Supplemental Material for ASYMPTOTIC SIZE OF KLEIBERGEN'S LM AND CONDITIONAL LR TESTS FOR MOMENT CONDITION MODELS

By

Donald W. K. Andrews and Patrik Guggenberger

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1977S

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Supplemental Material

for

Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Patrik Guggenberger
Department of Economics
Pennsylvania State University

First Version: March 25, 2011
Revised: December 19, 2014

Contents

13 Outline 2

14 Proof of Lemma 8.2 3

15 Proof of Lemma 8.3 4

16 Proof of Theorem 8.4 11

17 Proofs of Sufficiency of Several Conditions for the $\lambda_{p-j}(\cdot)$ Condition in $\mathcal{F}_{0j}$ 26

18 Asymptotic Size of Kleibergen's CLR Test with Jacobian-Variance Weighting and the Proof of Theorem 5.1 29

19 Proof of Theorem 7.1 52

13 Outline

We let AG1 abbreviate the main paper "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models" and its Appendix. References to Sections with Section numbers less than 13 refer to Sections of AG1. Similarly, all theorems and lemmas with Section numbers less than 13 refer to results in AG1.

This Supplemental Material provides proofs of some of the results stated in AG1. It also provides some complementary results to those in AG1.

Sections 14, 15, and 16 prove Lemma 8.2, Lemma 8.3, and Theorem 8.4, respectively, which appear in Section 8 in the Appendix to AG1. Section 17 proves that the conditions in (3.9) and (3.10) are sufficient for the second condition in $\mathcal{F}_{0j}$.

Section 18 proves Theorem 5.1. Section 18 also determines the asymptotic size of Kleibergen's (2005) CLR test with Jacobian-variance weighting that employs the Robin and Smith (2000) rank statistic, defined in Section 5, for the general case of $p \geq 1$. When $p = 1$, the asymptotic size of this test is correct. But, when $p \geq 2$, we cannot show that its asymptotic size is necessarily correct (because the sample moments and the rank statistic can be asymptotically dependent under some sequences of distributions). Section 18 provides some simulation results for this test.

Section 19 proves Theorem 7.1, which provides results for time series observations.

For notational simplicity, throughout the Supplemental Material, we often suppress the argument θ_0 for various quantities that depend on the null value θ_0. Throughout the Supplemental Material, the quantities B_F, C_F, and $(\tau_{1F}, ..., \tau_{pF})$ are defined using the general definitions given in (8.6)-(8.8), rather than the definitions given in Section 3, which are a special case of the former definitions.

For notational simplicity, the proofs in Sections 14-16 are for the sequence $\{n\}$, rather than a subsequence $\{w_n : n \geq 1\}$. The same proofs hold for any subsequence $\{w_n : n \geq 1\}$. The proofs in these three sections use the following simplified notation. Define

$$\begin{aligned}
&D_n := E_{F_n} G_i,\ \Omega_n := \Omega_{F_n},\ B_n := B_{F_n},\ C_n := C_{F_n},\ B_n = (B_{n,q}, B_{n,p-q}),\ C_n = (C_{n,q}, C_{n,k-q}),\\
&W_n := W_{F_n},\ W_{2n} := W_{2F_n},\ U_n := U_{F_n},\ \text{and } U_{2n} := U_{2F_n}, \qquad (13.1)
\end{aligned}$$

where $q = q_h$ is defined in (8.16), $B_{n,q} \in R^{p \times q}$, $B_{n,p-q} \in R^{p \times (p-q)}$, $C_{n,q} \in R^{k \times q}$, and $C_{n,k-q} \in$

$R^{k\times(k-q)}$. Define

$$\Upsilon_{n,q} := Diag\{\tau_{1F_n},...,\tau_{qF_n}\} \in R^{q\times q},\ \Upsilon_{n,p-q} := Diag\{\tau_{(q+1)F_n},...,\tau_{pF_n}\} \in R^{(p-q)\times(p-q)},\ \text{and}$$

$$\Upsilon_n := \left[\begin{array}{cc} \Upsilon_{n,q} & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & \Upsilon_{n,p-q} \\ 0^{(k-p)\times q} & 0^{(k-p)\times(p-q)} \end{array}\right] \in R^{k\times p}. \tag{13.2}$$

Note that Υ_n is the diagonal matrix of singular values of $W_nD_nU_n$, see (8.8).

14 Proof of Lemma 8.2

Lemma 8.2 of AG1. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$,

$$n^{1/2}\left(\begin{array}{c} \widehat{g}_n \\ vec(\widehat{D}_n - E_{F_n}G_i) \end{array}\right) \rightarrow_d \left(\begin{array}{c} \overline{g}_h \\ vec(\overline{D}_h) \end{array}\right) \sim N\left(0^{(p+1)k}, \left(\begin{array}{cc} h_{5,g} & 0^{k\times pk} \\ 0^{pk\times k} & \Phi_h^{vec(G_i)} \end{array}\right)\right).$$

Under all subsequences $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$, *the same result holds with* n *replaced with* w_n.

Proof of Lemma 8.2. We have

$$\begin{aligned} n^{1/2}vec(\widehat{D}_n - D_n) &= n^{-1/2}\sum_{i=1}^n vec(G_i - D_n) - \left(\begin{array}{c} \widehat{\Gamma}_{1n} \\ \vdots \\ \widehat{\Gamma}_{pn} \end{array}\right)\widehat{\Omega}_n^{-1}n^{1/2}\widehat{g}_n \\ &= n^{-1/2}\sum_{i=1}^n \left[vec(G_i - D_n) - \left(\begin{array}{c} E_{F_n}G_{\ell 1}g_\ell' \\ \vdots \\ E_{F_n}G_{\ell p}g_\ell' \end{array}\right)\Omega_{F_n}^{-1}g_i\right] + o_p(1), \end{aligned} \tag{14.1}$$

where the second equality holds by (i) the weak law of large numbers (WLLN) applied to $n^{-1}\sum_{\ell=1}^n G_{\ell j}g_\ell'$ for $j = 1,...,p$, $n^{-1}\sum_{\ell=1}^n vec(G_\ell)$, and $n^{-1}\sum_{\ell=1}^n g_\ell g_\ell'$, (ii) $E_{F_n}g_i = 0^k$, (iii) $h_{5,g} = \lim \Omega_{F_n}$ is pd, and (iv) the CLT, which implies that $n^{1/2}\widehat{g}_n = O_p(1)$.

Using (14.1), the convergence result of Lemma 8.2 holds (with n in place of w_n) by the Lyapunov triangular-array multivariate CLT using the moment restrictions in $\mathcal{F}$. The limiting covariance matrix between $n^{1/2}vec(\widehat{D}_n - D_n)$ and $n^{1/2}\widehat{g}_n$ in Lemma 8.2 is a zero matrix because

$$E_{F_n}[G_{ij} - D_{nj} - (E_{F_n}G_{\ell j}g_\ell')\Omega_{F_n}^{-1}g_i]g_i' = 0^{k\times k}, \tag{14.2}$$

where D_{nj} denotes the jth column of D_n, using $E_{F_n} g_i = 0^k$ for $j = 1, ..., p$. By the CLT, the limiting variance matrix of $n^{1/2} vec(\widehat{D}_n - D_n)$ in Lemma 8.2 equals

$$\lim Var_{F_n}(vec(G_i) - (E_{F_n} vec(G_\ell) g_\ell')\Omega_{F_n}^{-1} g_i) = \lim \Phi_{F_n}^{vec(G_i)} = \Phi_h^{vec(G_i)}, \tag{14.3}$$

see (8.15), and the limit exists because (i) the components of $\Phi_{F_n}^{vec(G_i)}$ are comprised of λ_{4,F_n} and submatrices of λ_{5,F_n} and (ii) $\lambda_{s,F_n} \to h_s$ for $s = 4, 5$. By the CLT, the limiting variance matrix of $n^{1/2}\widehat{g}_n$ equals $\lim E_{F_n} g_i g_i' = h_{5,g}$. $\square$

15 Proof of Lemma 8.3

Lemma 8.3 of AG1. *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

$$n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n} G_i, W_{F_n} \widehat{D}_n U_{F_n} T_n) \to_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h),$$

where (a) $(\overline{g}_h, \overline{D}_h)$ *are defined in Lemma* 8.2, (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (8.17), (c) $(\overline{D}_h, \overline{\Delta}_h)$ *and* $\overline{g}_h$ *are independent,* (d) *if Assumption* WU *holds with* $\Lambda_* = \Lambda_0$, $W_F = \Omega_F^{-1/2}$, *and* $U_F = I_p$, *then* $\overline{\Delta}_h$ *has full column rank* p *with probability one and* (e) *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$, *the convergence result above and the results of parts* (a)-(d) *hold with* n *replaced with* w_n.

The proof of part (d) of Lemma 8.3 uses the following two lemmas and corollary.

Lemma 15.1 *Suppose* $\Delta \in R^{k \times p}$ *has a multivariate normal distribution* (*with possibly singular variance matrix*), $k \geq p$, *and the variance matrix of* $\Delta\xi \in R^k$ *has rank at least* p *for all nonrandom vectors* $\xi \in R^p$ *with* $||\xi|| = 1$. *Then,* $P(\Delta$ *has full column rank* $p) = 1$.

Comments: (i) Let Condition Δ denote the condition of the lemma on the variance of $\Delta\xi$. A sufficient condition for Condition Δ is that $vec(\Delta)$ has a pd variance matrix (because $\Delta\xi = (\xi' \otimes I_k) vec(\Delta)$). The converse is not true. This is proved in Comment (iii) below.

(ii) A weaker sufficient condition for Condition Δ is that the variance matrix of $\Delta\xi \in R^k$ has rank k for all constant vectors $\xi \in R^p$ with $||\xi|| = 1$. The latter condition holds iff $Var(\zeta' vec(\Delta)) > 0$ for all $\zeta \in R^{pk}$ of the form $\zeta = \xi \otimes \mu$ for some $\xi \in R^p$ and $\mu \in R^k$ with $||\xi|| = 1$ and $||\mu|| = 1$ (because $(\xi' \otimes \mu') vec(\Delta) = vec(\mu'\Delta\xi) = \mu'\Delta\xi$). In contrast, $vec(\Delta)$ has a pd variance matrix iff $Var(\zeta' vec(\Delta)) > 0$ for all $\zeta \in R^{pk}$ with $||\zeta|| = 1$.

(iii) For example, the following matrix Δ satisfies the sufficient condition given in Comment (ii) for Condition Δ (and hence Condition Δ holds), but not the sufficient condition given in Comment (i). Let Z_j for $j = 1, 2, 3$ be independent standard normal random variables. Define

$$\Delta = \begin{pmatrix} Z_1 & Z_2 \\ Z_3 & Z_1 \end{pmatrix}. \tag{15.1}$$

Obviously, $Var(vec(\Delta))$ is not pd. On the other hand, writing $\xi = (\xi_1, \xi_2)'$ and $\mu = (\mu_1, \mu_2)'$, we have

$$\begin{aligned} Var(\mu'\Delta\xi) &= Var(\mu_1[Z_1\xi_1 + Z_2\xi_2] + \mu_2[Z_3\xi_1 + Z_1\xi_2]) \\ &= Var((\mu_1\xi_1 + \mu_2\xi_2)Z_1 + \mu_1\xi_2 Z_2 + \mu_2\xi_1 Z_3) \\ &= (\mu_1\xi_1 + \mu_2\xi_2)^2 + (\mu_1\xi_2)^2 + (\mu_2\xi_1)^2. \end{aligned} \tag{15.2}$$

Now, $(\mu_1\xi_2)^2 = 0$ implies $\mu_1 = 0$ or $\xi_2 = 0$ and $(\mu_2\xi_1)^2 = 0$ implies $\mu_2 = 0$ or $\xi_1 = 0$. In addition, $\mu_1 = 0$ implies $\mu_2 \neq 0$, $\xi_2 = 0$ implies $\xi_1 \neq 0$, etc. So, the two cases where $(\mu_1\xi_2)^2 = (\mu_2\xi_1)^2 = 0$ are: $(\mu_1, \xi_1) = (0, 0)$ and $(\mu_2, \xi_2) = (0, 0)$. But, $(\mu_1, \xi_1) = (0, 0)$ implies $(\mu_1\xi_1 + \mu_2\xi_2)^2 = (\mu_2\xi_2)^2 > 0$ and $(\mu_2, \xi_2) = (0, 0)$ implies $(\mu_1\xi_1 + \mu_2\xi_2)^2 = (\mu_1\xi_1)^2 > 0$. Hence, $Var(\mu'\Delta\xi) > 0$ for all μ and ξ with $||\mu|| = ||\xi|| = 1$, $Var(\Delta\xi)$ is pd for all $\xi \in R^2$ with $||\xi||^2 = 1$, and the sufficient condition given in Comment (ii) for Condition Δ holds.

(iv) Condition Δ allows for redundant rows in Δ, which corresponds to redundant moment conditions in the application of Lemma 15.1. Suppose a matrix Δ satisfies Condition Δ. Then, one adds one or more rows to Δ, which consist of one or more of the existing rows of Δ or some linear combinations of them. (In fact, the added rows can be arbitrary provided the resulting matrix has a multivariate normal distribution.) Call the new matrix Δ_+. The matrix Δ_+ also satisfies Condition Δ (because the rank of the variance of $\Delta_+\xi$ is at least as large as the rank of the variance of $\Delta\xi$, which is p).

Corollary 15.2 *Suppose $\Delta_{q_*} \in R^{k \times q_*}$ is a nonrandom matrix with full column rank q_* and $\Delta_{p-q_*} \in R^{k \times (p-q_*)}$ has a multivariate normal distribution (with possibly singular variance matrix) and $k \geq p$. Let $M \in R^{k \times k}$ be a nonsingular matrix such that $M\Delta_{q_*} = (e_1, ..., e_{q_*})$, where e_l denotes the l-th coordinate vector in R^k. Decompose $M = (M_1', M_2')'$ with $M_1 \in R^{q_* \times k}$ and $M_2 \in R^{(k-q_*) \times k}$. Suppose the variance matrix of $M_2\Delta_{p-q_*}\xi_2 \in R^{k-q_*}$ has rank at least $p - q_*$ for all nonrandom vectors $\xi_2 \in R^{p-q_*}$ with $||\xi_2|| = 1$. Then, for $\Delta = (\Delta_{q_*}, \Delta_{p-q_*}) \in R^{k \times p}$, we have $P(\Delta$ has full column rank $p) = 1$.*

Comment: Corollary 15.2 follows from Lemma 15.1 by the following argument. We have

$$M\Delta = \begin{pmatrix} M_1\Delta_{q_*} & M_1\Delta_{p-q_*} \\ M_2\Delta_{q_*} & M_2\Delta_{p-q_*} \end{pmatrix} = \begin{pmatrix} I_{q_*} & M_1\Delta_{p-q_*} \\ 0^{(k-q_*)\times q_*} & M_2\Delta_{p-q_*} \end{pmatrix}. \tag{15.3}$$

The matrix Δ has full column rank p iff $M\Delta$ has full column rank p iff $M_2\Delta_{p-q_*}$ has full column rank $p-q_*$. The Corollary now follows from Lemma 15.1 applied with Δ, k, p, and ξ replaced by $M_2\Delta_{p-q_*}$, $k-q_*$, $p-q_*$, and ξ_2, respectively.

The following lemma is a special case of Cauchy's interlacing eigenvalues result, e.g., see Hwang (2004). As above, for a symmetric matrix A, let $\lambda_1(A) \geq \lambda_2(A) \geq ...$ denote the eigenvalues of A. Let A_{-r} denote a principal submatrix of A of order $r \geq 1$. That is, A_{-r} denotes A with some choice of r rows and the same r columns deleted.

Proposition 15.3 *Let A by a symmetric $k \times k$ matrix. Then,* $\lambda_k(A) \leq \lambda_{k-1}(A_{-1}) \leq \lambda_{k-1}(A) \leq ... \leq \lambda_2(A) \leq \lambda_1(A_{-1}) \leq \lambda_1(A)$.

The following is a straightforward corollary of Proposition 15.3.

Corollary 15.4 *Let A by a symmetric $k \times k$ matrix and let $r \in \{1, ..., k-1\}$. Then,* (a) $\lambda_m(A) \geq \lambda_m(A_{-r})$ *for* $m = 1, ..., k-r$ *and* (b) $\lambda_m(A) \leq \lambda_{m-r}(A_{-r})$ *for* $m = r+1, ..., k$.

Proof of Lemma 8.3. First, we prove the convergence result in Lemma 8.3. The singular value decomposition of $W_nD_nU_n$ is

$$W_nD_nU_n = C_n\Upsilon_nB_n', \tag{15.4}$$

because B_n is a matrix of eigenvectors of $U_n'D_n'W_n'W_nD_nU_n$, C_n is a matrix of eigenvectors of $W_nD_nU_nU_n'D_n'W_n'$, and Υ_n is the $k \times p$ matrix with the singular values $\{\tau_{jF_n} : j \leq p\}$ of $W_nD_nU_n$ on the diagonal (ordered so that $\tau_{jF_n} \geq 0$ is nonincreasing in j).

Using (15.4), we get

$$W_nD_nU_nB_{n,q}\Upsilon_{n,q}^{-1} = C_n\Upsilon_nB_n'B_{n,q}\Upsilon_{n,q}^{-1} = C_n\Upsilon_n\begin{pmatrix} I_q \\ 0^{(p-q)\times q} \end{pmatrix}\Upsilon_{n,q}^{-1} = C_n\begin{pmatrix} I_q \\ 0^{(k-q)\times q} \end{pmatrix} = C_{n,q}, \tag{15.5}$$

where the second equality uses $B_n'B_n = I_p$. Hence, we obtain

$$\begin{aligned} W_n\widehat{D}_nU_nB_{n,q}\Upsilon_{n,q}^{-1} &= W_nD_nU_nB_{n,q}\Upsilon_{n,q}^{-1} + W_nn^{1/2}(\widehat{D}_n - D_n)U_nB_{n,q}(n^{1/2}\Upsilon_{n,q})^{-1} \\ &= C_{n,q} + o_p(1) \rightarrow_p h_{3,q} = \overline{\Delta}_{h,q}, \end{aligned} \tag{15.6}$$

where the second equality uses $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$ (by the definition of q in (8.16)), $W_n = O(1)$ (by the condition $||W_F|| \leq M_1 < \infty \ \forall F \in \mathcal{F}_{WU}$, see (8.5)), $n^{1/2}(\widehat{D}_n - D_n) = O_p(1)$ (by Lemma 8.2), $U_n = O(1)$ (by the condition $||U_F|| \leq M_1 < \infty \ \forall F \in \mathcal{F}_{WU}$, see (8.5)), and $B_{n,q} \to h_{2,q}$ with $||vec(h_{2,q})|| < \infty$ (by (8.12) using the definitions in (8.17) and (13.1)). The convergence in (15.6) holds by (8.12), (8.17), and (13.1), and the last equality in (15.6) holds by the definition of $\overline{\Delta}_{h,q}$ in (8.17).

Using (15.4) again, we have

$$n^{1/2}W_nD_nU_nB_{n,p-q} = n^{1/2}C_n\Upsilon_nB_n'B_{n,p-q} = n^{1/2}C_n\Upsilon_n\begin{pmatrix} 0^{q\times(p-q)} \\ I_{p-q}\end{pmatrix}$$

$$= C_n\begin{pmatrix} 0^{q\times(p-q)} \\ n^{1/2}\Upsilon_{n,p-q} \\ 0^{(k-p)\times(p-q)}\end{pmatrix} \to h_3\begin{pmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p)\times(p-q)}\end{pmatrix} = h_3h_{1,p-q}^{\diamond}, \qquad (15.7)$$

where the second equality uses $B_n'B_n = I_p$, the convergence holds by (8.12) using the definitions in (8.17) and (13.2), and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (8.17).

Using (15.7) and Lemma 8.2, we get

$$\begin{aligned} n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} &= n^{1/2}W_nD_nU_nB_{n,p-q} + W_nn^{1/2}(\widehat{D}_n - D_n)U_nB_{n,p-q} \\ &\to_d h_3h_{1,p-q}^{\diamond} + h_{71}\overline{D}_hh_{81}h_{2,p-q} = \overline{\Delta}_{h,p-q}, \end{aligned} \qquad (15.8)$$

where $B_{n,p-q} \to h_{2,p-q}$, $W_n \to h_{71}$, and $U_n \to h_{81}$ by (8.3), (8.12), (8.17), and Assumption WU using the definitions in (13.1) and the last equality holds by the definition of $\overline{\Delta}_{h,p-q}$ in (8.17).

Equations (15.6) and (15.8) combine to prove

$$\begin{aligned} n^{1/2}W_n\widehat{D}_nU_nT_n &= n^{1/2}W_n\widehat{D}_nU_nB_nS_n = (W_n\widehat{D}_nU_nB_{n,q}\Upsilon_{n,q}^{-1}, n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q}) \\ &\to_d (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) = \overline{\Delta}_h \end{aligned} \qquad (15.9)$$

using the definition of S_n in (8.19). The convergence is joint with that in Lemma 8.2 because it just relies on the convergence of $n^{1/2}(\widehat{D}_n - D_n)$, which is part of the former. This establishes the convergence result of Lemma 8.3.

Properties (a) and (b) in Lemma 8.3 hold by definition. Property (c) in Lemma 8.3 holds by Lemma 8.2 and property (b) in Lemma 8.3.

Now, we prove property (d). We have

$$h'_{2,p-q}h_{2,p-q} = \lim B'_{n,p-q}B_{n,p-q} = I_{p-q} \text{ and } h'_{3,q}h_{3,q} = \lim C'_{n,q}C_{n,q} = I_q \tag{15.10}$$

because B_n and C_n are orthogonal matrices by (8.6) and (8.7). Hence, if $q = p$, then $\overline{\Delta}_h = \overline{\Delta}_{h,q} = h_{3,q}$, $\overline{\Delta}'_h\overline{\Delta}_h = I_p$, and $\overline{\Delta}_h$ has full column rank.

Hence, it suffices to consider the case where $q < p$ and $\lambda_{n,h} \in \Lambda_0\ \forall n \geq 1$, which is assumed in part (d). We prove part (d) for this case by applying Corollary 15.2 with $q_* = q$, $\Delta_{q_*} = \overline{\Delta}_{h,q}$ $(= h_{3,q})$, $\Delta_{p-q_*} = \overline{\Delta}_{h,p-q}$, $M = h'_3$, $M_1 = h'_{3,q}$, $M_2 = h'_{3,k-q}$, $\xi_2 \in R^{p-q}$, and $\Delta = \overline{\Delta}_h$. Corollary 15.2 gives the desired result that $P(\overline{\Delta}_h$ has full column rank $p) = 1$. The condition in Corollary 15.2 that "$M\Delta_{q_*} = (e_1, ..., e_{q_*})$" holds in this case because $h'_3\overline{\Delta}_{h,q} = h'_3h_{3,q} = (e_1, ..., e_q)$. The condition in Corollary 15.2 that "the variance matrix of $M_2\Delta_{p-q_*}\xi_2 \in R^{k-q_*}$ has rank at least $p - q_*$ for all nonrandom vectors $\xi_2 \in R^{p-q_*}$ with $||\xi_2|| = 1$" in this case becomes "the variance matrix of $h'_{3,k-q}\overline{\Delta}_{h,p-q}\xi_2 \in R^{k-q}$ has rank at least $p-q$ for all nonrandom vectors $\xi_2 \in R^{p-q}$ with $||\xi_2|| = 1$." It remains to establish the latter property, which is equivalent to

$$\lambda_{p-q}\left(Var(h'_{3,k-q}\overline{\Delta}_{h,p-q}\xi_2)\right) > 0\ \forall\xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{15.11}$$

We have

$$\begin{aligned} Var(h'_{3,k-q}\overline{\Delta}_{h,p-q}\xi_2) &= Var(h'_{3,k-q}h_{5,g}^{-1/2}\overline{D}_h h_{2,p-q}\xi_2) \\ &= ((h_{2,p-q}\xi_2)' \otimes (h'_{3,k-q}h_{5,g}^{-1/2}))Var(vec(\overline{D}_h))((h_{2,p-q}\xi_2) \otimes (h'_{3,k-q}h_{5,g}^{-1/2})') \\ &= ((h_{2,p-q}\xi_2)' \otimes (h'_{3,k-q}h_{5,g}^{-1/2}))\Phi_h^{vec(G_i)}((h_{2,p-q}\xi_2) \otimes (h'_{3,k-q}h_{5,g}^{-1/2})') \\ &= \Phi_h^{h'_{3,k-q}h_{5,g}^{-1/2}G_ih_{2,p-q}\xi_2}, \end{aligned} \tag{15.12}$$

where the first equality holds by the definition of $\overline{\Delta}_{h,p-q}$ in (8.17) and the fact that $h_{71} = h_{5,g}^{-1/2}$ and $h_{81} = I_p$ by the conditions in part (d) of Lemma 8.3, the second and fourth equalities use the general formula $vec(ABC) = (C' \otimes A)vec(B)$, the third equality holds because $vec(\overline{D}_h) \sim N(0^{pk}, \Phi_h^{vec(G_i)})$ by Lemma 8.2, and the fourth equality uses the definition of the variance matrix $\Phi_h^{a_i}$ in (8.15) for an arbitrary random vector a_i.

