Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Precise asymptotics for linear mixed models with crossed random effects

Jiming Jiang, Matt P. Wand & Swarnadip Ghosh

To cite this article: Jiming Jiang, Matt P. Wand & Swarnadip Ghosh (10 Mar 2026): Precise asymptotics for linear mixed models with crossed random effects, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2026.2633805

To link to this article: https://doi.org/10.1080/24754269.2026.2633805

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2026.2633805

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Precise asymptotics for linear mixed models with crossed random effects

Jiming Jiang[a], Matt P. Wand[b] and Swarnadip Ghosh[c]

[a]Department of Statistics, University of California, Davis, USA; [b]School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, Australia; [c]Radix Trading, New York, USA

ABSTRACT
We obtain an asymptotic normality result that reveals the precise asymptotic behaviour of the maximum likelihood estimators of parameters for a very general class of linear mixed models containing cross random effects. In achieving the result, we overcome theoretical difficulties that arise from random effects being crossed as opposed to the simpler nested random effects case. Our new theory is for a class of Gaussian response linear mixed models which include crossed random slopes that partner arbitrary multivariate predictor effects and do not require the cell counts to be balanced. Statistical utilities include the confidence interval construction, Wald hypothesis test and sample size calculations.

ARTICLE HISTORY
Received 29 April 2025
Revised 10 September 2025
Accepted 12 February 2026

KEYWORDS
Asymptotic normality; maximum likelihood estimation; sample size calculations

1. Introduction

Linear mixed models with crossed random effects are useful for the analysis of regression-type data that are cross-classified according to two or more grouping mechanisms. Baayen et al. (2008), for example, use the terms *subjects* and *items* for groupings that are typical in psychology studies. Specific examples discussed in Baayen et al. (2008) have subjects corresponding to human participants in a psycholinguistic experiment and items corresponding to words in a particular language. Gao and Owen (2020) and Ghosh et al. (2022) are concerned with electronic commerce and related applications involving crossed random effects, and make subjects and items correspond to customers and products.

Despite the widespread use of linear mixed models with crossed random effects, the theory concerning the asymptotic behaviours of model parameter estimators is scant. This is largely due to the complicated mathematical forms that arise from random effects being crossed. Unlike the nested random effects case, the marginal covariance matrix of the response vector does not have a block diagonal form, which makes theoretical analyses significantly more challenging. For Gaussian response linear mixed models with nested random effects precise asymptotics are relatively straightforward as conveyed by, for example, Section 3.5 of McCulloch et al. (2008). Recently Jiang et al. (2022) obtained a precise asymptotic normality result for the joint distribution of all model parameters in a generalized linear mixed model with

CONTACT Matt P. Wand matt.wand@uts.edu.au School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123, Broadway, Ultimo, Sydney NSW 2007, Australia

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/24754269.2026.2633805.

nested random effects. In this article we derive an analogous result for Gaussian response linear mixed models with crossed random effects.

Some early contributions to asymptotic theory for linear mixed models with crossed random effects structures are Hartley and Rao (1967) and Miller (1977). Indeed, the second example in Section 4 of Miller (1977) corresponds to a special case of the class of linear mixed models considered in the present article when his c_{ij} term is omitted. Further details concerning this example are in Sections 6.1 and 6.2 of Miller (1973), and include an expression for the asymptotic covariance matrix of the maximum likelihood estimator of the vector of variance parameters. Asymptotic normality of the maximum likelihood estimators is also established in Miller (1973,1977). However, the explicit results in these seminal articles are confined to balanced linear mixed models that are devoid of predictor data. Jiang (1996) focussed on restricted maximum likelihood (REML) estimation of variance parameters in a wide class of linear mixed models that include those containing crossed random effects and obtained conditions under which asymptotic normality of the REML estimators holds. The results in Jiang (1996) are expressed in terms of generic Fisher information matrices rather than the explicit asymptotic forms provided by Jiang et al. (2022). Lyu et al. (2024) is a recent article that is also concerned with asymptotic normality of estimators in a crossed random effects setting. Connections between Lyu et al. (2024) and this paper are described below.

In this article we obtain precise asymptotics, in a similar vein to those of Jiang et al. (2022), for Gaussian response linear mixed models with crossed random effects. Our results apply to a wide class of situations that include unbalanced designs, predictor data and multivariate crossed random effects. They reveal that asymptotic covariance matrices of the estimators parameter vectors are quite similar to those that arise for nested random effects despite inherent differences due to effects being crossed. For example, the estimates of fixed effects parameters that are unaccompanied by random effects have the same asymptotic variances regardless of whether the model contains nested or crossed random effects. However, as we shall see, the pathway towards establishing such results for the crossed random effects case is much longer and involved.

The majority of the research in this article was done concurrently with and independently of the Lyu et al. (2024) research and we became aware of their article after devising Result 3.1. The linear mixed model treated by Lyu et al. (2024) does not assume that the responses are Gaussian. They also include a random interaction term, which our model does not have. In the case of Gaussian responses and additive crossed random effects, our main result extends the theoretical findings of Lyu et al. (2024) in the following two ways: (1) multivariate random slopes are included and (2) unbalanced cell counts are accommodated. Each of (1) and (2) is quite important in practice, but requires lengthy matrix algebraic and convergence in probability arguments since the deterministic Kronecker product forms used in Miller (1973) and Lyu et al. (2024) no longer apply.

Contemporary data sets for which linear mixed models with crossed random effects provide a useful vehicle for analysis vastly differ in terms of the density of the observations. For some applications, the cell counts arising from subject/item cross-classification are all nonzero. As an example, the illustration given in Section 6 of Menictas et al. (2023) for the U.S. National Education Longitudinal Study has $8564 \times 24 = 205{,}488$ cells with a few observations per cell. The rows correspond to 8564 U.S. school students and the columns correspond to 24 items such as reading, mathematics and science ability. The responses correspond to the scores for each student/item combination. The students were followed longitudinally, which resulted in higher cell counts. Predictor data such as gender, time spent on homework and parental education were also recorded. Other data sets, such as those that motivate Ghosh et al., 2022, have total number of observations much lower than the number of cells. Ghosh et al. (2022) describe an example concerning customer ratings from the clothing company Stitch Fix with 762,752 × 6318 cells. The rows correspond to 762,752 customers and the columns correspond to 6318 clothing items. There are five million ratings, which means that the average cell count is approximately 0.001. In this article we focus on dense data situations where the cell counts are non-zero and growing in our asymptotic analyses. Relaxation to various sparse data situations is certainly of interest but, with conciseness and closure in mind, this is left aside in this article's theoretical study.

