Submitted to Biometrika (2018), *pp.* 1–37
Printed in Great Britain

Supplementary material for "Accounting for unobserved covariates with varying degrees of estimability in high dimensional biological data"

BY CHRIS MCKENNAN, DAN NICOLAE
Department of Statistics, University of Chicago, 5747 S. Ellis Avenue, Chicago, IL
cgm29@galton.uchicago.edu nicolae@galton.uchicago.edu

S1. ADDITIONAL SIMULATIONS

We first include an empirical verification of the surprising results from Proposition 2 and Lemma 2 that we can accurately estimate L, but the naive estimate for Ω in (12) is biased. The data were simulated as follows:

$$\begin{aligned}
&d = K = 1,\, n = 100,\, p = 10^5\\
&X \sim N_n\left(0, I_n\right)\\
&B = 0\\
&L_g \sim \left(1 - n^{-1}\right)\delta_0 + n^{-1}N_1(0,1) \;\; (g = 1, \ldots, p)\\
&X = \left(1_{n/2}^{\mathrm{T}}, 0_{n/2}^{\mathrm{T}}\right)^{\mathrm{T}}\\
&C_2 \sim N_{n-d}\left(0, I_{n-d}\right)\\
&C = X + Q_X C_2\\
&\mathcal{E} \sim MN_{p\times n}\left(0, I_p, I_n\right).
\end{aligned}$$

where $Q_X \in \mathbb{R}^{n\times(n-d)}$ is a matrix whose columns form an orthonormal basis for the null sapce of X. Therefore, $\Omega = 1$ for every simulation. L and Σ were simulated so that $\lambda_1 \approx 1$ and $\rho = 1$. If (18) from Lemma 2 was correct, then $n^{1/2}\left(\hat{L}_g - L_g\right) \approx N(0,1)$ for each $g \in [p]$. If we let $W = n^{1/2}\left(\hat{L} - L\right) \in \mathbb{R}^p$, we validate Lemma 2 by partitioning the components of W by whether or not the corresponding component of L was non-zero. If Lemma 2 were true, both histograms in Fig. S1 should look as if they were sampled from a $N(0,1)$ random variable, which they clearly do.

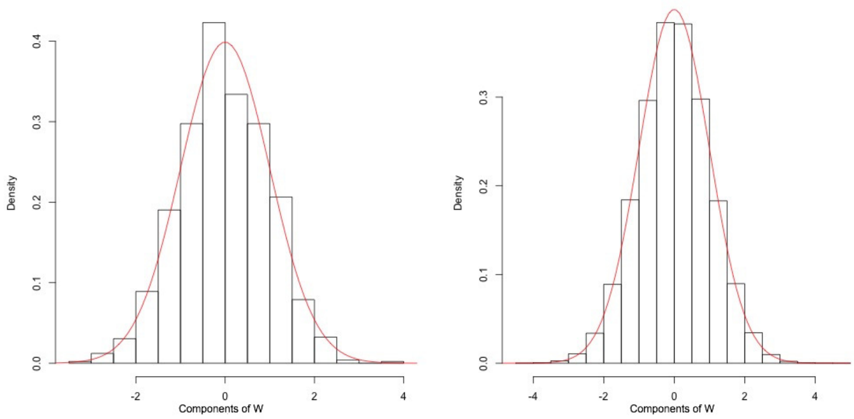


Fig. S1: $W = n^{1/2}\left(\hat{L} - L\right) \in \mathbb{R}^p$ for one simulation for components of L that are non-zero (left) or 0 (right). The overlaid red curve is the density of a $N(0,1)$ random variable.

We next empirically verified (19) from Proposition 2 using 20 simulations. Figure S2 contains the results of the 20 simulations, which clearly shows that $n^{1/2}\|\hat{\Omega}^{\text{shrunk}} - \Omega\lambda_1(\lambda_1+\rho)^{-1}\|_2 \approx 0$.

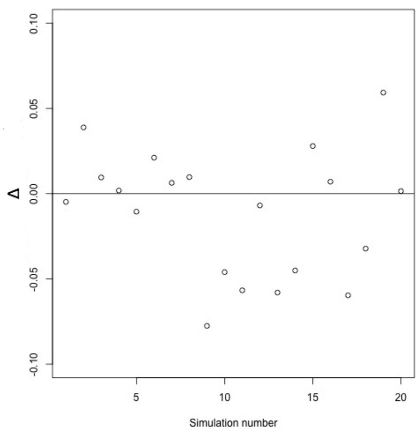
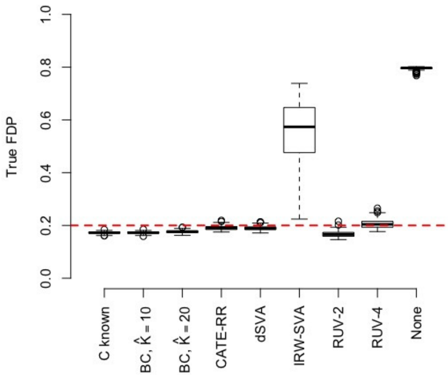


Fig. S2: $\Delta = n^{1/2}\left\{\hat{\Omega}^{\text{shrunk}} - \Omega\lambda_1(\lambda_1+\rho)^{-1}\right\}$ for 20 simulations.

Lastly, Fig. S3 gives the simulation results from Section 4.1, with $B_g \sim 0.80\delta_0 + 0.20N\left(0, 0.4^2\right)$. The average power for the simulations on the left panel for C known, BC $\hat{K} = 10$, BC $\hat{K} = 20$ was $23.3\%, 23.3\%, 22.1\%$ and $23.0\%, 23.0\%, 21.9\%$ for the simulations on the right panel.



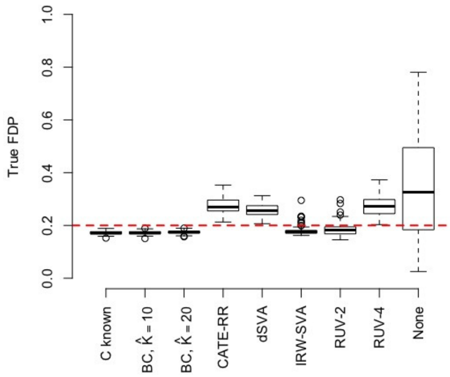


Fig. S3: Simulations with $B_g \sim 0.80\delta_0 + 0.20N\left(0, 0.4^2\right)$ for $A = A_1$ (left) and $A = A_2$ (right). All other parameters are the same as the simulations in Section 4.1.

S2. PROOFS OF ALL PROPOSITIONS, LEMMAS, AND THEOREMS

S2.1. *Proof of Proposition* 1 *and the identifiability of* $\Omega\left(n^{-1}C_2^TC_2\right)^{-1}\Omega^T$

For the remainder of the Supplement, we define $L_{g*}=\ell_g$ for all $g\in[p]$. Let X be a matrix, vector or scalar and let $\|X\|_2$ be the spectral norm, Euclidean norm or magnitude of X. We use the notation that $X=O_{\text{p}}(a_n)$ if for some sequence a_n, $\|X\|_2/a_n=O_{\text{p}}(1)$. Similarly, $X=o_{\text{p}}(a_n)$ if $\|X\|_2/a_n=o_{\text{p}}(1)$. Lastly, for any vector $v\in\mathbb{R}^m$, we define v_j to be the jth element of v for all $j=1,\ldots,m$. If the vector has a subscript r, then we define v_{rj} to be the jth elements of $v_r\in\mathbb{R}^m$ for all $j=1,\ldots,m$.

We first prove Proposition 1.

Proof of Proposition 1. Under Assumptions 1 and 2, we can find an $L\in\mathbb{R}^{p\times K}$, $C\in\mathbb{R}^{n\times K}$ such that for $C_2=P_X^{\perp}C$, $E(Y_2)=LC_2^{\text{T}}$ and

$$np^{-1}L^{\text{T}}L=\text{diag}\left(\lambda_1,\ldots,\lambda_K\right),\quad n^{-1}C_2^{\text{T}}C_2=I_K \tag{S1}$$

by taking the singular value decomposition of $E(Y_2)$. The columns of L and C_2 are unique up to sign by the uniqueness of the singular value decomposition, since $\lambda_k>\lambda_{k+1}$ for all $k=1,\ldots,K$ (where $\lambda_{K+1}=0$). That is, if $\tilde{L}$ and $\tilde{C}$ also satisfy (S1), then $\tilde{L}=L\Pi$ and $\tilde{C}_2=C_2\Pi$ where $\Pi=\text{diag}\left(a_1,\ldots,a_K\right)$ and $a_k\in\{-1,1\}$ for all $k\in[K]$.

Next, suppose Assumption 3(a) holds and let $B^{(a)},L^{(a)},C^{(a)}$ and $B^{(b)},L^{(b)},C^{(b)}$ be such that $L^{(a)}\left\{C_2^{(a)}\right\}^{\text{T}}=E(Y_2)=L^{(b)}\left\{C_2^{(b)}\right\}^{\text{T}}$ and

$$B^{(a)}+L^{(a)}\left\{\Omega^{(a)}\right\}^{\text{T}}=E(Y_1)=B^{(b)}+L^{(b)}\left\{\Omega^{(b)}\right\}^{\text{T}}.$$

We can find invertible matrices $R^{(a)},R^{(b)}\in\mathbb{R}^{K\times K}$ such that $L^{(a)}R^{(a)},C^{(a)}\left\{R^{(a)}\right\}^{-\text{T}}$ and $L^{(b)}R^{(b)},C^{(b)}\left\{R^{(b)}\right\}^{-\text{T}}$ satisfy (S1), where

$$L^{(i)}\left\{\Omega^{(i)}\right\}^{\text{T}}=\left\{L^{(i)}R^{(i)}\right\}\left[\Omega^{(i)}\left\{R^{(i)}\right\}^{-\text{T}}\right]^{\text{T}}\quad(i=a,b).$$

Therefore, to prove the identifiability of B, it suffices to assume $L^{(a)},C^{(a)}$ and $L^{(b)},C^{(b)}$ satisfy (S1), meaning $L^{(b)}=L^{(a)}\Pi$ and $C_2^{(b)}=C_2^{(a)}\Pi$ for some $\Pi=\text{diag}\left(a_1,\ldots,a_K\right)$ where $a_k\in\{-1,1\}$ for all $k\in[K]$. Define $L=L^{(a)}$ (for notational convenience), $\mathcal{S}=\left\{g\in[p]:B_{g*}^{(a)}=B_{g*}^{(b)}=0\right\}$ and $L_{\mathcal{S}}\in\mathbb{R}^{|\mathcal{S}|\times K}$ to be the submatrix of L restricted to the rows $g\in\mathcal{S}$. To prove that $B^{(a)}=B^{(b)}$ and $C^{(b)}=C^{(a)}\Pi$, it suffices to show that $L_{\mathcal{S}}^{\text{T}}L_{\mathcal{S}}\succ 0$, i.e. $L_{\mathcal{S}}$ has full column rank.

$$n\left(p\lambda_K\right)^{-1}L_{\mathcal{S}}^{\text{T}}L_{\mathcal{S}}=\text{diag}\left(\lambda_1\lambda_K^{-1},\cdots,\lambda_{K-1}\lambda_K^{-1},1\right)-n\left(p\lambda_K\right)^{-1}\sum_{g\in[p]\setminus\mathcal{S}}\ell_g\ell_g^{\text{T}}$$

where for any $r,s\in[K]$,

$$\begin{aligned}|n\left(p\lambda_K\right)^{-1}\sum_{g\in[p]\setminus\mathcal{S}}\ell_{gr}\ell_{gs}|&\leq n\left(p\lambda_K\right)^{-1}\sum_{g\in[p]\setminus\mathcal{S}}|\ell_{gr}||\ell_{gs}|\\&\leq c_2^2n\lambda_K^{-1}\left[p^{-1}\sum_{g=1}^{p}I\left\{B_{g*}^{(a)}\neq 0\right\}+p^{-1}\sum_{g=1}^{p}I\left\{B_{g*}^{(b)}\neq 0\right\}\right]=o\left(n^{-1/2}\right),\end{aligned}$$

which completes the proof. □

We next state and prove a proposition regarding the identifiability of $\Omega\left(n^{-1}C_2^{\text{T}}C_2\right)^{-1}\Omega^{\text{T}}$.

PROPOSITION S1. *Suppose Assumptions* 1, 2 *and* 3*(a) hold. Then* $\Omega\left(n^{-1}C_2^{\text{T}}C_2\right)^{-1}\Omega^{\text{T}}$ *is identifiable for all* $n\geq c_4$*, where* c_4 *is defined in the statement of Proposition* 1.

Proof. Under these assumptions, Proposition 1 proves that B is identifiable for all $n \geq c_4$, meaning $E(Y) - BX^{\mathrm{T}} = LC^{\mathrm{T}}$ is identifiable for all $n \geq c_4$. Suppose $C_{(a)}, C_{(b)} \in \mathbb{R}^{n \times K}$ and $L_{(a)}, L_{(b)} \in \mathbb{R}^{p \times K}$ are such that

$$L_{(a)} C_{(a)}^{\mathrm{T}} = E(Y) - BX^{\mathrm{T}} = L_{(b)} C_{(b)}^{\mathrm{T}}.$$

Under Assumptions 1 and 2, $L_{(a)}$ and $L_{(b)}$ have full column rank, meaning we may define

$$R = L_{(a)}^{\mathrm{T}} L_{(b)} \left\{ L_{(b)}^{\mathrm{T}} L_{(b)} \right\}^{-1},$$

where $C_{(b)} = C_{(a)} R$. Since $C_{(a)}$ and $C_{(b)}$ have full column rank by Assumptions 1 and 2, R must be invertible. Therefore, for $\Omega_{(i)} = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} C_{(i)}$ ($i = a, b$),

$$\begin{aligned} \Omega_{(b)} \left\{ n^{-1} C_{(b)}^{\mathrm{T}} P_X^{\perp} C_{(b)} \right\}^{-1} \Omega_{(b)}^{\mathrm{T}} &= \Omega_{(a)} R R^{-1} \left\{ n^{-1} C_{(a)}^{\mathrm{T}} P_X^{\perp} C_{(a)} \right\}^{-1} R^{-\mathrm{T}} R^{\mathrm{T}} \Omega_{(a)}^{\mathrm{T}} \\ &= \Omega_{(a)} \left\{ n^{-1} C_{(a)}^{\mathrm{T}} P_X^{\perp} C_{(a)} \right\}^{-1} \Omega_{(a)}^{\mathrm{T}}, \end{aligned}$$

which completes the proof. □

S2.2. *The behavior of the off-diagonal elements of $np^{-1} L^{\mathrm{T}} \Sigma L$*

Let $m_k \in \mathbb{R}^p$ be the kth left singular vector of LC_2^{T} ($k = 1, \ldots, K$). In this section, we state and prove a proposition regarding the generality of the condition that $(\lambda_r \lambda_s)^{1/2} |m_r^{\mathrm{T}} \Sigma m_s| \leq c_8 \lambda_{\max(r,s)}$ for all $r, s \in [K]$, which is used in the statements of Theorems 2 and 3. To do so, we note that $(\lambda_r \lambda_s)^{1/2} |m_r^{\mathrm{T}} \Sigma m_s| = |np^{-1} L_{*r}^{\mathrm{T}} \Sigma L_{*s}|$ for some L such that $(L, C) \in \Theta_{(0)}$.

PROPOSITION S2. *Let $\bar{L} = \left[\bar{\ell}_1 \cdots \bar{\ell}_p\right]^{\mathrm{T}} \in \mathbb{R}^{p \times K}$ and $\Sigma = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_p^2\right)$ where for each $g \in [p]$,*

$$\bar{\ell}_g \sim F_{\bar{\ell}}$$
$$\sigma_g^2 \sim F_{\sigma^2}$$

where the distributions $F_{\bar{\ell}}$ and F_{σ^2} have compact support. Suppose $C \in \mathbb{R}^{n \times K}$ and $X \in \mathbb{R}^{n \times d}$ are non-random matrices and define $R \in \mathbb{R}^{K \times K}$ such that $R^2 = n^{-1} C^{\mathrm{T}} P_X^{\perp} C$. In addition, let

(a) $\gamma_K \leq \cdots \leq \gamma_1$ *be the eigenvalues of* $np^{-1} \bar{L}^{\mathrm{T}} \bar{L}$
(b) $\lambda_K \leq \cdots \leq \lambda_1$ *be the first K eigenvalues of* $P_X^{\perp} C \left(p^{-1} \bar{L}^{\mathrm{T}} \bar{L}\right) C^{\mathrm{T}} P_X^{\perp}$.
(c) $L = \bar{L} R U$, *where $U \in \mathbb{R}^{K \times K}$ is a unitary matrix such that* $np^{-1} L^{\mathrm{T}} L = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_K\right)$.

Suppose the following assumptions hold:

(i) $\|n^{-1} C^{\mathrm{T}} P_X^{\perp} C\|_2, \|\left(n^{-1} C^{\mathrm{T}} P_X^{\perp} C\right)^{-1}\|_2 \leq c^2$ *for some constant* $c \geq 1$.
(ii) *For any $\epsilon > 0$, there exists a $\delta_\epsilon > 0$ such that* $\mathrm{pr}\left(\gamma_K p / n \geq \delta_\epsilon\right) \geq 1 - \epsilon$ *for all* n, p.

Then for any $r, s \in [K]$,

$$n \left\{ p \lambda_{\max(r,s)} \right\}^{-1} L_{*r}^{\mathrm{T}} \Sigma L_{*s} = O_{\mathrm{p}}(1)$$

as $n, p \to \infty$.

Proof. First, $n\left(\gamma_K p\right)^{-1} = O_{\mathrm{p}}(1)$ by Item (ii). Next, by the sampling mechanism used to draw $\bar{L}$ and Σ, $\bar{L}$ and Σ are independent. Suppose $r \leq s$ and define $\ell_g = L_{g*}$ for all $g \in [p]$. First,

$$E\left[n\left\{p\lambda_{\max(r,s)}\right\}^{-1} L_{*r}^{\mathrm{T}} \Sigma L_{*s} \mid \bar{L}\right] = n\left\{p\lambda_{\max(r,s)}\right\}^{-1} \sum_{g=1}^{p} \ell_{gr}\ell_{gs} E\left(\sigma_g^2 \mid \bar{L}\right)$$

$$= E\left(\sigma_1^2\right) n\left\{p\lambda_{\max(r,s)}\right\}^{-1} \sum_{g=1}^{p} \ell_{gr}\ell_{gs} = 0.$$

Next, $\|\ell_g\|_2 = \|U^{\mathrm{T}} R \bar{\ell}_g\|_2 \leq c\|\bar{\ell}_g\|_2$, meaning $\|\ell_g\|_2^2, \sigma_g^2 \leq a$ for all $g \in [p]$ for some constant $a > 0$ not dependent on n or p (since $F_{\bar{\ell}}$ and F_{σ^2} have compact support). Let $\bar{\Psi} = np^{-1}\bar{L}^T\bar{L}$ and $\Psi = np^{-1}L^T L$. Then

$$\lambda_K^{-1} = \|\Psi^{-1}\|_2 \leq c^2\|\bar{\Psi}^{-1}\|_2 = c^2\gamma_K^{-1}$$

and

$$\gamma_K^{-1} = \|\bar{\Psi}^{-1}\|_2 \leq c^2\|\Psi^{-1}\|_2 = c^2\lambda_K^{-1},$$

which implies $c^{-2}\gamma_K \leq \lambda_K \leq c^2\gamma_K$. Therefore, $n\left(p\lambda_K\right)^{-1} = O_{\mathrm{p}}(1)$ and

$$\operatorname{var}\left[n\left\{p\lambda_{\max(r,s)}\right\}^{-1} L_{*r}^{\mathrm{T}} \Sigma L_{*s} \mid \bar{L}\right] = n^2\left(p\lambda_s\right)^{-2} \sum_{g=1}^{p} \ell_{gr}^2 \ell_{gs}^2 \operatorname{var}\left(\sigma_g^2 \mid \bar{L}\right)$$

$$= \operatorname{var}\left(\sigma_1^2\right) n^2\left(p\lambda_s\right)^{-2} \sum_{g=1}^{p} \ell_{gr}^2 \ell_{gs}^2$$

$$\leq a^2 n^2\left(p\lambda_s\right)^{-2} \sum_{g=1}^{p} \ell_{gs}^2 = a^2 n\left(p\lambda_s\right)^{-1}$$

$$\leq a^2 n\left(p\lambda_K\right)^{-1} = O_{\mathrm{p}}(1).$$

This proves the claim. □

Remark 1. Item (ii) is a weak assumption because $p\gamma_K/n$ is the smallest eigenvalue of $\bar{L}^{\mathrm{T}}\bar{L}$. Further, we assume $p\lambda_K/n \to \infty$ as $n \to \infty$ (where $\gamma_K \asymp \lambda_K$) in Assumption 2.

Remark 2. We can extend to Proposition S2 to include the case that C is random if we assume C is independent of Σ and if for all $\epsilon > 0$, there exists an $M > 0$ such that $\operatorname{pr}\left(\|n^{-1}C^{\mathrm{T}}P_X^{\perp}C\|_2 \leq M\right), \operatorname{pr}\left\{\|\left(n^{-1}C^{\mathrm{T}}P_X^{\perp}C\right)^{-1}\|_2 \leq M\right\} \geq 1-\epsilon$. To prove the claim, we would simply condition on $\bar{L}$ and C instead of just $\bar{L}$.

S2.3. *A re-definition of C_2 and Y_2*

Let $Q_X \in \mathbb{R}^{n\times(n-d)}$ be a matrix whose columns form an orthonormal basis for the null space of X^{T}. For the remainder of the Supplement, we re-define C_2 to be

$$C_2 = Q_X^{\mathrm{T}} C \in \mathbb{R}^{(n-d)\times K}. \tag{S2}$$

Note that since $P_X^{\perp} = Q_X Q_X^{\mathrm{T}}$, $C_2^{\mathrm{T}} C_2 = C^{\mathrm{T}} P_X^{\perp} C$ and the C_2 defined in (3) is simply $P_X^{\perp} C = Q_X\left(Q_X^{\mathrm{T}} C\right) = Q_X C_2$. This implies that the first $n-d$ singular values and left singular vectors of $YP_X^{\perp}$ and YQ_X are the same and if we let $V_1 \in \mathbb{R}^{n\times(n-d)}$ and $V_2 \in \mathbb{R}^{(n-d)\times(n-d)}$ be the right singular vectors of $YP_X^{\perp}$ and YQ_X corresponding to the non-zero singular values, then $V_1 = Q_X V_2$. We therefore replace C_2 defined in (3) with that defined in (S2) in the statements of all remaining propositions, lemmas and theorems, as well as the proofs of all propositions, lemmas and theorems stated in the main text.