Next, we show that $\Phi_h^{h'_{3,k-q}h_{5,g}^{-1/2}G_ih_{2,p-q}\xi_2}$ equals the expected outer-product matrix

$\lim \Psi_{F_n}^{C'_{n,k-q}\Omega_n^{-1/2}G_iB_{n,p-q}\xi_2}$:

$$
\begin{aligned}
&\Phi_h^{h'_{3,k-q}h_{5,g}^{-1/2}G_ih_{2,p-q}\xi_2} \\
&= ((h_{2,p-q}\xi_2)' \otimes (h'_{3,k-q}h_{5,g}^{-1/2}))\Phi_h^{vec(G_i)}((h_{2,p-q}\xi_2) \otimes (h'_{3,k-q}h_{5,g}^{-1/2})') \\
&= \lim((B_{n,p-q}\xi_2)' \otimes (C'_{n,k-q}\Omega_n^{-1/2}))\Phi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C'_{n,k-q}\Omega_n^{-1/2})') \\
&= \lim((B_{n,p-q}\xi_2)' \otimes (C'_{n,k-q}\Omega_n^{-1/2}))\Psi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C'_{n,k-q}\Omega_n^{-1/2})') \\
&\quad - \lim((B_{n,p-q}\xi_2)' \otimes (C'_{n,k-q}\Omega_n^{-1/2}))E_{F_n}vec(G_i) \cdot E_{F_n}vec(G_i)'((B_{n,p-q}\xi_2) \otimes (C'_{n,k-q}\Omega_n^{-1/2})') \\
&= \lim((B_{n,p-q}\xi_2)' \otimes (C'_{n,k-q}\Omega_n^{-1/2}))\Psi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C'_{n,k-q}\Omega_n^{-1/2})') \\
&\quad - \lim E_{F_n}vec(C'_{n,k-q}\Omega_n^{-1/2}G_iB_{n,p-q}\xi_2) \cdot E_{F_n}vec(C'_{n,k-q}\Omega_n^{-1/2}G_iB_{n,p-q}\xi_2)' \\
&= \lim \Psi_{F_n}^{C'_{n,k-q}\Omega_n^{-1/2}G_iB_{n,p-q}\xi_2},
\end{aligned}
\tag{15.13}
$$

where the general formula $vec(ABC) = (C' \otimes A)vec(B)$ is used multiple times, the limits exist by the conditions imposed on the sequence $\{\lambda_{n,h} : n \geq 1\}$, the second equality uses $B_{n,p-j} \to h_{2,p-j}$, $C_{n,k-q} \to h_{3,k-q}$, and $\Omega_n^{-1/2} \to h_{5,g}^{-1/2}$, the third equality uses the definitions of $\Psi_F^{a_i}$ and $\Phi_F^{a_i}$ given in (3.2) and (8.15), respectively, and the last equality uses $E_{F_n}vec(C'_{n,k-q}\Omega_n^{-1/2}G_iB_{n,p-q}) = vec(C'_{n,k-q}\Omega_n^{-1/2}D_nB_{n,p-q}) = O(n^{-1/2})$ by (15.7) with $W_n = \Omega_n^{-1/2}$.

We can write $\lim \Psi_{F_n}^{vec(C'_n\Omega_n^{-1/2}G_iB_n)}$ as the limit of a subsequence $\{n_m : m \geq 1\}$ of matrices $\Psi_{F_{n_m}}^{vec(C'_{n_m}\Omega_{n_m}^{-1/2}G_iB_{n_m})}$ for which $F_{n_m} \in \mathcal{F}_{0j}$ for all $m \geq 1$ for some $j = 0, ..., q$. It cannot be the case that $j > q$, because if $j > q$, then we obtain a contradiction because $n_m^{1/2}\tau_{jF_{n_m}} \to \infty$ as $m \to \infty$ by the first condition of $\mathcal{F}_{0j}$ and $n_m^{1/2}\tau_{jF_{n_m}} \nrightarrow \infty$ as $m \to \infty$ by the definition of q in (8.16).

Now, we fix an arbitrary $j \in \{0, ..., q\}$. The continuity of the $\lambda_{p-j}(\cdot)$ function and the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$ imply that, for all $\xi \in R^{p-j}$ with $||\xi|| = 1$,

$$
\lambda_{p-j}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-j}\xi}\right) = \lim \lambda_{p-j}\left(\Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-j}\xi}\right) > 0. \tag{15.14}
$$

For all $\xi_2 \in R^{p-q}$ with $||\xi_2|| = 1$, let $\xi = (0^{q-j\prime}, \xi_2')' \in R^{p-j}$. Then, $B_{n_m,p-j}\xi = B_{n_m,p-q}\xi_2$ and, by (15.14),

$$
\lambda_{p-j}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}\right) > 0 \ \forall \xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{15.15}
$$

Next, we apply Corollary 15.4(b) with $A = \lim \Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}$ and $A_{-(q-j)} = \lim \Psi_{F_{n_m}}^{C'_{n_m,k-q}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}$, $m = p-j$, $r = q-j$, where $A_{-(q-j)}$ equals A with its first $q-j$ rows and columns deleted in the present case and $p > q$ implies that $m = p-j \geq 1$ for all $j = 0, ..., q$.

Corollary 15.4 and (15.15) give

$$\lambda_{p-q}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-q}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}\right) > 0 \ \forall \xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{15.16}$$

Equations (15.12), (15.13), and (15.16) combine to establish (15.11) and the proof of part (d) is complete.

Part (e) of the Lemma holds by replacing n by the subsequence value w_n throughout the arguments given above. □

Proof of Lemma 15.1. It suffices to show that $P(\Delta\xi = 0^k$ for some $\xi \in R^p$ with $||\xi|| = 1) = 0$.

For any constant $\gamma > 0$, there exists a constant $K_\gamma < \infty$ such that $P(||vec(\Delta)|| > K_\gamma) \leq \gamma$.

Given $\varepsilon > 0$, let $\{B(\xi_s, \varepsilon) : s = 1, ..., N_\varepsilon\}$ be a finite cover of $\{\xi \in R^p : ||\xi|| = 1\}$, where $||\xi_s|| = 1$ and $B(\xi_s, \varepsilon)$ is a ball in R^p centered at ξ_s of radius ε. It is possible to choose $\{\xi_s : s = 1, ..., N_\varepsilon\}$ such that the number, N_ε, of balls in the cover is of order ε^{-p+1}. That is, $N_\varepsilon \leq C_1\varepsilon^{-p+1}$ for some constant $C_1 < \infty$.

Let Δ_r denote the rth row of Δ for $r = 1, ..., k$ written as a column vector. If $\xi \in B(\xi_s, \varepsilon)$, we have

$$||\Delta\xi - \Delta\xi_s|| = \left(\sum_{r=1}^{k}(\Delta'_r(\xi - \xi_s))^2\right)^{1/2} \leq \left(\sum_{r=1}^{k}||\Delta_r||^2||\xi - \xi_s||^2\right)^{1/2} = \varepsilon||vec(\Delta)||, \tag{15.17}$$

where the inequality holds by the Cauchy-Bunyakovsky-Schwarz inequality. If $\xi \in B(\xi_s, \varepsilon)$ and $\Delta\xi = 0^k$, this gives

$$||\Delta\xi_s|| \leq \varepsilon||vec(\Delta)||. \tag{15.18}$$

Suppose $Z_* \in R^p$ has a multivariate normal distribution with pd variance matrix. Then, for any $\varepsilon > 0$,

$$P(||Z_*|| \leq \varepsilon) = \int_{\{||z|| \leq \varepsilon\}} f_{Z_*}(z)dz \leq \sup_{z \in R^k} f_{Z_*}(z)\int_{\{||z|| \leq \varepsilon\}} dz \leq C_2\varepsilon^p \tag{15.19}$$

for some constant $C_2 < \infty$, where $f_{Z_*}(z)$ denotes the density of Z_* with respect to Lebesgue measure, which exists because the variance matrix of Z_* is pd, and the inequalities hold because the density of a multivariate normal is bounded and the volume of a sphere in R^p of radius ε is proportional to ε^p.

For any $\xi \in R^p$ with $||\xi|| = 1$, let $B_\xi\Lambda_\xi B'_\xi$ be a spectral decomposition of $Var(\Delta\xi)$, where Λ_ξ is the diagonal $k \times k$ matrix with the eigenvalues of $Var(\Delta\xi)$ on its diagonal in nonincreasing order and B_ξ is an orthogonal $k \times k$ matrix whose columns are eigenvectors of $Var(\Delta\xi)$ that correspond to the eigenvalues in Λ_ξ. By assumption, the rank of $Var(\Delta\xi)$ is p or larger. In consequence, the first p diagonal elements of Λ_ξ are positive. We have $||\Delta\xi|| = ||B'_\xi \Delta\xi||$ and $Var(B'_\xi \Delta\xi) = B'_\xi Var(\Delta\xi) B_\xi = \Lambda_\xi$. Let $(B'_\xi \Delta\xi)_p$ denote the p vector that contains the first p elements of the k vector $B'_\xi \Delta\xi$. Let $\Lambda_{\xi p}$ denote the upper left $p \times p$ submatrix of Λ_ξ. We have $Var((B'_\xi \Delta\xi)_p) = \Lambda_{\xi p}$ and $\Lambda_{\xi p}$ is pd (because the first p diagonal elements of Λ_ξ are positive).

Now, given any $\gamma > 0$ and $\varepsilon > 0$, we have

$$\begin{aligned}
& P(\Delta\xi = 0^k \text{ for some } \xi \in R^p \text{ with } ||\xi|| = 1) \\
= \; & P\left(\cup_{s=1}^{N_\varepsilon} \cup_{\xi \in B(\xi_s, \varepsilon): ||\xi||=1} \{\Delta\xi = 0^k\}\right) \\
\leq \; & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon ||vec(\Delta)||\}\right) \\
\leq \; & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon ||vec(\Delta)||\} \cap \{||vec(\Delta)|| \leq K_\gamma\}\right) + P(||vec(\Delta)|| > K_\gamma) \\
\leq \; & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon K_\gamma\}\right) + \gamma \\
\leq \; & \sum_{s=1}^{N_\varepsilon} P(||\Delta\xi_s|| \leq \varepsilon K_\gamma) + \gamma \\
\leq \; & \sum_{s=1}^{N_\varepsilon} P(||(B'_{\xi_s} \Delta\xi_s)_p|| \leq \varepsilon K_\gamma) + \gamma \\
\leq \; & N_\varepsilon C_2 K_\gamma^p \varepsilon^p + \gamma \\
\leq \; & C_1 \varepsilon^{-p+1} C_2 K_\gamma^p \varepsilon^p + \gamma \\
\to \; & \gamma \text{ as } \varepsilon \to 0,
\end{aligned} \tag{15.20}$$

where the first inequality holds by (15.18) using $\xi \in B(\xi_s, \varepsilon)$, the third inequality uses the definition of K_γ, the third last inequality holds because $||(B'_{\xi_s} \Delta\xi_s)_p|| \leq ||B'_{\xi_s} \Delta\xi_s|| = ||\Delta\xi_s||$ using the definitions in the paragraph that follows the paragraph that contains (15.19), the second last inequality holds by (15.19) with $Z_* = (B'_{\xi_s} \Delta\xi_s)_p$ and the fact that the variance matrix of $(B'_{\xi_s} \Delta\xi_s)_p$ is pd by the argument given in the paragraph following (15.19), and the last inequality holds by the bound given above on N_ε.

Because $\gamma > 0$ is arbitrary, (15.20) implies that $P(\Delta\xi = 0^k$ for some $\xi \in R^p$ with $||\xi|| = 1) = 0$, which completes the proof. □

16 Proof of Theorem 8.4

Theorem 8.4 of AG1. *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

(a) $\widehat{\kappa}_{pn} \to_p \infty$ *if* $q = p$,

(b) $\widehat{\kappa}_{pn} \to_d \lambda_{\min}(\overline{\Delta}'_{h,p-q} h_{3,k-q} h'_{3,k-q} \overline{\Delta}_{h,p-q})$ *if* $q < p$,

(c) $\widehat{\kappa}_{jn} \to_p \infty$ *for all* $j \leq q$,

(d) *the* (*ordered*) *vector of the smallest* $p-q$ *eigenvalues of* $n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n$, *i.e.*, $(\widehat{\kappa}_{(q+1)n}, ..., \widehat{\kappa}_{pn})'$, *converges in distribution to the* (*ordered*) $p-q$ *vector of the eigenvalues of* $\overline{\Delta}_{h,p-q}'h_{3,k-q}h_{3,k-q}' \times\overline{\Delta}_{h,p-q} \in R^{(p-q)\times(p-q)}$,

(e) *the convergence in parts* (a)-(d) *holds jointly with the convergence in Lemma* 8.3, *and*

(f) *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$, *the results in parts* (a)-(e) *hold with* n *replaced with* w_n.

The proof of Theorem 8.4 uses the following rate of convergence lemma. This lemma is a key technical contribution of the paper.

Lemma 16.1 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$ *and for which* q *defined in* (8.16) *satisfies* $q \geq 1$, *we have* (a) $\widehat{\kappa}_{jn} \to_p \infty$ *for* $j = 1, ..., q$ *and* (b) *when* $p > q$, $\widehat{\kappa}_{jn} = o_p((n^{1/2}\tau_{\ell F_n})^2)$ *for all* $\ell \leq q$ *and* $j = q+1, ..., p$. *Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$, *the same result holds with* n *replaced with* w_n.

Proof of Lemma 16.1. By the definitions in (8.9) and (8.12), $h_{6,j} := \lim \tau_{(j+1)F_n}/\tau_{jF_n}$ for $j = 1, ..., p-1$. By the definition of q in (8.16), $h_{6,q} = 0$ if $q < p$. If $q = p$, $h_{6,q}$ is not defined by (8.9) and (8.12) and we define it here to equal zero. Because τ_{jF} is nonnegative and nonincreasing in j, $h_{6,j} \in [0, 1]$. If $h_{6,j} > 0$, then $\{\tau_{jF_n} : n \geq 1\}$ and $\{\tau_{(j+1)F_n} : n \geq 1\}$ are of the same order of magnitude, i.e., $0 < \lim \tau_{(j+1)F_n}/\tau_{jF_n} \leq 1$.[50] We group the first q singular values into groups that have the same order of magnitude within each group. Let G_h ($\in \{1, ..., q\}$) denote the number of groups. (We have $G_h \geq 1$ because $q \geq 1$ is assumed in the statement of the lemma.) Note that G_h equals the number of values in $\{h_{6,1}, ..., h_{6,q}\}$ that equal zero. Let r_g and $r_g^\diamond$ denote the indices of the first and last singular values, respectively, in the gth group for $g = 1, ..., G_h$. Thus, $r_1 = 1$, $r_g^\diamond = r_{g+1} - 1$, where r_{G_h+1} is defined to equal $q+1$, and $r_{G_h}^\diamond = q$. Note that r_g and $r_g^\diamond$ depend on h. By definition, the singular values in the gth group, which have the gth largest order of magnitude, are $\{\tau_{r_gF_n} : n \geq 1\}, ..., \{\tau_{r_g^\diamond F_n} : n \geq 1\}$. By construction, $h_{6,j} > 0$ for all $j \in \{r_g, ..., r_g^\diamond - 1\}$ for $g = 1, ..., G_h$. (The reason is: if $h_{6,j}$ is equal to zero for some $j \in \{r_g, ..., r_g^\diamond - 1\}$, then $\{\tau_{r_g^\diamond F_n} : n \geq 1\}$ is of smaller order of magnitude than $\{\tau_{r_gF_n} : n \geq 1\}$, which contradicts the definition of $r_g^\diamond$.) Also by construction, $\lim \tau_{j'F_n}/\tau_{jF_n} = 0$ for any (j, j') in groups (g, g'), respectively, with $g < g'$. Note that when $p = 1$ we have $G_h = 1$ and $r_1 = r_1^\diamond = 1$.

[50] Note that $\sup_{j\geq 1, F\in\mathcal{F}_{WU}} \tau_{jF} < \infty$ by the conditions $||W_F|| \leq M_1$ and $||U_F|| \leq M_1$ in $\mathcal{F}_{WU}$ and the moment conditions in $\mathcal{F}$. Thus, $\{\tau_{jF_n} : n \geq 1\}$ does not diverge to infinity, and the "order of magnitude" of $\{\tau_{jF_n} : n \geq 1\}$ refers to whether this sequence converges to zero, and how slowly or quickly it does, when it does converge to zero.

The eigenvalues $\{\widehat{\kappa}_{jn} : j \leq p\}$ of $n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n$ are solutions to the determinantal equation $|n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n - \kappa I_p| = 0$. Equivalently, by multiplying this equation by $\tau_{r_1F_n}^{-2} n^{-1}|B_n'U_n'\widehat{U}_n^{-1\prime}| \times |\widehat{U}_n^{-1}U_nB_n|$, they are solutions to

$$|\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - (n^{1/2}\tau_{r_1F_n})^{-2}\kappa B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n| = 0 \tag{16.1}$$

wp$\to$1, using $|A_1A_2| = |A_1|\cdot|A_2|$ for any conformable square matrices A_1 and A_2, $|B_n| > 0$, $|U_n| > 0$ (by the conditions in $\mathcal{F}_{WU}$ in (8.5) because $\Lambda_* \subset \Lambda_2$ and Λ_2 only contains distributions in $\mathcal{F}_{WU}$), $|\widehat{U}_n^{-1}| > 0$ wp$\to$1 (because $\widehat{U}_n \to_p h_{81}$ by (8.2), (8.12), (8.17), and Assumption WU(b) and (c) and h_{81} is pd), and $\tau_{r_1F_n} > 0$ for n large (because $n^{1/2}\tau_{r_1F_n} \to \infty$ for $r_1 \leq q$). (For simplicity, we omit the qualifier wp$\to$1 from some statements below.) Thus, $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j \leq p\}$ solve

$$\begin{aligned}
&|\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - \kappa(I_p + \widehat{A}_n)| = 0 \text{ or} \\
&|(I_p + \widehat{A}_n)^{-1}\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - \kappa I_p| = 0, \text{ where} \\
&\widehat{A}_n = \begin{bmatrix} \widehat{A}_{1n} & \widehat{A}_{2n} \\ \widehat{A}_{2n}' & \widehat{A}_{3n} \end{bmatrix} := B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n - I_p
\end{aligned} \tag{16.2}$$

for $\widehat{A}_{1n} \in R^{r_1^\diamond \times r_1^\diamond}$, $\widehat{A}_{2n} \in R^{r_1^\diamond \times (p-r_1^\diamond)}$, and $\widehat{A}_{3n} \in R^{(p-r_1^\diamond)\times(p-r_1^\diamond)}$ and the second line is obtained by multiplying the first line by $|(I_p + \widehat{A}_n)^{-1}|$.

We have

$$\begin{aligned}
&\tau_{r_1F_n}^{-1}\widehat{W}_n\widehat{D}_nU_nB_n \\
=\ &\tau_{r_1F_n}^{-1}(\widehat{W}_nW_n^{-1})W_nD_nU_nB_n - (n^{1/2}\tau_{r_1F_n})^{-1}\widehat{W}_nn^{1/2}(\widehat{D}_n - D_n)U_nB_n \\
=\ &\tau_{r_1F_n}^{-1}(\widehat{W}_nW_n^{-1})C_n\Upsilon_n + O_p((n^{1/2}\tau_{r_1F_n})^{-1}) \qquad (16.3)\\
=\ &(I_k + o_p(1))C_n \begin{bmatrix} h_{6,r_1^\diamond}^\diamond + o(1) & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(p-r_1^\diamond)\times r_1^\diamond} & O(\tau_{r_2F_n}/\tau_{r_1F_n})^{(p-r_1^\diamond)\times(p-r_1^\diamond)} \\ 0^{(k-p)\times r_1^\diamond} & 0^{(k-p)\times(p-r_1^\diamond)} \end{bmatrix} + O_p((n^{1/2}\tau_{r_1F_n})^{-1}) \\
\to_p\ &h_3 \begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix}, \text{ where } h_{6,r_1^\diamond}^\diamond := Diag\{1, h_{6,1}, h_{6,1}h_{6,2}, ..., \prod_{\ell=1}^{r_1^\diamond-1} h_{6,\ell}\},
\end{aligned}$$

$h_{6,r_1^\diamond}^\diamond \in R^{r_1^\diamond\times r_1^\diamond}$, $h_{6,r_1^\diamond}^\diamond := 1$ when $r_1^\diamond = 1$, $O(\tau_{r_2F_n}/\tau_{r_1F_n})^{(p-r_1^\diamond)\times(p-r_1^\diamond)}$ denotes a diagonal $(p-r_1^\diamond)\times(p-r_1^\diamond)$ matrix whose diagonal elements are $O(\tau_{r_2F_n}/\tau_{r_1F_n})$, the second equality uses (15.4), $\widehat{W}_n \to_p h_{71}$ (by Assumption WU(a) and (c)), $||h_{71}|| = ||\lim W_n|| < \infty$ (by the conditions in $\mathcal{F}_{WU}$ defined in (8.5)), $n^{1/2}(\widehat{D}_n - D_n) = O_p(1)$ (by Lemma 8.2), $U_n = O(1)$ (by the conditions in $\mathcal{F}_{WU}$), and $B_n =$

$O(1)$ (because B_n is orthogonal), the third equality uses $\widehat{W}_n W_n^{-1} \to_p I_k$ (because $\widehat{W}_n \to_p h_{71}$, $h_{71} := \lim W_n$, and h_{71} is pd by the conditions in $\mathcal{F}_{WU}$), $\tau_{jF_n}/\tau_{r_1F_n} = \prod_{\ell=1}^{j-1}(\tau_{(\ell+1)F_n}/\tau_{\ell F_n}) = \prod_{\ell=1}^{j-1} h_{6,\ell} + o(1)$ for $j = 2, ..., r_1^\diamond$, and $\tau_{jF_n}/\tau_{r_1F_n} = O(\tau_{r_2F_n}/\tau_{r_1F_n})$ for $j = r_2, ..., p$ (because $\{\tau_{jF_n} : j \leq p\}$ are nonincreasing in j), and the convergence uses $C_n \to h_3$, $\tau_{r_2F_n}/\tau_{r_1F_n} \to 0$ (by the definition of r_2), and $n^{1/2}\tau_{r_1F_n} \to \infty$ (by (8.16) because $r_1 \leq q$).[51]

Equation (16.3) yields

$$\tau_{r_1F_n}^{-2} B_n' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_n \to_p \begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond \times (p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix}' h_3' h_3 \begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond \times (p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix}$$
$$= \begin{bmatrix} h_{6,r_1^\diamond}^{\diamond 2} & 0^{r_1^\diamond \times (p-r_1^\diamond)} \\ 0^{(p-r_1^\diamond)\times r_1^\diamond} & 0^{(p-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix}, \tag{16.4}$$

where the equality holds because $h_3'h_3 = \lim C_n'C_n = I_k$ using (8.7).

In addition, we have

$$\widehat{A}_n := B_n' U_n' \widehat{U}_n^{-1\prime} \widehat{U}_n^{-1} U_n B_n - I_p \to_p 0^{p\times p} \tag{16.5}$$

using $\widehat{U}_n^{-1} U_n \to_p I_p$ (because $\widehat{U}_n \to_p h_{81}$ by Assumption WU(b) and (c), $h_{81} := \lim U_n$, and h_{81} is pd by the conditions in $\mathcal{F}_{WU}$), $B_n \to h_2$, and $h_2'h_2 = I_p$ (because B_n is orthogonal for all $n \geq 1$).

The ordered vector of eigenvalues of a matrix is a continuous function of the matrix by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38). Hence, by the second line of (16.2), (16.4), (16.5), and Slutsky's Theorem, the largest $r_1^\diamond$ eigenvalues of $\tau_{r_1F_n}^{-2} B_n' \widehat{U}_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n B_n$ (i.e., $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j \leq r_1^\diamond\}$ by the definition of $\widehat{\kappa}_{jn}$), satisfy

$$((n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{1n}, ..., (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{r_1^\diamond n}) \to_p (1, h_{6,1}^2, h_{6,1}^2 h_{6,2}^2, ..., \prod_{\ell=1}^{r_1^\diamond - 1} h_{6,\ell}^2) \text{ and so}$$
$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = 1, ..., r_1^\diamond \tag{16.6}$$

because $n^{1/2}\tau_{r_1F_n} \to \infty$ (by (8.16) since $r_1 \leq q$) and $h_{6,\ell} > 0$ for all $\ell \in \{1, ..., r_1^\diamond - 1\}$ (as noted above). By the same argument, the smallest $p - r_1^\diamond$ eigenvalues of $\tau_{r_1F_n}^{-2} B_n' \widehat{U}_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n B_n$, i.e., $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j = r_1^\diamond + 1, ..., p\}$, satisfy

$$(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} \to_p 0 \ \forall j = r_1^\diamond + 1, ..., p. \tag{16.7}$$

If $G_h = 1$, (16.6) proves part (a) of the lemma and (16.7) proves part (b) of the lemma (because

[51] For matrices that are written as $O(\cdot)$, we sometimes provide the dimensions of the matrix as superscripts for clarity, and sometimes we do not provide the dimensions for simplicity.

in this case $r_1^\diamond = q$ and $\tau_{r_1 F_n}/\tau_{\ell F_n} = O(1)$ for all $\ell \leq q$ by the definitions of q and G_h). Hence, from here on, we assume that $G_h \geq 2$.

Next, define B_{n,j_1,j_2} to be the $p \times (j_2 - j_1)$ matrix that consists of the $j_1 + 1, ..., j_2$ columns of B_n for $0 \leq j_1 < j_2 \leq p$. Note that the difference between the two subscripts j_1 and j_2 equals the number of columns of B_{n,j_1,j_2}, which is useful for keeping track of the dimensions of the B_{n,j_1,j_2} matrices that appear below. By definition, $B_n = (B_{n,0,r_1^\diamond}, B_{n,r_1^\diamond,p})$.

By (16.3) (excluding the convergence part) applied once with $B_{n,r_1^\diamond,p}$ in place of B_n as the far-right multiplicand and applied a second time with $B_{n,0,r_1^\diamond}$ in place of B_n as the far-right multiplicand, we have

$$\begin{aligned} \varrho_n &:= \tau_{r_1 F_n}^{-2} B_{n,0,r_1^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} \\ &= \begin{bmatrix} h_{6,r_1^\diamond}^\diamond + o(1) \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} \end{bmatrix}' C_n'(I_k + o_p(1))C_n \begin{bmatrix} 0^{r_1^\diamond \times (p-r_1^\diamond)} \\ O(\tau_{r_2 F_n}/\tau_{r_1 F_n})^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix} \\ &\quad + O_p((n^{1/2}\tau_{r_1 F_n})^{-1}) \\ &= o_p(\tau_{r_2 F_n}/\tau_{r_1 F_n}) + O_p((n^{1/2}\tau_{r_1 F_n})^{-1}), \end{aligned} \tag{16.8}$$

where the last equality holds because (i) $C_n'(I_k + o_p(1))C_n = I_k + o_p(1)$, (ii) when I_k appears in place of $C_n'(I_k + o_p(1))C_n$, the first summand on the left-hand side (lhs) of the last equality equals $0^{r_1^\diamond \times (p-r_1^\diamond)}$, and (iii) when $o_p(1)$ appears in place of $C_n'(I_k + o_p(1))C_n$, the first summand on the lhs of the last equality equals an $r_1^\diamond \times (p - r_1^\diamond)$ matrix with elements that are $o_p(\tau_{r_2 F_n}/\tau_{r_1 F_n})$.

