Supplement to “Confidence Regions Near Singular Information and Boundary Points With Applications to Mixed Models”

Karl Oskar Ekvall* — karl.oskar.ekvall@ki.se

Matteo Bottai — matteo.bottai@ki.se

Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institute
Nobels väg 13, 17177 Stockholm, Sweden

July 6, 2021

Details for Example 1

Recall, the $Y_i = [Y_1, \dots, Y_r]^\mathsf{T}$, $i = 1, \dots, n$, are independent and multivariate normally distributed with mean 0 and common covariance matrix $\Sigma(\theta) = \theta^2 1_r 1_r^\mathsf{T} + I_r$. Thus, the log-likelihood is for one observation is

$$\log f_\theta(y_i) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2} y_i^\mathsf{T}\Sigma(\theta)^{-1} y_i.$$

Since $1_r 1_r^\mathsf{T}$ has eigenvalues r and 0, $\Sigma(\theta)$ has eigenvalues $1 + r\theta^2$ and 1, the latter with multiplicity $r - 1$. Thus, $|\Sigma(\theta)| = (1 + r\theta^2)1^{r-1} = 1 + r\theta^2$. Applying the Sherman–Morrison formula to $\Sigma(\theta)^{-1}$ gives$(I_r + \theta^2 1_r 1_r^\mathsf{T})^{-1} = I_r - \theta^2 1_r 1_r^\mathsf{T}(1 + r\theta^2)^{-1}$, and hence $2\log f_\theta(y_i) = -\log(1 + r\theta^2) - y_i^\mathsf{T} y_i + (y_i^\mathsf{T} 1_r)^2\theta^2(1 + r\theta^2)^{-1}$. Differentiating $\log f_\theta(y_i)$ with respect to θ gives $s^i(\theta; y_i) = -r\theta(1 + r\theta^2)^{-1} + \theta(y_i^\mathsf{T} 1_r)^2(1 + r\theta^2)^{-2}$. Differentiating again we get $h^i(\theta; y_i) = -(r - r^2\theta^2)\{(1 + r\theta^2)^2\}^{-1} + (y_i^\mathsf{T} 1_r)^2(1 - 3r\theta^2)(1 + r\theta^2)^{-3}$. Thus, using that $\mathbb{E}_\theta\{(Y_i^\mathsf{T} 1_r)^2\} = \mathrm{var}_\theta(Y_i^\mathsf{T} 1_r) = 1_r^\mathsf{T}\Sigma(\theta)1_r = 1_r^\mathsf{T} 1_r(1 + r\theta^2) = r(1 + r\theta^2)$ we find $\mathcal{I}^i(\theta) = -\mathbb{E}_\theta[h^i(\theta; Y_i)] = (r - r^2\theta^2)(1 + r\theta^2)^{-2} - r(1 + r\theta^2)(1 - 3r\theta^2)(1 + r\theta^2)^{-3} = 2r^2\theta^2(1 + r\theta^2)^{-2}$.

*Corresponding author

Consequently, for $\theta > 0$,

$$T_n(\theta; y^n)^{1/2} = n^{-1/2} \sum_{i=1}^{n} \frac{s^i(\theta; y_i)}{\sqrt{\mathcal{I}^1(\theta)}} = (2n)^{-1/2} \sum_{i=1}^{n} \left\{ -1 + \frac{(y_i^\mathsf{T} 1_r)^2}{r(1 + r\theta^2)} \right\}.$$

Define the score test statistic standardized by observed information

$$T_n^O(\theta; y^n) = s_n(\theta; y^n)^\mathsf{T} \left\{ -\nabla^2 \ell_n(\theta; y^n) \right\}^{-1} s_n(\theta; y^n).$$

Theorem 0.1. *In Example 1, with known $\psi = 0$ and $r = 1$ it holds as $n \to \infty$, with $Z \sim \mathcal{N}(0, 1)$,*

$$T_n^O(\theta_n; Y_n^n) \rightsquigarrow \begin{cases} Z^2 & \text{if } n^{1/4}|\theta_n| \to \infty \\ \frac{2a^2 Z^2}{2a^2 - \sqrt{2}Z} & \text{if } \theta_n = a n^{-1/4}, \quad a \in \mathbb{R} \\ 0 & \text{if } \theta_n = o(n^{-1/4}) \end{cases},$$

where $Y_n^n = (Y_{n1}, \ldots, Y_{nn})$ has the distribution indexed by θ_n.