Using this definition of C_2, we define Y_1 and Y_2 from to be

$$Y_1 = B + L\Omega^{\mathrm{T}} + \mathcal{E}_1 \tag{S3}$$

$$Y_2 = YQ_X = LC_2^{\mathrm{T}} + \mathcal{E}_2 \tag{S4}$$

where $\mathcal{E}_1 \sim MN_{p\times d}\left\{0, \Sigma, (X^{\mathrm{T}}X)^{-1}\right\}$ and $\mathcal{E}_2 \sim MN_{p\times(n-d)}\left(0, \Sigma, I_{n-d}\right)$ are independent. Note that this Y_1 is the same as the one defined in the main text. To get back to the Y_2 defined in the main text, we simply multiply the Y_2 defined in (S4) on the right by Q_X^{T}. It is easy to see that because Q_X has orthonormal columns and $Q_X Q_X^{\mathrm{T}} = P_X^{\perp}$, Assumptions 1, 2 and 3 are equivalent with this redefinition of C_2 and Y_2. Further, Propositions 1 and S1 hold with $C_2 = Q_X^{\mathrm{T}} C$.

We also define y_{2_g} and y_{1_g} to be the gth rows of Y_1 and Y_2 defined in (S3) and (S4), respectively. If $V \in \mathbb{R}^{(n-d)\times K}$ are the first K right singular vectors of Y_2 defined in (S4), then $\hat{C}_2 = n^{1/2}V$, $\hat{L} = Y_2\hat{C}_2\left(\hat{C}_2^{\mathrm{T}}\hat{C}_2\right)^{-1}$ and $\hat{\sigma}_g^2 = (n-d-K)^{-1} y_{g_2}^{\mathrm{T}} P_{\hat{C}_2}^{\perp} y_{g_2}$. Therefore, none of our estimators for L, Σ, Ω or B change when we use this definition of C_2 and Y_2.

Since Σ, LC_2^{T} and $\lambda_1, \ldots, \lambda_K$ are identifiable under Assumptions 1(a) and 1(b) and $B, L\Omega^{\mathrm{T}}$ and $\Omega^{\mathrm{T}}\left(n^{-1}C_2^{\mathrm{T}}C_2\right)^{-1}\Omega$ are identifiable under Assumptions 1, 2 and 3(a) (see Propositions 1 and S1), then Lemma 1, (17) in Lemma 2 and Theorems 1 and 2 hold regardless of the parametrization of L and C. Therefore, we will assume $(L, C) \in \Theta_{(0)}$ when Assumptions 1 and 2 hold, and will assume $(L, C) \in \Theta_{(1)}$ when 1, 2 and 3(a) hold (again, where $C_2 = Q_X^{\mathrm{T}}C$). The first goal is to understand the asymptotic properties of $\hat{L}$ and $\hat{C}_2$, which are essential to all of the proofs that follow.

S2.4. *Understanding the behavior of $\hat{C}_2$ and $\hat{L}$*

We start by stating and proving Lemmas S1 and S2 and use their results to prove theoretical statements made in the main text. For ease of notation, we assume for the statements and proofs in this subsection (Section S2.4) that

$$Y_{p\times n} = L_{p\times K} C_{K\times n}^{\mathrm{T}} + \mathcal{E}_{p\times n}, \quad \mathcal{E} \sim MN_{p\times n}\left(0, \Sigma, I_n\right) \tag{S5}$$

where $n^{-1}C^{\mathrm{T}}C = I_K$. We also define

$$\tilde{C} = n^{-1/2}C \tag{S6}$$

$$\tilde{L} = n^{1/2}p^{-1/2}L. \tag{S7}$$

We will lastly define a matrix $Q \in \mathbb{R}^{n\times n-K}$ such that $Q^{\mathrm{T}}Q = I_{n-K}$ and $Q_T\tilde{C} = 0_{(n-K)\times K}$. We use a technique developed in Paul (2007) to define the rotated matrix $F_{n\times n}$ to be

$$\begin{aligned} F &= \begin{pmatrix} \tilde{C}^{\mathrm{T}} \\ Q^{\mathrm{T}} \end{pmatrix} p^{-1} Y^{\mathrm{T}} Y \left(\tilde{C}\ Q\right) \\ &= \begin{bmatrix} \left(\tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\left(\tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1\right) & \left(\tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}} p^{-1/2}\tilde{\mathcal{E}}_2 \\ p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(\tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1\right) & p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 \end{bmatrix} \end{aligned} \tag{S8}$$

where $\tilde{\mathcal{E}}_1 = \mathcal{E}\,\tilde{C}$ and $\tilde{\mathcal{E}}_2 = \mathcal{E}\,Q$ are independent. Since $\left(\tilde{C}\ Q\right)$ is a unitary matrix, the eigenvalues of F are also the eigenvalues of $p^{-1}Y^{\mathrm{T}}Y$. For the remainder of the Supplement, we assume

$$\begin{pmatrix} \hat{V}_{K\times K} \\ \hat{Z}_{(n-K)\times K} \end{pmatrix}$$

are the first K eigenvectors of F, meaning $\tilde{C}\hat{V} + Q\hat{Z}$ are the first K eigenvectors of $p^{-1}Y^{\mathrm{T}}Y$. Further, since $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$ are independent, the upper left block of F is independent of $\tilde{\mathcal{E}}_2$. We exploit this by first studying the eigen-structure of the upper left block in Lemma S1, and then using those results to enumerate the asymptotic properties of the first K eigenvalues and eigenvectors of F in Lemma S2. In order to avoid confusing subscripts and superscripts, we define the scalar $v[s]$ to be the sth component of the vector v.

LEMMA S1. *Let* $\tilde{L} \in \mathbb{R}^{p\times K}$, $\tilde{\mathcal{E}}_1 \sim MN_{p\times K}(0, \Sigma, I_K)$ *and* $\tilde{N} = \tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1$. *Assume* $\tilde{L}^{\mathrm{T}}\tilde{L} = \text{diag}(\lambda_1, \ldots, \lambda_K)$ *where the* λ_k*'s are the same as those given in Assumption* 2 *and* $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$ *follows Assumption* 1*(c). If* $d_k^2 = \lambda_k\left(\tilde{N}^{\mathrm{T}}\tilde{N}\right)$ *and* v_k *are the* k^{th} *eigenvalue and eigenvector of* $\tilde{N}^{\mathrm{T}}\tilde{N}$, *then*

$$d_k^2\lambda_k^{-1} = 1 + \rho\lambda_k^{-1} + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\} \tag{S9}$$

and

$$\begin{aligned} v_k = &\left[1 + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1}\right\}\right] e_k + O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1/2}\right\} e_1 + \cdots + O_{\mathrm{p}}\left\{(\lambda_{k-1} p)^{-1/2}\right\} e_{k-1} \\ &+ O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\} e_{k+1} + \cdots + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\} e_K \end{aligned} \tag{S10}$$

where e_k, $k = 1, \ldots, K$, *are the standard basis vectors in* $\mathbb{R}^K$.

Proof. First, $\tilde{N}^{\mathrm{T}}\tilde{N} = \tilde{L}^{\mathrm{T}}\tilde{L} + \rho I_K + p^{-1/2}\tilde{L}^{\mathrm{T}}\tilde{\mathcal{E}}_1 + p^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\tilde{L} + B$ where the entries of B are $O_{\mathrm{p}}\left(p^{-1/2}\right)$. Let $RR^{\mathrm{T}} = \tilde{L}^{\mathrm{T}}\Sigma\tilde{L}$ where R is a lower triangular matrix. By Cauchy-Schwartz, the kth row of R is $R_k^{\mathrm{T}} = O\left(\lambda_k^{1/2}\right)$. We also note that $p^{-1/2}\tilde{L}^{\mathrm{T}}\tilde{\mathcal{E}}_1 \sim RM$ where the entries of $M \in \mathbb{R}^{K\times K}$ are $O_{\mathrm{p}}\left(p^{-1/2}\right)$. If we let the columns of M be M_s ($s \in [K]$), then $[RM]_{ks} = R_k^{\mathrm{T}}M_s = O_{\mathrm{p}}\left\{\left(\lambda_k p^{-1}\right)^{1/2}\right\}$ ($k, s \in [K]$). Next, define the matrix $A^{(1)} \in \mathbb{R}^{K\times K}$ to be

$$A^{(1)} = \lambda_1^{-1}\tilde{N}^{\mathrm{T}}\tilde{N} = \begin{pmatrix} \mu_1 & a_{12} & \cdots & a_{1K} \\ a_{21} & \mu_2 & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & \mu_K \end{pmatrix}$$

where

$$\begin{aligned} \mu_k &= (\lambda_k + \rho)\lambda_1^{-1} + 2\lambda_1^{-1}R_k^{\mathrm{T}}M_k + \lambda_1^{-1}B_{kk} \\ a_{ks} &= \lambda_1^{-1}R_k^{\mathrm{T}}M_s + \lambda_1^{-1}R_s^{\mathrm{T}}M_k + \lambda_1^{-1}B_{sk} = O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_1^{-1}p^{-1/2}\right) \end{aligned}$$

for $k < s$. Our goal is to break $A^{(1)}$ into K rank one pieces, each of which are approximately orthogonal. The procedure is enumerated in four steps:

1. Define $A_1 = A_1^{(1)}, A_2 = \left(0, A_{22}^{(1)}, \ldots, A_{2K}^{(1)}\right), \ldots, A_K = \left(\underbrace{0, \ldots, 0}_{K-1\text{ 0's}}, A_{KK}^{(1)}\right)^{\mathrm{T}}$.
2. We wish to first modify A_1 and A_2 so that they are orthogonal. To do this, we will add ϵ_2 to $A_2[1]$ and remove ϵ_2 from $A_1[2]$. That is, we define $A_{1_2} = A_1 + \epsilon_2 e_2$ and $A_{2_2} = A_2 - \epsilon_2 e_1$ such that
$$0 = A_{1_2}^{\mathrm{T}}A_{2_2} = A_1^{\mathrm{T}}A_2 + \epsilon_2\mu_2 - \epsilon_2\mu_1 = a_{12}\mu_2 + \epsilon_2\mu_2 - \epsilon_2\mu_1 + O_{\mathrm{p}}\left(\lambda_2^{1/2}\lambda_1^{-3/2}p^{-1}\right)$$
meaning $\epsilon_2 = a_{12}\mu_2(\mu_1 - \mu_2)^{-1} + O_{\mathrm{p}}\left(\lambda_2^{1/2}\lambda_1^{-3/2}p^{-1}\right) = O_{\mathrm{p}}\left(\lambda_2\lambda_1^{-3/2}p^{-1/2}\right)$. We now have $A_{1_2}^{\mathrm{T}}A_{2_2} = 0$.
3. Define $A_{1_k} = A_{1_{k-1}} + \epsilon_k e_k$ and $A_{k_2} = A_k - \epsilon_k e_1$ inductively:
$$\begin{aligned} 0 = \left(A_{1_{k-1}} + \epsilon_k e_k\right)^{\mathrm{T}}(A_k - \epsilon_k e_1) &= A_{1_{k-1}}^{\mathrm{T}}A_k + \epsilon_k\mu_k - \epsilon_k\mu_1 = a_{1k}\mu_k + \epsilon_k\mu_k - \epsilon_k\mu_1 \\ &+ O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_1^{-3/2}p^{-1}\right) \end{aligned}$$
meaning $\epsilon_k = a_{1k}\mu_k(\mu_1 - \mu_k)^{-1} + O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_1^{-3/2}p^{-1}\right) = O_{\mathrm{p}}\left(\lambda_k\lambda_1^{-3/2}p^{-1/2}\right)$.

4. After we complete this process $K-1$ times to get A_{1_K}, we now have for $s<K$
$$\begin{aligned}A_{1_K}^{\mathrm{T}}A_{s_2} =&\left(A_1+\epsilon_2e_2+\cdots+\epsilon_Ke_K\right)^{\mathrm{T}}\left(A_s-\epsilon_se_1\right)=\left(A_1+\epsilon_2e_2+\cdots+\epsilon_se_s\right)^{\mathrm{T}}\left(A_s-\epsilon_se_1\right)\\ &+\left(\epsilon_{s+1}e_{s+1}+\cdots+\epsilon_Ke_K\right)^{\mathrm{T}}\left(A_s-\epsilon_se_1\right)=0+\epsilon_{s+1}a_{s,s+1}+\cdots\epsilon_Ka_{s,K}\\ =&O_{\mathrm{p}}\left(\lambda_{s+1}\lambda_1^{-3/2}p^{-1/2}\lambda_s^{1/2}\lambda_1^{-1}p^{-1/2}\right)=O_{\mathrm{p}}\left\{\left(\lambda_s\lambda_1^{-1}\right)^{3/2}\left(\lambda_1p\right)^{-1}\right\}\end{aligned}$$
and $A_{1_K}^{\mathrm{T}}A_{1_K}=\mu_1^2+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$, meaning $\|A_{1_K}\|_2=\mu_1+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$.

We now have
$$A^{(1)}=\underbrace{\begin{pmatrix}A_{1_K} & \rightarrow\\ \downarrow & 0_{(K-1)\times(K-1)}\end{pmatrix}}_{B^{(1)}}+\underbrace{\begin{pmatrix}0 & \uparrow & 0_{1\times(K-2)}\\ \leftarrow & A_{2_2} & \rightarrow\\ 0_{(K-2)\times 1} & \downarrow & 0_{(K-2)\times(K-2)}\end{pmatrix}}_{B^{(2)}}+\cdots+\underbrace{\begin{pmatrix}0_{(K-1)\times(K-1)} & \uparrow\\ \leftarrow & A_{K_2}\end{pmatrix}}_{B^{(K)}}$$
$$=\begin{pmatrix}\mu_1 & a_{12}+\epsilon_2 & \cdots & a_{1K}+\epsilon_K\\ a_{12}+\epsilon_2 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ a_{1K}+\epsilon_K & 0 & \cdots & 0\end{pmatrix}+\begin{pmatrix}0 & -\epsilon_2 & 0 & \cdots & 0\\ -\epsilon_2 & \mu_2 & a_{23} & \cdots & a_{2K}\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ 0 & a_{2K} & 0 & \cdots & 0\end{pmatrix}+\cdots+\begin{pmatrix}0 & \cdots & 0 & -\epsilon_K\\ \vdots & \cdots & \vdots & \vdots\\ 0 & \cdots & 0 & 0\\ -\epsilon_K & \cdots & 0 & \mu_K\end{pmatrix}$$

Define $u_{1_K}=\|A_{1_K}\|_2^{-1}A_{1_K}=\left\{1,\left(a_{12}+\epsilon_2\right)\mu_1^{-1},\ldots,\left(a_{1K}+\epsilon_K\right)\mu_1^{-1}\right\}^{\mathrm{T}}+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$. Then $B^{(1)}=\mu_1u_{1_K}u_{1_K}^{\mathrm{T}}+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$. Further, for $s\in[K]$,
$$\|B^{(s)}u_{1_K}\|_2=\|\begin{pmatrix}-\epsilon_s\left(a_{1s}+\epsilon_s\right)\|A_{1_K}\|_2^{-1}\\ 0\\ \vdots\\ 0\\ \|A_{1_K}\|_2^{-1}A_{s_2}^{\mathrm{T}}A_{1_K}\\ \|A_{1_K}\|_2^{-1}a_{s,s+1}\left(a_{1s}+\epsilon_s\right)\\ \vdots\\ \|A_{1_K}\|_2^{-1}a_{s,K}\left(a_{1s}+\epsilon_s\right)\end{pmatrix}\|_2=O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$$
which means $A^{(1)}u_{1_K}=\mu_1u_{1_K}+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$. We then define
$$\delta=u_{1_K}^{\mathrm{T}}A^{(1)}u_{1_K}=\mu_1+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$$
$$\gamma=\|A^{(1)}u_{1_K}-\delta u_{1_K}\|_2=O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}.$$
By Weyl's Theorem, the eigenvalues of $A^{(1)}$ are $\mu_k+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1/2}\right\}$, so if ξ is the second largest eigenvalue of $A^{(1)}$, $\xi=\mu_2+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1/2}\right\}$, meaning $f=\delta-\xi=\left(\lambda_1-\lambda_2\right)\lambda_1^{-1}+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1/2}\right\}$. By Theorem 3.6 in Auffinger & Tang (2015), we have

1. There exists an eigenvalue λ_γ of $A^{(1)}$ s.t. $\lambda_\gamma\in[\delta-\gamma,\delta+\gamma]$, i.e. $\lambda_\gamma=\mu_1+O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}$.
2. If λ_γ is the only eigenvalue in $[\delta-\gamma,\delta+\gamma]$ and v_γ is the eigenvalue corresponding to λ_γ and $f>\gamma$,
$$\|v_\gamma-u_{1_K}^{\mathrm{T}}v_\gamma u_{1_K}\|_2\le 2\gamma\left(f-\gamma\right)^{-1}=O_{\mathrm{p}}\left\{\left(\lambda_1p\right)^{-1}\right\}.$$

Let $G_{\lambda_\gamma,n,p} = \left\{\lambda_\gamma \text{ is the maximum eigenvalue of } A^{(1)}\right\}$. Then

$$\begin{aligned}\mathrm{pr}\left(\left|\lambda_1\left(A^{(1)}\right) - \mu_1\right| \geq M\right) &\leq \mathrm{pr}\left(\left|\lambda_\gamma - \mu_1\right| \geq M, G_{\lambda_\gamma,n,p}\right) + \mathrm{pr}\left(G^c_{\lambda_\gamma,n,p}\right)\\ &\leq \mathrm{pr}\left(\left|\delta - \mu_1\right| \geq M\right) + \mathrm{pr}\left(G^c_{\lambda_\gamma,n,p}\right)\end{aligned}$$

Since $\mathrm{pr}\left(G^c_{\lambda_\gamma,n,p}\right) \to 0$ and $|\lambda_\gamma - \mu_1| = O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1}\right\}$, $d_1^2\lambda_1^{-1} = \lambda_1\left(A^{(1)}\right) = \mu_1 + O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1}\right\}$. We can apply an identical procedure to show that $\|v_1 - u_{1_K}^{\mathrm{T}} v_1 u_{1_K}\|_2 = O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1}\right\}$ since on the event that λ_γ is the largest eigenvalue of $A^{(1)}$, $\lambda_\gamma - \xi > c + o_{\mathrm{p}}(1)$, where c is a constant that does not depend on n or p (i.e. λ_γ is the only eigenvalue in $[\delta-\gamma, \delta+\gamma]$ and $f > \delta$ with probability tending to 1). Since v_1 and u_{1_K} are unit vectors, we must have $u_{1_K}^{\mathrm{T}} v_1 = \pm 1 + O_{\mathrm{p}}\left\{(\lambda_1 p)^{-2}\right\}$. That is, we know v_1 up to sign parity. □

We then have

$$A^{(2)} = \lambda_2^{-1}\left(\lambda_1 A^{(1)} - d_1^2 v_1 v_1^{\mathrm{T}}\right) = \lambda_1\lambda_2^{-1}B^{(2)} + \cdots + \lambda_1\lambda_2^{-1}B^{(K)} + O_{\mathrm{p}}\left\{(\lambda_2 p)^{-1}\right\}.$$

Since $\epsilon_k\lambda_1\lambda_2^{-1} = O_{\mathrm{p}}\left\{\lambda_k\lambda_2^{-1}(\lambda_1 p)^{-1/2}\right\}$, all off-diagonal entries of the above matrix at most $O_{\mathrm{p}}\left\{(\lambda_2 p)^{-1/2}\right\}$. We can then apply the exact same procedure as we did above to show that for all $k \in [K]$,

$$d_k^2\lambda_k^{-1} = 1 + \rho\lambda_k^{-1} + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\}$$

and

$$v_k = \begin{bmatrix} O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\} \\ \vdots \\ 1 + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1}\right\} \\ \vdots \\ O_{\mathrm{p}}\left\{(\lambda_k p)^{-1/2}\right\}. \end{bmatrix}$$

Lastly, for $s < k$,

$$0 = v_s^{\mathrm{T}} v_k = v_k[s]v_s[s] + O_{\mathrm{p}}\left\{(\lambda_k p)^{-1}\right\} + v_s[k]v_k[k] = v_k[s] + O_{\mathrm{p}}\left\{(\lambda_s p)^{-1/2}\right\}$$

meaning $v_k[s] = O_{\mathrm{p}}\left\{(\lambda_s p)^{-1/2}\right\}$ since $\lambda_s^{1/2}\lambda_k^{-1}p^{-1/2} \to 0$ by assumption. This completes the proof.

We use $\tilde{\mathcal{E}}_1$, $\tilde{\mathcal{E}}_2$, $\tilde{N}$, d_k and v_k defined in Lemma S1 in the remainder of the paper. We also define

$$R = p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 - \rho I_{n-K} \tag{S11}$$

and let $V = \left[v_1 \cdots v_K\right]$, $U = \left[u_1 \cdots u_K\right]$ be the first K right and left singular values of $\tilde{N}$. That is

$$\tilde{N} = \tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1 = UDV^{\mathrm{T}} \tag{S12}$$

is the singular value decomposition of $\tilde{N}$, where $D_{kk} = d_k$. By Theorem 5.39 in Eldar & Kutyniok (2012), $\|R\|_2 = O_{\mathrm{p}}\left\{\left(np^{-1}\right)^{1/2}\right\}$ under Assumptions 1(c) and 2(c). The next lemma uses what we have established in Lemma S1 to prove convergence properties of the first K eigenvalues and eigenvectors of F (see (S8)).

