Supplemental Material for
IDENTIFICATION- AND SINGULARITY-ROBUST INFERENCE FOR MOMENT CONDITION MODELS

By

Donald W. K. Andrews and Patrik Guggenberger

January 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1978S

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

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

Supplemental Material
for
Identification- and Singularity-Robust Inference for Moment Condition Models

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

Patrik Guggenberger
Department of Economics
Pennsylvania State University

First Version: March 25, 2011
Revised: December 31, 2014

Contents

11 Outline

We let AG2 abbreviate the main paper "Identification- and Singularity-Robust Inference for Moment Condition Models." References to sections with section numbers less than 11 refer to sections of AG2. All theorems, lemmas, and equations with section numbers less than 11 refer to results and equations in AG2.

We let SM abbreviate Supplemental Material. We let AG1 abbreviate the paper Andrews and Guggenberger (2014a). The SM to AG1 is given in Andrews and Guggenberger (2014b).

Section 12 generalizes the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests from i.i.d. observations to strictly stationary strong mixing observations.

Section 13 provides finite-sample null rejection probability simulation results for the SR-AR and SR-CQLR$_2$ tests for cases where the variance matrix of the moment functions is singular and near singular.

Section 14 compares the test statistics and conditioning statistics of the SR-CQLR$_1$, SR-CQLR$_2$, and Kleibergen's (2005, 2007) CLR tests to those of Moreira's (2003) LR statistic and conditioning statistic in the homoskedastic linear IV model with fixed (i.e., nonrandom) IV's.

Section 15 provides finite-sample simulation results that illustrate that Kleibergen's CLR test with moment-variance weighting can have low power in certain linear IV models with a single right-hand side (rhs) endogenous variable, as the theoretical results in Section 14 suggest.

Section 16 provides asymptotic power comparisons based on the estimated linear IV models (with one rhs endogenous variable) in Yogo (2004). The tests considered are the AR test, Kleibergen's (2005) LM, JVW-CLR, and MVW-CLR tests, the SR-CQLR$_2$ test, I. Andrews's (2014) plug-in conditional linear combination (PI-CLC) test, and Moreira and Moreira's (2013) MM1-SU and MM2-SU tests.

Section 17 establishes some properties of the eigenvalue-adjustment procedure defined in Section 6.1 and used in the definitions of the two SR-CQLR tests.

Section 18 defines a new SR-LM test.

The remainder of the SM, in conjunction with the Appendix to AG2, provides the proofs of the results stated in AG2 and the SM. Section 19 proves Lemmas 6.1 and 6.2. Section 20 proves Lemma 10.3 and Proposition 10.4. Section 21 proves Theorem 10.5. Section 22 proves Theorem 10.1 (using Theorem 10.5). Section 23 proves Theorem 9.1. Section 24 proves Lemmas 14.1, 14.2, and 14.3. Section 25 proves Theorem 12.1.

For notational simplicity, throughout the SM, we often suppress the argument θ_0 for various quantities that depend on the null value θ_0.

12 Time Series Observations

In this section, we define the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests for observations that are strictly stationary strong mixing. We also generalize the asymptotic size results of Theorem 8.1 from i.i.d. observations to strictly stationary strong mixing observations. In the time series case, F denotes the distribution of the stationary infinite sequence $\{W_i : i = ..., 0, 1, ...\}$.[56]

We define

$$V_{F,n}(\theta) := Var_F\left(n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} g_i(\theta) \\ vec(G_i(\theta))\end{pmatrix}\right),$$
$$\Omega_{F,n}(\theta) := Var_F(n^{-1/2}\sum_{i=1}^{n} g_i(\theta)), \text{ and } r_{F,n}(\theta) := rk(\Omega_{F,n}(\theta)). \quad (12.1)$$

Note that $V_{F,n}(\theta)$, $\Omega_{F,n}(\theta)$, and $r_{F,n}(\theta)$ depend on n in the time series case, but not in the i.i.d. case. We define $A_{F,n}(\theta)$ and $\Pi_{1F,n}(\theta)$ as $A_F(\theta)$ and $\Pi_{1F}(\theta)$ are defined in (4.7), (4.8), and the paragraph following (4.8), but with $\Omega_{F,n}(\theta)$ in place of $\Omega_F(\theta)$.

For the SR-AR test, the parameter space of time series distributions F for the null hypothesis $H_0 : \theta = \theta_0$ is taken to be

$$\begin{aligned}\mathcal{F}_{TS,AR}^{SR} := \{F : &\{W_i : i = ..., 0, 1, ...\} \text{ are stationary and strong mixing under } F \text{ with} \\ &\text{strong mixing numbers } \{\alpha_F(m) : m \geq 1\} \text{ that satisfy } \alpha_F(m) \leq Cm^{-d}, \\ &E_F g_i = 0^k, \text{ and } \sup_{n\geq 1} E_F||\Pi_{1F,n}^{-1/2} A'_{F,n} g_i||^{2+\gamma} \leq M\}\end{aligned} \quad (12.2)$$

for some $\gamma > 0$, $d > (2+\gamma)/\gamma$, and $C, M < \infty$, where the dependence of g_i, $\Pi_{1F,n}$, and $A_{F,n}$ on θ_0 is suppressed. For CS's, we use the corresponding parameter space $\mathcal{F}_{TS,\Theta,AR}^{SR} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,AR}^{SR}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{TS,AR}^{SR}(\theta_0)$ denotes $\mathcal{F}_{TS,AR}^{SR}$ with its dependence on θ_0 made explicit. The moment conditions in $\mathcal{F}_{TS,AR}^{SR}$ are placed on the normalized moment functions $\Pi_{1F,n}^{-1/2} A'_{F,n} g_i$ that satisfy $Var_F(n^{-1/2}\sum_{i=1}^{n} \Pi_{1F,n}^{-1/2} A'_{F,n} g_i) = I_k$ for all $n \geq 1$.

For the SR-CQLR$_1$ and SR-CQLR$_2$ tests, we use the null parameter spaces $\mathcal{F}_{TS,1}^{SR}$ and $\mathcal{F}_{TS,2}^{SR}$, respectively, which are defined as $\mathcal{F}_1^{SR}$ and $\mathcal{F}_2^{SR}$ are defined in (4.9), but with (i) $\mathcal{F}_{TS,AR}^{SR}$ in place of $\mathcal{F}_{AR}^{SR}$, (ii) A_F and Π_{1F} replaced by $A_{F,n}$ and $\Pi_{1F,n}$, respectively, and (iii) $\sup_{n\geq 1}$ added before the quantities $\mathcal{F}_1^{SR}$ and $\mathcal{F}_2^{SR}$ that depend on $A_{F,n}$ and $\Pi_{1F,n}$. For SR-CQLR$_1$ and SR-CQLR$_2$ CS's, we use the parameter spaces $\mathcal{F}_{TS,\Theta,1}^{SR}$ and $\mathcal{F}_{TS,\Theta,2}^{SR}$, respectively, which are defined as $\mathcal{F}_{TS,\Theta,AR}^{SR}$ is

[56] Asymptotics under drifting sequences of true distributions $\{F_n : n \geq 1\}$ are used to establish the correct asymptotic size of the SR-AR and SR-CQLR tests and CS's. Under such sequences, the observations form a triangular array of row-wise strictly stationary observations.

defined, but with $\mathcal{F}_{TS,1}^{SR}(\theta_0)$ and $\mathcal{F}_{TS,2}^{SR}(\theta_0)$ in place of $\mathcal{F}_{TS,AR}^{SR}(\theta_0)$, where $\mathcal{F}_{TS,1}^{SR}(\theta_0)$ and $\mathcal{F}_{TS,2}^{SR}(\theta_0)$ denote $\mathcal{F}_{TS,1}^{SR}$ and $\mathcal{F}_{TS,2}^{SR}$ with their dependence on θ_0 made explicit.

The SR-CQLR test statistics depend on some estimators $\widehat{V}_n$ $(= \widehat{V}_n(\theta_0))$ of $V_{F,n}$. The SR-AR test statistic only depends on an estimator $\widehat{\Omega}_n$ $(= \widehat{\Omega}_n(\theta_0))$ of the submatrix $\Omega_{F,n}$ of $V_{F,n}$. For the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests, these estimators are heteroskedasticity and autocorrelation consistent (HAC) variance matrix estimators based on $\{g_i - \widehat{g}_n : i \leq n\}$, $\{(u_i^* - \widehat{u}_{in}^*) \otimes Z_i : i \leq n\}$ (defined in (6.3)), and $\{f_i - \widehat{f}_n : i \leq n\}$ (defined in (7.1)), respectively. There are a number of HAC estimators available in the literature, e.g., see Newey and West (1987) and Andrews (1991).

We say that $\widehat{V}_n$ is *equivariant* if the replacement of g_i and G_i by $A'g_i$ and $A'G_i$, respectively, in the definition of $\widehat{V}_n$ transforms $\widehat{V}_n$ into $(I_{p+1} \otimes A')\widehat{V}_n(I_{p+1} \otimes A)$, for any matrix $A \in R^{r \times k}$ with full row rank $r \leq k$ for any $r = \{1, ..., k\}$. Equivariance of $\widehat{\Omega}_n$ means that the replacement of g_i by $A'g_i$ transforms $\widehat{\Omega}_n$ into $A'\widehat{\Omega}_n A$. Equivariance holds quite generally for HAC estimators in the literature.

We write the $(p+1)k \times (p+1)k$ matrix $\widehat{V}_n$ in terms of its $k \times k$ submatrices:

$$\widehat{V}_n = \begin{bmatrix} \widehat{\Omega}_n & \widehat{\Gamma}_{1n}' & \cdots & \widehat{\Gamma}_{pn}' \\ \widehat{\Gamma}_{1n} & \widehat{V}_{G_{11}n} & \cdots & \widehat{V}_{G_{p1}n}' \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\Gamma}_{pn} & \widehat{V}_{G_{p1}n} & \cdots & \widehat{V}_{G_{pp}n} \end{bmatrix}. \tag{12.3}$$

We define $\widehat{r}_n$ $(= \widehat{r}_n(\theta_0))$ and $\widehat{A}_n$ $(= \widehat{A}_n(\theta_0))$ as in (5.3) and (5.4) with $\theta = \theta_0$, but with $\widehat{\Omega}_n$ defined in (12.3), rather than in (5.1).

The asymptotic size and similarity properties of the tests considered here are the same for any consistent HAC estimator. Hence, for generality, we do not specify a particular estimator $\widehat{V}_n$ (or $\widehat{\Omega}_n$). Rather, we state results that hold for any estimator $\widehat{V}_n$ (or $\widehat{\Omega}_n$) that satisfies one the following assumptions when the null value θ_0 is the true value. The following assumptions are used with the SR-CQLR$_2$ test and CS, respectively.

Assumption SR-V$_2$: (a) $[I_{p+1} \otimes (\Pi_{1F_n,n}^{-1/2}(\theta_0)A_{F_n,n}'(\theta_0))][\widehat{V}_n(\theta_0) - V_{F_n,n}(\theta_0)][I_{p+1} \otimes (A_{F_n,n}(\theta_0)\Pi_{1F_n,n}^{-1/2}(\theta_0))] \to_p 0^{(p+1)k \times (p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{TS,2}^{SR} : n \geq 1\}$ for which $V_{F_n,n}(\theta_0) \to V$ for some matrix V and $r_{F_n,n}(\theta_0) = r$ for all n large, for any $r \in \{1, ..., k\}$.

(b) $\widehat{V}_n(\theta_0)$ is equivariant.

(c) $\lambda' g_i(\theta_0) = 0$ a.s.$[F]$ implies that $\lambda'\widehat{\Omega}_n(\theta_0)\lambda = 0$ a.s.$[F]$ for all $\lambda \in R^k$ and $F \in \mathcal{F}_{TS,2}^{SR}$.

For SR-CQLR$_2$ CS's, we use the following assumption that allows both the null parameter θ_{0n}, as well as the distribution F_n, to drift with n.

Assumption SR-V$_2$-CS: $[I_{p+1} \otimes (\Pi_{1F_n,n}^{-1/2}(\theta_{0n})A'_{F_n,n}(\theta_{0n}))][\widehat{V}_n(\theta_{0n}) - V_{F_n,n}(\theta_{0n})][I_{p+1} \otimes (A_{F_n,n}(\theta_{0n})\Pi_{1F_n,n}^{-1/2}(\theta_{0n}))] \to_p 0^{(p+1)k\times(p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{(F_n, \theta_{0n}) \in \mathcal{F}_{TS,\Theta,2}^{SR} : n \geq 1\}$ for which $V_{F_n,n}(\theta_{0n}) \to V$ for some matrix V and $r_{F_n,n}(\theta_{0n}) = r$ for all n large, for any $r \in \{1, ..., k\}$.

(b) $\widehat{V}_n(\theta_0)$ is equivariant for all $\theta_0 \in \Theta$.

(c) $\lambda' g_i(\theta_0) = 0$ a.s.$[F]$ implies that $\lambda'\widehat{\Omega}_n(\theta_0)\lambda = 0$ a.s.$[F]$ for all $\lambda \in R^k$ and $(F, \theta_0) \in \mathcal{F}_{TS,\Theta,2}^{SR}$.

Assumptions SR-V$_2$(a) and SR-V$_2$-CS(a) require the HAC estimator based on the normalized moments and Jacobian (i.e., $\Pi_{1F_n,n}^{-1/2}(\theta_{0n})A'_{F_n,n}(\theta_{0n})g_i(\theta_{0n})$ and $\Pi_{1F_n,n}^{-1/2}(\theta_{0n})A'_{F_n,n}(\theta_{0n})G_i(\theta_{0n})$, respectively) to be consistent. This can be verified using standard methods. For typical HAC estimators, equivariance and Assumptions SR-V$_2$(c) and SR-V$_2$-CS(c) can be shown easily.

For the SR-CQLR$_1$ test and CS, we use **Assumptions SR-V$_1$** and **SR-V$_1$-CS**, which are defined as Assumptions SR-V$_2$ and SR-V$_2$-CS are defined, respectively, but with $\mathcal{F}_{TS,1}^{SR}$ and $\mathcal{F}_{TS,\Theta,1}^{SR}$ in place of $\mathcal{F}_{TS,2}^{SR}$ and $\mathcal{F}_{TS,\Theta,2}^{SR}$.

For the SR-AR test and CS, we use **Assumptions SR-Ω** and **SR-Ω-CS**, which are defined as Assumptions SR-V$_2$ and SR-V$_2$-CS are defined, respectively, but with (i) Assumption SR-Ω(a) being: $\Pi_{1F_n,n}^{-1/2}(\theta_0)A'_{F_n,n}(\theta_0))[\widehat{\Omega}_n(\theta_0) - \Omega_{F_n,n}(\theta_0)]A_{F_n,n}(\theta_0)\Pi_{1F_n,n}^{-1/2}(\theta_0) \to_p 0^{k\times k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{TS,AR}^{SR} : n \geq 1\}$ for which $\Omega_{F_n,n}(\theta_0) \to \Omega$ for some matrix Ω and $r_{F_n,n}(\theta_0) = r$ for all n large, for any $r \in \{1, ..., k\}$, (ii) Assumption SR-Ω-CS(a) being as in (i), but with θ_{0n} and $\mathcal{F}_{TS,\Theta,AR}^{SR}$ in place of θ_0 and $\mathcal{F}_{TS,AR}^{SR}$, (iii) $\widehat{\Omega}_n(\theta_0)$ in place of $\widehat{V}_n(\theta_0)$ in part (b) of each assumption, and (iv) $\mathcal{F}_{TS,AR}^{SR}$ in place of $\mathcal{F}_{TS,2}^{SR}$ in part (c) of each assumption.

Now we define the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests in the time series context. The definitions are the same as in the i.i.d. context given in Sections 5, 6, and 7 with the following changes. For all three tests, $\widehat{r}_n$ and $\widehat{A}_n^{\perp}$ in the condition $\widehat{A}_n^{\perp\prime}\widehat{g}_n \neq 0^{k-\widehat{r}_n}$ in (5.7) are defined as in (5.3) and (5.4), but with $\widehat{\Omega}_n$ defined to satisfy Assumption SR-Ω, rather than being defined in (5.1). The SR-AR statistic is defined as in Section 5, but with $\widehat{\Omega}_n$ defined to satisfy Assumption SR-Ω. This affects the definitions of $\widehat{r}_n$ and $\widehat{A}_n$, given in (5.3) and (5.4). With these changes, the critical value for the SR-AR test in the time series case is defined in the same way as in the i.i.d. case.

In the time series case, the SR-QLR$_1$ statistic is defined as in Section 6, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined to satisfy Assumption SR-V$_1$ and (12.3) based on $\{(u_i^* - \widehat{u}_{in}^*) \otimes Z_i : i \leq n\}$, rather than in (6.3) and (5.1), respectively. In turn, this affects the definitions of $\widehat{R}_n$, $\widehat{\Sigma}_n$, $\widehat{L}_n$, $\widehat{D}_n^*$, $\widehat{Q}_n$, $\widehat{r}_n$, $\widehat{A}_n$, and $SR\text{-}AR_n$ (which appears in (6.7)). Given the changes described above, the definition of the SR-CQLR$_1$ critical value is unchanged.

In the time series case, the SR-QLR$_2$ statistic is defined as in Section 7, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$

defined to satisfy Assumption SR-V_2 and (12.3) based on $\{f_i - \widehat{f}_n : i \le n\}$, in place of $\widetilde{V}_n$ and $\widehat{\Omega}_n$ defined in (7.1) and (5.1), respectively. This affects the definitions of $\widetilde{R}_n$, $\widetilde{\Sigma}_n$, $\widetilde{L}_n$, $\widetilde{D}_n^*$, $\widehat{r}_n$, $\widehat{A}_n$, and $SR\text{-}AR_n$. Given the previous changes, the definition of the SR-CQLR_2 critical value is unchanged.

In the time series context,

$$\begin{aligned} V_F &:= \lim Var_F\left(n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} g_i \\ vec(G_i)\end{pmatrix}\right) \\ &= \sum_{m=-\infty}^{\infty} E_F \begin{pmatrix} g_i \\ vec(G_i - E_F G_i)\end{pmatrix}\begin{pmatrix} g_{i-m} \\ vec(G_{i-m} - E_F G_{i-m})\end{pmatrix}' \text{ and} \\ \Omega_F &:= \sum_{m=-\infty}^{\infty} E_F g_i g_{i-m}', \end{aligned} \tag{12.4}$$

where the dependence of various quantities on the null value θ_0 is suppressed for notational simplicity. The second equality holds for $F \in \mathcal{F}_{TS,2}^{SR}$.[57]

For the time series case, the asymptotic size and similarity results for the tests described above are as follows.

Theorem 12.1 *Suppose the* SR-AR, SR-CQLR_1, *and* SR-CQLR_2 *tests are defined as in this section, the null parameter spaces for F are $\mathcal{F}_{TS,AR}^{SR}$, $\mathcal{F}_{TS,1}^{SR}$, and $\mathcal{F}_{TS,2}^{SR}$, respectively, and the corresponding Assumption* SR-Ω, SR-V_1, *or* SR-V_2 *holds for each test. Then, these tests have asymptotic sizes equal to their nominal size $\alpha \in (0,1)$. These tests also are asymptotically similar (in a uniform sense) for the subsets of these parameter spaces that exclude distributions F under which $g_i = 0^k$ a.s. Analogous results hold for the* SR-AR, SR-CQLR_1, *and* SR-CQLR_2 CS's *for the parameter spaces $\mathcal{F}_{TS,\Theta,AR}^{SR}$, $\mathcal{F}_{TS,\Theta,1}^{SR}$, and $\mathcal{F}_{TS,\Theta,2}^{SR}$, respectively, provided the corresponding Assumption* SR-Ω-CS, SR-V_1-CS, *or* SR-V_2-CS *holds for each* CS*, rather than Assumption* SR-Ω, SR-V_1, *or* SR-V_2.

13 Simulation Results for Singular and Near-Singular Variance Matrices

Here, we provide some finite-sample simulations of the null rejection probabilities of the nominal 5% SR-AR and SR-CQLR_2 tests when the variance matrix of the moments is singular and near singular.[58] The model we consider is the second example discussed in Section 4.2 in AG2 in which the reduced-form equations are $y_{1i} = Z_i'\pi\beta + V_{1i}$ and $Y_{2i} = Z_i'\pi + V_{2i}$ and the moment functions are

[57] This is shown in the proof of Lemma 19.1 in Section 19 in the SM to AG1.

[58] Analogous results for the SR-CQLR_1 test are not provided because the moment functions considered are not of the form in (4.4) in AG2, which is necessary to apply the SR-CQLR_1 test.

Table I. Null Rejection Probabilities (×100) of Nominal 5% SR-AR and SR-CQLR$_2$ Tests with Singular and Near Singular Variance Matrices of the Moment Functions and $k = 8$

		SR-AR			SR-CQLR$_2$		
n	ρ_V:	.95	.999,999	1.0	.95	.999,999	1.0
250		6.0	6.0	5.4	5.8	5.8	5.3
500		5.5	5.5	5.2	5.3	5.3	5.1
1,000		5.5	5.5	5.2	5.3	5.3	5.1
2,000		5.0	5.0	4.9	4.8	4.8	4.8
4,000		5.0	5.0	5.1	4.8	4.8	5.0
8,000		5.1	5.1	5.0	4.8	4.8	4.9
16,000		5.0	5.0	5.1	4.9	4.9	5.0

$g_i(\theta) = ((y_{1i} - Z_i'\pi\beta)Z_i', (Y_{2i} - Z_i'\pi)Z_i')' \in R^k$, where $k = 2d_Z$ and d_Z is the dimension of Z_i. We take $(V_{1i}, V_{2i}) \sim N(0^2, \Sigma_V)$, where Σ_V has unit variances and correlation ρ_V, $Z_i \sim N(0^2, I_{d_Z})$, (V_{1i}, V_{2i}) and Z_i are independent, and the observations are i.i.d. across i. The null hypothesis is $H_0 : (\beta, \pi) = (\beta_0, \pi_0)$. We consider the values: $\rho_V = .95$, $.999,999$, and 1.0; $n = 250$, $500, 1,000$, $2,000$, $4,000$, $8,000$, and $16,000$; $\pi_0 = (\pi_{10}, 0, 0, 0)'$, where $\pi_{10} = \pi_{10n} = C/n^{1/2}$ and $C = \sqrt{10}$, which yields a concentration parameter of $\lambda = \pi' E Z_i Z_i' \pi = 10$ for all $n \geq 1$; and $\beta_0 = 0$. The variance matrix Ω_F of the moment functions is singular when $\rho_V = 1$ (because $g_i(\theta_0) = (V_{1i}Z_i', V_{1i}Z_i')'$ a.s.) and near singular when ρ_V is close to one. Under H_0, with probability one, the extra rejection condition in (5.7) is: reject H_0 if $[I_4, -I_4]\widehat{g}_n(\theta_0) \neq 0^4$, which fails to hold a.s. and, hence, can be ignored in probability calculations made under H_0. Forty thousand simulation repetitions are employed.

Tables I-III report results for $k = 8$ (which corresponds to $d_Z = 4$), $k = 4$, and $k = 12$, respectively. Table I shows that the SR-AR and SR-CQLR$_2$ tests have null rejection probabilities that are close to the nominal 5% level for singular and near singular variance matrices as measured by ρ_V. As expected, the deviations from 5% decrease with n. For all $40,000$ simulation repetitions, all values of n considered, and $k = 8$, we obtain $\widehat{r}_n(\theta_0) = 8$ when $\rho_V < 1.0$ and $\widehat{r}_n(\theta_0) = 4$ when $\rho_V = 1$. The estimator $\widehat{r}_n(\theta_0)$ also makes no errors when $k = 4$ and 12. Tables II and III show that the deviations of the null rejection probabilities from 5% are somewhat smaller when $k = 4$ and $n \leq 1000$ than when $k = 8$, and somewhat larger when $k = 12$ and $n \leq 500$. Results for $k = 8$ and $C = 0$, 2, $\sqrt{30}$, and 10 produced similar results. For brevity, these results are not reported.

We conclude that the method introduced in Section 5 to make the SR-AR and SR-CQLR$_2$ tests robust to singularity works very well in the model that is considered in the simulations.

Table II. Null Rejection Probabilities ($\times 100$) of Nominal 5% SR-AR and SR-CQLR$_2$ Tests with Singular and Near Singular Variance Matrices of the Moment Functions and $k = 4$

		SR-AR			SR-CQLR$_2$		
n	ρ_V:	.95	.999,999	1.0	.95	.999,999	1.0
250		5.5	5.5	5.2	5.4	5.4	4.9
500		5.1	5.1	5.2	5.0	5.0	5.0
1,000		4.9	4.9	5.1	4.8	4.8	4.8
2,000		5.1	5.1	5.2	5.0	5.0	5.0
4,000		5.1	5.1	5.1	5.0	5.0	4.9
8,000		5.1	5.1	5.1	5.0	5.0	4.8
16,000		5.1	5.1	5.0	4.9	4.9	4.8

Table III. Null Rejection Probabilities ($\times 100$) of Nominal 5% SR-AR and SR-CQLR$_2$ Tests with Singular and Near Singular Variance Matrices of the Moment Functions and $k = 12$

		SR-AR			SR-CQLR$_2$		
n	ρ_V:	.95	.999,999	1.0	.95	.999,999	1.0
250		7.0	7.0	5.6	7.0	7.0	5.5
500		6.0	6.0	5.4	6.0	6.0	5.4
1,000		5.5	5.5	5.3	5.5	5.5	5.3
2,000		5.2	5.2	5.1	5.2	5.2	5.1
4,000		5.1	5.1	5.1	5.1	5.1	5.1
8,000		5.0	5.0	4.9	5.0	5.0	4.8
16,000		4.9	4.9	5.0	4.9	4.9	5.0

14 SR-CQLR$_1$, SR-CQLR$_2$, and Kleibergen's Nonlinear CLR Tests in the Homoskedastic Linear IV Model

It is desirable for tests to reduce asymptotically to Moreira's (2003) CLR test in the homoskedastic linear IV regression model with fixed (i.e., nonrandom) IV's when $p = 1$, where p is the number of endogenous rhs variables, which equals the dimension of θ. The reason is that the latter test has been shown to have some (approximate) optimality properties under normality of the errors, see Andrews, Moreira, and Stock (2006, 2008) and Chernozhukov, Hansen, and Jansson (2009).[59]

In this section, we show that the components of the SR-QLR$_1$ statistic and its corresponding conditioning matrix are asymptotically equivalent to those of Moreira's (2003) LR statistic and its conditioning statistic, respectively, in the homoskedastic linear IV model with $k \geq p$ fixed (i.e., nonrandom) IV's and nonsingular moments variance matrix (whether or not the errors are Gaussian). This holds for all values of $p \geq 1$.

We also show that the same is true for the SR-QLR$_2$ statistic and its conditioning matrix in some, but not in all cases (where the cases depend on the behavior of the reduced-form parameter matrix $\pi \in R^{k \times p}$ as $n \to \infty$.) Nevertheless, when $p = 1$, the SR-CQLR$_2$ test and Moreira's (2003) CLR test are asymptotically equivalent. When $p \geq 2$, for the cases where asymptotic equivalence of these tests does not hold, the difference is due only to the IV's being fixed, whereas the SR-QLR$_2$ statistic and its conditioning matrix are designed (essentially) for random IV's.

We also evaluate the behavior of Kleibergen's (2005, 2007) nonlinear CLR tests in the homoskedastic linear IV model with fixed IV's. Kleibergen's tests depend on the choice of a weight matrix for the conditioning statistic (which enters both the CLR test statistic and the critical value function). We find that when $p = 1$ Kleibergen's CLR test statistic and conditioning statistic reduce asymptotically to those of Moreira (2003) when one employs the Jacobian-variance weighted conditioning statistic suggested by Kleibergen (2005, 2007) and Smith (2007). However, they do not when one employs the moments-variance weighted conditioning statistic suggested by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012). Notably, the scale of the scalar conditioning statistic can differ from the desired value of one by a factor that can be arbitrarily close to zero or infinity (depending on the value of the reduced-form error matrix Σ_V and null hypothesis value θ_0), see Lemma 14.3 and Comment (iv) following it. Kleibergen's nonlinear CLR tests depend on the form of a rank statistic. When $p \geq 2$, we find that no choice of rank statistic makes Kleibergen's CLR test statistic and conditioning statistic reduce asymptotically to those of Moreira (2003) (when Jacobian- or moments-variance weighting is employed).

[59] Whether this also holds for $p \geq 2$ is an open question.

Section 15 below provides finite-sample simulation results that illustrate the results of the previous paragraph for Kleibergen's CLR test with moment-variance weighting.

14.1 Homoskedastic Linear IV Model

The model we consider is the homoskedastic linear IV model introduced in Section 3 but without the assumption of normality of the reduced-form errors V_i. Specifically, we use the following assumption.

Assumption HLIV: (a) $\{V_i \in R^{p+1} : i \geq 1\}$ are i.i.d., $\{Z_i \in R^k : i \geq 1\}$ are fixed, not random, and $k \geq p$.

(b) $EV_i = 0$, $\Sigma_V := EV_iV_i'$ is pd, and $E||V_i||^4 < \infty$.[60]

(c) $n^{-1}\sum_{i=1}^n Z_iZ_i' \to K_Z$ for some pd matrix $K_Z \in R^{k\times k}$, $n^{-1}\sum_{i=1}^n ||Z_i||^6 = o(n)$, and $\sup_{i\leq n}(c'Z_i)^2/\sum_{i=1}^n(c'Z_i)^2 \to 0\ \forall c \neq 0^k$.

