndal, M.P. Singh, eds) (1987) 103–137.

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

[18] T. Kubokawa, Corrected empirical Bayes confidence intervals in nested error regression models, J. Korean Statist. Soc. 39 (2010) 221–236.

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

[20] H. Li, P.Lahiri, An adjusted maximum likelihood method for solving small area estimation problems, J. Multivariate Anal. 101 (2010) 882–892.

[21] I. Ngaruye, D. von Rosen, M. Singull, Crop yield estimation at district level for agricultural seasons 2014 in Rwanda, African J. Applied Statist. 3 (2016) 69-90.

[22] D. Pfeffermann, New important developments in small area estimation, Statist. Science 28 (2013) 40–68.

[23] A.T. Porter, C.K. Wikle, S.H. Holan, Small area estimation via multivariate Fay-Herriot models with latent spatial dependence, Australian and New Zealand J. Statist. 57 (2015) 15–29.

[24] N.G.N. Prasad, J.N.K. Rao, The estimation of mean squared errors of small area estimators, J. Am. Statist. Assoc. 85 (1990) 163–171.

[25] J.N.K. Rao, I. Molina, Small Area Estimation, 2nd Edition (2015) Wiley.

[26] S. Sugasawa, T. Kubokawa, Parametric transformed Fay-Herriot model for small area estimation, J. Multivariate Anal. 139 (2015) 295–311.

[27] M. Yoshimori, P. Lahiri, A second-order efficient empirical Bayes confidence interval, Ann. Statist. 42 (2014) 1233–1261.