Generalized linear mixed models with crossed random effects are particularly challenging theoretically and it was not until Jiang (2013) that a consistency proof was established. Associated asymptotic distribution theory was established recently in Jiang (2025) for an intercepts-only special case. In this article we treat the Gaussian response case for a very general class of crossed random effects models.

The linear mixed model with crossed random effects that we study is described in Section 2, as well as maximum likelihood estimation of the model parameters. An asymptotic normality result that reveals the precise asymptotic behaviour of all maximum likelihood estimators is given in Section 3. A key finding in Section 3 is that the leading terms are very similar to those arising in nested random effects models. In Section 4 we provide some heuristic arguments that help explain these similarities. Section 5 discusses statistical utility of the new theory. Some concluding remarks are made in Section 6. An online supplement provides derivational details of the central result.

2. Model description and maximum likelihood estimation

Consider the following crossed random effects linear mixed models:

$$\begin{aligned} &\boldsymbol{Y}_{ii'} \mid \boldsymbol{U}_i, \boldsymbol{U}'_{i'}, \boldsymbol{X}_{\mathrm{A}ii'}, \boldsymbol{X}_{\mathrm{B}ii'} \overset{\text{ind.}}{\sim} N(\boldsymbol{X}_{\mathrm{A}ii'}(\boldsymbol{\beta}_{\mathrm{A}}^0 + \boldsymbol{U}_i + \boldsymbol{U}'_{i'}) + \boldsymbol{X}_{\mathrm{B}ii'}\boldsymbol{\beta}_{\mathrm{B}}^0, (\sigma^2)^0 I),\\ &\boldsymbol{U}_i \overset{\text{ind.}}{\sim} N(0, \boldsymbol{\Sigma}^0), \quad 1 \le i \le m, \quad \boldsymbol{U}'_{i'} \overset{\text{ind.}}{\sim} N(0, (\boldsymbol{\Sigma}')^0), \quad 1 \le i' \le m' \end{aligned} \tag{1}$$

where here, and throughout this article, $\overset{\text{ind.}}{\sim}$ stands for 'independently distributed as'.

The dimensions of the matrices in (1) are as follows:

$$\begin{aligned} &\boldsymbol{Y}_{ii'} \text{ is } n_{ii'} \times 1, \quad \boldsymbol{X}_{\mathrm{A}ii'} \text{ is } n_{ii'} \times d_{\mathrm{A}}, \quad \boldsymbol{\beta}_{\mathrm{A}}^0 \text{ is } d_{\mathrm{A}} \times 1, \quad \boldsymbol{U}_i \text{ is } d_{\mathrm{A}} \times 1, \quad \boldsymbol{U}'_{i'} \text{ is } d_{\mathrm{A}} \times 1,\\ &\boldsymbol{X}_{\mathrm{B}ii'} \text{ is } n_{ii'} \times d_{\mathrm{B}}, \quad \boldsymbol{\beta}_{\mathrm{B}}^0 \text{ is } d_{\mathrm{B}} \times 1, \quad \boldsymbol{\Sigma}^0 \text{ is } d_{\mathrm{A}} \times d_{\mathrm{A}} \quad \text{and} \quad (\boldsymbol{\Sigma}')^0 \text{ is } d_{\mathrm{A}} \times d_{\mathrm{A}}. \end{aligned}$$

Here $n_{ii'}$ is the number of response measurements in the (i,i')th cell. If $n_{ii'} = 0$ then each of $\boldsymbol{Y}_{ii'}$, $\boldsymbol{X}_{\mathrm{A}ii'}$ and $\boldsymbol{X}_{\mathrm{B}ii'}$ are null. The focus of this article is the precise asymptotic properties of the maximum likelihood estimators of the model parameters when m, m' and the $n_{ii'}$ all diverge to ∞. Therefore, from now onwards, we assume that $n_{ii'} > 0$ for all $1 \le i \le m$ and $1 \le i' \le m'$.

In (1), let the rows of $\boldsymbol{X}_{\mathrm{A}ii'}$ and $\boldsymbol{X}_{\mathrm{B}ii'}$ be defined according to the notation

$$\boldsymbol{X}_{\mathrm{A}ii'} = \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}ii'1}^{\top} \\ \vdots \\ \boldsymbol{X}_{\mathrm{A}ii'n_{ii'}}^{\top} \end{bmatrix} \quad \text{and} \quad \boldsymbol{X}_{\mathrm{B}ii'} = \begin{bmatrix} \boldsymbol{X}_{\mathrm{B}ii'1}^{\top} \\ \vdots \\ \boldsymbol{X}_{\mathrm{B}ii'n_{ii'}}^{\top} \end{bmatrix}.$$

We assume that the $\boldsymbol{X}_{\mathrm{A}ii'j}$, $1 \le i \le m$, $1 \le i' \le m'$, $1 \le j \le n_{ii'}$ are independent and identically distributed $d_{\mathrm{A}} \times 1$ random vectors having the same distribution as $\boldsymbol{X}_{\mathrm{A}\circ}$. Similarly, the $\boldsymbol{X}_{\mathrm{B}ii'j}$ over the same index set are independent and identically distributed $d_{\mathrm{B}} \times 1$ random vectors having the same distribution as $\boldsymbol{X}_{\mathrm{B}\circ}$.

The following matrix assembly notation is useful for describing the maximum likelihood estimators and their asymptotic properties. Firstly,

$$\underset{1\le i\le d}{\mathrm{stack}}(\boldsymbol{A}_i) = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_d \end{bmatrix} \quad \text{and} \quad \underset{1\le i\le d}{\mathrm{blockdiag}}(\boldsymbol{A}_i) = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{A}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_d \end{bmatrix}$$

for matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_d$. The first of these definitions requires that $\boldsymbol{A}_i$, $1 \le i \le d$ have the same number of columns. Next, define

$$\underset{1\le i,\underset{\sim}{i}\le d}{\mathrm{blockmatrix}}(\boldsymbol{B}_{i\underset{\sim}{i}}) = \begin{bmatrix} \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1d} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B}_{d1} & \cdots & \boldsymbol{B}_{dd} \end{bmatrix}$$

for matrices $\boldsymbol{B}_{i\underset{\sim}{i}}$, $1 \le i, \underset{\sim}{i} \le d$, each having the same numbers of rows and columns. If we then define

$$\begin{gathered} n_{\bullet\bullet} = \sum_{i=1}^{m}\sum_{i'=1}^{m'} n_{ii'}, \quad \boldsymbol{Y} \equiv \underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{Y}_{ii'})\right\}, \\ \boldsymbol{X}_{\mathrm{A}} \equiv \underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{X}_{\mathrm{A}ii'})\right\} \quad \text{and} \quad \boldsymbol{X}_{\mathrm{B}} \equiv \underset{1\le i\le m}{\mathrm{stack}}\left\{\underset{1\le i'\le m'}{\mathrm{stack}}(\boldsymbol{X}_{\mathrm{B}ii'})\right\}, \end{gathered} \tag{2}$$

