Supplementary Material for: Modified Likelihood Root in High Dimensions

Yanbo Tang
University of Toronto, Toronto, Canada.
Vector Institute, Toronto, Canada.
E-mail: yanbo@utstat.toronto.edu
Nancy Reid
University of Toronto, Toronto, Canada.
E-mail: reid@utstat.toronto.edu

1. Assumptions

A list of the Assumptions used in the main text, reproduced here for convenience. We assume that $p = O(n^\alpha)$ for some $0 \leq \alpha < 1/2$. Let $N_{\theta_0,\delta} = \{\theta : \|\theta - \theta_0\|_2 < \delta\}$ for $\delta > 0$ denote a neighbourhood of radius δ centered around θ_0.

1.1. Assumptions in §3

ASSUMPTION 1. $\|\hat{\theta} - \theta_0\|_2 = o_p(1)$ *and* $\sup_{\psi \in A_n} \|\hat{\theta}_\psi - \theta_0\|_2 = o_p(1)$, *where* $A_n = \{\psi : |\psi - \psi_0| \leq |\hat{\psi} - \psi_0|\}$.

ASSUMPTION 2. $j_{\psi\lambda_r}(\theta) = O_p(n^{1/2})$ *uniformly in* r, *for* $\theta \in N_{\theta_0,\delta}$.

ASSUMPTION 3. *The eigenvalues of* $j(\theta)/n$ *and* $\{j(\theta)/n\}^{-1}$ *are bounded in probability, for* $\theta \in N_{\theta_0,\delta}$.

ASSUMPTION 4. *The log-likelihood derivatives* $l_{\theta_r\theta_s\theta_t}(\theta)$, $l_{\theta_r\theta_s\theta_t\theta_o}(\theta)$ *and* $l_{\theta_r\theta_s;\hat{\theta}_t}(\theta)$ *are continuous and uniformly* $O_p(n)$ *in* r, s, t, o, *for* $\theta \in N_{\theta_0,\delta}$.

ASSUMPTION 5. *The log-likelihood root,* $r \xrightarrow{D} Z$, *for some random variable* Z, *whose distribution has no point mass at* 0. *The Wald statistic* $t = j_{\mathrm{p}}^{1/2}(\hat{\psi})(\hat{\psi} - \psi_0) \xrightarrow{D} \tilde{Z}$ *for some random variable* $\tilde{Z}$.

Assumptions in §5.1

ASSUMPTION 6. *The eigenvalues of the Gram matrix satisfy* $0 < a_1 n < \eta_i(X^T X) < a_2 n < \infty$, *and* $\sum_{i=1}^n x_{ij}x_{ik} = O(n)$ *for each* j, k *in* $(1, \ldots, p)$.

ASSUMPTION 7. $\max_{i=1,\ldots,n} K''(x_i^\top \theta) = O(1)$, $\max_{i=1,\ldots,n} \left\{K''(x_i^\top \theta)\right\}^{-1} = O(1)$, *and* $\Sigma_i K'''(x_i^\top \theta) x_{i1}^3 = O(n)$ *for* $\theta \in N_{\theta_0,\delta}$.

ASSUMPTION 8. *The third log-likelihood derivative* $l_{\psi\psi\psi}(\theta) = O_p(n)$, *for* $\theta \in N_{\theta_0,\delta}$.

ASSUMPTION 9. *The derivative of the observed Fisher information matrix under the* (ψ, τ) *parameterization with respect to* ψ *satisfies* $\|j_{\psi\tau\tau}(\theta)\|_{op} = O_p(n)$, *for* $\theta \in N_{\theta_0,\delta}$.

1.2. Assumptions in §5.2

ASSUMPTION 10. $\max_{j=1,\ldots,p} \left\| j_{\theta_j\lambda\lambda}(\theta) \right\|_{op} = O_p(n)$, *for* $\theta \in N_{\theta_0,\delta}$.

2. Preliminary Results

LEMMA 1. *Under Assumptions 1–3,*

$$\left\| \frac{\partial \hat{\lambda}_\psi}{\partial \psi} |_{\psi=\tilde{\psi}} \right\|_2 = O_p\left(\frac{p^{1/2}}{n^{1/2}}\right).$$

PROOF. By differentiating the score equation for the constrained maximum likelihood estimator, $l_\lambda(\hat{\theta}_\psi) = 0$, we obtain

$$j_{\lambda\lambda}(\hat{\theta}_\psi)\frac{\partial \hat{\lambda}_\psi}{\partial \psi} = -j_{\psi\lambda}(\hat{\theta}_\psi), \tag{1}$$

so

$$\frac{\partial \hat{\lambda}_\psi}{\partial \psi} = -j_{\lambda\lambda}^{-1}(\hat{\theta}_\psi)j_{\psi\lambda}(\hat{\theta}_\psi).$$

From this we have

$$\begin{aligned}\left\| \frac{\partial \hat{\lambda}_\psi}{\partial \psi} |_{\psi=\tilde{\psi}} \right\|_2 &= [j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}})\{j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}})\}^2 j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}})]^{1/2} \le \left(\left\| j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op}^2 \left\| j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_2^2 \right)^{1/2}, \\ &= \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} \left\| j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_2 = O_p\left(\frac{p^{1/2}}{n^{1/2}}\right),\end{aligned}$$

where the inequality follows from an application of Rayleigh quotient, and we have used Assumptions 1 and 3. The final equality follows from

$$\left\| j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_2 = \left\{ \sum_{j=2}^{p} j_{\psi\lambda_j}(\hat{\theta}_{\tilde{\psi}}) \right\}^{1/2} = O_p\{(pn)^{1/2}\},$$

which is the sum of $p-1$ elements that are uniformly $O_p(n^{1/2})$ by Assumptions 1 and 2.

PROPOSITION 1. *Under Assumptions 1–4,*

(i) $\left\| j_{\psi\lambda(\hat\theta_{\tilde\psi})} \right\|_2 = O_p\{(pn)^{1/2}\}$, *(ii)* $\left\| j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\|_F = O_p\left(\frac{p^{1/2}}{n}\right)$,

(iii) $\left\| l_{\psi\lambda\lambda}(\hat\theta_{\tilde\psi}) \right\|_F = O_p(pn)$, $\left\| l_{\lambda\lambda;\hat\psi}(\hat\theta_{\tilde\psi}) \right\|_F = O_p(pn)$,

(iv) $\left\| \frac{d}{d\psi} l_{\lambda;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\tilde\psi}} \right\|_F = O_p(pn)$, $\left\| \frac{d}{d\psi} j_{\lambda\lambda}(\hat\theta_\psi)|_{\hat\theta_{\tilde\psi}} \right\|_F = O_p(pn)$,

(v) $\left\| l_{\psi\psi\lambda}(\hat\theta_{\tilde\psi}) \right\|_2 = O_p(p^{1/2}n)$, $\left\| l_{\lambda\psi;\hat\psi}(\hat\theta_{\tilde\psi}) \right\|_2 = O_p(p^{1/2}n)$, $\left\| l_{\lambda;\hat\psi}(\hat\theta_{\tilde\psi}) \right\|_2 = O_p(p^{1/2}n)$,

(vi) $\left\| \frac{d}{d\psi} l_{\lambda;\hat\psi}(\hat\theta_\psi)|_{\hat\theta_{\tilde\psi}} \right\|_2 = O_p(p^{1/2}n)$, $\left\| \frac{d}{d\psi} l_{;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\tilde\psi}} \right\|_2 = O_p(p^{1/2}n)$.