(d) $\sup_{\pi\in\Pi}||\pi|| < \infty$, where Π is the parameter space for π.

(e) $\lambda_{\max}(\Sigma_V)/\lambda_{\min}(\Sigma_V) \leq 1/\varepsilon$ for $\varepsilon > 0$ as in the definition of the SR-QLR$_1$ or SR-QLR$_2$ statistic.

Here HLIV abbreviates "homoskedastic linear IV model." Assumption HLIV(b) specifies that the reduced-form errors are homoskedastic (because their variance matrix does not depend on i or Z_i). Assumptions HLIV(c) and (d) are used to obtain a weak law of large numbers (WLLN) and central limit theorem (CLT) for certain quantities under drifting sequences of reduced-form parameters $\{\pi_n : n \geq 1\}$. These assumptions are not very restrictive. Note that Assumptions HLIV(a)-(c) imply that the variance matrix of the sample moments is pd. This implies that $\widehat{r}_n$ $(= \widehat{r}_n(\theta_0)) = k$ wp$\to$1 (by Lemma 14.1(b) below) and no SR adjustment of the SR-CQLR tests occurs (wp$\to$1). Assumption HLIV(e) guarantees that the eigenvalue adjustment used in the definition of the SR-QLR statistics does not have any effect asymptotically. One could analyze the properties of the SR-CQLR tests when this condition is eliminated. One would still obtain asymptotic null rejection probabilities equal to α, but the eigenvalue adjustment would render the SR-CQLR tests to behave somewhat differently than Moreira's CLR test, because the latter test does not employ an eigenvalue adjustment.

[60] In this section, the underlying i.i.d. random variables $\{V_i : i \geq 1\}$ have a distribution that does not depend on n. Hence, for notational simplicity, we denote expectations by E, rather than E_{F_n}. Nevetheless, it should be kept in mind that the reduced-form parameters π_n may depend on n.

14.2 SR-CQLR$_1$ Test

The components of the SR-QLR$_1$ statistic and its conditioning matrix are $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ and $n^{1/2}\widehat{D}_n^*$ (see (5.2) and (6.7)) when $\widehat{r}_n = k$, which holds wp→1 under Assumption HLIV. Those of Moreira (2003) are $\overline{S}_n$ and $\overline{T}_n$ (see (3.4)). The asymptotic equivalence of these components in the model specified by (3.1)-(3.2) and Assumption HLIV is established in parts (e) and (f) of the following lemma. Parts (a)-(d) of the lemma establish the asymptotic behavior of the components $\widehat{\Omega}_n$ and $\widehat{\Sigma}_n$ of the test statistic $SR\text{-}QLR_{1n}$ and its conditioning statistic.

Lemma 14.1 *Suppose Assumption* HLIV *holds. Under the null hypothesis* $H_0 : \theta = \theta_0$, *for any sequence of reduced-form parameters* $\{\pi_n \in \Pi : n \geq 1\}$ *and any* $p \geq 1$, *we have*

(a) $\widehat{R}_n \rightarrow_p \Sigma_V \otimes K_Z$,

(b) $\widehat{\Omega}_n \rightarrow_p (b_0'\Sigma_V b_0)K_Z$, *where* $b_0 := (1, -\theta_0')'$,

(c) $\widehat{\Sigma}_n \rightarrow_p (b_0'\Sigma_V b_0)^{-1}\Sigma_V$,

(d) $\widehat{\Sigma}_n^\varepsilon \rightarrow_p (b_0'\Sigma_V b_0)^{-1}\Sigma_V$,

(e) $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = \overline{S}_n + o_p(1)$, *and*

(f) $n^{1/2}\widehat{D}_n^* = -(I_k + o_p(1))\overline{T}_n(I_p + o_p(1)) + o_p(1)$.

Comments: (i) The minus sign in Lemma 14.1(f) is not important because QLR_{1n} in (6.7) is unchanged if $\widehat{D}_n^*$ is replaced by $-\widehat{D}_n^*$ in the definition of $\widehat{Q}_n$ (and $SR\text{-}QLR_{1n} = QLR_{1n}$ wp→1 under Assumption HLIV).[61]

(ii) The results of Lemma 14.1 hold under the null hypothesis. Statistics that differ by $o_p(1)$ under sequences of null distributions also differ by $o_p(1)$ under sequences of contiguous alternatives. Hence, the asymptotic equivalence results of Lemma 14.1(e) and (f) also hold under contiguous alternatives to the null.

Note that in the linear IV regression model the alternative parameter values $\{\theta_n : n \geq 1\}$ that yield contiguous sequences of distributions from a sequence of null distributions depend on the strength of identification as measured by π_n. The reduced-form equation (3.2) states that $y_{1i} = Z_i'\pi_n\theta_n + V_{1i}$ when π_n and θ_n are the true values of π and θ. Contiguous alternatives to the null distributions with parameters π_n and θ_0 are obtained for parameter values π_n and θ_n $(\neq \theta_0)$ that satisfy $\pi_n\theta_n - \pi_n\theta_0 = \pi_n(\theta_n - \theta_0) = O(n^{-1/2})$. If the IV's are strong, i.e., $\liminf_{n\to\infty} \pi_n' n^{-1}\sum_{i=1}^n Z_i Z_i'\pi_n > 0$, then contiguous alternatives have true θ_n values of distance $O(n^{-1/2})$ from the null value θ_0. If the IV's are weak in the standard sense, e.g., $\pi_n = \pi n^{-1/2}$ for

[61] This holds because for $a_1 \in R^k$ and $A_2 \in R^{k\times p}$ we have $\lambda_{\min}((a_1, -A_2)'(a_1, -A_2)) = \inf_{\lambda=(\lambda_1,\lambda_2')':||\lambda||=1}(a_1\lambda_1 - A_2\lambda_2)'(a_1\lambda_1 - A_2\lambda_2) = \inf_{\lambda=(\lambda_1,-\lambda_2')':||\lambda||=1}(a_1\lambda_1 + A_2\lambda_2)'(a_1\lambda_1 + A_2\lambda_2) = \inf_{\lambda=(\lambda_1,\lambda_2')':||\lambda||=1}(a_1\lambda_1 + A_2\lambda_2)'(a_1\lambda_1 + A_2\lambda_2) = \lambda_{\min}((a_1, A_2)'(a_1, A_2))$.

some fixed matrix π, then all θ values not equal θ_0 yield contiguous alternatives. For semi-strong identification in the standard sense, e.g., $\pi_n = \pi n^{-\delta}$ for some $\delta \in (0, 1/2)$ and some fixed full-column-rank matrix π, the contiguous alternatives have $\theta_n - \theta_0 = O(n^{-(1/2-\delta)})$. For joint weak identification, contiguity occurs when $\pi_n = (\pi_{1n}, ..., \pi_{pn}) \in R^{k\times p}$, $n^{1/2}||\pi_{jn}|| \to \infty$ for all $j \le p$, $\limsup_{n\to\infty} \lambda_{\min}(n\pi_n'\pi_n) < \infty$, and θ_n is such that $\pi_n(\theta_n - \theta_0) = O(n^{-1/2})$.

(iii) The proofs of Lemma 14.1 and Lemmas 14.2 and 14.3 below are given in Section 24 below.

14.3 SR-CQLR$_2$ Test

The components of the SR-QLR$_2$ statistic and its conditioning matrix are $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ and $n^{1/2}\widetilde{D}_n^*$ (see (5.2), (6.7), and (7.2)) when $\widehat{r}_n = k$, which holds wp$\to$1 under Assumption HLIV. Here we show that the conditioning statistic $n^{1/2}\widetilde{D}_n^*$ is asymptotically equivalent to Moreira's (2003) conditioning statistic $\overline{T}_n$ (in the homoskedastic linear IV model with fixed IV's) when $\pi_n \to 0^{k\times p}$. This includes the cases of standard weak identification and semi-strong identification. It is not asymptotically equivalent in other circumstances. (See Comment (ii) to Lemma 14.2 below.) Nevertheless, under strong and semi-strong IV's, the SR-CQLR$_2$ test and Moreira's CLR test are asymptotically equivalent.[62] In consequence, when $p = 1$, the SR-CQLR$_2$ test and Moreira's CLR test are asymptotically equivalent (because standard weak, strong, and semi-strong identification cover all possible cases). When $p \ge 2$, this is not true (because weak identification can occur even when $\pi_n \nrightarrow 0^{k\times p}$, if $n^{1/2}$ times the smallest singular value of π_n is $O(1)$). Although asymptotic equivalence of the tests fails in some cases when $p \ge 2$, the differences appear to be small because they are due only to the differences between fixed IV's and random IV's (which cause Σ_V to differ somewhat from Σ_{V*} defined below).

For $\pi \in R^{k\times p}$, define

$$\zeta_n(\pi) := n^{-1}\sum_{i=1}^n (\pi' \otimes Z_i)Z_iZ_i'(\pi \otimes Z_i') - \left(n^{-1}\sum_{i=1}^n (\pi' \otimes Z_i)Z_i\right)\left(n^{-1}\sum_{i=1}^n (\pi' \otimes Z_i)Z_i\right)' \in R^{kp\times kp}. \tag{14.1}$$

If $\lim n^{-1}\sum_{i=1}^n vec(Z_iZ_i')vec(Z_iZ_i')'$ exists, then $\zeta(\pi) := \lim \zeta_n(\pi)$ exists for all $\pi \in R^{k\times p}$. Define

$$R(\pi) := \Sigma_V \otimes K_Z + (B' \otimes I_k)\begin{bmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi) \end{bmatrix}(B \otimes I_k) \in R^{k(p+1)\times k(p+1)}, \tag{14.2}$$

[62] This holds because, under strong and semi-strong IV's, the SR-QLR$_2$ statistic and Moreira's CLR statistic behave asymptotically like LM statistics that project onto $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ (or equivalently, $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widehat{L}_n^{1/2}$) and $\overline{T}_n$, respectively, see Theorem 9.1 for the SR-QLR$_2$ statistic, and $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widehat{L}_n^{1/2}$ and $\overline{T}_n$ are asymptotically equivalent (up to multiplication by -1) by Lemma 14.1(f). Furthermore, the conditional critical values of the two tests both converge in probability to $\chi^2_{p,1-\alpha}$ under strong and semi-strong identification, see Theorem 9.1 for the SR-CQLR$_2$ critical value.

where $B = B(\theta_0)$ is defined in (6.3).

The probability limit of $\widetilde{\Sigma}_n$ is shown below to be the symmetric matrix $(b_0'\Sigma_V b_0)^{-1}\Sigma_{V*} \in R^{(p+1)\times(p+1)}$, where Σ_{V*} is defined as follows. The (j,ℓ) element of Σ_{V*} is

$$\Sigma_{V*j\ell} := tr(R_{j\ell}(\pi_*)'K_Z^{-1})/k, \tag{14.3}$$

where $R_{j\ell}(\pi_*)$ denotes the (j,ℓ) $k \times k$ submatrix of $R(\pi_*)$ for $j,\ell = 1,...,p+1$ and $\pi_* = \lim \pi_n$. Equivalently, Σ_{V*} is the unique minimizer of $||[I_{p+1} \otimes ((b_0'\Sigma_V b_0)^{-1/2}K_Z^{-1/2})][\Sigma \otimes K_Z - R(\pi_*)]$ $[I_{p+1} \otimes ((b_0'\Sigma_V b_0)^{-1/2}K_Z^{-1/2})]||$ over all symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$. Note that when $\zeta(\pi_*) = 0$ (as occurs when $\pi_* = 0^{k\times p}$), $\Sigma_{V*} = \Sigma_V$ (because $R(\pi_*) = \Sigma_V \otimes K_Z$ in this case).

We use the following assumption.

Assumption HLIV2: (a) $\lim n^{-1}\sum_{i=1}^n vec(Z_iZ_i')vec(Z_iZ_i')'$ exists and is finite,

(b) $\pi_n \to \pi_*$ for some $\pi_* \in R^{k\times p}$, and

(c) $\lambda_{\max}(\Sigma_{V*})/\lambda_{\min}(\Sigma_{V*}) \leq 1/\varepsilon$ for $\varepsilon > 0$ as in the definition of the SR-QLR$_2$ statistic.

Assumption HLIV2(c) implies that the eigenvalue adjustment to $\widetilde{\Sigma}_n$ employed in the SR-QLR$_2$ statistic has no effect asymptotically. One could analyze the behavior of the SR-CQLR$_2$ test when this condition is eliminated. This would not affect the asymptotic null rejection probabilities, but it would affect the form of the asymptotic distribution when the condition is violated. For brevity, we do not do so here.

The asymptotic behavior of $n^{1/2}\widetilde{D}_n^*$ is given in the following lemma. Under Assumption HLIV, $n^{1/2}\widetilde{D}_n^*$ equals the SR-CQLR$_2$ conditioning statistic $n^{1/2}\widetilde{D}_{An}^*$ wp→1 (because $\widehat{r}_n = k$ wp→1).

Lemma 14.2 *Suppose Assumptions* HLIV *and* HLIV2 *hold. Under the null hypothesis* $H_0 : \theta = \theta_0$ *and any* $p \geq 1$, *we have*

(a) $\widetilde{R}_n \to_p R(\pi_*)$,

(b) $\widetilde{\Sigma}_n \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$,

(c) $\widetilde{\Sigma}_n^\varepsilon \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$, *and*

(d) $n^{1/2}\widetilde{D}_n^* = -(I_k + o_p(1))\overline{T}_n(L_{V0}^{-1/2}L_{V*}^{1/2} + o_p(1)) + o_p(1)$, *where* $L_{V0} := (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' \in R^{p\times p}$ *and* $L_{V*} := (\theta_0, I_p)\Sigma_{V*}^{-1}(\theta_0, I_p)' \in R^{p\times p}$.

Comments: **(i)** If $\pi_* = 0^{k\times p}$, which occurs when all θ parameters are either weakly identified in the standard sense or semi-strongly identified, then $\zeta(\pi_*) = 0^{kp\times kp}$, $R(\pi_*) = \Sigma_V \otimes K_Z$, and $\Sigma_{V*} = \Sigma_V$. In this case, Lemma 14.2(d) yields

$$n^{1/2}\widetilde{D}_n^* = -(I_k + o_p(1))\overline{T}_n(I_p + o_p(1)) + o_p(1) \tag{14.4}$$

and $n^{1/2}\widetilde{D}_n^*$ is asymptotically equivalent to $\overline{T}_n$ (up to multiplication by -1).

(ii) On the other hand, if $\pi_* \neq 0^{k \times p}$, then $n^{1/2}\widetilde{D}_n^*$ is not asymptotically equivalent to $\overline{T}_n$ in general due to the $\zeta(\pi_*)$ factor that appears in the second summand of $R(\pi_*)$ in (14.2). This factor arises because the IV's are fixed in the linear IV model (by assumption), but the variance estimator $\widetilde{V}_n$, which appears in $\widetilde{R}_n$, see (7.1), and which determines $\widetilde{\Sigma}_n$ and Σ_{V*}, treats the IV's as though they are random.

14.4 Kleibergen's Nonlinear CLR Tests

This section analyzes the behavior of Kleibergen's (2005, 2007) nonlinear CLR tests in the homoskedastic linear IV regression model with $k \geq p$ fixed IV's. The behavior of Kleibergen's nonlinear CLR tests is found to depend on the choice of weighting matrix for the conditioning statistic. We find that when $p = 1$ (where p is the dimension of θ) and one employs the Jacobian-variance weighted conditioning statistic, Kleibergen's CLR test and conditioning statistics reduce asymptotically to those of Moreira's (2003) CLR test, as desired. This type of weighting has been suggested by Kleibergen's (2005, 2007) and Smith (2007). On the other hand, Kleibergen's CLR test and conditioning statistics do not reduce asymptotically to those of Moreira (2003) when $p = 1$ and one employs the moments-variance weighted conditioning statistic. The latter has been suggested by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012). Furthermore, the scale of the scalar conditioning statistic can differ from the desired value of one by a factor that can be arbitrarily close to zero or infinity (depending on the value of the reduced-form error matrix Σ_V and null hypothesis value θ_0). This has adverse effects on the power of the moment-variance weighted CLR test.

When $p \geq 2$, Kleibergen's nonlinear CLR tests depend on the form of a rank statistic. In this case, we find that no choice of rank statistic makes Kleibergen's CLR test statistic and conditioning statistic reduce asymptotically to those of Moreira (2003).

Kleibergen's test statistic takes the form:

$$CLR_n(\theta) := \frac{1}{2}\left(AR_n(\theta) - rk_n(\theta) + \sqrt{(AR_n(\theta) - rk_n(\theta))^2 + 4LM_n(\theta) \cdot rk_n(\theta)}\right), \text{ where}$$
$$LM_n(\theta) := n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1/2}(\theta)P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta) \quad (14.5)$$

and $rk_n(\theta)$ is a real-valued rank statistic, which is a conditioning statistic (i.e., the critical value may depend on $rk_n(\theta)$).

The critical value of Kleibergen's CLR test is $c(1-\alpha, rk_n(\theta))$, where $c(1-\alpha, r)$ is the $1-\alpha$

quantile of the distribution of

$$clr(r) := \frac{1}{2}\left(\chi_p^2 + \chi_{k-p}^2 - r + \sqrt{(\chi_p^2 + \chi_{k-p}^2 - r)^2 + 4\chi_p^2 r}\right) \tag{14.6}$$

for $0 \leq r < \infty$ and the chi-square random variables χ_p^2 and χ_{k-p}^2 in (14.6) are independent. The CLR test rejects the null hypothesis $H_0 : \theta = \theta_0$ if $CLR_n > c(1-\alpha, rk_n)$ (where, as elsewhere, the dependence of these statistics on θ_0 is suppressed for simplicity).

Kleibergen's CLR test depends on the choice of the rank statistic $rk_n(\theta)$. Kleibergen (2005, p. 1114, 2007, eqn. (37)) and Smith (2007, p. 7, footnote 4) propose to take $rk_n(\theta)$ to be a function of $\widetilde{V}_{Dn}^{-1/2}(\theta)vec(\widehat{D}_n(\theta))$, where $\widetilde{V}_{Dn}(\theta) \in R^{kp\times kp}$ is a consistent estimator of the covariance matrix of the asymptotic distribution of $vec(\widehat{D}_n(\theta))$ (after suitable normalization). We refer to $\widetilde{V}_{Dn}^{-1/2}(\theta)vec(\widehat{D}_n(\theta))$ as the orthogonalized sample Jacobian with Jacobian-variance weighting. In the i.i.d. case considered here, we have

$$\begin{aligned}\widetilde{V}_{Dn}(\theta) &:= n^{-1}\sum_{i=1}^{n} vec(G_i(\theta) - \widehat{G}_n(\theta))vec(G_i(\theta) - \widehat{G}_n(\theta))' - \widehat{\Gamma}_n(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{\Gamma}_n(\theta)', \text{ where} \\ \widehat{\Gamma}_n(\theta) &:= (\widehat{\Gamma}_{1n}(\theta)', ..., \widehat{\Gamma}_{pn}(\theta)')' \in R^{pk\times k}\end{aligned} \tag{14.7}$$

and $\widehat{\Gamma}_{1n}(\theta), ..., \widehat{\Gamma}_{pn}(\theta)$ are defined in (6.2).

Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) propose to take $rk_n(\theta)$ to be a function of $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$. We refer to $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$ as the orthogonalized sample Jacobian with moment-variance weighting. Below we consider both choices. For reasons that will become apparent, we treat the cases $p = 1$ and $p \geq 2$ separately.

14.5 p = 1 Case

Whether Kleibergen's nonlinear CLR test reduces asymptotically to Moreira's CLR test in the homoskedastic linear IV regression model depends on the rank statistic chosen. Here we consider the two choices of rank statistic that have been considered in the literature. We find that Kleibergen's nonlinear CLR test reduces asymptotically to Moreira's CLR test with a rank statistic based on $\widetilde{V}_{Dn}(\theta)$, but not with a rank statistic based on $\widehat{\Omega}_n(\theta)$. This illustrates that the flexibility in the choice of the rank statistic for Kleibergen's CLR test can have drawbacks. It may lead to a test that has reduced power.

When $p = 1$, some calculations (based on the closed-form expression for the minimum eigenvalue of a 2×2 matrix) show that

$$CLR_n(\theta) = AR_n(\theta) - \lambda_{\min}((n^{1/2}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), r_n(\theta))'(n^{1/2}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), r_n(\theta))) \text{ provided}$$
$$rk_n(\theta) = r_n(\theta)'r_n(\theta) \text{ for some random vector } r_n(\theta) \in R^k. \quad (14.8)$$

This equivalence is the origin of the $p = 1$ formula for the LR statistic in Moreira (2003). Hence, when $p = 1$, for testing $H_0 : \theta = \theta_0$, Kleibergen's test statistic with $rk_n(\theta) = r_n(\theta)'r_n(\theta)$ is of the same form as Moreira's (2003) LR statistic with $r_n(\theta_0)$ in place of $\overline{T}_n$ and with $n^{1/2}\widehat{\Omega}_n^{-1/2}(\theta_0)\widehat{g}_n(\theta_0)$ in place of $\overline{S}_n$, where θ_0 is the null value of θ.[63] The two choices for $rk_n(\theta)$ that we consider when $p = 1$ are

$$rk_{1n}(\theta) := n\widehat{D}_n(\theta)'\widetilde{V}_{Dn}^{-1}(\theta)\widehat{D}_n(\theta) \text{ and } rk_{2n}(\theta) := n\widehat{D}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{D}_n(\theta). \quad (14.9)$$

The statistic $rk_{1n}(\theta)$ has been proposed by Kleibergen (2005, 2007) and Smith (2007) and $rk_{2n}(\theta)$ has been proposed by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012).

Let

$$\zeta_n(\pi) := n^{-1}\sum_{i=1}^n Z_i Z_i'(Z_i'\pi)^2 - \left(n^{-1}\sum_{i=1}^n Z_i Z_i'\pi\right)\left(n^{-1}\sum_{\ell=1}^n Z_\ell Z_\ell'\pi\right)'. \quad (14.10)$$

This definition of $\zeta_n(\pi)$ is the same as in (14.1) when $p = 1$.

Lemma 14.3 *Suppose Assumption* HLIV *holds and $p = 1$. Under the null hypothesis $H_0 : \theta = \theta_0$, for any sequence of reduced-form parameters $\{\pi_n \in \Pi : n \geq 1\}$, we have*

(a) $rk_{1n}(\theta_0) = \overline{T}_n'[I_k + L_{V0}K_Z^{-1/2}\zeta_n(\pi_n)K_Z^{-1/2} + o_p(1)]^{-1}\overline{T}_n \cdot (1 + o_p(1)) + o_p(1),$

(b) $rk_{2n}(\theta_0) = \overline{T}_n'\overline{T}_n(L_{V0}b_0'\Sigma_V b_0)^{-1} \cdot (1 + o_p(1)) + o_p(1),$ *where* $L_{V0} := (\theta_0, 1)\Sigma_V^{-1}(\theta_0, 1)' \in R,$ *and*

(c) $L_{V0}b_0'\Sigma_V b_0 = \frac{(1-2\theta_0\rho c+\theta_0^2 c^2)^2}{c^2(1-\rho^2)}$, *where* $c^2 := Var(V_{2i})/Var(V_{1i}) > 0$ *and* $\rho = Corr(V_{1i}, V_{2i}) \in (-1, 1).$

Comments: **(i)** If $\pi_n \to 0$, then $\zeta_n(\pi_n) \to 0$ and Lemma 14.3(a) shows that $rk_{1n}(\theta_0)$ equals $\overline{T}_n'\overline{T}_n(1 + o_p(1)) + o_p(1)$. That is, under weak IV's and semi-strong IV's, $rk_{1n}(\theta_0)$ reduces asymptotically to Moreira's (2003) conditioning statistic. Under strong IV's, this does not occur. However, under strong IV's, we have $rk_{1n}(\theta_0) \to_p \infty$, just as $\overline{T}_n'\overline{T}_n \to_p \infty$. In consequence, the test constructed using $rk_{1n}(\theta_0)$ has the same asymptotic properties as Moreira's (2003) CLR test under the null and contiguous alternative distributions.

[63] The functional form of the rank statistics that have been considered in the literature, such as the statistics of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006) all reduce to the same function when $p = 1$. Specifically, $rk_n(\theta)$ equals the squared length of some k vector $r_n(\theta)$.

(ii) Simple calculations show that $\zeta_n(\pi_n)$ is positive semi-definite (psd). Hence, $rk_{1n}(\theta_0)$ is smaller than it would be if the second summand in the square brackets in Lemma 14.3(a) was zero.

(iii) Lemma 14.3(b) shows that the rank statistic $rk_{2n}(\theta_0)$ differs asymptotically from Moreira's conditioning statistic $\overline{T}_n'\overline{T}_n$ by the scale factor $(L_{V0}b_0'\Sigma_V b_0)^{-1}$. Thus, the nonlinear CLR test considered by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) does not reduce asymptotically to Moreira's (2003) CLR test in the homoskedastic linear IV regression model with fixed IV's under weak IV's. This has negative consequences for its power. Under strong or semi-strong IV's, this test does reduce asymptotically to Moreira's (2003) CLR test because $rk_{1n}(\theta_0) \to_p \infty$, just as $\overline{T}_n'\overline{T}_n \to_p \infty$, which is sufficient for asymptotic equivalence in these case.

(iv) For example, if $\rho = 0$ and $c = 1$ in Lemma 14.3(c), then $(L_{V0}b_0'\Sigma_V b_0)^{-1} = (1+\theta_0^2)^{-2} \leq 1$. In this case, if $|\theta_0| = 1$, then $(L_{V0}b_0'\Sigma_V b_0)^{-1} = 1/4$ and $rk_{2n}(\theta_0)$ is 1/4 as large as $\overline{T}_n'\overline{T}_n$ asymptotically. On the other hand, if $\rho = 0$ and $\theta_0 = 0$, then $(L_{V0}b_0'\Sigma_V b_0)^{-1} = c^2$, which can be arbitrarily close to zero or infinity depending on c.

(v) When $(L_{V0}b_0'\Sigma_V b_0)^{-1}$ is large (small), the $rk_{2n}(\theta_0)$ statistic is larger (smaller) than desired and it behaves as though the IV's are stronger (weaker) than they really are, which sacrifices power unless the IV's are quite strong (weak). Note that the inappropriate scale of $rk_{2n}(\theta_0)$ does not cause asymptotic size problems, only power reductions.

14.6 p $\geq$ 2 Case

When $p \geq 2$, Kleibergen's (2005) nonlinear CLR test does not reduce asymptotically to Moreira's (2003) CLR test for any choice of rank statistic $rk_n(\theta_0)$ for several reasons.

First, Moreira's (2003) LR statistic is given in (3.4), whereas Kleibergen's (2005) nonlinear LR statistic is defined in (14.5). By Lemma 14.1(e), $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = \overline{S}_n + o_p(1)$, where, here and below, we suppress the dependence of various quantities on θ_0. Hence, $AR_n = \overline{S}_n'\overline{S}_n + o_p(1)$. Even if rk_n takes the form $r_n'r_n$ for some random k vector r_n, it is not the case that

$$CLR_n = AR_n - \lambda_{\min}((n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, r_n)'(n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, r_n)) \qquad (14.11)$$

when $p \geq 2$. Hence, the functional form of Kleibergen's test statistic differs from that of Moreira's LR statistic when $p \geq 2$.

Second, for the rank statistics that have been suggested in the literature, viz., those of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006), rk_n is not of the form $r_n'r_n$, when $p \geq 2$.

Third, Moreira's conditioning statistic is the $k \times p$ matrix $\overline{T}_n$. Conditioning on this random matrix is equivalent asymptotically to conditioning on the $k \times p$ matrix $n^{1/2}\widehat{D}_n^*$ by Lemma 14.1(f). But, it is not equivalent asymptotically to conditioning on any of the *scalar* rank statistics considered in the literature when $p \geq 2$.

Fourth, if one weights the conditioning statistic in the way suggested by Kleibergen (2005) and Smith (2007), then the resulting CLR test is not guaranteed to have correct asymptotic size, see Section 5 of AG1. If one weights the conditioning statistic by $\widehat{\Omega}_n^{-1}$, as suggested by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012), then the CLR test is guaranteed to have correct asymptotic size under the conditions given in AG1, but the conditioning statistic is not asymptotically equivalent to Moreira's (2003) conditioning statistic and the difference can be substantial, see Lemma 14.3(b) and (c) for the $p = 1$ case.

15 Simulation Results for Kleibergen's MVW-CLR Test

This section presents finite-sample simulation results that show that Kleibergen's (2005) CLR test with moment-variance weighting (MVW-CLR) has low power in some scenarios in the homoskedastic linear IV model with normal errors, relative to the power of the SR-CQLR$_1$ and SR-CQLR$_2$ tests, Kleibergen's CLR test with Jacobian-variance weighting (JVW-CLR), and the CLR test of Moreira (2003) (Mor-CLR).[64] As noted at the beginning of Section 14.4, Lemma 14.3 and Comment (iv) following it show that the scale (denoted by *scale* below) of the moment-variance weighting conditioning statistic can be far from the optimal value of one.[65] We provide results for one scenario where *scale* is too large and one scenario where it is too small. These scenarios are chosen based on the formula given in Lemma 14.3.

The model is the homoskedastic normal linear IV model introduced in Section 3 with unknown error variance matrix Σ_V and $p = 1$. The IV's are fixed—they are generated once from a $N(0^k, I_k)$ distribution. The sample size n equals $1,000$. The hypotheses are $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$. The tests have nominal size .05. The power results are based on $40,000$ simulation repetitions and $1,000$ critical value repetitions and are size-corrected (by adding non-negative constants to the critical values of those tests that over-reject under the null). The reduced-form error variances and correlation are denoted by Σ_{V11}, Σ_{V22}, and ρ, respectively, and $\lambda := \pi' Z' Z \pi$. The number of IV's is k. The MVW-CLR and JVW-CLR tests employ the Robin and Smith (2000) rank statistic.

