Supplement to "Improved inference on the rank of a matrix"

(*Quantitative Economics*, Vol. 10, No. 4, November 2019, 1787–1824)

Qihui Chen
School of Management and Economics and Shenzhen Finance Institute,
The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen)

Zheng Fang
Department of Economics, Texas A&M University

For convenience of the reader, we commence by gathering some notation that appear in the paper, most of which are standard in the literature.

$\mathbf{M}^{m\times k}$	The space of $m \times k$ real matrices for $m, k \in \mathbf{N}$.
I_k	The identity matrix of size k.
$\mathbf{0}_k, \mathbf{1}_k$	The $k \times 1$ vectors of zeros and ones.
$A^\intercal$	The transpose of a matrix $A \in \mathbf{M}^{m\times k}$.
$\mathrm{tr}(A)$	The trace of a square matrix $A \in \mathbf{M}^{k\times k}$.
$\mathrm{vec}(A)$	The column vectorization of $A \in \mathbf{M}^{m\times k}$.
$\lVert A\rVert$	The Frobenius norm of a matrix $A \in \mathbf{M}^{m\times k}$.
$\sigma_j(A)$	The jth largest singular value of a matrix $A \in \mathbf{M}^{m\times k}$.
$\mathbb{S}^{m\times k}$	A subset of $\mathbf{M}^{m\times k}$: $\mathbb{S}^{m\times k} \equiv \{U \in \mathbf{M}^{m\times k} : U^\intercal U = I_k\}$.
$C(T)$	The space of continuous functions on a (topological) space T.
$\varphi : \mathbb{D} \twoheadrightarrow \mathbb{E}$	A correspondence from a set $\mathbb{D}$ to another set $\mathbb{E}$.

Due to the fundamental role played by the singular value decomposition in the paper, we next provide a brief review and emphasize facts that are relevant to our development. Conceptually, the singular value decomposition generalizes the spectral decomposition to arbitrary (possibly rectangular) matrices. Let $\Pi \in \mathbf{M}^{m\times k}$ with $m \geq k$. Then the singular value decomposition of Π is

$$\Pi = P\Sigma Q^\intercal,$$

where $P \in \mathbb{S}^{m\times m}$ and $Q \in \mathbb{S}^{k\times k}$ are orthornormal, and $\Sigma \in \mathbf{M}^{m\times k}$ is a diagonal matrix with its diagonal entries in descending order—throughout the paper such a decomposition format is silently understood. The columns of P, called the left singular vectors

Qihui Chen: qihuichen@cuhk.edu.cn
Zheng Fang: zfang@tamu.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE1139

of Π, are eigenvectors of $\Pi\Pi^\intercal$ (which is symmetric), the columns of Q, called the right singular vectors of Π, are eigenvectors of $\Pi^\intercal\Pi$ (which is also symmetric), and the diagonal entries of Σ, called the singular values, are the corresponding square roots of the eigenvalues of $\Pi\Pi^\intercal$ and also of $\Pi^\intercal\Pi$. Such a decomposition allows us to conclude that $\operatorname{rank}(\Pi)$ is precisely equal to the number of nonzero singular values.

The matrix Σ is uniquely determined, though not the matrices P and Q. If $\operatorname{rank}(\Pi) = r_0$, then we may partition P as $P = [P_1, P_2]$ such that P_1 consists of precisely the first r_0 columns of P that are associated with the nonzero singular values of Π; similarly, we may partition Q as $Q = [Q_1, Q_2]$. Then the null space of Π is precisely the column space of Q_2, and the null space of $\Pi^\intercal$ is precisely the column space of P_2. Moreover, P_2 and Q_2 are uniquely determined respectively up to postmultiplication by $(m-r_0)\times(m-r_0)$ and $(k-r_0)\times(k-r_0)$ orthonormal matrices. Fortunately, the singular values $\sigma_j(P_2^\intercal M Q_2)$ (as in (14)) for any $M \in \mathbf{M}^{m\times k}$ are invariant to such transformations.

For convenience of applied researchers who work with Stata, we have developed a command `bootranktest` that may be used to test whether a matrix of the form $E[VZ^\intercal]$ has full rank based on our two-step test. In the first step, we use the KP test to obtain the rank estimator by choosing $\beta = 0.05/15$. Its syntax is as follows:

```
bootranktest (varlist1) (varlist2) [if] [in]
```

where `varlist1` should have more variables than `varlist2`. As of now, this command is designed for i.i.d. data and employs Efron's (1979) empirical bootstrap with 500 bootstrap repetitions. We plan to refine it by adding more features in future.

The remainder of the supplement is organized as follows. Appendix A presents the proofs of our main results. Appendix B provides additional details and discussions regarding comparisons with Kleibergen and Paap (2006), while Appendix C derives some estimation results based on a sequential testing procedure. Appendix D contains some supporting lemmas. Additional examples are presented in Appendix E where special attention is paid to inference on cointegration rank.

Appendix A: Proofs of main results

Proof of Lemma 3.1. The proof is based on a simple application of the representation of extremal partial trace. Recall that $\sigma_1^2(\Pi), \ldots, \sigma_k^2(\Pi)$ are eigenvalues of $\Pi^\intercal\Pi$ in descending order. Let $d \equiv k - r$. It follows by Proposition 1.3.4 in Tao (2012) that

$$\phi_r(\Pi) = \sum_{j=r+1}^{k} \sigma_j^2(\Pi) = \inf_{u_1,\ldots,u_d} \sum_{j=1}^{d} u_j^\intercal \Pi^\intercal \Pi u_j, \tag{A.1}$$

where the infimum is taken over all $u_1, \ldots, u_d \in \mathbf{R}^k$ that are orthonormal. Noting $U \equiv [u_1, \ldots, u_d] \in \mathbb{S}^{k\times d}$, we obtain by (A.1) and the definition of Frobenius norm that

$$\phi_r(\Pi) = \inf_{U\in\mathbb{S}^{k\times d}} \operatorname{tr}\big(U^\intercal \Pi^\intercal \Pi U\big) = \inf_{U\in\mathbb{S}^{k\times d}} \|\Pi U\|^2. \tag{A.2}$$

The infimum in (A.2) is achieved on $\mathbb{S}^{k\times d}$ because $U \mapsto \|\Pi U\|^2$ is continuous, and $\mathbb{S}^{k\times d}$ is compact since it is closed and bounded. This completes the proof of the lemma. □

PROOF OF PROPOSITION 3.1. Let $d \equiv k - r$, and define $\psi_1 : \mathbf{M}^{m\times k} \to C(\mathbb{S}^{k\times d})$ by $\psi_1(\Pi)(U) = \|\Pi U\|^2$, and $\psi_2 : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ by $\psi_2(f) = \min\{f(U) : U \in \mathbb{S}^{k\times d}\}$, so that $\phi_r = \psi_2 \circ \psi_1$ by Lemma 3.1. For part (i), we proceed by verifying first-order Hadamard directional differentiability of ψ_1 and ψ_2, and then conclude by the chain rule.

Let $\{M_n\} \subset \mathbf{M}^{m\times k}$ be a sequence satisfying $M_n \to M \in \mathbf{M}^{m\times k}$, and $t_n \downarrow 0$ as $n \to \infty$. For each $n \in \mathbf{N}$, define $g_n : \mathbb{S}^{k\times d} \to \mathbf{R}$ by

$$g_n(U) = \frac{\big\|(\Pi + t_n M_n)U\big\|^2 - \|\Pi U\|^2}{t_n} = \frac{\|\Pi U + t_n M_n U\|^2 - \|\Pi U\|^2}{t_n},$$

and $g : \mathbb{S}^{k\times d} \to \mathbf{R}$ by $g(U) = 2\,\mathrm{tr}((\Pi U)^\intercal M U)$. Then by simple algebra we have

$$\begin{aligned}\sup_{U\in\mathbb{S}^{k\times d}} \big|g_n(U) - g(U)\big| &= \sup_{U\in\mathbb{S}^{k\times d}} \big|2\,\mathrm{tr}\big((\Pi U)^\intercal (M_n - M)U\big) + t_n\|M_n U\|^2\big| \\ &\le \sup_{U\in\mathbb{S}^{k\times d}} \big\{2\|\Pi U\|\,\big\|(M_n - M)U\big\| + t_n\|M_n U\|^2\big\}, \qquad \text{(A.3)}\end{aligned}$$

where the inequality follows by the triangle inequality and the Cauchy–Schwarz inequality for the trace operator. For the right-hand side of (A.3), we further have

$$\begin{aligned}&\sup_{U\in\mathbb{S}^{k\times d}} \big\{2\|\Pi U\|\,\big\|(M_n - M)U\big\| + t_n\|M_n U\|^2\big\} \\ &\le \sup_{U\in\mathbb{S}^{k\times d}} \big\{2\|\Pi\|\|U\|\|M_n - M\|\|U\| + t_n\|M_n\|^2\|U\|^2\big\} = o(1), \qquad \text{(A.4)}\end{aligned}$$

where we exploited the submultiplicativity of Frobenius norm and the facts that $\|U\| = \sqrt{d}$, $M_n \to M$ and $t_n \downarrow 0$ as $n \to \infty$. We thus conclude from (A.3) and (A.4) that $g_n \to g$ uniformly in $C(\mathbb{S}^{k\times d})$, or equivalently ψ_1 is first-order Hadamard directionally differentiable at Π with derivative $\psi'_{1,\Pi} : \mathbf{M}^{m\times k} \to C(\mathbb{S}^{k\times d})$ given by

$$\psi'_{1,\Pi}(M)(U) = 2\,\mathrm{tr}\big((\Pi U)^\intercal M U\big). \qquad \text{(A.5)}$$

On the other hand, Theorem 3.1 in Shapiro (1991) implies that $\psi_2 : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ is first-order Hadamard directionally differentiable at any $f \in C(\mathbb{S}^{k\times d})$ with derivative $\psi'_{2,f} : C(\mathbb{S}^{k\times d}) \to \mathbf{R}$ given by: for $\Psi(f) \equiv \arg\min_{U\in\mathbb{S}^{k\times d}} f(U)$,

$$\psi'_{2,f}(h) = \min_{U\in\Psi(f)} h(U). \qquad \text{(A.6)}$$

Combining (A.5), (A.6), and the chain rule (Shapiro (1990, Proposition 3.6)), we may now conclude that $\phi_r : \mathbf{M}^{m\times k} \to \mathbf{R}$ is first-order Hadamard directionally differentiable at any $\Pi \in \mathbf{M}^{m\times k}$ with the derivative $\phi'_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ given by

$$\phi'_{r,\Pi}(M) = \psi'_{2,\psi_1(\Pi)} \circ \psi'_{1,\Pi}(M) = \min_{U\in\Psi(\Pi)} 2\,\mathrm{tr}\big((\Pi U)^\intercal M U\big).$$

This completes the proof of part (i) of the proposition.

For part (ii), note that $\phi_r(\Pi) = 0$ implies that $\Pi U = 0$ for all $U \in \Psi(\Pi)$, and hence $\phi'_{r,\Pi}(M) = 0$ for all $M \in \mathbf{M}^{m\times k}$. Recall that $\{M_n\} \subset \mathbf{M}^{m\times k}$ with $M_n \to M \in \mathbf{M}^{m\times k}$ and $t_n \downarrow 0$ as $n \to \infty$. By Lemma 3.1, we have

$$\begin{aligned}
&\big|\phi_r(\Pi + t_n M_n) - \phi_r(\Pi + t_n M)\big| \\
&= \Big|\min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M_n)U\big\| - \min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M)U\big\|\Big| \\
&\quad \times \Big(\min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M_n)U\big\| + \min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M)U\big\|\Big),
\end{aligned} \tag{A.7}$$

where the equality also exploited the elementary formula $a^2 - b^2 = (a+b)(a-b)$. For the first term on the right-hand side of (A.7), we have

$$\Big|\min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M_n)U\big\| - \min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M)U\big\|\Big| \le t_n\sqrt{d}\|M_n - M\| = o(t_n), \tag{A.8}$$

where the inequality follows by the Lipschitz continuity of the infimum operator, the triangle inequality, $\|\cdot\|$ being submultiplicative, and $\|U\| = \sqrt{d}$ for $U \in \mathbb{S}^{k\times d}$. For the second term on the right-hand side of (A.7), we have: for any fixed $U^* \in \Psi(\Pi)$,

$$\begin{aligned}
&\min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M_n)U\big\| + \min_{U\in\mathbb{S}^{k\times d}} \big\|(\Pi + t_n M)U\big\| \\
&\le \big\|(\Pi + t_n M_n)U^*\big\| \big\|(\Pi + t_n M)U^*\big\| \le t_n\|M_n\|\big\|U^*\big\| + t_n\|M\|\big\|U^*\big\| \\
&= O(t_n),
\end{aligned} \tag{A.9}$$

where we exploited $\Pi U^* = 0$, the submultiplicativity of Frobenius norm, $\|U^*\| = \sqrt{d}$ and $M_n \to M$ as $n \to \infty$. Combining (A.7)–(A.9), we thus obtain

$$\big|\phi_r(\Pi + t_n M_n) - \phi_r(\Pi + t_n M)\big| = o\big(t_n^2\big). \tag{A.10}$$