LEMMA S2. *Suppose the probability model for Y is given by* (S5) *and that Assumptions* 1 *and* 2 *hold for $d = 0$ (d is the number of columns in X). Then*

$$\hat{\lambda}_k = \lambda_k(F) = d_k^2 + O_{\mathrm{p}}\left(np^{-1}\right). \tag{S13}$$

Define $\begin{bmatrix}\hat{v}_k\\ \hat{z}_k\end{bmatrix}$, $\hat{v}_k \in \mathbb{R}^K$ *and* $\hat{z}_k \in \mathbb{R}^{n-K}$ *to be the* k^{th} *eigenvector of* F. *Then*

$$\hat{v}_k = v_k + \epsilon_k, \quad \|\epsilon_k\|_2 = O_{\mathrm{p}}\left\{n\left(\lambda_k p\right)^{-1}\right\}. \tag{S14}$$

and

$$\hat{z}_k = d_k\lambda_k^{-1}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}u_k + d_k\lambda_k^{-1}p^{-1/2}R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_k + O_{\mathrm{p}}\left\{n^{3/2}\left(\lambda_k p\right)^{-3/2} + n^{1/2}\left(p\lambda_k\right)^{-1}\right\} \tag{S15}$$

where d_k and v_k are defined (S9) *and* (S10) *and u_k is the k^{th} left singular vector of $Y\tilde{C}$. Further, if $np^{-1}L_{*k}^{\mathrm{T}}\Sigma L_{*s} \leq c\lambda_{\max(k,s)}$ for all $k, s \in [K]$, then for any $s < k$,*

$$\epsilon_k[s] = o_{\mathrm{p}}\left(\lambda_k\lambda_s^{-1}n^{-1/2}\right). \tag{S16}$$

Proof. First, define

$$F^{(1)} = F = \lambda_1\begin{bmatrix}\hat{A}_1 & H_1\\ H_1^{\mathrm{T}} & J_1\end{bmatrix}.$$

We immediately observe from the expression for F in (S8) that

$$\hat{\lambda}_1\lambda_1^{-1} = d_1^2\lambda_1^{-1} + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_1 p\right)^{-1/2}\right\} = \left(\lambda_1 + \rho\right)\lambda_1^{-1} + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_1 p\right)^{-1/2}\right\}$$

by Weyl's Theorem. The eigenvalue equations for $F^{(1)}$ are

$$\hat{\lambda}_1\lambda_1^{-1}\hat{v}_1 = \hat{A}_1\hat{v}_1 + H_1\hat{z}_1$$
$$\hat{\lambda}_1\lambda_1^{-1}\hat{z}_1 = H_1^{\mathrm{T}}\hat{v}_1 + J_1\hat{z}_1$$

which then implies

$$\hat{z}_1 = \left(\hat{\lambda}_1\lambda_1^{-1}I_{n-K} - J_1\right)^{-1}H_1^{\mathrm{T}}\hat{v}_1$$
$$\hat{\lambda}_1\lambda_1^{-1}\hat{v}_1 = \hat{A}_1\hat{v}_1 + H_1\left(\hat{\lambda}_1\lambda_1^{-1}I_{n-K} - J_1\right)^{-1}H_1^{\mathrm{T}}\hat{v}_1$$

where

$$H_1 = \lambda_1^{-1}p^{-1/2}\left(\tilde{L} + p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{\mathcal{E}}_2$$
$$\hat{\lambda}_1\lambda_1^{-1}I_{n-K} - J_1 = \left(\hat{\lambda}_1 - \rho\right)\lambda_1^{-1}I_{n-K} - \lambda_1^{-1}R.$$

The latter is invertible with eigenvalues that are uniformly bounded away from 0 with high probability, since

$$\hat{\lambda}_1\lambda_1^{-1} = \left(\lambda_1 + \rho\right)\lambda_1^{-1} + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_1 p\right)^{-1/2}\right\}$$

and $\|R\|_2 = O_{\mathrm{p}}\left(n^{1/2}p^{-1/2}\right)$. Therefore,

$$\|H_1\left(\hat{\lambda}_1\lambda_1^{-1}I_{n-K} - J_1\right)^{-1}H_1^{\mathrm{T}}\|_2 = O_{\mathrm{p}}\left\{n\left(\lambda_1 p\right)^{-1}\right\}.$$

Since $\hat{A}_1 = A^{(1)}$ (see Lemma S1),

$$\hat{\lambda}_1 \lambda_1^{-1} = \lambda_1 \left\{ A^{(1)} \right\} + O_{\mathrm{p}} \left\{ n \left(\lambda_1 p \right)^{-1} \right\} = d_1^2 \lambda_1^{-1} + O_{\mathrm{p}} \left\{ n \left(\lambda_1 p \right)^{-1} \right\}$$

by Weyl's Theorem. To determine the behavior of $\hat{v}_1$, we first notice that since $\hat{z}_1^{\mathrm{T}} \hat{z}_1 = O_{\mathrm{p}} \left\{ n \left(\lambda_1 p \right)^{-1} \right\}$ and $\|\hat{v}_1\|_2^2 + \|\hat{z}_1\|_2^2 = 1$, $\|\hat{v}_1\|_2 = 1 - O_{\mathrm{p}} \left\{ n \left(\lambda_1 p \right)^{-1} \right\}$. This shows that,

$$\hat{v}_1 = v_1 + O_{\mathrm{p}} \left\{ n \left(\lambda_1 p \right)^{-1} \right\}.$$

Recall from (S12) that $UDV^{\mathrm{T}} = \tilde{L} + p^{-1/2} \tilde{\mathcal{E}}_1$ is the singular value decomposition of $\tilde{L} + p^{-1/2} \tilde{\mathcal{E}}_1$. Using these above relations and the fact that

$$\left(\hat{\lambda}_1 - \rho \right) \lambda_1^{-1} = 1 + O_{\mathrm{p}} \left\{ \left(\lambda_1 p \right)^{-1/2} + n \left(\lambda_1 p \right)^{-1} \right\},$$

we can get an expression for $\hat{z}_1$:

$$\begin{aligned}
\hat{z}_1 &= \left(\hat{\lambda}_1 \lambda_1^{-1} I_{n-K} - J_1 \right)^{-1} H_1^{\mathrm{T}} \hat{v}_1 \\
&= \lambda_1^{-1} p^{-1/2} \left(\hat{\lambda}_1 \lambda_1^{-1} I_{n-K} - \left(\lambda_1 p \right)^{-1} \tilde{\mathcal{E}}_2^{\mathrm{T}} \tilde{\mathcal{E}}_2 \right)^{-1} \tilde{\mathcal{E}}_2^{\mathrm{T}} \left(\tilde{L} + p^{-1/2} \tilde{\mathcal{E}}_1 \right) \hat{v}_1 \\
&= \lambda_1^{-1} p^{-1/2} \left\{ \left(\hat{\lambda}_1 - \rho \right) \lambda_1^{-1} I_{n-K} - \lambda_1^{-1} R \right\}^{-1} \tilde{\mathcal{E}}_2^{\mathrm{T}} UDV^{\mathrm{T}} v_1 + O_{\mathrm{p}} \left\{ n^{3/2} \left(\lambda_1 p \right)^{-3/2} \right\} \\
&= d_1 \lambda_1^{-1} p^{-1/2} \left(I_{n-K} - \lambda_1^{-1} R \right)^{-1} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_1 + O_{\mathrm{p}} \left\{ n^{3/2} \left(\lambda_1 p \right)^{-3/2} + n^{1/2} \left(p \lambda_1 \right)^{-1} \right\} \\
&= d_1 \lambda_1^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_1 + d_1 \lambda_1^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_1 + O_{\mathrm{p}} \left\{ n^{3/2} \left(\lambda_1 p \right)^{-3/2} + n^{1/2} \left(p \lambda_1 \right)^{-1} \right\}
\end{aligned}$$

since

$$\| \left(I_{n-K} - \lambda_1^{-1} R \right)^{-1} - \left(I_{n-K} + \lambda_1^{-1} R \right) \|_2 = O \left(\| \lambda_1^{-2} R^2 \|_2 \right) = O_{\mathrm{p}} \left\{ n \left(\lambda_1^2 p \right)^{-1} \right\}.$$

We can then find expressions for $\hat{\lambda}_k$, $\hat{v}_k$ and $\hat{z}_k$ by induction. First, assume the following three conditions hold for all $s \leq k$, where $k < K$.

$$\hat{\lambda}_s = d_s^2 + O_{\mathrm{p}} \left(n p^{-1} \right) \tag{S17a}$$

$$\hat{v}_s = v_s + O_{\mathrm{p}} \left\{ n \left(\lambda_s p \right)^{-1} \right\} \tag{S17b}$$

$$\hat{z}_s = d_s \lambda_s^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s + d_s \lambda_s^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s + O_{\mathrm{p}} \left\{ n^{3/2} \left(\lambda_s p \right)^{-3/2} + n^{1/2} \left(p \lambda_s \right)^{-1} \right\} \tag{S17c}$$

$$\begin{aligned}
\lambda_s H_s^{\mathrm{T}} &= p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} \tilde{N} - \hat{\lambda}_1 \hat{z}_1 \hat{v}_1^{\mathrm{T}} - \cdots - \hat{\lambda}_{s-1} \hat{z}_{s-1} \hat{v}_{s-1}^{\mathrm{T}} \\
&= O_{\mathrm{p}} \left\{ n^{1/2} \left(\lambda_1 p \right)^{-1/2} \right\} v_1^{\mathrm{T}} + \cdots + O_{\mathrm{p}} \left\{ n^{1/2} \left(\lambda_{s-1} p \right)^{-1/2} \right\} v_{s-1}^{\mathrm{T}} \\
&\quad + p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} \sum_{\ell=s}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}} + O_{\mathrm{p}} \left\{ \lambda_{s-1}^{-1/2} \left(n p^{-1} \right)^{3/2} + n^{1/2} p^{-1} \right\}.
\end{aligned} \tag{S17d}$$

If we can show that these hold for $k+1$, this would prove (S13), (S14) and (S15). To show that the above hold for $k+1$, we first show that (S17d) holds, and then use the result to show that (S17a), (S17b) and then (S17c) hold. Due to the lengthy calculations, we break the proof into four steps for convenience.

1.

$$
\begin{aligned}
\lambda_{k+1} H_{k+1}^{\mathrm{T}} =& p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{N} - \hat{\lambda}_1\hat{z}_1\hat{v}_1^{\mathrm{T}} - \cdots - \hat{\lambda}_k\hat{z}_k\hat{v}_k^{\mathrm{T}} = \lambda_k H_k - \hat{\lambda}_k\hat{z}_k\hat{v}_k^{\mathrm{T}} \\
=& O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_1 p\right)^{-1/2}\right\} v_1^{\mathrm{T}} + \cdots + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k-1} p\right)^{-1/2}\right\} v_{k-1}^{\mathrm{T}} + d_k p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k v_k^{\mathrm{T}} \\
&+ p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}} + O_{\mathrm{p}}\left\{\lambda_{k-1}^{-1/2}\left(np^{-1}\right)^{3/2} + n^{1/2}p^{-1}\right\} \\
&- \left(\hat{\lambda}_k\lambda_k^{-1}\right) d_k p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k\hat{v}_k^{\mathrm{T}} - \left(\hat{\lambda}_k\lambda_k^{-1}\right) R d_k \lambda_k^{-1} p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k \hat{v}_k^{\mathrm{T}} \\
=& O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_1 p\right)^{-1/2}\right\} v_1^{\mathrm{T}} + \cdots + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_k p\right)^{-1/2}\right\} v_k^{\mathrm{T}} \\
&+ O_{\mathrm{p}}\left\{\lambda_k^{-1/2}\left(np^{-1}\right)^{3/2} + n^{1/2}p^{-1}\right\} + p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}}.
\end{aligned}
$$

where third equality follows because

$$
\begin{aligned}
&\left(\hat{\lambda}_k\lambda_k^{-1}\right) = 1 + \rho\lambda_k^{-1} + O_{\mathrm{p}}\left\{\left(\lambda_k p\right)^{-1/2} + n\left(p\lambda_k\right)^{-1}\right\} \\
&d_k p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k\hat{v}_k^{\mathrm{T}} = d_k p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k v_k^{\mathrm{T}} + O_{\mathrm{p}}\left\{\lambda_k^{-1/2}\left(np^{-1}\right)^{3/2}\right\} \\
&\left(\hat{\lambda}_k\lambda_k^{-1}\right) R d_k\lambda_k^{-1} p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}} u_k\hat{v}_k^{\mathrm{T}} = O_{\mathrm{p}}\left(np^{-1}\lambda_k^{-1/2}\right) v_k^{\mathrm{T}} + O_{\mathrm{p}}\left\{\lambda_k^{-1/2}\left(np^{-1}\right)^{3/2}\right\}.
\end{aligned}
$$

This shows (S17d) in the inductive hypothesis also holds for $k+1$, and shows $\|H_{k+1}\|_2 = O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k+1}p\right)^{-1/2}\right\}$.

2. We next see that

$$
\begin{aligned}
\lambda_{k+1}\hat{A}_{k+1} &= \tilde{N}^{\mathrm{T}}\tilde{N} - \hat{\lambda}_1\hat{v}_1\hat{v}_1^{\mathrm{T}} - \cdots - \hat{\lambda}_k\hat{v}_k\hat{v}_k^{\mathrm{T}} = \tilde{N}^{\mathrm{T}}\tilde{N} - d_1^2 v_1 v_1^{\mathrm{T}} - \cdots - d_k^2 v_k v_k^{\mathrm{T}} + O_{\mathrm{p}}\left(np^{-1}\right) \\
&= \lambda_{k+1}A^{(k+1)} + O_{\mathrm{p}}\left(np^{-1}\right)
\end{aligned}
$$

3. Lastly,

$$
\lambda_{k+1}J_{k+1} = p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 - \hat{\lambda}_1\hat{z}_1\hat{z}_1^{\mathrm{T}} - \cdots - \hat{\lambda}_k\hat{z}_k\hat{z}_k^{\mathrm{T}} = p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 + O_{\mathrm{p}}\left(np^{-1}\right).
$$

By the above expressions for $\hat{A}_{k+1}$, H_{k+1} and J_{k+1},

$$
\left(\hat{\lambda}_{k+1} - \rho\right)\lambda_{k+1}^{-1} = \left(d_{k+1}^2 - \rho\right)\lambda_{k+1}^{-1} + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k+1}p\right)^{-1/2}\right\} = 1 + O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k+1}p\right)^{-1/2}\right\}
$$

by Weyl's Theorem. Therefore,

$$
\hat{\lambda}_{k+1}\lambda_{k+1}^{-1}I_{n-K} - J_{k+1} = \left(\hat{\lambda}_{k+1} - \rho\right)\lambda_{k+1}^{-1}I_{n-K} - \lambda_{k+1}^{-1}R + O_{\mathrm{p}}\left\{n\left(\lambda_{k+1}p\right)^{-1}\right\}
$$

is invertible with high probability. We can now compute the eigenvalue equations to get:

4. We can then put parts 1., 2. and 3. together to find expressions for the eigenvalue $\hat{\lambda}_{k+1}$ and components of the eigenvector $\hat{v}_{k+1}$, $\hat{z}_{k+1}$.

   (a)

$$
\begin{aligned}
\hat{\lambda}_{k+1}\lambda_{k+1}^{-1}\hat{v}_{k+1} &= \hat{A}_{k+1}\hat{v}_{k+1} + H_{k+1}^{\mathrm{T}}\left(\hat{\lambda}_{k+1}\lambda_{k+1}^{-1}I_{n-K} - J_{k+1}\right)^{-1} H_{k+1}\hat{v}_{k+1} \\
&= A^{(k+1)}\hat{v}_{k+1} + O_{\mathrm{p}}\left\{n\left(\lambda_{k+1}p\right)^{-1}\right\}
\end{aligned}
$$

(b)

$$\begin{aligned}\hat{z}_{k+1} &= \left(\hat{\lambda}_{k+1}\lambda_{k+1}^{-1}I_{n-K} - J_{k+1}\right)H_{k+1}\hat{v}_{k+1}\\ &= \left[\left(\hat{\lambda}_{k+1}-\rho\right)\lambda_{k+1}^{-1}I_{n-K} - \lambda_{k+1}^{-1}R + O_{\mathrm{p}}\left\{n\left(\lambda_{k+1}p\right)^{-1}\right\}\right]^{-1}\lambda_{k+1}^{-1}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\times\\ &\times\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}}\hat{v}_{k+1} + O_{\mathrm{p}}\left(\lambda_{k+1}^{-1}\lambda_1^{-1/2}n^{1/2}p^{-1/2}\right)v_1^{\mathrm{T}}\hat{v}_{k+1} + \cdots\\ &+ O_{\mathrm{p}}\left(\lambda_{k+1}^{-1}\lambda_k^{-1/2}n^{1/2}p^{-1/2}\right)v_k^{\mathrm{T}}\hat{v}_{k+1} + O_{\mathrm{p}}\left\{\lambda_k^{-1/2}\lambda_{k+1}^{-1}\left(np^{-1}\right)^{3/2}\right\}\\ &+ O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k+1}p\right)^{-1}\right\}\end{aligned}$$

Therefore

$$\|\hat{z}_{k+1}\|_2 = O_{\mathrm{p}}\left\{n^{1/2}\left(\lambda_{k+1}p\right)^{-1/2}\right\},$$

meaning

$$\|\hat{v}_{k+1}\|_2 = 1 - O_{\mathrm{p}}\left\{n\left(\lambda_{k+1}p\right)^{-1}\right\}.$$

We can then use this and what we showed in part a. to get that

$$\begin{aligned}\hat{v}_{k+1} &= v_{k+1} + O_{\mathrm{p}}\left\{n\left(\lambda_{k+1}p\right)^{-1}\right\}\\ \hat{\lambda}_{k+1} &= d_{k+1}^2 + O_{\mathrm{p}}\left(np^{-1}\right)\end{aligned}$$

which means the 1. of the inductive hypothesis applies for $k+1$. Using the fact that for any $s \le k$

$$\begin{aligned}v_s\hat{v}_{k+1} &= O_{\mathrm{p}}\left\{n\left(p\lambda_{k+1}\right)^{-1}\right\}\\ \left(\hat{\lambda}_{k+1}-\rho\right)\lambda_{k+1}^{-1} &= 1 + O_{\mathrm{p}}\left\{\left(\lambda_{k+1}p\right)^{-1/2} + n\left(\lambda_{k+1}p\right)^{-1}\right\},\end{aligned}$$

we can then modify our expression for $\hat{z}_{k+1}$ to get

(c)

$$\begin{aligned}\hat{z}_{k+1} =&\lambda_{k+1}^{-1}p^{-1/2}\left\{\left(\hat{\lambda}_{k+1}-\rho\right)\lambda_{k+1}^{-1}I_{n-K} - \lambda_{k+1}^{-1}R\right\}^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}}\hat{v}_{k+1}\\ &+ O_{\mathrm{p}}\left\{\left(\lambda_{k+1}^2\lambda_1^{1/2}\right)^{-1}\left(np^{-1}\right)^{3/2}\right\} + \cdots + O_{\mathrm{p}}\left\{\left(\lambda_{k+1}^2\lambda_k^{1/2}\right)^{-1}\left(np^{-1}\right)^{3/2}\right\}\\ &+ O_{\mathrm{p}}\left\{n^{3/2}\left(p\lambda_{k+1}\right)^{-3/2} + n^{1/2}\left(\lambda_{k+1}p\right)^{-1}\right\}\\ =&\lambda_{k+1}^{-1}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}}\hat{v}_{k+1} + \lambda_{k+1}^{-2}p^{-1/2}R\tilde{\mathcal{E}}_2^{\mathrm{T}}\sum_{\ell=k+1}^{K} d_\ell u_\ell v_\ell^{\mathrm{T}}\hat{v}_{k+1}\\ &+ O_{\mathrm{p}}\left\{n^{3/2}\left(p\lambda_{k+1}\right)^{-3/2} + n^{1/2}\left(\lambda_{k+1}p\right)^{-1}\right\}\\ =&d_{k+1}\lambda_{k+1}^{-1}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}u_{k+1} + d_{k+1}\lambda_{k+1}^{-2}p^{-1/2}R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_{k+1}\\ &+ O_{\mathrm{p}}\left\{n^{3/2}\left(p\lambda_{k+1}\right)^{-3/2} + n^{1/2}\left(\lambda_{k+1}p\right)^{-1}\right\}.\end{aligned}$$

This completes the proof by induction and therefore proves (S13), (S14) and (S15). It remains to show (S16).

Since F is symmetric with distinct eigenvalues (with probability 1), for $s < k$ (i.e. $\lambda_s > \lambda_k$),

$$0 = \hat{v}_s^{\mathrm{T}} \hat{v}_k + \hat{z}_s^{\mathrm{T}} \hat{z}_k = (v_s + \epsilon_s)^{\mathrm{T}} (v_k + \epsilon_k) + \hat{z}_s^{\mathrm{T}} \hat{z}_k = 0 + \epsilon_s^{\mathrm{T}} \hat{v}_k + v_s^{\mathrm{T}} \epsilon_k + \hat{z}_s^{\mathrm{T}} \hat{z}_k.$$

where

$$\epsilon_s^{\mathrm{T}} \hat{v}_k = O_{\mathrm{p}} \left\{ n (p\lambda_s)^{-1} \right\} = o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$$
$$v_s^{\mathrm{T}} \epsilon_k = \epsilon_k[s] + O_{\mathrm{p}} \left\{ (\lambda_s p)^{-1/2} n (p\lambda_k)^{-1} + n \left(p^2 \lambda_s \lambda_k \right)^{-1} \right\} = \epsilon_k[s] + o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$$

Therefore, if we can show $\hat{z}_s^{\mathrm{T}} \hat{z}_k = o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$, we must have $\epsilon_k[s] = o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$. By our above expression for $\hat{z}_k$,

$$\hat{z}_s^{\mathrm{T}} \hat{z}_k = \left[d_s \lambda_s^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s + d_s \lambda_s^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s + O_{\mathrm{p}} \left\{ n^{3/2} (p\lambda_s)^{-3/2} + n^{1/2} (\lambda_s p)^{-1} \right\} \right]^{\mathrm{T}} \times$$
$$\times \left[d_k \lambda_k^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k + d_k \lambda_k^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k + O_{\mathrm{p}} \left\{ n^{3/2} (p\lambda_k)^{-3/2} + n^{1/2} (\lambda_k p)^{-1} \right\} \right]$$

We see that

$$O_{\mathrm{p}} \left\{ n^{3/2} (p\lambda_k)^{-3/2} + n^{1/2} (\lambda_k p)^{-1} \right\} \|\hat{z}_s\|_2 = O_{\mathrm{p}} \left\{ \left(np^{-1} \right)^2 \lambda_k^{-3/2} \lambda_s^{-1/2} + n (p\lambda_k)^{-1} (p\lambda_s)^{-1/2} \right\}$$
$$= o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$$

$$\| d_s \lambda_s^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s + O_{\mathrm{p}} \left\{ n^{3/2} (p\lambda_s)^{-3/2} + n^{1/2} (\lambda_s p)^{-1} \right\} \|_2 \|\hat{z}_k\|_2 = O_{\mathrm{p}} \left\{ n (p\lambda_s)^{-1} \right\}$$
$$= o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$$

Therefore,

$$\hat{z}_s^{\mathrm{T}} \hat{z}_k = \left(d_s \lambda_s^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_s \right)^{\mathrm{T}} \left(d_k \lambda_k^{-1} p^{-1/2} \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k + d_k \lambda_k^{-2} p^{-1/2} R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k \right) + o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$$
$$= d_s d_k (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k + d_s d_k \left(\lambda_s \lambda_k^2 p \right)^{-1} u_s^{\mathrm{T}} \tilde{\mathcal{E}}_2 R \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k + o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right).$$

We analyze the two terms in the above equation in 1. and 2. below.

