# Supplementary Material to "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

## A Proofs

*Proof of Lemma 1.* Define a sequence $(g_n)$ of functions by $g_n(\psi) = \mathrm{P}_\psi\{\psi \in \mathbb{C}_n(\alpha)\}$, and define the constant function $g(\psi) = 1 - \alpha$. By the equivalence of compact and continuous convergence (Remmert, 1991, p.98), $\sup_{\psi \in \mathbb{A}} |g_n(\psi) - g(\psi)| \to 0$ for every compact $\mathbb{A} \subseteq \mathbb{P}$ if and only if, for every convergent sequence $(\psi_n)$ with limit $\psi_0 \in \mathbb{P}$, it holds that $g_n(\psi_n) \to g(\psi_0) = 1 - \alpha$. □

*Proof of Lemma 2.* Pick an arbitrary $\alpha$ and sequence $(\psi_n)$ convergent in $\mathbb{P}$, and observe $\mathrm{P}_{\psi_n}\{\psi_n \in \mathbb{C}_n(\alpha)\} = \mathrm{P}_{\psi_n}\{T_n(\psi_n) \leq F^{-}(1-\alpha)\} = F_n(F^{-}(1-\alpha)) \to F(F^{-}(1-\alpha)) = 1 - \alpha$. The last equality uses that the assumed continuity of $F$ ensures $F(t) = 1 - \alpha$ for some $t \in \mathbb{R}$. Lemma 1 in the main text completes the proof. □

*Proof of Proposition 1.* For every $n$, $nM_n/(1 + \psi_n) \sim \chi^2_n$. Thus, $N_n = M_n/(1 + \psi_n) \to 1$ in probability under $\psi_n$ by the law of large numbers, and $W_n^S(\psi_n) = (n/2)^{1/2}\{M_n/(1+\psi_n)-1\} \to W_1$ in distribution by the central limit theorem, which proves Equation (6) in the main text.

The Wald test-statistic is

$$W_n^W(\psi_n) = \frac{(n/2)^{1/2}(\hat{\psi}_n - \psi_n)}{1 + \psi_n} = \max\left(\frac{(n/2)^{1/2}(M_n - 1)}{1 + \psi_n}, \frac{-(n/2)^{1/2}\psi_n}{1 + \psi_n}\right).$$

The first argument to the maximum is $W_n^S(\psi_n)$, so Equation (6) in the main text, continuity of $(x_1, x_2) \mapsto \max(x_1, x_2)$, and the continuous mapping theorem give Equation (7) in the main text.

The likelihood ratio statistic is

$$\begin{aligned} T_n^L(\psi_n) &= -n \left\{ \log(1+\hat{\psi}_n) - \log(1+\psi_n) + \frac{M_n}{1+\psi_n}\left(\frac{1+\psi_n}{1+\hat{\psi}_n} - 1\right) \right\} \\ &= -n\{\log(L_n) + N_n(1/L_n - 1)\}, \end{aligned}$$

where $L_n = (1+\hat{\psi}_n)/(1+\psi_n)$. Note that $L_n - 1 = (\hat{\psi}_n - \psi_n)/(1+\psi_n) = \max\{N_n - 1, -\psi_n/(1+\psi_n)\}$, which tends to 0 in probability under $\psi_n$ since $N_n \to 1$ in probability and $-\psi_n/(1+\psi_n) \le 0$. Thus, we get in probability, by second order Taylor expansion around 1,

$$\begin{aligned} T_n^L(\psi_n) &= -n\{(L_n-1) - (L_n-1)^2/2 - N_n(L_n-1) + N_n(L_n-1)^2\} + o_p(n|L_n-1|^2) \\ &= n(L_n-1)(N_n-1) - n(N_n - 1/2)(L_n-1)^2 + o_p(n|L_n-1|^2) \\ &= n(L_n-1)(N_n-1) - (n/2)(L_n-1)^2 - n(N_n-1)(L_n-1)^2 + o_p(n|L_n-1|^2). \end{aligned}$$

The first term is

$$n(L_n-1)(N_n-1) = 2\{(n/2)^{1/2}\max(N_n-1, -\psi_n/(1+\psi_n))\}\{(n/2)^{1/2}(N_n-1)\},$$

which, by Equations (6)–(7) in the main text, and the continuous mapping theorem, tends in distribution to $2\max(W_1, -a2^{-1/2})W_1$. By similar arguments, the term $(n/2)(L_n-1)^2$ tends in distribution to $\max(W_1, -a2^{-1/2})^2$, and the remaining terms tend to zero. This establishes Equation (8) in the main text. □

*Proof of Lemma 3.* The first claim is due to $\mathsf{E}(R^{\mathrm{T}}A_1R) = \mathsf{E}\{\mathrm{tr}(RR^{\mathrm{T}}A_1)\} = \mathrm{tr}\{\mathsf{E}(RR^{\mathrm{T}})A_1\} = \mathrm{tr}(A_1)$. The second follows from the fact that, for any $i \in \{1, \dots, n\}$,

$$\mathsf{E}(R_iR^{\mathrm{T}}A_1R) = \sum_{j=1}^n\sum_{k=1}^n \mathsf{E}(R_iR_jR_k[A_1]_{jk}) = [A_1]_{ii}\mathsf{E}(R_i^3) = 0.$$

The last equality uses $\mathsf{E}(R_i^3) = 0$ and the previous one uses $\mathsf{E}(R) = 0$ and independence of the elements of $R$. The final claim is established by writing $\mathsf{E}(R^{\mathrm{T}}A_1RR^{\mathrm{T}}A_2R)$ as

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}[A_1]_{ij}[A_2]_{kl}\mathsf{E}(R_iR_jR_kR_l)\\
&=\sum_{i=1}^{n}[A_1]_{ii}[A_2]_{ii}\mathsf{E}(R_i^4)+\sum_{i=1}^{n}\sum_{k\neq i}[A_1]_{ii}[A_2]_{kk}\mathsf{E}(R_i^2R_k^2)\\
&\quad+\sum_{i=1}^{n}\sum_{j\neq i}[A_1]_{ij}[A_2]_{ij}\mathsf{E}(R_i^2)\mathsf{E}(R_j^2)+\sum_{i=1}^{n}\sum_{j\neq i}[A_1]_{ij}[A_2]_{ji}\mathsf{E}(R_i^2)\mathsf{E}(R_j^2)\\
&=\mathsf{E}(R_i^4)\operatorname{tr}(A_1\circ A_2)+\sum_{i=1}^{n}\sum_{k\neq i}[A_1]_{ii}[A_2]_{kk}+2\sum_{i=1}^{n}\sum_{j\neq i}[A_1]_{ij}[A_2]_{ij}\\
&=\mathsf{E}(R_i^4)\operatorname{tr}(A_1\circ A_2)+\{\operatorname{tr}(A_1)\operatorname{tr}(A_2)-\operatorname{tr}(A_1\circ A_2)\}+2\{\operatorname{tr}(A_1A_2)-\operatorname{tr}(A_1\circ A_2)\}\\
&=\{\mathsf{E}(R_i^4)-3\}\operatorname{tr}(A_1\circ A_2)+2\operatorname{tr}(A_1A_2)+\operatorname{tr}(A_1)\operatorname{tr}(A_2),
\end{aligned}
$$

where $\circ$ denotes elementwise product. □

*Proof of Theorem 1.* A block diagonal matrix is positive definite if and only if all of its diagonal blocks are. Under $\theta$, $R=\Sigma^{-1/2}(Y-X\beta)$ has mean zero and identity covariance matrix, so the covariance matrix of $S(\beta)=X^{\mathrm{T}}\Sigma^{-1}(Y-X\beta)$ is $X^{\mathrm{T}}\Sigma^{-1}X$, which establishes $\mathcal{I}(\beta;\psi)$. Because $\Sigma$ is positive definite for all $\theta$, $X^{\mathrm{T}}\Sigma^{-1}X$ is positive definite if and only if $X$ has full column rank. Next, for any $v\in\mathbb{R}^r$,

$$v^{\mathrm{T}}\mathcal{I}(\psi)v=\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}v_iv_j\operatorname{tr}\{A_i(\psi)A_j(\psi)\}=\frac{1}{2}\operatorname{tr}\left\{\left(\sum_{i=1}^{r}v_iA_i(\psi)\right)^2\right\}=\frac{1}{2}\|A(v,\psi)\|_F^2,$$

where

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

Since $\Sigma(\psi)$ is positive definite for every $\theta$, $\|A(v,\psi)\|_F=0$ if and only if $\Sigma(v)=0$, which completes the proof. □

