Annals of the Institute of Statistical Mathematics (2022) 74:1143–1161
https://doi.org/10.1007/s10463-022-00825-7

Inference of random effects for linear mixed-effects models with a fixed number of clusters

Chih-Hao Chang[1] · **Hsin-Cheng Huang**[2] · **Ching-Kang Ing**[3]

Received: 14 June 2021 / Revised: 9 March 2022 / Accepted: 22 March 2022 /
Published online: 14 May 2022

Abstract
We consider a linear mixed-effects model with a clustered structure, where the parameters are estimated using maximum likelihood (ML) based on possibly unbalanced data. Inference with this model is typically done based on asymptotic theory, assuming that the number of clusters tends to infinity with the sample size. However, when the number of clusters is fixed, classical asymptotic theory developed under a divergent number of clusters is no longer valid and can lead to erroneous conclusions. In this paper, we establish the asymptotic properties of the ML estimators of random-effects parameters under a general setting, which can be applied to conduct valid statistical inference with fixed numbers of clusters. Our asymptotic theorems allow both fixed effects and random effects to be misspecified, and the dimensions of both effects to go to infinity with the sample size.

Keywords Confidence interval · Consistency · Maximum likelihood

✉ Chih-Hao Chang
jhow@nuk.edu.tw

Hsin-Cheng Huang
hchuang@stat.sinica.edu.tw

Ching-Kang Ing
cking@stat.nthu.edu.tw

1 Institute of Statistics, National University of Kaohsiung, No.700, Kaohsiung University Rd., Kaohsiung, Taiwan

2 Institute of Statistical Science, Academia Sinica, No.128, Academia Road, Section 2, Taipei, Taiwan

3 Institute of Statistics, National Tsing Hua University, No.101, Kuang-Fu Road, Section 2, HsinChu, Taiwan

1 Introduction

Over the past several decades, linear mixed-effects models have been broadly applied to clustered data (Longford, 1993), longitudinal data (Laird and Ware, 1982; Verbeke and Molenberghs, 2000), spatial data (Mardia and Marshall, 1984), and data in scientific fields (Jiang, 2007, 2017), particularly due to their usefulness in modeling data with clustered structures. Model parameters are traditionally estimated, for example, via minimum norm quadratic, maximum likelihood (ML), and restricted ML (REML) methods. ML and REML estimators are compared in Gumedze and Dunne (2011).

Estimating random-effects variances in mixed-effects models is usually more challenging than estimating fixed-effects parameters. Although desired asymptotic properties have been developed for ML and REML estimators of random-effects variances (Hartley and Rao, 1967; Harville, 1977; Miller, 1977), these are mainly obtained under the mathematical device of requiring the number of clusters (denoted as m) to grow to infinity with the sample size (denoted as N) and the numbers of fixed effects and random effects (denoted as p and q) to be fixed. In fact, most asymptotic results for likelihood ratio tests and model selection in linear mixed-effects models are established under a similar mathematical device; see Self and Liang (1987), Stram and Lee (1994), Crainiceanu and Ruppert (2004), Pu and Niu (2006), Fan and Li (2012), and Peng and Lu (2012). However, in many practical situations, we are faced with a small m, which does not grow to infinity with N. As pointed out by McNeish and Stapleton (2016a) and Huang (2018), data collected in the fields of education or developmental psychology typically have a small number of clusters, corresponding, for example, to classrooms or schools. Unfortunately, to the best of our knowledge, no theoretical justification has been provided for random-effects estimators when m is fixed.

As shown by Maas and Hox (2004), Bell et al. (2014), and McNeish and Stapleton (2016b), for a linear mixed-effects model with few clusters, random-effects variances are not well estimated by either ML or REML. This is because when m is fixed, the Fisher information for random-effects variances fails to grow with N, and hence, the corresponding ML estimators do not achieve consistency. A similar difficulty arises in a spatial-regression model of Chang et al. (2017) under the fixed domain asymptotics, in which the spatial covariance parameters cannot be consistently estimated. A direct impact of this difficulty is that the classical central limit theorem established under $m \to \infty$ for the ML (or REML) estimators (Hartley and Rao, 1967; Harville, 1977; Miller, 1977) is no longer valid. Consequently, statistical inference based on the asymptotic results for $m \to \infty$ can be misleading.

In this article, we focus on the ML estimators in linear mixed-effects models with possibly unbalanced data. We first develop the asymptotic properties of the ML estimators, without assuming that fixed- and random-effects models are correctly specified, p and q are fixed, or $m \to \infty$. Based on the asymptotic properties of the ML estimators, we provide, for the first time in the mixed-effects models literature, the asymptotic valid confidence intervals for random-effects variances when m is fixed. In addition, we present an example illustrating that empirical best linear unbiased predictors (BLUPs) of random effects (which are the BLUPs with the unknown parameters replaced by their ML estimators) compare favorably to least squares (LS) predictors even when the ML estimators are not consistent; see Sect. 3.1 for details. Also note that our asymptotic theorems allow both fixed- and random-effects models to be misspecified. Consequently, our results are crucial to facilitate further studies on model selection for linear mixed-effects models with fixed m, in which investigating the impact of model misspecification is indispensable.

This article is organized as follows. Section 2 introduces the linear mixed-effects model and the regularity conditions. The asymptotic results for the ML estimators are given in Sect. 3. Section 4 describes simulation studies that confirm our asymptotic theory, and a real-data example, including a comparison between the conventional confidence intervals and the proposed ones for random-effects variances. A brief discussion is given in Sect. 5. The proofs of all the theoretical results are deferred to the online supplementary material.

2 Linear mixed-effects models

Consider a set of observations with m clusters, $\{(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)\}_{i=1}^m$, where $\boldsymbol{y}_i = (y_{i,1}, \dots, y_{i,n_i})'$ is the response vector, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are $n_i \times p$ and $n_i \times q$ design matrices of p and q covariates with the (j, k)-th entries $x_{i,j,k}$ and $z_{i,j,k}$, respectively, and n_i is the number of observations in cluster i. A general linear mixed-effects model can be written as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i; \quad i = 1, \dots, m, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is the p-vector of fixed effects, $\boldsymbol{b}_i = (b_{i,1}, \dots, b_{i,q})' \sim N(\boldsymbol{0}, \operatorname{diag}(\sigma_1^2, \dots, \sigma_q^2))$ is the q-vector of random effects, $\boldsymbol{\epsilon}_i \sim N(\boldsymbol{0}, v^2\boldsymbol{I}_{n_i})$, and $\boldsymbol{I}_{n_i}$ is the n_i-dimensional identity matrix. Here, $\{\boldsymbol{b}_i\}$ and $\{\boldsymbol{\epsilon}_i\}$ are mutually independent. Let $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{b}$, and $\boldsymbol{\epsilon}$ be obtained by stacking $\{\boldsymbol{y}_i\}$, $\{\boldsymbol{X}_i\}$, $\{\boldsymbol{b}_i\}$, and $\{\boldsymbol{\epsilon}_i\}$. Also let $\boldsymbol{Z} = \operatorname{diag}(\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_m)$ be the block diagonal matrix with diagonal blocks $\{\boldsymbol{Z}_i\}$ and dimension $N \times (mq)$, where $N = n_1 + \cdots + n_m$ is the total sample size. Let $\theta_k = \sigma_k^2/v^2$; $k = 1, \dots, q$ and $\boldsymbol{D} = \operatorname{diag}(\theta_1, \dots, \theta_q)$. Then, we can rewrite (1) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon} \sim N(\boldsymbol{X}\boldsymbol{\beta}, v^2\boldsymbol{H}), \tag{2}$$

where $\boldsymbol{H} = \boldsymbol{R} + \boldsymbol{I}_N$, $\boldsymbol{R} = \operatorname{diag}(\boldsymbol{R}_1, \dots, \boldsymbol{R}_m)$, and $\boldsymbol{R}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i'$; $i = 1, \dots, m$.

