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: August 22, 2016

Contents

We let AG1 abbreviate the main paper Andrews and Guggenberger (2017) "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," *Econometric Theory,* forthcoming. References to Sections with Section numbers less than 9 refer to Sections of AG1. In consequence, the Section numbers in this Supplemental Material (SM) follow on from the main paper, starting with Section 9. Similarly, all theorems and lemmas with Section numbers less than 9 refer to results in AG1.

We let ACG abbreviate Andrews, Cheng, and Guggenberger (2009) "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper No. 1813, Yale University. This SM makes use of some results in ACG.

We let AG2 abbreviate Andrews and Guggenberger (2014) "Identification- and Singularity-Robust Inference for Moment Condition Models," Cowles Foundation Discussion Paper No. 1978, Yale University. AG2 utilizes some of the results in this SM.

9 Outline

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

Section 10 states some basic results that are used in all of the proofs. These results also are used in AG2 and should be useful for establishing the asymptotic sizes of other tests for moment condition models when strong identification is not assumed. Given the results in Section 10, Section 11 proves Theorem 4.1, Section 12 proves Theorem 6.1, and Section 13 proves Theorem 5.3.

Section 14 shows that the eigenvalue condition in $\mathcal{F}_0$, defined in (3.9), is not redundant in Theorems 4.1, 5.3, and 6.1.

Sections 15, 16, and 17 prove Lemma 10.2, Lemma 10.3, and Theorem 10.4, respectively, which appear in Section 10.

Section 18 proves that the conditions in (3.10) and (3.11) are sufficient for the second condition in $\mathcal{F}_{0j}$.

Section 19 proves Theorem 5.1 and Lemma 5.2. Section 19 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 19 provides some simulation results for this test.

Section 20 proves Theorem 7.1, which provides results for time series observations.

For notational simplicity, throughout the SM, we often suppress the argument θ_0 for various quantities that depend on the null value θ_0. Throughout the SM, the quantities B_F, C_F, and $(\tau_{1F}, ..., \tau_{pF})$ are defined using the general definitions given in (10.6)-(10.8), rather than the definitions given in Section 3, which are a special case of the former definitions.

For notational simplicity, the proofs in Sections 15-17 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

$$D_n := E_{F_n} G_i,\ \Omega_n := \Omega_{F_n},\ B_n := B_{F_n},\ C_n := C_{F_n},\ B_n = (B_{n,q}, B_{n,p-q}),\ C_n = (C_{n,q}, C_{n,k-q}),$$
$$W_n := W_{F_n},\ W_{2n} := W_{2F_n},\ U_n := U_{F_n},\ \text{and } U_{2n} := U_{2F_n}, \tag{9.1}$$

where $q = q_h$ is defined in (10.16), $B_{n,q} \in R^{p \times q}$, $B_{n,p-q} \in R^{p \times (p-q)}$, $C_{n,q} \in R^{k \times q}$, and $C_{n,k-q} \in$

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

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

$$\Upsilon_n := \begin{bmatrix} \Upsilon_{n,q} & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & \Upsilon_{n,p-q} \\ 0^{(k-p)\times q} & 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times p}. \tag{9.2}$$

Note that Υ_n is the diagonal matrix of singular values of $W_n D_n U_n$, see (10.8).

10 Basic Framework and Results for the Proofs

10.1 Uniformity

The proofs of Theorems 4.1, 5.3, and 6.1 use Corollary 2.1(c) in ACG. The latter result provides general sufficient conditions for the correct asymptotic size and (uniform) asymptotic similarity of a sequence of tests.

We now state Corollary 2.1(c) of ACG. 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

$$\begin{aligned} H := \{h \in (R \cup \{\pm\infty\})^J &: h_{w_n}(\lambda_{w_n}) \to h \text{ for some subsequence } \{w_n\} \\ &\text{of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\}. \end{aligned} \tag{10.1}$$

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

Proposition 10.1 (ACG, Corollary 2.1(c)) *Under Assumption* B*, *the tests* $\{\phi_n : n \geq 1\}$ *have asymptotic size* α *and are asymptotically similar* (*in a uniform sense*). *That is,* $AsySz := \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda) = \alpha$ *and* $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} RP_n(\lambda) = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda)$.

Comments: (i) By Comment 4 to Theorem 2.1 of ACG, Proposition 10.1 provides asymptotic size and similarity results for nominal $1-\alpha$ confidence sets (CS's), rather than tests, by defining λ as one would for a test, but having it depend also on the parameter that is restricted by the null hypothesis, by enlarging the parameter space Λ correspondingly (so it includes all possible values of the parameter that is restricted by the null hypothesis), and by replacing (i) ϕ_n by a CS based on a sample of size n, (ii) α by $1-\alpha$, (iii) $RP_n(\lambda)$ by $CP_n(\lambda)$, where $CP_n(\lambda)$ denotes the coverage probability of the CS under λ when the sample size is n, and (iv) the first $\limsup_{n\to\infty}\sup_{\lambda\in\Lambda}$ that appears by $\liminf_{n\to\infty}\inf_{\lambda\in\Lambda}$. In the present case, where the null hypotheses are of the form $H_0 : \theta = \theta_0$ for $\theta \in \Theta$, for CS's, θ_0 is taken to be a subvector of λ and Λ is specified so that the value of this subvector ranges over Θ.

(ii) In the application of Proposition 10.1 to prove Theorems 4.1 and 6.1, one takes Λ to be a one-to-one transformation of $\mathcal{F}_0$ for tests, and one takes Λ to be a one-to-one transformation of $\mathcal{F}_{\Theta,0}$ for CS's. With these changes, the proofs for tests and CS's are the same. In consequence, we provide explicit proofs for tests only and obtain the proofs for CS's by analogous applications of Proposition 10.1. In the application of Proposition 10.1 to prove Theorem 5.3, the same is done but with $\mathcal{F}_{JVW,p=1}$ in place of $\mathcal{F}_0$.

(iii) We prove the test results in Theorems 4.1, 5.3, and 6.1 using Proposition 10.1 by verifying Assumption B* for suitable choices of λ and $h_n(\lambda)$.

10.2 Random Weight Matrices $\widehat{\mathbf{W}}_{\mathbf{n}}$ and $\widehat{\mathbf{U}}_{\mathbf{n}}$

We prove results for statistics that depend on random weight matrices $\widehat{W}_n \in R^{k\times k}$ and $\widehat{U}_n \in R^{p\times p}$. In particular, we consider statistics of the form $\widehat{W}_n\widehat{D}_n\widehat{U}_n$ and functions of this statistic, where $\widehat{D}_n$ is defined in (3.2). The definitions of the random weight matrices $\widehat{W}_n$ and $\widehat{U}_n$ depend upon the statistic that is of interest. They are taken to be of the form

$$\widehat{W}_n := W_1(\widehat{W}_{2n}) \in R^{k\times k} \text{ and } \widehat{U}_n := U_1(\widehat{U}_{2n}) \in R^{p\times p}, \tag{10.2}$$

where $\widehat{W}_{2n}$ and $\widehat{U}_{2n}$ are random finite-dimensional quantities, such as matrices, and $W_1(\cdot)$ and $U_1(\cdot)$ are nonrandom functions that are assumed below to be continuous on certain sets. The estimators $\widehat{W}_{2n}$ and $\widehat{U}_{2n}$ have corresponding population quantities W_{2F} and U_{2F}, respectively. For examples, see Examples 1-3 immediately below. Thus, the population quantities corresponding to $\widehat{W}_n$ and $\widehat{U}_n$ are

$$W_F := W_1(W_{2F}) \text{ and } U_F := U_1(U_{2F}), \tag{10.3}$$

respectively.

Example 1: With Kleibergen's (2005) LM test and the CLR test with moment-variance weighting, which are considered in Sections 4 and 6, respectively, we take

$$\widehat{W}_n = \widehat{\Omega}_n^{-1/2} \text{ and } \widehat{U}_n = I_p. \tag{10.4}$$

In this case, the functions $W_1(\cdot)$ and $U_1(\cdot)$ are the identity functions, and the corresponding population quantities are $W_F = W_{2F} = \Omega_F^{-1/2}$, where $\Omega_F := E_F g_i g_i'$, see (3.6), and $U_F = U_{2F} = I_p$.

Example 2: For a CLR test based on an equally-weighted statistic other than $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$, such as $\widetilde{W}_n\widehat{D}_n$, as in Comment (ii) to Theorem 6.1, one defines a pd matrix $\widetilde{W}_n$ as desired and one takes $\widehat{W}_n = \widetilde{W}_n$ and $\widehat{U}_n = U_F = U_{2F} = I_p$.

Example 3: With Kleibergen's (2005) CLR test with Jacobian-variance weighting and $p = 1$, which is considered in Section 5, we determine the asymptotic distribution of the rank statistic in (5.10) by taking $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$ and $\widehat{U}_n = I_p$. In this case, the functions $W_1(\cdot)$ and $U_1(\cdot)$ are as in Example 1, and the corresponding population quantities are $W_F = W_{2F} = (Var_F(vec(G_i)) - \Gamma_F^{vec(G_i)}\Omega_F^{-1}\Gamma_F^{vec(G_i)\prime})^{-1/2} = (\Psi_F^{vec(G_i)} - E_F G_i E_F G_i')^{-1/2}$, and $U_F = U_{2F} = I_p$. For this test, we need the asymptotic distribution of the LM statistic. In consequence, for this test, we also establish some asymptotic results with $\widehat{W}_n$ and $\widehat{U}_n$ defined as in Example 1.

Examples 4 & 5: The results of this section are used in AG2 when the asymptotic sizes of two new SR-CQLR tests are determined. For the SR-CQLR tests, $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and it is convenient to take $W_1(\cdot) = (\cdot)^{-1/2}$ and $\widehat{W}_{2n} = \widehat{\Omega}_n$, and the matrix $\widehat{U}_n$ is a nonlinear transformation $U_1(\cdot)$ of a matrix estimator, which is different for the two tests. For brevity, we do not define the nonlinear transformation or the two matrix estimators here.

We provide results for distributions F in the following set of null distributions:

$$\mathcal{F}_{WU} := \{F \in \mathcal{F} : \lambda_{\min}(W_F) \geq \delta_{WU}, \lambda_{\min}(U_F) \geq \delta_{WU}, ||W_F|| \leq M_{WU}, \text{ and } ||U_F|| \leq M_{WU}\} \tag{10.5}$$

for some constants $\delta_{WU} > 0$ and $M_{WU} < \infty$, where $\mathcal{F}$ is defined in (3.3). The set $\mathcal{F}_{WU} \cap \mathcal{F}_0$ is used to establish results for Kleibergen's LM and the CLR test with moment-variance weighting, considered in Section 6, using the fact that $\mathcal{F}_0 = \mathcal{F}_{WU} \cap \mathcal{F}_0$ for $\delta_{WU} > 0$ sufficiently small and $M_{WU} < \infty$ sufficiently large. This holds because for all $F \in \mathcal{F}_0$, $\lambda_{\min}(W_F) = \lambda_{\min}(\Omega_F^{-1/2}) = \lambda_{\max}^{-1/2}(\Omega_F) \geq ||\Omega_F||^{-1/2} \geq M_*^{-1/2}$ for some $M_* < \infty$ (because $||\Omega_F|| = ||E_F g_i g_i'|| \leq M_*$ for some $M_* < \infty$ by the moment conditions in $\mathcal{F}$), $||W_F|| = ||\Omega_F^{-1/2}|| \leq \lambda_{\min}^{-1/2}(\Omega_F) \leq \delta^{-1/2}$ (using the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}$), where $\delta > 0$, $\lambda_{\min}(U_F) = \lambda_{\min}(I_p) = 1$, and $||U_F|| = ||I_p|| = p$.

10.3 Reparametrization

To apply Proposition 10.1, we reparametrize the null distribution F to a vector λ. The vector λ is chosen such that for a subvector of λ convergence of a drifting subsequence of the subvector (after suitable renormalization) yields convergence in distribution of the test statistic and convergence in distribution of the critical value in the case of the CLR tests.

To be consistent with the use of general weight matrices $\widehat{W}_n$ and $\widehat{U}_n$ in this section, we provide more general definitions of τ_{jF}, B_F, and C_F here than are given in Section 3. These general definitions reduce to the definitions given in Section 3 when $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$.

The vector λ depends on the following quantities. Let

$$B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } U_F'(E_F G_i)' W_F' W_F (E_F G_i) U_F \tag{10.6}$$

ordered so that the corresponding eigenvalues $(\kappa_{1F}, ..., \kappa_{pF})$ are nonincreasing. The matrix B_F is such that the columns of $W_F(E_F G_i)U_F B_F$ are orthogonal. Let

$$C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } W_F(E_F G_i)U_F U_F'(E_F G_i)' W_F' \tag{10.7}$$

ordered so that the corresponding eigenvalues are $(\kappa_{1F}, ..., \kappa_{pF}, 0, ..., 0) \in R^k$. The matrices B_F and C_F are not uniquely defined. We let B_F denote one choice of the matrix of eigenvectors of $U_F'(E_F G_i)' W_F' W_F (E_F G_i) U_F$ and analogously for C_F. Let

$$(\tau_{1F}, ..., \tau_{pF}) \text{ denote the } p \text{ singular values of } W_F(E_F G_i)U_F, \tag{10.8}$$

which are nonnegative, ordered so that τ_{jF} is nonincreasing. (Some of these singular values may be zero.) As is well-known, the squares of the p singular values of a $k \times p$ matrix A with $k \geq p$ equal the p eigenvalues of $A'A$ and the largest p eigenvalues of AA'. In consequence, $\kappa_{jF} = \tau_{jF}^2$ for $j = 1, ..., p$.

Define the elements of λ to be

$$\begin{aligned}
\lambda_{1,F} &:= (\tau_{1F}, ..., \tau_{pF})' \in R^p, \\
\lambda_{2,F} &:= B_F \in R^{p \times p}, \\
\lambda_{3,F} &:= C_F \in R^{k \times k}, \\
\lambda_{4,F} &:= (E_F G_{i1}, ..., E_F G_{ip}) \in R^{k \times p}, \\
\lambda_{5,F} &:= E_F \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix} \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}' \in R^{(p+1)k \times (p+1)k}, \\
\lambda_{6,F} &= (\lambda_{6,1F}, ..., \lambda_{6,(p-1)F})' := (\frac{\tau_{2F}}{\tau_{1F}}, ..., \frac{\tau_{pF}}{\tau_{(p-1)F}})' \in R^{p-1}, \text{ where } 0/0 := 0, \\
\lambda_{7,F} &:= W_{2F}, \\
\lambda_{8,F} &:= U_{2F}, \\
\lambda_{9,F} &:= F, \text{ and} \\
\lambda &= \lambda_F := (\lambda_{1,F}, ..., \lambda_{9,F}). \qquad (10.9)
\end{aligned}$$

For simplicity, when writing $\lambda = (\lambda_{1,F}, ..., \lambda_{9,F})$, we allow the elements to be scalars, vectors, matrices, and distributions and likewise in similar expressions. If $p = 1$, no vector $\lambda_{6,F}$ appears in λ because $\lambda_{1,F}$ only contains a single element. The vector $\lambda_{6,F}$ is only used in the proofs for CLR tests. It could be deleted when considering only an LM test. The dimensions of W_{2F} and U_{2F} depend on the choices of $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$. We let $\lambda_{5,gF}$ denote the upper left $k \times k$ submatrix of $\lambda_{5,F}$. Thus, $\lambda_{5,gF} = E_F g_i g_i' = \Omega_F$.

We consider the parameter space Λ_0 for λ, which corresponds to $\mathcal{F}_{WU} \cap \mathcal{F}_0$, where $\mathcal{F}_{WU}$ and $\mathcal{F}_0$ are defined in (10.5) and (3.9), respectively. The parameter space Λ_0 and the function $h_n(\lambda)$ are defined by

$$\begin{aligned}
\Lambda_0 &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU} \cap \mathcal{F}_0\} \text{ and} \\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}). \qquad (10.10)
\end{aligned}$$

By the definition of $\mathcal{F}$, Λ_0 indexes distributions that satisfy the null hypothesis $H_0 : \theta = \theta_0$. The dimension J of $h_n(\lambda)$ equals the number of elements in $(\lambda_{1,F}, ..., \lambda_{8,F})$. Redundant elements in $(\lambda_{1,F}, ..., \lambda_{8,F})$, such as the redundant off-diagonal elements of the symmetric matrix $\lambda_{5,F}$, are not needed, but do not cause any problem. Note that two parameter spaces denoted by Λ_1 and Λ_2, which are larger than Λ_0, are considered for the two SR-CQLR tests analyzed in AG2. (We also use Λ_2 in this paper, see (10.11) below.)

We define λ and $h_n(\lambda)$ as in (10.9) and (10.10) because, as shown below, the asymptotic distributions of the test statistics under a sequence $\{F_n : n \geq 1\}$ for which $h_n(\lambda_{F_n}) \rightarrow h \in H$ depend on the behavior of $\lim n^{1/2}\lambda_{1,F_n}$, as well as $\lim \lambda_{m,F_n}$ for $m = 2, ..., 8$. For example, the LM statistic in (4.2) depends on $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$, or equivalently, on $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n B_{F_n} S_n$ (because projections are invariant to rescaling and right-hand side (rhs) transformations by nonsingular matrices), where S_n is a pd diagonal matrix that is designed to make this quantity $O_p(1)$ and not $o_p(1)$. We show that this quantity is asymptotically equivalent to $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{D}_n B_{F_n} S_n$. In turn, the latter quantity depends on $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{G}_n B_{F_n} = n^{1/2}\Omega_{F_n}^{-1/2}(\widehat{G}_n B_{F_n} - E_{F_n} G_i B_{F_n}) + n^{1/2}\Omega_{F_n}^{-1/2} E_{F_n} G_i B_{F_n}$. The quantity $vec(n^{1/2}\Omega_{F_n}^{-1/2}(\widehat{G}_n B_{F_n} - E_{F_n} G_i B_{F_n}))$ has a nondegenerate asymptotic normal distribution by the central limit theorem (CLT), using the behavior of $\lim \lambda_{s,F_n}$ for $s = 2, 4, 5$, the fact that B_{F_n} is an orthogonal matrix, and the restriction in $\mathcal{F}_0$. Hence, the asymptotic behavior of $vec(n^{1/2}\Omega_{F_n}^{-1/2}\widehat{G}_n B_{F_n})$ depends on that of $n^{1/2}\Omega_{F_n}^{-1/2} E_{F_n} G_i B_{F_n}$. Using the SVD of $\Omega_{F_n}^{-1/2} E_{F_n} G_i$, the latter is shown below to equal $\lambda_{3,F_n} Diag\{n^{1/2}\lambda_{1,F_n}\}$, where $Diag\{n^{1/2}\lambda_{1,F_n}\}$ denotes the $k \times p$ matrix with $n^{1/2}\lambda_{1,F_n}$ on the main diagonal and zeros elsewhere.

In Example 1 of Section 10.2 applied to the linear model (2.2), we have $W_F = \Omega_F^{-1/2}$ and τ_{jF} is the jth singular value of $-\Omega_F^{-1/2} E_F Z_i Y_{2i}' = -\Omega_F^{-1/2} E_F Z_i Z_i' \pi$, where $\Omega_F = E_F u_i^2 Z_i Z_i'$ for $j = 1, ..., p$. As is well known, if π is close to zero, weak instrument problems occur. But, as we show, matrices π that are close to being singular, without their columns being close to zero, also lead to weak IV problems. This is captured in the present set-up by τ_{pF} being close to zero in the sense that $\lim n^{1/2}\tau_{pF_n} < \infty$. If this occurs, then weak identification problems arise.

For notational convenience,

$$\begin{aligned} &\{\lambda_{n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_n \in \Lambda_2 : n \geq 1\} \text{ for which } h_n(\lambda_n) \rightarrow h \in H, \text{ where} \\ &\qquad \Lambda_2 := \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU}\} \end{aligned} \tag{10.11}$$

and H is defined in (10.1) with Λ replaced by Λ_2. Analogously, for any subsequence $\{w_n : n \geq 1\}$, $\{\lambda_{w_n,h} : n \geq 1\}$ denotes a sequence $\{\lambda_{w_n} \in \Lambda_2 : n \geq 1\}$ for which $h_{w_n}(\lambda_{w_n}) \rightarrow h \in H$. By definition, $\Lambda_0 \subset \Lambda_2$. We use the parameter space Λ_2 in many places in the paper, rather than Λ_0, for two reasons. First, this makes it clear where the conditions specified in $\mathcal{F}_0$ (and Λ_0) are really needed. Second, some of the results given here are used in AG2, which does not employ the smaller set Λ_0, but does use Λ_2. By the definitions of Λ_2 and $\mathcal{F}_{WU}$, $\{\lambda_{n,h} : n \geq 1\}$ is a sequence of distributions that satisfies the null hypothesis $H_0 : \theta = \theta_0$.

We decompose h (defined by (10.1), (10.9), and (10.10)) analogously to the decomposition of the first eight components of λ: $h = (h_1, ..., h_8)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, ..., 8$. We further decompose the vector h_1 as $h_1 = (h_{1,1}, ..., h_{1,p})'$, where the elements of h_1 could equal ∞. We decompose h_6 as $h_6 = (h_{6,1}, ..., h_{6,p-1})'$. In addition, we let $h_{5,g}$ denote the upper left $k \times k$ submatrix of h_5. In consequence, under a sequence $\{\lambda_{n,h} : n \geq 1\}$, we have

$$\begin{aligned} n^{1/2}\tau_{jF_n} &\to h_{1,j} \geq 0 \ \forall j \leq p, \ \lambda_{m,F_n} \to h_m \ \forall m = 2, ..., 8, \\ \lambda_{5,gF_n} &= \Omega_{F_n} = E_{F_n} g_i g_i' \to h_{5,g}, \text{ and } \lambda_{6,jF_n} \to h_{6,j} \ \forall j = 1, ..., p-1. \end{aligned} \tag{10.12}$$

By the conditions in $\mathcal{F}$, defined in (3.3), $h_{5,g}$ is pd.

The smallest and largest singular values of $W_F(E_F G_i)U_F$ (i.e., τ_{pF} and τ_{1F}) can be related to those of $E_F G_i$ (i.e., s_{pF} and s_{1F}) for $F \in \mathcal{F}_{WU}$ via

$$c_1 s_{jF} \leq \tau_{jF} \leq c_2 s_{jF} \text{ for } j = 1 \text{ and } j = p \text{ for some constants } 0 < c_1 < c_2 < \infty \tag{10.13}$$

that do not depend on F. As shown below, the parameter θ is strongly or semi-strongly identified under $\{\lambda_{n,h} : n \geq 1\}$ if $\lim n^{1/2}\tau_{pF_n} = \infty$. In consequence of (10.13), this holds iff $\lim n^{1/2} s_{pF_n} = \infty$. The parameters are weakly identified in the standard sense if $\lim n^{1/2}\tau_{jF_n} < \infty \ \forall j \leq p$ or, equivalently, if $\lim n^{1/2}\tau_{1F_n} < \infty$, which holds by (10.13) iff $\lim n^{1/2} s_{1F_n} < \infty$. The parameters are weakly identified in the non-standard sense if $\lim n^{1/2}\tau_{1F_n} = \infty$ and $\lim n^{1/2}\tau_{pF_n} < \infty$, which holds by (10.13) iff $\lim n^{1/2} s_{1F_n} = \infty$ and $\lim n^{1/2} s_{pF_n} < \infty$.

The proof of (10.13) is as follows. For notational simplicity, we drop the subscript F in some of the calculations. We have

$$\begin{aligned} & \lambda_{\min}(U'EG_i'W'WEG_iU) \\ &= \min_{\lambda:||\lambda||=1} (U\lambda/||U\lambda||)'EG_i'W'WEG_i(U\lambda/||U\lambda||) \cdot ||U\lambda||^2 \\ &\leq \min_{\lambda:||\lambda||=1} \lambda'EG_i'W'WEG_i\lambda \cdot \lambda_{\max}(U'U) \\ &= \min_{\lambda:||\lambda||=1} (EG_i\lambda/||EG_i\lambda||)'W'W(EG_i\lambda/||EG_i\lambda||) \cdot ||EG_i\lambda||^2 \cdot \lambda_{\max}(U'U) \\ &\leq \lambda_{\max}(W'W)\lambda_{\min}(EG_i'EG_i)\lambda_{\max}(U'U) \\ &\leq c_2^2 \lambda_{\min}(EG_i'EG_i), \text{ where} \\ c_2 &:= \sup_{F \in \mathcal{F}_{WU}} [\lambda_{\max}(W_F'W_F)\lambda_{\max}(U_F'U_F)]^{1/2} < \infty \end{aligned} \tag{10.14}$$

and the last inequality holds by the conditions in $\mathcal{F}_{WU}$ (defined in (10.5)). Because the smallest eigenvalues of $U'EG_i'W'WEG_iU$ and $EG_i'EG_i$ equal the squares of the smallest singular values of WEG_iU and EG_i, respectively, (10.14) establishes the second inequality in (10.13) for $j = p$. Analogous calculations establish the lower bound in (10.14) for $j = p$ and the bounds for $j = 1$

by replacing min and $\leq$ by max and $\geq$, respectively, in the appropriate places and taking $c_1 := \inf_{F\in\mathcal{F}_{WU}}[\lambda_{\min}(W_F'W_F)\lambda_{\min}(U_F'U_F)]^{1/2} > 0$.

10.4 Assumption WU

We assume that the random weight matrices $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$ defined in (10.2) satisfy the following assumption that depends on a suitably chosen parameter space Λ_* $(\subset \Lambda_2)$, such as Λ_2, Λ_0, or Λ_1.

Assumption WU for the parameter space $\Lambda_* \subset \Lambda_2$: Under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_*$,

(a) $\widehat{W}_{2w_n} \to_p h_7$ $(:= \lim W_{2F_{w_n}})$,

(b) $\widehat{U}_{2w_n} \to_p h_8$ $(:= \lim U_{2F_{w_n}})$, and

(c) $W_1(\cdot)$ is a continuous function at h_7 on some set $\mathcal{W}_2$ that contains $\{\lambda_{7,F}$ $(= W_{2F}) : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \in \Lambda_*\}$ and contains $\widehat{W}_{2w_n}$ wp$\to$1 and $U_1(\cdot)$ is a continuous function at h_8 on some set $\mathcal{U}_2$ that contains $\{\lambda_{8,F}$ $(= U_{2F}) : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \in \Lambda_*\}$ and contains $\widehat{U}_{2w_n}$ wp$\to$1.

In Assumption WU and elsewhere below, "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$" means "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ for any $h \in H$" and likewise with n in place of w_n. Note that, by definition, a sequence $\{\lambda_{w_n,h} : n \geq 1\}$ determines a sequence of distributions $\{F_{w_n} : n \geq 1\}$, see (10.9).