*Proof of Corollary 1.* It suffices to verify Equation (12) in the main text. Pick an abitrary $v\in\mathbb{S}^{r-1}$. If $v_r\neq 0$, then $Z\Psi(v)Z^{\mathrm{T}}+v_rI_n\neq 0$ since $v_rI_n$ has rank $n$ while $-Z\Psi(v)Z^{\mathrm{T}}$ has rank at most $q<n$, so they cannot be equal.

If instead $v_r=0$, then suppose for contradiction $Z\Psi(v)Z^{\mathrm{T}}=0$. Then left- and right-multiplying that identity by $Z^{\mathrm{T}}$ and $Z$, respectively, gives $Z^{\mathrm{T}}Z\Psi(v)Z^{\mathrm{T}}Z=0$. But $Z^{\mathrm{T}}Z$ is positive definite since $Z$ has full column rank, so left- and right-multiplying the latest identity by $(Z^{\mathrm{T}}Z)^{-1}$ gives $\Psi(v) = 0$, which can only happen if $v_{-r} = 0$. Thus, since also $v_r = 0$, we have $v = 0$, which contradicts $v \in \mathbb{S}^{r-1}$. □

*Proof of Lemma 4.* First,

$$v^{\mathrm{T}} W^S(\psi) = \tilde{v}^{\mathrm{T}} S(\psi) = \frac{1}{2} Y^{\mathrm{T}} \Sigma^{-1/2} A(\tilde{v}, \psi) \Sigma^{-1/2} Y - \frac{1}{2} \operatorname{tr}\{A(\tilde{v}, \psi)\}.$$

Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of $A(\tilde{v}, \psi)$. Then, by orthogonal invariance of the multivariate normal distribution, the right-hand side in the last display has the same distribution as $(1/2)\sum_{i=1}^n \lambda_i (W_i^2 - 1)$, where the $W_i$ are independent standard normal. Theorem 1(a) of Zhang (2005) therefore gives, for any $t \in \mathbb{R}$,

$$|g(t; v, \psi) - \phi(t)| \leq 0.14 \left(4 + \frac{0.29}{(1 - 8\Delta)^2}\right) (d^*)^{-1/2},$$

where $\Delta = \max_{i \in \{1,\dots,n\}} \lambda_i^2 / \sum_{i=1}^n \lambda_i^2 = a(\tilde{v}, \psi)^2$ and

$$d^* = \frac{\{\sum_{i=1}^n \lambda_i^2\}^3}{\{\sum_{i=1}^n |\lambda_i|^3\}^2} \geq \frac{\{\sum_{i=1}^n \lambda_i^2\}^3}{\max_{i \in \{1,\dots,n\}} \lambda_i^2 \{\sum_{i=1}^n \lambda_i^2\}^2} = \frac{\sum_{i=1}^n \lambda_i^2}{\max_{i \in \{1,\dots,n\}} \lambda_i^2} = a(\tilde{v}, \psi)^{-2},$$

which completes the proof. □

*Proof of Lemma 5.* For any $\psi \in \mathbb{P}$, by absolute homogeneity of norms, $a(\tilde{v}, \psi) = a(\tilde{v}/\|\tilde{v}\|, \psi) \leq \bar{a}$ since $(\tilde{v}/\|\tilde{v}\|) \in \mathbb{S}^{r-1}$. The conclusion now follows from the right-hand side of Equation (15) in the main text being increasing in $a$ on $(0, 8^{-1/2})$. □

*Proof of Theorem 2.* The claim about convergence when $\beta_n = 0$ is almost immediate from Lemma 5 in the main text. Indeed, the lemma says the density of $v_n^{\mathrm{T}} W_n^S(\psi_n)$ tends to the standard normal density as $n \to \infty$ since $\bar{a}_n \to 0$. Thus, Scheffé's theorem (Billingsley, 1995, Theorem 16.12) says $v_n^{\mathrm{T}} W_n^S(\psi_n)$ tends to a standard normal in total variation, and hence in distribution.

It remains to show the claim when $\beta_n$ is unknown. Partition $u_n = [w_n, v_n]^{\mathrm{T}}$ with $w_n \in \mathbb{R}^{p_n}$ and let $R_n = \Sigma_n^{-1/2}(Y - X\beta_n)$. Then, since $\mathcal{I}_n(\theta_n)$ is block diagonal with leading block

$\mathcal{I}_n(\beta_n;\psi_n) \in \mathbb{R}^{p_n\times p_n}$ and trailing block $\mathcal{I}_n(\psi_n) \in \mathbb{R}^{r_n\times r_n}$,

$$\begin{aligned} u_n^{\mathrm{T}} W_n^S(\theta_n) &= w_n^{\mathrm{T}} \mathcal{I}_n(\beta_n;\psi_n)^{-1/2} S_n(\beta_n;\psi_n) + v_n^{\mathrm{T}} W_n^S(\psi_n) \\ &= w_n^{\mathrm{T}} \mathcal{I}_n(\beta_n;\psi_n)^{-1/2} X^{\mathrm{T}} \Sigma_n^{-1/2} R_n + \frac{1}{2} R_n^{\mathrm{T}} A_n(\tilde{v}_n, \psi_n) R_n - \frac{1}{2}\operatorname{tr}\{A_n(\tilde{v}_n,\psi_n)\} \\ &= \tilde{w}_n^{\mathrm{T}} R_n + \frac{1}{2} R_n^{\mathrm{T}} A_n(\tilde{v}_n,\psi_n) R_n - \frac{1}{2}\operatorname{tr}\{A_n(\tilde{v}_n,\psi_n)\} \end{aligned}$$

where $\tilde{w}_n = \Sigma_n^{-1/2} X \mathcal{I}_n(\beta_n;\psi_n)^{-1/2} w_n = \Sigma^{-1/2} X (X^{\mathrm{T}}\Sigma_n^{-1} X)^{-1/2} w_n$ and $R_n \sim \mathrm{N}(0, I_n)$ under $\theta_n$. By the subsequence principle (Billingsley, 1999, Theorem 2.6), it suffices to show every subsequence has a further subsequence along which the quantity in the last display is asymptotically standard normal. Thus, since $\|v_n\| \leq 1$, we may by the Bolzano–Weierstrass property assume $\|v_n\| \to v_* \in [0,1]$. If $v_* = 0$, then $u_n^{\mathrm{T}} W_n^S(\theta_n) = \tilde{w}_n^{\mathrm{T}} R_n + o_p(1) \to \mathrm{N}(0,1)$ in distribution under $\theta_n$ by Slutsky's theorem since $\|\tilde{w}_n\| = \|w_n\| \to 1$.