Next, for $\epsilon > 0$, let $\Psi(\Pi)^\epsilon \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \le \epsilon\}$ and $\Psi(\Pi)_1^\epsilon \equiv \{U \in \mathbb{S}^{k\times d} : \min_{U'\in\Psi(\Pi)} \|U' - U\| \ge \epsilon\}$. In what follows, we consider the nontrivial case when $\Pi \ne 0$ and $M \ne 0$. Then we must have $\Psi(\Pi) \subsetneq \mathbb{S}^{k\times d}$, and hence $\Psi(\Pi)_1^\epsilon \ne \emptyset$ for ϵ sufficiently small. Let $\sigma^+_{\min}(\Pi)$ denote the smallest positive singular value of Π which exists since $\Pi \ne 0$, and $\Delta \equiv 3\sqrt{2}[\sigma^+_{\min}(\Pi)]^{-1} \max_{U\in\mathbb{S}^{k\times d}} \|MU\| > 0$ since $M \ne 0$. Then it follows that for all n sufficiently large

$$\begin{aligned}
\min_{U\in\Psi(\Pi)_1^{t_n\Delta}} \big\|(\Pi + t_n M)U\big\| &\ge \min_{U\in\Psi(\Pi)_1^{t_n\Delta}} \|\Pi U\| - t_n \max_{U\in\mathbb{S}^{k\times d}} \|MU\| \\
&\ge \frac{\sqrt{2}}{2} t_n \sigma^+_{\min}(\Pi)\Delta - t_n \max_{U\in\mathbb{S}^{k\times d}} \|MU\| > t_n \max_{U\in\mathbb{S}^{k\times d}} \|MU\| \\
&\ge \min_{U\in\Psi(\Pi)} \big\|(\Pi + t_n M)U\big\| \ge \sqrt{\phi_r(\Pi + t_n M)},
\end{aligned} \tag{A.11}$$

where the first inequality follows by the triangle inequality and the fact that $\Psi(\Pi)_1^{t_n\Delta} \subset S^{k\times d}$, the second inequality follows by Lemma D.1, the third inequality is due to the definition of Δ, and the fourth inequality holds by the fact that $\Pi U = 0$ for $U \in \Psi(\Pi)$. By (A.11), we thus obtain that, for all n sufficiently large

$$\phi_r(\Pi + t_n M) = \min_{U \in \Psi(\Pi)^{t_n\Delta}} \|(\Pi + t_n M)U\|^2. \tag{A.12}$$

Now, for fixed $U \in \Psi(\Pi)$, $\Delta > 0$ and $t \in \mathbf{R}$, let $\Gamma^\Delta \equiv \{V \in \mathbf{M}^{k\times d} : \|V\| \le \Delta\}$ and $\Gamma_U^\Delta(t) \equiv \{V \in \Gamma^\Delta : U + tV \in \mathbb{S}^{k\times d}\} = \{V \in \Gamma^\Delta : V^\intercal U + U^\intercal V = -tV^\intercal V\}$. Define a correspondence $\varphi : \mathbf{R} \twoheadrightarrow \mathbb{S}^{k\times d} \times \Gamma^\Delta$ by $\varphi(t) = \{(U, V) : U \in \Psi(\Pi), V \in \Gamma_U^\Delta(t)\}$. Then the right-hand side of (A.12) can be written as

$$\begin{aligned}\min_{U \in \Psi(\Pi)^{t_n\Delta}} \|(\Pi + t_n M)U\|^2 &= \min_{(U,V)\in\varphi(t_n)} \|(\Pi + t_n M)(U + t_n V)\|^2 \\ &= t_n^2 \min_{(U,V)\in\varphi(t_n)} \|\Pi V + MU\|^2 + o(t_n^2),\end{aligned} \tag{A.13}$$

where we exploited $\Pi U = 0$ for all $U \in \Psi(\Pi)$ and $\|MV\| \le \|M\|\Delta$ for all $V \in \Gamma^\Delta$. By Lemma D.2, $\varphi(t)$ is continuous at $t = 0$. Since φ is obviously compact-valued, we may then obtain by Theorem 17.31 in Aliprantis and Border (2006) that

$$\begin{aligned}\min_{(U,V)\in\varphi(t_n)} \|\Pi V + MU\|^2 &= \min_{(U,V)\in\varphi(0)} \|\Pi V + MU\|^2 + o(1) \\ &= \min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 + o(1),\end{aligned} \tag{A.14}$$

where the second equality holds by letting Δ sufficiently large in view of Lemma D.3. Combining (A.12), (A.13), and (A.14) then yields

$$\phi_r(\Pi + t_n M) = t_n^2 \min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 + o(t_n^2). \tag{A.15}$$

The proposition now follows from result (A.15) and Lemma D.4. □

Proof of Theorem 3.1. The first and second results are respectively straightforward implications of Theorems 2.1 in Fang and Santos (2018) and Chen and Fang (2019) by noting that $\phi'_{r,\Pi_0} = 0$ under H_0. In particular, their Assumptions 2.1 are satisfied in view of Proposition 3.1 and their Assumptions 2.2 are satisfied by Assumption 3.1. □

Proof of Theorem 3.2. By the rate conditions on $\{\kappa_n\}$ and Assumption 3.1, the numerical estimator (16) satisfies the condition (15) by Proposition 3.1 in Chen and Fang (2019), while the analytic estimator in (17) and (9) does so by Lemma D.6. In turn, following exactly the same proof of Corollary 3.2 in Fang and Santos (2018), we obtain that $\hat{c}_{n,1-\alpha} \xrightarrow{p} c_{1-\alpha}$ by Assumption 3.2 and the quantile restrictions on $c_{1-\alpha}$. Thus, under H_0, the first claim follows from combining Theorem 3.1, Slutsky's lemma, $c_{1-\alpha}$ being a continuity point of the limiting law and the Portmanteau theorem.

For the second claim, Consider first the numerical estimator (16). Note that by Assumption 3.2, $\hat{\mathcal{M}}_n^* = O_{P_W}(1)$ in P_X-probability. Together with Assumption 3.1, $\kappa_n = o(1)$ as $n \to \infty$ and continuity of ϕ_r, we in turn see that, in P_X-probability,

$$\phi_r(\hat{\Pi}_n + \kappa_n \hat{\mathcal{M}}_n^*) = O_{P_W}(1). \tag{A.16}$$

By the definition of $\hat{c}_{n,1-\alpha}$, it follows from (A.16) and $\phi_r(\hat{\Pi}_n) \geq 0$ that

$$\kappa_n^2 \hat{c}_{n,1-\alpha} \leq O_{P_W}(1) \tag{A.17}$$

in P_X-probability. By Assumption 3.1 and continuity of ϕ_r at Π_0, we have: under H_1,

$$\phi_r(\hat{\Pi}_n) \xrightarrow{p} \phi_r(\Pi_0) > 0. \tag{A.18}$$

Combining results (A.17) and (A.18), together with $\tau_n \kappa_n \to \infty$, we thus conclude that

$$P\big(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}\big) = P\big((\tau_n \kappa_n)^2 \phi_r(\hat{\Pi}_n) > \kappa_n^2 \hat{c}_{n,1-\alpha}\big) = 1. \tag{A.19}$$

For the analytic estimator, let $\hat{d}_n \equiv k - \hat{r}_n$ and $d \equiv k - r$. By Lemma 3.1, we have

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) = \min_{U \in \mathbb{S}^{\hat{d}_n \times d}} \big\| \hat{P}_{2,n}^{\intercal} \hat{\mathcal{M}}_n^* \hat{Q}_{2,n} U \big\|^2 \leq \big\| \hat{\mathcal{M}}_n^* \big\|^2 mkd, \tag{A.20}$$

where the second inequality exploited $\|\hat{P}_{2,n}^{\intercal}\|^2 \|\hat{Q}_{2,n}\|^2 \leq mk$ and $\|U\|^2 = d$. Since $\hat{\mathcal{M}}_n^* = O_{P_W}(1)$ in P_X-probability by Assumption 3.2, it follows from (A.20) that

$$\hat{c}_{n,1-\alpha} \leq O_{P_W}(1) \tag{A.21}$$

in P_X-probability. Combining (A.18) and (A.21), together with $\tau_n \to \infty$, we thus obtain

$$P\big(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}\big) = 1. \tag{A.22}$$

This completes the proof of the second claim. □

PROOF OF THEOREM 3.3. For notational simplicity, define

$$A_n = \{\hat{r}_n > r\}, \qquad B_n = \big\{\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}\big\}, \qquad C_n = \{\hat{r}_n = r_0\}. \tag{A.23}$$

It follows that, under H_0,

$$\begin{aligned}
\limsup_{n\to\infty} E[\psi_n] &\leq \limsup_{n\to\infty} P\big((A_n \cup B_n) \cap C_n\big) + \limsup_{n\to\infty} P\big((A_n \cup B_n) \cap C_n^c\big) \\
&\leq \limsup_{n\to\infty} P(A_n \cap C_n) + \limsup_{n\to\infty} P(B_n \cap C_n) + \limsup_{n\to\infty} P\big(C_n^c\big) \\
&\leq 0 + \alpha - \beta + \beta = \alpha,
\end{aligned} \tag{A.24}$$

where we exploited $A_n \cap C_n = \emptyset$ under H_0, $\limsup_{n\to\infty} P(B_n \cap C_n) \leq \alpha - \beta$ by Theorem 3.2, and $\limsup_{n\to\infty} P(C_n^c) \leq \beta$. This completes the proof of the first claim. For the second claim of the theorem, note that

$$\liminf_{n\to\infty} E[\psi_n] \geq \liminf_{n\to\infty} P\big(\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha+\beta}\big) = 1, \tag{A.25}$$

where the equality follows by the proof of Theorem 3.2. □

Proof of Proposition 3.2. By Assumption 3.1′(ii)(iii), we have

$$\tau\{\hat{\Pi}_n - \Pi_0\} = \tau_n\{\hat{\Pi}_n - \Pi_{0,n}\} + \tau_n\{\Pi_{0,n} - \Pi_0\} \xrightarrow{L} \mathcal{M} + \Delta. \tag{A.26}$$

This in turn allows us to conclude by Proposition 3.1 and $\phi_r(\Pi_0) = 0$. □

Appendix B: Comparisons with the KP test

In this section, we first review the KP test for the reader's convenience, and then provide additional results regarding comparisons with Kleibergen and Paap (2006).

To describe the KP test, let $\hat{\Pi}_n$ be an estimator for $\Pi_0 \in \mathbf{M}^{m\times k}$ such that

$$\sqrt{n}\big\{\mathrm{vec}(\hat{\Pi}_n) - \mathrm{vec}(\Pi_0)\big\} \xrightarrow{L} N(0, \Omega_0), \tag{B.1}$$

where the covariance matrix Ω_0 admits a consistent estimator $\hat{\Omega}_n$. Let $\hat{\Pi}_n = \hat{P}_n\hat{\Sigma}_n\hat{Q}_n^{\intercal}$ be a singular value decomposition of $\hat{\Pi}_n$, where $\hat{P}_n \in \mathbb{S}^{m\times m}$, $\hat{Q}_n \in \mathbb{S}^{k\times k}$, and $\hat{\Sigma}_n \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. For r the hypothesized value in (2), rewrite $\hat{P}_n = [\hat{P}_{1,n}, \hat{P}_{2,n}]$ and $\hat{Q}_n = [\hat{Q}_{1,n}, \hat{Q}_{2,n}]$ with $\hat{P}_{1,n} \in \mathbf{M}^{m\times r}$ and $\hat{Q}_{1,n} \in \mathbf{M}^{k\times r}$, and let $\hat{\Sigma}_{2,n}$ be the right bottom $(m-r)\times(k-r)$ submatrix of $\hat{\Sigma}_n$. Then the testing statistic proposed by Kleibergen and Paap (2006) for the hypotheses (2) is

$$T_{n,\mathrm{kp}} = n \cdot \mathrm{vec}(\hat{\Sigma}_{2,n})^{\intercal}\big[(\hat{Q}_{2,n} \otimes \hat{P}_{2,n})^{\intercal}\hat{\Omega}_n(\hat{Q}_{2,n} \otimes \hat{P}_{2,n})\big]^{-1}\mathrm{vec}(\hat{\Sigma}_{2,n}), \tag{B.2}$$

where $\otimes$ signifies the Kronecker product, and the inverse is assumed to exist asymptotically. A special case of the testing statistic designed by Robin and Smith (2000) shares exactly the same form but without the weighting matrix,[1] that is,

$$T_{n,\mathrm{rs}} = n \cdot \mathrm{vec}(\hat{\Sigma}_{2,n})^{\intercal}\,\mathrm{vec}(\hat{\Sigma}_{2,n}). \tag{B.3}$$

Kleibergen and Paap (2006) show that if $\mathrm{rank}(\Pi_0) = r$, then

$$T_{n,\mathrm{kp}} \xrightarrow{L} \chi^2\big((m-r)(k-r)\big). \tag{B.4}$$

Thus, the KP test rejects the null H_0' in (2) at the significance level α if $T_{n,\mathrm{kp}}$ is larger than the $(1-\alpha)$-quantile of $\chi^2((m-r)(k-r))$.

