Japanese Journal of Statistics and Data Science (2021) 4:841–859
https://doi.org/10.1007/s42081-021-00138-8

ORIGINAL PAPER

General unbiased estimating equations for variance components in linear mixed models

T. Kubokawa[1] · S. Sugasawa[2] · H. Tamae[3] · S. Chaudhuri[4]

Received: 16 May 2021 / Revised: 25 August 2021 / Accepted: 29 August 2021 /
Published online: 8 September 2021

Abstract

This paper introduces a general framework for estimating variance components in the linear mixed models via general unbiased estimating equations, which include some well-used estimators such as the restricted maximum likelihood estimator. We derive the asymptotic covariance matrices and second-order biases under general estimating equations without assuming the normality of the underlying distributions and identify a class of second-order unbiased estimators of variance components. It is also shown that the asymptotic covariance matrices and second-order biases do not depend on whether the regression coefficients are estimated by the generalized or ordinary least squares methods. We carry out numerical studies to check the performance of the proposed methods based on typical linear mixed models.

Keywords Estimating equation · Linear mixed model · Restricted maximum likelihood · Second-order approximation · Variance component

✉ T. Kubokawa
tatsuya@e.u-tokyo.ac.jp

S. Sugasawa
sugasawa@csis.u-tokyo.ac.jp

H. Tamae
taisho.1603@gmail.com

S. Chaudhuri
sanjay@stat.nus.edu.sg

1 Faculty of Economics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

2 Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

3 Nospare Inc., 2-7-13, Kita-aoyama, Minato-ku, Tokyo 107-0061, Japan

4 Department of Statistics and Applied Probability, National University of Singapore, Block S16, Level 7, 6 Science Drive 2, Singapore 117546, Singapore

1 Introduction

Linear mixed models are widely used in a variety of scientific areas such as small area estimation (Rao & Molina, 2015), longitudinal data analysis (Verbeke & Molenberghs, 2006) and meta-analysis (Boreinstein et al., 2009), and estimation of variance components play an essential role in fitting the models. Estimation of variance components has a long history, and various methods have been suggested in the literature. For example, the analysis of variance estimation (ANOVA), the minimum norm quadratic unbiased estimation (MINQUE), the maximum likelihood estimation (ML), and the restricted maximum likelihood estimation (REML) are well-known methods. See Rao and Kleffe (1988) and Searle et al. (1992) for the details.

This paper is motivated by the derivation of the restricted maximum likelihood estimator. In the linear mixed models, the ML estimator of variance components $\boldsymbol{\psi}$ is the solution of the likelihood equation $S(\boldsymbol{\psi}, \widehat{\boldsymbol{\beta}}^{\mathrm{G}}) = 0$ where $S(\cdot, \cdot)$ is the score function and $\widehat{\boldsymbol{\beta}}^{\mathrm{G}}$ is the generalized least squares (GLS) estimator of regression coefficients $\boldsymbol{\beta}$. Although $E\{S(\boldsymbol{\psi}, \boldsymbol{\beta})\} = 0$ because $S(\boldsymbol{\psi}, \boldsymbol{\beta})$ is the score function, after substituting the estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{G}}$ we have $E\{S(\boldsymbol{\psi}, \widehat{\boldsymbol{\beta}}^{\mathrm{G}})\} = h(\boldsymbol{\psi})$, which is not zero. Despite $S(\boldsymbol{\psi}, \widehat{\boldsymbol{\beta}}^{\mathrm{G}})$ is asymptotically unbiased, the bias is not negligible under moderate sample sizes, which may lead to undesirable estimation performance. To overcome the issue, the corrected equation is $S(\boldsymbol{\psi}, \widehat{\boldsymbol{\beta}}^{\mathrm{G}}) - h(\boldsymbol{\psi}) = 0$, and the solution of the equation gives the REML estimator, which is known to have better performance than the ML estimator. As noted later, the estimating equation is still valid without normality as long as some standard moment assumptions are met.

In this paper, we extend the idea of the unbiased estimating equations to more general situations, where $S(\cdot, \cdot)$ is not necessarily the score function and the underlying distribution is not necessarily normal. We suggest the general class of estimating equations for estimating parameters in covariance matrices of random effects and error terms without assuming the normality. This class includes the REML estimator and the Fay–Herriot estimator (Fay & Herriot, 1979), and the Prasad–Rao estimator (Prasad & Rao, 1990), which have widely used in the small area estimation. We first provide unified formulas of the asymptotic covariance matrices and second-order biases without assuming the normality. The resulting important observation is that the asymptotic covariance matrices and second-order biases do not depend on whether the regression coefficients are estimated by the generalized or ordinary least squares methods, suggesting constructing a simpler estimating equation by using the ordinary least squares estimator. Moreover, owing to the explicit formula for the second-order bias, we derive conditions to ensure that the resulting estimator is second-order unbiased without normality assumption. This is the main contribution of this work since the detailed derivation of the second-order asymptotic properties is quite tricky. We also apply the general theory to two important classes of linear mixed models, Fay–Herriot (Fay & Herriot, 1979) and nested error regression (Battese et al., 1988) models, and the numerical performance of the resulting estimators is investigated through simulation studies.

In this paper, $(\boldsymbol{V})_{ab}$ and $(\boldsymbol{V})^{ab}$ denote the (a, b)-th element of matrix $\boldsymbol{V}$ and the inverse $\boldsymbol{V}^{-1}$. Letting $\partial_a = \partial/\partial\psi_a$ for $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)^\top$, we use the simple notations $\boldsymbol{V}_{(a)} = \partial_a \boldsymbol{V}$ and $\boldsymbol{V}_{(ab)} = \partial_a \partial_b \boldsymbol{V}$ for $a, b = 1, \ldots, k$.

This paper is organized as follows. The general unbiased estimating equations are introduced in Sect. 2 with the second-order biases and asymptotic covariance matrices of the resulting estimators. Some specific estimators and their asymptotic properties are given in Sect. 3. A numerical investigation is given in Sect. 4, and all the proofs are given in the Appendix.

2 General estimating equations for variance components

2.1 Settings and restricted maximum likelihood estimator

Consider the linear mixed model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an observable N-variate vector, $\boldsymbol{\beta}$ is a p-variate vector of unknown regression coefficients, and $\boldsymbol{X}$ is an $N \times p$ known matrix of covariates, and $\boldsymbol{Z}$ is an $N \times m$ design matrix. Here, $\boldsymbol{v}$ is a vector of random effects, and $\boldsymbol{\varepsilon}$ is a vector of sampling errors. It is only assumed that $\boldsymbol{v}$ and $\boldsymbol{\varepsilon}$ are mutually independent and distributed as $E(\boldsymbol{v}) = \boldsymbol{0}$, $\mathbf{Cov}\,(\boldsymbol{v}) = \boldsymbol{R}_v(\boldsymbol{\psi}) = \boldsymbol{R}_v$, $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $\mathbf{Cov}\,(\boldsymbol{\varepsilon}) = \boldsymbol{R}_e(\boldsymbol{\psi}) = \boldsymbol{R}_e$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_k)^\top$ is a vector of unknown parameters including variance components. Also, the fourth moments are described as $E[\{(\boldsymbol{R}_e^{-1/2}\boldsymbol{\varepsilon})_i\}^4] = K_e + 3$ and $E[\{(\boldsymbol{R}_v^{-1/2}\boldsymbol{v})_i\}^4] = K_v + 3$, where $(\boldsymbol{a})_i$ is the i-th element of vector $\boldsymbol{a}$, and $\boldsymbol{A}^{1/2}$ is the symmetric root matrix of matrix $\boldsymbol{A}$. Then, $E(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $\mathbf{Cov}\,(\boldsymbol{y}) = \boldsymbol{\Sigma} = \boldsymbol{R}_e(\boldsymbol{\psi}) + \boldsymbol{Z}\boldsymbol{R}_v(\boldsymbol{\psi})\boldsymbol{Z}^\top$.

Under the normality, the maximum likelihood estimator of $\boldsymbol{\psi}$ is the solution of the equations:

$$(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) - \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}) = 0, \quad a = 1, \ldots, k,$$

where $\widehat{\boldsymbol{\beta}}^{\mathrm{G}} = (\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$ is the GLS estimator. The above estimating equations are approximately unbiased under large N, but the bias is not necessarily negligible under moderate N, leading to bias in the resulting estimator of $\boldsymbol{\psi}$. As a solution, the REML estimator of $\boldsymbol{\psi}$ has been widely used as the solution of the equations:

$$(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) - \mathrm{tr}\,(\boldsymbol{P}\boldsymbol{\Sigma}_{(a)}) = 0, \quad a = 1, \ldots, k,$$

where $\boldsymbol{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}$. An essential property of the above estimating equations is that they are exactly unbiased. Moreover, a key observation for the equations is that the unbiasedness property is still valid without the normality, and it only requires the moment assumptions. In this work, we generalize the REML method for estimating the variance components $\boldsymbol{\psi}$, that is, we consider the general class of unbiased estimating equations and develop a unified asymptotic theory for the resulting estimator of $\boldsymbol{\psi}$.