1. Define $U_{s,k} = (u_s \; u_k)$, $W = \left(U_{s,k}^{\mathrm{T}} \Sigma U_{s,k} \right)^{1/2}$ and let $M \sim MN_{(n-K)\times 2}(0, I_{n-K}, I_2)$. Then

$$d_s d_k (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} u_k = \left[U_{s,k}^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} U_{s,k} \right]_{1,2} \overset{\mathcal{D}}{=} \left[d_s d_k n (\lambda_s \lambda_k p)^{-1} W \left(n^{-1} M^{\mathrm{T}} M \right) W \right]_{1,2}$$
$$= d_s d_k n (\lambda_s \lambda_k p)^{-1} \left[W^2 + O_{\mathrm{p}} \left(n^{-1/2} \right) \right]_{1,2}$$
$$= d_s d_k n (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \Sigma u_k + O_{\mathrm{p}} \left(n^{1/2} \lambda_s^{-1/2} \lambda_k^{-1/2} p^{-1} \right)$$
$$= d_s d_k n (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \Sigma u_k + o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right).$$

If $\Sigma = \sigma^2 I_p$, we would be done. However, if Σ were arbitrary then under no assumptions $u_s^{\mathrm{T}} \Sigma u_k = O_{\mathrm{p}}(1)$, meaning $d_s d_k n (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \Sigma u_k = O_{\mathrm{p}} \left(\lambda_s^{-1/2} \lambda_k^{-1/2} n p^{-1} \right)$ which is not necessarily $o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$. To see this, if $\lambda_s = n$ and $\lambda_k = 1$ then $O_{\mathrm{p}} \left(\lambda_s^{-1/2} \lambda_k^{-1/2} n p^{-1} \right) = O_{\mathrm{p}} \left(n^{1/2} p^{-1} \right)$, which is not $o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$. We will use the assumption that $np^{-1} L_{*k}^{\mathrm{T}} \Sigma L_{*s} = O_{\mathrm{p}} \left\{ \lambda_{\max(k,s)} \right\}$ in the statement of the lemma to show that $u_s^{\mathrm{T}} \Sigma u_k = O_{\mathrm{p}} \left(\lambda_k^{1/2} \lambda_s^{-1/2} \right)$. If this were the case, we would have $d_s d_k n (\lambda_s \lambda_k p)^{-1} u_s^{\mathrm{T}} \Sigma u_k = O_{\mathrm{p}} \left\{ n (\lambda_s p)^{-1} \right\} = o_{\mathrm{p}} \left(\lambda_k \lambda_s^{-1} n^{-1/2} \right)$.

Lemma S5 in Section S2.10 proves $u_s^{\mathrm{T}}\Sigma u_k = O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_s^{-1/2}\right)$ under the assumption that $np^{-1}L_{*k}^{\mathrm{T}}\Sigma L_{*s} = O_{\mathrm{p}}\left\{\lambda_{\max(k,s)}\right\}$.

2. Recall that $R = p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 - \rho I_{n-K}$. We will prove a lemma that shows $p^{-1}u_s^{\mathrm{T}}\tilde{\mathcal{E}}_2 R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_k = O_{\mathrm{p}}\left\{\left(np^{-1}\right)^2 + np^{-3/2}\right\}$. Once we prove the lemma, we will have $d_s d_k\left(\lambda_s\lambda_k^2 p\right)^{-1}u_s^{\mathrm{T}}\tilde{\mathcal{E}}_2 R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_k = o_{\mathrm{p}}\left(\lambda_k\lambda_s^{-1}n^{-1/2}\right)$. We prove this in Lemma S6 in Section S2.10.

This proves (S16) and completes the proof.

S2.5. *Proof of Lemmas* 1 *and* 2

In this section, we prove Lemmas 1 and 2. To do so, we first prove a modified version of Lemma 2 in which we modify (18) to be

$$n^{1/2}\left(\hat{\ell}_g - \ell_g\right) \overset{\mathcal{D}}{=} \sigma_g W + o_{\mathrm{p}}(1),$$

where $W \sim N_K\left(0, I_K\right)$. We then prove Lemma 1. (18) from Lemma 2 then follows. For ease of notation, we use the definition of Y from Section S2.4 defined in (S5).

Proof of Lemma 2. We first note that (17) is a direct consequence of (S9) in Lemma S1 and (S13) in Lemma S2. It therefore remains to prove (18), the asymptotic distribution of $\hat{\ell}_g$. Define y_g and $\tilde{e}_{i,g}$ to be the g^{th} row of Y and $\tilde{\mathcal{E}}_i$ $(i = 1, 2)$.

$$n^{1/2}\hat{\ell}_g = \hat{\tilde{C}}^{\mathrm{T}}y_g = \left(\hat{V}^{\mathrm{T}}\tilde{C}^{\mathrm{T}} + \hat{Z}^{\mathrm{T}}Q^{\mathrm{T}}\right)y_g = n^{1/2}\hat{V}^{\mathrm{T}}\ell_g + \hat{V}^{\mathrm{T}}\tilde{e}_{g,1} + \hat{Z}^{\mathrm{T}}\tilde{e}_{2,g}.$$

We then have

$$n^{1/2}\hat{V}^{\mathrm{T}}\ell_g = n^{1/2}\ell_g + n^{1/2}O_{\mathrm{p}}\left\{(p\lambda_K)^{-1/2} + n\,(p\lambda_K)^{-1}\right\}$$

$$\hat{V}^{\mathrm{T}}\tilde{e}_{g,1} \sim N\left(0, \sigma_g^2 I_K\right) + O_{\mathrm{p}}\left\{(p\lambda_K)^{-1/2} + n\,(p\lambda_K)^{-1}\right\}$$

$$\hat{z}_k^{\mathrm{T}}\tilde{e}_{2,g} = d_k\lambda_k^{-1}p^{-1/2}u_k[g]\tilde{e}_{g,2}^{\mathrm{T}}\tilde{e}_{g,2} + d_k\lambda_k^{-1}p^{-1/2}u_k[-g]^{\mathrm{T}}\tilde{\mathcal{E}}_2[-g,]\tilde{e}_{g,2} + O_{\mathrm{p}}\left\{n^{3/2}\,(p\lambda_k)^{-1}\right\}$$

where

$$u_k[g] = n^{1/2}d_k^{-1}p^{-1/2}\left(\ell_g + n^{-1/2}\tilde{e}_{g,1}\right)^{\mathrm{T}}v_k = O_{\mathrm{p}}\left(n^{1/2}\lambda_k^{-1/2}p^{-1/2}\right).$$

Therefore, $d_k\lambda_k^{-1}p^{-1/2}u_k[g]\tilde{e}_{g,2}^{\mathrm{T}}\tilde{e}_{g,2} = O_{\mathrm{p}}\left\{n^{3/2}\,(p\lambda_k)^{-1}\right\}$. Lastly,

$$u_k[-g]^{\mathrm{T}}\tilde{\mathcal{E}}_2[-g,]\tilde{e}_{g,2} \sim N\left(0, u_k[-g]^{\mathrm{T}}\Sigma[-g]u_k[-g]\tilde{e}_{g,2}^{\mathrm{T}}\tilde{e}_{g,2}\right) = O_{\mathrm{p}}\left(n^{1/2}\right).$$

Therefore, $\hat{Z}^{\mathrm{T}}\tilde{e}_{2,g} = O_{\mathrm{p}}\left\{n^{3/2}\,(p\lambda_K)^{-1} + n^{1/2}\,(p\lambda_K)^{-1/2}\right\}$, which means $n^{1/2}\left(\hat{\ell}_g - \ell_g\right) \overset{\mathcal{D}}{\to} N_K\left(0, \sigma_g^2 I_K\right)$.

We also note that this also shows that

$$n^{1/2}\|\hat{\ell}_g^{\mathrm{OLS}} - \hat{\ell}_g\|_2 = o_{\mathrm{p}}(1)$$

where $\hat{\ell}_g^{\mathrm{OLS}} = \ell_g + n^{-1/2}\tilde{e}_{g,1}$ is the ordinary least squares estimate for ℓ_g when C is known, since

$$n^{1/2}\|\hat{V}^{\mathrm{T}}\ell_g - \ell_g\|_2, \|\hat{V}^{\mathrm{T}}\tilde{e}_{g,1} - \tilde{e}_{g,1}\|_2 = o_{\mathrm{p}}(1).$$

Proof of Lemma 1. Once we estimate C by singular value decomposition, we let

$$\hat{\sigma}_g^2 = (n-K)^{-1} y_{2,g}^{\mathrm{T}} P_{\hat{C}}^{\perp} y_{2,g}$$

for each site $g = 1, \ldots, p$. We will prove (15) and (16) by showing the following:

(a) $\hat{\sigma}_g^2 = \sigma_g^2 + O_{\mathrm{P}}\left\{n^{-1/2} + n^{1/2}\left(p\lambda_K\right)^{-1/2}\right\} = \sigma_g^2 + o_{\mathrm{P}}(1)$.

(b) $\hat{\rho} = p^{-1}\sum_{g=1}^{p}\hat{\sigma}_g^2 = \rho + O_{\mathrm{P}}\left\{\left(p\lambda_K\right)^{-1/2} + n\left(p\lambda_K\right)^{-1}\right\} = \rho + o_{\mathrm{P}}\left(n^{-1/2}\right)$. □

We first define the estimated scaled covariates $\hat{W} = n^{-1/2}\hat{C} = \tilde{C}\hat{V} + Q\hat{Z} \in \mathbb{R}^{n\times K}$, where $\hat{V}$, $\hat{Z}$, $\tilde{C}^{\mathrm{T}}$ and Q^{T} are given in Lemmas S1 and S2. Also, define $\epsilon = \left[\epsilon_1 \cdots \epsilon_K\right]$, where $\epsilon_k \in \mathbb{R}^K$ is as defined in (S14) of Lemma S2. We then see that

$$\begin{aligned}(n-K)\,\hat{\sigma}_g^2 &= y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}P_{\hat{W}}y_g = y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\hat{W}\hat{W}^{\mathrm{T}}y_g = y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\left(\tilde{C}\hat{V} + Q\hat{Z}\right)\left(\hat{V}^{\mathrm{T}}\tilde{C}^{\mathrm{T}} + \hat{Z}^{\mathrm{T}}Q^{\mathrm{T}}\right)y_g \\ &= \underbrace{\left(y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\tilde{C}\hat{V}\hat{V}^{\mathrm{T}}\tilde{C}^{\mathrm{T}}y_g\right)}_{(1)} - 2\underbrace{y_g^{\mathrm{T}}\tilde{C}\hat{V}\hat{Z}^{\mathrm{T}}Q^{\mathrm{T}}y_g}_{(2)} - \underbrace{y_g^{\mathrm{T}}Q\hat{Z}\hat{Z}^{\mathrm{T}}Q^{\mathrm{T}}y_g}_{(3)}\end{aligned}$$

We define $\tilde{e}_{g,1}$ and $\tilde{e}_{g,2}$ to be the gth rows of $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$, respectively, and derive the asymptotic properties of (1), (2) and (3) to show (a) and (b) above.

(1)

$$y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\tilde{C}\hat{V}\hat{V}^{\mathrm{T}}\tilde{C}^{\mathrm{T}}y_g = \underbrace{y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\tilde{C}\tilde{C}^{\mathrm{T}}y_g}_{\text{(i)}} + 2\underbrace{y_g^{\mathrm{T}}\tilde{C}\delta^{\mathrm{T}}\tilde{C}^{\mathrm{T}}y_g}_{\text{(ii)}} + \underbrace{y_g^{\mathrm{T}}\tilde{C}\delta^{\mathrm{T}}\delta\tilde{C}^{\mathrm{T}}y_g}_{\text{(iii)}}$$

where $\delta = \hat{V} - I_K$.

(a)

$$\begin{aligned}(n-K)^{-1}\left(y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\tilde{C}\hat{V}\hat{V}^{\mathrm{T}}\tilde{C}^{\mathrm{T}}y_g\right) &= \hat{\sigma}_{g,\mathrm{OLS}}^2 + O_{\mathrm{P}}\left\{\left(\lambda_k p\right)^{-1/2} + n\left(\lambda_k p\right)^{-1}\right\} \\ &= \sigma_g^2 + O_{\mathrm{P}}\left(n^{-1/2}\right)\end{aligned}$$

(b) (i)

$$(n-K)^{-1}p^{-1}\sum_{g=1}^{p}\left(y_g^{\mathrm{T}}y_g - y_g^{\mathrm{T}}\tilde{C}\tilde{C}^{\mathrm{T}}y_g\right) = p^{-1}\sum_{g=1}^{p}\hat{\sigma}_{g,\mathrm{OLS}}^2 = \rho + O_{\mathrm{P}}\left\{(np)^{-1/2}\right\}$$

(ii)

$$\begin{aligned}|(np)^{-1}\sum_{g=1}^{p}y_g^{\mathrm{T}}\tilde{C}\delta^{\mathrm{T}}\tilde{C}^{\mathrm{T}}y_g| &\le \|\delta\|_2 p^{-1}\sum_{g=1}^{p}\left(\ell_g + n^{-1/2}\tilde{e}_{g,1}\right)^{\mathrm{T}}\left(\ell_g + n^{-1/2}\tilde{e}_{g,1}\right) \\ &= O_{\mathrm{P}}\left\{\left(\lambda_k p\right)^{-1/2} + n\left(\lambda_k p\right)^{-1}\right\}\end{aligned}$$

(iii)

$$(np)^{-1}\sum_{g=1}^{p}y_g^{\mathrm{T}}\tilde{C}\delta^{\mathrm{T}}\delta\tilde{C}^{\mathrm{T}}y_g = o_{\mathrm{P}}\left\{\left(\lambda_k p\right)^{-1/2} + n\left(\lambda_k p\right)^{-1}\right\}$$

(2)

$$(n-K)^{-1}y_g^{\mathrm{T}}Q\hat{Z}\hat{Z}^{\mathrm{T}}Q^{\mathrm{T}}y_g = (n-K)^{-1}\tilde{e}_{g,2}^{\mathrm{T}}\hat{Z}\hat{Z}^{\mathrm{T}}\tilde{e}_{g,2} \le \|\hat{Z}\|_2^2(n-K)^{-1}\tilde{e}_{g,2}^{\mathrm{T}}\tilde{e}_{g,2}$$

where $\|\hat{Z}\|_2^2 = O_{\mathrm{P}}\left\{n\left(\lambda_K p\right)^{-1}\right\}$ and $(n-K)^{-1}\tilde{e}_{g,2}^{\mathrm{T}}\tilde{e}_{g,2} = O_{\mathrm{P}}(1)$.

(a)

$$(n-K)^{-1} y_g^{\mathrm{T}} Q \hat{Z} \hat{Z}^{\mathrm{T}} Q^{\mathrm{T}} y_g = O_{\mathrm{p}} \left\{ n \left(\lambda_K p \right)^{-1} \right\}.$$

(b)

$$(n-K)^{-1} p^{-1} \sum_{g=1}^{p} y_g^{\mathrm{T}} Q \hat{Z} \hat{Z}^{\mathrm{T}} Q^{\mathrm{T}} y_g \leq \|\hat{Z}\|_2^2 p^{-1} \sum_{g=1}^{p} (n-K)^{-1} \tilde{e}_{g,2}^{\mathrm{T}} \tilde{e}_{g,2} = O_{\mathrm{p}} \left\{ n \left(\lambda_K p \right)^{-1} \right\}.$$

(3)

$$\begin{aligned} n^{-1} y_g^{\mathrm{T}} \tilde{C} \hat{V} \hat{Z}^{\mathrm{T}} Q^{\mathrm{T}} y_g &= \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \hat{V} \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \\ &= \underbrace{\left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} V \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2}}_{\text{(i)}} + \underbrace{\left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \epsilon \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2}}_{\text{(ii)}}. \end{aligned}$$

(a)

$$\begin{aligned} \left| \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \hat{V} \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \right| &\leq \left\| \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \hat{V} \right\|_2 \|\hat{Z}^{\mathrm{T}}\|_2 \|n^{-1/2} \tilde{e}_{g,2}\|_2 \\ &= O_{\mathrm{p}} \left\{ n^{1/2} \left(p \lambda_K \right)^{-1/2} \right\} \end{aligned}$$

(b) (i)

$$\begin{aligned} & \left| p^{-1} \sum_{g=1}^{p} \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \epsilon \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \right| \\ & \leq \underbrace{\left\{ p^{-1} \sum_{g=1}^{p} \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} \epsilon \epsilon^{\mathrm{T}} \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right) \right\}^{1/2}}_{O_{\mathrm{p}} \left\{ n (p \lambda_K)^{-1} \right\}} \underbrace{\left(p^{-1} \sum_{g=1}^{p} n^{-1} \tilde{e}_{g,2}^{\mathrm{T}} \hat{Z} \hat{Z}^{\mathrm{T}} \tilde{e}_{g,2} \right)^{1/2}}_{O_{\mathrm{p}} \left\{ n^{1/2} (p \lambda_K)^{-1/2} \right\}} \\ & = o_{\mathrm{p}} \left\{ n \left(p \lambda_K \right)^{-1} \right\} \end{aligned}$$

(ii)

$$\begin{aligned} \left(\ell_g + n^{-1/2} \tilde{e}_{g,1} \right)^{\mathrm{T}} V \hat{Z}^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} &= p^{1/2} n^{-1/2} \left(d_1 u_1[g] \cdots d_K u_K[g] \right) \begin{pmatrix} \hat{z}_1^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \\ \vdots \\ \hat{z}_K^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \end{pmatrix} \\ &= p^{1/2} n^{-1/2} d_1 u_1[g] \hat{z}_1^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} + \cdots \\ &\quad + p^{1/2} n^{-1/2} d_K u_K[g] \hat{z}_K^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} \end{aligned}$$

and for any $k \in [K]$,

$$\begin{aligned} p^{-1} \sum_{g=1}^{p} p^{1/2} n^{-1/2} d_k u_k[g] \hat{z}_k^{\mathrm{T}} n^{-1/2} \tilde{e}_{g,2} &= (np)^{-1/2} d_k \hat{z}_k^{\mathrm{T}} \sum_{g=1}^{p} u_k[g] n^{-1/2} \tilde{e}_{g,2} \\ &= O_{\mathrm{p}} \left(p^{-1} \right). \end{aligned}$$

The second equality follows because

$$\begin{aligned} & (np)^{-1/2} d_k \hat{z}_k^{\mathrm{T}} = O_{\mathrm{p}} \left(p^{-1} \right) \\ & \sum_{g=1}^{p} u_k[g] n^{-1/2} \tilde{e}_{g,2} \sim n^{-1/2} N_n \left(0, u_k^{\mathrm{T}} \Sigma u_K I_n \right) = O_{\mathrm{p}} \left(1 \right). \end{aligned}$$

This completes the proof.

S2.6. *Proof of* (20) *from Lemma* 3 *under the conditions of Theorem* 2

In this section, and for the remainder of the Supplement, we return to assuming Y is distributed according (1a). However, we continue to use $\tilde{\mathcal{E}}_1,\ \tilde{\mathcal{E}}_2 \in \mathbb{R}^{p\times(n-d-K)},\ \tilde{N},\ v_k,\ V,\ \hat{v}_k,\ \hat{V},\ \hat{z}_k \in \mathbb{R}^{n-d-K}$, $\hat{Z} \in \mathbb{R}^{(n-d-K)\times K}$ and $R \in \mathbb{R}^{(n-d-K)\times(n-d-K)}$ defined in Lemmas S1 and S2 in Section S2.4 in what follows.

We now prove Lemma S3, which will be useful in the proof of Theorems 1 and 2, and also acts as a proof of Lemma 3.

LEMMA S3. *Suppose the conditions of Theorem* 2 *hold and the diagonal elements of* $\hat{C}_2^{\mathrm{T}} C_2$ *are non-negative. Then*

$$n^{1/2}\|\hat{\Omega}-\Omega\|_2 = o_{\mathrm{P}}(1)$$

where $\hat{\Omega}$ *is defined in* (10).

Proof. Recall

$$\hat{\Omega}^{\mathrm{T}} = \mathrm{diag}\left\{\hat{\lambda}_1\left(\hat{\lambda}_1-\hat{\rho}\right)^{-1},\ldots,\hat{\lambda}_K\left(\hat{\lambda}_K-\hat{\rho}\right)^{-1}\right\}\left(\hat{L}^{\mathrm{T}}\hat{L}\right)^{-1}\hat{L}^{\mathrm{T}}Y_1 =$$

$$\begin{pmatrix}\frac{\hat{\lambda}_1}{\hat{\lambda}_1-\hat{\rho}} & & \\ & \ddots & \\ & & \frac{\hat{\lambda}_K}{\hat{\lambda}_K-\hat{\rho}}\end{pmatrix}\begin{pmatrix}\frac{\lambda_1}{\hat{\lambda}_1} & & \\ & \ddots & \\ & & \frac{\lambda_K}{\hat{\lambda}_K}\end{pmatrix}\left\{\underbrace{\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}B}_{(a)}+\underbrace{\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}L\,\Omega^{\mathrm{T}}}_{(b)}+\underbrace{\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}\mathcal{E}_1}_{(c)}\right\}.$$

We will go through each one of these terms to prove $\|\hat{\Omega}-\Omega\|_2 = o_p\left(n^{-1/2}\right)$.