In Section 2, we have shown that the KP test may be invalid since the χ^2-limit of the KP statistic is derived under H_0', ignoring the possibility $\mathrm{rank}(\Pi_0) < r$. As an alternative, one may construct a valid test for (1) by a multiple test on $\mathrm{rank}(\Pi_0) = 0, 1, \ldots, r$. Indeed, to show the validity of a multiple test, let $\psi_{n,r}$ be a nonrandomized test for hypotheses of the form (2) that rejects the null if $\psi_{n,r} = 1$ and fails to reject if $\psi_{n,r} = 0$. Moreover, suppose that $\psi_{n,r}$ is a consistent test that has asymptotic null rejection rates exactly equal to α. Then one may design a valid multiple test ψ_n for (1) by setting $\psi_n = \prod_{j=0}^{r}\psi_{n,j}$, that

[1] Robin and Smith (2000) proposed a class of testing statistics (indexed by functions h in their paper) which are asymptotically equivalent.

is, ψ_n rejects H_0 if and only if all $\psi_{n,j}$'s reject. It follows that ψ_n has size control because, under H_0 and for $r_0 \equiv \mathrm{rank}(\Pi_0)$,

$$\limsup_{n\to\infty} E[\psi_n] = \limsup_{n\to\infty} P(\psi_{n,0}=1,\ldots,\psi_{n,r}=1) \le \limsup_{n\to\infty} P(\psi_{n,r_0}=1) = \alpha, \tag{B.5}$$

and that ψ_n is also consistent because, under H_1,

$$\begin{aligned}\liminf_{n\to\infty} E[\psi_n] &= \liminf_{n\to\infty} P(\psi_{n,0}=1,\ldots,\psi_{n,r}=1)\\ &\ge 1-\sum_{j=1}^{r}\Big[1-\liminf_{n\to\infty} P(\psi_{n,j}=1)\Big] = 1,\end{aligned} \tag{B.6}$$

where the inequality holds by the Boole's inequality and consistency of each $\psi_{n,j}$. This shows that ψ_n is valid, in fact consistent but may be conservative. The source of conservativeness of ψ_n is inherent in the inequality of (B.5) which is generically strict. Moreover, ψ_n is conservative whenever $\psi_{n,r}$ is, because

$$\limsup_{n\to\infty} E[\psi_n] = \limsup_{n\to\infty} P(\psi_{n,0}=1,\ldots,\psi_{n,r}=1) \le \limsup_{n\to\infty} P(\psi_{n,r}=1) < \alpha. \tag{B.7}$$

The remainder of this section is devoted to additional comparisons of our tests with the KP test based on the simulation designs and empirical application in Kleibergen and Paap (2006). First, following those authors, we consider, for $R_t \in \mathbf{R}^{10}$, $F_t \in \mathbf{R}^4$,

$$R_t = \Pi_0 F_t + \varepsilon_t, \tag{B.8}$$

where $\{F_t\} \overset{i.i.d.}{\sim} N(0,\Sigma_F)$ and $\{\epsilon_t\}$ are independently generated according to

$$\varepsilon_t = v_t + \Gamma v_{t-1} \tag{B.9}$$

with $\{v_t\} \overset{i.i.d.}{\sim} N(0,\Sigma_v)$. We are interested in Π_0 which is specified as

$$\Pi_0 = \beta\alpha^\intercal + \delta\Pi_1, \tag{B.10}$$

where $\delta \in \mathbf{R}$, $\alpha \in \mathbf{R}^4$, $\beta \in \mathbf{R}^{10}$ and $\Pi_1 \in \mathbf{M}^{10\times 4}$. Kleibergen and Paap (2006) try a wide range of values for δ; we shall focus on $\delta = 0, 0.01, \ldots, 0.1$ since we are concerned with local power. Other unknown parameters involved are configured to be exactly the same as those in Kleibergen and Paap (2006):

- Σ_F is specified as the sample correlation matrix of $\{F_t\}_{t=1}^n$, where $\{F_t\}_{t=1}^n$ is the real data to be studied for the empirical application;
- $\alpha = (0.0813, -0.0271, -0.6203, -0.0460)^\intercal$;
- $\beta = (-0.3411, -0.1277, -0.3838, -0.5312, -0.2728, -0.3527, -0.2188, -0.293, -0.2035, -0.3427)^\intercal$;
- $\Pi_1 = \bar{\Pi}_n - \beta\alpha^\intercal$, where $\bar{\Pi}_n = \sum_{t=1}^n R_t F_t^\intercal (\sum_{t=1}^n F_t F_t^\intercal)^{-1}$ with $\{R_t, F_t\}_{t=1}^n$ being the real data in the empirical application;

- Γ is specified as

$$\Gamma = \begin{bmatrix} 0.0312 & 0.0255 & -0.0185 & 0.0591 & 0.0389 & 0.0953 & -0.1515 & 0.2286 & -0.0806 & -0.1659 \\ 0.0346 & -0.0166 & -0.0608 & 0.0743 & 0.0794 & -0.0043 & -0.2194 & 0.2959 & -0.0043 & 0.0016 \\ -0.0304 & 0.0624 & -0.1347 & 0.1054 & -0.0369 & -0.0187 & -0.0989 & 0.3571 & 0.0133 & -0.1731 \\ -0.0414 & 0.0951 & 0.0029 & -0.0497 & -0.0586 & 0.0910 & -0.0903 & 0.1850 & 0.0616 & -0.0865 \\ -0.0570 & -0.0845 & 0.0606 & -0.0143 & -0.1971 & 0.0528 & 0.0403 & 0.1935 & -0.0114 & 0.1141 \\ -0.0649 & -0.0738 & 0.0030 & 0.0335 & 0.0346 & -0.0432 & -0.0787 & 0.2199 & -0.0266 & -0.0013 \\ -0.0334 & -0.1163 & -0.0139 & -0.0218 & -0.0390 & 0.0128 & -0.0645 & 0.1299 & 0.1105 & 0.0097 \\ -0.1029 & 0.0368 & 0.0737 & -0.0005 & -0.1686 & 0.0254 & 0.0184 & 0.0966 & -0.0176 & 0.0596 \\ -0.1153 & 0.0008 & 0.0373 & 0.0185 & -0.0927 & 0.1029 & 0.0546 & 0.0529 & -0.1792 & 0.0798 \\ -0.0737 & -0.0669 & 0.0500 & 0.1466 & -0.1359 & 0.0617 & 0.1090 & 0.0402 & -0.0659 & -0.0440 \end{bmatrix};$$

- Σ_v is specified as

$$\Sigma_v = \frac{1}{100} \begin{bmatrix} 0.19 & 0.09 & 0.07 & 0.05 & 0.04 & 0.03 & 0.02 & -0.01 & 0.00 & -0.01 \\ 0.09 & 0.11 & 0.06 & 0.05 & 0.04 & 0.04 & 0.03 & 0.01 & 0.02 & 0.01 \\ 0.07 & 0.06 & 0.10 & 0.05 & 0.04 & 0.04 & 0.03 & 0.03 & 0.02 & 0.01 \\ 0.05 & 0.05 & 0.05 & 0.08 & 0.04 & 0.04 & 0.04 & 0.03 & 0.02 & 0.01 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.08 & 0.05 & 0.05 & 0.05 & 0.04 & 0.03 \\ 0.03 & 0.04 & 0.04 & 0.04 & 0.05 & 0.08 & 0.06 & 0.05 & 0.05 & 0.03 \\ 0.02 & 0.03 & 0.03 & 0.04 & 0.05 & 0.06 & 0.08 & 0.06 & 0.05 & 0.03 \\ -0.01 & 0.01 & 0.03 & 0.03 & 0.05 & 0.05 & 0.06 & 0.10 & 0.07 & 0.05 \\ 0.00 & 0.02 & 0.02 & 0.02 & 0.04 & 0.05 & 0.05 & 0.07 & 0.09 & 0.04 \\ -0.01 & 0.01 & 0.01 & 0.01 & 0.03 & 0.03 & 0.03 & 0.05 & 0.04 & 0.07 \end{bmatrix}.$$

Given the above configurations, we test the hypotheses $\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \leq r$ vs. $\mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r$ for $r = 3$ at $\alpha = 5\%$. Thus, H_0 holds if and only if $\delta = 0$, in which case $\mathrm{rank}(\Pi_0) < r$. Note that Kleibergen and Paap (2006) instead consider $\mathrm{H}_0' : \mathrm{rank}(\Pi_0) = 1$ vs. $\mathrm{H}_1' : \mathrm{rank}(\Pi_0) > 1$ so that the possibility $\mathrm{rank}(\Pi_0) < 1$ is excluded. We estimate Π_0 based a sample $\{R_t, F_t\}_{t=1}^n$ of size $n = 330$ (as in Kleibergen and Paap (2006)) that is generated according to the process (B.8). The number of simulation replications is set to be 5000, while the number of block bootstrap repetitions (with block size 2) is 500 for each simulation replication. We implement the three of our tests in same manner as we did in Section 4, and compare with the multiple KP test (based on the HACC estimator for the long run variance), although the results for the direct application of the KP test are similar and available upon request.

Table B.1 summarizes the simulation results. We find patterns similar to those exhibited in Table 2. In particular, the multiple KP test is severely undersized, and its local power is overall dominated by our tests, though again the test based on numerical derivative estimators (CF-N) is somewhat sensitive to the choices of the step size. The two-step test (CF-T) and the test based on numerical derivative estimators (CF-A), on the other hand, show strong insensitivity to the choices of the tuning parameters.

Finally, following Kleibergen and Paap (2006), we study a stochastic discount factor model based on the conditional capital asset pricing model proposed in the influential work of Jagannathan and Wang (1996). Suppose that $R_t \in \mathbf{R}^m$ is a vector of returns on m assets at time t and $F_t \in \mathbf{R}^k$ is a vector of k common factors at time t. According to the stochastic discount factor model, R_t and F_t are related through

$$E\big[R_{t+1}F_{t+1}^{\intercal}\gamma_0 | \mathcal{I}_t\big] = \mathbf{1}_m, \tag{B.11}$$

TABLE B.1. Rejection rates of rank tests for the model (B.8) with $r = 3$, at $\alpha = 5\%$.

	CF-T						CF-A							
δ	$\alpha/5$	$\alpha/10$	$\alpha/15$	$\alpha/20$	$\alpha/25$	$\alpha/30$	$n^{-1/5}$	$1.5n^{-1/5}$	$n^{-1/4}$	$1.5n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$
0.00	0.03	0.03	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.01	0.06	0.07	0.07	0.07	0.08	0.08	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
0.02	0.10	0.10	0.10	0.10	0.10	0.10	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
0.03	0.20	0.20	0.20	0.20	0.20	0.20	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.59
0.04	0.35	0.35	0.35	0.35	0.35	0.35	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
0.05	0.54	0.54	0.53	0.53	0.53	0.53	0.87	0.87	0.87	0.87	0.87	0.87	0.79	0.87
0.06	0.71	0.71	0.71	0.71	0.71	0.71	0.95	0.95	0.95	0.95	0.95	0.95	0.89	0.95
0.07	0.85	0.85	0.85	0.84	0.84	0.84	0.98	0.98	0.98	0.98	0.96	0.98	0.96	0.96
0.08	0.93	0.93	0.93	0.93	0.92	0.92	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
0.09	0.97	0.97	0.97	0.97	0.97	0.97	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99
0.10	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

		CF-N							
δ	KP-M	$n^{-1/5}$	$1.5n^{-1/5}$	$n^{-1/4}$	$1.5n^{-1/4}$	$n^{-1/3}$	$1.5n^{-1/3}$	$n^{-2/5}$	$1.5n^{-2/5}$
0.00	0.00	0.05	0.05	0.05	0.05	0.04	0.05	0.03	0.04
0.01	0.05	0.21	0.25	0.18	0.22	0.11	0.17	0.06	0.11
0.02	0.10	0.22	0.29	0.16	0.24	0.09	0.15	0.03	0.09
0.03	0.20	0.28	0.37	0.22	0.31	0.12	0.20	0.07	0.13
0.04	0.36	0.39	0.49	0.33	0.42	0.24	0.32	0.14	0.24
0.05	0.53	0.54	0.62	0.49	0.56	0.40	0.48	0.28	0.41
0.06	0.69	0.69	0.75	0.66	0.71	0.58	0.65	0.45	0.58
0.07	0.80	0.83	0.86	0.81	0.84	0.74	0.80	0.63	0.74
0.08	0.87	0.92	0.93	0.90	0.92	0.86	0.90	0.77	0.87
0.09	0.91	0.96	0.97	0.96	0.97	0.94	0.96	0.88	0.94
0.10	0.93	0.99	0.99	0.98	0.99	0.98	0.98	0.95	0.98

Note: The six values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in (9) and (27), respectively.

TABLE B.2. The p-values for different tests.

Block Size	CF-T			CF-A			CF-N		
	$\alpha/10$	$\alpha/15$	$\alpha/20$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$	$n^{-1/5}$	$n^{-1/4}$	$n^{-1/3}$
				Panel A: Our tests†					
$b=1$	0.15	0.15	0.15	0.08	0.08	0.08	0.11	0.12	0.12
$b=2$	0.15	0.15	0.15	0.10	0.09	0.09	0.11	0.11	0.12
$b=3$	0.18	0.18	0.18	0.10	0.10	0.10	0.13	0.13	0.14
$b=4$	0.16	0.16	0.16	0.08	0.08	0.08	0.13	0.13	0.14
				Panel B: the KP-M test‡					
				0.91					

Note: † The three values under CF-T are the choices of β, and those under CF-A and CF-N are the choices of κ_n as in (9) and (16), respectively.
‡ The p-value for KP-M is given by the smallest significance level such that the null hypothesis is rejected, which is equal to the maximum p-value of all Kleibergen and Paap's (2006) tests implemented by the multiple testing method.

where $\mathcal{I}_t$ represents information at time t, and $\gamma_0 \in \mathbf{R}^k$ is a vector of risk premia. If $\{R_t, F_t\}$ is governed by a stationary linear process:

$$R_t = \Pi_0 F_t + \varepsilon_t, \tag{B.12}$$

where $E[\epsilon_{t+1} F_{t+1} | \mathcal{I}_t] = 0$ and $E[F_{t+1} F_{t+1}^\intercal]$ is nonsingular, then γ_0 is identified if and only if the coefficient matrix Π_0 is of full rank. For this, we may test $\mathrm{H}_0 : \mathrm{rank}(\Pi_0) \le r$ vs. $\mathrm{H}_1 : \mathrm{rank}(\Pi_0) > r$ with $r = k - 1$.

We use the same data set as in Kleibergen and Paap (2006). There are returns R_t on 10 portfolios and 4 factors in F_t with monthly observations from July 1963 to December 1990, so $m = 10$, $k = 4$, and $n = 330$. The factors in F_t consist of constant, the return on a value-weighted portfolio, a corporate bond yield spread and a measure of per capita labor income growth. We estimate Π_0 by

$$\hat{\Pi}_n = \sum_{t=1}^{n} R_t F_t^\intercal \left(\sum_{t=1}^{n} F_t F_t^\intercal \right)^{-1}. \tag{B.13}$$

Since the return sequence $\{R_t\}$ exhibits first-order autocorrelation, we thus follow Kleibergen and Paap (2006) and compute the KP statistic by employing the HACC estimator with one lag (West (1997)) for the long run covariance matrix. We implement our CF-T, CF-A, and CF-N tests by adopting the block bootstrap (Lahiri (2003)) with block size $b = 1, 2, 3, 4$, employing the same choices of tuning parameters as before, and setting the number of bootstrap repetitions to be 1000.

Table B.2 reports the p-values of CF-T, CF-A, and CF-N, as well as that of the KP-M test. The differences between our p-values and those of the KP-M tests are substantial: ours are uniformly less than 20% while the latter are over 90%. Thus, while the KP-M test strongly support the null, our tests are inconclusive depending on the significance levels and of course also the choices of the tuning parameters. It is worth noting that our three tests are quite insensitive across all choices of tuning parameters and the block size; in particular, the p-values of CF-T and CF-A are invariant to these choices.