Assumption WU for the parameter space Λ_0 is verified in Comment (ii) to Theorem 12.1 given below for the CLR test with moment-variance weighting, which is considered in Section 6. It also holds for Kleibergen's LM test (for the same parameter space Λ_0) by the same argument (because $\widehat{W}_{2n}$, $\widehat{U}_{2n}$, $W_1(\cdot)$, and $U_1(\cdot)$ are the same for these two tests, see (10.4)).

10.5 Basic Results

For any square-integrable random vector a_i and $F, F_n \in \mathcal{F}$, define

$$\Phi_F^{a_i} := Var_F(a_i - (E_F a_\ell g_\ell')\Omega_F^{-1} g_i) \text{ and } \Phi_h^{a_i} := \lim \Phi_{F_{w_n}}^{a_i} \tag{10.15}$$

whenever the limit exists, where the distributions $\{F_{w_n} : n \geq 1\}$ correspond to $\{\lambda_{w_n,h} : n \geq 1\}$ for any subsequence $\{w_n : n \geq 1\}$. Note that $\Phi_F^{a_i} = \Psi_F^{a_i} - E_F a_i E_F a_i'$ (because $\Psi_F^{a_i} = E_F b_i b_i'$ for $b_i = a_i - (E_F a_\ell g_\ell')\Omega_F^{-1} g_i$ and $E_F g_i = 0^k$).

A basic result that is used in the proofs of results for all of the tests considered in this paper and AG2 is the following.

Lemma 10.2 *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$,

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

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.

Comments: (i) The variance matrix $\Phi_h^{vec(G_i)}$ depends on h only through h_4 and h_5. The assumptions allow $\Phi_h^{vec(G_i)}$ to be singular.

(ii) Suppose one eliminates the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}$ and one defines $\widehat{D}_n$ in (3.2) with $\widehat{\Omega}_n$ replaced by an eigenvalue-adjusted matrix, denoted by $\widehat{\Omega}_n^{\varepsilon}$, which is constructed to have its smallest eigenvalue greater than or equal to $\varepsilon > 0$ multiplied by its largest eigenvalue, see AG2 for the details of such a construction. In this case, the result of Lemma 10.2 still holds and all of the other asymptotic results following from Lemma 10.2 still hold, except the independence of $\overline{g}_h$ and $\overline{D}_h$. However, this independence is key because it is used in the conditioning argument that establishes the correct asymptotic size of all of the tests that are shown to have correct asymptotic size. Without it, these tests do not necessarily have correct asymptotic size. In consequence, we define $\widehat{D}_n$ in (3.2) using $\widehat{\Omega}_n$, not $\widehat{\Omega}_n^{\varepsilon}$.

The reason that independence does not necessarily hold when $\widehat{D}_n$ is defined using $\widehat{\Omega}_n^{\varepsilon}$, rather than $\widehat{\Omega}_n$, is that the covariance term $E_{F_n}[G_{ij} - E_{F_n} G_{ij} - (E_{F_n} G_{\ell j} g_\ell')(\Omega_{F_n}^{\varepsilon})^{-1} g_i] g_i'$ typically does not equal $0^{k \times k}$ when $\Omega_{F_n}^{\varepsilon} \neq \Omega_{F_n}$, whereas $E_{F_n}[G_{ij} - E_{F_n} G_{ij} - (E_{F_n} G_{\ell j} g_\ell') \Omega_{F_n}^{-1} g_i] g_i'$ necessarily equals $0^{k \times k}$, see the proof of Lemma 10.2 in Section 15 below for more details.

(iii) The proofs of Lemma 10.2 and other results in this section are given in Sections 15-17 below.

The following is a key definition. Consider a sequence $\{\lambda_{n,h} : n \geq 1\}$. Let $q = q_h$ $(\in \{0, ..., p\})$ be such that

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq p, \tag{10.16}$$

where $h_{1,j} := \lim n^{1/2} \tau_{jF_n} \geq 0$ for $j = 1, ..., p$ by (10.12) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (10.11). Such a q exists because $\{h_{1,j} : j \leq p\}$ are nonincreasing in j (since $\{\tau_{jF} : j \leq p\}$ are the ordered singular values of $W_F(E_F G_i) U_F$, as defined in (10.8)). As defined, q is the number of singular values of $W_{F_n}(E_{F_n} G_i) U_{F_n}$ that diverge to infinity when multiplied by $n^{1/2}$. Roughly speaking, q is the number of parameters, or one-to-one transformations of the parameters, that are strongly or semi-strongly identified.

The following quantities appear in Lemma 10.3 below, which gives the asymptotic distribution of $\widehat{D}_n$ after suitable rotations and rescaling, but without the recentering (by subtracting $E_{F_n}G_i$) that appears in Lemma 10.2. We partition h_2 and h_3 and define $\overline{\Delta}_h$ as follows:

$$h_2 = (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}),\ h^{\diamond}_{1,p-q} := \begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times(p-q)},$$
$$\overline{\Delta}_h = (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q},\ \overline{\Delta}_{h,p-q} := h_3 h^{\diamond}_{1,p-q} + h_{71}\overline{D}_h h_{81} h_{2,p-q},$$
$$h_{71} := W_1(h_7),\ \text{and}\ h_{81} := U_1(h_8), \tag{10.17}$$

where $h_{2,q} \in R^{p\times q}$, $h_{2,p-q} \in R^{p\times(p-q)}$, $h_{3,q} \in R^{k\times q}$, $h_{3,k-q} \in R^{k\times(k-q)}$, $\overline{\Delta}_{h,q} \in R^{k\times q}$, $\overline{\Delta}_{h,p-q} \in R^{k\times(p-q)}$, $h_{71} \in R^{k\times k}$, $h_{81} \in R^{p\times p}$, and $\overline{D}_h$ is defined in Lemma 10.2. For simplicity, there is some abuse of notation here, e.g., $h_{2,q}$ and $h_{2,p-q}$ denote different matrices even if $p-q$ happens to equal q. Note that when Assumption WU holds $h_{71} = \lim W_{F_n} = \lim W_1(W_{2F_n})$ and $h_{81} = \lim U_{F_n} = \lim U_1(U_{2F_n})$ under $\{\lambda_{n,h} : n \geq 1\}$.

The case where $q = p$ (i.e., $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq p$) is the strong or semi-strong identification case. In this case, no $h_{2,p-q}$, $h^{\diamond}_{1,p-q}$, and $\overline{\Delta}_{h,p-q}$ matrices appear in (10.17), $\overline{\Delta}_h = h_{3,q} = h_{3,p}$, and $\overline{\Delta}_h$ is non-random. In consequence, the limit in distribution (or probability) of the normalized matrix $n^{1/2}W_{F_n}\widehat{D}_n U_{F_n} T_n$, where $T_n \in R^{p\times p}$ is defined below, is non-random, see Lemma 10.3 below. When $q < p$, identification is weak and the limit of this matrix is random.

Now we provide some motivation for Lemma 10.3, which is stated below. To show that the LM statistic has a χ^2_p asymptotic distribution we need to determine the asymptotic behavior of $\widehat{D}_n$ without the recentering by $E_{F_n}G_i$ that occurs in Lemma 10.2. In addition, to determine the asymptotic distribution of the rk_n statistic in (6.2), we need to determine the asymptotic distribution of $W_{F_n}\widehat{D}_n U_{F_n}$ without recentering by $E_{F_n}G_i$. (Furthermore, to determine the asymptotic distributions of the two SR-CQLR test statistics and conditional critical values considered in AG2, we need to determine the asymptotic distribution of $W_{F_n}\widehat{D}_n U_{F_n}$ without recentering by $E_{F_n}G_i$.) To do so, we post-multiply $W_{F_n}\widehat{D}_n U_{F_n}$ first by B_{F_n} and then by a nonrandom diagonal matrix $S_n \in R^{p\times p}$ (which may depend on F_n and h). The matrix S_n rescales the columns of $W_{F_n}\widehat{D}_n U_{F_n} B_{F_n}$ to ensure that $n^{1/2}W_{F_n}\widehat{D}_n U_{F_n} B_{F_n} S_n$ converges in distribution to a (possibly) random matrix that is finite a.s. and not almost surely zero. For $F \in \mathcal{F}_{WU} \cap \mathcal{F}_0$, it ensures that the (possibly) random limit matrix has full column rank with probability one. For example, in the case of the LM statistic, these transformations are applied with $W_{F_n} = \Omega^{-1/2}_{F_n}$ and $U_{F_n} = I_p$.

For the LM statistic and the CLR statistics that employ it, we need the full column rank property of the limit random matrix in order to apply the continuous mapping theorem (CMT).

For the LM statistic, the full rank property ensures that the quantity $\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n$ (whose inverse appears in the expression for LM_n, see (4.2)), is nonsingular asymptotically with probability one after $\widehat{D}_n$ has been transformed and rescaled to yield $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{D}_nB_{F_n}S_n$. Note that $P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}$, which appears in the definition of LM_n in (4.2), can be written as

$$\begin{aligned}
P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} &:= \widehat{\Omega}_n^{-1/2}\widehat{D}_n(\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n)^{-1}\widehat{D}_n'\widehat{\Omega}_n^{-1/2} \\
&= (\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})(n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n)\left[(n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n)'(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})'(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})\right. \\
&\quad \left.\times\, (n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n)\right]^{-1}(n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n)'(\Omega_n^{1/2}\widehat{\Omega}_n^{-1/2}), \text{ where} \\
T_n &:= B_{F_n}S_n \in R^{p\times p} \text{ and } \Omega_n := \Omega_{F_n}\ (= E_{F_n}g_ig_i'),
\end{aligned} \tag{10.18}$$

provided T_n has full rank and Ω_n is pd. In consequence, these transformations do not affect the value or distribution of the LM statistic.

Note that the two SR-CQLR test statistics considered in AG2 do not depend on an LM statistic and do not require the asymptotic distribution of $n^{1/2}W_{F_n}\widehat{D}_nU_{F_n}B_{F_n}S_n$ to have full column rank a.s.

Define

$$S_n := Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}, 1, ..., 1\} \in R^{p\times p}, \tag{10.19}$$

where $q = q_h$ is defined in (10.16). Note that $\tau_{jF_n} > 0$ for n large for $j \leq q$ and, hence, S_n is well defined for n large, because $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$.

The proof of Theorem 11.1 for the LM test, the proofs of Theorems 10.4 and 12.1 for the CLR test with moment-variance weighting, and the proofs for the two SR-CQLR tests in AG2 use the following lemma. The $p \times p$ matrix T_n is defined in (10.18).

Lemma 10.3 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

where (a) $(\overline{g}_h, \overline{D}_h)$ *are defined in Lemma* 10.2, (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (10.17), (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.

Comments: (i) Lemma 10.3(c)-(d) are key properties of the asymptotic distribution of $n^{1/2}(\widehat{g}_n,$

$W_{F_n}\widehat{D}_n U_{F_n} T_n$) that lead to the LM statistic having a χ^2_p asymptotic distribution and the CLR test with moment-variance weighting having correct asymptotic size. Lemma 10.3(c) is a key property that leads to the correct asymptotic size of the two SR-CQLR tests in AG2. Lemma 10.3(d) is not needed for these tests because they do not rely on an LM statistic.

(ii) The conditions in $\mathcal{F}_0$ are used in the proofs to obtain the result of Lemma 10.3(d) and are not used elsewhere in the proofs, except where Lemma 10.3(d) is used.

The following theorems are used only for the CLR tests. For the proof of Theorem 4.1 concerning Kleibergen's (2005) LM test, one can go from here to Section 11.

Let

$$\widehat{\kappa}_{jn} \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n,\ \forall j = 1, ..., p, \tag{10.20}$$

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

Theorem 10.4 *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) $\widehat{\kappa}_{pn} \rightarrow_p \infty$ *if* $q = p$,

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

(c) $\widehat{\kappa}_{jn} \rightarrow_p \infty$ *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* 10.3, *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.

Comments: (i) The statistic $\widehat{\kappa}_{pn} = \lambda_{\min}(n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n)$ in Theorem 10.4(a) and (b) is a Robin and Smith (2000)-type rank statistic.

(ii) Theorem 10.4(a) and (b) is used to determine the asymptotic behavior of the statistic rk_n defined in (6.2) (which is employed by the CLR test with moment-variance weighting that is considered in Section 6). More specifically, Theorem 10.4(a) and (b) is used to verify Assumption R in Section 12 below.

(iii) Theorem 10.4(c) and (d) is used to determine the asymptotic behavior of the critical value functions for the two SR-CQLR tests considered in AG2 (with $\widehat{W}_n$ and $\widehat{U}_n$ defined suitably). Because Theorem 10.4(c) and (d) are immediate by-products of the proofs of Theorem 10.4(a) and (b), they are stated and proved here, rather than in AG2.

(iv) The statement of Theorem 3 in Kleibergen (2005) is difficult to interpret because the expression given for the conditional asymptotic distribution of the CLR statistic involves Kleibergen's (2005) statistic rk(θ_0), which is a finite-sample object. Based on Theorem 10.4, (12.7) below provides the asymptotic distribution of a class of CLR statistics in terms of an asymptotic version of the rank statistic employed, which is necessary for a precise statement of the asymptotic distribution. The class of CLR statistics considered are those defined in (5.1) and based on the rank statistic in Theorem 10.4 for some choices of $\widehat{W}_n$ and $\widehat{U}_n$, which is a Robin and Smith (2000)-type rank statistic. In particular, taking $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$ gives the rank statistic defined in (6.2).

11 Asymptotic Size of the Nonlinear LM Test

In this section, we prove Theorem 4.1 for the LM test.

We state a theorem that verifies Assumption B* of ACG (stated in Section 10) for the LM test. The following theorem applies with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $W_F = \Omega_F^{-1/2}$, and $\widehat{U}_n = U_F = I_p$. (These definitions affect the definition of $\lambda_{n,h}$, which appears in the theorem).

Theorem 11.1 *The asymptotic null rejection probabilities of the nominal size* $\alpha \in (0,1)$ LM *test equal* α *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$.

Comments: (i) The requirement that $\lambda_{w_n,h} \in \Lambda_0$ (defined in (10.10)) implies that the parameter space for F is $\mathcal{F}_0$ (defined in (3.9)) for the results given in Theorems 4.1 and 11.1 (because the restrictions in $\mathcal{F}_{WU}$ are not binding, see the discussion in the paragraph containing (10.5)).

(ii) Proposition 10.1 and Theorem 11.1 prove Theorem 4.1 for the LM test. The proof of Theorem 4.1 for the LM CS is analogous, see Comments (i) and (ii) to Proposition 10.1.

For notational simplicity, we prove Theorem 11.1 for the sequence $\{n\}$, rather than a subsequence $\{w_n : n \geq 1\}$. We note here that the same proof holds for any subsequence $\{w_n : n \geq 1\}$.

Proof of Theorem 11.1. Let $\Omega_n := \Omega_{F_n}$. We derive the limiting distribution of the statistic LM_n using the CMT applied to $\Omega_n^{-1/2} n^{1/2} \widehat{g}_n$, $\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2}$, and $n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n$, where the latter two quantities appear in the expression on the rhs of (10.18). Note that $\widehat{\Omega}_n \to_p h_{5,g}$ by the WLLN, $\Omega_n \to h_{5,g}$, and $h_{5,g}$ is pd. Thus, $\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2} \to_p I_k$. By Lemma 10.3 applied with $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$ (which results from taking $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$), we get $(\Omega_n^{-1/2} n^{1/2} \widehat{g}_n, n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n) \to_d (h_{5,g}^{-1/2}\overline{g}_h, \overline{\Delta}_h)$. For the CMT to apply, it is enough to show that the function $f : R^{k\times p} \to R^{k\times k}$ defined by $f(D) := D(D'D)^{-1}D'$ for $D \in R^{k\times p}$ is continuous on a set $C \subset R^{k\times p}$ with $P(\overline{\Delta}_h \in C) = 1$. This holds because the function $f_2(D, L) := LD((LD)'(LD))^{-1}D'L'$ for a nonsingular

$k \times k$ matrix L is continuous at (D, I_k) if $f(D)$ is continuous at D. Note that f is continuous at each D that has full column rank. And, by Lemma 10.3(d), $\overline{\Delta}_h$ has full column rank a.s. because $\lambda_{n,h} \in \Lambda_0$, $F_n \in \mathcal{F}_0$, $W_F = \Omega_F^{-1/2}$, and $U_F = I_p$. Hence, f is continuous a.s. By $\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2} \rightarrow_p I_k$, the convergence result in Lemma 10.3, and the CMT, we have

$$P_{D_n^\diamond}\widehat{\Omega}_n^{-1/2}n^{1/2}\widehat{g}_n = D_n^\diamond(D_n^{\diamond\prime}D_n^\diamond)^{-1}D_n^{\diamond\prime}\widehat{\Omega}_n^{-1/2}n^{1/2}\widehat{g}_n \rightarrow_d \overline{v}_h := P_{\overline{\Delta}_h}h_{5,g}^{-1/2}\overline{g}_h, \tag{11.1}$$

where $D_n^\diamond := (\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n$.

Conditional on $\overline{\Delta}_h$, $\overline{v}_h^\prime\overline{v}_h$ is distributed as χ_p^2 because (i) $\overline{\Delta}_h$ and $\overline{g}_h$ are independent by property (c) in Lemma 10.3, (ii) $h_{5,g}^{-1/2}\overline{g}_h$ is conditionally distributed as $N(0^k, I_k)$ by $\overline{g}_h \sim N(0^k, h_{5,g})$ and (i), and (iii) $P_{\overline{\Delta}_h}$ is fixed given $\overline{\Delta}_h$ and projects onto a space of dimension p a.s. by property (d) in Lemma 10.3. Because the χ_p^2 distribution does not depend on $\overline{\Delta}_h$, $\overline{v}_h^\prime\overline{v}_h$ is unconditionally distributed as χ_p^2 as well. In consequence, using the CMT again, we have

$$LM_n \rightarrow_d \overline{LM}_h := \overline{v}_h^\prime\overline{v}_h \sim \chi_p^2. \tag{11.2}$$

Given this result and the use of the $\chi_{p,1-\alpha}^2$ critical value by the LM test, we obtain the conclusion of Theorem 11.1 for the LM test: $\lim P_{F_n}(LM_n > \chi_{p,1-\alpha}^2) = \alpha$. $\square$

12 Asymptotic Size of the CLR Test with Moment-Variance Weighting

In this section, we prove Theorem 6.1, which concerns the CLR test (and CS) with moment-variance weighting based on the Robin-Smith rank statistic. In fact, for the CLR test defined by (5.1)-(5.2), we prove a stronger result than that given in Theorem 6.1. We establish Theorem 6.1 for a CLR test that is based on any rank statistic rk_n that satisfies a high-level assumption, denoted Assumption R, not just the rank statistic $rk_n(\theta_0)$ defined in (6.2). Then, we verify Assumption R for the moment-variance-weighted Robin-Smith rank statistic $rk_n(\theta_0)$ in (6.2). Note that Assumption R does not hold for the rank statistic in (5.5) when $p \geq 2$.

Section 19.5 below provides additional asymptotic size results for equally-weighted CLR tests (and CS's), which are CLR tests that are based on rk_n statistics that depend on $\widehat{D}_n$ only through $\widetilde{W}_n\widehat{D}_n$ for some $k \times k$ weighting matrix $\widetilde{W}_n$. These results show that equally-weighted CLR tests (and CS's) based on the Robin and Smith (2000) rank statistic with a general weight matrix $\widetilde{W}_n$ ($\in R^{k\times k}$) have correct asymptotic size under suitable conditions on $\widetilde{W}_n$. One can view these results as verifying Assumption R for a broad class of rk_n statistics. In contrast, the results in the present section establish the correct asymptotic size of CLR tests (and CS's) under the high-level condition Assumption R and for the Robin and Smith (2000) rank statistic when $\widetilde{W}_n$ is the moment-variance weighting matrix $\widehat{\Omega}_n^{-1/2}$, see Comment (ii) to Theorem 12.1 below.

The high-level condition on the rank statistic rk_n is the following.

Assumption R: For any subsequence $\{w_n\}$ and any sequence $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$ either (a) $rk_{w_n} \to_p r_h = \infty$ or (b) $rk_{w_n} \to_d r_h(\overline{D}_h)$ for some nonrandom function $r_h : R^{k \times p} \to R$, where $\overline{D}_h$ is defined in Lemma 10.2, and the convergence is joint with that in Lemma 10.2.

In Assumption R, by $rk_{w_n} \to_p \infty$, we mean that for every $K < \infty$ we have $P_{\theta_0,\lambda_{w_n}}(rk_{w_n} > K) \to 1$, where $P_{\theta_0,\lambda_{w_n}}(\cdot)$ denotes probability under λ_{w_n} when the true parameter vector equals θ_0.

The following theorem applies when the LM statistic is defined as in (4.2) with projection onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$. In consequence, the quantities in (10.2) in the present case are $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $W_F = \Omega_F^{-1/2}$, and $\widehat{U}_n = U_F = I_p$. (These definitions affect the definition of $\lambda_{n,h}$, which appears in the theorem).

Theorem 12.1 *For any statistic rk_n that satisfies Assumption* R, *the asymptotic null rejection probabilities of the nominal size $\alpha \in (0,1)$* CLR *test defined in* (4.2)-(5.2) *based on rk_n equal α under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$.*

Comments: (i) Theorem 12.1 and Proposition 10.1 imply that a nominal size α CLR test based on any rank statistic that satisfies Assumption R has asymptotic size α and is asymptotically similar. Analogous CS results (to the test results stated in Theorem 12.1) hold for a parameter space $\Lambda_{\Theta,0}$ that is a reparametrization of $\mathcal{F}_{\Theta,0}$ and is defined as Λ_0 is defined, but with the adjustments outlined in Comments (i) and (ii) to Proposition 10.1.

(ii) Theorems 10.4 and 12.1 and Proposition 10.1 establish the test results of Theorem 6.1. This holds because Theorem 10.4(a), (b), (e), and (f) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$ imply that Assumption R holds for the CLR test with moment-variance weighting, that is considered in Section 6, which uses the Robin and Smith (2000) rk_n statistic defined in (6.2). (In the present context, Theorem 10.4 requires that Assumption WU holds for the parameter space Λ_0. It holds with $\widehat{W}_n = \widehat{W}_{2n}$, $W_1(w) = w$ for $w \in R^{k \times k}$, $\mathcal{W}_2 = R^{k \times k}$, $\widehat{U}_n = \widehat{U}_{2n}$, $U_1(u) = u$ for $u \in R^{p \times p}$, and $\mathcal{U}_2 = R^{p \times p}$, because $\widehat{W}_n = \widehat{\Omega}_n^{-1/2} \to_p h_{5,g}^{-1/2}$ under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_0$ and $\widehat{U}_n = I_p$ for all $n \geq 1$.) In particular, Assumption R holds with $r_h = \infty$ if $q = p$ and with $r_h(\overline{D}_h)$ equal to the smallest eigenvalue of $\overline{\Delta}'_{h,p-q} h_{3,k-q} h'_{3,k-q} \overline{\Delta}_{h,p-q}$ if $q < p$ (where $\overline{\Delta}_{h,p-q}$ and $h_{3,k-q}$ are defined in (10.17) based on $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$). The CS results of Theorem 6.1 hold by Theorem 10.4, Comment (i) to Theorem 12.1, and Comment (i) to Proposition 10.1.

(iii) Theorem 5.1 shows that Assumption R does not hold in general for rank statistics based on $\widetilde{V}_{Dn}$ and $\widehat{D}_n^\dagger$, defined in (5.3)-(5.4), when $p \geq 2$. The reason is that for some sequences of distributions the asymptotic distribution of $\widehat{D}_n^\dagger$ and, hence, the rank statistic rk_n depends on $\overline{D}_h$ and $\overline{M}_h^\dagger \neq 0^{k \times p}$, not just on $\overline{D}_h$ alone.

For notational simplicity, the following proof is for the sequence $\{n\}$, rather than a subsequence $\{w_n : n \geq 1\}$. The same proof holds for any subsequence $\{w_n : n \geq 1\}$.

Proof of Theorem 12.1. Let

$$J_n := n\widehat{g}_n' \widehat{\Omega}_n^{-1/2} M_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} \widehat{\Omega}_n^{-1/2} \widehat{g}_n. \tag{12.1}$$

It follows from (4.2) that

$$AR_n = LM_n + J_n. \tag{12.2}$$

We now distinguish two cases. First, suppose Assumption R(a) holds: $rk_n \rightarrow_p \infty$. By (12.2) and some algebra, we have $(AR_n - rk_n)^2 + 4LM_n \cdot rk_n = (LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n$. Therefore,

$$CLR_n = \frac{1}{2}\left(LM_n + J_n - rk_n + \sqrt{(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n}\right). \tag{12.3}$$

Using a mean-value expansion of the square-root expression in (12.3) about $(LM_n - J_n + rk_n)^2$, we have

$$\sqrt{(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n} = LM_n - J_n + rk_n + (2\sqrt{\zeta_n})^{-1} 4LM_n \cdot J_n \tag{12.4}$$

for an intermediate value ζ_n between $(LM_n - J_n + rk_n)^2$ and $(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n$. It follows that $CLR_n = LM_n + o_p(1) \rightarrow_d \chi_p^2$ using (11.2) and $(\sqrt{\zeta_n})^{-1} = o_p(1)$ (which holds because $rk_n \rightarrow_p \infty$, $LM_n = O_p(1)$, and $J_n = O_p(1)$ by (12.6) below). Analogously, it can be shown that the critical value $c(1-\alpha, rk_n)$, defined above (5.2), of the CLR test converges in probability to $\chi^2_{p,1-\alpha}$. The result of Theorem 12.1 then follows by the definition of convergence in distribution.

Second, suppose Assumption R(b) holds. Then, using Lemma 10.2, we have $(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i), rk_n) \rightarrow_d (\overline{g}_h, \overline{D}_h, r_h(\overline{D}_h))$. By the proof of Lemma 10.3 applied with $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$ (which correspond to $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$), using the former result in place of $(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i)) \rightarrow_d (\overline{g}_h, \overline{D}_h)$ gives

$$(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i), n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n, rk_n) \rightarrow_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h, r_h(\overline{D}_h)), \tag{12.5}$$

where $\Omega_n := \Omega_{F_n}$, $(\overline{D}_h, \overline{\Delta}_h)$ and $\overline{g}_h$ are independent, and $\overline{\Delta}_h$ has full column rank p with probability one by Lemma 10.3(d) (because we are considering sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$, $W_F = \Omega_F^{-1/2}$, and $U_F = I_p$). In addition, $\widehat{\Omega}_n \to_p h_{5,g}$, $h_{5,g}$ is pd, and $M_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} = M_{\left(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2}\right)n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n}$ because T_n (defined in (10.18)) and $\Omega_n^{-1/2}$ are nonsingular. These results and the CMT imply that

$$J_n \to_d \overline{J}_h := \overline{g}_h' h_{5,g}^{-1/2} M_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h. \tag{12.6}$$

The convergence results in (11.2) and (12.6) and $rk_n \to_d r_h(\overline{D}_h)$ hold jointly by (12.5) and the definitions of LM_n and J_n in (4.2) and (12.1).