Define

$$\begin{aligned} \widehat{\xi}_{1n}(\kappa) &:= \tau_{r_1 F_n}^{-2} B_{n,0,r_1^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,0,r_1^\diamond} - \kappa(I_{r_1^\diamond} + \widehat{A}_{1n}) \in R^{r_1^\diamond \times r_1^\diamond}, \\ \widehat{\xi}_{2n}(\kappa) &:= \varrho_n - \kappa \widehat{A}_{2n} \in R^{r_1^\diamond \times (p-r_1^\diamond)}, \text{ and} \\ \widehat{\xi}_{3n}(\kappa) &:= \tau_{r_1 F_n}^{-2} B_{n,r_1^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} - \kappa(I_{p-r_1^\diamond} + \widehat{A}_{3n}) \in R^{(p-r_1^\diamond)\times(p-r_1^\diamond)}. \end{aligned} \tag{16.9}$$

As in the first line of (16.2), $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j \le p\}$ solve

$$\begin{aligned}
0 &= |\tau_{r_1F_n}^{-2} B_n' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_n - \kappa(I_p + \widehat{A}_n)| \\
&= \left| \begin{bmatrix} \widehat{\xi}_{1n}(\kappa) & \widehat{\xi}_{2n}(\kappa) \\ \widehat{\xi}_{2n}(\kappa)' & \widehat{\xi}_{3n}(\kappa) \end{bmatrix} \right| \\
&= |\widehat{\xi}_{1n}(\kappa)| \cdot |\widehat{\xi}_{3n}(\kappa) - \widehat{\xi}_{2n}(\kappa)' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{\xi}_{2n}(\kappa)| \\
&= |\widehat{\xi}_{1n}(\kappa)| \cdot |\tau_{r_1F_n}^{-2} B_{n,r_1^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} - \varrho_n' \widehat{\xi}_{1n}^{-1}(\kappa) \varrho_n \\
&\quad -\kappa(I_{p-r_1^\diamond} + \widehat{A}_{3n} - \widehat{A}_{2n}' \widehat{\xi}_{1n}^{-1}(\kappa) \varrho_n - \varrho_n' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{A}_{2n} + \kappa \widehat{A}_{2n}' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{A}_{2n})|, \qquad (16.10)
\end{aligned}$$

where the third equality uses the standard formula for the determinant of a partitioned matrix and the result given in (16.11) below, which shows that $\widehat{\xi}_{1n}(\kappa)$ is nonsingular wp$\to$1 for κ equal to any solution $(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}$ to the first equality in (16.10) for $j \le p$, and the last equality holds by algebra.[52]

Now we show that, for $j = r_1^\diamond+1, ..., p$, $(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}$ cannot solve the determinantal equation $|\widehat{\xi}_{1n}(\kappa)| = 0$, wp$\to$1, where this determinant is the first multiplicand on the right-hand side (rhs) of (16.10). This implies that $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j = r_1^\diamond + 1, ..., p\}$ must solve the determinantal equation based on the second multiplicand on the rhs of (16.10) wp$\to$1. For $j = r_1^\diamond + 1, ..., p$, we have

$$\begin{aligned}
\widetilde{\xi}_{j1n} &:= \widehat{\xi}_{1n}((n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}) \\
&= \tau_{r_1F_n}^{-2} B_{n,0,r_1^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,0,r_1^\diamond} - (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}(I_{r_1^\diamond} + \widehat{A}_{1n}) \\
&= h_{6,r_1^\diamond}^{\diamond 2} + o_p(1) - o_p(1)(I_{r_1^\diamond} + o_p(1)) \\
&= h_{6,r_1^\diamond}^{\diamond 2} + o_p(1), \qquad (16.11)
\end{aligned}$$

where the second last equality holds by (16.4), (16.5), and (16.7). Equation (16.11) and $\lambda_{\min}(h_{6,r_1^\diamond}^{\diamond 2}) > 0$ (which follows from the definition of $h_{6,r_1^\diamond}^{\diamond}$ in (16.3) and the fact that $h_{6,\ell} > 0$ for all $\ell \in \{1, ..., r_1^\diamond - 1\}$) establish the result stated in the first sentence of this paragraph.

For $j = r_1^\diamond+1, ..., p$, plugging $(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}$ into the second multiplicand on the rhs of (16.10)

[52] The determinant of the partitioned matrix $\xi = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_2' & \xi_3 \end{bmatrix}$ equals $|\xi| = |\xi_1| \cdot |\xi_3 - \xi_2' \xi_1^{-1} \xi_2|$ provided ξ_1 is nonsingular, e.g., see Rao (1973, p. 32).

gives

$$\begin{aligned} 0 &= |\tau^{-2}_{r_1F_n} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p((\tau_{r_2F_n}/\tau_{r_1F_n})^2) + O_p((n^{1/2}\tau_{r_1F_n})^{-2}) \\ &\quad -(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r_1^\diamond} + \widehat{A}_{j2n})|, \text{ where} \qquad (16.12) \\ \widehat{A}_{j2n} &:= \widehat{A}_{3n} - \widehat{A}'_{2n}\widetilde{\xi}^{-1}_{j1n}\varrho_n - \varrho'_n\widetilde{\xi}^{-1}_{j1n}\widehat{A}_{2n} + (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}\widehat{A}'_{2n}\widetilde{\xi}^{-1}_{j1n}\widehat{A}_{2n} \in R^{(p-r_1^\diamond)\times(p-r_1^\diamond)} \end{aligned}$$

using (16.8) and (16.11). Multiplying (16.12) by $\tau^2_{r_1F_n}/\tau^2_{r_2F_n}$ gives

$$0 = |\tau^{-2}_{r_2F_n} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p(1) - (n^{1/2}\tau_{r_2F_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r_1^\diamond} + \widehat{A}_{j2n})| \qquad (16.13)$$

using $O_p((n^{1/2}\tau_{r_2F_n})^{-2}) = o_p(1)$ (because $r_2 \leq q$ by the definition of r_2 and $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$ by the definition of q in (8.16)).

Thus, $\{(n^{1/2}\tau_{r_2F_n})^{-2}\widehat{\kappa}_{jn} : j = r_1^\diamond + 1, ..., p\}$ solve

$$0 = |\tau^{-2}_{r_2F_n} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p(1) - \kappa(I_{p-r_1^\diamond} + \widehat{A}_{j2n})|. \qquad (16.14)$$

For $j = r_1^\diamond + 1, ..., p$, we have

$$\widehat{A}_{j2n} = o_p(1), \qquad (16.15)$$

because $\widehat{A}_{2n} = o_p(1)$ and $\widehat{A}_{3n} = o_p(1)$ (by (16.5)), $\widetilde{\xi}^{-1}_{j1n} = O_p(1)$ (by (16.11)), $\varrho_n = o_p(1)$ (by (16.8) since $\tau_{r_2F_n} \leq \tau_{r_1F_n}$ and $n^{1/2}\tau_{r_1F_n} \to \infty$), and $(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1)$ for $j = r_1^\diamond + 1, ..., p$ (by (16.7)).

Now, we repeat the argument from (16.2) to (16.15) with the expression in (16.14) replacing that in the first line of (16.2), with (16.15) replacing (16.5), and with $j = r_2^\diamond + 1, ..., p$, $\widehat{A}_{j2n}$, $B_{n,p-r_1^\diamond}$, $\tau_{r_2F_n}$, $\tau_{r_3F_n}$, $r_2^\diamond - r_1^\diamond$, $p - r_2^\diamond$, and $h^\diamond_{6,r_2^\diamond} = Diag\{1, h_{6,r_1^\diamond+1}, h_{6,r_1^\diamond+1}h_{6,r_1^\diamond+2}, ..., \prod_{\ell=r_1^\diamond+1}^{r_2^\diamond-1} h_{6,\ell}\} \in R^{(r_2^\diamond-r_1^\diamond)\times(r_2^\diamond-r_1^\diamond)}$ in place of $j = r_1^\diamond + 1, ..., p$, $\widehat{A}_n$, B_n, $\tau_{r_1F_n}$, $\tau_{r_2F_n}$, $r_1^\diamond$, $p - r_1^\diamond$, and $h^\diamond_{6,r_1^\diamond}$, respectively. (The fact that $\widehat{A}_{j2n}$ depends on j, whereas $\widehat{A}_n$ does not, does not affect the argument.) In addition, $B_{n,0,r_1^\diamond}$ and $B_{n,r_1^\diamond,p}$ in (16.8)-(16.10) are replaced by the matrices $B_{n,r_1^\diamond,r_2^\diamond}$ and $B_{n,r_2^\diamond,p}$ (which consist of the $r_1^\diamond + 1, ..., r_2^\diamond$ columns of B_n and the last $p - r_2^\diamond$ columns of B_n, respectively.) This argument gives the analogues of (16.6) and (16.7), which are

$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = r_2, ..., r_2^\diamond \text{ and } (n^{1/2}\tau_{r_2F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \ \forall j = r_2^\diamond + 1, ..., p. \qquad (16.16)$$

In addition, the analogue of (16.14) is the same as (16.14) but with $\widehat{A}_{j3n}$ in place of $\widehat{A}_{j2n}$, where $\widehat{A}_{j3n}$ is defined just as $\widehat{A}_{j2n}$ is defined in (16.12) but with $\widehat{A}_{2j2n}$ and $\widehat{A}_{3j2n}$ in place of $\widehat{A}_{2n}$ and $\widehat{A}_{3n}$, respectively, where

$$\widehat{A}_{j2n} = \left[\begin{array}{cc} \widehat{A}_{1j2n} & \widehat{A}_{2j2n} \\ \widehat{A}'_{2j2n} & \widehat{A}_{3j2n} \end{array} \right] \tag{16.17}$$

for $\widehat{A}_{1j2n} \in R^{r_2^\diamond \times r_2^\diamond}$, $\widehat{A}_{2j2n} \in R^{r_2^\diamond \times (p - r_1^\diamond - r_2^\diamond)}$, and $\widehat{A}_{3j2n} \in R^{(p - r_1^\diamond - r_2^\diamond) \times (p - r_1^\diamond - r_2^\diamond)}$.

Repeating the argument $G_h - 2$ more times yields

$$\widehat{\kappa}_{jn} \rightarrow_p \infty \ \forall j = 1, ..., r_{G_h}^\diamond \text{ and } (n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \ \forall j = r_g^\diamond + 1, ..., p, \forall g = 1, ..., G_h. \tag{16.18}$$

A formal proof of this "repetition of the argument $G_h - 2$ more times" is given below using induction. Because $r_{G_h}^\diamond = q$, the first result in (16.18) proves part (a) of the lemma.

The second result in (16.18) with $g = G_h$ implies: for all $j = q + 1, ..., p$,

$$(n^{1/2}\tau_{r_{G_h} F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \tag{16.19}$$

because $r_{G_h}^\diamond = q$. Either $r_{G_h} = r_{G_h}^\diamond = q$ or $r_{G_h} < r_{G_h}^\diamond = q$. In the former case, $(n^{1/2}\tau_{qF_n})^{-2}\widehat{\kappa}_{jn} = o_p(1)$ for $j = q + 1, ..., p$ by (16.19). In the latter case, we have

$$\lim \frac{\tau_{qF_n}}{\tau_{r_{G_h} F_n}} = \lim \frac{\tau_{r_{G_h}^\diamond F_n}}{\tau_{r_{G_h} F_n}} = \prod_{j = r_{G_h}}^{r_{G_h}^\diamond - 1} h_{6,j} > 0, \tag{16.20}$$

where the inequality holds because $h_{6,\ell} > 0$ for all $\ell \in \{r_{G_h}, ..., r_{G_h}^\diamond - 1\}$, as noted at the beginning of the proof. Hence, in this case too, $(n^{1/2}\tau_{qF_n})^{-2}\widehat{\kappa}_{jn} = o_p(1)$ for $j = q + 1, ..., p$ by (16.19) and (16.20). Because $\tau_{\ell F_n} \geq \tau_{qF_n}$ for all $\ell \leq q$, this establishes part (b) of the lemma.

Now we establish by induction the results given in (16.18) that are obtained heuristically by "repeating the argument $G_h - 2$ more times." The induction proof shows that subtleties arise when establishing the asymptotic negligibility of certain terms.

Let o_{gp} denote a symmetric $(p - r_{g-1}^\diamond) \times (p - r_{g-1}^\diamond)$ matrix whose (ℓ, m) element for $\ell, m = 1, ..., p - r_{g-1}^\diamond$ is $o_p(\tau_{(r_{g-1}^\diamond + \ell)F_n}\tau_{(r_{g-1}^\diamond + m)F_n}/\tau_{r_g F_n}^2) + O_p((n^{1/2}\tau_{r_g F_n})^{-1})$. Note that $o_{gp} = o_p(1)$ because $r_{g-1}^\diamond + \ell \geq r_g$ for $\ell \geq 1$ (since τ_{jF_n} are nonincreasing in j) and $n^{1/2}\tau_{r_g F_n} \rightarrow \infty$ for $g = 1, ..., G_h$.

We now show by induction over $g = 1, ..., G_h$ that wp→1 $\{(n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn} : j = r_{g-1}^\diamond + 1, ..., p\}$ solve

$$|\tau_{r_g F_n}^{-2} B'_{n, r_{g-1}^\diamond, p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n, r_{g-1}^\diamond, p} + o_{gp} - \kappa(I_{p - r_{g-1}^\diamond} + \widehat{A}_{jgn})| = 0 \tag{16.21}$$

for some $(p - r_{g-1}^\diamond) \times (p - r_{g-1}^\diamond)$ symmetric matrices $\widehat{A}_{jgn} = o_p(1)$ and o_{gp} (where the matrices that are o_{gp} may depend on j).

The initiation step of the induction proof holds because (16.21) holds with $g = 1$ by the first line of (16.2) with $\widehat{A}_{jgn} := \widehat{A}_n$ and $o_{gp} = 0$ for $g = 1$ (and using the fact that, for $g = 1$, $r^\diamond_{g-1} = r^\diamond_0 := 0$ and $B_{n,r^\diamond_{g-1},p} = B_{n,0,p} = B_n$).

For the induction step of the proof, we assume that (16.21) holds for some $g \in \{1, ..., G_h - 1\}$ and show that it then also holds for $g + 1$. By an argument analogous to that in (16.3), we have

$$\tau^{-1}_{r_gF_n}\widehat{W}_n\widehat{D}_nU_nB_{n,r^\diamond_{g-1},p} = (I_k + o_p(1))C_n \begin{bmatrix} 0^{r^\diamond_{g-1}\times(p-r^\diamond_{g-1})} \\ Diag\{\tau_{r_gF_n}, ..., \tau_{pF_n}\}/\tau_{r_gF_n} \\ 0^{(k-p)\times(p-r^\diamond_{g-1})} \end{bmatrix} + O_p((n^{1/2}\tau_{r_gF_n})^{-1})$$

$$\to_p h_3 \left(\begin{bmatrix} 0^{r^\diamond_{g-1}\times(r^\diamond_g-r^\diamond_{g-1})} \\ h^\diamond_{6,r^\diamond_g} \\ 0^{(k-r^\diamond_g)\times(r^\diamond_g-r^\diamond_{g-1})} \end{bmatrix}, 0^{k\times(p-r^\diamond_g)} \right), \text{ where } h^\diamond_{6,r^\diamond_g} := Diag\{1, h_{6,r_g}, ..., \prod_{j=r^\diamond_{g-1}+1}^{r^\diamond_g-1} h_{6,j}\}, \tag{16.22}$$

$h^\diamond_{6,r^\diamond_g} \in R^{(r^\diamond_g-r^\diamond_{g-1})\times(r^\diamond_g-r^\diamond_{g-1})}$, and $h^\diamond_{6,r^\diamond_g} := 1$ when $r^\diamond_g = 1$.

Equation (16.22) and $h'_3h_3 = \lim C'_nC_n = I_k$ yield

$$\tau^{-2}_{r_gF_n}B'_{n,r^\diamond_{g-1},p}U'_n\widehat{D}'_n\widehat{W}'_n\widehat{W}_n\widehat{D}_nU_nB_{n,r^\diamond_{g-1},p} \to_p \begin{bmatrix} h^{\diamond 2}_{6,r^\diamond_g} & 0^{(r^\diamond_g-r^\diamond_{g-1})\times(p-r^\diamond_g)} \\ 0^{(p-r^\diamond_g)\times(r^\diamond_g-r^\diamond_{g-1})} & 0^{(p-r^\diamond_g)\times(p-r^\diamond_g)} \end{bmatrix}. \tag{16.23}$$

By (16.21) and $o_{gp} = o_p(1)$, we have wp$\to$1 $\{(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn} : j = r^\diamond_{g-1} + 1, ..., p\}$ solve $|(I_{p-r^\diamond_{g-1}} + \widehat{A}_{jgn})^{-1}\tau^{-2}_{r_gF_n}B'_{n,r^\diamond_{g-1},p}U'_n\widehat{D}'_n\widehat{W}'_n\widehat{W}_n\widehat{D}_nU_nB_{n,r^\diamond_{g-1},p} + o_p(1) - \kappa I_{p-r^\diamond_{g-1}}| = 0$. Hence, by (16.23), $\widehat{A}_{jgn} = o_p(1)$ (which holds by the induction assumption), and the same argument as used to establish (16.6) and (16.7), we obtain

$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = r^\diamond_{g-1} + 1, ..., r^\diamond_g \text{ and } (n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn} \to_p 0 \ \forall j = r^\diamond_g + 1, ..., p. \tag{16.24}$$

Let o^*_{gp} denote an $(r^\diamond_g - r^\diamond_{g-1}) \times (p - r^\diamond_g)$ matrix whose elements in column j for $j = 1, ..., p - r^\diamond_g$ are $o_p(\tau_{(r^\diamond_g+j)F_n}/\tau_{r_gF_n}) + O_p((n^{1/2}\tau_{r_gF_n})^{-1})$. Note that $o^*_{gp} = o_p(1)$.

By (16.22) applied once with $B_{n,r^\diamond_g,p}$ in place of $B_{n,r^\diamond_{g-1},p}$ as the far-right multiplicand and applied a second time with $B_{n,r^\diamond_{g-1},r^\diamond_g}$ in place of $B_{n,r^\diamond_{g-1},p}$ as the far-right multiplicand, we have

$$\begin{aligned}
&\varrho_{gn} \\
:=\ & \tau^{-2}_{r_g F_n} B'_{n,r^\diamond_{g-1},r^\diamond_g} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_g,p} \\
=\ & \left[\begin{array}{c} 0^{r^\diamond_{g-1}\times(r^\diamond_g - r^\diamond_{g-1})} \\ Diag\{\tau_{(r^\diamond_{g-1}+1)F_n},...,\tau_{r^\diamond_g F_n}\}/\tau_{r_g F_n} \\ 0^{(k-r^\diamond_g)\times(r^\diamond_g - r^\diamond_{g-1})} \end{array}\right]' C'_n(I_k + o_p(1))C_n \left[\begin{array}{c} 0^{r^\diamond_g\times(p-r^\diamond_g)} \\ Diag\{\tau_{(r^\diamond_g+1)F_n},...,\tau_{pF_n}\}/\tau_{r_g F_n} \\ 0^{(k-p)\times(p-r^\diamond_g)} \end{array}\right] \\
& +O_p((n^{1/2}\tau_{r_g F_n})^{-1}) \\
=\ & o^*_{gp}, \qquad (16.25)
\end{aligned}$$

where $\varrho_{gn} \in R^{(r^\diamond_g - r^\diamond_{g-1})\times(p-r^\diamond_g)}$, $Diag\{\tau_{(r^\diamond_{g-1}+1)F_n},...,\tau_{r^\diamond_g F_n}\}/\tau_{r_g F_n} = h^\diamond_{6,r^\diamond_g} + o(1) = O(1)$ and the last equality holds because (i) $C'_n(I_k + o_p(1))C_n = I_k + o_p(1)$, (ii) when I_k appears in place of $C'_n(I_k + o_p(1))C_n$, then the contribution from the first summand on the lhs of the last equality in (16.25) equals $0^{(r^\diamond_g - r^\diamond_{g-1})\times(p-r^\diamond_g)}$, and (iii) when $o_p(1)$ appears in place of $C'_n(I_k + o_p(1))C_n$, the contribution from the first summand on the lhs of the last inequality in (16.25) equals an o^*_{gp} matrix.

We partition the $(p - r^\diamond_{g-1}) \times (p - r^\diamond_{g-1})$ matrices o_{gp} and $\widehat{A}_{jgn}$ as follows:

$$o_{gp} = \begin{pmatrix} o_{1gp} & o_{2gp} \\ o'_{2gp} & o_{3gp} \end{pmatrix} \text{ and } \widehat{A}_{jgn} = \begin{bmatrix} \widehat{A}_{1jgn} & \widehat{A}_{2jgn} \\ \widehat{A}'_{2jgn} & \widehat{A}_{3jgn} \end{bmatrix}, \qquad (16.26)$$

where $o_{1gp}, \widehat{A}_{1jgn} \in R^{(r^\diamond_g - r^\diamond_{g-1})\times(r^\diamond_g - r^\diamond_{g-1})}$, $o_{2gp}, \widehat{A}_{2jgn} \in R^{(r^\diamond_g - r^\diamond_{g-1})\times(p-r^\diamond_g)}$, and $o_{3gp}, \widehat{A}_{3jgn} \in R^{(p-r^\diamond_g)\times(p-r^\diamond_g)}$, for $j = r^\diamond_{g-1}+1, ..., p$ and $g = 1, ..., G_h$. Define

$$\begin{aligned}
\widehat{\xi}_{1jgn}(\kappa) &:= \tau^{-2}_{rg} B'_{n,r^\diamond_{g-1},r^\diamond_g} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_{g-1},r^\diamond_g} + o_{1gp} - \kappa(I_{r^\diamond_g - r^\diamond_{g-1}} + \widehat{A}_{1jgn}), \\
\widehat{\xi}_{2jgn}(\kappa) &:= \varrho_{gn} + o_{2gp} - \kappa \widehat{A}_{2jgn}, \text{ and} \qquad (16.27) \\
\widehat{\xi}_{3jgn}(\kappa) &:= \tau^{-2}_{r_g F_n} B'_{n,r^\diamond_g,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_g,p} + o_{3gp} - \kappa(I_{p-r^\diamond_g} + \widehat{A}_{3jgn}),
\end{aligned}$$

where $\widehat{\xi}_{1jgn}(\kappa)$, $\widehat{\xi}_{2jgn}(\kappa)$, and $\widehat{\xi}_{3jgn}(\kappa)$ have the same dimensions as o_{1gp}, o_{2gp}, and o_{3gp}, respectively.

From (16.21), we have wp→1 $\{(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn} : j = r_{g-1}^{\diamond} + 1, ..., p\}$ solve

$$\begin{aligned}
0 &= |\tau_{r_gF_n}^{-2} B'_{n,r_{g-1}^{\diamond},p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_{g-1}^{\diamond},p} + o_{gp} - \kappa(I_{p-r_{g-1}^{\diamond}} + \widehat{A}_{jgn})| \\
&= |\widehat{\xi}_{1jgn}(\kappa)| \cdot |\widehat{\xi}_{3jgn}(\kappa) - \widehat{\xi}_{2jgn}(\kappa)'\widehat{\xi}_{1jgn}^{-1}(\kappa)\widehat{\xi}_{2jgn}(\kappa)| \\
&= |\widehat{\xi}_{1jgn}(\kappa)| \cdot |\tau_{r_gF_n}^{-2} B'_{n,r_g^{\diamond},p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^{\diamond},p} + o_{3gp} - (\varrho_{gn} + o_{2gp})'\widehat{\xi}_{1jgn}^{-1}(\kappa)(\varrho_{gn} + o_{2gp}) \\
&\quad -\kappa[I_{p-r_g^{\diamond}} + \widehat{A}_{3jgn} - \widehat{A}'_{2jgn}\widehat{\xi}_{1jgn}^{-1}(\kappa)(\varrho_{gn} + o_{2gp}) - (\varrho_{gn} + o_{2gp})'\widehat{\xi}_{1jgn}^{-1}(\kappa)\widehat{A}_{2jgn} \\
&\qquad +\kappa\widehat{A}'_{2jgn}\widehat{\xi}_{1jgn}^{-1}(\kappa)\widehat{A}_{2jgn}]|,
\end{aligned} \tag{16.28}$$

where the second equality holds by the same argument as for (16.10) and uses the result given in (16.29) below which shows that $\widehat{\xi}_{1jgn}(\kappa)$ is nonsingular wp→1 when κ equals $(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}$ for $j = r_g^{\diamond} + 1, ..., p$.

Now we show that, for $j = r_g^{\diamond}+1, ..., p$, $(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}$ cannot solve the determinantal equation $|\widehat{\xi}_{1jgn}(\kappa)| = 0$ for n large, where this determinant is the first multiplicand on the rhs of (16.28) and, hence, it must solve the determinantal equation based on the second multiplicand on the rhs of (16.28). For $j = r_g^{\diamond} + 1, ..., p$, we have

$$\widetilde{\xi}_{1jgn} := \widehat{\xi}_{1jgn}((n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}) = h_{6,r_g^{\diamond}}^{\diamond 2} + o_p(1), \tag{16.29}$$

by the same argument as in (16.11), using $o_{1gp} = o_p(1)$ and $\widehat{A}_{1jgn} = o_p(1)$ (which holds by the definition of $\widehat{A}_{1jgn}$ following (16.21)). Equation (16.29) and $\lambda_{\min}(h_{6,r_g^{\diamond}}^{\diamond 2}) > 0$ establish the result stated in the first sentence of this paragraph.