then standard manipulations show that

$$\boldsymbol{Y} \mid \boldsymbol{X}_{\mathrm{A}}, \boldsymbol{X}_{\mathrm{B}} \sim N\left(\boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}}^0 + \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}^0, \boldsymbol{V}(\boldsymbol{\Sigma}^0, (\boldsymbol{\Sigma}')^0, (\sigma^2)^0)\right)$$

where

$$\begin{aligned} \boldsymbol{V}(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2) \equiv\ & \underset{1\le i\le m}{\mathrm{blockdiag}}\left\{\underset{1\le i'\le \underset{\sim}{i}'\le m'}{\mathrm{blockmatrix}}(\boldsymbol{X}_{\mathrm{A}ii'}\boldsymbol{\Sigma}\boldsymbol{X}_{\mathrm{A}ii'}^{\top})\right\} \\ & + \underset{1\le i,\underset{\sim}{i}\le m}{\mathrm{blockmatrix}}\left\{\underset{1\le i'\le m'}{\mathrm{blockdiag}}(\boldsymbol{X}_{\mathrm{A}ii'}\boldsymbol{\Sigma}'\boldsymbol{X}_{\mathrm{A}ii'}^{\top})\right\} + \sigma^2\boldsymbol{I}_{n_{\bullet\bullet}}. \end{aligned} \tag{3}$$

Therefore, the conditional log-likelihood is

$$\begin{aligned} & \ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2) \\ & \quad = -\tfrac{1}{2}n_{\bullet\bullet}\log(2\pi) - \tfrac{1}{2}\log|\boldsymbol{V}(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2)| \\ & \qquad - \tfrac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}})^{\top}\boldsymbol{V}(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2)^{-1}(\boldsymbol{Y} - \boldsymbol{X}_{\mathrm{A}}\boldsymbol{\beta}_{\mathrm{A}} - \boldsymbol{X}_{\mathrm{B}}\boldsymbol{\beta}_{\mathrm{B}}). \end{aligned} \tag{4}$$

The maximum likelihood estimator of $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}^0, (\boldsymbol{\Sigma}')^0, (\sigma^2)^0)$ is

$$(\hat{\boldsymbol{\beta}}_{\mathrm{A}}, \hat{\boldsymbol{\beta}}_{\mathrm{B}}, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\Sigma}}', \hat{\sigma}^2) \equiv \underset{\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2}{\mathrm{argmax}}\ \ell(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}', \sigma^2).$$

3. Asymptotic normality result

We now present the article's main centerpiece: an asymptotic normality result that reveals the precise asymptotic behaviour of the maximum likelihood estimation of $(\boldsymbol{\beta}_{\mathrm{A}}^0, \boldsymbol{\beta}_{\mathrm{B}}^0, \boldsymbol{\Sigma}^0, (\boldsymbol{\Sigma}')^0, (\sigma^2)^0)$ for data corresponding to (1).

Define

$$n \equiv \frac{n_{\bullet\bullet}}{mm'} = \text{average of the within-cell sample sizes}$$

and

$$\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}} \equiv \text{lower right } d_{\mathrm{B}} \times d_{\mathrm{B}} \text{ block of } \{E(\boldsymbol{X}_0\boldsymbol{X}_0^{\top})\}^{-1} \quad \text{where } \boldsymbol{X}_0 \equiv \begin{bmatrix} \boldsymbol{X}_{\mathrm{A}\circ} \\ \boldsymbol{X}_{\mathrm{B}\circ} \end{bmatrix}.$$

Let $\boldsymbol{D}_d$ denote the matrix of zeroes and ones such that $\boldsymbol{D}_d\text{vech}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A})$ for all $d \times d$ symmetric matrices $\boldsymbol{A}$. The Moore-Penrose inverse of $\boldsymbol{D}_d$ is $\boldsymbol{D}_d^+ = (\boldsymbol{D}_d^{\top}\boldsymbol{D}_d)^{-1}\boldsymbol{D}_d^{\top}$.

The result relies on the following assumptions.

(A1) The cell dimensions m and m' diverge to ∞ in such a way that $m = O(m')$ and $m' = O(m)$.

(A2) The within-cell sample sizes $n_{ii'}$ diverge to ∞ in such a way that

$$\max_{1\le i\le m,\ 1\le i'\le m'} |n_{ii'}/n - C_{ii'}| \to 0 \quad \text{as } m, m' \to \infty$$

for positive constants $C_{ii'}$, $1 \le i \le m$, $1 \le i' \le m'$, that are bounded above and away from zero. Also, $n/m \to 0$ as m and n diverge.

(A3) All entries of both $\boldsymbol{X}_{\mathrm{A}\circ}$ and $\boldsymbol{X}_{\mathrm{B}\circ}$ are not degenerate at zero and have finite second moment.

Result 3.1: Assume that (A1)–(A3) and some additional regularity conditions hold. Then

$$\begin{bmatrix} \left\{\dfrac{\boldsymbol{\Sigma}^0}{m} + \dfrac{(\boldsymbol{\Sigma}')^0}{m'}\right\}^{-1/2} \left(\hat{\boldsymbol{\beta}}_{\mathrm{A}} - \boldsymbol{\beta}_{\mathrm{A}}^0\right) \\ \left\{\dfrac{(\sigma^2)^0\boldsymbol{C}_{\boldsymbol{\beta}_{\mathrm{B}}}}{mm'n}\right\}^{-1/2} \left(\hat{\boldsymbol{\beta}}_{\mathrm{B}} - \boldsymbol{\beta}_{\mathrm{B}}^0\right) \\ \left\{\dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^+(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\mathrm{A}}}^{+\top}}{m}\right\}^{-1/2} \text{vech}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^0) \\ \left\{\dfrac{2\boldsymbol{D}_{d_{\mathrm{A}}}^+((\boldsymbol{\Sigma}')^0 \otimes (\boldsymbol{\Sigma}')^0)\boldsymbol{D}_{d_{\mathrm{A}}}^{+\top}}{m'}\right\}^{-1/2} \text{vech}(\hat{\boldsymbol{\Sigma}}' - (\boldsymbol{\Sigma}')^0) \\ \left[\dfrac{2\{(\sigma^2)^0\}^2}{mm'n}\right]^{-1/2} (\hat{\sigma}^2 - (\sigma^2)^0) \end{bmatrix} \xrightarrow{D} N(\boldsymbol{0}, \boldsymbol{I}).$$

Some remarks concerning Result 3.1 are as follows.