(a) $M_a = \left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}B = n^{1/2}p^{-1/2}\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\hat{\tilde{L}}^{\mathrm{T}}B$. Define $n^{1/2}p^{-1/2}B = \tilde{B}$ and let $M_a[k,]$ be the kth row of M_a.

$$M_a[k,] = \lambda_k^{-1}\hat{v}_k^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{B}+\lambda_k^{-1}p^{-1/2}\hat{z}_k^{\mathrm{T}}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{B} = \lambda_k^{-1}\hat{v}_k^{\mathrm{T}}\tilde{L}^{\mathrm{T}}\tilde{B}+o_{\mathrm{P}}\left(n^{-1/2}\right)$$

where the second equality follows because $B_{*j}^{\mathrm{T}}B_{*j} = o\left(n^{-3/2}p\lambda_K\right)$ for all $j=1,\ldots,d$ by Assumption 3 and

$$\lambda_k^{-1}p^{-1/2}\hat{z}_k^{\mathrm{T}}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{B} = O_{\mathrm{P}}\left\{n\left(\lambda_k p\right)^{-1}\|n\left(\lambda_k p\right)^{-1}B^{\mathrm{T}}B\|_2^{1/2}\right\} = o_{\mathrm{P}}\left(n^{-1/2}\right)$$
$$\hat{v}_k^{\mathrm{T}}\left(\lambda_k p\right)^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\left(\lambda_k^{-1/2}\tilde{B}\right) = O_{\mathrm{P}}\left\{\left(\lambda_k p\right)^{-1/2}\|n\left(\lambda_k p\right)^{-1}B^{\mathrm{T}}B\|_2^{1/2}\right\} = o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Lastly, the s,j element of of $\lambda_k^{-1}\tilde{L}^{\mathrm{T}}\tilde{B} \in \mathbb{R}^{K\times d}$ is such that

$$\begin{aligned}|n\left(\lambda_k p\right)^{-1}\sum_{g=1}^{p}\ell_{gs}\beta_{gj}| &\leq n\left(\lambda_k p\right)^{-1}\left\{c+o_{\mathrm{P}}(1)\right\}\sum_{g=1}^{p}I\left(\beta_{gj}\neq 0\right) = n\lambda_k^{-1}\left\{c+o_{\mathrm{P}}(1)\right\}\delta_j\\ &= o_{\mathrm{P}}\left(n^{-1/2}\right)\end{aligned}$$

by Assumption 3, where $\delta_j = p^{-1}\sum_{g=1}^{p}I\left(B_{gj}\neq 0\right)$ and $c>0$ is a constant that does not depend on n or p. The first inequality above is because the magnitude of the entries of B and L are bounded by a constant by Assumptions 2 and 3. Therefore, $\|\lambda_k^{-1}\hat{v}_k^{\mathrm{T}}\tilde{L}^{\mathrm{T}}\tilde{B}\|_2 = o_{\mathrm{P}}\left(n^{-1/2}\right)$ for all $k=1,\ldots,K$.

(b) $\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}L = \left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\hat{\tilde{L}}^{\mathrm{T}}\tilde{L} = \begin{pmatrix} \lambda_1^{-1} & & \\ & \ddots & \\ & & \lambda_K^{-1} \end{pmatrix}\hat{\tilde{L}}^{\mathrm{T}}\tilde{L}$ where

$$\begin{aligned}\hat{\tilde{L}}^{\mathrm{T}}\tilde{L} &= \hat{V}^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L} + \hat{Z}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{L}\\ &= \underbrace{\epsilon^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L}}_{\text{(i)}} + \underbrace{V^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L}}_{\text{(ii)}} + \underbrace{\hat{Z}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)}_{\text{(iii)}} + O_{\mathrm{p}}\left(np^{-1}\right)\end{aligned}$$

(i) Suppose $\epsilon = \left(\epsilon_1 \cdots \epsilon_K\right)$ where $\epsilon_k \in \mathbb{R}^K$ was defined in Lemma S2 as $\hat{v}_k - v_k$. Since $\epsilon = O_{\mathrm{p}}\left\{n\left(\lambda_K p\right)^{-1}\right\}$ and $p^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\tilde{L} = O_{\mathrm{p}}\left(\lambda_1^{1/2}p^{-1/2}\right)$, then

$$\|\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\epsilon^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\tilde{L}\|_2 = \lambda_K^{-1/2}O_{\mathrm{p}}\left\{\frac{n}{p\lambda_K}\left(\frac{\lambda_1}{\lambda_K p}\right)^{1/2}\right\} = o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Next,

$$\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\epsilon^{\mathrm{T}}\tilde{L}^{\mathrm{T}}\tilde{L} = \begin{pmatrix} \epsilon_1[1] & \frac{\lambda_2}{\lambda_1}\epsilon_1[2] & \cdots & \frac{\lambda_K}{\lambda_1}\epsilon_1[K] \\ \vdots & \ddots & \cdots & \vdots \\ \frac{\lambda_1}{\lambda_K}\epsilon_K[1] & \frac{\lambda_2}{\lambda_K}\epsilon_K[2] & \cdots & \epsilon_K[K] \end{pmatrix} \underbrace{=}_{\text{Lemma S2}} o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Therefore, $\epsilon^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L} = o_{\mathrm{p}}\left(n^{-1/2}\right)$.

(ii) $\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}V^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L}$

$$\begin{aligned}V^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L} =& V^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right) - V^{\mathrm{T}}\tilde{L}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_1\\ &- \rho V^{\mathrm{T}} + O_{\mathrm{p}}\left(p^{-1/2}\right)\\ =& \operatorname{diag}\left(d_1^2-\rho,\ldots,d_K^2-\rho\right)V^{\mathrm{T}} - V^{\mathrm{T}}\tilde{L}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_1\\ &+ O_{\mathrm{p}}\left(p^{-1/2}\right)\\ =& \operatorname{diag}\left(d_1^2-\rho,\ldots,d_K^2-\rho\right)V^{\mathrm{T}} - V^{\mathrm{T}}\begin{bmatrix} \leftarrow O_{\mathrm{p}}\left(\frac{\lambda_1^{1/2}}{p^{1/2}}\right) \rightarrow \\ \vdots \quad \ddots \quad \vdots \\ \leftarrow O_{\mathrm{p}}\left(\frac{\lambda_K^{1/2}}{p^{1/2}}\right) \rightarrow \end{bmatrix}\\ &+ O_{\mathrm{p}}\left(p^{-1/2}\right)\\ =& \operatorname{diag}\left(\lambda_1,\ldots,\lambda_K\right) + \operatorname{diag}\left\{O_{\mathrm{p}}\left(\frac{\lambda_1^{1/2}}{p^{1/2}}\right),\ldots,O_{\mathrm{p}}\left(\frac{\lambda_K^{1/2}}{p^{1/2}}\right)\right\}\\ &- \begin{bmatrix} \leftarrow O_{\mathrm{p}}\left(\frac{\lambda_1^{1/2}}{p^{1/2}}\right) \rightarrow \\ \vdots \quad \ddots \quad \vdots \\ \leftarrow O_{\mathrm{p}}\left(\frac{\lambda_K^{1/2}}{p^{1/2}}\right) \rightarrow \end{bmatrix} + O_{\mathrm{p}}\left(p^{-1/2}\right)\end{aligned}$$

Therefore,

$$\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}V^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)^{\mathrm{T}}\tilde{L}=I_K+O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\}.$$

(iii) $\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\hat{Z}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)$

$$\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\hat{Z}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)=\begin{pmatrix}\lambda_1^{-1}p^{-1/2}\hat{z}_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\sum_{k=1}^{K}d_k u_k v_k^{\mathrm{T}}\\ \vdots\\ \lambda_K^{-1}p^{-1/2}\hat{z}_K^{\mathrm{T}}\tilde{\mathcal{E}}_2\sum_{k=1}^{K}d_k u_k v_k^{\mathrm{T}}\end{pmatrix}$$

The largest row (in magnitude) in the above matrix will obviously be the K^{th} row, so we need only focus on that row. By the expression for $\hat{z}_K$ given in (S15) and Lemmas S5 and S6,

$$\begin{aligned}\frac{d_1}{\lambda_K p^{1/2}}\hat{z}_K^{\mathrm{T}}\tilde{\mathcal{E}}_2 u_1=&\frac{d_1 d_K}{\lambda_K^2 p}u_K^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}u_1+\frac{d_1 d_K}{\lambda_K^3 p}u_K^{\mathrm{T}}\tilde{\mathcal{E}}_2 R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_1\\&+O_{\mathrm{p}}\left[\frac{\lambda_1^{1/2}}{\lambda_K^{3/2}}\frac{n^{1/2}}{p^{1/2}}\left\{\left(\frac{n}{\lambda_K p}\right)^{3/2}+\frac{n^{1/2}}{\lambda_K p}\right\}\right]\end{aligned}$$

where

$$\begin{aligned}&\frac{d_1 d_K}{\lambda_K^2 p}u_K^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}u_1=O_{\mathrm{p}}\left(\frac{n}{p\lambda_K}+\frac{n^{1/2}d_1}{\lambda_K^{3/2}p}\right)=o_{\mathrm{p}}\left(n^{-1/2}\right)\\&\frac{d_1 d_K}{\lambda_K^3 p}u_K^{\mathrm{T}}\tilde{\mathcal{E}}_2 R\tilde{\mathcal{E}}_2^{\mathrm{T}}u_1=O_{\mathrm{p}}\left(\frac{d_1 n}{\lambda_K p}\frac{n}{p\lambda_K^{3/2}}+\frac{d_1}{\lambda_K^{1/2}p^{1/2}}\frac{n}{p\lambda_K^2}\right)=o_{\mathrm{p}}\left(n^{-1/2}\right)\\&\frac{\lambda_1^{1/2}}{\lambda_K^{3/2}}\frac{n^{1/2}}{p^{1/2}}\left\{\left(\frac{n}{\lambda_K p}\right)^{3/2}+\frac{n^{1/2}}{\lambda_K p}\right\}=\frac{n}{\lambda_K^2 p}\frac{\lambda_1^{1/2}n}{p}+\frac{\lambda_1^{1/2}}{\lambda_K^{1/2}p^{1/2}}\frac{n}{\lambda_K^2 p}=o\left(n^{-1/2}\right)\end{aligned}$$

Second,

$$\frac{d_K}{\lambda_K p^{1/2}}\hat{z}_K^{\mathrm{T}}\tilde{\mathcal{E}}_2 u_K=O_{\mathrm{p}}\left(\frac{n}{\lambda_K p}\right)=o_{\mathrm{p}}\left(n^{-1/2}\right)$$

Therefore, $\left(\tilde{L}^{\mathrm{T}}\tilde{L}\right)^{-1}\hat{Z}^{\mathrm{T}}p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(\tilde{L}+p^{-1/2}\tilde{\mathcal{E}}_1\right)=o_{\mathrm{p}}\left(n^{-1/2}\right)$.

We have shown that $\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}L=I_K+o_{\mathrm{p}}\left(n^{-1/2}\right)$.

(c) Recall that $Y_1=YX^{\mathrm{T}}\left(XX\right)^{-1}$ and $Y_2=YA$ where $A^{\mathrm{T}}X=0_{(n-d)\times d}$. Since the residuals $\mathcal{E}\sim MN_{p\times n}\left(0,\Sigma,I_n\right)$, $\mathcal{E}_1=\mathcal{E}\,X^{\mathrm{T}}\left(XX\right)^{-1}$ and $\mathcal{E}_2=\mathcal{E}\,Q_X$ are independent. And since we use Y_2 to estimate $\hat{L}$, $\hat{L}$ and $\mathcal{E}_1$ are independent. (We abuse notation here. $\tilde{\mathcal{E}}_1$ and $\mathcal{E}_1$ are different. $\tilde{\mathcal{E}}_1$ is defined using the second set of data in part 1). Therefore,

$$\begin{aligned}\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}\mathcal{E}_1\sim p^{-1/2}\mathrm{diag}\left(\lambda_1^{-1/2},\ldots,\lambda_K^{-1/2}\right)&MN_{K\times d}\left\{0,\mathrm{diag}\left(\lambda_1^{-1/2},\ldots,\lambda_K^{-1/2}\right)\hat{\hat{L}}^{\mathrm{T}}\Sigma\hat{\hat{L}}\times\right.\\&\left.\times\mathrm{diag}\left(\lambda_1^{-1/2},\ldots,\lambda_K^{-1/2}\right),\left(n^{-1}XX^{\mathrm{T}}\right)^{-1}\right\}\\&=O_{\mathrm{p}}\left(\lambda_K^{-1/2}p^{-1/2}\right)=o_{\mathrm{p}}\left(n^{-1/2}\right).\end{aligned}$$

The above work shows that

$$\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}B+\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}L\Omega+\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}\mathcal{E}_1=\Omega+o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Our last task is to understand $\left\{\hat{\lambda}_k\left(\hat{\lambda}_k-\hat{\rho}\right)^{-1}\right\}\left(\lambda_k\hat{\lambda}_k^{-1}\right)$ for $k\in[K]$.

$$\frac{\hat{\lambda}_k}{\hat{\lambda}_k-\hat{\rho}}\frac{\lambda_k}{\hat{\lambda}_k}=\left(\frac{\hat{\lambda}_k-\hat{\rho}}{\lambda_k}\right)^{-1}\underbrace{=}_{\text{Lemmas S1 and S2}}\left[1+(\rho-\hat{\rho})\,\lambda_k^{-1}+O_{\mathrm{P}}\left\{\lambda_K^{-1/2}p^{-1/2}+n\left(\lambda_k p\right)^{-1}\right\}\right]^{-1}$$
$$\underbrace{=}_{\text{Lemma 1}}\left\{1+o_{\mathrm{P}}\left(n^{-1/2}\right)\right\}^{-1}=1+o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Therefore,

$$\hat{\Omega}^{\mathrm{T}}=\begin{pmatrix}\frac{\hat{\lambda}_1}{\hat{\lambda}_1-\hat{\rho}}&&\\&\ddots&\\&&\frac{\hat{\lambda}_K}{\hat{\lambda}_K-\hat{\rho}}\end{pmatrix}\begin{pmatrix}\frac{\lambda_1}{\hat{\lambda}_1}&&\\&\ddots&\\&&\frac{\lambda_K}{\hat{\lambda}_K}\end{pmatrix}\left\{\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}B+\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}L\Omega^{\mathrm{T}}\right.$$
$$\left.+\left(L^{\mathrm{T}}L\right)^{-1}\hat{L}^{\mathrm{T}}\mathcal{E}_1\right\}$$
$$=\Omega^{\mathrm{T}}+o_{\mathrm{P}}\left(n^{-1/2}\right).$$

S2.7. *Proof of the remaining theory from Sections* 3.2, 3.3 *and* 3.4

In this section, we prove Theorem 2, Proposition 2 and Corollary 1 (in that order). We need not prove Theorem 1, since Theorem 1 is a special case of Theorem 2.

Proof of Theorem 2. Define $e_{g,1}$ to be the g^{th} row of $\mathcal{E}_1$. Then

$$\hat{\beta}_g-\beta_g=\Omega\left(\ell_g-\hat{\ell}_g\right)+e_{g,1}+\left(\Omega-\hat{\Omega}\right)\hat{\ell}_g$$
$$\hat{\beta}_g^{\mathrm{OLS}}-\beta_g=\Omega\left(\ell_g-\hat{\ell}_g^{\mathrm{OLS}}\right)+e_{g,1}$$

where $\hat{\ell}_g^{\mathrm{OLS}}$ is the ordinary least squares estimate for ℓ_g, assuming C was known. By the proof of Lemma 2, $n^{1/2}\|\Omega\hat{\ell}_g^{\mathrm{OLS}}-\Omega\hat{\ell}_g\|_2=o_{\mathrm{P}}(1)$. Equation (21) and its equivalent in the statement of Theorem 2 follow because $n^{1/2}\|\Omega-\hat{\Omega}\|_2=o_{\mathrm{P}}(1)$. Equation (22) and its equivalent in the statement of Theorem 2 then follows because $\hat{\sigma}_g=\sigma_g+o_{\mathrm{P}}(1)$ and

$$n^{1/2}\sigma_g^{-1}\left(\hat{\beta}_g^{\mathrm{OLS}}-\beta_g\right)\sim N_d\left\{0,\left(n^{-1}X^{\mathrm{T}}X\right)^{-1}+\Omega\Omega^{\mathrm{T}}\right\}.$$

Proof of Proposition 2. Define

$$\hat{\Gamma}=\operatorname{diag}\left\{\left(\hat{\lambda}_1-\hat{\rho}\right)/\hat{\lambda}_1,\ldots,\left(\hat{\lambda}_K-\hat{\rho}\right)/\hat{\lambda}_K\right\}$$
$$\Gamma=\left\{\lambda_1/\left(\lambda_1+\rho\right),\ldots,\lambda_K/\left(\lambda_K+\rho\right)\right\}.$$

By Lemmas 1 and 2,

$$\hat{\Gamma}=\Gamma+o_{\mathrm{P}}\left(n^{-1/2}\right).$$

And by Lemma S3,

$$\|\hat{\Omega}^{\mathrm{shrunk}}-\Omega\Gamma\|_2=\|\hat{\Omega}\hat{\Gamma}-\Omega\Gamma\|_2\leq\|\hat{\Omega}-\Omega\|_2+o_{\mathrm{P}}\left(n^{-1/2}\right)=o_{\mathrm{P}}\left(n^{-1/2}\right).$$

Proof of Corollary 1. Define $\Omega^{\mathrm{shrunk}}=\Omega\operatorname{diag}\left\{\lambda_1\left(\lambda_1+\rho\right)^{-1},\ldots,\lambda_K\left(\lambda_K+\rho\right)^{-1}\right\}$ and let ω_k be the kth element of $\Omega\in\mathbb{R}^{1\times K}$. Using Proposition 2 and the definition of $\hat{\beta}_g^{\mathrm{shrunk}}$ from the statement of

Corollary 1,

$$\hat{\beta}_g^{\text{shrunk}} - \beta_g = \Omega^{\text{shrunk}} \left(\ell_g - \hat{\ell}_g \right) + e_{1_g} + \rho \Omega \operatorname{diag} \left\{ (\lambda_1 + \rho)^{-1}, \ldots, (\lambda_K + \rho)^{-1} \right\} \ell_g + o_{\text{p}} \left(n^{-1/2} \right). \tag{S18}$$

where e_{1_g} is the gth row of $\mathcal{E}_1$. By Lemma 2,

$$n^{1/2} \left\{ \Omega^{\text{shrunk}} \left(\ell_g - \hat{\ell}_g \right) + e_{1_g} \right\} \overset{\mathcal{D}}{=} Z + o_{\text{p}}(1),$$

where $\sigma_g^{-1} Z \sim N \left\{ 0, \left(n^{-1} \|X\|_2^2 \right)^{-1} + \|\Omega^{\text{shrunk}}\|_2^2 \right\}$. Define

$$s_g = \hat{\sigma}_g \left\{ \left(n^{-1} \|X\|_2^2 \right)^{-1} + \|\hat{\Omega}^{\text{shrunk}}\|_2^2 \right\}^{1/2}.$$

If $\lambda_K^{-1} n^{1/2} \to 0$, then clearly $n^{1/2} s_g^{-1} \left(\hat{\beta}_g^{\text{shrunk}} - \beta_g \right) \overset{\mathcal{D}}{=} W + o_{\text{p}}(1)$, where $W \sim N(0,1)$. Next, we can write

$$\begin{aligned} |z_g| = s_g^{-1} |n^{1/2} \left(\hat{\beta}_g^{\text{shrunk}} - \beta_g \right)| &= s_g^{-1} |O_{\text{p}}(1) + \rho n^{1/2} \sum_{k=1}^{K} \omega_k \ell_{gk} \left(\rho + \lambda_k \right)^{-1}| \\ &\geq s_g^{-1} \left\{ \rho n^{1/2} |\sum_{k=1}^{K} \omega_k \ell_{gk} \left(\rho + \lambda_k \right)^{-1}| - |O_{\text{p}}(1)| \right\} \end{aligned} \tag{S19}$$

where $s_g^{-1} \geq c + o_{\text{p}}(1)$ for some constant $c > 0$. If $\lambda_K^{-1} n^{1/2} \to \infty$, then by Item (ii) in the statement of Corollary 1, $\text{pr} \left(|z_g| \geq q_{1-\alpha/2} \right) \to 1$ for any $q_{1-\alpha/2} > 0$ because

$$n^{1/2} |\sum_{k=1}^{K} \omega_k \ell_{gk} \left(\rho + \lambda_k \right)^{-1}| \geq n^{1/2} \left(\rho + \lambda_K \right)^{-1} \epsilon \asymp n^{1/2} \lambda_K^{-1} \epsilon \to \infty.$$

Next, assume $\lambda_K^{-1} n^{1/2 - c_6} \to \infty$ for some small constant $c_6 > 0$. Then

$$n^{1/2} |\sum_{k=1}^{K} \omega_k \ell_{gk} \left(\rho + \lambda_k \right)^{-1}| \geq n^{1/2} \left(\rho + \lambda_K \right)^{-1} \epsilon \asymp n^{c_6} \left(n^{1/2 - c_6} \lambda_K^{-1} \epsilon \right)$$

where for Φ the cumulative distribution function for the standard normal and $|z|_g$ large enough,

$$\log \left\{ 2\Phi \left(-|z_g| \right) \right\} \leq -z_g^2/2 \leq -\tilde{c} n^{2c_6} \left(n^{1/2 - c_6} \lambda_K^{-1} \right)^2 \left\{ 1 + o_{\text{p}}(1) \right\}$$

for some constant $\tilde{c} > 0$. If $n^{-r} p \to 0$ for some $r > 0$ as $n, p \to \infty$, then $\exp \left(-\tilde{c} n^{2c_6} \right) p \to 0$ as $n, p \to \infty$. Therefore, for any $\alpha \in (0,1)$,

$$\text{pr} \left\{ |z_g| \geq q_{1-(p^{-1}\alpha)/2} \right\} = \text{pr} \left\{ 2p\Phi \left(-|z_g| \right) \leq \alpha \right\} \to 1$$

as $n, p \to \infty$.