Proof. Recall $s^i(\theta; y_i) = \theta\{-1 + y_i^2/(1 + \theta^2)\}/(1 + \theta^2)$. Some algebra gives that $\nabla^2 \ell^i(\theta; y_i) = (\theta^4 - 3\theta^2 y_i^2 + y_i^2 - 1)/(1 + \theta^2)^3$ and hence

$$T_n^O(\theta; y^n) = \theta^2(1 + \theta^2) \frac{\left[\sum_{i=1}^n \{y_i/(1 + \theta^2) - 1\}\right]^2}{\sum_{j=1}^n \{1 - y_i^2 + 3\theta^2 y_i^2 - \theta^4\}}$$

Let $x_n = \sum_{i=1}^n \{y_i^2/(1 + \theta^2) - 1\}$, or $\sum_{i=1}^n y_i^2 = (1 + \theta^2)(x_n + n)$, to get

$$\begin{aligned} T_n^O(\theta; y^n) &= \theta^2(1 + \theta^2) \frac{x_n^2}{n(1 + \theta^2)(1 - \theta^2) - (1 - 3\theta^2)(x_n + n)(1 + \theta^2)} \\ &= \frac{x_n^2/n}{2 - (1 - 3\theta^2)(\theta^2 n)^{-1} x_n} \end{aligned}$$

Observe that $X_n \sim (\chi_n^2 - n)$ regardless of θ, where X_n is defined as x_n but with Y_i in place of y_i. Thus, $n^{-1/2} X_n \rightsquigarrow \sqrt{2}Z$ by the central limit theorem. Thus, if $\theta_n^2 n = a^2\sqrt{n}$, or $\theta_n = a n^{-1/4}$, then $T_n^O(\theta_n) \rightsquigarrow 2a^2 Z^2/(2a^2 - \sqrt{2}Z)$ by Slutsky and mapping theorems. The other cases now follow by routine arguments. □

Additional results

Lemma 0.2. *If Assumptions 1–2 and 4–5 hold; then for any $n = 1, 2, \dots$ and $\{\theta_m\} \in \Theta$ tending to some $\theta \in \Theta$, with Y_m^n and Y^n having the distributions indexed by θ_m and θ, respectively, as $m \to \infty$ with n fixed:*

$$T_n(\theta_m; Y_m^n) \rightsquigarrow T_n(\theta; Y^n).$$

Proof. Since $f_{\theta_m}^i \to f_\theta^i$ for every i pointwise by Assumption 2, $Y_n^n \rightsquigarrow Y^n$ (see Proof of Theorem 2.1). Thus, by Slutsky's theorem, $(\theta_m, Y_m^n) \rightsquigarrow (\theta, Y^n)$. The result now follows from the continuous mapping theorem and Theorem 2.1. □

Lemma 0.3. *If Assumption 2 holds, then $\mathbb{E}_\theta\{s_n(\theta; Y^n)\} = 0$ for all $\theta \in \Theta$.*