2.2 General estimating equations for variance parameters

Let $\widehat{\boldsymbol{\beta}} = \boldsymbol{L}\boldsymbol{y}$ be a linear unbiased estimator of $\boldsymbol{\beta}$, where $\boldsymbol{L} = \boldsymbol{L}(\boldsymbol{\psi})$ is a $p \times N$ matrix of functions of $\boldsymbol{\psi}$ and satisfies $\boldsymbol{L}\boldsymbol{X} = \boldsymbol{I}$. Let $\boldsymbol{W}_a = \boldsymbol{W}_a(\boldsymbol{\psi})$ be an $N \times N$ matrix of functions of $\boldsymbol{\psi}$ for $a = 1, \ldots, k$. The expectation $E\{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top \boldsymbol{W}_a(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})\}$ is $\mathrm{tr}\,(\boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q}\boldsymbol{\Sigma})$ for $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{X}\boldsymbol{L}$, which gives the general estimating equations:

$$\boldsymbol{y}^\top \boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q}\boldsymbol{y} - \mathrm{tr}\,(\boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q}\boldsymbol{\Sigma}) = 0, \quad a = 1, \ldots, k. \tag{1}$$

For example, the choice of $\boldsymbol{W}_a = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}$ leads to the REML estimation, and other choices of $\boldsymbol{W}_a$ lead to different estimators of $\boldsymbol{\psi}$. In the following theorem, we provide the second-order bias and asymptotic covariance matrix of the general estimator $\widehat{\boldsymbol{\psi}}$ as the solution of (1). Define $k \times k$ matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\widetilde{\boldsymbol{B}}$ by

$$\begin{aligned}
&(\boldsymbol{A})_{ab} = \mathrm{tr}\,(\boldsymbol{W}_a \boldsymbol{\Sigma}_{(b)}), \quad (\boldsymbol{B})_{ab} = \mathrm{tr}\,(\boldsymbol{W}_a \boldsymbol{\Sigma} \boldsymbol{W}_b \boldsymbol{\Sigma}),\\
&(\widetilde{\boldsymbol{B}})_{ab} = K_e h_e(\boldsymbol{W}_a, \boldsymbol{W}_b) + K_v h_v(\boldsymbol{W}_a, \boldsymbol{W}_b),
\end{aligned} \tag{2}$$

where for matrices $\boldsymbol{C}$ and $\boldsymbol{D}$,

$$\begin{aligned}
h_e(\boldsymbol{C}, \boldsymbol{D}) &= \sum_{i=1}^{N} (\boldsymbol{R}_e^{1/2}\boldsymbol{C}\boldsymbol{R}_e^{1/2})_{ii} \cdot (\boldsymbol{R}_e^{1/2}\boldsymbol{D}\boldsymbol{R}_e^{1/2})_{ii},\\
h_v(\boldsymbol{C}, \boldsymbol{D}) &= \sum_{i=1}^{m} (\boldsymbol{R}_v^{1/2}\boldsymbol{Z}^\top\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v^{1/2})_{ii} \cdot (\boldsymbol{R}_v^{1/2}\boldsymbol{Z}^\top\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v^{1/2})_{ii}.
\end{aligned}$$

Theorem 2.1 *Assume that* $(\boldsymbol{L}\boldsymbol{X}^\top \boldsymbol{W}_a \boldsymbol{X}\boldsymbol{L})_{ij} = O(N^{-1})$, $\boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{L}^\top = O(N^{-1})$, *and* $(\boldsymbol{X}\boldsymbol{L})_{ij} = O(N^{-1})$ *as* $N \to \infty$. *Then,* $\mathbf{Cov}\,(\widehat{\boldsymbol{\psi}}) = 2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1} + O(N^{-3/2})$ *and*

$$\begin{aligned}
E(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = {}& 2\boldsymbol{A}^{-1}\mathbf{col}_a(\boldsymbol{K}_a\boldsymbol{A}^{-1} - \boldsymbol{H}_a\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})\\
&+ \boldsymbol{A}^{-1}\mathbf{col}_a(\widetilde{\boldsymbol{K}}_a\boldsymbol{A}^{-1} - \boldsymbol{H}_a\boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1}) + O(N^{-3/2}),
\end{aligned} \tag{3}$$

where $\mathbf{col}_a(x_a) = (x_1, \ldots, x_k)^\top$ *is a column vector with the a-th element* x_a, *and the* $k \times k$ *matrices* $\boldsymbol{K}$, $\boldsymbol{H}$ *and* $\widetilde{\boldsymbol{K}}$ *are defined by*

$$\begin{aligned}
(\boldsymbol{K}_a)_{bc} &= \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}\boldsymbol{W}_c\boldsymbol{\Sigma}),\\
(\boldsymbol{H}_a)_{bc} &= \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) + 2^{-1}\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(bc)}),\\
(\widetilde{\boldsymbol{K}}_a)_{bc} &= K_e h_e(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c) + K_v h_v(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c).
\end{aligned}$$

Two typical choices of $\boldsymbol{L}$ are $\boldsymbol{L}^{\mathrm{G}} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{L}^{\mathrm{O}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$, which correspond to the GLS estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{G}}$ and ordinary least squares (OLS) estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{O}}$. However, Theorem 2.1 tells us that the second-order bias and the asymptotic covariance matrix do not depend on such a choice of $\boldsymbol{L}$. This is an essential observation from Theorem 2.1, and the specific form of $\widehat{\boldsymbol{\beta}}$ in the estimating equation (1) is irrelevant to the asymptotic properties of $\widehat{\boldsymbol{\psi}}$ as long as $\widehat{\boldsymbol{\beta}}$ is unbiased. Hence, it would be better to use a simpler form of $\widehat{\boldsymbol{\beta}}$, so in what follows, we employ $\boldsymbol{L} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$, corresponding to the ordinary least squares estimators of $\boldsymbol{\beta}$. On the other hand, the choice of $\boldsymbol{W}_a$ affects the asymptotic properties.

The second-order unbiasedness is one of the desirable properties of estimators $\widehat{\boldsymbol{\psi}}$. From Theorem 2.1, we need to use $\boldsymbol{W}_a$ such that the leading term in (3) is 0 to achieve second-order unbiasedness of $\widehat{\boldsymbol{\psi}}$. In typical linear mixed models such as the Fay–Herriot and nested error regression models, the covariance matrix $\boldsymbol{\Sigma}$ is a linear function of $\boldsymbol{\psi}$. In this case, $\boldsymbol{\Sigma}_{(bc)} = 0$, which simplifies the condition for the second-order unbiasedness in (3), because $(\boldsymbol{H}_a)_{bc} = \mathrm{tr}\,(\boldsymbol{W}_{a(b)} \boldsymbol{\Sigma}_{(c)})$. When $K_e = K_v = 0$, the estimator $\widehat{\boldsymbol{\psi}}$ is second-order unbiased if

$$\boldsymbol{K}_a = \boldsymbol{H}_a \boldsymbol{A}^{-1} \boldsymbol{B}. \tag{4}$$

This condition is investigated in the next section for some specific choices of $\boldsymbol{W}_a$.

3 Specific estimators and their asymptotic properties

3.1 Three estimators

We now describe some specific estimators of $\boldsymbol{\psi}$ and provide their asymptotic variances and biases. In what follows, we assume that $\boldsymbol{\Sigma}$ is a linear function of $\boldsymbol{\psi}$, which are satisfied in typical linear mixed models such as the Fay–Herriot and nested error regression models. We here consider the three candidates for $\boldsymbol{W}_a$; $\boldsymbol{W}_a^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{(a)} \boldsymbol{\Sigma}^{-1}$, $\boldsymbol{W}_a^{\mathrm{FH}} = (\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{(a)} + \boldsymbol{\Sigma}_{(a)} \boldsymbol{\Sigma}^{-1})/2$ and $\boldsymbol{W}_a^{\mathrm{Q}} = \boldsymbol{\Sigma}_{(a)}$, which are motivated from the REML estimator, the Fay–Herriot moment estimator (Fay & Herriot, 1979) and the Prasad–Rao unbiased estimator (Prasad & Rao, 1990) under the Fay–Herriot model. The estimators induced from $\boldsymbol{W}_a^{\mathrm{RE}}$, $\boldsymbol{W}_a^{\mathrm{FH}}$ and $\boldsymbol{W}_a^{\mathrm{Q}}$ are called here the REML-type, FH-type and PR-type estimators, respectively. From Theorem 2.1, we can derive the asymptotic properties of the three estimators. When $\boldsymbol{\Sigma}$ is a linear function of $\boldsymbol{\psi}$, the asymptotic variances and second-order biases are simplified in the case of $K_e = K_v = 0$, which is satisfied in the normal distributions.

Proposition 3.1 *Assume the conditions in Theorem 2.1 and that $\boldsymbol{\Sigma}$ is a linear function of $\boldsymbol{\psi}$. Also assume that $K_e = K_v = 0$. Let $\widehat{\boldsymbol{\psi}}^{\mathrm{RE}}$, $\widehat{\boldsymbol{\psi}}^{\mathrm{FH}}$ and $\widehat{\boldsymbol{\psi}}^{\mathrm{Q}}$ be the estimators based on $\boldsymbol{W}_a^{\mathrm{RE}}$, $\boldsymbol{W}_a^{\mathrm{FH}}$ and $\boldsymbol{W}_a^{\mathrm{Q}}$, respectively. Then the following results hold.*

(a) *REML-type estimator $\widehat{\boldsymbol{\psi}}^{\mathrm{RE}}$ is second-order unbiased and has the asymptotic covariance matrix $2\boldsymbol{A}_{\mathrm{RE}}^{-1}$, where $(\boldsymbol{A}_{\mathrm{RE}})_{ij} = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{(i)} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{(j)})$.*

(b) *FH-type estimator* $\widehat{\boldsymbol{\psi}}^{\mathrm{FH}}$ *is not second-order unbiased. The asymptotic covariance matrix is*

$$2\boldsymbol{A}_{\mathrm{FH}}^{-1}\boldsymbol{B}_{\mathrm{FH}}\boldsymbol{A}_{\mathrm{FH}}^{-1},$$

for $(\boldsymbol{A}_{\mathrm{FH}})_{ij} = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}_{(j)})$ *and* $(\boldsymbol{B}_{\mathrm{FH}})_{ij} = \{\mathrm{tr}\,(\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}_{(j)}) + \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(j)})\}/2$. *The second-order bias is*

$$2\boldsymbol{A}_{\mathrm{FH}}^{-1}\mathbf{col}_a(\boldsymbol{K}_a\boldsymbol{A}_{\mathrm{FH}}^{-1} - \boldsymbol{H}_a\boldsymbol{A}_{\mathrm{FH}}^{-1}\boldsymbol{B}_{\mathrm{FH}}\boldsymbol{A}_{\mathrm{FH}}^{-1}),$$

for $(\boldsymbol{K}_a)_{bc} = -\mathrm{tr}\,\{\boldsymbol{\Sigma}_{(a)}(\boldsymbol{\Sigma}_{b)} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}_{(c)}\boldsymbol{\Sigma}^{-1}\}/2$ *and* $(\boldsymbol{H}_a)_{bc} = -\mathrm{tr}\,\{\boldsymbol{\Sigma}_{a)}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}_{(c)}\boldsymbol{\Sigma}^{-1}$.