[64] The MVW-CLR and JVW-CLR tests denote Kleibergen's (2005) CLR test with the rank statistic given by the Robin and Smith (2000) statistics $rk_n = \lambda_{\min}(n\widehat{D}_n'\widehat{\Omega}_n^{-1/2}\widehat{D}_n)$ and $rk_n = \lambda_{\min}(n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n)$, respectively, where $\widehat{\Omega}_n$ and $\widehat{D}_n$ are defined in (5.1) and (6.2) with $\theta = \theta_0$ and $\widetilde{V}_{Dn}$ is an estimator of the asymptotic variance of $\widehat{D}_n$ (after suitable normalization) and is defined in (14.7). Note that the second formula for rk_n is appropriate only for the case $p = 1$, which is the case considered here. The estimators $\widehat{\Omega}_n$ and $\widetilde{V}_{Dn}$ are estimators of the asymptotic variances of the sample moments and Jacobian, respectively, which leads to the MVW and JVW terminology.

[65] The constant *scale* is the constant $(L_{V0} b_0' \Sigma_V b_0)^{-1}$ in Lemma 14.3(b) and (c).

Results are reported for the tests discussed above, as well as Kleibergen's LM test and the AR test.

Design 1 takes $\Sigma_{V11} = 1.0$, $\Sigma_{V22} = 4.0$, $\rho = 0.5$, $\pi = 0.044$, $\lambda = 2.009$, and $k = 5$. These parameter values yield $scale = 30.0$, which results in the MVW-CLR test behaving like Kleibergen's LM test even though the LM test has low power in this scenario. Design 2 takes $\Sigma_{V11} = 3.0$, $\Sigma_{V22} = 0.1$, $\rho = 0.95$, $\pi = 0.073$, $\lambda = 4.995$, and $k = 10$. These parameter values yield $scale = 0.0033$, which results in the MVW-CLR test behaving like the AR test even though the AR test has low power in this scenario.

The power functions of the tests are reported in Figure 1 (with $\theta\lambda^{1/2}$ on the horizontal axes with $\lambda^{1/2}$ fixed). Figure 1(a) shows that, for Design 1, the MVW-CLR and LM tests have very similar power functions and both are substantially below the power functions of the SR-CQLR$_1$, SR-CQLR$_2$, JVW-CLR, and Mor-CLR tests, which have essentially equal and optimal power. The AR test has high power, like that of the SR-CQLR$_1$, SR-CQLR$_2$, JVW-CLR, and Mor-CLR tests, for positive θ, and low power, like that of the MVW-CLR and LM tests, for negative θ.

Figure 1(b) shows that, for Design 2, the MVW-CLR and AR tests have similar power functions and both are substantially below the power functions of the SR-CQLR$_1$, SR-CQLR$_2$, JVW-CLR, Mor-CLR, and LM tests, which have essentially equal and optimal power.

16 Power Comparisons in Heteroskedastic/Autocorrelated Linear IV Models with $\mathbf{p = 1}$

In this section, we present some power comparisons for the AR test, Kleibergen's (2005) LM, JVW-CLR, and MVW-CLR tests, and the SR-CQLR$_2$ test introduced in AG2.[66] We also consider the plug-in conditional linear combination (PI-CLC) test introduced in I. Andrews (2014), as well as the MM1-SU and MM2-SU tests introduced in Moreira and Moreira (2013). The PI-CLC test aims to approximate the test that has minimum regret among conditional tests constructed using linear combinations of the LM and AR test statistics (with coefficients that depend on the conditioning statistic), see I. Andrews (2014) for details.[67] The MM1-SU and MM2-SU tests have optimal weighted average power for two different weight functions (over the alternative parameter values θ and the strength of identification parameter vector μ, given in (16.1) below) among tests that satisfy a sufficient condition for local unbiasedness.[68]

[66] See (5.2), (9.1), and a footnote in Section 15 for the definitions of AR test and Kleibergen's LM, MVW-CLR, and JVW-CLR tests. The AR test is called the S test in Stock and Wright (2000). The LM and JVW-CLR tests are denoted by K and QCLR, respectively, in I. Andrews (2014).

[67] The PI-CLC test does not possess an optimality property because it does not actually equal the minimum regret test.

[68] The weight functions considered depend on the variance parameters Σ_{gG} and Σ_{GG} in (16.1) below.

Figure 1(a): Design 1

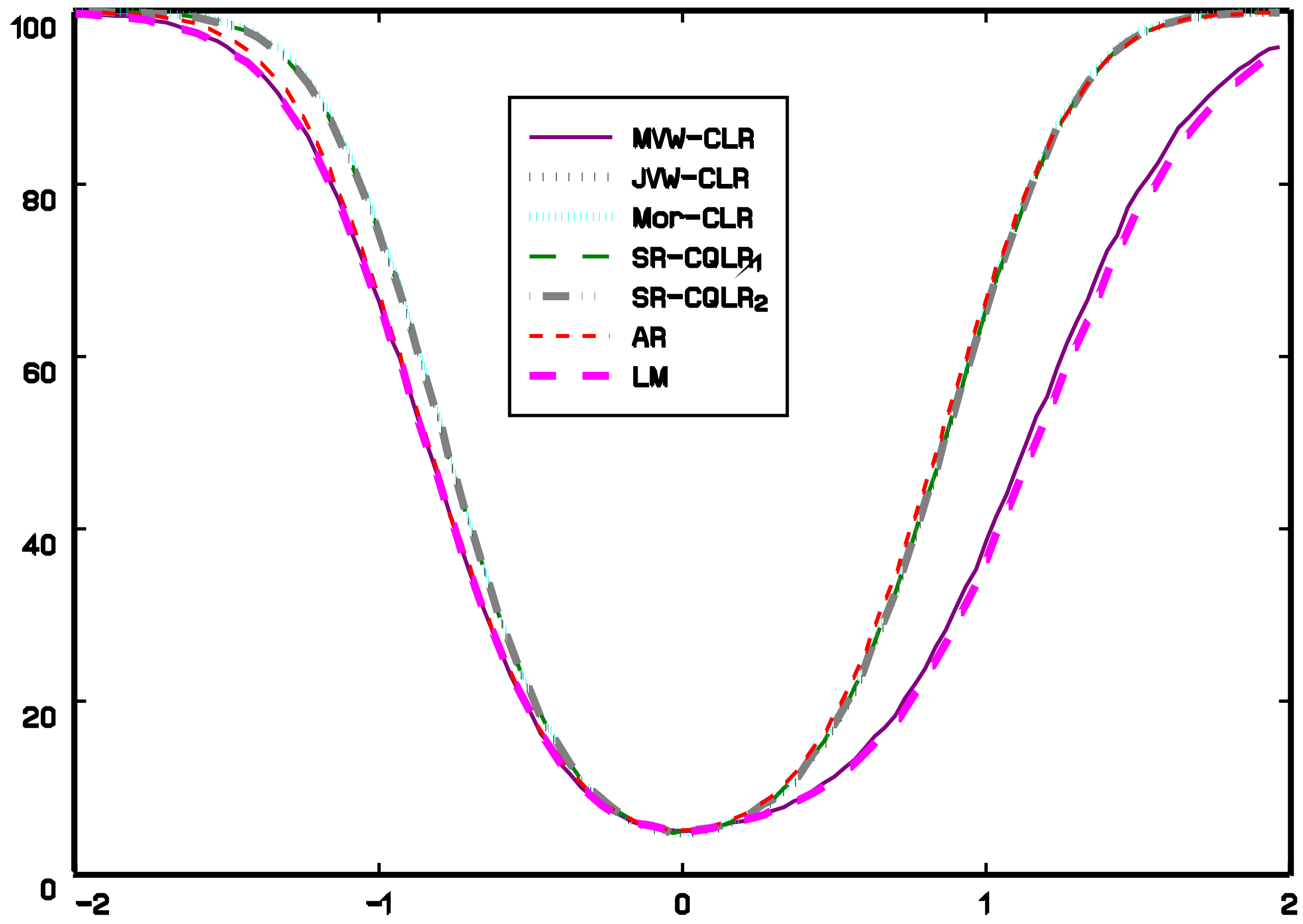
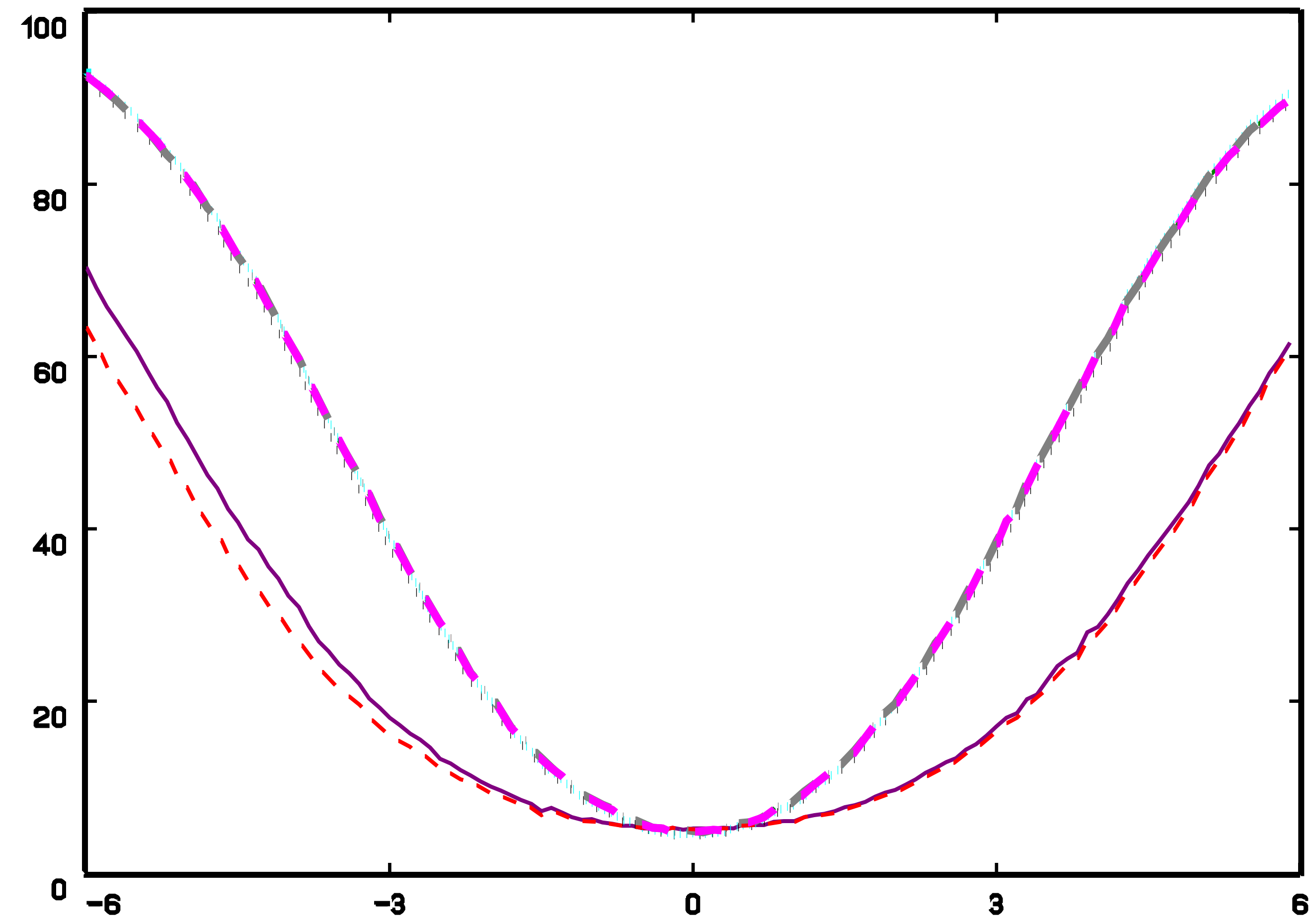


Figure 1(b): Design 2



We consider the same designs as in I. Andrews (2014, Sec. 6.2). These designs are for heteroskedastic and/or autocorrelated linear IV models with $p = 1$ and $k = 4$. The designs are calibrated to mimic the linear IV models for the elasticity of inter-temporal substitution estimated by Yogo (2004) for eleven countries using quarterly data from the early 1970's to the late 1990's. The power comparisons are for the limiting experiment under standard weak identification asymptotics. In consequence, for the simulations, the observations are drawn from the following model:

$$\begin{pmatrix} \widehat{\Omega}_n^{-1/2} n^{1/2} \widehat{g}_n(\theta_0) \\ \widehat{\Omega}_n^{-1/2} n^{1/2} \widehat{G}_n(\theta_0) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu\theta \\ \mu \end{pmatrix}, \begin{pmatrix} I_k & \Sigma_{gG} \\ \Sigma_{gG}' & \Sigma_{GG} \end{pmatrix} \right) \tag{16.1}$$

for $\theta \in R$, $\mu \in R^k$, and $\Sigma_{gG}, \Sigma_{GG} \in R^{k \times k}$, where Σ_{gG} and Σ_{GG} are assumed to be known.[69,70] The values of μ, Σ_{gG}, and Σ_{GG} are taken to be equal to the estimated values using the data from Yogo (2004).[71] A sample is a single observation from the distribution in (16.1) and the tests are constructed using the known values Σ_{gG} and Σ_{GG}.[72] The hypotheses are $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$.

Power is computed using $10,000$ simulation repetitions for the rejection probabilities, $10,000$ simulation repetitions for the data-dependent critical values of the MVW-CLR, JVW-CLR, and SR-CQLR$_2$ tests, and two million simulation repetitions for the critical values for the PI-CLC tests (which are taken from a look-up table that is simulated just one time).

Some details concerning the computation and definitions of the SR-CQLR$_2$, PI-CLC, MM1-SU, and MM2-SU tests are as follows. The SR-CQLR$_2$ test uses $\varepsilon = .05$, where ε appears in the definition of $\widetilde{L}_n(\theta)$ in (7.2) of AG2. For the PI-CLC test, the number of values "a" considered in the search over $[0, 1]$ is 100, the number of simulation repetitions used to determine the best choice of "a" is 2000, and the number of alternative parameter values considered in the search for the best "a" is 41. For the MM1-SU and MM2-SU tests, the number of variables in the discretization of maximization problem is 1000, the number of points used in the numerical approximations of the integrals $h1$ and $h2$ that appear in the definitions of these tests is 1000, and when approximating integrals $h1$ and $h2$ by sums of 1000 rectangles these rectangles cover $[-4, 4]$.

[69] In linear IV models with i.i.d. observations, the matrix Σ_{gG} is necessarily symmetric. However, with autocorrelation, it need not be. In the eleven countries considered here, it is not.

[70] The variance matrix in the limit experiment varies slightly depending on whether one treats the IV's as fixed or random. For example, the asymptotic variance of $n^{1/2}\widehat{G}_n(\theta_0)$ under standard weak IV asymptotics varies slightly in these two cases. Power results for the SR-CQLR$_1$ test when the limiting variance is computed using fixed IV's are equivalent to those computed for the SR-CQLR$_2$ test for the case where the limiting variance is computed using random IV's. In consequence, we do not separately report power results for the SR-CQLR$_1$ test.

[71] See I. Andrews (2014, Appendices D.3 and D.4) for details on the calculations of the simulation designs based on Yogo's (2004) data, as well as for details on the computation of I. Andrews' PI test, referred to here as PI-CLC, and the two tests of Moreira and Moreira (2013), referred to here and in I. Andrews (2014) as MM1-SU and MM2-SU. The JVW-CLR and LM tests here are the same as the QCLR and K tests, respectively, in I. Andrews (2014).

[72] For example, $\widehat{\Gamma}_{jn}(\theta_0)$ in (6.2) is taken to be known and equal to Σ_{gG}', and $\widetilde{V}_n(\theta_0)$ in (7.1) is taken to be known and equal to the variance matrix in (16.1).

Table IV. Shortfalls in Average-Power (×100)

Country	$\mu'\mu$	non-Kron	SR-CQLR	JVW	MVW	PI-CLC	MM1	MM2	LM	AR
Australia	138	17	.0	.1	.1	.2	2.4	.1	.1	6.9
Canada	48	5	.0	.0	.2	.0	1.4	.5	.3	6.8
France	79	6	.1	.2	.0	.3	.7	.3	.0	8.0
Germany	10	3	.0	.1	.4	.0	.2	.1	2.3	6.5
Italy	84	15	.5	1.1	2.0	.2	1.1	.0	2.6	5.5
Japan	17	14	3.3	3.2	8.9	.4	.0	2.4	17.4	.6
Netherlands	25	3	.0	.2	.1	.2	.9	.5	1.6	6.6
Sweden	174	9	.3	.2	.3	.2	1.5	.0	.3	7.5
Switzerland	31	4	.1	.0	.0	.4	1.3	1.1	.5	7.2
U. K.	53	38	.7	6.0	5.4	.8	2.5	.0	7.8	3.8
U.S.	81	10	.8	2.0	2.9	.0	7.3	.8	3.5	3.2
Average	over	Countries	.5	1.2	1.8	.2	1.8	.5	3.3	5.7

The asymptotic power functions are given in Figure 2. Each graph is based on 41 equi-spaced values on the x axis covering $[-6, 6]$. The x axis variable is the parameter θ scaled by a fixed value of $||\mu||$ for a given country, thus $\theta||\mu|| \in [-6, 6]$, where θ is the alternative parameter value (when $\theta \neq 0$) defined in (16.1) of AG2 and μ is the mean vector that determines the strength of identification. The y axis variable is power ×100.

Table IV provides the *shortfall in average-power* (×100) of each test for each country relative to the other seven tests considered, where average power is an unweighted average over the 40 alternative parameter values. Table V provides the *maximum power shortfall* (×100) of each test for each country relative to the other seven tests considered, where the maximum is taken over the 40 alternative parameter values.[73] The shortfall in average-power is an unweighted average power criterion, whereas the maximum power shortfall is a minimax regret criterion.

The last row of Table IV shows the average (across countries) of the shortfall in average-power (×100) of each test. This provides a summary measure. Similarly, the last row of Table V shows the average (across countries) of the maximum power shortfall (×100) of each test.

The second and third columns of Table IV provide the concentration parameter, $\mu'\mu$, which measures of the strength of identification, and a non-Kronecker index, abbreviated by non-Kron, which measures the deviation of the variance matrix in (16.1), call it Ψ, from a Kronecker matrix.

[73] More precisely, let AP_{tc} denote the average power of test t for country c, where the average is taken over the 40 parameter values in the alternative hypothesis. By definition, the *shortfall in average-power* of test t for country c is $\max_{s \leq 8} AP_{sc} - AP_{tc}$, where the maximum is taken over the eight tests considered.

Let $P_{tc}(\theta)$ denote the power of test t in country c against the alternative θ. By definition, the power shortfall of test t in country c for alternative θ is $\max_{s \leq 8} P_{sc}(\theta) - P_{tc}(\theta)$ and the *maximum power shortfall* of test t in country c is $\max_{\theta \in \Theta_{40}}(\max_{s \leq 8} P_{sc}(\theta) - P_{tc}(\theta))$, where Θ_{40} contains the 40 alternative parameter values considered.

Note that, as defined, the shortfall in average-power is not equal to the average of the power shortfalls over $\theta \in \Theta_{40}$.