For $j = r_g^{\diamond}+1, ..., p$, plugging $(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}$ into the second multiplicand on the rhs of (16.28) gives

$$\begin{aligned}
0 &= |\tau_{r_gF_n}^{-2} B'_{n,r_g^{\diamond},p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^{\diamond},p} + o_{3gp} - (\varrho_{gn} + o_{2gp})'\widetilde{\xi}_{1jgn}^{-1}(\varrho_{gn} + o_{2gp}) \\
&\quad -(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r_g^{\diamond}} + \widehat{A}_{j(g+1)n})|, \text{ where} \\
\widehat{A}_{j(g+1)n} &:= \widehat{A}_{3jgn} - \widehat{A}'_{2jgn}\widetilde{\xi}_{1jgn}^{-1}(\varrho_{gn} + o_{2gp}) - (\varrho_{gn} + o_{2gp})'\widetilde{\xi}_{1jgn}^{-1}\widehat{A}_{2jgn} \\
&\quad +(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn}\widehat{A}'_{2jgn}\widetilde{\xi}_{1jgn}^{-1}\widehat{A}_{2jgn}
\end{aligned} \tag{16.30}$$

and $\widehat{A}_{j(g+1)n} \in R^{(p-r_g^{\diamond})\times(p-r_g^{\diamond})}$. The last two summands on the rhs of the first line of (16.30) satisfy

$$\begin{aligned}
& o_{3gp} - (\varrho_{gn} + o_{2gp})'\widetilde{\xi}_{1jgn}^{-1}(\varrho_{gn} + o_{2gp}) = o_{3gp} - (o_{gp}^{*} + o_{2gp})'(h_{6,r_g^{\diamond}}^{\diamond -2} + o_p(1))(o_{gp}^{*} + o_{2gp}) \\
&= o_{3gp} - o_{gp}^{*\prime}o_{gp}^{*} = (\tau_{r_{g+1}F_n}^2/\tau_{r_gF_n}^2)o_{(g+1)p},
\end{aligned} \tag{16.31}$$

where (i) the first equality uses (16.25) and (16.29), (ii) the second equality uses $o_{2gp} = o^*_{gp}$ (which holds because the (j,m) element of o_{2gp} for $j = 1, ..., r^\diamond_g - r^\diamond_{g-1}$ and $m = 1, ..., p - r^\diamond_g$ is $o_p(\tau_{(r^\diamond_{g-1}+j)F_n} \times \tau_{(r^\diamond_g+m)F_n}/\tau^2_{r_gF_n}) + O_p((n^{1/2}\tau_{r_gF_n})^{-1}) = o_p(\tau_{(r^\diamond_g+m)F_n}/\tau_{r_gF_n}) + O_p((n^{1/2}\tau_{r_gF_n})^{-1})$ since $r^\diamond_{g-1} + j \geq r_g$) and $(h^{\diamond -2}_{6,r^\diamond_g} + o_p(1))o^*_{gp} = o^*_{gp}$ (which holds because $h^\diamond_{6,r^\diamond_g}$ is diagonal and $\lambda_{\min}(h^{\diamond 2}_{6,r^\diamond_g}) > 0$), (iii) the last equality uses the fact that the (j,m) element of $(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n})o^{*\prime}_{gp}o^*_{gp}$ for $j, m = 1, ..., p - r^\diamond_g$ is the sum of a term that is $o_p(\tau_{(r^\diamond_g+j)F_n}\tau_{(r^\diamond_g+m)F_n}/\tau^2_{r_gF_n})(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}) = o_p(\tau_{(r^\diamond_g+j)F_n}\tau_{(r^\diamond_g+m)F_n} /\tau^2_{r_{g+1}F_n})$ and a term that is $O_p((n^{1/2}\tau_{r_gF_n})^{-2})(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}) = O_p((n^{1/2}\tau_{r_{g+1}F_n})^{-2})$ and, hence, $(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n})o^{*\prime}_{gp}o^*_{gp}$ is $o_{(g+1)p}$ (using the definition of $o_{(g+1)p}$), and (iv) the last equality uses the fact that the (j,m) element of $(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n})o_{3gp}$ for $j, m = 1, ..., p - r^\diamond_g$ is $o_p(\tau_{(r^\diamond_g+j)F_n}\tau_{(r^\diamond_g+m)F_n} /\tau^2_{r_gF_n})(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}) + O_p((n^{1/2}\tau_{r_gF_n})^{-1})(\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}) = o_p(\tau_{(r^\diamond_g+j)F_n}\tau_{(r^\diamond_g+m)F_n}/\tau^2_{r_{g+1}F_n}) + O_p((n^{1/2}\tau_{r_{g+1}F_n})^{-1})(\tau_{r_gF_n}/\tau_{r_{g+1}F_n})$, which again is the same order as the (j,m) element of $o_{(g+1)p}$ (using $\tau_{r_gF_n}/\tau_{r_{g+1}F_n} \leq 1$).

The calculations in (16.31) are a key part of the induction proof. The definitions of the terms o_{gp} and o^*_{gp} (given preceding (16.21) and (16.25), respectively) are chosen so that the results in (16.31) hold.

For $j = r^\diamond_g + 1, ..., p$, we have

$$\widehat{A}_{j(g+1)n} = o_p(1), \tag{16.32}$$

because $\widehat{A}_{2jgn} = o_p(1)$ and $\widehat{A}_{3jgn} = o_p(1)$ by (16.21), $\widetilde{\xi}^{-1}_{1jgn} = O_p(1)$ (by (16.29)), $\varrho_{gn} + o_{2gp} = o_p(1)$ (by (16.25) since $o^*_{gp} = o_p(1)$), and $(n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn} = o_p(1)$ (by (16.24)).

Inserting (16.31) and (16.32) into (16.30) and multiplying by $\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}$ gives

$$0 = |\tau^{-2}_{r_{g+1}F_n} B'_{n,r^\diamond_g,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_g,p} + o_{(g+1)p} - (n^{1/2}\tau_{r_{g+1}F_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r^\diamond_g} + \widehat{A}_{j(g+1)n})|. \tag{16.33}$$

Thus, wp$\rightarrow$1, $\{(n^{1/2}\tau_{r_{g+1}F_n})^{-2}\widehat{\kappa}_{jn} : j = r_{g+1}, ..., p\}$ solve

$$0 = |\tau^{-2}_{r_{g+1}F_n} B'_{n,r^\diamond_g,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_g,p} + o_{(g+1)p} - \kappa(I_{p-r^\diamond_g} + \widehat{A}_{j(g+1)n})|. \tag{16.34}$$

This establishes the induction step and concludes the proof that (16.21) holds for all $g = 1, ..., G_h$.

Finally, given that (16.21) holds for all $g = 1, ..., G_h$, (16.24) gives the results stated in (16.18) and (16.18) gives the results stated in the Lemma by the argument in (16.18)-(16.20). $\square$

Now we use the approach in Johansen (1991, pp. 1569-1571) and Robin and Smith (2000, pp. 172-173) to prove Theorem 8.4. In these papers, asymptotic results are established under a fixed true distribution under which certain population eigenvalues are either positive or zero. Here we need to deal with drifting sequences of distributions under which these population eigenvalues may be positive or zero for any given n, but the positive ones may drift to zero as $n \to \infty$, possibly at different rates. This complicates the proof. In particular, the rate of convergence result of Lemma 16.1(b) is needed in the present context, but not in the fixed distribution scenario considered in Johansen (1991) and Robin and Smith (2000).

Proof of Theorem 8.4. Theorem 8.4(a) and (c) follow immediately from Lemma 16.1(a).

Next, we assume $q < p$ and we prove part (b). The eigenvalues $\{\widehat{\kappa}_{jn} : j \leq p\}$ of $n\widehat{U}_n\widehat{D}_n'\widehat{W}_n'\widehat{W}_n \times \widehat{D}_n\widehat{U}_n$ are the ordered solutions to the determinantal equation $|n\widehat{U}_n\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n - \kappa I_p| = 0$. Equivalently, with probability that goes to one (wp$\to$1), they are the solutions to

$$|Q_n^{\diamond}(\kappa)| = 0, \text{ where } Q_n^{\diamond}(\kappa) := nS_nB_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_nS_n - \kappa S_n'B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_nS_n, \tag{16.35}$$

because $|S_n| > 0$, $|B_n| > 0$, $|U_n| > 0$, and $|\widehat{U}_n| > 0$ wp$\to$1. Thus, $\lambda_{\min}(n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n)$ equals the smallest solution, $\widehat{\kappa}_{pn}$, to $|Q_n^{\diamond}(\kappa)| = 0$ wp$\to$1. (For simplicity, we omit the qualifier wp$\to$1 that applies to several statements below.)

We write $Q_n^{\diamond}(\kappa)$ in partitioned form using

$$\begin{aligned} B_nS_n &= (B_{n,q}S_{n,q}, B_{n,p-q}), \text{ where} \\ S_{n,q} &:= Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}\} \in R^{q \times q}. \end{aligned} \tag{16.36}$$

The convergence result of Lemma 8.3 for $n^{1/2}W_n\widehat{D}_nU_nT_n$ $(= n^{1/2}W_n\widehat{D}_nU_nB_nS_n)$ can be written as

$$n^{1/2}W_n\widehat{D}_nU_nB_{n,q}S_{n,q} \to_p \overline{\Delta}_{h,q} := h_{3,q} \text{ and } n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} \to_d \overline{\Delta}_{h,p-q}, \tag{16.37}$$

where $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are defined in (8.17).

We have

$$\widehat{W}_nW_n^{-1} \to_p I_k \text{ and } \widehat{U}_nU_n^{-1} \to_p I_p \tag{16.38}$$

because $\widehat{W}_n \to_p h_{71} := \lim W_n$ (by Assumption WU(a) and (c)), $\widehat{U}_n \to_p h_{81} := \lim U_n$ (by Assumption WU(b) and (c)), and h_{71} and h_{81} are pd (by the conditions in $\mathcal{F}_{WU}$).

By (16.35)-(16.38), we have

$$
\begin{aligned}
Q_n^{\diamond}(\kappa) &= \left[\begin{array}{cc} I_q + o_p(1) & h_{3,q}' n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) \\ n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' h_{3,q} + o_p(1) & n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' W_n n^{1/2} \widehat{D}_n U_n B_{n,p-q} + o_p(1) \end{array}\right] \\
&\quad -\kappa \left[\begin{array}{cc} S_{n,q}^2 & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & I_{p-q} \end{array}\right] - \kappa \left[\begin{array}{cc} S_{n,q} A_{1n} S_{n,q} & S_{n,q} A_{2n} \\ A_{2n}' S_{n,q} & A_{3n} \end{array}\right], \text{ where} \qquad (16.39) \\
\widehat{A}_n &= \left[\begin{array}{cc} A_{1n} & A_{2n} \\ A_{2n}' & A_{3n} \end{array}\right] := B_n' U_n' \widehat{U}_n^{-1\prime} \widehat{U}_n^{-1} U_n B_n - I_p = o_p(1) \text{ for } A_{1n} \in R^{q\times q},\ A_{2n} \in R^{q\times(p-q)},
\end{aligned}
$$

and $A_{3n} \in R^{(p-q)\times(p-q)}$, $\widehat{A}_n$ is defined in (16.39) just as in (16.5), and the first equality uses $\overline{\Delta}_{h,q} := h_{3,q}$ and $\overline{\Delta}_{h,q}' \overline{\Delta}_{h,q} = h_{3,q}' h_{3,q} = \lim C_{n,q}' C_{n,q} = I_q$ (by (8.7), (8.9), (8.12), and (8.17)). Note that A_{jn} and $\widehat{A}_{jn}$ (defined in (16.2)) are not the same in general for $j = 1, 2, 3$, because their dimensions differ. For example, $A_{1n} \in R^{q\times q}$, whereas $\widehat{A}_{1n} \in R^{r_1^{\diamond}\times r_1^{\diamond}}$.

If $q = 0$ $(< p)$, then $B_n = B_{n,p-q}$ and

$$
\begin{aligned}
& n B_n' \widehat{U}_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n B_n \\
= \ & n B_n' (U_n^{-1} \widehat{U}_n)' B_n^{-1\prime} B_n' U_n' \widehat{D}_n' W_n' \left(\widehat{W}_n W_n^{-1}\right)' \left(\widehat{W}_n W_n^{-1}\right) (W_n \widehat{D}_n U_n B_n) B_n^{-1} (U_n^{-1} \widehat{U}_n) B_n \\
\to_d \ & \overline{\Delta}_{h,p-q}' \overline{\Delta}_{h,p-q}, \qquad (16.40)
\end{aligned}
$$

where the convergence holds by (16.37) and (16.38) and $\overline{\Delta}_{h,p-q}$ is defined as in (8.17) with $q = 0$. The smallest eigenvalue of a matrix is a continuous function of the matrix (by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). Hence, the smallest eigenvalue of $n B_n' \widehat{U}_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n B_n$ converges in distribution to the smallest eigenvalue of $\overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \overline{\Delta}_{h,p-q}$ (using $h_{3,k-q} h_{3,k-q}' = h_3 h_3' = I_k$ when $q = 0$), which proves part (b) of Theorem 8.4 when $q = 0$.

In the remainder of the proof of part (b), we assume $1 \le q < p$, which is the remaining case to be considered in the proof of part (b). The formula for the determinant of a partitioned matrix and (16.39) give

$$
\begin{aligned}
|Q_n^{\diamond}(\kappa)| &= |Q_{1n}^{\diamond}(\kappa)| \cdot |Q_{2n}^{\diamond}(\kappa)|, \text{ where} \\
Q_{1n}^{\diamond}(\kappa) &:= I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n} S_{n,q}, \\
Q_{2n}^{\diamond}(\kappa) &:= n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' W_n n^{1/2} \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \kappa I_{p-q} - \kappa A_{3n} \\
&\quad -[n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' h_{3,q} + o_p(1) - \kappa A_{2n}' S_{n,q}](I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n} S_{n,q})^{-1} \\
&\quad \times [h_{3,q}' n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \kappa S_{n,q} A_{2n}], \qquad (16.41)
\end{aligned}
$$

none of the $o_p(1)$ terms depend on κ, and the equation in the first line holds provided $Q^{\diamond}_{1n}(\kappa)$ is nonsingular.

By Lemma 16.1(b) (which applies for $1 \leq q < p$), for $j = q+1, ..., p$, we have $\widehat{\kappa}_{jn} S^2_{n,q} = o_p(1)$ and $\widehat{\kappa}_{jn} S_{n,q} A_{1n} S_{n,q} = o_p(1)$. Thus,

$$Q^{\diamond}_{1n}(\widehat{\kappa}_{jn}) = I_q + o_p(1) - \widehat{\kappa}_{jn} S^2_{n,q} - \widehat{\kappa}_{jn} S_{n,q} A_{1n} S_{n,q} = I_q + o_p(1). \tag{16.42}$$

By (16.35) and (16.41), $|Q^{\diamond}_n(\widehat{\kappa}_{jn})| = |Q^{\diamond}_{1n}(\widehat{\kappa}_{jn})| \cdot |Q^{\diamond}_{2n}(\widehat{\kappa}_{jn})| = 0$ for $j = 1, ..., p$. By (16.42), $|Q^{\diamond}_{1n}(\widehat{\kappa}_{jn})| \neq 0$ for $j = q+1, ..., p$ wp$\to$1. Hence, wp$\to$1,

$$|Q^{\diamond}_{2n}(\widehat{\kappa}_{jn})| = 0 \text{ for } j = q+1, ..., p. \tag{16.43}$$

Now we plug in $\widehat{\kappa}_{jn}$ for $j = q+1, ..., p$ into $Q^{\diamond}_{2n}(\kappa)$ in (16.41) and use (16.42). We have

$$\begin{aligned} Q^{\diamond}_{2n}(\widehat{\kappa}_{jn}) = \; & n B'_{n,p-q} U'_n \widehat{D}'_n W'_n W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) \\ & -[n^{1/2} B'_{n,p-q} U'_n \widehat{D}'_n W'_n h_{3,q} + o_p(1)](I_q + o_p(1))[h'_{3,q} n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1)] \\ & -\widehat{\kappa}_{jn}[I_{p-q} + A_{3n} - (n^{1/2} B'_{n,p-q} U'_n \widehat{D}'_n W'_n h_{3,q} + o_p(1))(I_q + o_p(1)) S_{n,q} A_{2n} \\ & -A'_{2n} S_{n,q}(I_q + o_p(1))(h'_{3,q} n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1)) \\ & +\widehat{\kappa}_{jn} A'_{2n} S_{n,q}(I_q + o_p(1)) S_{n,q} A_{2n}]. \end{aligned} \tag{16.44}$$

The term in square brackets on the last three lines of (16.44) that multiplies $\widehat{\kappa}_{jn}$ equals

$$I_{p-q} + o_p(1), \tag{16.45}$$

because $A_{3n} = o_p(1)$ (by (16.39)), $n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} = O_p(1)$ (by (16.37)), $S_{n,q} = o(1)$ (by the definitions of q and $S_{n,q}$ in (8.16) and (16.36), respectively, and $h_{1,j} := \lim n^{1/2} \tau_{jF_n}$), $A_{2n} = o_p(1)$ (by (16.39)), and $\widehat{\kappa}_{jn} A'_{2n} S_{n,q}(I_q + o_p(1)) S_{n,q} A_{2n} = A'_{2n} \widehat{\kappa}_{jn} S^2_{n,q} A_{2n} + A'_{2n} \widehat{\kappa}_{jn} S_{n,q} o_p(1) S_{n,q} A_{2n} = o_p(1)$ (using $\widehat{\kappa}_{jn} S^2_{n,q} = o_p(1)$ and $A_{2n} = o_p(1)$).

Equations (16.44) and (16.45) give

$$\begin{aligned} Q^{\diamond}_{2n}(\widehat{\kappa}_{jn}) &= n^{1/2} B'_{n,p-q} U'_n \widehat{D}'_n W'_n [I_k - h_{3,q} h'_{3,q}] n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)] \\ &= n^{1/2} B'_{n,p-q} U'_n \widehat{D}'_n W'_n h_{3,k-q} h'_{3,k-q} n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)] \\ &:= M_{n,p-q} - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)], \end{aligned} \tag{16.46}$$

where the second equality uses $I_k = h_3 h'_3 = h_{3,q} h'_{3,q} + h_{3,k-q} h'_{3,k-q}$ (because $h_3 = \lim C_n$ is an orthogonal matrix) and the last line defines the $(p-q)\times(p-q)$ matrix $M_{n,p-q}$.

Equations (16.43) and (16.46) imply that $\{\widehat{\kappa}_{jn}: j=q+1,...,p\}$ are the $p-q$ eigenvalues of the matrix

$$M_{n,p-q}^{\diamond} := [I_{p-q}+o_p(1)]^{-1/2} M_{n,p-q} [I_{p-q}+o_p(1)]^{-1/2} \tag{16.47}$$

by pre- and post-multiplying the quantities in (16.46) by the rhs quantity $[I_{p-q}+o_p(1)]^{-1/2}$ in (16.46). By (16.37),

$$M_{n,p-q}^{\diamond} \to_d \overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \overline{\Delta}_{h,p-q}. \tag{16.48}$$

The vector of (ordered) eigenvalues of a matrix is a continuous function of the matrix (by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). By (16.48), the matrix $M_{n,p-q}^{\diamond}$ converges in distribution. In consequence, by the CMT, the vector of eigenvalues of $M_{n,p-q}^{\diamond}$, viz., $\{\widehat{\kappa}_{jn}: j=q+1,...,p\}$, converges in distribution to the vector of eigenvalues of the limit matrix $\overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \overline{\Delta}_{h,p-q}$, which proves part (d) of Theorem 8.4. In addition, $\lambda_{\min}(n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n' \times \widehat{W}_n\widehat{D}_n\widehat{U}_n)$, which equals the smallest eigenvalue, $\widehat{\kappa}_{pn}$, converges in distribution to the smallest eigenvalue of $\overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \overline{\Delta}_{h,p-q}$, which completes the proof of part (b) of Theorem 8.4.

The convergence in parts (a)-(d) of Theorem 8.4 is joint with that in Lemma 8.3 because it just relies on the convergence in distribution of $n^{1/2}W_n\widehat{D}_nU_nT_n$, which is part of the former. This establishes part (e) of Theorem 8.4.

Part (f) of Theorem 8.4 holds by the same proof as used for parts (a)-(e) with n replaced by w_n. $\square$

17 Proofs of Sufficiency of Several Conditions for the $\boldsymbol{\lambda_{p-j}}(\cdot)$ Condition in $\mathcal{F}_{\mathbf{0j}}$

In this section, we show that the conditions in (3.9) and (3.10) are sufficient for the second condition in $\mathcal{F}_{0j}$, which is $\lambda_{p-j}(\Psi_F^{C_{F,k-j}'\Omega_F^{-1/2}G_iB_{F,p-j}\xi}) \geq \delta_1\ \forall \xi \in R^{p-j}$ with $||\xi|| = 1$.

Condition (i) in (3.9) is sufficient by the following argument:

$$
\begin{aligned}
&\lambda_{p-j}\left(\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right)\\
\geq\ &\lambda_{p-j}\left(\Psi_F^{\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right)\\
=\ &\lambda_{\min}\left((\xi'\otimes I_{p-j})\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}(\xi\otimes I_{p-j})\right)\\
=\ &\min_{\lambda\in R^{p-j}:||\lambda||=1}\left(\frac{(\xi\otimes I_{p-j})\lambda}{||(\xi\otimes I_{p-j})\lambda||}\right)'\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\frac{(\xi\otimes I_{p-j})\lambda}{||(\xi\otimes I_{p-j})\lambda||}\times||(\xi\otimes I_{p-j})\lambda||^2\\
\geq\ &\min_{\eta\in R^{(p-j)^2}:||\eta||=1}\eta'\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\eta\times\min_{\lambda\in R^{p-j}:||\lambda||=1}||(\xi\otimes I_{p-j})\lambda||^2\\
=\ &\lambda_{\min}\left(\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\right),
\end{aligned}
\tag{17.1}
$$

where the first inequality holds by Corollary 15.4(a) with $m=p-j$ and $r=k-p$ (because $\Psi_F^{\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}$ is a submatrix of $\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}$, since $\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}=C'_{F,k-j}\Psi_F^{\Omega_F^{-1/2}G_iB_{F,p-j}\xi}C_{F,k-j}$, likewise with $C'_{F,k-j}$ replaced by $\overline{C}'_{F,p-j}$, and by definition the rows of $\overline{C}'_{F,p-j}$ are a collection of $p-j$ rows of $C'_{F,k-j}$), the first equality holds because the $(p-j)$-th largest eigenvalue of a $(p-j)\times(p-j)$ matrix equals its minimum eigenvalue and by the general formula $vec(ABC)=(C'\otimes A)vec(B)$, and the last equality holds because $||(\xi\otimes I_{p-j})\lambda||^2=\lambda'(\xi'\xi\otimes I_{p-j})\lambda=\lambda'\lambda=1$ using $||\xi||=||\lambda||=1$.

Condition (ii) in (3.9) is sufficient by sufficient condition (i) in (3.9) and the following:

$$
\begin{aligned}
&\lambda_{\min}\left(\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\right)\\
=\ &\min_{\eta\in R^{(p-j)^2}:||\eta||=1}\left(\frac{(I_{p-j}\otimes\overline{C}_{F,p-j})\eta}{||(I_{p-j}\otimes\overline{C}_{F,p-j})\eta||}\right)'\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\frac{(I_{p-j}\otimes\overline{C}_{F,p-j})\eta}{||(I_{p-j}\otimes\overline{C}_{F,p-j})\eta||}\\
&\times||(I_{p-j}\otimes\overline{C}_{F,p-j})\eta||^2\\
\geq\ &\min_{\zeta\in R^{(p-j)k}:||\zeta||=1}\zeta'\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\zeta\times\min_{\eta\in R^{(p-j)^2}:||\eta||=1}||(I_{p-j}\otimes\overline{C}_{F,p-j})\eta||^2\\
=\ &\lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\right),
\end{aligned}
\tag{17.2}
$$

where the last equality uses $||(I_{p-j}\otimes\overline{C}_{F,p-j})\eta||^2=\eta'(I_{p-j}\otimes\overline{C}'_{F,p-j}\overline{C}_{F,p-j})\eta=1$ because the rows of $\overline{C}'_{F,p-j}$ are orthonormal and $||\eta||=1$.

Condition (iii) in (3.9) is sufficient by sufficient condition (ii) in (3.9) and a similar argument to that given in (17.2) using the fact that $\min_{\psi\in R^{pk}:||\psi||=1}||(B'_{F,p-j}\otimes I_k)\psi||^2=1$ because the columns of $B_{F,p-j}$ are orthonormal.

Condition (iv) in (3.9) is sufficient by sufficient condition (iii) in (3.9) and a similar argument to that given in (17.2) using $\min_{\phi \in R^{pk}:||\phi||=1} ||(I_p \otimes \Omega_F^{-1/2})\phi||^2 \geq M^{-2/(2+\gamma)}$ for M as in the definition of $\mathcal{F}$ in place of $\min_{\eta \in R^{(p-j)^2}:||\eta||=1} ||(I_{p-j} \otimes \overline{C}_{F,p-j})\eta||^2 = 1$. The latter inequality holds by the following calculations:

$$
\begin{aligned}
\phi'(I_p \otimes \Omega_F^{-1})\phi &= \sum_{j=1}^{p} (\phi_j/||\phi_j||)'\Omega_F^{-1}(\phi_j/||\phi_j||) \times ||\phi_j||^2 \\
&\geq \sum_{j=1}^{p} \lambda_{\min}(\Omega_F^{-1}) \times ||\phi_j||^2 = 1/\lambda_{\max}(\Omega_F) \geq M^{-2/(2+\gamma)},
\end{aligned} \tag{17.3}
$$

where $\phi = (\phi_1', ..., \phi_p')'$ for $\phi_j \in R^k \ \forall j \leq p$, the sums are over j for which $\phi_j \neq 0^k$, the second equality uses $||\phi|| = 1$, and the last inequality holds because $\lambda_{\max}(\Omega_F) = \max_{\lambda \in R^k:||\lambda||=1} E_F(\lambda' g_i)^2 \leq E_F||g_i||^2 = ((E_F||g_i||^2)^{1/2})^2 \leq ((E_F||g_i||^{2+\gamma})^{1/(2+\gamma)})^2 \leq M^{2/(2+\gamma)}$ by successively applying the Cauchy-Bunyakovsky-Schwarz inequality, Lyapunov's inequality, and the moment bound $E_F||g_i||^{2+\gamma} \leq M$ in $\mathcal{F}$.

Conditions (v) and (vi) in (3.9) are sufficient by the following argument. Write

$$
\Psi_F^{vec(G_i)} = (M_F, I_{pk})\Sigma_F^{f_i}(M_F, I_{pk})', \text{ where } M_F = -(E_F vec(G_i)g_i')(E_F g_i g_i')^{-1} \in R^{pk \times k}. \tag{17.4}
$$

We have

$$
\begin{aligned}
\lambda_{\min}(\Psi_F^{vec(G_i)}) &= \min_{\lambda \in R^{pk}:||\lambda||=1} \lambda'(M_F, I_{pk})\Sigma_F^{f_i}(M_F, I_{pk})'\lambda \\
&= \min_{\lambda \in R^{pk}:||\lambda||=1} \left(\frac{(M_F, I_{pk})'\lambda}{||(M_F, I_{pk})'\lambda||}\right)' \Sigma_F^{f_i} \left(\frac{(M_F, I_{pk})'\lambda}{||(M_F, I_{pk})'\lambda||}\right) \times ||(M_F, I_{pk})'\lambda||^2 \\
&\geq \min_{\mu \in R^{(p+1)k}:||\mu||=1} \mu'\Sigma_F^{f_i}\mu \\
&= \lambda_{\min}(\Sigma_F^{f_i}),
\end{aligned} \tag{17.5}
$$

where the inequality uses $||(M_F, I_{pk})'\lambda||^2 = \lambda'\lambda + \lambda' M_F' M_F \lambda \geq 1$ for $\lambda \in R^{pk}$ with $||\lambda|| = 1$. This shows that condition (v) is sufficient for sufficient condition (iv) in (3.9). Since $\Sigma_F^{f_i} = Var_F(f_i) + E_F f_i E_F f_i'$, condition (vi) is sufficient for sufficient condition (v) in (3.9).