(c) *PR-type estimator* $\widehat{\boldsymbol{\psi}}^{\mathrm{Q}}$ *is second-order unbiased and has the asymptotic covariance matrix* $2\boldsymbol{A}_{\mathrm{Q}}^{-1}\boldsymbol{B}_{\mathrm{Q}}\boldsymbol{A}_{\mathrm{Q}}^{-1}$, *where* $(\boldsymbol{A}_{\mathrm{Q}})_{ij} = \mathrm{tr}\,(\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}_{(j)})$ *and* $(\boldsymbol{B}_{\mathrm{Q}})_{ij} = \mathrm{tr}\,(\boldsymbol{\Sigma}_{(i)}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(j)}\boldsymbol{\Sigma})$.

In Proposition 3.1, the linearity of $\boldsymbol{\Sigma}(\boldsymbol{\psi})$ on $\boldsymbol{\psi}$ is only used to compute the second-order bias. The expressions for the asymptotic covariances hold, in general, without such constraints. Without assuming $K_e = K_v = 0$, the estimator $\widehat{\boldsymbol{\psi}}^{\mathrm{RE}}$ has the second-order bias, while $\widehat{\boldsymbol{\psi}}^{\mathrm{Q}}$ remains second-order unbiased.

It is noted that the REML-type is the most efficient in the normal distributions, which corresponds to the case of $K_e = K_v = 0$. This implies that the following inequality holds for any $\boldsymbol{W}_a$:

$$\begin{aligned}&[\mathbf{mat}_{a,b}\{\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)})\}]^{-1}\mathbf{mat}_{a,b}\{\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}\boldsymbol{W}_b\boldsymbol{\Sigma})\}[\mathbf{mat}_{a,b}\{\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)})\}]^{-1}\\ &\quad\geq [\mathbf{mat}_{a,b}\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)})\}]^{-1},\end{aligned} \tag{5}$$

where $\mathbf{mat}_{ab}\{x_{ab}\}$ is a $k \times k$ matrix with the (a, b)-th element x_{ab}. However, it should be remarked that REML is not necessarily efficient without assuming $K_e = 0$ and $K_v = 0$.

3.2 Detailed properties under two useful models

We provide more detailed formulas for the asymptotic covariances (or variances) and second-order biases under the Fay–Herriot and nested error regression models.

The first model is the Fay–Herriot model, which corresponds to $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$, $N = m$, $\boldsymbol{R}_v = \psi_1\boldsymbol{I}_m$, $\boldsymbol{R}_e = \boldsymbol{D} = \mathrm{diag}\,(D_1, \ldots, D_m)$ and $\boldsymbol{\Sigma} = \psi_1\boldsymbol{I}_m + \boldsymbol{D}$ for known D_i's. The following proposition can be derived from Theorem 2.1.

Proposition 3.2 *In the Fay–Herriot model, estimator* $\widehat{\psi}_1$ *is the solution of* (1) *for diagonal matrix* $\boldsymbol{W}_1$. *Without assuming* $K_e = K_v = 0$, *the asymptotic variance of* $\widehat{\psi}_1$ *is*

$$\mathrm{Var}(\widehat{\psi}_1) \approx 2\frac{\mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma})}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2} + \frac{K_e\mathrm{tr}\,(\boldsymbol{W}_1^2\boldsymbol{D}^2) + \psi_1^2K_v\mathrm{tr}\,(\boldsymbol{W}_1^2)}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2},$$

and the second-order bias is

$$\mathrm{Bias}(\widehat{\psi}_1) \approx 2\frac{\mathrm{tr}\,(\boldsymbol{W}_{1(1)}\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma}) - \mathrm{tr}\,(\boldsymbol{W}_{1(1)})\mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma})}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2} + \frac{K_e\mathrm{tr}\,(\boldsymbol{W}_{1(1)}\boldsymbol{W}_1\boldsymbol{D}^2) + \psi_1^2 K_v\mathrm{tr}\,(\boldsymbol{W}_{1(1)}\boldsymbol{W}_1)}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2} - \frac{\mathrm{tr}\,(\boldsymbol{W}_{1(1)})\{K_e\mathrm{tr}\,(\boldsymbol{W}_1^2\boldsymbol{D}^2) + \psi_1^2 K_v\mathrm{tr}\,(\boldsymbol{W}_1^2)\}}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^3}.$$

In this model, the inequality (5) is expressed as

$$\frac{\mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma})}{\{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2} \geq \frac{1}{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})},$$

or $\mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma})\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}) \geq \{\mathrm{tr}\,(\boldsymbol{W}_1)\}^2$. This inequality can be directly proved by using the Cauchy–Schwarz inequality.

The second example is the nested error regression model. Let $\boldsymbol{Z} = \mathrm{block\,diag}(\boldsymbol{j}_{n_1}, \ldots, \boldsymbol{j}_{n_m})$ for $\boldsymbol{j}_{n_i} = (1, \ldots, 1)^\top \in \mathbb{R}^{n_i}$, and let $\boldsymbol{G} = \mathrm{block\,diag}(\boldsymbol{J}_{n_1}, \ldots, \boldsymbol{J}_{n_m})$ for $\boldsymbol{J}_{n_i} = \boldsymbol{j}_{n_i}\boldsymbol{j}_{n_i}^\top$. This model corresponds to $N = \sum_{i=1}^m n_i$, $\boldsymbol{R}_v = \psi_1\boldsymbol{G}$, $\boldsymbol{R}_e = \psi_2\boldsymbol{I}_N$ and $\boldsymbol{\Sigma} = \psi_1\boldsymbol{G} + \psi_2\boldsymbol{I}_N$. Note that $\boldsymbol{\Sigma}_{(1)} = \boldsymbol{G}$ and $\boldsymbol{\Sigma}_{(2)} = \boldsymbol{I}_N$. Then,

$$\boldsymbol{A} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{W}_1) \\ \mathrm{tr}\,(\boldsymbol{W}_2\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{W}_2) \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_1\boldsymbol{\Sigma}) & \mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_2\boldsymbol{\Sigma}) \\ \mathrm{tr}\,(\boldsymbol{W}_1\boldsymbol{\Sigma}\boldsymbol{W}_2\boldsymbol{\Sigma}) & \mathrm{tr}\,(\boldsymbol{W}_2\boldsymbol{\Sigma}\boldsymbol{W}_2\boldsymbol{\Sigma}) \end{pmatrix},$$

and $(\widetilde{\boldsymbol{B}})_{ab} = \psi_2^2 K_e \sum_{i=1}^N (\boldsymbol{W}_a)_{ii}(\boldsymbol{W}_b)_{ii} + \psi_1^2 K_v \sum_{i=1}^m (\boldsymbol{Z}^\top\boldsymbol{W}_a\boldsymbol{Z})_{ii}(\boldsymbol{Z}^\top\boldsymbol{W}_b\boldsymbol{Z})_{ii}$. The following proposition is provided from Theorem 2.1.

Proposition 3.3 *In the nested error regression model, estimator* $\widehat{\boldsymbol{\psi}}$ *is the solution of* (1). *Without assuming* $K_e = K_v = 0$, *the asymptotic covariance matrix of* $\widehat{\boldsymbol{\psi}}$ *is*

$$\mathbf{Cov}\,(\widehat{\boldsymbol{\psi}}) \approx 2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1}, \tag{6}$$

and the second-order bias of $\widehat{\boldsymbol{\psi}}$ *is*

$$\mathbf{Bias}(\widehat{\boldsymbol{\psi}}) \approx 2\boldsymbol{A}^{-1}\begin{pmatrix} \mathrm{tr}\,(\boldsymbol{K}_1\boldsymbol{A}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_1\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}) \\ \mathrm{tr}\,(\boldsymbol{K}_2\boldsymbol{A}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}) \end{pmatrix} + \boldsymbol{A}^{-1}\begin{pmatrix} \mathrm{tr}\,(\widetilde{\boldsymbol{K}}_1\boldsymbol{A}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_1\boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1}) \\ \mathrm{tr}\,(\widetilde{\boldsymbol{K}}_2\boldsymbol{A}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_2\boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1}) \end{pmatrix}, \tag{7}$$

where $(\boldsymbol{K}_a)_{bc} = \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}\boldsymbol{W}_c\boldsymbol{\Sigma})$, $(\boldsymbol{H}_a)_{bc} = \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)})$ *and* $(\widetilde{\boldsymbol{K}}_a)_{bc} = \psi_2^2 K_e \sum_{i=1}^N (\boldsymbol{W}_{a(b)})_{ii}(\boldsymbol{W}_c)_{ii} + \psi_1^2 K_v \sum_{i=1}^m (\boldsymbol{Z}^\top\boldsymbol{W}_{a(b)}\boldsymbol{Z})_{ii}(\boldsymbol{Z}^\top\boldsymbol{W}_c\boldsymbol{Z})_{ii}$ *for* $a = 1, 2$.

4 Simulation studies

We here investigate the finite-sample performance of some estimators obtained from the estimating equation (1) in the two typical linear mixed models, the Fay–Herriot model and the nested error regression model. As distributions of the random effects and the error terms in the both models, we treat the normal distribution and the t-distribution with 6 degrees of freedom.

We first consider the Fay–Herriot model studied in Fay and Herriot (1979) as a simple area-level linear mixed model, which is described as $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + v_i + \varepsilon_i$ for $i = 1, \dots, m$. There are five groups $G_1, \dots, G_5$ and six small areas in each group, that is, we have $m = 5 \times 6 = 30$. The sampling variances D_i's are the same for area within the same group, and we consider the D_i-pattern (1.4, 1.2, 1.0, 0.8, 0.6). For $p = 3$, we set $\boldsymbol{\beta} = \boldsymbol{j}_p$ and construct $\boldsymbol{x}_i$ as $\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{u}_i$ and fix it, where $\boldsymbol{A}$ is the Cholesky decomposition of $0.8\boldsymbol{I}_p + 0.2\boldsymbol{J}_p$ and $\boldsymbol{u}_i$ is a p-variate value generated from $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I}_p)$ for $i = 1, \dots, m$. In this model, $\boldsymbol{\Sigma} = \psi_1 \boldsymbol{I}_m + \boldsymbol{D}$. Since the average of D_i's is 1.0, we consider the three cases of $\psi_1 = 0.2$, 1 and 5.