Note that since the Frobenius norm is an upper bound for the maximum singular value, the rates above also apply to the maximum singular values.

Proof. *i)* This result is obtained as an intermediate step in the proof of Lemma 1.
ii) By Assumptions 1 and 3,

$$\left\| j_{\lambda\lambda}^{-1}(\hat\theta) \right\|_F = \mathrm{Tr}\left[\left\{ j_{\lambda\lambda}^{-1}(\hat\theta) \right\}^2 \right]^{1/2} \le p^{1/2} \left\| \left\{ j_{\lambda\lambda}^{-1}(\hat\theta) \right\} \right\|_2 = O_p\left(\frac{p^{1/2}}{n} \right). \tag{2}$$

iii) By Assumptions 1 and 4, the elements of these matrices are uniformly $O_p(n)$ giving

$$\left\| l_{\lambda\lambda;\hat\psi}(\hat\theta_{\tilde\psi}) \right\|_F = \left\{ \sum_{r=1}^{p-1} \sum_{s=1}^{p-1} l_{\lambda_r\lambda_s;\hat\psi}(\hat\theta_{\tilde\psi})^2 \right\}^{1/2} = O_p(pn).$$

The same argument applies to the other matrix $l_{\psi\lambda\lambda}$.
iv) By the chain rule and the triangle inequality we have

$$\begin{aligned}\left\| \frac{d}{d\psi} l_{\lambda;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\hat\psi}} \right\|_F &= \left\| l_{\psi\lambda;\hat\lambda}(\hat\theta_{\tilde\psi}) + \sum_{j=1}^{p-1} l_{\lambda_j\lambda;\hat\lambda}(\hat\theta_{\tilde\psi}) \frac{\partial\hat\lambda_{\psi,j}}{\partial\psi}|_{\tilde\psi} \right\|_F, \\ &\le \left\| l_{\psi\lambda;\hat\lambda}(\hat\theta_{\tilde\psi}) \right\|_F + \left\| \sum_{j=1}^{p-1} l_{\lambda_j\lambda;\hat\lambda}(\hat\theta_{\tilde\psi}) \frac{\partial\hat\lambda_{\psi,j}}{\partial\psi}|_{\tilde\psi} \right\|_F.\end{aligned}$$

By Proposition 3 iii), $\left\| l_{\psi\lambda;\hat\lambda}(\hat\theta_{\tilde\psi}) \right\|_F = O_p(pn)$. We now obtain the order of the second term by considering the order of each entry. The absolute value of the (r,s) entry of the matrix is

$$\left| \sum_{j=1}^{p-1} \frac{\partial\hat\lambda_{\psi,j}}{\partial\psi}|_{\tilde\psi}\, l_{\lambda_j\lambda_r;\hat\lambda_s}(\hat\theta_{\tilde\psi}) \right| \le \left\| \frac{\partial\hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} \right\|_2 \left\| l_{\lambda\lambda_r;\hat\lambda_s}(\hat\theta_{\tilde\psi}) \right\|_2 = O_p(pn^{1/2}) \le O_p(n),$$

where we have used the Cauchy-Schwartz inequality, Lemma 1, and $\left\| l_{\lambda\lambda_r;\hat{\lambda}_s}(\hat{\theta}_{\tilde{\psi}}) \right\|_F = O_p(p^{1/2}n)$, since it is the Euclidean norm of a vector of length $p-1$ whose elements are uniformly $O_p(n)$ by Assumptions 1 and 4. Therefore

$$\left\| \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \; l_{\lambda_j\lambda;\hat{\lambda}}(\hat{\theta}_{\tilde{\psi}}) \right\|_F = O_p(pn), \tag{3}$$

which proves the result. The proof for the other matrix is similar.

v) This follows as we are computing the Euclidean norm of a vector of length $p-1$ whose entries are uniformly $O_p(n)$ by Assumptions 1 and 4.

vi) This can be proved in the same manner as Proposition 3 iv); we simply need to note that we are now taking the Euclidean norm of a vector of length $p-1$, hence the factor of $p^{1/2}$ instead of p.

LEMMA 2. *Under Assumptions 1–4*

$$\left\| \frac{\partial^2 \hat{\lambda}_\psi}{\partial \psi^2}|_{\psi=\tilde{\psi}} \right\|_2 = O_p(p^{1/2}).$$

PROOF. We obtain an expression for the second derivative of the constrained maximum likelihood estimate by differentiating (1),

$$\left\{ \frac{d}{d\psi} j_{\lambda\lambda}(\hat{\theta}_\psi) \right\} \frac{\partial \hat{\lambda}_\psi}{\partial \psi} + j_{\lambda\lambda}(\hat{\theta}_\psi) \frac{\partial^2 \hat{\lambda}_\psi}{\partial \psi^2} = l_{\psi\psi\lambda}(\hat{\theta}_\psi) + l_{\psi\lambda\lambda}(\hat{\theta}_\psi) \frac{\partial \hat{\lambda}_\psi}{\partial \psi}. \tag{4}$$

Substituting in the expression for the first derivative and rearranging terms we obtain

$$\frac{\partial^2 \hat{\lambda}_\psi}{\partial \psi^2}|_\psi = j_{\lambda\lambda}^{-1}(\hat{\theta}_\psi) \Big[l_{\psi\psi\lambda}(\hat{\theta}_\psi) - l_{\psi\lambda\lambda}(\hat{\theta}_\psi) j_{\lambda\lambda}^{-1}(\hat{\theta}_\psi) j_{\psi\lambda}(\hat{\theta}_\psi) + \left\{ \frac{d}{d\psi} j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\hat{\theta}_\psi} \right\} j_{\lambda\lambda}^{-1}(\hat{\theta}_\psi) j_{\psi\lambda}(\hat{\theta}_\psi) \Big].$$

Thus

$$\begin{aligned}
\left\| \frac{\partial^2 \hat{\lambda}_\psi}{\partial \psi^2}|_{\tilde{\psi}} \right\|_2 &= \Big[l_{\lambda\psi\psi}(\hat{\theta}_{\tilde{\psi}}) \Big\{ j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \Big\}^2 l_{\psi\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \Big]^{1/2} + \Big[j_{\lambda\psi}(\hat{\theta}_{\tilde{\psi}}) \Big\{ j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) l_{\psi\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \Big\}^2 j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \Big]^{1/2} \\
&+ \Big(j_{\lambda\psi}(\hat{\theta}_{\tilde{\psi}}) \Big[j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \Big\{ \frac{d}{d\psi} j_{\lambda\lambda}(\hat{\theta}_\psi)|_{\tilde{\psi}} \Big\} j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}}) \Big]^2 j_{\psi\lambda}(\hat{\theta}_{\tilde{\psi}}) \Big)^{1/2}, \\
&= A_1 + A_2 + A_3,
\end{aligned}$$

The orders of A_1, A_2 and A_3 are obtained by combining the Rayleigh quotient with Assumptions 1 and 3, Proposition 1, and by noting that $p/n^{1/2} = o(1)$.