The condition in (3.10) is sufficient by the following argument:

$$
\lambda_{p-j}\left(\Psi_F^{C_{F,k-j}'\Omega_F^{-1/2}G_i B_{F,p-j}\xi}\right) \geq \lambda_p\left(\Psi_F^{C_F'\Omega_F^{-1/2}G_i B_{F,p-j}\xi}\right) = \lambda_p\left(\Psi_F^{\Omega_F^{-1/2}G_i B_{F,p-j}\xi}\right), \tag{17.6}
$$

where the first inequality holds by Corollary 15.4(b) with $m = p$ and $r = j$ and the equality holds because $\Psi_F^{C_F' \Omega_F^{-1/2} G_i B_{F,p-j}\xi} = C_F' \Psi_F^{\Omega_F^{-1/2} G_i B_{F,p-j}\xi} C_F$ and C_F is orthogonal.

18 Asymptotic Size of Kleibergen's CLR Test with Jacobian-Variance Weighting and the Proof of Theorem 5.1

In this section, we establish the asymptotic size of Kleibergen's CLR test with Jacobian-variance weighting when the Robin and Smith (2000) rank statistic (defined in (5.5)) is employed. This rank statistic depends on a variance matrix estimator $\widetilde{V}_{Dn}$. See Section 5 for the definition of the test. We provide a formula for the asymptotic size of the test that depends on the specifics of the moment conditions considered and does not necessarily equal its nominal size α. First, in Section 18.1, we provide an example that illustrates the results in Theorem 5.1 and Comment (v) to Theorem 5.1. In Section 18.2, we establish the asymptotic size of the test based on $\widetilde{V}_{Dn}$ defined as in (5.3). In Section 18.3, we report some simulation results for a linear instrumental variable (IV) model with two rhs endogenous variables. In Section 18.4, we establish the asymptotic size of Kleibergen's CLR test with Jacobian-variance weighting under a general assumption that allows for other definitions of $\widetilde{V}_{Dn}$.

In Section 18.5, we show that equally-weighted versions of Kleibergen's CLR test have correct asymptotic size when the Robin and Smith (2000) rank statistic is employed and a general equal-weighting matrix $\widetilde{W}_n$ is employed. This result extends the result given in Theorem 6.1 in Section 6, which applies to the specific case where $\widetilde{W}_n = \widehat{\Omega}_n^{-1/2}$, as in (6.2). The results of Section 18.5 are a relatively simple by-product of the results in Section 18.4.

Proofs of the results stated in this section are given in Section 18.6.

Theorem 5.1 follows from Lemma 18.2 and Theorem 18.3, which are stated in Section 18.4.

18.1 An Example

Here we provide a simple example that illustrates the result of Theorem 5.1. In this example, the true distribution F does not depend on n. Suppose $p = 2$, $E_F G_i = (1^k, 0^k)$, where $c^k = (c, ..., c)' \in R^k$ for $c = 0, 1$, $n^{1/2}(\widehat{D}_n - E_F G_i) \to_d \overline{D}_h$ under F for some random matrix $\overline{D}_h = (\overline{D}_{1h}, \overline{D}_{2h}) \in R^{k\times 2}$. Suppose for $\widetilde{M}_n = \widetilde{V}_{Dn}^{-1/2}$ and $M_F = I_{2k}$, we have $n^{1/2}(\widetilde{M}_n - M_F) \to_d \overline{M}_h$ under F for some random matrix $\overline{M}_h \in R^{2k\times 2k}$.[53] We have

$$\widehat{D}_n^{\dagger} = vec_{k,p}^{-1}(\widetilde{V}_{Dn}^{-1/2} vec(\widehat{D}_n)) = \left(\widetilde{M}_{11n}\widehat{D}_{1n} + \widetilde{M}_{12n}\widehat{D}_{2n}, \widetilde{M}_{21n}\widehat{D}_{1n} + \widetilde{M}_{22n}\widehat{D}_{2n}\right), \qquad (18.1)$$

[53] The convergence results $n^{1/2}(\widehat{D}_n - E_F G_i) \to_d \overline{D}_h$ and $n^{1/2}(\widetilde{M}_n - M_F) \to_d \overline{M}_h$ are established in Lemmas 8.2 and 18.2, respectively, in Section 8 of AG1 and Section 18 in this Supplemental Material under general conditions.

where $\widehat{D}_n = (\widehat{D}_{1n}, \widehat{D}_{2n})$, $\widetilde{M}_{j\ell n}$ for $j, \ell = 1, 2$ are the four $k \times k$ submatrices of $\widetilde{M}_n$, and likewise for $M_{j\ell F}$ for $j, \ell = 1, 2$. Let $\overline{M}_{j\ell h}$ for $j, \ell = 1, 2$ denote the four $k \times k$ submatrices of $\overline{M}_h$. We let $T_n^\dagger = Diag\{n^{-1/2}, 1\}$. Then, we have

$$\begin{aligned} n^{1/2}\widehat{D}_n^\dagger T_n^\dagger &= \left(\widetilde{M}_{11n}\widehat{D}_{1n} + \widetilde{M}_{12n}\widehat{D}_{2n},\ n^{1/2}\widetilde{M}_{21n}\widehat{D}_{1n} + \widetilde{M}_{22n}n^{1/2}\widehat{D}_{2n}\right) \\ &\to_d \left(I_k 1^k + 0^{k\times k}0^k,\ \overline{M}_{21h}1^k + I_k\overline{D}_{2h}\right) = \left(1^k, \overline{M}_{21h}1^k + \overline{D}_{2h}\right), \end{aligned} \tag{18.2}$$

where the convergence uses $n^{1/2}\widetilde{M}_{21n} \to_d \overline{M}_{21h}$ (because $M_{21F} = 0^{k\times k}$) and $n^{1/2}\widehat{D}_{2n} \to_d \overline{D}_{2h}$ (because $E_F G_{i2} = 0^k$). Equation (18.2) shows that the asymptotic distribution of $n^{1/2}\widehat{D}_n^\dagger T_n^\dagger$ depends on the randomness of the variance estimator $\widetilde{V}_{Dn}$ through $\overline{M}_{21h}$.

It may appear that this example is quite special and the asymptotic behavior in (18.2) only arises in special circumstances, because $E_F G_i = (1^k, 0^k)$, $M_{21F} = 0^{k\times k}$, and $M_F = I_{2k}$ in this example. But this is not true. The asymptotic behavior in (18.2) arises quite generally, as shown in Theorem 5.1, whenever $p \geq 2$.[54]

If one replaces $\widetilde{V}_{Dn}^{-1/2}$ by its probability limit, M_F, in the definition of $\widehat{D}_n^\dagger$, then the calculations in (18.2) hold but with $n^{1/2}\widetilde{M}_{21n}$ replaced by $n^{1/2}M_{21F} = 0^{k\times k}$ in the first line and, hence, $\overline{M}_{21h}$ replaced by $0^{k\times k}$ in the second line. Hence, in this case, the asymptotic distribution only depends on $\overline{D}_h$. Hence, Comment (iv) to Theorem 5.1 holds in this example.

Suppose one defines $\widehat{D}_n^\dagger$ by $\widetilde{W}_n\widehat{D}_n$ as in Comment (v) to Theorem 5.1. This yields equal weighting of each column of $\widehat{D}_n$. This is equivalent to replacing $\widetilde{V}_{Dn}^{-1/2}$ by $I_2 \otimes \widetilde{W}_n$ in the definition of $\widehat{D}_n^\dagger$ in (18.1). In this case, the off-diagonal $k \times k$ blocks of $I_2 \otimes \widetilde{W}_n$ are $0^{k\times k}$ and, hence, $\widetilde{M}_{21n}$ in the first line of (18.2) equals $0^{k\times k}$, which implies that $\overline{M}_{21h} = 0^{k\times k}$ in the second line of (18.2). Thus, the asymptotic distribution of $\widehat{D}_n^\dagger$ does not depend on the asymptotic distribution of the (normalized) weight matrix estimator $\widetilde{W}_n$. It only depends on the probability limit of $\widetilde{W}_n$, as stated in Comment (v) to Theorem 5.1.

18.2 Asymptotic Size of Kleibergen's CLR Test with Jacobian-Variance Weighting

In this subsection, we determine the asymptotic size of Kleibergen's CLR test when $\widehat{D}_n$ is weighted by $\widetilde{V}_{Dn}$, defined in (5.3), which yields what we call Jacobian-variance weighting, and the Robin and Smith (2000) rank statistic is employed. This rank statistic is defined in (5.5) with

[54] When the matrix $M_{21F} \neq 0^{k\times k}$, the argument in (18.2) does not go through because $n^{1/2}\widetilde{M}_{21n}$ does not converge in distribution (since $n^{1/2}(\widetilde{M}_{21n} - M_{21F}) \to_d \overline{M}_{21h}$ by assumption). In this case, one has to alter the definition of $T_n^\dagger$ so that it rotates the columns of $\widehat{D}_n$ before rescaling them. The rotation required depends on both M_F and $E_F G_i$.

$\theta = \theta_0$. For convenience, we restate the definition here:

$$rk_n = rk_n^\dagger := \lambda_{\min}(n(\widehat{D}_n^\dagger)'\widehat{D}_n^\dagger), \text{ where } \widehat{D}_n^\dagger := vec_{k,p}^{-1}(\widetilde{V}_{Dn}^{-1/2}vec(\widehat{D}_n)) \tag{18.3}$$

(so $\widehat{D}_n^\dagger$ is as in (5.4) with $\theta = \theta_0$).[55] Let

$$\widehat{\kappa}_{jn}^\dagger \text{ denote the } j\text{th eigenvalue of } n(\widehat{D}_n^\dagger)'\widehat{D}_n^\dagger, \text{ for } j = 1, ..., p, \tag{18.4}$$

ordered to be nonincreasing in j. By definition, $\lambda_{\min}(n(\widehat{D}_n^\dagger)'\widehat{D}_n^\dagger) = \widehat{\kappa}_{pn}^\dagger$. Also, the jth singular value of $n^{1/2}\widehat{D}_n^\dagger$ equals $(\widehat{\kappa}_{jn}^\dagger)^{1/2}$.

Define the parameter space $\mathcal{F}_{KCLR}$ for the distribution F by

$$\mathcal{F}_{KCLR} := \{F \in \mathcal{F} : \lambda_{\min}(Var_F((g_i', vec(G_i)')')) \geq \delta_2, E_F||(g_i', vec(G_i)')'||^{4+\gamma} \leq M\}, \tag{18.5}$$

where $\delta_2 > 0$ and $\gamma > 0$ and $M < \infty$ are as in the definition of $\mathcal{F}$ in (3.1). Note that $\mathcal{F}_{KCLR} \subset \mathcal{F}_0$ when δ_1 in $\mathcal{F}_0$ satisfies $\delta_1 \leq M^{-2/(2+\gamma)}\delta_2$, by condition (vi) in (3.9). Let $vech(\cdot)$ denote the half vectorization operator that vectorizes the nonredundant elements in the columns of a symmetric matrix (that is, the elements on or below the main diagonal). The moment condition in $\mathcal{F}_{KCLR}$ is imposed because the asymptotic distribution of the rank statistic $rk_n^\dagger$ depends on a triangular array CLT for $vech(f_i^* f_i^{*\prime})$, which employs $4 + \gamma$ moments for f_i^*, where $f_i^* := (g_i', vec(G_i - E_{F_n}G_i)')'$ as in (5.6). The $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$ ensures that $\widetilde{V}_{Dn}$ is positive definite wp$\to$1, which is needed because $\widetilde{V}_{Dn}$ enters the rank statistic $rk_n^\dagger$ via $\widetilde{V}_{Dn}^{-1/2}$, see (18.3).

For a fixed distribution F, $\widetilde{V}_{Dn}$ estimates $\Phi_F^{vec(G_i)}$ defined in (8.15), where $\Phi_F^{vec(G_i)}$ is pd by its definition in (8.15) and the $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$.[56] Let

$$M_F = \begin{bmatrix} M_{11F} & \cdots & M_{1pF} \\ \vdots & \ddots & \vdots \\ M_{p1F} & \cdots & M_{ppF} \end{bmatrix} := (\Phi_F^{vec(G_i)})^{-1/2} \text{ and} \tag{18.6}$$
$$D_F^\dagger := \sum_{j=1}^{p}(M_{1jF}E_FG_{ij}, ..., M_{pjF}E_FG_{ij}) \in R^{k \times p}, \text{ where } G_i = (G_{i1}, ..., G_{ip}) \in R^{k \times p}.$$

[55] As in Section 5, the function $vec_{k,p}^{-1}(\cdot)$ is the inverse of the $vec(\cdot)$ function for $k \times p$ matrices. Thus, the domain of $vec_{k,p}^{-1}(\cdot)$ consists of kp-vectors and its range consists of $k \times p$ matrices.

[56] More specifically, $\Phi_F^{vec(G_i)}$ is pd because by (8.15) $\Phi_F^{vec(G_i)} := Var_F(vec(G_i) - (E_F vec(G_\ell)g_\ell')\Omega_F^{-1}g_i) = (-(E_F vec(G_\ell)g_\ell')\Omega_F^{-1}, I_{pk})Var_F((g_i', vec(G_i)')')(-(E_F vec(G_\ell)g_\ell')\Omega_F^{-1}, I_{pk})'$, where $(-(E_F vec(G_\ell)g_\ell')\Omega_F^{-1}, I_{pk}) \in R^{pk \times (p+1)k}$ has full row rank pk and $Var_F((g_i', vec(G_i)')')$ is pd by the $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$.

Let $(\tau^{\dagger}_{1F}, ..., \tau^{\dagger}_{pF})$ denote the singular values of $D^{\dagger}_F$. Define

$$\begin{aligned} &B^{\dagger}_F \in R^{p\times p} \text{ to be an orthogonal matrix of eigenvectors of } D^{\dagger\prime}_F D^{\dagger}_F \text{ and} \\ &C^{\dagger}_F \in R^{k\times k} \text{ to be an orthogonal matrix of eigenvectors of } D^{\dagger}_F D^{\dagger\prime}_F \end{aligned} \tag{18.7}$$

ordered so that the corresponding eigenvalues $(\kappa^{\dagger}_{1F}, ..., \kappa^{\dagger}_{pF})$ and $(\kappa^{\dagger}_{1F}, ..., \kappa^{\dagger}_{pF}, 0, ..., 0) \in R^k$, respectively, are nonincreasing. We have $\kappa^{\dagger}_{jF} = (\tau^{\dagger}_{jF})^2$ for $j = 1, ..., p$. Note that (18.7) gives definitions of B_F and C_F that are similar to the definitions in (8.6) and (8.7), but differ because $D^{\dagger}_F$ replaces $W_F(E_F G_i)U_F$ in the definitions.

Define $(\lambda_{1,F}, ..., \lambda_{9,F})$ as in (8.9) with $\lambda_{7,F} = W_F = \Omega_F^{-1/2}$, $\lambda_{8,F} = I_p$, and $W_1(\cdot)$ and $U_1(\cdot)$ equal to identity functions. Define

$$\lambda_{10,F} = Var_F \left(\begin{array}{c} f^*_i \\ vech\,(f^*_i f^{*\prime}_i) \end{array} \right) \in R^{d^* \times d^*}, \tag{18.8}$$

where $d^* := (p+1)k + (p+1)k((p+1)k+1)/2$. Define $(\lambda^{\dagger}_{1,F}, \lambda^{\dagger}_{2,F}, \lambda^{\dagger}_{3,F}, \lambda^{\dagger}_{6,F})$ as $(\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{6,F})$ are defined in (8.9) but with $\{\tau^{\dagger}_{jF} : j \leq p\}$, $B^{\dagger}_F$, and $C^{\dagger}_F$ in place of $\{\tau_{jF} : j \leq p\}$, B_F, and C_F, respectively.

Define

$$\begin{aligned} \lambda &= \lambda_F := (\lambda_{1,F}, ..., \lambda_{10,F}, \lambda^{\dagger}_{1,F}, \lambda^{\dagger}_{2,F}, \lambda^{\dagger}_{3,F}, \lambda^{\dagger}_{6,F}), \qquad (18.9) \\ \Lambda_{KCLR} &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{10,F}, \lambda^{\dagger}_{1,F}, \lambda^{\dagger}_{2,F}, \lambda^{\dagger}_{3,F}, \lambda^{\dagger}_{6,F}) \text{ for some } F \in \mathcal{F}_{KCLR}\}, \text{ and} \\ h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}, \lambda_{10,F}, n^{1/2}\lambda^{\dagger}_{1,F}, \lambda^{\dagger}_{2,F}, \lambda^{\dagger}_{3,F}, \lambda^{\dagger}_{6,F}). \end{aligned}$$

Let $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ denote a sequence $\{\lambda_n \in \Lambda_{KCLR} : n \geq 1\}$ for which $h_n(\lambda_n) \to h \in H$, for H as in (8.1). The asymptotic variance of $n^{1/2} vec(\widehat{D}_n - E_{F_n} G_i)$ is $\Phi_h^{vec(G_i)}$ under $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ by Lemma 8.2.

Define $h_{1,j}$ for $j \leq p$ and h_s for $s = 2, ..., 8$ as in (8.12), $q = q_h$ as in (8.16), $h_{2,q}$, $h_{2,p-q}$, $h_{3,q}$, $h_{3,p-q}$, and $h^{\diamond}_{1,p-q}$ as in (8.17), and Υ_n, $\Upsilon_{n,q}$, and $\Upsilon_{n,p-q}$ as in (13.2). Note that $h_7 = h_{5,g}^{-1/2}$ and $h_8 = I_p$ due to the definitions of $\lambda_{7,F}$ and $\lambda_{8,F}$ given above, where $h_{5,g}$ $(= \lim E_{F_n} g_i g_i^{\prime})$ denotes the upper left $k \times k$ submatrix of h_5, as in Section 8.

For a sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$, we have

$$h_{10} = \left[\begin{array}{cc} h_{10,f^*} & h_{10,f^* f^{*2}} \\ h_{10,f^{*2} f^*} & h_{10,f^{*2} f^{*2}} \end{array} \right] := \lim Var_{F_n} \left(\begin{array}{c} f^*_i \\ vech\,(f^*_i f^{*\prime}_i) \end{array} \right) \in R^{d^* \times d^*}. \tag{18.10}$$

Note that $h_{10,f^*} \in R^{(p+1)k \times (p+1)k}$ is pd by the definition of $\mathcal{F}_{KCLR}$ in (18.5).

With $\tau^\dagger_{jF}$, $B^\dagger_F$, and $C^\dagger_F$ in place of τ_{jF}, B_F, and C_F, respectively, define $h^\dagger_{1,j}$ for $j \leq p$ and $h^\dagger_s$ for $s = 2, 3, 6$ as in (8.12) as analogues to the quantities without the $\dagger$ superscript, define $q^\dagger = q^\dagger_h$ as in (8.16), define $h^\dagger_{2,q^\dagger}$, $h^\dagger_{2,p-q^\dagger}$, $h^\dagger_{3,q^\dagger}$, $h^\dagger_{3,k-q^\dagger}$, and $h^{\dagger\diamond}_{1,p-q^\dagger}$ as in (8.17), and define $\Upsilon^\dagger_n$, $\Upsilon^\dagger_{n,q^\dagger}$, and $\Upsilon^\dagger_{n,p-q^\dagger}$ as in (13.2). The quantity $q^\dagger$ determines the asymptotic behavior of $rk^\dagger_n$. By definition, $q^\dagger$ is the largest value j $(\leq p)$ for which $\lim n^{1/2}\tau^\dagger_{jF_n} = \infty$ under $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$. It is shown below that if $q^\dagger = p$, then $rk^\dagger_n \to_p \infty$, whereas if $q^\dagger < p$, then $rk^\dagger_n$ converges in distribution to a nondegenerate random variable, see Lemma 18.4.

By the CLT, for any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$,

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} f_i^* \\ vech\,(f_i^* f_i^{*\prime} - E_{F_n} f_i^* f_i^{*\prime}) \end{pmatrix} \to_d \overline{L}_h \sim N(0^{d^*}, h_{10}), \text{ where} \\
& \overline{L}_h = (\overline{L}'_{h,1}, \overline{L}'_{h,2}, \overline{L}'_{h,3})' \text{ for } \overline{L}_{h,1} \in R^k,\ \overline{L}_{h,2} \in R^{kp}, \text{ and } \overline{L}_{h,3} \in R^{(p+1)k((p+1)k+1)/2}
\end{aligned} \tag{18.11}
$$

and the CLT holds using the moment conditions in $\mathcal{F}_{KCLR}$. Note that by the definitions of $h_4 := \lim E_{F_n} G_i$ and $h_5 := \lim E_{F_n}(g_i', vec(G_i)')'(g_i', vec(G_i)')$, we have

$$
h_{10,f^*} = \begin{bmatrix} h_{5,g} & h_{5,gG} \\ h_{5,Gg} & h_{5,G} - vec(h_4)vec(h_4)' \end{bmatrix}, \text{ where } h_5 = \begin{bmatrix} h_{5,g} & h_{5,gG} \\ h_{5,Gg} & h_{5,G} \end{bmatrix} \tag{18.12}
$$

for $h_{5,g} \in R^{k \times k}$, $h_{5,Gg} \in R^{kp \times k}$, and $h_{5,G} \in R^{kp \times kp}$.

We now provide new, but distributionally equivalent, definitions of $\overline{g}_h$ and $\overline{D}_h$:

$$
\overline{g}_h := \overline{L}_{h,1} \text{ and } vec(\overline{D}_h) := \overline{L}_{h,2} - h_{5,Gg} h_{5,g}^{-1} \overline{L}_{h,1}. \tag{18.13}
$$

These definitions are distributionally equivalent to the previous definitions of $\overline{g}_h$ and $\overline{D}_h$ given in Lemma 8.2, because by either set of definitions $\overline{g}_h$ and $vec(\overline{D}_h)$ are independent mean zero random vectors with variance matrices $h_{5,g}$ and $\Phi^{vec(G_i)}_h$ $(= h_{5,G} - vec(h_4)vec(h_4)' - h_{5,Gg}h^{-1}_{5,g}h'_{5,Gg})$, respectively, where $\Phi^{vec(G_i)}_h$ is defined in (8.15) and is pd (because $\Phi^{vec(G_i)}_h = \lim \Phi^{vec(G_i)}_{F_n}$ and $\lambda_{\min}(\Phi^{vec(G_i)}_{F_n})$ is bounded away from zero by its definition in (8.15) and the $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$).

Define

$$\overline{D}_h^\dagger := \sum_{j=1}^p (M_{1jh}\overline{D}_{jh}, ..., M_{pjh}\overline{D}_{jh}) \in R^{k\times p}, \text{ where } \begin{bmatrix} M_{11h} & \cdots & M_{1ph} \\ \vdots & \ddots & \vdots \\ M_{p1h} & \cdots & M_{pph} \end{bmatrix} := (\Phi_h^{vec(G_i)})^{-1/2}, \tag{18.14}$$

$\overline{D}_h = (\overline{D}_{1h}, ..., \overline{D}_{ph})$, and $\overline{D}_h$ is defined in (18.13). Define

$$\begin{aligned} \overline{\Delta}_h^\dagger &= (\overline{\Delta}_{h,q^\dagger}^\dagger, \overline{\Delta}_{h,p-q^\dagger}^\dagger) \in R^{k\times p}, \ \overline{\Delta}_{h,q^\dagger}^\dagger := h_{3,q^\dagger}^\dagger \in R^{k\times q^\dagger}, \text{ and} \\ \overline{\Delta}_{h,p-q^\dagger}^\dagger &:= h_3^\dagger h_{1,p-q^\dagger}^{\dagger\diamond} + \overline{D}_h^\dagger h_{2,p-q^\dagger}^\dagger \in R^{k\times(p-q^\dagger)}. \end{aligned} \tag{18.15}$$

Let $a(\cdot)$ be the function from R^{d^*} to $R^{kp(kp+1)/2}$ that maps

$$n^{-1}\sum_{i=1}^n \begin{pmatrix} f_i^* \\ vech\,(f_i^* f_i^{*\prime}) \end{pmatrix} \text{ into} \tag{18.16}$$

$$A_n := vech\left(\left(n^{-1}\sum_{i=1}^n vec(G_i - E_{F_n}G_i)vec(G_i - E_{F_n}G_i)' - \widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}\widetilde{\Gamma}_n'\right)^{-1/2}\right), \text{ where}$$

$$\widetilde{\Omega}_n := n^{-1}\sum_{i=1}^n g_i g_i' \in R^{k\times k} \text{ and } \widetilde{\Gamma}_n := n^{-1}\sum_{i=1}^n vec(G_i - E_{F_n}G_i)g_i' \in R^{pk\times k}.$$

Note that $a(\cdot)$ does not depend on the $n^{-1}\sum_{i=1}^n f_i^*$ part of its argument. Also, $a(\cdot)$ is well defined and continuously partially differentiable at any value of its argument for which $n^{-1}\sum_{i=1}^n f_i^* f_i^{*\prime}$ is pd.[57] We define $\overline{A}_h$ as follows:

$$\begin{aligned} &\overline{A}_h \text{ denotes the } (kp)(kp+1)/2 \times d^* \text{ matrix of partial derivatives of } a(\cdot) \\ &\text{evaluated at } (0^{(p+1)k\prime}, vech(h_{10,f^*})')', \end{aligned} \tag{18.17}$$

where the latter vector is the limit of the mean vector of $(f_i^{*\prime}, vech\,(f_i^* f_i^{*\prime})')'$ under $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$.

Define

$$\overline{M}_h := vech_{kp,kp}^{-1}(\overline{A}_h\overline{L}_h) \in R^{kp\times kp}, \tag{18.18}$$

where $vech_{kp,kp}^{-1}(\cdot)$ denotes the inverse of the $vech(\cdot)$ operator applied to symmetric $kp\times kp$ matrices.

[57] The function $a(\cdot)$ is well defined in this case because $n^{-1}\sum_{i=1}^n vec(G_i - E_{F_n}G_i)vec(G_i - E_{F_n}G_i)' - \widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}\widetilde{\Gamma}_n'$ $= (-\widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}, I_{pk})n^{-1}\sum_{i=1}^n f_i^* f_i^{*\prime}(-\widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}, I_{pk})'$ and $(-\widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}, I_{pk}) \in R^{pk\times(p+1)k}$ has full row rank pk.