Proof. Let γ^n denote the product measure $\otimes_{i=1}^n \gamma_i$. Pick a sequence $t_m \downarrow 0$ as $m \to \infty$ and use the mean value theorem, applicable by Assumption 2, to write for some $\tilde{t}_m \in [0, t_m]$,

$$\begin{aligned}
0 &= t_m^{-1} \int \left\{ f_{\theta + t_m e_j}(y^n) - f_\theta(y^m) \right\} \gamma^n(\mathrm{d}y^n) \\
&= \int \nabla_j f_{\theta + \tilde{t}_m e_j}(y^n) \gamma^n(\mathrm{d}y^n) \\
&= \int s_{nj}(\theta + \tilde{t}_m e_j; y^n) f_{\theta + \tilde{t}_m e_j}(y^n) \gamma^n(\mathrm{d}y^n) \\
&= \mathbb{E}\{s_{nj}(\theta_m, Y_m^n)\},
\end{aligned}$$

where $\theta_m = \theta + \tilde{t}_m e_j \to \theta$ as $m \to \infty$ and Y_m^n has the distribution indexed by θ_m. By Slutsky's theorem, $(\theta_m, Y_m^n) \rightsquigarrow (\theta, Y^n)$, where Y^n has the distribution indexed by θ. Thus, by Assumption 2 and the continuous mapping theorem, $s_{nj}(\theta_m, Y_m^n) \rightsquigarrow s_{nj}(\theta, Y^n)$. Moreover, by Assumption 2 there exists an $M < \infty$ such that $\mathbb{E}\{s_{nj}(\theta_m; Y_m^n)^2\} \leq M$ for all large enough m. Thus, the sequence $\{s_{nj}(\theta_m, Y_m^n)\}$ is uniformly integrable and, consequently, $0 = \mathbb{E}\{s_{nj}(\theta_m; Y_m^n)\} \to \mathbb{E}\{s_{nj}(\theta; Y^n)\}$, which completes the proof. □

Proofs of results in main text

Proof of Lemma 2.2. The assumptions of the lemma imply $T_n(\cdot; \cdot)$ is continuous on $\{\theta : \mathcal{I}(\theta) > 0\} \times \mathcal{Y}^n$. They also say we may, for any critical θ and $y^n \in \mathcal{Y}^n$, unambiguously define $T_n(\theta; y^n) = \lim_{m \to \infty} T_n(\theta_m; y^n)$, where $\{\theta_m\}$ is any sequence of non-critical points tending to

θ; Assumption 5 says at least one such sequence exists. To verify this extension is continuous on $\Theta \times \mathcal{Y}^n$, let instead $\{\theta_m\} \in \Theta$ be an arbitrary sequence, possibly including critical points, tending to θ. Let also $\{y^n_m\} \in \mathcal{Y}^n$ be an arbitrary sequence tending to y^n. By the assumptions of the lemma, we can find, for every fixed m, a non-critical $\tilde{\theta}_m$ such that

$$|T_n(\theta_m; y^n_m) - T_n(\tilde{\theta}_m; y^n_m)| \leq 1/m \quad \text{and} \quad \|\theta_m - \tilde{\theta}_m\| \leq 1/m.$$

Thus, by the triangle inequality,

$$|T_n(\theta_m; y^n_m) - T_n(\theta; y^n)| \leq 1/m + |T_n(\tilde{\theta}_m; y^n_m) - T_n(\theta; y^n)|,$$

which tends to zero by the assumptions of the lemma since $\{\tilde{\theta}_m\}$ is a sequence of non-critical points tending to θ. $\square$

Proof of Lemma 2.5. We first prove Equation (5) implies Equation (4) in the main manuscript. For contradiction, suppose (5) holds and that there exist a compact $C \subseteq \Theta$ and an $\epsilon > 0$ such that, for infinitely many n, $\sup_{\theta \in C} |\mathsf{P}_\theta \{\theta \in \mathcal{R}_n(\alpha)\} - (1 - \alpha)| > \epsilon$. Let N be the set of such n and pick, for every $n \in N$, a $\theta_n \in C$ such that $|\mathsf{P}_{\theta_n} \{\theta_n \in \mathcal{R}_n(\alpha)\} - (1 - \alpha)| > \epsilon$ Because C is compact, it is bounded and hence $\{\theta_n : n \in N\}$ is a bounded sequence. Thus, it contains a convergent subsequence. But by (5), along this subsequence, $T_n(\theta_n; Y^n_n) \rightsquigarrow \chi^2_d$; in particular, since χ^2_d has a continuous cumulative distribution function, $\mathsf{P}_{\theta_n}\{\theta_n \in \mathcal{R}_n(\alpha)\} = \mathsf{P}\{T_n(\theta_n; Y^n_n) \leq q_{d,1-\alpha}\} \to 1 - \alpha$ along the subsequence, which is the desired contradiction.