Appendix C: Estimation of the rank

There are settings as evident in Examples E.1 and E.3–E.5 in Appendix E.2 where one would like to construct an estimate of the rank. The need of rank estimation is further reinforced should one deem our test based on (17) desirable. Following Cragg and Donald (1997) and Robin and Smith (2000), we adopt a sequential testing procedure that has been previously employed in the literature of model selection (Pötscher (1983), Bauer, Pötscher, and Hackl (1988), Hosoya (1989)).[2]

Specifically, one may progressively test if the true rank is equal to $0, 1, \ldots, k-1$ and set the estimator $\hat{r}_n$ to be the smallest $r \in \{0, 1, \ldots, k-1\}$ that cannot be rejected if such a r exists and to be k if it does not. The conventional setup (2) then suits well to this end because the possibility of $\mathrm{rank}(\Pi_0)$ strictly smaller than the hypothesized value is "ruled out" in each step by previous test(s). However, we argue that accommodating the possibility $\mathrm{rank}(\Pi_0) < r$, as we do in what follows, may once again lead to more reliable results. Heuristically, there are two possible errors involved in the procedure, namely, falsely rejecting a true null (i.e., type I error) and not rejecting a false null (i.e., type II error). Sequentially, testing nulls of the form (2) ignore type I errors potentially made in previous steps, and may have trivial or poor power when Π_0 is local to a matrix whose rank is "small," that is, the capability of controlling type II error is limited. These are the two channels through which our rank estimator improves upon existing ones. Given a confidence level $1-\alpha$, we formally define the rank estimator $\hat{r}_n$ as

$$\hat{r}_n = \min\bigl\{r = 0, \ldots, k-1 : \tau_n^2 \phi_r(\hat{\Pi}_n) \le \hat{c}_{n,1-\alpha}(r)\bigr\} \tag{C.1}$$

if the set is nonempty, and $\hat{r}_n = k$ if the set is empty, where $\hat{c}_{n,1-\alpha}(r)$ is defined by (18) for which we also make its dependence on r explicit.

The following theorem shows that the estimator $\hat{r}_n$ in (C.1) picks up the true rank with probability at least $1-\alpha$ (asymptotically).

Theorem C.1. *Let Assumptions* 3.1 *and* 3.2 *hold, and the cdf of the limiting law in* (14) *when $r = r_0$ be continuous and strictly increasing at its $(1-\alpha)$-quantile for $\alpha \in (0,1)$. Then the rank estimator $\hat{r}_n$ defined by* (C.1) *satisfies*

$$\lim_{n\to\infty} P(\hat{r}_n = r_0) = \begin{cases} 1-\alpha & \text{if } r_0 < k, \\ 1 & \text{if } r_0 = k, \end{cases} \tag{C.2}$$

$\lim_{n\to\infty} P(\hat{r}_n < r_0) = 0$, *and* $\lim_{n\to\infty} P(\hat{r}_n > r_0) = \alpha$ (*for* $r_0 < k$).

Theorem C.1 implies that the procedure will select an estimator that is no smaller than the truth (asymptotically), and the probability of choosing a larger value (i.e., false selection) is controlled by the significance level α; see Johansen (1995) for related results in cointegration settings. These properties are intrinsically connected to the size control and consistency of our test. Moreover, by Theorem C.1, the sequential procedure can be

[2] One may alternatively employ information criteria as in Cragg and Donald (1997). We do not pursue this possibility here in order to coherently present what is essential to our paper.

utilized in our two-step test to provide a preliminary rank estimator, although we stress that existing tests can also be employed in this regard—the downside of these tests is that they may yield less accurate rank estimators as argued previously.

While the construction of a "confidence set/singleton" is of interest in its own right, one may also be interested in obtaining a consistent estimator, for which the probability of false selection should be negligible. One such an estimator is given by (9) or Lemma D.7, where a tuning parameter is involved. This estimator is somewhat crude in that the probability of false selection is unclear and appears challenging to control. Employing the sequential procedure, we may achieve consistency while controlling the estimation error. As suggested by (C.2) and noted in the literature (Pötscher (1983)), we must adjust the significance level $\alpha = \alpha_n$ according to the sample size so that $\alpha_n \to 0$ at a suitable rate, in order to obtain a consistent estimator. This turns out to be nontrivial in the current setup (where $\operatorname{rank}(\Pi_0) \leq r$ is tested in each step) as we elaborate next.

If one sequentially tests $\operatorname{rank}(\Pi_0) = r$ for $r = 0, \ldots, k-1$ based on, for example, Cragg and Donald (1997) or Kleibergen and Paap (2006), the critical values are then obtained from chi-squared distributions. The rate at which α_n should tend to zero in order to deliver consistency has been well understood in this case by exploiting the analytic expansions of the cdfs of chi-squares; see Theorem 5.8 in Pötscher (1983) for this result, Cragg and Donald (1997) for an application of it in rank estimation, and Andrews (1999) in moment selection. There are, unfortunately, two challenges for us. First, the limiting distributions whose critical values we aim to approximate is highly nonstandard in general, and as a result, deriving rate conditions on α_n through analytic expansions appears challenging to us. Second, our critical values are obtained through bootstrap, and we believe that it is nontrivial to control the sample uncertainty embodied in these critical values as $\alpha_n \downarrow 0$. Nonetheless, we show that the our rank estimator is consistent under the same rate conditions on α_n as in Cragg and Donald (1997) and Robin and Smith (2000). To formalize our discussions below, we thus impose the following.

Assumption C.1. *$\{\alpha_n\}_{n=1}^{\infty}$ satisfy* (i) $\alpha_n \downarrow 0$, *and* (ii) $\tau_n^{-2} \log \alpha_n \to 0$.

Assumption C.1 is quite mild in that it merely requires that, loosely speaking, α_n approach zero slower than exponentially decaying rates (not too fast). In this way, it encompasses a wide range of choices for α_n. Given the adjusted significance level α_n, we may now formally define the rank estimator to be

$$\tilde{r}_n = \min\big\{r = 0, \ldots, k-1 : \tau_n^2 \phi_r(\hat{\Pi}_n) \leq \hat{c}_{n,1-\alpha_n}(r)\big\} \tag{C.3}$$

if the set is nonempty, and $\tilde{r}_n = k$ if the set is empty.

The next theorem establishes if Assumption C.1 holds and $\mathcal{M}$ is Gaussian (in addition to previous assumptions), then the estimator $\tilde{r}_n$ is indeed consistent.

Theorem C.2. *Suppose that Assumptions* 3.1, 3.2, *and* C.1 *hold. Let $\tilde{r}_n$ be given by* (C.3). *If $\mathcal{M}$ is Gaussian but not constant* (*in* $\mathbf{M}^{m\times k}$), *then* $\lim_{n\to\infty} P(\tilde{r}_n = r_0) = 1$.

We reiterate that Theorem C.2 may be of use not only in estimation problems but also in conducting our rank test based on the analytic derivative estimator; see (17) and Lemma D.6. On a technical note, the Gaussianity condition plays an instrumental but not essential role. Concretely, it allows us to relate the significance levels to the corresponding critical values through a concentration inequality for Gaussian random vectors/matrices; see Lemmas D.8–D.10. Thus, this condition can be relaxed whenever a suitable concentration inequality for $\mathcal{M}$ is available (Ledoux (2001)).

Proof of Theorem C.1. For notational simplicity, define: for $r = 0, \ldots, k-1$,

$$A_{n,r} = \big\{\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha}(r)\big\}, \tag{C.4}$$

that is, $A_{n,r}$ are the events of rejecting the nulls. Consider the first the case when $r_0 = k$. Then we must have $\{\hat{r}_n = r_0\} = A_{n,0} \cap A_{n,1} \cap \cdots \cap A_{n,k-1}$, and hence

$$\begin{aligned}\liminf_{n\to\infty} P(\hat{r}_n = r_0) &= \liminf_{n\to\infty} P(A_{n,0} \cap A_{n,1} \cap \cdots \cap A_{n,k-1}) \\ &\ge 1 - \sum_{r=0}^{k-1}\Big[1 - \liminf_{n\to\infty} P(A_{n,r})\Big] = 1,\end{aligned} \tag{C.5}$$

where the inequality follows from the Boole's inequality, and the last step is because of the consistency result of Theorem 3.2.

Next, suppose $r_0 < k$. Then $\{\hat{r}_n = r_0\} = A_{n,0} \cap \cdots \cap A_{n,r_0-1} \cap A^c_{n,r_0}$, and hence

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n = r_0) &= \limsup_{n\to\infty} P\big(A_{n,0} \cap \cdots \cap A_{n,r_0-1} \cap A^c_{n,r_0}\big) \\ &\le \limsup_{n\to\infty} P\big(A^c_{n,r_0}\big) = 1 - \liminf_{n\to\infty} P(A_{n,r_0}) = 1-\alpha,\end{aligned} \tag{C.6}$$

where the last step follows from the first claim of Theorem 3.2. Moreover,

$$\begin{aligned}\liminf_{n\to\infty} P(\hat{r}_n = r_0) &= \liminf_{n\to\infty} P\big(A_{n,0} \cap \cdots \cap A_{n,r_0-1} \cap A^c_{n,r_0}\big) \\ &\ge 1 - \sum_{r=0}^{r_0-1}\Big[1 - \liminf_{n\to\infty} P(A_{n,r})\Big] - \limsup_{n\to\infty} P(A_{n,r_0}) \\ &= 1-\alpha,\end{aligned} \tag{C.7}$$

where we exploited the size control and the consistency results in Theorem 3.2.

Turning to the second claim, note that if $\hat{r}_n < r_0$, then $r_0 > 0$ and $\{\hat{r}_n < r_0\} \subset A^c_{n,0} \cup \cdots \cup A^c_{n,r_0-1}$. It follows that

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n < r_0) &\le \limsup_{n\to\infty} P\big(A^c_{n,0} \cup \cdots \cup A^c_{n,r_0-1}\big) \\ &\le \sum_{r=0}^{r_0-1} \limsup_{n\to\infty} P\big(A^c_{n,r}\big) = \sum_{r=0}^{r_0-1}\Big[1 - \liminf_{n\to\infty} P(A_{n,r})\Big] = 0,\end{aligned} \tag{C.8}$$

where the last step is because of the consistency result of Theorem 3.2. The last claim is a simple implication of the first two claims. We are thus done. □

Proof of Theorem C.2. For notational simplicity, define: for $r = 0, \ldots, k-1$,

$$A_{n,r} = \{\tau_n^2 \phi_r(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha_n}(r)\}. \tag{C.9}$$

First, note that $\tilde{r}_n < r_0$ if and only if $r_0 \geq 1$ and

$$\{\tilde{r}_n = r\} = A_{n,0} \cap \cdots \cap A_{n,r-1} \cap A_{n,r}^c \tag{C.10}$$

for some $r = 0, \ldots, r_0 - 1$. Fix $r \in \{0, 1, \ldots, r_0 - 1\}$. It follows from (C.10) that

$$P(\tilde{r}_n = r) \leq P(A_{n,r}^c) = 1 - P\left(\phi_r(\hat{\Pi}_n) > \frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2}\right) \to 0, \tag{C.11}$$

where we exploited $\hat{c}_{n,1-\alpha_n}/\tau_n^2 = o_p(1)$ by Assumption C.1(ii), Lemma D.10 and $\phi_r(\hat{\Pi}_n) \xrightarrow{p} \phi_r(\Pi_0) > 0$ by the continuous mapping theorem and $\operatorname{rank}(\Pi_0) \equiv r_0 > r$. Since the result (C.11) is true for any $r = 0, \ldots, r_0 - 1$, we thus obtain

$$\limsup_{n\to\infty} P(\tilde{r}_n < r_0) = 0. \tag{C.12}$$

Next, note that $\tilde{r}_n > r_0$ if and only if $r_0 \leq k-1$ and either the relation (C.10) holds for some $r = r_0 + 1, \ldots, k-1$ or the following event occurs:

$$\{\tilde{r}_n = k\} = A_{n,0} \cap \cdots \cap A_{n,k-1} \cap A_{n,k}. \tag{C.13}$$

Hence, $\{\tilde{r}_n = r\} \subset A_{n,r_0}$ for all $r = r_0 + 1, \ldots, k$. Fix $r = \{r_0 + 1, \ldots, k\}$. We thus have

$$P(\tilde{r}_n = r) \leq P(A_{n,r_0}) = P(\tau_n^2 \phi_{r_0}(\hat{\Pi}_n) > \hat{c}_{n,1-\alpha_n}). \tag{C.14}$$

Fix $\epsilon \in (0,1)$ so that $c_{1-\epsilon}$ is a continuity point of the cdf F of $\phi''_{r,\Pi_0}(\mathcal{M})$. This can be done without loss of generality because the set of discontinuity points is countable. By Assumption C.1(i), it holds that: for all n sufficiently large,

$$F(c_{1-\epsilon}) = 1 - \epsilon < 1 - \alpha_n, \tag{C.15}$$

and hence $c_{1-\alpha_n} > c_{1-\epsilon}$. In turn, we obtain from (C.15) and Assumption C.1(i) that there exists some $\delta > 0$ satisfying: for all n sufficiently large,

$$F(c_{1-\epsilon}) + \delta < 1 - \alpha_n. \tag{C.16}$$

Note that if $\hat{c}_{n,1-\alpha_n} \leq c_{1-\epsilon}$, then we obtain from (C.16) that

$$F(c_{1-\epsilon}) + \delta < 1 - \alpha_n \leq \hat{F}_n(\hat{c}_{n,1-\alpha_n}) \leq \hat{F}_n(c_{1-\epsilon}). \tag{C.17}$$

By Lemma 10.11 in Kosorok (2008), we may thus conclude that

$$\limsup_{n\to\infty} P(\hat{c}_{n,1-\alpha_n} \leq c_{1-\epsilon}) \leq \limsup_{n\to\infty} P(\hat{F}_n(c_{1-\epsilon}) - F(c_{1-\epsilon}) > \delta) = 0. \tag{C.18}$$