Define

$$\overline{M}_h^\dagger := (\overline{M}_{h,q^\dagger}^\dagger, \overline{M}_{h,p-q^\dagger}^\dagger) := (0^{k\times q^\dagger}, \overline{M}_{h,p-q^\dagger}^\dagger) \in R^{k\times p}, \text{ where} \tag{18.19}$$

$$\overline{M}_{h,p-q^\dagger}^\dagger := \sum_{j=1}^{p} (\overline{M}_{1jh} h_{4,j}, ..., \overline{M}_{pjh} h_{4,j}) h_{2,p-q^\dagger}^\dagger \in R^{k\times(p-q^\dagger)},\ \overline{M}_h = \begin{bmatrix} \overline{M}_{11h} & \cdots & \overline{M}_{1ph} \\ \vdots & \ddots & \vdots \\ \overline{M}_{p1h} & \cdots & \overline{M}_{pph} \end{bmatrix},$$

and $h_4 = (h_{4,1}, ..., h_{4,p}) \in R^{k\times p}$.

Below (in Lemma 18.4), we show that the asymptotic distribution of $rk_n^\dagger$ under sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with $q^\dagger < p$ is given by

$$r_h(\overline{D}_h, \overline{M}_h) := \lambda_{\min}((\overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger)' h_{3,k-q^\dagger}^\dagger h_{3,k-q^\dagger}^{\dagger\prime} (\overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger)), \tag{18.20}$$

where $\overline{\Delta}_{h,p-q^\dagger}^\dagger$ is a nonrandom function of $\overline{D}_h$ by (18.14) and (18.15) and $\overline{M}_{h,p-q^\dagger}^\dagger$ is a nonrandom function of $\overline{M}_h$ by (18.19). For sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with $q^\dagger = p$, we show that $rk_n \to_p \overline{r}_h := \infty$.

We define $\overline{\Delta}_h$, as in (8.17), as follows:

$$\overline{\Delta}_h = (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q},\ \text{and } \overline{\Delta}_{h,p-q} := h_3 h_{1,p-q}^\diamond + h_7 \overline{D}_h h_8 h_{2,p-q}, \text{ where}$$

$$h_2 = (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}),\ h_{1,p-q}^\diamond := \begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times(p-q)}. \tag{18.21}$$

In the present case, $h_7 = h_{5,g}^{-1/2}$ and $h_8 = I_p$ because the CLR_n statistic depends on $\widehat{D}_n$ through $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$, which appears in the LM_n statistic.[58] This means that Assumption WU for the parameter space Λ_{KCLR} (defined in Section 8.4) holds with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $\widehat{U}_n = I_p$, $h_7 = h_{5,g}^{-1/2}$, and $h_8 = I_p$. Thus, the distribution of $\overline{\Delta}_h$ depends on $\overline{D}_h$, q, and h_s for $s = 1, 2, 3, 5$.

Below (in Lemma 18.5), we show that the asymptotic distribution of the CLR_n statistic under

[58] The CLR_n statistic also depends on $\widehat{D}_n$ through the rank statistic.

sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with $q^\dagger < p$ is given by[59]

$$\begin{aligned}
\overline{CLR}_h &:= \frac{1}{2}\left(\overline{LM}_h + \overline{J}_h - \overline{r}_h + \sqrt{(\overline{LM}_h + \overline{J}_h - \overline{r}_h)^2 + 4\overline{LM}\overline{r}_h}\right), \text{ where} \\
\overline{LM}_h &:= \overline{v}_h'\overline{v}_h \sim \chi^2_p,\ \overline{v}_h := P_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h,\ \overline{J}_h := \overline{g}_h' h_{5,g}^{-1/2} M_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h \sim \chi^2_{k-p}, \text{ and} \\
\overline{r}_h &:= r_h(\overline{D}_h, \overline{M}_h). \qquad (18.22)
\end{aligned}$$

The quantities $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ are specified in (18.13) and (18.18) (and $(\overline{g}_h, \overline{D}_h)$ are the same as in Lemma 8.2). Conditional on $\overline{D}_h$, $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ^2_p and χ^2_{k-p}, respectively (see the paragraph following (10.6)). For sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with $q^\dagger = p$, we show that the asymptotic distribution of the CLR_n statistic is $\overline{CLR}_h := \overline{LM}_h := \overline{v}_h'\overline{v}_h \sim \chi^2_p$, where $\overline{v}_h := P_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h$.

The critical value function $c(1-\alpha, r)$ is defined in (5.2) for $0 \leq r < \infty$. For $r = \infty$, we define $c(1-\alpha, r)$ to be the $1-\alpha$ quantile of the χ^2_p distribution.

Now we state the asymptotic size of Kleibergen's CLR test based on Robin and Smith (2000) statistic with $\widetilde{V}_{Dn}$ defined in (5.3).

Theorem 18.1 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the variance matrix estimator $\widetilde{V}_{Dn}$ employed by the rank statistic $rk_n^\dagger$ (defined in (18.3)) is defined by (5.3). Then, the asymptotic size of Kleibergen's* CLR *test based on the rank statistic $rk_n^\dagger$ is*

$$AsySz = \max\{\alpha, \sup_{h \in H} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))\}$$

provided $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$ for all $h \in H$.

Comments: (i) The proviso in Theorem 18.1 is a continuity condition on the distribution function of $\overline{CLR}_h - c(1-\alpha, \overline{r}_h)$ at zero. If the proviso in Theorem 18.1 does not hold, then the following weaker conclusion holds:

$$\begin{aligned}
&AsySz \qquad (18.23) \\
&\in [\max\{\alpha, \sup_{h \in H} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))\}, \max\{\alpha, \sup_{h \in H} \lim_{x \uparrow 0} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) + x)\}].
\end{aligned}$$

(ii) Conditional on $(\overline{D}_h, \overline{M}_h)$, $\overline{g}_h$ has a multivariate normal distribution a.s. (because $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ has a multivariate normal distribution unconditionally).[60] The proviso in Theorem 18.1 holds

[59] The definitions of $\overline{v}_h$, $\overline{LM}_h$, $\overline{J}_h$, and $\overline{CLR}_h$ in (18.22) are the same as in (9.1), (9.2), (10.6), and (10.7), respectively.

[60] Note that $\overline{g}_h$ is independent of $\overline{D}_h$.

whenever $\overline{g}_h$ has a non-zero variance matrix conditional on $(\overline{D}_h, \overline{M}_h)$ a.s. for all $h \in H$. This holds because (a) $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = E_{(\overline{D}_h,\overline{M}_h)} P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)|\overline{D}_h, \overline{M}_h)$ by the law of iterated expectations, (b) some calculations show that $\overline{CLR}_h = c(1-\alpha, \overline{r}_h)$ iff $(\overline{r}_h + c)\overline{LM}_h = -c\overline{J}_h + c^2 + c\overline{r}_h$ iff $\overline{X}_h'\overline{X}_h = c^2 + c\overline{r}_h$, where $c := c(1-\alpha, \overline{r}_h)$ and $\overline{X}_h := ((\overline{r}_h + c)^{1/2}(P_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h)', c^{1/2}(M_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h)')'$ using (18.22), (c) $P_{\overline{\Delta}_h} + M_{\overline{\Delta}_h} = I_k$ and $P_{\overline{\Delta}_h} M_{\overline{\Delta}_h} = 0^{k\times k}$, and (d) conditional on $(\overline{D}_h, \overline{M}_h)$, $\overline{r}_h$, c, and $\overline{\Delta}_h$ are constants.

(iii) When $p = 1$, the formula for $AsySz$ in Theorem 18.1 reduces to α and the proviso holds automatically. That is, Kleibergen's CLR test has correct asymptotic size when $p = 1$. This holds because when $p = 1$ the quantity $\overline{M}_h^{\dagger}$ in (18.19) equals $0^{k\times p}$ by Comment (ii) to Theorem 18.3 below. This implies that $r_h(\overline{D}_h, \overline{M}_h)$ in (18.20) does not depend on $\overline{M}_h$. Given this, the proof that $P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) = \alpha$ for all $h \in H$ and that the proviso holds is the same as in (10.9)-(10.10) in the proof of Theorem 10.1.

(iv) Theorem 18.1 is proved by showing that it is a special case of Theorem 18.6 below, which is similar but applies not to $\widetilde{V}_{Dn}$ defined in (5.3), but to an arbitrary estimator $\widetilde{V}_{Dn}$ (of the asymptotic variance $\Phi_h^{vec(G_i)}$ of $n^{1/2}vec(\widehat{D}_n - E_{F_n}G_i))$ that satisfies an Assumption VD (which is stated below). Lemma 18.2 below shows that the estimator $\widetilde{V}_{Dn}$ defined in (5.3) satisfies Assumption VD.

(v) A CS version of Theorem 18.1 holds with the parameter space $\mathcal{F}_{\Theta,KCLR}$ in place of $\mathcal{F}_{KCLR}$, where $\mathcal{F}_{\Theta,KCLR} := \{(F, \theta_0) : F \in \mathcal{F}_{KCLR}(\theta_0), \theta_0 \in \Theta\}$ and $\mathcal{F}_{KCLR}(\theta_0)$ is the set $\mathcal{F}_{KCLR}$ defined in (18.5) with its dependence on θ_0 made explicit. The proof of this CS result is as outlined in the Comment to Proposition 8.1. For the CS result, the h index and its parameter space H are as defined above, but h also includes θ_0 as a subvector, and H allows this subvector to range over Θ.

18.3 Simulation Results

In this section, for a particular linear IV regression model, we simulate (i) the correlations between $\overline{M}_{h,p-q^{\dagger}}^{\dagger}$ (defined in (18.19)) and $\overline{g}_h$ and (ii) some asymptotic null rejection probabilities (NRP's) of Kleibergen's CLR test that uses Jacobian-variance weighting and employs the Robin and Smith (2000) rank statistic. The model has $p = 2$ rhs endogenous variables, $k = 5$ IV's, and an error structure that yields simplified asymptotic formulae for some key quantities. The model is

$$y_{1i} = Y_{2i}'\theta_0 + u_i \text{ and } Y_{2i} = \pi' Z_i + V_{2i}, \tag{18.24}$$

where $y_{1i}, u_i \in R$, $Y_{2i}, V_{2i} = (V_{21i}, V_{22i})', \theta \in R^2$, $Z_i = (Z_{i1}, ..., Z_{i5})' \in R^5$, and $\pi \in R^{5\times 2}$. We take $Z_{ij} \sim N(.05, (.05)^2)$ for $j = 1, ..., 5$, $u_i \sim N(0, 1)$, $V_{1i} \sim N(0, 1)$, and $V_{2i} = u_i V_{21i}$. The random variables $Z_{i1}, ..., Z_{i5}$, u_i, and V_{1i} are taken to be mutually independent. We take $\pi =$

$\pi_n = (e_1, e_2 cn^{-1/2})$, where $e_1 = (1, 0, ..., 0)' \in R^5$ and $e_2 = (0, 1, 0, ..., 0)' \in R^5$. We consider 26 values of the constant c lying between 0 and 60.1 (viz., $0.0, 0.1, ..., 1.0, 1.1, ..., 10.1, 20.1, ..., 60.1$), as well as 707.1, 1414.2, and $1,000,000$. Given these definitions, $h_{1,1} = \infty$, $h_{1,2} = c$, and $\overline{M}_h^{\dagger} = (0^5, \overline{M}_{h,p-q^{\dagger}}^{\dagger}) \in R^{5 \times 2}$, see (18.19).

In this model, we have $g_i = -Z_i u_i$ and $G_i = -Z_i Y_{2i}'$. The specified error distribution leads to $E_F G_i g_i' = 0^{k \times k}$. In consequence, the matrix $\Phi_h^{vec(G_i)}$ (defined in (8.15)), which is the asymptotic variance of the Jacobian-variance matrix estimator $\widetilde{V}_{Dn}$ (defined in (5.3)), simplifies as follows:

$$\begin{aligned} \Phi_h^{vec(G_i)} &= \lim Var_{F_n}\left(vec(D_i - E_{F_n} D_i) vec(D_i - E_{F_n} D_i)'\right) \\ &= \lim Var_{F_n}\left(vec(G_i - E_{F_n} G_i) vec(G_i - E_{F_n} G_i)'\right), \text{ where} \qquad (18.25) \\ D_i &:= \left(G_{i1} - \Gamma_{1F}\Omega_F^{-1} g_i, G_{i2} - \Gamma_{2F}\Omega_F^{-1} g_i\right), \ \Gamma_{jF} = E_F G_{ij} g_i' \text{ for } j = 1, 2, \text{ and } \Omega_F = E_F g_i g_i'. \end{aligned}$$

In addition, in the present model, G_{i1} and G_{i2} are uncorrelated, where $G_i = (G_{i1}, G_{i2})$. In consequence, $\Phi_h^{vec(G_i)}$ is block diagonal. In turn, this implies that $\lim M_{F_n} := (\Phi_h^{vec(G_i)})^{-1/2}$ is block diagonal with off-diagonal block $\lim M_{12F_n} = 0^{5 \times 5}$.

The quantities $h_{1,j}^{\dagger}$ for $j = 1, ..., 5$ (defined just below (18.10)) are not available in closed form, so we simulate them using a very large value of n, viz., $n = 2,000,000$. We use $4,000,000$ simulation repetitions to compute the correlations between the jth elements of $\overline{M}_{h,p-q^{\dagger}}^{\dagger}$ and $\overline{g}_h$ for $j = 1, ..., 5$ and the asymptotic NRP's of the CLR test.[61] The data-dependent critical values for the test are computed using a look-up table that gives the critical values for each fixed value r of the rank statistic in a grid from 0 to 100 with a step size of .005. These critical values are computed using $4,000,000$ simulation repetitions.

Results are obtained for each of the 29 values of c listed above. The simulated correlations between the jth elements of $\overline{M}_{h,p-q^{\dagger}}^{\dagger}$ and $\overline{g}_h$ for $j = 1, ..., 5$ take the following values

$$-.33, \ -.38, \ -.38, \ -.38, \text{ and } -.38 \qquad (18.26)$$

for all values of $c \leq 60.1$. For $c = 707.1$, the correlations are $-.32$, $-.36$, $-.36$, $-.36$, and $-.36$. For $c = 1414.2$, the correlations are $-.24$, $-.27$, $-.27$, $-.27$, and $-.27$. For $c = 1,000,000$, the correlations are $-.01$, $-.01$, $-.01$, $-.01$, and $-.01$. These results corroborate the findings given in Theorem 5.1 that $\overline{M}_{h,p-q^{\dagger}}^{\dagger}$ and $\overline{g}_h$ are correlated asymptotically in some models under some sequences of distributions. In consequence, it is not possible to show the Jacobian-variance weighted CLR test has correct asymptotic size via a conditioning argument that relies on the independence

[61] The correlations between the jth and kth elements of these vectors for $j \neq k$ are zero by analytic calculation. Hence, they are not reported here.

of $\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}$ and $\overline{g}_h$.

Next, we report the asymptotic NRP results for Kleibergen's CLR test that uses Jacobian-variance weighting and the Robin and Smith (2000) rank statistic. The asymptotic NRP's are found to be between 4.95% and 5.01% for the 29 values of c considered. These values are very close to the nominal size of 5.00%. Whether the difference is due to simulation noise or not is not clear. The simulation standard error based on the formula $100 * (\alpha(1-\alpha)/reps)^{1/2}$, where $reps = 4,000,000$ is the number of simulation repetitions, is .01. However, this formula does not take into account simulation error from the computation of the critical values.

We conclude that, for the model and error distribution considered, the asymptotic NRP's of the Kleibergen's CLR test with Jacobian-variance weighting is equal to, or very close to, its nominal size. This occurs even though there are non-negligible correlations between $\overline{M}^{\dagger}_{h,p-q^{\dagger}}$ and $\overline{g}_h$. Whether this occurs for all parameters and distributions in the linear IV model, and whether it occurs in other moment condition model, is an open question. It appears to be a question that can only be answered on a case by case basis.

18.4 Asymptotic Size of Kleibergen's CLR Test for General $\widetilde{\mathbf{V}}_{\mathbf{Dn}}$ Estimators

In this section, we determine the asymptotic size of Kleibergen's CLR test (defined in Section 5) using the Robin and Smith (2000) rank statistic based on a general "Jacobian-variance" estimator $\widetilde{V}_{Dn}$ $(= \widetilde{V}_{Dn}(\theta_0))$ that satisfies the following Assumption VD.

The first two results of this section, viz., Lemma 18.2 and Theorem 18.3, combine to establish Theorem 5.1, see Comment (i) to Theorem 18.3. The first and last results of this section, viz., Lemma 18.2 and Theorem 18.6, combine to prove Theorem 18.1.

The proofs of the results in this section are given in Section 18.6.

Assumption VD: For any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$, the estimator $\widetilde{V}_{Dn}$ is such that $n^{1/2}(\widetilde{M}_n - M_{F_n}) \to_d \overline{M}_h$ for some random matrix $\overline{M}_h \in R^{kp \times kp}$ (where $\widetilde{M}_n = \widetilde{V}_{Dn}^{-1/2}$ and M_{F_n} is defined in (18.6)), the convergence is joint with

$$n^{1/2}\begin{pmatrix} \widehat{g}_n \\ vec(\widehat{D}_n - E_{F_n}G_i) \end{pmatrix} \to_d \begin{pmatrix} \overline{g}_h \\ vec(\overline{D}_h) \end{pmatrix} \sim N\left(0^{(p+1)k}, \begin{pmatrix} h_{5,g} & 0^{k \times pk} \\ 0^{pk \times k} & \Phi_h^{vec(G_i)} \end{pmatrix}\right), \quad (18.27)$$

and $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ has a mean zero multivariate normal distribution with pd variance matrix. The same condition holds for any subsequence $\{w_n\}$ and any sequence $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with w_n in place of n throughout.

Note that the convergence in (18.27) holds by Lemma 8.2.

The following lemma verifies Assumption VD for the estimator $\widetilde{V}_{Dn}$ defined in (5.3).

Lemma 18.2 *The estimator* $\widetilde{V}_{Dn}$ *defined in* (5.3) *satisfies Assumption* VD. *Specifically,* $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i, \widetilde{M}_n - M_{F_n}) \rightarrow_d (\overline{g}_h, \overline{D}_h, \overline{M}_h)$, *where* $\widetilde{M}_n := \widetilde{V}_{Dn}^{-1/2}$, $M_{F_n} := (\Phi_{F_n}^{vec(G_i)})^{-1/2}$, *and* $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ *has a mean zero multivariate normal distribution defined by* (18.11) *and* (18.13)-(18.18) *with pd variance matrix.*

Comment: As stated in the paragraph containing (18.21), $\widehat{D}_n$ is defined in Lemma 18.2 and Theorem 18.3 below with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$.

Define

$$S_n^\dagger := Diag\{(n^{1/2}\tau_{1F_n}^\dagger)^{-1}, ..., (n^{1/2}\tau_{qF_n}^\dagger)^{-1}, 1, ..., 1\} \in R^{p \times p} \text{ and } T_n^\dagger := B_n^\dagger S_n^\dagger, \tag{18.28}$$

where $B_n^\dagger$ is defined in (18.7).

The asymptotic distribution of $n^{1/2}\widehat{D}_n^\dagger T_n^\dagger$ is given in the following theorem.

Theorem 18.3 *Suppose Assumption* VD *holds. For all sequences* $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$, $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i, \widehat{D}_n^\dagger T_n^\dagger) \rightarrow_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h^\dagger + \overline{M}_h^\dagger)$, *where* $\overline{\Delta}_h^\dagger$ *is a nonrandom affine function of* $\overline{D}_h$ *defined in* (18.14) *and* (18.15), $\overline{M}_h^\dagger$ *is a nonrandom linear* (*i.e., affine and homogeneous of degree one*) *function of* $\overline{M}_h$ *defined in* (18.19), $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ *has a mean zero multivariate normal distribution, and* $\overline{g}_h$ *and* $\overline{D}_h$ *are independent. Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$, *the same result holds with* n *replaced with* w_n.

Comments: (i) Note that the random variables $(\overline{g}_h, \overline{\Delta}_h^\dagger, \overline{M}_h^\dagger)$ in Theorem 5.1 have a multivariate normal distribution whose mean and variance matrix depend on $\lim Var_{F_n}((f_i^{*\prime}, vec(f_i^* f_i^{*\prime})^\prime)$ and on the limits of certain functions of $E_{F_n}G_i$ by (18.11)-(18.19). This, Lemma 18.2, and Theorem 18.3 combine to prove Theorem 5.1 of AG1.

(ii) From (18.19), $\overline{M}_h^\dagger = 0^{k \times p}$ if $p = 1$ (because $q^\dagger = 0$ implies $q = 0$ which, in turn, implies $h_4 = 0^k$ and $q^\dagger = 1$ implies $\overline{M}_{h,p-q^\dagger}^\dagger$ has no columns).[62] For $p \geq 2$, $\overline{M}_h^\dagger = 0^{k \times p}$ if $p = q^\dagger$ (because $\overline{M}_{h,p-q^\dagger}^\dagger$ has no columns) or if $h_{4,j} = 0^k$ for all $j \leq p$. The former holds if the singular values $(\tau_{1F_n}, ..., \tau_{pF_n})$ of $D_{F_n}^\dagger$ satisfy $n^{1/2}\tau_{jF_n} \rightarrow \infty$ for all $j \leq p$ (i.e., all parameters are strongly or semi-strongly identified). The latter occurs if $E_{F_n}G_i \rightarrow 0^{k \times p}$ (i.e., all parameters are either weakly identified in the standard sense or semi-strongly identified). These two condition fail to hold when

[62] Note that $q^\dagger = 0$ implies $q = 0$ when $p = 1$ because $n^{1/2}D_{F_n}^\dagger = n^{1/2}M_{F_n}E_{F_n}G_i = O(1)$ when $q^\dagger = 0$ (by the definition of $q^\dagger$) and this implies that $n^{1/2}E_{F_n}G_i = O(1)$ using the first condition in $\mathcal{F}_{KCLR}$. In turn, the latter implies that $n^{1/2}\Omega_{F_n}^{-1/2}E_{F_n}G_i = O(1)$ using the last condition in $\mathcal{F}$. That is, $q = 0$ (since $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$ because $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$ in the present case, see the Comment to Lemma 18.2).

one or more parameters are strongly identified and one or more parameters are weakly identified or jointly weakly identified.

(iii) For example, when $p = 2$ the conditions in Comment (ii) (under which $\overline{M}_h^\dagger = 0^{k \times p}$) fail to hold if $E_{F_n} G_{i1} \neq 0^k$ does not depend on n and $n^{1/2} E_{F_n} G_{i2} \to c$ for some $c \in R^k$.

The following lemma establishes the asymptotic distribution of $rk_n^\dagger$.

Lemma 18.4 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the variance matrix estimator $\widetilde{V}_{Dn}$ employed by the rank statistic $rk_n^\dagger$ (defined in* (18.3)*) satisfies Assumption* VD*. Then, under all sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$,*

(a) $rk_n^\dagger := \widehat{\kappa}_{pn}^\dagger \to_p \infty$ *if* $q^\dagger = p$,

(b) $rk_n^\dagger := \widehat{\kappa}_{pn}^\dagger \to_d r_h(\overline{D}_h, \overline{M}_h)$ *if* $q^\dagger < p$, *where* $r_h(\overline{D}_h, \overline{M}_h)$ *is defined in* (18.20) *using* (18.19) *with* $\overline{M}_h$ *defined in Assumption* VD (*rather than in* (18.18)),

(c) $\widehat{\kappa}_{jn}^\dagger \to_p \infty$ *for all* $j \leq q^\dagger$,

(d) *the* (*ordered*) *vector of the smallest* $p - q^\dagger$ *singular values of* $n^{1/2} \widehat{D}_n^\dagger$, *i.e.,* $((\widehat{\kappa}_{(q^\dagger+1)n}^\dagger)^{1/2}, ..., (\widehat{\kappa}_{pn}^\dagger)^{1/2})'$, *converges in distribution to the* (*ordered*) $p - q^\dagger$ *vector of the singular values of* $h_{3,k-q^\dagger}^{\dagger\prime} (\overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger) \in R^{(k-q^\dagger) \times (p-q^\dagger)}$, *where* $\overline{M}_{h,p-q^\dagger}^\dagger$ *is defined in* (18.19) *with* $\overline{M}_h$ *defined in Assumption* VD (*rather than in* (18.18)),

(e) *the convergence in parts* (a)-(d) *holds jointly with the convergence in Theorem* 18.3, *and*

(f) *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$, *parts* (a)-(e) *hold with* n *replaced with* w_n.

The following lemma gives the joint asymptotic distribution of CLR_n and $rk_n^\dagger$ and the asymptotic null rejection probabilities of Kleibergen's CLR test.

Lemma 18.5 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the variance matrix estimator $\widetilde{V}_{Dn}$ employed by the rank statistic $rk_n^\dagger$ (defined in* (18.3)*) satisfies Assumption* VD*. Then, under all sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$,*

(a) $CLR_n = LM_n + o_p(1) \to_d \chi_p^2$ *and* $rk_n^\dagger \to_p \infty$ *if* $q^\dagger = p$,

(b) $\lim_{n \to \infty} P(CLR_n > c(1 - \alpha, rk_n^\dagger)) = \alpha$ *if* $q^\dagger = p$,

(c) $(CLR_n, rk_n^\dagger) \to_d (\overline{CLR}_h, \overline{r}_h)$ *if* $q^\dagger < p$, *and*

(d) $\lim_{n \to \infty} P(CLR_n > c(1 - \alpha, rk_n^\dagger)) = P(\overline{CLR}_h > c(1 - \alpha, \overline{r}_h))$ *if* $q^\dagger < p$, *provided* $P(\overline{CLR}_h = c(1 - \alpha, \overline{r}_h)) = 0$.

Under all subsequences $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} \in \Lambda_{KCLR} \geq 1\}$, *parts* (a)-(d) *hold with* n *replaced with* w_n.

Comments: **(i)** The CLR critical value function $c(1-\alpha, r)$ is the $1-\alpha$ quantile of $clr(r)$. By definition,

$$clr(r) := \frac{1}{2}\left(\chi_p^2 + \chi_{k-p}^2 - r + \sqrt{(\chi_p^2 + \chi_{k-p}^2 - r)^2 + 4\chi_p^2 r}\right), \tag{18.29}$$

where the chi-square random variables χ_p^2 and χ_{k-p}^2 are independent. If $\overline{r}_h := r_h(\overline{D}_h, \overline{M}_h)$ does not depend on $\overline{M}_h$, then, conditional on $\overline{D}_h$, $\overline{r}_h$ is a constant and $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2 (see the paragraph following (10.6)). In this case, even when $q^\dagger = p$,

$$P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)) = E_{\overline{D}_h} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) | \overline{D}_h) = \alpha, \tag{18.30}$$

as desired, where the first equality holds by the law of iterated expectations and the second equality holds because $\overline{r}_h$ is a constant conditional on $\overline{D}_h$ and $c(1-\alpha, \overline{r}_h)$ is the $1-\alpha$ quantile of the conditional distribution of $clr(\overline{r}_h)$ given $\overline{D}_h$, which equals that of $\overline{CLR}_h$ given $\overline{D}_h$.