To prove Equation (4) implies Equation (5) in the main manuscript, note that if $\theta_n \to \theta$, then for all large enough n, θ_n is in a compact neighborhood C of θ. Thus, for those n and any $\alpha \in (0, 1)$, $|\mathsf{P}\{T_n(\theta_n; Y^n_n) \leq q_{d,1-\alpha}\} - (1-\alpha)| = |\mathsf{P}_{\theta_n}\{\theta_n \in \mathcal{R}_n(\alpha)\} - (1-\alpha)| \leq \sup_{\theta \in C} \mathsf{P}_\theta\{\theta \in \mathcal{R}_n(\alpha) - (1 - \alpha)\}|$, which tends to zero by (4). Thus, since the range of $\alpha \mapsto q_{d,1-\alpha}$ is $\mathbb{R}$, the cumulative distribution function of $T_n(\theta_n; Y^n_n)$ tends to that of χ^2_d at every point in $\mathbb{R}$, which completes the proof. $\square$

Proof of Lemma 3.7. Differentiating $\log f_\theta(y_i)$ with respect to ψ gives $s^i_\psi(\theta; y_i, X_i) = X_i^\mathsf{T} \Sigma_i^{-1}(y_i - X_i\psi)$. Differentiating this with respect to λ and taking expectations shows $\mathcal{I}^i(\theta)$ is block-diagonal. The trailing $d_2 \times d_2$ block is $\mathcal{I}^i_\psi(\theta) = \mathrm{cov}_\theta\{s^i_\psi(\theta; Y_i, X_i)\} = \mathbb{E}(X_i^\mathsf{T} \Sigma_i^{-1} X_i)$, which is positive definite for all θ since $\underline{\mathrm{e}}(\Sigma) \geq \sigma^2 > 0$ and $\underline{\mathrm{e}}\{\mathbb{E}(X_i^\mathsf{T} X_i)\} > 0$ by assumption; the result follows. $\square$

Proof of Lemma 3.8. For $j = 1, \ldots, d_1$, let

$$\zeta_j^i(\theta; y_i, X_i) = \operatorname{tr}\left\{\Sigma_i^{-1} H_j^i - \Sigma_i^{-1}(y_i - X_i\psi)(y_i - X_i\psi)^\mathsf{T}\Sigma_i^{-1} H_j^i\right\},$$

which are the first d_1 elements of $\xi^i(\theta; y_i, X_i)$ defined in the proof of Theorem 3.9. In particular, for $\lambda_j > 0$, $\zeta_j^i(\theta; y_i, X_i) = s_j^i(\theta; y_i, X_i)/\lambda_j$, and hence the claim to be proved is equivalent to, for any $v \in \mathbb{R}^{d_1}$,

$$\bar{\mathrm{e}}(\Sigma_i)^{-2}\, \underline{\mathrm{e}}(Z_i^\mathsf{T} Z_i)^2 \max_j (v_j)^2 \leq \frac{1}{2}\operatorname{var}_\theta\{v^\mathsf{T}\zeta^i(\theta; Y_i, X_i) \mid X_i\} \leq r_i\, \underline{\mathrm{e}}(\Sigma_i)^{-2}\, \bar{\mathrm{e}}(Z_i^\mathsf{T} Z_i)^2 \max_j (v_j)^2.$$

With $G_i = \sum_{j=1}^{d_1} v_j H_j^i \in \mathbb{R}^{r_i \times r_i}$, we have

$$v^\mathsf{T}\zeta^i(\theta; Y_i, X_i) = \operatorname{tr}\left[\left\{\Sigma_i^{-1} - \Sigma_i^{-1}(Y_i - X_i\psi)(Y_i - X_i\psi)^\mathsf{T}\Sigma_i^{-1}\right\} G_i\right].$$