(1) Result 3.1 provides following asymptotic covariance matrices of the maximum likelihood estimators:

$$\text{Asy.Cov}(\hat{\boldsymbol{\beta}}_{\text{A}}) = \frac{\boldsymbol{\Sigma}^0}{m} + \frac{(\boldsymbol{\Sigma}')^0}{m'},$$

$$\text{Asy.Cov}(\hat{\boldsymbol{\beta}}_{\text{B}}) = \frac{(\sigma^2)^0 \boldsymbol{C}_{\boldsymbol{\beta}_{\text{B}}}}{mm'n},$$

$$\text{Asy.Cov}(\hat{\boldsymbol{\Sigma}}) = \frac{2\boldsymbol{D}_{d_{\text{A}}}^{+}(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_{\text{A}}}^{+\top}}{m},$$

$$\text{Asy.Cov}(\hat{\boldsymbol{\Sigma}}') = \frac{2\boldsymbol{D}_{d_{\text{A}}}^{+}((\boldsymbol{\Sigma}')^0 \otimes (\boldsymbol{\Sigma}')^0)\boldsymbol{D}_{d_{\text{A}}}^{+\top}}{m'}, \quad \text{and}$$

$$\text{Asy.Var}(\hat{\sigma}^2) = \frac{2\{(\sigma^2)^0\}^2}{mm'n}.$$

Note that $\text{Asy.Var}(\hat{\sigma}^2)$ is based on the fact that, for large m, m' and n, $\hat{\sigma}^2$ has an approximate normal distribution with mean $(\sigma^2)^0$ and variance $\text{Asy.Var}(\hat{\sigma}^2)$. There are marked differences in the rates of convergence. For example, the entries of $\hat{\boldsymbol{\beta}}_{\text{A}}$ have order m^{-1} asymptotic variances, whilst those of $\hat{\boldsymbol{\beta}}_{\text{B}}$ have order $(mm'n)^{-1}$ asymptotic variances. Note that $\boldsymbol{\beta}_{\text{A}}^0$ and $\boldsymbol{\beta}_{\text{B}}^0$ differ in that the former is partnered by crossed random effects in (1).

(2) The asymptotic normality results for $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}'$ can be converted to forms that are more amenable to interpretation and confidence interval construction using the Multivariate Delta Method (e.g. Agresti, 2013, Section 16.1.3). For example, if $d_{\text{A}} = 2$ and the entries of $\boldsymbol{\Sigma}$ are parameterized as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

then Result 3.1 implies the following asymptotic normality results for standard transformations of the first standard deviation parameter and correlation parameter:

$$\sqrt{m}\left\{\log(\hat{\sigma}_1) - \log(\sigma_1^0)\right\} \xrightarrow{D} N\left(0, \tfrac{1}{2}\right)$$

and

$$\sqrt{m}\left\{\tanh^{-1}(\hat{\rho}) - \tanh^{-1}(\rho^0)\right\} \xrightarrow{D} N(0, 1).$$

Analogous results hold for $\hat{\sigma}_2$ and $\hat{\boldsymbol{\Sigma}}'$.

(3) There is asymptotic orthogonality between each pair of random vectors within the set

$$\left\{\hat{\boldsymbol{\beta}}_{\text{A}}, \hat{\boldsymbol{\beta}}_{\text{B}}, \text{vech}(\hat{\boldsymbol{\Sigma}}), \text{vech}(\hat{\boldsymbol{\Sigma}}'), \hat{\sigma}^2\right\}.$$

(4) Outside of Result 3.1 and Lyu et al. (2024), we are not aware of results for linear mixed models with crossed random effects that provide the precise asymptotic covariances given by Result 3.1 for estimation of fixed effects, even for simplified versions of (1) such as those having $\boldsymbol{X}_{\text{A}ii'} = \mathbf{1}_{n_{ii'}}$ and $\boldsymbol{X}_{\text{B}ii'}$ null. In this special case, in which the only fixed effect is the intercept parameter, the $\boldsymbol{\Sigma}/m + \boldsymbol{\Sigma}'/m'$ leading term behaviour is also apparent from Theorem 1 of Lyu et al. (2024) when their variable η is in the interior of the

positive half-line. The predictor set-ups differ between the two articles, which hinders succinct comparison of the fixed effects results for more general cases.

(5) Result 3.1 extends the results of Miller (1973) and Lyu et al. (2024), concerning asymptotic distributions of variance component estimators, to covariance matrices of arbitrary dimension.

(6) Under (A1) m and m' diverge to ∞ at the same rate. In some circumstances this assumption may not be realistic and other assumptions concerning m and m' divergence may be more appropriate. The subsequent modification of Result 3.1 is straightforward. For example, if $m' = o(m)$ then the component concerning $\hat{\boldsymbol{\beta}}_{\mathrm{A}}$ becomes

$$\left\{\frac{(\boldsymbol{\Sigma}')^0}{m'}\right\}^{-1/2}\left(\hat{\boldsymbol{\beta}}_{\mathrm{A}}-\boldsymbol{\beta}_{\mathrm{A}}^0\right)\xrightarrow{D}N(\mathbf{0},\boldsymbol{I})\quad\text{leading to}\quad\text{Asy.Cov}(\hat{\boldsymbol{\beta}}_{\mathrm{A}})=\frac{(\boldsymbol{\Sigma}')^0}{m'}.$$

(7) The asymptotic covariances for linear mixed models with crossed random effects have forms that are very similar to those with two-level nested random effects. See, for example, the Gaussian special case of Theorem 1 of Jiang et al. (2022). At first glance, this result is somewhat surprising and intriguing since the two types of linear mixed models have fundamental differences. In Section 4 we provide some heuristic arguments that help explain this interesting phenomenon.

(8) For the special case $\boldsymbol{X}_{\mathrm{A}\circ} = 1$ and $\boldsymbol{X}_{\mathrm{B}\circ} = X_\circ$ we have

$$\text{Asy.Var}(\hat{\boldsymbol{\beta}}_{\mathrm{B}})=\frac{(\sigma^2)^0}{\text{Var}(X_\circ)(\text{total sample size})}.$$

This matches the well-known expression for the asymptotic variance of the slope parameter in the simple linear regression model. Analogous results arise when $\boldsymbol{X}_{\mathrm{B}\circ}$ is multivariate. Despite the presence of crossed random effects, the asymptotic behaviours of the estimators of slope parameters that are unaccompanied by random effects are the same as in the ordinary multiple regression situation. The heuristics in Section 4 provide some insight into this phenomenon.

(9) The presence of multivariate random slopes in the crossed random effects model (1) leads to considerable challenges in the establishment of the Result 3.1 precise asymptotic normality statement. Detailed and delicate arguments, not given here, would be required to obtain sufficient regularity conditions under which Result 3.1 holds.