(ii) However, when $\overline{r}_h := r_h(\overline{D}_h, \overline{M}_h)$ depends on $\overline{M}_h$, the distribution of $\overline{r}_h$ conditional on $\overline{D}_h$ is not a pointmass distribution. Rather, conditional on $\overline{D}_h$, $\overline{r}_h$ is a random variable that is not independent of $\overline{LM}_h$, $\overline{J}_h$, and $\overline{CLR}_h$. In consequence, the second equality in (18.30) does not hold and the asymptotic null rejection probability of Kleibergen's CLR test may be larger or smaller than α depending upon the sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ (or $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$) when $q^\dagger < p$.

Next, we use Lemma 18.5 to provide an expression for the asymptotic size of Kleibergen's CLR test based on the Robin and Smith (2000) rank statistic with Jacobian-variance weighting.

Theorem 18.6 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the variance matrix estimator $\widetilde{V}_{Dn}$ employed by the rank statistic $rk_n^\dagger$ (defined in* (18.3)*) satisfies Assumption* VD*. Then, the asymptotic size of Kleibergen's* CLR *test based on $rk_n^\dagger$ is*

$$AsySz = \max\{\alpha, \sup_{h \in H} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))\}$$

provided $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$ for all $h \in H$.

Comments: **(i)** Comment (i) to Theorem 18.1 also applies to Theorem 18.6.

(ii) Theorem 18.6 and Lemma 18.2 combine to prove Theorem 18.1.

(iii) A CS version of Theorem 18.6 holds with the parameter space $\mathcal{F}_{\Theta,KCLR}$ in place of $\mathcal{F}_{KCLR}$, see Comment (v) to Theorem 18.1 and the Comment to Proposition 8.1.

18.5 Correct Asymptotic Size of Equally-Weighted CLR Tests Based on the Robin-Smith Rank Statistic

In this subsection, we consider equally-weighted CLR tests, a special case of which is considered in Section 6. By definition, an equally-weighted CLR test is a CLR test that is based on a rk_n statistic that depends on $\widehat{D}_n$ only through $\widetilde{W}_n \widehat{D}_n$ for some general $k \times k$ weighting matrix $\widetilde{W}_n$. We show that such tests have correct asymptotic size when they are based on the rank statistic of Robin and Smith (2000) and employ a general weight matrix $\widetilde{W}_n \in R^{k \times k}$ that satisfies certain conditions. In contrast, the results in Section 6 consider the specific weight matrix $\widehat{\Omega}_n^{-1/2} \in R^{k \times k}$. The reason for considering these tests in this section is that the asymptotic results can be obtained as a relatively simple by-product of the results in Section 18.4. All that is required is a slight change in Assumption VD.

The rank statistic that we consider here is

$$rk_n^{\dagger} := \lambda_{\min}(n \widehat{D}_n' \widetilde{W}_n' \widetilde{W}_n \widehat{D}_n). \tag{18.31}$$

We replace Assumption VD in Section 18.4 by the following assumption.

Assumption W: For any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$, the random $k \times k$ weight matrix $\widetilde{W}_n$ is such that $n^{1/2}(\widetilde{W}_n - W_{F_n}^{\dagger}) \to_d \overline{W}_h$ for some non-random $k \times k$ matrices $\{W_{F_n}^{\dagger} : n \geq 1\}$ and some random $k \times k$ matrix $\overline{W}_h \in R^{k \times k}$, $W_{F_n}^{\dagger} \to W_h^{\dagger}$ for some nonrandom pd $k \times k$ matrix $W_h^{\dagger}$, the convergence is joint with the convergence in (18.27), and $(\overline{g}_h, \overline{D}_h, \overline{W}_h)$ has a mean zero multivariate normal distribution with pd variance matrix. The same condition holds for any subsequence $\{w_n\}$ and any sequence $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with w_n in place of n throughout.

If one takes $\widetilde{M}_n$ $(= \widetilde{V}_{Dn}^{-1/2}) = I_p \otimes \widetilde{W}_n$ in Assumption VD, then $\widehat{D}_n^{\dagger} = \widetilde{W}_n \widehat{D}_n$ and the rank statistics in (18.3) and (18.31) are the same. Thus, Assumption W is analogous to Assumption VD with $\widetilde{M}_n = I_p \otimes \widetilde{W}_n$ and $M_{F_n} = I_p \otimes W_{F_n}^{\dagger}$. Note, however, that the latter matrix does not typically satisfy the condition in Assumption VD that M_{F_n} is defined in (18.6), i.e., the condition that $M_{F_n} = (\Phi_{F_n}^{vec(G_i)})^{-1/2}$. Nevertheless, the results in Section 18.4 hold with Assumption VD replaced by Assumption W and with $M_F = I_p \otimes W_F^{\dagger}$, $D_F^{\dagger} = W_F^{\dagger} E_F G_i$, and $\overline{M}_h = I_p \otimes \overline{W}_h$. With these changes, $\overline{D}_h^{\dagger} = W_h^{\dagger} \overline{D}_h$ in (18.14) (because $(\Phi_h^{vec(G_i)})^{-1/2}$ is replaced by $I_p \otimes W_h^{\dagger}$), $\overline{\Delta}_h^{\dagger}$ is defined as in (18.15) with $\overline{D}_h^{\dagger}$ as just given, and $\overline{M}_h^{\dagger}$ is defined as in (18.19) with $\overline{M}_{h,p-q^{\dagger}}^{\dagger} = \overline{W}_h h_4 h_{2,p-q^{\dagger}}^{\dagger}$.

Below we show the key result that $\overline{M}_{h,p-q^{\dagger}}^{\dagger} = 0^{k \times (p-q^{\dagger})}$ for $rk_n^{\dagger}$ defined in (18.31). By (18.20), this implies that

$$r_h(\overline{D}_h, \overline{M}_h) := \lambda_{\min}((\overline{\Delta}_{h,p-q^{\dagger}}^{\dagger})' h_{3,k-q^{\dagger}}^{\dagger} h_{3,k-q^{\dagger}}^{\dagger\prime} (\overline{\Delta}_{h,p-q^{\dagger}}^{\dagger})) \tag{18.32}$$

when $q^\dagger < p$. Note that the rhs in (18.32) does not depend on $\overline{M}_h$ and, hence, is a function only of $\overline{D}_h$. That is, $r_h(\overline{D}_h, \overline{M}_h) = r_h(\overline{D}_h)$. Given that $r_h(\overline{D}_h, \overline{M}_h)$ does not depend on $\overline{M}_h$, Comment (i) to Lemma 18.5 implies that $P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)) = \alpha$ under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} \in \Lambda_{KCLR} : n \geq 1\}$. This and Theorem 18.6 give the following result.

Corollary 18.7 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the rank statistic $rk_n^\dagger$ (defined in* (18.31)) *is based on a weight matrix $\widetilde{W}_n$ that satisfies Assumption* W. *Then, the asymptotic size of the corresponding equally-weighted version of Kleibergen's* CLR *test (defined in Section* 5 *with $rk_n(\theta) = rk_n^\dagger$) equals α.*

Comment: A CS version of Corollary 18.7 holds with the parameter space $\mathcal{F}_{\Theta,KCLR}$ in place of $\mathcal{F}_{KCLR}$, see Comment (v) to Theorem 18.1 and the Comment to Proposition 8.1.

Now, we establish that $\overline{M}^\dagger_{h,p-q^\dagger}$ $(= \overline{W}_h h_4 h^\dagger_{2,p-q^\dagger}) = 0^{k \times (p-q^\dagger)}$. We have

$$W_h^\dagger h_4 := \lim W_{F_n}^\dagger E_{F_n} G_i = \lim C_{F_n}^\dagger \Upsilon_{F_n}^\dagger B_{F_n}^{\dagger\prime} = h_3^\dagger \lim \Upsilon_{F_n}^\dagger h_2^{\dagger\prime}, \tag{18.33}$$

where $C_{F_n}^\dagger \Upsilon_{F_n}^\dagger (B_{F_n}^\dagger)'$ is the singular value decomposition of $W_{F_n}^\dagger E_{F_n} G_i$, $\Upsilon_{F_n}^\dagger$ is the $k \times p$ matrix with the singular values of $W_{F_n}^\dagger E_{F_n} G_i$, denoted by $\{\tau_{jF_n}^\dagger : n \geq 1\}$ for $j \leq p$, on the main diagonal and zeroes elsewhere, and $C_{F_n}^\dagger$ and $B_{F_n}^\dagger$ are the corresponding $k \times k$ and $p \times p$ orthogonal matrices of singular vectors, as defined in (18.7). Hence, $\lim \Upsilon_n^\dagger$ exists, call it $\Upsilon_h^\dagger$, and equals $h_3^{\dagger\prime} h_4 h_2^\dagger$. That is, the singular value decomposition of $W_h^\dagger h_4$ is

$$W_h^\dagger h_4 = h_3^\dagger \Upsilon_h^\dagger h_2^{\dagger\prime}. \tag{18.34}$$

The $k \times p$ matrix $\Upsilon_h^\dagger$ has the limits of the singular values of $W_{F_n}^\dagger E_{F_n} G_i$ on its main diagonal and zeroes elsewhere. Let $\tau_{h,j}^\dagger$ for $j \leq p$ denote the limits of these singular values. By the definition of $q^\dagger$, $\tau_{h,j}^\dagger = 0$ for $j = q^\dagger + 1, ..., p$ (because $n^{1/2} \tau_{jF_n}^\dagger \to h_{1,j}^\dagger < \infty$). In consequence, $\Upsilon_h^\dagger$ can be written as

$$\Upsilon_h^\dagger = \begin{bmatrix} \Upsilon_{h,q^\dagger}^\dagger & 0^{q^\dagger \times (p-q^\dagger)} \\ 0^{(k-q^\dagger) \times q^\dagger} & 0^{(k-q^\dagger) \times (p-q^\dagger)} \end{bmatrix}, \text{ where } \Upsilon_{h,q^\dagger}^\dagger := Diag\{\tau_{h,1}^\dagger, ..., \tau_{h,q^\dagger}^\dagger\}. \tag{18.35}$$

In addition,

$$h_2^{\dagger\prime} h_{2,p-q^\dagger}^\dagger = \begin{pmatrix} 0^{q^\dagger \times (p-q^\dagger)} \\ I_{p-q^\dagger} \end{pmatrix}. \tag{18.36}$$

Thus, we have

$$
\begin{aligned}
\overline{M}^{\dagger}_{h,p-q^{\dagger}} &:= \overline{W}_h (W_h^{\dagger})^{-1} W_h^{\dagger} h_4 h_{2,p-q^{\dagger}}^{\dagger} = \overline{W}_h (W_h^{\dagger})^{-1} h_3^{\dagger} \Upsilon_h^{\dagger} h_2^{\dagger\prime} h_{2,p-q^{\dagger}}^{\dagger} \\
&= \overline{W}_h (W_h^{\dagger})^{-1} h_3^{\dagger} \left[\begin{array}{cc} \Upsilon^{\dagger}_{h,p-q^{\dagger}} & 0^{q^{\dagger} \times (p-q^{\dagger})} \\ 0^{(k-q^{\dagger}) \times q^{\dagger}} & 0^{(k-q^{\dagger}) \times (p-q^{\dagger})} \end{array} \right] \left(\begin{array}{c} 0^{q^{\dagger} \times (p-q^{\dagger})} \\ I_{p-q^{\dagger}} \end{array} \right) \\
&= 0^{k \times (p-q^{\dagger})},
\end{aligned}
\tag{18.37}
$$

where the first equality holds by the paragraph following Assumption W and uses the condition in Assumption W that $W_h^{\dagger}$ is pd and the second equality holds by (18.35) and (18.36). This completes the proof of Corollary 18.7.

18.6 Proofs of Results Stated in Sections 18.2 and 18.4

For notational simplicity, the proofs in this section are for the sequence $\{n\}$, rather than a subsequence $\{w_n : n \geq 1\}$. The same proofs hold for any subsequence $\{w_n : n \geq 1\}$.

Proof of Theorem 18.1. Theorem 18.1 follows from Theorem 18.6, which imposes Assumption VD, and Lemma 18.2, which verifies Assumption VD when $\widetilde{V}_{Dn}$ is defined by (5.3). $\square$

Proof of Lemma 18.2. Consider any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$. By the CLT result in (18.11), the linear expansion of $n^{1/2}(\widehat{D}_n - E_{F_n} G_i)$ in (14.1), and the definitions of $\overline{g}_h$ and $\overline{D}_h$ in (18.13), we have

$$
n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n} G_i) \rightarrow_d (\overline{g}_h, \overline{D}_h). \tag{18.38}
$$

Next, we apply the delta method to the CLT result in (18.11) and the function $a(\cdot)$ defined in (18.16). The mean component in the lhs quantity in (18.11) is $(0^{(p+1)k\prime}, vech(E_{F_n} f_i^* f_i^{*\prime})')'$. We have

$$
\begin{aligned}
& a \left(\left(\begin{array}{c} 0^{(p+1)k} \\ vech(E_{F_n} f_i^* f_i^{*\prime}) \end{array} \right) \right) \\
&= vech \left(\left(E_{F_n} vec(G_i - E_{F_n} G_i) vec(G_i - E_{F_n} G_i)' - \Gamma_{F_n}^{vec(G_i)} \Omega_{F_n}^{-1} \Gamma_{F_n}^{vec(G_i)\prime} \right)^{-1/2} \right) \\
&= vech \left(\left(\Phi_{F_n}^{vec(G_i)} \right)^{-1/2} \right) = vech(M_{F_n}),
\end{aligned}
\tag{18.39}
$$

where $\Gamma_{F_n}^{vec(G_i)}$ and Ω_{F_n} are defined in (3.2), the first equality uses the definitions of $a(\cdot)$ and f_i^* (given in (18.16) and (5.6), respectively), the second equality holds by the definition of $\Phi_{F_n}^{vec(G_i)}$ in (8.15), and the third equality holds by the definition of M_{F_n} in (18.6). Also, $E_{F_n} f_i^* f_i^{*\prime} \rightarrow h_{10,f^*}$ and h_{10,f^*} is pd. Hence, $a(\cdot)$ is well defined and continuously partially differentiable at

$\lim(0^{(p+1)k\prime}, vech(E_{F_n} f_i^* f_i^{*\prime})')' = (0^{(p+1)k\prime}, vech(h_{10,f^*})')'$, as required for the application of the delta method.

The delta method gives

$$\begin{aligned} n^{1/2}(A_n - vech(M_{F_n})) &= n^{1/2}\left(a\left(n^{-1}\sum_{i=1}^{n}\left(\begin{array}{c} f_i^* \\ vech\left(f_i^* f_i^{*\prime}\right)\end{array}\right)\right) - a\left(\begin{array}{c} 0^{(p+1)k} \\ vech(E_{F_n} f_i^* f_i^{*\prime})\end{array}\right)\right) \\ &\to_d \overline{A}_h \overline{L}_h, \end{aligned} \quad (18.40)$$

where the first equality holds by (18.39) and the definitions of $a(\cdot)$ and A_n in (18.16), the convergence holds by the delta method using the CLT result in (18.11) and the definition of $\overline{A}_h$ following (18.16).

Applying the inverse $vech(\cdot)$ operator, namely, $vech_{kp,kp}^{-1}(\cdot)$, to both sides of (18.40) gives the reconfigured convergence result

$$n^{1/2}(vech_{kp,kp}^{-1}(A_n)) - M_{F_n}) \to_d vech_{kp,kp}^{-1}(\overline{A}_h \overline{L}_h) = \overline{M}_h, \quad (18.41)$$

where the last equality holds by the definition of $\overline{M}_h$ in (18.18).

The convergence results in (18.38) and (18.41) hold jointly because both rely on the convergence result in (18.11).

We show below that

$$n^{1/2}(\widetilde{V}_{Dn} - (vech_{kp,kp}^{-1}(A_n))^{-2}) = o_p(1). \quad (18.42)$$

This and the delta method applied again (using the function $\ell(A) = A^{-1/2}$ for a pd $kp \times kp$ matrix A) give

$$n^{1/2}(\widetilde{V}_{Dn}^{-1/2} - vech_{kp,kp}^{-1}(A_n)) = o_p(1) \quad (18.43)$$

because $vech_{kp,kp}^{-1}(A_n) = (\Phi_h^{vec(G_i)})^{-1/2} + o_p(1)$ and $\Phi_h^{vec(G_i)}$ is pd (because h_{10,f^*} is pd and $\Phi_h^{vec(G_i)} = Qh_{10,f^*}Q'$ for some full row rank matrix Q). Equations (18.38), (18.41), and (18.43) establish the result of the lemma.

Now we prove (18.42). We have

$$
\begin{aligned}
\widetilde{V}_{Dn} &:= n^{-1}\sum_{i=1}^{n} vec(G_i-\widehat{G}_n)vec(G_i-\widehat{G}_n)' - \widehat{\Gamma}_n\widehat{\Omega}_n^{-1}\widehat{\Gamma}_n' \\
&= \left(n^{-1}\sum_{i=1}^{n} vec(G_i-E_{F_n}G_i)vec(G_i-E_{F_n}G_i)'\right) - \left(vec(\widehat{G}_n-E_{F_n}G_i)vec(\widehat{G}_n-E_{F_n}G_i)'\right) \\
&\quad - \left(\widetilde{\Gamma}_n - vec(\widehat{G}_n-E_{F_n}G_i)\widehat{g}_n'\right)\left(\widetilde{\Omega}_n-\widehat{g}_n\widehat{g}_n'\right)^{-1}\left(\widetilde{\Gamma}_n - vec(\widehat{G}_n-E_{F_n}G_i)\widehat{g}_n'\right)' \\
&= n^{-1}\sum_{i=1}^{n} vec(G_i-E_{F_n}G_i)vec(G_i-E_{F_n}G_i)' - \widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}\widetilde{\Gamma}_n' + O_p(n^{-1}), \qquad (18.44)
\end{aligned}
$$

where the second equality holds by subtracting and adding $E_{F_n}G_i$ and some algebra, by the definitions of $\widehat{\Omega}_n$ and $\widehat{\Gamma}_n$ in (4.1), (4.3), and (5.3), and by the definitions of $\widetilde{\Omega}_n$ and $\widetilde{\Gamma}_n$ in (18.16) and the third equality holds because (i) the second summand on the lhs of the third equality is $O_p(n^{-1})$ because $n^{1/2}vec(\widehat{G}_n - E_{F_n}G_i) = O_p(1)$ (by the CLT using the moment conditions in $\mathcal{F}$, defined in (3.1)) and (ii) $n^{1/2}\widehat{g}_n = O_p(1)$ (by Lemma 8.3)), $n^{1/2}vec(\widehat{G}_n - E_{F_n}G_i) = O_p(1)$, and $\widehat{\Gamma}_n = O_p(1)$, $\widehat{\Omega}_n^{-1} = O_p(1)$, $\widetilde{\Gamma}_n = O_p(1)$, and $\widetilde{\Omega}_n^{-1} = O_p(1)$ (by the justification given for (14.1)).

Excluding the $O_p(n^{-1})$ term, the rhs in (18.44) equals $(vech_{kp,kp}^{-1}(A_n))^{-2}$. Hence, (18.42) holds and the proof is complete. $\square$

Proof of Theorem 18.3. The proof is similar to that of Lemma 8.3 in Section 8 with $\widehat{W}_n = W_n = I_k$, $\widehat{U}_n = U_n = I_p$, and the following quantities q, $\widehat{D}_n$, D_n $(= E_{F_n}G_i)$, $B_{n,q}$, $\Upsilon_{n,q}$, C_n, and Υ_n replaced by $q^\dagger$, $\widehat{D}_n^\dagger$, $D_n^\dagger$ $(= D_{F_n}^\dagger)$, $B_{n,q^\dagger}^\dagger$, $\Upsilon_{n,q^\dagger}^\dagger$, $C_n^\dagger$, and $\Upsilon_n^\dagger$, respectively. The proof employs the notational simplifications in (13.1). We can write

$$\widehat{D}_n^\dagger B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} = D_n^\dagger B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} + n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger)B_{n,q^\dagger}^\dagger(n^{1/2}\Upsilon_{n,q^\dagger}^\dagger)^{-1}. \qquad (18.45)$$

By the singular value decomposition, $D_n^\dagger = C_n^\dagger\Upsilon_n^\dagger B_n^{\dagger\prime}$. Thus, we obtain

$$
\begin{aligned}
D_n^\dagger B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} &= C_n^\dagger\Upsilon_n^\dagger B_n^{\dagger\prime}B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} = C_n^\dagger\Upsilon_n^\dagger\begin{pmatrix} I_{q^\dagger} \\ 0^{(p-q^\dagger)\times q^\dagger}\end{pmatrix}(\Upsilon_{n,q^\dagger}^\dagger)^{-1} \\
&= C_n^\dagger\begin{pmatrix} I_{q^\dagger} \\ 0^{(k-q^\dagger)\times q^\dagger}\end{pmatrix} = C_{n,q^\dagger}^\dagger. \qquad (18.46)
\end{aligned}
$$

Let $\widehat{D}_n = (\widehat{D}_{1n}, ..., \widehat{D}_{pn}) \in R^{k \times p}$ and $\overline{D}_h = (\overline{D}_{1h}, ..., \overline{D}_{ph}) \in R^{k \times p}$. We have

$$\begin{aligned} n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger) &= n^{1/2} \sum_{j=1}^p (\widetilde{M}_{1jn} \widehat{D}_{jn} - M_{1jF_n} E_{F_n} G_{ij}, ..., \widetilde{M}_{pjn} \widehat{D}_{jn} - M_{pjF_n} E_{F_n} G_{ij}) \\ &= \sum_{j=1}^p [\widetilde{M}_{1jn} n^{1/2} (\widehat{D}_{jn} - E_{F_n} G_{ij}) + n^{1/2} (\widetilde{M}_{1jn} - M_{1jF_n}) E_{F_n} G_{ij}, ..., \\ &\qquad \widetilde{M}_{pjn} n^{1/2} (\widehat{D}_{jn} - E_{F_n} G_{ij}) + n^{1/2} (\widetilde{M}_{pjn} - M_{pjF_n}) E_{F_n} G_{ij}] \\ &\to_d \sum_{j=1}^p (M_{1jh} \overline{D}_{jh} + \overline{M}_{1jh} h_{4,j}, ..., M_{pjh} \overline{D}_{jh} + \overline{M}_{pjh} h_{4,j}), \end{aligned} \tag{18.47}$$

where the convergence holds by Lemma 8.2 in Section 8, Assumption VD, and $E_{F_n} G_{ij} \to h_{4,j}$ (by the definition of $h_{4,j}$).

Combining (18.45)-(18.47) gives

$$\widehat{D}_n^\dagger B_{n,q^\dagger}^\dagger (\Upsilon_{n,q^\dagger}^\dagger)^{-1} = C_{n,q^\dagger}^\dagger + o_p(1) \to_p h_{3,q^\dagger}^\dagger = \overline{\Delta}_{h,q^\dagger}^\dagger, \tag{18.48}$$

where the equality uses $n^{1/2} \tau_{jF_n}^\dagger \to \infty$ for all $j \leq q^\dagger$ by the definition of $q^\dagger$ and $B_{n,q^\dagger}' B_{n,q^\dagger} = I_{q^\dagger}$, the convergence holds by the definition of $h_{3,q^\dagger}^\dagger$, and the last equality holds by the definition of $\overline{\Delta}_{h,q^\dagger}^\dagger$ in (18.15).

Using the singular value decomposition $D_n^\dagger = C_n^\dagger \Upsilon_n^\dagger B_n^{\dagger\prime}$ again, we obtain

$$\begin{aligned} n^{1/2} D_n^\dagger B_{n,p-q^\dagger}^\dagger &= n^{1/2} C_n^\dagger \Upsilon_n^\dagger B_n^{\dagger\prime} B_{n,p-q^\dagger}^\dagger = n^{1/2} C_n^\dagger \Upsilon_n^\dagger \begin{pmatrix} 0^{q^\dagger \times (p-q^\dagger)} \\ I_{p-q^\dagger} \end{pmatrix} \\ &= C_n^\dagger \begin{pmatrix} 0^{q^\dagger \times (p-q^\dagger)} \\ n^{1/2} \Upsilon_{n,p-q^\dagger}^\dagger \\ 0^{(k-p) \times (p-q^\dagger)} \end{pmatrix} \to h_3^\dagger \begin{pmatrix} 0^{q^\dagger \times (p-q^\dagger)} \\ Diag\{h_{1,q^\dagger+1}^\dagger, ..., h_{1,p}^\dagger\} \\ 0^{(k-p) \times (p-q^\dagger)} \end{pmatrix} = h_3^\dagger h_{1,p-q^\dagger}^{\dagger\diamond}, \end{aligned} \tag{18.49}$$

where the second equality uses $B_n^{\dagger\prime} B_n^\dagger = I_p$, the convergence holds by the definitions of $h_3^\dagger$ and $h_{1,j}^\dagger$ for $j = 1, ..., p$, and the last equality holds by the definition of $h_{1,p-q^\dagger}^{\dagger\diamond}$ in the paragraph following (18.10), which uses (8.17).

By (18.47) and $B_{n,p-q^\dagger}^\dagger \to h_{2,p-q^\dagger}^\dagger$, we have

$$n^{1/2} (\widehat{D}_n^\dagger - D_n^\dagger) B_{n,p-q^\dagger}^\dagger \to_d \overline{D}_h^\dagger h_{2,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger, \tag{18.50}$$

using the definitions of $\overline{D}_h^\dagger$ and $\overline{M}_{h,p-q^\dagger}^\dagger$ in (18.14) and (18.19), respectively.

Using (18.49) and (18.50), we get

$$\begin{aligned} n^{1/2}\widehat{D}_n^\dagger B_{n,p-q^\dagger}^\dagger &= n^{1/2}D_n^\dagger B_{n,p-q^\dagger}^\dagger + n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger)B_{n,p-q^\dagger}^\dagger \\ &\to_d h_3^\dagger h_{1,p-q^\dagger}^{\dagger\diamond} + \overline{D}_h^\dagger h_{2,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger = \overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger, \end{aligned} \tag{18.51}$$

where the last equality holds by the definition of $\overline{\Delta}_{h,p-q^\dagger}^\dagger$ in (18.15).