If instead $v^* > 0$, then for large enough $n$, $\|v_n\| > 0$, and hence $\tilde{v}_n = \mathcal{I}_n(\psi_n)^{-1/2} v \neq 0$. Here, we used that $a(\psi_n, v) \leq \bar{a}_n < 8^{-1/2}$ implies $a(\psi_n, v)$ is defined for all $v \in \mathbb{S}^{r-1}$, so $\mathcal{I}_n(\psi_n)$ is positive definite by the discussion preceding the definition of $a(v,\psi)$. Now let $U_n \in \mathbb{R}^{r_n\times r_n}$ be an orthogonal matrix with eigenvectors of $A_n(\tilde{v}_n,\psi_n)$ as columns, and let $\Gamma_n = \operatorname{diag}(\gamma_{n1},\dots,\gamma_{nn})$ be a matrix with the corresponding eigenvalues, $n' \geq 1$ of which are non-zero since $A_n(\tilde{v},\psi)$ is symmetric and not identically zero. Suppose without loss of generality the eigenvalues are sorted so that $|\gamma_{n1}| \geq \cdots \geq |\gamma_{nn}|$. By orthogonal invariance of the multivariate standard normal distribution, $U_n R_n$ has the same distribution as $R_n$. Thus, with $\bar{w}_n = U_n^{\mathrm{T}} \tilde{w}_n$ the right-hand side in the last display equation has the same distribution as

$$\begin{aligned} &\tilde{w}_n^{\mathrm{T}} U R_n + \frac{1}{2} R_n^{\mathrm{T}} U^{\mathrm{T}} A_n(\tilde{v}_n,\psi_n) U_n R_n - \frac{1}{2}\operatorname{tr}\{U_n^{\mathrm{T}} A_n(\tilde{v}_n,\psi_n) U\} \\ &= \sum_{i=1}^n \bar{w}_{ni} R_{ni} + \frac{1}{2}\sum_{i=1}^n R_{ni}^2 \gamma_{ni} - \frac{1}{2}\sum_{i=1}^n \gamma_{ni} \\ &= \sum_{i=n'+1}^n \bar{w}_{ni} R_{ni} + \frac{1}{2}\sum_{i=1}^{n'} \gamma_{ni}(R_{ni}^2 + 2\bar{w}_{ni} R_{ni}/\gamma_{ni} - 1) \\ &= \sum_{i=n'+1}^n \bar{w}_{ni} R_{ni} + \frac{1}{2}\sum_{i=1}^{n'} \gamma_{ni}\{(R_{ni} + \bar{w}_{ni}/\gamma_{ni})^2 - \bar{w}_{ni}^2/\gamma_{ni}^2 - 1\}. \end{aligned}$$

Observe the first and second sum are independent because the elements of $R_n$ are, the first sum is normally distributed with mean zero, and the second sum is a weighted sum of independent noncentral chi-squared random variables, with respective noncentrality parameters $\bar{w}_{ni}/\gamma_{ni}$,

$i \in \{1, \dots, n'\}$; centered to have mean zero. Let $h_n^2 = \sum_{i=n'+1}^{n} \bar{w}_{ni}^2$ be the variance of the first sum. Using that $R_{ni}$ and $R_{ni}^2$ are uncorrelated, one gets that the variance of the second sum is $k_n^2 = 2^{-1} \sum_{i=1}^{n'} (\gamma_{ni}^2 + 2\bar{w}_{ni}^2)$.

By construction, $h_n^2 + k_n^2 = 1$, so we may again pass to a subsequence and assume $k_n \to k_* \in [0, 1]$. If $k_* = 0$, convergence follows as in the case where $v_* = 0$. Suppose lastly that $k_* > 0$. Then Theorem 2 of Zhang (2005) says the second sum is asymptotically normal with mean zero and variance $k^*$ if $\gamma_{n1}^2 / \sum_{j=1}^{n'} \gamma_{ni}^2 \to 0$. But that is equivalent to $a(\tilde{v}, \psi_n)^2 = \|A_n(\tilde{v}, \psi_n)\|^2 / \|A_n(\tilde{v}, \psi_n)\|_F^2 \to 0$, which follows from the assumption $\bar{a}_n \to 0$. Thus, using independence of the first and second sum in the last display, they converge jointly to a bivariate normal with mean zero and covariance matrix $\operatorname{diag}(1 - k_*^2, k_*^2)$. Thus, their sum is asymptotically standard normal by the continuous mapping theorem, which completes the proof. □

*Proof of Lemma 6.* Submultiplicativity of the spectral norm gives

$$\|A(v, \psi)\| \leq \|\Sigma(\psi)^{-1/2}\|^2 \|\Sigma(v)\| = \|\Sigma(\psi)^{-1}\| \|\Sigma(v)\|.$$

Moreover, using the minimax principle for eigenvalues of symmetric matrices (Bhatia, 2012, Corollary III.1.2), with $e_j \in \mathbb{R}^n$ denoting the $j$th standard basis vector,

$$\begin{aligned}
\|A(v, \psi)\|_F^2 &= \operatorname{tr}\{\Sigma(\psi)^{-1/2} \Sigma(v) \Sigma(\psi)^{-1} \Sigma(v) \Sigma(\psi)^{-1/2}\} \\
&= \sum_{j=1}^{n} e_j^{\mathrm{T}} \Sigma(\psi)^{-1/2} \Sigma(v) \Sigma(\psi)^{-1} \Sigma(v) \Sigma(\psi)^{-1/2} e_j \\
&\geq \gamma_{\min}\{\Sigma(\psi)^{-1}\} \sum_{j=1}^{n} e_j^{\mathrm{T}} \Sigma(\psi)^{-1/2} \Sigma(v) \Sigma(v) \Sigma(\psi)^{-1/2} e_j.
\end{aligned}$$

Using cyclical invariance of the trace, that is, of the matrices in the sum in the last line, and then the minimax principle again, gives that $\|A(v, \psi)\|_F^2$ is no smaller than

$$\gamma_{\min}\{\Sigma(\psi)^{-1}\}^2 \operatorname{tr}\{\Sigma(v)^2\} = \gamma_{\min}\{\Sigma(\psi)^{-1}\}^2 \|\Sigma(v)\|_F^2 = \|\Sigma(\psi)\|^{-2} \|\Sigma(v)\|_F^2.$$

Combining this lower bound with the upper bound on $\|A(v, \psi)\|$ completes the proof. □

*Proof of Theorem 3.* Let us first suppose $p = 0$ and consider the special case $Z_i = Z_1$ for all $i$. That is, observations are independent and identically distributed, which substantially simplifies the proof. To apply Lemma 6, note $\|\Sigma(\psi)\| = \max_{1 \leq i \leq m} \|\Sigma_i(\psi)\|$, where $\Sigma_i(\psi) =$

$Z_i\Psi_1(\psi)Z_i^{\mathrm{T}} + \psi_r I_{n_i}$. Thus, by the triangle inequality and submultiplicativity of the spectral norm, $\|\Sigma(\psi)\| \leq \|Z_i\|^2\|\Psi_1(\psi)\| + \psi_r \leq c_2\|\Psi_1(\psi)\| + \psi_r$. Moreover, $\|\Psi_1(\psi)\|^2 \leq \|\Psi_1(\psi)\|_F^2 \leq 2\|\psi_{-r}\|^2$ since $\mathrm{vech}(\Psi_1) = \psi_{-r}$. Thus, we have established $\|\Sigma(\psi)\| \leq 2^{1/2}c_2\|\psi_{-r}\| + \psi_r$, and hence by using $\|\Sigma(\psi)^{-1}\| = 1/\gamma_{\min}\{\Sigma(\psi)\} \leq 1/\psi_r$,

$$\|\Sigma(\psi)^{-1}\|\|\Sigma(\psi)\| \leq 2^{1/2}c_2\|\psi_{-r}/\psi_r\| + 1 \leq 2^{1/2}c_2(1 + \|\psi_{-r}/\psi_r\|).$$