$$A_1 \le \left\| l_{\lambda\psi\psi}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\|_{op} = O_p(p^{1/2} n) O_p\left(n^{-1}\right) = O_p(p^{1/2}).$$

$$A_2 \le \left\| j_{\lambda\psi}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| \left\{ j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) l_{\psi\lambda\lambda}(\hat\theta_{\tilde\psi}) j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\} \right\|_{op} \le \left\| j_{\lambda\psi}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\|_{op}^2 \left\| l_{\psi\lambda\lambda}(\hat\theta_{\tilde\psi}) \right\|_{op},$$

$$= O_p\{(pn)^{1/2}\} O_p\left(\frac{1}{n^2}\right) O_p(pn) = O_p\left(\frac{p^{3/2}}{n^{1/2}}\right) \le O_p(p^{1/2}).$$

$$A_3 = \left\| j_{\lambda\psi}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \left\{ \frac{d}{d\psi} j_{\lambda\lambda}(\hat\theta_\psi)|_{\tilde\psi} \right\} j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\|_{op} = \left\| j_{\lambda\psi}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| j_{\lambda\lambda}^{-1}(\hat\theta_{\tilde\psi}) \right\|_{op}^2 \left\| \frac{d}{d\psi} j_{\lambda\lambda}(\hat\theta_\psi)|_{\tilde\psi} \right\|_{op},$$

$$= O_p\{(pn)^{1/2}\} O_p\left(n^{-2}\right) O_p(pn) = O_p\left(\frac{p^{3/2}}{n^{1/2}}\right) = O_p(p^{1/2}),$$

LEMMA 3. *Under Assumptions 1–5*

$$t = r\left\{1 + O_p\left(n^{-1/2}\right)\right\}.$$

PROOF.

$$\begin{aligned} r^2 &= 2\Big\{ l_{\rm p}(\hat\psi) - l_{\rm p}(\psi_o) \Big\}, \\ &= 2\Big\{ l_{\rm p}(\hat\psi) - l_{\rm p}(\hat\psi) + (\hat\psi - \psi_0)\zeta_1(\hat\psi) - \frac{(\hat\psi - \psi_0)^2}{2}\zeta_2(\hat\psi) + \frac{(\hat\psi - \psi_0)^3}{6}\zeta_3(\tilde\psi) \Big\}, \\ &= t^2 \left\{ 1 + \frac{t}{3} \frac{\zeta_3(\tilde\psi)}{j_{\rm p}^{3/2}(\hat\theta)} \right\}, \end{aligned}$$

where $t = j_{\rm p}^{1/2}(\hat\psi)(\hat\psi - \psi_0)$, and $\tilde\psi$ lies on the line segment between ψ_0 and $\hat\psi$. We now show that $\zeta_3(\tilde\psi) = O_p(n)$. By differentiating the profile log-likelihood,

$$\left| \frac{d^3}{d\psi^3} l_{\rm p}(\tilde\psi) \right| = \left| l_{\psi\psi\psi}(\hat\theta_{\tilde\psi}) + 2 l_{\psi\psi\lambda}(\hat\theta_{\tilde\psi}) \frac{\partial \hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} - j_{\psi\lambda}(\hat\theta_{\tilde\psi}) \frac{\partial^2 \hat\lambda_\psi}{\partial\psi^2}|_{\tilde\psi} + \left(\frac{\partial \hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} \right)^T l_{\psi\lambda\lambda}(\hat\theta_{\tilde\psi}) \frac{\partial \hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} \right|, \tag{5}$$

$$\le \left| l_{\psi\psi\psi}(\hat\theta_{\tilde\psi}) \right| + 2 \left\| l_{\psi\psi\lambda}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| \frac{\partial \hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} \right\|_2 + \left\| j_{\psi\lambda}(\hat\theta_{\tilde\psi}) \right\|_2 \left\| \frac{\partial^2 \hat\lambda_\psi}{\partial\psi^2}|_{\tilde\psi} \right\|_2 \tag{6}$$

$$+ \left\| l_{\psi\lambda\lambda}(\hat\theta_{\tilde\psi}) \right\|_{op} \left\| \frac{\partial \hat\lambda_\psi}{\partial\psi}|_{\tilde\psi} \right\|_2^2 = O_p(n) + O_p\{(np)^{1/2}\} + O_p(pn^{1/2}) + O_p(p^2) \tag{7}$$

$$\le O_p(n), \tag{8}$$

from an application of Lemma 1 and Lemma 2, as well as Proposition 1. This implies that

$$\frac{\zeta_3(\tilde\psi)}{j_{\rm p}^{3/2}(\hat\theta)} = O_p\left(n^{-1/2}\right),$$

combining this with Assumption 5, we have

$$r = t\left\{1 + O_p\left(n^{-1/2}\right)\right\}^{1/2},$$

which further implies that,

$$t = r\left\{1 + O_p\left(n^{-1/2}\right)\right\}^{-1/2} = r\left\{1 + O_p\left(n^{-1/2}\right)\right\},$$

by the following inequality for square roots,

$$1 - x/2 - x^2/2 \leq (1-x)^{1/2} \leq 1 - x/2,$$

and combining the above with

$$\left\{1 + O_p\left(n^{-1/2}\right)\right\}^{-1} = \left\{1 + O_p\left(n^{-1/2}\right)\right\}.$$

LEMMA 4. *Under Assumption 5, $r^{-1} = O_p(1)$.*

PROOF. Since $r \xrightarrow{D} Z$ for some random variable Z, the continuous mapping theorem implies that $r^{-1} \xrightarrow{D} Z'$ for some random variable Z' since $P(Z = 0) = 0$. Prokhorov's theorem then implies that sequence r^{-1} is tight and therefore bounded in probability.

3. Proof of Theorem 1 and Corollary 1

LEMMA 5. *Under Assumptions 1–5,*

$$\sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op}^{k} / (k+2) = O_p(1),$$

where

$$R_1 = -(\hat{\psi} - \psi_0)\frac{d}{d\psi} l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|_{\hat{\theta}_{\bar{\psi}}} = -\frac{t}{j_p^{1/2}(\hat{\psi})}\frac{d}{d\psi} l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|_{\hat{\theta}_{\bar{\psi}}}.$$

PROOF. Note that

$$\left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op} \leq \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op} \| R_1 \|_{op} = O_p(n^{-1}) O_p(pn^{1/2}) = O_p\left(\frac{p}{n^{1/2}}\right), \tag{9}$$

by Proposition 1, noting that by Assumptions 1 and 3, $j_{\mathrm{p}}^{-1/2}(\hat{\psi}) = O_p(n^{-1/2})$ and by Assumption 5, $t = O_p(1)$. Therefore for every fixed $\epsilon > 0$, there exists an M such that for all $n \geq n_0$

$$P\left(\left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op} \leq \frac{Mp}{n^{1/2}} \right) \geq 1 - \epsilon. \tag{10}$$

Therefore with probability greater than $1 - \epsilon$,

$$\sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op}^{k} / (k+2) \leq \sum_{k=0}^{\infty} \frac{1}{k+2} \left\{ \frac{Mp}{n^{1/2}} \right\}^{k}.$$