(10) *Restricted* maximum likelihood estimation is a commonly used alternative to maximum likelihood estimation in linear mixed models-based analyses. For model (1), it involves replacement of (4) by the restricted log-likelihood

$$\begin{aligned}&\ell_R(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)\\&\quad\equiv\ell(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)-\tfrac{1}{2}\log|[\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}]^\top\boldsymbol{V}(\boldsymbol{\Sigma},\boldsymbol{\Sigma}',\sigma^2)^{-1}[\boldsymbol{X}_{\mathrm{A}}\,\boldsymbol{X}_{\mathrm{B}}]|.\end{aligned}$$

The extra term invokes a finite sample adjustment to the estimators. Result 3.1, which is concerned with large sample behaviour, also applies to the restricted maximum likelihood estimators of the parameters in (1).

(11) The establishment of Result 3.1 requires complicated and long-winded arguments, and is deferred to an online supplement.

4. Heuristics on nested/crossed asymptotics similarities

We now address the fact that the asymptotic covariance expressions in Result 3.1 are quite similar to those arising in the two-level nested case. This involves heuristic arguments that show that the fixed effects maximum likelihood estimators admit quite similar forms when sample means are replaced by population means. Throughout this section we write $\boldsymbol{\beta}$ rather than $\boldsymbol{\beta}^0$. A similar convention is used for $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}'$ and σ^2. This suppression of the 'true value' notation is to aid exposition.

Gaussian response linear mixed models have the following general form:

$$\boldsymbol{Y} \,|\, \boldsymbol{U} \sim N(\boldsymbol{X\beta} + \boldsymbol{ZU}, \boldsymbol{R}), \quad \boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{G}). \tag{5}$$

For the crossed random effects model (1)

$$\begin{aligned} \boldsymbol{X} &= [\boldsymbol{X}_{\mathrm{A}}\ \boldsymbol{X}_{\mathrm{B}}], \\ \boldsymbol{Z} &= \left[\underset{1\le i\le m}{\text{blockdiag}} \left\{ \underset{1\le i'\le m'}{\text{stack}} (\boldsymbol{X}_{\mathrm{A}ii'}) \right\} \ \underset{1\le i\le m}{\text{stack}} \left\{ \underset{1\le i'\le m'}{\text{blockdiag}} (\boldsymbol{X}_{\mathrm{A}ii'}) \right\} \right], \\ \boldsymbol{G} &= \text{blockdiag}(\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}, \boldsymbol{I}_{m'} \otimes \boldsymbol{\Sigma}') \quad \text{and} \quad \boldsymbol{R} = \sigma^2 \boldsymbol{I}, \end{aligned}$$

where $\boldsymbol{X}_{\mathrm{A}}$ and $\boldsymbol{X}_{\mathrm{B}}$ are given by (2).

The Gaussian version of the class of *nested* linear mixed models studied by Jiang et al. (2022) is

$$\begin{aligned} \boldsymbol{Y}_i \,|\, \boldsymbol{U}_i, \boldsymbol{X}_{\mathrm{A}i}, \boldsymbol{X}_{\mathrm{B}i} &\overset{\text{ind.}}{\sim} N(\boldsymbol{X}_{\mathrm{A}i}(\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U}_i) + \boldsymbol{X}_{\mathrm{B}i}\boldsymbol{\beta}_{\mathrm{B}}, \sigma^2 \boldsymbol{I}), \\ \boldsymbol{U}_i &\overset{\text{ind.}}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad 1 \le i \le m, \end{aligned} \tag{6}$$

which is a special case of (5) with

$$\begin{aligned} \boldsymbol{X} &= \underset{1\le i\le m}{\text{stack}} [\boldsymbol{X}_{\mathrm{A}i}\ \boldsymbol{X}_{\mathrm{B}i}], \\ \boldsymbol{Z} &= \underset{1\le i\le m}{\text{blockdiag}} (\boldsymbol{X}_{\mathrm{A}i}), \\ \boldsymbol{G} &= \boldsymbol{I}_m \otimes \boldsymbol{\Sigma} \quad \text{and} \quad \boldsymbol{R} = \sigma^2 \boldsymbol{I}. \end{aligned}$$

Analogous to the set-up for model (1), we assume that the transposes of the rows of $\boldsymbol{X}_{\mathrm{A}i}$, $1 \le i \le m$, are independent and identically distributed $d_{\mathrm{A}} \times 1$ random vectors having the same distribution as $\boldsymbol{X}_{\mathrm{A}\circ}$. A similar assumption applies to the $\boldsymbol{X}_{\mathrm{B}i}$.

In terms of the notation in (5), the fixed effects maximum likelihood estimator has the following generalized least squares form:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{Y} \quad \text{where } \boldsymbol{V} \equiv \boldsymbol{ZGZ}^\top + \boldsymbol{R}.$$

If $\mathcal{X}$ denotes the predictor data in the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices then the conditional covariance matrix of the fixed effects estimator is

$$\text{Cov}(\hat{\boldsymbol{\beta}} | \mathcal{X}) = (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}.$$

For the remainder of this section we assume that the data are balanced. In the crossed case this corresponds to $n_{ii'} = n$ for all $1 \le i \le m$ and $1 \le i' \le m'$. For the nested case $n_i = n$ for all $1 \le i \le m$.

4.1. The $X = 1$ special case

Consider the following special case of (5):

$$\boldsymbol{Y} \mid \boldsymbol{U} \sim N(\boldsymbol{1}\beta_0 + \boldsymbol{Z}\boldsymbol{U}, \boldsymbol{R}), \quad \boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{G})$$

for which $\boldsymbol{X} = \boldsymbol{1}$, such that the only fixed effect is the intercept parameter β_0.