Thus, applying the well-known expression for the variance of a quadratic form in multivariate normal vectors (Seber and Lee, 2003, Theorem 1.6),

$$\begin{aligned}
\operatorname{var}_\theta\left\{v^\mathsf{T}\zeta^i(\theta; Y_i, X_i) \mid X_i\right\} &= \operatorname{var}_\theta\left[\operatorname{tr}\left\{\Sigma_i^{-1}(Y_i - X_i\psi)(Y_i - X_i\psi)^\mathsf{T}\Sigma_i^{-1} G_i\right\} \mid X_i\right] \\
&= \operatorname{var}_\theta\left[(Y_i - X_i\psi)^\mathsf{T}\Sigma_i^{-1} G_i \Sigma_i^{-1}(Y_i - X_i\psi) \mid X_i\right] \\
&= 2\operatorname{tr}[(\Sigma_i^{-1/2} G_i \Sigma_i^{-1/2})^2].
\end{aligned}$$

We start with the lower bound. Observe that since $\Sigma_i^{-1/2} G_i \Sigma_i^{-1/2}$ is symmetric, its eigenvalues are real, and hence the eigenvalues of its square are non-negative as the squares of real numbers. Thus, the trace upper bounds the maximum eigenvalue, and hence $\operatorname{var}_\theta\left\{v^\mathsf{T}\zeta^i(\theta; Y_i, X_i) \mid X_i\right\}$ is lower bounded by

$$2\|(\Sigma_i^{-1/2} G_i \Sigma_i^{-1/2})^2\| \geq 2\, \underline{\mathrm{e}}(\Sigma_i^{-1})\|\Sigma_i^{-1/2} G_i^2 \Sigma_i^{-1/2}\| \geq 2\, \underline{\mathrm{e}}(\Sigma_i^{-1})^2\, \bar{\mathrm{e}}(G_i^2).$$

Now write

$$G_i = \sum_{j=1}^{d_1}\sum_{k \in [j]} v_j Z_i^k (Z_i^k)^\mathsf{T} = \sum_{k=1}^{q} v_{j(k)} Z_i^k (Z_i^k)^\mathsf{T} = Z_i \tilde{V} Z_i^\mathsf{T},$$

where $v_{j(k)}$ is the v_j scaling $Z_i^k (Z_i^k)^\mathsf{T}$ in the double sum and $\tilde{V}$ is diagonal with the elements of v on the diagonal, ordered so that the last equality holds. Then $\|G_i^2\| = \|Z_i\tilde{V} Z_i^\mathsf{T} Z_i \tilde{V} Z_i^\mathsf{T}\| \geq \underline{\mathrm{e}}(Z_i^\mathsf{T} Z_i)\|Z_i \tilde{V}^2 Z_i^\mathsf{T}\|$. The last norm is $\|Z_i \tilde{V}^2 Z_i^\mathsf{T}\| = \bar{\mathrm{e}}(Z_i \tilde{V}^2 Z_i^\mathsf{T}) = \max_{\|b\|=1} b^\mathsf{T} Z_i \tilde{V}^2 Z_i^\mathsf{T} b$ which by considering $b = Z_i^l/\|Z_i^l\|$ is lower bounded by, for every $l = 1, \dots, q$,

$$\left(\frac{Z_i^l}{\|Z_i^l\|}\right)^{\mathsf{T}} \sum_{k=1}^{q} v_{j(k)}^2 Z_i^k (Z_i^k)^{\mathsf{T}} \left(\frac{Z_i^l}{\|Z_i^l\|}\right) \geq v_{j(l)} \|Z_i^l\|^2 \geq v_{j(l)}^2 \min_k \|Z_i^k\|^2.$$