Let $\mathcal{A} \times \mathcal{G} \subset 2^{\{1,\dots,p\}} \times 2^{\{1,\dots,q\}}$ be the set of candidate models with $\alpha \in \mathcal{A}$ and $\gamma \in \mathcal{G}$ corresponding to the fixed-effects and random-effects covariates indexed by α and γ, respectively. Then, a linear mixed-effects model corresponding to $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$ can be written as

$$\boldsymbol{y} = \boldsymbol{X}(\alpha)\boldsymbol{\beta}(\alpha) + \boldsymbol{Z}(\gamma)\boldsymbol{b}(\gamma) + \boldsymbol{\epsilon}. \tag{3}$$

For $i = 1, \dots, m$, let $\boldsymbol{Z}_i(\gamma)$ be the sub-matrix of $\boldsymbol{Z}_i$ and $\boldsymbol{b}_i(\gamma)$ be the sub-vector of $\boldsymbol{b}_i$ corresponding to γ. Then for $\gamma \in \mathcal{G}$,

$$\boldsymbol{R}_i(\gamma, \boldsymbol{\theta}(\gamma)) \equiv \frac{1}{v^2}\mathrm{var}(\boldsymbol{Z}_i(\gamma)\boldsymbol{b}_i(\gamma)) = \sum_{k\in\gamma} \theta_k \boldsymbol{z}_{i,k}\boldsymbol{z}'_{i,k},$$

where $\boldsymbol{z}_{i,k}$ is the k-th column of $\boldsymbol{Z}_i$ and $\boldsymbol{\theta}(\gamma)$ is the parameter vector of θ_k; $k \in \gamma$. In other words, under $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$,

$$\boldsymbol{y} \sim N(\boldsymbol{X}(\alpha)\boldsymbol{\beta}(\alpha), v^2\boldsymbol{H}(\gamma, \boldsymbol{\theta})), \tag{4}$$

where

$$\begin{aligned} \boldsymbol{H}(\gamma, \boldsymbol{\theta}) &= \boldsymbol{R}(\gamma, \boldsymbol{\theta}) + \boldsymbol{I}_N, \\ \boldsymbol{R}(\gamma, \boldsymbol{\theta}) &= \mathrm{diag}(\boldsymbol{R}_1(\gamma, \boldsymbol{\theta}), \dots, \boldsymbol{R}_m(\gamma, \boldsymbol{\theta})) = \sum_{i=1}^{m}\sum_{k\in\gamma} \theta_k \boldsymbol{h}_{i,k}\boldsymbol{h}'_{i,k}, \end{aligned} \tag{5}$$

$\boldsymbol{h}_{i,k} = (\boldsymbol{0}'_{n_1}, \dots, \boldsymbol{0}'_{n_{k-1}}, \boldsymbol{z}'_{i,k}, \boldsymbol{0}'_{n_{k+1}}, \dots, \boldsymbol{0}'_{n_m})'$, and $\boldsymbol{0}_{n_i}$ is the n_i-vector of zeros. Here, for notational simplicity, we suppress the dependence of $\boldsymbol{\theta}$ on γ.

For $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$, let $p(\alpha)$ be the dimension of α and let $q(\gamma)$ be the dimension of γ. Assume that the true model of $\boldsymbol{y}$ is

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}_0, v_0^2\boldsymbol{H}_0), \tag{6}$$

where $\boldsymbol{\mu}_0$ is the underlying mean trend, $v_0^2 > 0$ is the true value of v^2, $\boldsymbol{H}_0 = \boldsymbol{R}_0 + \boldsymbol{I}_N$, $\boldsymbol{R}_0 = \mathrm{diag}(\boldsymbol{Z}_1\boldsymbol{D}_0\boldsymbol{Z}'_1, \dots, \boldsymbol{Z}_m\boldsymbol{D}_0\boldsymbol{Z}'_m)$, and $\boldsymbol{D}_0 = \mathrm{diag}(\theta_{1,0}, \dots, \theta_{q,0})$ for some $\theta_{k,0} \geq 0$; $k = 1, \dots, q$. Similarly, let $v_0^2\boldsymbol{D}_0 = \mathrm{diag}(\sigma^2_{1,0}, \dots, \sigma^2_{q,0})$ with $\sigma^2_{k,0} \geq 0$ being the true values of σ_k^2, for $k = 1, \dots, q$. We say that a fixed-effects model α is correct if there exists $\boldsymbol{\beta}(\alpha) \in \mathbb{R}^{p(\alpha)}$ such that $\boldsymbol{\mu}_0 = \boldsymbol{X}(\alpha)\boldsymbol{\beta}(\alpha)$. Similarly, a random-effects model γ is correct if $\{k : \theta_{k,0} > 0, k = 1, \dots, q\} \subset \gamma$. Let $\mathcal{A}_0$ and $\mathcal{G}_0$ denote the sets of all correct fixed-effects and random-effects models, respectively. A linear mixed-effects model (α, γ) is said to be correct if $(\alpha, \gamma) \in \mathcal{A}_0 \times \mathcal{G}_0$. We denote the smallest correct model by (α_0, γ_0), which satisfies

$$\begin{aligned} p_0 &\equiv p(\alpha_0) = \inf_{\alpha\in\mathcal{A}_0} p(\alpha), \\ q_0 &\equiv q(\gamma_0) = \inf_{\gamma\in\mathcal{G}_0} q(\gamma), \end{aligned}$$

where $p_0 > 0$ and $q_0 > 0$ are assumed fixed.

Given a model $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$, the covariance parameters consist of $\boldsymbol{\theta}$ and v^2. We estimate these by ML. We assume that $\boldsymbol{X}$ and $\boldsymbol{Z}$ are of full column rank. The ML estimators $\hat{\boldsymbol{\theta}}(\alpha, \gamma)$ and $\hat{v}^2(\alpha, \gamma)$ of $\boldsymbol{\theta}$ and v^2 based on model $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$ can be obtained by minimizing the negative twice profile log-likelihood function:

$$-2\log L(\boldsymbol{\theta}, v^2;\alpha,\gamma) = N\log(2\pi) + N\log(v^2) + \log\det(\boldsymbol{H}(\gamma,\boldsymbol{\theta})) + \frac{\boldsymbol{y}'\boldsymbol{H}^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{A}(\alpha,\gamma;\boldsymbol{\theta})\boldsymbol{y}}{v^2}, \tag{7}$$

where

$$\boldsymbol{A}(\alpha,\gamma;\boldsymbol{\theta}) \equiv \boldsymbol{I}_N - \boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta}), \tag{8}$$

$$\boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta}) \equiv \boldsymbol{X}(\alpha)(\boldsymbol{X}(\alpha)'\boldsymbol{H}^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{H}^{-1}(\gamma,\boldsymbol{\theta}). \tag{9}$$

Note that $\boldsymbol{M}^2(\alpha,\gamma;\boldsymbol{\theta}) = \boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta})$, $\boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta})\boldsymbol{X}(\alpha) = \boldsymbol{X}(\alpha)$ and

$$\boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta})'\boldsymbol{H}^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta}) = \boldsymbol{H}^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{M}(\alpha,\gamma;\boldsymbol{\theta}).$$

For model $(\alpha,\gamma) \in \mathcal{A}\times\mathcal{G}$, the ML estimator of $\boldsymbol{\beta}(\alpha)$ is given by

$$\hat{\boldsymbol{\beta}}(\alpha,\gamma;\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}(\alpha)'\boldsymbol{H}^{-1}(\gamma,\hat{\boldsymbol{\theta}})\boldsymbol{X}(\alpha))^{-1}\boldsymbol{X}(\alpha)'\boldsymbol{H}^{-1}(\gamma,\hat{\boldsymbol{\theta}})\boldsymbol{y}, \tag{10}$$

where $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\alpha,\gamma)$ satisfies

$$(\hat{\boldsymbol{\theta}}(\alpha,\gamma), \hat{v}^2(\alpha,\gamma)) = * \underset{\boldsymbol{\theta}\in[0,\infty)^{q(\gamma)}, v^2\in(0,\infty)}{arg\,min} \{-2\log L(\boldsymbol{\theta}, v^2;\alpha,\gamma)\}.$$

Then, the ML estimator of σ_k^2 is

$$\hat{\sigma}_k^2(\alpha,\gamma) = \hat{\theta}_k(\alpha,\gamma)\hat{v}^2(\alpha,\gamma); \quad k\in\gamma,$$

where $\hat{\theta}_k(\alpha,\gamma)$ is the ML estimator of θ_k based on model (α,γ).

To study the asymptotic properties for the ML estimators of σ_k^2 and v^2, we provide a novel decomposition of the likelihood equations based on the following quadratic forms: $\boldsymbol{x}_{i,j}'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{x}_{i,j^*}$, $\boldsymbol{z}_{i,k}'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{z}_{i,k^*}$, $\boldsymbol{\epsilon}_i'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{\epsilon}_i$, $\boldsymbol{x}_{i,j}'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{z}_{i,k}$, $\boldsymbol{x}_{i,j}'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{\epsilon}_i$, and $\boldsymbol{z}_{i,k}'\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})\boldsymbol{\epsilon}_i$, where $\boldsymbol{H}_i(\gamma,\boldsymbol{\theta}) = \boldsymbol{I}_{n_i} + \boldsymbol{R}_i(\gamma,\boldsymbol{\theta})$ and $\boldsymbol{x}_{i,j}$ is the j-th column of $\boldsymbol{X}_i$, for $i = 1,\dots,m$ and $j = 1,\dots,p$. The main difficulty lies in how to build a suitable representation of the precision matrix $\boldsymbol{H}_i^{-1}(\gamma,\boldsymbol{\theta})$ so that the desired convergent rates of these quadratic forms are derived. We shall also require the following regularity conditions, which play the key role in establishing the asymptotic theory for the ML estimators of the parameters in linear mixed-effects models.