Note that $\overline{LM}_h = \overline{g}_h' h_{5,g}^{-1/2} P_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ by (11.1) and (11.2). Conditional on $\overline{\Delta}_h$, $P_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ and $M_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ have a joint normal distribution with zero covariance (because $Var(h_{5,g}^{-1/2}\overline{g}_h) = I_k$ and $P_{\overline{\Delta}_h} M_{\overline{\Delta}_h} = 0^{k \times k}$) and, hence, are independent. The same holds true conditional on $\overline{D}_h$, because $\overline{\Delta}_h$ is a nonrandom function of $\overline{D}_h$ and $\overline{D}_h$ is independent of $\overline{g}_h$. In consequence, conditional on $\overline{D}_h$, $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2, respectively.

Using the convergence results in (12.5) and (12.6), the definition of CLR_n in (5.1) with $AR_n = LM_n + J_n$ substituted in, and the CMT, we obtain

$$CLR_n \to_d \overline{CLR}_h := \frac{1}{2}\left(\overline{LM}_h + \overline{J}_h - \overline{r}_h + \sqrt{(\overline{LM}_h + \overline{J}_h - \overline{r}_h)^2 + 4\overline{LM}\overline{r}_h}\right), \tag{12.7}$$

where $\overline{r}_h := r_h(\overline{D}_h)$.

The function $c(1-\alpha, r)$ (defined in (5.2)) is continuous in r on R_+ by the absolute continuity of the distributions of χ_p^2 and χ_{k-p}^2, which appear in $clr(r)$ (also defined in (5.2)), and the continuity of $clr(r)$ in r a.s. This, $rk_n \to_d \overline{r}_h$, and (12.7) yield

$$CLR_n - c(1-\alpha, rk_n) \to_d \overline{CLR}_h - c(1-\alpha, \overline{r}_h). \tag{12.8}$$

Therefore, by the definition of convergence in distribution, we have

$$P_{\theta_0,\lambda_n}(CLR_n > c(1-\alpha, rk_n)) \to P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)) \tag{12.9}$$

provided $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$, which holds because $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)|\overline{D}_h) = 0$ a.s. The latter holds because conditional on $\overline{D}_h$, $\overline{CLR}_h$ is absolutely continuous (by (12.7) since $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2 and $\overline{r}_h$ is a nonrandom function of $\overline{D}_h$) and $c(1-\alpha, \overline{r}_h)$ is a constant.

From above, conditional on $\overline{D}_h$, $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2, respectively, and $\overline{r}_h$ is a constant. Thus, conditional on $\overline{D}_h$, $\overline{CLR}_h$ and $clr(\overline{r}_h)$ have the same distribution. By definition, $c(1-\alpha, \overline{r}_h)$ is the $1-\alpha$ quantile of the absolutely continuous random variable $clr(\overline{r}_h)$ for any constant $\overline{r}_h$. Hence,

$$P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) | \overline{D}_h) = \alpha \text{ a.s.} \tag{12.10}$$

Because the left-hand side conditional probability equals α a.s. and α does not depend on $\overline{D}_h$, the unconditional probability $P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))$ equals α as well. Combined with (12.9), this gives the desired result. □

13 Asymptotic Size of the CLR Test with Jacobian-Variance Weighting when p = 1

In this section, we prove the test results of Theorem 5.3, which concerns Kleibergen's CLR test (and CS) with Jacobian-variance weighting when $p = 1$. The CS results of Theorem 5.3 hold by an analogous argument, see Comments (i) and (ii) to Proposition 10.1.

Proof of Theorem 5.3. We prove the test results of Theorem 5.3 using Proposition 10.1 and results (or variants of results) in Lemma 10.3 and Theorems 10.4, 11.1, and 12.1. The proof is made more complicated by the fact that we need to use two different definitions of $\widehat{W}_n$. To obtain the asymptotic distribution of the LM statistic (which is a component of the CLR statistic), we need to take $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = 1$, because the LM statistic (defined in (4.2)) depends on $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$. But, to obtain the asymptotic distribution of the rank statistic $rk_n := n\widehat{D}_n{}'\widetilde{V}_{Dn}^{-1}\widehat{D}_n$ (defined in (5.10)), we need to take $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$ and $\widehat{U}_n = 1$, because rk_n depends on $\widetilde{V}_{Dn}^{-1/2}\widehat{D}_n$.

For notational simplicity, we establish results below for sequences $\{n\}$, rather than subsequences $\{w_n\}$ of $\{n\}$. Subsequence results hold by replacing n by w_n in the proofs.

We proceed as follows. First, we apply Lemma 10.3 exactly as in the proof of Theorem 11.1 with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $\widehat{U}_n = 1$, $W_F = \Omega_F^{-1/2}$, and $U_F = 1$. This yields $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i, W_{F_n}\widehat{D}_nU_{F_n}T_n) \to_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h)$ for sequences $\{\lambda_{n,h} : n \geq 1\}$ that correspond to distributions F in $\mathcal{F}_{WU} \cap \mathcal{F}_0$ based on these definitions of W_F and U_F. As discussed in the paragraph containing (10.5), $\mathcal{F}_0 = \mathcal{F}_{WU} \cap \mathcal{F}_0$ for δ_{WU} sufficiently small and M_{WU} sufficiently large. We employ constants δ_{WU} and M_{WU} for which this holds. The joint convergence result above yields the asymptotic distributions of the AR_n, LM_n, and J_n statistics via the calculations in (11.1), (11.2), (12.1), (12.2), and (12.6).

Next, we take $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$, $\widehat{U}_n = 1$, $W_F = W_{2F} = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$, where $\Gamma_F^{G_i}$ and Ω_F are defined in (3.6), $W_1(\cdot)$ equals the identity function on $\mathcal{W}_2 := R^{k\times k}$, $U_F = U_{2F} = 1$, and $U_1(\cdot)$ equals the identity function on $\mathcal{U}_2 := R$. We consider distributions in $\mathcal{F}_{JVW,p=1}$ (which is a subset of $\mathcal{F}_0$ when $\delta_3 = \delta_2$ by the paragraph following (5.9)). We obtain the asymptotic distribution of rk_n

under the corresponding sequences $\{\lambda_{n,h} : n \geq 1\}$ (which differ from the sequences $\{\lambda_{n,h} : n \geq 1\}$ in the previous paragraph due to the difference between the two definitions of W_F). More specifically, we verify the convergence results in Assumption R for $rk_n := n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n$ (defined in (5.10)) for the $\{\lambda_{n,h} : n \geq 1\}$ sequences of this paragraph. The result of Theorem 10.4(a), (b), (e), and (f) verifies the convergence results in Assumption R for sequences $\{\lambda_{n,h} : n \geq 1\}$ for which $F_n \in \mathcal{F}_{JVW,p=1}$ $\forall n \geq 1$ provided Assumption WU holds for such sequences with $\widehat{W}_{2n} = \widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$, $W_1(\cdot)$ equal to the identity function, $\widehat{U}_{2n} = \widehat{U}_n = 1$, $U_1(\cdot)$ equal to the identity function, and the parameter space Λ_* being equal to $\Lambda_{JVW,p=1} := \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F})$ for some $F \in \mathcal{F}_{WU} \cap \mathcal{F}_{JVW,p=1}\}$. Here $\mathcal{F}_{WU}$ is defined in (10.5) with $W_F = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$ and $U_F = 1$. Note that $\mathcal{F}_{JVW,p=1} = \mathcal{F}_{WU} \cap \mathcal{F}_{JVW,p=1}$ for $\delta_{WU} > 0$ sufficiently small and $M_{WU} < \infty$ sufficiently large (and we employ constants δ_{WU} and M_{WU} that satisfy these conditions). This holds because for all $F \in \mathcal{F}_{JVW,p=1}$, $\lambda_{\min}(W_F) = \lambda_{\min}((Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}) = \lambda_{\max}^{-1/2}(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})$ $\geq \lambda_{\max}^{-1/2}(E_F G_i G_i') \geq M_+^{-1/2}$ for some $M_+ < \infty$ (because $E_F G_i G_i' - (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}) = E_F G_i E_F G_i' + \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}$ is psd and $||E_F G_i G_i'|| \leq M_+$ for some $M_+ < \infty$ by the moment conditions in $\mathcal{F}$), $||W_F|| = ||(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}|| \leq \lambda_{\min}^{-1/2}(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}) \leq \delta_3^{-1/2}$ (using the condition in $\mathcal{F}_{JVW,p=1}$ and the fact that $Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime} = \Psi_F^{G_i} - E_F G_i E_F G_i'$ using the definition of $\Psi_F^{G_i}$ in (3.6)), where $\delta_3 > 0$, and $||U_F|| = \lambda_{\min}(U_F) = 1$.

Assumption WU(b) holds automatically with $h_8 = 1$ because $\widehat{U}_{2n} := 1$. The requirement of Assumption WU(c) that $W_1(\cdot)$ is continuous at h_7 and $U_1(\cdot)$ is continuous at h_8 also holds automatically because $W_1(\cdot)$ and $U_1(\cdot)$ are identity functions.

Assumption WU(a) for the parameter space $\Lambda_{JVW,p=1}$ requires that $\widehat{W}_{2n} \rightarrow_p h_7$ $(:= \lim W_{2F_n})$. For sequences $\{\lambda_{n,h} : n \geq 1\}$, we have

$$\begin{aligned} \widetilde{V}_{Dn} &:= n^{-1}\sum_{i=1}^{n}(G_i - \widehat{G}_n)(G_i - \widehat{G}_n)' - \widehat{\Gamma}_n\widehat{\Omega}_n^{-1}\widehat{\Gamma}_n' \\ &= E_{F_n}(G_i - E_{F_n}G_i)(G_i - E_{F_n}G_i)' - \Gamma_{F_n}^{G_i}\Omega_{F_n}^{-1}\Gamma_{F_n}^{G_i\prime} + o_p(1) \\ &= W_{2F_n}^{-2} + o_p(1) \\ &\rightarrow_p h_7^{-2}, \end{aligned} \tag{13.1}$$

where the first equality holds by (5.3), the second equality holds by the WLLN's applied multiple times and Slutsky's Theorem using the conditions in $\mathcal{F}$, the third equality holds by the definition of W_{2F}, and the convergence holds because $W_{2F_n} = \lambda_{7,F_n} \rightarrow h_7$ by the definition of the sequence $\{\lambda_{n,h} : n \geq 1\}$ and h_7 is pd (since $h_7 = \lim W_{2F_n}$ and the eigenvalues of W_{2F}^{-2} are bounded above for $F \in \mathcal{F}$). Equation (13.1) and Slutsky's Theorem give $\widetilde{V}_{Dn}^{-1/2} \rightarrow_p h_7$ because h_7^{-2} is pd using the condition in $\mathcal{F}_{JVW,p=1}$ that $\lambda_{\min}(\Psi_F^{G_i} - E_F G_i E_F G_i') \geq \delta$. In consequence, Assumption WU(a) holds.

This completes the verification of Assumption WU for the parameter space $\Lambda_{JVW,p=1}$ and, in consequence, the verification of the convergence results of Assumption R for rk_n for sequences $\{\lambda_{n,h} : n \geq 1\}$ defined in the fourth paragraph of this proof.

Now we consider sequences $\{\lambda_{n,h} : n \geq 1\}$ that satisfy the conditions on $\{\lambda_{n,h} : n \geq 1\}$ given in both the third and fourth paragraphs of this proof. These sequences correspond to distributions F in $\mathcal{F}_{JVW,p=1}$. These sequences satisfy the convergence conditions in (8.11) using the definitions in (8.9) and (8.10) with τ_{jF}, B_F, C_F, and W_{2F} defined based on $W_F = \Omega_F^{-1/2}$ and with these quantities based on $W_F = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$. In consequence, for these sequences of distributions $\{\lambda_{n,h} : n \geq 1\}$, the results above establish the asymptotic distributions of the AR_n, LM_n, J_n, and rk_n statistics and the convergence is joint because all of the convergence results are based on the underlying CLT result in Lemma 10.2. Given this joint convergence, by the same arguments as given in the proof of Theorem 12.1, we obtain that the CLR test with Jacobian-variance weighting has asymptotic null rejection probabilities equal to α under all such sequences $\{\lambda_{n,h} : n \geq 1\}$ (and all subsequences of such sequences).

Finally, we apply Proposition 8.1 with λ and $h_n(\theta)$ given by the concatenation of the λ vectors and $h_n(\lambda)$ functions used in the third and fourth paragraphs above and with Λ given by the product space of the Λ spaces used in these paragraphs. (Redundant elements of λ and $h_n(\lambda)$ do not cause any problems.) The result of the previous paragraph verifies Assumption B* for this choice λ, $h_n(\lambda)$, and Λ. In consequence, Proposition 8.1 implies that the Jacobian-variance weighted CLR test has correct asymptotic size and is asymptotically similar when $p = 1$. $\square$

14 The Eigenvalue Condition in $\mathcal{F}_0$

In this section, we show that the restriction $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_{0j}$, defined in (3.9), is not redundant. If this restriction is weakened to $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$, we show that, for some models, some sequences of distributions, and some (consistent) choices of variance and covariance estimators, the LM statistic in (4.2) has a χ^2_k asymptotic distribution. This leads to over-rejection of the null when the standard χ^2_p critical value is used and the parameters are over-identified (i.e., $k > p$). On the other hand, we show that the LM statistic equals zero a.s. for some models and some distributions F if the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ is removed entirely. This implies that the LM test also under-rejects the null hypothesis and is nonsimilar in both finite samples and asymptotically for some F.

All of the CLR tests considered in Sections 5 and 6, except that of Smith (2007), are functions of the LM statistic in (4.2) (and other statistics). In consequence, the aberrant behavior of the LM statistic and test demonstrated in this section, when the restriction $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_0$ is weakened or eliminated, carries over to the CLR statistics and tests in Sections 5 and 6. Smith's (2007) CLR test is a function of the LM statistic in (4.2) but with $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ replaced by $\widehat{D}_n^{\dagger}$.

14.1 Eigenvalue Condition Counter-Examples

For simplicity, we consider the case $p = 1$ in this section. As above, the null hypothesis is $H_0 : \theta = \theta_0$.

Lemma 14.1 (a) *Suppose $\mathcal{F}_0$ is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$ in place of $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_{0j}$ for all $j \in \{0, ..., p\}$, where $p = 1$. Suppose $\widehat{\Omega}_n(\theta)$ is defined in* (3.1) *and $\widehat{\Gamma}_{1n}(\theta) = n^{-1}\sum_{i=1}^n G_i(\theta)g_i(\theta)'$ (which differs from its definition in* (3.2)*). Then, there exist moment functions $g(W_i, \theta)$ and a sequence of null distributions $\{F_n \in \mathcal{F}_0 : n \geq 1\}$ for which $\widehat{\Omega}_n = \widehat{\Omega}_n(\theta_0)$ and $\widehat{\Gamma}_{1n} = \widehat{\Gamma}_{1n}(\theta_0)$ are well-behaved (in the sense that $\widehat{\Omega}_n - E_{F_n}g_ig_i' \to_p 0^{k\times k}$ and $\widehat{\Gamma}_{1n} - E_{F_n}G_ig_i' \to_p 0^{k\times k}$) and $LM_n(\theta_0) = AR_n(\theta_0) + o_p(1) \to_d \chi^2_k$.*

(b) *Suppose $\mathcal{F}_0$ is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ deleted in $\mathcal{F}_{0j}$ for all $j \in \{0, ..., p\}$, where $p = 1$. Suppose $\widehat{\Omega}_n(\theta)$ and $\widehat{\Gamma}_{1n}(\theta)$ are defined in* (3.1) *and* (3.2), *respectively. Then, there exists moment functions and a null distribution $F \in \mathcal{F}_0$ for which $LM_n(\theta_0) = 0$ a.s. for all $n \geq 1$.*

Comments: (i) The model we use to prove Lemma 14.1(a) is the linear IV regression model with one endogenous rhs variable and (for simplicity) no exogenous variables. Specifically, the model is

$$y_{1i} = y_{2i}\theta + u_i \text{ and } y_{2i} = Z_i'\pi + v_{2i}, \tag{14.1}$$

where $y_{1i}, \theta, y_{2i}, v_{2i} \in R$, $Z_i, \pi \in R^k$, $v_{2i} = \rho u_i + \delta\xi_i$ for some random variable ξ_i, $\delta = (1-\rho^2)^{1/2}$, and the observations are i.i.d. across i for any given n. The parameter space $\mathcal{F}^*$ for the distribution F of the random vector $W_i = (y_{1i}, y_{2i}, Z_i')'$ is

$$\begin{aligned}\mathcal{F}^* := \{F &: (14.1) \text{ holds with } \theta = \theta_0,\ \pi = \pi_F \in R^k,\ \rho = \rho_F \in (-1,1), \\ &Z_i,\ u_i, \text{ and } \xi_i \text{ are mutually independent, } E_F u_i = E_F\xi_i = 0, \\ &E_F u_i^2 = E_F\xi_i^2 = 1,\ E_F||(u_i, \xi_i, Z_i'Z_i)||^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(E_F Z_i Z_i') \geq \delta\}\end{aligned} \tag{14.2}$$

for some $\gamma, \delta > 0$ and $M < \infty$. As defined, ρ is the correlation between u_i and v_{2i}.

The moment functions are $g(W_i,\theta) = Z_i(y_{1i} - y_{2i}\theta)$. When the null value θ_0 is the true value, this gives $g_i = g_i(\theta_0) = Z_i u_i$ and $G_i = G_i(\theta_0) = -Z_i y_{2i}$. The set $\mathcal{F}^*$ is a subset of $\mathcal{F}_0$ when the latter is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$ in place of $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$. This holds because (i) for all $F \in \mathcal{F}^*$, $\lambda_{\min}(\Psi_F^{vec(G_i)}) > 0$ (by the argument in the paragraph that contains (3.12) because $\lambda_{\min}(E_F Z_i Z_i') > 0$ and $\lambda_{\min}(E_F \varepsilon_i \varepsilon_i') > 0$, where $\varepsilon_i = (u_i, -\rho u_i - \delta \xi_i)'$ for $\rho \in (-1,1)$), (ii) $\lambda_{\min}(E_F g_i g_i') = E_F u_i^2 \lambda_{\min}(E_F Z_i Z_i') \geq \delta > 0$, and (iii) $\lambda_{p-j}(\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_i B_{F,p-j}\xi}) \geq \lambda_{\min}(\Psi_F^{vec(G_i)}) M^{-2/(2+\gamma)}$ for all $\xi \in R^{p-j}$ with $||\xi|| = 1$ and all $j \in \{0,...,p\}$ (by the results and arguments in the paragraphs that contain (18.1)-(18.3), which verify that condition (iv), stated in (3.10), is a sufficient condition for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$). The quantity $\lambda_{\min}(\Psi_F^{vec(G_i)})$ is arbitrarily close to zero for ρ arbitrarily close to one.

We consider a sequence of distributions $\{F_n \in \mathcal{F}^* : n \geq 1\}$ for which $\pi_{F_n} = 0^k$ for all $n \geq 1$, ρ_n $(= \rho_{F_n}) \to 1$, and $E_{F_n} Z_i Z_i'$ does not depend on n. For these distributions,

$$G_i = -\rho_n g_i + \delta_n G_i^*, \text{ where } G_i^* := -Z_i \xi_i \text{ and } \delta_n := (1-\rho_n^2)^{1/2}. \tag{14.3}$$

In this case, the IV's are irrelevant and the degree of endogeneity is close to perfect for n large.

(ii) The model we consider in Lemma 14.1(b) is the same as that in part (a) except that $\mathcal{F}^*$ allows for $\rho = \rho_F \in (-1,1]$ and we consider a single distribution F with $\pi = 0^k$ and $\rho = 1$, rather than a drifting sequence of distributions. For this distribution, $\lambda_{\min}(\Psi_F^{vec(G_i)}) = 0$.

(iii) The intuition for the results in Lemma 14.1(a) and (b) is as follows. As (14.3) shows, G_i is close to being proportional to g_i when $\pi_{F_n} = 0^k$ and ρ_n is close to one. And, when $\pi_{F_n} = 0^k$ and $\rho_n = 1$, they are exactly proportional. By averaging over $i = 1,...,n$ and by taking expectations, the same properties are seen to hold for $\widehat{G}_n$ and $\widehat{g}_n$ and their population counterparts. In consequence, $\widehat{D}_n$ $(:= \widehat{G}_n - \widehat{\Gamma}_n \widehat{\Omega}_n^{-1} \widehat{g}_n$ when $p = 1)$ is close to 0^k (because it is a sample version of the $L^2(F)$ projection of G_i on g_i) and the same is true of the population counterpart of $\widehat{D}_n$ (because it is the $L^2(F)$ projection of G_i on g_i). The latter implies that the direction of the k-vector $\widehat{D}_n$ is primarily random. In consequence, this direction turns out to be sensitive to the specification of the sample matrices $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ even within the class of consistent estimators of their population counterparts.

One consistent choice of $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ (used in Lemma 14.1(a)) yields $\widehat{D}_n$ to be very close to being proportional to $\widehat{g}_n$. In this case, the projection of $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ is asymptotically equivalent to $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ itself. The LM statistic is a quadratic form in this projection k-vector (i.e., $P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} \widehat{\Omega}_n^{-1/2}\widehat{g}_n$) multiplied by n. Hence, it behaves asymptotically like a quadratic form in $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ multiplied by n, which is just the AR statistic. This explains the result in Lemma 14.1(a).

On the other hand, when $\rho_n = 1$ (which implies that $\widehat{G}_n = -\widehat{g}_n$ by (14.3)), another consistent choice of $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ (used in Lemma 14.1(b)) yields $\widehat{D}_n = 0^k$ a.s. In this case, the projection of $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ equals 0^k a.s. Hence, the LM statistic (which is a quadratic form in this projection times n) equals zero a.s. This explains the result in Lemma 14.1(b).

(iv) The result of Lemma 14.1(a) also holds for the model described in Comment (ii). Hence, drifting sequences of distributions are not required to show the result of Lemma 14.1(a) if one removes the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ entirely from $\mathcal{F}_{0j}$. Furthermore, the result of Lemma 14.1(a) can be extended to cover weak IV cases (in which $\pi = \pi_n \neq 0^k$, but $\pi_n \to 0^k$ sufficiently quickly as $n \to \infty$), rather than the irrelevant IV case (in which $\pi = 0^k$).

(v) In the extreme case of the model, where $\rho = 1$ and $\pi = 0$, the endogenous variables y_{1i} and y_{2i} are identical, which is similar to perfect multicollinearity in linear regression. However, the result of Lemma 14.1(a) does not require either ρ to be exactly equal to one or π to be exactly equal to zero.

(vi) Finite sample simulations corroborate the asymptotic result given in Lemma 14.1(a). For the model and LM test described in Comment (i) with $k = 5$, $\pi = 0^k$, $\rho = 1$, $Z_i \sim N(0^5, I_5)$, $(u_i, \xi_i) \sim N(0^2, I_2)$, and Z_i independent of (u_i, ξ_i), the null rejection rate of the nominal 5% LM test is 59.4% when $n = 200$ and 57.6% when $n = 1000$. However, when ρ deviates from 1 even by a small amount, the magnitude of over-rejection drops very quickly. The null rejection rate of this nominal 5% LM test is 10.1% when $\rho = 0.99$ and $n = 200$ and 12.9% when $\rho = 0.998$ and $n = 1000$. (These simulation results are based on 50,000 simulation repetitions.)

(vii) The conditions of Lemma 14.1(a) and (b) are consistent with those of Theorem 1 of Kleibergen (2005). This implies that the χ^2_p asymptotic distribution of the LM statistic obtained in the latter only holds under additional conditions, such as those in $\mathcal{F}_0$.

14.2 Proof of Lemma 14.1

Proof of Lemma 14.1. To prove part (a), we use the model defined in (14.1)-(14.3). We have

$$\begin{aligned}
\widehat{G}_n &= -\rho_n \widehat{g}_n + \delta_n \widehat{G}_n^*, \text{ where } \widehat{G}_n^* := n^{-1}\sum_{i=1}^n G_i^*, \text{ and} \\
\widehat{\Gamma}_{1n} &= n^{-1}\sum_{i=1}^n G_i g_i' = n^{-1}\sum_{i=1}^n (-\rho_n g_i + \delta_n G_i^*) g_i' = -\rho_n \widehat{\Omega}_n - \rho_n \widehat{g}_n \widehat{g}_n' + \delta_n \widehat{\Gamma}_{1n}^*, \text{ where} \\
\widehat{\Gamma}_{1n}^* &:= n^{-1}\sum_{i=1}^n G_i^* g_i'. \qquad (14.4)
\end{aligned}$$

We choose $\{\rho_n : n \geq 1\}$ to converge to one sufficiently fast that $n\delta_n \to 0$, where $\delta_n = (1-\rho_n^2)^{1/2}$

by (14.3). For example, we can take $\rho_n = (1 - n^{-3})^{1/2}$. Using the results above, we obtain

$$\begin{aligned}
\widehat{D}_n &= \widehat{G}_n - \widehat{\Gamma}_{1n}\widehat{\Omega}_n^{-1}\widehat{g}_n \\
&= -\rho_n\widehat{g}_n + \delta_n\widehat{G}_n^* - [-\rho_n\widehat{\Omega}_n - \rho_n\widehat{g}_n\widehat{g}_n' + \delta_n\widehat{\Gamma}_{1n}^*]\widehat{\Omega}_n^{-1}\widehat{g}_n \\
&= \rho_n(\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)\widehat{g}_n + \delta_n(\widehat{G}_n^* - \widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1}\widehat{g}_n). \qquad (14.5)
\end{aligned}$$

This gives

$$\begin{aligned}
&\widetilde{g}_n := \widehat{g}_n + n\delta_n\zeta_n = \widehat{D}_n/(\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n), \text{ where} \\
&\zeta_n := (\widehat{G}_n^* - \widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1}\widehat{g}_n)/(\rho_n n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n) = O_p(n^{-1/2}) \text{ and } \widetilde{g}_n = \widehat{g}_n + o_p(n^{-1/2}), \qquad (14.6)
\end{aligned}$$

where $\zeta_n = O_p(n^{-1/2})$ because $\rho_n \to 1$, $\widehat{G}_n^* = O_p(n^{-1/2})$ by the CLT since $E_{F_n}G_i^* = -E_{F_n}Z_i \cdot E_{F_n}\xi_i = 0^k$, $\widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1} = O_p(1)$ by the WLLN applied twice and $\lambda_{\min}(E_{F_n}g_ig_i') = \lambda_{\min}(E_{F_n}Z_iZ_i') \geq \delta > 0$, $\widehat{g}_n = O_p(n^{-1/2})$ by the CLT, and $(n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)^{-1} = O_p(1)$, which holds by the CMT because $AR_n = n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \to_d \chi_k^2$ (by the CLT, WLLN, and CMT) and $\chi_k^2 > 0$ a.s., and lastly the result for $\widetilde{g}_n$ in the second line of (14.6) holds by $\zeta_n = O_p(n^{-1/2})$ and $n\delta_n = o(1)$.