The estimators which we compare are the ML estimator, the REML estimator, the REML-type estimator based on OLS of $\boldsymbol{\beta}$, the Fay–Herriot estimator, the Fay–Herriot-type estimator based on OLS of $\boldsymbol{\beta}$, the bias-corrected Fay–Herriot estimator and the Prasad–Rao estimator, which are denoted by ML, RE, ORE, FH, OFH, UFH and PR, respectively. Note that the Prasad–Rao estimator is identical to the second-order unbiased estimator constructed from $\boldsymbol{W}_a^{\mathrm{Q}}$ in this model under normality. The detailed forms of estimators are provided in the Appendix. In particular, the solutions of the estimating equations are computed by the solver of non-linear equations in Ox (Doornik, 2007). The values of their biases (Bias), standard deviations (SD) and square roots of mean squared errors (RMSE) are calculated by simulation with 10,000 replications. Those values are reported in Table 1 when the standard normal distributions are assumed for $v_i/\sqrt{\psi_1}$ and $\varepsilon_i/\sqrt{D_i}$, and in Table 2 when we assume $t_\nu/\sqrt{\nu/(\nu-2)}$ instead of the normality, where t_ν is a random variable having the t-distribution with ν degrees of freedom. We here set $\nu = 6$.

The REML estimator is a bias-corrected procedure of ML up to second order under the normality. We can confirm this fact, because RE has a smaller bias than ML in Table 1. On the other hand, RE has a larger SD, which turns into a larger RMSE than ML. Although UFH is a bias-corrected procedure of FH up to second order, there are little difference in bias, SD and RMSE between FH and UFH. It is also revealed from Table 1 that RE and FH have similar performances to ORE and OFH, respectively. This supports the results of Theorem 2.1, namely, the second-order bias and the asymptotic covariance matrix do not depend on whether $\boldsymbol{\beta}$ is estimated by the GLS or OLS estimators of $\boldsymbol{\beta}$. In light of values of RMSE, ML$\succ$(RE, ORE)$\succ$(FH, OFH, UFH)$\succ$PR for all the cases of ψ_1, where ML$\succ$RE means that ML is better than RE.

As a non-normal distribution, we treat the t-distribution with 6 degrees of freedom, and the simulation results are reported in Table 2. Comparing Tables 1 and 2, the values of SD and RMSE under the t-distributions are larger than those under the normality. However, relative performances of the estimators under the t-distributions

Table 1 Values of bias, standard deviation (SD) and square-root of MSE for the seven estimators of ψ_1 in the Fay–Herriot model under normality

	ML	RE	ORE	FH	OFH	UFH	PR
$\psi_1 = 0.2$							
Bias	−0.0498	0.0383	0.0387	0.0427	0.0427	0.0395	0.0485
SD	0.2003	0.2543	0.2548	0.2609	0.2610	0.2603	0.2740
RMSE	0.2064	0.2572	0.2577	0.2644	0.2645	0.2633	0.2783
ψ_1=1							
Bias	−0.1966	−0.0021	−0.0022	−0.0002	−0.0002	−0.0031	−0.0031
SD	0.4694	0.5255	0.5255	0.5286	0.5286	0.5293	0.5413
RMSE	0.5089	0.5255	0.5255	0.5286	0.5286	0.5293	0.5413
ψ_1=5							
Bias	−0.5765	0.0200	0.0199	0.0220	0.0220	0.0210	0.0221
SD	1.4755	1.6398	1.6398	1.6420	1.6420	1.6423	1.6478
RMSE	1.5841	1.6399	1.6400	1.6422	1.6422	1.6424	1.6480

Table 2 Values of bias, standard deviation (SD) and square-root of MSE for the seven estimators of ψ_1 in the Fay–Herriot model under *t*-distributions

	ML	RE	ORE	FH	OFH	UFH	PR
$\psi_1 = 0.2$							
Bias	−0.0128	0.0705	0.0718	0.0729	0.0729	0.0702	0.0788
SD	0.3078	0.3711	0.3717	0.3824	0.3824	0.3819	0.3990
RMSE	0.3080	0.3778	0.3786	0.3893	0.3893	0.3883	0.4067
ψ_1=1							
Bias	−0.1868	0.0012	0.0010	0.0037	0.0037	0.0008	0.0008
SD	0.6101	0.6837	0.6839	0.6888	0.6888	0.6895	0.70203
RMSE	0.6381	0.6837	0.6839	0.6888	0.6888	0.6895	0.7020
ψ_1=5							
Bias	−0.6299	−0.0259	−0.0258	−0.0252	−0.0252	−0.0262	−0.0270
SD	1.9817	2.2017	2.2017	2.2009	2.2009	2.2012	2.2036
RMSE	2.0795	2.2019	2.2018	2.2011	2.2011	2.2014	2.2038

are quite similar to the performances under the normality. Under the t-distributions, RE and UFH are not second-order unbiased, and the bias of ML is smaller than RE for $\psi_1 = 0.2$, while RE has smaller biases for $\psi_1 = 1, 5$.

We next consider the nested error regression model studied in Battese et al. (1988) as a unit-level random intercept model, which is described as $y_{ij} = x_{ij}^\top \beta + v_i + \varepsilon_{ij}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$ with $E(v_i) = 0$, $\mathrm{Var}(v_i) = \psi_1$, $E(\varepsilon_{ij}) = 0$, $\mathrm{Var}(\varepsilon_{ij}) = \psi_2$. We set up $\boldsymbol{\beta}$ and $\boldsymbol{x}_{ij}$ as in the Fay–Herriot model for $p = 3$. The number of clusters is $m = 20$, and 20 clusters are equally divided into five groups, and the sample sizes n_i are the same for clusters within the same group. We consider the n_i-pattern (4, 4, 5, 6, 6), so that the total sample size is $N = 100$. As estimators of the two variance components ψ_1 and ψ_2, we investigate the ML estimator, the REML estimator, the REML estimator based on OLS of $\boldsymbol{\beta}$, the Fay–Herriot estimator, the Prasad–Rao estimator and the PR-type second-order unbiased estimator constructed from $\boldsymbol{W}_a^{\mathrm{Q}}$, which are denoted by ML, RE, ORE, FH, PR and Q, respectively. The detailed forms of estimators are provided in the Appendix. The values of their biases (Bias), standard deviations (SD) and square roots of mean squared errors (RMSE) are calculated by simulation with 1000 replications for $\psi_1 = 0.2, 1, 5$ and $\psi_2 = 5$. Those values are reported in Table 3 under the normality and in Table 4 under the t-distribution with ν degrees of freedom for $\nu = 6$, where the first three columns are for estimation of ψ_1 and the second three columns are for estimation of ψ_2. Since the variance of the sample mean $n_i^{-1} \sum_{j=1}^{n_i} y_{ij}$ is ψ_2/n_i, the average of the variances is $m^{-1} \sum_{i=1}^{m} \psi_2/n_i$, which is close to one for the above n_i-pattern and $\psi_2 = 5$. This suggests that our setup of ψ_1 and ψ_2 is close to the setup in the Fay–Herriot model.

The overall features of the estimators in Tables 3 and 4 are similar to those in Tables 1 and 2, that is, RE has smaller biases, but larger SD than ML, and the values of SD and RMSE in the non-normal distributions are larger than those under the normality. Although RE and ORE have similar performances, ORE has smaller values for $\psi_1 = 0.2$, but larger values for $\psi_1 = 5$. Since PR and Q are second-order unbiased, their biases are small. However, their SD and RMSE are large and give the worst values in estimation of ψ_2 at $\psi_1 = 5$. In light of RMSE, the ML, RE and FH estimators are recommendable.

Acknowledgements We would like to thank the Associate Editor and the two reviewers for many valuable comments and helpful suggestions, which led to an improved version of this paper. This research was supported in part by Grant-in-Aid for Scientific Research (18K11188) from the Japan Society for the Promotion of Science.

Appendix A: Proofs

A.1 A preliminary lemma

For the proof, we use the following lemma:

Lemma A.1 *Let $\boldsymbol{u} = \boldsymbol{\varepsilon} + \boldsymbol{Z}\boldsymbol{v}$. Then, for matrices $\boldsymbol{C}$ and $\boldsymbol{D}$, it holds that*

$$E[\boldsymbol{u}^\top \boldsymbol{C}\boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{D}\boldsymbol{u}] = 2\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{D}\boldsymbol{\Sigma}) + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{\Sigma})\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{\Sigma}) + K_e h_e(\boldsymbol{C}, \boldsymbol{D}) + K_v h_v(\boldsymbol{C}, \boldsymbol{D}), \tag{8}$$

where $h_e(\boldsymbol{C}, \boldsymbol{D})$ and $h_v(\boldsymbol{C}, \boldsymbol{D})$ are given in Theorem 2.1.

Proof It is demonstrated that $E[\boldsymbol{u}^\top \boldsymbol{C}\boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{D}\boldsymbol{u}] = E[\boldsymbol{\varepsilon}^\top \boldsymbol{C}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \boldsymbol{D}\boldsymbol{\varepsilon}] + E[\boldsymbol{v}^\top \boldsymbol{Z}^\top \boldsymbol{C}\boldsymbol{Z}\boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z}\boldsymbol{v}] + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + \mathrm{tr}\,(\boldsymbol{D}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + 4\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top)$. Let $\boldsymbol{x} = (x_1, \ldots, x_N)^\top = \boldsymbol{R}_e^{-1/2}\boldsymbol{\varepsilon}$, $\widetilde{\boldsymbol{C}} = \boldsymbol{R}_e^{1/2}\boldsymbol{C}\boldsymbol{R}_e^{1/2}$ and $\widetilde{\boldsymbol{D}} = \boldsymbol{R}_e^{1/2}\boldsymbol{D}\boldsymbol{R}_e^{1/2}$. Then, $E[\boldsymbol{x}] = \boldsymbol{0}$, $E[\boldsymbol{x}\boldsymbol{x}^\top] = \boldsymbol{I}_N$, $E[x_a^4] = K_e + 3$, $a = 1, \ldots, N$, and $E[\boldsymbol{\varepsilon}^\top \boldsymbol{C}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \boldsymbol{D}\boldsymbol{\varepsilon}] = E[\boldsymbol{x}^\top \widetilde{\boldsymbol{C}}\boldsymbol{x}\boldsymbol{x}^\top \widetilde{\boldsymbol{D}}\boldsymbol{x}]$. Let $\delta_{a=b=c=d} = 1$ for $a = b = c = d$, and

Table 3 Values of bias, standard deviation (SD) and square-root of MSE for the six estimators of ψ_1 and ψ_2 in the nested error regression model under normality