(A0) Let $n_{\min} = \min_{i=1,\dots,m} n_i$. Assume that $p = c_p + o(n_{\min}^{\tau})$ and $q = c_q + o(n_{\min}^{\tau})$, for some constant $\tau\in[0,1/2)$, where $c_p > 0$ and $c_q > 0$.

(A1) With τ given in (A0), there exist constants $\xi\in(2\tau,1]$ and $d_{i,j} > 0$; $i = 1,\dots,m$, $j = 1,\dots,p$, with $0 < \inf\{d_{i,j}\} \le \sup\{d_{i,j}\} < \infty$ such that for $i = 1,\dots,m$ and $1 \le j, j^* \le p$,

$$\boldsymbol{x}_{i,j}'\boldsymbol{x}_{i,j^*} = \begin{cases} d_{i,j}n_i^{\xi} + o(n_i^{\xi}); & \text{if } j = j^*, \\ o(n_i^{\xi-\tau}); & \text{if } j \ne j^*. \end{cases}$$

(A2) With τ given in (A0), there exist constants $\ell \in (2\tau, 1]$ and $c_{i,k} > 0; i = 1, \dots, m$, $k = 1, \dots, q$, with $0 < \inf\{c_{i,k}\} \le \sup\{c_{i,k}\} < \infty$ such that for $i = 1, \dots, m$ and $1 \le k, k^* \le q$,

$$z'_{i,k} z_{i,k^*} = \begin{cases} c_{i,k} n_i^{\ell} + o(n_i^{\ell}); & \text{if } k = k^*, \\ o(n_i^{\ell-\tau}); & \text{if } k \ne k^* . \end{cases}$$

(A3) For $i = 1, \dots, m$, $j = 1, \dots, p$, and $k = 1, \dots, q$,

$$\boldsymbol{x}'_{i,j} z_{i,k} = o(n_i^{(\xi+\ell)/2-\tau}),$$

where τ, ξ, and ℓ are given in (A0), (A1), and (A2), respectively.

Condition (A0) allows the numbers of fixed effects and random effects (i.e., p and q) to go to infinity with $n_{\min}$ at a certain rate. Conditions (A1)–(A3) impose correlation constraints on $\{\boldsymbol{x}_{i,j}\}$ and $\{z_{i,k}\}$. For example, Condition (A2) implies that the maximum eigenvalue satisfies $\lambda_{\max}(\boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}'_i) = O(n_i^{\ell})$, which is similar to an assumption given in Condition 3 of Fan and Li (2012).

3 Asymptotic properties

In this section, we investigate the asymptotic properties of the ML estimators of v^2 and $\{\sigma_k^2 : k \in \gamma\}$ for any $(\alpha, \gamma) \in \mathcal{A} \times \mathcal{G}$. We allow p and q to go to infinity with the sample size N. In addition, as we allow m to be fixed, we must account for the fact that $\{\sigma_k^2 : k \in \gamma\}$ may not be estimated consistently.

3.1 Asymptotics under correct specification

In this subsection, we consider a correct (but possibly overfitted) model $(\alpha, \gamma) \in \mathcal{A}_0 \times \mathcal{G}_0$. We derive not only the convergence rates for the ML estimators of v^2 and $\{\sigma_k^2 : k \in \gamma\}$, but also their asymptotic distributions.

Theorem 1 *Consider the data generated from* (2) *with the true parameters given by* (6). *Let* $(\alpha, \gamma) \in \mathcal{A}_0 \times \mathcal{G}_0$ *be a correct model defined in* (4). *Denote* $\hat{\sigma}_k^2(\alpha, \gamma)$ *and* $\hat{v}^2(\alpha, \gamma)$ *to be the ML estimators of* σ_k^2 *and* v^2, *respectively. Assume that* (A0)–(A3) *hold. Then,*

$$\hat{v}^2(\alpha, \gamma) = v_0^2 + O_p\left(\frac{p + mq}{N}\right) + O_p(N^{-1/2}), \tag{11}$$

$$\hat{\sigma}_k^2(\alpha, \gamma) = \begin{cases} \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} b_{i,k}^2 + O_p\left(\dfrac{1}{m} \displaystyle\sum_{i=1}^{m} n_i^{-\ell/2}\right); & \text{if } k \in \gamma \cap \gamma_0, \\ O_p\left(n_{\max}^{-\ell}\right); & \text{if } k \in \gamma \setminus \gamma_0, \end{cases} \tag{12}$$

where $n_{\max} = \max\limits_{i=1,\dots,m} n_i$. *In addition, if* $p + mq = o\left(N^{1/2}\right)$, *then*

$$N^{1/2}\big(\hat{v}^2(\alpha,\gamma)-v_0^2\big)\xrightarrow{d} N\big(0,2v_0^4\big),\qquad \text{as } N\to\infty. \tag{13}$$

Remark 1 When m is fixed, for $(\alpha,\gamma)\in\mathcal{A}_0\times\mathcal{G}_0$ and $k\in\gamma\cap\gamma_0$, (12) implies that

$$m\hat{\sigma}_k^2(\alpha,\gamma)\xrightarrow{d}\sigma_{k,0}^2\chi_m^2,$$

which further shows that $\hat{\sigma}_k^2(\alpha,\gamma)$ does not converge to $\sigma_{k,0}^2$. This is because the data do not contain enough information about $\{\sigma_k^2 : k\in\gamma\cap\gamma_0\}$. On the other hand, for $k\in\gamma\setminus\gamma_0$, $\hat{\sigma}_k^2(\alpha,\gamma)$ converges to $\sigma_{k,0}^2=0$ at a rate $n_{\max}^{-\ell}$, which can be faster than $N^{-1/2}$.

When $m\to\infty$, by applying the law of large numbers and the central limit theorem to $b_{i,k}^2$; $i=1,\ldots,m$, $k\in\gamma_0$, we immediately have the following corollary.

Corollary 1 *Under the assumptions of Theorem 1, $\hat{\sigma}_k^2(\alpha,\gamma)\xrightarrow{p}\sigma_{k,0}^2$ as $m\to\infty$, for $k\in\gamma$. If, in addition, $m=o(n_{\min}^{\ell})$, then*

$$m^{1/2}(\hat{\sigma}_k^2(\alpha,\gamma)-\sigma_{k,0}^2)\xrightarrow{d} N(0,2\sigma_{k,0}^4);\quad k\in\gamma\cap\gamma_0,\qquad \text{as } N\to\infty.$$

From Corollary 1, for $k\in\gamma_0$, we obtain a $100(1-\alpha)\%$ confidence interval of $\sigma_{k,0}^2$:

$$\left(\hat{\sigma}_k^2(\alpha,\gamma)-\left(\frac{2\hat{\sigma}_k^4(\alpha,\gamma)}{m}\right)^{1/2}\zeta_{1-\alpha/2},\ \hat{\sigma}_k^2(\alpha,\gamma)-\left(\frac{2\hat{\sigma}_k^4(\alpha,\gamma)}{m}\right)^{1/2}\zeta_{\alpha/2}\right), \tag{14}$$

where ζ_a is the $(100a)$-th percentile of the standard normal distribution. Although this confidence interval is commonly applied in practice (e.g., Maas and Hox, 2004; McNeish and Stapleton, 2016b), it is only valid when m is large, as detailed in a simulation experiment of Sect. 4.2. Thanks to Theorem 1, we can derive a $100(1-\alpha)\%$ confidence interval of $\sigma_{k,0}^2$ valid for a fixed m.

Theorem 2 *Under the assumptions of Theorem 1, suppose that m is fixed. Then for $k\in\gamma\cap\gamma_0$, a $100(1-\alpha)\%$ confidence interval of σ_k^2 is*

$$\left(\frac{m\hat{\sigma}_k^2(\alpha,\gamma)}{\chi_{m,1-\alpha/2}^2},\frac{m\hat{\sigma}_k^2(\alpha,\gamma)}{\chi_{m,\alpha/2}^2}\right), \tag{15}$$

where $\chi_{m,a}^2$ denotes the $(100a)$-th percentile of the chi-square distribution on m degrees of freedom.