Figure 2(a): Australia

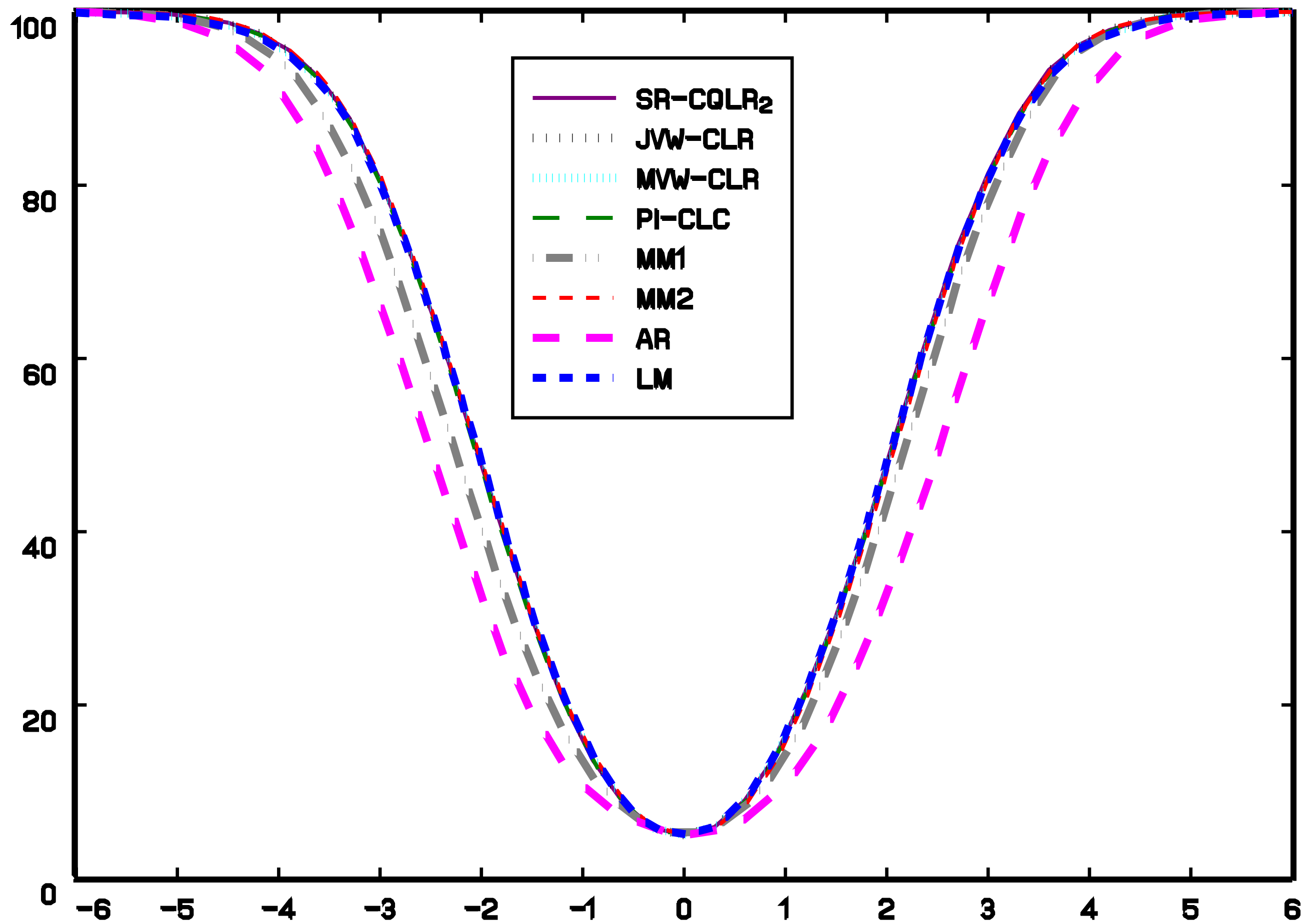


Figure 2(b): Canada

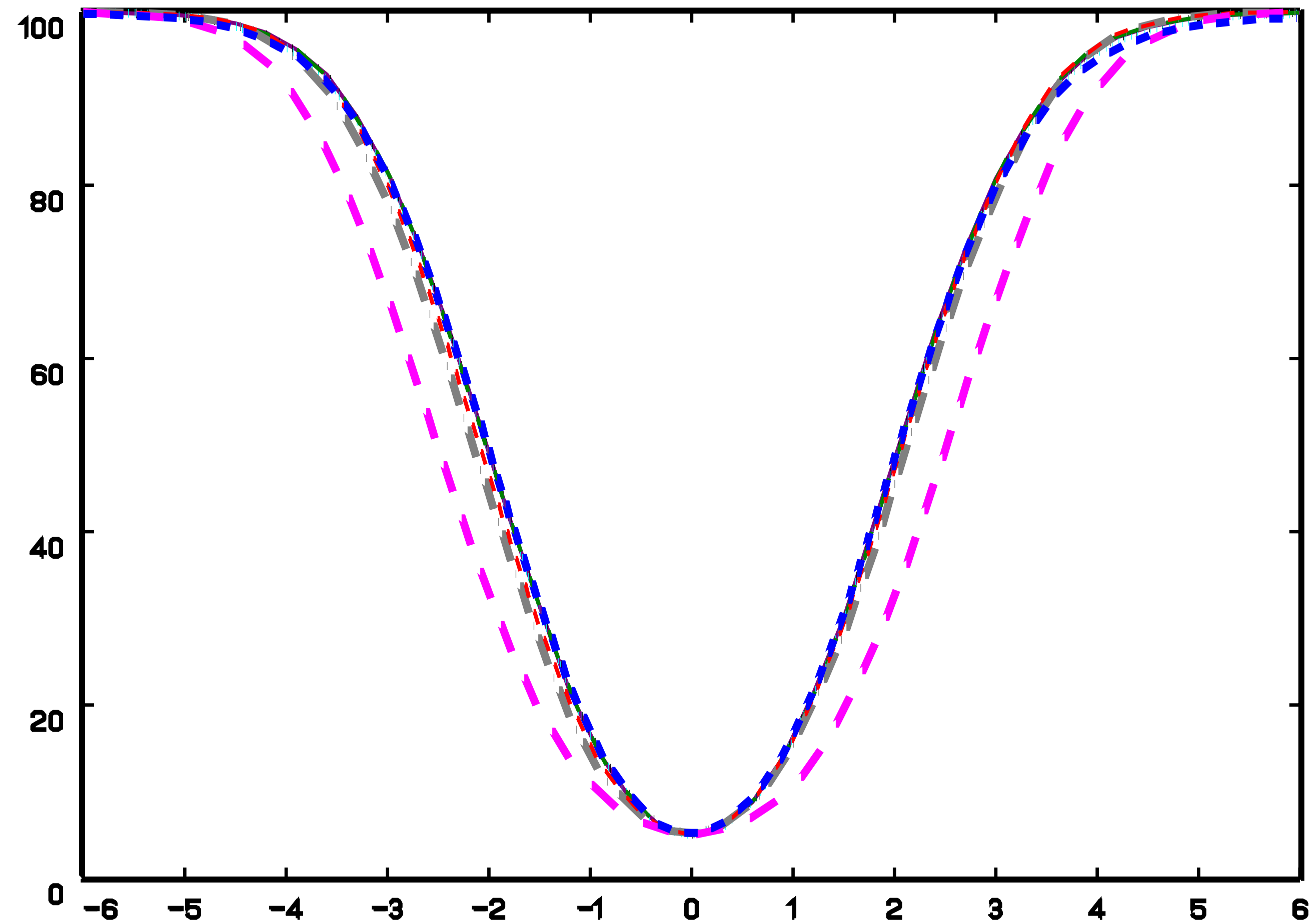


Figure 2(c): France

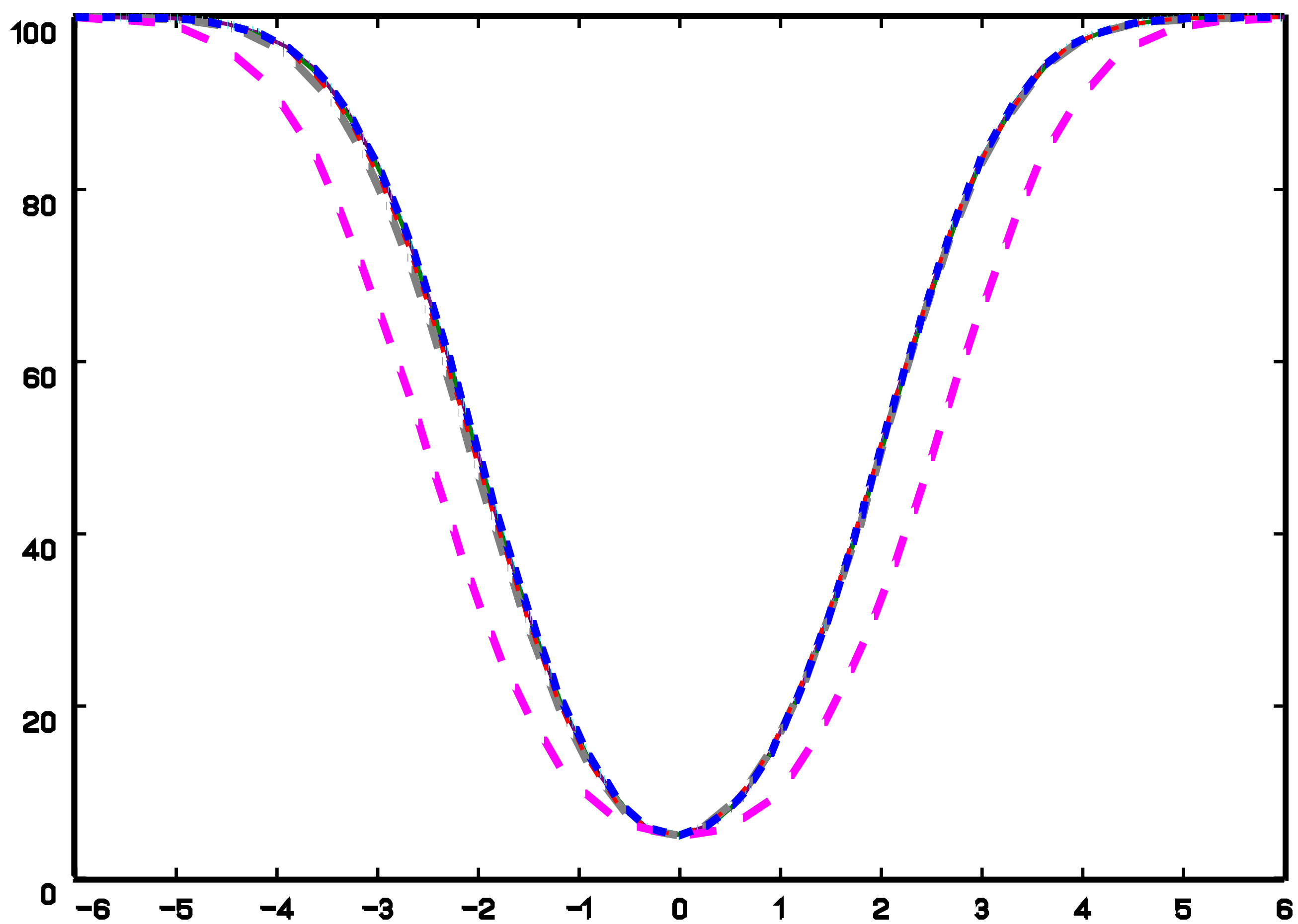


Figure 2(d): Germany

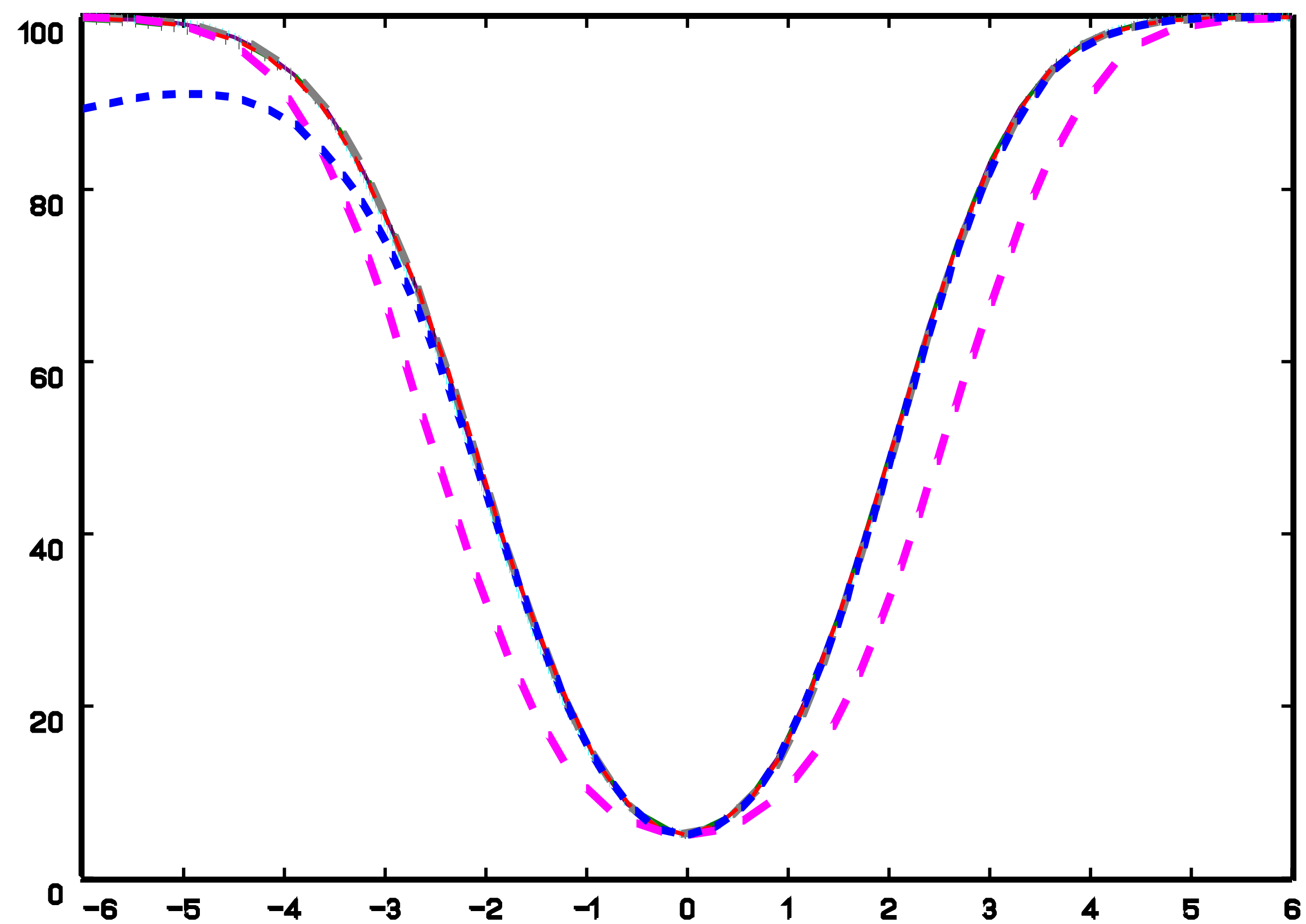


Figure 2(e): Italy

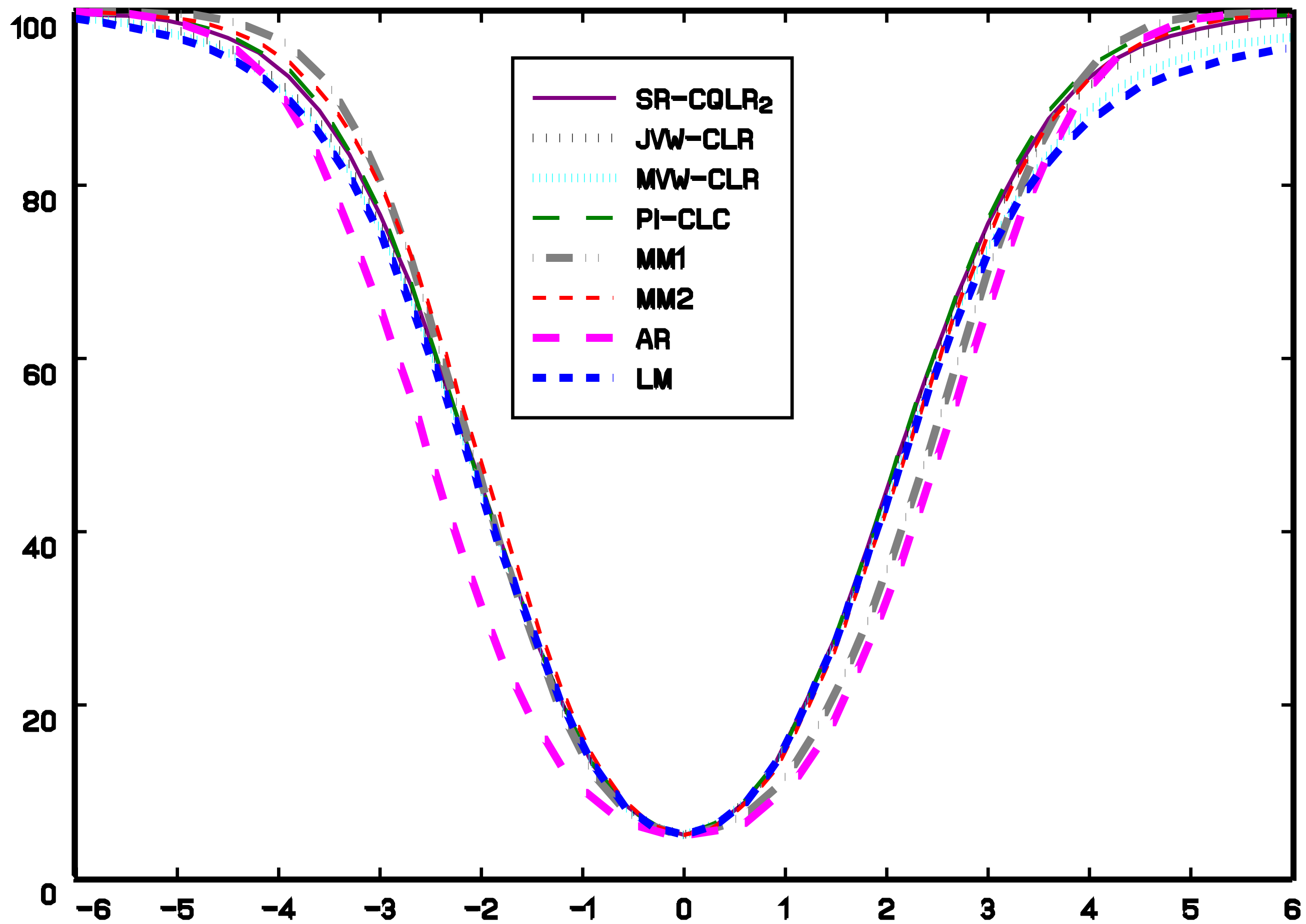


Figure 2(f): Japan

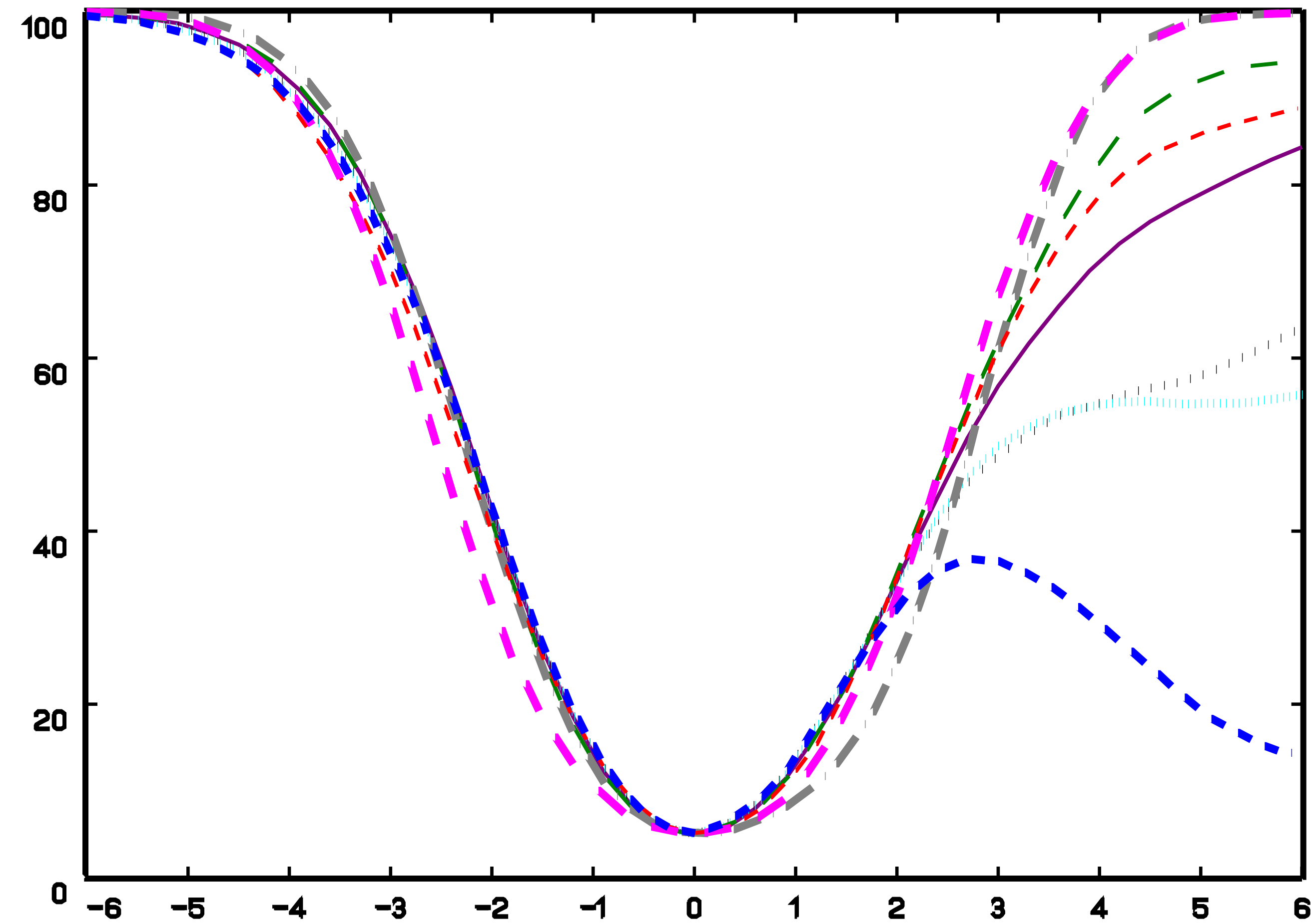


Figure 2(g): Netherlands

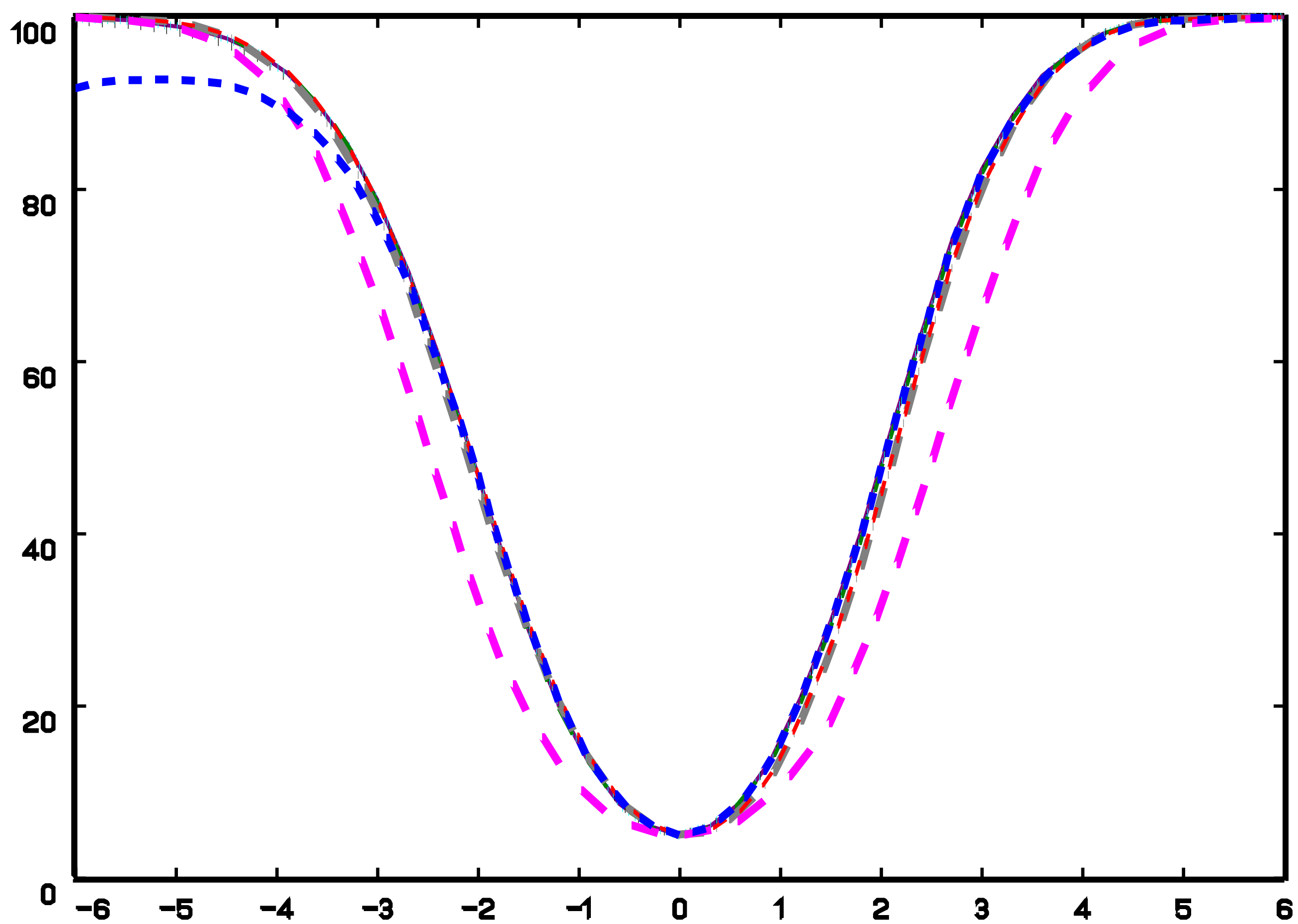


Figure 2(h): Sweden

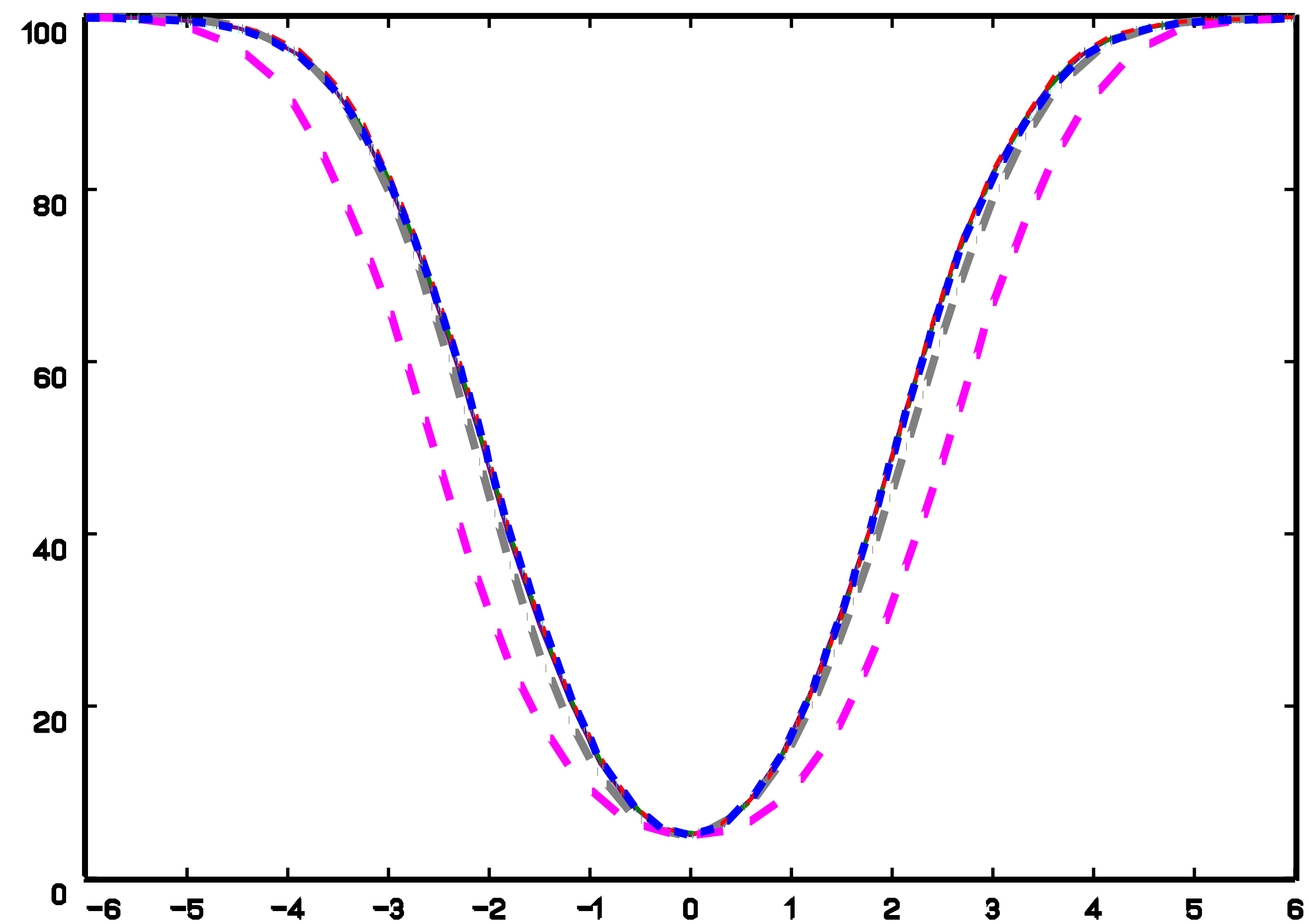


Figure 2(i): Switzerland

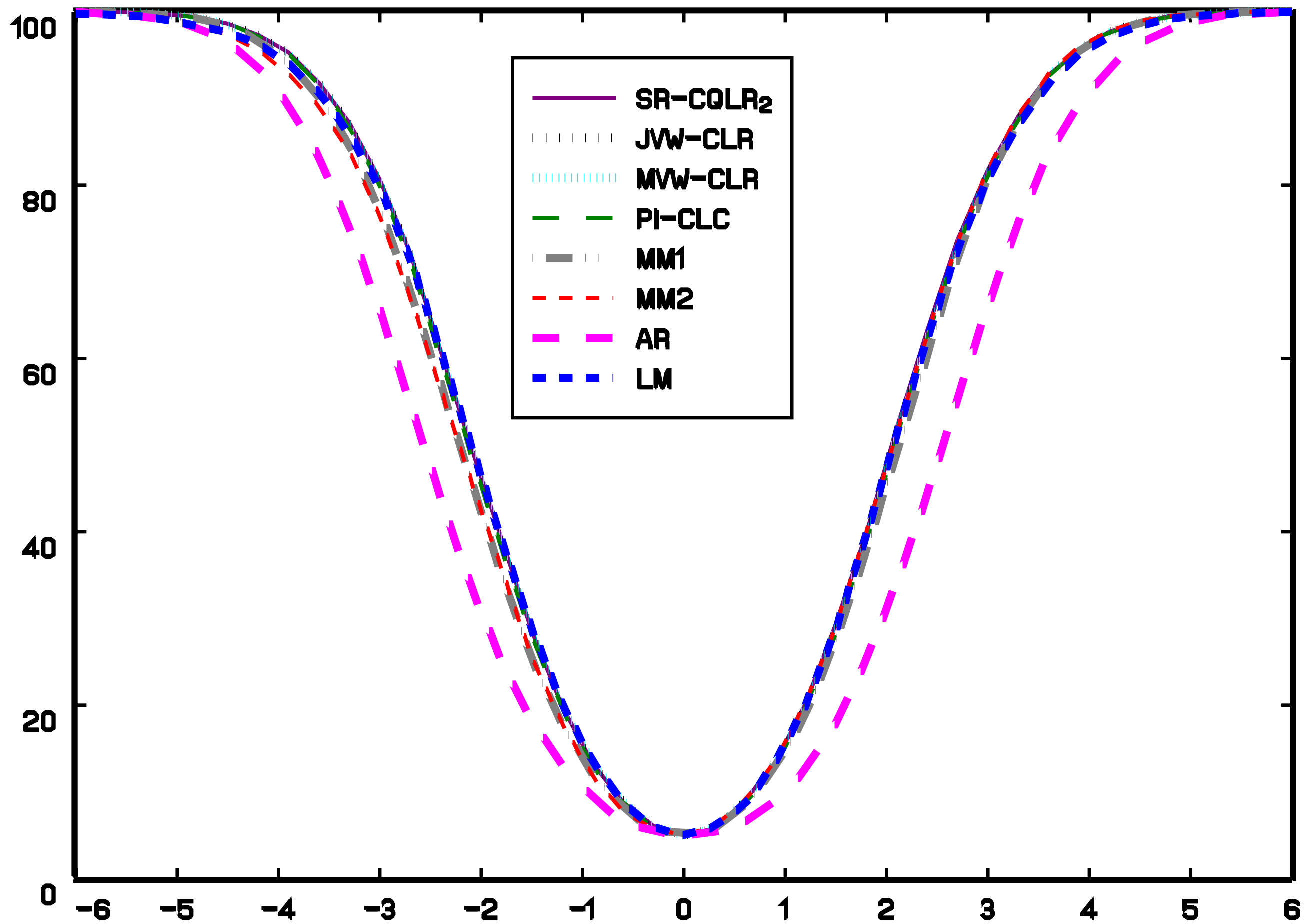


Figure 2(j): United Kingdom

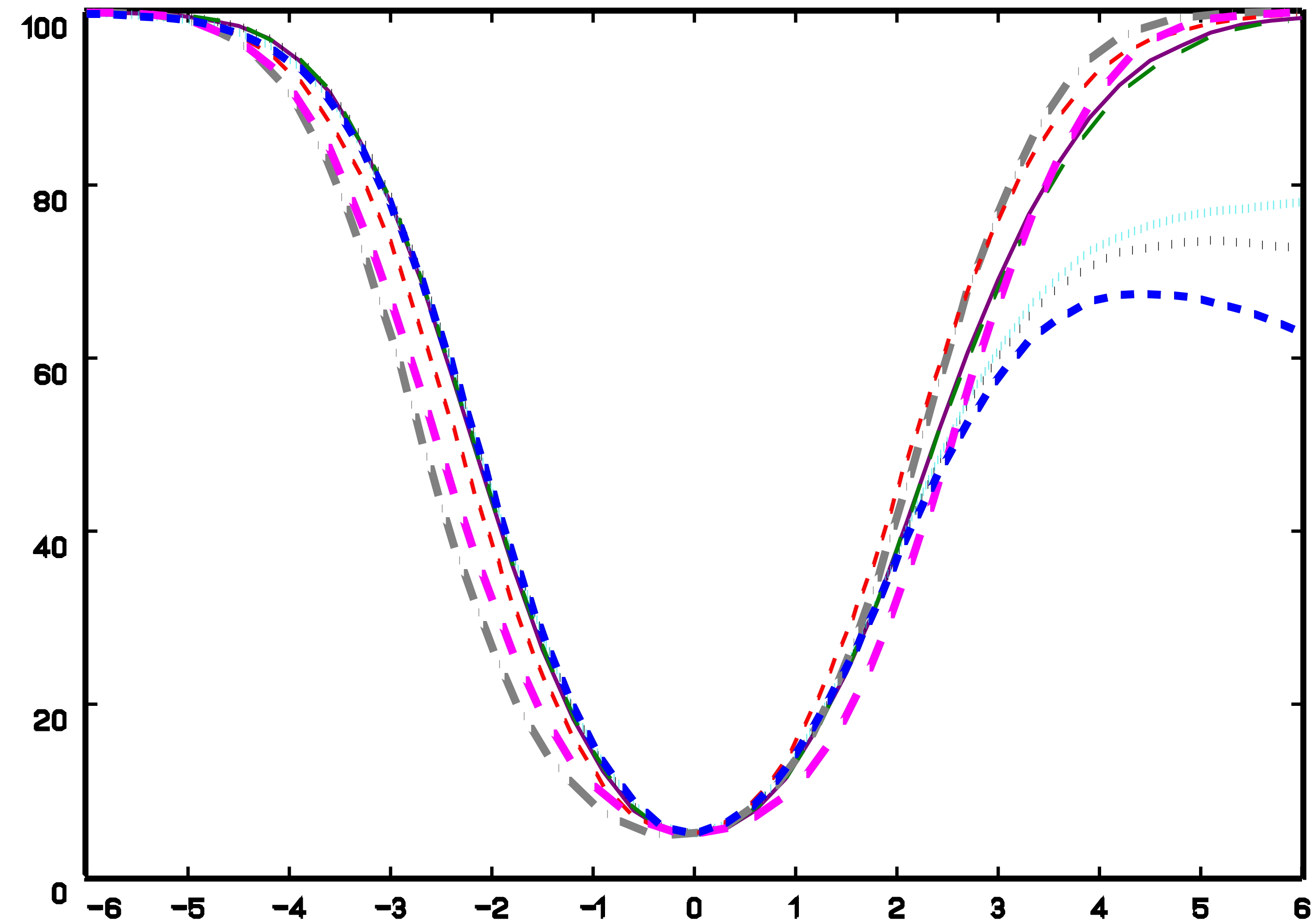


Figure 2(k): United States

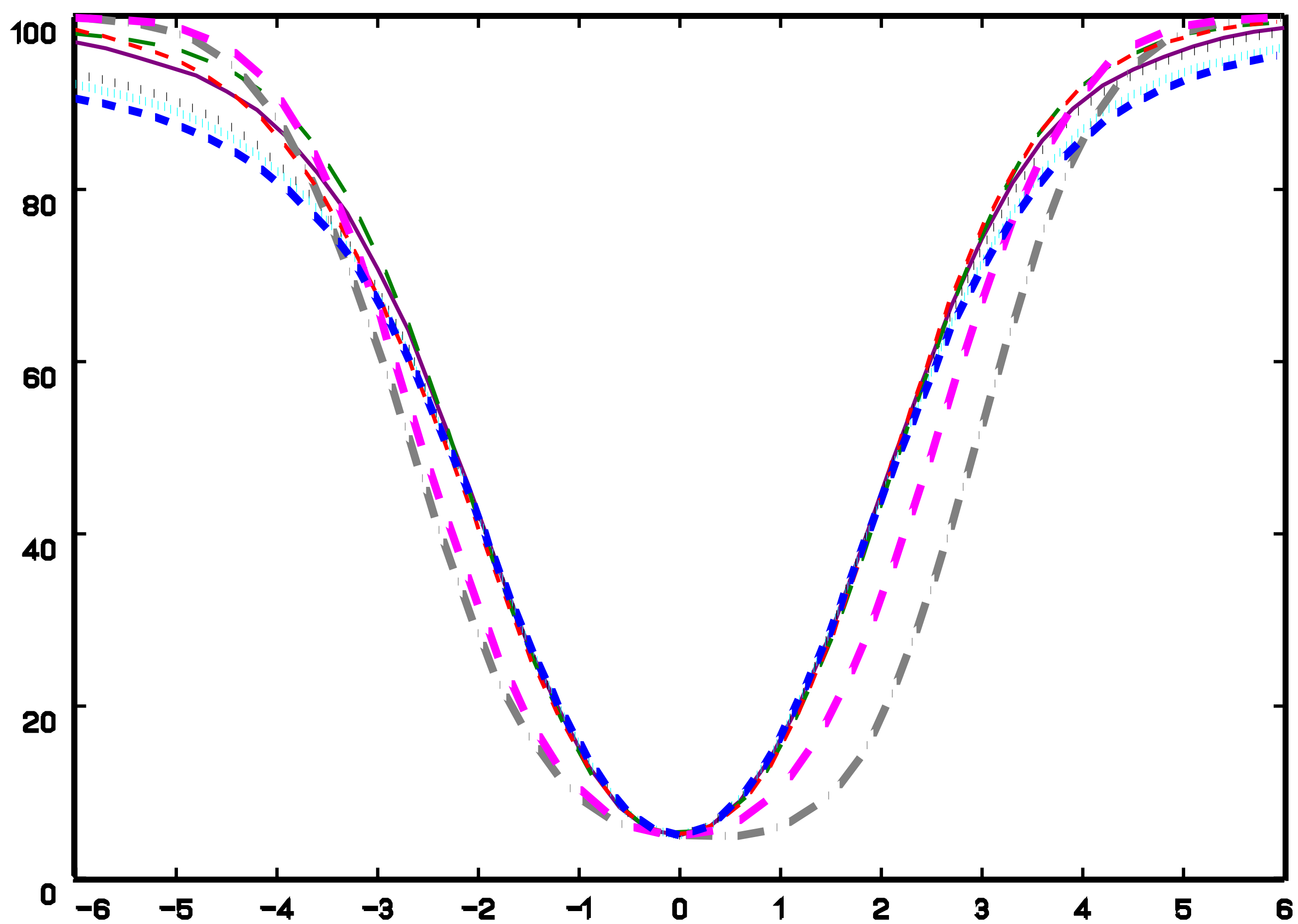


Table V. Maximum Power Shortfalls ($\times 100$)

Country	$\mu'\mu$	non-Kron	SR-CQLR	JVW	MVW	PI	MM1	MM2	LM	AR
Australia	138	17	.5	.6	.8	1.0	8.2	1.3	.9	17.2
Canada	48	5	.6	.5	.9	.7	5.4	3.0	1.7	17.7
France	79	6	.7	.8	.5	1.0	3.0	1.6	.4	19.9
Germany	10	3	.8	.8	2.2	.6	1.0	.8	10.6	18.4
Italy	84	15	4.4	5.7	6.5	3.9	9.7	2.3	7.1	17.7
Japan	17	14	21.3	41.4	44.9	8.6	10.1	13.6	85.8	11.9
Netherlands	25	3	.9	1.1	.9	1.4	3.9	3.3	8.2	18.6
Sweden	174	9	1.0	.6	1.0	.7	4.9	.4	1.1	19.6
Switzerland	31	4	.5	.3	.5	1.6	4.8	5.5	1.4	18.8
U. K.	53	38	8.4	27.3	23.2	9.0	20.6	7.1	37.0	14.7
U.S.	81	10	5.2	9.0	10.2	2.6	27.7	5.1	11.7	12.4
Average	over	Countries	4.0	8.0	8.3	2.8	9.0	4.0	14.9	17.0

This deviation is given by the formula $1,000 \times \min_{B,C} ||B \otimes C - \Psi||$, where the minimum is taken over symmetric pd matrices B and C of dimensions 2×2 and 4×4, respectively, $|| \cdot ||$ denotes the Frobenius norm, and the rescaling by $1,000$ is for convenience.[74] Germany, Japan, and the Netherlands exhibit the weakest identification, while Sweden and Australia exhibit the strongest. The U.K., Australia, Italy, and Japan have variance matrices that are farthest from Kronecker-product form, while Germany, the Netherlands, and Switzerland have variance matrices that are closest to Kronecker-product form.