By assumption $p/n^{1/2} \to 0$, so that there exists an n_1 such that for all $n \geq n_1$, $Mp/n^{1/2} \leq 1/2$ and

$$\sum_{k=0}^{\infty} \frac{1}{(k+2)} \left\{ \frac{Mp}{n^{1/2}} \right\}^k \leq 2,$$

which implies that for arbitrary $\epsilon > 0$, there exists an $n' = \max(n_0, n_1)$ such that for all $n \geq n'$

$$P\Big(\sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op}^k /(k+2) \leq 2 \Big) \geq 1 - \epsilon.$$

LEMMA 6. *Under Assumptions 1–5,*

$$\log(|I + j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1|) = O_p \left\{ \max\left(\frac{p^{3/2}}{n^{1/2}}, \frac{p^3}{n} \right) \right\}.$$

PROOF. We express $\log(|I + j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1|)$ as a trace,

$$\log(|I + j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1|) = \mathrm{Tr}[\log\{I + j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1\}] = \mathrm{Tr}\Big[\sum_{k=1}^{\infty} (-1)^{k+1} \frac{\{j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1\}^k}{k} \Big],$$

$$= \mathrm{Tr}[j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1] + \mathrm{Tr}\Big[\sum_{k=2}^{\infty} (-1)^{k+1} \frac{\{j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1\}^k}{k} \Big], \tag{11}$$

where the first equality follows from $|A| = \exp(\mathrm{Tr}[\log A])$ and the second equality from $\log(I + A) = \sum_{k=1}^{\infty} (-1)^{k+1} A^k / k$. This expansion is valid if the maximum singular value of the matrix A is less than 1. Under our assumption that $p = o(n^{1/2})$, and by (14) we have that the maximal singular value of $j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1$ is $o_p(1)$, so the expansion is valid with probability tending to 1.

We first examine the order of

$$\left| \mathrm{Tr}[j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1] \right| \leq \left| \frac{t}{j_{\mathrm{p}}^{1/2}(\hat{\psi})} \mathrm{Tr}\left[j_{\lambda\lambda}^{-1}(\hat{\theta}) \frac{d}{d\psi} l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|_{\hat{\theta}_{\tilde{\psi}}} \right] \right|, \tag{12}$$

$$\leq \frac{|t|}{j_{\mathrm{p}}^{1/2}(\hat{\psi})} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_F \left\| \frac{d}{d\psi} l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|_{\hat{\theta}_{\tilde{\psi}}} \right\|_F = O_p\left(\frac{p^{3/2}}{n^{1/2}} \right), \tag{13}$$

by Proposition 1 and noting that by Assumptions 1 and 3, $j_{\mathrm{p}}^{-1/2}(\hat{\psi}) = O_p(n^{-1/2})$ and by Assumption 5, $t = O_p(1)$.

We now examine the magnitude of the second term in (11),

$$\mathrm{Tr}\Big[\sum_{k=2}^{\infty} (-1)^{k+1} \frac{\{j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1\}^k}{k} \Big] \leq p \left\| \sum_{k=2}^{\infty} (-1)^{k+1} \frac{\{j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1\}^k}{k} \right\|_{op},$$

$$\leq p \sum_{k=2}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) R_1 \right\|_{op}^k / k,$$

by using the maximum singular value to bound the trace using von Neumann's inequality and the triangle inequality.

The maximum singular value of $j_{\lambda\lambda}^{-1}(\hat{\theta})R_1$ satisfies

$$\left\| j_{\lambda\lambda}^{-1}(\hat{\theta})R_1 \right\|_{op} = O_p\left(\frac{p}{n^{1/2}}\right), \tag{14}$$

by (9) in the proof of Lemma 5. Note that

$$p\sum_{k=2}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta})R_1 \right\|_{op}^{k} /k = p\left\| j_{\lambda\lambda}^{-1}(\hat{\theta})R_1 \right\|_{op}^{2} \sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta})R_1 \right\|_{op}^{k} /(k+2), \tag{15}$$

$$= pO_p\left(\frac{p^2}{n}\right)O_p(1) = O_p\left(\frac{p^3}{n}\right), \tag{16}$$

by equation (14) and by Lemma 5. The result follows by combining (11), (13) and (16).

Corollary 1. *Under a p-fixed regime, Assumptions 1, 3–5 and the further assumption $\left\| dl_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)/d\psi|_\theta \right\|_2 = o_p(n)$, for θ in a neighbourhood of θ_0*

$$r_{np} - r^{-1}\log\rho = r^{-1}\log\left(|I + j_{\lambda\lambda}^{-1}(\hat{\theta})R_1|\right) = o_p\left(n^{-1/2}\right).$$

Proof. We note that in the p-fixed regime Assumption 2 is always satisfied by using an orthogonal parametrization at θ_0.

$$\log(|I + j_{\lambda\lambda}^{-1}(\hat{\theta})R_1|) = \mathrm{Tr}\left[\sum_{k=1}^{\infty}(-1)^{k+1}\frac{\{j_{\lambda\lambda}^{-1}(\hat{\theta})R_1\}^k}{k}\right]$$

$$\leq p\left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op} \|R_1\|_{op} \sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op}^{k} \|R_1\|_{op}^{k} /(k+1),$$

and by Assumption 3 and the additional assumption stated in Corollary 1,

$$\left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op} = O_p\left(n^{-1}\right),$$

$$\|R_1\|_{op} = \frac{t}{j_{\mathrm{p}}^{1/2}(\hat{\theta})} \left\| \frac{d}{d\psi} l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|_{\hat{\theta}_{\hat{\psi}}} \right\|_{op} = o_p(n^{1/2}).$$

Finally

$$\sum_{k=0}^{\infty} \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op}^{k} \|R_1\|_{op}^{k} /(k+1) = O_p(1),$$

by the argument used in the proof of Lemma 5. Combining the above with Lemma 4 we have,

$$r^{-1}\log\left(|I + j_{\lambda\lambda}^{-1}(\hat{\theta})R_1|\right) = o_p\left(n^{-1/2}\right).$$

4. Proof of Theorem 2 and Corollary 2

LEMMA 7. *Under Assumptions 1–5,*

$$R_2 = \frac{1}{j_p(\hat{\psi})} \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\hat{\psi}} \; j_{\lambda_j \psi}(\hat{\theta}) = O_p\left(\frac{p}{n^{1/2}}\right), \tag{17}$$

$$R_3 = \frac{t}{2 j_p^{3/2}(\hat{\psi})}\Big[l_{\psi\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial^2 \hat{\lambda}_{\psi,j}}{\partial \psi^2}|_{\tilde{\psi}} \; l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \; l_{\lambda_j \psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})$$
$$+ \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\tilde{\psi}} \; l_{\lambda_j \lambda_i;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})\Big] = O_p\left(\frac{p}{n^{1/2}}\right), \tag{18}$$