Equations (18.48) and (18.51) combine to give

$$\begin{aligned} n^{1/2}\widehat{D}_n^\dagger T_n^\dagger &= n^{1/2}\widehat{D}_n^\dagger B_n^\dagger S_n^\dagger = (\widehat{D}_n^\dagger B_{n,q^\dagger}^\dagger (\Upsilon_{n,q^\dagger}^\dagger)^{-1}, n^{1/2}\widehat{D}_n^\dagger B_{n,p-q^\dagger}^\dagger) \\ &\to_d (\overline{\Delta}_{h,q^\dagger}^\dagger, \overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger) = \overline{\Delta}_h^\dagger + \overline{M}_h^\dagger \end{aligned} \tag{18.52}$$

using the definitions of $S_n^\dagger$ and $T_n^\dagger$ in (18.28), $\overline{\Delta}_h^\dagger$ in (18.15), and $\overline{M}_h^\dagger$ in (18.19).

By Lemma 8.2, $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i) \to_d (\overline{g}_h, \overline{D}_h)$. This convergence is joint with that in (18.52) because the latter just relies on the convergence of $n^{1/2}(\widehat{D}_n - E_{F_n}G_i)$, which is part of the former, and of $n^{1/2}(\widetilde{M}_n - M_{F_n}) \to_d \overline{M}_h$, which holds jointly with the former by Assumption VD. This establishes the convergence result of Theorem 18.3.

The independence of $\overline{g}_h$ and $(\overline{D}_h, \overline{\Delta}_h^\dagger)$ follows from the independence of $\overline{g}_h$ and $\overline{D}_h$, which holds by Lemma 8.2, and the fact that $\overline{\Delta}_h^\dagger$ is a nonrandom function of $\overline{D}_h$. $\square$

Proof of Lemma 18.4. The proof of Lemma 18.4 is analogous to the proof of Theorem 8.4 with $\widehat{W}_n = W_n = I_k$, $\widehat{U}_n = U_n = I_p$, and the following quantities q, $\widehat{D}_n$, D_n $(= E_{F_n}G_i)$, $\widehat{\kappa}_{jn}$, B_n, $B_{n,q}$, S_n, $S_{n,q}$, τ_{jF_n}, and $h_{3,q}$ replaced by $q^\dagger$, $\widehat{D}_n^\dagger$, $D_n^\dagger$ $(= D_{F_n}^\dagger)$, $\widehat{\kappa}_{jn}^\dagger$, $B_n^\dagger$, $B_{n,q^\dagger}^\dagger$, $S_n^\dagger$, $S_{n,q^\dagger}^\dagger$, $\tau_{jF_n}^\dagger$, and $h_{3,q^\dagger}^\dagger$, respectively. Theorem 18.3, rather than Lemma 8.3, is employed to obtain the results in (16.37). In consequence, $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are replaced by $\overline{\Delta}_{h,q^\dagger}^\dagger + \overline{M}_{h,q^\dagger}^\dagger$ and $\overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger$, respectively, where $\overline{\Delta}_{h,q^\dagger}^\dagger + \overline{M}_{h,q^\dagger}^\dagger = \overline{\Delta}_{h,q^\dagger}^\dagger$ (because $\overline{M}_{h,q^\dagger}^\dagger := 0^{k\times q^\dagger}$ by (18.19)). The quantities $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are replaced by $\overline{\Delta}_{h,q^\dagger}^\dagger$ and $\overline{\Delta}_{h,p-q^\dagger}^\dagger + \overline{M}_{h,p-q^\dagger}^\dagger$ in (16.37) and in the rest of the proof of Theorem 8.4. Note that (16.39) holds with $h_{3,q}$ replaced by $h_{3,q^\dagger}^\dagger$ because $\overline{\Delta}_{h,q^\dagger}^\dagger = h_{3,q^\dagger}^\dagger$ by (18.15) (just as $\overline{\Delta}_{h,q} = h_{3,q}$). Because $\widehat{U}_n = U_n$, the matrices $\widehat{A}_n$ and A_{jn} for $j = 1, 2, 3$ (defined in (16.39)) are all zero matrices, which simplifies the expressions in (16.41)-(16.44) considerably.

The proof of Theorem 8.4 uses Lemma 16.1 to obtain (16.42). Hence, an analogue of Lemma 16.1 is needed, where the changes listed in the first paragraph of this proof are made and $h_{6,j}$ and C_n are replaced by $h_{6,j}^\dagger$ and $C_n^\dagger$, respectively. In addition, $\mathcal{F}_{WU}$ is replaced by $\mathcal{F}_{KCLR}$ (because $\mathcal{F}_{KCLR} \subset \mathcal{F}_{WU}$ for δ_{WU} sufficiently small and M_{WU} sufficiently large using the facts that $\mathcal{F}_0 \cap \mathcal{F}_{WU}$ equals $\mathcal{F}_0$ for δ_{WU} sufficiently small and M_{WU} sufficiently large by the argument following (8.5) and $\mathcal{F}_{KCLR} \subset \mathcal{F}_0$ by the argument following (18.5)). Because $\widehat{U}_n = U_n$, the matrices $\widehat{A}_{jn}$ for

$j = 1, 2, 3$ (defined in (16.2)) are all zero matrices, which simplifies the expressions in (16.9)-(16.12) considerably. For (16.3) to go through with the changes listed above (in particular, with $\widehat{W}_n$, $\widehat{D}_n$, D_n, and U_n replaced by I_k, $\widehat{D}_n^\dagger$, $D_n^\dagger$, and I_p, respectively), we need to show that

$$n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger) = O_p(1). \tag{18.53}$$

By (5.4) with $\theta = \theta_0$ (and with the dependence of various quantities on θ_0 suppressed for notational simplicity), we have

$$\widehat{D}_n^\dagger = \sum_{j=1}^{p}(\widetilde{M}_{1jn}\widehat{D}_{jn}, ..., \widetilde{M}_{pjn}\widehat{D}_{jn}), \text{ where } \widetilde{M}_n = \begin{bmatrix} \widetilde{M}_{11n} & \cdots & \widetilde{M}_{1pn} \\ \vdots & \ddots & \vdots \\ \widetilde{M}_{p1n} & \cdots & \widetilde{M}_{ppn} \end{bmatrix} := \widetilde{V}_{Dn}^{-1/2} \in R^{kp\times kp}. \tag{18.54}$$

By (18.6), we have

$$D_n^\dagger = \sum_{j=1}^{p}(M_{1jF_n}D_{jn}, ..., M_{pjF_n}D_{jn}) \tag{18.55}$$

using $D_n = (D_{1n}, ..., D_{pn})$, and $D_{jn} := E_{F_n}G_{ij}$ for $j = 1, ..., p$.

For $s = 1, ..., p$, we have

$$n^{1/2}(\widetilde{M}_{sjn}\widehat{D}_{jn} - M_{sjF_n}D_{jn}) = \widetilde{M}_{sjn}n^{1/2}(\widehat{D}_{jn} - D_{jn}) + n^{1/2}(\widetilde{M}_{sjn} - M_{sjF_n})D_{jn} = O_p(1), \tag{18.56}$$

where $n^{1/2}(\widehat{D}_{jn} - D_{jn}) = O_p(1)$ (by Lemma 8.2), $n^{1/2}(\widetilde{M}_{sjn} - M_{sjF_n}) = O_p(1)$ (because $n^{1/2}(\widetilde{M}_n - M_{F_n}) \to_d \overline{M}_h$ by Assumption VD), $M_{sjF_n} = O(1)$ (because $M_F = (\Phi_F^{vec(G_i)})^{-1/2}$, $\Phi_F^{vec(G_i)}$ defined in (8.15) satisfies $\Phi_F^{vec(G_i)} := Var_F(vec(G_i) - \Gamma_F^{vec(G_i)}\Omega_F^{-1}g_i) = [-E_F vec(G_i)g_i'\Omega_F^{-1} : I_{pk}]Var_F(f_i^*)$, and $\lambda_{\min}(Var_F(f_i^*)) \geq \delta_2$ by the definition of $\mathcal{F}_{KCLR}$ in (18.5)), and $D_{jn} = O(1)$ (by the moment conditions in $\mathcal{F}$, defined in (3.1)).

Hence,

$$n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger) = \sum_{j=1}^{p} n^{1/2}[(\widetilde{M}_{1jn}\widehat{D}_{jn}, ..., \widetilde{M}_{pjn}\widehat{D}_{jn}) - (M_{1jF_n}D_{jn}, ..., M_{pjF_n}D_{jn})] = O_p(1). \tag{18.57}$$

This completes the proof of the analogue of Lemma 16.1, which completes the proof of parts (a)-(d) of Lemma 18.4.

For part (e) of Lemma 18.4, the results of parts (a)-(d) hold jointly with those in Theorem 18.3, rather than those in Lemma 8.3, because Theorem 18.3 is used to obtain the results in (16.37), rather than Lemma 8.3. This completes the proof. □

Proof of Lemma 18.5. The proof of parts (a) and (b) is the same as the proof of Theorem 10.1 for the case where Assumption R(a) holds (which states that $rk_n \to_p \infty$) using Lemma 18.4(a), which shows that $rk_n^\dagger \to_d \infty$ if $q^\dagger = p$.

The proofs of parts (c) and (d) are the same as in (10.5)-(10.9) in the proof of Theorem 10.1 for the case where Assumption R(b) holds, using Theorem 18.3 and Lemma 18.4(b) in place of Lemma 8.3, with $r_h(\overline{D}_h, \overline{M}_h)$ (defined in (18.20)) in place of $r_h(\overline{D}_h)$, and for part (d), with the proviso that $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$. (The proof in Theorem 10.1 that $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$ does not go through in the present case because $\overline{r}_h = r_h(\overline{D}_h, \overline{M}_h)$ is not necessarily a constant conditional on $\overline{D}_h$ and alternatively, conditional on $(\overline{D}_h, \overline{M}_h)$, $\overline{LM}_h$ and $\overline{J}_h$ are not necessarily independent and distributed as χ^2_p and χ^2_{k-p}.) Note that (10.10) does not necessarily hold in the present case, because $\overline{r}_h = r_h(\overline{D}_h, \overline{M}_h)$ is not necessarily a constant conditional on $\overline{D}_h$. $\square$

The proof of Theorem 18.6 given below uses Corollary 2.1(a) of ACG, which is stated below as Proposition 18.8. It is a generic asymptotic size result. Unlike Proposition 8.1 above, Proposition 18.8 applies when the asymptotic size is not necessarily equal to the nominal size α. Let $\{\phi_n : n \geq 1\}$ be a sequence of tests of some null hypothesis whose null distributions are indexed by a parameter λ with parameter space Λ. Let $RP_n(\lambda)$ denote the null rejection probability of ϕ_n under λ. For a finite nonnegative integer J, let $\{h_n(\lambda) = (h_{1n}(\lambda), ..., h_{Jn}(\lambda))' \in R^J : n \geq 1\}$ be a sequence of functions on Λ. Define H as in (8.1).

For a sequence of scalar constants $\{C_n : n \geq 1\}$, let $C_n \to [C_{1,\infty}, C_{2,\infty}]$ denote that $C_{1,\infty} \leq \liminf_{n\to\infty} C_n \leq \limsup_{n\to\infty} C_n \leq C_{2,\infty}$.

Assumption B: For any subsequence $\{w_n\}$ of $\{n\}$ and any sequence $\{\lambda_{w_n} \in \Lambda : n \geq 1\}$ for which $h_{w_n}(\lambda_{w_n}) \to h \in H$, $RP_{w_n}(\lambda_{w_n}) \to [RP^-(h), RP^+(h)]$ for some $RP^-(h)$, $RP^+(h) \in [0, 1]$.

Proposition 18.8 (ACG, Corollary 2.1(a)) *Under Assumption* B*, the tests* $\{\phi_n : n \geq 1\}$ *have* $AsySz := \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda) \in [\sup_{h\in H} RP^-(h), \sup_{h\in H} RP^+(h)]$.

Comments: (i) Corollary 2.1(a) of ACG is stated for confidence sets, rather than tests. But, following Comment 4 to Theorem 2.1 of ACG, with suitable adjustments (as in Proposition 18.8 above) it applies to tests as well.

(ii) Under Assumption B, if $RP^-(h) = RP^+(h)$ for all $h \in H$, then $AsySz = \sup_{h\in H} RP^+(h)$. We use this to prove Theorem 18.6. The result of Proposition 18.8 for the case where $RP^-(h) \neq RP^+(h)$ for some $h \in H$ is used when proving Comment (i) to Theorem 18.1 and the Comment to Theorem 18.6.

Proof of Theorem 18.6. Theorem 18.6 follows from Lemma 18.5 and Proposition 18.8 because

Lemma 18.5 verifies Assumption B with $RP^{-}(h) = RP^{+}(h) = \alpha$ when $q^{\dagger} = p$ and with $RP^{-}(h) = RP^{+}(h) = P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))$ when $q^{\dagger} < p$. $\square$

19 Proof of Theorem 7.1

Theorem 7.1 of AG1. *Suppose the* LM *test, the* CLR *test with moment-variance weighting, and when* $p = 1$ *the* CLR *test with Jacobian-variance weighting are defined as in this section, the parameter space for* F *is* $\mathcal{F}_{TS,0}$ *for the first two tests and* $\mathcal{F}_{TS,JVW,p=1}$ *for the third test, and Assumption* V *holds. Then, these tests have asymptotic sizes equal to their nominal size* $\alpha \in (0,1)$ *and are asymptotically similar (in a uniform sense). Analogous results hold for the corresponding* CS's *for the parameter spaces* $\mathcal{F}_{\Theta,TS,0}$ *and* $\mathcal{F}_{\Theta,TS,JVW,p=1}$.

The proof of Theorem 7.1 is analogous to that of Theorems 4.1, 5.2, and 6.1. In the time series case, for tests, we define $\lambda = (\lambda_{1,F}, ..., \lambda_{9,F})$ and $\{\lambda_{n,h} : n \geq 1\}$ as in (8.9) and (8.11), respectively, but with $\lambda_{5,F}$ defined differently than in the i.i.d. case. (For CS's in the time series case, we make the adjustments outlined in the Comment to Proposition 8.1.) We define[63]

$$\lambda_{5,F} := V_F = \sum_{m=-\infty}^{\infty} E_F \begin{pmatrix} g_i \\ vec(G_i - E_F G_i) \end{pmatrix} \begin{pmatrix} g_{i-m} \\ vec(G_{i-m} - E_F G_{i-m}) \end{pmatrix}'. \qquad (19.1)$$

In consequence, $\lambda_{5,F_n} \to h_5$ implies that $V_{F_n} \to h_5$ and the condition in Assumption V holds with $V = h_5$.

The proof of Theorem 7.1 uses the CLT given in the following lemma.

Lemma 19.1 *Let* $f_i := (g_i', vec(G_i)')'$. *We have:* $w_n^{-1/2} \sum_{i=1}^{w_n} (f_i - E_{F_n} f_i) \to_d N(0^{(p+1)k}, h_5)$ *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$.

Proof of Theorem 7.1. The proof is the same as the proofs of Theorems 4.1, 5.2, and 6.1 (given in Sections 9, 10, and 11, respectively, in the Appendix to AG1) and the proofs of Lemmas 8.2 and 8.3 and Theorem 8.4 (given in Sections 14, 15, and 16 in this Supplemental Material), upon which the former proofs rely, for the i.i.d. case with some modifications. The modifications affect the proofs of Lemmas 8.2 and 8.3 and the proof of Theorem 5.2. No modifications are needed elsewhere.

The first modification is the change in the definition of $\lambda_{5,F}$ described in (19.1).

[63] The difference in the definitions of $\lambda_{5,F}$ in the i.i.d. and time series cases reflects the difference in the definitions of $\Sigma_F^{vec(G_i)}$ in these two cases. See the footnote at (7.1) above regarding the latter.

The second modification is that $\widehat{\Omega}_n = \widehat{\Omega}_n(\theta_0) \to_p h_{5,g}$ not by the WLLN but by Assumption V and the definition of $\widehat{\Omega}_n(\theta)$ in (7.4). In the time series case, by definition, $\lambda_{5,F} := V_F$, so $h_5 := \lim \lambda_{5,F_n} = \lim V_{F_n}$. By definition, $h_{5,g}$ is the upper left $k \times k$ submatrix of h_5 and Ω_F is the upper left $k \times k$ submatrix of V_F by (7.1) and (19.1). Hence, $h_{5,g} = \lim \Omega_{F_n}$. By the definition of $\mathcal{F}_{TS}$, $\lambda_{\min}(\Omega_F) \geq \delta \; \forall F \in \mathcal{F}_{TS}$. Hence, $h_{5,g}$ is pd.

Let $h_{5,G_j g}$ be the $k \times k$ submatrix of h_5 that corresponds to the submatrix $\widehat{\Gamma}_{jn}(\theta)$ of $\widehat{V}_n(\theta)$ in (7.4) for $j = 1, ..., p$. The third modification is that $\widehat{\Gamma}_{jn} = \widehat{\Gamma}_{jn}(\theta_0) = h_{5,G_j g} + o_p(1)$ in (14.1) in the proof of Lemma 8.2 (rather than $\widehat{\Gamma}_{jn} = E_{F_n} G_{ij} g_i' + o_p(1)$) for $j = 1, ..., p$ and this holds by Assumption V and the definition of $\widehat{\Gamma}_{jn}(\theta)$ in (7.4) (rather than by the WLLN).

We write

$$h_5 = \begin{pmatrix} h_{5,g} & h_{5,Gg}' \\ h_{5,Gg} & h_{5,G} \end{pmatrix} \text{ for } h_{5,g} \in R^{k \times k}, \; h_{5,Gg} = \begin{pmatrix} h_{5,G_1 g} \\ \vdots \\ h_{5,G_p g} \end{pmatrix} \in R^{pk \times k}, \text{ and } h_{5,G} \in R^{pk \times pk}. \tag{19.2}$$

The fourth modification is that $\widetilde{V}_{Dn}$ in (11.1) in the proof of Theorem 5.2 is defined as described in Section 7, rather than as in (5.3). In addition, $\widetilde{V}_{Dn} \to_p h_7$ in (11.1) holds with $h_7 = h_{5,G} - h_{5,Gg}(h_{5,g})^{-1} h_{5,Gg}'$ by Assumption V, rather than by the WLLN.

The fifth modification is the use of a WLLN and CLT for triangular arrays of strong mixing random vectors, rather than i.i.d. random vectors, for the quantities in the proof of Lemma 8.2 and elsewhere. For the WLLN, we use Example 4 of Andrews (1988), which shows that for a strong mixing row-wise-stationary triangular array $\{W_i : i \leq n\}$ we have $n^{-1} \sum_{i=1}^n (\xi(W_i) - E_{F_n} \xi(W_i)) \to_p 0$ for any real-valued function $\xi(\cdot)$ (that may depend on n) for which $\sup_{n \geq 1} E_{F_n} ||\xi(W_i)||^{1+\delta} < \infty$ for some $\delta > 0$. For the CLT, we use Lemma 19.1 as follows. The joint convergence of $n^{1/2} \widehat{g}_n$ and $n^{1/2}(\widehat{D}_n - E_{F_n} G_i)$ in the proof of Lemma 8.2 is obtained from (14.1), modified by the second and third modifications above, and the following result:

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n (\zeta(W_i) - E_{F_n} \zeta(W_i)) = \begin{pmatrix} I_k & 0^{k \times pk} \\ -h_{5,Gg} h_{5,g}^{-1} & I_{pk} \end{pmatrix} n^{-1/2} \sum_{i=1}^n (f_i - E_{F_n} f_i) \\ & \to_d N(0^{(p+1)k}, L_{h_5}), \text{ where} \\ & \zeta(W_i) := \begin{pmatrix} g_i \\ vec(G_i) - h_{5,Gg} h_{5,g}^{-1} g_i \end{pmatrix} = \begin{pmatrix} I_k & 0^{k \times pk} \\ -h_{5,Gg} h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}, \end{aligned} \tag{19.3}$$

$f_i = (g_i', vec(G_i)')'$, and the convergence holds by Lemma 19.1. Using (19.2), the variance matrix

L_{h_5} in (19.3) takes the form:

$$\begin{aligned} L_{h_5} &= \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} h_{5,g} & h_{5,Gg'} \\ h_{5,Gg} & h_{5,G} \end{pmatrix} \begin{pmatrix} I_k & -h_{5,g}^{-1}h_{5,Gg}' \\ 0^{pk\times k} & I_{pk} \end{pmatrix} \\ &= \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} h_{5,g} & 0^{k\times pk} \\ h_{5,Gg} & \Phi_h^{vec(G_i)} \end{pmatrix} = \begin{pmatrix} h_{5,g} & 0^{k\times pk} \\ 0^{pk\times k} & \Phi_h^{vec(G_i)} \end{pmatrix}, \text{ where} \\ \Phi_h^{vec(G_i)} &= h_{5,G} - h_{5,Gg}h_{5,g}^{-1}h_{5,Gg}'. \end{aligned} \tag{19.4}$$

Equations (14.1) (modified as described above), (19.3), and (19.4) combine to give the result of Lemma 8.2 for the time series case.

The sixth modification occurs in the proof of Lemma 8.3(d) in Section 15 in this Supplemental Material. In the time series case, the proof goes through as is, except that the calculations in (15.13) are not needed because $\Sigma_F^{a_i}$ (and, hence, $\Psi_F^{a_i}$ as well) is defined with its underlying components re-centered at their means (which is needed to ensure that $\Sigma_F^{a_i}$ is a convergent sum). The latter implies that $\lim \Psi_{F_n}^{vec(G_i)} = \Phi_h^{vec(G_i)}$ automatically holds and $\lim \Psi_{F_n}^{vec(C_{F_n,k-q}'\Omega_{F_n}^{-1/2}G_iB_{F_n,p-q}\xi_2)} = \Phi_h^{vec(h_{3,k-q}'h_{5,g}^{-1/2}G_ih_{2,p-q}\xi_2)}$ (which, in the i.i.d. case, is proved in (15.13).

This completes the proof of Theorem 7.1. $\square$

Proof of Lemma 19.1. For notational simplicity, we prove the result for the sequence $\{n\}$ rather than a subsequence $\{w_n : n \geq 1\}$. The same proof applies for any subsequence. By the Cramér-Wold device, it suffices to prove the result with $f_i - E_{F_n}f_i$ and h_5 replaced by $s(W_i) = b'(f_i - E_{F_n}f_i)$ and $b'h_5b$, respectively, for arbitrary $b \in R^{(p+1)k}$. First, we show

$$\lim Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) = b'h_5b, \tag{19.5}$$

where by assumption $\lambda_{5,F_n} = \sum_{m=-\infty}^{\infty} E_{F_n}s(W_i)s(W_{i-m}) \to h_5$. By change of variables, we have

$$Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) = \sum_{m=-n+1}^{n-1} Cov_{F_n}(s(W_i), s(W_{i-m})) - \sum_{m=-n+1}^{n-1} \frac{|m|}{n} Cov_{F_n}(s(W_i), s(W_{i-m})). \tag{19.6}$$

This gives

$$\begin{aligned} &\left\| Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) - b'\lambda_{5,F_n}b \right\| \\ &\leq 2\sum_{m=n}^{\infty} ||Cov_{F_n}(s(W_i), s(W_{i-m}))|| + \sum_{m=-n+1}^{n-1} \frac{|m|}{n} ||Cov_{F_n}(s(W_i), s(W_{i-m}))||. \end{aligned} \tag{19.7}$$

By a standard strong mixing covariance inequality, e.g., see Davidson (1994, p. 212),

$$\sup_{F \in \mathcal{F}_{TS}} ||Cov_F(s(W_i), s(W_{i-m}))|| \leq C_1 \alpha_F^{\gamma/(2+\gamma)}(m) \leq C_1 C^{\gamma/(2+\gamma)} m^{-d\gamma/(2+\gamma)}, \text{ where } d\gamma/(2+\gamma) > 1, \tag{19.8}$$

for some $C_1 < \infty$, where the second inequality uses the definition of $\mathcal{F}_{TS}$ in (7.2). In consequence, both terms on the rhs of (19.7) converge to zero. This and $b'\lambda_{5,F_n} b \to b'h_5 b$ establish (19.5).

When $b'h_5 b = 0$, we have $\lim_{n\to\infty} Var_{F_n}(n^{-1/2}\sum_{i=1}^n s(W_i)) = 0$, which implies that $n^{-1/2}\sum_{i=1}^n s(W_i) \to_d N(0, b'h_5 b) = 0$. When $b'h_5 b > 0$, we can assume $\sigma_n^2 = Var_{F_n}(n^{-1/2}\sum_{i=1}^n s(W_i)) \geq c$ for some $c > 0$ $\forall n \geq 1$ without loss of generality. We apply the triangular array CLT in Corollary 1 of de Jong (1997) with (using de Jong's notation) $\beta = \gamma = 0$, $c_{ni} := n^{-1/2}\sigma_n^{-1}$, and $X_{ni} := n^{-1/2}s(W_i)\sigma_n^{-1}$. Now we verify conditions (a)-(c) of Assumption 2 of de Jong (1997). Condition (a) holds automatically. Condition (b) holds because $c_{ni} > 0$ and $E_{F_n}|X_{ni}/c_{ni}|^{2+\gamma} = E_{F_n}|s(W_i)|^{2+\gamma} \leq 2||b||^{2+\gamma}M < \infty$ $\forall F_n \in \mathcal{F}_{TS}$. Condition (c) holds by taking $V_{ni} = X_{ni}$ (where V_{ni} is the random variable that appears in the definition of near epoch dependence in Definition 2 of de Jong (1997)), $d_{ni} = 0$, and using $\alpha_{F_n}(m) \leq Cm^{-d}$ $\forall F_n \in \mathcal{F}_{TS}$ for $d > (2+\gamma)/\gamma$ and $C < \infty$. By Corollary 1 of de Jong (1997), we have $X_{ni} \to_d N(0,1)$. This and (19.5) give

$$n^{-1/2}\sum_{i=1}^n s(W_i) \to_d N(0, b'h_5 b), \tag{19.9}$$

as desired. $\square$

References

Andrews, D. W. K. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory,* 4, 458–467.

Davidson, J. (1994): *Stochastic Limit Theory.* Oxford: Oxford University Press.

de Jong, R. M. (1997): "Central Limit Theorems for Dependent Heterogeneous Random Variables," *Econometric Theory,* 13, 353–367.

Hwang, S.-G. (2004): "Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices," *American Mathematical Monthly,* 111, 157–159.

Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica,* 59, 1551–1580.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica,* 73, 1103–1123.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley.

Robin, J.-M., and R. J. Smith (2000): "Tests of Rank," *Econometric Theory,* 16, 151–175.

Stewart, G. W. (2001): *Matrix Algorithms Volume II*: *Eigensystems.* Philadelphia: SIAM.