The test that performs best in Tables IV and V is the PI-CLC test, followed by the SR-CQLR$_2$ and MM2-SU tests. The difference between these tests is not large. For example, the difference in the average (across countries) shortfall in average-power (not rescaled by multiplication by 100 in contrast to the results in Table IV) of the PI-CLC test and the SR-CQLR$_2$ and MM2-SU tests is .003. This small power advantage is almost entirely due to the relative performances for Japan, which exhibits very weak identification and moderately large non-Kronecker index.

The remaining tests in decreasing order of power (in an overall sense) are the JVW-CLR, MVW-CLR, MM1-SU, LM, and AR tests. Not surprisingly, the LM and AR tests have noticeably lower power than the other tests in an overall sense, and the AR test has noticeably lower power than the LM test.

We conclude that the SR-CQLR$_2$ test has asymptotic power that is competitive with, or better than, that of other tests in the literature for the particular parameters considered here in the particular model considered here. The SR-CQLR$_2$ test has advantages compared to the PI-CLC,

[74] The non-Kronecker index is computed using the Framework 2 method given in Section 4 of Van Loan and Pitsianis (1993) with symmetry of C imposed by replacing $\widehat{A}_{ij}$ by $(\widehat{A}_{ij} + \widehat{A}_{ji})/2$ in equation (9) of that paper.

MM1-SU, and MM2-SU tests of (i) being applicable in almost any moment condition model, whereas the latter tests are not, (ii) being easy to implement (i.e., program), and (iii) being fast to compute.

17 Eigenvalue-Adjustment Procedure

Eigenvalue adjustments are made to two sample matrices that appear in the two SR-CQLR test statistics. These adjustments guarantee that the adjusted sample matrices have minimum eigenvalues that are not too close to zero even if the corresponding population matrices are singular or near singular. These adjustments improve the asymptotic and finite-sample performance of the tests by improving their robustness to singularities or near singularities.

The eigenvalue-adjustment procedure can be applied to any non-zero positive semi-definite (psd) matrix $H \in R^{d_H \times d_H}$ for some positive integer d_H. Let ε be a positive constant. Let $A_H \Lambda_H A_H'$ be a spectral decomposition of H, where $\Lambda_H = Diag\{\lambda_{H1}, ..., \lambda_{Hd_H}\} \in R^{d_H \times d_H}$ is the diagonal matrix of eigenvalues of H with nonnegative nonincreasing diagonal elements and A_H is a corresponding orthogonal matrix of eigenvectors of H. The eigenvalue-adjusted matrix $H^\varepsilon \in R^{d_H \times d_H}$ is

$$H^\varepsilon := A_H \Lambda_H^\varepsilon A_H', \text{ where } \Lambda_H^\varepsilon := Diag\{\max\{\lambda_{H1}, \lambda_{\max}(H)\varepsilon\}, ..., \max\{\lambda_{Hd_H}, \lambda_{\max}(H)\varepsilon\}\}. \quad (17.1)$$

We have $\lambda_{\max}(H) = \lambda_{H1}$, and $\lambda_{\max}(H) > 0$ provided the psd matrix H is non-zero.

The following lemma provides some useful properties of this eigenvalue adjustment procedure.

Lemma 17.1 *Let d_H be a positive integer, let ε be a positive constant, and let $H \in R^{d_H \times d_H}$ be a non-zero positive semi-definite non-random matrix. Then,*

(a) (*uniqueness*) H^ε, *defined in* (17.1), *is uniquely defined.* (*That is, every choice of spectral decomposition of H yields the same matrix H^ε*),

(b) (*eigenvalue lower bound*) $\lambda_{\min}(H^\varepsilon) \geq \lambda_{\max}(H)\varepsilon$,

(c) (*condition number upper bound*) $\lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) \leq \max\{1/\varepsilon, 1\}$,

(d) (*scale equivariance*) *For all* $c > 0$, $(cH)^\varepsilon = cH^\varepsilon$, *and*

(e) (*continuity*) $H_n^\varepsilon \to H^\varepsilon$ *for any sequence of psd matrices* $\{H_n \in R^{d_H \times d_H} : n \geq 1\}$ *that satisfies* $H_n \to H$.

Comments: **(i)** The lower bound $\lambda_{\max}(H)\varepsilon$ for $\lambda_{\min}(H^\varepsilon)$ given in Lemma 17.1(b) is positive provided $H \neq 0^{d_H \times d_H}$.

(ii) Lemma 17.1(c) shows that one can choose ε to control the condition number of H^ε. The latter is a common measure of how ill-conditioned a matrix is. If $\varepsilon \leq 1$, which is a typical choice, then the upper bound is $1/\varepsilon$. Note that $H^\varepsilon = H$ iff $\lambda_{\min}(H) \geq \lambda_{\max}(H)\varepsilon$ iff the condition number of H is less than or equal to $1/\varepsilon$.

(iii) Scale equivariance of $(\cdot)^\varepsilon$ established in Lemma 17.1(d) is an important property. For example, one does not want the choice of measurements in \$ or \$1,000 to affect inference.

(iv) Continuity of $(\cdot)^\varepsilon$ established in Lemma 17.1(e) is an important property because it implies that for random matrices $\{\widehat{H}_n : n \geq 1\}$ for which $\widehat{H}_n \rightarrow_p H$, one has $\widehat{H}_n^\varepsilon \rightarrow_p H^\varepsilon$.

Proof of Lemma 17.1. For notational simplicity, we drop the H subscript on A_H, Λ_H , and Λ_H^ε . We prove part (a) first. The eigenvectors of H^ε $(= A\Lambda^\varepsilon A')$ defined in (6.6) are unique up to the choice of vectors that span the eigenspace that corresponds to any eigenvalue. Suppose the $j, ..., j+d$ eigenvalues of H are equal for some $d \geq 0$ and $1 \leq j < d_H$. We can write $A = (A_1, A_2, A_3)$, where $A_1 \in R^{d_H \times (j-1)}$, $A_2 \in R^{d_H \times (d+1)}$, and $A_3 \in R^{d_H \times (d_H - j - d)}$. In addition, H can be written as $H = A_* \Lambda A_*'$, where $A_* = (A_1, A_{2*}, A_3)$, the column space of A_{2*} equals that of A_2, and A_* is an orthogonal matrix. As above, $H^\varepsilon = A\Lambda^\varepsilon A'$. To establish part (a), if suffices to show that $H^\varepsilon = A_* \Lambda^\varepsilon A_*'$, or equivalently, $A\Lambda^\varepsilon A'\xi = A_* \Lambda^\varepsilon A_*' \xi$ for any $\xi \in R^{d_H}$.

For any $\xi \in R^{d_H}$, we can write $\xi = \xi_1 + \xi_2$, where ξ_1 belongs to the column space of A_2 (and A_{2*}) and ξ_2 is orthogonal to this column space. We have

$$
\begin{aligned}
A\Lambda^\varepsilon A'\xi &= A\Lambda^\varepsilon (A_1, A_2, A_3)'(\xi_1 + \xi_2) \\
&= A\Lambda^\varepsilon (0^{j-1\prime}, (A_2'\xi_1)', 0^{d_H - j - d\prime})' + A\Lambda^\varepsilon ((A_1'\xi_2)', 0^{d+1\prime}, (A_3'\xi_2)')' \\
&= A\lambda_j^\varepsilon (0^{j-1\prime}, (A_2'\xi_1)', 0^{d_H - j - d\prime})' + (A_1, A_2, A_3)\Lambda^\varepsilon ((A_1'\xi_2)', 0^{d+1\prime}, (A_3'\xi_2)')' \\
&= A_2 A_2' \xi_1 \lambda_j^\varepsilon + (A_1, A_3)\Lambda_-^\varepsilon ((A_1'\xi_2)', (A_3'\xi_2)')' \\
&= A_{2*} A_{2*}' \xi_1 \lambda_j^\varepsilon + (A_1, A_3)\Lambda_-^\varepsilon ((A_1'\xi_2)', (A_3'\xi_2)')' \\
&= A_* \Lambda^\varepsilon A_*' \xi, \qquad (17.2)
\end{aligned}
$$

where $\Lambda_-^\varepsilon \in R^{(d_H - d - 1) \times (d_H - d - 1)}$ is the diagonal matrix equal to Λ^ε with its $j, ..., j+d$ rows and columns deleted, $\lambda_j^\varepsilon = \max\{\lambda_j, \lambda_{\max}(H)\varepsilon\}$, λ_j is the jth eigenvalue of Λ, the second equality uses $A_1'\xi_1 = 0^{j-1}$, $A_3'\xi_1 = 0^{d_H - j - d}$, and $A_2'\xi_2 = 0^{d+1}$, the third equality holds because $\lambda_j = ... = \lambda_{j+d}$ implies that $\lambda_j^\varepsilon = ... = \lambda_{j+d}^\varepsilon$, the fourth equality holds using the definition of Λ_-^ε, the fifth equality holds because $A_2 A_2' = A_{2*} A_{2*}'$ (since both equal the projection matrix onto the column space of A_2 (and A_{2*})), and the last equality holds by reversing the steps in the previous equalities with $A_* = (A_1, A_{2*}, A_3)$ in place of $A = (A_1, A_2, A_3)$. Because (17.2) holds for any matrix A_{2*} defined as above and any feasible j and d, part (a) holds.

To prove parts (b) and (c), we note that the eigenvalues of H^ε are $\{\max\{\lambda_{Hj}, \lambda_{\max}(H)\varepsilon\}$:

$j = 1, ..., d_H\}$ because $H^\varepsilon = A\Lambda^\varepsilon A'$ and A is an orthogonal matrix. In consequence, $\lambda_{\min}(H^\varepsilon) \geq \lambda_{\max}(H)\varepsilon$, which establishes part (b). If $\lambda_{\min}(H) > \lambda_{\max}(H)\varepsilon$, then $H^\varepsilon = H$, $\lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H)/\lambda_{\min}(H) < 1/\varepsilon$, and the result of part (c) holds. Alternatively, if $\lambda_{\min}(H) \leq \lambda_{\max}(H)\varepsilon$, then $\lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H)\varepsilon$. In addition, we have $\lambda_{\max}(H^\varepsilon) = \max\{\lambda_{H1}, \lambda_{\max}(H)\varepsilon\} = \lambda_{\max}(H) \times \max\{1, \varepsilon\}$ using $\lambda_{H1} = \lambda_{\max}(H)$. Combining these two results gives $\lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H) \max\{1, \varepsilon\}/(\lambda_{\max}(H)\varepsilon) = \max\{1/\varepsilon, 1\}$, where the second equality uses the assumption that H is non-zero, which implies that $\lambda_{\max}(H) > 0$. This gives the result of part (c).

We now prove part (d) and for clarity make the H subscripts on A_H and Λ_H explicit in this paragraph. We have $\Lambda_{cH} = c\Lambda_H$ and we can take $A_{cH} = A_H$ by the definition of eigenvalues and eigenvectors. This implies that $\Lambda^\varepsilon_{cH} = c\Lambda^\varepsilon_H$ (using the definition of Λ^ε_H in (6.6)) and $(cH)^\varepsilon = A_{cH}\Lambda^\varepsilon_{cH}A'_{cH} = cA_H\Lambda^\varepsilon_H A'_H = cH^\varepsilon$, which establishes part (d).

Now we prove part (e). Let $A_n\Lambda_n A'_n$ be a spectral decomposition of H_n for $n \geq 1$. Let $H^\varepsilon_n = A_n\Lambda^\varepsilon_n A'_n$ for $n \geq 1$, where Λ^ε_n is the diagonal matrix with jth diagonal element given by $\lambda^\varepsilon_{nj} = \max\{\lambda_{nj}, \lambda_{\max}(H_n)\varepsilon\}$ and λ_{nj} is the jth largest eigenvalue of H_n. (By part (a) of the Lemma, H^ε_n is invariant to the choice of eigenvector matrix A_n used in its definition.)

Given any subsequence $\{n_\ell\}$ of $\{n\}$, let $\{n_m\}$ be a subsubsequence such that $A_{n_m} \to A$ for some orthogonal matrix A that may depend on the subsubsequence $\{n_m\}$. (Such a subsubsequence exists because the set of orthogonal $d_H \times d_H$ matrices is compact.) By assumption, $H_n \to H$. This implies that $\Lambda_n \to \Lambda$, where Λ is the diagonal matrix of eigenvalues of H in nonincreasing order (by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). In turn, this gives $\Lambda^\varepsilon_n \to \Lambda^\varepsilon$, where Λ^ε is the diagonal matrix with jth diagonal element given by $\lambda^\varepsilon_j = \max\{\lambda_j, \lambda_{\max}(H)\varepsilon\}$ and λ_j is the jth largest eigenvalue of H, because $\lambda_{\max}(\cdot)$ is a continuous function (by Elsner's Theorem again). The previous results imply that $H_{n_m} = A_{n_m}\Lambda_{n_m}A'_{n_m} \to A\Lambda A'$, $H = A\Lambda A'$, $H^\varepsilon_{n_m} = A_{n_m}\Lambda^\varepsilon_{n_m}A'_{n_m} \to A\Lambda^\varepsilon A'$, and $A\Lambda^\varepsilon A' = H^\varepsilon$. Because every subsequence $\{n_\ell\}$ of $\{n\}$ has a subsubsequence $\{n_m\}$ for which $H^\varepsilon_{n_m} \to H^\varepsilon$, we obtain $H^\varepsilon_n \to H^\varepsilon$, which completes the proof of part (e). $\square$

18 Singularity-Robust LM Test

SR-LM versions of Kleibergen's LM test and CS can be defined analogously to the SR-AR and SR-CQLR tests and CS's. However, these procedures are only partially singularity robust, see the discussion below. In addition, LM tests have low power in some circumstances under weak identification.

The SR-LM test statistic is

$$SR\text{-}LM_n(\theta) := n\widehat{g}_{An}(\theta)' P_{\widehat{\Omega}_{An}^{-1/2}(\theta)\widehat{D}_{An}(\theta)} \widehat{g}_{An}(\theta), \tag{18.1}$$

where P_M denotes the projection matrix onto the column space of the matrix M. For testing $H_0 : \theta = \theta_0$, the SR-LM test rejects the null hypothesis if

$$SR\text{-}LM_n(\theta_0) > \chi^2_{\min\{\widehat{r}_n(\theta_0),p\},1-\alpha}, \tag{18.2}$$

where $\chi^2_{\min\{\widehat{r}_n(\theta_0),p\},1-\alpha}$ denotes the $1-\alpha$ quantile of a chi-squared distribution with $\min\{\widehat{r}_n(\theta_0), p\}$ degrees of freedom. This test can be shown to have correct asymptotic size and to be asymptotically similar for the parameter space $\mathcal{F}^{SR}_{LM}$, which is a generalization of the parameter space $\mathcal{F}_0$ in AG1 and has a similar (rather complicated) form to $\mathcal{F}_0$. It is defined as follows: for some $\delta_1 > 0$,

$$\begin{aligned}
\mathcal{F}^{SR}_{LM} &:= \cup_{j=0}^{\min\{r_F,p\}} \mathcal{F}^{SR}_{LMj}, \text{ where} \\
\mathcal{F}^{SR}_{LMj} &:= \{F \in \mathcal{F}^{SR}_2 : \tau^*_{jF} \geq \delta_1 \text{ and } \lambda_{p-j}\left(\Psi_F^{C^{*\prime}_{F,k-j} G^*_i B^*_{F,p-j}\xi}\right) \geq \delta_1 \ \forall \xi \in R^{p-j} \text{ with } ||\xi|| = 1\}, \\
G^*_i &:= \Pi_{1F}^{-1/2} A'_F G_i \in R^{r_F \times p}, \ r_F := rk(\Omega_F), \ g^*_i := \Pi_{1F}^{-1/2} A'_F g_i \in R^{r_F}, \\
\Psi_F^{a_i} &:= E_F a_i a_i' - E_F a_i g_i^{*\prime} (E_F g^*_i g^*_i)^{-1} E_F g^*_i a_i' \text{ for any random vector } a_i,
\end{aligned} \tag{18.3}$$

τ^*_{jF} is the jth largest singular value of $E_F G^*_i$ for $j = 1, ..., \min\{r_F, p\}$, $\tau^*_{0F} := \delta_1$, B^*_F is a $p \times p$ orthogonal matrix of eigenvalues of $(E_F G^*_i)'(E_F G^*_i)$ ordered so that the corresponding eigenvalues $(\kappa^*_{1F}, ..., \kappa^*_{pF})$ are nonincreasing, C^*_F is an $r_F \times r_F$ orthogonal matrix of eigenvalues of $(E_F G^*_i)(E_F G^*_i)'$ ordered so that the corresponding eigenvalues $(\kappa^*_{1F}, ..., \kappa^*_{r_F F})$ are nonincreasing, $B^*_F := (B^*_{F,j}, B^*_{F,p-j})$ for $B^*_{F,j} \in R^{p\times j}$ and $B^*_{F,k-j} \in R^{p\times(p-j)}$, and $C^*_F := (C^*_{F,j}, C^*_{F,k-j})$ for $C^*_{F,j} \in R^{r_F \times j}$ and $C^*_{F,k-j} \in R^{r_F \times (r_F - j)}$.[75,76] See Section 3 of AG1 for a discussion of the form of this parameter space and the quantities upon which it depends. Note that $\Psi_F^{a_i}$ is the expected outer-product matrix of the vector of residuals, $a_i - E_F a_i g_i^{*\prime}(E_F g^*_i g^*_i)^{-1} g^*_i$, from the $L^2(F)$ projections of a_i onto the space spanned by the components of g^*_i, see AG1 for further discussion.

The conditions in $\mathcal{F}^{SR}_{LM}$ (beyond those in $\mathcal{F}^{SR}_2$) are used to guarantee that the conditioning matrix $\widehat{D}_{An} \in R^{\widehat{r}_n \times p}$ has full rank $\min\{\widehat{r}_n, p\}$ asymptotically with probability one (after pre- and post-multiplication by suitable matrices). AG1 shows that these conditions are not redundant.

[75] The first $\min\{r_F, p\}$ eigenvalues of $(E_F G^*_i)'(E_F G^*_i)$ and $(E_F G^*_i)(E_F G^*_i)'$ are the same. If $r_F > p$, the remaining $r_F - p$ eigenvalues of $(E_F G^*_i)(E_F G^*_i)'$ are all zeros. If $r_F < p$, the remaining $p - r_F$ eigenvalues of $(E_F G^*_i)'(E_F G^*_i)$ are all zeros.

[76] The matrices B^*_F and C^*_F are not necessarily uniquely defined. But, this is not of consequence because the $\lambda_{p-j}(\cdot)$ condition is invariant to the choice of B^*_F and C^*_F.

Given the need for these conditions, the SR-LM test is not fully singularity robust. The asymptotic size and similarity result for the SR-LM test stated above can be proved using Theorem 4.1 of AG1 combined with the argument given in Section 10.2 below. For brevity, we do not provide the details. Extensions of the asymptotic size and similarity results to SR-LM CS's are analogous to those for the SR-AR and SR-CQLR CS's.

A theoretical advantage of the SR-AR and SR-CQLR tests and CS's considered in this paper, relative to tests and CS's that make use of the LM statistic, is that they avoid the complicated conditions that appear in $\mathcal{F}_{LM}^{SR}$.

19 Proofs of Lemmas 6.1 and 6.2

Lemma 6.1 of AG2. *Let D be a $k \times p$ matrix with the singular value decomposition $D = C\Upsilon B'$, where C is a $k \times k$ orthogonal matrix of eigenvectors of DD', B is a $p \times p$ orthogonal matrix of eigenvectors of $D'D$, and Υ is the $k \times p$ matrix with the $\min\{k,p\}$ singular values $\{\tau_j : j \leq \min\{k,p\}\}$ of D as its first $\min\{k,p\}$ diagonal elements and zeros elsewhere, where τ_j is nonincreasing in j. Then, $c_{k,p}(D, 1-\alpha) = c_{k,p}(\Upsilon, 1-\alpha)$.*

Proof of Lemma 6.1. Define

$$B^+ := \begin{bmatrix} B & 0^p \\ 0^{p\prime} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}. \tag{19.1}$$

The matrix B^+ is orthogonal because B is, where B is as in the statement of the lemma. The eigenvalues of $(D,Z)'(D,Z)$ are solutions $\{\kappa_j : j \leq p+1\}$ to

$$\begin{aligned} |(D,Z)'(D,Z) - \kappa I_{p+1}| &= 0 \text{ or} \\ |B^{+\prime}(D,Z)'(D,Z)B^+ - \kappa I_{p+1}| &= 0 \text{ or} \\ |(DB,Z)'(DB,Z) - \kappa I_{p+1}| &= 0, \text{ or} \\ |(C\Upsilon,Z)'CC'(C\Upsilon,Z) - \kappa I_{p+1}| &= 0, \text{ or,} \\ |(\Upsilon,Z^*)'(\Upsilon,Z^*) - \kappa I_{p+1}| &= 0, \text{ where } Z^* := C'Z \sim N(0^k, I_k), \end{aligned} \tag{19.2}$$

the equivalence of the first and second lines holds because $|A_1A_2| = |A_1| \cdot |A_2|$, $|B^+| = 1$, and $B^{+\prime}B^+ = I_{p+1}$, the equivalence of the second and third lines holds by matrix algebra, the equivalence of the third and fourth lines holds because $DB = C\Upsilon B'B = C\Upsilon$ and $CC' = I_k$, and the equivalence of the last two lines holds by $CC' = I_k$ and the definition of Z^*. Equation (19.2) implies that $\lambda_{\min}((D,Z)'(D,Z))$ equals $\lambda_{\min}((\Upsilon,Z^*)'(\Upsilon,Z^*))$. In addition, $Z'Z = Z^{*\prime}Z^*$. Hence,[77]

$$CLR_{k,p}(D) = Z'Z - \lambda_{\min}((D,Z)'(D,Z)) = Z^{*\prime}Z^* - \lambda_{\min}((\Upsilon,Z^*)'(\Upsilon,Z^*)). \tag{19.3}$$

Since Z and Z^* have the same distribution, $CLR_{k,p}(D)$ $(= Z^{*\prime}Z^* - \lambda_{\min}((\Upsilon,Z^*)'(\Upsilon,Z^*)))$ and $CLR_{k,p}(\Upsilon) := Z'Z - \lambda_{\min}((\Upsilon,Z)'(\Upsilon,Z))$ have the same distribution and the same $1-\alpha$ quantile. That is, $c_{k,p}(D,1-a) = c_{k,p}(\Upsilon,1-\alpha)$. $\square$

Lemma 6.2 of AG2. *The statistics QLR_{1n}, $c_{k,p}(n^{1/2}\widehat{D}_n^*,1-\alpha)$, $\widehat{D}_n^{*\prime}\widehat{D}_n^*$, AR_n, $\widehat{u}_{in}^*$, $\widehat{\Sigma}_n$, and $\widehat{L}_n$ are invariant to the transformation $(Z_i,u_i^*) \rightsquigarrow (MZ_i,u_i^*)$ for any $k\times k$ nonsingular matrix M. This transformation induces the following transformations: $g_i \rightsquigarrow Mg_i$, $G_i \rightsquigarrow MG_i$, $\widehat{g}_n \rightsquigarrow M\widehat{g}_n$, $\widehat{G}_n \rightsquigarrow M\widehat{G}_n$, $\widehat{\Omega}_n \rightsquigarrow M\widehat{\Omega}_nM'$, $\widehat{\Gamma}_{jn} \rightsquigarrow M\widehat{\Gamma}_{jn}M'$, $\widehat{D}_n \rightsquigarrow M\widehat{D}_n$, $Z_{n\times k} \rightsquigarrow Z_{n\times k}M'$, $\widehat{\Xi}_n \rightsquigarrow {M'}^{-1}\widehat{\Xi}_n$, $\widehat{V}_n \rightsquigarrow (I_{p+1}\otimes M)\widehat{V}_n(I_{p+1}\otimes M')$, and $\widehat{R}_n \rightsquigarrow (I_{p+1}\otimes M)\widehat{R}_n(I_{p+1}\otimes M')$.*

Proof of Lemma 6.2. We will refer to the results of the Lemma for $g_i, G_i, ..., \widehat{R}_n$ as equivariance results. The equivariance results are immediate for g_i, G_i, $\widehat{g}_n$, $\widehat{G}_n$, $\widehat{\Omega}_n$, $\widehat{\Gamma}_{jn}$, and $Z_{n\times k}$. For $\widehat{D}_n = (\widehat{D}_{1n},...,\widehat{D}_{pn})$, we have

$$\widehat{D}_{jn} := \widehat{G}_{jn} - \widehat{\Gamma}_{jn}\widehat{\Omega}_n^{-1}\widehat{g}_n \rightsquigarrow M\widehat{G}_{jn} - M\widehat{\Gamma}_{jn}M'(M\widehat{\Omega}_nM')^{-1}M\widehat{g}_n = M\widehat{D}_{jn} \tag{19.4}$$

for $j=1,...,p$. We have $\widehat{\Xi}_n := (Z_{n\times k}'Z_{n\times k})^{-1}Z_{n\times k}'U^* \rightsquigarrow (MZ_{n\times k}'Z_{n\times k}M')^{-1}MZ_{n\times k}'U^* = {M'}^{-1}\widehat{\Xi}_n$. We have $\widehat{u}_{in}^* := \widehat{\Xi}_n'Z_i \rightsquigarrow ({M'}^{-1}\widehat{\Xi}_n)'MZ_i = \widehat{u}_{in}^*$. We have $\widehat{V}_n := n^{-1}\sum_{i=1}^n[(u_i^*-\widehat{u}_{in}^*) \times (u_i^*-\widehat{u}_{in}^*)'\otimes Z_iZ_i'] \rightsquigarrow n^{-1}\sum_{i=1}^n[(u_i^*-\widehat{u}_{in}^*)(u_i^*-\widehat{u}_{in}^*)'\otimes MZ_iZ_i'M'] = (I_{p+1}\otimes M)\widehat{V}_n(I_{p+1}\otimes M')$ using the invariance of $\widehat{u}_{in}^*$. We have $\widehat{R}_n := (B'\otimes I_k)\widehat{V}_n(B\otimes I_k) \rightsquigarrow (B'\otimes M)\widehat{V}_n(B\otimes M') = (I_{p+1}\otimes M)\widehat{R}_n(I_{p+1}\otimes M')$ using the equivariance result for $\widehat{V}_n$.