PROOF. We begin with R_2,

$$|R_2| = \left| \frac{1}{j_{\mathrm{p}}(\hat{\psi})} \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\hat{\psi}} \; j_{\lambda_j \psi}(\hat{\theta}) \right| \leq \frac{1}{j_{\mathrm{p}}(\hat{\psi})} \left\| \frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}|_{\hat{\psi}} \right\|_2 \left\| j_{\lambda\psi}(\hat{\theta}) \right\|_2,$$
$$= O_p\Big(n^{-1}\Big) O_p\Big(\frac{p^{1/2}}{n^{1/2}}\Big) O_p\{(np)^{1/2}\} = O_p\Big(\frac{p}{n^{1/2}}\Big),$$

by Proposition 1 and Lemma 1. Now for R_3

$$|R_3| = \Bigg| \frac{t}{2 j_{\mathrm{p}}^{3/2}(\hat{\psi})}\Big\{l_{\psi\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial^2 \hat{\lambda}_{\psi,j}}{\partial \psi^2}|_{\tilde{\psi}} \; l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \; l_{\lambda_j \psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})$$
$$+ \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\tilde{\psi}} \; l_{\lambda_j \lambda_i;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})\Big\}\Bigg|,$$
$$\leq \frac{|t|}{2 j_{\mathrm{p}}^{3/2}(\hat{\psi})}\Big\{|l_{\psi\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})| + \left\|\frac{\partial^2 \hat{\lambda}_{\psi}}{\partial \psi^2}|_{\tilde{\psi}}\right\|_2 \left\|l_{\lambda;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})\right\|_2 + \left\|\frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}|_{\tilde{\psi}}\right\|_2 \left\|l_{\lambda\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})\right\|_2$$
$$+ \left\|\frac{\partial \hat{\lambda}_{\psi}}{\partial \psi}|_{\tilde{\psi}}\right\|_2^2 \left\| \, l_{\lambda\lambda;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})\right\|_F \Big\},$$
$$= O_p\left(n^{-3/2}\right)\Big\{O_p(n) + O_p(p^{1/2})O_p(p^{1/2}n) + O_p\Big(\frac{p^{1/2}}{n^{1/2}}\Big)O_p(p^{1/2}n) + O_p\Big(\frac{p^{1/2}}{n^{1/2}}\Big)^2 O_p(pn)\Big\},$$
$$= O_p\Big(n^{-1/2}\Big) + O_p\Big(\frac{p}{n^{1/2}}\Big) + O_p\Big(\frac{p}{n}\Big) + O_p\Big(\frac{p^2}{n^{3/2}}\Big) = O_p\Big(\frac{p}{n^{1/2}}\Big),$$

by Lemma 1 and 2, Proposition 1 and the fact that $p/n^{1/2} = o(1)$.

THEOREM 2. *Under Assumptions 1–5,* $r_{inf} = O_p(p/n^{1/2})$.

PROOF. Recall the definition of r_{inf} in Equation (6) of the main text

$$\begin{aligned} r_{inf} &= r^{-1} \log \left[\frac{l_{;\hat{\psi}}(\hat{\theta}_{\psi_0}) - l_{;\hat{\psi}}(\hat{\theta})}{j_{\mathrm{p}}^{1/2}(\hat{\psi}) r} - \frac{l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0})\{l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0})\}^{-1}\left\{l_{;\hat{\lambda}}(\hat{\theta}_{\psi_0}) - l_{;\hat{\lambda}}(\hat{\theta})\right\}}{j_{\mathrm{p}}^{1/2}(\hat{\psi}) r} \right], \\ &= r^{-1} \log(C + D), \end{aligned}$$

say.

Order of C: By a second order Taylor expansion,

$$C = \left\{ r j_{\mathrm{p}}^{1/2}(\hat{\psi}) \right\}^{-1} \Big[\frac{t}{j_{\mathrm{p}}^{1/2}(\hat{\psi})} j_{\psi\psi}(\hat{\theta}) + \frac{t}{j_{\mathrm{p}}^{1/2}(\hat{\psi})} \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi} l_{\lambda_j;\hat{\psi}}(\hat{\theta}) \tag{19}$$

$$\frac{t^2}{2 j_{\mathrm{p}}(\hat{\psi})} \Big\{ l_{\psi\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial^2 \hat{\lambda}_{\psi,j}}{\partial \psi^2}|_{\tilde{\psi}}\, l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}}\, l_{\lambda_j \psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) \tag{20}$$

$$+ \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi}|_{\tilde{\psi}} \frac{\partial \hat{\lambda}_{\psi,i}}{\partial \psi}|_{\tilde{\psi}}\, l_{\lambda_j \lambda_i;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) \Big\} \Big], \tag{21}$$

$$= \frac{t}{r} \Big\{ \frac{j_{\mathrm{p}}(\hat{\psi}) + j_{\psi\lambda}(\hat{\theta}) j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})} + R_2 + R_3 \Big\}, \tag{22}$$

$$= \frac{t}{r} \Big\{ 1 + \frac{j_{\psi\lambda}(\hat{\theta}) j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})} + R_2 + R_3 \Big\}, \tag{23}$$

$$= \left\{ 1 + O_p\left(n^{-1/2}\right) \right\} \left\{ 1 + O_p\left(\frac{p}{n^{1/2}}\right) \right\} = 1 + O_p\left(\frac{p}{n^{1/2}}\right), \tag{24}$$

where (22) uses $j_{\mathrm{p}}(\hat{\psi}) = j_{\psi\psi}(\hat{\theta}) - j_{\psi\lambda}(\hat{\theta}) j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\lambda\psi}(\hat{\theta})$. The final rate on line (24) is obtained by Lemmas 3 and 7, and by noting that the ratio

$$\frac{j_{\psi\lambda}(\hat{\theta}) j_{\lambda\lambda}^{-1}(\hat{\theta}) j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})} \le \left\| j_{\lambda\lambda}^{-1}(\hat{\theta}) \right\|_{op} \left\| j_{\lambda\psi}(\hat{\theta}) \right\|_2^2 / j_{\mathrm{p}}(\hat{\psi}) = O_p\left(\frac{p}{n}\right), \tag{25}$$

using Rayleigh's quotient on the numerator and Assumptions 1–3.

Order of D:

$$|D| \le \left\{ j_{\mathrm{p}}^{1/2}(\hat{\psi})\, |r| \right\}^{-1} \left\| l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0}) \{ l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0}) \}^{-1} \right\|_2 \left\| l_{;\hat{\lambda}}(\hat{\theta}_{\psi_0}) - l_{;\hat{\lambda}}(\hat{\theta}) \right\|_2 .$$

We first expand

$$
\begin{aligned}
&- l_{\lambda;\hat\psi}(\hat\theta_{\psi_0})\Big\{l_{\lambda;\hat\lambda}(\hat\theta_{\psi_0})\Big\}^{-1}\\
&= -\Big\{j_{\lambda\psi}(\hat\theta) + (\psi_0-\hat\psi)\frac{d}{d\psi}l_{\lambda;\hat\psi}(\hat\theta_\psi)|_{\hat\theta_{\bar\psi}}\Big\}j_{\lambda\lambda}^{-1}(\hat\theta)\Big\{I + (\psi_0-\hat\psi)\frac{d}{d\psi}l_{\lambda;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\bar\psi}}\, j_{\lambda\lambda}^{-1}(\hat\theta)\Big\}^{-1},\\
&= -\Big\{j_{\lambda\psi}(\hat\theta) + R_4\Big\}j_{\lambda\lambda}^{-1}(\hat\theta)\Big\{I + R_5\Big\}^{-1},\\
&= -\Big\{j_{\lambda\psi}(\hat\theta) + R_4\Big\}j_{\lambda\lambda}^{-1}(\hat\theta)\Big\{I + \sum_{k=1}^{\infty}(-1)^{k+1}R_5^k\Big\}.
\end{aligned}
$$

