Uniform inference in linear mixed models

Karl Oskar Ekvall[⋆,†] Matteo Bottai[†]

[⋆]Department of Statistics, University of Florida

[†]Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet

k.ekvall@ufl.edu matteo.bottai@ki.se

July 25, 2025

Abstract

We provide finite-sample distribution approximations, that are uniform in the parameter, for inference in linear mixed models. Focus is on variances and covariances of random effects in cases where existing theory fails because their covariance matrix is nearly or exactly singular, and hence near or at the boundary of the parameter set. Quantitative bounds on the differences between the standard normal density and those of linear combinations of the score function enable, for example, the assessment of sufficient sample size. The bounds also lead to useful asymptotic theory in settings where both the number of parameters and the number of random effects grow with the sample size. We consider models with independent clusters and ones with a possibly diverging number of crossed random effects, which are notoriously complicated. Simulations indicate the theory leads to practically relevant methods. In particular, the studied confidence regions, which are straightforward to implement, have near-nominal coverage in finite samples even when some random effects have variances near or equal to zero, or correlations near or equal to ± 1.

1 Introduction

Linear mixed models, and random effects in particular, are used routinely to model dependence and effect heterogeneity. However, while random effects are convenient for specifying a model, they often complicate inference. For example, it is well known that, when testing if the variance of a random effect is zero, common test-statistics have nonstandard asymptotic distributions

because the parameter is on the boundary of the parameter set; see for example Self and Liang (1987) and Geyer (1994) and the references therein. In general, the distributions, and hence the appropriate critical values for tests, depend on the particular boundary point, the structure of the parameter set, and the test-statistic. By contrast, at any fixed interior point of the parameter set, test-statistics such as score, Wald, and likelihood ratio all have asymptotic chi-squared distributions under regularity conditions. It is common, therefore, to use chi-squared quantiles as critical values for interior points. Unfortunately, doing so does not lead to reliable inference in general. Confidence regions obtained by inverting the tests often have uniform coverage probabilities quite different from nominal, even asymptotically. The regions are often overly conservative, but they can also be invalid. To date, the issues have been addressed only in a few special cases of the settings considered here. Importantly, existing results on near-boundary inference preclude many dependence structures common in mixed models.

Let $X \in \mathbb{R}^{n\times p}$ and $Z \in \mathbb{R}^{n\times q}$ be non-stochastic matrices corresponding to a vector $\beta \in \mathbb{R}^p$ of fixed effects and a vector $U \in \mathbb{R}^q$ of random effects, respectively. Suppose U is multivariate normal with mean zero and covariance matrix $\Psi \in \mathbb{R}^{q\times q}$, and that a vector $Y \in \mathbb{R}^n$ of responses satisfies

$$Y = X\beta + ZU + E, \tag{1}$$

where $E \sim \mathrm{N}(0, \psi_r I_n)$ and $U \sim \mathrm{N}(0, \Psi)$, independently, and Ψ is parameterized by the first $r-1$ elements of the vector $\psi = [\psi_1, \ldots, \psi_r]^{\mathrm{T}}$; see Section 3 for details. A main goal is reliable inference on ψ, and in particular confidence regions with good coverage properties. Focus is on parameters near or at the boundary of the parameter set, because those are often of interest in practice, and many common methods fail near the boundary. The boundary includes points where Ψ is singular, which happens, for example, if some random effects have perfect correlation or vanishing variances. Thus, to reliably assess the practical significance of random effects, confidence regions need to have good coverage properties near the boundary.

A key issue is that the distributions of test-statistics in general depend not only on whether a parameter is on the boundary or not, but on its proximity to the boundary (Moran, 1971; Rotnitzky et al., 2000; Stern and Welsh, 2000; Bottai, 2003; Crainiceanu and Ruppert, 2004; Ekvall and Bottai, 2022; Zhang et al., 2025). Figure 1 illustrates this in a special case of (1). Clearly, even though the parameter is interior, the distributions of Wald and likelihood ratio statistics are quite different from chi-squared. Moreover, the distributions are different from each other, as also noted in other recent research (Battey and McCullagh, 2024). By contrast, the figure shows the score statistic is approximately chi-squared distributed. Evidently, it

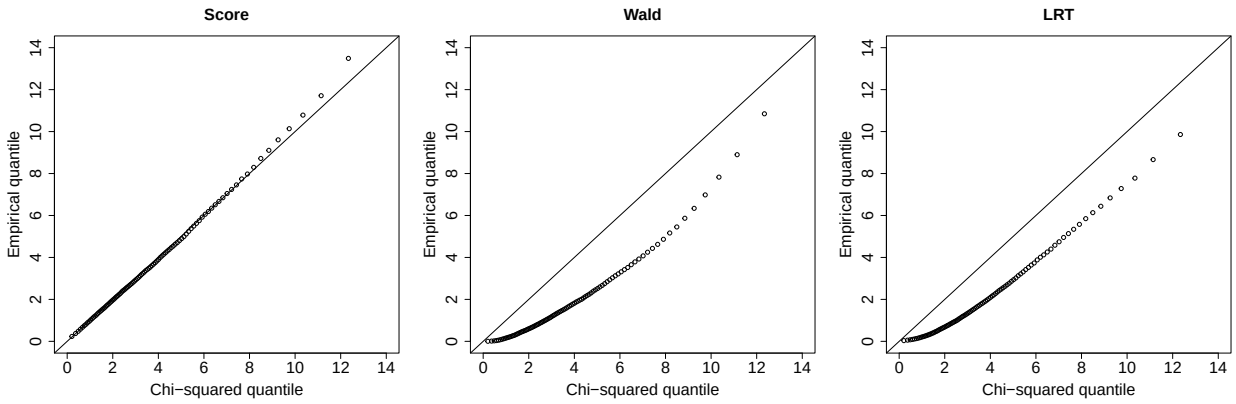

Figure 1: Quantiles of test-statistics evaluated at a true $\psi \in \mathbb{R}^4$ that is near the boundary. For $i \in \{1, \ldots, 50\}$, $Y_i = Z_i U_i + E_i$, where $U_i \sim \mathrm{N}(0, \Psi_1)$, $\Psi_1 \in \mathbb{R}^{2\times 2}$ has diagonal elements $\psi_1 = \psi_3 = 10^{-3}$ and off-diagonal $\psi_2 = 0$, and $E_i \sim \mathrm{N}(0, \psi_4)$, $\psi_4 = 1$. Elements of $Z_i \in \mathbb{R}^{5\times 2}$ were drawn prior to simulations from a Bernoulli distribution with mean $1/2$. Empirical quantiles based on 10 000 replications.

is inappropriate to use the same critical values for all test-statistics, even at interior points. These nuances are not reflected in pointwise asymptotic theory, but they are revealed by uniform results (Section 2).

Suppose for now that $\beta = 0$ is known and let $S(\psi) \in \mathbb{R}^r$ be the score, or the gradient of the log-likelihood at ψ, and $\mathcal{I}(\psi) = \mathrm{cov}_\psi\{S(\psi)\} \in \mathbb{R}^{r\times r}$ the Fisher information matrix. Here and elsewhere, the subscript ψ indicates the covariance is with respect to the distribution indexed by ψ. Let also $W^S(\psi) = \mathcal{I}(\psi)^{-1/2} S(\psi)$ and, for non-zero $v \in \mathbb{R}^r$, let $g(\cdot; v, \psi)$ be the density of $v^{\mathrm{T}} W^S(\psi)$ under ψ; that is, when ψ is the true parameter. Our main finite-sample results in Section 4 upper bound $|g(t; v, \psi) - \phi(t)|$, where $\phi(\cdot)$ is the standard normal density. The bounds are uniform over suitably chosen sets of t, v, and ψ, including ones with boundary points. Upon letting n tend to infinity, possibly along with r and p, the bounds also lead to uniform asymptotic results.

We treat both settings with independent clusters and ones with crossed random effects. The latter are particularly challenging and not handled by existing theory for inference near the boundary. Moreover, to the best of our knowledge, there are no existing finite-sample results similar to ours. In fact, the first results on pointwise asymptotic normality of maximum likelihood estimators of interior parameters of fixed dimension only appeared shortly before submission of our work (Jiang et al., 2024; Lyu et al., 2024). However, while impressive and useful for other purposes, those results are not uniform, and cannot be made so near the boundary. Consequently, they do not lead to reliable confidence regions for ψ in general; see

Section 2 for details and an example illustrating the limitations of pointwise convergence in distribution.