We have $\widehat{\Sigma}_{j\ell n} := tr(\widehat{R}_{j\ell n}'\widehat{\Omega}_n^{-1})/k \rightsquigarrow tr((M\widehat{R}_{j\ell n}M')'(M\widehat{\Omega}_nM')^{-1})/k = tr(M\widehat{R}_{j\ell n}'M'M'^{-1}\widehat{\Omega}_n^{-1} \times M^{-1})/k = \widehat{\Sigma}_{j\ell n}$ for $j,\ell = 1,...,p+1$ using the equivariance result for $\widehat{R}_n$. We have $\widehat{L}_n := (\theta,I_p)(\widehat{\Sigma}_n^{\varepsilon})^{-1}(\theta,I_p)' \rightsquigarrow \widehat{L}_n$ using the invariance result for $\widehat{\Sigma}_n$. We have $\widehat{D}_n^{*\prime}\widehat{D}_n^* := \widehat{L}_n^{1/2}\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n\widehat{L}_n^{1/2} \rightsquigarrow \widehat{L}_n^{1/2}\widehat{D}_n'M'(M\widehat{\Omega}_nM')^{-1}M\widehat{D}_n\widehat{L}_n^{1/2} = \widehat{D}_n^{*\prime}\widehat{D}_n^*$. This implies that $c_{k,p}(n^{1/2}\widehat{D}_n^*,1-\alpha) \rightsquigarrow c_{k,p}(n^{1/2}\widehat{D}_n^*, 1-\alpha)$ because $c_{k,p}(n^{1/2}\widehat{D}_n^*,1-\alpha)$ only depends on $\widehat{D}_n^*$ through $\widehat{D}_n^{*\prime}\widehat{D}_n^*$ by the Comment to Lemma 6.1.

[77] The quantity $CLR_{k,p}(D)$ is written in terms of (D,Z) in (19.3), whereas it is written in terms of (Z,D) in (3.5). Both expressions give the same value.

We have $AR_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \rightsquigarrow n\widehat{g}_n' M'(M\widehat{\Omega}_n M')^{-1}M\widehat{g}_n = AR_n$. We have

$$\begin{aligned}
&QLR_{1n} := AR_n - \lambda_{\min}\left(n\left(\widehat{g}_n, \widehat{D}_n\widehat{L}_n^{1/2}\right)'\widehat{\Omega}_n^{-1}\left(\widehat{g}_n, \widehat{D}_n\widehat{L}_n^{1/2}\right)\right) \\
&\rightsquigarrow AR_n - \lambda_{\min}\left(n\left(M\widehat{g}_n, M\widehat{D}_n\widehat{L}_n^{1/2}\right)'(M\widehat{\Omega}_n M')^{-1}\left(M\widehat{g}_n, M\widehat{D}_n\widehat{L}_n^{1/2}\right)\right) = QLR_{1n}, \quad (19.5)
\end{aligned}$$

using the invariance of AR_n and $\widehat{L}_n$ and the equivariance of the other statistics that appear. □

20 Proofs of Lemma 10.3 and Proposition 10.4

Lemma 10.3 of AG2. *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

where (a) $(\overline{g}_h, \overline{D}_h)$ *are defined in* (10.21), (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (10.24), (c) $(\overline{D}_h, \overline{\Delta}_h)$ *and* $\overline{g}_h$ *are independent, and* (d) *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 results of parts* (a)-(c) *hold with* n *replaced with* w_n.

Here and below, we use the following simplified notation:

$$\begin{aligned}
&D_n := E_{F_n}G_i,\ B_n := B_{F_n},\ C_n := C_{F_n},\ B_n = (B_{n,q}, B_{n,p-q}),\ C_n = (C_{n,q}, C_{n,k-q}), \\
&W_n := W_{F_n},\ W_{2n} := W_{2F_n},\ U_n := U_{F_n},\ \text{and } U_{2n} := U_{2F_n}, \quad (20.1)
\end{aligned}$$

where $q = q_h$ is defined in (10.22), $B_{n,q} \in R^{p\times q}$, $B_{n,p-q} \in R^{p\times(p-q)}$, $C_{n,q} \in R^{k\times q}$, and $C_{n,k-q} \in R^{k\times(k-q)}$. Let

$$\begin{aligned}
\Upsilon_{n,q} &:= Diag\{\tau_{1F_n}, ..., \tau_{qF_n}\} \in R^{q\times q}, \\
\Upsilon_{n,p-q} &:= Diag\{\tau_{(q+1)F_n}, ..., \tau_{pF_n}\} \in R^{(p-q)\times(p-q)} \text{ if } k \geq p, \\
\Upsilon_{n,k-q} &:= Diag\{\tau_{(q+1)F_n}, ..., \tau_{kF_n}\} \in R^{(k-q)\times(k-q)} \text{ if } k < p, \\
\Upsilon_n &:= \begin{bmatrix} \Upsilon_{n,q} & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & \Upsilon_{n,p-q} \\ 0^{(k-p)\times q} & 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times p} \text{ if } k \geq p, \text{ and} \\
\Upsilon_n &:= \begin{bmatrix} \Upsilon_{n,q} & 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ 0^{(k-q)\times q} & \Upsilon_{n,k-q} & 0^{(k-q)\times(p-k)} \end{bmatrix} \in R^{k\times p} \text{ if } k < p. \quad (20.2)
\end{aligned}$$

As defined, Υ_n is the diagonal matrix of singular values of $W_n D_n U_n$, see (10.15).

Proof of Lemma 10.3. The asymptotic distribution of $n^{1/2}(\widehat{g}_n, vec(\widehat{D}_n - E_{F_n}G_i))$ given in Lemma 10.3 follows from the Lyapunov triangular-array multivariate CLT (using the moment restrictions in $\mathcal{F}_2$) and the following:

$$\begin{aligned} n^{1/2}vec(\widehat{D}_n - E_{F_n}G_i) &= n^{-1/2}\sum_{i=1}^{n} vec(G_i - E_{F_n}G_i) - \begin{pmatrix} \widehat{\Gamma}_{1n} \\ \vdots \\ \widehat{\Gamma}_{pn} \end{pmatrix} \widehat{\Omega}_n^{-1} n^{1/2}\widehat{g}_n \qquad (20.3) \\ &= n^{-1/2}\sum_{i=1}^{n}\left[vec(G_i - E_{F_n}G_i) - \begin{pmatrix} E_{F_n}G_{\ell 1}g_\ell' \\ \vdots \\ E_{F_n}G_{\ell p}g_\ell' \end{pmatrix} \Omega_{F_n}^{-1} g_i \right] + o_p(1), \end{aligned}$$

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 $n^{1/2}\widehat{g}_n = O_p(1)$.

The limiting covariance matrix between $n^{1/2}vec(\widehat{D}_n - E_{F_n}G_i)$ and $n^{1/2}\widehat{g}_n$ is a zero matrix because $E_{F_n}[G_{ij} - E_{F_n}G_{ij} - (E_{F_n}G_{\ell j}g_\ell')\Omega_{F_n}^{-1}g_i]g_i' = 0^{k\times k}$, where G_{ij} denotes the jth column of G_i. By the CLT, the limiting variance matrix of $n^{1/2}vec(\widehat{D}_n - D_n)$ equals $\lim Var_{F_n}(vec(G_i) - (E_{F_n}vec(G_\ell)g_\ell')\Omega_{F_n}^{-1}g_i) = \lim \Phi_{F_n}^{vec(G_i)} = \Phi_h^{vec(G_i)}$, see (10.20), and the limit exists because (i) the components of $\Phi_{F_n}^{vec(G_i)}$ are comprised of λ_{4,F_n} and submatrices of λ_{5,F_n} and (ii) $\lambda_{s,F_n} \to h_s$ for $s = 4, 5$. By the CLT, the limiting variance matrix of $n^{1/2}\widehat{g}_n$ equals $\lim E_{F_n}g_ig_i' = h_{5,g}$.

The asymptotic distribution of $n^{1/2}W_{F_n}\widehat{D}_nU_{F_n}T_n$ is obtained as follows. Using (10.13)-(10.15), the singular value decomposition of $W_nD_nU_n$ is $W_nD_nU_n = C_n\Upsilon_nB_n'$. Using this, we get

$$W_nD_nU_nB_{n,q}\Upsilon_{n,q}^{-1} = C_n\Upsilon_nB_n'B_{n,q}\Upsilon_{n,q}^{-1} = C_n\Upsilon_n \begin{pmatrix} I_q \\ 0^{(p-q)\times q} \end{pmatrix} \Upsilon_{n,q}^{-1} = C_n \begin{pmatrix} I_q \\ 0^{(k-q)\times q} \end{pmatrix} = C_{n,q}, \qquad (20.4)$$

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

$$\begin{aligned} W_n\widehat{D}_nU_nB_{n,q}\Upsilon_{n,q}^{-1} &= W_nD_nU_nB_{n,q}\Upsilon_{n,q}^{-1} + W_nn^{1/2}(\widehat{D}_n - D_n)U_nB_{n,q}(n^{1/2}\Upsilon_{n,q})^{-1} \\ &= C_{n,q} + o_p(1) \to_p h_{3,q} = \overline{\Delta}_{h,q}, \qquad (20.5) \end{aligned}$$

where the second equality uses (among other things) $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$ (by the definition of q in (10.22)). The convergence in (20.5) holds by (10.19), (10.24), and (20.1), and the last equality in (20.5) holds by the definition of $\overline{\Delta}_{h,q}$ in (10.24).

Using the singular value decomposition of $W_n D_n U_n$ again, we obtain: if $k \geq p$,

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

$$= C_n \begin{pmatrix} 0^{q \times (p-q)} \\ n^{1/2} \Upsilon_{n,p-q} \\ 0^{(k-p) \times (p-q)} \end{pmatrix} \to h_3 \begin{pmatrix} 0^{q \times (p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p) \times (p-q)} \end{pmatrix} = h_3 h_{1,p-q}^{\diamond}, \quad (20.6)$$

where the second equality uses $B_n' B_n = I_p$, the third equality and the convergence hold by (10.19) using the definitions in (10.24) and (20.2) with $k \geq p$, and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (10.24) with $k \geq p$. Analogously, if $k < p$, we have

$$n^{1/2} W_n D_n U_n B_{n,p-q} = n^{1/2} C_n \Upsilon_n \begin{pmatrix} 0^{q \times (p-q)} \\ I_{p-q} \end{pmatrix} = C_n \begin{pmatrix} 0^{q \times (k-q)} & 0^{q \times (p-k)} \\ n^{1/2} \Upsilon_{n,k-q} & 0^{(k-q) \times (p-k)} \end{pmatrix}$$

$$\to h_3 \begin{pmatrix} 0^{q \times (k-q)} & 0^{q \times (p-k)} \\ Diag\{h_{1,q+1}, ..., h_{1,k}\} & 0^{(k-q) \times (p-k)} \end{pmatrix} = h_3 h_{1,p-q}^{\diamond}, \quad (20.7)$$

where the third equality holds by (20.2) with $k < p$ and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (10.24) with $k < p$.

Using (20.6), (20.7), and $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n} G_i) \to_d (\overline{g}_h, \overline{D}_h)$, we get

$$\begin{aligned} n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} &= n^{1/2} W_n D_n U_n B_{n,p-q} + W_n n^{1/2} (\widehat{D}_n - D_n) U_n B_{n,p-q} \\ &\to_d h_3 h_{1,p-q}^{\diamond} + h_{71} \overline{D}_h h_{81} h_{2,p-q} = \overline{\Delta}_{h,p-q}, \end{aligned} \quad (20.8)$$

where $B_{n,p-q} \to h_{2,p-q}$, $W_n \to h_{71}$, and $U_n \to h_{81}$, and the last equality holds by the definition of $\overline{\Delta}_{h,p-q}$ in (10.24).

Equations (20.5) and (20.8) combine to establish

$$\begin{aligned} n^{1/2} W_n \widehat{D}_n U_n T_n &= n^{1/2} W_n \widehat{D}_n U_n B_n S_n = (W_n \widehat{D}_n U_n B_{n,q} \Upsilon_{n,q}^{-1}, n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q}) \\ &\to_d (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) = \overline{\Delta}_h \end{aligned} \quad (20.9)$$

using the definition of S_n in (10.23). This completes the proof of the convergence result of Lemma 10.3.

Parts (a) and (b) of the lemma hold by the definitions of $(\overline{g}_h, \overline{D}_h)$ and $\overline{\Delta}_h$. The independence of $(\overline{D}_h, \overline{\Delta}_h)$ and $\overline{g}_h$, stated in part (c) of the lemma, holds by the independence of $\overline{g}_h$ and $\overline{D}_h$ (which follows from (10.21)), and part (b) of the lemma. Part (d) is proved by replacing n by w_n in the proofs above. □

Proposition 10.4 of AG2. *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) $\widehat{\kappa}_{jn} \rightarrow_p \infty$ *for all* $j \leq q$,

(b) *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* $\overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \times \overline{\Delta}_{h,p-q} \in R^{(p-q)\times(p-q)}$,

(c) *the convergence in parts* (a) *and* (b) *holds jointly with the convergence in Lemma* 10.3, *and*

(d) *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)-(c) *hold with* n *replaced with* w_n.

Proof of Proposition 10.4. For the case where $k \geq p$, Proposition 10.4 is the same as Theorem 8.4(c)-(f) given in the Appendix to AG1, which is proved in Section 16 in the SM to AG1. For brevity, we only describe the changes that need to be made to that proof to cover the case where $k < p$. Note that the proof of Theorem 8.4(c)-(f) in AG1 is similar to, but simpler than, the proof of Theorem 10.5, which is given in Section 21 below.

In the second line of the proof of Lemma 16.1 in the SM to AG1, p needs to be replaced by $\min\{k, p\}$ three times.

In the fourth line of (16.3) in the SM to AG1, the $k \times p$ matrix that contains six submatrices needs to be replaced by the following matrix when $k < p$:

$$\begin{bmatrix} h_{6,r_1^\diamond}^\diamond + o(1) & 0^{r_1^\diamond \times (k-r_1^\diamond)} & 0^{r_1^\diamond \times (p-k)} \\ 0^{(k-r_1^\diamond)\times r_1^\diamond} & O(\tau_{r_2 F_n}/\tau_{r_1 F_n})^{(k-r_1^\diamond)\times(k-r_1^\diamond)} & 0^{(k-r_1^\diamond)\times(p-k)} \end{bmatrix} \in R^{k\times p}. \quad (20.10)$$

In the first line of (16.22) in the SM to AG1, the $k \times (p - r_{g-1}^\diamond)$ matrix that contains three submatrices needs to be replaced by the following matrix when $k < p$:

$$\begin{bmatrix} 0^{r_{g-1}^\diamond \times (k-r_{g-1}^\diamond)} & 0^{r_{g-1}^\diamond \times (p-k)} \\ Diag\{\tau_{r_g F_n}, ..., \tau_{kF_n}\}/\tau_{r_g F_n} & 0^{(k-r_{g-1}^\diamond)\times(p-k)} \end{bmatrix} \in R^{k\times(p-r_{g-1}^\diamond)}. \quad (20.11)$$

The limit of this matrix as $n \to \infty$ equals the matrix given in the second line of (16.22) that contains three submatrices. Thus, the limit of the matrix on the first line of (16.22) is the same for the cases where $k \geq p$ and $k < p$.

In the third line of (16.25) in the SM to AG1, the second matrix that contains three submatrices (which is a $k \times (p - r_g^\diamond)$ matrix) is the same as the matrix in the first line of (16.22) in the SM to

AG1, but with $r_g^\diamond$ in place of $r_{g-1}^\diamond$ (using $r_{g+1} = r_g^\diamond + 1$ and $r_g = r_{g-1}^\diamond + 1$). When $k < p$, this matrix needs to be changed just as the matrix in the first line of (16.22) is changed in (20.11), but with $r_g^\diamond$ in place of $r_{g-1}^\diamond$.

No other changes are needed. □

21 Proof of Theorem 10.5

Theorem 10.5 of AG2. *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

$$QLR_n \rightarrow_d \overline{g}_h' h_{5,g}^{-1} \overline{g}_h - \lambda_{\min}((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h))$$

and the convergence holds jointly with the convergence in Lemma 10.3 *and Proposition* 10.4. *When* $q = p$ *(which can only hold if* $k \geq p$ *because* $q \leq \min\{k, p\}$*),* $\overline{\Delta}_{h,p-q}$ *does not appear in the limit random variable and the limit random variable reduces to* $(h_{5,g}^{-1/2} \overline{g}_h)' h_{3,p} h_{3,p}' h_{5,g}^{-1/2} \overline{g}_h \sim \chi_p^2$. *When* $q = k$ *(which can only hold if* $k \leq p$*), the* $\lambda_{\min}(\cdot)$ *expression does not appear in the limit random variable and the limit random variable reduces to* $\overline{g}_h' h_{5,g}^{-1} \overline{g}_h \sim \chi_k^2$. *When* $k \leq p$ *and* $q < k$*, the* $\lambda_{\min}(\cdot)$ *expression equals zero and the limit random variable reduces to* $\overline{g}_h' h_{5,g}^{-1} \overline{g}_h \sim \chi_k^2$. *Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$*, the same results hold with* n *replaced with* w_n.

The proof of Theorem 10.5 uses the approach in Johansen (1991, pp. 1569-1571) and Robin and Smith (2000, pp. 172-173). 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 $n \rightarrow \infty$, possibly at different rates. This complicates the proof considerably. For example, the rate of convergence result of Lemma 21.1(b) below is needed in the present context, but not in the fixed distribution scenario considered in Johansen (1991) and Robin and Smith (2000).

The proof uses the notation given in (20.1) and (20.2) above. The following definitions are used:

$$\begin{aligned}
\widehat{D}_n^+ &:= (\widehat{D}_n, \widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n) \in R^{k\times(p+1)},\ \widehat{U}_n^+ := \begin{bmatrix} \widehat{U}_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
U_n^+ &:= \begin{bmatrix} U_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},\ h_{81}^+ := \begin{bmatrix} h_{81} & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
B_n^+ &: \begin{bmatrix} B_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
B_n^+ &= (B_{n,q}^+, B_{n,p+1-q}^+) \text{ for } B_{n,q}^+ \in R^{(p+1)\times q} \text{ and } B_{n,p+1-q}^+ \in R^{(p+1)\times(p+1-q)}, \qquad (21.1) \\
D_n^+ &:= (D_n, 0^k) \in R^{k\times(p+1)},\ \Upsilon_n^+ := (\Upsilon_n, 0^k) \in R^{k\times(p+1)}, \\
S_n^+ &:= Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}, 1, ..., 1\} = \begin{bmatrix} S_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},
\end{aligned}$$

where $\widehat{g}_n$ and $\widehat{\Omega}_n$ are defined in (5.1) with $\theta = \theta_0$, $\widehat{D}_n$ is defined in (6.2) with $\theta = \theta_0$, $\widehat{W}_n$, $\widehat{U}_n$, U_n $(:= U_{F_n})$, and W_n $(:= W_{F_n})$ are defined in (10.4), h_{81} is defined in (10.24), B_n $(:= B_{F_n})$ is defined in (10.13), D_n is defined in (20.1), Υ_n is defined in (20.2), and S_n is defined in (10.23).

Let

$$\widehat{\kappa}_{jn}^+ \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n^{\prime}\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+,\ \forall j = 1, ..., p+1, \qquad (21.2)$$

ordered to be nonincreasing in j. We have[78]

$$\begin{aligned}
&\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+ = (\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n) \text{ and} \qquad (21.3) \\
&\lambda_{\min}(n(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)'(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)) = \lambda_{\min}(n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n^{\prime}\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+) = \widehat{\kappa}_{(p+1)n}^+.
\end{aligned}$$

The proof of Theorem 10.5 uses the following rate of convergence lemma, which is analogous to Lemma 16.1 in Section 16 of the SM to AG1.

Lemma 21.1 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$ *for which* q *defined in* (10.22) *satisfies* $q \geq 1$, *we have* (a) $\widehat{\kappa}_{jn}^+ \to_p \infty$ *for* $j = 1, ..., q$ *and* (b) $\widehat{\kappa}_{jn}^+ = o_p((n^{1/2}\tau_{\ell F_n})^2)$ *for all* $\ell \leq q$ *and* $j = q+1, ..., p+1$. *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*

[78] In (21.3), we write $(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$, whereas we write its analogue $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$ in (6.7) with its columns in the reverse order. Both ways give the same value for the minimum eigenvalue of the inner product of the matrix with itself, which is the statistic of interest. We use the order $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$ in AG2 because it is consistent with the order in Moreira (2003) and Andrews, Moreira, and Stock (2006). We use the order $(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$ here (and elsewhere in the SM) because it has significant notational advantages in the proofs, especially in the proof of Theorem 10.5 in this Section.

holds with n replaced with w_n.

Proof of Theorem 10.5. We have $n^{1/2}\widehat{g}_n \to_d \overline{g}_h$ (by Lemma 10.3) and $\widehat{\Omega}_n^{-1/2} \to_p h_{5,g}^{-1/2}$ (because $\widehat{\Omega}_n - \Omega_{F_n} \to_p 0^{k\times k}$ by the WLLN, $\Omega_{F_n} \to h_{5,g}$, and $h_{5,g}$ is pd). In consequence, $AR_n \to_d \overline{g}_h' h_{5,g}^{-1}\overline{g}_h$. Given this, the definition of QLR_n in (10.3), and (21.3), to prove the convergence result in Theorem 10.5, it suffices to show that

$$\lambda_{\min}(n\widehat{U}_n^{+}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^{+}\widehat{U}_n^{+}) \to_d \lambda_{\min}((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)' h_{3,k-q}h_{3,k-q}'(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)). \quad (21.4)$$

Now we establish (21.4). The eigenvalues $\{\widehat{\kappa}_{jn}^{+} : j \leq p+1\}$ of $n\widehat{U}_n^{+}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^{+}\widehat{U}_n^{+}$ are the ordered solutions to the determinantal equation $|n\widehat{U}_n^{+}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^{+}\widehat{U}_n^{+} - \kappa I_{p+1}| = 0$. Equivalently, with probability that goes to one (wp→1), they are the solutions to

$$\begin{aligned} &|Q_n^{+}(\kappa)| = 0, \text{ where} && (21.5)\\ &Q_n^{+}(\kappa) := nS_n^{+}B_n^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^{+}U_n^{+}B_n^{+}S_n^{+} - \kappa S_n^{+\prime}B_n^{+\prime}U_n^{+\prime}(\widehat{U}_n^{+})^{-1\prime}(\widehat{U}_n^{+})^{-1}U_n^{+}B_n^{+}S_n^{+}, \end{aligned}$$

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

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

$$\begin{aligned} B_n^{+}S_n^{+} &= (B_{n,q}^{+}S_{n,q}, B_{n,p+1-q}^{+}), \text{ where}\\ S_{n,q} &:= Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}\} \in R^{q\times q}. \end{aligned} \quad (21.6)$$

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

$$\begin{aligned} n^{1/2}W_n\widehat{D}_n^{+}U_n^{+}B_{n,q}^{+}S_{n,q} &= n^{1/2}W_n\widehat{D}_nU_nB_{n,q}S_{n,q} \to_p \overline{\Delta}_{h,q} := h_{3,q} \text{ and}\\ n^{1/2}W_n\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} &= n^{1/2}W_n(\widehat{D}_n, \widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n)U_n^{+}B_{n,p+1-q}^{+}\\ &= n^{1/2}(W_n\widehat{D}_nU_nB_{n,p-q}, W_n\widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n)\\ &\to_d (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h), \end{aligned} \quad (21.7)$$

where $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are defined in (10.24) and $B_{n,p-q}$ is defined in (20.1).

We have

$$\widehat{W}_nW_n^{-1} \to_p I_k \text{ and } \widehat{U}_n^{+}(U_n^{+})^{-1} \to_p I_{p+1} \quad (21.8)$$

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

By (21.5)-(21.8), we have

$$\begin{aligned}
&Q_n^+(\kappa) \\
&= \begin{bmatrix} I_q + o_p(1) & h_{3,q}' n^{1/2} W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) \\ n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' h_{3,q} + o_p(1) & n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' W_n n^{1/2} \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) \end{bmatrix} \\
&\quad -\kappa \begin{bmatrix} S_{n,q}^2 & 0^{q \times (p+1-q)} \\ 0^{(p+1-q) \times q} & I_{p+1-q} \end{bmatrix} - \kappa \begin{bmatrix} S_{n,q} A_{1n}^+ S_{n,q} & S_{n,q} A_{2n}^+ \\ A_{2n}^{+\prime} S_{n,q} & A_{3n}^+ \end{bmatrix}, \text{ where} \qquad (21.9) \\
&\widehat{A}_n^+ = \begin{bmatrix} A_{1n}^+ & A_{2n}^+ \\ A_{2n}^{+\prime} & A_{3n}^+ \end{bmatrix} := B_n^{+\prime} U_n^{+\prime} (\widehat{U}_n^+)^{-1\prime} (\widehat{U}_n^+)^{-1} U_n^+ B_n^+ - I_{p+1} = o_p(1)
\end{aligned}$$

for $A_{1n}^+ \in R^{q \times q}$, $A_{2n}^+ \in R^{q \times (p+1-q)}$, and $A_{3n}^+ \in R^{(p+1-q) \times (p+1-q)}$, 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 (10.14), (10.16), (10.19), and (10.24)). Note that A_{jn}^+ and $\widehat{A}_{jn}^+$ (defined in (21.19) below) 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 $q = 0$, then $B_n^+ = B_{n,p+1-q}^+$ and

$$\begin{aligned}
& n B_n^{+\prime} \widehat{U}_n^{+\prime} \widehat{D}_n^{+\prime} \widehat{W}_n' \widehat{W}_n \widehat{D}_n^+ \widehat{U}_n^+ B_n^+ \\
&= n B_n^{+\prime} ((U_n^+)^{-1} \widehat{U}_n^+)' (B_n^+)^{-1\prime} B_n^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' \left(\widehat{W}_n W_n^{-1}\right)' \\
&\quad \times \left(\widehat{W}_n W_n^{-1}\right) (W_n \widehat{D}_n^+ U_n^+ B_n^+) (B_n^+)^{-1} ((U_n^+)^{-1} \widehat{U}_n^+) B_n^+ \\
&\rightarrow_d (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h), \qquad (21.10)
\end{aligned}$$

where the convergence holds by (21.7) and (21.8) and $\overline{\Delta}_{h,p-q}$ is defined as in (10.24) with $q = 0$. The smallest eigenvalue of a matrix is a continuous function of the matrix (by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). Hence, the smallest eigenvalue of $n B_n^{+\prime} \widehat{U}_n^{+\prime} \widehat{D}_n^{+\prime} \widehat{W}_n' \widehat{W}_n \widehat{D}_n^+ \widehat{U}_n^+ B_n^+$ converges in distribution to the smallest eigenvalue of $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)$ (using $h_{3,k-q} h_{3,k-q}' = h_3 h_3' = I_k$ when $q = 0$), which proves (21.4) when $q = 0$.