	ML	RE	ORE	FH	PR	Q
Estimation of ψ_1 for $\psi_1 = 0.2$, $\psi_2 = 5$						
Bias	0.0903	0.0891	0.0856	0.0796	0.0790	−0.0011
SD	0.3413	0.3405	0.3342	0.3393	0.3387	0.4359
RMSE	0.3530	0.3520	0.3450	0.3485	0.3478	0.4359
Estimation of ψ_1 for $\psi_1 = 1$, $\psi_2 = 5$						
Bias	−0.0108	−0.0119	−0.0101	−0.0131	0.0672	−0.0266
SD	0.6565	0.6557	0.6680	0.6576	0.7515	0.8037
RMSE	0.6566	0.6558	0.6681	0.6577	0.7545	0.8041
Estimation of ψ_1 for $\psi_1 = 5$, $\psi_2 = 5$						
Bias	−0.0755	−0.0782	−0.0730	−0.0766	−0.0057	−0.1044
SD	1.9456	1.9452	2.0324	1.9635	2.1691	2.1078
RMSE	1.9471	1.9468	2.0338	1.9650	2.1691	2.1104
Estimation of ψ_2 for $\psi_1 = 0.2$, $\psi_2 = 5$						
Bias	−0.2477	−0.0946	0.0412	−0.0194	−0.0881	−0.0079
SD	0.7714	0.7928	0.7823	0.8353	0.7955	0.8453
RMSE	0.8102	0.7985	0.7834	0.8355	0.8003	0.8454
Estimation of ψ_2 for $\psi_1 = 1$, $\psi_2 = 5$						
Bias	−0.1732	−0.0084	0.2696	0.0025	−0.0984	−0.0045
SD	0.7712	0.7943	0.8106	0.8080	1.1290	1.1687
RMSE	0.7904	0.7943	0.85427	0.8080	1.1333	1.1688
Estimation of ψ_2 for $\psi_1 = 5$, $\psi_2 = 5$						
Bias	−0.1701	0.0097	0.3756	0.0123	−0.0934	0.0051
SD	0.7505	0.7772	0.8446	0.7803	1.4337	1.4034
RMSE	0.7695	0.7772	0.9244	0.7804	1.4368	1.4034

otherwise, $\delta_{a=b=c=d} = 0$. The notation $\delta_{a=b\neq c=d}$ is defined similarly. It is observed that for $a, b, c, d = 1, \dots, N$,

$$
\begin{aligned}
&E[x_a(\widetilde{\boldsymbol{C}})_{ab}x_b x_c(\widetilde{\boldsymbol{D}})_{cd}x_d] \\
&\quad = E[x_a^4(\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{aa}\delta_{a=b=c=d} + x_a^2x_c^2(\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{cc}\delta_{a=b\neq c=d} + 2x_a^2x_b^2(\widetilde{\boldsymbol{C}})_{ab}(\widetilde{\boldsymbol{D}})_{ab}\delta_{a=c\neq b=d}] \\
&\quad = (K_e+3)(\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{aa}\delta_{a=b=c=d} + (\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{cc}\delta_{a=b\neq c=d} + 2(\widetilde{\boldsymbol{C}})_{ab}(\widetilde{\boldsymbol{D}})_{ab}\delta_{a=c\neq b=d} \\
&\quad = K_e(\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{aa}\delta_{a=b=c=d} + (\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{cc}\delta_{a=b}\delta_{c=d} + 2(\widetilde{\boldsymbol{C}})_{ab}(\widetilde{\boldsymbol{D}})_{ab}\delta_{a=c}\delta_{b=d},
\end{aligned}
$$

which implies that

$$
\sum_{a,b,c,d} E[x_a(\widetilde{\boldsymbol{C}})_{ab}x_b x_c(\widetilde{\boldsymbol{D}})_{cd}x_d] = K_e\sum_{a=1}^{N}(\widetilde{\boldsymbol{C}})_{aa}(\widetilde{\boldsymbol{D}})_{aa} + \sum_{a=1}^{N}(\widetilde{\boldsymbol{C}})_{aa}\sum_{c=1}^{N}(\widetilde{\boldsymbol{D}})_{cc}
$$

Table 4 Values of bias, standard deviation (SD) and square-root of MSE for the six estimators of ψ_1 and ψ_2 in the nested error regression model under t-distributions

	ML	RE	ORE	FH	PR	Q
Estimation of ψ_1 for $\psi_1 = 0.2$, $\psi_2 = 5$						
Bias	0.0988	0.0973	0.0920	0.0877	0.0860	0.0062
SD	0.3296	0.3288	0.3234	0.3287	0.3278	0.4363
RMSE	0.3441	0.3429	0.3363	0.3402	0.3389	0.4363
Estimation of ψ_1 for $\psi_1 = 1$, $\psi_2 = 5$						
Bias	0.0429	0.0419	0.0478	0.0412	0.1323	0.0368
SD	0.7700	0.7694	0.8083	0.7847	0.8688	0.9042
RMSE	0.7712	0.7705	0.8097	0.7857	0.8788	0.9049
Estimation of ψ_1 for $\psi_1 = 5$, $\psi_2 = 5$						
Bias	0.0557	0.0528	0.0308	0.0345	0.1630	0.0828
SD	2.9158	2.9156	2.9072	2.8519	3.0328	2.9880
RMSE	2.9164	2.9161	2.9074	2.8521	3.0372	2.9891
Estimation of ψ_2 for $\psi_1 = 0.2$, $\psi_2 = 5$						
Bias	−0.2474	−0.0942	0.0502	−0.0170	−0.0852	−0.0054
SD	1.0553	1.0856	1.0905	1.1317	1.0868	1.1506
RMSE	1.0840	1.0897	1.0916	1.1318	1.0901	1.1506
Estimation of ψ_2 for $\psi_1 = 1$, $\psi_2 = 5$						
Bias	−0.1779	−0.0131	0.2643	−0.0048	−0.1038	−0.0083
SD	1.1481	1.1839	1.2174	1.1917	1.6101	1.6592
RMSE	1.1618	1.1840	1.2458	1.1917	1.6134	1.6592
Estimation of ψ_2 for $\psi_1 = 5$, $\psi_2 = 5$						
Bias	−0.1840	−0.0050	0.3663	0.0036	−0.0948	−0.0146
SD	1.0967	1.1358	1.2315	1.1414	2.0755	2.0095
RMSE	1.1120	1.1358	1.2849	1.1414	2.0777	2.0095

$$+2\sum_{a=1}^{N}\sum_{b=1}^{N}(\widetilde{\boldsymbol{C}})_{ab}(\widetilde{\boldsymbol{D}})_{ab},$$

or

$$E[\boldsymbol{\varepsilon}^\top \boldsymbol{C}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \boldsymbol{D}\boldsymbol{\varepsilon}] = 2\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e\boldsymbol{D}\boldsymbol{R}_e) + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{R}_e) + K_e h_e(\boldsymbol{C}, \boldsymbol{D}).$$

Similarly,

$$\begin{aligned}&E[\boldsymbol{v}^\top \boldsymbol{Z}^\top \boldsymbol{C}\boldsymbol{Z}\boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z}\boldsymbol{v}]\\ &\quad = 2\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + K_v h_v(\boldsymbol{C}, \boldsymbol{D}).\end{aligned}$$

Thus, we have

$$
\begin{aligned}
E[\boldsymbol{u}^\top \boldsymbol{C}\boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{D}\boldsymbol{u}] = & \, 2\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e\boldsymbol{D}\boldsymbol{R}_e) + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{R}_e) + 2\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) \\
& + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + \mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) \\
& + \mathrm{tr}\,(\boldsymbol{D}\boldsymbol{R}_e)\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) + 4\mathrm{tr}\,(\boldsymbol{C}\boldsymbol{R}_e\boldsymbol{D}\boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top) \\
& + K_e h_e(\boldsymbol{C}, \boldsymbol{D}) + K_v h_v(\boldsymbol{C}, \boldsymbol{D}),
\end{aligned}
$$

which can be rewritten as the expression in (9) for $\boldsymbol{\Sigma} = \boldsymbol{R}_e + \boldsymbol{Z}\boldsymbol{R}_v\boldsymbol{Z}^\top$. □

A.2 Proof of Theorem 2.1

For $a = 1, \ldots, k$, let $\ell_a = \boldsymbol{y}^\top \boldsymbol{C}_a \boldsymbol{y} - \mathrm{tr}\,(\boldsymbol{D}_a)$ for $\boldsymbol{C}_a = \boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q}$ and $\boldsymbol{D}_a = \boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q}\boldsymbol{\Sigma}$. For $\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\varepsilon} + \boldsymbol{Z}\boldsymbol{v}$, ℓ_a is rewritten as $\ell_a = \boldsymbol{u}^\top \boldsymbol{C}_a \boldsymbol{u} - \mathrm{tr}\,(\boldsymbol{D}_a)$. By the Taylor series expansion,

$$
\begin{aligned}
0 = & \, \mathbf{col}_a(\ell_a) + \mathbf{mat}_{ab}(\ell_{a(b)})(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \\
& + \frac{1}{2}\mathbf{col}_a \left\{ \sum_{b=1}^k \sum_{c=1}^k \ell_{a(bc)}(\widehat{\psi}_b - \psi_b)(\widehat{\psi}_c - \psi_c) \right\} + O_p(N^{-1/2}),
\end{aligned}
$$

where $\mathbf{mat}_{ab}(x_{ab})$ is a $k \times k$ matrix with the (a, b)-th element x_{ab}. Then,

$$
\begin{aligned}
\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi} = & -\{\mathbf{mat}_{ab}(\ell_{a(b)})\}^{-1} \left[\mathbf{col}_a(\ell_a) + \frac{1}{2}\mathbf{col}_a \left\{ \sum_{b=1}^k \sum_{c=1}^k \ell_{a(bc)}(\widehat{\psi}_b - \psi_b)(\widehat{\psi}_c - \psi_c) \right\} \right] \\
& + O_p(N^{-3/2}).
\end{aligned}
$$