Lastly, suppose $\lambda_K^{-1} n^{1/2} \geq c_6 > 0$. By (S19), for any $\delta > 0$, there exists an M large enough such that if $\lambda_K^{-1} n^{1/2} \geq M$, $\text{pr} \left(|z_g| \geq q_{1-\alpha/2} \right) \geq 1 - \delta$ for all n large enough. Therefore, it suffices to assume $\lambda_K^{-1} n^{1/2}$ is bounded from above by a constant. By (S18), this implies

$$n^{1/2} s_g^{-1} \left(\hat{\beta}_g^{\text{shrunk}} - \beta_g \right) \overset{\mathcal{D}}{=} W + c_{n,p} + o_{\text{p}}(1)$$

where $W \sim N(0,1)$, $c_{n,p}$ is non-random and

$$|c_{n,p}| = \sigma_g^{-1} \left\{ \left(n^{-1} \|X\|_2^2 \right)^{-1} + \|\Omega^{\text{shrunk}}\|_2^2 \right\}^{-1/2} \rho |n^{1/2} \Omega \operatorname{diag} \left\{ (\lambda_1 + \rho)^{-1}, \ldots, (\lambda_K + \rho)^{-1} \right\} | \geq c$$

for all n, p large enough, where $c > 0$ is a constant not dependent on n or p. Since

$$\text{pr}\left(|W + c_{n,p}| \geq q_{1-\alpha/2}\right) \geq \text{pr}\left(|W + c| \geq q_{1-\alpha/2}\right) > \alpha$$

for all n, p large enough, this proves the claim. □

S2.8. *CATE-RR and dSVA inflate test statistics*

We now state and prove results similar to Proposition 2 and Corollary 1, except for the estimators used in dSVA (Lee et al., 2017) and CATE-RR (Wang et al., 2017).

PROPOSITION S3 (ESTIMATE FOR Ω USED IN dSVA). *Suppose the assumptions of Proposition 2 hold but we estimate Ω as*

$$\hat{\Omega}^{\text{dSVA}} = Y_1^{\mathrm{T}} P_{1_p}^{\perp} \hat{L} \left(\hat{L}^{\mathrm{T}} P_{1_p}^{\perp} \hat{L}\right)^{-1}.$$

Then if the smallest eigenvalue of $np^{-1} L^{\mathrm{T}} P_{1_p}^{\perp} L$ is greater than $\delta \lambda_K$ where $\delta > 0$ is a constant,

$$\|\hat{\Omega}^{\text{dSVA}} - \Omega \left(L^{\mathrm{T}} P_{1_p}^{\perp} L\right) \left(L^{\mathrm{T}} P_{1_p}^{\perp} L + p n^{-1} \rho I_K\right)^{-1} \|_2 = o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Proof. Define $\hat{V} = (\hat{v}_1 \cdots \hat{v}_K) \in \mathbb{R}^{K \times K}$ and $\hat{Z} = (\hat{z}_1 \cdots \hat{z}_K) \in \mathbb{R}^{(n-d-K) \times K}$, where $\hat{v}_1, \ldots, \hat{v}_K$ and $\hat{z}_1, \ldots, \hat{z}_K$ are defined in (S14) and (S15) in the statement of Lemma S2. By Lemmas S1 and S2,

$$\hat{L} = n^{-1/2} Y_2 \left(\tilde{C} \hat{V} + Q \hat{Z}\right) = n^{-1/2} L \hat{V} + n^{-1/2} \tilde{\mathcal{E}}_1 \hat{V} + n^{-1/2} \tilde{\mathcal{E}}_2 \hat{Z}$$

where, $\tilde{C} = n^{-1/2} C_2$, $Q = Q_{C_2}$, $\tilde{\mathcal{E}}_1 = \mathcal{E} Q_X \tilde{C}$ and $\tilde{\mathcal{E}}_2 = \mathcal{E} Q_X Q$. Note that $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$ are independent by Craig's Theorem. Therefore,

$$\begin{aligned}
n (\lambda_K p)^{-1} \hat{L}^{\mathrm{T}} P_{1_p}^{\perp} \hat{L} = & n (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} L \hat{V} + n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} \\
& + \left\{n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V}\right\}^{\mathrm{T}} + n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z} \\
& + \left\{n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z}\right\}^{\mathrm{T}} + (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} \tilde{\mathcal{E}}_1^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} \\
& + (\lambda_K p)^{-1} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z} + (\lambda_K p)^{-1} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} \\
& + \left\{(\lambda_K p)^{-1} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V}\right\}^{\mathrm{T}}.
\end{aligned}$$

By the Lemmas S1 and S2,

$$n (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} L \hat{V} = m (\lambda_K p)^{-1} L^{\mathrm{T}} P_{1_p}^{\perp} L + o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Next,

$$\begin{aligned}
n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} &= (\lambda_K p)^{-1/2} \hat{V}^{\mathrm{T}} \left\{n^{1/2} (\lambda_K p)^{-1/2} L\right\}^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} = O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\} \\
&= o_{\mathrm{p}}\left(n^{-1/2}\right)
\end{aligned}$$

$$\begin{aligned}
n^{1/2} (\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} L^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z} &= (\lambda_K p)^{-1/2} \hat{V}^{\mathrm{T}} \left\{n^{1/2} (\lambda_K p)^{-1/2} L\right\}^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z} = O_{\mathrm{p}}\left\{n (\lambda_K p)^{-1}\right\} \\
&= o_{\mathrm{p}}\left(n^{-1/2}\right).
\end{aligned}$$

$$(\lambda_K p)^{-1} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_2 \hat{Z} = O_{\mathrm{p}}\left(n p^{-1} \lambda_K^{-2}\right) = o_{\mathrm{p}}\left(n^{-1/2}\right)$$

$$(\lambda_K p)^{-1} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} = \lambda_K^{-1} p^{-1/2} \hat{Z}^{\mathrm{T}} \tilde{\mathcal{E}}_2^{\mathrm{T}} P_{1_p}^{\perp} \left(p^{-1/2} \tilde{\mathcal{E}}_1\right) \hat{V} = O_{\mathrm{p}}\left(n p^{-1} \lambda_K^{-3/2}\right) = o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Lastly,

$$(\lambda_K p)^{-1} \hat{V}^{\mathrm{T}} \tilde{\mathcal{E}}_1^{\mathrm{T}} P_{1_p}^{\perp} \tilde{\mathcal{E}}_1 \hat{V} = \lambda_K^{-1} p^{-1} \operatorname{tr}\left(\Sigma P_{1_p}^{\perp}\right) I_K + o_{\mathrm{p}}\left(n^{-1/2}\right) = \lambda_K^{-1} \rho I_K + o_{\mathrm{p}}\left(n^{-1/2}\right).$$

An identical calculation shows that

$$n\,(\lambda_K p)^{-1} L^{\mathrm{T}} P_{1_p}^{\perp} \hat{L} = n\,(\lambda_K p)^{-1} L^{\mathrm{T}} P_{1_p}^{\perp} L + o_{\mathrm{p}}\left(n^{-1/2}\right).$$

Lastly, for $\mathcal{E}_1 = \mathcal{E}\, X\, (X^{\mathrm{T}} X)^{-1}$,

$$(\lambda_K p)^{-1/2} n^{1/2} \mathcal{E}_1^{\mathrm{T}}\, P_{1_p}^{\perp} \hat{L} \left(\hat{L}^{\mathrm{T}} P_{1_p}^{\perp} \hat{L}\right)^{-1} = O_{\mathrm{p}}\left\{p^{-1/2} \lambda_K^{-1}\right\} = o_{\mathrm{p}}\left(n^{-1/2}\right).$$

This completes the proof. □

PROPOSITION S4 (ESTIMATE FOR Ω USED IN CATE-RR). *Suppose the assumptions of Proposition 2 hold with $d = 1$ but we estimate Ω as*

$$\hat{\Omega}^{\mathrm{cate}} = \underset{\alpha \in \mathbb{R}^{1\times K}}{\arg\min}\, \Psi\left(y_{1_g} - \alpha \hat{\ell}_g\right)$$

where for some constant $c > 0$,

$$\Psi(x) = \begin{cases} x^2/2 & \textit{if } |x| \le c \\ c|x| - c^2/2 & \textit{if } |x| > c \end{cases}. \tag{S20}$$

Note that Ψ is Huber's loss. Suppose further that $pn^{-r} \to 0$ for some $r > 0$. Then if $\lambda_K \to \infty$, the results of Proposition 2 hold. If $\lambda_1 = O(1)$, then there exists a constant $\epsilon > 0$ such that

$$\lim_{n,p\to\infty} \operatorname{pr}\left(\|\hat{\Omega}^{\mathrm{cate}} - \Omega\|_2 \ge \epsilon \|\Omega\|_2\right) = 1.$$

Proof. Let $d/dx\Psi(x) = \dot{\Psi}(x)$. Since Ψ is a convex function, $\hat{\Omega}^{\mathrm{cate}}$ solves

$$0 = \sum_{g=1}^{p} \dot{\Psi}\left(y_{1_g} - \hat{\ell}_g \hat{\Omega}^{\mathrm{cate}}\right) \hat{\ell}_g^{\mathrm{T}} = \left(Y_1 - \hat{L} \hat{\Omega}^{\mathrm{cate}^{\mathrm{T}}}\right)^{\mathrm{T}} \hat{A} \hat{L}$$

where $\hat{A} \in \mathbb{R}^{p\times p}$ is a diagonal matrix with

$$\hat{A}_{gg} = \begin{cases} 1 & \text{if } |y_{1_g} - \hat{\Omega}^{\mathrm{cate}} \hat{\ell}_g| \le c \\ |y_{1_g} - \hat{\Omega}^{\mathrm{cate}} \hat{\ell}_g|^{-1} & \text{if } |y_{1_g} - \hat{\Omega}^{\mathrm{cate}} \hat{\ell}_g| > c \end{cases} \quad (g = 1, \ldots, p).$$

We start by assuming $\lambda_K \to \infty$. When this is true, it suffices to show that

$$\max_{g\in[p]} |y_{1_g} - \hat{\Omega}^{\mathrm{shrunk}} \hat{\ell}_g| = o_{\mathrm{p}}(1).$$

We see that

$$y_{1_g} - \hat{\Omega}^{\mathrm{shrunk}} \hat{\ell}_g = \hat{\Omega}^{\mathrm{shrunk}} \left(\ell_g - \hat{\ell}_g\right) + \left(\Omega - \hat{\Omega}^{\mathrm{shrunk}}\right) \ell_g + e_{1_g}$$

where

$$\max_{g\in[p]} |e_{1_g}| = O_{\mathrm{p}}\left\{n^{-1/2} \log(p)\right\} = o_{\mathrm{p}}(1)$$

and

$$\max_{g\in[p]} |\left(\Omega - \hat{\Omega}^{\mathrm{shrunk}}\right) \ell_g| = o_{\mathrm{p}}(1)$$

because the entries of L are uniformly bounded and $\|\Omega - \hat{\Omega}^{\text{shrunk}}\|_2 = o_{\text{p}}(1)$. To complete the proof, we need only show that

$$\|\hat{L} - L\|_\infty = o_{\text{p}}(1).$$

By the proof of Proposition S3,

$$\hat{L} - L = L\left(I_K - \hat{V}\right) + n^{-1/2}\tilde{\mathcal{E}}_1\hat{V} + n^{-1/2}\tilde{\mathcal{E}}_2\hat{Z}.$$

Since $\|I_K - \hat{V}\|_2 = o_{\text{p}}(1)$, $\|L\left(I_K - \hat{V}\right)\|_\infty = o_{\text{p}}(1)$. Next, $\|n^{-1/2}\tilde{\mathcal{E}}_1\|_\infty = O_{\text{p}}\left\{n^{-1/2}\log(p)\right\} = o_{\text{p}}(1)$. For the last term, define the random variable $Z_g = \sigma_g^{-2}\tilde{\mathcal{E}}_{2_g}^{\text{T}}\tilde{\mathcal{E}}_{2_g}$. Since this is a sub-exponential random variable with parameters $\{4(n-d-K), 4\}$,

$$\text{pr}\left\{Z_g \geq (n-d-K) + t\sqrt{n}\right\} \leq \exp\left(-bt^2\right)$$

for some constant $b > 0$, provided $t = o\left(n^{1/2}\right)$. If we let $t = b'\{\log(p)\}^{1/2}$ for some constant $b' > b^{-1}$, then

$$\text{pr}\left\{\max_{g\in[p]} Z_g \geq (n-d-K) + t\sqrt{n}\right\} \leq 1 - \left\{1 - \exp\left(-bt^2\right)\right\}^p$$

where

$$p\log\left\{1 - \exp\left(-bt^2\right)\right\} = -\exp\left\{\log(p)\left(1 - bb'\right)\right\}\left\{1 + o(1)\right\} = o(1).$$

Therefore, for each $k \in [K]$,

$$\max_{g\in[p]}|n^{-1/2}\tilde{\mathcal{E}}_{2_g}^{\text{T}}\hat{z}_k| \leq O_{\text{p}}\left\{n^{1/2}\left(\lambda_K p\right)^{-1/2}\right\}\max_{g\in[p]}\left(n^{-1}\tilde{\mathcal{E}}_{2_g}^{\text{T}}\tilde{\mathcal{E}}_{2_g}\right)^{1/2} = o_{\text{p}}(1)$$

which completes the proof when $\lambda_K \to \infty$.

When $\lambda_1 = O(1)$, the results $\|L - \hat{L}\|_\infty, \|\mathcal{E}_1\|_\infty = o_{\text{p}}(1)$ still hold. We therefore need only understand how $\left(\Omega - \hat{\Omega}^{\text{shrunk}}\right)\ell_g$ behaves. By assumption, there exists a constant $m > 0$ that does not depend on p such that

$$\max_{g\in[p]}\|\ell_g\|_2 \leq m.$$

Define $\delta_1 = c/(2m)$, where c was defined in (S20). If $\|\Omega\|_2 \leq \delta_1$, then because

$$\hat{\Omega}^{\text{shrunk}} = \Omega\,\text{diag}\left\{\lambda_1/\left(\lambda_1+\rho\right), \cdots, \lambda_K/\left(\lambda_K+\rho\right)\right\} + o_{\text{p}}\left(n^{-1/2}\right),$$

we get that

$$\Omega - \hat{\Omega}^{\text{cate}} = \Omega\,\text{diag}\left\{\rho/\left(\lambda_1+\rho\right), \cdots, \rho/\left(\lambda_K+\rho\right)\right\} + o_{\text{p}}\left(n^{-1/2}\right).$$

If $\|\Omega\|_2 > \delta_1$, suppose we initialize the optimization problem with $\alpha_1 \in \mathbb{R}^{1\times K}$ such that $\|\Omega - \alpha_1\|_2 \leq \delta_1$. Then the next iteration will be

$$\alpha_2 = Y_1^{\text{T}}\hat{L}\left(\hat{L}^{\text{T}}\hat{L}\right)^{-1} = \Omega\,\text{diag}\left\{\lambda_1/\left(\lambda_1+\rho\right), \cdots, \lambda_K/\left(\lambda_K+\rho\right)\right\} + o_{\text{p}}\left(n^{-1/2}\right)$$

with probability tending to 1, where

$$\Omega - \alpha_2 = \Omega\,\text{diag}\left\{\rho/\left(\lambda_1+\rho\right), \cdots, \rho/\left(\lambda_K+\rho\right)\right\} + o_{\text{p}}\left(n^{-1/2}\right).$$

Therefore,

$$\|\Omega - \alpha_2\|_2 \geq \|\Omega\|_2\rho/\left(\lambda_1+\rho\right)\left\{1 + o_{\text{p}}\left(n^{-1/2}\right)\right\} \geq \delta_2 + o_{\text{p}}\left(n^{-1/2}\right)$$

for some constant $\delta_2 > 0$ not dependent n or p, since $\lambda_1 = O(1)$ by assumption. Note that we may assume $\delta_1 > \delta_2$. Therefore,

$$\|\hat{\Omega}^{\mathrm{cate}} - \Omega\|_2 \geq \delta_2 + o_{\mathrm{p}}(1)$$

which completes the proof. □

Remark 3. The above proof shows that the behavior of Huber's loss function is very dependent on the constant c used in (S20) when $\lambda_1 = O(1)$, meaning we cannot predict its behavior. This is an additional reason why this loss function should not be used to estimate Ω when the data are only moderately informative for C.

COROLLARY S1 (THE RESULTS OF COROLLARY 1 HOLD USING DSVA AND CATE-RR). *Suppose the assumptions of Proposition* 2 *hold with* $d = 1$ *and for some fixed* $g \in [p]$*, define*

$$\hat{\beta}_g^{\mathrm{dSVA}} = y_{1_g} - \hat{\Omega}^{\mathrm{dSVA}}\hat{\ell}_g$$
$$\hat{\beta}_g^{\mathrm{cate}} = y_{1_g} - \hat{\Omega}^{\mathrm{cate}}\hat{\ell}_g.$$

In addition, suppose $K = 1$ *and*

(i) $n^{-r}p \to 0$ *for some* $r > 0$ *as* $n \to \infty$.
(ii) $np^{-1}L^{\mathrm{T}}P_{1_p}^{\perp}L \geq \delta\lambda_1$ *for some constant* $\delta > 0$
(iii) $|\Omega\ell_g| \geq \epsilon$ *for some constant* $\epsilon > 0$.

Then the results of Corollary 1 *hold for the z-score*

$$z_g^{\mathrm{dSVA}} = \sigma_g^{-1}\left(\|X\|_2^{-2} + n^{-1}\|\hat{\Omega}^{\mathrm{dSVA}}\|_2^2\right)^{-1/2}\hat{\beta}_g^{\mathrm{dSVA}}.$$

If $\lambda_1 \to \infty$*, then the results of Corollary* 1 *hold for the z-score*

$$z_g^{\mathrm{cate}} = \sigma_g^{-1}\left(\|X\|_2^{-2} + n^{-1}\|\hat{\Omega}^{\mathrm{cate}}\|_2^2\right)^{-1/2}\hat{\beta}_g^{\mathrm{cate}}.$$

Proof. The proof is identical to the proof of Corollary 1 and is omitted. □

Remark 4. We require $\lambda_1 \to \infty$ to prove Proposition S1 for z-scores returned by CATE-RR because the behavior of $\hat{\Omega}^{\mathrm{cate}}$ depends heavily on the constant c chosen in (S20) when $\lambda_1 = O(1)$.

S2.9. *A framework for when C is treated as a random variable and the proof of Theorem* 3

Next, we provide a framework to extend all of our theoretical results to the case when C is treated as a random variable. We then prove Theorem 3 at the end of this section. First, we prove a proposition regarding the identifiability of factor models when C is random.

PROPOSITION S5. *Suppose* $Y = BX^{\mathrm{T}} + \bar{L}\bar{C}^{\mathrm{T}} + \mathcal{E}$ *where* $B \in \mathbb{R}^{p\times d}$ *and* $\bar{L} \in \mathbb{R}^{p\times K}$ *are fixed effects,* $X \in \mathbb{R}^{n\times d}$ *is observed and*

(i) X *has full column rank.*
(ii) $\bar{C} \in \mathbb{R}^{n\times K}$ *is such that* $E\left(\bar{C}\right) = X\bar{A}$ *for some non-random* $\bar{A} \in \mathbb{R}^{d\times K}$ *and* $\mathrm{var}\left\{\mathrm{vec}\left(\bar{C}\right)\right\} = \bar{\Psi} \otimes I_n$ *where* $\bar{\Psi} \succ 0$.
(iii) $\mathcal{E} \in \mathbb{R}^{p\times n}$ *is independent of* $\bar{C}$ *and* $\mathrm{var}\left\{\mathrm{vec}\left(\mathcal{E}\right)\right\} = I_n \otimes \Sigma$*, where* $\Sigma = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_p^2\right) \succ 0$.
(iv) *If any row is removed from* $\bar{L}$*, there exists two sub-matrices with rank* K.

Then $\bar{L}\bar{\Psi}\bar{L}^{\mathrm{T}}$ *and* Σ *are identifiable.*

Proof. Define $\bar{C}_2 = Q_X^{\mathrm{T}}\bar{C}$ and $Y_2 = YQ_X$, where the columns of $Q_X \in \mathbb{R}^{n\times(n-d)}$ form an orthonormal basis for the null space of X^{T}. Then $E\left(\bar{C}_2\right) = 0$, $\mathrm{var}\left\{\mathrm{vec}\left(\bar{C}\right)\right\} = \bar{\Psi} \otimes I_{n-d}$ and $\mathrm{var}\left\{\mathrm{vec}\left(Y_2\right)\right\} = I_{n-d} \otimes \left(\Sigma + \bar{L}\bar{\Psi}\bar{L}^{\mathrm{T}}\right)$. The identifiability of $\bar{L}\bar{\Psi}\bar{L}^{\mathrm{T}}$ and Σ then follows from Theorem 5.1 of (Anderson & Rubin, 1956). □

COROLLARY S2. *Let $c > 1$ be a constant. Suppose that in addition to the assumptions in Proposition S5, the following hold*

(i) *p is a non-decreasing function of n.*
(ii) *$\bar{L}_{g*}^{\mathrm{T}} \bar{\Psi} \bar{L}_{g*} \leq c$ for all $g \in [p]$.*
(iii) *There are K non-zero eigenvalues of $\bar{L}\bar{\Psi}\bar{L}^{\mathrm{T}}$ $\gamma_1, \ldots, \gamma_K$ such that $c^{-1} \leq \gamma_1 \leq \cdots \leq \gamma_K \leq cn$.*
(iv) *For all $r \in [d]$, $p^{-1} \sum_{g=1}^{p} I\left(B_{gr} \neq 0\right) = o\left(n^{-1}\gamma_K\right)$.*

Then B is identifiable for all $n \geq c'$, where $c' > 0$ is a constant.

Proof. Define $Y_1 = YX\left(X^{\mathrm{T}}X\right)^{-1}$, where

$$E\left(Y_1\right) = B + \bar{L}\bar{A}^{\mathrm{T}} = B + \left(\bar{L}\bar{\Psi}^{1/2}\right)\left(\bar{A}\bar{\Psi}^{-1/2}\right)^{\mathrm{T}}.$$

The identical method used to prove B was identifiable in Proposition 1 can then be used to show B is identifiable here. □

Remark 5. If Items (ii) and (iii) from the statement of Corollary S2 hold for some $\bar{L} \in \mathbb{R}^{p \times K}$ and $\bar{\Psi} \succ 0$, then Item (iv) from the statement of Proposition S5 holds for all n, p suitably large. Therefore, another way to express Item (ii) from the statement of Theorem 3 is to first assume Item (iv) from the statement of Proposition S5 holds to identify Σ and LL^{T} (and therefore $\lambda_1, \ldots, \lambda_K$), and then assume $L^{\mathrm{T}}L$ is orthogonal with decreasing elements, since this will not affect the isotropic distribution assumption on C (or any uniform bound on the fourth moments of its entries).

Remark 6. We do not need Corollary S2 to prove Theorem 3. We state it to show that our theoretical results from Sections 3.2, 3.3 and 3.4 can be extended to the case when C is treated as a random variable.