A further simplification is

$$\boldsymbol{Z} = \begin{cases} [\boldsymbol{I}_m \otimes \boldsymbol{1}_{m'n} \quad \boldsymbol{1}_m \otimes \boldsymbol{I}_{m'} \otimes \boldsymbol{1}_n], & \text{for the crossed case,} \\ \boldsymbol{I}_m \otimes \boldsymbol{1}_n, & \text{for the nested case,} \end{cases} \tag{7}$$

which corresponds to the random intercept-only models. Let $\boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V}_{\text{nest}}$ respectively denote the $\boldsymbol{V}$ matrix for the crossed and nested cases based on the versions of $\boldsymbol{Z}$ given in (7). Bringing in the commonly used notation $\boldsymbol{J}_d \equiv \boldsymbol{1}_d \boldsymbol{1}_d^\top$ we then have

$$\boldsymbol{V}_{\text{cross}} = \Sigma(\boldsymbol{I}_m \otimes \boldsymbol{J}_{m'n}) + \Sigma'(\boldsymbol{J}_m \otimes \boldsymbol{I}_{m'} \otimes \boldsymbol{J}_n) + \sigma^2 \boldsymbol{I}_{mm'n} \quad \text{and}$$
$$\boldsymbol{V}_{\text{nest}} = \Sigma(\boldsymbol{I}_m \otimes \boldsymbol{J}_n) + \sigma^2 \boldsymbol{I}_{mn},$$

where $\Sigma \equiv \boldsymbol{\Sigma}$ and $\Sigma' \equiv \boldsymbol{\Sigma}'$ are scalars in the current random intercept special cases. The following results are key:

$$\boldsymbol{V}_{\text{cross}} \boldsymbol{1} = \lambda_{\text{cross}} \boldsymbol{1} \quad \text{and} \quad \boldsymbol{V}_{\text{nest}} \boldsymbol{1} = \lambda_{\text{nest}} \boldsymbol{1}, \tag{8}$$

where $\boldsymbol{1}$ denotes a vector of ones with appropriate size,

$$\lambda_{\text{cross}} \equiv \Sigma m'n + \Sigma' mn + \sigma^2 \quad \text{and} \quad \lambda_{\text{nest}} \equiv n\Sigma + \sigma^2. \tag{9}$$

The fact that $\boldsymbol{1}$ is an eigenvector of both $\boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V}_{\text{nest}}$ leads to the fixed effects estimators having simpler and similar forms. A key step involves the inverse eigenvalue results

$$\boldsymbol{V}_{\text{cross}}^{-1} \boldsymbol{1} = (1/\lambda_{\text{cross}}) \boldsymbol{1} \quad \text{and} \quad \boldsymbol{V}_{\text{nest}}^{-1} \boldsymbol{1} = (1/\lambda_{\text{nest}}) \boldsymbol{1}.$$

We then obtain

$$\hat{\beta}_0 = (\boldsymbol{1}^\top \boldsymbol{V}^{-1} \boldsymbol{1})^{-1} \boldsymbol{1}^\top \boldsymbol{V}^{-1} \boldsymbol{Y} = (\boldsymbol{1}^\top \boldsymbol{1})^{-1} \boldsymbol{1}^\top \boldsymbol{Y} = \text{average of the response data}$$

for both $\boldsymbol{V} = \boldsymbol{V}_{\text{cross}}$ and $\boldsymbol{V} = \boldsymbol{V}_{\text{nest}}$. We also have

$$\text{Var}(\hat{\beta}_0) = \frac{\lambda}{\text{total sample size}}, \tag{10}$$

where $\lambda = \lambda_{\text{cross}}$ in the crossed case and $\lambda = \lambda_{\text{nest}}$ in the nested case. Results (9) and (10) then lead to the exact expressions

$$\text{Var}(\hat{\beta}_0) = \begin{cases} \dfrac{\Sigma}{m} + \dfrac{\Sigma'}{m'} + \dfrac{\sigma^2}{mm'n}, & \text{in the crossed case,} \\ \dfrac{\Sigma}{m} + \dfrac{\sigma^2}{mn}, & \text{in the nested case,} \end{cases}$$

which are in keeping with the leading term expression in (1) and the analogous result in Jiang et al. (2022).

In this subsection, we have seen that the eigenvalue/eigenvector results given by (8) lead to the fixed effects estimator reducing to ordinary least squares form in both cases. Therefore, the β_0 estimators behave quite similarly despite the ostensible differences between the crossed and nested cases.

4.2. Heuristics for the general X crossed case

We commence by noting the following exact result:

$$\boldsymbol{V}_{\text{cross}}\boldsymbol{X} = \left[\operatorname*{stack}_{1\le i\le m}\left[\left\{\operatorname*{stack}_{1\le i'\le m'}(\boldsymbol{X}_{\text{A}ii'})\right\}\left(\boldsymbol{\Sigma}\sum_{i'=1}^{m'}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{A}ii'} + \boldsymbol{\Sigma}'\sum_{i=1}^{m}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{A}ii'}\right)\right]\right.$$
$$\left.\times \operatorname*{stack}_{1\le i\le m}\left[\left\{\operatorname*{stack}_{1\le i'\le m'}(\boldsymbol{X}_{\text{A}ii'})\right\}\left(\boldsymbol{\Sigma}\sum_{i'=1}^{m'}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{B}ii'} + \boldsymbol{\Sigma}'\sum_{i=1}^{m}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{B}ii'}\right)\right]\right] + \sigma^2\boldsymbol{X}.$$

Then results such as

$$\frac{1}{mn}\sum_{i=1}^{m}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{A}ii'} \xrightarrow{P} E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{A}\circ}^{\top})$$

and

$$\frac{1}{mn}\sum_{i=1}^{m}\boldsymbol{X}_{\text{A}ii'}^{\top}\boldsymbol{X}_{\text{B}ii'} \xrightarrow{P} E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{B}\circ}^{\top})$$

for all $1 \le i' \le m'$ lead to the approximation

$$\boldsymbol{V}_{\text{cross}}\boldsymbol{X} \approx \boldsymbol{X}\boldsymbol{\Lambda}_{\text{cross}}$$

where

$$\boldsymbol{\Lambda}_{\text{cross}} \equiv \begin{bmatrix} n(m'\boldsymbol{\Sigma} + m\boldsymbol{\Sigma}')E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{A}\circ}^{\top}) + \sigma^2\boldsymbol{I}_{d_{\text{A}}} & n(m'\boldsymbol{\Sigma} + m\boldsymbol{\Sigma}')E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{B}\circ}^{\top}) \\ \boldsymbol{O} & \sigma^2\boldsymbol{I}_{d_{\text{B}}} \end{bmatrix}. \quad (11)$$

We then have

$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\Lambda}_{\text{cross}}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{\Lambda}_{\text{cross}}^{-\top}\boldsymbol{X}^{\top}\boldsymbol{Y} \quad \text{and} \quad \text{Cov}(\hat{\boldsymbol{\beta}}|\mathcal{X}) \approx \boldsymbol{\Lambda}_{\text{cross}}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}. \quad (12)$$

A simple consequence of (11) and (12) is

$$\begin{aligned}\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{B}}\,|\,\mathcal{X}) &\approx \sigma^2\left\{\text{lower right } d_{\text{B}} \times d_{\text{B}} \text{ block of } (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\right\}\\ &\approx \left(\frac{\sigma^2}{mm'n}\right)\left[\text{lower right } d_{\text{B}} \times d_{\text{B}} \text{ block of } \{E(\boldsymbol{X}_{\circ}\boldsymbol{X}_{\circ}^{\top})\}^{-1}\right].\\ &= \left(\frac{\sigma^2}{\text{total sample size}}\right)\left[\text{lower right } d_{\text{B}} \times d_{\text{B}} \text{ block of } \{E(\boldsymbol{X}_{\circ}\boldsymbol{X}_{\circ}^{\top})\}^{-1}\right]. \quad (13)\end{aligned}$$

The asymptotic covariance matrix of $\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{A}}\,|\,\mathcal{X})$ has a similar derivation based on (11) and (12).