Thus, because it holds for every l it holds for $l \in \arg\max_k v_{j(k)}^2$, and the proof of the lower bound is completed by observing $\|Z_i^k\|^2 = e_k^{\mathsf{T}} Z_i^{\mathsf{T}} Z_i e_k \geq \underline{\mathrm{e}}(Z_i^{\mathsf{T}} Z_i)$. For the upper bound, note

$$2\,\mathrm{tr}[(\Sigma_i^{-1/2} G_i \Sigma_i^{-1/2})^2] \leq 2r_i \|\Sigma_i^{-1}\|^2 \|G_i\|^2 \leq 2r_i \|\Sigma_i^{-1}\|^2 \|Z_i\|^4 \|\tilde{V}\|^2,$$

which is equal to the stated upper bound and hence the proof is completed. □

Proof of Theorem 3.10. The claim about $\mathcal{R}_n^\lambda(\alpha)$ is almost immediate from Theorem 3.9 and the fact that $\mathcal{I}_n(\theta)$ is block diagonal so we omit the proof. To prove the second claim, observe that $T_n^\lambda(\lambda; \psi, Y^n, X^n)$ is equal to

$$\left\{n^{-1/2} \sum_{i=1}^{n} \zeta^i(\theta; Y_i, X_i)^{\mathsf{T}}\right\} \mathrm{cov}_\theta \left\{\zeta^1(\theta; Y_1, X_1)\right\}^{-1} \left\{n^{-1/2} \sum_{i=1}^{n} \zeta^i(\theta; Y_i, X_i)\right\},$$

where ζ^i is defined in the proof of Lemma 3.8. Let $C(\theta)$ be the covariance matrix in the middle term. We showed in the proof of Lemma 3.8 that $C(\theta)$ is positive definite at any θ; in particular, $v^{\mathsf{T}} C(\theta) v \geq 2 \max_j v_j^2\, \bar{\mathrm{e}}(\Sigma_1)^{-2}\, \underline{\mathrm{e}}(Z_1^{\mathsf{T}} Z_1)^2$. Moreover, it is straightforward to show C is continuous using uniform integrability of $\{\zeta^1(\theta_m; Y_{m1}, X_{m1})\zeta^1(\theta_m; Y_{m1}, X_{m1})^{\mathsf{T}}\}$, where (Y_{m1}, X_{m1}) has the distribution indexed by a θ_m tending to some θ; the arguments are very similar to those in the proof of Theorem 2.1 and hence omitted. Thus, it suffices (Billingsley, 1999, Theorems 2.7 and 3.1) to show

$$\left\| n^{-1/2} \sum_{i=1}^{n} \zeta^i(\lambda_n; \psi_n, Y_{ni}, X_{ni}) - n^{-1/2} \sum_{i=1}^{n} \zeta^i(\lambda_n; \hat{\psi}_n, Y_{ni}, X_{ni}) \right\| = o_{\mathsf{P}}(1),$$

where (Y_{ni}, X_{ni}) has the distribution indexed by θ_n. We show the equivalent result that every element of the vector in the norm is $o_{\mathsf{P}}(1)$. The jth element is

$$n^{-1/2} \sum_{i=1}^{n} \left\{(Y_{ni} - X_{ni}\psi_n)^{\mathsf{T}} \Omega_{nj} (Y_{ni} - X_{ni}\psi_n) - (Y_{ni} - X_{ni}\hat{\psi}_n)^{\mathsf{T}} \Omega_{nj} (Y_{ni} - X_{ni}\hat{\psi}_n)\right\},$$

where $\Omega_{nj} = \Sigma_n^{-1} H_j \Sigma_n^{-1}$. Let $\varepsilon_{ni} = Y_{ni} - X_{ni}\psi_n \sim \mathcal{N}(0, \Sigma_n)$ to get that the last display is equal to

$$n^{-1/2}\sum_{i=1}^{n}\left[\varepsilon_{ni}^{\mathsf{T}}\Omega_{nj}\varepsilon_{ni}-\{\varepsilon_{ni}+X_{ni}(\psi_n-\hat{\psi}_n)\}^{\mathsf{T}}\Omega_{nj}\{\varepsilon_{ni}+X_{ni}(\psi_n-\hat{\psi}_n)\}\right],$$

which in turn is equal to

$$-n^{-1/2}2(\psi_n-\hat{\psi}_n)^{\mathsf{T}}\sum_{i=1}^{n}X_{ni}^{\mathsf{T}}\Omega_{nj}\varepsilon_{ni}-n^{-1/2}(\psi_n-\hat{\psi}_n)^{\mathsf{T}}\left(\sum_{i=1}^{n}X_{ni}^{\mathsf{T}}\Omega_{nj}X_{ni}\right)(\psi_n-\hat{\psi}_n).$$