Combination of results (C.14) and (C.18), together with Assumption 3.1, now yields

$$\limsup_{n\to\infty} P(\tilde{r}_n = r) \le \limsup_{n\to\infty} P\big(\tau_n^2 \phi_{r_0}(\hat{\Pi}_n) > c_{1-\epsilon}\big) = 1 - F(c_{1-\epsilon}) \le \epsilon. \tag{C.19}$$

Since $\epsilon > 0$ and $r \in \{r_0+1, \ldots, k\}$ are both arbitrary, it follows from (C.19) that

$$\limsup_{n\to\infty} P(\tilde{r}_n > r_0) = 0. \tag{C.20}$$

The theorem now follows from results (C.12) and (C.20) since

$$\liminf_{n\to\infty} P(\tilde{r}_n = r_0) \ge 1 - \limsup_{n\to\infty} P(\tilde{r}_n < r_0) - \limsup_{n\to\infty} P(\tilde{r}_n > r_0) = 1. \qquad \square$$

Appendix D: Auxiliary lemmas

Lemma D.1. *Suppose $\Pi \in \mathbf{M}^{m\times k}$ with $\Pi \neq 0$ and $\mathrm{rank}(\Pi) \le r$. For $\epsilon > 0$, let $\Psi(\Pi)_1^\epsilon$ be given as in the proof of Proposition 3.1. Let $\sigma_{\min}^+(\Pi)$ be the smallest positive singular value of Π. Then for all sufficiently small $\epsilon > 0$, we have*

$$\min_{U\in\Psi(\Pi)_1^\epsilon} \|\Pi U\| \ge \frac{\sqrt{2}}{2}\sigma_{\min}^+(\Pi)\epsilon.$$

Proof. Let $\Pi = P\Sigma Q^\intercal$ be a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$, $Q \in \mathbb{S}^{k\times k}$, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $d \equiv k - r$ and $d_0 \equiv k - r_0$ with $r_0 \equiv \mathrm{rank}(\Pi)$. For $U \in \mathbb{S}^{k\times d}$, let $U_Q \equiv Q^\intercal U$ and write $U_Q^\intercal = [U_Q^{(1)\intercal}, U_Q^{(2)\intercal}]$ such that $U_Q^{(1)} \in \mathbf{M}^{r_0\times d}$. Then we have that for $U \in \mathbb{S}^{k\times d}$,

$$\|\Pi U\| = \big\|P\Sigma Q^\intercal U\big\| = \|\Sigma U_Q\| \ge \sigma_{\min}^+(\Pi)\big\|U_Q^{(1)}\big\|, \tag{D.1}$$

where the second equality follows by $P^\intercal P = I_m$, and the inequality follows by the fact that Σ is diagonal with diagonal entries in descending order with $\sigma_{\min}^+(\Pi) = \sigma_{r_0}(\Pi)$ the smallest positive entry. Let $U_Q^{(2)} = P_U^{(2)}\Sigma_U^{(2)}Q_U^{(2)\intercal}$ be a singular value decomposition of $U_Q^{(2)}$ where $Q_U^{(2)} \in \mathbb{S}^{d\times d}$, $P_U^{(2)} \in \mathbb{S}^{d_0\times d_0}$ and $\Sigma_U^{(2)} \in \mathbf{M}^{d_0\times d}$. Since $r_0 \le r$, and hence $d_0 \ge d$, it follows that, for $U \in \mathbb{S}^{k\times d}$,

$$\|U_Q^{(2)}\|^2 = \sum_{j=1}^d \sigma_j^2\big(U_Q^{(2)}\big) \le \sum_{j=1}^d \sigma_j\big(U_Q^{(2)}\big) = \mathrm{tr}\big([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)}\big), \tag{D.2}$$

where the inequality follows by the fact that $\sigma_j(U_Q^{(2)}) \in [0,1]$ as singular values of $U_Q^{(2)}$ due to $U_Q^{(2)\intercal}U_Q^{(2)} + U_Q^{(1)\intercal}U_Q^{(1)} = I_d$, and the second equality follows by noting that the diagonal entries of $\Sigma_U^{(2)}$ are singular values of $U_Q^{(2)}$. Since $\|U_Q^{(1)}\|^2 + \|U_Q^{(2)}\|^2 = \|U_Q\|^2 = d$, thus combining (D.1) and (D.2) yields that for $U \in \mathbb{S}^{k\times d}$,

$$\|\Pi U\| \ge \sigma_{\min}^+(\Pi)\sqrt{d - \mathrm{tr}\big([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)}\big)}. \tag{D.3}$$

Since $\|U_Q^{(1)}\|^2 + \|\Sigma_U^{(2)}\|^2 = \|U_Q^{(1)}\|^2 + \|U_Q^{(2)}\|^2 = d$ and $\|[I_d, \mathbf{0}_{r-r_0}]^\intercal\|^2 = d$, then simple algebra yields that for $U \in \mathbb{S}^{k\times d}$,

$$2\big(d - \operatorname{tr}\big([I_d, \mathbf{0}_{d-r_0}]\Sigma_U^{(2)}\big)\big) = \big\|U_Q^{(1)}\big\|^2 + \big\|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal\big\|^2. \tag{D.4}$$

Write $Q = [Q_1, Q_2]$ such that $Q_1 \in \mathbf{M}^{k\times r_0}$. Since $Q_1^\intercal Q_1 = I_{r_0}$, $Q_2^\intercal Q_2 = I_{d_0}$ and $Q_1^\intercal Q_2 = 0$ as well as $P_U^{(2)}$ and $Q_U^{(2)}$ are orthonormal, we then have that for $U \in \mathbb{S}^{k\times d}$,

$$\big\|U_Q^{(1)}\big\|^2 + \big\|\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal\big\|^2 = \big\|Q_1 U_Q^{(1)} + Q_2 P_U^{(2)}\big(\Sigma_U^{(2)} - [I_d, \mathbf{0}_{r-r_0}]^\intercal\big)Q_U^{(2)\intercal}\big\|^2. \tag{D.5}$$

Since $U_Q^{(1)} = Q_1^\intercal U$ and $U_Q^{(2)} = Q_2^\intercal U$ by construction and $Q_1 Q_1^\intercal U + Q_2 Q_2^\intercal U = U$ by $QQ^\intercal = I_k$, we then have that, for $U \in \mathbb{S}^{k\times d}$,

$$|Q_1 U_Q^{(1)} + Q_2 P_U^{(2)}\big(\Sigma_2 - [I_d, \mathbf{0}_{r-r_0}]^\intercal\big)Q_U^{(2)\intercal}\|^2 = \big\|U - Q_2 P_U^{(2)}[I_d, \mathbf{0}_{r-r_0}]^\intercal Q_U^{(2)\intercal}\big\|^2. \tag{D.6}$$

Noting that $Q_2 P_U^{(2)}[I_d, \mathbf{0}_{r-r_0}]^\intercal Q_U^{(2)\intercal} \in \Psi(\Pi)$, we have by (D.4)–(D.6) that, for $U \in \mathbb{S}^{k\times d}$,

$$2\big(d - \operatorname{tr}\big([I_d, \mathbf{0}_{r-r_0}]\Sigma_U^{(2)}\big)\big) \geq \min_{U' \in \Psi(\Pi)} \big\|U - U'\big\|^2. \tag{D.7}$$

Since $\Pi \neq 0$, then $\Psi(\Pi)_1^\epsilon \neq \emptyset$ for all sufficiently small $\epsilon > 0$. Fix such an $\epsilon > 0$. By the definition of $\Psi(\Pi)_1^\epsilon$, combining (D.3) and (D.7) yields that for all $U \in \Psi(\Pi)_1^\epsilon$,

$$\|\Pi U\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi) \min_{U' \in \Psi(\Pi)} \big\|U - U'\big\| \geq \frac{\sqrt{2}}{2}\sigma_{\min}^{+}(\Pi)\epsilon. \tag{D.8}$$

Then the lemma follows by applying minimum over $\Psi(\Pi)_1^\epsilon$ to both sides of (D.8) and noting that the result continues to hold for all sufficiently small $\epsilon > 0$. □

Lemma D.2. *The correspondence φ in the proof of Proposition* 3.1 *is continuous at* 0.

Proof. Fix $U_0 \in \Psi(\Pi)$, and define the correspondence $\bar{\varphi} : \mathbf{R} \twoheadrightarrow \Gamma^\Delta$ by $\bar{\varphi}(t) = \Gamma_{U_0}^\Delta(t)$, where $\Psi(\Pi)$, Γ^Δ and $\Gamma_{U_0}^\Delta(t)$ are given in the proof of Proposition 3.1. Let $d \equiv k - r$. For each t_n and each $V_0 \in \bar{\varphi}(0)$, define $f : \Gamma^\Delta \to \mathbf{M}^{k\times d}$ by

$$f(V) = V_0 - \frac{t_n}{2} U_0 V^\intercal V.$$

Since f is continuous and Γ^Δ is compact, f is a compact map in the sense of Granas and Dugundji (2003). By Theorem 0.2.3 in Granas and Dugundji (2003), one of the following two cases must happen: (i) f has a fixed point $V_{1n} \in \Gamma^\Delta$, and (ii) there exists some $V_{2n} \in \Gamma^\Delta$ such that $\|V_{2n}\| = \Delta$ and $V_{2n} = \lambda_n f(V_{2n})$ with $\lambda_n \equiv \frac{\Delta}{\|f(V_{2n})\|} \in (0,1)$. In case (i), since $U_0 \in \Psi(\Pi)$, $V_0 \in \bar{\varphi}(0)$ and $f(V_{1n}) = V_{1n}$, we have by simple algebra:

$$V_{1n}^\intercal U_0 + U_0^\intercal V_{1n} = \left(V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}\right)^\intercal U_0 + U_0^\intercal \left(V_0 - \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n}\right) = -t_n V_{1n}^\intercal V_{1n}. \tag{D.9}$$

This together with $V_{1n} \in \Gamma^\Delta$ implies that $V_{1n} \in \bar{\varphi}(t_n)$. Moreover, since $f(V_{1n}) = V_{1n}$, $\|U_0\| = \sqrt{d}$ and $V_{1n} \in \Gamma^\Delta$, then by the submultiplicativity of Frobenius norm we have

$$\|V_{1n} - V_0\| = \left\| \frac{t_n}{2} U_0 V_{1n}^\intercal V_{1n} \right\| \leq \frac{t_n}{2} \sqrt{d} \Delta^2. \tag{D.10}$$

In case (ii), since $U_0 \in \Psi(\Pi)$, $\lambda_n^2 V_0 \in \bar{\varphi}(0)$ and $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, then by analogous calculations as in (D.9), we have

$$(\lambda_n V_{2n})^\intercal U_0 + U_0^\intercal (\lambda_n V_{2n}) = -t_n (\lambda_n V_{2n})^\intercal (\lambda_n V_{2n}).$$

This together with $\lambda_n V_{2n} \in \Gamma^\Delta$ due to $\lambda_n \in (0,1)$ and $V_{2n} \in \Gamma^\Delta$ implies that $\lambda_n V_{2n} \in \bar{\varphi}(t_n)$. Moreover, since $\lambda_n V_{2n} = \lambda_n^2 f(V_{2n})$, similar to (D.10) we have

$$\|\lambda_n V_{2n} - V_0\| \leq \left\|\lambda_n^2 f(V_{2n}) - \lambda_n^2 V_0\right\| + \left|\lambda_n^2 - 1\right| \|V_0\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 + \left|\lambda_n^2 - 1\right|\Delta, \tag{D.11}$$

where the first inequality follows the triangle inequality and the second inequality follows since $\lambda_n \in (0,1)$. Now, for each $n \in \mathbf{N}$, define V_n^* to be V_{1n} if case (i) happens and $\lambda_n V_{2n}$ otherwise. Let $\delta_n \equiv 1$ if case (i) happens and $\delta_n \equiv \lambda_n$ otherwise. Then $V_n^* \in \Gamma_{U_0}^\Delta(t_n)$ for all $n \in \mathbf{N}$, and combination of (D.10) and (D.11) yields

$$\left\|V_n^* - V_0\right\| \leq \frac{t_n}{2}\sqrt{d}\Delta^2 + \left|\delta_n^2 - 1\right|\Delta \to 0,$$

where we exploited the fact that if V_{2n} exists infinitely often, $\delta_n = \lambda_n = \frac{\Delta}{\|f(V_{2n})\|} \to 1$ due to $f(V_{2n}) \to V_0$ as $n \to \infty$ and $\|V_0\| \leq \Delta$, and $t_n \to 0$ as $n \to \infty$. It follows that $\bar{\varphi}(t)$ is lower hemicontinuous at $t = 0$ by Theorem 17.21 in Aliprantis and Border (2006).