We next show that, when $Z_i = Z_1$ for all $i$, for any $v \in \mathbb{S}^{r-1}$,

$$\|\Sigma(v)\|/\|\Sigma(v)\|_F \leq m^{-1/2}.$$

By block-diagonality, $\|\Sigma(v)\| = \|\Sigma_1(v)\| \leq \|\Sigma_1(v)\|_F$ and $\|\Sigma(v)\|_F^2 = \sum_{i=1}^m \|\Sigma_i(v)\|_F^2 = m\|\Sigma_1(v)\|_F^2$, so $\|\Sigma(v)\|_F \geq m^{1/2}\|\Sigma_1(v)\|_F$. Thus, the bound is established if $\|\Sigma_1(v)\|_F \neq 0$ for all $v \in \mathbb{S}^{r-1}$. But by block-diagonality, that is equivalent to $\|\Sigma(v)\|_F > 0$, which holds by Corollary 1 since $Z = \mathrm{bdiag}(Z_1, \dots, Z_m)$ has full column rank. Indeed, $\gamma_{\min}(Z_i^{\mathrm{T}}Z_i) > 0$ for all $i$ implies each $Z_i$ has full column rank.

Now suppose the $Z_i$ need not be the same, but still $p = 0$. The bound on $\|\Sigma(\psi)^{-1}\|\|\Sigma(\psi)\|$ still holds because we did not use the assumption that $Z_i = Z_1$, so it suffices to bound $\|\Sigma(v)\|/\|\Sigma(v)\|_F$. To that end, note

$$\|\Sigma(v)\| = \max\left(\max_{1\leq i\leq m}\|Z_i\Psi_1(v)Z_i^{\mathrm{T}} + v_r I_{n_i}\|, |v_r|\right).$$

Moreover, using the triangle inequality, submultiplicativity of the spectral norm, $|v_r| \leq 1$, and $\|\Psi_1(v)\|^2 \leq \|\Psi_1(v)\|_F^2 \leq 2\|v\|^2 \leq 2$, we get $\|Z_i\Psi_1(v)Z_i^{\mathrm{T}} + v_r I_{n_i}\| \leq 2^{1/2}\|Z_i\|^2 + 1 \leq 2^{1/2}c_2 + 1$. Thus, $\|\Sigma(v)\| \leq 1 + 2^{1/2}c_2$.

Next, $\|\Sigma(v)\|_F^2 = \sum_{i=1}^m \|Z_i\Psi_1(v)Z_i^{\mathrm{T}} + v_r I_{n_i}\|_F^2$. By cyclic invariance of the trace, the $i$th term in the sum satisfies, with $C_i = Z_i^{\mathrm{T}}Z_i$,

$$\begin{aligned}
\|Z_i\Psi_1(v)Z_i^{\mathrm{T}} + v_r I_{n_i}\|_F^2 &= \mathrm{tr}\{Z_i\Psi_1(v)Z_i^{\mathrm{T}}Z_i\Psi(v)Z_i^{\mathrm{T}} + 2v_r Z_i\Psi_1(v)Z_i^{\mathrm{T}} + v_r^2 I_{n_i}\} \\
&= \mathrm{tr}\{C_i^{1/2}\Psi_1(v)C_i^{1/2} + 2v_r C_i^{1/2}\Psi_1(v)C_i^{1/2} + v_r^2 I_{q_1}\} + v_r^2(n_i - q_1) \\
&= \|C_i^{1/2}\Psi_1(v)C_i^{1/2} + v_r I_{q_1}\|_F^2 + v_r^2(n_i - q_1).
\end{aligned}$$

To lower bound this, we use, in order, the triangle inequality, $\gamma_{\min}(C_i) \geq c_2^{-1}$, $\|\Psi_1(v)\|_F^2 \geq$

$\|v_{-r}\|^2 = 1 - v_r^2$, and Jensen's inequality applied to the concave square-root to write

$$\begin{aligned}\|C_i^{1/2}\Psi_1(v)C_i^{1/2} + v_r I_{q_1}\|_F &\geq \|C_i^{1/2}\Psi_1(v)C_i^{1/2}\|_F - |v_r| q_1^{1/2} \\ &\geq c_2^{-1}\|\Psi_1(v)\|_F - |v_r| q_1^{1/2} \\ &\geq c_2^{-1}(1 - v_r^2)^{1/2} - |v_r| q_1^{1/2} \\ &\geq c_2^{-1} 2^{-1/2}(1 - |v_r|) - |v_r| q_1^{1/2} \\ &= 2^{-1/2}\{c_2^{-1} - |v_r|(c_2^{-1} + (2q_1)^{1/2})\}.\end{aligned}$$

Thus, if $|v_r| \leq 2^{-1}c_2^{-1}(c_2^{-1} + (2q_1)^{1/2})^{-1} = 1/(2 + c_2(8q_1)^{1/2})$, say; then $\|C_i^{1/2}\Psi_1(v)C_i^{1/2} + v_r I_{q_1}\|_F \geq 1/(c_2 8^{1/2})$ and, therefore, $\|\Sigma(v)\|_F \geq m^{1/2}/(c_2 8^{1/2})$. If on the other hand $|v_r| > 1/(2 + c_2(8q_1)^{1/2})$, then

$$\|\Sigma(v)\|_F \geq |v_r|(n - mq_1)^{1/2} \geq m^{1/2}(\bar{n} - q_1)^{1/2}/(2 + c_1(8q_1)^{1/2}) \geq m^{1/2}(c_1 - 1)^{1/2}/(2 + c_2 8^{1/2}).$$

Consequently,

$$\|\Sigma(v)\|_F \geq m^{1/2} \min\{1/(c_2 8^{1/2}), (c_1 - 1)^{1/2}/(2 + c_2 8^{1/2})\} = m^{1/2}\tilde{c}_3,$$

where $\tilde{c}_3$ is defined by the last equality. In summary, Lemma 6 gives

$$a(v, \psi) \leq (1 + 2^{1/2}c_2)2^{1/2}c_2\tilde{c}_3^{-1}m^{-1/2}(1 + \|\psi_{-r}/\psi_r\|),$$

which upon taking, for example, $c_3 = (1 + 2^{1/2}c_2)^2\tilde{c}_3^{-1}$ completes the case $p = 0$.

Suppose finally that $p > 0$ and let $\tilde{\Sigma}(v) = V^{\mathrm{T}}\Sigma(v)V$. Observe $\|\tilde{\Sigma}(v)\| = \|V^{\mathrm{T}}\Sigma(v)V\| \leq \|V\|^2\|\Sigma(v)\| = \|\Sigma(v)\|$ by submultiplicativity of the spectral norm. Moreover, by H.1.h. of Marshall et al. (2011, p.341) and the claim following it,

$$\|\tilde{\Sigma}(v)\|_F^2 = \sum_{j=1}^{n} \lambda_j\{VV^{\mathrm{T}}\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\} \geq \sum_{j=1}^{n} \lambda_j\{\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\}\lambda_{n-j+1}(VV^{\mathrm{T}}),$$

which, since $VV^{\mathrm{T}}$ has 1 and 0 as eigenvalues with respective multiplicities $n - p$ and $p$, equals $\sum_{j=p+1}^{n} \lambda_j\{\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\} = \mathrm{tr}\{\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\} - \sum_{j=1}^{p} \lambda_j\{\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\}$. For the first of these terms, cyclical invariance of the trace and H.1.h. of Marshall et al. (2011) gives

$$\mathrm{tr}\{\Sigma(v)VV^{\mathrm{T}}\Sigma(v)\} = \mathrm{tr}\{\Sigma(v)^2VV^{\mathrm{T}}\} \geq \sum_{j=p+1}^{n} \lambda_j\{\Sigma(v)^2\} = \|\Sigma(v)\|_F^2 - \sum_{j=1}^{p} \lambda_j\{\Sigma(v)^2\}.$$