Projections are invariant to nonzero scalar multiplications of the matrix that defines the projection. That is, $P_A = P_{cA}$ for any matrix A and any scalar $c \neq 0$. We have $\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \neq 0$ wp$\to$1 because $(n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)^{-1} = O_p(1)$ and $\rho_n \to 1$. So, the LM statistic is unchanged wp$\to$1 when $\widehat{D}_n$ is replaced by $\widehat{D}_n/(\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n) = \widetilde{g}_n = \widehat{g}_n + o_p(n^{-1/2})$ using (14.6). Thus, we have

$$\begin{aligned}
LM_n &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \\
&= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widetilde{g}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1) \\
&= n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widetilde{g}_n(\widetilde{g}_n'\widehat{\Omega}_n^{-1}\widetilde{g}_n)^{-1}\widetilde{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n + o_p(1) \\
&= n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n + o_p(1) = AR_n + o_p(1) \to_d \chi_k^2, \qquad (14.7)
\end{aligned}$$

which completes the proof of part (a).

Next, we prove part (b). In this case, we use the model in (14.1)-(14.3) with $\rho_n = 1$ and $\delta_n = 0$ for all $n \geq 1$. In consequence, $G_i = -g_i$ and $\widehat{G}_n = -\widehat{g}_n$. Given the definitions of $\widehat{\Omega}_n$ and $\widehat{\Gamma}_{1n}$ in (3.1) and (3.2), this yields

$$\begin{aligned}
\widehat{\Gamma}_{1n} &= n^{-1}\sum_{i=1}^n G_ig_i' - \widehat{G}_n\widehat{g}_n' = -n^{-1}\sum_{i=1}^n g_ig_i' + \widehat{g}_n\widehat{g}_n' = -\widehat{\Omega}_n, \\
\widehat{D}_n &= \widehat{G}_n - \widehat{\Gamma}_{1n}\widehat{\Omega}_n^{-1}\widehat{g}_n = 0^k, \text{ and} \\
LM_n &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{0^k}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = 0 \qquad (14.8)
\end{aligned}$$

for all $n \geq 1$, where the projection matrix, P_{0^k}, onto 0^k equals $0^{k\times k}$. $\square$

15 Proof of Lemma 10.2

Lemma 10.2 of AG1. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$,

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

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 10.2. We have

$$\begin{aligned} n^{1/2}vec(\widehat{D}_n - D_n) &= n^{-1/2}\sum_{i=1}^{n} vec(G_i - D_n) - \begin{pmatrix} \widehat{\Gamma}_{1n} \\ \vdots \\ \widehat{\Gamma}_{pn} \end{pmatrix} \widehat{\Omega}_n^{-1} n^{1/2}\widehat{g}_n \\ &= n^{-1/2}\sum_{i=1}^{n}\left[vec(G_i - D_n) - \begin{pmatrix} E_{F_n}G_{\ell 1}g_\ell' \\ \vdots \\ E_{F_n}G_{\ell p}g_\ell' \end{pmatrix}\Omega_{F_n}^{-1}g_i\right] + o_p(1), \end{aligned} \tag{15.1}$$

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 (15.1), the convergence result of Lemma 10.2 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 10.2 is a zero matrix because

$$E_{F_n}[G_{ij} - D_{nj} - (E_{F_n}G_{\ell j}g_\ell')\Omega_{F_n}^{-1}g_i]g_i' = 0^{k\times k}, \tag{15.2}$$

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 10.2 equals

$$\lim Var_{F_n}(vec(G_i) - (E_{F_n}vec(G_\ell)g_\ell')\Omega_{F_n}^{-1}g_i) = \lim \Phi_{F_n}^{vec(G_i)} = \Phi_h^{vec(G_i)}, \tag{15.3}$$

see (10.15), 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) $\lambda_{s,F_n} \rightarrow h_s$ for $s = 4, 5$. By the CLT, the limiting variance matrix of

$n^{1/2}\widehat{g}_n$ equals $\lim E_{F_n} g_i g_i' = h_{5,g}$. $\square$

16 Proof of Lemma 10.3

Lemma 10.3 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* $\lambda_{n,h} \in \Lambda_*$,

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

where (a) $(\overline{g}_h, \overline{D}_h)$ *are defined in Lemma* 10.2, (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (10.17), (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 10.3 uses the following two lemmas and corollary.

Lemma 16.1 *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

$$\Delta = \begin{pmatrix} Z_1 & Z_2 \\ Z_3 & Z_1 \end{pmatrix}. \tag{16.1}$$

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

$$
\begin{aligned}
Var(\mu'\Delta\xi) &= Var(\mu_1[Z_1\xi_1 + Z_2\xi_2] + \mu_2[Z_3\xi_1 + Z_1\xi_2]) \\
&= Var((\mu_1\xi_1 + \mu_2\xi_2)Z_1 + \mu_1\xi_2 Z_2 + \mu_2\xi_1 Z_3) \\
&= (\mu_1\xi_1 + \mu_2\xi_2)^2 + (\mu_1\xi_2)^2 + (\mu_2\xi_1)^2. \qquad (16.2)
\end{aligned}
$$

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 16.1. 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 16.2 *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 16.2 follows from Lemma 16.1 by the following argument. We have

$$
M\Delta = \begin{pmatrix} M_1\Delta_{q_*} & M_1\Delta_{p-q_*} \\ M_2\Delta_{q_*} & M_2\Delta_{p-q_*} \end{pmatrix} = \begin{pmatrix} I_{q_*} & M_1\Delta_{p-q_*} \\ 0^{(k-q_*)\times q_*} & M_2\Delta_{p-q_*} \end{pmatrix}. \qquad (16.3)
$$

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 16.1 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 16.3 *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 16.3.

Corollary 16.4 *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 10.3. First, we prove the convergence result in Lemma 10.3. The singular value decomposition of $W_n D_n U_n$ is

$$W_n D_n U_n = C_n \Upsilon_n B_n', \tag{16.4}$$

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

Using (16.4), we get

$$W_n D_n U_n B_{n,q} \Upsilon_{n,q}^{-1} = C_n \Upsilon_n B_n' B_{n,q} \Upsilon_{n,q}^{-1} = C_n \Upsilon_n \begin{pmatrix} I_q \\ 0^{(p-q)\times q} \end{pmatrix} \Upsilon_{n,q}^{-1} = C_n \begin{pmatrix} I_q \\ 0^{(k-q)\times q} \end{pmatrix} = C_{n,q}, \tag{16.5}$$

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

$$\begin{aligned} W_n \widehat{D}_n U_n B_{n,q} \Upsilon_{n,q}^{-1} &= W_n D_n U_n B_{n,q} \Upsilon_{n,q}^{-1} + W_n n^{1/2}(\widehat{D}_n - D_n) U_n B_{n,q} (n^{1/2} \Upsilon_{n,q})^{-1} \\ &= C_{n,q} + o_p(1) \rightarrow_p h_{3,q} = \overline{\Delta}_{h,q}, \end{aligned} \tag{16.6}$$

where the second equality uses $n^{1/2} \tau_{jF_n} \rightarrow \infty$ for all $j \leq q$ (by the definition of q in (10.16)), $W_n = O(1)$ (by the condition $||W_F|| \leq M_1 < \infty \; \forall F \in \mathcal{F}_{WU}$, see (10.5)), $n^{1/2}(\widehat{D}_n - D_n) = O_p(1)$ (by Lemma 10.2), $U_n = O(1)$ (by the condition $||U_F|| \leq M_1 < \infty \; \forall F \in \mathcal{F}_{WU}$, see (10.5)), and $B_{n,q} \rightarrow h_{2,q}$ with $||vec(h_{2,q})|| < \infty$ (by (10.12) using the definitions in (10.17) and (9.1)). The convergence in (16.6) holds by (10.12), (10.17), and (9.1), and the last equality in (16.6) holds by the definition of $\overline{\Delta}_{h,q}$ in (10.17).

Using (16.4) again, we have

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

$$= C_n\begin{pmatrix} 0^{q\times(p-q)} \\ n^{1/2}\Upsilon_{n,p-q} \\ 0^{(k-p)\times(p-q)}\end{pmatrix} \to h_3\begin{pmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1},...,h_{1,p}\} \\ 0^{(k-p)\times(p-q)}\end{pmatrix} = h_3h_{1,p-q}^{\diamond}, \tag{16.7}$$

where the second equality uses $B_n'B_n = I_p$, the convergence holds by (10.12) using the definitions in (10.17) and (9.2), and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (10.17).

Using (16.7) and Lemma 10.2, we get

$$\begin{aligned} n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} \;&=\; n^{1/2}W_nD_nU_nB_{n,p-q} + W_nn^{1/2}(\widehat{D}_n - D_n)U_nB_{n,p-q} \\ &\to_d h_3h_{1,p-q}^{\diamond} + h_{71}\overline{D}_hh_{81}h_{2,p-q} = \overline{\Delta}_{h,p-q}, \end{aligned} \tag{16.8}$$

where $B_{n,p-q} \to h_{2,p-q}$, $W_n \to h_{71}$, and $U_n \to h_{81}$ by (10.3), (10.12), (10.17), and Assumption WU using the definitions in (9.1) and the last equality holds by the definition of $\overline{\Delta}_{h,p-q}$ in (10.17).

Equations (16.6) and (16.8) combine to prove

$$\begin{aligned} n^{1/2}W_n\widehat{D}_nU_nT_n \;&=\; n^{1/2}W_n\widehat{D}_nU_nB_nS_n = (W_n\widehat{D}_nU_nB_{n,q}\Upsilon_{n,q}^{-1}, n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q}) \\ &\to_d (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) = \overline{\Delta}_h \end{aligned} \tag{16.9}$$

using the definition of S_n in (10.19). The convergence is joint with that in Lemma 10.2 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 10.3.

Properties (a) and (b) in Lemma 10.3 hold by definition. Property (c) in Lemma 10.3 holds by Lemma 10.2 and property (b) in Lemma 10.3.

Now, we prove property (d). We have

$$h_{2,p-q}'h_{2,p-q} = \lim B_{n,p-q}'B_{n,p-q} = I_{p-q} \text{ and } h_{3,q}'h_{3,q} = \lim C_{n,q}'C_{n,q} = I_q \tag{16.10}$$

because B_n and C_n are orthogonal matrices by (10.6) and (10.7). 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 16.2 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 16.2 gives the desired result that $P(\overline{\Delta}_h$ has full column rank $p) = 1$. The condition in Corollary 16.2 that "$M\Delta_{q_*} = (e_1, ..., e_{q_*})$" holds in this case because $h_3'\overline{\Delta}_{h,q} = h_3' h_{3,q} = (e_1, ..., e_q)$. The condition in Corollary 16.2 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

$$\lambda_{p-q}\left(Var(h_{3,k-q}'\overline{\Delta}_{h,p-q}\xi_2)\right) > 0 \ \forall \xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{16.11}$$

We have

$$\begin{aligned} Var(h_{3,k-q}'\overline{\Delta}_{h,p-q}\xi_2) &= Var(h_{3,k-q}' h_{5,g}^{-1/2}\overline{D}_h h_{2,p-q}\xi_2) \\ &= ((h_{2,p-q}\xi_2)' \otimes (h_{3,k-q}' h_{5,g}^{-1/2}))Var(vec(\overline{D}_h))((h_{2,p-q}\xi_2) \otimes (h_{3,k-q}' h_{5,g}^{-1/2})') \\ &= ((h_{2,p-q}\xi_2)' \otimes (h_{3,k-q}' h_{5,g}^{-1/2}))\Phi_h^{vec(G_i)}((h_{2,p-q}\xi_2) \otimes (h_{3,k-q}' h_{5,g}^{-1/2})') \\ &= \Phi_h^{h_{3,k-q}' h_{5,g}^{-1/2} G_i h_{2,p-q}\xi_2}, \end{aligned} \tag{16.12}$$

where the first equality holds by the definition of $\overline{\Delta}_{h,p-q}$ in (10.17) and the fact that $h_{71} = h_{5,g}^{-1/2}$ and $h_{81} = I_p$ by the conditions in part (d) of Lemma 10.3, 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 10.2, and the fourth equality uses the definition of the variance matrix $\Phi_h^{a_i}$ in (10.15) for an arbitrary random vector a_i.

Next, we show that $\Phi_h^{h_{3,k-q}' h_{5,g}^{-1/2} G_i h_{2,p-q}\xi_2}$ equals the expected outer-product matrix $\lim \Psi_{F_n}^{C_{n,k-q}'\Omega_n^{-1/2} G_i B_{n,p-q}\xi_2}$:

$$\begin{aligned} &\Phi_h^{h_{3,k-q}' h_{5,g}^{-1/2} G_i h_{2,p-q}\xi_2} \\ &= ((h_{2,p-q}\xi_2)' \otimes (h_{3,k-q}' h_{5,g}^{-1/2}))\Phi_h^{vec(G_i)}((h_{2,p-q}\xi_2) \otimes (h_{3,k-q}' h_{5,g}^{-1/2})') \\ &= \lim((B_{n,p-q}\xi_2)' \otimes (C_{n,k-q}'\Omega_n^{-1/2}))\Phi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C_{n,k-q}'\Omega_n^{-1/2})') \\ &= \lim((B_{n,p-q}\xi_2)' \otimes (C_{n,k-q}'\Omega_n^{-1/2}))\Psi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C_{n,k-q}'\Omega_n^{-1/2})') \\ &\quad - \lim((B_{n,p-q}\xi_2)' \otimes (C_{n,k-q}'\Omega_n^{-1/2}))E_{F_n}vec(G_i) \cdot E_{F_n}vec(G_i)'((B_{n,p-q}\xi_2) \otimes (C_{n,k-q}'\Omega_n^{-1/2})') \\ &= \lim((B_{n,p-q}\xi_2)' \otimes (C_{n,k-q}'\Omega_n^{-1/2}))\Psi_{F_n}^{vec(G_i)}((B_{n,p-q}\xi_2) \otimes (C_{n,k-q}'\Omega_n^{-1/2})') \\ &\quad - \lim E_{F_n}vec(C_{n,k-q}'\Omega_n^{-1/2} G_i B_{n,p-q}\xi_2) \cdot E_{F_n}vec(C_{n,k-q}'\Omega_n^{-1/2} G_i B_{n,p-q}\xi_2)' \\ &= \lim \Psi_{F_n}^{C_{n,k-q}'\Omega_n^{-1/2} G_i B_{n,p-q}\xi_2}, \end{aligned} \tag{16.13}$$

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 (3.6) and (10.15), 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 (16.7) 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 (10.16).

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 $||\xi|| = 1$,

$$\lambda_{p-j}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-j}\xi}\right) = \lim \lambda_{p-j}\left(\Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-j}\xi}\right) > 0. \tag{16.14}$$

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 (16.14),

$$\lambda_{p-j}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-j}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}\right) > 0 \ \forall \xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{16.15}$$

Next, we apply Corollary 16.4(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 16.4 and (16.15) give

$$\lambda_{p-q}\left(\lim \Psi_{F_{n_m}}^{C'_{n_m,k-q}\Omega_{n_m}^{-1/2}G_iB_{n_m,p-q}\xi_2}\right) > 0 \ \forall \xi_2 \in R^{p-q} \text{ with } ||\xi_2|| = 1. \tag{16.16}$$

Equations (16.12), (16.13), and (16.16) combine to establish (16.11) 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. $\square$

Proof of Lemma 16.1. 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

$$||\Delta\xi - \Delta\xi_s|| = \left(\sum_{r=1}^{k}(\Delta_r'(\xi - \xi_s))^2\right)^{1/2} \leq \left(\sum_{r=1}^{k}||\Delta_r||^2||\xi - \xi_s||^2\right)^{1/2} = \varepsilon||vec(\Delta)||, \qquad (16.17)$$

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

$$||\Delta\xi_s|| \leq \varepsilon||vec(\Delta)||. \qquad (16.18)$$

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

$$P(||Z_*|| \leq \varepsilon) = \int_{\{||z|| \leq \varepsilon\}} f_{Z_*}(z)dz \leq \sup_{z \in R^k} f_{Z_*}(z) \int_{\{||z|| \leq \varepsilon\}} dz \leq C_2\varepsilon^p \qquad (16.19)$$

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

$$\begin{aligned}
& P(\Delta\xi = 0^k \text{ for some } \xi \in R^p \text{ with } ||\xi|| = 1) \\
= \ & P\left(\cup_{s=1}^{N_\varepsilon} \cup_{\xi \in B(\xi_s, \varepsilon): ||\xi||=1} \{\Delta\xi = 0^k\}\right) \\
\leq \ & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon ||vec(\Delta)||\}\right) \\
\leq \ & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon ||vec(\Delta)||\} \cap \{||vec(\Delta)|| \leq K_\gamma\}\right) + P(||vec(\Delta)|| > K_\gamma) \\
\leq \ & P\left(\cup_{s=1}^{N_\varepsilon} \{||\Delta\xi_s|| \leq \varepsilon K_\gamma\}\right) + \gamma \\
\leq \ & \sum_{s=1}^{N_\varepsilon} P(||\Delta\xi_s|| \leq \varepsilon K_\gamma) + \gamma \\
\leq \ & \sum_{s=1}^{N_\varepsilon} P(||(B_{\xi_s}'\Delta\xi_s)_p|| \leq \varepsilon K_\gamma) + \gamma \\
\leq \ & N_\varepsilon C_2 K_\gamma^p \varepsilon^p + \gamma \\
\leq \ & C_1 \varepsilon^{-p+1} C_2 K_\gamma^p \varepsilon^p + \gamma \\
\to \ & \gamma \text{ as } \varepsilon \to 0, \qquad (16.20)
\end{aligned}$$

where the first inequality holds by (16.18) 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 (16.19), the second last inequality holds by (16.19) 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 (16.19), and the last inequality holds by the bound given above on N_ε.

Because $\gamma > 0$ is arbitrary, (16.20) implies that $P(\Delta\xi = 0^k$ for some $\xi \in R^p$ with $||\xi|| = 1) = 0$, which completes the proof. □

17 Proof of Theorem 10.4

Theorem 10.4 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* 10.3, *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 10.4 uses the following rate of convergence lemma. This lemma is a key technical contribution of the paper.

Lemma 17.1 *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* (10.16) *satisfies* $q \geq 1$, *we have* (a) $\widehat{\kappa}_{jn} \rightarrow_p \infty$ *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 17.1. By the definitions in (10.9) and (10.12), $h_{6,j} := \lim \tau_{(j+1)F_n}/\tau_{jF_n}$ for $j = 1, ..., p-1$. By the definition of q in (10.16), $h_{6,q} = 0$ if $q < p$. If $q = p$, $h_{6,q}$ is not defined by (10.9) and (10.12) 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., $0 < \lim \tau_{(j+1)F_n}/\tau_{jF_n} \leq 1$.[11] 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 $\{\tau_{r_g F_n} : n \geq 1\}, ..., \{\tau_{r_g^\diamond F_n} : n \geq 1\}$. 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 $\{\tau_{r_g F_n} : n \geq 1\}$, 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 $r_1 = r_1^\diamond = 1$.

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

$$|\tau_{r_1F_n}^{-2} B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - (n^{1/2}\tau_{r_1F_n})^{-2}\kappa B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n| = 0 \tag{17.1}$$

[11] 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.

wp→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 (10.5) because $\Lambda_* \subset \Lambda_2$ and Λ_2 only contains distributions in $\mathcal{F}_{WU}$), $|\widehat{U}_n^{-1}| > 0$ wp→1 (because $\widehat{U}_n \to_p h_{81}$ by (10.2), (10.12), (10.17), 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→1 from some statements below.) Thus, $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j \leq p\}$ solve

$$\begin{aligned}
&|\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - \kappa(I_p + \widehat{A}_n)| = 0 \text{ or}\\
&|(I_p + \widehat{A}_n)^{-1}\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n - \kappa I_p| = 0, \text{ where}\\
&\widehat{A}_n = \begin{bmatrix} \widehat{A}_{1n} & \widehat{A}_{2n} \\ \widehat{A}_{2n}' & \widehat{A}_{3n} \end{bmatrix} := B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n - I_p
\end{aligned} \tag{17.2}$$

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

$$\begin{aligned}
&\tau_{r_1F_n}^{-1}\widehat{W}_n\widehat{D}_nU_nB_n\\
&= \tau_{r_1F_n}^{-1}(\widehat{W}_nW_n^{-1})W_nD_nU_nB_n - (n^{1/2}\tau_{r_1F_n})^{-1}\widehat{W}_nn^{1/2}(\widehat{D}_n - D_n)U_nB_n\\
&= \tau_{r_1F_n}^{-1}(\widehat{W}_nW_n^{-1})C_n\Upsilon_n + O_p((n^{1/2}\tau_{r_1F_n})^{-1}) \qquad (17.3)\\
&= (I_k + o_p(1))C_n\begin{bmatrix} h_{6,r_1^\diamond}^\diamond + o(1) & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(p-r_1^\diamond)\times r_1^\diamond} & O(\tau_{r_2F_n}/\tau_{r_1F_n})^{(p-r_1^\diamond)\times(p-r_1^\diamond)} \\ 0^{(k-p)\times r_1^\diamond} & 0^{(k-p)\times(p-r_1^\diamond)} \end{bmatrix} + O_p((n^{1/2}\tau_{r_1F_n})^{-1})\\
&\to_p h_3\begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix}, \text{ where } h_{6,r_1^\diamond}^\diamond := Diag\{1, h_{6,1}, h_{6,1}h_{6,2}, ..., \prod_{\ell=1}^{r_1^\diamond-1} h_{6,\ell}\},
\end{aligned}$$

$h_{6,r_1^\diamond}^\diamond \in R^{r_1^\diamond\times r_1^\diamond}$, $h_{6,r_1^\diamond}^\diamond := 1$ 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 (16.4), $\widehat{W}_n \to_p h_{71}$ (by Assumption WU(a) and (c)), $||h_{71}|| = ||\lim W_n|| < \infty$ (by the conditions in $\mathcal{F}_{WU}$ defined in (10.5)), $n^{1/2}(\widehat{D}_n - D_n) = O_p(1)$ (by Lemma 10.2), $U_n = O(1)$ (by the conditions in $\mathcal{F}_{WU}$), and $B_n = O(1)$ (because B_n is orthogonal), the third equality uses $\widehat{W}_nW_n^{-1} \to_p I_k$ (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 (10.16) because $r_1 \leq q$). Note that, 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.

Equation (17.3) yields

$$\begin{aligned}\tau_{r_1F_n}^{-2}B_n'U_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_nU_nB_n &\to_p \begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)}\end{bmatrix}' h_3'h_3 \begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & 0^{(k-r_1^\diamond)\times(p-r_1^\diamond)}\end{bmatrix} \\ &= \begin{bmatrix} h_{6,r_1^\diamond}^{\diamond 2} & 0^{r_1^\diamond\times(p-r_1^\diamond)} \\ 0^{(p-r_1^\diamond)\times r_1^\diamond} & 0^{(p-r_1^\diamond)\times(p-r_1^\diamond)}\end{bmatrix}, \qquad (17.4)\end{aligned}$$

where the equality holds because $h_3'h_3 = \lim C_n'C_n = I_k$ using (10.7).

In addition, we have

$$\widehat{A}_n := B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n - I_p \to_p 0^{p\times p} \qquad (17.5)$$

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 (17.2), (17.4), (17.5), and Slutsky's Theorem, the largest $r_1^\diamond$ eigenvalues of $\tau_{r_1F_n}^{-2}B_n'\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_nB_n$ (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

$$\begin{aligned}&((n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{1n}, ..., (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{r_1^\diamond n}) \to_p (1, h_{6,1}^2, h_{6,1}^2h_{6,2}^2, ..., \prod_{\ell=1}^{r_1^\diamond-1} h_{6,\ell}^2) \text{ and so} \\ &\widehat{\kappa}_{jn} \to_p \infty \; \forall j = 1, ..., r_1^\diamond \qquad (17.6)\end{aligned}$$

because $n^{1/2}\tau_{r_1F_n} \to \infty$ (by (10.16) 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 $\tau_{r_1F_n}^{-2}B_n'\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_nB_n$, i.e., $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j = r_1^\diamond + 1, ..., p\}$, satisfy

$$(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} \to_p 0 \; \forall j = r_1^\diamond + 1, ..., p. \qquad (17.7)$$

If $G_h = 1$, (17.6) proves part (a) of the lemma and (17.7) proves part (b) of the lemma (because in this case $r_1^\diamond = q$ and $\tau_{r_1F_n}/\tau_{\ell F_n} = O(1)$ 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 (17.3) (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

$$\begin{aligned}
\varrho_n &:= \tau_{r_1F_n}^{-2} B_{n,0,r_1^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} \\
&= \begin{bmatrix} h_{6,r_1^\diamond}^\diamond + o(1) \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} \end{bmatrix}' C_n'(I_k + o_p(1))C_n \begin{bmatrix} 0^{r_1^\diamond \times (p-r_1^\diamond)} \\ O(\tau_{r_2F_n}/\tau_{r_1F_n})^{(k-r_1^\diamond)\times(p-r_1^\diamond)} \end{bmatrix} \\
&\quad + O_p((n^{1/2}\tau_{r_1F_n})^{-1}) \\
&= o_p(\tau_{r_2F_n}/\tau_{r_1F_n}) + O_p((n^{1/2}\tau_{r_1F_n})^{-1}),
\end{aligned} \tag{17.8}$$

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 $o_p(\tau_{r_2F_n}/\tau_{r_1F_n})$.