In the remainder of (21.4), we assume $q \geq 1$, which is the remaining case to be considered in the proof of (21.4). The formula for the determinant of a partitioned matrix and (21.9) give

$$\begin{aligned}
|Q_n^+(\kappa)| &= |Q_{1n}^+(\kappa)| \cdot |Q_{2n}^+(\kappa)|, \text{ where} \\
Q_{1n}^+(\kappa) &:= I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n}^+ S_{n,q}, \\
Q_{2n}^+(\kappa) &:= n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' W_n n^{1/2} \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) - \kappa I_{p+1-q} - \kappa A_{3n}^+ \\
&\quad -[n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' h_{3,q} + o_p(1) - \kappa A_{2n}^{+\prime} S_{n,q}](I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q} A_{1n}^+ S_{n,q})^{-1} \\
&\quad \times [h_{3,q}' n^{1/2} W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) - \kappa S_{n,q} A_{2n}^+], \qquad (21.11)
\end{aligned}$$

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

By Lemma 21.1(b) (which applies for $q \geq 1$), for $j = q+1, ..., p+1$, and $A_{1n}^+ = o_p(1)$ (by (21.9)), 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,

$$Q_{1n}^+(\widehat{\kappa}_{jn}^+) = I_q + o_p(1) - \widehat{\kappa}_{jn}^+ S_{n,q}^2 - \widehat{\kappa}_{jn}^+ S_{n,q} A_{1n}^+ S_{n,q} = I_q + o_p(1). \qquad (21.12)$$

By (21.5) and (21.11), $|Q_n^+(\widehat{\kappa}_{jn}^+)| = |Q_{1n}^+(\widehat{\kappa}_{jn}^+)| \cdot |Q_{2n}^+(\widehat{\kappa}_{jn}^+)| = 0$ for $j = 1, ..., p+1$. By (21.12), $|Q_{1n}^+(\widehat{\kappa}_{jn}^+)| \neq 0$ for $j = q+1, ..., p+1$ wp$\to$1. Hence, wp$\to$1,

$$|Q_{2n}^+(\widehat{\kappa}_{jn}^+)| = 0 \text{ for } j = q+1, ..., p+1. \qquad (21.13)$$

Now we plug in $\widehat{\kappa}_{jn}^+$ for $j = q+1, ..., p+1$ into $Q_{2n}^+(\kappa)$ in (21.11) and use (21.12). We have

$$\begin{aligned}
Q_{2n}^+(\widehat{\kappa}_{jn}^+) &= n B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) \\
&\quad -[n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' h_{3,q} + o_p(1)](I_q + o_p(1))[h_{3,q}' n^{1/2} W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1)] \\
&\quad -\widehat{\kappa}_{jn}^+[I_{p+1-q} + A_{3n}^+ - (n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' h_{3,q} + o_p(1))(I_q + o_p(1)) S_{n,q} A_{2n}^+ \\
&\quad -A_{2n}^{+\prime} S_{n,q}(I_q + o_p(1))(h_{3,q}' n^{1/2} W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1)) \\
&\quad +\widehat{\kappa}_{jn}^+ A_{2n}^{+\prime} S_{n,q}(I_q + o_p(1)) S_{n,q} A_{2n}^+]. \qquad (21.14)
\end{aligned}$$

The term in square brackets on the last three lines of (21.14) that multiplies $\widehat{\kappa}_{jn}^+$ equals

$$I_{p+1-q} + o_p(1), \qquad (21.15)$$

because $A_{3n}^+ = o_p(1)$ (by (21.9)), $n^{1/2} W_n \widehat{D}_n^+ U_n^+ B_{n,p+1-q}^+ = O_p(1)$ (by (21.7)), $S_{n,q} = o(1)$ (by the definitions of q and $S_{n,q}$ in (10.22) and (21.6), respectively, and $h_{1,j} := \lim n^{1/2} \tau_{jF_n}$), $A_{2n}^+ = o_p(1)$ (by (21.9)), and $\widehat{\kappa}_{jn}^+ A_{2n}^{+\prime} S_{n,q}(I_q + o_p(1)) S_{n,q} A_{2n}^+ = A_{2n}^{+\prime} \widehat{\kappa}_{jn}^+ S_{n,q}^2 A_{2n}^+ + A_{2n}^{+\prime} \widehat{\kappa}_{jn}^+ S_{n,q} o_p(1) S_{n,q} A_{2n}^+ = o_p(1)$

(using $\widehat{\kappa}_{jn}^{+} S_{n,q}^{2} = o_p(1)$ and $A_{2n}^{+} = o_p(1)$).

Equations (21.14) and (21.15) give

$$\begin{aligned}
&Q_{2n}^{+}(\widehat{\kappa}_{jn}^{+}) \\
&= n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n^{\prime} [I_k - h_{3,q} h_{3,q}^{\prime}] n^{1/2} W_n \widehat{D}_n^{+} U_n^{+} B_{n,p+1-q}^{+} + o_p(1) - \widehat{\kappa}_{jn}^{+}[I_{p+1-q} + o_p(1)] \\
&= n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n^{\prime} h_{3,k-q} h_{3,k-q}^{\prime} n^{1/2} W_n \widehat{D}_n^{+} U_n^{+} B_{n,p+1-q}^{+} + o_p(1) - \widehat{\kappa}_{jn}^{+}[I_{p+1-q} + o_p(1)] \\
&:= M_{n,p+1-q}^{+} - \widehat{\kappa}_{jn}^{+}[I_{p+1-q} + o_p(1)], \qquad (21.16)
\end{aligned}$$

where the second equality uses $I_k = h_3 h_3^{\prime} = h_{3,q} h_{3,q}^{\prime} + h_{3,k-q} h_{3,k-q}^{\prime}$ (because $h_3 = \lim C_n$ is an orthogonal matrix) and the last line defines the $(p+1-q) \times (p+1-q)$ matrix $M_{n,p+1-q}^{+}$.

Equations (21.13) and (21.16) imply that $\{\widehat{\kappa}_{jn}^{+} : j = q+1, ..., p+1\}$ are the $p+1-q$ eigenvalues of the matrix

$$M_{n,p+1-q}^{++} := [I_{p+1-q} + o_p(1)]^{-1/2} M_{n,p+1-q}^{+} [I_{p+1-q} + o_p(1)]^{-1/2} \qquad (21.17)$$

by pre- and post-multiplying the quantities in (21.16) by the rhs quantity $[I_{p+1-q} + o_p(1)]^{-1/2}$ in (21.16). By (21.7),

$$M_{n,p+1-q}^{++} \to_d (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)^{\prime} h_{3,k-q} h_{3,k-q}^{\prime} (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h). \qquad (21.18)$$

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 (21.18), the matrix $M_{n,p+1-q}^{++}$ converges in distribution. In consequence, by the CMT, the vector of eigenvalues of $M_{n,p+1-q}^{++}$, viz., $\{\widehat{\kappa}_{jn}^{+} : j = q+1, ..., p+1\}$, converges in distribution to the vector of eigenvalues of the limit matrix $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)^{\prime} h_{3,k-q} h_{3,k-q}^{\prime} (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)$. Hence, $\lambda_{\min}(n \widehat{U}_n^{+\prime} \widehat{D}_n^{+\prime} \widehat{W}_n^{\prime} \widehat{W}_n \widehat{D}_n^{+} \widehat{U}_n^{+})$, which equals the smallest eigenvalue, $\widehat{\kappa}_{(p+1)n}^{+}$, converges in distribution to the smallest eigenvalue of $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)^{\prime} h_{3,k-q} h_{3,k-q}^{\prime} \overline{\Delta}_{h,p-q}^{\prime}$, which completes the proof of (21.4).

The previous paragraph proves Comment (v) to Theorem 10.5 for the smallest $p+1-q$ eigenvalues of $n(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)^{\prime} (\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)$. In addition, by Lemma 21.1(a), the largest q eigenvalues of this matrix diverge to infinity in probability, which completes the proof of Comment (v) to Theorem 10.5.

When $q = p$, the third and fourth lines in (21.7) become $n^{1/2} W_n \widehat{W}_n^{-1} \widehat{\Omega}_n^{-1/2} \widehat{g}_n$ and $h_{5,g}^{-1/2} \overline{g}_h$, respectively, i.e., $n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q}$ and $\overline{\Delta}_{h,p-q}$ drop out (because $U_n^{+} B_{n,p+1-q}^{+} = (0^{p\prime}, 1)^{\prime}$ in this case). In consequence, the limit in (21.18) becomes $(h_{5,g}^{-1/2} \overline{g}_h)^{\prime} h_{3,k-q} h_{3,k-q}^{\prime} h_{5,g}^{-1/2} \overline{g}_h$, which has a χ_{k-p}^2 distribution (because $h_{5,g}^{-1/2} \overline{g}_h \sim N(0^k, I_k)$, $h_3 = (h_{3,q}, h_{3,k-q}) \in R^{k \times k}$ is an orthogonal matrix, and $h_{3,k-q}$ has $k-p$ columns when $q = p$).

The convergence in Theorem 10.5 holds jointly with that in Lemma 10.3 and Proposition 10.4 because the results in Proposition 10.4 and Theorem 10.5 just rely on the convergence in distribution of $n^{1/2}W_n\widehat{D}_nU_nT_n$, which is part of Lemma 10.3.

When $q = k$, the $\lambda_{\min}(\cdot)$ expression does not appear in the limit random variable in the statement of Theorem 10.5 because, in the second line of (21.16) above, the term $I_k - h_{3,q}h_{3,q}'$ equals $0^{k\times k}$, which implies that $M_{n,p+1-q}^{+} = 0^{(p+1-q)\times(p+1-q)} + o_p(1)$ and $M_{n,p+1-q}^{++} = 0^{(p+1-q)\times(p+1-q)} + o_p(1) \to_p 0^{(p+1-q)\times(p+1-q)}$ in (21.17) and (21.18).

When $k \leq p$ and $q < k$, the $\lambda_{\min}(\cdot)$ expression (in the limit random variable in the statement of Theorem 10.5) equals zero because $h_{3,k-q}'(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)$ is a $(k-q)\times(p+1-q)$ matrix, which has fewer rows than columns when $k < p+1$.

The convergence in Theorem 10.5 holds for a subsequence $\{w_n : n \geq 1\}$ of $\{n\}$ by the same proof as given above with n replaced by w_n. $\square$

Proof of Lemma 21.1. The proof of Lemma 21.1 is the same as the proof of Lemma 16.1 in Section 16 in the SM to AG1, but with p replaced by $p+1$ (so $p+1$ is always at least two), with $\tau_{(p+1)F_n} := 0$, with $h_{6,p} := \lim \tau_{(p+1)F_n}/\tau_{pF_n} = 0$ (using $0/0 := 0$), and with $\widehat{D}_n, \widehat{U}_n, B_n, \widehat{\kappa}_{jn}, \widehat{A}_n, D_n, U_n, h_{81}, \Upsilon_n$, $B_{n,r_1^\diamond}$, and $B_{n,p-r_1^\diamond}$ replaced by $\widehat{D}_n^+, \widehat{U}_n^+, B_n^+, \widehat{\kappa}_{jn}^+, \widehat{A}_n^+, D_n^+, U_n^+, h_{81}^+, \Upsilon_n^+, B_{n,r_1^\diamond}^+$, and $B_{n,p+1-r_1^\diamond}^+$, respectively, where

$$\widehat{A}_n^+ = \begin{bmatrix} \widehat{A}_{1n}^+ & \widehat{A}_{2n}^+ \\ \widehat{A}_{2n}^{+\prime} & \widehat{A}_{3n}^+ \end{bmatrix} := (B_n^+)'(U_n^+)'(\widehat{U}_n^+)^{-1\prime}(\widehat{U}_n^+)^{-1}U_n^+B_n^+ - I_{p+1}, \tag{21.19}$$

where $\widehat{A}_{1n}^+ \in R^{r_1^\diamond\times r_1^\diamond}$, $\widehat{A}_{2n}^+ \in R^{r_1^\diamond\times(p+1-r_1^\diamond)}$, $\widehat{A}_{3n}^+ \in R^{(p+1-r_1^\diamond)\times(p+1-r_1^\diamond)}$, and $r_1^\diamond$ is defined as in the proof of Lemma 13.1 in the SM to AG1. Note that the quantities $\widehat{A}_{\ell n}$ for $\ell = 1,2,3$, which depend on $\widehat{A}_n$ (see (13.18) in the SM to AG1), differ between the two proofs (because $\widehat{A}_n$ differs from $\widehat{A}_n^+$). Similarly, the quantities ϱ_n (defined in (13.24) in the SM to AG1), $\widehat{\xi}_{\ell n}(\kappa)$ for $\ell = 1,2,3$ (defined in (13.25) in the SM to AG1), and $\widehat{A}_{j2n}$ (defined in (13.28) in the SM to AG1) differ between the two proofs (because the quantities on which they depend differ between the two proofs).

The following quantities are the same in both proofs: $\{\tau_{jF_n} : j \leq p\}, q, \{h_{6,j} : j \leq p-1\}, G_h, \{r_j : j \leq G_h\}, \{r_j^\diamond : j \leq G_h\}, h_{6,r_1^\diamond}^\diamond, \widehat{W}_n, W_n, h_{71}, C_n$, and h_3. Note that the first p singular values of $W_nD_nU_n$ (i.e., $\{\tau_{jF_n} : j \leq p\}$) and the first p singular values of $W_nD_n^+U_n^+$ are the same. This holds because $\tau_{jF_n} = \kappa_{jF_n}^{1/2}$, where κ_{jF_n} is the jth eigenvalue of $W_nD_nU_nU_n'D_n'W_n'$, $W_nD_n^+U_n^+ = W_n(D_n, 0^k)U_n^+ = (W_nD_nU_n, 0^k)$, and hence, $W_nD_n^+U_n^+U_n^{+\prime}D_n^{+\prime}W_n' = W_nD_nU_nU_n'D_n'W_n'$.

The second equality in (13.19) in the SM to AG1, which states that $W_nD_nU_nB_n = C_n\Upsilon_n$, is a key equality in the proof of Lemma 13.1 in the SM to AG1. The analogue in the proof of the current lemma is

$$W_n D_n^+ U_n^+ B_n^+ = (W_n D_n, 0^k) \begin{bmatrix} U_n B_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} = (W_n D_n U_n B_n, 0^k) = (C_n \Upsilon_n, 0^k) = C_n \Upsilon_n^+. \tag{21.20}$$

Hence, this part of the proof goes through when D_n, U_n, B_n, and Υ_n are replaced by D_n^+, U_n^+, B_n^+, and Υ_n^+, respectively. □

22 Proof of the Asymptotic Size Results

In this section we prove Theorem 10.1. For the reader's convenience, we restate this theorem here.

Theorem 10.1 of AG2. *The* AR, CQLR$_1$, *and* CQLR$_2$ *tests* (*without the* SR *extensions*), *defined in* (5.2), (6.8), *and* (7.3), *respectively, have asymptotic sizes equal to their nominal size* $\alpha \in (0,1)$ *and are asymptotically similar* (*in a uniform sense*) *for the parameter spaces* $\mathcal{F}_{AR}$, $\mathcal{F}_1$, *and* $\mathcal{F}_2$, *respectively. Analogous results hold for the corresponding* AR, CQLR$_1$, *and* CQLR$_2$ CS's *for the parameter spaces* $\mathcal{F}_{\Theta,AR}$, $\mathcal{F}_{\Theta,1}$, *and* $\mathcal{F}_{\Theta,2}$, *respectively.*

Theorem 10.1 is proved first for the CQLR tests and CS's. For the CQLR test results, we actually prove a more general result that applies to a CQLR test that is defined as the CQLR$_1$ test is defined in Section 6, but with the weight matrices $(\widehat{\Omega}_n^{-1/2}, \widehat{L}_n^{1/2})$ replaced by any matrices $(\widehat{W}_n, \widehat{U}_n)$ that satisfy Assumption WU for some parameter space $\Lambda_* \subset \Lambda_2$ (stated in Section 10.1.5). Then, we show that Assumption WU holds for the parameter spaces Λ_1 and Λ_2 for the weight matrices employed by the CQLR$_1$ and CQLR$_2$ tests, respectively, defined in Sections 6 and 7. These results combine to establish the CQLR test results of Theorem 10.1. The CQLR CS results of Theorem 10.1 are proved analogously to those for the tests, see the Comment to Proposition 10.2 for details.

In Section 22.6, we prove Theorem 10.1 for the AR test and CS.

22.1 Statement of Results

A general QLR test statistic for testing $H_0 : \theta = \theta_0$ is defined in (10.3) as

$$\begin{aligned} QLR_n &:= AR_n - \lambda_{\min}(n\widehat{Q}_{WU,n}), \text{ where} \\ \widehat{Q}_{WU,n} &:= (\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)'(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n), \end{aligned} \tag{22.1}$$

AR_n is defined in (6.2), and the dependence of $QLR_n, \widehat{Q}_{WU,n}, \widehat{W}_n, \widehat{D}_n, \widehat{U}_n, \widehat{\Omega}_n$, and $\widehat{g}_n$ on θ_0 is suppressed for notational simplicity.

The general CQLR test rejects the null hypothesis if

$$QLR_n > c_{k,p}(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha), \tag{22.2}$$

where $c_{k,p}(D, 1-\alpha)$ is defined just below (3.5).

The correct asymptotic size of the general QLR test is established using the following theorem.

Theorem 22.1 *Suppose Assumption* WU (*defined in Section* 10.1.5) *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$*. Then, the asymptotic null rejection probabilities of the nominal size* α CQLR *test based on* $(\widehat{W}_{w_n}, \widehat{U}_{w_n})$ *equal* α *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$.

Comments: **(i)** Theorem 22.1 and Proposition 10.2 imply that any nominal size α CQLR test based on matrices $(\widehat{W}_n, \widehat{U}_n)$ that satisfy Assumption WU for some parameter space Λ_* has correct asymptotic size α and is asymptotically similar (in a uniform sense) for the parameter space Λ_*.

(ii) In Lemma 22.4 below, we show that the choice of matrices $(\widehat{W}_n, \widehat{U}_n)$ for the CQLR$_1$ and CQLR$_2$ tests (defined in Sections 6 and 7, respectively) satisfy Assumption WU for the parameter spaces Λ_1 and Λ_2 (defined in (10.17)), respectively. In addition, Lemma 22.4 shows that $\mathcal{F}_1 \subset \mathcal{F}_{WU}$ and $\mathcal{F}_2 \subset \mathcal{F}_{WU}$ when δ_{WU} and M_{WU} that appear in the definition of $\mathcal{F}_{WU}$ are sufficiently small and large, respectively.[79] In consequence, the CQLR$_1$ and CQLR$_2$ tests have correct asymptotic size α and are asymptotically similar (in a uniform sense) for the parameter spaces $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, as stated in Theorem 10.1.

The proof of Theorem 22.1 uses Proposition 10.4 and Theorem 10.5, as well as the following lemmas.

Let $\{D_n^c : n \geq 1\}$ be a sequence of constant (i.e., nonrandom) $k \times p$ matrices. Here, we determine the limit as $n \to \infty$ of $c_{k,p}(D_n^c, 1-\alpha)$ under certain assumptions on the singular values of D_n^c.

Lemma 22.2 *Suppose* $\{D_n^c : n \geq 1\}$ *is a sequence of constant* (*i.e., nonrandom*) $k \times p$ *matrices with singular values* $\{\tau_{jn}^c \geq 0 : j \leq \min\{k,p\}\}$ *for* $n \geq 1$ *that satisfy* (i) $\{\tau_{jn}^c \geq 0 : j \leq \min\{k,p\}\}$ *are nonincreasing in* j *for* $n \geq 1$, (ii) $\tau_{jn}^c \to \infty$ *for* $j \leq q$ *for some* $0 \leq q \leq \min\{k,p\}$ *and* (iii)

[79] Note that the set of distributions $\mathcal{F}_{WU}$ depends on the definitions of (W_F, U_F), see (10.12), and (W_F, U_F) are defined differently for the QLR$_1$ and QLR$_2$ statistics, see (10.6)-(10.8) and (10.9)-(10.11), respectively. Hence, the set of distributions $\mathcal{F}_{WU}$ differs for the CQLR$_1$ and CQLR$_2$ tests.

$\tau^c_{jn} \to \tau^c_{j\infty} < \infty$ *for* $j = q+1, ..., \min\{k,p\}$. *Then,*

$$c_{k,p}(D^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha), \text{ where } \tau^c_\infty := (\tau^c_{(q+1)\infty}, ..., \tau^c_{\min\{k,p\}\infty})' \in R^{\min\{k,p\}-q},$$

$$\Upsilon(\tau^c_\infty) := \begin{pmatrix} Diag\{\tau^c_\infty\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \in R^{(k-q)\times(p-q)} \text{ if } k \geq p,$$

$$\Upsilon(\tau^c_\infty) := \left(Diag\{\tau^c_\infty\}, 0^{(k-q)\times(p-k)}\right) \in R^{(k-q)\times(p-q)} \text{ if } k < p,$$

$c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ *denotes the* $1-\alpha$ *quantile of*

$$ACLR_{k,p,q}(\tau^c_\infty) := Z'Z - \lambda_{\min}((\Upsilon(\tau^c_\infty), Z_2)'(\Upsilon(\tau^c_\infty), Z_2)), \text{ and}$$

$$Z := \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N(0^k, I_k) \text{ for } Z_1 \in R^q \text{ and } Z_2 \in R^{k-q}.$$

Comments: (i) The matrix $\Upsilon(\tau^c_\infty)$ is the diagonal matrix containing the $\min\{k,p\} - q$ finite limiting eigenvalues of D^c_n. Note that $\Upsilon(\tau^c_\infty)$ has only $k - q$ rows, not k rows.

(ii) If $q = p$ (which requires that $k \geq p$), then $\Upsilon(\tau^c_\infty)$ has no columns, $ACLR_{k,p,q}(\tau^c_\infty) = Z_1'Z_1 \sim \chi^2_p$, and $c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ equals the $1-\alpha$ quantile of the χ^2_p distribution.

(iii) If $q = k$ (which requires that $k \leq p$), then $\Upsilon(\tau^c_\infty)$ and Z_2 have no rows, the $\lambda_{\min}(\cdot)$ expression in $ACLR_{k,p,q}(\tau^c_\infty)$ disappears, $ACLR_{k,p,q}(\tau^c_\infty) = Z'Z \sim \chi^2_k$, and $c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ is the $1-\alpha$ quantile of the χ^2_k distribution.

(iv) If $k \leq p$ and $q < k$, then $(\Upsilon(\tau^c_\infty), Z_2)$ has fewer rows $(k-q)$ than columns $(p-q+1)$ and, hence, the $\lambda_{\min}(\cdot)$ expression in $ACLR_{k,p,q}(\tau^c_\infty)$ equals zero, $ACLR_{k,p,q}(\tau^c_\infty) = Z'Z \sim \chi^2_k$, and $c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ is the $1-\alpha$ quantile of the χ^2_k distribution.

(v) The distribution function (df) of $ACLR_{k,p,q}(\tau^c_\infty)$ is shown in Lemma 22.3 below to be continuous and strictly increasing at its $1-\alpha$ quantile for all possible (k,p,q,τ^c_∞) values, which is required in the proof of Lemma 22.2.

The following lemma proves that the df of $ACLR_{k,p,q}(\tau^c_\infty)$, defined in Lemma 22.2, is continuous and strictly increasing at its $1-\alpha$ quantile. This is a key lemma for showing that the CQLR$_1$ and CQLR$_2$ tests have correct asymptotic size and are asymptotically similar.

Lemma 22.3 *Let* τ^c_∞ *and* $\Upsilon(\tau^c_\infty)$ *be defined as in Lemma* 22.2. *For all admissible integers* (k,p,q) *(i.e.,* $k \geq 1$, $p \geq 1$, *and* $0 \leq q \leq \min\{k,p\}$*) and all* $\min\{k,p\} - q$ (≥ 0) *vectors* τ^c_∞ *with non-negative elements in non-increasing order, the df of* $ACLR_{k,p,q}(\tau^c_\infty) := Z'Z - \lambda_{\min}((\Upsilon(\tau^c_\infty), Z_2)'(\Upsilon(\tau^c_\infty), Z_2))$ *is continuous and strictly increasing at its* $1-\alpha$ *quantile* $c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ *for all* $\alpha \in (0,1)$, *where* $Z := (Z_1', Z_2')' \sim N(0^k, I_k)$ *for* $Z_1 \in R^q$ *and* $Z_2 \in R^{k-q}$.

The next lemma verifies Assumption WU for the choices of $(\widehat{W}_n, \widehat{U}_n)$ that are used to construct the CQLR_1 and CQLR_2 tests. Part (a) of the lemma shows that the parameter space $\mathcal{F}_{WU}$, when defined for $(\widehat{W}_n, \widehat{U}_n)$ as in the CQLR_1 test, contains the parameter space $\mathcal{F}_1$ that appears in the statement of Theorem 10.1 (for suitable choices of the constants δ_1 and M_1 that appear in the definition of $\mathcal{F}_{WU}$). Part (b) of the lemma shows that $\mathcal{F}_{WU}$, when defined for $(\widehat{W}_n, \widehat{U}_n)$ as in the CQLR_2 test, contains $\mathcal{F}_2$ for suitable δ_1 and M_1.

Lemma 22.4 (a) *Suppose* $g_i(\theta) = u_i(\theta)Z_i$, *as in* (4.4), *and* $(\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widehat{L}_n^{1/2})$, *where* $\widehat{\Omega}_n$ $(= \widehat{\Omega}_n(\theta_0))$ *and* $\widehat{L}_n$ $(= \widehat{L}_n(\theta_0))$ *are defined in* (5.1) *and* (6.7), *respectively. Then,* (i) *Assumption* WU *holds for the parameter space* Λ_1 *with* $(\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widehat{R}_n))$, $W_1(W_2) = W_2^{-1/2}$ *for* $W_2 \in R^{k \times k}$, $U_1(U_{2F}) = ((\theta_0, I_p)\Sigma^{-1}(\Omega_F, R_F)(\theta_0, I_p)')^{1/2}$ *for* $U_{2F} = (\Omega_F, R_F)$, $h_7 = \lim W_{2F_{w_n}} := \lim \Omega_{F_{w_n}}$, *and* $h_8 = \lim U_{2F_{w_n}} := \lim(\Omega_{F_{w_n}}, R_{F_{w_n}})$, *where* $\Sigma_F := \Sigma(\Omega_F, R_F)$ *is defined in* (10.8), $\Omega_F := E_F g_i g_i'$, *and* R_F *is defined in* (10.7), *and* (ii) $\mathcal{F}_1 \subset \mathcal{F}_{WU}$ *for* δ_1 *sufficiently small and* M_1 *sufficiently large in the definition of* $\mathcal{F}_{WU}$, *where* $\mathcal{F}_1$ *is defined in* (10.1) *and* $\mathcal{F}_{WU}$ *is defined in* (10.12).

(b) *Suppose* $(\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widetilde{L}_n^{1/2})$, *where* $\widehat{\Omega}_n$ $(= \widehat{\Omega}_n(\theta_0))$ *and* $\widetilde{L}_n$ $(= \widetilde{L}_n(\theta_0))$ *are defined in* (5.1) *and* (7.2). *Then,* (i) *Assumption* WU *holds for the parameter space* Λ_2 *with* $(\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widetilde{R}_n))$, $W_1(\cdot)$ *and* $U_1(\cdot)$ *are defined as in part* (a) *of the lemma,* $h_7 = \lim W_{2F_{w_n}} := \lim \Omega_{F_{w_n}}$, *and* $h_8 = \lim U_{2F_{w_n}} := \lim(\Omega_{F_{w_n}}, \widetilde{R}_{F_{w_n}})$, *where* $\Omega_F := E_F g_i g_i'$ *and* $\widetilde{R}_F$ *is defined in* (10.10), *and* (ii) $\mathcal{F}_2 = \mathcal{F}_{WU}$ *for* δ_1 *sufficiently small and* M_1 *sufficiently large in the definition of* $\mathcal{F}_{WU}$, *where* $\mathcal{F}_2$ *is defined in* (10.1) *and* $\mathcal{F}_{WU}$ *is defined in* (10.12).

Comment: Theorem 22.1, Lemma 22.4, and Proposition 10.2 combine to prove the CQLR test results of Theorem 10.1, which state that the CQLR_1 and CQLR_2 tests have correct asymptotic size and are asymptotically similar (in a uniform sense) for the parameter spaces $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. As stated at the beginning of this section, the proofs of the CQLR CS results of Theorem 10.1 are analogous to those for the tests, see the Comment to Proposition 10.2 and, hence, are not stated explicitly.

22.2 Proof of Theorem 22.1

Theorem 22.1 is stated in Section 22.1.

For notational simplicity, the proof below is given 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 22.1. Let

$$\overline{Z}_h = \begin{pmatrix} \overline{Z}_{h1} \\ \overline{Z}_{h2} \end{pmatrix} := \begin{pmatrix} h_{3,q}' h_{5,g}^{-1/2} \overline{g}_h \\ h_{3,k-q}' h_{5,g}^{-1/2} \overline{g}_h \end{pmatrix} = h_3' h_{5,g}^{-1/2} \overline{g}_h \sim N(0^k, I_k), \tag{22.3}$$

where $\overline{Z}_{h1} \in R^q$ and $\overline{Z}_{h2} \in R^{k-q}$ and the distributional result holds because $\overline{g}_h \sim N(0^k, h_{5,g})$ (by (10.21)) and $h_3' h_3 = \lim C_n' C_n = I_k$. Note that $\overline{Z}_h$ and $(\overline{D}_h, \overline{\Delta}_h)$ are independent because $\overline{g}_h$ and $(\overline{D}_h, \overline{\Delta}_h)$ are independent (by Lemma 10.3(c)).

By Theorem 10.5,

$$\begin{aligned} QLR_n &\rightarrow_d \overline{g}_h' h_{5,g}^{-1} \overline{g}_h - \lambda_{\min}((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)) \\ &= \overline{Z}_h' \overline{Z}_h - \lambda_{\min}((h_{3,k-q}' \overline{\Delta}_{h,p-q}, \overline{Z}_{h2})'(h_{3,k-q}' \overline{\Delta}_{h,p-q}, \overline{Z}_{h2})) =: \overline{QLR}_h, \end{aligned} \tag{22.4}$$

where the equality uses $h_3' h_3 = I_k$. When $q = p$, the term $\overline{\Delta}_{h,p-q}$ does not appear and $\overline{QLR}_h := \overline{Z}_h' \overline{Z}_h - \overline{Z}_{h2}' \overline{Z}_{h2} = \overline{Z}_{h1}' \overline{Z}_{h1}$.