Since $\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_a) = \mathrm{tr}\,(\boldsymbol{D}_a)$, we have $\ell_a = \mathrm{tr}\,\{\boldsymbol{C}_a(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}$. In addition, $\ell_{a(b)} = \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(b)} - \boldsymbol{D}_{a(b)}) + \mathrm{tr}\,\{\boldsymbol{C}_{a(b)}(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}$ and $\ell_{a(bc)} = \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(bc)} - \boldsymbol{D}_{a(bc)}) + \mathrm{tr}\,\{\boldsymbol{C}_{a(bc)}(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}$. Let $\boldsymbol{A}_1 = \mathbf{mat}_{ab}\{\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(b)} - \boldsymbol{D}_{a(b)})\}$ and $\boldsymbol{A}_0 = \mathbf{mat}_{ab}[\mathrm{tr}\,\{\boldsymbol{C}_{a(b)}(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}]$. It is noted that $\boldsymbol{A}_1 = O(N)$, $\boldsymbol{A}_0 = O_p(N^{1/2})$, $\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(bc)} - \boldsymbol{D}_{a(bc)}) = O(N)$ and $\mathrm{tr}\,\{\boldsymbol{C}_{a(bc)}(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\} = O_p(N^{1/2})$. Then it can be seen that

$$
\{\mathbf{mat}_{ab}(\ell_{a(b)})\}^{-1} = (\boldsymbol{A}_1 + \boldsymbol{A}_0)^{-1} = \boldsymbol{A}_1^{-1} - \boldsymbol{A}_1^{-1}\boldsymbol{A}_0\boldsymbol{A}_1^{-1} + O_p(N^{-2}),
$$

so that

$$
\begin{aligned}
\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi} = & - \boldsymbol{A}_1^{-1}\mathbf{col}_a[\mathrm{tr}\,\{\boldsymbol{C}_a(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}] + \boldsymbol{A}_1^{-1}\boldsymbol{A}_0\boldsymbol{A}_1^{-1}\mathbf{col}_a[\mathrm{tr}\,\{\boldsymbol{C}_a(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}] \\
& - \frac{1}{2}\boldsymbol{A}_1^{-1}\mathbf{col}_a \left\{ \sum_{b=1}^k \sum_{c=1}^k \mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(bc)} - \boldsymbol{D}_{a(bc)})(\widehat{\psi}_b - \psi_b)(\widehat{\psi}_c - \psi_c) \right\} + O_p(N^{-3/2}).
\end{aligned}
$$

It is noted that $(\boldsymbol{C}_a)_{ij} = (\boldsymbol{Q}^\top \boldsymbol{W}_a \boldsymbol{Q})_{ij} = (\boldsymbol{W}_a)_{ij} + O(N^{-1})$, $(\boldsymbol{C}_{a(b)})_{ij} = (\boldsymbol{W}_{a(b)})_{ij} + O(N^{-1})$ and $(\boldsymbol{C}_{a(bc)})_{ij} = (\boldsymbol{W}_{a(bc)})_{ij} + O(N^{-1})$. Then, $\mathrm{tr}\,(\boldsymbol{C}_a\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}) + O(1)$,

$\mathrm{tr}\,(\boldsymbol{C}_{a(b)}\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}) + O(1)$ and $\mathrm{tr}\,(\boldsymbol{C}_{a(bc)}\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{W}_{a(bc)}\boldsymbol{\Sigma}) + O(1)$. Since $\boldsymbol{D}_a = \boldsymbol{C}_a\boldsymbol{\Sigma}$, $\boldsymbol{D}_{a(b)} = \boldsymbol{C}_{a(b)}\boldsymbol{\Sigma} + \boldsymbol{C}_a\boldsymbol{\Sigma}_{(b)}$ and $\boldsymbol{D}_{a(bc)} = \boldsymbol{C}_{a(bc)}\boldsymbol{\Sigma} + \boldsymbol{C}_{a(b)}\boldsymbol{\Sigma}_{(c)} + \boldsymbol{C}_{a(c)}\boldsymbol{\Sigma}_{(b)} + \boldsymbol{C}_a\boldsymbol{\Sigma}_{(bc)}$, it is seen that $\mathrm{tr}\,(\boldsymbol{D}_{a(b)}) = \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}) + \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)}) + O(1)$ and $\mathrm{tr}\,(\boldsymbol{D}_{a(bc)}) = \mathrm{tr}\,(\boldsymbol{W}_{a(bc)}\boldsymbol{\Sigma}) + \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) + \mathrm{tr}\,(\boldsymbol{W}_{a(c)}\boldsymbol{\Sigma}_{(b)}) + \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(bc)}) + O(1)$. Thus,

$$
\begin{aligned}
\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(b)} - \boldsymbol{D}_{a(b)}) &= -\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)}) + O(1),\\
\mathrm{tr}\,(\boldsymbol{\Sigma}\boldsymbol{C}_{a(bc)} - \boldsymbol{D}_{a(bc)}) &= -\mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) - \mathrm{tr}\,(\boldsymbol{W}_{a(c)}\boldsymbol{\Sigma}_{(b)}) - \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(bc)}) + O(1).
\end{aligned}
\tag{19}
$$

Letting $\boldsymbol{A} = \mathbf{mat}_{ab}\{\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)})\}$, we have $\boldsymbol{A}_1 = -\boldsymbol{A} + O(1)$. Using Lemma A.1, we can approximate the covariance matrix of $\widehat{\boldsymbol{\psi}}$ as

$$
\begin{aligned}
E[(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})^\top] &= \boldsymbol{A}_1^{-1}\mathbf{mat}_{ab}(E[\mathrm{tr}\,\{\boldsymbol{C}_a(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}\mathrm{tr}\,\{\boldsymbol{C}_b(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}])\boldsymbol{A}_1^{-1} + O(N^{-3/2})\\
&= 2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\widetilde{\boldsymbol{B}}\boldsymbol{A}^{-1} + O(N^{-3/2}),
\end{aligned}
$$

for $\boldsymbol{B} = \mathbf{mat}_{ab}\{\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}\boldsymbol{W}_b\boldsymbol{\Sigma})\}$ and $\widetilde{\boldsymbol{B}} = \mathbf{mat}_{ab}\{K_e h_e(\boldsymbol{W}_a, \boldsymbol{W}_b) + K_v h_v(\boldsymbol{W}_a, \boldsymbol{W}_b)\}$.

The bias of $\widehat{\boldsymbol{\psi}}$ is

$$
\begin{aligned}
E(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = &-\frac{1}{2}\boldsymbol{A}^{-1}\mathbf{col}_a\left[\sum_{b=1}^{k}\sum_{c=1}^{k}\{2\mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) + \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(bc)})\}(\boldsymbol{A}^{-1}(2\boldsymbol{B} + \widetilde{\boldsymbol{B}})\boldsymbol{A}^{-1})_{bc}\right]\\
&+ E(\boldsymbol{A}^{-1}\boldsymbol{A}_0\boldsymbol{A}^{-1}\mathbf{col}_a[\mathrm{tr}\,\{\boldsymbol{C}_a(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}]) + O(N^{-3/2}).
\end{aligned}
$$

Concerning the second term in RHS, the a-th element of $E\{(\boldsymbol{A}_0\boldsymbol{A}^{-1}\mathbf{col}_c[\mathrm{tr}\,\{\boldsymbol{C}_c(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}])\}$ is

$$
\begin{aligned}
&E\{(\boldsymbol{A}_0\boldsymbol{A}^{-1}\mathbf{col}_c[\mathrm{tr}\,\{\boldsymbol{C}_c(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}])_a\}\\
&= \sum_{b=1}^{k}\sum_{c=1}^{k} E[\mathrm{tr}\,\{\boldsymbol{C}_{a(b)}(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}(\boldsymbol{A})^{bc}\mathrm{tr}\,\{\boldsymbol{C}_c(\boldsymbol{u}\boldsymbol{u}^\top - \boldsymbol{\Sigma})\}]\\
&= \sum_{b=1}^{k}\sum_{c=1}^{k}\{2\mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}\boldsymbol{W}_c\boldsymbol{\Sigma}) + K_e h_e(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c) + K_v h_v(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c)\}(\boldsymbol{A})^{bc} + O(N^{-1}).
\end{aligned}
$$

Then,

$$
\begin{aligned}
&E(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})\\
&= \boldsymbol{A}^{-1}\mathbf{col}_a\left(\sum_{b=1}^{k}\sum_{c=1}^{k}\{2\mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}\boldsymbol{W}_c\boldsymbol{\Sigma}) + K_e h_e(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c) + K_v h_v(\boldsymbol{W}_{a(b)}, \boldsymbol{W}_c)\}(\boldsymbol{A})^{bc}\right)\\
&\quad - \frac{1}{2}\boldsymbol{A}^{-1}\mathbf{col}_a\left[\sum_{b=1}^{k}\sum_{c=1}^{k}\{2\mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) + \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(bc)})\}(\boldsymbol{A}^{-1}(2\boldsymbol{B} + \widetilde{\boldsymbol{B}})\boldsymbol{A}^{-1})_{bc}\right] + O(N^{-3/2}),
\end{aligned}
$$

which provides the expression in (3) in Theorem 2.1.

A.3 Proof of Proposition 3.1

Case of $\boldsymbol{W}_a = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}$. We have $\boldsymbol{W}_{a(b)} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(ab)}\boldsymbol{\Sigma}^{-1}$, which yields that $\mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}_{(b)}) = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}) = (\boldsymbol{A})_{ab}$ and $(\boldsymbol{B})_{ab} = \mathrm{tr}\,(\boldsymbol{W}_a\boldsymbol{\Sigma}\boldsymbol{W}_b\boldsymbol{\Sigma}) = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}) = (\boldsymbol{A})_{ab}$. Thus, $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}$ and the covariance matrix of $\widehat{\boldsymbol{\psi}}$ is $2\boldsymbol{A}^{-1} + O(N^{-3/2})$. Moreover, note that

$$
\begin{aligned}
(\boldsymbol{K}_a)_{bc} =& \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}\boldsymbol{W}_c\boldsymbol{\Sigma}) = -2\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(c)}) + \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(ab)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(c)}),\\
(\boldsymbol{H}_a)_{bc} =& \mathrm{tr}\,(\boldsymbol{W}_{a(b)}\boldsymbol{\Sigma}_{(c)}) = -2\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(c)}) + \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(ab)}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(c)}),
\end{aligned}
$$

which shows that $\boldsymbol{W}_a^{\mathrm{REML}}$ satisfies (4).