Thus, since $\|\psi_n-\hat{\psi}_n\|=O_{\mathsf{P}}(1/\sqrt{n})$ it suffices to show that

$$\left\|n^{-1}\sum_{i=1}^{n}X_{ni}^{\mathsf{T}}\Omega_{nj}\varepsilon_{ni}\right\|=o_{\mathsf{P}}(1)\quad\text{and}\quad\left\|n^{-1}\sum_{i=1}^{n}X_{ni}^{\mathsf{T}}\Omega_{nj}X_{ni}\right\|=O_{\mathsf{P}}(1).$$

For the former we show the elements are $o_{\mathsf{P}}(1)$ and for the latter it suffices, since the matrix in the norm is positive semi-definite, to show the diagonal elements are $O_{\mathsf{P}}(1)$. First, then, condition on $\{X_1,\dots,X_n\}$, apply Chebyshev's inequality, and take expectations to get, for any $s\geq 0$ and standard basis vector e_l,

$$\begin{aligned}\mathsf{P}\left\{\left|e_l^{\mathsf{T}}n^{-1}\sum_{i=1}^{n}X_{ni}^{\mathsf{T}}\Omega_{nj}\varepsilon_{ni}\right|\geq s\right\}&\leq\frac{1}{s^2n^2}\mathbb{E}\left(\sum_{i=1}^{n}e_l^{\mathsf{T}}X_{ni}^{\mathsf{T}}\Omega_{nj}\Sigma_n\Omega_{nj}X_{ni}e_l\right)\\&\leq\frac{1}{s^2n}\|H_j\|^2\sigma^{-6}\mathbb{E}(\|X_1\|^2),\end{aligned}$$

which tends to zero since $\|H_j\|\leq\|Z^{\mathsf{T}}Z\|$ and $\mathbb{E}(\|X_1\|^2)$ are bounded by assumption. The second holds since, for any standard basis vector e_l, $e_l^{\mathsf{T}}X_{ni}^{\mathsf{T}}\Omega_{nj}X_{ni}e_l\leq\|\Omega_{nj}\|e_l^{\mathsf{T}}X_{ni}^{\mathsf{T}}X_{ni}e_l$, $\|\Omega_{nj}\|\leq\|H_j\|\|\Sigma^{-1}\|\leq\|H_j\|\sigma^{-4}$, and $n^{-1}\sum_{i=1}^{n}e_l^{\mathsf{T}}X_{ni}^{\mathsf{T}}X_{ni}e_l\to e_l^{\mathsf{T}}\mathbb{E}(X_1^{\mathsf{T}}X_1)e_l<\infty$ by the law of large numbers. $\square$

Additional simulations

Figure 1 summarizes the results of a Monte Carlo experiment with 10,000 replications and compares coverage probabilities for our method with ψ known and ψ estimated, i.e. a nuisance parameter. Other than that, the settings are like those for producing Figure 2 in the main text. Notably, coverage is slightly higher when estimating ψ. In general, however, both methods have near-nominal coverage and the differences are especially small for larger n.