Define

$$\begin{aligned}
\widehat{\xi}_{1n}(\kappa) &:= \tau_{r_1F_n}^{-2} B_{n,0,r_1^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,0,r_1^\diamond} - \kappa(I_{r_1^\diamond} + \widehat{A}_{1n}) \in R^{r_1^\diamond \times r_1^\diamond}, \\
\widehat{\xi}_{2n}(\kappa) &:= \varrho_n - \kappa \widehat{A}_{2n} \in R^{r_1^\diamond \times (p-r_1^\diamond)}, \text{ and} \\
\widehat{\xi}_{3n}(\kappa) &:= \tau_{r_1F_n}^{-2} B_{n,r_1^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} - \kappa(I_{p-r_1^\diamond} + \widehat{A}_{3n}) \in R^{(p-r_1^\diamond)\times(p-r_1^\diamond)}.
\end{aligned} \tag{17.9}$$

As in the first line of (17.2), $\{(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn} : j \le p\}$ solve

$$\begin{aligned}
0 &= |\tau_{r_1F_n}^{-2} B_n' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_n - \kappa(I_p + \widehat{A}_n)| \\
&= \left| \begin{bmatrix} \widehat{\xi}_{1n}(\kappa) & \widehat{\xi}_{2n}(\kappa) \\ \widehat{\xi}_{2n}(\kappa)' & \widehat{\xi}_{3n}(\kappa) \end{bmatrix} \right| \\
&= |\widehat{\xi}_{1n}(\kappa)| \cdot |\widehat{\xi}_{3n}(\kappa) - \widehat{\xi}_{2n}(\kappa)' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{\xi}_{2n}(\kappa)| \\
&= |\widehat{\xi}_{1n}(\kappa)| \cdot |\tau_{r_1F_n}^{-2} B_{n,r_1^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} - \varrho_n' \widehat{\xi}_{1n}^{-1}(\kappa) \varrho_n \\
&\quad - \kappa(I_{p-r_1^\diamond} + \widehat{A}_{3n} - \widehat{A}_{2n}' \widehat{\xi}_{1n}^{-1}(\kappa) \varrho_n - \varrho_n' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{A}_{2n} + \kappa \widehat{A}_{2n}' \widehat{\xi}_{1n}^{-1}(\kappa) \widehat{A}_{2n})|,
\end{aligned} \tag{17.10}$$

where the third equality uses the standard formula for the determinant of a partitioned matrix (i.e., the determinant of $\xi = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_2' & \xi_3 \end{bmatrix}$ 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)) and the result given in (17.11) below, which shows that $\widehat{\xi}_{1n}(\kappa)$ is nonsingular wp→1 for κ equal to any solution $(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}$ to the first equality in (17.10) for $j \leq p$, and the last equality holds by algebra.

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→1, where this determinant is the first multiplicand on the rhs of (17.10). 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 (17.10) wp→1. For $j = r_1^\diamond + 1, ..., p$, we have

$$\begin{aligned}
\widetilde{\xi}_{j1n} &:= \widehat{\xi}_{1n}((n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}) \\
&= \tau_{r_1F_n}^{-2} B'_{n,0,r_1^\diamond} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,0,r_1^\diamond} - (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}(I_{r_1^\diamond} + \widehat{A}_{1n}) \\
&= h_{6,r_1^\diamond}^{\diamond 2} + o_p(1) - o_p(1)(I_{r_1^\diamond} + o_p(1)) \\
&= h_{6,r_1^\diamond}^{\diamond 2} + o_p(1), \qquad (17.11)
\end{aligned}$$

where the second last equality holds by (17.4), (17.5), and (17.7). Equation (17.11) and $\lambda_{\min}(h_{6,r_1^\diamond}^{\diamond 2}) > 0$ (which follows from the definition of $h_{6,r_1^\diamond}^{\diamond}$ in (17.3) 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 (17.10) gives

$$\begin{aligned}
0 &= |\tau_{r_1F_n}^{-2} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p((\tau_{r_2F_n}/\tau_{r_1F_n})^2) + O_p((n^{1/2}\tau_{r_1F_n})^{-2}) \\
&\quad -(n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r_1^\diamond} + \widehat{A}_{j2n})|, \text{ where} \qquad (17.12) \\
\widehat{A}_{j2n} &:= \widehat{A}_{3n} - \widehat{A}'_{2n}\widetilde{\xi}_{j1n}^{-1}\varrho_n - \varrho'_n\widetilde{\xi}_{j1n}^{-1}\widehat{A}_{2n} + (n^{1/2}\tau_{r_1F_n})^{-2}\widehat{\kappa}_{jn}\widehat{A}'_{2n}\widetilde{\xi}_{j1n}^{-1}\widehat{A}_{2n} \in R^{(p-r_1^\diamond)\times(p-r_1^\diamond)}
\end{aligned}$$

using (17.8) and (17.11). Multiplying (17.12) by $\tau_{r_1F_n}^2/\tau_{r_2F_n}^2$ gives

$$0 = |\tau_{r_2F_n}^{-2} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p(1) - (n^{1/2}\tau_{r_2F_n})^{-2}\widehat{\kappa}_{jn}(I_{p-r_1^\diamond} + \widehat{A}_{j2n})| \qquad (17.13)$$

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 (10.16)).

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

$$0 = |\tau_{r_2F_n}^{-2} B'_{n,r_1^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_1^\diamond,p} + o_p(1) - \kappa(I_{p-r_1^\diamond} + \widehat{A}_{j2n})|. \qquad (17.14)$$

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

$$\widehat{A}_{j2n} = o_p(1), \qquad (17.15)$$

because $\widehat{A}_{2n} = o_p(1)$ and $\widehat{A}_{3n} = o_p(1)$ (by (17.5)), $\widetilde{\xi}_{j1n}^{-1} = O_p(1)$ (by (17.11)), $\varrho_n = o_p(1)$ (by (17.8) 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 (17.7)).

Now, we repeat the argument from (17.2) to (17.15) with the expression in (17.14) replacing that in the first line of (17.2), with (17.15) replacing (17.5), 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 $h_{6,r_2^\diamond}^\diamond = Diag\{1, h_{6,r_1^\diamond+1}, h_{6,r_1^\diamond+1}h_{6,r_1^\diamond+2}, ..., \prod_{\ell=r_1^\diamond+1}^{r_2^\diamond-1} h_{6,\ell}\} \in R^{(r_2^\diamond-r_1^\diamond)\times(r_2^\diamond-r_1^\diamond)}$ 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 $h_{6,r_1^\diamond}^\diamond$, 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 (17.8)-(17.10) 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 (17.6) and (17.7), which are

$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = r_2, ..., r_2^\diamond \text{ and } (n^{1/2}\tau_{r_2F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \ \forall j = r_2^\diamond + 1, ..., p. \tag{17.16}$$

In addition, the analogue of (17.14) is the same as (17.14) 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 (17.12) but with $\widehat{A}_{2j2n}$ and $\widehat{A}_{3j2n}$ in place of $\widehat{A}_{2n}$ and $\widehat{A}_{3n}$, respectively, where

$$\widehat{A}_{j2n} = \begin{bmatrix} \widehat{A}_{1j2n} & \widehat{A}_{2j2n} \\ \widehat{A}_{2j2n}' & \widehat{A}_{3j2n} \end{bmatrix} \tag{17.17}$$

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

$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = 1, ..., r_{G_h}^\diamond \text{ and } (n^{1/2}\tau_{r_gF_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \ \forall j = r_g^\diamond + 1, ..., p, \forall g = 1, ..., G_h. \tag{17.18}$$

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 (17.18) proves part (a) of the lemma.

The second result in (17.18) with $g = G_h$ implies: for all $j = q + 1, ..., p$,

$$(n^{1/2}\tau_{r_{G_h}F_n})^{-2}\widehat{\kappa}_{jn} = o_p(1) \tag{17.19}$$

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 (17.19). In the latter case, we have

$$\lim \frac{\tau_{qF_n}}{\tau_{r_{G_h}F_n}} = \lim \frac{\tau_{r^\diamond_{G_h}F_n}}{\tau_{r_{G_h}F_n}} = \prod_{j=r_{G_h}}^{r^\diamond_{G_h}-1} h_{6,j} > 0, \tag{17.20}$$

where the inequality holds because $h_{6,\ell} > 0$ for all $\ell \in \{r_{G_h}, ..., r^\diamond_{G_h} - 1\}$, 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 (17.19) and (17.20). 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 (17.18) 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 $(p - r^\diamond_{g-1}) \times (p - r^\diamond_{g-1})$ matrix whose (ℓ, m) element for $\ell, m = 1, ..., p - r^\diamond_{g-1}$ is $o_p(\tau_{(r^\diamond_{g-1}+\ell)F_n}\tau_{(r^\diamond_{g-1}+m)F_n}/\tau^2_{r_gF_n}) + O_p((n^{1/2}\tau_{r_gF_n})^{-1})$. Note that $o_{gp} = o_p(1)$ because $r^\diamond_{g-1} + \ell \geq r_g$ for $\ell \geq 1$ (since τ_{jF_n} are nonincreasing in j) and $n^{1/2}\tau_{r_gF_n} \to \infty$ for $g = 1, ..., G_h$.

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

$$|\tau^{-2}_{r_gF_n} B'_{n,r^\diamond_{g-1},p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_{g-1},p} + o_{gp} - \kappa(I_{p-r^\diamond_{g-1}} + \widehat{A}_{jgn})| = 0 \tag{17.21}$$

for some $(p - r^\diamond_{g-1}) \times (p - r^\diamond_{g-1})$ 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 (17.21) holds with $g = 1$ by the first line of (17.2) 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 (17.21) 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 (17.3), we have

$$\begin{aligned} \tau^{-1}_{r_gF_n} \widehat{W}_n \widehat{D}_n U_n B_{n,r^\diamond_{g-1},p} &= (I_k + o_p(1)) C_n \begin{bmatrix} 0^{r^\diamond_{g-1} \times (p - r^\diamond_{g-1})} \\ Diag\{\tau_{r_gF_n}, ..., \tau_{pF_n}\}/\tau_{r_gF_n} \\ 0^{(k-p)\times(p-r^\diamond_{g-1})} \end{bmatrix} + O_p((n^{1/2}\tau_{r_gF_n})^{-1}) \\ &\to_p h_3 \left(\begin{bmatrix} 0^{r^\diamond_{g-1} \times (r^\diamond_g - r^\diamond_{g-1})} \\ h^\diamond_{6,r^\diamond_g} \\ 0^{(k-r^\diamond_g)\times(r^\diamond_g - r^\diamond_{g-1})} \end{bmatrix}, 0^{k \times (p - r^\diamond_g)} \right), \text{ where } h^\diamond_{6,r^\diamond_g} := Diag\{1, h_{6,r_g}, ..., \prod_{j=r^\diamond_{g-1}+1}^{r^\diamond_g - 1} h_{6,j}\}, \end{aligned} \tag{17.22}$$

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

Equation (17.22) and $h_3' h_3 = \lim C_n' C_n = I_k$ yield

$$\tau_{r_g F_n}^{-2} B_{n,r_{g-1}^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_{g-1}^\diamond,p} \to_p \left[\begin{array}{cc} h_{6,r_g^\diamond}^{\diamond 2} & 0^{(r_g^\diamond - r_{g-1}^\diamond)\times(p-r_g^\diamond)} \\ 0^{(p-r_g^\diamond)\times(r_g^\diamond - r_{g-1}^\diamond)} & 0^{(p-r_g^\diamond)\times(p-r_g^\diamond)} \end{array} \right]. \quad (17.23)$$

By (17.21) and $o_{gp} = o_p(1)$, we have wp→1 $\{(n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn} : j = r_{g-1}^\diamond + 1, ..., p\}$ solve $|(I_{p-r_{g-1}^\diamond} + \widehat{A}_{jgn})^{-1}\tau_{r_g F_n}^{-2} B_{n,r_{g-1}^\diamond,p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_{g-1}^\diamond,p} + o_p(1) - \kappa I_{p-r_{g-1}^\diamond}| = 0$. Hence, by (17.23), $\widehat{A}_{jgn} = o_p(1)$ (which holds by the induction assumption), and the same argument as used to establish (17.6) and (17.7), we obtain

$$\widehat{\kappa}_{jn} \to_p \infty \ \forall j = r_{g-1}^\diamond + 1, ..., r_g^\diamond \text{ and } (n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn} \to_p 0 \ \forall j = r_g^\diamond + 1, ..., p. \quad (17.24)$$

Let o_{gp}^* denote an $(r_g^\diamond - r_{g-1}^\diamond) \times (p - r_g^\diamond)$ matrix whose elements in column j for $j = 1, ..., p - r_g^\diamond$ are $o_p(\tau_{(r_g^\diamond + j)F_n}/\tau_{r_g F_n}) + O_p((n^{1/2}\tau_{r_g F_n})^{-1})$. Note that $o_{gp}^* = o_p(1)$.

By (17.22) applied once with $B_{n,r_g^\diamond,p}$ in place of $B_{n,r_{g-1}^\diamond,p}$ as the far-right multiplicand and applied a second time with $B_{n,r_{g-1}^\diamond,r_g^\diamond}$ in place of $B_{n,r_{g-1}^\diamond,p}$ as the far-right multiplicand, we have

$$\begin{aligned} & \varrho_{gn} \\ := \ & \tau_{r_g F_n}^{-2} B_{n,r_{g-1}^\diamond,r_g^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond,p} \\ = \ & \left[\begin{array}{c} 0^{r_{g-1}^\diamond \times (r_g^\diamond - r_{g-1}^\diamond)} \\ Diag\{\tau_{(r_{g-1}^\diamond+1)F_n}, ..., \tau_{r_g^\diamond F_n}\}/\tau_{r_g F_n} \\ 0^{(k-r_g^\diamond)\times(r_g^\diamond - r_{g-1}^\diamond)} \end{array} \right]' C_n'(I_k + o_p(1))C_n \left[\begin{array}{c} 0^{r_g^\diamond \times (p - r_g^\diamond)} \\ Diag\{\tau_{(r_g^\diamond+1)F_n}, ..., \tau_{pF_n}\}/\tau_{r_g F_n} \\ 0^{(k-p)\times(p-r_g^\diamond)} \end{array} \right] \\ & + O_p((n^{1/2}\tau_{r_g F_n})^{-1}) \\ = \ & o_{gp}^*, \end{aligned} \quad (17.25)$$

where $\varrho_{gn} \in R^{(r_g^\diamond - r_{g-1}^\diamond)\times(p-r_g^\diamond)}$, $Diag\{\tau_{(r_{g-1}^\diamond+1)F_n}, ..., \tau_{r_g^\diamond F_n}\}/\tau_{r_g F_n} = h_{6,r_g^\diamond}^\diamond + o(1) = O(1)$ 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 (17.25) equals $0^{(r_g^\diamond - r_{g-1}^\diamond)\times(p-r_g^\diamond)}$, 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 (17.25) equals an o_{gp}^* matrix.

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

$$o_{gp} = \left(\begin{array}{cc} o_{1gp} & o_{2gp} \\ o_{2gp}' & o_{3gp} \end{array} \right) \text{ and } \widehat{A}_{jgn} = \left[\begin{array}{cc} \widehat{A}_{1jgn} & \widehat{A}_{2jgn} \\ \widehat{A}_{2jgn}' & \widehat{A}_{3jgn} \end{array} \right], \quad (17.26)$$

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

$$
\begin{aligned}
\widehat{\xi}_{1jgn}(\kappa) &:= \tau_{rg}^{-2} B_{n,r_{g-1}^\diamond, r_g^\diamond}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_{g-1}^\diamond, r_g^\diamond} + o_{1gp} - \kappa(I_{r_g^\diamond - r_{g-1}^\diamond} + \widehat{A}_{1jgn}),\\
\widehat{\xi}_{2jgn}(\kappa) &:= \varrho_{gn} + o_{2gp} - \kappa \widehat{A}_{2jgn}, \text{ and} \qquad (17.27)\\
\widehat{\xi}_{3jgn}(\kappa) &:= \tau_{r_g F_n}^{-2} B_{n,r_g^\diamond, p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond, p} + o_{3gp} - \kappa(I_{p-r_g^\diamond} + \widehat{A}_{3jgn}),
\end{aligned}
$$

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

$$
\begin{aligned}
0 &= |\tau_{r_g F_n}^{-2} B_{n,r_{g-1}^\diamond, p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_{g-1}^\diamond, p} + o_{gp} - \kappa(I_{p-r_{g-1}^\diamond} + \widehat{A}_{jgn})|\\
&= |\widehat{\xi}_{1jgn}(\kappa)| \cdot |\widehat{\xi}_{3jgn}(\kappa) - \widehat{\xi}_{2jgn}(\kappa)' \widehat{\xi}_{1jgn}^{-1}(\kappa) \widehat{\xi}_{2jgn}(\kappa)|\\
&= |\widehat{\xi}_{1jgn}(\kappa)| \cdot |\tau_{r_g F_n}^{-2} B_{n,r_g^\diamond, p}' U_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond, p} + o_{3gp} - (\varrho_{gn} + o_{2gp})' \widehat{\xi}_{1jgn}^{-1}(\kappa)(\varrho_{gn} + o_{2gp})\\
&\quad -\kappa[I_{p-r_g^\diamond} + \widehat{A}_{3jgn} - \widehat{A}_{2jgn}' \widehat{\xi}_{1jgn}^{-1}(\kappa)(\varrho_{gn} + o_{2gp}) - (\varrho_{gn} + o_{2gp})' \widehat{\xi}_{1jgn}^{-1}(\kappa) \widehat{A}_{2jgn}\\
&\qquad +\kappa \widehat{A}_{2jgn}' \widehat{\xi}_{1jgn}^{-1}(\kappa) \widehat{A}_{2jgn}]|, \qquad (17.28)
\end{aligned}
$$

where the second equality holds by the same argument as for (17.10) and uses the result given in (17.29) below which shows that $\widehat{\xi}_{1jgn}(\kappa)$ is nonsingular wp→1 when κ equals $(n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn}$ for $j = r_g^\diamond + 1, ..., p$.

Now we show that, for $j = r_g^\diamond + 1, ..., p$, $(n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn}$ cannot solve the determinantal equation $|\widehat{\xi}_{1jgn}(\kappa)| = 0$ for n large, where this determinant is the first multiplicand on the rhs of (17.28) and, hence, it must solve the determinantal equation based on the second multiplicand on the rhs of (17.28). For $j = r_g^\diamond + 1, ..., p$, we have

$$
\widetilde{\xi}_{1jgn} := \widehat{\xi}_{1jgn}((n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn}) = h_{6,r_g^\diamond}^{\diamond 2} + o_p(1), \qquad (17.29)
$$

by the same argument as in (17.11), using $o_{1gp} = o_p(1)$ and $\widehat{A}_{1jgn} = o_p(1)$ (which holds by the definition of $\widehat{A}_{1jgn}$ following (17.21)). Equation (17.29) 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 $(n^{1/2}\tau_{r_g F_n})^{-2}\widehat{\kappa}_{jn}$ into the second multiplicand on the rhs of (17.28) gives

$$
\begin{aligned}
0 &= |\tau_{r_g F_n}^{-2} B'_{n,r_g^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond,p} + o_{3gp} - (\varrho_{gn} + o_{2gp})' \widetilde{\xi}_{1jgn}^{-1} (\varrho_{gn} + o_{2gp}) \\
&\quad -(n^{1/2}\tau_{r_g F_n})^{-2} \widehat{\kappa}_{jn} (I_{p-r_g^\diamond} + \widehat{A}_{j(g+1)n})|, \text{ where} \\
\widehat{A}_{j(g+1)n} &:= \widehat{A}_{3jgn} - \widehat{A}'_{2jgn} \widetilde{\xi}_{1jgn}^{-1} (\varrho_{gn} + o_{2gp}) - (\varrho_{gn} + o_{2gp})' \widetilde{\xi}_{1jgn}^{-1} \widehat{A}_{2jgn} \\
&\quad + (n^{1/2}\tau_{r_g F_n})^{-2} \widehat{\kappa}_{jn} \widehat{A}'_{2jgn} \widetilde{\xi}_{1jgn}^{-1} \widehat{A}_{2jgn} \qquad (17.30)
\end{aligned}
$$

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 (17.30) satisfy

$$
\begin{aligned}
& o_{3gp} - (\varrho_{gn} + o_{2gp})' \widetilde{\xi}_{1jgn}^{-1} (\varrho_{gn} + o_{2gp}) = o_{3gp} - (o_{gp}^* + o_{2gp})' (h_{6,r_g^\diamond}^{\diamond -2} + o_p(1))(o_{gp}^* + o_{2gp}) \\
&= o_{3gp} - o_{gp}^{*\prime} o_{gp}^* = (\tau_{r_{g+1}F_n}^2 / \tau_{r_g F_n}^2) o_{(g+1)p}, \qquad (17.31)
\end{aligned}
$$

where (i) the first equality uses (17.25) and (17.29), (ii) the second equality uses $o_{2gp} = o_{gp}^*$ (which holds because the (j,m) element of o_{2gp} for $j = 1, ..., r_g^\diamond - r_{g-1}^\diamond$ and $m = 1, ..., p - r_g^\diamond$ is $o_p(\tau_{(r_{g-1}^\diamond + j)F_n} \times \tau_{(r_g^\diamond + m)F_n} / \tau_{r_g F_n}^2) + O_p((n^{1/2}\tau_{r_g F_n})^{-1}) = o_p(\tau_{(r_g^\diamond + m)F_n} / \tau_{r_g F_n}) + O_p((n^{1/2}\tau_{r_g F_n})^{-1})$ since $r_{g-1}^\diamond + j \geq r_g$) and $(h_{6,r_g^\diamond}^{\diamond -2} + o_p(1)) o_{gp}^* = o_{gp}^*$ (which holds because $h_{6,r_g^\diamond}^\diamond$ is diagonal and $\lambda_{\min}(h_{6,r_g^\diamond}^{\diamond 2}) > 0$), (iii) the last equality uses the fact that the (j,m) element of $(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) o_{gp}^{*\prime} o_{gp}^*$ for $j, m = 1, ..., p - r_g^\diamond$ is the sum of a term that is $o_p(\tau_{(r_g^\diamond + j)F_n} \tau_{(r_g^\diamond + m)F_n} / \tau_{r_g F_n}^2)(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) = o_p(\tau_{(r_g^\diamond + j)F_n} \tau_{(r_g^\diamond + m)F_n} / \tau_{r_{g+1}F_n}^2)$ and a term that is $O_p((n^{1/2}\tau_{r_g F_n})^{-2})(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) = O_p((n^{1/2}\tau_{r_{g+1}F_n})^{-2})$ and, hence, $(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) o_{gp}^{*\prime} o_{gp}^*$ 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 $(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) o_{3gp}$ for $j, m = 1, ..., p - r_g^\diamond$ is $o_p(\tau_{(r_g^\diamond + j)F_n} \tau_{(r_g^\diamond + m)F_n} / \tau_{r_g F_n}^2)(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) + O_p((n^{1/2}\tau_{r_g F_n})^{-1})(\tau_{r_g F_n}^2 / \tau_{r_{g+1}F_n}^2) = o_p(\tau_{(r_g^\diamond + j)F_n} \tau_{(r_g^\diamond + m)F_n} / \tau_{r_{g+1}F_n}^2) + O_p((n^{1/2}\tau_{r_{g+1}F_n})^{-1})(\tau_{r_g F_n} / \tau_{r_{g+1}F_n})$, which again is the same order as the (j,m) element of $o_{(g+1)p}$ (using $\tau_{r_g F_n} / \tau_{r_{g+1}F_n} \leq 1$).

The calculations in (17.31) are a key part of the induction proof. The definitions of the terms o_{gp} and o_{gp}^* (given preceding (17.21) and (17.25), respectively) are chosen so that the results in (17.31) hold.

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

$$
\widehat{A}_{j(g+1)n} = o_p(1), \qquad (17.32)
$$

because $\widehat{A}_{2jgn} = o_p(1)$ and $\widehat{A}_{3jgn} = o_p(1)$ by (17.21), $\widetilde{\xi}_{1jgn}^{-1} = O_p(1)$ (by (17.29)), $\varrho_{gn} + o_{2gp} = o_p(1)$ (by (17.25) since $o_{gp}^* = o_p(1)$), and $(n^{1/2}\tau_{r_g F_n})^{-2} \widehat{\kappa}_{jn} = o_p(1)$ (by (17.24)).

Inserting (17.31) and (17.32) into (17.30) and multiplying by $\tau^2_{r_gF_n}/\tau^2_{r_{g+1}F_n}$ gives

$$0 = |\tau^{-2}_{r_{g+1}F_n} B'_{n,r_g^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond,p} + o_{(g+1)p} - (n^{1/2}\tau_{r_{g+1}F_n})^{-2} \widehat{\kappa}_{jn} (I_{p-r_g^\diamond} + \widehat{A}_{j(g+1)n})|. \tag{17.33}$$

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

$$0 = |\tau^{-2}_{r_{g+1}F_n} B'_{n,r_g^\diamond,p} U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_{n,r_g^\diamond,p} + o_{(g+1)p} - \kappa (I_{p-r_g^\diamond} + \widehat{A}_{j(g+1)n})|. \tag{17.34}$$

This establishes the induction step and concludes the proof that (17.21) holds for all $g = 1, ..., G_h$.

Finally, given that (17.21) holds for all $g = 1, ..., G_h$, (17.24) gives the results stated in (17.18) and (17.18) gives the results stated in the Lemma by the argument in (17.18)-(17.20). □

Now we use the approach in Johansen (1991, pp. 1569-1571) and Robin and Smith (2000, pp. 172-173) to prove Theorem 10.4. 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 17.1(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 10.4. Theorem 10.4(a) and (c) follow immediately from Lemma 17.1(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→1), they are the solutions to

$$|Q_n^\diamond(\kappa)| = 0, \text{ where } Q_n^\diamond(\kappa) := nS_n B'_n U'_n \widehat{D}'_n \widehat{W}'_n \widehat{W}_n \widehat{D}_n U_n B_n S_n - \kappa S'_n B'_n U'_n \widehat{U}_n^{-1\prime} \widehat{U}_n^{-1} U_n B_n S_n, \tag{17.35}$$

because $|S_n| > 0$, $|B_n| > 0$, $|U_n| > 0$, and $|\widehat{U}_n| > 0$ wp→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→1. (For simplicity, we omit the qualifier wp→1 that applies to several statements below.)

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

$$B_n S_n = (B_{n,q} S_{n,q}, B_{n,p-q}), \text{ where}$$
$$S_{n,q} := Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}\} \in R^{q \times q}. \tag{17.36}$$

The convergence result of Lemma 10.3 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

$$n^{1/2}W_n\widehat{D}_nU_nB_{n,q}S_{n,q} \to_p \overline{\Delta}_{h,q} := h_{3,q} \text{ and } n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} \to_d \overline{\Delta}_{h,p-q}, \tag{17.37}$$

where $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are defined in (10.17).