The lower hemicontinuity of $\varphi(t)$ at $t = 0$ follows easily from that of $\bar{\varphi}(t)$ again by Theorem 17.21 in Aliprantis and Border (2006). To see this, let $t_n \to 0$ and $(U_0, V_0) \in \varphi(0)$. Define (U_n^*, V_n^*) to be $U_n^* = U_0$ and V_n^* be as in previous construction for all $n \in \mathbf{N}$. Clearly, $(U_n^*, V_n^*) \to (U_0, V_0)$, implying that $\varphi(t)$ is lower hemicontinuous at $t = 0$. Since $\varphi(t)$ is contained in the compact set $\mathbb{S}^{k\times d} \times \Gamma^\Delta$ for all t, $\varphi(t)$ is upper hemicontinuous at $t = 0$ by Theorem 17.20 in Aliprantis and Border (2006). We have therefore showed that $\varphi(t)$ is continuous at $t = 0$. □

LEMMA D.3. *Suppose $\Pi \in \mathbf{M}^{m\times k}$ with* $\text{rank}(\Pi) \leq r$, *and $M \in \mathbf{M}^{m\times k}$ with $M \neq 0$. Let $\Psi(\Pi) = \arg\min_{U\in\mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2$, and for $U \in \Psi(\Pi)$ and $\Delta > 0$ let $\Gamma_U^\Delta(0)$ be as in the proof of Proposition* 3.1. *For Δ sufficiently large, it follows that for all $U \in \Psi(\Pi)$,*

$$\min_{V\in\Gamma_U^\Delta(0)} \|\Pi V + MU\|^2 = \min_{V\in\mathbf{M}^{k\times(k-r)}} \|\Pi V + MU\|^2.$$

PROOF. The conclusion is trivial if $\Pi = 0$. Suppose that $\Pi \neq 0$ and let $d \equiv k - r$. Let $r_0 = \text{rank}(\Pi)$ and $\Pi = P\Sigma Q^\intercal$ be a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$, $Q \in \mathbb{S}^{k\times k}$, and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Since $\Pi \neq 0$ and $r_0 \leq r$, we may write $\Sigma = [\Sigma_1, 0]$ with $\Sigma_1 \in \mathbf{M}^{m\times r_0}$ of full rank so that

$$\min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 = \min_{V\in\mathbf{M}^{r_0\times d}} \|[P\Sigma_1 V + MU\|^2. \tag{D.12}$$

By the projection theorem, the minimum on the right-hand side of (D.12) is attained at some point, say $V_1^* \in \mathbf{M}^{r_0\times d}$. Moreover, V_1^* is uniformly bounded over $U \in \Psi(\Pi)$. Let $V^* \equiv Q[V_1^{*\intercal}, 0]^\intercal \in \mathbf{M}^{k\times d}$, then the minimum on the left-hand side of (D.12) is attained at V^*. Decompose Q as $Q = [Q_1, Q_2]$, where $Q_1 \in \mathbf{M}^{k\times r_0}$. Then $V^* = Q_1 V_1^* \in \Gamma_U^\Delta(0)$ for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. It implies that the minimum on the right-hand side of (D.12) is attained within $\Gamma_U^\Delta(0)$ as well for all $U \in \Psi(\Pi)$, when Δ is sufficiently large. This implies that when Δ is sufficiently large,

$$\min_{V\in\Gamma_U^\Delta(0)} \|\Pi V + MU\|^2 \le \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2$$

for all $U \in \Psi(\Pi)$. The reverse inequality is simply true since $\Gamma_U^\Delta(0) \subset \mathbf{M}^{k\times d}$ all $U \in \Psi(\Pi)$ and all $\Delta > 0$. This completes the proof of the lemma. □

LEMMA D.4. *If* $r_0 \equiv \text{rank}(\Pi) \le r$, *then for any* $M \in \mathbf{M}^{m\times k}$,

$$\min_{U\in\Psi(\Pi)} \min_{V\in\mathbf{M}^{k\times d}} \|\Pi V + MU\|^2 = \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2\big(P_2^\intercal M Q_2\big), \tag{D.13}$$

where $\Psi(\Pi) = \arg\min_{U\in\mathbb{S}^{k\times(k-r)}} \|\Pi U\|^2$.

PROOF. Let $d \equiv k - r$ and $d_0 \equiv k - r_0$. Noting that the column vectors in Q_2 form a orthonormal basis for the null space of Π_0, we may rewrite $\Psi(\Pi)$ as $\Psi(\Pi) = \{Q_2 V : V \in \mathbb{S}^{d_0\times d}\}$. This, together with the projection theorem, implies

$$\phi''_{r,\Pi}(M) = \min_{V\in\mathbb{S}^{d_0\times d}} \big\|\big(I - \Pi\big(\Pi^\intercal\Pi\big)^-\Pi^\intercal\big)MQ_2V\big\|^2, \tag{D.14}$$

where A^- denotes the Moore–Penrose inverse of a generic matrix A. By the singular value decomposition of Π, we have

$$\begin{aligned}\big(I - \Pi\big(\Pi^\intercal\Pi\big)^-\Pi^\intercal\big)P &= P - P\Sigma Q^\intercal\big(Q\Sigma^\intercal P^\intercal P\Sigma Q^\intercal\big)^- Q\Sigma^\intercal P^\intercal P\\ &= P - P\Sigma Q^\intercal Q\big(\Sigma^\intercal P^\intercal P\Sigma\big)^- Q^\intercal Q\Sigma^\intercal P^\intercal P = P - P\Sigma\big(\Sigma^\intercal\Sigma\big)^-\Sigma^\intercal\\ &= [0, P_2],\end{aligned} \tag{D.15}$$

where the second equality exploited Theorem 20.5.6 in Harville (2008), the third equality follows from P and Q being orthonormal, and the fourth equality is obtained by carrying out the Moore–Penrose inverse by Exercise 2.7.4 in Magnus and Neudecker (2007) and noting that Σ is diagonal. In view of (D.15), we have

$$\begin{aligned}&\min_{V\in\mathbb{S}^{d_0\times d}} \big\|\big(I - \Pi\big(\Pi^\intercal\Pi\big)^-\Pi^\intercal\big)MQ_2V\big\|^2\\ &\quad= \min_{V\in\mathbb{S}^{d_0\times d}} \big\|[0, P_2]P^\intercal MQ_2V\big\|^2\\ &\quad= \min_{V\in\mathbb{S}^{d_0\times d}} \big\|P_2P_2^\intercal MQ_2V\big\|^2 = \min_{V\in\mathbb{S}^{d_0\times d}} \big\|P_2^\intercal MQ_2V\big\|^2\end{aligned}$$

$$= \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2 (P_2^\intercal M Q_2), \tag{D.16}$$

where the third equality follows from $P_2^\intercal P_2 = I_{m-r_0}$ and the final equality follows from Lemma 3.1. Combining (D.14) and (D.16) concludes the proof of the lemma. □

LEMMA D.5. *Suppose* $\text{rank}(\Pi) \le r$ *and let* $\phi''_{r,\Pi} : \mathbf{M}^{m\times k} \to \mathbf{R}$ *be given as in Proposition* 3.1. *If* $\text{rank}(\Pi) = r$, *there exists a bilinear map* $\Phi''_{r,\Pi} : \mathbf{M}^{m\times k} \times \mathbf{M}^{m\times k} \to \mathbf{R}$ *such that* $\phi''_{r,\Pi}(M) = \Phi''_{r,\Pi}(M, M)$ *for all* $M \in \mathbf{M}^{m\times k}$; *if* $\text{rank}(\Pi) < r$, *such a* $\Phi''_{r,\Pi}$ *does not exist.*

PROOF. Let $\Pi = P\Sigma Q^\intercal$ is a singular value decomposition of Π, where $P \in \mathbb{S}^{m\times m}$ whose last $m - r$ columns constitutes P_2, $Q \in \mathbb{S}^{k\times k}$ whose last $k - r$ columns constitutes Q_2„ and $\Sigma \in \mathbf{M}^{m\times k}$ is diagonal with diagonal entries in descending order. Let $d \equiv k - r$. If $\text{rank}(\Pi) = r$, then Lemma D.4 and Lemma 3.1 imply

$$\phi''_{r,\Pi}(M) = \min_{V \in \mathbb{S}^{d\times d}} \|P_2^\intercal M Q_2 V\|^2 = \|P_2^\intercal M Q_2\|^2,$$

for all $M \in \mathbf{M}^{m\times k}$, which is a quadratic form corresponding to the bilinear form $\Phi''_{r,\Pi}(M_1, M_2) \equiv \text{tr}(Q_2^\intercal M_1^\intercal P_2 P_2^\intercal M_2 Q_2)$ for $M_1 \in \mathbf{M}^{m\times k}$ and $M_2 \in \mathbf{M}^{m\times k}$.

Next, assume that $\text{rank}(\Pi) < r$. Suppose for the sake of a contradiction that there exists a bilinear map $\Phi''_{r,\Pi}$ corresponding to $\phi''_{r,\Pi}$. Bilinearity of $\Phi''_{r,\Pi}$ then implies that

$$\phi''_{r,\Pi}(M_1) + \phi''_{r,\Pi}(M_2) = \frac{\phi''_{r,\Pi}(M_1 + M_2) + \phi''_{r,\Pi}(M_1 - M_2)}{2} \tag{D.17}$$

for all $M_1 \in \mathbf{M}^{m\times k}$ and $M_2 \in \mathbf{M}^{m\times k}$. Let $r_0 \equiv \text{rank}(\Pi)$ and $d_0 \equiv k - r_0$. If $M = P_2 H Q_2^\intercal$ for some $H \in \mathbf{M}^{(m-r_0)\times d_0}$, then Lemma D.4 and Lemma 3.1 imply

$$\phi''_{r,\Pi}(M) = \sigma^2_{r-r_0+1}(H) + \cdots + \sigma^2_{d_0}(H). \tag{D.18}$$

Now, let $H_1 \in \mathbf{M}^{(m-r_0)\times d_0}$ be diagonal with the (j, j)th entry equal to 1 for $j = 1, \ldots, d_0$ and $H_2 \in \mathbf{M}^{(m-r_0)\times d_0}$ be diagonal with the (j, j)th entry equal to -1 for $j = 1$ and 1 for $j = 2, \ldots, d_0$. Set $M_i = P_2 H_i Q_2^\intercal$ for $i = 1, 2$, the result in (D.18) implies $\phi''_{r,\Pi}(M_1) = \phi''_{r,\Pi}(M_2) = k - r$, $\phi''_{r,\Pi}(M_1 + M_2) = 4(k - r) - 4$ and $\phi''_{r,\Pi}(M_1 - M_2) = 0$. It follows that

$$2(k - r) = \phi''_{r,\Pi}(M_1) + \phi''_{r,\Pi}(M_2) \neq \frac{\phi''_{r,\Pi}(M_1 + M_2) + \phi''_{r,\Pi}(M_1 - M_2)}{2} = 2(k - r) - 2,$$

which contradicts the result (D.17). Thus, the second result of the lemma follows. □

LEMMA D.6. *Suppose Assumption* 3.1 *holds. Let* $\hat{\phi}''_{r,n}$ *be the analytic estimator given by* (17). *If* $\hat{r}_n \xrightarrow{p} r_0 \equiv \text{rank}(\Pi_0)$ *and* $r_0 \le r < k$, *then condition* (15) *holds.*

Proof. For notational simplicity, let $d \equiv k - r$ and $\hat{d}_n \equiv k - \hat{r}_n$. Fix a sequence $\{M_n\}$ such that $M_n \to M$ as $n \to \infty$. By Lemma 3.1, we have

$$\begin{aligned}
&\big|\hat{\phi}''_{r,n}(M_n) - \hat{\phi}''_{r,n}(M)\big| \\
&\quad = \Big|\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M_n \hat{Q}_{2,n} U\big\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M \hat{Q}_{2,n} U\big\|\Big| \\
&\qquad \times \Big(\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M_n \hat{Q}_{2,n} U\big\| + \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M \hat{Q}_{2,n} U\big\|\Big), \qquad \text{(D.19)}
\end{aligned}$$

where the inequality follows by the formula $(a^2 - b^2) = (a+b)(a-b)$. For the first term on the right-hand side of (D.19), we have

$$\begin{aligned}
&\Big|\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M_n \hat{Q}_{2,n} U\big\| - \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M \hat{Q}_{2,n} U\big\|\Big| \\
&\quad \leq \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} (M_n - M) \hat{Q}_{2,n} U\big\| \leq \sqrt{kmd}\|M_n - M\| = o_p(1), \qquad \text{(D.20)}
\end{aligned}$$

where the first inequality follows by the Lipschitz continuity of the min operator and the triangle inequality, the second inequality holds by the submultiplicativity of Frobenius norm, $\|\hat{P}_{2,n}\| \leq \sqrt{m}$, $\|\hat{Q}_{2,n}\| \leq \sqrt{k}$, and $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n\times d}$, and the equality is because $M_n \to M$. For the second term on the right-hand side of (D.19), once again exploiting the submultiplicability of the Frobenius norm, $\|\hat{P}_{2,n}\| \leq \sqrt{m}$, $\|\hat{Q}_{2,n}\| \leq \sqrt{k}$, $\|U\| = \sqrt{r}$ for all $U \in \mathbb{S}^{\hat{d}_n\times d}$ and $M_n \to M$, we have that

$$\begin{aligned}
&\min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M_n \hat{Q}_{2,n} U\big\| + \min_{U\in\mathbb{S}^{\hat{d}_n\times d}} \big\|\hat{P}^\intercal_{2,n} M \hat{Q}_{2,n} U\big\| \\
&\quad \leq \sqrt{kmd}\|M_n\| + \sqrt{kmd}\|M\| = O(1). \qquad \text{(D.21)}
\end{aligned}$$

Combining results (D.19)–(D.21), then we obtain

$$\big|\hat{\phi}''_{r,n}(M_n) - \hat{\phi}''_{r,n}(M)\big| = o_p(1). \qquad \text{(D.22)}$$

In view of (D.22), it thus suffices to show that

$$\begin{aligned}
&\big|\hat{\phi}''_{r,n}(M) - \phi''_{r,\Pi_0}(M)\big| \\
&\quad \equiv \Big|\sum_{j=r-\hat{r}_n+1}^{k-\hat{r}_n} \sigma_j^2\big(\hat{P}^\intercal_{2,n} M \hat{Q}_{2,n}\big) - \sum_{j=r-r_0+1}^{k-r_0} \sigma_j^2\big(P^\intercal_{0,2} M Q_{0,2}\big)\Big| = o_p(1). \qquad \text{(D.23)}
\end{aligned}$$

Let $\hat{q}_j$ be the jth column of $\hat{Q}_{2,n}$. Since $Q_0 \in \mathbb{S}^{k\times k}$, we may write $\hat{q}_j = Q_0\hat{u}_j$ for some (random) $\hat{u}_j \in \mathbb{S}^{k\times 1}$. Noting that $\hat{q}_j$ is an eigenvector of $\hat{\Pi}^\intercal_n\hat{\Pi}_n$ associated with the eigenvalue $\sigma^2_{r_0+j}(\hat{\Pi}_n)$ when $\hat{r}_n = r_0$ and that $P(\hat{r}_n = r_0) \to 1$ as given, we have

$$\begin{aligned}
&\big[\hat{\Pi}^\intercal_n\hat{\Pi}_n - \Pi^\intercal_0\Pi_0 - \big(\sigma^2_{r_0+j}(\hat{\Pi}_n) - \sigma^2_{r_0+j}(\Pi_0)\big)I_k + \Pi^\intercal_0\Pi_0 - \sigma^2_{r_0+j}(\Pi_0)I_k\big]Q_0\hat{u}_j \\
&\quad = \big[\hat{\Pi}^\intercal_n\hat{\Pi}_n - \sigma^2_{r_0+j}(\hat{\Pi}_n)I_k\big]\hat{q}_j = o_p(1). \qquad \text{(D.24)}
\end{aligned}$$