Next, we state and prove a technical lemma to be used in the proof of Theorem 3. This lemma is also important because it shows that we can generalize Assumption 2 to the case when C is a random variable.

LEMMA S4. *Let $a > 1$ be a constant not dependent on n or p, suppose $Y = \bar{L}\bar{C}^{\mathrm{T}} + \mathcal{E}$ where $\bar{L} \in \mathbb{R}^{p \times K}$, $\bar{C} \in \mathbb{R}^{n \times K}$ and $\mathcal{E} \in \mathbb{R}^{p \times n}$ and assume Items* (ii), (iii) *and* (iv) *from the statement of Proposition S5 hold. Define $\gamma_1, \ldots, \gamma_K$ to be the eigenvalues of $np^{-1}\bar{\Psi}^{1/2}\bar{L}^{\mathrm{T}}\bar{L}\bar{\Psi}^{1/2}$ with eigenvectors $u_1, \ldots, u_K \in \mathbb{R}^K$ and assume the following hold*

(i) *$\mathcal{E} \sim MN_{p \times n}\left(0, \Sigma, I_n\right)$ where $\sigma_g^2 \in \left[a^{-1}, a\right]$ for all $g \in [p]$.*
(ii) *$\|n^{-1}\bar{C}^{\mathrm{T}}\bar{C} - \bar{\Psi}\|_2 = O_{\mathrm{p}}\left(n^{-1/2}\right)$.*
(iii) *The magnitude of the entries of $\bar{L}$ are uniformly bounded by a.*
(iv) *$a^{-1} \leq \gamma_K < \cdots < \gamma_1 \leq an$ and $\left(\gamma_k - \gamma_{k+1}\right)\gamma_k^{-1} \geq a^{-1}$ $(k = 1, \ldots, K)$ where γ_{K+1} is defined to be 0.*
(v) *$|u_r^{\mathrm{T}}\left(np^{-1}\bar{\Psi}^{1/2}\bar{L}^{\mathrm{T}}\Sigma\bar{L}\bar{\Psi}^{1/2}\right)u_s| \leq a\gamma_{\max(r,s)}$ (r=1,...,K; s=1,...,K).*

Then there exists an $L \in \mathbb{R}^{p \times K}$, $C \in \mathbb{R}^{n \times K}$ and constant $c > 1$ such that the following hold:

1. *$\bar{L}\bar{C}^{\mathrm{T}} = LC^{\mathrm{T}}$ such that $P_{\bar{C}} = P_C$, $n^{-1}C_2^{\mathrm{T}}C_2 = I_K$ and $\sup_{g \in [p], k \in [K]} |L_{gk}| \leq c + o_{\mathrm{p}}(1)$.*
2. *$L^{\mathrm{T}}L$ is a diagonal matrix with decreasing entries $\lambda_1, \ldots, \lambda_K$ such that λ_k is the kth largest eigenvalue of $\bar{C}\left(p^{-1}\bar{L}^{\mathrm{T}}\bar{L}\right)\bar{C}$ $(k = 1, \ldots, K)$.*
3. *$1 - \lambda_k\gamma_k^{-1} = O_{\mathrm{p}}\left(n^{-1/2}\right)$ and $\left(\lambda_k - \lambda_{k+1}\right)\lambda_k^{-1} \geq c^{-1} + O_{\mathrm{p}}\left(n^{-1/2}\right)$ $(k = 1, \ldots, K)$ where λ_{K+1} is defined to be 0.*
4. *$n\left\{p\lambda_{\max(r,s)}\right\}^{-1} L_{*r}^{\mathrm{T}}\Sigma L_{*s} = O_{\mathrm{p}}(1)$ (r=1,...,K; s=1,...,K).*

Proof. We first re-define $\bar{L}$ as $\bar{L}\bar{\Psi}^{1/2}$ and $\bar{C}$ as $\bar{C}\bar{\Psi}^{-1/2}$, meaning we now have $\|n^{-1}\bar{C}^{\mathrm{T}}\bar{C} - I_K\|_2 = O_{\mathrm{p}}\left(n^{-1/2}\right)$. Define $\hat{R}$ such that $\hat{R}^2 = n^{-1}\bar{C}^{\mathrm{T}}\bar{C}$ and

$$L = \bar{L}\hat{R}\hat{U}$$
$$C = \bar{C}\hat{R}^{-1}\hat{U}$$

where the columns of $\hat{U} \in \mathbb{R}^{K\times K}$ contain the right singular vectors of $\bar{L}\hat{R}$. Since $n^{-1}C^{\mathrm{T}}P_X^{\perp}C = I_K$, this proves 1 and 2.

To then prove 3 and 4, we study the eigenvalues and eigenvectors of $np^{-1}\hat{R}^{\mathrm{T}}\bar{L}^{\mathrm{T}}\bar{L}^{\mathrm{T}}\hat{R}$. We can write $np^{-1}L^{\mathrm{T}}L$ (whose diagonal elements are the eigenvalues of $np^{-1}\hat{R}^{\mathrm{T}}\bar{L}^{\mathrm{T}}\bar{L}^{\mathrm{T}}\hat{R}$) as

$$np^{-1}L^{\mathrm{T}}L = \hat{U}^{\mathrm{T}}UU^{\mathrm{T}}\hat{R}U\mathrm{diag}\left(\gamma_1,\ldots,\gamma_K\right)U^{\mathrm{T}}\hat{R}UU^{\mathrm{T}}\hat{U} = \hat{U}^{\mathrm{T}}\hat{F}\mathrm{diag}\left(\gamma,\ldots,\gamma_K\right)\hat{F}\hat{U}$$

where $U = (u_1 \cdots u_K) \in \mathbb{R}^{K\times K}$ and $\hat{F} = U^{\mathrm{T}}\hat{R}U$ where the diagonal entries of $\hat{F}$ are $1 + O_{\mathrm{p}}\left(n^{-1/2}\right)$ and the off-diagonal entries are $O_{\mathrm{p}}\left(n^{-1/2}\right)$. We have also re-defined $\hat{U}$ as $U^{\mathrm{T}}\hat{U}$, which is still a random unitary matrix. Define the matrix $A = \hat{F}\mathrm{diag}\left(\gamma_1,\ldots,\gamma_K\right)\hat{F} \in \mathbb{R}^{K\times K}$ where

$$A_{kk} = \gamma_k\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\} + \sum_{r\neq k}\gamma_r O_{\mathrm{p}}\left(n^{-1}\right) \quad (k = 1,\ldots,K)$$
$$A_{rs} = \left(\gamma_r + \gamma_s\right)O_{\mathrm{p}}\left(n^{-1/2}\right) + \sum_{k\neq r,s}\gamma_k O_{\mathrm{p}}\left(n^{-1}\right) \quad (r = 1,\ldots,K; s = 1,\ldots,K; r\neq s).$$

Next, define $A^{(1)} = \gamma_1^{-1}A$ where

$$A^{(1)}_{kk} = \frac{\gamma_k}{\gamma_1}\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\} + \sum_{r\neq k}\frac{\gamma_r}{\gamma_1}O_{\mathrm{p}}\left(n^{-1}\right) \quad (k = 1,\ldots,K)$$
$$A^{(1)}_{rs} = \frac{\gamma_r + \gamma_s}{\gamma_1}O_{\mathrm{p}}\left(n^{-1/2}\right) + \sum_{k\neq r,s}\frac{\gamma_k}{\gamma_1}O_{\mathrm{p}}\left(n^{-1}\right) \quad (r = 2,\ldots,K; s = 2,\ldots,K; r\neq s).$$

We first decompose $A^{(1)}$ into K rank (approximately) 1 matrices to study the behavior of the eigenvalues and eigenvectors of A. We see that

$$A^{(1)} = \underbrace{\begin{bmatrix} A^{(1)}_{11} & A^{(1)}_{12} + \omega_2 & \cdots & A^{(1)}_{1K} + \omega_K \\ A^{(1)}_{12} + \omega_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A^{(1)}_{1K} + \omega_K & 0 & \cdots & 0 \end{bmatrix}}_{=D_1} + \underbrace{\begin{bmatrix} 0 & -\omega_2 & 0 & \cdots & 0 \\ -\omega_2 & A^{(1)}_{22} & A^{(1)}_{23} & \cdots & A^{(1)}_{2K} \\ 0 & A^{(1)}_{23} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & A^{(1)}_{2K} & 0 & \cdots & 0 \end{bmatrix}}_{=D_2} + \cdots$$
$$+ \underbrace{\begin{bmatrix} 0 & 0 \cdots & -\omega_K \\ 0 & 0 \cdots & 0 \\ \vdots & \vdots \ddots & \vdots \\ -\omega_K & 0 \cdots & A^{(1)}_{KK} \end{bmatrix}}_{=D_K}$$

where we define

$$\omega_k = \frac{A^{(1)}_{kk}A^{(1)}_{1k}}{A^{(1)}_{11} - A^{(1)}_{kk}} = O_{\mathrm{p}}\left(\frac{\gamma_k}{\gamma_1}n^{-1/2}\right) \quad (k = 2,\ldots,K).$$

Let

$$\begin{aligned} v_1 &= \left[\left\{A_{11}^{(1)}\right\}^2 + \left\{A_{12}^{(1)}+\omega_2\right\}^2 + \cdots + \left\{A_{1K}^{(1)}+\omega_K\right\}^2\right]^{-1/2} \left\{A_{11}^{(1)} \; A_{12}^{(1)}+\omega_2 \; \cdots \; A_{1K}^{(1)}+\omega_K\right\}^{\mathrm{T}} \\ &= \left\{1 \; \frac{A_{12}^{(1)}+\omega_2}{A_{11}^{(1)}} \; \cdots \; \frac{A_{1K}^{(1)}+\omega_K}{A_{11}^{(1)}}\right\}^{\mathrm{T}} + O_{\mathrm{p}}\left(n^{-1}\right). \end{aligned}$$

Then

$$\begin{aligned} A^{(1)}v_1 = D_1 v_1 + D_2 v_1 + \cdots + D_K v_1 = &\left\{\begin{array}{c} A_{11}^{(1)} + O_{\mathrm{p}}\left(n^{-1}\right) \\ \frac{A_{12}^{(1)}+\omega_2}{A_{11}^{(1)}} + O_{\mathrm{p}}\left(n^{-1}\right) \\ \cdots \\ \frac{A_{1K}^{(1)}+\omega_K}{A_{11}^{(1)}} + O_{\mathrm{p}}\left(n^{-1}\right) \end{array}\right\} \\ &+ \left\{\begin{array}{c} O_{\mathrm{p}}\left(n^{-1}\right) \\ \underbrace{-\omega_2 + A_{22}^{(1)}\frac{A_{12}^{(1)}+\omega_2}{A_{11}^{(1)}}}_{=0} + O_{\mathrm{p}}\left(n^{-1}\right) \\ \cdots \\ O_{\mathrm{p}}\left(n^{-1}\right) \end{array}\right\} \\ &+ \cdots + \left\{\begin{array}{c} O_{\mathrm{p}}\left(n^{-1}\right) \\ \cdots \\ O_{\mathrm{p}}\left(n^{-1}\right) \\ \underbrace{-\omega_K + A_{KK}^{(1)}\frac{A_{1K}^{(1)}+\omega_K}{A_{11}^{(1)}}}_{=0} + O_{\mathrm{p}}\left(n^{-1}\right) \end{array}\right\} \end{aligned}$$

and

$$\begin{aligned} \delta_1 &= v_1^{\mathrm{T}} A^{(1)} v_1 = A_{11}^{(1)} + O_{\mathrm{p}}\left(n^{-1}\right) \\ \|A^{(1)}v_1 - \delta_1 v_1\|_2 &= O_{\mathrm{p}}\left(n^{-1}\right). \end{aligned}$$

By Weyl's Theorem and Theorem 3.6 in Auffinger & Tang (2015) the largest eigenvalue of $A^{(1)}$ is $\hat{\mu}_1 = A_{11}^{(1)} + O_{\mathrm{p}}\left(n^{-1}\right)$ with corresponding eigenvector $\hat{u}_1$ such that $\|\hat{u}_1 - v_1\|_2 = O_{\mathrm{p}}\left(n^{-1}\right)$. To find the next eigenvalue and eigenvector of A, we first have to remove the principal direction from $A^{(1)}$:

$$A^{(1)} - \hat{\mu}_1 \hat{u}_1 \hat{u}_1^{\mathrm{T}} = D_2 + \cdots + D_K + O_{\mathrm{p}}\left(n^{-1}\right)$$

and we define

$$\begin{aligned} A^{(2)} &= \frac{\gamma_1}{\gamma_2}\left\{A^{(1)} - \hat{\mu}_1 \hat{u}_1 \hat{u}_1^{\mathrm{T}}\right\} = \frac{\gamma_1}{\gamma_2} D_2 + \cdots + \frac{\gamma_1}{\gamma_2} D_K + O_{\mathrm{p}}\left(\frac{\gamma_1}{\gamma_2 n}\right) \\ &= \begin{bmatrix} 0 & -\frac{\gamma_1}{\gamma_2}\omega_2 & 0 & \cdots & 0 \\ -\frac{\gamma_1}{\gamma_2}\omega_2 & A_{22}^{(2)} & A_{23}^{(2)} & \cdot & A_{2K}^{(2)} \\ 0 & A_{23}^{(2)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & A_{2K}^{(2)} & 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & 0 \cdots & -\frac{\gamma_1}{\gamma_2}\omega_K \\ 0 & 0 \cdots & 0 \\ \vdots & \vdots \; \ddots & \vdots \\ -\frac{\gamma_1}{\gamma_2}\omega_K & 0 \cdots & A_{KK}^{(2)} \end{bmatrix} + O_{\mathrm{p}}\left(\frac{\gamma_1}{\gamma_2 n}\right) \end{aligned}$$

where

$$\frac{\gamma_1}{\gamma_2}\omega_k = O_{\mathrm{p}}\left(\frac{\gamma_k}{\gamma_2}n^{-1/2}\right) \quad (k = 2, \ldots, K)$$

$$A_{kk}^{(2)} = \frac{\gamma_k}{\gamma_2}\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\} + \sum_{r \neq k}\frac{\gamma_r}{\gamma_2}O_{\mathrm{p}}\left(n^{-1}\right) \quad (k = 2, \ldots, K)$$

$$A_{rs}^{(2)} = \frac{\gamma_r + \gamma_s}{\gamma_2}O_{\mathrm{p}}\left(n^{-1/2}\right) + \sum_{k \neq r,s}\frac{\gamma_k}{\gamma_2}O_{\mathrm{p}}\left(n^{-1}\right) \quad (r = 2, \ldots, K; s = 2, \ldots, K; r \neq s).$$

A subsequent application of the above procedure will show that the largest eigenvalue of $A^{(2)}$ is

$$\hat{\mu}_2 = A_{22}^{(2)} + O_{\mathrm{p}}\left(\frac{\gamma_1}{\gamma_2 n}\right)$$

with eigenvalue $\hat{u}_2$ such that

$$\|\hat{u}_2 - v_2\|_2 = O_{\mathrm{p}}\left(\frac{\gamma_1}{\gamma_2 n}\right),$$

where v_2 is the second column of $\gamma_1\gamma_2^{-1}D_2$. When we subsequently remove the second principal direction, we will remove $\gamma_1\gamma_2^{-1}D_2$ and the $O_{\mathrm{p}}\left\{\gamma_1\left(\gamma_2 n\right)^{-1}\right\}$ error term will become $O_{\mathrm{p}}\left\{\gamma_1\left(\gamma_3 n\right)^{-1}\right\}$. Provided $\gamma_1\left(\gamma_k n\right)^{-1} \lesssim n^{-1/2}$, this procedure will give us estimates $\hat{\mu}_k$ and $\hat{u}_k$ such that

$$\lambda_k = \gamma_k\hat{\mu}_k = \gamma_k\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\} \tag{S21}$$

$$\|\hat{u}_k - e_k\|_2 = O_{\mathrm{p}}\left(n^{-1/2}\right) \tag{S22}$$

where here $e_k \in \mathbb{R}^K$ is the standard basis vector with 1 in the k^{th} position and 0's everywhere else.

We next handle the case when $\gamma_1\left(\gamma_k n\right)^{-1} \gtrsim n^{-1/2}$. Let $r \leq K$ be such that $\gamma_1\left(\gamma_k n\right)^{-1} \lesssim n^{-1/2}$ for $k \leq r$ and $\gamma_1\left(\gamma_k n\right)^{-1} \gtrsim n^{-1/2}$ for $k > r$. For these eigenvalues, we note that we can study the smallest eigenvalues and their eigenvectors of A by studying the largest eigenvalues of A^{-1}. If λ_k is an eigenvalue of A with eigenvector $\hat{u}_k$, then λ_k^{-1} is an eigenvalue of A^{-1} with the same eigenvector. We note that

$$A^{-1} = \hat{F}^{-1}\text{diag}\left(\gamma_1^{-1}, \ldots, \gamma_K^{-1}\right)\hat{F}^{-1} = \gamma_1^{-1}\hat{F}^{-1}\text{diag}\left(1, \gamma_1\gamma_2^{-1}, \ldots, \gamma_1\gamma_K^{-1}\right)\hat{F}^{-1}$$

where the diagonal entries of $\hat{F}^{-1}$ are $1 + O_{\mathrm{p}}\left(n^{-1/2}\right)$ and the off-diagonal entries are $O_{\mathrm{p}}\left(n^{-1/2}\right)$. If $k > r$, then $\gamma_1\gamma_k^{-1} \gtrsim n^{1/2}$, meaning $\gamma_k \lesssim n^{1/2}$, since $\gamma_1 \lesssim n$. Therefore,

$$\frac{\gamma_1\gamma_K^{-1}}{\gamma_1\gamma_k^{-1}} = \frac{\gamma_k}{\gamma_K} \lesssim n^{1/2}$$

for all $k > r$. By what we have shown above, the $K - k + 1$ eigenvalue of $\gamma_1 A^{-1}$ is $\gamma_1\gamma_k^{-1}\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\}$ with eigenvectors that satisfies (S22). Therefore, the k^{th} eigenvalue of A is $\gamma_k\left\{1 + O_{\mathrm{p}}\left(n^{-1/2}\right)\right\}$ with eigenvector that satisfies (S22). This proves item 3.

To prove item 4,

$$np^{-1}L^{\mathrm{T}}\Sigma L = \hat{M}^{\mathrm{T}}\left\{np^{-1}U^{\mathrm{T}}\bar{L}^{\mathrm{T}}\Sigma\bar{L}U\right\}\hat{M} \tag{S23}$$

where $\hat{M} = \hat{F}\hat{U}$ is such that $\|\hat{M} - I_K\|_2 = O_{\mathrm{p}}\left(n^{-1/2}\right)$ by the analysis above. To evaluate (S23), we first see that

$$np^{-1}U^{\mathrm{T}}\bar{L}^{\mathrm{T}}\Sigma\bar{L}U = \begin{bmatrix} O(\gamma_1) & O(\gamma_2) & \cdots & O(\gamma_K) \\ O(\gamma_2) & O(\gamma_2) & \cdots & O(\gamma_K) \\ \vdots & \vdots & \ddots & \vdots \\ O(\gamma_K) & O(\gamma_K) & \cdots & O(\gamma_K) \end{bmatrix} = \begin{bmatrix} O(\gamma_1) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & O(\gamma_2) & \cdots & 0 \\ O(\gamma_2) & O(\gamma_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & O(\gamma_K) \\ 0 & 0 & \cdots & O(\gamma_K) \\ \vdots & \vdots & \ddots & \vdots \\ O(\gamma_K) & O(\gamma_K) & \cdots & O(\gamma_K) \end{bmatrix}.$$

Fix some $r, s \leq K$ such that $r < s$. If $\hat{m}_k$ is the k^{th} column of $\hat{M}$, then

$$\begin{aligned} \hat{m}_r^{\mathrm{T}}\left(np^{-1}U^{\mathrm{T}}\bar{L}^{\mathrm{T}}\Sigma\bar{L}U\right)\hat{m}_s =& O_{\mathrm{p}}\left(\gamma_1 n^{-1}\right) + \cdots + O_{\mathrm{p}}\left(\gamma_{r-1}n^{-1}\right) \\ &+ \left\{O(\gamma_r)\hat{m}_{r_r}\hat{m}_{s_1} + O(\gamma_r)\hat{m}_{r_r}\hat{m}_{s_2} + \cdots + O(\gamma_r)\hat{m}_{r_r}\hat{m}_{s_r}\right\} \\ &+ O(\gamma_{r+1})\hat{m}_{r_r}\hat{m}_{s_{r+1}} + \cdots + O(\gamma_{s-1})\hat{m}_{r_r}\hat{m}_{s_{s-1}} + O_{\mathrm{p}}(\gamma_s). \end{aligned}$$

Next, note that for any $k < s$, we have

$$0 = \hat{m}_k^{\mathrm{T}}\mathrm{diag}(\gamma_1, \ldots, \gamma_K)\hat{m}_s = \underbrace{\gamma_1\hat{m}_{k_1}\hat{m}_{s_1}}_{=O_{\mathrm{p}}(1)} + \cdots + \gamma_k\hat{m}_{k_k}\hat{m}_{s_k} + \cdots + \underbrace{\gamma_s\hat{m}_{k_s}\hat{m}_{s_s}}_{=O_{\mathrm{p}}\left(\gamma_s n^{-1/2}\right)} + O_{\mathrm{p}}(1).$$

Therefore,

$$\gamma_k\hat{m}_{s_k} = O_{\mathrm{p}}\left\{\max\left(\gamma_s n^{-1/2}, 1\right)\right\}$$

for all $k < s$. This also shows that

$$\gamma_r\hat{m}_{r_r}\hat{m}_{s_k} = O_{\mathrm{p}}\left\{\max\left(\gamma_s n^{-1/2}, 1\right)\right\}$$

for $k = 1, 2, \ldots, r$ and completes the proof.

We now prove Theorem 3.