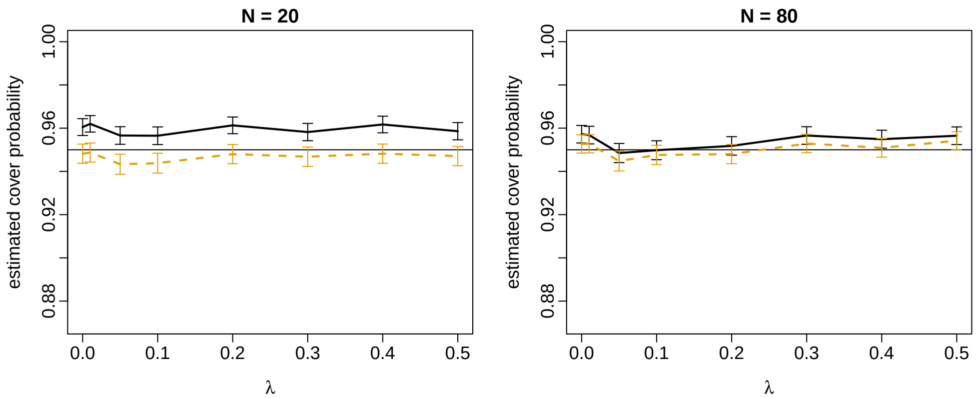

Figure 1: Monte Carlo estimates of coverage probabilities of confidence regions from inverting the modified score with ψ estimated (solid) or known (dashed). The straight horizontal line indicates the nominal 0.95 coverage probability and vertical bars denote ± 2 times Monte Carlo standard errors.

Figure 2 here is also similar to Figure 2 in the main text, but here σ is treated as unknown and the boundary point $(\lambda_1, \lambda_2, \sigma) = (0, 0, 1)$ is included instead of the near-boundary point $(\lambda_1, \lambda_2) = (10^{-12}, 10^{-12})$. In the simulations, the unknown $\sigma = 1$ and coverage probabilities are for a range of $(\lambda_1, \lambda_2, \sigma) = (\lambda_1, \lambda_2, 1)$ where the values of $\lambda_1 = \lambda_2$ are on the horizontal axis in Figure 2. All confidence regions use the chi-square distribution with 3 degrees of freedom as reference. Notably, the proposed method's coverage probability of the boundary point is indistinguishable from that of the near-boundary point presented in Figure 2 in the main text. As noted in the main text, the other methods' coverage probabilities for the boundary point could be improved by using a different reference distribution, motivated by asymptotic theory with a fixed, true parameter at the boundary.

References

Billingsley, P. (1999). *Convergence of Probability Measures.* Wiley series in probability and statistics. Probability and statistics section. Wiley, New York, second edition.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis.* Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J, 2nd ed edition.

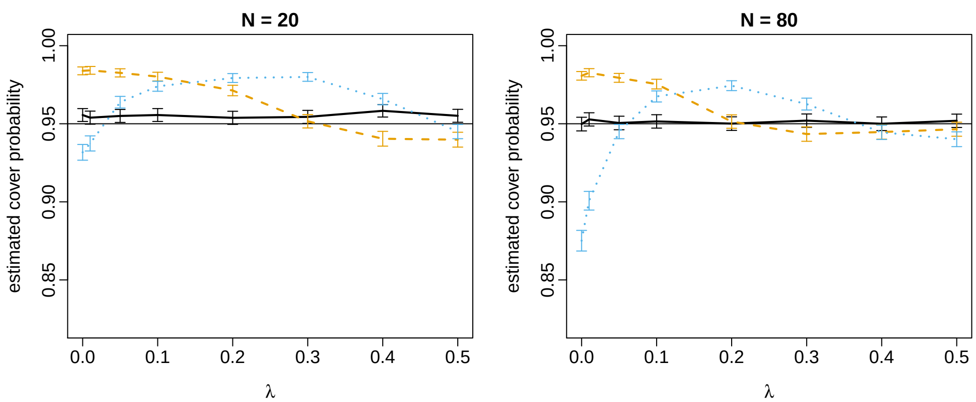

Figure 2: Monte Carlo estimates of coverage probabilities of confidence regions from inverting the modified score (solid), likelihood ratio (dashed), and Wald (dotted) test statistics. The straight horizontal line indicates the nominal 0.95 coverage probability and vertical bars denote ± 2 times Monte Carlo standard errors.