Observe that $\|\hat{\Pi}_n^\intercal \hat{\Pi}_n - \Pi_0^\intercal \Pi_0\| = o_p(1)$ and $|\sigma^2_{r_0+j}(\hat{\Pi}_n) - \sigma^2_{r_0+j}(\Pi_0)| = o_p(1)$ by the continuous mapping theorem, the Weyl inequality (Tao (2012, Exercise 1.3.22(iv))) and Assumption 3.1, we then conclude from (D.24) that

$$o_p(1) = [\Pi_0^\intercal \Pi_0 - \sigma^2_{r_0+j}(\Pi_0) I_k] Q_0 \hat{u}_j = Q_0 \Sigma_0^\intercal \Sigma_0 \hat{u}_j, \tag{D.25}$$

where we exploited the singular value decomposition $\Pi_0 = P_0 \Sigma_0 Q_0^\intercal$, and the fact that $\sigma^2_{r_0+j}(\Pi_0) = 0$. Since the first r_0 diagonal elements of $\Sigma_0^\intercal \Sigma_0$ are positive and Q_0 is nonsingular, we may conclude from result (D.25) that the first r_0 elements of $\hat{u}_j$ are $o_p(1)$ and moreover by the definition of $\hat{q}_j$ that for some random $U_2 \in \mathbb{S}^{(k-r_0)\times(k-r_0)}$,

$$\hat{Q}_{2,n} = Q_{0,2} U_2 + o_p(1). \tag{D.26}$$

By an analogous argument, we have that for some random $V_2 \in \mathbb{S}^{(m-r_0)\times(m-r_0)}$,

$$\hat{P}_{2,n} = P_{0,2} V_2 + o_p(1). \tag{D.27}$$

Combining results (D.26) and (D.27) and the continuous mapping theorem yields

$$\|\hat{P}_{2,n}^\intercal M \hat{Q}_{2,n} - V_2^\intercal P_{0,2}^\intercal M Q_{0,2} U_2\| = o_p(1). \tag{D.28}$$

Thus, (D.23) follows from (D.28), the continuous mapping theorem and the fact that the singular values of $V_2^\intercal P_{0,2}^\intercal M Q_{0,2} U_2$ are equal to those of $P_{0,2}^\intercal M Q_{0,2}$. □

LEMMA D.7. *Suppose Assumption* 3.1 *holds. Let* $\hat{r}_n$ *be the maximal* $j \in \{1, \ldots, k\}$ *such that* $\sigma_j(\hat{\Pi}_n) \geq \kappa_n$ *if such a* j *exists and* $\hat{r}_n = 0$ *otherwise. If* $\kappa_n \downarrow 0$ *and* $\tau_n \kappa_n \to \infty$, *then it follows that*

$$\lim_{n\to\infty} P(\hat{r}_n = r_0) = 1.$$

PROOF. On the one hand, note that if $\hat{r}_n > r_0$, then we must have $r_0 \leq k-1$, $\sigma_{r_0+1}(\hat{\Pi}_n) \geq \kappa_n$ and $\sigma_{r_0+1}(\Pi_0) = 0$. In turn, it follows that

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n > r_0) &\leq \limsup_{n\to\infty} P\big(|\sigma_{r_0+1}(\hat{\Pi}_n) - \sigma_{r_0+1}(\Pi_0)| \geq \kappa_n\big) \\ &\leq \limsup_{n\to\infty} P\big(\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| \geq \tau_n \kappa_n\big) = 0,\end{aligned} \tag{D.29}$$

where the second inequality is by the Weyl inequality (Tao (2012, Exercise 1.3.22(iv))), and the equality follows from $\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| = O_p(1)$ by Assumption 3.1 and $\tau_n\kappa_n \to \infty$ as given. On the other hand, if $\hat{r}_n < r_0$, then $r_0 > 0$ and $\sigma_{r_0}(\hat{\Pi}_n) < \kappa_n$. Hence,

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{r}_n < r_0) &\leq \limsup_{n\to\infty} P\big(|\sigma_{r_0}(\hat{\Pi}_n) - \sigma_{r_0}(\Pi_0)| > -\kappa_n + \sigma_{r_0}(\Pi_0)\big) \\ &\leq \limsup_{n\to\infty} P(\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| \geq \tau_n \sigma_{r_0}(\Pi_0)\big(1 - \kappa_n/\sigma_{r_0}(\Pi_0)\big) \\ &= 0,\end{aligned} \tag{D.30}$$

where the first inequality exploited $\kappa_n < \sigma_{r_0}(\Pi_0)$ for all n sufficiently large by $\kappa_n \downarrow 0$, the second inequality again follows by the Weyl inequality (Tao (2012, Exercise 1.3.22(iv))) and also $\sigma_{r_0}(\Pi_0) > 0$, and the equality is because $\|\tau_n\{\hat{\Pi}_n - \Pi_0\}\| = O_p(1)$ by Assumption 3.1, $\tau_n \to \infty$ and $\kappa_n \downarrow 0$. Combining (D.29) and (D.30) yields

$$\limsup_{n\to\infty} P(\hat{r}_n \neq r_0) \leq \limsup_{n\to\infty} P(\hat{r}_n < r_0) + \limsup_{n\to\infty} P(\hat{r}_n > r_0) = 0.$$

This completes the proof of the lemma. □

LEMMA D.8. *Let $\mathbb{G} \in \mathbf{R}^k$ follow $N(\mu, \Omega_0)$ and $g : \mathbf{R}^k \to \mathbf{R}$ be a Lipschitz map with Lipschitz constant L. Then, for M the median of $g(\mathbb{G})$ and any $x > 0$*

$$P\big(g(\mathbb{G}) - M > x\big) \leq \frac{1}{2}\exp\left\{-\frac{1}{2}\frac{x^2}{C^2}\right\} \tag{D.31}$$

for some $C > 0$ depending on L and $\|\Omega_0\|$.

PROOF. This is a mild extension of Lemma A.2.2 in van der Vaart and Wellner (1996), and we include a proof here only for completeness. Since $\mathbb{G} \sim N(\mu, \Omega_0)$, we may write $\mathbb{G} \overset{d}{=} \Omega_0^{1/2} Z + \mu$ for some $Z \sim N(0, I_k)$. Define a map $h : \mathbf{R}^k \to \mathbf{R}$ by $h(z) = g(\Omega_0^{1/2} z + \mu)$ for any $z \in \mathbf{R}^k$. Then by Lipschitz continuity of g we have: for any $z_1, z_2 \in \mathbf{R}^k$,

$$\begin{aligned}\big|h(z_1) - h(z_2)\big| &= \big|g\big(\Omega_0^{1/2} z_1 + \mu\big) - g\big(\Omega_0^{1/2} z_2 + \mu\big)\big| \leq L\big\|\Omega_0^{1/2} z_1 - \Omega_0^{1/2} z_2\big\| \\ &\leq L\big\|\Omega_0^{1/2}\big\|\|z_1 - z_2\| \leq L\|\Omega_0\|^{1/2}\|z_1 - z_2\|,\end{aligned} \tag{D.32}$$

where the fact $\|\Omega_0^{1/2}\| \leq \|\Omega_0\|^{1/2}$ follows from Theorem X.1.1 in Bhatia (1997). By replacing $L\|\Omega_0\|^{1/2}$ with $(L\|\Omega_0\|^{1/2}) \vee 1$ if necessary, we may assume $C \equiv L\|\Omega_0\|^{1/2} > 0$ without loss of generality. Since M is the median of $g(\mathbb{G})$ and hence also of $h(Z)$, we conclude that M/C is the median of $h(Z)/C$. It follows from Lemma A.2.2 in van der Vaart and Wellner (1996) that: for any $x > 0$,

$$P\big(g(\mathbb{G}) - M > x\big) = P\left(\frac{h(Z)}{C} - \frac{M}{C} > \frac{x}{C}\right) \leq \frac{1}{2}\exp\left\{-\frac{1}{2}\frac{x^2}{C^2}\right\}. \tag{D.33}$$

This completes the proof of the lemma. □

For the next two lemmas, we let $\mathrm{BL}_1(\mathbf{R})$ be the set of real-valued Lipschitz functions on $\mathbf{R}$ with levels and Lipschitz constants both bounded by one.

LEMMA D.9. *Let $T_n^* : \{X_i, W_{ni}\}_{i=1}^n \to \mathbf{R}$ be a bootstrap estimator for the distribution of $g(\mathbb{G})$ such that $\mathbb{G} \in \mathbf{R}^k$ is Gaussian, $g : \mathbf{R}^k \to \mathbf{R}$ is a Lipschitz map, and*

$$\sup_{f \in \mathrm{BL}_1(\mathbf{R})} \big|E_W\big[f\big(T_n^*\big)\big] - E\big[f\big(g(\mathbb{G})\big)\big]\big| = o_p(1). \tag{D.34}$$

Suppose Assumption C.1 holds. Let $\hat{c}_{n,1-\alpha_n}$ be $(1-\alpha_n)$ conditional quantiles of T_n^* given the data. If the cdf of $g(\mathbb{G})$ is continuous and strictly increasing on $[r_0, \infty)$ for some $r_0 \in \mathbf{R}$, then $\hat{c}_{n,1-\alpha_n}/\tau_n \overset{p}{\to} 0$.

PROOF. Let $\hat{F}_n$ be the conditional cdf of T_n^* given $\{X_i\}_{i=1}^n$, and F be the cdf of $g(\mathbb{G})$. By Lemma 10.11 in Kosorok (2008), we have

$$\sup_{t\in[r_0,\infty)} \left|\hat{F}_n(t) - F(t)\right| = o_p(1). \tag{D.35}$$

By the definition of quantiles, we thus obtain from (D.35) that, for any $r \in [r_0, \infty)$,

$$\begin{aligned}\limsup_{n\to\infty} P(\hat{c}_{n,1-\alpha_n} \le r) &\le \limsup_{n\to\infty} P\big(\hat{F}_n(r) \ge 1-\alpha_n\big)\\ &= \limsup_{n\to\infty} P\big(o_p(1) + F(r) \ge 1-\alpha_n\big) = 0,\end{aligned} \tag{D.36}$$

where we exploited the facts that $F(r) < 1$ by strict monotonicity of F on $[r_0, \infty)$ and that $\alpha_n \downarrow 0$. Next, fix $\epsilon > 0$. Combination of (D.36) and Lemma D.8 yields

$$\begin{aligned}\alpha_n < 1 - \hat{F}_n(\hat{c}_{n,1-\alpha_n} - \epsilon) &= P\big(g(\mathbb{G}) > \hat{c}_{n,1-\alpha_n} - \epsilon\big) + o_p(1)\\ &\le \frac{1}{2}\exp\left\{-\frac{1}{2}\frac{(\hat{c}_{n,1-\alpha_n} - \epsilon - c_{0.5})^2}{C^2}\right\} + o_p(1),\end{aligned} \tag{D.37}$$

for some $C > 0$ and $c_{0.5}$ the 0.5-quantile of $g(\mathbb{G})$. It follows from (D.37) that

$$\left(\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n} - \frac{\epsilon}{\tau_n} - \frac{c_{0.5}}{\tau_n}\right)^2 \le 2C^2\left(-\frac{\log\alpha_n}{\tau_n^2} + \frac{\log o_p(1)}{\tau_n^2} - \frac{\log 2}{\tau_n^2}\right). \tag{D.38}$$

By Assumption C.1(ii), $\tau_n \uparrow \infty$ and $\log o_p(1) \xrightarrow{p} -\infty$ as $n \to \infty$, we may then conclude the proof of the lemma from result (D.38). □

LEMMA D.10. *Suppose Assumptions* 3.1, 3.2, *and* C.1 *hold. Let* $\hat{c}_{n,1-\alpha}$ *be defined by* (18) *for* $\alpha \in (0,1)$ *where* $\kappa_n \to 0$ *and* $\tau_n\kappa_n \to \infty$ *if* $\hat{\phi}''_{r,n}$ *is defined by* (16) *but no restrictions on* $\hat{r}_n$ *if* $\hat{\phi}''_{r,n}$ *is defined by* (17). *If* $\mathcal{M}$ *is Gaussian but not constant, then* $\hat{c}_{n,1-\alpha_n}/\tau_n^2 \xrightarrow{p} 0$.