Thus, since $|\sum_{j=1}^{p}\lambda_j\{\Sigma(v)VV^{\mathsf{T}}\Sigma(v)\}| \leq p\|\Sigma(v)VV^{\mathsf{T}}\Sigma(v)\| \leq p\|\Sigma(v)\|^2$ by submultiplicativity of the spectral norm, we find

$$\begin{aligned}\|\tilde{\Sigma}(v)\|_F^2 &\geq \|\Sigma(v)\|_F^2 - 2p\|\Sigma(v)\|^2\\ &\geq m\left[\min\{c_2^{-2}8^{-1}, (c_1-1)^{-1}(2+c_2 8^{1/2})^{-1}\} - 2c_4(1+c_2 2^{1/2})^2\right]\\ &= m\tilde{c}_4,\end{aligned}$$

where $\tilde{c}_4$, defined by the last equality, is strictly positive for small enough $c_4$. Thus, by Lemma 6, $\tilde{a}(v,\psi) \leq (1+2^{1/2}c_2)2^{1/2}c_2\tilde{c}_4^{-1/2}m^{-1/2}(1+\|\psi_{-r}/\psi_r\|)$, so we are done upon taking, for example, $c_3 = (1+2^{1/2}c_2)^2\tilde{c}_4^{-1/2}$. □

*Proof of Corollary 3.* By Lemmas 3 and 4, it suffices to consider an arbitrary sequence $(\psi_n)$ converging to some $\psi_0 \in \mathbb{P}$ and show that, under that sequence, $\tilde{T}_n^S(\psi_n)$ has as asymptotic chi-squared distribution with $r$ degrees of freedom. For this, it suffices that $\tilde{W}_n^S(\psi_n) \to \mathrm{N}(0, I_r)$ in distribution. This, in turn, follows from the Cramér–Wold theorem and the second conclusion of Corollary 2 if we verify its condition (v). To that end, note $\psi_0 \in \mathbb{P}$ implies $\psi_{0r} > 0$, and hence $\psi_{nr} > 0$ for all large enough $n$. Thus, since $\|\psi_{-nr}\|$ is bounded due to $\psi_n$ being convergent, $\|\psi_{-nr}\|/\psi_{nr} = O(1) = o(m^{1/2})$, so the proof is completed. □

*Proof of Theorem 4.* Let $Q_j = I_{n_j} - P_j$ and define $\mathcal{P}_j^Q$ the same way as $\mathcal{P}_j$ but with $Q_j$ in place of $I_{n_j}$, $j \in \{1,\dots,r-1\}$. Let also $\mathcal{P}_r = P_1 \otimes \cdots \otimes P_{r-1} = 1_n 1_n^{\mathsf{T}}/n$. Then $\mathcal{P}_j = \mathcal{P}_r + \mathcal{P}_j^Q$ for $j \in \{1,\dots,r-1\}$ and, consequently,

$$\begin{aligned}\Sigma &= \sum_{j=1}^{r-1}\psi_j n_{(j)}(\mathcal{P}_r + \mathcal{P}_j^Q) + \psi_r I_n\\ &= \sum_{j=1}^{r-1}(\psi_r + \psi_j n_{(j)})\mathcal{P}_j^Q + \mathcal{P}_r\left(\psi_r + \sum_{j=1}^{r-1}\psi_j n_{(j)}\right) + \psi_r(I_n - \mathcal{P}),\end{aligned} \tag{1}$$

where $\mathcal{P} = \mathcal{P}_r + \sum_{j=1}^{r-1}\mathcal{P}_j^Q$. Here, the last step uses that $\mathcal{P}_r\mathcal{P}_j^Q = 0$ for every $j \in \{1,\dots,r-1\}$, and $\mathcal{P}_i^Q\mathcal{P}_j^Q = 0$ for $i \neq j$, both of which follow from the mixed-product property (Magnus and Neudecker, 2002, Equation (4), p.32) and $Q_jP_j = 0$ for $j \in \{1,\dots,r-1\}$. Thus, $\mathcal{P}$ is a projection matrix satisfying $(I_n - \mathcal{P})\mathcal{P}_j^Q = 0$ for $j \in \{1,\dots,r-1\}$ and $(I_n - \mathcal{P})\mathcal{P}_r = 0$. Consequently, (1) gives a spectral decomposition of $\Sigma$ where each of the $r+1$ addends is a projection onto an eigenspace, scaled by the corresponding eigenvalue.

Recalling the definition $A_j = \Sigma^{-1/2}(\partial\Sigma/\partial\psi_j)\Sigma^{-1/2}$, we get

$$\begin{aligned} A(v,\psi) &= \Sigma(\psi)^{-1/2}\Sigma(v)\Sigma(\psi)^{-1/2} \\ &= \sum_{j=1}^{r-1} \frac{v_r + v_j n_{(j)}}{\psi_r + \psi_j n_{(j)}} \mathcal{P}_j^Q + \mathcal{P}_r \left( \frac{v_r + \sum_{j=1}^{r-1} v_j n_{(j)}}{\psi_r + \sum_{j=1}^{r-1} \psi_j n_{(j)}} \right) + \frac{v_r}{\psi_r}(I_n - \mathcal{P}). \end{aligned}$$

Write $s_1^2, \ldots, s_{r+1}^2$ for the squared eigenvalues of $A(v,\psi)$, ordered as in the last display. That is, $s_j^2 = (v_r + v_j n_{(j)})^2/(\psi_r + \psi_j n_{(j)})^2$, $j \in \{1,\ldots,r-1\}$; $s_r^2 = (v_r + \sum_{j=1}^{r-1} v_j n_{(j)})^2/(\psi_r + \sum_{j=1}^{r-1} \psi_j n_{(j)})^2$; and $s_{r+1}^2 = v_r^2/\psi_r^2$. Thus, $\|A(v,\psi)\|^2 = \max\{s_1^2,\ldots,s_{r+1}^2\}$. If this maximum is $s_j^2$ for some $j \in \{1,\ldots,r-1\}$, then since $\|A(v,\psi)\|_F^2 \geq s_j^2 \|\mathcal{P}_j^Q\|_F^2 = s_j^2(n_j - 1)$, we get $a(v,\psi)^2 \leq 1/(n_j - 1)$, where we used that, for a projection matrix, its squared Frobenius norm is its rank. Similarly, if $\|A(v,\psi)\|^2 = s_{r+1}^2$, then $a(v,\psi)^2 \leq 1/\|I_n - \mathcal{P}_r\|_F^2 = 1/\tilde{n}$. It remains only to consider the case where $\|A(v,\psi)\|^2 = s_r^2$. This case is more complicated because $\|\mathcal{P}_r\|_F = 1$.