To be sure, we are not the first to consider score-based inference in mixed models (see for example Verbeke and Molenberghs, 2003; Qu et al., 2013; Zhu and Zhang, 2006). However, we are aware of no results similar to ours. Zhang et al. (2025) developed asymptotic theory for score-based confidence regions for a variance component, but only for the special case of (1) where $\Psi = \psi_1 I_q$; that is, with a single variance component. That setting is substantially simpler, partly because there is only one variance component, but also because observations are independent after transforming the data by the left singular vectors of Z. Thus, theory can be developed assuming independence, which is not the case in general under (1). Indeed, dependence is often complicated, especially with crossed random effects, and there can be few independent vectors even as n grows (see for example Sung and Geyer, 2007; Greven et al., 2008). Ekvall and Bottai (2022) provided asymptotic results in a related setting, but with a diverging number of independent observations, a fixed number of parameters, diagonal Ψ, and singular Fisher information, neither of which is true here in general.

2 Background on inference near the boundary

We will often first state results where $\beta = 0$ is known, so that ψ is the only parameter. This lets us focus on the key issues, which arise when making inferences about the covariance parameters, and it simplifies notation. Then we address settings where β is unknown.

Let $T(\psi) = T(\psi; Y, X, Z)$ denote a generic test-statistic, defined for ψ in some parameter set $\mathbb{P}$. Define a $(1-\alpha) \in (0,1)$ confidence region obtained by inverting T by

$$\mathbb{C}(\alpha) = \{\psi \in \mathbb{P} : T(\psi) \leq q_{1-\alpha}(\psi)\}. \tag{2}$$

If the critical value $q_{1-\alpha}(\psi)$ is the $(1-\alpha)$th quantile of the exact distribution of $T(\psi)$, for every $\psi \in \mathbb{P}$, then the confidence region has uniformly correct coverage probability. That is, $\sup_{\psi \in \mathbb{P}} |\mathrm{P}_\psi\{\psi \in \mathbb{C}(\alpha)\} - (1-\alpha)| = 0$. However, because exact distributions are typically unavailable, we instead look for useful approximations.Assume for simplicity, for the remainder of Section 2, that $\mathbb{P}$ does not depend on n. Then we would like to pick $q_{1-\alpha}(\psi)$ such that, for every compact $\mathbb{A} \subseteq \mathbb{P}$,

$$\lim_{n\to\infty} \sup_{\psi \in \mathbb{A}} |\mathrm{P}_\psi\{\psi \in \mathbb{C}_n(\alpha)\} - (1-\alpha)| = 0. \tag{3}$$

The added subscript n indicates the confidence region depends on n; we omit these subscript when stating finite-sample results. When (3) holds, we say $\mathbb{C}_n(\alpha)$ has asymptotically correct coverage probability on compact sets. We will see that the compactness restriction can sometimes be relaxed.

The following lemma, which is not specific to mixed models, gives a useful characterization of (3). The lemma is somewhat similar to, but applies more generally than, Lemma 2.5 in Ekvall and Bottai (2022). Proofs are in the Supplementary Material unless otherwise stated.

Lemma 1. *Equation* (3) *holds for every compact* $\mathbb{A} \subseteq \mathbb{P}$ *if and only if, for every sequence* (ψ_n) *convergent in* $\mathbb{P}$*,*

$$\lim_{n\to\infty} |\mathrm{P}_{\psi_n}\{\psi_n \in \mathbb{C}_n(\alpha)\} - (1-\alpha)| = 0. \tag{4}$$

We will use that (4) implies (3) in our proofs, but the other direction of the equivalence is also important: It tells us that pointwise convergence in distribution cannot be used to establish (3) in general–one must consider sequences of parameters. The next result, which essentially says (4) holds if a test-statistic has the same asymptotic distribution under any convergent sequence of parameters, is also not specific to mixed models; see for example Mikusheva (2007) for a similar result.

Lemma 2. *Suppose that, for a test-statistic* T_n *and continuous cumulative distribution function* F*, it holds for every sequence* (ψ_n) *convergent in* $\mathbb{P}$ *and* $t \in \mathbb{R}$*, that* $F_n(t) = \mathrm{P}_{\psi_n}\{T_n(\psi_n) \leq t\} \to F(t)$ *as* $n \to \infty$*. Then, for any* $\alpha \in (0,1)$*,* $\mathbb{C}_n(\alpha)$ *defined by* (2) *with* $T = T_n$ *and* $q_{1-\alpha}(\psi) = F^{-}(1-\alpha) = \min\{t : F(1-\alpha) \geq t\}$*, satisfies* (4)*.*

In later sections we will verify the conditions of Lemma 2 for score-based confidence regions. Conversely, the conditions of the lemma do not hold for likelihood ratio and Wald statistics in general. To see this it suffices to consider constant sequences with $\psi_n = \psi$, which are trivially convergent. Under such sequences, classical results say likelihood ratio and Wald statistics have different asymptotic distributions depending on whether ψ is an interior or boundary point. This does not imply that those statistics cannot give a confidence region satisfying (3), but Lemma 2 does not apply to them, so $q_{1-\alpha}(\psi)$ would have to depend on ψ in some nontrivial way in general. It is not enough to use different critical values for boundary and interior points, as the following example illustrates. We have made the example as simple as possible while still illustrating key issues to be addressed more generally.

Example 1. Suppose that, independently for $i \in \{1, \ldots, n\}$,

$$Y_i \sim \mathrm{N}(0, 1 + \psi_1). \tag{5}$$

This is a special case of (1) with known $\beta = 0$, $Z = I_n$, $\Psi = \psi_1 I_n$, and $\psi \in \mathbb{P} = [0, \infty) \times \{1\}$. Because $\psi_2 = 1$, we simplify notation and write $\psi = \psi_1 \in [0, \infty)$ for the remainder of the example. Up to a constant, the log-likelihood is $\ell_n(\psi) = -n\{\log(1 + \psi) + M_n/(1 + \psi)\}/2$, where $M_n = \sum_{i=1}^n Y_i^2/n$. The score is $S_n(\psi) = \nabla \ell_n(\psi) = -n\{(1 + \psi)^{-1} - M_n(1 + \psi)^{-2}\}/2$ and the Fisher information is $\mathcal{I}_n(\psi) = \mathrm{var}_\psi\{S_n(\psi)\} = n/\{2(1 + \psi)^2\}$. Define score and Wald statistics by, respectively,

$$W_n^S(\psi) = \mathcal{I}_n(\psi)^{-1/2} S_n(\psi) = (n/2)^{1/2}\{M_n/(1 + \psi) - 1\}$$

and $W_n^W(\psi) = (\hat{\psi}_n - \psi)\mathcal{I}_n(\psi)^{1/2}$, where $\hat{\psi}_n = \max(M_n - 1, 0)$ is the maximum likelihood estimator. The likelihood ratio test-statistic is $T_n^L(\psi) = 2\{\ell_n(\hat{\psi}_n) - \ell_n(\psi)\}$. It is clear that $\hat{\psi}_n$ can be zero with substantial probability. More specifically, since $nM_n/(1+\psi) \sim \chi_n^2$, a normal approximation gives $\mathrm{P}_\psi(\hat{\psi} = 0) \approx \Phi\{-(n/2)^{1/2}\psi/(1 + \psi)\}$, where Φ is the standard normal cumulative distribution function. This probability is $1/2$ if $\psi = 0$, and is close to that if $n^{1/2}\psi$ is small. This suggests a normality approximation for the maximum likelihood estimator may perform poorly near the boundary. Intuitively, then, a chi-squared approximation for the likelihood ratio statistic may also be inappropriate. The following result formalizes this intuition.

Proposition 1. *Suppose* (5) *holds with* $\psi = \psi_n \in [0, \infty)$ *satisfying* $\psi_n n^{1/2} \to a \in [0, \infty]$ *as* $n \to \infty$. *Then, with* $W_1 \sim \mathrm{N}(0, 1)$, *in distribution,*

$$W_n^S(\psi_n) \to W_1, \tag{6}$$

$$W_n^W(\psi_n) \to \max(W_1, -a2^{-1/2}), \tag{7}$$

$$T_n^L(\psi_n) \to 2W_1 \max(W_1, -a2^{-1/2}) - \max(W_1, -a2^{-1/2})^2. \tag{8}$$

Results similar to Proposition 1 are given by Moran (1971) and Stern and Welsh (2000). Clearly, the test-statistics are not asymptotically equivalent under sequences of parameters. Statistics which use the maximum likelihood estimator behave irregularly since there is an appreciable probability the estimator is on the boundary, as is illustrated by (7)–(8).

Proposition 1 implies a confidence region based on $T_n^S(\cdot) = \{W_n^S(\cdot)\}^2$ satisfies (3). Indeed, by (6), Lemma 2 applies with $T_n = T_n^S$ upon taking F to be the cumulative distribution

function of χ^2_1. We will see that this generalizes to more complex settings.