We have

$$\widehat{W}_nW_n^{-1} \to_p I_k \text{ and } \widehat{U}_nU_n^{-1} \to_p I_p \tag{17.38}$$

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 (17.35)-(17.38), we have

$$\begin{aligned} Q_n^{\diamond}(\kappa) = & \begin{bmatrix} I_q + o_p(1) & h_{3,q}'n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} + o_p(1) \\ n^{1/2}B_{n,p-q}'U_n'\widehat{D}_n'W_n'h_{3,q} + o_p(1) & n^{1/2}B_{n,p-q}'U_n'\widehat{D}_n'W_n'W_nn^{1/2}\widehat{D}_nU_nB_{n,p-q} + o_p(1) \end{bmatrix} \\ & -\kappa \begin{bmatrix} S_{n,q}^2 & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & I_{p-q} \end{bmatrix} - \kappa \begin{bmatrix} S_{n,q}A_{1n}S_{n,q} & S_{n,q}A_{2n} \\ A_{2n}'S_{n,q} & A_{3n} \end{bmatrix}, \text{ where} \\ \widehat{A}_n = & \begin{bmatrix} A_{1n} & A_{2n} \\ A_{2n}' & A_{3n} \end{bmatrix} := B_n'U_n'\widehat{U}_n^{-1\prime}\widehat{U}_n^{-1}U_nB_n - I_p = o_p(1) \text{ for } A_{1n} \in R^{q\times q},\ A_{2n} \in R^{q\times(p-q)}, \end{aligned} \tag{17.39}$$

and $A_{3n} \in R^{(p-q)\times(p-q)}$, $\widehat{A}_n$ is defined in (17.39) just as in (17.5), 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 (10.7), (10.9), (10.12), and (10.17)). Note that A_{jn} and $\widehat{A}_{jn}$ (defined in (17.2)) 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

$$\begin{aligned} & nB_n'\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_nB_n \\ = \ & nB_n'(U_n^{-1}\widehat{U}_n)'B_n^{-1\prime}B_n'U_n'\widehat{D}_n'W_n'\left(\widehat{W}_nW_n^{-1}\right)'\left(\widehat{W}_nW_n^{-1}\right)(W_n\widehat{D}_nU_nB_n)B_n^{-1}(U_n^{-1}\widehat{U}_n)B_n \\ \to_d \ & \overline{\Delta}_{h,p-q}'\overline{\Delta}_{h,p-q}, \end{aligned} \tag{17.40}$$

where the convergence holds by (17.37) and (17.38) and $\overline{\Delta}_{h,p-q}$ is defined as in (10.17) 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 $nB_n'\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_nB_n$ 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 $h_{3,k-q}h_{3,k-q}' = h_3h_3' = I_k$ when $q = 0$), which proves part (b) of Theorem 10.4 when $q = 0$.

In the remainder of the proof of part (b), we assume $1 \leq q < p$, which is the remaining case to be considered in the proof of part (b). The formula for the determinant of a partitioned matrix and (17.39) give

$$\begin{aligned}
|Q_n^\diamond(\kappa)| &= |Q_{1n}^\diamond(\kappa)| \cdot |Q_{2n}^\diamond(\kappa)|, \text{ where} \\
Q_{1n}^\diamond(\kappa) &:= I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n} S_{n,q}, \\
Q_{2n}^\diamond(\kappa) &:= n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' W_n n^{1/2} \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \kappa I_{p-q} - \kappa A_{3n} \\
&\quad -[n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' h_{3,q} + o_p(1) - \kappa A_{2n}' S_{n,q}](I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n} S_{n,q})^{-1} \\
&\quad \times [h_{3,q}' n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) - \kappa S_{n,q} A_{2n}], \qquad (17.41)
\end{aligned}$$

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

By Lemma 17.1(b) (which applies for $1 \leq q < p$), for $j = q+1, ..., p$, we have $\widehat{\kappa}_{jn} S_{n,q}^2 = o_p(1)$ and $\widehat{\kappa}_{jn} S_{n,q} A_{1n} S_{n,q} = o_p(1)$. Thus,

$$Q_{1n}^\diamond(\widehat{\kappa}_{jn}) = I_q + o_p(1) - \widehat{\kappa}_{jn} S_{n,q}^2 - \widehat{\kappa}_{jn} S_{n,q} A_{1n} S_{n,q} = I_q + o_p(1). \qquad (17.42)$$

By (17.35) and (17.41), $|Q_n^\diamond(\widehat{\kappa}_{jn})| = |Q_{1n}^\diamond(\widehat{\kappa}_{jn})| \cdot |Q_{2n}^\diamond(\widehat{\kappa}_{jn})| = 0$ for $j = 1, ..., p$. By (17.42), $|Q_{1n}^\diamond(\widehat{\kappa}_{jn})| \neq 0$ for $j = q+1, ..., p$ wp→1. Hence, wp→1,

$$|Q_{2n}^\diamond(\widehat{\kappa}_{jn})| = 0 \text{ for } j = q+1, ..., p. \qquad (17.43)$$

Now we plug in $\widehat{\kappa}_{jn}$ for $j = q+1, ..., p$ into $Q_{2n}^\diamond(\kappa)$ in (17.41) and use (17.42). We have

$$\begin{aligned}
Q_{2n}^\diamond(\widehat{\kappa}_{jn}) &= n B_{n,p-q}' U_n' \widehat{D}_n' W_n' W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1) \\
&\quad -[n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' h_{3,q} + o_p(1)](I_q + o_p(1))[h_{3,q}' n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1)] \\
&\quad -\widehat{\kappa}_{jn}[I_{p-q} + A_{3n} - (n^{1/2} B_{n,p-q}' U_n' \widehat{D}_n' W_n' h_{3,q} + o_p(1))(I_q + o_p(1)) S_{n,q} A_{2n} \\
&\quad -A_{2n}' S_{n,q}(I_q + o_p(1))(h_{3,q}' n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} + o_p(1)) \\
&\quad +\widehat{\kappa}_{jn} A_{2n}' S_{n,q}(I_q + o_p(1)) S_{n,q} A_{2n}]. \qquad (17.44)
\end{aligned}$$

The term in square brackets on the last three lines of (17.44) that multiplies $\widehat{\kappa}_{jn}$ equals

$$I_{p-q} + o_p(1), \qquad (17.45)$$

because $A_{3n} = o_p(1)$ (by (17.39)), $n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} = O_p(1)$ (by (17.37)), $S_{n,q} = o(1)$ (by the definitions of q and $S_{n,q}$ in (10.16) and (17.36), respectively, and $h_{1,j} := \lim n^{1/2} \tau_{jF_n}$), $A_{2n} = o_p(1)$

(by (17.39)), 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 (17.44) and (17.45) give

$$\begin{aligned} Q^{\diamond}_{2n}(\widehat{\kappa}_{jn}) &= n^{1/2}B'_{n,p-q}U'_n\widehat{D}'_nW'_n[I_k - h_{3,q}h'_{3,q}]n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} + o_p(1) - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)] \\ &= n^{1/2}B'_{n,p-q}U'_n\widehat{D}'_nW'_nh_{3,k-q}h'_{3,k-q}n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q} + o_p(1) - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)] \\ &:= M_{n,p-q} - \widehat{\kappa}_{jn}[I_{p-q} + o_p(1)], \end{aligned} \quad (17.46)$$

where the second equality uses $I_k = h_3h'_3 = h_{3,q}h'_{3,q} + h_{3,k-q}h'_{3,k-q}$ (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 (17.43) and (17.46) imply that $\{\widehat{\kappa}_{jn} : j = q+1, ..., p\}$ are the $p-q$ eigenvalues of the matrix

$$M^{\diamond}_{n,p-q} := [I_{p-q} + o_p(1)]^{-1/2}M_{n,p-q}[I_{p-q} + o_p(1)]^{-1/2} \quad (17.47)$$

by pre- and post-multiplying the quantities in (17.46) by the rhs quantity $[I_{p-q} + o_p(1)]^{-1/2}$ in (17.46). By (17.37),

$$M^{\diamond}_{n,p-q} \rightarrow_d \overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q}. \quad (17.48)$$

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 (17.48), the matrix $M^{\diamond}_{n,p-q}$ converges in distribution. In consequence, by the CMT, the vector of eigenvalues of $M^{\diamond}_{n,p-q}$, 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 10.4. 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 10.4.

The convergence in parts (a)-(d) of Theorem 10.4 is joint with that in Lemma 10.3 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 10.4.

Part (f) of Theorem 10.4 holds by the same proof as used for parts (a)-(e) with n replaced by w_n. □

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

In this section, we show that the conditions in (3.10) and (3.11) 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 (3.10) is sufficient by the following argument:

$$
\begin{aligned}
& \lambda_{p-j}\left(\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right) \\
\geq\ & \lambda_{p-j}\left(\Psi_F^{\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right) \\
=\ & \lambda_{\min}\left((\xi' \otimes I_{p-j})\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}(\xi \otimes I_{p-j})\right) \\
=\ & \min_{\lambda \in R^{p-j}:||\lambda||=1}\left(\frac{(\xi \otimes I_{p-j})\lambda}{||(\xi \otimes I_{p-j})\lambda||}\right)'\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\frac{(\xi \otimes I_{p-j})\lambda}{||(\xi \otimes I_{p-j})\lambda||} \times ||(\xi \otimes I_{p-j})\lambda||^2 \\
\geq\ & \min_{\eta \in R^{(p-j)^2}:||\eta||=1}\eta'\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\eta \times \min_{\lambda \in R^{p-j}:||\lambda||=1}||(\xi \otimes I_{p-j})\lambda||^2 \\
=\ & \lambda_{\min}\left(\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\right),
\end{aligned}
\tag{18.1}
$$

where the first inequality holds by Corollary 16.4(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 (3.10) is sufficient by sufficient condition (i) in (3.10) and the following:

$$
\begin{aligned}
& \lambda_{\min}\left(\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\right) \\
=\ & \min_{\eta \in R^{(p-j)^2}:||\eta||=1}\left(\frac{(I_{p-j} \otimes \overline{C}_{F,p-j})\eta}{||(I_{p-j} \otimes \overline{C}_{F,p-j})\eta||}\right)'\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\frac{(I_{p-j} \otimes \overline{C}_{F,p-j})\eta}{||(I_{p-j} \otimes \overline{C}_{F,p-j})\eta||} \\
& \times||(I_{p-j} \otimes \overline{C}_{F,p-j})\eta||^2 \\
\geq\ & \min_{\zeta \in R^{(p-j)k}:||\zeta||=1}\zeta'\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\zeta \times \min_{\eta \in R^{(p-j)^2}:||\eta||=1}||(I_{p-j} \otimes \overline{C}_{F,p-j})\eta||^2 \\
=\ & \lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\right),
\end{aligned}
\tag{18.2}
$$

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 (3.10) is sufficient by sufficient condition (ii) in (3.10) and a similar argument to that given in (18.2) 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 (3.10) is sufficient by sufficient condition (iii) in (3.10) and a similar argument to that given in (18.2) 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:

$$\begin{aligned}
\phi'(I_p \otimes \Omega_F^{-1})\phi &= \sum_{j=1}^{p} (\phi_j/||\phi_j||)'\Omega_F^{-1}(\phi_j/||\phi_j||) \times ||\phi_j||^2 \\
&\geq \sum_{j=1}^{p} \lambda_{\min}(\Omega_F^{-1}) \times ||\phi_j||^2 = 1/\lambda_{\max}(\Omega_F) \geq M^{-2/(2+\gamma)}, \qquad (18.3)
\end{aligned}$$

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 (3.10) are sufficient by the following argument. Write

$$\Psi_F^{vec(G_i)} = (M_F, I_{pk})\Sigma_F^{f_i}(M_F, I_{pk})', \text{ where } M_F = -(E_F vec(G_i)g_i')(E_F g_i g_i')^{-1} \in R^{pk \times k}. \qquad (18.4)$$

We have

$$\begin{aligned}
\lambda_{\min}(\Psi_F^{vec(G_i)}) &= \min_{\lambda \in R^{pk}:||\lambda||=1} \lambda'(M_F, I_{pk})\Sigma_F^{f_i}(M_F, I_{pk})'\lambda \\
&= \min_{\lambda \in R^{pk}:||\lambda||=1} \left(\frac{(M_F, I_{pk})'\lambda}{||(M_F, I_{pk})'\lambda||}\right)' \Sigma_F^{f_i} \left(\frac{(M_F, I_{pk})'\lambda}{||(M_F, I_{pk})'\lambda||}\right) \times ||(M_F, I_{pk})'\lambda||^2 \\
&\geq \min_{\mu \in R^{(p+1)k}:||\mu||=1} \mu'\Sigma_F^{f_i}\mu \\
&= \lambda_{\min}(\Sigma_F^{f_i}), \qquad (18.5)
\end{aligned}$$

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 (3.10). 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 (3.10).

The condition in (3.11) is sufficient by the following argument:

$$\lambda_{p-j}\left(\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right) \geq \lambda_p\left(\Psi_F^{C'_F\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right) = \lambda_p\left(\Psi_F^{\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right), \qquad (18.6)$$

where the first inequality holds by Corollary 16.4(b) with $m = p$ and $r = j$ and the equality holds because $\Psi_F^{C'_F\Omega_F^{-1/2}G_iB_{F,p-j}\xi} = C'_F\Psi_F^{\Omega_F^{-1/2}G_iB_{F,p-j}\xi}C_F$ and C_F is orthogonal.

19 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 19.1, we provide an example that illustrates the results in Theorem 5.1 and Comment (v) to Theorem 5.1. In Section 19.2, we establish the asymptotic size of the test based on $\widetilde{V}_{Dn}$ defined as in (5.3). In Section 19.3, we report some simulation results for a linear instrumental variable (IV) model with two rhs endogenous variables. In Section 19.4, 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 19.5, 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 19.5 are a relatively simple by-product of the results in Section 19.4.

Proofs of the results stated in this section are given in Section 19.6.

Lemma 5.2 is proved in Section 19.7.

Theorem 5.1 follows from Lemma 19.2 and Theorem 19.3, which are stated in Section 19.4.

As stated in footnote 4 in Section 2 of AG1, "under sequences F_n such that $n^{1/2}E_{F_n}G(W_i,\theta)$ converges to a finite matrix, $n^{1/2}\widehat{D}_n(\theta)$ and $n^{1/2}\widehat{g}_n(\theta)$ $(= n^{-1/2}\sum_{i=1}^n g(W_i,\theta))$ are asymptotically independent (see Lemmas 10.2 and 10.3 in Section 10 in this SM). Therefore, if $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ is a continuous function of $n^{1/2}\widehat{D}_n(\theta)$ and a weighting matrix $\widehat{V}_n$ (that converges in probability to a positive definite matrix), then by the continuous mapping theorem (CMT),

$n^{1/2}\widehat{g}_n(\theta)$ and $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ are also asymptotically independent."

Footnote 4 of AG1 also states "however, under sequences for which a component of $n^{1/2}E_{F_n}G(W_i,\theta)$ diverges to plus or minus infinity, the CMT cannot be applied because $n^{1/2}\widehat{D}_n(\theta)$ does not converge in distribution, but rather, some component of it diverges to plus or minus infinity in probability (see Lemma 10.3 in Section 10 in this SM when $h_{1,j} = \infty$ for some $j \leq p$). In this case, $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ may not have an asymptotic distribution, and if it does, $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ and $n^{1/2}\widehat{g}_n(\theta)$ are not necessarily asymptotically independent."

The following is a simple example of the latter situation when $p = 2$. Let $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta)) = \widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)||$, where $\widehat{V}_{12n}$ is the (1, 2) component of $\widehat{V}_n$ and $\widehat{D}_{1n}(\theta)$ is the first column of $\widehat{D}_n(\theta)$. Assume $\widehat{V}_n - V \to_p 0$ for some matrix V and $n^{1/2}(\widehat{V}_n - V) \to_d \xi$, where ξ is a mean zero normal random matrix. Assume that under F_n the first column $E_{F_n}G_1(W_i,\theta)$ of $E_{F_n}G(W_i,\theta)$ is a fixed nonzero vector, G_1^e say. Assume that the $(1,2)$ element of V, denoted by V_{12}, equals zero under F_n. Then, $\widehat{D}_{1n}(\theta) \to_p G_1^e$ (see Lemma 10.2 in Section 10 in this SM) and $\widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)|| = n^{1/2}(\widehat{V}_{12n} - V_{12})||\widehat{D}_{1n}(\theta)|| \to_d \xi_{12}||G_1^e||$. But, in general there is no reason why ξ_{12} and the random limit of $n^{1/2}\widehat{g}_n(\theta)$ are independent. For simplicity, the previous example is somewhat contrived, because rank statistics typically are not of the form $\widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)||$. But, components of rank statistics may be of this form. Section 19.1, which follows, provides a more concrete example of this type of situation.

19.1 An Example

Here we provide an example that illustrates the result of Theorem 5.1. In this example, the true distribution F does not depend on n. Suppose $p = 2$, $E_FG_i = (1^k, 0^k)$, where $c^k = (c, ..., c)' \in R^k$ for $c = 0, 1$, $n^{1/2}(\widehat{D}_n - E_FG_i) \to_d \overline{D}_h$ 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 $\overline{M}_h \in R^{2k\times 2k}$. (The convergence results $n^{1/2}(\widehat{D}_n - E_FG_i) \to_d \overline{D}_h$ and $n^{1/2}(\widetilde{M}_n - M_F) \to_d \overline{M}_h$ are established in Lemmas 10.2 and 19.2, respectively, in Section 10 and Section 19.4 in this SM under general conditions.) We have

$$\widehat{D}_n^\dagger = vec_{k,p}^{-1}(\widetilde{V}_{Dn}^{-1/2}vec(\widehat{D}_n)) = \left(\widetilde{M}_{11n}\widehat{D}_{1n} + \widetilde{M}_{12n}\widehat{D}_{2n}, \widetilde{M}_{21n}\widehat{D}_{1n} + \widetilde{M}_{22n}\widehat{D}_{2n}\right), \tag{19.1}$$

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

$$\begin{aligned} n^{1/2}\widehat{D}_n^\dagger T_n^\dagger &= \left(\widetilde{M}_{11n}\widehat{D}_{1n} + \widetilde{M}_{12n}\widehat{D}_{2n},\ n^{1/2}\widetilde{M}_{21n}\widehat{D}_{1n} + \widetilde{M}_{22n}n^{1/2}\widehat{D}_{2n}\right) \\ &\to_d \left(I_k 1^k + 0^{k\times k}0^k,\ \overline{M}_{21h}1^k + I_k\overline{D}_{2h}\right) = \left(1^k, \overline{M}_{21h}1^k + \overline{D}_{2h}\right), \end{aligned} \tag{19.2}$$

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 (19.2) 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 (19.2) 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 (19.2) arises quite generally, as shown in Theorem 5.1, whenever $p \geq 2$.[12]

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 (19.2) 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 (19.1). 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 (19.2) equals $0^{k\times k}$, which implies that $\overline{M}_{21h} = 0^{k\times k}$ in the second line of (19.2). 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.

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

$$rk_n = rk_n^\dagger := \lambda_{\min}(n(\widehat{D}_n^\dagger)'\widehat{D}_n^\dagger), \text{ where } \widehat{D}_n^\dagger := vec_{k,p}^{-1}(\widetilde{V}_{Dn}^{-1/2}vec(\widehat{D}_n)) \tag{19.3}$$

[12] When the matrix $M_{21F} \neq 0^{k\times k}$, the argument in (19.2) 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 $E_F G_i$.

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

$$\widehat{\kappa}_{jn}^\dagger \text{ denote the } j\text{th eigenvalue of } n(\widehat{D}_n^\dagger)'\widehat{D}_n^\dagger, \text{ for } j = 1, ..., p, \tag{19.4}$$

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

$$\mathcal{F}_{KCLR} := \{F \in \mathcal{F} : \lambda_{\min}(Var_F((g_i', vec(G_i)')')) \geq \delta_2, E_F||(g_i', vec(G_i)')'||^{4+\gamma} \leq M\}, \tag{19.5}$$

where $\delta_2 > 0$ and $\gamma > 0$ and $M < \infty$ are as in the definition of $\mathcal{F}$ in (3.3). 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 (3.10). 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 (19.3).

For a fixed distribution F, $\widetilde{V}_{Dn}$ estimates $\Phi_F^{vec(G_i)}$ defined in (10.15), where $\Phi_F^{vec(G_i)}$ is pd by its definition in (10.15) and the $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$. More specifically, $\Phi_F^{vec(G_i)}$ is pd because by (10.15) $\Phi_F^{vec(G_i)} := Var_F(vec(G_i) - (E_F vec(G_\ell)g_\ell')\Omega_F^{-1}g_i) = (-(E_F vec(G_\ell)g_\ell')\Omega_F^{-1}, I_{pk})$ $Var_F((g_i', vec(G_i)')')(-(E_F vec(G_\ell)g_\ell')\Omega_F^{-1}, I_{pk})'$, 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

$$M_F = \begin{bmatrix} M_{11F} & \cdots & M_{1pF} \\ \vdots & \ddots & \vdots \\ M_{p1F} & \cdots & M_{ppF} \end{bmatrix} := (\Phi_F^{vec(G_i)})^{-1/2} \text{ and} \tag{19.6}$$

$$D_F^\dagger := \sum_{j=1}^{p} (M_{1jF}E_F G_{ij}, ..., M_{pjF}E_F G_{ij}) \in R^{k \times p}, \text{ where } G_i = (G_{i1}, ..., G_{ip}) \in R^{k \times p}.$$

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

$$\begin{aligned} &B^\dagger_F \in R^{p \times p} \text{ to be an orthogonal matrix of eigenvectors of } D^{\dagger\prime}_F D^\dagger_F \text{ and} \\ &C^\dagger_F \in R^{k \times k} \text{ to be an orthogonal matrix of eigenvectors of } D^\dagger_F D^{\dagger\prime}_F \end{aligned} \tag{19.7}$$

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 (19.7) gives definitions of B_F and C_F that are similar to the definitions in (10.6) and (10.7), 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 (10.9) 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

$$\lambda_{10,F} = Var_F \begin{pmatrix} f^*_i \\ vech\,(f^*_i f^{*\prime}_i) \end{pmatrix} \in R^{d^* \times d^*}, \tag{19.8}$$

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 (10.9) 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

$$\begin{aligned} \lambda &= \lambda_F := (\lambda_{1,F}, ..., \lambda_{10,F}, \lambda^\dagger_{1,F}, \lambda^\dagger_{2,F}, \lambda^\dagger_{3,F}, \lambda^\dagger_{6,F}), \\ \Lambda_{KCLR} &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{10,F}, \lambda^\dagger_{1,F}, \lambda^\dagger_{2,F}, \lambda^\dagger_{3,F}, \lambda^\dagger_{6,F}) \text{ for some } F \in \mathcal{F}_{KCLR}\}, \text{ and} \\ h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}, \lambda_{10,F}, n^{1/2}\lambda^\dagger_{1,F}, \lambda^\dagger_{2,F}, \lambda^\dagger_{3,F}, \lambda^\dagger_{6,F}). \end{aligned} \tag{19.9}$$

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 (10.1). The asymptotic variance of $n^{1/2} vec(\widehat{D}_n - E_{F_n} G_i)$ is $\Phi^{vec(G_i)}_h$ under $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ by Lemma 10.2.

Define $h_{1,j}$ for $j \leq p$ and h_s for $s = 2, ..., 8$ as in (10.12), $q = q_h$ as in (10.16), $h_{2,q}$, $h_{2,p-q}$, $h_{3,q}$, $h_{3,p-q}$, and $h^\diamond_{1,p-q}$ as in (10.17), and Υ_n, $\Upsilon_{n,q}$, and $\Upsilon_{n,p-q}$ as in (9.2). Note that $h_7 = h^{-1/2}_{5,g}$ and $h_8 = I_p$ due to the definitions of $\lambda_{7,F}$ and $\lambda_{8,F}$ given above, where $h_{5,g}$ $(= \lim E_{F_n} g_i g_i')$ denotes the upper left $k \times k$ submatrix of h_5, as in Section 10.

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

$$h_{10} = \begin{bmatrix} h_{10,f^*} & h_{10,f^* f^{*2}} \\ h_{10,f^{*2} f^*} & h_{10,f^{*2} f^{*2}} \end{bmatrix} := \lim Var_{F_n} \begin{pmatrix} f^*_i \\ vech\,(f^*_i f^{*\prime}_i) \end{pmatrix} \in R^{d^* \times d^*}. \tag{19.10}$$

Note that $h_{10,f^*} \in R^{(p+1)k \times (p+1)k}$ is pd by the definition of $\mathcal{F}_{KCLR}$ in (19.5).

With $\tau_{jF}^\dagger$, $B_F^\dagger$, and $C_F^\dagger$ in place of τ_{jF}, B_F, and C_F, respectively, define $h_{1,j}^\dagger$ for $j \leq p$ and $h_s^\dagger$ for $s = 2, 3, 6$ as in (10.12) as analogues to the quantities without the † superscript, define $q^\dagger = q_h^\dagger$ as in (10.16), define $h_{2,q^\dagger}^\dagger$, $h_{2,p-q^\dagger}^\dagger$, $h_{3,q^\dagger}^\dagger$, $h_{3,k-q^\dagger}^\dagger$, and $h_{1,p-q^\dagger}^{\dagger\diamond}$ as in (10.17), and define $\Upsilon_n^\dagger$, $\Upsilon_{n,q^\dagger}^\dagger$, and $\Upsilon_{n,p-q^\dagger}^\dagger$ as in (9.2). The quantity $q^\dagger$ determines the asymptotic behavior of $rk_n^\dagger$. By definition, $q^\dagger$ is the largest value j ($\leq p$) for which $\lim n^{1/2}\tau_{jF_n}^\dagger = \infty$ under $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$. It is shown below that if $q^\dagger = p$, then $rk_n^\dagger \rightarrow_p \infty$, whereas if $q^\dagger < p$, then $rk_n^\dagger$ converges in distribution to a nondegenerate random variable, see Lemma 19.4.