Since the squared Frobenius norm of a symmetric matrix is the sum of squared eigenvalues, $\|A(v,\psi)\|_F^2$ is

$$\sum_{j=1}^{r-1} \frac{(v_r + v_j n_{(j)})^2}{(\psi_r + \psi_j n_{(j)})^2}(n_j - 1) + \left( \frac{v_r + \sum_{j=1}^{r-1} v_j n_{(j)}}{\psi_r + \sum_{j=1}^{r-1} \psi_j n_{(j)}} \right)^2 + \frac{v_r^2}{\psi_r^2}\tilde{n}.$$

Since the $\psi_j$ are all non-negative, the denominator for each ratio in the last display is upper bounded by $(\psi_r + \sum_{j=1}^{r-1} \psi_j n_{(j)})^2$, which is also the denominator of $s_r^2$. Thus, when $\|A(v,\psi)\|^2 = s_r^2$, $a(v,\psi)^2$ is upper bounded by

$$\frac{(v_r + \sum_{j=1}^{r-1} v_j n_{(j)})^2}{\sum_{j=1}^{r-1}(v_r + v_j n_{(j)})^2(n_j - 1) + (v_r + \sum_{j=1}^{r-1} v_j n_{(j)})^2 + v_r^2\tilde{n}}.$$

Let $b_j = v_r + v_j n_{(j)}$, $j \in \{1,\ldots,r-1\}$, and $b_r = v_r$. Then the last display is a ratio of quadratic forms in $b = [b_1,\ldots,b_r]^{\mathrm{T}}$. Specifically, since $v_r + \sum_{j=1}^{r-1} v_j n_{(j)} = \sum_{j=1}^{r-1}(v_r + n_{(j)} v_j) - (r-2)v_r = \sum_{j=1}^{r-1} b_j - (r-2)b_r$, it is

$$\frac{(b^{\mathrm{T}}u)^2}{b^{\mathrm{T}}Db + (b^{\mathrm{T}}u)^2} = \frac{b^{\mathrm{T}}uu^{\mathrm{T}}b}{b^{\mathrm{T}}(D + uu^{\mathrm{T}})b}.$$

where $u = [1,\ldots,1,-(r-2)]^{\mathrm{T}}$ and $D = \mathrm{diag}(n_1 - 1,\ldots,n_{r-1} - 1,\tilde{n})$. Since $D$ is positive definite, we can let $\tilde{b} = (D + uu^{\mathrm{T}})^{1/2} b$ and get by Cauchy–Schwartz's inequality that

$$\frac{\tilde{b}^{\mathrm{T}}(D + uu^{\mathrm{T}})^{-1/2} uu^{\mathrm{T}} (D + uu^{\mathrm{T}})^{-1/2} \tilde{b}}{\|\tilde{b}\|^2} \leq u^{\mathrm{T}}(D + uu^{\mathrm{T}})^{-1} u.$$

Using the Sherman–Morrison formula, $u^{\mathrm{T}}(D+uu^{\mathrm{T}})^{-1}u = u^{\mathrm{T}}D^{-1}u - (u^{\mathrm{T}}D^{-1}u)^2(1+u^{\mathrm{T}}D^{-1}u)^{-1} = u^{\mathrm{T}}D^{-1}u\{1 - (u^{\mathrm{T}}D^{-1}u)(1 + u^{\mathrm{T}}D^{-1}u)^{-1}\} \leq u^{\mathrm{T}}D^{-1}u = \sum_{j=1}^{r-1}(n_j - 1)^{-1} + (r-2)^2/\tilde{n}$, which completes the proof of the first claim. We note the last inequality is asymptotically strict in the sense that when $u^{\mathrm{T}}D^{-1}u$ tends to zero, the left-hand side is $u^{\mathrm{T}}D^{-1}u(1 + o(1))$.

For the second claim, let $\tilde{\Sigma}(v) = V^{\mathrm{T}}\Sigma(v)V$ and $\tilde{A}(v, \psi) = \tilde{\Sigma}(\psi)^{-1/2}\tilde{\Sigma}(v)\tilde{\Sigma}(\psi)^{-1/2}$. Observe that $\mathcal{P}_r = 1_n 1_n^{\mathrm{T}}/n = I_n - VV^{\mathrm{T}}$. Therefore, by (1),

$$\begin{aligned}
\tilde{\Sigma} &= \sum_{j=1}^{r-1}(\psi_r + \psi_j n_{(j)}) V^{\mathrm{T}}\mathcal{P}_j^Q V + \psi_r V^{\mathrm{T}}(I_n - \mathcal{P})V \\
&= \sum_{j=1}^{r-1}(\psi_r + \psi_j n_{(j)}) V^{\mathrm{T}}\mathcal{P}_j^Q V + \psi_r I_{n-1} - \sum_{j=1}^{r-1} \psi_r V^{\mathrm{T}}\mathcal{P}_j^Q V \\
&= \sum_{j=1}^{r-1} \psi_j n_{(j)} V^{\mathrm{T}}\mathcal{P}_j^Q V + \psi_r I_{n-1}.
\end{aligned}$$

Now $V^{\mathrm{T}}\mathcal{P}_i^Q V V^{\mathrm{T}}\mathcal{P}_j^Q V = V^{\mathrm{T}}\mathcal{P}_i^Q (I_n - \mathcal{P}_r)\mathcal{P}_j^Q V = V^{\mathrm{T}}\mathcal{P}_i^Q \mathcal{P}_j^Q V$, which is zero if $i \neq j$ and $V^{\mathrm{T}}\mathcal{P}_i^Q V$ otherwise. That is, the leading $r-1$ terms in the sum in last display are projection matrices onto orthogonal spaces. Therefore,

$$\tilde{\Sigma} = \sum_{j=1}^{r-1}(\psi_r + \psi_j n_{(j)}) V^{\mathrm{T}}\mathcal{P}_j^Q V + \psi_r (I_{n-1} - \tilde{\mathcal{P}}),$$

where $\tilde{\mathcal{P}} = \sum_{j=1}^{r-1} V^{\mathrm{T}}\mathcal{P}_j^Q V = V^{\mathrm{T}}\mathcal{P}V$ is also a projection matrix. Now by arguments similar to those for the case with $V = I_n$,

$$\begin{aligned}
\tilde{a}(v, \psi)^2 &\leq \max\left(\|I_{n-1} - \tilde{\mathcal{P}}\|_F^{-2}, \|V^{\mathrm{T}}\mathcal{P}_1^Q V\|_F^{-2}, \ldots, \|V^{\mathrm{T}}\mathcal{P}_{r-1}^Q V\|_F^{-2}\right) \\
&\leq \max\left((\tilde{n} - 1)^{-1}, (n_j - 2)^{-1}, \ldots, (n_{r-1} - 2)^{-1}\right),
\end{aligned}$$

which completes the proof.

□

*Proof of Corollary 4.* Recall $n = n(k)$ depends on $k$ but let us suppress that for simplicity.

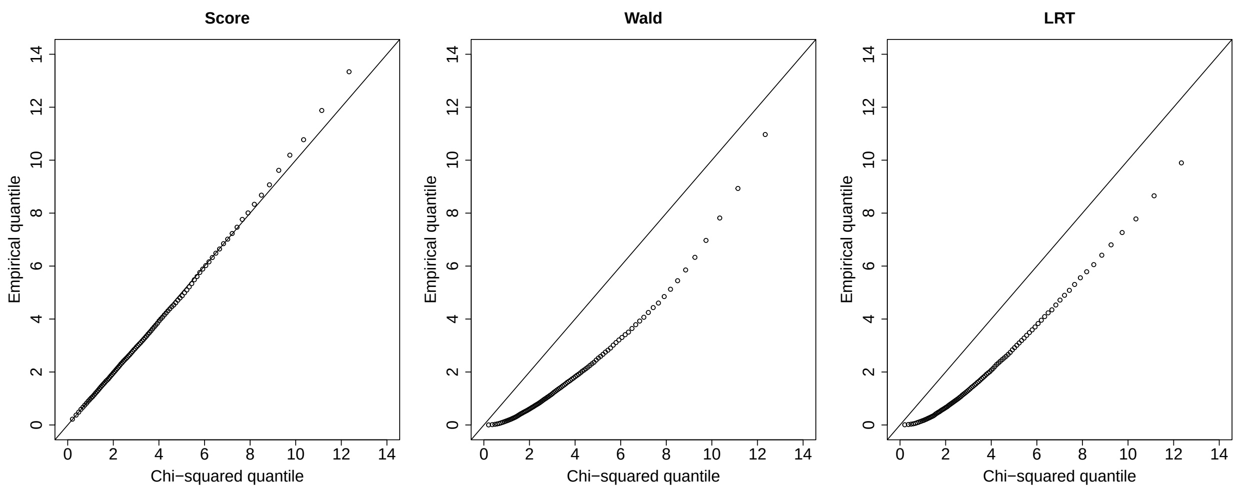


Figure A: Quantiles of test-statistics evaluated at a true $\psi \in \mathbb{R}^4$ that is near the boundary. For $i \in \{1, \dots, 50\}$, $Y_i = Z_i U_i + E_i$, where $U_i \sim \mathrm{N}(0, \Psi_1)$, $\Psi_1 \in \mathbb{R}^{2\times 2}$ has diagonal elements $\psi_1 = \psi_3 = 10^{-3}$ and off-diagonal $\psi_2 = 0$, and $E_i \sim \mathrm{N}(0, \psi_4)$, $\psi_4 = 10$. 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.