Suppose $a = 0$ in Proposition 1, including $\psi_n = 0$ for every n as a special case. Then we recover familiar asymptotic distributions for inference on boundary points. The asymptotic distribution of $T^L_n(\psi_n)$ simplifies to that of $\max(W_1, 0)^2$, a mixture of χ^2_1 and a point mass at zero. Conversely, if $a = \infty$, which happens for example if $\psi_n = \psi > 0$, we recover the classical result that $T^S_n(\psi)$, $T^W_n(\psi) = \{W^W_n(\psi)\}^2$, and $T^L_n(\psi)$ are all asymptotically χ^2_1.

To see why a likelihood ratio confidence region can be unreliable in our settings, let $\mathbb{C}^L_n(\alpha)$ be defined by (2) with $T = T^L_n$ and, for interior $\psi > 0$, $q_{1-\alpha}(\psi) = c_{1,1-\alpha}$, the $(1-\alpha)$th quantile of χ^2_1. Then, regardless of which critical value is used for the boundary point $\psi = 0$, we get by (8) with $\psi_n = 1/n$,

$$\begin{aligned}\sup_{\psi\in[0,1]} |\mathrm{P}_\psi\{\psi \in \mathbb{C}^L_n(\alpha)\} - (1-\alpha)| &\geq |\mathrm{P}_{1/n}\{T^L_n(1/n) \leq c_{1,1-\alpha}\} - (1-\alpha)| \\ &\to |\mathrm{P}\{\max(W_1, 0)^2 \leq c_{1,1-\alpha}\} - (1-\alpha)|,\end{aligned}$$

which is greater than zero. For example, with $\alpha = 0.05$, $\mathrm{P}\{\max(W_1, 0)^2 \leq c_{1,0.95}\} = \mathrm{P}(W_1 \leq 1.96) = 0.975 \neq 0.95$. Thus, $\mathbb{C}^L_n(\alpha)$ does not satisfy (3). Intuitively, this happens because $T^L_n(\psi)$ behaves as if ψ is on the boundary when it is close enough to it. Similar arguments apply to the Wald statistic. For reasons illustrated by this example, we shall focus on score-statistics.

3 Model and test-statistic

Equation (1) implies

$$Y \sim \mathrm{N}(X\beta, \Sigma), \quad \Sigma = Z\Psi Z^{\mathrm{T}} + \psi_r I_n, \tag{9}$$

where $\Psi \in \mathbb{R}^{q\times q}$ is parameterized by $\psi_{-r} = [\psi_1, \ldots, \psi_{r-1}]^{\mathrm{T}}$. We assume each element of Ψ is either known to be zero or equal to one of the elements of ψ_{-r}. Thus, the elements of ψ_{-r} are variances or covariances of random effects. To simplify notation, we consider $\Psi = \Psi(\psi)$ a function of ψ and let $H_j = \partial\Psi/\partial\psi_j$, $j \in \{1, \ldots, r\}$. The elements of H_j are all zeros and ones, with at least one non-zero element for each $j < r$; $H_r = 0$; and $\Psi(\psi) = \sum_{j=1}^r \psi_j H_j$. The parameter set for ψ is $\mathbb{P} = \{\psi \in \mathbb{R}^r : \Psi(\psi) \geq 0, \psi_r > 0\}$, where $\Psi(\psi) \geq 0$ indicates positive semi-definiteness. For example, in models with a single random effect, $\Psi(\psi) = \psi_1 I_q$, $H_1 = I_q$, and $\mathbb{P} = [0, \infty) \times (0, \infty)$. We often omit the argument to Ψ for simplicity. Similarly,

we write $\Sigma(\psi) = Z\Psi(\psi)Z^{\mathrm{T}} + \psi_r I_n$ when the argument needs to be emphasized, and omit it otherwise.

The log-likelihood for $\theta = (\beta, \psi) \in \mathbb{R}^p \times \mathbb{P}$ corresponding to (9) is, up to a constant,

$$\ell(\theta) = \ell(\theta; Y, X, Z) = -\frac{1}{2}\{\log|\Sigma(\psi)| + (Y - X\beta)^{\mathrm{T}}\Sigma(\psi)^{-1}(Y - X\beta)\}.$$

We abuse notation somewhat and define score functions for θ, β, and ψ, respectively, by

$$S(\theta) = \frac{\partial\ell(\theta)}{\partial\theta} \in \mathbb{R}^{p+r}, \quad S(\beta;\psi) = \frac{\partial\ell(\beta,\psi)}{\partial\beta} \in \mathbb{R}^p, \quad S(\psi;\beta) = \frac{\partial\ell(\beta,\psi)}{\partial\psi} \in \mathbb{R}^r.$$

Similarly, $S(\psi_j) = \partial\ell_n(\beta,\psi)/\partial\psi_j$, $j \in \{1,\ldots,r\}$. Let $A_j = \Sigma^{-1/2}(\partial\Sigma/\partial\psi_j)\Sigma^{-1/2}$, so $A_j = \Sigma^{-1/2}ZH_jZ^{\mathrm{T}}\Sigma^{-1/2}$ for $j \in \{1,\ldots,r-1\}$, and $A_r = \Sigma^{-1}$. Then, with $R = \Sigma^{-1/2}(Y - X\beta)$,

$$S(\psi_j) = \frac{1}{2}\{R^{\mathrm{T}}A_jR - \mathrm{tr}(A_j)\}, \quad j \in \{1,\ldots,r\}, \tag{10}$$

and $S(\beta) = X^{\mathrm{T}}\Sigma^{-1/2}R$. The Fisher information matrix is $\mathcal{I}(\theta) = \mathrm{cov}_\theta\{S(\theta)\}$. Since the scores for the ψ_j are quadratic forms in R and the score for β is linear in R, an expression for $\mathcal{I}(\theta)$ follows from the following routine result. The result does not require normality.

Lemma 3. *Suppose $A_1, A_2 \in \mathbb{R}^{n\times n}$ are symmetric and that $R \in \mathbb{R}^n$ is a random vector with mean zero and identity covariance matrix. Then $\mathsf{E}(R^{\mathrm{T}}A_1R) = \mathrm{tr}(A_1)$. If in addition $\mathsf{E}(R_i^3) = 0$ for all i, then $\mathsf{E}\{(R^{\mathrm{T}}A_1R)R\} = 0$; and if also $\mathsf{E}(R_i^4) = 3$ for all i, then $\mathrm{cov}(R^{\mathrm{T}}A_1R, R^{\mathrm{T}}A_2R) = 2\,\mathrm{tr}(A_1A_2)$.*

Lemma 3 is essentially well known, but a proof is in the Supplementary Material for completeness. The expressions in the proof give $\mathrm{cov}_\theta\{S(\theta)\}$ also in settings where the model is misspecified so that some conditions in Lemma 3 do not hold, for example when Y is not normally distributed. It follows from the lemma that $\mathcal{I}(\theta)$ is block-diagonal with leading $p \times p$ block and trailing $r \times r$ block given by, respectively,

$$\mathcal{I}(\beta;\psi) = X^{\mathrm{T}}\Sigma^{-1}X, \quad \mathcal{I}_{ij}(\psi) = \mathrm{tr}(A_iA_j)/2, \quad i,j \in \{1,\ldots,r\}. \tag{11}$$

As the notation suggests, the information for ψ does not depend on β. To state the next result, let $\Psi(v) = \sum_{j=1}^r v_jH_j$ and $\Sigma(v) = Z\Psi(v)Z^{\mathrm{T}} + v_rI_n$ for any $v \in \mathbb{R}^r$. Both $\Psi(v)$ and $\Sigma(v)$ are symmetric, but they need not be positive semi-definite. Let also $\mathbb{S}^{r-1} = \{v \in \mathbb{R}^r : \|v\| = 1\}$.

Theorem 1. *The matrix $\mathcal{I}(\theta)$ is positive definite if and only if $\mathcal{I}(\beta;\psi)$ and $\mathcal{I}(\psi)$ are. The matrix $\mathcal{I}(\beta;\psi)$ is positive definite if and only if X has full column rank $p \leq n$, and $\mathcal{I}(\psi)$ is positive definite if and only if, for every $v \in \mathbb{S}^{r-1}$,*

$$\|\Sigma(v)\| > 0. \tag{12}$$

The condition involving (12) can only hold if $r \leq n(n+1)/2$, and r is typically substantially smaller than that. It is almost trivial to show that the conditions in Theorem 1 ensure identifiability. However, identifiability does not imply a positive definite information matrix in general (Rothenberg, 1971). Indeed, if (1) were parameterized in terms of a matrix $\Lambda \in \mathbb{R}^{r\times r}$ such that $\Psi = \Lambda\Lambda^{\mathrm{T}}$, which can be done without losing identifiability, then the information matrix would be singular at points where Λ is singular (Ekvall and Bottai, 2022; Guédon et al., 2024); see also Chesher (1984); Cox and Hinkley (2000); Lee and Chesher (1986). By contrast, since the conditions in Theorem 1 do not say anything about θ, here $\mathcal{I}(\theta)$ is either positive definite for every θ or for none. Consequently, here, singular information implies an unidentifiable parameter.