Remark 2 Theorem 2 provides a proper confidence interval for $\hat{\sigma}_k^2(\alpha,\gamma)$ in practice when m is small. Note that the length of the confidence interval of $\sigma_{k,0}^2$ in (15) does not shrink to 0 as $N\to\infty$, which is owing to the fact that $\hat{\sigma}_k^2(\alpha,\gamma)$ is not a consistent estimator of σ_k^2 when m is fixed and $k\in\gamma\cap\gamma_0$.

We close this section by mentioning that although a fixed m hinders us from consistently estimating σ_k^2, the empirical BLUPs of random effects, based on the ML estimator of σ_k^2, are still asymptotically more efficient than the LS predictors, as illustrated in the following example.

Example 1 Consider model (2) with $p = 0, q = 1, n_1 = \cdots = n_m = n$ and $m > 1$ fixed. Assume that (A2) holds with $c_{1,1} = \cdots = c_{m,1} = 1$ and $\ell = 1$. Let $\tilde{\boldsymbol{b}}_i$ be the LS predictor of $\boldsymbol{b}_i$ and $\hat{\boldsymbol{b}}_i(\sigma_1^2, v^2)$ be the BLUP of $\boldsymbol{b}_i$ given (σ_1^2, v^2). Define

$$D(\sigma_1^2, v^2) \equiv \sum_{i=1}^{m} \left\| \boldsymbol{Z}_i\big(\tilde{\boldsymbol{b}}_i - \boldsymbol{b}_i\big) \right\|^2 - \sum_{i=1}^{m} \left\| \boldsymbol{Z}_i\big(\hat{\boldsymbol{b}}_i(\sigma_1^2, v^2) - \boldsymbol{b}_i\big) \right\|^2.$$

Then, we show in Appendix B of the supplementary material that

$$nD(\hat{\sigma}_1^2, \hat{v}^2) = G_{n,m} + o_p(1),$$

where $\hat{\sigma}_1^2$ and $\hat{v}^2$ are the ML estimators of σ_1^2 and v^2, and $G_{n,m}$ is some random variable depending on n, m. Moreover, it is shown in Appendix B that the moments of $G_{n,m}$ do not exist for $m \leq 4$ and

$$\mathrm{E}(G_{n,m}) = \frac{m(m-4)v_0^4}{(m-2)\sigma_{1,0}^2} \tag{16}$$

for $m > 4$. Equation (16) reveals that for any fixed $m > 4$, the empirical BLUP, $\boldsymbol{Z}_i\hat{\boldsymbol{b}}_i(\hat{\sigma}_1^2, \hat{v}^2)$ of $\boldsymbol{Z}_i\boldsymbol{b}_i$, is asymptotically more efficient than its LS counterpart, $\boldsymbol{Z}_i\tilde{\boldsymbol{b}}_i$, even when $\hat{\sigma}_1^2$ is not a consistent estimator of σ_1^2. In addition, the advantage of the former over the latter rapidly increases with m.

3.2 Asymptotics under misspecification

In this subsection, we consider a misspecified model $(\alpha, \gamma) \in (\mathcal{A} \times \mathcal{G}) \setminus (\mathcal{A}_0 \times \mathcal{G}_0)$. We derive not only the convergence rates for $\hat{v}^2(\alpha, \gamma)$ and $\{\hat{\sigma}_k^2(\alpha, \gamma) : k \in \gamma\}$, but also their asymptotic distributions. These results are crucial for developing model selection consistency and efficiency in linear mixed-effects models under fixed m; see Chang et al. (2022).

We start by investigating the asymptotic properties for the ML estimators of v^2 and $\{\sigma_k^2 : k \in \gamma\}$ for $(\alpha, \gamma) \in \mathcal{A}_0 \times (\mathcal{G} \setminus \mathcal{G}_0)$ under a misspecified random-effects model.

Theorem 3 *Under the assumptions of Theorem 1, except that* $(\alpha, \gamma) \in \mathcal{A}_0 \times (\mathcal{G} \setminus \mathcal{G}_0)$,

$$\begin{aligned} \hat{v}^2(\alpha, \gamma) = v_0^2 + \frac{1}{N} \sum_{i=1}^{m} \left(n_i^{\ell} \sum_{k \in \gamma_0 \setminus \gamma} c_{i,k} b_{i,k}^2 \right) + o_p\left(\frac{1}{N} \sum_{i=1}^{m} n_i^{\ell} \right) \\ + O_p\left(\frac{p + mq}{N} \right) + O_p(N^{-1/2}) \end{aligned} \tag{17}$$

and

$$\hat{\sigma}_k^2(\alpha,\gamma)=\begin{cases}\dfrac{1}{m}\displaystyle\sum_{i=1}^{m}b_{i,k}^2+o_p(a_N(\xi,\ell))+o_p(1); & \text{if } k\in\gamma\cap\gamma_0,\\ o_p(a_N(\xi,\ell))+o_p(1); & \text{if } k\in\gamma\setminus\gamma_0,\end{cases}\tag{18}$$

where $a_N(\xi,\ell)=\left(\dfrac{\sum_{i=1}^m n_i^\ell}{\sum_{i=1}^m n_i^\xi}\right)\left(\dfrac{\sum_{i=1}^m n_i^{\xi-\ell}}{m}\right)$. *In addition, if* $\ell<1$, *then*

$$\hat{v}^2(\alpha,\gamma)\xrightarrow{p}v_0^2,\qquad \text{as } N\to\infty.$$

Furthermore, if $\ell\in(0,1/2)$ *and* $p+mq=o\big(N^{1/2}\big)$, *then*

$$N^{1/2}(\hat{v}^2(\alpha,\gamma)-v_0^2)\xrightarrow{d}N(0,2v_0^4),\qquad \text{as } N\to\infty.$$

Remark 3 Note that $\dfrac{1}{N}\displaystyle\sum_{i=1}^{m}\left(n_i^\ell\sum_{k\in\gamma_0\setminus\gamma}c_{i,k}b_{i,k}^2\right)$ in (17) is the dominant bias term contributed by the non-negligible random effects missed by model γ. When $\ell=1$, this term is asymptotically positive and hence $\hat{v}^2(\alpha,\gamma)$ suffers from a upward bias.

For $\xi\geq\ell$ or nearly balanced data, the following corollary shows that $\hat{\sigma}_k^2(\alpha,\gamma)\xrightarrow{p}\sigma_k^2;\ k\in\gamma$, as $m\to\infty$, even though $(\alpha,\gamma)\in\mathcal{A}_0\times(\mathcal{G}\setminus\mathcal{G}_0)$ is misspecified.

Corollary 2 *Under the assumptions of Theorem 3, with* $\xi\geq\ell$ *or* $n_{\max}=O(n_{\min})$,

$$\hat{\sigma}_k^2(\alpha,\gamma)=\begin{cases}\dfrac{1}{m}\displaystyle\sum_{i=1}^{m}b_{i,k}^2+o_p(1); & \text{if } k\in\gamma\cap\gamma_0,\\ o_p(1); & \text{if } k\in\gamma\setminus\gamma_0.\end{cases}$$

If $m\to\infty$, *then*

$$\hat{\sigma}_k^2(\alpha,\gamma)\xrightarrow{p}\sigma_{k,0}^2;\quad k\in\gamma,\qquad \text{as } N\to\infty.$$

The following theorem presents the asymptotic properties of $\hat{v}^2(\alpha,\gamma)$ and $\{\hat{\sigma}_k^2(\alpha,\gamma):k\in\gamma\}$ for $(\alpha,\gamma)\in(\mathcal{A}\setminus\mathcal{A}_0)\times\mathcal{G}$ under a misspecified fixed-effects model.