The final equality above uses $(I-A)^{-1} = \sum_{k=1}^{\infty} A^k$. Under our assumption that $p = o(n^{1/2})$ and (26), we have that the maximal singular value of R_5 is $o_p(1)$, so expansion is valid with probability tending to 1. Recalling that the Euclidean norm and maximum singular value of a vector coincide and using the sub-additivity of the induced matrix norm and the triangle inequality we obtain

$$
\begin{aligned}
\Big\|l_{\lambda;\hat\psi}(\hat\theta_{\psi_0})\{l_{\lambda;\hat\lambda}(\hat\theta_{\psi_0})\}^{-1}\Big\|_2 &= \Big\|l_{\lambda;\hat\psi}(\hat\theta_{\psi_0})\{l_{\lambda;\hat\lambda}(\hat\theta_{\psi_0})\}^{-1}\Big\|_{op}\\
&\le \Big\{\Big\|j_{\lambda\psi}(\hat\theta)\Big\|_2 + \|R_4\|_2\Big\}\Big\|j_{\lambda\lambda}^{-1}(\hat\theta)\Big\|_{op}\Big\{1 + \|R_5\|_{op}\sum_{k=0}^{\infty}(-1)^{k+2}\|R_5\|_{op}^k\Big\}.
\end{aligned}
$$

Rates of growth can be obtained for the maximum singular values for the above matrices by using the Frobenius norm as an upper bound for the maximum singular value,

$$
\begin{aligned}
\Big\|j_{\lambda\psi}(\hat\theta)\Big\|_2 &= O_p\{(np)^{1/2}\},\\
\|R_4\|_2 &\le \frac{|t|}{j_{\rm p}^{1/2}(\hat\psi)}\Big\|\frac{d}{d\psi}l_{\lambda;\hat\psi}(\hat\theta_\psi)|_{\hat\theta_{\bar\psi}}\Big\|_F = O_p\{(np)^{1/2}\},\\
\|R_5\|_{op} &\le \frac{|t|}{j_{\rm p}^{1/2}(\hat\psi)}\Big\|\frac{d}{d\psi}l_{\lambda;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\bar\psi}}\Big\|_F\Big\|l_{\lambda\lambda}^{-1}(\hat\theta)\Big\|_{op} = O_p\left(\frac{p}{n^{1/2}}\right),
\end{aligned}
\tag{26}
$$

by Proposition 1 and Assumptions 1, 3, and 5. Finally,

$$
\sum_{k=0}^{\infty}(-1)^{k+2}\|R_5\|_{op}^k = O_p(1),
$$

by the same arguments as in Lemma 5. Combining the above, we obtain

$$
\Big\|l_{\lambda;\hat\psi}(\hat\theta_{\psi_0})\{l_{\lambda;\hat\lambda}(\hat\theta_{\psi_0})\}^{-1}\Big\|_2 = O_p\Big\{\max\left(\frac{p^{1/2}}{n^{1/2}}, \frac{p^{3/2}}{n}\right)\Big\}. \tag{27}
$$

Now consider

$$
\begin{aligned}
\Big\|l_{;\hat\lambda}(\hat\theta_{\psi_0}) - l_{;\hat\lambda}(\hat\theta)\Big\|_2 &= \frac{|t|}{j_{\rm p}^{1/2}(\hat\psi)}\Big\|\frac{d}{d\psi}l_{;\hat\lambda}(\hat\theta_\psi)|_{\hat\theta_{\bar\psi}}\Big\|_2\\
&= |t|O_p\{(np)^{1/2}\}.
\end{aligned}
$$

We combine this with (27) to get

$$\left\{j_{\mathrm{p}}^{1/2}(\hat{\psi})|r|\right\}^{-1}\left\|l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0})\{l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0})\}^{-1}\right\|_2\left\|l_{;\hat{\lambda}}(\hat{\theta}_{\psi_0})-l_{;\hat{\lambda}}(\hat{\theta})\right\|_2$$
$$\leq\left|\frac{t}{r}\right|j_{\mathrm{p}}^{-1/2}(\hat{\psi})O_p\Big\{\max\left(\frac{p^{1/2}}{n^{1/2}},\frac{p^{3/2}}{n}\right)\Big\}O_p\{(np)^{1/2}\}=O_p\Big\{\max\left(\frac{p}{n^{1/2}},\frac{p^2}{n}\right)\Big\}=O_p\left(\frac{p}{n^{1/2}}\right),$$

by Lemmas 1 and 3 and noting that $j_{\mathrm{p}}^{-1/2}(\hat{\psi})=O_p(n^{-1/2})$ under Assumptions 1 and 3. Therefore we may express r_{inf} as

$$r_{inf}=\frac{1}{r}\log\left\{1+O_p\left(\frac{p}{n^{1/2}}\right)\right\}=O_p\left(\frac{p}{n^{1/2}}\right),$$

using Lemma 4 and $\log\{1+O_p(p/n^{1/2})\}=O_p(p/n^{1/2})$, as $x/(1+x)\leq\log(1+x)\leq x$.

COROLLARY 2. *Under a p-fixed regime and Assumptions 1, 3–5, and if* $l_{\psi\psi;\hat{\psi}}(\theta)=o_p(n)$ *in a neighbourhood of* θ_0,

$$r_{inf}-\frac{1}{r}\log\left(\frac{t}{r}\right)=o_p\left(n^{-1/2}\right).$$

PROOF. We note that Assumption 2 is always satisfied in the p-fixed regime, as we may without loss of generality assume that the parametrization is orthogonal at θ_0. We use the same decomposition of r_{inf} as given in Theorem 2.

$$r_{inf}=\frac{1}{r}\log(C)+\frac{1}{r}\log(D).$$

First,

$$\frac{1}{r}\log(C)=\frac{1}{r}\log\frac{t}{r}+\frac{1}{r}\log\left(1+\frac{j_{\psi\lambda}(\hat{\theta})j_{\lambda\lambda}^{-1}(\hat{\theta})j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})}+R_2+R_3\right);$$

the first term is the leading term, we now bound the second term. Using the Rayleigh quotient and Assumptions 1 to 3

$$\frac{j_{\psi\lambda}(\hat{\theta})j_{\lambda\lambda}^{-1}(\hat{\theta})j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})}=O_p\left(n^{-1}\right).$$

As for R_2,

$$R_2=\frac{1}{j_{\mathrm{p}}(\hat{\psi})}\sum_{j=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,j}}{\partial\psi}|_{\hat{\psi}}\;j_{\lambda_j\psi}(\hat{\theta})=O_p\left(n^{-1}\right),$$

by Assumptions 1 and 2 and noting that in the p-fixed case, the derivative of the constrained maximum likelihood estimator is $O_p(n^{-1/2})$ under the orthogonal paramterization, as follows from Lemma 1 with p fixed. We examine the components of R_3:

$$R_3 = \frac{t}{2j_{\mathrm{p}}^{3/2}(\hat{\psi})}\Big[\underbrace{l_{\psi\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})}_{(a)} + \underbrace{\sum_{j=1}^{p-1}\frac{\partial^2\hat{\lambda}_{\psi,j}}{\partial\psi^2}|_{\tilde{\psi}}\, l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})}_{(b)}$$
$$\underbrace{+\sum_{j=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,j}}{\partial\psi}|_{\tilde{\psi}}\, l_{\lambda_j\psi;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) + \sum_{i=1}^{p-1}\sum_{j=1}^{p-1}\frac{\partial\hat{\lambda}_{\psi,j}}{\partial\psi}|_{\tilde{\psi}}\frac{\partial\hat{\lambda}_{\psi,i}}{\partial\psi}|_{\tilde{\psi}}\, l_{\lambda_j\lambda_i;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}})}_{(c)}\Big].$$