The following corollary is useful in some examples.

Corollary 1. *If Z has full column rank $0 < q < n$, then $\mathcal{I}(\psi)$ is positive definite.*

The condition in Corollary 1 is not necessary. For example, the proof reveals that if $\Psi = \psi_1 I_q$, $\mathcal{I}(\psi)$ is positive definite unless ZZ^{T} is proportional to the identity. Settings with $\Psi(\psi) = \psi_1 I_q$ are fairly common in applications, and were studied by Zhang et al. (2025). More generally, a diagonal Ψ leads to a variance components model where $\Sigma = \sum_{j=1}^{r} \psi_j K_j$, $K_r = I_n$, and K_j positive semi-definite for $j < r$. For that model, Theorem 1 says $\mathcal{I}(\psi)$ is positive definite if and only if $K_1, \ldots, K_r$ are linearly independent.

When $\mathcal{I}(\theta)$ is invertible, define

$$W^S(\theta) = \mathcal{I}(\theta)^{-1/2} S(\theta), \quad T^S(\theta) = \|W^S(\theta)\|^2, \tag{13}$$

where $\|\cdot\|$ is the Euclidean norm. Unlike Wald statistics, for example, T^S is invariant under differentiable reparameterizations with full rank Jacobian. Consequently, our results are not specific to the considered parameterization.

For inference on ψ only it is common to use the restricted likelihood (Patterson and Thompson, 1971). It often gives estimators that are less biased, and less likely to be on the boundary (Stern and Welsh, 2000). Suppose X has column rank $p < n$, and let $V \in \mathbb{R}^{n\times(n-p)}$

be a semi-orthogonal matrix such that $V^{\mathrm{T}}X = 0$. Let also $\tilde{Y} = V^{\mathrm{T}}Y$, $\tilde{Z} = V^{\mathrm{T}}Z$, and $\tilde{\Sigma} = V^{\mathrm{T}}\Sigma V$. Then

$$\tilde{Y} \sim \mathrm{N}(0, \tilde{\Sigma}), \quad \tilde{\Sigma} = \tilde{Z}\Psi\tilde{Z}^{\mathrm{T}} + \psi_r I_{n-p}. \tag{14}$$

The restricted likelihood is the likelihood for (14). Having not specified a particular Z or X, (14) is essentially a special case of (9) with known $\beta = 0$ and sample size $n - p$. Therefore, many results for the usual likelihood apply to the restricted likelihood after minor changes to notation. For this reason, we often first state results assuming (9) with known $\beta = 0$ and then apply these to the restricted likelihood.

We write $\tilde{S}(\psi)$ and $\tilde{\mathcal{I}}(\psi)$ for the restricted score and information matrix, respectively. Using the similarity of (9) and (14), explicit expressions are immediate from (10)–(11). In particular, R and A_j, are replaced by, respectively, $\tilde{R} = \tilde{\Sigma}^{-1/2}\tilde{Y}$ and $\tilde{A}_j = \tilde{\Sigma}^{-1/2}\tilde{Z}H_j\tilde{Z}^{\mathrm{T}}\tilde{\Sigma}^{-1/2}$, if $j < r$, or $\tilde{A}_r = \tilde{\Sigma}^{-1}$. Similarly, $\tilde{W}^S(\psi) = \tilde{\mathcal{I}}(\psi)^{-1/2}\tilde{S}(\psi)$ and $\tilde{T}^S(\psi) = \|\tilde{W}^S(\psi)\|^2$. The Supplementary Material contains additional remarks how inference based on the different likelihoods can be implemented in practice.

4 Approximate distributions

4.1 Finite-sample bounds and asymptotic normality

We now turn to finite-sample distribution approximations for score statistics. Recall $A_j(\psi) = \Sigma(\psi)^{-1/2}\{\partial\Sigma(\psi)/\partial\psi_j\}\Sigma(\psi)^{-1/2}$, $j \in \{1, \ldots, r\}$. For $v \in \mathbb{R}^r$ and $\psi \in \mathbb{P}$, let

$$A(v, \psi) = \sum_{j=1}^{r} v_j A_j(\psi) = \Sigma(\psi)^{-1/2}\Sigma(v)\Sigma(\psi)^{-1/2}.$$

For v and ψ such that $\|A(v, \psi)\|_F > 0$, define also

$$a(v, \psi) = \frac{\|A(v, \psi)\|}{\|A(v, \psi)\|_F} \leq 1,$$

where $\|\cdot\|$ and $\|\cdot\|_F$ are the spectral and Frobenius norms, respectively. Recall, for symmetric matrices, the squared Frobenius norm is the sum of squared eigenvalues, and the spectral norm is the largest absolute eigenvalue. For v and ψ such that $\|A(v, \psi)\|_F = 0$, define $a(v, \psi) = 1$. Note $\|A(v, \psi)\|_F > 0$ if and only if $\|\Sigma(v)\| > 0$ since $\gamma_{\min}\{\Sigma(\psi)\} \geq \psi_r > 0$, where $\gamma_{\min}(\cdot)$ is the smallest eigenvalue. Thus, by Theorem 1 and the discussion following it, $\|A(v, \psi)\|_F > 0$ for all non-zero $v \in \mathbb{R}^r$ if $\mathcal{I}(\psi)$ positive definite and ψ identifiable. Shortly, we

will see the score is close to normal if $a(v, \psi)$ is small for appropriate v and ψ. For intuition, recall that in Example 1, $\Sigma(v) = (1 + v)I_n$, and hence $a(v, \psi) = n^{-1/2}$. In more complicated examples given later there is an interplay between the identifiability of ψ, controlled by $\Sigma(v)$, and the regularity of $\Sigma(\psi)$.

The following result builds on a normal approximation for quadratic forms due to Zhang et al. (2025). Recall $g(\cdot; v, \psi)$ is the density of $v^{\mathrm{T}} W^S(\psi)$ under ψ when (9) holds with known $\beta = 0$, so that $\theta = \psi$, and $\phi(\cdot)$ is the standard normal density.

Lemma 4. *For any $t \in \mathbb{R}$, $\psi \in \mathbb{P}$, and $v \in \mathbb{S}^{r-1}$ such that $\tilde{v} = \mathcal{I}^{-1/2}(\psi)v \neq 0$ and $a(\tilde{v}, \psi)^2 < 1/8$, it holds that*

$$|g(t; v, \psi) - \phi(t)| \leq 0.14 \left(4 + \frac{0.29}{\{1 - 8a(\tilde{v}, \psi)^2\}^2}\right) a(\tilde{v}, \psi). \tag{15}$$

Note the bound decreases approximately linearly if $a(\tilde{v}, \psi) \to 0$. Finding an exact expression for $a(\tilde{v}, \psi)$ in Lemma 4 is complicated in general, especially since $\tilde{v}$ depends on $\mathcal{I}(\psi)$. However, an upper bound can often be established, leading to a bound on the density difference in (15) that is uniform in v.

Lemma 5. *If, for a fixed $\psi \in \mathbb{P}$ and $\bar{a}^2 < 1/8$, it holds for every $v \in \mathbb{S}^{r-1}$ that $a(v, \psi) \leq \bar{a}$, then $|g(t; v, \psi) - \phi(t)| \leq 0.14\{4 + 0.29(1 - 8\bar{a}^2)^{-2}\}\bar{a}$ for every $t \in \mathbb{R}$ and $v \in \mathbb{S}^{r-1}$.*

The lemma allows $\bar{a}$ to depend on ψ, but the uniform results we state later result from finding $\bar{a}$ that work for all ψ in some set of interest. Since $W^S(\psi)$ would be multivariate standard normal under ψ if and only if $g(t; v, \psi) = \phi(t)$ for every $t \in \mathbb{R}$ and $v \in \mathbb{S}^{r-1}$, an interpretation of Lemma 5 is that $W^S(\psi)$ is close to multivariate normal under ψ if $\bar{a}$ is small. By the arguments following (14), Lemma 5 can be applied to the restricted likelihood, which we give examples of shortly.