Theorem 4 *Under the assumptions of Theorem 1 except that* $(\alpha,\gamma)\in(\mathcal{A}\setminus\mathcal{A}_0)\times\mathcal{G}_0$,

$$\begin{aligned}\hat{v}^2(\alpha,\gamma)=v_0^2+\frac{1}{N}\sum_{i=1}^{m}\left(n_i^\xi\sum_{j\in\alpha_0\setminus\alpha}d_{i,j}\beta_{j,0}^2\right)+o_p\left(\frac{1}{N}\sum_{i=1}^{m}n_i^\xi\right)\\+O_p\left(\frac{p+mq}{N}\right)+O_p(N^{-1/2})\end{aligned}\tag{19}$$

and

$$\hat{\sigma}_k^2(\alpha,\gamma) = \begin{cases} \dfrac{1}{m}\displaystyle\sum_{i=1}^{m} b_{i,k}^2 + o_p\left(\dfrac{1}{m}\sum_{i=1}^{m} n_i^{\xi-\ell}\right) + o_p(1); & \text{if } k \in \gamma \cap \gamma_0, \\ o_p\left(\dfrac{1}{m}\displaystyle\sum_{i=1}^{m} n_i^{\xi-\ell}\right) + o_p(1); & \text{if } k \in \gamma \setminus \gamma_0. \end{cases} \tag{20}$$

In addition, if $\xi < 1$, *then*

$$\hat{v}^2(\alpha,\gamma) \xrightarrow{p} v_0^2, \quad \text{as } N \to \infty.$$

Furthermore, if $\xi \in (0, 1/2)$ *and* $p + mq = o(N^{1/2})$, *then*

$$N^{1/2}(\hat{v}^2(\alpha,\gamma) - v_0^2) \xrightarrow{d} N(0, 2v_0^4), \quad \text{as } N \to \infty.$$

Remark 4

(a) When $\xi = 1$, $\dfrac{1}{N}\displaystyle\sum_{i=1}^{m}\left(n_i^{\xi}\sum_{j\in\alpha_0\setminus\alpha} d_{i,j}\beta_{j,0}^2\right)$ in (19) is asymptotically positive, yielding an upward bias in $\hat{v}^2(\alpha,\gamma)$.

(b) For $\gamma \in \mathcal{G}_0$, $\hat{\sigma}_k^2(\alpha,\gamma)$ is consistent when $\xi \le \ell$, as $m \to \infty$.

The following theorem establishes the asymptotic properties of $\hat{v}^2(\alpha,\gamma)$ and $\{\hat{\sigma}_k^2(\alpha,\gamma) : k \in \gamma\}$ for $(\alpha,\gamma) \in (\mathcal{A} \setminus \mathcal{A}_0) \times (\mathcal{G} \setminus \mathcal{G}_0)$ when both the fixed-effects model and the random-effects model are misspecified.

Theorem 5 *Under the assumptions of Theorem 1 except that* $(\alpha,\gamma) \in (\mathcal{A} \setminus \mathcal{A}_0) \times (\mathcal{G} \setminus \mathcal{G}_0)$,

$$\begin{aligned} \hat{v}^2(\alpha,\gamma) = v_0^2 &+ \frac{1}{N}\sum_{i=1}^{m}\left(n_i^{\xi}\sum_{j\in\alpha_0\setminus\alpha} d_{i,j}\beta_{j,0}^2 + n_i^{\ell}\sum_{k\in\gamma_0\setminus\gamma} c_{i,k}b_{i,k}^2\right) \\ &+ o_p\left(\frac{1}{N}\sum_{i=1}^{m}(n_i^{\xi} + n_i^{\ell})\right) + O_p\left(\frac{p+mq}{N}\right) + O_p(N^{-1/2}) \end{aligned} \tag{21}$$

and

$$\hat{\sigma}_k^2(\alpha,\gamma) = \begin{cases} \dfrac{1}{m}\displaystyle\sum_{i=1}^{m} b_{i,k}^2 + o_p(a_N^*(\xi,\ell)) + o_p(1); & \text{if } k \in \gamma \cap \gamma_0, \\ o_p(a_N^*(\xi,\ell)) + o_p(1); & \text{if } k \in \gamma \setminus \gamma_0, \end{cases} \tag{22}$$

where $a_N^*(\xi,\ell) = \left(1 + \dfrac{\sum_{i=1}^{m} n_i^{\ell}}{\sum_{i=1}^{m} n_i^{\xi}}\right)\left(\dfrac{\sum_{i=1}^{m} n_i^{\xi-\ell}}{m}\right)$. *In addition, if* $\max\{\xi,\ell\} < 1$, *then*

$$\hat{v}^2(\alpha,\gamma) \xrightarrow{p} v_0^2, \quad \text{as } N \to \infty.$$

Furthermore, if $(\xi, \ell) \in (0, 1/2) \times (0, 1/2)$ *and* $p + mq = o(N^{1/2})$, *then*

$$N^{1/2}(\hat{v}^2(\alpha, \gamma) - v_0^2) \xrightarrow{d} N(0, 2v_0^4), \quad \text{as } N \to \infty.$$

Remark 5 When $\xi = 1$ or $\ell = 1$, $\frac{1}{N}\sum_{i=1}^{m}\left(n_i^{\xi}\sum_{j\in\alpha_0\setminus\alpha} d_{i,j}\beta_{j,0}^2 + n_i^{\ell}\sum_{k\in\gamma_0\setminus\gamma} c_{i,k}b_{i,k}^2\right)$ in (21) is asymptotically positive, and hence $\hat{v}^2(\alpha, \gamma)$ has a non-negligible positive bias.

Similar to Corollary 2, we also have the following corollary.

Corollary 3 *Under the assumptions of Theorem 5, with either (i)* $\xi = \ell$ *or (ii)* $\xi < \ell$ *and* $n_{\max} = O(n_{\min})$,

$$\hat{\sigma}_k^2(\alpha, \gamma) = \begin{cases} \frac{1}{m}\sum_{i=1}^{m} b_{i,k}^2 + o_p(1); & \text{if } k \in \gamma \cap \gamma_0, \\ o_p(1); & \text{if } k \in \gamma \setminus \gamma_0. \end{cases}$$

If $m \to \infty$, *then*

$$\hat{\sigma}_k^2(\alpha, \gamma) \xrightarrow{p} \sigma_{k,0}^2; \quad k \in \gamma, \quad \text{as } N \to \infty.$$

4 Numerical examples

We conduct two simulation experiments and a real-data example for linear mixed-effects models. The first one examines estimation of mixed-effects models, and the second concerns confidence intervals.

4.1 Experiment 1

We generate data according to (1) with $(p, q) \in \{(5, 5), (15, 15)\}$, where

$$(\sigma_{1,0}^2, \sigma_{2,0}^2, \sigma_{3,0}^2, \sigma_{4,0}^2, \sigma_{5,0}^2, \dots, \sigma_{q,0}^2)' = (0, 0.5, 1, 1.5, \mathbf{0}_{q-4}')',$$
$$(\beta_{1,0}, \beta_{2,0}, \beta_{3,0}, \beta_{4,0}, \dots, \beta_{p,0})' = (1.2, -0.7, 0.8, \mathbf{0}_{p-3}')',$$

$v^2 = 1$, and $\boldsymbol{x}_{i,j} \sim N(\mathbf{0}, \boldsymbol{I}_{n_i})$ and $\boldsymbol{z}_{i,k} \sim N(\mathbf{0}, \boldsymbol{I}_{n_i})$ are independent, for $i = 1, \dots, m$, $j = 1, \dots, p$ and $k = 1, \dots, q$. This setup satisfies (A1)–(A3) with $\xi = \ell = 1$ and $d_{i,j} = c_{i,k} = 1$, for $i = 1, \dots, m$, $j = 1 \dots, p$ and $k = 1, \dots, q$. We consider parameter estimation under two scenarios corresponding to balanced data and unbalanced data. We apply the lme4 package (Bates et al., 2015) in R to obtain the ML estimators of fixed and random effects for our linear mixed-effects models.

For parameter estimation, we consider balanced data with $m \in \{10, 20, 30\}$, $n_1 = \cdots = n_m = m$, and hence $N = m^2$. Since $\beta_{4,0} = \beta_{5,0} = \cdots = \beta_{p,0} = 0$ and $\sigma_{1,0}^2 = \sigma_{5,0}^2 = \sigma_{6,0}^2 = \cdots = \sigma_{q,0}^2 = 0$, we only show results for $\hat{\beta}_1, \dots, \hat{\beta}_4, \hat{\sigma}_2^2, \dots, \hat{\sigma}_5^2$ and

$\hat{\nu}^2$, which includes all the estimators corresponding to $(\alpha_0, \gamma_0) = (\{1,2,3\}, \{2,3,4\})$. The results for the full model $(\{1,\ldots,p\},\{1,\ldots,q\}) \in \mathcal{A}_0 \times \mathcal{G}_0$ based on 100 simulated replicates are summarized in Table 1. The results for the model $(\{1,\ldots,p\},\{3,\ldots,q\}) \in \mathcal{A}_0 \times (\mathcal{G} \setminus \mathcal{G}_0)$ with correct fixed effects but misspecified random effects based on 100 simulated replicates are summarized in Table 2. The results for the model $(\{2,\ldots,p\},\{4,\ldots,q\}) \in (\mathcal{A} \setminus \mathcal{A}_0) \times (\mathcal{G} \setminus \mathcal{G}_0)$ with both misspecified fixed and random effects based on 100 simulated replicates are summarized in Table 3.