By the assumption in the statement (a) $= o_p(n)$. That $(c) = O_p(n^{1/2})$ follows by noting that the derivative of the constrained maximum likelihood estimator is $O_p(n^{-1/2})$ combined with Assumptions 1 and 4 since,

$$(c) = \sum_{j=1}^{p-1} O_p\left(n^{-1/2}\right) O_p(n) + \sum_{i=1}^{p-1}\sum_{j=1}^{p-1} O_p\left(n^{-1/2}\right) O_p\left(n^{-1/2}\right) O_p(n) = O_p(n^{1/2}).$$

Finally for (b),

$$\sum_{j=1}^{p-1}\frac{\partial^2\hat{\lambda}_{\psi,j}}{\partial\psi^2}|_{\tilde{\psi}}\left[\{l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) - l_{\lambda_j;\hat{\psi}}(\hat{\theta})\} + j_{\lambda_j\psi}(\hat{\theta})\right] = O_p(n^{1/2}), \tag{28}$$

using Lemma 2, Assumption 1, noting that the second derivative of the constrained maximum likelihood estimate is $O_p(1)$ in a p-fixed asymptotic regime, and noting that

$$l_{\lambda_j;\hat{\psi}}(\hat{\theta}_{\tilde{\psi}}) - l_{\lambda_j;\hat{\psi}}(\hat{\theta}) = l_{\theta\lambda_j;\hat{\psi}}(\theta')(\hat{\theta}_{\tilde{\psi}} - \hat{\theta}) = O_p(n^{1/2}), \tag{29}$$

for some θ' lying on a line segment between $\hat{\theta}$ and $\hat{\theta}_{\tilde{\psi}}$ by a first-order Taylor expansion. Result (29) follows from $l_{\theta\lambda_i;\hat{\psi}}(\theta') = O_p(n)$ by Assumptions 1 and 4 and the fact that $(\hat{\theta}_{\tilde{\psi}} - \hat{\theta}) = O_p(n^{-1/2})$ in p-fixed asymptotic regime. Therefore,

$$\frac{1}{r}\log\left(1 + \frac{j_{\psi\lambda}(\hat{\theta})j_{\lambda\lambda}^{-1}(\hat{\theta})j_{\lambda\psi}(\hat{\theta})}{j_{\mathrm{p}}(\hat{\psi})} + R_2 + R_3\right) = o_p\left(n^{-1/2}\right).$$

We now show that $r^{-1}\log(D) = O_p(n^{-1})$, which completes the proof.

$$\begin{aligned}
|D| &= \left| \left\{ j_{\mathrm{p}}^{1/2}(\hat{\psi}) r \right\}^{-1} l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0}) \{ l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0}) \}^{-1} \left\{ l_{;\hat{\lambda}}(\hat{\theta}_{\psi_0}) - l_{;\hat{\lambda}}(\hat{\theta}) \right\} \right|, \\
&= \left| \left\{ j_{\mathrm{p}}^{1/2}(\hat{\psi}) r \right\}^{-1} l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0}) \{ l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0}) \}^{-1} \left[(\psi_0 - \hat{\psi}) \{ l_{\psi;\hat{\lambda}}(\theta') + \sum_{j=1}^{p} l_{\lambda_i;\hat{\lambda}}(\theta') \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi} |_{\psi'} \} \right] \right|, \\
&\leq \left\{ j_{\mathrm{p}}(\hat{\psi})^{1/2} |r| \right\}^{-1} \left\| l_{\lambda;\hat{\psi}}(\hat{\theta}_{\psi_0}) \{ l_{\lambda;\hat{\lambda}}(\hat{\theta}_{\psi_0}) \}^{-1} \right\|_2 \left[|\psi_0 - \hat{\psi}| \left\{ \left\| l_{\psi;\hat{\lambda}}(\theta') \right\|_2 + \left\| \sum_{j=1}^{p} l_{\lambda_i;\hat{\lambda}}(\theta') \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi} |_{\psi'} \right\|_2 \right\} \right], \\
&= O_p\left(n^{-1}\right),
\end{aligned}$$

where θ' lies on the line segment between $\hat{\theta}_{\psi_0}$ and $\hat{\theta}$. Noting that $(\hat{\psi} - \psi_0) = O_p(n^{-1/2})$, that $l_{\psi;\hat{\lambda}}(\theta') = O_p(n^{1/2})$ by the same argument as for C, and that

$$\sum_{j=1}^{p} l_{\lambda_i;\hat{\lambda}}(\theta') \frac{\partial \hat{\lambda}_{\psi,j}}{\partial \psi} |_{\psi'} = O_p(n^{1/2}),$$

with (27) shows $D = O_p(n^{-1})$.

5. Proof of Proposition 1

LEMMA 8. *Under the orthogonal parameterization of a linear exponential family, and Assumptions 1, 5–9,*

$$t = r\left\{1 + O_p\left(n^{-1/2}\right)\right\},$$

and

$$j^{-1}_{\psi\psi}(\hat{\theta}_{\tilde{\psi}}) = O_p\left(n^{-1}\right), \quad \left\| j^{-1}_{\tau\tau}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} = O_p\left(n^{-1}\right).$$

PROOF. We first show that

$$\left\| j^{-1}_{\tau\tau}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} = O_p\left(n^{-1}\right).$$

We begin by considering the information matrix under the canonical parameterization for generalized linear models, which can be written as

$$\tilde{\jmath}_{\lambda\lambda}(\theta) = X_\lambda^\top D X_\lambda,$$

where D is a diagonal matrix with ith entry $K''(x_i^\top \theta)$, and X_λ is the design matrix with the column of covariates associated with ψ removed. Then

$$\left\| \tilde{\jmath}_{\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} \leq \max_{i=1,\dots,n} K''\left(x_i^\top \hat{\theta}_{\tilde{\psi}}\right) \left\| X_\lambda^\top X_\lambda \right\|_{op} = O_p(n),$$

by Assumptions 1, 6 and 7, and noting the relationship between the observed information functions under the two parameterizations,

$$\left\| j_{\tau\tau}^{-1}(\tilde{\psi}, \hat{\tau}) \right\|_{op} = \frac{1}{n^2} \left\| \tilde{\jmath}_{\lambda\lambda}(\tilde{\psi}, \hat{\lambda}_{\tilde{\psi}}) \right\|_{op} = O_p\left(n^{-1}\right),$$

gives us the desired result.