Case of $\boldsymbol{W}_a = (\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)} + \boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}^{-1})/2$. From (2), it follows that $(\boldsymbol{A})_{ab} = \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}_{(b)})$ and $(\boldsymbol{B})_{ab} = \{\mathrm{tr}\,(\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}_{(b)}) + \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(b)})\}/2$. The asymptotic covariance matrix of $\widehat{\boldsymbol{\psi}}$ is $2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, and the bias is derived from (3).

Case of $\boldsymbol{W}_a = \boldsymbol{\Sigma}_{(a)}$. Straightforward calculation shows that $(\boldsymbol{A})_{ab} = \mathrm{tr}\,(\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}_{(b)})$ and $(\boldsymbol{B})_{ab} = \mathrm{tr}\,(\boldsymbol{\Sigma}_{(a)}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{(b)}\boldsymbol{\Sigma})$. The asymptotic covariance matrix of $\widehat{\boldsymbol{\psi}}$ is $2\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} + O(N^{-3/2})$. Moreover, since $W_{a(b)} = 0$, the condition (4) holds.

Appendix B: Summary of estimation methods in specific models

Here, we provide specific forms of the REML-type, FH-type, and their OLS-based estimators, the PR-type estimator and the Prasad–Rao estimator in the Fay–Herriot model and the nested error regression model.

B.1 Fay–Herriot model

The marginal distribution of $\boldsymbol{y} = (y_1, \ldots, y_m)^\top$ in the Fay–Herriot model has $E[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\beta}$ and $\mathbf{Cov}\,(\boldsymbol{y}) = \boldsymbol{\Sigma} = \psi_1\boldsymbol{I}_m + \boldsymbol{D}$, where p is a dimension of $\boldsymbol{\beta}$ and $\boldsymbol{D} = \mathrm{diag}\,(D_1, \ldots, D_m)$.

REML $\widehat{\psi}_1^{\mathrm{RE}}$ corresponds to $\boldsymbol{W}_1^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-2}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{G}}$ and the estimating equation is $(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top\boldsymbol{\Sigma}^{-2}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) = \mathrm{tr}\,(\boldsymbol{P})$ for $\boldsymbol{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}$.

OLS-based REML $\widehat{\psi}_1^{\mathrm{ORM}}$ corresponds to $\boldsymbol{W}_1^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-2}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{O}}$ and the estimating equation is $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}})^\top\boldsymbol{\Sigma}^{-2}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}}) = \mathrm{tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}^{-2}\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma})$ for $\widetilde{\boldsymbol{P}} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$.

Fay–Herriot estimator $\widehat{\psi}_1^{\mathrm{FH}}$ corresponds to $\boldsymbol{W}_1^{\mathrm{FH}} = \boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{G}}$ and the estimating equation is $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) = m - p$.

OLS-based FH estimator $\widehat{\psi}_1^{\mathrm{OFH}}$ corresponds to $\boldsymbol{W}_1^{\mathrm{FH}} = \boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{O}}$ and the estimating equation is $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}}) = m - 2p + \mathrm{tr}\,\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\}$.

Prasad–Rao estimator $\widehat{\psi}_1^{\mathrm{PR}}$ corresponds to $\boldsymbol{W}_1^{\mathrm{Q}} = \boldsymbol{I}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{O}}$ and it is given by $\widehat{\psi}_1^{\mathrm{PR}} = [\boldsymbol{y}^\top\widetilde{\boldsymbol{P}}\boldsymbol{y} - \mathrm{tr}\,(\boldsymbol{D}) + \mathrm{tr}\,\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}\boldsymbol{X}\}]/(m - p)$.

The asymptotic variances and second-order biases can be provided from Proposition 3.2 as follows: REML $\widehat{\psi}_1^{\mathrm{RE}}$ and OLS-based REML $\widehat{\psi}_1^{\mathrm{ORM}}$ have the same asymptotic variance and the second-order bias

$$\mathrm{Var}(\widehat{\psi}_1^{\mathrm{RE}}) \approx \frac{2}{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})} + \frac{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-4}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-4})}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})\}^2},$$

$$\mathrm{Bias}(\widehat{\psi}_1^{\mathrm{RE}}) \approx -2\frac{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-5}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-5})}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})\}^2} + 2\frac{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-3})\{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-4}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-4})\}}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})\}^3}.$$

Fay–Herriot estimator $\widehat{\psi}_1^{\mathrm{FH}}$ and OLS-based FH estimator $\widehat{\psi}_1^{\mathrm{OFH}}$ have the same asymptotic variance and the second-order bias

$$\mathrm{Var}(\widehat{\psi}_1^{\mathrm{FH}}) \approx \frac{2m}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^2} + \frac{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^2},$$

$$\mathrm{Bias}(\widehat{\psi}_1^{\mathrm{FH}}) \approx 2\frac{m\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}) - \{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^2}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^3} - \frac{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-3}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-3})}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^2} + \frac{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})\{K_e \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{D}^2) + \psi_1^2 K_v \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2})\}}{\{\mathrm{tr}\,(\boldsymbol{\Sigma}^{-1})\}^3},$$

which implies that $\widehat{\psi}_1^{\mathrm{UFH}} = \widehat{\psi}_1^{\mathrm{FH}} - 2[m\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}^{-2}) - \{\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}^{-1})\}^2]/\{\mathrm{tr}\,(\widehat{\boldsymbol{\Sigma}}^{-1})\}^3$ is unbiased up to second order under normality, where $\widehat{\boldsymbol{\Sigma}} = \widehat{\psi}_1^{\mathrm{FH}}\boldsymbol{I}_m + \boldsymbol{D}$.

Prasad–Rao estimator $\widehat{\psi}_1^{\mathrm{PR}}$ is second-order unbiased and has the asymptotic variance $\mathrm{Var}(\widehat{\psi}_1^{\mathrm{PR}}) \approx \{2\mathrm{tr}\,(\boldsymbol{\Sigma}^2) + K_e \mathrm{tr}\,(\boldsymbol{D}^2) + m\psi_1^2 K_v\}/m^2$.

B.2 Nested error regression model

The NER model is written as $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{j}_{n_i} v_i + \boldsymbol{\varepsilon}_i$ for $i = 1, \ldots, m$, where $\boldsymbol{y}_i$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}_i$ are n_i, p and n_i dimensional vectors, $\boldsymbol{X}_i$ is an $n_i \times p$ matrix, v_i is scalar and $\boldsymbol{j}_{n_i} = (1, \ldots, 1)^\top \in \mathbb{R}^{n_i}$. Here, v_i and $\boldsymbol{\varepsilon}_i$ are independent random variables such that $E[v_i] = 0$, $\mathrm{Var}(v_i) = \psi_1$, $E[\boldsymbol{\varepsilon}_i] = \boldsymbol{0}$ and $\mathbf{Cov}\,(\boldsymbol{\varepsilon}_i) = \psi_2 \boldsymbol{I}_{n_i}$. Let $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$, $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_m^\top)^\top$, $N = \sum_{i=1}^m n_i$ and $\boldsymbol{G} = \text{block diag}(\boldsymbol{J}_{n_1}, \ldots, \boldsymbol{J}_{n_m})$ for $\boldsymbol{J}_{n_i} = \boldsymbol{j}_{n_i}\boldsymbol{j}_{n_i}^\top$. In addition, let $\boldsymbol{\Sigma} = \text{block diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_m)$ for $\boldsymbol{\Sigma}_i = \psi_1 \boldsymbol{J}_{n_i} + \psi_2 \boldsymbol{I}_{n_i}$. Then, $\boldsymbol{\Sigma} = \psi_1 \boldsymbol{G} + \psi_2 \boldsymbol{I}_N$, $\boldsymbol{\Sigma}_{(1)} = \boldsymbol{G}$ and $\boldsymbol{\Sigma}_{(2)} = \boldsymbol{I}_N$.

REML $\widehat{\psi}_1^{\mathrm{RE}}$ and $\widehat{\psi}_1^{\mathrm{RE}}$ correspond to $\boldsymbol{W}_1^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{W}_2^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-2}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{G}}$, and the estimating equations are $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) = \mathrm{tr}\,(\boldsymbol{P}\boldsymbol{G})$ and $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}})^\top \boldsymbol{\Sigma}^{-2}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{G}}) = \mathrm{tr}\,(\boldsymbol{P})$.

OLS-based REML $\widehat{\psi}_1^{\mathrm{ORM}}$ and $\widehat{\psi}_2^{\mathrm{ORM}}$ correspond to $\boldsymbol{W}_1^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}$, $\boldsymbol{W}_2^{\mathrm{RE}} = \boldsymbol{\Sigma}^{-2}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{O}}$, and the estimating equations are $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}}) = \mathrm{tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1})$ and $(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}})^\top \boldsymbol{\Sigma}^{-2}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^{\mathrm{O}}) = \mathrm{tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}^{-2})$.

FH-type estimators $\widehat{\psi}_1^{\rm FH}$ and $\widehat{\psi}_2^{\rm FH}$ correspond to $\boldsymbol{W}_1^{\rm FH} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{G} + \boldsymbol{G}\boldsymbol{\Sigma}^{-1})/2$, $\boldsymbol{W}_2^{\rm FH} = \boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\rm G}$, and the estimating equations are

$$\sum_{i=1}^m \frac{n_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top \widehat{\boldsymbol{\beta}}^{\rm G})^2}{n_i\psi_1 + \psi_2} = N - \sum_{i=1}^m \frac{n_i^2\overline{\boldsymbol{x}}_i^\top(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_i}{n_i\psi_1 + \psi_2},$$

$$\psi_2 = \frac{1}{N-p}\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^\top\widehat{\boldsymbol{\beta}}^{\rm G})^2 - \frac{1}{N-p}\sum_{i=1}^m \frac{n_i^2\psi_1}{n_i\psi_1+\psi_2}(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top\widehat{\boldsymbol{\beta}}^{\rm G})^2.$$

OLS-based FH estimators $\widehat{\psi}_1^{\rm OFH}$ and $\widehat{\psi}_2^{\rm OFH}$ correspond to $\boldsymbol{W}_1^{\rm FH} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{G} + \boldsymbol{G}\boldsymbol{\Sigma}^{-1})/2$, $\boldsymbol{W}_2^{\rm FH} = \boldsymbol{\Sigma}^{-1}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\rm O}$, and the estimating equations are

$$\sum_{i=1}^m \frac{n_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2}{n_i\psi_1+\psi_2} = N - 2\sum_{i=1}^m n_i^2\overline{\boldsymbol{x}}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_i$$
$$+ \sum_{i=1}^m \frac{n_i^2\overline{\boldsymbol{x}}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_i}{n_i\psi_1+\psi_2},$$
$$\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2 - \sum_{i=1}^m\frac{n_i^2\psi_1}{n_i\psi_1+\psi_2}(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2 = {\rm tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{P}}\boldsymbol{\Sigma}^{-1}).$$