To state asymptotic results, we suppose that for each $n \in \{1, 2, \dots\}$, a version of (9) holds. Specifically, $X \in \mathbb{R}^{n \times p_n}$, $Z \in \mathbb{R}^{n \times q_n}$, $\beta = \beta_n \in \mathbb{R}^{p_n}$, $\psi = \psi_n \in \mathbb{P} = \mathbb{P}_n = \{\psi \in \mathbb{R}^{r_n} : \Psi_n(\psi) \geq 0, \psi_r > 0\}$, and $\Psi(\cdot) = \Psi_n(\cdot)$ can depend on n in any way consistent with (9). We omit indices on Y, X, and Z for simplicity. Let $a_n(\cdot, \cdot)$ and $W_n^S(\cdot)$ be defined as $a(\cdot, \cdot)$ and $W^S(\cdot)$, respectively, with the dependence on n made explicit.

Theorem 2. *Assume that, for $n \in \{1, 2, \dots\}$, (9) holds with known $\beta = 0$ and $\psi = \psi_n \in \mathbb{P}_n \subseteq \mathbb{R}^{r_n \times r_n}$. If there exists $\bar{a}_n$ with $\sup_{v \in \mathbb{S}^{r_n - 1}} a_n(v, \psi_n) \leq \bar{a}_n$ for every n and $\lim_{n \to \infty} \bar{a}_n = 0$, then, for any $v_n \in \mathbb{S}^{r_n - 1}$, $v_n^{\mathrm{T}} W_n^S(\psi_n) \to \mathrm{N}(0, 1)$. Moreover, if in addition X has full column*

rank for every n, and $\beta = \beta_n \in \mathbb{R}^{p_n \times p_n}$ is unknown, it holds for any $u_n \in \mathbb{S}^{p_n + r_n - 1}$ that $u_n^{\mathrm{T}} W_n^S(\theta_n) \to \mathrm{N}(0,1)$.

Like Lemma 5, Theorem 2 holds if $\bar{a}_n$ depends on ψ_n, and the result with known $\beta = 0$ can be applied to the restricted likelihood. The asymptotic normality can hold even if $r_n \to \infty$, indicating a normality approximation can be useful even if there are many random effect parameters. What sample size is sufficient in practice will depend on the setting, including in particular the dependence between observations. The part of the theorem about asymptotic normality when β_n is unknown is not immediate from Lemma 5. The result is suggested, but not implied, by the fact that the score for β_n is multivariate normal for every n.

We next apply the results of this section in two common settings.

4.2 Independent clusters

Many existing results for mixed models assume a large number of independent clusters. We consider such a setting here, first stating finite-sample results and then asymptotic ones. The first result is a separable bound on $a(v, \psi)$, which does not assume a particular version of (9) but is often useful when there are many independent observations or $\Sigma(\psi)$ is well-conditioned.

Lemma 6. *For any $v \in \mathbb{R}^r$ and $\psi \in \mathbb{P}$ such that $\|\Sigma(v)\|_F \neq 0$,*

$$a(v, \psi) \leq \|\Sigma(\psi)^{-1}\| \|\Sigma(\psi)\| \frac{\|\Sigma(v)\|}{\|\Sigma(v)\|_F}.$$

Lemma 6 lets one work separately with v and ψ to find bounds on $a(v, \psi)$. For example, when there are m independent clusters of bounded size, then the first product in Lemma 6 is often bounded uniformly over certain sets of ψ, while the ratio is of order $m^{-1/2}$, uniformly in $v \in \mathbb{S}^{r-1}$.

To be more specific, suppose, independently for $i \in \{1, \ldots, m\}$, for non-stochastic $X_i \in \mathbb{R}^{n_i \times p}$ and $Z_i \in \mathbb{R}^{n_i \times q_1}$,

$$Y_i = X_i \beta + Z_i U_i + E_i, \tag{16}$$

where $E_i \sim \mathrm{N}(0, \psi_r I_{n_i})$, $U_i \sim \mathrm{N}(0, \Psi_1)$, and $\Psi_1 \in \mathbb{R}^{q_1 \times q_1}$. Now (9) holds for $Y = [Y_1^{\mathrm{T}}, \ldots, Y_m^{\mathrm{T}}]^{\mathrm{T}}$ with $X = [X_1^{\mathrm{T}}, \ldots, X_m^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{n \times p}$, and $Z = \mathrm{bdiag}(Z_1, \ldots, Z_m) \in \mathbb{R}^{n \times q}$, where bdiag evaluates to a block-diagonal matrix with the arguments as diagonal blocks. Note $n = \sum_{i=1}^m n_i = m\bar{n}$ and $q = mq_1$. We get $\Psi = I_m \otimes \Psi_1$, where $\otimes$ is the Kronecker product, and

$$\Sigma = \mathrm{bdiag}(Z_1 \Psi_1 Z_1^{\mathrm{T}}, \ldots, Z_m \Psi_1 Z_m^{\mathrm{T}}) + \psi_r I_n. \tag{17}$$

Let $\psi_{-r} = \text{vech}(\Psi_1) \in \mathbb{R}^{q_1(q_1+1)/2}$, the half-vectorization of Ψ_1 obtained by stacking its lower triangular part. Thus, $r = q_1(q_1+1)/2 + 1$ and $\mathbb{P} = \{\psi \in \mathbb{R}^r : \Psi_1(\psi) \geq 0, \psi_r > 0\}$, where $\Psi_1(\cdot)$ maps $\psi \in \mathbb{R}^r$ to the $q_1 \times q_1$ symmetric matrix whose half-vectorization is ψ_{-r}.

To state a result also for the restricted likelihood, define $\tilde{a}(\cdot,\cdot)$ as $a(\cdot,\cdot)$, replacing the A_j with the $\tilde{A}_j$ defined following (14). Let also $c_1, \ldots, c_4$ denote arbitrary constants that do not depend on any model parameters or quantities.

Theorem 3. *Assume* (16) *with (i)* $\bar{n}/q_1 \geq c_1 > 1$ *and (ii), for every* $i \in \{1, \ldots, m\}$,

$$c_2^{-1} \leq \gamma_{\min}(Z_i^{\mathrm{T}} Z_i) \leq \gamma_{\max}(Z_i^{\mathrm{T}} Z_i) \leq c_2, \quad c_2 \in (1, \infty).$$

Then for a $c_3 < \infty$,

$$a(v, \psi) \leq c_3 m^{-1/2}(1 + \|\psi_{-r}\|/\psi_r). \tag{18}$$

Moreover, if in addition (iii) $p/m \leq c_4$ *for a small enough* $c_4 > 0$, *then* (18) *holds with* $\tilde{a}(\cdot,\cdot)$ *in place of* $a(\cdot,\cdot)$, *with a different* c_3.

Explicit expressions for c_3 and c_4, as functions of c_1 and c_2, are available in the proof, but they are somewhat unwieldy. In the special case that $Z_i = Z_1$ for all i, we can take $c_3 = 2^{1/2} c_2$.

The $m^{-1/2}$ is intuitive in this setting as m is the number of independent clusters while observations can be arbitrarily dependent within clusters. Better bounds are possible by restricting within-cluster dependence, but the details will depend on the dependence structure.

That ψ appears only through $\|\psi_{-r}\|/\psi_r$ indicates, by Lemma 5, that a normality approximation for $W^S(\psi)$ is useful as long as random effect variances are small in comparison to $m^{1/2}$ times the error variance. In particular, it is not a problem if $\|\psi_{-r}\|$ is small, including as an extreme special case $\psi_{-r} = 0$, which corresponds to testing the existence of any random effects. More generally, testing random effect variances equal to zero or creating confidence regions for small variances is unproblematic using the proposed test-statistic. Similar arguments apply to correlations near unity.

It is common to reparameterize the model as $\Sigma = \sigma^2\{Z\Psi(\tau)Z^{\mathrm{T}} + I_n\}$, where $\sigma^2 = \psi_r$ and $\tau = \psi_{-r}/\psi_r$ (see for example Crainiceanu and Ruppert, 2004; Bates et al., 2005). For that parameterization, Theorem 3 indicates reliable inference is possible even if σ^2 is close to zero, as long as $\|\tau\|$ is bounded.

The condition that the eigenvalues of $Z_i^{\mathrm{T}} Z_i$ be bounded from above can be ensured in practice without changing the model, for example by replacing Z_i by $Z_i/\|Z\|$. The lower bound on the eigenvalues rules out, for example, the possibility that some clusters are affected

only by a strict subset of the random effects, as that would mean some Z_i have columns of zeros. This could be allowed by instead restricting the proportion of clusters effected by each random effect to be bounded away from zero. However, doing so in full generality would substantially complicate notation, and we do not believe it would lead to any fundamental insights.

Theorem 3 has the following asymptotic result as a corollary. We omit the proof because it is almost immediate from Theorems 2 and 3. Recall the notation of Theorem 2, but let us here index by m instead of n. Note $m \to \infty$ implies $n \to \infty$ since $n = \sum_{i=1}^{m} n_i \geq q_1 m \geq m$.