As seen in Table 1, all the ML estimators based on the full model have small biases except for $\hat{\nu}^2$ with $m = 10$, where $\hat{\nu}^2$ is under-estimated, particularly when $(p,q) = (15,15)$. Comparing between the two settings of (p, q), the sample standard deviations of the ML estimators are larger when $(p,q) = (15,15)$. This phenomenon can also be found in Tables 2 and 3. We note that biases and standard deviations of ML estimators in Table 1 tend to be smaller when m is larger. In particular, the standard deviation of $\hat{\sigma}_5^2$ is much smaller than the others, which echoes a faster convergence rate of $\hat{\sigma}_k^2$ of Theorem 1, when it converges to zero. Although our main concern is on the estimation of ν^2 and $\hat{\sigma}_k^2$s', the fixed-effects parameters are all well estimated as expected. For model $(\alpha,\gamma) = (\{1,\ldots,p\},\{3,\ldots,q\})$ with misspecified random effects, Table 2 shows that the ML estimator $\hat{\nu}^2$

Table 1 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameters for different values of m obtained from the full model in Experiment 1 with balanced data based on 100 simulated replicates. Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$	$\hat{\nu}^2$
$(p,q) = (5,5)$									
10	1.188	− 0.695	0.790	0.011	0.529	0.972	1.463	0.024	0.871
	(0.116)	(0.124)	(0.125)	(0.140)	(0.311)	(0.564)	(0.683)	(0.043)	(0.159)
20	1.197	− 0.698	0.808	0.001	0.508	0.966	1.481	0.006	0.972
	(0.049)	(0.048)	(0.057)	(0.048)	(0.169)	(0.339)	(0.456)	(0.010)	(0.085)
30	1.206	− 0.696	0.803	− 0.002	0.491	0.988	1.442	0.003	0.992
	(0.039)	(0.034)	(0.036)	(0.035)	(0.153)	(0.250)	(0.302)	(0.004)	(0.047)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
True	1.200	− 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
$(p,q) = (15,15)$									
10	1.203	− 0.683	0.822	0.013	0.572	1.056	1.460	0.059	0.327
	(0.178)	(0.157)	(0.170)	(0.176)	(0.412)	(0.626)	(0.788)	(0.103)	(0.258)
20	1.205	− 0.696	0.800	0.010	0.498	0.989	1.510	0.006	0.868
	(0.055)	(0.054)	(0.057)	(0.051)	(0.201)	(0.275)	(0.490)	(0.012)	(0.075)
30	1.199	− 0.702	0.796	− 0.000	0.485	1.044	1.474	0.004	0.940
	(0.038)	(0.034)	(0.034)	(0.033)	(0.131)	(0.268)	(0.427)	(0.006)	(0.054)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
True	1.200	− 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000

Table 2 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameter for different values of m obtained from model $(\alpha, \gamma) = (\{1, \dots, p\}, \{3, \dots, q\})$ in Experiment 1 with balanced data based on 100 simulated replicates. Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$	$\hat{\nu}^2$
$(p, q) = (5, 5)$								
10	1.198	− 0.698	0.783	0.020	0.962	1.445	0.037	1.373
	(0.130)	(0.147)	(0.144)	(0.149)	(0.555)	(0.714)	(0.074)	(0.359)
20	1.197	− 0.700	0.807	0.004	0.955	1.457	0.013	1.472
	(0.067)	(0.067)	(0.070)	(0.068)	(0.345)	(0.475)	(0.023)	(0.190)
30	1.206	− 0.696	0.800	0.001	0.997	1.447	0.007	1.481
	(0.047)	(0.042)	(0.038)	(0.038)	(0.265)	(0.306)	(0.011)	(0.171)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.500
True	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.000
$(p, q) = (15, 15)$								
10	1.204	− 0.672	0.805	0.002	1.052	1.409	0.053	0.728
	(0.160)	(0.153)	(0.175)	(0.169)	(0.619)	(0.753)	(0.094)	(0.318)
20	1.205	− 0.699	0.796	0.011	0.984	1.525	0.012	1.305
	(0.062)	(0.071)	(0.064)	(0.064)	(0.276)	(0.504)	(0.018)	(0.199)
30	1.198	− 0.705	0.800	0.001	1.049	1.482	0.005	1.394
	(0.042)	(0.042)	(0.040)	(0.043)	(0.274)	(0.436)	(0.009)	(0.128)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.500
True	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.000

overestimates $\nu_0^2 = 1$ by about $\sigma_{2,0}^2 = 0.5$ on average, particularly for larger values of m. This is in line with our conclusion in Theorem 3. Finally, for model $(\alpha, \gamma) = (\{2, \dots, p\}, \{4, \dots, p\})$ with both fixed and random effects misspecified, Table 3 confirms that $\hat{\nu}^2$ is far from its true value and reasonably close to its theoretical limit, $\nu_0^2 + \sigma_{2,0}^2 + \sigma_{3,0}^2 + \beta_{1,0}^2 = 3.69$, derived in Theorem 5. In addition, $\hat{\sigma}_4^2$ tends to be closer to $\sigma_{4,0}^2$ when m is larger, as expected from Theorem 5.

Next, we consider unbalanced data with $m \in \{10, 20, 30\}$ and $N = m^2$. We set $n_1 = [N^{1/4}]$, $n_2 = [N^{3/4}]$, $n_3 = \cdots = n_{m-1} = [(N - n_1 - n_2)/(m - 2)]$, and hence $n_m = N - \sum_{i=1}^{m-1} n_i$. The ML estimators of $\beta_1, \dots, \beta_4, \sigma_2^2, \dots, \sigma_5^2$ and ν^2 under the full model $(\{1, \dots, p\}, \{1, \dots, q\}) \in \mathcal{A}_0 \times \mathcal{G}_0$ based on 100 simulated replicates are summarized in Table 4. The ML estimators of $\beta_1, \dots, \beta_4$, $\sigma_3^2, \sigma_4^2, \sigma_5^2$ and ν^2 under model $(\{1, \dots, p\}, \{3, \dots, q\}) \in \mathcal{A}_0 \times (\mathcal{G} \setminus \mathcal{G}_0)$ with correct fixed effects but misspecified random effects based on 100 simulation runs are summarized in Table 5. The ML estimators of $\beta_2, \beta_3, \beta_4, \sigma_4^2, \sigma_5^2$ and ν^2 under model $(\{2, \dots, p\}, \{4, \dots, q\}) \in (\mathcal{A} \setminus \mathcal{A}_0) \times (\mathcal{G} \setminus \mathcal{G}_0)$ with both misspecified fixed and random effects based on 100 simulated replicates are summarized in Table 6. The results based on unbalanced data can be seen to perform similarly to those based on balanced data.

Table 3 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameters for different values of m obtained from model $(\alpha, \gamma) = (\{2, , \dots, p\}, \{4, \dots, q\})$ in Experiment 1 with balanced data based on 100 simulated replicates

m	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$	$\hat{v}^2$
$(p, q) = (5, 5)$						
10	− 0.666	0.766	0.035	1.452	0.095	3.610
	(0.222)	(0.225)	(0.209)	(0.828)	(0.158)	(0.885)
20	− 0.712	0.804	− 0.007	1.474	0.026	3.847
	(0.109)	(0.109)	(0.103)	(0.535)	(0.046)	(0.427)
30	− 0.701	0.803	− 0.001	1.426	0.011	3.919
	(0.059)	(0.058)	(0.065)	(0.308)	(0.018)	(0.352)
∞	− 0.700	0.800	0.000	1.500	0.000	3.940
True	− 0.700	0.800	0.000	1.500	0.000	1.000
$(p, q) = (15, 15)$						
10	− 0.670	0.837	0.018	1.431	0.144	2.186
	(0.263)	(0.260)	(0.247)	(0.866)	(0.328)	(0.774)
20	− 0.709	0.799	0.012	1.543	0.028	3.479
	(0.108)	(0.112)	(0.100)	(0.526)	(0.046)	(0.382)
30	− 0.705	0.783	0.001	1.456	0.042	3.728
	(0.067)	(0.064)	(0.065)	(0.462)	(0.025)	(0.339)
∞	− 0.700	0.800	0.000	1.500	0.000	3.940
True	− 0.700	0.800	0.000	1.500	0.000	1.000

Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

4.2 Experiment 2

In the second experiment, we compare the conventional confidence interval given by (14) with the proposed confidence interval given by (15). Similar to Experiment 1, we generate data according to (1) with $p = q = 5$, $\boldsymbol{\beta} = (1.2, -0.7, 0.8, 0, 0)'$, $v^2 = 1$, and $(\sigma_{1,0}^2, \sigma_{2,0}^2, \sigma_{3,0}^2, \sigma_{4,0}^2, \sigma_{5,0}^2)' = (0, 0.5, 1, 1.5, 0)'$, where $\boldsymbol{x}_{i,j} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_x)$ and $z_{i,k} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_z)$ are independent, for $i = 1, \dots, m$ and $j, k = 1, \dots, 5$. Here, we consider a more challenging situation of dependent covariates. Specifically, we assume that $\boldsymbol{\Sigma}_x$ is a 5×5 matrix with the (i, j)-th entry $0.4^{|i-j|}$, and $\boldsymbol{\Sigma}_z$ is a 5×5 matrix with the (i, j)-th entry $0.6^{|i-j|}$. We consider balanced data with $n = n_1 = \dots = n_m \in \{10, 50, 100\}$ and three numbers of clusters, $m \in \{2, 5, 10\}$, resulting in a total of nine different combinations.