PR-type estimators $\widehat{\psi}_1^{\rm Q}$ and $\widehat{\psi}_2^{\rm Q}$ correspond to $\boldsymbol{W}_1^{\rm Q} = \boldsymbol{G}$, $\boldsymbol{W}_2^{\rm Q} = \boldsymbol{I}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\rm O}$, and the estimators are $\widehat{\psi}_1^{\rm Q} = \{\sum_{i=1}^m n_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2 - \widehat{\psi}_2{\rm tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{G})\}/{\rm tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{G})^2$ and

$$\widehat{\psi}_2^{\rm Q} = \frac{\sum_{i=1}^m\sum_{j=1}^{n_i}(y_{ij} - \boldsymbol{x}_{ij}^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2 - [{\rm tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{G})/{\rm tr}\,\{(\widetilde{\boldsymbol{P}}\boldsymbol{G})^2\}]\sum_{i=1}^m n_i^2(\overline{y}_i - \overline{\boldsymbol{x}}_i^\top\widehat{\boldsymbol{\beta}}^{\rm O})^2}{N - p - \{{\rm tr}\,(\widetilde{\boldsymbol{P}}\boldsymbol{G})\}^2{\rm tr}\,\{(\widetilde{\boldsymbol{P}}\boldsymbol{G})^2\}}.$$

Prasad–Rao estimators are $\widehat{\psi}_1^{\rm PR} = \{\boldsymbol{y}^\top\widetilde{\boldsymbol{P}}\boldsymbol{y} - (N-p)\widehat{\psi}_2\}/\{N - \sum_{i=1}^m n_i^2\overline{\boldsymbol{x}}_i^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\overline{\boldsymbol{x}}_i\}$ and $\widehat{\psi}_2^{\rm PR} = \{\boldsymbol{y}^\top\{\boldsymbol{E} - \boldsymbol{E}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{E}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{E}\}\boldsymbol{y}\}/(N-k-p)$, where $\boldsymbol{E} =$ block diag$(\boldsymbol{I}_{n_1} - n_1^{-1}\boldsymbol{J}_{n_1}, \ldots, \boldsymbol{I}_{n_m} - n_m^{-1}\boldsymbol{J}_{n_m})$.

Hereafter we assume that $K_e = K_v = 0$ for simplicity. Note that $\boldsymbol{\Sigma}\boldsymbol{G} = \boldsymbol{G}\boldsymbol{\Sigma}$, $\psi_1\boldsymbol{G} = \boldsymbol{\Sigma} - \psi_2\boldsymbol{I}_N$, $\psi_2\boldsymbol{\Sigma}^{-1} = \boldsymbol{I}_N - \psi_1$block diag$(\gamma_1\boldsymbol{J}_{n_1}, \ldots, \gamma_m\boldsymbol{J}_{n_m})$, $\psi_2^2\boldsymbol{\Sigma}^{-2} = \boldsymbol{I}_N - \psi_1$block diag$((1+\psi_2\gamma_1)\gamma_1\boldsymbol{J}_{n_1}, \ldots, (1+\psi_2\gamma_m)\gamma_m\boldsymbol{J}_{n_m})$ for $\gamma_i = 1/(\psi_2 + n_i\psi_1)$. Then the asymptotic variances and second-order biases can be provided from Proposition 3.3 as follows: REML $\widehat{\boldsymbol{\psi}}^{\rm RE}$ and OLS-based REML $\widehat{\boldsymbol{\psi}}^{\rm ORM}$ are second-order unbiased and have the same asymptotic variance

$$\textbf{Cov}\,(\widehat{\boldsymbol{\psi}}^{\rm RE}) \approx 2\begin{pmatrix} {\rm tr}\,\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{G})^2\} & {\rm tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}) \\ {\rm tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}) & {\rm tr}\,(\boldsymbol{\Sigma}^{-2}) \end{pmatrix}^{-1} = 2\begin{pmatrix} \sum_{i=1}^m n_i^2\gamma_i^2 & \sum_{i=1}^m n_i\gamma_i^2 \\ \sum_{i=1}^m n_i\gamma_i^2 & (N-m)/\psi_2^2 + \sum_{i=1}^m \gamma_i^2 \end{pmatrix}^{-1},$$

which was given in Datta and Lahiri (2000).

Fay–Herriot estimator $\widehat{\boldsymbol{\psi}}^{\mathrm{FH}}$ and OLS-based FH estimator $\widehat{\boldsymbol{\psi}}^{\mathrm{OFH}}$ have the same asymptotic covariance matrix $\mathbf{Cov}\,(\widehat{\boldsymbol{\psi}}^{\mathrm{FH}}) \approx 2\boldsymbol{A}_{\mathrm{FH}}^{-1}\boldsymbol{B}_{\mathrm{FH}}\boldsymbol{A}_{\mathrm{FH}}^{-1}$, where

$$\boldsymbol{A}_{\mathrm{FH}} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m n_i^2\gamma_i & \sum_{i=1}^m n_i\gamma_i \\ \sum_{i=1}^m n_i\gamma_i & (N-m)/\psi_2 + \sum_{i=1}^m \gamma_i \end{pmatrix},$$
$$\boldsymbol{B}_{\mathrm{FH}} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G} + \boldsymbol{G}^2) & N \\ N & N \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m n_i^2 & N \\ N & N \end{pmatrix},$$

and the same second-order bias

$$\mathbf{Bias}(\widehat{\boldsymbol{\psi}}^{\mathrm{FH}}) \approx 2\boldsymbol{A}_{\mathrm{FH}}^{-1} \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{K}_1\boldsymbol{A}_{\mathrm{FH}}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_1\boldsymbol{A}_{\mathrm{FH}}^{-1}\boldsymbol{B}_{\mathrm{FH}}\boldsymbol{A}_{\mathrm{FH}}^{-1}) \\ \mathrm{tr}\,(\boldsymbol{K}_2\boldsymbol{A}_{\mathrm{FH}}^{-1}) - \mathrm{tr}\,(\boldsymbol{H}_2\boldsymbol{A}_{\mathrm{FH}}^{-1}\boldsymbol{B}_{\mathrm{FH}}\boldsymbol{A}_{\mathrm{FH}}^{-1}) \end{pmatrix},$$

where

$$\boldsymbol{K}_1 = -\begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^3) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^2) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) \end{pmatrix} = -\begin{pmatrix} \sum_i n_i^3\gamma_i & \sum_i n_i^2\gamma_i \\ \sum_i n_i^2\gamma_i & \sum_i n_i\gamma_i \end{pmatrix},$$
$$\boldsymbol{K}_2 = -\begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}) \end{pmatrix} = -\begin{pmatrix} \sum_i n_i^2\gamma_i & \sum_i n_i\gamma_i \\ \sum_i n_i\gamma_i & (N-m)/\psi_2 + \sum_i \gamma_i \end{pmatrix},$$
$$\boldsymbol{H}_1 = -\begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}) \end{pmatrix} = -\begin{pmatrix} \sum_i n_i^3\gamma_i^2 & \sum_i n_i^2\gamma_i^2 \\ \sum_i n_i^2\gamma_i^2 & \sum_i n_i\gamma_i^2 \end{pmatrix},$$
$$\boldsymbol{H}_2 = -\begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{G}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{\Sigma}^{-2}) \end{pmatrix} = -\begin{pmatrix} \sum_i n_i^2\gamma_i^2 & \sum_i n_i\gamma_i^2 \\ \sum_i n_i\gamma_i^2 & (N-m)/\psi_2 + \sum_i \gamma_i^2 \end{pmatrix}.$$

PR-type estimator $\widehat{\boldsymbol{\psi}}^{\mathrm{Q}}$ is second-order unbiased and has the same asymptotic covariance matrix $\mathbf{Cov}\,(\widehat{\boldsymbol{\psi}}^{\mathrm{Q}}) \approx 2\boldsymbol{A}_{\mathrm{Q}}^{-1}\boldsymbol{B}_{\mathrm{Q}}\boldsymbol{A}_{\mathrm{Q}}^{-1}$, where

$$\boldsymbol{A}_{\mathrm{Q}} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{I}_N) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m n_i^2 & N \\ N & N \end{pmatrix},$$
$$\boldsymbol{B}_{\mathrm{Q}} = \begin{pmatrix} \mathrm{tr}\,(\boldsymbol{\Sigma}^2\boldsymbol{G}^2) & \mathrm{tr}\,(\boldsymbol{\Sigma}^2\boldsymbol{G}) \\ \mathrm{tr}\,(\boldsymbol{\Sigma}^2\boldsymbol{G}) & \mathrm{tr}\,(\boldsymbol{\Sigma}^2) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m n_i^2/\gamma_i^2 & \sum_{i=1}^m n_i/\gamma_i^2 \\ \sum_{i=1}^m n_i/\gamma_i^2 & (N-m)\psi_2^2 + \sum_{i=1}^m 1/\gamma_i^2 \end{pmatrix}.$$

References

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

Boreinstein, M., Hedges, L. V., & Higgins, J. P. T. (2009). *Introduction to meta-analysis*. Wiley.

Datta, G. S., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica, 10*, 613–627.

Doornik, J. A. (2007). *Object-oriented matrix programming using Ox* (3rd ed.). Timberlake Consultants Press and Oxford.

Fay, R. E., & Herriot, R. (1979). Estimates of income for small places: An application of James–Stein procedures to census data. *Journal of the American Statistical Association, 74*, 269–277.
Prasad, N. G. N., & Rao, J. N. K. (1990). The estimation of the mean squared error of small area estimators. *Journal of the American Statistical Association, 85*, 163–171.
Rao, C. R., & Kleffe, J. (1988). *Estimation of variance components and applications*. North-Holland.
Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). Wiley.
Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. Wiley.
Verbeke, G., & Molenberghs, G. (2006). *Linear mixed models for longitudinal data*. Springer.