Corollary 2. *Assume that, for $m \in \{1, 2, \dots\}$, (16) holds with $\beta = \beta_m \in \mathbb{R}^{p_m}$ and $\psi = \psi_m \in \mathbb{P}_m \subseteq \mathbb{R}^{r_m \times r_m}$. If (i) and (ii) of Theorem 3 hold for every m, with c_1, c_2 not depending on m; (iv) $X \in \mathbb{R}^{n \times p}$ has full column rank for every m; and (v) $\|\psi_m\|/\psi_{mr} = o(m^{1/2})$ as $m \to \infty$; then it holds for any $u_m \in \mathbb{S}^{r_m + p_m - 1}$ that $u_m^{\mathrm{T}} W_m^S(\theta_m) \to \mathrm{N}(0, 1)$. Moreover, if in addition (iii') $p_m/m \to 0$, then for any $v_m \in \mathbb{S}^{r_m - 1}$, $v_m^{\mathrm{T}} \tilde{W}_m^S(\psi_m) \to \mathrm{N}(0, 1)$.*

Quantities in the definition of (16) that do not appear explicitly in Corollary 2 can depend on m in any way consistent with (16) and the assumptions of the corollary. For example, q_1 can grow with m as long as (i) holds for every m.

In settings where the n_i grow, which is not necessary for Corollary 2, it may be more natural to state bounds for $Z_i^{\mathrm{T}} Z_i / n_i$, but as noted above, Z_i can be standardized to ensure the upper bound holds. Then, if $n_i \to \infty$, the bounds effectively say the standardized $Z_i^{\mathrm{T}} Z_i$ should be well-conditioned, for example by tending some positive definite limit. Such a condition often holds if the Z_i are drawn randomly, but it rules out, for example, within-cluster crossed random effects.

If $r_m = r$ is fixed, Corollary 2, through Lemma 2, implies asymptotically correct uniform coverage probability on compact sets; the key observation is that the conditions of Corollary 2 are compatible with any sequence (ψ_m) convergent in $\mathbb{P}$. In this setting, r being fixed implies q_1 is also fixed.

We state a result for a restricted score confidence region, which does not require p_m to be fixed, but a similar result holds for the score confidence region for θ with unknown β if $p_m = p$ is also fixed. Recall that $c_{r,1-\alpha}$ denotes the $(1-\alpha)$th quantile of χ^2_r, and define $\tilde{\mathbb{C}}_n^S(\alpha)$ by (2) with $T(\psi) = \tilde{T}_n^S(\psi)$ and $q_{1-\alpha}(\psi) = c_{r,1-\alpha}$.

Corollary 3. *Assume conditions (i), (ii), (iii') and (iv) of Corollary 2. Then if $r_m = r$ is fixed as $m \to \infty$, $\tilde{\mathbb{C}}_n^S(\alpha)$ satisfies (3) for any $\alpha \in (0, 1)$*

4.3 Crossed random effects

With crossed random effects it is impossible to split the data into several independent vectors, and Lemma 6 is typically not useful because $\Sigma(\psi)$ is not well-conditioned. Specifically, even for fixed ψ, some eigenvalues usually grow without bound as $n \to \infty$ while others are bounded. Thus, more intricate arguments are needed to get relevant asymptotic theory.

To understand the setting, suppose there is one mean parameter $\beta \in \mathbb{R}$, and only two random effects, which is common (Jiang, 2013; Ekvall and Jones, 2020; Jiang et al., 2024; Lyu et al., 2024). Specifically, for $i \in \{1, \dots, n_1\}$ and $j \in \{1, \dots, n_2\}$,

$$Y_{ij} = \beta + U_{1i} + U_{2j} + E_{ij},$$

where, independently, $U_{i1} \sim \mathrm{N}(0, \psi_1)$ and $U_{2j} \sim \mathrm{N}(0, \psi_2)$, and $E_{ij} \sim \mathrm{N}(0, \psi_3)$. Let $Y = [Y_{11}, Y_{12}, \dots, Y_{n_1 n_2}]^{\mathrm{T}}$, $Z^{(1)} = I_{n_1} \otimes 1_{n_2} \in \mathbb{R}^{n \times n_1}$, $Z^{(2)} = 1_{n_1} \otimes I_{n_2} \in \mathbb{R}^{n \times n_2}$, and $Z = [Z^{(1)}, Z^{(2)}]$, where $1_{n_j} \in \mathbb{R}^{n_j}$ is a vector of ones. Thus, $Z^{(1)}$ and $Z^{(2)}$ correspond to the first and second random effect, respectively. Now, with $U = [U_{11}, U_{12}, \dots, U_{n_1 n_2}]^{\mathrm{T}} \in \mathbb{R}^{n_1 + n_2}$ and $\Psi = \mathrm{bdiag}(\psi_1 I_{n_1}, \psi_2 I_{n_2})$,

$$\Sigma = Z \Psi Z^{\mathrm{T}} + \psi_3 I_n = \psi_1 Z^{(1)} Z^{(1)\mathrm{T}} + \psi_2 Z^{(2)} Z^{(2)\mathrm{T}} + \psi_3 I_n,$$

where $Z^{(1)} Z^{(1)\mathrm{T}} = I_{n_1} \otimes 1_{n_2} 1_{n_2}^{\mathrm{T}}$ and $Z^{(2)} Z^{(2)^{\mathrm{T}}} = 1_{n_1} 1_{n_1}^{\mathrm{T}} \otimes I_{n_2}$. One reason crossed random effects are complicated is that Σ is not block-diagonal, and cannot be made so by reordering observations. However, a key observation for the subsequent analysis is that Σ is a linear combination of three projection matrices, neither of which depends on ψ. Thus, the eigenvalues of Σ depend on ψ, but the eigenvectors do not. Specifically, since $P_j = 1_{n_j} 1_{n_j}^{\mathrm{T}} / n_j$, $j \in \{1, 2\}$, is a projection matrix, so are $\mathcal{P}_1 = Z^{(1)} Z^{(1)\mathrm{T}} / n_2 = I_{n_1} \otimes P_2$ and $\mathcal{P}_2 = Z^{(2)} Z^{(2)\mathrm{T}} / n_2 = P_1 \otimes I_{n_2}$. With these definitions, $\Sigma = \psi_1 n_2 \mathcal{P}_1 + \psi_2 n_1 \mathcal{P}_2 + \psi_3 I_n$.

Similar arguments apply when there are $r - 1 \geq 2$ crossed random effects. Let $Z^{(j)} = 1_{n_1} \otimes \cdots \otimes 1_{n_{j-1}} \otimes I_{n_j} \otimes 1_{n_{j+1}} \otimes \cdots \otimes 1_{n_{r-1}}$, $j \in \{1, \dots, r-1\}$, and $Z = [Z^{(1)}, \dots, Z^{(r-1)}]$. Define $\mathcal{P}_j = P_1 \otimes \cdots \otimes P_{j-1} \otimes I_{n_j} \otimes P_{j+1} \otimes \cdots \otimes P_{r-1}$, $j \in \{1, \dots, r-1\}$. Then with $\Psi = \mathrm{bdiag}(\psi_1 I_{n_1}, \dots, \psi_{r-1} I_{r-1})$,

$$Y \sim \mathrm{N}(1_n \beta, \Sigma), \quad \Sigma = \sum_{j=1}^{r-1} \psi_j n_{(j)} \mathcal{P}_j + \psi_r I_n, \tag{19}$$

where $n_{(j)} = \prod_{i \neq j} n_i$ and $n = \prod_{i=1}^{r-1} n_i$. To state the main result for crossed random effects,

let also $n_{\min} = \min_{1 \le j \le r-1} n_j$ and $\tilde{n} = n - 1 - \sum_{j=1}^{r-1}(n_j - 1)$.

Theorem 4. *Assume* (19). *If* $r \ge 3$ *and* $n_{\min} \ge 2$, *then*

$$a(v, \psi)^2 \le \sum_{j=1}^{r-1} \frac{1}{n_j - 1} + \frac{(r-2)^2}{\tilde{n}} \le \frac{r-1}{n_{\min} - 1} + \frac{(r-2)^2}{\tilde{n}}.$$

Moreover, if also $n_{\min} \ge 3$, *then*

$$\tilde{a}(v, \psi)^2 \le \max\left\{(n_{\min} - 2)^{-1}, (\tilde{n} - 1)^{-1}\right\}.$$

The first part of the proof of Theorem 4 consists of finding a spectral decomposition of Σ (Supplementary Material). We provide a direct argument but note that the results of Henderson and Searle (1981), which were used by Lyu et al. (2024), could also be used for that particular part of the proof. We prefer the direct argument because it is short and sets up the rest of the proof.