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

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} f_i^* \\ vech\,(f_i^* f_i^{*\prime} - E_{F_n} f_i^* f_i^{*\prime}) \end{pmatrix} \rightarrow_d \overline{L}_h \sim N(0^{d^*}, h_{10}), \text{ where} \\
& \overline{L}_h = (\overline{L}_{h,1}^\prime, \overline{L}_{h,2}^\prime, \overline{L}_{h,3}^\prime)^\prime \text{ for } \overline{L}_{h,1} \in R^k,\ \overline{L}_{h,2} \in R^{kp},\ \text{and } \overline{L}_{h,3} \in R^{(p+1)k((p+1)k+1)/2}
\end{aligned} \tag{19.11}
$$

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 $h_5 := \lim E_{F_n}(g_i^\prime, vec(G_i)^\prime)^\prime (g_i^\prime, vec(G_i)^\prime)$, we have

$$
h_{10,f^*} = \begin{bmatrix} h_{5,g} & h_{5,gG} \\ h_{5,Gg} & h_{5,G} - vec(h_4)vec(h_4)^\prime \end{bmatrix}, \text{ where } h_5 = \begin{bmatrix} h_{5,g} & h_{5,gG} \\ h_{5,Gg} & h_{5,G} \end{bmatrix} \tag{19.12}
$$

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 $\overline{D}_h$:

$$
\overline{g}_h := \overline{L}_{h,1} \text{ and } vec(\overline{D}_h) := \overline{L}_{h,2} - h_{5,Gg} h_{5,g}^{-1} \overline{L}_{h,1}. \tag{19.13}
$$

These definitions are distributionally equivalent to the previous definitions of $\overline{g}_h$ and $\overline{D}_h$ given in Lemma 10.2, 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 $\Phi_h^{vec(G_i)}$ ($= h_{5,G} - vec(h_4)vec(h_4)^\prime - h_{5,Gg} h_{5,g}^{-1} h_{5,Gg}^\prime$), respectively, where $\Phi_h^{vec(G_i)}$ is defined in (10.15) and is pd (because $\Phi_h^{vec(G_i)} = \lim \Phi_{F_n}^{vec(G_i)}$ and $\lambda_{\min}(\Phi_{F_n}^{vec(G_i)})$ is bounded away from zero by its definition in (10.15) and the $\lambda_{\min}(\cdot)$ condition in $\mathcal{F}_{KCLR}$).

Define

$$\overline{D}_h^{\dagger} := \sum_{j=1}^{p}(M_{1jh}\overline{D}_{jh}, ..., M_{pjh}\overline{D}_{jh}) \in R^{k\times p}, \text{ where } \begin{bmatrix} M_{11h} & \cdots & M_{1ph} \\ \vdots & \ddots & \vdots \\ M_{p1h} & \cdots & M_{pph} \end{bmatrix} := (\Phi_h^{vec(G_i)})^{-1/2}, \tag{19.14}$$

$\overline{D}_h = (\overline{D}_{1h}, ..., \overline{D}_{ph})$, and $\overline{D}_h$ is defined in (19.13). Define

$$\begin{aligned} \overline{\Delta}_h^{\dagger} &= (\overline{\Delta}_{h,q^{\dagger}}^{\dagger}, \overline{\Delta}_{h,p-q^{\dagger}}^{\dagger}) \in R^{k\times p},\ \overline{\Delta}_{h,q^{\dagger}}^{\dagger} := h_{3,q^{\dagger}}^{\dagger} \in R^{k\times q^{\dagger}}, \text{ and} \\ \overline{\Delta}_{h,p-q^{\dagger}}^{\dagger} &:= h_3^{\dagger} h_{1,p-q^{\dagger}}^{\dagger\diamond} + \overline{D}_h^{\dagger} h_{2,p-q^{\dagger}}^{\dagger} \in R^{k\times(p-q^{\dagger})}. \end{aligned} \tag{19.15}$$

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

$$n^{-1}\sum_{i=1}^{n}\begin{pmatrix} f_i^* \\ vech\,(f_i^* f_i^{*\prime}) \end{pmatrix} \text{ into} \tag{19.16}$$

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

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

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. (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.) We define $\overline{A}_h$ as follows:

$$\begin{aligned} &\overline{A}_h \text{ denotes the } (kp)(kp+1)/2 \times d^* \text{ matrix of partial derivatives of } a(\cdot) \\ &\text{evaluated at } (0^{(p+1)k\prime}, vech(h_{10,f^*})')', \end{aligned} \tag{19.17}$$

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

$$\overline{M}_h := vech_{kp,kp}^{-1}(\overline{A}_h\overline{L}_h) \in R^{kp\times kp}, \tag{19.18}$$

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

Define

$$\overline{M}_h^{\dagger} := (\overline{M}_{h,q^{\dagger}}^{\dagger}, \overline{M}_{h,p-q^{\dagger}}^{\dagger}) := (0^{k\times q^{\dagger}}, \overline{M}_{h,p-q^{\dagger}}^{\dagger}) \in R^{k\times p}, \text{ where} \tag{19.19}$$
$$\overline{M}_{h,p-q^{\dagger}}^{\dagger} := \sum_{j=1}^{p}(\overline{M}_{1jh}h_{4,j}, ..., \overline{M}_{pjh}h_{4,j})h_{2,p-q^{\dagger}}^{\dagger} \in R^{k\times(p-q^{\dagger})},\ \overline{M}_h = \begin{bmatrix} \overline{M}_{11h} & \cdots & \overline{M}_{1ph} \\ \vdots & \ddots & \vdots \\ \overline{M}_{p1h} & \cdots & \overline{M}_{pph} \end{bmatrix},$$

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

Below (in Lemma 19.4), 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

$$r_h(\overline{D}_h, \overline{M}_h) := \lambda_{\min}((\overline{\Delta}_{h,p-q^{\dagger}}^{\dagger} + \overline{M}_{h,p-q^{\dagger}}^{\dagger})' h_{3,k-q^{\dagger}}^{\dagger} h_{3,k-q^{\dagger}}^{\dagger\prime}(\overline{\Delta}_{h,p-q^{\dagger}}^{\dagger} + \overline{M}_{h,p-q^{\dagger}}^{\dagger})), \tag{19.20}$$

where $\overline{\Delta}_{h,p-q^{\dagger}}^{\dagger}$ is a nonrandom function of $\overline{D}_h$ by (19.14) and (19.15) and $\overline{M}_{h,p-q^{\dagger}}^{\dagger}$ is a nonrandom function of $\overline{M}_h$ by (19.19). 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 (10.17), as follows:

$$\overline{\Delta}_h = (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q},\ \text{and}\ \overline{\Delta}_{h,p-q} := h_3 h_{1,p-q}^{\diamond} + h_7\overline{D}_h h_8 h_{2,p-q},\ \text{where}$$
$$h_2 = (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}),\ h_{1,p-q}^{\diamond} := \begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times(p-q)}. \tag{19.21}$$

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. (The CLR_n statistic also depends on $\widehat{D}_n$ through the rank statistic.) This means that Assumption WU for the parameter space Λ_{KCLR} (defined in Section 10.4) 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 19.5), we show that the asymptotic distribution of the CLR_n statistic under sequences $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$ with $q^{\dagger} < p$ is given by

$$\overline{CLR}_h := \frac{1}{2}\left(\overline{LM}_h + \overline{J}_h - \overline{r}_h + \sqrt{(\overline{LM}_h + \overline{J}_h - \overline{r}_h)^2 + 4\overline{LM}\overline{r}_h}\right),\ \text{where}$$
$$\overline{LM}_h := \overline{v}_h'\overline{v}_h \sim \chi_p^2,\ \overline{v}_h := P_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h,\ \overline{J}_h := \overline{g}_h' h_{5,g}^{-1/2} M_{\overline{\Delta}_h} h_{5,g}^{-1/2}\overline{g}_h \sim \chi_{k-p}^2,\ \text{and}$$
$$\overline{r}_h := r_h(\overline{D}_h, \overline{M}_h). \tag{19.22}$$

The quantities $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ are specified in (19.13) and (19.18) (and $(\overline{g}_h, \overline{D}_h)$ are the same as in Lemma 10.2). The definitions of $\overline{v}_h$, $\overline{LM}_h$, $\overline{J}_h$, and $\overline{CLR}_h$ in (19.22) are the same as in (11.1), (11.2), (12.6), and (12.7), respectively.

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 (12.6)). 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 19.1 *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 (19.3)) 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 19.1 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 19.1 does not hold, then the following weaker conclusion holds:

$$\begin{aligned} & AsySz \qquad (19.23) \\ & \in [\max\{\alpha, \sup_{h \in H} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))\}, \max\{\alpha, \sup_{h \in H} \lim_{x \uparrow 0} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) + x)\}]. \end{aligned}$$

(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). Note that $\overline{g}_h$ is independent of $\overline{D}_h$. The proviso in Theorem 19.1 holds 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 (19.22), (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 19.1 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 (19.19) equals $0^{k \times p}$ by Comment (ii) to Theorem 19.3 below. This implies that $r_h(\overline{D}_h, \overline{M}_h)$ in (19.20) 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 (12.9)-(12.10) in the proof of Theorem 12.1.

(iv) Theorem 19.1 is proved by showing that it is a special case of Theorem 19.6 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 19.2 below shows that the estimator $\widetilde{V}_{Dn}$ defined in (5.3) satisfies Assumption VD.

(v) A CS version of Theorem 19.1 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 (19.5) with its dependence on θ_0 made explicit. The proof of this CS result is as outlined in the Comment to Proposition 10.1. 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.

19.3 Simulation Results

In this section, for a particular linear IV regression model, we simulate (i) correlations between $\overline{M}_{h,p-q^\dagger}^\dagger$ (defined in (19.19)) 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 and $k = 15$ IV's. The model is

$$y_{1i} = Y_{2i}' \theta_0 + u_i \text{ and } Y_{2i} = \pi' Z_i + V_{2i}, \tag{19.24}$$

where $y_{1i}, u_i \in R$, $Y_{2i}, V_{2i} \in R^2$, $\theta_0 \in R^2$, $Z_i = (Z_{i1}, ..., Z_{ik})' \in R^k$, and $\pi \in R^{k \times 2}$. We take $Z_{ij} \sim \chi_1^2 - 1$ i.i.d. for $j = 1, ..., k$, $u_i \sim ||Z_i|| \widetilde{u}_i$, $(\widetilde{u}_i, V_{2i}')' \sim N(0, \Sigma_\rho)$, $(\widetilde{u}_i, V_{2i}')'$ independent of Z_i, and $\Sigma_\rho \in R^{3 \times 3}$ with diagonal elements 1 and off-diagonal elements ρ. This data generating process (DGP) involves an asymmetric distribution for Z_{ij} and conditional heteroskedasticity in u_i. We take $\pi = \pi_n = (e_1, e_2 c n^{-1/2})$, where $e_j \in R^k$ denote the jth coordinate vector for $j = 1, 2$. We consider integer values of the constant c in $[0, 30]$, $\rho = .5$, $\theta_0 = (0, 0)'$, and nominal size 5% for the tests. We also experimented with additional DGPs for $(u_i, V_{2i}', Z_i')'$ and $k \in \{5, 10\}$ and nominal size of 1% but no important additional insights were gained from these simulations.

In this model, we have $g_i = Z_i u_i$ and $G_i = -Z_i Y_{2i}'$. Furthermore, $h_{1,1} = \infty$ and $h_{1,2}$ is a finite nonnegative number that depends on c. The quantities $h_{1,j}^\dagger$ for $j = 1, 2$ (defined just below

(19.10)) 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 $\overline{M}^{\dagger}_{h,p-q^{\dagger}}$ and $\overline{g}_h$ for $j = 1,...,k$ and the asymptotic NRP's of the CLR test. To conserve space we do not report the correlations between the jth and kth elements of these vectors for $j \neq k$. 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 $10,000$ with a step size of .005, .05, and 1 for $r \in [0, 100], [100, 1000]$, and $[1000, 10000]$, respectively. These critical values are computed using $4,000,000$ simulation repetitions. Note that for $p = 2$, the dimension $d^* := (p+1)k + (p+1)k((p+1)k+1)/2$ in (19.8) equals 135, 495, and 1080, for $k = 5, 10, 15$, respectively, and simulation with 4 million repetitions becomes computationally involved for large k.

(i) The simulations provide evidence for the findings given in Theorem 5.1 that $\overline{M}^{\dagger}_{h,p-q^{\dagger}}$ (the second column of $\overline{M}^{\dagger}_{h} \in R^{k \times 2}$) and $\overline{g}_h$ are correlated asymptotically in some models under some sequences of distributions. For example, when $k = 15$ the simulated correlations between the jth elements of $\overline{M}^{\dagger}_{h,p-q^{\dagger}}$ and $\overline{g}_h$ for $j = 1, 8, 15$ take on the values .32, .11, and $-.06$, respectively, for all values $c \in [0, 30]$. 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 of $\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}$ and $\overline{g}_h$.

(ii) 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 4.99% and 5.11% for the values of c considered. These values are 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 and from error in approximation of $h^{\dagger}_{1,j}$. For comparison, we also simulated the asymptotic NRP of the LM test (that has asymptotic size equal to nominal size) and find them to be between 5.01% and 5.02% for the values of c considered.

We conclude that, for the model and error distribution considered, the asymptotic NRP's of Kleibergen's CLR test with Jacobian-variance weighting is quite 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.

19.4 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 19.2 and Theorem 19.3, combine to establish Theorem 5.1, see Comment (i) to Theorem 19.3. The first and last results of this section, viz., Lemma 19.2 and Theorem 19.6, combine to prove Theorem 19.1.

The proofs of the results in this section are given in Section 19.6.

Assumption VD: For any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$, the estimator $\widetilde{V}_{Dn}$ is such that $n^{1/2}(\widetilde{M}_n - M_{F_n}) \rightarrow_d \overline{M}_h$ 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 (19.6)), the convergence is joint with

$$n^{1/2}\begin{pmatrix} \widehat{g}_n \\ vec(\widehat{D}_n - E_{F_n}G_i) \end{pmatrix} \rightarrow_d \begin{pmatrix} \overline{g}_h \\ vec(\overline{D}_h) \end{pmatrix} \sim N\left(0^{(p+1)k}, \begin{pmatrix} h_{5,g} & 0^{k \times pk} \\ 0^{pk \times k} & \Phi_h^{vec(G_i)} \end{pmatrix}\right), \quad (19.25)$$

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 (19.25) holds by Lemma 10.2.

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

Lemma 19.2 *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* (19.11) *and* (19.13)-(19.18) *with pd variance matrix.*

Comment: As stated in the paragraph containing (19.21), $\widehat{D}_n$ is defined in Lemma 19.2 and Theorem 19.3 below with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$.

Define

$$S_n^\dagger := Diag\{(n^{1/2}\tau_{1F_n}^\dagger)^{-1}, ..., (n^{1/2}\tau_{qF_n}^\dagger)^{-1}, 1, ..., 1\} \in R^{p \times p} \text{ and } T_n^\dagger := B_n^\dagger S_n^\dagger, \quad (19.26)$$

where $B_n^\dagger$ is defined in (19.7).

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

Theorem 19.3 *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* (19.14) *and* (19.15), $\overline{M}_h^{\dagger}$ *is a nonrandom linear (i.e., affine and homogeneous of degree one) function of* $\overline{M}_h$ *defined in* (19.19), $(\overline{g}_h, \overline{D}_h, \overline{M}_h)$ *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 (19.11)-(19.19). This, Lemma 19.2, and Theorem 19.3 combine to prove Theorem 5.1 of AG1.

(ii) From (19.19), $\overline{M}_h^{\dagger} = 0^{k \times p}$ 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).[13] 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 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 $n^{1/2}E_{F_n}G_{i2} \rightarrow c$ for some $c \in R^k$.

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

Lemma 19.4 *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* (19.3)) *satisfies Assumption* VD. *Then, under all sequences* $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$,

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

(b) $rk_n^{\dagger} := \widehat{\kappa}_{pn}^{\dagger} \rightarrow_d r_h(\overline{D}_h, \overline{M}_h)$ *if* $q^{\dagger} < p$, *where* $r_h(\overline{D}_h, \overline{M}_h)$ *is defined in* (19.20) *using* (19.19) *with* $\overline{M}_h$ *defined in Assumption* VD *(rather than in* (19.18)),

(c) $\widehat{\kappa}_{jn}^{\dagger} \rightarrow_p \infty$ *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.,* $((\widehat{\kappa}_{(q^{\dagger}+1)n}^{\dagger})^{1/2}, ..., (\widehat{\kappa}_{pn}^{\dagger})^{1/2})^{\prime}$, *converges in distribution to the (ordered)* $p - q^{\dagger}$ *vector of the singular values of*

[13] 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 19.2).

$h^{\dagger\prime}_{3,k-q^\dagger}(\overline{\Delta}^\dagger_{h,p-q^\dagger} + \overline{M}^\dagger_{h,p-q^\dagger}) \in R^{(k-q^\dagger)\times(p-q^\dagger)}$, *where* $\overline{M}^\dagger_{h,p-q^\dagger}$ *is defined in* (19.19) *with* $\overline{M}_h$ *defined in Assumption* VD (*rather than in* (19.18)),

(e) *the convergence in parts* (a)-(d) *holds jointly with the convergence in Theorem* 19.3*, 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 $rk^\dagger_n$ and the asymptotic null rejection probabilities of Kleibergen's CLR test.

Lemma 19.5 *Let the parameter space for* F *be* $\mathcal{F}_{KCLR}$*. Suppose the variance matrix estimator* $\widetilde{V}_{Dn}$ *employed by the rank statistic* $rk^\dagger_n$ (*defined in* (19.3)) *satisfies Assumption* VD*. Then, under all sequences* $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$,

(a) $CLR_n = LM_n + o_p(1) \to_d \chi^2_p$ *and* $rk^\dagger_n \to_p \infty$ *if* $q^\dagger = p$,

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

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

(d) $\lim_{n\to\infty} P(CLR_n > c(1-\alpha, rk^\dagger_n)) = P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))$ *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,

$$clr(r) := \frac{1}{2}\left(\chi^2_p + \chi^2_{k-p} - r + \sqrt{(\chi^2_p + \chi^2_{k-p} - r)^2 + 4\chi^2_p r}\right), \tag{19.27}$$

where the chi-square random variables χ^2_p and χ^2_{k-p} 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^2_p and χ^2_{k-p} (see the paragraph following (12.6)). In this case, even when $q^\dagger = p$,

$$P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)) = E_{\overline{D}_h} P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h) | \overline{D}_h) = \alpha, \tag{19.28}$$

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 (19.28) 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 19.5 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 19.6 *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* (19.3)*) 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 19.1 also applies to Theorem 19.6.

(ii) Theorem 19.6 and Lemma 19.2 combine to prove Theorem 19.1.

(iii) A CS version of Theorem 19.6 holds with the parameter space $\mathcal{F}_{\Theta,KCLR}$ in place of $\mathcal{F}_{KCLR}$, see Comment (v) to Theorem 19.1 and the Comment to Proposition 10.1.

19.5 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 19.4. All that is required is a slight change in Assumption VD.

The rank statistic that we consider here is

$$rk_n^\dagger := \lambda_{\min}(n\widehat{D}_n'\widetilde{W}_n'\widetilde{W}_n\widehat{D}_n). \tag{19.29}$$

We replace Assumption VD in Section 19.4 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 $n^{1/2}(\widetilde{W}_n - W^{\dagger}_{F_n}) \to_d \overline{W}_h$ for some non-random $k \times k$ matrices $\{W^{\dagger}_{F_n} : n \geq 1\}$ and some random $k \times k$ matrix $\overline{W}_h \in R^{k \times k}$, $W^{\dagger}_{F_n} \to W^{\dagger}_h$ for some nonrandom pd $k \times k$ matrix $W^{\dagger}_h$, the convergence is joint with the convergence in (19.25), 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 $\widetilde{M}_n$ $(= \widetilde{V}^{-1/2}_{Dn}) = I_p \otimes \widetilde{W}_n$ in Assumption VD, then $\widehat{D}^{\dagger}_n = \widetilde{W}_n \widehat{D}_n$ and the rank statistics in (19.3) and (19.29) are the same. Thus, Assumption W is analogous to Assumption VD with $\widetilde{M}_n = I_p \otimes \widetilde{W}_n$ and $M_{F_n} = I_p \otimes W^{\dagger}_{F_n}$. Note, however, that the latter matrix does not typically satisfy the condition in Assumption VD that M_{F_n} is defined in (19.6), i.e., the condition that $M_{F_n} = (\Phi^{vec(G_i)}_{F_n})^{-1/2}$. Nevertheless, the results in Section 19.4 hold with Assumption VD replaced by Assumption W and with $M_F = I_p \otimes W^{\dagger}_F$, $D^{\dagger}_F = W^{\dagger}_F E_F G_i$, and $\overline{M}_h = I_p \otimes \overline{W}_h$. With these changes, $\overline{D}^{\dagger}_h = W^{\dagger}_h \overline{D}_h$ in (19.14) (because $(\Phi^{vec(G_i)}_h)^{-1/2}$ is replaced by $I_p \otimes W^{\dagger}_h$), $\overline{\Delta}^{\dagger}_h$ is defined as in (19.15) with $\overline{D}^{\dagger}_h$ as just given, and $\overline{M}^{\dagger}_h$ is defined as in (19.19) with $\overline{M}^{\dagger}_{h,p-q^{\dagger}} = \overline{W}_h h_4 h^{\dagger}_{2,p-q^{\dagger}}$.

Below we show the key result that $\overline{M}^{\dagger}_{h,p-q^{\dagger}} = 0^{k \times (p-q^{\dagger})}$ for $rk^{\dagger}_n$ defined in (19.29). By (19.20), this implies that

$$r_h(\overline{D}_h, \overline{M}_h) := \lambda_{\min}((\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}})' h^{\dagger}_{3,k-q^{\dagger}} h^{\dagger\prime}_{3,k-q^{\dagger}} (\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}})) \tag{19.30}$$

when $q^{\dagger} < p$. Note that the rhs in (19.30) 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 19.5 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 19.6 give the following result.

Corollary 19.7 *Let the parameter space for F be $\mathcal{F}_{KCLR}$. Suppose the rank statistic $rk^{\dagger}_n$ (defined in* (19.29)) *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 $rk_n(\theta) = rk^{\dagger}_n$*) *equals α.*

Comment: A CS version of Corollary 19.7 holds with the parameter space $\mathcal{F}_{\Theta,KCLR}$ in place of $\mathcal{F}_{KCLR}$, see Comment (v) to Theorem 19.1 and the Comment to Proposition 10.1.

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

$$W^{\dagger}_h h_4 := \lim W^{\dagger}_{F_n} E_{F_n} G_i = \lim C^{\dagger}_{F_n} \Upsilon^{\dagger}_{F_n} B^{\dagger\prime}_{F_n} = h^{\dagger}_3 \lim \Upsilon^{\dagger}_{F_n} h^{\dagger\prime}_2, \tag{19.31}$$

where $C^{\dagger}_{F_n} \Upsilon^{\dagger}_{F_n} (B^{\dagger}_{F_n})'$ is the singular value decomposition of $W^{\dagger}_{F_n} E_{F_n} G_i$, $\Upsilon^{\dagger}_{F_n}$ is the $k \times p$ matrix with the singular values of $W^{\dagger}_{F_n} E_{F_n} G_i$, denoted by $\{\tau^{\dagger}_{jF_n} : n \geq 1\}$ for $j \leq p$, on the main diagonal and zeroes elsewhere, and $C^{\dagger}_{F_n}$ and $B^{\dagger}_{F_n}$ are the corresponding $k \times k$ and $p \times p$ orthogonal matrices of singular vectors, as defined in (19.7). Hence, $\lim \Upsilon^{\dagger}_n$ exists, call it $\Upsilon^{\dagger}_h$, and equals $h^{\dagger\prime}_3 h_4 h^{\dagger}_2$. That is, the singular value decomposition of $W^{\dagger}_h h_4$ is

$$W^{\dagger}_h h_4 = h^{\dagger}_3 \Upsilon^{\dagger}_h h^{\dagger\prime}_2. \tag{19.32}$$

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

$$\Upsilon^{\dagger}_h = \begin{bmatrix} \Upsilon^{\dagger}_{h,q^{\dagger}} & 0^{q^{\dagger} \times (p-q^{\dagger})} \\ 0^{(k-q^{\dagger}) \times q^{\dagger}} & 0^{(k-q^{\dagger}) \times (p-q^{\dagger})} \end{bmatrix}, \text{ where } \Upsilon^{\dagger}_{h,q^{\dagger}} := Diag\{\tau^{\dagger}_{h,1}, ..., \tau^{\dagger}_{h,q^{\dagger}}\}. \tag{19.33}$$

In addition,

$$h^{\dagger\prime}_2 h^{\dagger}_{2,p-q^{\dagger}} = \begin{pmatrix} 0^{q^{\dagger} \times (p-q^{\dagger})} \\ I_{p-q^{\dagger}} \end{pmatrix}. \tag{19.34}$$

Thus, we have

$$\begin{aligned} \overline{M}^{\dagger}_{h,p-q^{\dagger}} &:= \overline{W}_h (W^{\dagger}_h)^{-1} W^{\dagger}_h h_4 h^{\dagger}_{2,p-q^{\dagger}} = \overline{W}_h (W^{\dagger}_h)^{-1} h^{\dagger}_3 \Upsilon^{\dagger}_h h^{\dagger\prime}_2 h^{\dagger}_{2,p-q^{\dagger}} \\ &= \overline{W}_h (W^{\dagger}_h)^{-1} h^{\dagger}_3 \begin{bmatrix} \Upsilon^{\dagger}_{h,p-q^{\dagger}} & 0^{q^{\dagger} \times (p-q^{\dagger})} \\ 0^{(k-q^{\dagger}) \times q^{\dagger}} & 0^{(k-q^{\dagger}) \times (p-q^{\dagger})} \end{bmatrix} \begin{pmatrix} 0^{q^{\dagger} \times (p-q^{\dagger})} \\ I_{p-q^{\dagger}} \end{pmatrix} \\ &= 0^{k \times (p-q^{\dagger})}, \end{aligned} \tag{19.35}$$

where the first equality holds by the paragraph following Assumption W and uses the condition in Assumption W that $W^{\dagger}_h$ is pd and the second equality holds by (19.33) and (19.34). This completes the proof of Corollary 19.7.