4.3. Heuristics for the general X nested case

For the nested model (6) we have the exact expression

$$V_{\text{nest}}\boldsymbol{X} = \boldsymbol{X}_{\text{A}}\boldsymbol{\Sigma}\left(\underset{1\le i\le m}{\text{stack}}\left[\boldsymbol{X}_{\text{A}_i}^\top\boldsymbol{X}_{\text{A}_i} \quad \boldsymbol{X}_{\text{A}_i}^\top\boldsymbol{X}_{\text{B}_i}\right]\right) + \sigma^2\boldsymbol{X}.$$

As $n \to \infty$ and for each $1 \le i \le m$ we have

$$\frac{1}{n}\boldsymbol{X}_{\text{A}_i}^\top\boldsymbol{X}_{\text{A}_i} \xrightarrow{P} E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{A}\circ}^\top)$$

and

$$\frac{1}{n}\boldsymbol{X}_{\text{A}_i}^\top\boldsymbol{X}_{\text{B}_i} \xrightarrow{P} E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{B}\circ}^\top) \quad \text{as } n \to \infty.$$

Therefore,

$$\boldsymbol{V}_{\text{nest}}\boldsymbol{X} \approx \boldsymbol{X}\boldsymbol{\Lambda}_{\text{nest}},$$

where

$$\boldsymbol{\Lambda}_{\text{nest}} \equiv \begin{bmatrix} n\boldsymbol{\Sigma}E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{A}\circ}^\top) + \sigma^2\boldsymbol{I}_{d_{\text{A}}} & n\boldsymbol{\Sigma}E(\boldsymbol{X}_{\text{A}\circ}\boldsymbol{X}_{\text{B}\circ}^\top) \\ \boldsymbol{O} & \sigma^2\boldsymbol{I}_{d_{\text{B}}} \end{bmatrix}$$

which then leads to

$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\Lambda}_{\text{nest}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{\Lambda}_{\text{nest}}^{-\top}\boldsymbol{X}^\top\boldsymbol{Y} \quad \text{and} \quad \text{Cov}(\hat{\boldsymbol{\beta}}|\mathcal{X}) \approx \boldsymbol{\Lambda}_{\text{nest}}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}.$$

The bottom d_{B} rows of $\boldsymbol{A}_{\text{nest}}$ have the same simple form as $\boldsymbol{\Lambda}_{\text{cross}}$ and we obtain

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{\text{B}}\,|\,\mathcal{X}) \approx \left(\frac{\sigma^2}{\text{total sample size}}\right)\left[\text{lower right } d_{\text{B}} \times d_{\text{B}} \text{ block of } \left\{E(\boldsymbol{X}_\circ\boldsymbol{X}_\circ^\top)\right\}^{-1}\right]$$

which matches (13) and, indeed, the asymptotic covariance matrix form that arises in ordinary multiple regression.

4.4. Closing discussion on the asymptotic similarities

In this section we have provided heuristic justifications for the fixed effects estimators and their asymptotic covariance matrices having the approximate forms

$$\hat{\boldsymbol{\beta}} \approx \boldsymbol{\Lambda}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{\Lambda}^{-\top}\boldsymbol{X}^\top\boldsymbol{Y} \quad \text{and} \quad \text{Cov}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Lambda}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} \tag{14}$$

for *both* the crossed random effects model (1) and the nested model (6). The common approximate forms in (14) provide a reasonable explanation for the asymptotic covariance matrices in Result 3.1 having forms similar to the nested case.

The approximate $\hat{\boldsymbol{\beta}}$ expression in (14) is intriguingly close to the well-known ordinary least squares expression. In the special case of $\boldsymbol{X}$ being a column vector, $\boldsymbol{\Lambda}$ is scalar and cancels to give the ordinary least squares form. Such reduction occurred in Section 4.1 for the $\boldsymbol{X} = \mathbf{1}$ case. However, there is no such cancellation in general.

The heuristics in the general $\boldsymbol{X}$ cases involve approximations having generic form

$$\boldsymbol{V}\boldsymbol{X} \approx \boldsymbol{X}\boldsymbol{\Lambda}. \tag{15}$$

In the special case where $\boldsymbol{X} = \boldsymbol{x}$ is a column vector and $\boldsymbol{\Lambda} = \lambda$ is scalar then (15) becomes $\boldsymbol{V}\boldsymbol{x} \approx \boldsymbol{x}\lambda$ which corresponds, approximately, to λ being an eigenvalue of $\boldsymbol{V}$ with eigenvector $\boldsymbol{x}$. For general $\boldsymbol{X}$ and $\boldsymbol{\Lambda}$, (15), with '=' instead of '≈', is an instance of *Sylvester's equation* (e.g. Stewart & Sun, 1990; Chapter V, Section 1.2).

5. Statistical utility

Result 3.1 provides a great deal of statistical utility concerning inference and design. Confidence intervals and Wald hypothesis tests based on studentization are immediate consequences. Another is sample size calculations, for which we provide some details in this section.

For illustration of sample size calculations arising from Result 3.1, consider the following special case of (1):

$$\begin{aligned} &Y_{ii'j} \mid B_{ii'j}, X_{ii'j}, U_i, U'_{i'} \overset{\text{ind.}}{\sim} N(\beta_0^0 + U_i + U'_{i'} + \beta_1^0 B_{ii'j} + \beta_2^0 X_{ii'j} + \beta_3^0 B_{ii'j} X_{ii'j}, \sigma^2),\\ &U_i \overset{\text{ind.}}{\sim} N(0, \Sigma^0), \quad U'_{i'} \overset{\text{ind.}}{\sim} N(0, (\Sigma')^0), \quad 1 \le i \le m,\ 1 \le i' \le m',\ 1 \le j \le n, \end{aligned} \tag{16}$$

where $B_{ii'j} \overset{\text{ind.}}{\sim}$ Bernoulli(p) and $X_{ii'j}$ being independently and identically distributed the same as a general random variable $X_\circ$ having finite second moment. Consider the one-sided hypotheses

$$H_0 : \beta_3^0 = 0 \quad \text{versus} \quad H_1 : \beta_3^0 > 0 \tag{17}$$

corresponding to a possibly positive interaction effect between the two predictors. Let $\Delta > 0$ be a particular alternative value of β_3^0 and let P be the corresponding power. Then Result 3.1 and standard arguments lead to the following sample size formula:

$$m = \left\lceil \frac{\{\Phi^{-1}(\alpha) + \Phi^{-1}(1-P)\}^2}{(\Delta/\sigma^0)^2 p(1-p)\mathrm{Var}(X_0) m' n} \right\rceil, \tag{18}$$

where, for any $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x and Φ^{-1} is the $N(0,1)$ quantile function.