Proof of Theorem 3. To make notation consistent with the statement of Lemma S4, we first redefine C, Ω, Ξ and L from the statement of Theorem 3 to be $\bar{C}$, $\bar{\Omega}$, $\bar{\Xi}$ and $\bar{L}$. Under the null hypothesis $\bar{\Omega} = 0$, we define

$$\hat{\bar{\Omega}} = \left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}\bar{C} = n^{-1/2}\left(n^{-1}X^{\mathrm{T}}X\right)^{-1}\hat{s}_n$$

$$\hat{s}_n = n^{-1/2}\sum_{i=1}^{n} x_i\bar{\xi}_i^{\mathrm{T}}$$

Define $a = \mathrm{vec}\left(1_d \times \bar{\xi}_1\right)$, where $1_d \in \mathbb{R}^d$ is the vector of all ones, and $\varphi_a(t)$, $t \in \mathbb{R}^{dK \times dK}$, to be the characteristic function of a. Under the null hypothesis, the gradient of $\varphi_a(t)$ is 0 and the Hessian is $-1_{d\times d} \otimes I_K$, where $1_{d\times d} \in \mathbb{R}^{d\times d}$ is the matrix of all ones. Lastly, let $t = \left(t_1^{\mathrm{T}}, \ldots, t_d^{\mathrm{T}}\right)^{\mathrm{T}}$, $t_j \in \mathbb{R}^K$. If the magnitude of the entries of X are bounded above by x, we then have that

$$\begin{aligned}\log\left\{\varphi_{\text{vec}(\hat{s}_n)}\left(t\right)\right\} &= \sum_{i=1}^{n}\log\left[\varphi_a\left\{n^{-1/2}\begin{pmatrix}x_i[1]t_1\\ \vdots\\ x_i[d]t_d\end{pmatrix}\right\}\right]\\ &= \sum_{i=1}^{n}\left[-(2n)^{-1}t^{\mathrm{T}}\left\{(x_i x_i^{\mathrm{T}})\otimes I_K\right\}t + o(n^{-1}x^2\|t\|_2^2)\right]\\ &= -2^{-1}t^{\mathrm{T}}\left(\Sigma_X\otimes I_K\right)t + o(1).\end{aligned}$$

where $\Sigma_X = \lim_{n\to\infty} n^{-1}X^{\mathrm{T}}X$. Therefore,

$$(X^{\mathrm{T}}X)^{1/2}\,\hat{\bar{\Omega}}\left\{n^{-1}\bar{\Xi}^{\mathrm{T}}P_X^{\perp}\bar{\Xi}\right\}^{-1/2} \xrightarrow{\mathcal{D}} MN_{d\times K}\left(0, I_d, I_K\right)$$

since $\|n^{-1}\bar{\Xi}^{\mathrm{T}}P_X^{\perp}\bar{\Xi} - I_K\|_2 = o_{\mathrm{p}}(1)$.

We next define Ω be the that from the statement and proof of Lemma S4, i.e.

$$\Omega = \hat{\bar{\Omega}}\left\{n^{-1}\bar{\Xi}^{\mathrm{T}}P_X^{\perp}\bar{\Xi}\right\}^{-1/2}\hat{U}$$

where $\hat{U}$ is a unitary matrix ensuring that

$$L^{\mathrm{T}}L = \hat{U}^{\mathrm{T}}\left\{n^{-1}\bar{\Xi}^{\mathrm{T}}P_X^{\perp}\bar{\Xi}\right\}^{1/2}\bar{L}^{\mathrm{T}}\bar{L}\left\{n^{-1}\bar{\Xi}^{\mathrm{T}}P_X^{\perp}\bar{\Xi}\right\}^{1/2}\hat{U}$$

is diagonal with decreasing elements. Since the assumptions of Lemma S4 hold with $\bar{\Psi} = I_K$, it is then straightforward to adapt the proof of Lemma S3 to show that $n^{1/2}\|\Omega - \hat{\Omega}\| = o_{\mathrm{p}}(1)$ under the assumptions of Theorem 3. The result then follows by an application of Slutsky's Theorem.

S2.10. *Two technical lemmas used in the proof of Lemma S*2

We now state and prove two technical lemmas are used in the proof of Lemma S2. For these two lemmas, we assume Y is distributed according to (S5) (as it is in Lemmas S1 and S2).

LEMMA S5. *Let* $U = \left(u_1 \cdots u_K\right)$, $V = \left(v_1 \cdots v_K\right)$, $D = diag\left(d_1,\ldots,d_K\right)$ *and* $\tilde{N}$ *be as defined in Lemmas S*1 *and S*2 *and suppose* $\frac{n}{p}L_s^{\mathrm{T}}\Sigma L_k = O_{\mathrm{p}}\left(\lambda_k\right)$ *for* $s\leq k$, *where* $s,k\in[K]$. *Then*

$$u_s^{\mathrm{T}}\Sigma u_k = O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_s^{-1/2}\right).$$

Proof. We need to understand how

$$U^{\mathrm{T}}\Sigma U = D^{-1}V^{\mathrm{T}}\tilde{N}^{\mathrm{T}}\Sigma\tilde{N}VD^{-1}$$

behaves. First, let $R_iR_i^{\mathrm{T}} = \tilde{L}^{\mathrm{T}}\Sigma^i\tilde{L}$ for $i = 1,2,3$ and define $\gamma = p^{-1}\operatorname{tr}\left(\Sigma^2\right)$. Then

$$R_i = \begin{bmatrix} O\left(\lambda_1^{1/2}\right) & 0 & \cdots & 0\\ O\left(\lambda_2^{1/2}\right) & O\left(\lambda_2^{1/2}\right) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ O\left(\lambda_K^{1/2}\right) & O\left(\lambda_K^{1/2}\right) & \cdots & O\left(\lambda_K^{1/2}\right)\end{bmatrix}$$

and

$$\tilde{N}^{\mathrm{T}}\Sigma\tilde{N} = \tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + p^{-1/2}\tilde{L}^{\mathrm{T}}\Sigma\tilde{\mathcal{E}}_1 + p^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K + O_{\mathrm{p}}\left(p^{-1/2}\right)$$

The next quantity we need to determine is VD^{-1}:

$$VD^{-1} = D^{-1} + \begin{bmatrix} O_{\mathrm{p}}\left(\lambda_1^{-3/2}p^{-1}\right) & O_{\mathrm{p}}\left\{(\lambda_1\lambda_2 p)^{-1/2}\right\} & \cdots & O_{\mathrm{p}}\left\{(\lambda_1\lambda_K p)^{-1/2}\right\} \\ O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1}\right\} & O_{\mathrm{p}}\left(\lambda_2^{-3/2}p^{-1}\right) & \cdots & O_{\mathrm{p}}\left\{(\lambda_2\lambda_K p)^{-1/2}\right\} \\ \vdots & \vdots & \ddots & \vdots \\ O_{\mathrm{p}}\left\{(\lambda_1 p)^{-1}\right\} & O_{\mathrm{p}}\left\{(\lambda_2 p)^{-1}\right\} & \cdots & O_{\mathrm{p}}\left(\lambda_K^{-3/2}p^{-1}\right) \end{bmatrix} = D^{-1} + e$$

and

$$R_i^{\mathrm{T}} VD^{-1} = O_{\mathrm{p}}(1) + O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\}.$$

Then for $M \sim MN_{K\times K}\left(0, I_K, I_K\right)$, we have

$$p^{-1/2}\tilde{\mathcal{E}}_1^{\mathrm{T}}\Sigma\tilde{L}VD^{-1} \sim p^{-1/2}MR_3^{\mathrm{T}}VD^{-1} = O_{\mathrm{p}}\left(p^{-1/2}\right).$$

Next

$$D^{-1}V^{\mathrm{T}}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)VD^{-1} = D^{-1}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)D^{-1} + e^{\mathrm{T}}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)D^{-1} + D^{-1}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)e + e^{\mathrm{T}}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)e$$

where

$$e^{\mathrm{T}}\tilde{L}^{\mathrm{T}}\Sigma\tilde{L}e = e^{\mathrm{T}}R_1R_1^{\mathrm{T}}e = O_{\mathrm{p}}\left\{(\lambda_K p)^{-1}\right\}$$

and

$$D^{-1}\left(\tilde{L}^{\mathrm{T}}\Sigma\tilde{L} + \gamma I_K\right)e = D^{-1}R_1R_1^{\mathrm{T}}e + O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\} = O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\}.$$

The second equality holds because $D^{-1}R_1 = O_{\mathrm{p}}(1)$ and $R_1^{\mathrm{T}}e = O_{\mathrm{p}}\left\{(\lambda_K p)^{-1/2}\right\}$. Next, let $A = \tilde{L}^{\mathrm{T}}\Sigma\tilde{L}$ and $B = D^{-1}\left(A + \gamma I_K\right)D^{-1}$. Then if $s \leq k$, $A_{sk} = O_{\mathrm{p}}\left(\lambda_k\right)$ by assumption and

$$B_{sk} = \frac{A_{sk} + \gamma\delta_{sk}}{d_s d_k} = O_{\mathrm{p}}\left(\frac{\lambda_k}{d_s d_k}\right) + \frac{\gamma}{d_s d_k}\delta_{sk} = O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_s^{-1/2}\right)$$

where $\delta_{sk} = I(s = k)$. Therefore, for $s \leq k$ $(s, k \in [K])$,

$$\left[U^{\mathrm{T}}\Sigma U\right]_{sk} = O_{\mathrm{p}}\left\{\lambda_k^{1/2}\lambda_s^{-1/2} + (\lambda_K p)^{-1/2}\right\} = O_{\mathrm{p}}\left(\lambda_k^{1/2}\lambda_s^{-1/2}\right).$$

LEMMA S6. *Let* $a_1, a_2 \in \mathbb{R}^p$ *be linearly independent unit vectors independent of* $\tilde{\mathcal{E}}_2 \sim MN_{p\times(n-K)}\left(0, \Sigma, I_{n-K}\right)$ *for* K *is a fixed constant. Recall from* (S11) *that* $R = p^{-1}\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2 - \rho I_{n-K}$ *where* $\rho = p^{-1}\operatorname{tr}(\Sigma)$. *Then*

$$p^{-1}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2R\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = O_{\mathrm{p}}\left\{\left(np^{-1}\right)^2 + np^{-3/2}\right\}.$$

Proof. Since K is a fixed constant not dependent on n or p, I will assume $\tilde{\mathcal{E}}_2 \sim MN_{p\times n}\left(0, \Sigma, I_n\right)$ for notational convenience.

$$p^{-1}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2R\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 - \rho p^{-1}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2$$

We will focus our efforts on understanding $p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2$. Define $A = \begin{pmatrix}a_1 & a_2\end{pmatrix}$, $\tilde{A} = \Sigma A$ and $Q \in \mathbb{R}^{p\times(p-2)}$ s.t. $A^{\mathrm{T}}\Sigma Q = 0_{2\times(p-2)}$. Let $P_{\tilde{A}} = GG^{\mathrm{T}}$ where $G \in \mathbb{R}^{p\times 2}$ and $P_{\tilde{A}}^{\perp} = QQ^{\mathrm{T}}$. Since $P_{\tilde{A}} + P_{\tilde{A}}^{\perp} =$

I_p, we have

$$p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}\left(P_{\tilde{A}} + P_{\tilde{A}}^{\perp}\right)\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2$$
$$= p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}P_{\tilde{A}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 + p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}P_{\tilde{A}}^{\perp}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2$$

Since $a_i^{\mathrm{T}}\Sigma a_i \le c$ and $\|G^{\mathrm{T}}\Sigma G\|_2 \le c$ for some constant $c > 0$, $\|\tilde{\mathcal{E}}_2^{\mathrm{T}}a_i\|_2 \sim \|MN_{n\times 1}\left(0, I_n, a_i^{\mathrm{T}}\Sigma a_i\right)\|_2 = O_{\mathrm{P}}\left(n^{1/2}\right)$ for $i = 1, 2$ and $\|\tilde{\mathcal{E}}_2^{\mathrm{T}}G\|_2 \sim \|MN_{n\times 2}\left(0, I_n, G^{\mathrm{T}}\Sigma G\right)\|_2 = O_{\mathrm{P}}\left(n^{1/2}\right)$. Then by Cauchy-Schwartz,

$$p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}P_{\tilde{A}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = p^{-2}\underbrace{a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2}_{1\times n}\underbrace{\tilde{\mathcal{E}}_2^{\mathrm{T}}G}_{n\times 2}G^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = O_{\mathrm{P}}\left\{\left(np^{-1}\right)^2\right\}$$

By Craig's Theorem, $\tilde{\mathcal{E}}^{\mathrm{T}}a_i$ and $\tilde{\mathcal{E}}^{\mathrm{T}}Q$ are independent, since $a_i^{\mathrm{T}}\Sigma Q = 0$. We then have

$$p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}P_{\tilde{A}}^{\perp}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 = p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}QQ^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2$$

Let $B = \Sigma^{1/2}QQ^{\mathrm{T}}\Sigma^{1/2}$ and let $H\Delta H^{\mathrm{T}}$ be its singular value decomposition. Note that $\max\Delta \le c$ for some constant $c > 0$ since QQ^{T} is just a projection matrix. Therefore, $\tilde{\mathcal{E}}_2^{\mathrm{T}}QQ^{\mathrm{T}}\tilde{\mathcal{E}}_2 \sim J^{\mathrm{T}}J$, where $J \sim MN_{p\times n}\left(0, \Delta, I_n\right)$ and is independent of $p^{-1/2}\tilde{\mathcal{E}}_2^{\mathrm{T}}a_i = \tilde{a}_i \in \mathbb{R}^{n\times 1}$ where $\|\tilde{a}_i\|_2 = O_{\mathrm{P}}\left(n^{1/2}p^{-1/2}\right)$. Define $\delta = p^{-1}\operatorname{tr}\left(\Delta\right) = \rho + O\left(p^{-1}\right)$, $\gamma = p^{-1}\operatorname{tr}\left(\Delta^2\right)$ and $b_i = \|\tilde{a}_i\|_2^{-1}\tilde{a}_i$. Then

$$p^{-2}a_1^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}QQ^{\mathrm{T}}\tilde{\mathcal{E}}_2\tilde{\mathcal{E}}_2^{\mathrm{T}}a_2 \sim \|\tilde{a}_1\|_2\|\tilde{a}_2\|_2 b_1^{\mathrm{T}}p^{-1}J^{\mathrm{T}}Jb_2 = \|\tilde{a}_1\|_2\|\tilde{a}_2\|_2 b_1^{\mathrm{T}}\begin{pmatrix} p^{-1}J_1^{\mathrm{T}}J_1 & \cdots & p^{-1}J_1^{\mathrm{T}}J_n \\ \vdots & \ddots & \vdots \\ p^{-1}J_1^{\mathrm{T}}J_n & \cdots & p^{-1}J_n^{\mathrm{T}}J_n \end{pmatrix} b_2$$

$$b_1^{\mathrm{T}}\begin{pmatrix} p^{-1}J_1^{\mathrm{T}}J_1 & \cdots & p^{-1}J_1^{\mathrm{T}}J_n \\ \vdots & \ddots & \vdots \\ p^{-1}J_1^{\mathrm{T}}J_n & \cdots & p^{-1}J_n^{\mathrm{T}}J_n \end{pmatrix} b_2 = \sum_{i=1}^{n} b_1[i]b_2[i]p^{-1}J_i^{\mathrm{T}}J_i + \sum_{i\neq q} b_1[i]b_2[q]p^{-1}J_i^{\mathrm{T}}J_q$$

$$\sum_{i=1}^{n} b_1[i]b_2[i]p^{-1}J_i^{\mathrm{T}}J_i \underbrace{=}_{X_i = \frac{1}{p}J_i^{\mathrm{T}}J_i - \delta} \delta b_1^{\mathrm{T}}b_2 + \underbrace{\sum_{i=1}^{n} b_1[i]b_2[i]X_i}_{=X}$$

$$\operatorname{var}(X) = \sum_{i=1}^{n} b_1[i]^2 b_2[i]^2 \operatorname{var}\left(X_i\right) = 2\gamma p^{-1}\sum_{i=1}^{n} b_1[i]^2 b_2[i]^2 \le 2\gamma p^{-1}$$

$$\Rightarrow \sum_{i=1}^{n} b_1[i]b_2[i]p^{-1}J_i^{\mathrm{T}}J_i = \delta b_1^{\mathrm{T}}b_2 + O_{\mathrm{P}}\left(p^{-1/2}\right) = \rho b_1^{\mathrm{T}}b_2 + O_{\mathrm{P}}\left(p^{-1/2}\right). \tag{S24}$$

Note that $E\left(\sum_{i\neq q} b_1[i]b_2[q]p^{-1}J_i^{\mathrm{T}}J_q\right) = 0$, meaning

$$\operatorname{var}\left(\sum_{i\neq q} b_1[i]b_2[q]p^{-1}J_i^{\mathrm{T}}J_q\right) = E\left\{\left(\sum_{i\neq q} b_1[i]b_2[q]p^{-1}J_i^{\mathrm{T}}J_q\right)^2\right\}.$$

Therefore,

$$\operatorname{var}\left(\sum_{i\neq q} b_1[i]b_2[q]p^{-1}J_i^{\mathrm{T}}J_q\right) = p^{-2}\sum_{i\neq q}\sum_{r\neq s} b_1[i]b_2[q]b_1[r]b_2[s]E\left\{\left(J_i^{\mathrm{T}}J_q\right)\left(J_r^{\mathrm{T}}J_s\right)\right\}.$$

We then need to go through various scenarios to evaluate the above expression.

1. $i \neq r, s$ and $q \neq r, s$. Then,
$$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_r^{\mathrm{T}} J_s\right)\right\} = 0$$
2. $i = r$.
   (a) $q \neq s$
   $$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_i^{\mathrm{T}} J_s\right)\right\} = E\left\{J_q^{\mathrm{T}} E\left(J_i J_i^{\mathrm{T}} \mid J_q, J_s\right) J_s\right\} = E\left(J_q^{\mathrm{T}} \Delta J_s\right) = \mathrm{tr}\left\{\Delta E\left(J_s J_q^{\mathrm{T}}\right)\right\} = 0$$
   (b) $q = s$
   $$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_i^{\mathrm{T}} J_q\right)\right\} = E\left\{J_q^{\mathrm{T}} E\left(J_i J_i^{\mathrm{T}} \mid J_q\right) J_q\right\} = E\left(J_q^{\mathrm{T}} \Delta J_q\right) = \mathrm{tr}\left(\Delta^2\right) = p\gamma$$
3. $i = s$
   (a) $q \neq r$
   $$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_r^{\mathrm{T}} J_i\right)\right\} = E\left\{J_q^{\mathrm{T}} E\left(J_i J_i^{\mathrm{T}} \mid J_q, J_r\right) J_r\right\} = E\left(J_q^{\mathrm{T}} \Delta J_r\right) = 0$$
   (b) $q = r$
   $$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_q^{\mathrm{T}} J_i\right)\right\} = p\gamma$$
4. $q = s, i \neq r$ (we already have the case $q = s, i = r$ above).
$$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_r^{\mathrm{T}} J_q\right)\right\} = 0$$
5. $q = r, i \neq s$ (we already have the case $q = r, i = s$ above).
$$E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_q^{\mathrm{T}} J_s\right)\right\} = 0$$

Therefore,
$$p^{-2} \sum_{i \neq q} \sum_{r \neq s} b_1[i] b_2[q] b_1[r] b_2[s] E\left\{\left(J_i^{\mathrm{T}} J_q\right)\left(J_r^{\mathrm{T}} J_s\right)\right\} = \gamma p^{-1} \sum_{i \neq q} b_1[i]^2 b_2[q]^2 + \gamma p^{-1} \sum_{i \neq q} b_1[i] b_2[i] b_1[q] b_2[q]$$

$$\sum_{i \neq q} b_1[i]^2 b_2[q]^2 \leq \sum_{i=1}^{n} b_1[i]^2 \sum_{q=1}^{n} b_2[q]^2 = 1$$

$$\sum_{i \neq q} b_1[i] b_2[i] b_1[q] b_2[q] = \sum_{i=1}^{n} b_1[i] b_2[i] \sum_{q \neq i}^{n} b_1[q] b_2[q], \quad |\sum_{q \neq i}^{n} b_1[q] b_2[q]| \leq \|b_{1,-i}\|_2 \|b_{2,-i}\|_2 \leq 1$$

$$\Rightarrow |\sum_{i=1}^{n} b_1[i] b_2[i] \sum_{q \neq i}^{n} b_1[q] b_2[q]| \leq \left\{\sum_{i=1}^{n} \left(\sum_{q \neq i}^{n} b_1[q] b_2[q]\right)^2 b_1[i]^2\right\}^{1/2} \|b_2\|_2 \leq \|b_1\|_2 \|b_2\|_2 = 1$$

Therefore $\mathrm{var}\left(\sum_{i \neq q} b_1[i] b_2[q] p^{-1} J_i^{\mathrm{T}} J_q\right) \leq \gamma p^{-1}$, meaning
$$\begin{aligned} p^{-2} a_1^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} a_2 &= \|\tilde{a}_1\|_2 \|\tilde{a}_2\|_2 \rho b_1^{\mathrm{T}} b_2 + \|\tilde{a}_1\|_2 \|\tilde{a}_2\|_2 O_{\mathrm{p}}\left(p^{-1/2}\right) + O_{\mathrm{p}}\left\{\left(np^{-1}\right)^2\right\} \\ &= \rho p^{-1} a_1^{\mathrm{T}} \tilde{\mathcal{E}}_2 \tilde{\mathcal{E}}_2^{\mathrm{T}} a_2 + O_{\mathrm{p}}\left(np^{-3/2}\right) + O_{\mathrm{p}}\left\{\left(np^{-1}\right)^2\right\}. \end{aligned}$$

Therefore,
$$p^{-1} a_1^{\mathrm{T}} \tilde{\mathcal{E}}_2 R \tilde{\mathcal{E}}_2^{\mathrm{T}} a_2 = O_{\mathrm{p}}\left(np^{-3/2}\right) + O_{\mathrm{p}}\left\{\left(np^{-1}\right)^2\right\}.$$

REFERENCES

ANDERSON, T. W. & RUBIN, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*. Berkeley, California: University of California Press.

AUFFINGER, A. & TANG, S. (2015). Extreme eigenvalues of sparse, heavy tailed random matrices. *arXiv:1506.06175v1* .

ELDAR, Y. & KUTYNIOK, G. (2012). *Compressed Sensing: Theory and Applications*. Cambridge University Press.

LEE, S., SUN, W., WRIGHT, F. A. & ZOU, F. (2017). An improved and explicit surrogate variable analysis procedure by coefficient adjustment. *Biometrika* **104**, 303–316.

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica* **17**, 1617–1642.

WANG, J., ZHAO, Q., HASTIE, T. & OWEN, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *The Annals of Statistics* **45**, 1863–1894.

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]