The bounds in Theorem 4 are remarkably simple. That they do not depend on ψ indicates reliable inference is possible uniformly over the parameter set. It is also notable that the bound for the restricted likelihood can be better, even if $\beta = 0$ is known. The technical reasons for this can be seen in the proof. The intuition is that centering can facilitate variance estimation even if the population mean is known to be zero. We only considered a simple mean structure here because our focus is on ψ. We expect similar results are possible for more general means, for example by using the idea of Lyu et al. (2024) to decompose the predictors into parts varying only along certain dimensions.

What is required in general is that $n_{\min}$ is large in comparison to r. When this holds, $\tilde{n}$ will also be large in comparison to r^2 under mild conditions. Thus, the result indicates the complicated dependence induced by crossed random effects decreases the effective number of observations from n to $n_{\min}$. For example, increasing r while keeping $n_{\min}$ fixed will increase n but may worsen the bound.

Corollaries with asymptotic results similar to those in Section 4.2 are straightforward from Theorem 4. For brevity, we only state one for the restricted score confidence region, $\tilde{\mathbb{C}}_n^S(\alpha)$. The result is remarkable because the uniformity is over the entire parameter set; this is an effect of the bounds in Theorem 4 not depending on ψ. Because $n = \prod_{j=1}^{r-1} n_j$, which can only take certain integer values, we index by $k \in \{1, 2, \dots\}$. As in previous asymptotic results, model quantities not appearing explicitly in the statement of the corollary can depend on k in any way consistent with the assumptions of the corollary.

Corollary 4. *Assume, for each* $k \in \{1, 2, \dots\}$, (19) *holds with an* $n_{\min} = n^k_{min}$ *such that* $n^k_{\min} \to \infty$ *as* $k \to \infty$, *and fixed* $r \geq 3$. *Then*

$$\sup_{\psi \in [0,\infty)^{r-1} \times (0,\infty)} \left| \mathrm{P}_\psi\{\psi \in \tilde{\mathbb{C}}^S_n(\alpha)\} - (1-\alpha) \right| \to 0.$$

5 Numerical experiments

We investigate fine-sample coverage probabilities for confidence regions for ψ in three different settings. In the first, we generate data from a version of the model with independent clusters and correlated random effects discussed in Section 4.2. For $i \in \{1, \dots, m\}$ and $j \in \{1, 2, 3\}$,

$$Y_{ij} = X_{ij}^{\mathrm{T}}\beta + Z_{ij}^{\mathrm{T}}U_i + E_{ij},$$

where, independently for all i and j, $E_{ij} \sim \mathrm{N}(0, \psi_4)$ and $U_i \sim \mathrm{N}(0, \Psi_1)$ with

$$\Psi_1 = \begin{bmatrix} \psi_1 & \psi_2 \\ \psi_2 & \psi_3 \end{bmatrix}.$$

We fix $\psi_1 = \psi_3 = \psi_4 = 1$ in the simulations and consider different values of ψ_2, which is then the correlation between the random effects. We consider the restricted score test-statistic, $\tilde{T}^S(\psi)$, and the corresponding confidence region, $\tilde{\mathbb{C}}^S(\alpha)$. For comparison, we include a confidence region based on the statistic $T^P(\psi) = S(\psi; \tilde{\beta})^{\mathrm{T}}\mathcal{I}(\psi)^{-1}S(\psi; \tilde{\beta})$, where $\tilde{\beta} = \{X^{\mathrm{T}}\Sigma(\psi)^{-1}X\}^{-1}X^{\mathrm{T}}\Sigma(\psi)^{-1}Y$, which is in effect using the profile score (Supplementary Material). We also include confidence regions based on the restricted likelihood ratio and a Wald statistic using restricted maximum likelihood estimates standardized by the Fisher information at those estimates. We focus on the restricted likelihood because it tends to give more accurate coverage probabilities for all considered methods.

Before starting the simulation, we drew the elements of β from a standard normal distribution, and created the $X_{ij} = [X_{ij1}, \dots, X_{ijp}]^{\mathrm{T}} \in \mathbb{R}^p$ by setting $X_{ij1} = 1$ and drawing the remaining X_{ijk} independently from a uniform distribution on $(-1, 1)$. We set $Z_{ij} = [X_{ij1}, X_{ij2}]^{\mathrm{T}} \in \mathbb{R}^2$, so there is a random intercept and a random slope.

Figure 2 shows coverage probabilities for different values of the random effect correlation, ψ_2 (horizontal axes), for three different choices of m and p. In the left plot, $m = 500$, so $n = 1500$, and $p = 2$. As suggested by our theory, the restricted score regions (RSCR) have approximately correct coverage probability regardless of how close ψ_1 is to the boundary

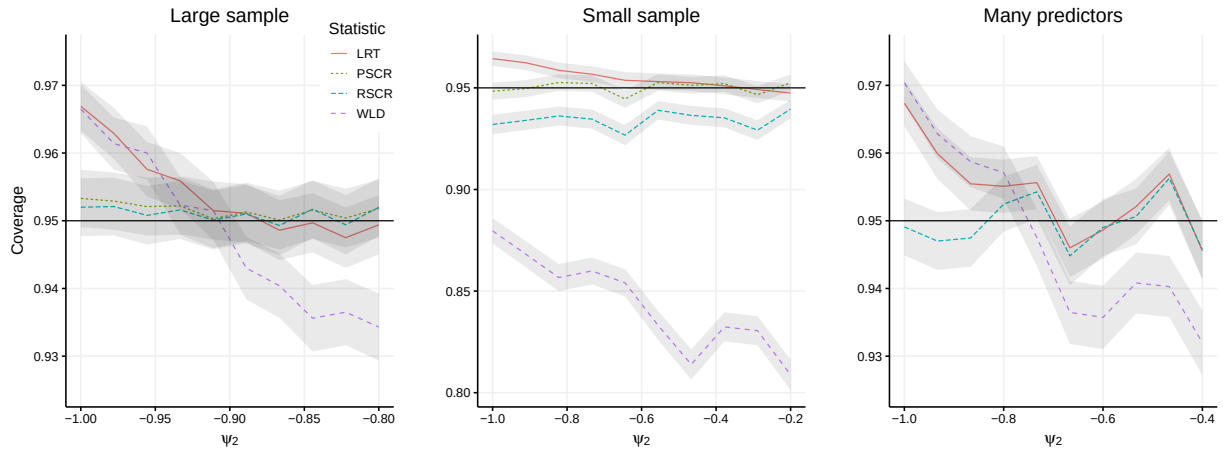

Figure 2: Coverage probabilities with independent clusters and correlated random effects, for different values of the random effect correlation $\psi_2(\psi_1\psi_3)^{-1/2} = \psi_2$ (horizontal axes). The settings are (m, n_i, p) equal to $(500, 3, 2)$ (left plot), $(20, 3, 2)$ (middle plot), and $(200, 3, 100)$ (right plot).

point -1. The regions based on the profile score (PSCR) also have good coverage properties. By contrast, the coverage probabilities for the Wald (WLD) and likelihood ratio (LRT) confidence regions depend substantially on the true ψ_2.

When the sample size is small (middle plot, Figure 2), the Wald regions are particularly unreliable while the other three are relatively similar. The profile score region's coverage is near-nominal for all values of ψ_2 while the restricted score region's is below nominal, around 0.92–0.94 for all values of ψ_2. The likelihood ratio region also has coverage close to nominal in this setting, with some distortion near the boundary. We caution that this should not be interpreted as the likelihood ratio performing well in small samples in general. We do not have a formal result for the likelihood ratio region in small samples, but simulations show its coverage depends on the setting. What is true more generally is that its coverage near the boundary tends to move away from the nominal level as the sample size increases.

Intuitively, there is little difference between restricted and profile score regions when m is large in comparison to p. Conversely, when $p = 100$ is relatively large in comparison to $m = 200$ (right plot, Figure 2), the profile likelihood has coverage probabilities so far from nominal, around 0.4, that we omit it from the plot. In this setting, too, the restricted score gives near-nominal coverage, but the restricted likelihood ratio or Wald statistics do not in general.

The second setting is a version of the crossed random effects model discussed in Section

4.3. Specifically, for $i \in \{1, \ldots, n_1\}$ and $j \in \{1, \ldots, n_2\}$,

$$Y_{ij} = X_{ij}^{\mathrm{T}}\beta + U_{1i} + U_{2j} + E_{ij},$$

with X_{ij}, β, and E_{ij} as in the first setting. The U_{1i} and U_{2j} are independent with variances ψ_1 and ψ_2, respectively. Data were generated with $\psi_1 = \psi_2$ indicated on the horizontal axis in Figure 3. We consider (n_1, n_2, p) equal to $(40, 40, 2)$, so $n = 1600$ is large relative to p; $(10, 10, 2)$, so both $n = 100$ and p are relatively small; and $(20, 20, 80)$, so $n = 400$ and p are both moderately large. Recall, because the random effects are crossed, neither of n, n_1, and n_2 is a number of independent observations. Nevertheless, Theorem 4 suggests $n_{\min} = \min(n_1, n_2)$ controls the quality of the distribution approximations.