Now consider a psychological study such that model (16) and hypotheses (17) apply with $m' = 25$ items and $n = 1$ observation per subject-item combination. How many subjects should be recruited to potentially detect the smallest meaningful interaction effect of $\Delta = 0.25$ with power 0.9 from a 0.05 level of significance test? If it is further be assumed that $p = \frac{1}{2}$ and $\mathrm{Var}(X) = \frac{1}{12}$ then from (18) we should recruit $m = 53$ subjects if the error standard deviation is $\sigma^0 = 0.4$. Table 1 below provides the required m values for some other values of σ^0.

In contemporary Gaussian response linear mixed model software, such as the function `lmer()` within the package `lme4` (Bates et al., 2015), standard errors are typically obtained using exact observed Fisher information rather than the approximation to the (expected) Fisher information on which (18) is based. This raises the question as to whether the number of subjects chosen according to the Result 3.1 approximation to the standard error of $\hat{\beta}_3$ leads

Table 1. The results from the illustrative sample size calculation and corresponding empirical power checks for the simulation study described in the text. The number of subjects (*m*) corresponds to an advertized power of 0.9.

Error standard deviation (σ^0):	0.2	0.4	0.8	1.6
Minimum number of subjects (*m*):	14	53	211	842
Empirical estimate of power:	0.889	0.902	0.878	0.885
95% confidence interval of power:	(0.870, 0.908)	(0.884, 0.920)	(0.858, 0.898)	(0.865, 0.905)

to the advertized power for exact Fisher information-based hypothesis tests. We addressed this question by running a simulation study that involved replication of 1000 simulated data sets corresponding to (16) with various noise levels according to $\sigma^0 \in \{0.2, 0.4, 0.8, 1.6\}$. The $B_{ii'j}$ and $X_{ii'j}$ data were generated from Bernoulli($\frac{1}{2}$) and Uniform$(0, 1)$ distributions, respectively. As above, we set $(m', n, \Delta, \alpha, P) = (20, 1, 0.25, 0.05, 0.9)$ and determined m using (18). For each simulated data set we carried out a test of (17) using calls to `lmer()`, with rejection of H_0 if the t-statistic corresponding to β_3^0 exceeded $\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95)$. Table 1 shows the empirical estimates of $P = 0.9$ and corresponding 95% confidence intervals. For this example we see that the sample size formula (18) performs well with regards to the actual power delivered.

The example in this section demonstrates the statistical utility of Result 3.1. We are not aware of previous results in the literature for linear mixed models with crossed random effects that readily provide the sample size formula (18).

6. Concluding remarks

Result 3.1 provides the precise leading term behaviours of the maximum likelihood estimators for a general class of linear mixed models containing crossed random effects and enables statistical utilities such as Wald tests for all model parameters and sample size calculations. It complements the recent contributions of Lyu et al. (2024) via extensions to random slopes and unbalanced designs. In comparison with the nested random effects situation, the establishment of leading term results in the presence of crossed random effects is lengthy and arduous – even when the responses are Gaussian. The leading terms have similar or identical forms to those arising in nested models, and we have provided some heuristic arguments for this phenomenon. We conjecture that the two-term asymptotic covariance matrices for $\hat{\boldsymbol{\beta}}_{\mathrm{A}}$, $\hat{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}'$ in the Section 2 set-up are similar or identical to those appearing in Section 3.3.1 of Maestrini et al. (2024) for the nested case, but such an investigation would require a great deal of additional efforts. Lastly, there are questions of what precise asymptotic results, if any, could be obtained for non-Gaussian and sparse data versions of linear mixed models containing crossed random effects. The current article may pave the way for such future endeavours.

Acknowledgments

We are grateful to two reviewers for their improvement suggestions. We also thank John Duchi, Iain Johnstone, Ziyang Lyu, Luca Maestrini, Art Owen, Nicola Sartori and Alan Welsh for advice related to this research. The second author is grateful for hospitality from the Department of Statistical Sciences, University of Padua, Italy, during part of this research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the U.S. National Science Foundation [grant number DMS-2210569] and the Australian Research Council [grant number DP230101179].

References

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Gao, K., & Owen, A. B. (2020). Estimation and inference for very large linear mixed effects models. *Statistica Sinica*, *30*(4), 1741–1771.

Ghosh, S., Hastie, T., & Owen, A. B. (2022). Backfitting for large scale crossed random effects regressions. *The Annals of Statistics*, *50*(1), 560–583. https://doi.org/10.1214/21-AOS2121

Hartley, H. O., & Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, *54*(1–2), 93–108. https://doi.org/10.1093/biomet/54.1-2.93

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics*, *24*(1), 255–286. https://doi.org/10.1214/aos/1033066209

Jiang, J. (2013). The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond. *The Annals of Statistics*, *41*(1), 177–195. https://doi.org/10.1214/13-AOS1084

Jiang, J. (2025). Asymptotic distribution of maximum likelihood estimator in generalized linear mixed models with crossed random effects. *The Annals of Statistics*, *53*(3), 1298–1318.

Jiang, J., Wand, M. P., & Bhaskaran, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *Journal of the Royal Statistical Society, Series B*, *84*(1), 55–82. https://doi.org/10.1111/rssb.12473

Lyu, Z., Sisson, S. A., & Welsh, A. H. (2024). Increasing dimension asymptotics for two-way crossed mixed effect models. *The Annals of Statistics*, *52*(6), 2956–2978. https://doi.org/10.1214/24-AOS2469

Maestrini, L., Bhaskaran, A., & Wand, M. P. (2024). Second term improvement to generalised linear mixed model asymptotics. *Biometrika*, *111*(3), 1077–1084. https://doi.org/10.1093/biomet/asad072

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models* (2nd ed.). Wiley-Interscience.

Menictas, M., Di Credico, G., & Wand, M. P. (2023). Streamlined variational inference for linear mixed models with crossed random effects. *Journal of Computational and Graphical Statistics*, *32*(1), 99–115. https://doi.org/10.1080/10618600.2022.2096622

Miller, J. J. (1973). *Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance* (Tech. Report No. 12). Department of Statistics, Stanford University.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, *5*(4), 746–762. https://doi.org/10.1214/aos/1176343897

Stewart, G. W., & Sun, J. (1990). *Matrix Perturbation Theory*. Academic Press.