We now show that $j_{\psi\psi}^{-1}(\hat{\theta}) = O_p(n^{-1})$. Since $j_{\psi\psi}^{-1}(\hat{\theta}) = \tilde{\jmath}_{\mathrm{p}}^{-1}(\hat{\theta})$, it is sufficient to show that the eigenvalues of $\tilde{\jmath}^{-1}(\hat{\theta})$ are $O_p(n^{-1})$ since $j_{\mathrm{p}}^{-1}(\hat{\theta})$ is an element on the diagonal of $\tilde{\jmath}^{-1}(\hat{\theta})$. By positive definiteness of the observed Fisher information matrix,

$$\left\| \tilde{\jmath}^{-1}(\hat{\theta}) \right\|_{op} \leq \left[\eta_p\{\tilde{\jmath}(\hat{\theta})\} \right]^{-1} \leq \max_{i=1,\dots,n} \left\{ K''\left(x_i^\top \hat{\theta}_{\tilde{\psi}} \right) \right\}^{-1} \eta_p \left(X_\lambda^\top X_\lambda \right)^{-1} = O_p\left(n^{-1}\right),$$

by Assumptions 1 and 7.

Lastly,

$$t = r\left\{ 1 + O_p\left(n^{-1/2} \right) \right\}, \tag{30}$$

follows as in the orthogonal parameterization of the linear exponential family

$$l_{\mathrm{p}}^{(3)}(\tilde{\psi}) = j_{\psi\psi\psi}(\hat{\theta}_{\tilde{\psi}}) = O_p(n),$$

by Assumption 8. We can then use the same argument as in Lemma 3 to show (30).

Proof of Proposition 1. *Under Assumptions 1 and 5–9, for the linear exponential model,*

$$r_{np} = O_p\left(\frac{p}{n^{1/2}} \right), \quad r_{inf} = O_p\left(n^{-1/2} \right).$$

Proof.

$$r_{np} = -\frac{1}{r} \log\left\{ \frac{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_{\psi_0})|^{1/2}} \right\} = \frac{1}{2r}(\hat{\psi} - \psi_0)\gamma_1(\tilde{\psi}) = \frac{t}{2r} \frac{\gamma_1(\tilde{\psi})}{j_{\mathrm{p}}(\hat{\psi})^{1/2}},$$

and by Neumann's inequality we have

$$\gamma_1(\tilde{\psi}) = \mathrm{Tr}[j_{\tau\tau}^{-1}(\hat{\theta}_{\tilde{\psi}}) j_{\psi\tau\tau}(\hat{\theta}_{\tilde{\psi}})] \leq p \left\| j_{\tau\tau}^{-1}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} \left\| j_{\psi\tau\tau}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} = pO_p\left(n^{-1}\right)O_p(n) = O_p(p),$$

by Lemma 8 and Assumptions 1 and 9. Combining this with Assumption 5 we obtain

$$r_{np} = O_p\left(\frac{p}{n^{1/2}} \right).$$

As for r_{inf}, note that $j_{\mathrm{p}}(\hat{\psi}) = j_{\psi\psi}(\hat{\theta})$

$$r_{inf} = \frac{1}{r} \log\left[\{ r j_{\psi\psi}^{1/2}(\hat{\psi}) \}^{-1} \{ \hat{\psi} - \psi_0 \} \right] = \frac{1}{r} \log\left(\frac{t}{r} \right),$$

therefore by Lemma 8, $r_{inf} = O_p(n^{-1/2})$.

6. Proof of Proposition 2

LEMMA 9. *Under Assumptions 1–5 and 10,*

$$\left\| \frac{d}{d\psi} j_{\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} = O_p(n), \tag{31}$$

PROOF. The maximum singular value of

$$\begin{aligned}
\left\| \frac{d}{d\psi} j_{\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} &= \left\| j_{\psi\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) + \sum_{j=1}^{p-1} \frac{d\hat{\lambda}_{\psi,j}}{d\psi} j_{\lambda_j\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op}, \\
&\leq \left\| j_{\psi\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} + \max_{j=1,\dots,p-1} \left\| j_{\lambda_j\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} \sum_{j=1}^{p-1} \left| \frac{d\hat{\lambda}_{\psi,j}}{d\psi} \right|, \\
&\leq \left\| j_{\psi\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} + p^{1/2} \max_{j=1,\dots,p-1} \left\| j_{\lambda_j\lambda\lambda}(\hat{\theta}_{\tilde{\psi}}) \right\|_{op} \left\| \frac{d\hat{\lambda}_{\psi}}{d\psi} \right\|_2, \\
&= O_p(n) + O_p(pn^{1/2}) \leq O_p(n),
\end{aligned}$$

where the last inequality follows from $\sum_{j=1}^{p} |a_j| \leq \{\sum_{j=1}^{p} a_j^2\}^{1/2}$. The result is obtained by using Lemma 1 and Assumptions 1 and 10.

LEMMA 10. *Under Assumptions 1–5 and 10,*

$$s = t\left\{1 + O_p\left(n^{-1/2}\right)\right\}.$$

PROOF.

$$\begin{aligned}
s = \frac{\zeta_1(\psi_0)}{j_{\mathrm{p}}^{1/2}(\hat{\psi})} &= \frac{1}{j_{\mathrm{p}}^{1/2}(\hat{\psi})}\Big\{\zeta_1(\hat{\psi}) - \zeta_2(\hat{\psi})(\hat{\psi} - \psi_0) + \frac{\zeta_3(\tilde{\psi})}{2}(\hat{\psi} - \psi_0)^2\Big\}, \\
&= t\Big\{1 + \frac{\kappa_3(\tilde{\psi})t}{2}\Big\},
\end{aligned}$$

for some $\tilde{\psi}$ lying on the line segment between $\hat{\psi}$ and ψ. The result follows by using proof of Lemma 3 to show that $\kappa_3(\tilde{\psi}) = O_p(n^{-1/2})$.

PROOF OF PROPOSITION 2. *Under Assumptions 1–5 and 10, for a location-scale model,*

$$r_{inf} = O_p\Big(n^{-1/2}\Big), \quad r_{np} = O_p\Big(\frac{p}{n^{1/2}}\Big).$$

PROOF. By Lemmas 3 and 10,

$$s = r\left\{1 + O_p\left(n^{-1/2}\right)\right\},$$

which implies that

$$r_{inf} = \frac{1}{r}\log\left(\frac{s}{r}\right) = \frac{1}{r}\log\left\{1 + O_p\left(n^{-1/2}\right)\right\} = O_p\left(n^{-1/2}\right).$$

As for r_{np}, the proof is similar to that of Proposition 1, although more terms are obtained from the differentiation of the constrained maximum likelihood estimate:

$$r_{np} = -\frac{1}{r}\log\left\{\frac{|j_{\lambda\lambda}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_{\psi_0})|^{1/2}}\right\} = \frac{1}{2r}(\hat{\psi}-\psi_0)\gamma_1(\tilde{\psi}),$$

where $\gamma_1(\tilde{\psi})$ is now,

$$\gamma_1(\tilde{\psi}) = \mathrm{Tr}[j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}})\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_{\tilde{\psi}})] \le p\left\|j_{\lambda\lambda}^{-1}(\hat{\theta}_{\tilde{\psi}})\right\|_{op}\left\|\frac{d}{d\psi}j_{\lambda\lambda}(\hat{\theta}_{\tilde{\psi}})\right\|_{op} = pO_p\left(n^{-1}\right)O_p(n) = O_p(p),$$

by Lemma 9. Combining this with Assumptions 3 and 5 shows that $r_{np} = O_p(p/n^{1/2})$.

References

Cox, D. R. and N. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B* **53**, 79–109.