PROOF. Consider first the case when $\hat{c}_{n,1-\alpha}$ is defined by the analytic derivative estimator. By Lemma 3.1 and simple manipulations, we have

$$\hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)^{1/2} \le \|\hat{P}_{2,n}^{\intercal}\hat{\mathcal{M}}_n^*\hat{Q}_{2,n}\| \le (mk)^{1/2}\|\hat{\mathcal{M}}_n^*\|. \tag{D.39}$$

Let $\tilde{c}_{n,1-\alpha}$ be the $(1-\alpha)$th conditional quantile of $\|\hat{\mathcal{M}}_n^*\|$ for each $\alpha \in (0,1)$. Since $\mathcal{M}$ is Gaussian and the variance of $\mathrm{vec}(\mathcal{M})$ is nonzero, $\|\mathcal{M}\|^2$ is equal in law to a weighted sum of independent $\chi^2(1)$ random variables. It follows that the cdf $\|\mathcal{M}\|$ is continuous and strictly increasing on $\mathbf{R}_+$. In turn, by Proposition 10.7 in Kosorok (2008), Assumptions 3.2 and C.1, we obtain from Lemma D.9 that $\tilde{c}_{n,1-\alpha}/\tau_n \xrightarrow{p} 0$. By result (D.39) and equivariance of quantiles to monotone transformations, we may then conclude that

$$\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2} \le \frac{\hat{c}^2_{n,1-\alpha_n}}{\tau_n^2} = o_p(1). \tag{D.40}$$

Next, turn to the case when $\hat{c}_{n,1-\alpha}$ is defined by the numerical derivative estimator. For each $\alpha \in (0,1)$, let $\bar{c}_{n,1-\alpha}$ be the conditional quantile (given the data) of

$$\kappa_n \hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*) = \frac{\phi_r(\hat{\Pi}_n + \kappa_n \hat{\mathcal{M}}_n^*) - \phi_r(\hat{\Pi}_n)}{\kappa_n}. \tag{D.41}$$

By Assumptions 3.1, 3.2, and the rates conditions on κ_n as given, we may employ Proposition 3.1 and Theorem 3.3 in Chen and Fang (2019) to conclude that

$$\sup_{f\in \mathrm{BL}_1(\mathbf{R})} \left|E_W\left[f\left(\kappa_n \hat{\phi}''_{r,n}(\hat{\mathcal{M}}_n^*)\right)\right] - E\left[f\left(\phi'_{r,\Pi_0}(\mathcal{M})\right)\right]\right| = o_p(1). \tag{D.42}$$

By simple algebra, we may obtain that: for any $M_1, M_2 \in \mathbf{M}^{m\times k}$,

$$\begin{aligned} \left|\phi'_{r,\Pi_0}(M_1) - \phi'_{r,\Pi_0}(M_2)\right| &= \left|\min_{U\in\Psi(\Pi)} 2\operatorname{tr}\left(U^\intercal \Pi^\intercal M_1 U\right) - \min_{U\in\Psi(\Pi)} 2\operatorname{tr}\left(U^\intercal \Pi^\intercal M_2 U\right)\right| \\ &\le \max_{U\in\Psi(\Pi_0)} 2\|\Pi_0 U\| \left\|(M_1 - M_2)U\right\| \le 2\sqrt{k}\|\Pi_0\|\|M_1 - M_2\|. \end{aligned} \tag{D.43}$$

By result (D.36) and Lemma D.9, we thus have $\bar{c}_{n,1-\alpha_n}/\tau_n = o_p(1)$, and hence

$$\frac{\hat{c}_{n,1-\alpha_n}}{\tau_n^2} \le \frac{\bar{c}_{n,1-\alpha_n}}{\tau_n} \frac{1}{\tau_n \kappa_n} = o_p(1), \tag{D.44}$$

since $\tau_n \kappa_n \to \infty$. This completes the proof of the lemma. □

We next present lemmas that are relevant to Section 5 and proceed by imposing the following.

Assumption D.1. (i) *The supports of X and Y are finite*; (ii) *the Jacobian matrix of* $\operatorname{vec}(E_{\pi(A,p,q)}[XY^\intercal])$ *with respect to* $\operatorname{vec}(A)$ *at* A_0 *is nonsingular.*

Assumption D.1(i) formalizes the setup that the matching attributes are finitely valued. Assumption D.1(ii) is a technical condition, as implicitly imposed in Galichon and Salanié (2010) and Dupuy and Galichon (2014) who showed that the Jacobian coincides with the Fisher information matrix for A_0.

Next, let the supports $\mathcal{X} = \{x_1, \dots, x_I\}$ and $\mathcal{Y} = \{y_1, \dots, y_J\}$. Then we may identify p_0 and q_0 as vectors in $(0,1)^I$ and $(0,1)^J$, respectively.

Lemma D.11. *If Assumption D.1 holds, then the implicit map* $A : (0,1)^I \times (0,1)^J \times \mathbf{M}^{m\times k} \to \mathbf{M}^{m\times k}$ *defined by* (26), *that is,* $A(p_0, q_0, E[XY^\intercal]) = A_0$, *is Hadamard differentiable on some open neighborhood of the truth* $(p_0, q_0, E[XY^\intercal])$.

Proof. First, note that A is uniquely defined by Lemma 3 in Dupuy and Galichon (2014). Next, define a map $\Psi : \mathbf{M}^{m\times k} \times (0,1)^I \times (0,1)^J \times \mathbf{M}^{m\times k} \to \mathbf{R}^{mk}$ by

$$\Psi(A, p, q, \Sigma) \equiv \operatorname{vec}\left(E_{\pi(A,p,q)}\left[X^\intercal Y\right] - \Sigma\right). \tag{D.45}$$

By Assumption D.1 and Lemma D.12, Ψ is continuously differentiable on some open neighborhood of the truth $(A_0, p_0, q_0, E[XY^\intercal])$—note in particular that X and Y are finitely supported. In turn, Assumption D.1(ii) allows us to invoke the implicit function theorem; see, for example, Theorem 9.28 in Rudin (1976), to conclude the proof. □

Lemma D.12. *If Assumption* D.1(i) *holds, then the map* $(A_0, p_0, q_0) \mapsto \pi(A_0, p_0, q_0)(x, y)$ *defined by* (24) *where* Φ *is specified as in* (25) *uniquely exists and is continuously differentiable on some open neighborhood of the truth* (A_0, p_0, q_0), *for each* $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Proof. First, we may rewrite the maximization problem (24) as

$$\max_{\pi} \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_{ij} x_i^\intercal A_0 y_j - \sum_{i=1}^{I}\sum_{j=1}^{J} \pi_{ij} \log \pi_{ij}, \tag{D.46}$$

subject to: for all $i = 1, \dots, I$ and all $j = 1, \dots, J$,

$$\sum_{j=1}^{J} \pi_{ij} = p_{0,i}, \qquad \sum_{i=1}^{I} \pi_{ij} = q_{0,j}, \tag{D.47}$$

where $p_{0,i} = P(X = x_i)$ and $q_{0,j} = P(Y = y_j)$ for all $i = 1, \dots, I$ and all $j = 1, \dots, J$. By defining $x \log x = 0$ if $x = 0$, it is simple to see that the objective function in (D.46) is continuous. Since the constraints define a compact domain for π, it follows that an optimal matching distribution π_0 always exists. The uniqueness of π_0 follows from strict concavity of the objective function since $x \mapsto x \log x$ is strictly convex. Moreover, the right derivative of the objective function at 0 is infinite (see equation (D.50) below or Galichon and Salanié (2010, p. 5)) implying that the optimal π_0 must satisfy $0 < \pi_{0,ij} < 1$ for all i and j. Exploiting the constraints in (D.47), together with the facts that p_0, q_0 and π are pmfs, the constrained optimization can be converted into an unconstrained one in which the objective function in (D.46) is a function of $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$ only, with $\pi_{0,iJ} = p_{0,i} - \sum_{j=1}^{J-1} \pi_{0,ij}$, $\pi_{0,Ij} = q_{0,j} - \sum_{i=1}^{I-1} \pi_{0,ij}$ for all $i = 1, \dots, I-1$ and $j = 1, \dots, J-1$, and

$$\begin{aligned}\pi_{0,IJ} &= 1 - \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} \pi_{0,ij} - \sum_{i=1}^{I-1} \pi_{0,iJ} - \sum_{j=1}^{J-1} \pi_{0,Ij} \\ &= 1 + \sum_{i=1}^{I-1}\sum_{j=1}^{J-1} \pi_{0,ij} - \sum_{i=1}^{I-1} p_{0,i} - \sum_{j=1}^{J-1} q_{0,j}.\end{aligned} \tag{D.48}$$

It follows that the unique maximizer π_0 must satisfy the first-order condition:

$$\begin{aligned}&x_i^\intercal A_0 y_j - x_I^\intercal A_0 y_j - x_i^\intercal A_0 y_J + x_I^\intercal A_0 y_J - 1 - \log \pi_{0,ij} \\ &\quad + 1 + \log \pi_{0,Ij} + 1 + \log \pi_{0,iJ} - 1 - \log \pi_{0,IJ} = 0,\end{aligned} \tag{D.49}$$

or equivalently

$$\begin{aligned} &x_i^\intercal A_0 y_j - x_I^\intercal A_0 y_j - x_i^\intercal A_0 y_J + x_I^\intercal A_0 y_J \\ &\quad - \log \pi_{0,ij} + \log \pi_{0,Ij} + \log \pi_{0,iJ} - \log \pi_{0,IJ} = 0, \end{aligned} \tag{D.50}$$

for all $i = 1, \dots, I-1$ and $j = 1, \dots, J-1$, where $\pi_{0,iJ}$, $\pi_{0,Ij}$ and $\pi_{0,IJ}$ are functions of $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$ as defined previously.

Let us stack the equations in (D.50) along $i = 1, \dots, m$ sequentially for fixed $j = 1, \dots, k$, and let $d^* \equiv (I-1)(J-1)$. The left side of (D.50) is then a $\mathbf{R}^{d^*}$-valued function of A_0, $\{p_{0,i}\}_{i=1}^{I-1}$, $\{q_{0,j}\}_{j=1}^{J-1}$ and $\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1}$, which is obviously continuously differentiable. Moreover, the derivative of the left side in (D.50) with respect to $\mathrm{vec}(\{\pi_{0,ij}\}_{i=1,j=1}^{I-1,J-1})$ is then a matrix of size $d^* \times d^*$ which is given by: for $\mathbb{J}_d$ a generic $d \times d$ matrix of ones,

$$-\underline{\pi}_0 - \underline{\pi}_{0,J} \otimes \mathbb{J}_{I-1} - \mathbb{J}_{J-1} \otimes \underline{\pi}_{0,I} - \pi_{0,IJ}^{-1} \mathbb{J}_{d^{*2}}, \tag{D.51}$$

with $\underline{\pi}_{0,I} \equiv \mathrm{diag}(\{\pi_{0,iJ}^{-1}\}_{i=1}^{I-1})$, $\underline{\pi}_{0,J} \equiv \mathrm{diag}(\{\pi_{0,Ij}^{-1}\}_{j=1}^{J-1})$ and $\underline{\pi}_0 \equiv \mathrm{diag}(\mathrm{vec}(\{\pi_{0,ij}^{-1}\}_{i=1,j=1}^{I-1,J-1}))$. Note that $\underline{\pi}_0$ is positive definite while $\underline{\pi}_{0,J} \otimes \mathbb{J}_{I-1}$, $\mathbb{J}_{J-1} \otimes \underline{\pi}_{0,I}$ and $\pi_{0,IJ}^{-1} \otimes \mathbb{J}_{d^{*2}}$ are positive semidefinite, so the matrix in (D.51) is invertible. The conclusion now follows from the implicit function theorem; see, for example, Theorem 9.28 in Rudin (1976). □

References

Aliprantis, C. D. and K. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition. Springer Verlag, Berlin. [5, 18]

Andrews, D. W. (1999), "Consistent moment selection procedures for generalized method of moments estimation." *Econometrica*, 67, 543–563. [13]

Bauer, P., B. M. Pötscher, and P. Hackl (1988), "Model selection by multiple test procedures." *Statistics*, 19, 39–44. [12]

Bhatia, R. (1997), *Matrix Analysis*. Springer-Verlag, New York, NY. [23]

Chen, Q. and Z. Fang (2019), "Inference on functionals under first order degeneracy." *Journal of Econometrics*, 210, 459–481. [5, 25]

Cragg, J. G. and S. G. Donald (1997), "Inferring the rank of a matrix." *Journal of Econometrics*, 76, 223–250. [12, 13]

Dupuy, A. and A. Galichon (2014), "Personality traits and the marriage market." *Journal of Political Economy*, 122, 1271–1319. [25]

Efron, B. (1979), "Bootstrap methods: Another look at the jackknife." *The Annals of Statistics*, 7, 1–26. [2]

Fang, Z. and A. Santos (2018), "Inference on directionally differentiable functions." *The Review of Economic Studies*, 86, 377–412. [5]

Galichon, A. and B. Salanié (2010), "Matching with trade-offs: Revealed preferences over competing characteristics." Available at SSRN, https://ssrn.com/abstract=1640380. [25, 26]

Granas, A. and J. Dugundji (2003), *Fixed Point Theory*. Springer, New York, NY. [17]

Harville, D. A. (2008), *Matrix Algebra From a Statistician's Perspective*. Springer, New York, NY. [19]

Hosoya, Y. (1989), "Hierarchical statistical models and a generalized likelihood ratio test." *Journal of the Royal Statistical Society. Series B. Methodological*, 51, 435–447. [12]

Jagannathan, R. and Z. Wang (1996), "The conditional CAPM and the cross-section of expected returns." *The Journal of Finance*, 51, 3–53. [9]

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York, NY. [12]

Kleibergen, F. and R. Paap (2006), "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133, 97–126. [2, 7, 8, 9, 11, 13]

Kosorok, M. (2008), *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, NY. [15, 24]

Lahiri, S. N. (2003), *Resampling Methods for Dependent Data*. Springer, New York, NY. [11]

Ledoux, M. (2001), *The Concentration of Measure Phenomenon*. American Mathematical Society, Providence. [14]

Magnus, J. R. and H. Neudecker (2007), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, third edition. John Wiley & Sons, England. [19]

Pötscher, B. M. (1983), "Order estimation in ARMA-models by Lagrangian multiplier tests." *The Annals of Statistics*, 11, 872–885. [12, 13]

Robin, J.-M. and R. J. Smith (2000), "Tests of rank." *Econometric Theory*, 16, 151–175. [7, 12, 13]

Rudin, W. (1976), *Principles of Mathematical Analysis*, third edition. McGraw-Hill, New York, NY. [26, 27]

Shapiro, A. (1990), "On concepts of directional differentiability." *Journal of Optimization Theory and Applications*, 66, 477–487. [3]

Shapiro, A. (1991), "Asymptotic analysis of stochastic programs." *Annals of Operations Research*, 30, 169–186. [3]

Tao, T. (2012), *Topics in Random Matrix Theory*. American Mathematical Society, Providence. [2, 22, 23]

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*. Springer Verlag, New York, NY. [23]

West, K. D. (1997), “Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator.” *Journal of Econometrics*, 76, 171–191. [11]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 22 May, 2018; final version accepted 9 May, 2019; available online 24 May, 2019.