Suppose for contradiction that

$$\limsup_{k\to\infty} \sup_{\psi \in \mathbb{P}} \left| \mathrm{P}_\psi\{\psi \in \tilde{\mathbb{C}}_n^S(\alpha)\} - (1-\alpha) \right| > 0.$$

Then for some $\epsilon > 0$, $\sup_{\psi\in\mathbb{P}} \left| \mathrm{P}_\psi\{\psi \in \tilde{\mathbb{C}}_n^S(\alpha)\} - (1-\alpha)\right| \geq \epsilon$ for infinitely many $k$. For each such $k$, we can pick a $\psi_k \in \mathbb{P}$ such that $\left| \mathrm{P}_{\psi_k}\{\psi_k \in \tilde{\mathbb{C}}_n^S(\alpha)\} - (1-\alpha)\right| \geq \epsilon/2$, say. But Theorem 4 says that as $k \to \infty$ and $n_{\min}^k \to \infty$, for any $v \in \mathbb{S}^{r-1}$, the density of $v^{\mathrm{T}} \tilde{W}_n^S(\psi_k)$ tends to the standard normal density. Therefore, as argued in the proof of Theorem 2, $\tilde{W}_n^S(\psi_k) \to \mathrm{N}(0, I_r)$ in distribution and $\tilde{T}_n^S(\psi_k) \to \chi^2_r$ in distribution. Thus, $\mathrm{P}_{\psi_k}\{\psi_k \in \tilde{\mathbb{C}}_n^S(\alpha)\} = \mathrm{P}_{\psi_k}\{\tilde{T}_n^S(\psi_k) \leq c_{r,1-\alpha}\} \to 1-\alpha$, which is the desired contradiction. □

# B Additional numerical results

Figure A is like Figure 1 in the main text but with error variance $\psi_4 = 10$ instead of $\psi_4 = 1$. Notably, though we have verified there are some small differences, the plots are almost identical. These results suggest that, at least in this setting, the conclusions are not particularly sensitive to the error variance.

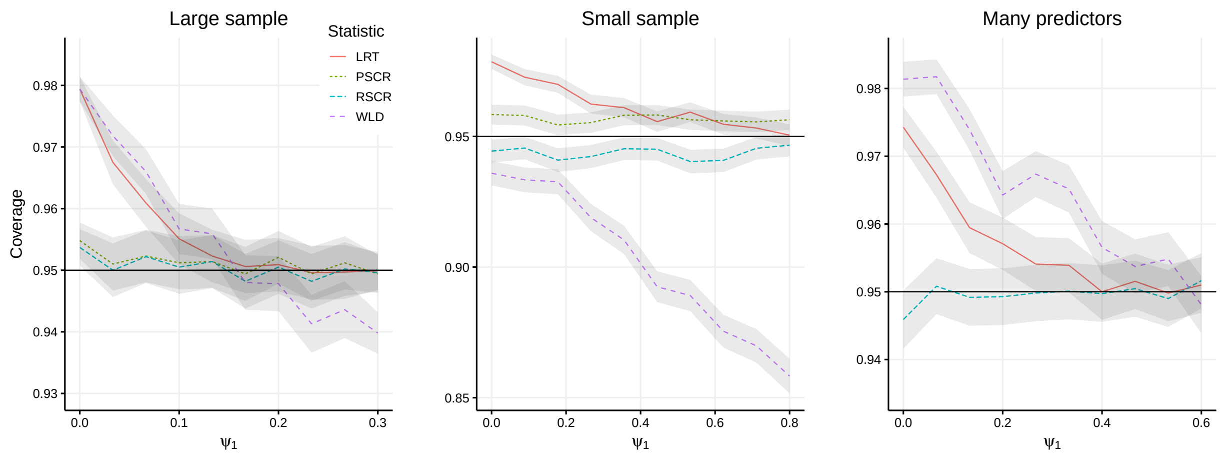


Figure B: Coverage probabilities with independent clusters and independent random effects, for different random effect variances (horizontal axes). The settings are $(m, n_i, p)$ equal to $(500, 3, 2)$ (left plot), $(20, 3, 3)$ (middle plot), and $(200, 3, 100)$ (right plot).

Figure B shows results from a simulation with the same settings as those for Figure 2 in the main text, except that, here, the random effects' covariance matrix is $\Psi_1 = \text{diag}(\psi_1, \psi_3)$, with $\psi_1 = \psi_3$ indicated on the horizontal axes. That is, the true covariance $\psi_2 = 0$ and the random effects are therefore independent. When $m = 500$ (left plot), the likelihood ratio and Wald regions are conservative near the boundary while both score-based methods have approximately nominal coverage probabilities. Further from the boundary, the Wald confidence regions are invalid while the other three have approximately nominal coverage probability. When $m = 20$ (middle plot), the Wald region is clearly invalid for all considered settings; the likelihood ratio region is conservative. The profile and restricted score regions have coverage probabilities above and below nominal, respectively. However, both are close to nominal for all considered parameter values. With many predictors (right plot), the likelihood ratio and Wald regions are conservative near the boundary. The restricted score-based region has approximately correct coverage. The profile score-based region is omitted because its coverage is so far below nominal that it would obscure the comparisons of the other methods.

Figure C shows results from a simulation with the same settings as that for Figure B, except that in the middle and right plot $\psi_3 \neq \psi_1$. The left plot, which is the same as the left plot in Figure B, is included for comparison. The middle plot shows that when $\psi_3$ is on the boundary, moving $\psi_1$ away from the boundary has only a small effect on the coverage probabilities of likelihood ratio and Wald regions, as expected. Conversely, the right plot shows that the coverage probabilities of likelihood ratio and Wald regions are closer to

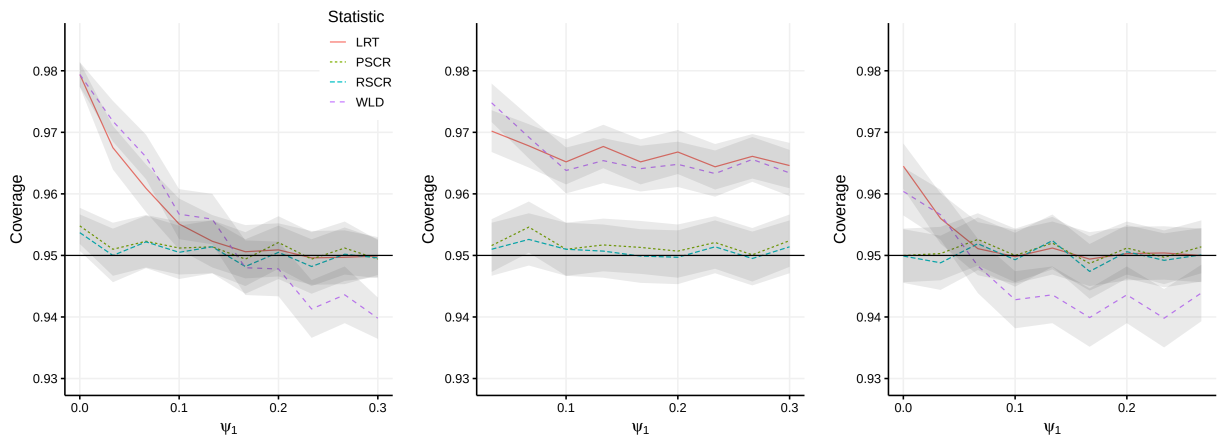


Figure C: Coverage probabilities with independent clusters and independent random effects with $(m, n_i, p) = (500, 3, 2)$. The random effect variance not on the horizontal axis is $\psi_3 = \psi_1$ (left plot), $\psi_3 = 0$ (middle plot), or $\psi_3 = 0.3$ (right plot).

nominal when only one of $\psi_1$ and $\psi_3$ are on the boundary. As before, the score-based regions have approximately nominal coverage probabilities in all settings.