We compare the 95% confidence intervals of (14) and (15) for σ_2^2 and σ_4^2 based on model $(\alpha, \gamma) = (\{1, 2, 3\}, \{2, 3, 4\})$. The coverage probabilities of both confidence intervals obtained from the two methods for various cases based on 1,000 simulated replicates are shown in Table 7. The proposed method has better coverage probabilities than the conventional ones in almost all cases. The coverage probabilities of our confidence interval tend to the nominal level (i.e., 0.95) as n increases for all cases even when m is very small. In contrast, the conventional method tends to be too optimistic for both σ_2^2 and σ_4^2. For example, the coverage probabilities are less than 0.73 when $m = 2$ regardless of n. Although the coverage probabilities are a bit closer to the nominal level when m is larger, they are

Table 4 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameters for different values of m obtained from the full model in Experiment 1 with unbalanced data based on 100 simulated replicates

m	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$	$\hat{\nu}^2$
$(p, q) = (5, 5)$									
10	1.191	– 0.694	0.795	0.002	0.503	0.960	1.533	0.031	0.880
	(0.120)	(0.129)	(0.131)	(0.132)	(0.312)	(0.537)	(0.782)	(0.068)	(0.170)
20	1.195	– 0.696	0.806	0.002	0.505	0.979	1.468	0.006	0.971
	(0.050)	(0.051)	(0.057)	(0.049)	(0.184)	(0.342)	(0.480)	(0.009)	(0.082)
30	1.206	– 0.697	0.803	– 0.002	0.497	0.989	1.458	0.003	0.992
	(0.038)	(0.034)	(0.036)	(0.034)	(0.151)	(0.232)	(0.323)	(0.005)	(0.045)
∞	1.200	– 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
True	1.200	– 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
$(p, q) = (15, 15)$									
10	1.199	– 0.673	0.808	– 0.001	0.536	0.976	1.411	0.044	0.511
	(0.145)	(0.136)	(0.136)	(0.138)	(0.378)	(0.573)	(0.836)	(0.100)	(0.181)
20	1.205	– 0.695	0.801	0.006	0.502	0.978	1.511	0.006	0.877
	(0.056)	(0.056)	(0.055)	(0.050)	(0.2010)	(0.283)	(0.487)	(0.011)	(0.080)
30	1.199	– 0.702	0.796	– 0.001	0.496	1.050	1.463	0.003	0.948
	(0.039)	(0.034)	(0.033)	(0.032)	(0.159)	(0.288)	(0.415)	(0.006)	(0.057)
∞	1.200	– 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000
True	1.200	– 0.700	0.800	0.000	0.500	1.000	1.500	0.000	1.000

Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

still in the range of (0.82, 0.87) when $m = 10$, showing that the conventional confidence interval is not valid for small m.

4.3 An application to preposition data

In this section, we applied the conventional confidence interval and the proposed one given in (14) and (15), respectively, to a preposition dataset that contains relative frequencies of prepositions in English texts written between 1150 and 1913. The dataset, downloaded from https://slcladal.github.io/data/lmm.rda, contains 537 texts with five variables in each text, including the name of the text, the genre of the text (16 types), the date it is written, the region it is written, and the relative frequency (i.e., the number of prepositions per 1000 words). Following an analysis studied in https://slcladal.github.io/mmws.html, we consider the following linear mixed-effects model:

$$y_{i,j} = \beta_0 + \beta_1 x_{i,j} + b_{i,0} + \epsilon_{i,j}, \quad i = 1, \dots, 16, j = 1, \dots, n_i, \tag{23}$$

where $y_{i,j}$ denotes the frequency of preposition in the jth text for the ith genre, $x_{i,j}$ denotes the centralized date the jth text of the ith genre was written, and $\sum_{i=1}^{16} n_i = 537$ is the total sample size. We fit the model using the lmer function in lme4 package of R. The ML estimator of $\mathrm{var}(b_{i,0})$ is equal to 148.8. Since the

Table 5 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameter for different values of m obtained from model $(\alpha,\gamma)=(\{1,\ldots,p\},\{3,\ldots,q\})$ in Experiment 1 with unbalanced data based on 100 simulated replicates

m	$\hat\beta_1$	$\hat\beta_2$	$\hat\beta_3$	$\hat\beta_4$	$\hat\sigma_3^2$	$\hat\sigma_4^2$	$\hat\sigma_5^2$	$\hat v^2$
$(p,q)=(5,5)$								
10	1.201	− 0.697	0.790	0.002	0.948	1.561	0.040	1.344
	(0.143)	(0.150)	(0.138)	(0.150)	(0.544)	(0.840)	(0.084)	(0.319)
20	1.195	− 0.703	0.806	0.004	0.984	1.471	0.010	1.469
	(0.058)	(0.062)	(0.073)	(0.057)	(0.350)	(0.495)	(0.018)	(0.237)
30	1.206	− 0.696	0.805	0.000	0.993	1.456	0.006	1.480
	(0.045)	(0.044)	(0.044)	(0.042)	(0.233)	(0.321)	(0.009)	(0.223)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.500
True	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.000
$(p,q)=(15,15)$								
10	1.201	− 0.675	0.812	0.007	0.983	1.443	0.065	0.825
	(0.169)	(0.143)	(0.153)	(0.161)	(0.604)	(0.900)	(0.152)	(0.324)
20	1.212	− 0.693	0.796	0.003	0.984	1.512	0.010	1.309
	(0.060)	(0.068)	(0.076)	(0.059)	(0.305)	(0.493)	(0.023)	(0.218)
30	1.198	− 0.704	0.796	− 0.000	1.051	1.476	0.005	1.394
	(0.040)	(0.039)	(0.039)	(0.038)	(0.283)	(0.412)	(0.009)	(0.150)
∞	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.500
True	1.200	− 0.700	0.800	0.000	0.500	1.000	0.000	1.000

Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

number of clusters is merely 16, as demonstrated in our simulation experiments, the conventional confidence interval given by (14) may be problematic. Applying our method in Theorem 2, we obtain an asymptotically valid 95% confidence interval of $\mathrm{var}(b_{i,0})$ to be (82.56, 344.8). Not surprisingly, the range of the interval is large due to a small number of clusters.

5 Discussion

In this article, we establish the asymptotic theory of the ML estimators of random-effects parameters in linear mixed-effects models for unbalanced data, without assuming that m grows to infinity with N. We not only allow the dimensions of both the fixed-effects and random-effects models to go to infinity with N, but also allow both models to be misspecified. In addition, we provide an asymptotic valid confidence interval for the random-effects parameters when m is fixed. These asymptotic results are essential for investigating the asymptotic properties of model-selection methods for linear mixed-effects models, which to the best of our knowledge, have only been developed under the assumption of $m \to \infty$.