When $n_{\min}$ is relatively large (left plot, Figure 3), both score-based confidence region have approximately correct coverage for every value of $\psi_1 = \psi_2$ while the likelihood ratio and Wald regions do not. The Wald regions are particularly unreliable; they are conservative near the boundary and invalid further from the boundary. The difference between the restricted and profile score-based regions is small since p is small.

When $n_{\min}$ is small (middle plot, Figure 3), the profile score region is conservative near the boundary. The restricted score region has coverage slightly below nominal. Nevertheless, both are in general closer to nominal than the likelihood ratio and Wald regions. The likelihood ratio is the most conservative near the boundary, while the Wald region is again conservative near the boundary and invalid further from it.

When the number of predictors is relatively large (right plot, Figure 3), the restricted score-based region has coverage slightly below but near nominal for all values of $\psi_1 = \psi_2$. The likelihood ratio region is conservative near the boundary, but the Wald region is even more so. As in the other settings, the Wald region is invalid further from the boundary.

The Supplementary Material includes additional simulation settings, and the results are qualitatively similar to those presented here.

6 Final remarks

Our results show the score standardized by expected information is often approximately multivariate normally distributed, even with complicated dependence and a large number of parameters. This leads to reliable inference. Such results are impossible for Wald and likelihood ratio statistics when maximum likelihood estimators can be on the boundary with appreciable probability. Nevertheless, asymptotically uniformly correct inference with those

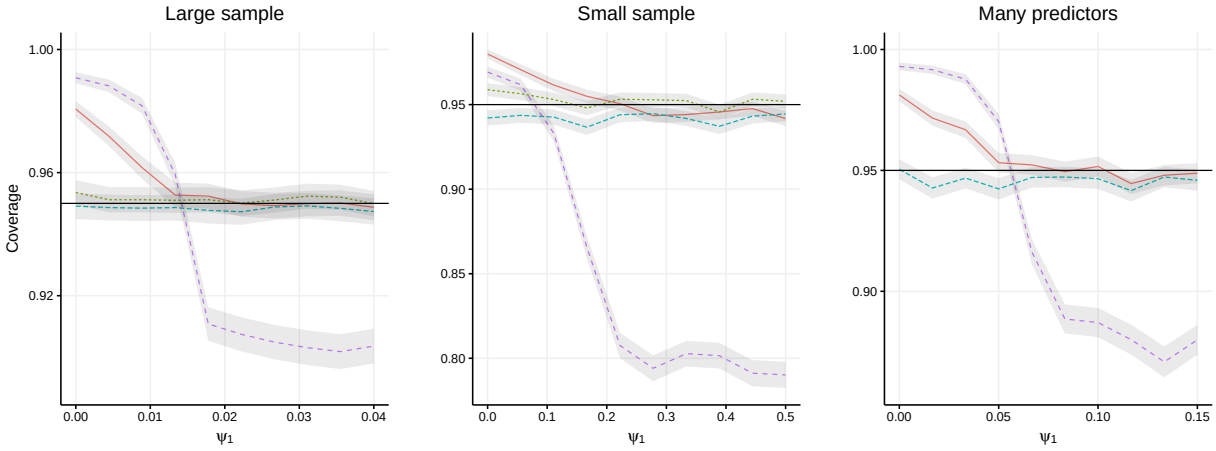

Figure 3: Coverage probabilities with crossed, independent random effects, for different random effect variances $\psi_1 = \psi_2$ (horizontal axes). The settings are (n_1, n_2, p) equal to $(40, 40, 2)$ (left plot), $(10, 10, 2)$ (middle plot), and $(20, 20, 80)$ (right plot).

statistics may be possible even in the settings we consider, but how to achieve that is currently unclear and an avenue for future research.

We expect our results can be extended to inference on other types of covariance matrices in linear models or, more generally, inference using multivariate normal likelihood as a pseudo-likelihood for possibly non-normal data. Non-normally distributed random effects or errors could be allowed by developing a distributional approximation similar to Lemma 5 for non-normal quadratic forms. One would also need to address the fact that the covariance matrix of the score depends on the third and fourth moment of the responses, possibly by using resampling or Monte Carlo-based methods. Similar asymptotic results could likely be obtained without assuming normality, either using Lindeberg's conditions directly on the quadratic forms or by adapting more specialized results (see Jiang, 1996, for example). For settings with crossed random effects, it may be possible to adapt some of the asymptotic results for the score function in Lyu et al. (2024) and Jiang et al. (2024) to boundary settings. However, substantial work may be needed to make them appropriately uniform.

The proposed confidence regions can be computed efficiently using the expressions for the score and information matrix in Section 3. Some additional remarks on implementation are in the Supplementary Material. Code for reproducing the simulation results is at `https://github.com/koekvall/uniform_lmm_suppl/`.

It is in principle possible to extend some of our results to generalized linear mixed models. However, both theory and computation will be substantially more difficult, especially for the

case with crossed random effects. Indeed, in that setting even pointwise asymptotic theory with interior parameters is challenging (Jiang, 2025).

Finally, while we have not focused on the testing on boundary points because that is a well-studied problem, we note tests with asymptotically correct size are immediate from our more general results.

Acknowledgement

The authors are grateful to two referees and an Associate Editor for comments that improved the manuscript substantially. The authors also thank James Hodges and Aaron Molstad for helpful discussions, and Yiqao Zhang and Matias Shedden for assisting with programming.

Supplementary material

Proofs of lemmas and theorems are in the Supplementary Material along with additional numerical results.

References

Bates, D. et al. (2005). Fitting linear mixed models in R. *R news*, 5(1):27–30.

Battey, H. S. and McCullagh, P. (2024). An anomaly arising in the analysis of processes with more than one source of variability. *Biometrika*, 111(2):677–689.

Bottai, M. (2003). Confidence regions when the Fisher information is zero. *Biometrika*, 90(1):73–84.

Chesher, A. (1984). Testing for neglected heterogeneity. *Econometrica*, 52(4):865–872.

Cox, D. R. and Hinkley, D. V. (2000). *Theoretical Statistics*. Chapman & Hall/CRC, Boca Raton.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):165–185.

Ekvall, K. O. and Bottai, M. (2022). Confidence regions near singular information and boundary points with applications to mixed models. *The Annals of Statistics*, 50(3):1806–1832.

Ekvall, K. O. and Jones, G. L. (2020). Consistent maximum likelihood estimation using subsets with applications to multivariate mixed models. *The Annals of Statistics*, 48(2):932–952.

Geyer, C. J. (1994). On the Asymptotics of Constrained M-Estimation. *The Annals of Statistics*, 22(4):1993–2010.

Greven, S., Crainiceanu, C. M., Küchenhoff, H., and Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17(4):870–891.

Guédon, T., Baey, C., and Kuhn, E. (2024). Bootstrap test procedure for variance components in nonlinear mixed effects models in the presence of nuisance parameters and a singular Fisher information matrix. *Biometrika*, 111(4):1331–1348.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60.

Jiang, J. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics*, 24(1).

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).

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., and Ghosh, S. (2024). Precise Asymptotics for Linear Mixed Models with Crossed Random Effects.

Lee, L.-F. and Chesher, A. (1986). Specification testing when score test statistics are identically zero. *Journal of Econometrics*, 31(2):121–149.

Lyu, Z., Sisson, S., and Welsh, A. (2024). Increasing dimension asymptotics for two-way crossed mixed effect models. *The Annals of Statistics*, 52(6):2956–2978.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452.

Moran, P. A. P. (1971). Maximum-likelihood estimation in non-standard conditions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 70(3):441–450.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554.

Qu, L., Guennel, T., and Marshall, S. L. (2013). Linear score tests for variance components in linear mixed models and applications to genetic association studies: Linear score tests for variance components. *Biometrics*, 69(4):883–892.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica*, 39(3):577.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6(2):243–284.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610.

Stern, S. E. and Welsh, A. H. (2000). Likelihood inference for small variance components. *Canadian Journal of Statistics*, 28(3):517–532.

Sung, Y. J. and Geyer, C. J. (2007). Monte Carlo likelihood inference for missing data models. *The Annals of Statistics*, 35(3):990–1011.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, 59(2):254–262.

Zhang, Y., Ekvall, K. O., and Molstad, A. J. (2025). Fast and reliable confidence intervals for a variance component. *Biometrika*, 112(2):asaf010.

Zhu, H. and Zhang, H. (2006). Generalized score test of homogeneity for mixed effects models. *The Annals of Statistics*, 34(3):1545–1569.