Figure D shows results from a simulation with the same settings as that for Figure 3 in the main text, except that in the right plot $n_1 = 20$ and $n_2 = 80$. That is, the total number of observations is still $n = 1600$, but the observations are unbalanced. The left plot, which is included for comparison, is the same as the left plot in Figure 3 of the main text. The theory suggests $\min_i n_i$ is an important quantity for the asymptotic approximations, and this is reflected in the coverage probabilities here. For example, in the right plot, the coverage issues for the likelihood ratio region are present for a wider range of $\psi_1 = \psi_2$ than in the left plot, where $n_1 = n_2 = 40$. Similarly, the coverage issues for the Wald region are more pronounced in the right plot. The score-based regions have approximately nominal coverage probabilities for all parameter values in both settings.

# C Practical considerations

We make a few observations that can help implement the proposed methods. Recall $\tilde{\beta} = \tilde{\beta}(\psi) = \{X^\mathsf{T}\Sigma(\psi)^{-1}X\}^{-1}X^\mathsf{T}\Sigma(\psi)^{-1}Y$ is a partial maximizer of $\ell(\beta, \psi)$ in $\beta$ and define the profile likelihood $\ell^P(\psi) = \ell(\tilde{\beta}, \psi)$. Then, the restricted log-likelihood can also be written

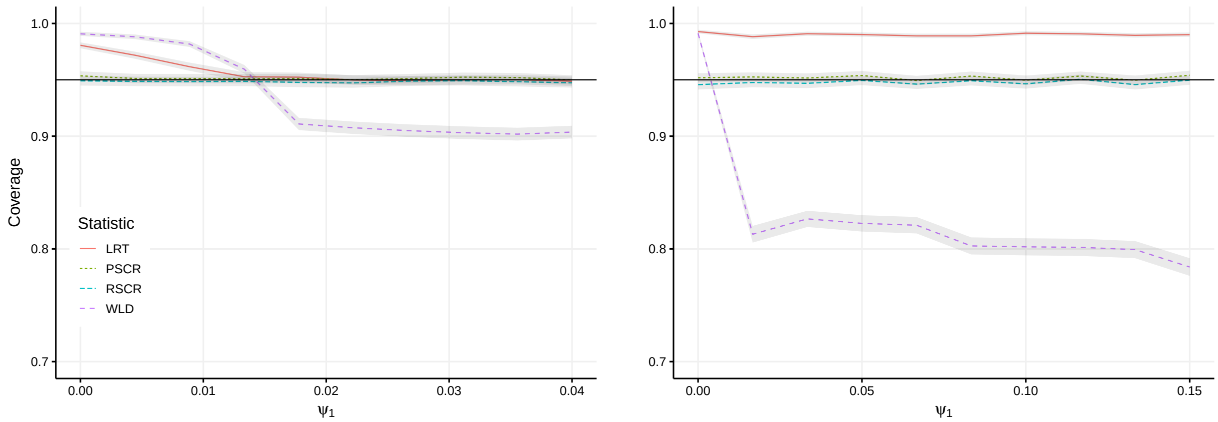


Figure D: Coverage probabilities with crossed, independent random effects, for different random effect variances $\psi_1 = \psi_2$ (horizontal axes). The settings are $(n_1, n_2, p)$ equal to $(40, 40, 2)$ (left plot) and $(20, 80, 2)$ (right plot).

(Harville, 1974), up to a constant,

$$\begin{aligned}\ell^R(\psi) &= \ell^P(\psi) - \frac{1}{2}\log|X^{\mathrm{T}}\Sigma(\psi)^{-1}X| \\ &= -\frac{1}{2}\left\{\log|\Sigma(\psi)| + \log|X^{\mathrm{T}}\Sigma(\psi)^{-1}X| + (Y - X\tilde{\beta})^{\mathrm{T}}\Sigma(\psi)^{-1}(Y - X\tilde{\beta})\right\}.\end{aligned}$$

By well-known envelope theorems (see for example Milgrom and Segal, 2002), the profile score satisfies $S^P(\psi) = \partial\ell^P(\psi)/\partial\psi = S\{\tilde{\beta}(\psi), \psi\}$, the usual score evaluated at the partial maximizer in $\beta$. Thus, the restricted score is, for $j < r$, $S^R(\psi_j) = S\{\tilde{\beta}(\psi), \psi\} - (1/2)\,\mathrm{tr}[\{X^{\mathrm{T}}\Sigma(\psi)^{-1}X\}^{-1}X^{\mathrm{T}}\Sigma^{-1}(\psi)ZH_jZ^{\mathrm{T}}\Sigma^{-1}(\psi)X]$, and similarly for $j = r$.

The profile score is straightforward to compute using the expressions for the usual score in the main text. Expressions involving $\Sigma(\psi)^{-1}$ can usually be computed efficiently by using the Woodbury identity to get $\Sigma^{-1} = \psi_r^{-1}I_n - \psi_r^{-2}Z(I_q + \psi_r^{-1}\Psi Z^{\mathrm{T}}Z)^{-1}\Psi Z^{\mathrm{T}}$. Using this, the additional term in the restricted score can also be computed efficiently. Similar arguments apply to the expression for the Fisher information given in the main text. For example, to compute $\mathrm{tr}(A_iA_j)$ for $i, j < r$, note

$$\mathrm{tr}(A_iA_j) = \mathrm{tr}(\Sigma^{-1/2}ZH_iZ^{\mathrm{T}}\Sigma^{-1}ZH_jZ^{\mathrm{T}}\Sigma^{-1/2}) = \mathrm{tr}(Z^{\mathrm{T}}\Sigma^{-1}ZH_iZ^{\mathrm{T}}\Sigma^{-1}ZH_j)$$

and $Z^{\mathrm{T}}\Sigma^{-1}Z$, which is shared for all $i, j < r$, can be computed efficiently using the Woodbury identity above.

# References

Bhatia, R. (2012). *Matrix Analysis*. Springer New York.

Billingsley, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 3rd ed edition.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Probability and Statistics Section. Wiley, New York, 2 edition.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61(2):383–385.

Magnus, J. R. and Neudecker, H. (2002). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. John Wiley, Chichester, rev. ed., reprinted edition.

Marshall, A. W., Olkin, I., and Arnold, B. C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer Science+Business Media, LLC, New York, 2nd ed edition.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601.

Remmert, R. (1991). *Theory of Complex Functions*. Number 122 in Graduate Texts in Mathematics. Springer-Verlag, New York.

Zhang, J.-T. (2005). Approximate and Asymptotic Distributions of Chi-Squared–Type Mixtures With Applications. *Journal of the American Statistical Association*, 100(469):273–285.