19.6 Proofs of Results Stated in Sections 19.2 and 19.4

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 19.1. Theorem 19.1 follows from Theorem 19.6, which imposes Assumption VD, and Lemma 19.2, which verifies Assumption VD when $\widetilde{V}_{Dn}$ is defined by (5.3). □

Proof of Lemma 19.2. Consider any sequence $\{\lambda_{n,h} \in \Lambda_{KCLR} : n \geq 1\}$. By the CLT result in (19.11), the linear expansion of $n^{1/2}(\widehat{D}_n - E_{F_n} G_i)$ in (15.1), and the definitions of $\overline{g}_h$ and $\overline{D}_h$ in

(19.13), we have

$$n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i) \rightarrow_d (\overline{g}_h, \overline{D}_h). \quad (19.36)$$

Next, we apply the delta method to the CLT result in (19.11) and the function $a(\cdot)$ defined in (19.16). The mean component in the lhs quantity in (19.11) is $(0^{(p+1)k\prime}, vech(E_{F_n}f_i^*f_i^{*\prime})')'$. We have

$$\begin{aligned} & a\left(\left(\begin{array}{c} 0^{(p+1)k} \\ vech(E_{F_n}f_i^*f_i^{*\prime}) \end{array}\right)\right) \\ = \;& vech\left(\left(E_{F_n}vec(G_i - E_{F_n}G_i)vec(G_i - E_{F_n}G_i)' - \Gamma_{F_n}^{vec(G_i)}\Omega_{F_n}^{-1}\Gamma_{F_n}^{vec(G_i)\prime}\right)^{-1/2}\right) \\ = \;& vech\left(\left(\Phi_{F_n}^{vec(G_i)}\right)^{-1/2}\right) = vech(M_{F_n}), \end{aligned} \quad (19.37)$$

where $\Gamma_{F_n}^{vec(G_i)}$ and Ω_{F_n} are defined in (3.6), the first equality uses the definitions of $a(\cdot)$ and f_i^* (given in (19.16) and (5.6), respectively), the second equality holds by the definition of $\Phi_{F_n}^{vec(G_i)}$ in (10.15), and the third equality holds by the definition of M_{F_n} in (19.6). 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

$$\begin{aligned} n^{1/2}(A_n - vech(M_{F_n})) \;&=\; n^{1/2}\left(a\left(n^{-1}\sum_{i=1}^{n}\left(\begin{array}{c} f_i^* \\ vech\left(f_i^*f_i^{*\prime}\right) \end{array}\right)\right) - a\left(\begin{array}{c} 0^{(p+1)k} \\ vech(E_{F_n}f_i^*f_i^{*\prime}) \end{array}\right)\right) \\ &\rightarrow_d \overline{A}_h\overline{L}_h, \end{aligned} \quad (19.38)$$

where the first equality holds by (19.37) and the definitions of $a(\cdot)$ and A_n in (19.16), the convergence holds by the delta method using the CLT result in (19.11) and the definition of $\overline{A}_h$ following (19.16).

Applying the inverse $vech(\cdot)$ operator, namely, $vech_{kp,kp}^{-1}(\cdot)$, to both sides of (19.38) gives the reconfigured convergence result

$$n^{1/2}(vech_{kp,kp}^{-1}(A_n)) - M_{F_n}) \rightarrow_d vech_{kp,kp}^{-1}(\overline{A}_h\overline{L}_h) = \overline{M}_h, \quad (19.39)$$

where the last equality holds by the definition of $\overline{M}_h$ in (19.18).

The convergence results in (19.36) and (19.39) hold jointly because both rely on the convergence result in (19.11).

We show below that

$$n^{1/2}(\widetilde{V}_{Dn} - (vech_{kp,kp}^{-1}(A_n))^{-2}) = o_p(1). \quad (19.40)$$

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

$$n^{1/2}(\widetilde{V}_{Dn}^{-1/2} - vech_{kp,kp}^{-1}(A_n)) = o_p(1) \tag{19.41}$$

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 (19.36), (19.39), and (19.41) establish the result of the lemma.

Now we prove (19.40). We have

$$\begin{aligned}
\widetilde{V}_{Dn} &:= n^{-1}\sum_{i=1}^{n} vec(G_i - \widehat{G}_n)vec(G_i - \widehat{G}_n)' - \widehat{\Gamma}_n\widehat{\Omega}_n^{-1}\widehat{\Gamma}_n' \\
&= \left(n^{-1}\sum_{i=1}^{n} vec(G_i - E_{F_n}G_i)vec(G_i - E_{F_n}G_i)'\right) - \left(vec(\widehat{G}_n - E_{F_n}G_i)vec(\widehat{G}_n - E_{F_n}G_i)'\right) \\
&\quad - \left(\widetilde{\Gamma}_n - vec(\widehat{G}_n - E_{F_n}G_i)\widehat{g}_n'\right)\left(\widetilde{\Omega}_n - \widehat{g}_n\widehat{g}_n'\right)^{-1}\left(\widetilde{\Gamma}_n - vec(\widehat{G}_n - E_{F_n}G_i)\widehat{g}_n'\right)' \\
&= n^{-1}\sum_{i=1}^{n} vec(G_i - E_{F_n}G_i)vec(G_i - E_{F_n}G_i)' - \widetilde{\Gamma}_n\widetilde{\Omega}_n^{-1}\widetilde{\Gamma}_n' + O_p(n^{-1}),
\end{aligned} \tag{19.42}$$

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 (3.1), (3.2), and (5.3), and by the definitions of $\widetilde{\Omega}_n$ and $\widetilde{\Gamma}_n$ in (19.16) 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 (3.3)) and (ii) $n^{1/2}\widehat{g}_n = O_p(1)$ (by Lemma 10.3)), $n^{1/2}vec(\widehat{G}_n - E_{F_n}G_i) = O_p(1)$, 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 (15.1)).

Excluding the $O_p(n^{-1})$ term, the rhs in (19.42) equals $(vech_{kp,kp}^{-1}(A_n))^{-2}$. Hence, (19.40) holds and the proof is complete. $\square$

Proof of Theorem 19.3. The proof is similar to that of Lemma 10.3 in Section 10 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 (9.1). We can write

$$\widehat{D}_n^\dagger B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} = D_n^\dagger B_{n,q^\dagger}^\dagger(\Upsilon_{n,q^\dagger}^\dagger)^{-1} + n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger)B_{n,q^\dagger}^\dagger(n^{1/2}\Upsilon_{n,q^\dagger}^\dagger)^{-1}. \tag{19.43}$$

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

$$\begin{aligned} D_n^\dagger B_{n,q^\dagger}^\dagger (\Upsilon_{n,q^\dagger}^\dagger)^{-1} &= C_n^\dagger \Upsilon_n^\dagger B_n^{\dagger\prime} B_{n,q^\dagger}^\dagger (\Upsilon_{n,q^\dagger}^\dagger)^{-1} = C_n^\dagger \Upsilon_n^\dagger \begin{pmatrix} I_{q^\dagger} \\ 0^{(p-q^\dagger)\times q^\dagger} \end{pmatrix} (\Upsilon_{n,q^\dagger}^\dagger)^{-1} \\ &= C_n^\dagger \begin{pmatrix} I_{q^\dagger} \\ 0^{(k-q^\dagger)\times q^\dagger} \end{pmatrix} = C_{n,q^\dagger}^\dagger. \end{aligned} \quad (19.44)$$

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

$$\begin{aligned} n^{1/2}(\widehat{D}_n^\dagger - D_n^\dagger) &= n^{1/2} \sum_{j=1}^p (\widetilde{M}_{1jn}\widehat{D}_{jn} - M_{1jF_n}E_{F_n}G_{ij}, ..., \widetilde{M}_{pjn}\widehat{D}_{jn} - M_{pjF_n}E_{F_n}G_{ij}) \\ &= \sum_{j=1}^p [\widetilde{M}_{1jn} n^{1/2}(\widehat{D}_{jn} - E_{F_n}G_{ij}) + n^{1/2}(\widetilde{M}_{1jn} - M_{1jF_n})E_{F_n}G_{ij}, ..., \\ &\qquad \widetilde{M}_{pjn} n^{1/2}(\widehat{D}_{jn} - E_{F_n}G_{ij}) + n^{1/2}(\widetilde{M}_{pjn} - M_{pjF_n})E_{F_n}G_{ij}] \\ &\to_d \sum_{j=1}^p (M_{1jh}\overline{D}_{jh} + \overline{M}_{1jh}h_{4,j}, ..., M_{pjh}\overline{D}_{jh} + \overline{M}_{pjh}h_{4,j}), \end{aligned} \quad (19.45)$$

where the convergence holds by Lemma 10.2 in Section 10, Assumption VD, and $E_{F_n}G_{ij} \to h_{4,j}$ (by the definition of $h_{4,j}$).

Combining (19.43)-(19.45) gives

$$\widehat{D}_n^\dagger B_{n,q^\dagger}^\dagger (\Upsilon_{n,q^\dagger}^\dagger)^{-1} = C_{n,q^\dagger}^\dagger + o_p(1) \to_p h_{3,q^\dagger}^\dagger = \overline{\Delta}_{h,q^\dagger}^\dagger, \quad (19.46)$$

where the equality uses $n^{1/2}\tau_{jF_n}^\dagger \to \infty$ for all $j \le q^\dagger$ 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 (19.15).

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

$$\begin{aligned} n^{1/2} D_n^\dagger B_{n,p-q^\dagger}^\dagger &= n^{1/2} C_n^\dagger \Upsilon_n^\dagger B_n^{\dagger\prime} B_{n,p-q^\dagger}^\dagger = n^{1/2} C_n^\dagger \Upsilon_n^\dagger \begin{pmatrix} 0^{q^\dagger\times(p-q^\dagger)} \\ I_{p-q^\dagger} \end{pmatrix} \\ &= C_n^\dagger \begin{pmatrix} 0^{q^\dagger\times(p-q^\dagger)} \\ n^{1/2}\Upsilon_{n,p-q^\dagger}^\dagger \\ 0^{(k-p)\times(p-q^\dagger)} \end{pmatrix} \to h_3^\dagger \begin{pmatrix} 0^{q^\dagger\times(p-q^\dagger)} \\ Diag\{h_{1,q^\dagger+1}^\dagger, ..., h_{1,p}^\dagger\} \\ 0^{(k-p)\times(p-q^\dagger)} \end{pmatrix} = h_3^\dagger h_{1,p-q^\dagger}^{\dagger\diamond}, \end{aligned} \quad (19.47)$$

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

(19.10), which uses (10.17).

By (19.45) and $B^{\dagger}_{n,p-q^{\dagger}} \to h^{\dagger}_{2,p-q^{\dagger}}$, we have

$$n^{1/2}(\widehat{D}^{\dagger}_n - D^{\dagger}_n)B^{\dagger}_{n,p-q^{\dagger}} \to_d \overline{D}^{\dagger}_h h^{\dagger}_{2,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}, \tag{19.48}$$

using the definitions of $\overline{D}^{\dagger}_h$ and $\overline{M}^{\dagger}_{h,p-q^{\dagger}}$ in (19.14) and (19.19), respectively.

Using (19.47) and (19.48), we get

$$\begin{aligned} n^{1/2}\widehat{D}^{\dagger}_n B^{\dagger}_{n,p-q^{\dagger}} &= n^{1/2}D^{\dagger}_n B^{\dagger}_{n,p-q^{\dagger}} + n^{1/2}(\widehat{D}^{\dagger}_n - D^{\dagger}_n)B^{\dagger}_{n,p-q^{\dagger}} \\ &\to_d h^{\dagger}_3 h^{\dagger\diamond}_{1,p-q^{\dagger}} + \overline{D}^{\dagger}_h h^{\dagger}_{2,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}} = \overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}, \end{aligned} \tag{19.49}$$

where the last equality holds by the definition of $\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}}$ in (19.15).

Equations (19.46) and (19.49) combine to give

$$\begin{aligned} n^{1/2}\widehat{D}^{\dagger}_n T^{\dagger}_n &= n^{1/2}\widehat{D}^{\dagger}_n B^{\dagger}_n S^{\dagger}_n = (\widehat{D}^{\dagger}_n B^{\dagger}_{n,q^{\dagger}}(\Upsilon^{\dagger}_{n,q^{\dagger}})^{-1}, n^{1/2}\widehat{D}^{\dagger}_n B^{\dagger}_{n,p-q^{\dagger}}) \\ &\to_d (\overline{\Delta}^{\dagger}_{h,q^{\dagger}}, \overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}) = \overline{\Delta}^{\dagger}_h + \overline{M}^{\dagger}_h \end{aligned} \tag{19.50}$$

using the definitions of $S^{\dagger}_n$ and $T^{\dagger}_n$ in (19.26), $\overline{\Delta}^{\dagger}_h$ in (19.15), and $\overline{M}^{\dagger}_h$ in (19.19).

By Lemma 10.2, $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i) \to_d (\overline{g}_h, \overline{D}_h)$. This convergence is joint with that in (19.50) 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 19.3.

The independence of $\overline{g}_h$ and $(\overline{D}_h, \overline{\Delta}^{\dagger}_h)$ follows from the independence of $\overline{g}_h$ and $\overline{D}_h$, which holds by Lemma 10.2, and the fact that $\overline{\Delta}^{\dagger}_h$ is a nonrandom function of $\overline{D}_h$. $\square$

Proof of Lemma 19.4. The proof of Lemma 19.4 is analogous to the proof of Theorem 10.4 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 $q^{\dagger}$, $\widehat{D}^{\dagger}_n$, $D^{\dagger}_n$ $(= D^{\dagger}_{F_n})$, $\widehat{\kappa}^{\dagger}_{jn}$, $B^{\dagger}_n$, $B^{\dagger}_{n,q^{\dagger}}$, $S^{\dagger}_n$, $S^{\dagger}_{n,q^{\dagger}}$, $\tau^{\dagger}_{jF_n}$, and $h^{\dagger}_{3,q^{\dagger}}$, respectively. Theorem 19.3, rather than Lemma 10.3, is employed to obtain the results in (17.37). In consequence, $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are replaced by $\overline{\Delta}^{\dagger}_{h,q^{\dagger}} + \overline{M}^{\dagger}_{h,q^{\dagger}}$ and $\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}$, respectively, where $\overline{\Delta}^{\dagger}_{h,q^{\dagger}} + \overline{M}^{\dagger}_{h,q^{\dagger}} = \overline{\Delta}^{\dagger}_{h,q^{\dagger}}$ (because $\overline{M}^{\dagger}_{h,q^{\dagger}} := 0^{k \times q^{\dagger}}$ by (19.19)). The quantities $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are replaced by $\overline{\Delta}^{\dagger}_{h,q^{\dagger}}$ and $\overline{\Delta}^{\dagger}_{h,p-q^{\dagger}} + \overline{M}^{\dagger}_{h,p-q^{\dagger}}$ in (17.37) and in the rest of the proof of Theorem 10.4. Note that (17.39) holds with $h_{3,q}$ replaced by $h^{\dagger}_{3,q^{\dagger}}$ because $\overline{\Delta}^{\dagger}_{h,q^{\dagger}} = h^{\dagger}_{3,q^{\dagger}}$ by (19.15) (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 (17.39)) are all zero matrices, which simplifies the expressions in (17.41)-(17.44) considerably.

The proof of Theorem 10.4 uses Lemma 17.1 to obtain (17.42). Hence, an analogue of Lemma 17.1 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 (10.5) and $\mathcal{F}_{KCLR} \subset \mathcal{F}_0$ by the argument following (19.5)). Because $\widehat{U}_n = U_n$, the matrices $\widehat{A}_{jn}$ for $j = 1, 2, 3$ (defined in (17.2)) are all zero matrices, which simplifies the expressions in (17.9)-(17.12) considerably. For (17.3) 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

$$n^{1/2}(\widehat{D}_n^{\dagger} - D_n^{\dagger}) = O_p(1). \tag{19.51}$$

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

$$\widehat{D}_n^{\dagger} = \sum_{j=1}^{p} (\widetilde{M}_{1jn} \widehat{D}_{jn}, ..., \widetilde{M}_{pjn} \widehat{D}_{jn}), \text{ where } \widetilde{M}_n = \begin{bmatrix} \widetilde{M}_{11n} & \cdots & \widetilde{M}_{1pn} \\ \vdots & \ddots & \vdots \\ \widetilde{M}_{p1n} & \cdots & \widetilde{M}_{ppn} \end{bmatrix} := \widetilde{V}_{Dn}^{-1/2} \in R^{kp \times kp}. \tag{19.52}$$

By (19.6), we have

$$D_n^{\dagger} = \sum_{j=1}^{p} (M_{1jF_n} D_{jn}, ..., M_{pjF_n} D_{jn}) \tag{19.53}$$

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

$$n^{1/2}(\widetilde{M}_{sjn} \widehat{D}_{jn} - M_{sjF_n} D_{jn}) = \widetilde{M}_{sjn} n^{1/2}(\widehat{D}_{jn} - D_{jn}) + n^{1/2}(\widetilde{M}_{sjn} - M_{sjF_n}) D_{jn} = O_p(1), \tag{19.54}$$

where $n^{1/2}(\widehat{D}_{jn} - D_{jn}) = O_p(1)$ (by Lemma 10.2), $n^{1/2}(\widetilde{M}_{sjn} - M_{sjF_n}) = O_p(1)$ (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 (10.15) 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 (19.5)), and $D_{jn} = O(1)$ (by the moment conditions in $\mathcal{F}$, defined in (3.3)).

Hence,

$$n^{1/2}(\widehat{D}_n^{\dagger} - D_n^{\dagger}) = \sum_{j=1}^{p} n^{1/2}[(\widetilde{M}_{1jn} \widehat{D}_{jn}, ..., \widetilde{M}_{pjn} \widehat{D}_{jn}) - (M_{1jF_n} D_{jn}, ..., M_{pjF_n} D_{jn})] = O_p(1). \tag{19.55}$$

This completes the proof of the analogue of Lemma 17.1, which completes the proof of parts (a)-(d) of Lemma 19.4.

For part (e) of Lemma 19.4, the results of parts (a)-(d) hold jointly with those in Theorem 19.3, rather than those in Lemma 10.3, because Theorem 19.3 is used to obtain the results in (17.37), rather than Lemma 10.3. This completes the proof. $\square$

Proof of Lemma 19.5. The proof of parts (a) and (b) is the same as the proof of Theorem 12.1 for the case where Assumption R(a) holds (which states that $rk_n \to_p \infty$) using Lemma 19.4(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 (12.5)-(12.9) in the proof of Theorem 12.1 for the case where Assumption R(b) holds, using Theorem 19.3 and Lemma 19.4(b) in place of Lemma 10.3, with $r_h(\overline{D}_h, \overline{M}_h)$ (defined in (19.20)) 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 12.1 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 (12.10) 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 19.6 given below uses Corollary 2.1(a) of ACG, which is stated below as Proposition 19.8. It is a generic asymptotic size result. Unlike Proposition 10.1 above, Proposition 19.8 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 (10.1).

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 19.8 (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 CS's, rather than tests. But, following Comment 4 to Theorem 2.1 of ACG, with suitable adjustments (as in Proposition 19.8 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 19.6. The result of Proposition 19.8 for the case where $RP^-(h) \neq RP^+(h)$ for some $h \in H$ is used when proving Comment (i) to Theorem 19.1 and the Comment to Theorem 19.6.

Proof of Theorem 19.6. Theorem 19.6 follows from Lemma 19.5 and Proposition 19.8 because Lemma 19.5 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 $q^\dagger < p$. $\square$

19.7 Proof of Lemma 5.2

Proof of Lemma 5.2. Define $J_n(\theta)$ by the decomposition $AR_n(\theta) = LM_n(\theta) + J_n(\theta)$. Under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ in (10.10), (11.2) and (12.6) imply that

$$\begin{pmatrix} J_{w_n}(\theta_0) \\ LM_{w_n}(\theta_0) \end{pmatrix} \rightarrow_d \begin{pmatrix} \overline{J}_h \\ \overline{LM}_h \end{pmatrix} \sim \begin{pmatrix} \overline{g}_h' h_{5,g}^{-1/2} M_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h \\ \overline{g}_h' h_{5,g}^{-1/2} P_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h, \end{pmatrix}, \tag{19.56}$$

where $\overline{\Delta}_h$ is defined in (10.17). Note that the parameter space Λ_0 for λ defined in (10.9) is equivalent to the parameter space $\mathcal{F}_0$, see Comment (i) to Theorem 11.1.

Equation (19.56) and the CMT imply that the test statistic in (5.7) converges in distribution to

$$\sup_{r \in [0,\infty]} \left[\frac{1}{2} \left(\overline{LM}_h + \overline{J}_h - r + \sqrt{(\overline{LM}_h + \overline{J}_h - r)^2 + 4\overline{LM}_h \cdot r} \right) - c(1-\alpha, r) \right]. \tag{19.57}$$

Conditional on $\overline{\Delta}_h$, $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ^2_p and χ^2_{k-p}, respectively. Therefore, the conditional distribution of the random variable in (19.57) given $\overline{\Delta}_h$ is the same as the distribution of the quantity in (5.8). Since the latter does not depend on $\overline{\Delta}_h$, the same statement holds for the unconditional distribution of the random variable in (19.57).

The results of the previous paragraph verify Assumption B* (stated just above Proposition 10.1) with the limit of the rejection probabilities in Assumption B*, i.e., $\lim_{n\to\infty} RP_{w_n}(\lambda_{w_n})$, equal to the probability that the random variable in (5.8) is positive. The asymptotic size result of the Lemma now follows by Proposition 10.1. $\square$

20 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, 5.3, 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 (10.9) and (10.11), 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 10.1.) We define

$$\lambda_{5,F} := V_F = \sum_{m=-\infty}^{\infty} E_F \begin{pmatrix} g_i \\ vec(G_i - E_F G_i) \end{pmatrix} \begin{pmatrix} g_{i-m} \\ vec(G_{i-m} - E_F G_{i-m}) \end{pmatrix}'. \tag{20.1}$$

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 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 discussion following (7.1) of AG1 above regarding the latter.

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

Lemma 20.1 *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, 5.3, and 6.1 (given in Sections 11, 12, and 13, respectively, above) and the proofs of Lemmas 10.2 and 10.3 and Theorem 10.4 (given in Sections 15, 16, and 17 above), upon which the former proofs rely, for the i.i.d. case with some modifications. The modifications affect the proofs of Lemmas 10.2 and 10.3 and the proof of Theorem 5.3. No modifications are needed elsewhere.

The first modification is the change in the definition of $\lambda_{5,F}$ described in (20.1).

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 (20.1). 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_jg} 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 $\widehat{\Gamma}_{jn} = \widehat{\Gamma}_{jn}(\theta_0) = h_{5,G_jg} + o_p(1)$ in (15.1) in the proof of Lemma 10.2 (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

$$h_5 = \begin{pmatrix} h_{5,g} & h'_{5,Gg} \\ h_{5,Gg} & h_{5,G} \end{pmatrix} \text{ for } h_{5,g} \in R^{k\times k},\ h_{5,Gg} = \begin{pmatrix} h_{5,G_1 g} \\ \vdots \\ h_{5,G_p g} \end{pmatrix} \in R^{pk\times k},\ \text{and } h_{5,G} \in R^{pk\times pk}. \tag{20.2}$$

The fourth modification is that $\widetilde{V}_{Dn}$ in (13.1) in the proof of Theorem 5.3 is defined as described in Section 7, rather than as in (5.3). In addition, $\widetilde{V}_{Dn} \to_p h_7$ in (13.1) 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 10.2 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 20.1 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 10.2 is obtained from (15.1), modified by the second and third modifications above, and the following result:

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^n(\zeta(W_i) - E_{F_n}\zeta(W_i)) = \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} n^{-1/2}\sum_{i=1}^n(f_i - E_{F_n}f_i) \\ & \to_d N(0^{(p+1)k}, L_{h_5}),\ \text{where} \\ & \zeta(W_i) := \begin{pmatrix} g_i \\ vec(G_i) - h_{5,Gg}h_{5,g}^{-1}g_i \end{pmatrix} = \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}, \end{aligned} \tag{20.3}$$

$f_i = (g'_i, vec(G_i)')'$, and the convergence holds by Lemma 20.1. Using (20.2), the variance matrix L_{h_5} in (20.3) takes the form:

$$\begin{aligned} L_{h_5} &= \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} h_{5,g} & h_{5,Gg'} \\ h_{5,Gg} & h_{5,G} \end{pmatrix} \begin{pmatrix} I_k & -h_{5,g}^{-1}h'_{5,Gg} \\ 0^{pk\times k} & I_{pk} \end{pmatrix} \\ &= \begin{pmatrix} I_k & 0^{k\times pk} \\ -h_{5,Gg}h_{5,g}^{-1} & I_{pk} \end{pmatrix} \begin{pmatrix} h_{5,g} & 0^{k\times pk} \\ h_{5,Gg} & \Phi_h^{vec(G_i)} \end{pmatrix} = \begin{pmatrix} h_{5,g} & 0^{k\times pk} \\ 0^{pk\times k} & \Phi_h^{vec(G_i)} \end{pmatrix},\ \text{where} \\ \Phi_h^{vec(G_i)} &= h_{5,G} - h_{5,Gg}h_{5,g}^{-1}h'_{5,Gg}. \end{aligned} \tag{20.4}$$

Equations (15.1) (modified as described above), (20.3), and (20.4) combine to give the result of Lemma 10.2 for the time series case.

The sixth modification occurs in the proof of Lemma 10.3(d) in Section 16 in this SM. In the time series case, the proof goes through as is, except that the calculations in (16.13) 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 $\lim \Psi_{F_n}^{vec(C'_{F_n,k-q}\Omega_{F_n}^{-1/2}G_i B_{F_n,p-q}\xi_2)} = \Phi_h^{vec(h'_{3,k-q}h_{5,g}^{-1/2}G_i h_{2,p-q}\xi_2)}$ (which, in the i.i.d. case, is proved in (16.13).

This completes the proof of Theorem 7.1. □

Proof of Lemma 20.1. 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 $b'h_5 b$, respectively, for arbitrary $b \in R^{(p+1)k}$. First, we show

$$\lim Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) = b'h_5 b, \tag{20.5}$$

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

$$Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) = \sum_{m=-n+1}^{n-1} Cov_{F_n}(s(W_i), s(W_{i-m})) - \sum_{m=-n+1}^{n-1} \frac{|m|}{n} Cov_{F_n}(s(W_i), s(W_{i-m})). \tag{20.6}$$

This gives

$$\begin{aligned} &\left\|Var_{F_n}\left(n^{-1/2}\sum_{i=1}^n s(W_i)\right) - b'\lambda_{5,F_n} b\right\| \\ &\leq 2\sum_{m=n}^{\infty} ||Cov_{F_n}(s(W_i), s(W_{i-m}))|| + \sum_{m=-n+1}^{n-1} \frac{|m|}{n} ||Cov_{F_n}(s(W_i), s(W_{i-m}))||. \end{aligned} \tag{20.7}$$

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

$$\sup_{F \in \mathcal{F}_{TS}} ||Cov_F(s(W_i), s(W_{i-m}))|| \leq C_1 \alpha_F^{\gamma/(2+\gamma)}(m) \leq C_1 C^{\gamma/(2+\gamma)} m^{-d\gamma/(2+\gamma)}, \text{ where } d\gamma/(2+\gamma) > 1, \tag{20.8}$$

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 (20.7) converge to zero. This and $b'\lambda_{5,F_n} b \to b'h_5 b$ establish (20.5).

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 (20.5) give

$$n^{-1/2}\sum_{i=1}^{n} s(W_i) \to_d N(0, b'h_5 b), \tag{20.9}$$

as desired. □

References

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

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2009): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper No. 1813, Yale University.

Andrews, D. W. K., and P. Guggenberger (2014): "Identification- and Singularity-Robust Inference for Moment Condition Models," Cowles Foundation Discussion Paper No. 1978, Yale University.

——— (2017): "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," *Econometric Theory,* forthcoming.

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.

Smith, R. J. (2007): "Weak Instruments and Empirical Likelihood: A Discussion of the Papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society,* Vol. III, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge, UK: Cambridge University Press. Also available as CEMMAP Working Paper No. 13/05, UCL.

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