Although it is common to assume the random effects to be uncorrelated as done in model (1), it is also of interest to consider correlated random effects with

Table 6 Sample means and sample standard deviations (in parentheses) of ML estimators of selected parameters for different values of m obtained from model $(\alpha,\gamma)=(\{2,,\ldots,p\},\{4,\ldots,q\})$ in Experiment 1 with unbalanced data based on 100 simulated replicates

m	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$	$\hat{v}^2$
$(p,q)=(5,5)$						
10	− 0.683	0.786	0.026	1.501	0.093	3.538
	(0.211)	(0.209)	(0.202)	(0.945)	(0.177)	(0.949)
20	− 0.709	0.813	− 0.001	1.442	0.003	3.865
	(0.106)	(0.105)	(0.106)	(0.530)	(0.041)	(0.522)
30	− 0.703	0.801	− 0.001	1.461	0.012	3.920
	(0.068)	(0.066)	(0.068)	(0.347)	(0.026)	(0.417)
∞	− 0.700	0.800	0.000	1.500	0.000	3.940
True	− 0.700	0.800	0.000	1.500	0.000	1.000
$(p,q)=(15,15)$						
10	− 0.701	0.859	0.001	1.521	0.086	2.426
	(0.238)	(0.216)	(0.227)	(1.142)	(0.202)	(0.677)
20	− 0.700	0.797	0.003	1.526	0.018	3.501
	(0.105)	(0.111)	(0.089)	(0.548)	(0.038)	(0.518)
30	− 0.707	0.788	− 0.001	1.481	0.016	3.742
	(0.072)	(0.069)	(0.069)	(0.439)	(0.031)	(0.401)
∞	− 0.700	0.800	0.000	1.500	0.000	3.940
True	− 0.700	0.800	0.000	1.500	0.000	1.000

Values corresponding to ∞ and True are the theoretical convergent values and the true parameter values, respectively

Table 7 Coverage probabilities (denoted by $\hat{P}$) for 95% confidence intervals of σ_2^2 and σ_4^2 obtained from two methods in Sect. 4.2 based on 1000 simulated replicates

m	n	Classical		Proposed	
		σ_2^2	σ_4^2	σ_2^2	σ_4^2
2	10	0.651 (0.015)	0.649 (0.015)	0.814 (0.012)	0.763 (0.013)
	50	0.724 (0.014)	0.703 (0.014)	0.932 (0.008)	0.935 (0.008)
	100	0.725 (0.014)	0.722 (0.014)	0.942 (0.007)	0.929 (0.008)
5	10	0.778 (0.013)	0.738 (0.014)	0.895 (0.010)	0.871 (0.011)
	50	0.809 (0.012)	0.818 (0.012)	0.936 (0.008)	0.937 (0.008)
	100	0.811 (0.012)	0.809 (0.012)	0.940 (0.008)	0.929 (0.008)
10	10	0.838 (0.012)	0.816 (0.012)	0.900 (0.009)	0.893 (0.010)
	50	0.874 (0.010)	0.849 (0.011)	0.952 (0.007)	0.946 (0.007)
	100	0.849 (0.011)	0.867 (0.011)	0.941 (0.007)	0.956 (0.006)

Values given in parentheses are standard errors of coverage probabilities (evaluated by $\sqrt{\hat{P}(1-\hat{P})/1000}$)

no structure imposed on $\boldsymbol{D}$. However, the technique developed in this article may not be directly applicable to the latter situation; further research in this direction is thus warranted.

Conditions (A1) and (A2) assume that the covariates are asymptotically uncorrelated. These restrictions can be relaxed. Here is a simple example.

Lemma 1 *Consider the data generated from* (2) *with* $m = 1$, $n_1 = N$, $p = q = 2$, *and the true parameters given in* (6). *Suppose that* $(\alpha_0, \gamma_0) = (\{1,2\},\{1,2\})$ *is the smallest true model and* $(\alpha_1, \gamma_1) = (\{1\},\{1\})$ *is a misspecified model defined in* (4). *Let* $\hat{\sigma}_k^2(\alpha,\gamma)$ *and* $\hat{\nu}^2(\alpha,\gamma)$ *be the ML estimators of* σ_k^2 *and* ν^2 *based on* (α,γ). *Assume that* (A1)–(A3) *hold except that* $\boldsymbol{z}_{1,1}'\boldsymbol{z}_{1,2} = c_{1,12}N + o(N)$ *and* $\boldsymbol{x}_{1,1}'\boldsymbol{x}_{1,2} = d_{1,12}N + o(N)$, *for some constants* $c_{1,12}, d_{1,12} \in \mathbb{R}$. *Then*

$$\hat{\nu}^2(\alpha_0,\gamma_0) = \nu_0^2 + O_p(N^{-1/2}),$$

$$\hat{\sigma}_k^2(\alpha_0,\gamma_0) = b_k^2 + O_p(N^{-1/2}); \quad k = 1,2,$$

$$\hat{\nu}^2(\alpha_1,\gamma_1) = \nu_0^2 + \left(d_{1,2} - \frac{d_{1,12}^2}{d_{1,1}}\right)\beta_{2,0}^2 + \left(c_{1,2} - \frac{c_{1,12}^2}{c_{1,1}}\right)b_2^2 + o_p(1),$$

$$\hat{\sigma}_1^2(\alpha_1,\gamma_1) = \left(b_1 + \frac{c_{1,12}}{c_{1,1}}b_2\right)^2 + o_p(1),$$

where $\beta_{2,0} \neq 0$ *is the true parameter of* β_2.

From Lemma 1, it is not surprising to see that $\hat{\nu}^2(\alpha_0,\gamma_0) \xrightarrow{p} \nu_0^2$. On the other hand, $\hat{\nu}^2(\alpha_1,\gamma_1)$ tends to overestimate ν_0^2 by $(d_{1,2} - d_{1,12}^2/d_{1,1})\beta_{2,0}^2 + (c_{1,2} - c_{1,12}^2/c_{1,1})b_2^2$. Since $d_{1,2} - d_{1,12}^2/d_{1,1} \geq 0$ and $c_{1,2} - c_{1,12}^2/c_{1,1} \geq 0$, the amount of overestimation is smaller when either $c_{1,12}^2$ or $d_{1,12}^2$ is larger. In contrast, $\hat{\sigma}_1^2(\alpha_1,\gamma_1)$ tends to be more upward biased when $c_{1,12}^2$ is larger, since $\mathrm{E}\big(b_1 + (c_{1,12}/c_{1,1})b_2\big)^2 = \sigma_1^2 + (c_{1,12}/c_{1,1})^2\sigma_2^2$. Lemma 1 demonstrates how the correlations between the two covariates affect the behavior of $\hat{\nu}^2(\alpha_1,\gamma_1)$ and $\hat{\sigma}_1^2(\alpha_1,\gamma_1)$. However, when the number of covariates is larger, the ML estimators of ν^2 and $\{\sigma_k^2\}$ become much more complicated. We leave this extension of Lemma 1 to the general case for future work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00825-7.

Acknowledgements The research of Chih-Hao Chang is supported by ROC Ministry of Science and Technology grant MOST 107-2118-M-390-001. The research of Hsin-Cheng Huang is supported by ROC Ministry of Science and Technology grant MOST 106-2118-M-001-002-MY3. The research of Ching-Kang Ing is supported by the Science Vanguard Research Program under the Ministry of Science and Technology, Taiwan, ROC.

References

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lm4. *Journal of Statistical Software, 67*, 1–48.

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Ferron, J. M. (2014). How low can you go? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 10,* 1–11.

Chang, C. H., Huang, H. C., Ing, C. K. (2017). Mixed domain asymptotics for a stochastic process model with time trend and measurement error. *Bernoulli, 23*, 159–190.
Chang, C. H., Huang, H. C., Ing, C. K. (2022). Selection of linear mixed-effects models with a small Number of clusters. *Submitted*.
Crainiceanu, C. M., Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of Royal Statistical Society Series B, 66,* 165–185.
Fan, Y., Li, R. (2012). Variable selection in linear mixed effects models. *The Annals of Statistics, 40*, 2043–2068.
Gumedze, F. N., Dunne, T. T. (2011). Parameter estimation and inference in the linear mixed model. *Linear Algebra and Its Applications, 435*, 1920–1944.
Hartley, H. O., Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika, 54*, 93–108.
Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and related problems. *Journal of the American Statistical Association, 7*, 320–338.
Huang, F. (2018). Using cluster bootstrapping to analyze nested data with a few clusters. *Educational and Psychological Measurement, 78*, 297–318.
Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. New York: Springers.
Jiang, J. (2017). *Asymptotic analysis of mixed effects models*: *Theory, applications, and open problems*. New York: Springers.
Laird, N. M., Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*, 963–974.
Longford, N. T. (1993). *Random coefficient models oxford statistical science series 11*. New York: Oxford University Press.
Maas, C. J. M., Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*, 127–137.
Mardia, K. V., Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika, 71*, 135–146.
McNeish, D., Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research, 51*, 495–518.
McNeish, D., Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*, 295–314.
Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics, 5*, 746–762.
Peng, H., Lu, Y. (2012). Model selection in linear mixed effect models. *Journal of Multivariate Analysis, 109*, 109–129.
Pu, W., Niu, X.-F. (2006). Selecting mixed-effects models based on a generalized information criterion. *Journal of Multivariate Analysis, 97*, 733–758.
Self, S. G., Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605–610.
Stram, D. O., Lee, J. W. (1994). Variance component testing in the longitudinal mixed effects model. *Biometrics, 50*, 1171–1177.
Verbeke, G., Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springers.