Let $\{\widehat{\tau}_{jn} : j \leq \min\{k,p\}\}$ denote the $\min\{k,p\}$ singular values of $n^{1/2} \widehat{W}_n \widehat{D}_n \widehat{U}_n$ in nonincreasing order. They equal the vector of square roots of the first $\min\{k,p\}$ eigenvalues of $n \widehat{U}_n' \widehat{D}_n \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n$ in nonincreasing order. Define

$$\begin{aligned} \widehat{\tau}_n &= (\widehat{\tau}_{[1]n}', \widehat{\tau}_{[2]n}')' \in R^{\min\{k,p\}}, \text{ where} \\ \widehat{\tau}_{[1]n} &= (\widehat{\tau}_{1n}, ..., \widehat{\tau}_{qn})' \in R^q \text{ and } \widehat{\tau}_{[2]n} = (\widehat{\tau}_{(q+1)n}, ..., \widehat{\tau}_{\min\{k,p\}n})' \in R^{\min\{k,p\}-q}. \end{aligned} \tag{22.5}$$

By Proposition 10.4(a) and (b), $\widehat{\tau}_{jn} \rightarrow_p \infty$ for $j \leq q$ (or, equivalently $Diag^{-1}\{\widehat{\tau}_{[1]n}\} \rightarrow_p 0^{q \times q}$) and

$$\widehat{\tau}_{[2]n} \rightarrow_d \overline{\tau}_{[2]h}, \tag{22.6}$$

where $\widehat{\tau}_{jn} = \widehat{\kappa}_{jn}^{1/2}$ for $j \leq q$ and $\overline{\tau}_{[2]h}$ is the vector of square roots of the first $\min\{k,p\} - q$ eigenvalues of $\overline{\Delta}_{h,p-q}' h_{3,k-q} h_{3,k-q}' \overline{\Delta}_{h,p-q} \in R^{p-q) \times (p-q)}$ in nonincreasing order. (When $q = \min\{k,p\}$, no vector $\overline{\tau}_{[2]h}$ appears.) By an almost sure representation argument, e.g., see Pollard (1990, Thm. 9.4, p. 45), there exists a probability space, say $(\Omega^0, \mathcal{F}^0, P^0)$, and random variables $(QLR_n^0, \widehat{\tau}_n^{0\prime}, \overline{QLR}_h^0, \overline{\tau}_{[2]h}^{0\prime})'$ defined on it such that $(QLR_n^0, \widehat{\tau}_n^{0\prime})'$ has the same distribution as $(QLR_n, \widehat{\tau}_n')'$ for all $n \geq 1$, $(\overline{QLR}_h^0, \overline{\tau}_{[2]h}^{0\prime})'$ has the same distribution as $(\overline{QLR}_h, \overline{\tau}_{[2]h}')'$, and

$$\begin{pmatrix} QLR_n^0 \\ Diag^{-1}\{\widehat{\tau}_{[1]n}^0\} \\ \widehat{\tau}_{[2]n}^0 \end{pmatrix} \rightarrow \begin{pmatrix} \overline{QLR}_h^0 \\ 0^{q \times q} \\ \overline{\tau}_{[2]h}^0 \end{pmatrix} \text{ a.s.}, \tag{22.7}$$

where $\overline{\tau}^0_{[2]h} \in R^{\min\{k,p\}-q}$. Let

$$\widehat{\Upsilon}^0_n := \begin{pmatrix} Diag\{\widehat{\tau}^0_n\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \text{ and } \widehat{\Upsilon}_n := \begin{pmatrix} Diag\{\widehat{\tau}_n\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \text{ if } k \geq p \text{ and} \tag{22.8}$$
$$\widehat{\Upsilon}^0_n := \left(Diag\{\widehat{\tau}^0_n\}, 0^{k\times(p-k)}\right) \in R^{k\times p} \text{ and } \widehat{\Upsilon}_n := \left(Diag\{\widehat{\tau}_n\}, 0^{k\times(p-k)}\right) \in R^{k\times p} \text{ if } k < p.$$

The distributions of $\widehat{\Upsilon}^0_n$ and $\widehat{\Upsilon}_n$ are the same. The matrix $\widehat{\Upsilon}^0_n$ has singular values given by the vector $\widehat{\tau}^0_n$ $(= (\widehat{\tau}^0_{1n}, ..., \widehat{\tau}^0_{\min\{k,p\}n})')$ whose first q elements all diverge to infinity a.s. and whose last $\min\{k,p\} - q$ elements written as the subvector $\widehat{\tau}^0_{[2]n}$ converge to $\overline{\tau}^0_{[2]h}$ a.s. Hence, for some set $C \in \mathcal{F}^0$ with $P^0(\omega \in C) = 1$, we have $\widehat{\tau}^0_{jn}(\omega) \to \infty$ for $j \leq q$ and $\widehat{\tau}^0_{[2]n}(\omega) \to \overline{\tau}^0_{[2]h}(\omega)$, where $\widehat{\tau}^0_{jn}(\omega)$, $\widehat{\tau}^0_{[2]n}(\omega)$, $\overline{\tau}^0_{[2]h}(\omega)$, and $\widehat{\Upsilon}^0_n(\omega)$ denote the realizations of the random quantities $\widehat{\tau}^0_{jn}$, $\widehat{\tau}^0_{[2]n}$, $\overline{\tau}^0_{[2]h}$, and $\widehat{\Upsilon}^0_n$, respectively, when ω occurs. Thus, using Lemma 22.2 with $D^c_n = \widehat{\Upsilon}^0_n(\omega)$ and $\tau^c_\infty = \overline{\tau}^0_{[2]h}(\omega)$, we have

$$c_{k,p}(\widehat{\Upsilon}^0_n(\omega), 1-\alpha) \to c_{k,p,q}(\overline{\tau}^0_{[2]h}(\omega), 1-\alpha) \text{ for all } \omega \in C \text{ with } P^0(\omega \in C) = 1, \tag{22.9}$$

where $c_{k,p,q}(\cdot, 1-\alpha)$ is defined in Lemma 22.2. When $q = \min\{k,p\}$, no vector $\overline{\tau}^0_{[2]h}(\omega)$ appears and by Comments (ii) and (iii) to Lemma 22.2 $c_{k,p,q}(\overline{\tau}^0_{[2]h}(\omega), 1-\alpha)$ equals the $1-\alpha$ quantile of the $\chi^2_{\min\{k,p\}}$ distribution.

Almost sure convergence implies convergence in distribution, so (22.7) and (22.9) also hold (jointly) with convergence in distribution in place of convergence a.s. These convergence in distribution results, coupled with the equality of the distributions of $(QLR^0_n, \widehat{\Upsilon}^0_n)$ and $(QLR_n, \widehat{\Upsilon}_n)$ for all $n \geq 1$ and of $(\overline{QLR}^0_h, \overline{\tau}^{0\prime}_{[2]h})'$ and $(\overline{QLR}_h, \overline{\tau}'_{[2]h})'$, yield the following convergence result:

$$\begin{pmatrix} QLR_n \\ c_{k,p}(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha) \end{pmatrix} = \begin{pmatrix} QLR_n \\ c_{k,p}(\widehat{\Upsilon}_n, 1-\alpha) \end{pmatrix} \to_d \begin{pmatrix} \overline{QLR}_h \\ c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) \end{pmatrix}, \tag{22.10}$$

where the first equality holds using Lemma 6.1.

Equation (22.10) and the continuous mapping theorem give

$$P(QLR_n > c_{k,p}(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha)) \to P(\overline{QLR}_h > c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)) \tag{22.11}$$

provided $P(\overline{QLR}_h = c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)) = 0$. The latter holds because $P(\overline{QLR}_h = c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)|\overline{D}_h) = 0$ a.s. In turn, the latter holds because, conditional on $\overline{D}_h$, the df of $\overline{QLR}_h$ is continuous at its $1-\alpha$ quantile (by Lemma 22.3, where $\overline{QLR}_h$ conditional on $\overline{D}_h$ and $ACLR_{k,p,q}(\tau^c_\infty)$, which appears in Lemma 22.3, have the same structure with the former being based on $h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}$, which is nonrandom conditional on $\overline{D}_h$, and the latter being based on $\Upsilon(\tau_{\infty}^c)$, which is nonrandom, and the former only depends on $h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}$ through its singular values, see (19.3)) and $c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)$ is a constant (because $\overline{\tau}_{[2]h}$ is random only through $\overline{D}_h$).

By the same argument as in the proof of Lemma 6.1,

$$c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) = c_{k,p,q}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, 1-\alpha), \tag{22.12}$$

where (with some abuse of notation) $c_{k,p,q}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, 1-\alpha)$ denotes the $1-\alpha$ quantile of $Z^{\prime}Z - \lambda_{\min}((h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, Z_2)^{\prime}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, Z_2))$ for Z as in Lemma 22.2, because $\overline{\tau}_{[2]h} \in R^{p-q}$ are the singular values of $h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q} \in R^{(k-q)\times(p-q)}$ and $\Upsilon(\overline{\tau}_{[2]h})$ (which appears in $ACLR_{k,p,q}(\overline{\tau}_{[2]h}) = Z^{\prime}Z - \lambda_{\min}((\Upsilon(\overline{\tau}_{[2]h}), Z_2)^{\prime}(\Upsilon(\overline{\tau}_{[2]h}), Z_2)))$ is the $(k-q)\times(p-q)$ matrix with $\overline{\tau}_{[2]h}$ on the main diagonal and zeros elsewhere.

Thus, we have

$$\begin{aligned}
& P(\overline{QLR}_h > c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)) \\
= \ & P(\overline{QLR}_h > c_{k,p,q}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, 1-\alpha)) \\
= \ & EP(\overline{QLR}_h > c_{k,p,q}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, 1-\alpha)|\overline{\Delta}_{h,p-q}) \\
= \ & E\alpha = \alpha,
\end{aligned} \tag{22.13}$$

where the second equality holds by the law of iterated expectations and the third equality holds because, conditional on $\overline{\Delta}_{h,p-q}$, $c_{k,p,q}(h_{3,k-q}^{\prime}\overline{\Delta}_{h,p-q}, 1-\alpha)$ is the $1-\alpha$ quantile of $\overline{QLR}_h$ (by the definitions of $c_{k,p,q}(\cdot, 1-\alpha)$ in Lemma 22.2 and $\overline{QLR}_h$ in (22.4)) and the df of $\overline{QLR}_h$ is continuous at its $1-\alpha$ quantile (see the explanation following (22.11)). $\square$

22.3 Proof of Lemma 22.2

Lemma 22.2 is stated in Section 22.1.

The proof of Lemma 22.2 uses the following two lemmas. Let $\{\tau_{jn}^c : j \leq \min\{k,p\}\}$ be the singular values of D_n^c, as in Lemma 22.2. Define

$$\begin{aligned}
\Upsilon_n^c &:= \begin{pmatrix} Diag\{\tau_{1n}^c, ..., \tau_{pn}^c\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \text{ if } k \geq p \text{ and} \\
\Upsilon_n^c &:= \left(Diag\{\tau_{1n}^c, ..., \tau_{kn}^c\}, 0^{k\times(p-k)}\right) \in R^{k\times p} \text{ if } k < p.
\end{aligned} \tag{22.14}$$

Lemma 22.5 *Suppose the scalar constants* $\{\tau_{jn}^c \geq 0 : j \leq \min\{k,p\}\}$ *for* $n \geq 1$ *satisfy* (i) $\{\tau_{jn}^c \geq 0 : j \leq \min\{k,p\}\}$ *are nonincreasing in* j *for* $n \geq 1$, (ii) $\tau_{jn}^c \to \infty$ *for* $j \leq q$ *for some* $1 \leq q \leq \min\{k,p\}$, (iii) $\tau_{jn}^c \to \tau_{j\infty}^c < \infty$ *for* $j = q+1, ..., \min\{k,p\}$, *and* (iv) *when* $p \geq 2$, $\tau_{(j+1)n}^c / \tau_{jn}^c \to h_{6,j}^c$ *for some* $h_{6,j}^c \in [0,1]$ *for all* $j \leq \min\{k,p\} - 1$. *Let* Υ_n^c *be defined as in* (22.14). *Let* $\{\kappa_{jn}^Z : j \leq p+1\}$ *denote the* $p+1$ *eigenvalues of* $(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)$, *ordered to be nonincreasing in* j, *where* $Z \sim N(0^k, I_k)$. *Then,*

(a) $\kappa_{jn}^Z \to \infty$ $\forall j \leq q$ *for all realizations of* Z *and*

(b) $\kappa_{jn}^Z = o((\tau_{\ell n}^c)^2)$ $\forall \ell \leq q$ *and* $\forall j = q+1, ..., p+1$ *for all realizations of* Z.

Comment: Lemma 22.5 only applies when $q \geq 1$, whereas Lemma 22.2 applies when $q \geq 0$.

Lemma 22.6 *Let* $\{F_n^*(x) : n \geq 1\}$ *and* $F^*(x)$ *be df's on* R *and let* $\alpha \in (0,1)$ *be given. Suppose* (i) $F_n^*(x) \to F^*(x)$ *for all continuity points* x *of* $F^*(x)$ *and* (ii) $F^*(q_\infty + \varepsilon) > 1 - \alpha$ *for all* $\varepsilon > 0$, *where* $q_\infty := \inf\{x : F^*(x) \geq 1 - \alpha\}$ *is the* $1 - \alpha$ *quantile of* $F^*(x)$. *Then, the* $1 - \alpha$ *quantile of* $F_n^*(x)$, *viz.,* $q_n := \inf\{x : F_n^*(x) \geq 1 - \alpha\}$, *satisfies* $q_n \to q_\infty$.

Comment: Condition (ii) of Lemma 22.6 requires that $F^*(x)$ is increasing at its $1 - \alpha$ quantile.

Proof of Lemma 22.2. By Lemma 6.1, $c_{k,p}(D_n^c, 1-\alpha) = c_{k,p}(\Upsilon_n^c, 1-\alpha)$, where Υ_n^c is defined in (22.14). Hence, it suffices to show that $c_{k,p}(\Upsilon_n^c, 1-\alpha) \to c_{k,p,q}(\tau_\infty^c, 1-\alpha)$. To prove the latter, it suffices to show that for any subsequence $\{w_n\}$ of $\{n\}$ there exists a subsubsequence $\{u_n\}$ such that $c_{k,p}(\Upsilon_{u_n}^c, 1-\alpha) \to c_{k,p,q}(\tau_\infty^c, 1-\alpha)$. When $p \geq 2$, given $\{w_n\}$, we select a subsubsequence $\{u_n\}$ for which $\tau_{(j+1)u_n}^c / \tau_{ju_n}^c \to h_{6,j}^c$ for some constant $h_{6,j}^c \in [0,1]$ for all $j = 1, ..., \min\{k,p\} - 1$ (where $0/0 := 0$). We can select a subsubsequence with this property because every sequence of numbers in $[0,1]$ has a convergent subsequence by the compactness of $[0,1]$.

For notational simplicity, when $p \geq 2$, we prove the full sequence result that $c_{k,p}(\Upsilon_n^c, 1-\alpha) \to c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ under the assumption that

$$\tau_{(j+1)n}^c / \tau_{jn}^c \to h_{6,j}^c \text{ for all } j \leq \min\{k,p\} - 1 \tag{22.15}$$

(as well as the other assumptions on the singular values stated in the theorem).[80] The same argument holds with n replaced by u_n below, which is the result that is needed to complete the proof. When $p = 1$, we prove the full sequence result that $c_{k,p}(\Upsilon_n^c, 1-\alpha) \to c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ without the condition in (22.15) (which is meaningless in this case because there is only one value $\tau_{ju_n}^c$, namely $\tau_{1u_n}^c$, for each n). In this case too, the same argument holds with n replaced by u_n

[80] The condition in (22.15) is required by Lemma 22.5, which is used in the proof of Lemma 22.2 below.

below, which is the result that is needed to complete the proof. We treat the cases $p \geq 2$ and $p = 1$ simultaneously from here on.

First, we show that

$$\begin{aligned} CLR_{k,p}(\Upsilon_n^c) &:= Z'Z - \lambda_{\min}((\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)) \\ &\rightarrow Z'Z - \lambda_{\min}((\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2)) := ACLR_{k,p,q}(\tau_\infty^c) \end{aligned} \tag{22.16}$$

for all realizations of Z. If $q = 0$, then (22.16) holds because $\Upsilon_n^c \rightarrow \Upsilon(\tau_\infty^c)$ (by the definition of Υ_n^c in (22.14), the definition of $\Upsilon(\tau_\infty^c)$ in the statement of the Lemma 22.2, and assumption (iii) of Lemma 22.2) and the minimum eigenvalue of a matrix is a continuous function of the matrix (by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)).

Now, we establish (22.16) when $q \geq 1$. The (ordered) eigenvalues $\{\kappa_{jn}^Z : j \leq p+1\}$ of $(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)$ are solutions to

$$\begin{aligned} &|(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) - \kappa I_{p+1}| = 0 \text{ or} \\ &|Q_n^c(\kappa)| = 0, \text{ where } Q_n^c(\kappa) := S_n^c(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)S_n^c - \kappa(S_n^c)^2 \text{ and} \\ &S_n^c := Diag\{(\tau_{1n}^c)^{-1}, ..., (\tau_{qn}^c)^{-1}, 1, ..., 1\} \in R^{(p+1)\times(p+1)}. \end{aligned} \tag{22.17}$$

Define

$$S_{n,q}^c := Diag\{(\tau_{1n}^c)^{-1}, ..., (\tau_{qn}^c)^{-1}\} \in R^{q\times q}. \tag{22.18}$$

We have

$$\begin{aligned} (\Upsilon_n^c, Z)S_n^c &= \left((\Upsilon_n^c, Z) \begin{pmatrix} I_q \\ 0^{(p+1-q)\times q} \end{pmatrix} S_{n,q}^c, (\Upsilon_n^c, Z) \begin{pmatrix} 0^{q\times(p+1-q)} \\ I_{p+1-q} \end{pmatrix} \right) \\ &= (I_{k,q}, \Upsilon_{n,p-q}^c, Z) \in R^{k\times(p+1)}, \text{ where} \\ I_{k,q} &:= \begin{pmatrix} I_q \\ 0^{(k-q)\times q} \end{pmatrix} \in R^{k\times q}, \\ \Upsilon_{n,p-q}^c &:= \begin{pmatrix} 0^{q\times(p-q)} \\ Diag\{\tau_{(q+1)n}^c, ..., \tau_{pn}^c\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \in R^{k\times(p-q)} \text{ if } k \geq p, \text{ and} \\ \Upsilon_{n,p-q}^c &:= \begin{pmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ Diag\{\tau_{(q+1)n}^c, ..., \tau_{kn}^c\} & 0^{(k-q)\times(p-k)} \end{pmatrix} \in R^{k\times(p-q)} \text{ if } k < p. \end{aligned} \tag{22.19}$$

By (22.17) and (22.19), we have

$$Q_n^c(\kappa) = \left[\begin{array}{cc} I_q & I_{k,q}'(\Upsilon_{n,p-q}^c, Z) \\ (\Upsilon_{n,p-q}^c, Z)' I_{k,q} & (\Upsilon_{n,p-q}^c, Z)'(\Upsilon_{n,p-q}^c, Z) \end{array} \right] - \kappa \left[\begin{array}{cc} (S_{n,q}^c)^2 & 0^{q \times (p+1-q)} \\ 0^{(p+1-q) \times q} & I_{p+1-q} \end{array} \right]. \quad (22.20)$$

By the formula for the determinant of a partitioned inverse (see the footnote above),

$$\begin{aligned} & |Q_n^c(\kappa)| = |Q_{n,1}^c(\kappa)| \cdot |Q_{n,2}^c(\kappa)|, \text{ where} \\ & Q_{n,1}^c(\kappa) := I_q - \kappa (S_{n,q}^c)^2 \in R^{q \times q} \text{ and} \\ & Q_{n,2}^c(\kappa) := (\Upsilon_{n,p-q}^c, Z)'(\Upsilon_{n,p-q}^c, Z) - \kappa I_{p+1-q} \qquad (22.21) \\ & \qquad -(\Upsilon_{n,p-q}^c, Z)' I_{k,q} (I_q - \kappa (S_{n,q}^c)^2)^{-1} I_{k,q}'(\Upsilon_{n,p-q}^c, Z) \in R^{(p+1-q) \times (p+1-q)}. \end{aligned}$$

For $j = q+1, ..., p+1$, we have

$$Q_{n,1}^c(\kappa_{jn}^Z) = I_q - \kappa_{jn}^Z (S_{n,q}^c)^2 = I_q - Diag\{\kappa_{jn}^Z (\tau_{1n}^c)^{-2}, ..., \kappa_{jn}^Z (\tau_{qn}^c)^{-2}\} = I_q + o(1) \qquad (22.22)$$

for all realizations of Z, where the last equality holds by Lemma 22.5 (which applies for $q \geq 1$). This implies that $|Q_{n,1}^c(\kappa_{jn}^Z)| \neq 0$ for $j = q+1, ..., p+1$ for n large. Hence, for n large,

$$|Q_{n,2}^c(\kappa_{jn}^Z)| = 0 \text{ for } j = q+1, ..., p+1. \qquad (22.23)$$

We write

$$I_k = (I_{k,q}, I_{k,k-q}), \text{ where } I_{k,k-q} := \left(\begin{array}{c} 0^{q \times (k-q)} \\ I_{k-q} \end{array} \right) \in R^{k \times (k-q)} \qquad (22.24)$$

and $I_{k,q}$ is defined in (22.19).[81]

For $j = q+1, ..., p+1$, we have

$$\begin{aligned} Q_{n,2}^c(\kappa_{jn}^Z) &= (\Upsilon_{n,p-q}^c, Z)'(\Upsilon_{n,p-q}^c, Z) - \kappa_{jn}^Z I_{p+1-q} - (\Upsilon_{n,p-q}^c, Z)' I_{k,q} (I_q + o(1)) I_{k,q}'(\Upsilon_{n,p-q}^c, Z) \\ &= (\Upsilon_{n,p-q}^c, Z)' I_{k,k-q} I_{k,k-q}'(\Upsilon_{n,p-q}^c, Z) + o(1) - \kappa_{jn}^Z I_{p+1-q} \\ &:= M_{n,p+1-q}^c - \kappa_{jn}^Z I_{p+1-q}, \qquad (22.25) \end{aligned}$$

where the first equality holds by (22.22) and the definition of $Q_{n,2}^c(\kappa)$ in (22.21) and the second equality holds because $I_k = (I_{k,q}, I_{k,k-q})(I_{k,q}, I_{k,k-q})' = I_{k,q} I_{k,q}' + I_{k,k-q} I_{k,k-q}'$ and $\Upsilon_{n,p-q}^c = O(1)$ by its definition in (22.19) and the condition (iii) of Lemma 22.2 on $\{\tau_{jn}^c : j = q+1, ..., \min\{k, p\}\}$

[81] There is some abuse of notation here because $I_{k,q}$ does not equal $I_{k,k-q}$ even if q equals $k - q$.

for $n \geq 1$.

Equations (22.23) and (22.25) imply that $\{\kappa^Z_{jn} : j = q+1, ..., p+1\}$ are the $p+1-q$ eigenvalues of the matrix $M^c_{n,p+1-q}$. By the definition of $\Upsilon^c_{n,p-q}$ in (22.19) and the conditions of the theorem on $\{\tau^c_{jn} : j = q+1, ..., \min\{k, p\}\}$ for $n \geq 1$, we have

$$\begin{aligned} M^c_{n,p+1-q} &\to \left(\begin{pmatrix} 0^{q\times(p-q)} \\ \Upsilon(\tau^c_\infty) \end{pmatrix}, Z \right)' I_{k,k-q} I'_{k,k-q} \left(\begin{pmatrix} 0^{q\times(p-q)} \\ \Upsilon(\tau^c_\infty) \end{pmatrix}, Z \right) \\ &= (\Upsilon(\tau^c_\infty), Z_2)'(\Upsilon(\tau^c_\infty), Z_2) \end{aligned} \tag{22.26}$$

for all realizations of Z, where the equality uses the definitions of $\Upsilon(\tau^c_\infty)$ and Z_2 in the statement of the theorem.

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)). Hence, by (22.26), the eigenvalues $\{\kappa^Z_{jn} : j = q+1, ..., p+1\}$ of $M^c_{n,p+1-q}$ converge (for all realizations of Z) to the vector of eigenvalues of $(\Upsilon(\tau^c_\infty), Z_2)'(\Upsilon(\tau^c_\infty), Z_2)$. In consequence, the smallest eigenvalue $\kappa^Z_{(p+1)n}$ (of both $M^c_{n,p+1-q}$ and $(\Upsilon^c_n, Z)'(\Upsilon^c_n, Z)$) satisfies

$$\lambda_{\min}((\Upsilon^c_n, Z)'(\Upsilon^c_n, Z)) = \kappa^Z_{(p+1)n} \to \lambda_{\min}((\Upsilon(\tau^c_\infty), Z_2)'(\Upsilon(\tau^c_\infty), Z_2)), \tag{22.27}$$

where the equality holds by the definition of $\kappa^Z_{(p+1)n}$ in (22.17). This establishes (22.16).

Now we use (22.16) to establish that $c_{k,p}(\Upsilon^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$, which proves the theorem. Let

$$F_{k,p,q,\tau^c_\infty}(x) = P(ACLR_{k,p,q}(\tau^c_\infty) \leq x). \tag{22.28}$$

By (22.16), for any $x \in R$ that is a continuity point of $F_{k,p,q,\tau^c_\infty}(x)$, we have

$$1(CLR_{k,p}(\Upsilon^c_n) \leq x) \to 1(ACLR_{k,p,q}(\tau^c_\infty) \leq x) \text{ a.s.} \tag{22.29}$$

Equation (22.29) and the bounded convergence theorem give

$$P(CLR_{k,p}(\Upsilon^c_n) \leq x) \to P(ACLR_{k,p,q}(\tau^c_\infty) \leq x) = F_{k,p,q,\tau^c_\infty}(x). \tag{22.30}$$

Now Lemma 22.6 gives the desired result, because (22.30) verifies assumption (i) of Lemma 22.6 and the df of $ACLR_{k,p,q}(\tau^c_\infty)$ is strictly increasing at its $1-\alpha$ quantile (by Lemma 22.3), which verifies assumption (ii) of Lemma 22.6. $\square$

Proof of Lemma 22.5. The proof is similar to the proof of Lemma 16.1 given in Section 16 in the SM of AG1. But there are enough differences that we provide a proof.

By the definition of q (≥ 1) in the statement of Lemma 22.5, $h_{6,q}^c = 0$ if $q < \min\{k,p\}$. If $q = \min\{k,p\}$, then $h_{6,q}^c$ is not defined in the statement of Lemma 22.5 and we define it here to equal zero. If $h_{6,j}^c > 0$, then $\{\tau_{jn}^c : n \geq 1\}$ and $\{\tau_{(j+1)n}^c : n \geq 1\}$ are of the same order of magnitude, i.e., $0 < \lim \tau_{(j+1)n}^c/\tau_{jn}^c \leq 1$. We group the first q values of τ_{jn}^c into groups that have the same order of magnitude within each group. Let G ($\in \{1,...,q\}$) denote the number of groups. Note that G equals the number of values in $\{h_{6,1}^c,...,h_{6,q}^c\}$ that equal zero. Let r_g and r_g^c denote the indices of the first and last values in the gth group, respectively, for $g = 1,...,G$. Thus, $r_1 = 1$, $r_g^c = r_{g+1} - 1$, where by definition $r_{G+1} = q+1$, and $r_G^c = q$. By definition, the τ_{jn}^c values in the gth group, which have the gth largest order of magnitude, are $\{\tau_{r_g n}^c : n \geq 1\},...,\{\tau_{r_g^c n}^c : n \geq 1\}$. By construction, $h_{6,j}^c > 0$ for all $j \in \{r_g,...,r_g^c - 1\}$ for $g = 1,...,G$. (The reason is: if $h_{6,j}^c$ is equal to zero for some $j \leq r_g^c - 1$, then $\{\tau_{r_g^c n}^c : n \geq 1\}$ is of smaller order of magnitude than $\{\tau_{r_g n}^c : n \geq 1\}$, which contradicts the definition of r_g^c.) Also by construction, $\lim \tau_{j'n}^c/\tau_{jn}^c = 0$ for any (j,j') in groups (g,g'), respectively, with $g < g'$.

The (ordered) eigenvalues $\{\kappa_{jn}^Z : j \leq p+1\}$ of $(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)$ are solutions to the determinantal equation $|(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) - \kappa I_{p+1}| = 0$. Equivalently, they are solutions to

$$|(\tau_{r_1 F_n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) - (\tau_{r_1 F_n}^c)^{-2}\kappa I_{p+1}| = 0. \tag{22.31}$$

Thus, $\{(\tau_{r_1 n}^c)^{-2}\kappa_{jn}^Z : j \leq p+1\}$ solve

$$|(\tau_{r_1 n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) - \kappa I_{p+1}| = 0. \tag{22.32}$$

Let

$$h_{6,r_1^c}^{cc} := Diag\{1, h_{6,1}^c, h_{6,1}^c h_{6,2}^c, ..., \prod_{\ell=1}^{r_1^c - 1} h_{6,\ell}^c\} \in R^{r_1^c \times r_1^c}. \tag{22.33}$$

When $k \geq p$, we have

$$\begin{aligned}
&(\tau_{r_1 n}^c)^{-1}(\Upsilon_n^c, Z) \\
&= \begin{bmatrix} h_{6,r_1^c}^{cc} + o(1) & 0^{r_1^c \times (q - r_1^c)} & 0^{r_1^c \times (p-q)} & O(1/\tau_{r_1 n}^c)^{r_1^c \times 1} \\ 0^{(q-r_1^c) \times r_1^c} & O(\tau_{r_2 n}^c/\tau_{r_1 n}^c)^{(q-r_1^c)\times(q-r_1^c)} & 0^{(q-r_1^c)\times(p-q)} & O(1/\tau_{r_1 n}^c)^{(q-r_1^c)\times 1} \\ 0^{(p-q)\times r_1^c} & 0^{(p-q)\times(q-r_1^c)} & O(1/\tau_{r_1 n}^c)^{(p-q)\times(p-q)} & O(1/\tau_{r_1 n}^c)^{(p-q)\times 1} \\ 0^{(k-p)\times r_1^c} & 0^{(k-p)\times(q-r_1^c)} & 0^{(k-p)\times(p-q)} & O(1/\tau_{r_1 n}^c)^{(k-p)\times 1} \end{bmatrix} \\
&\to \begin{bmatrix} h_{6,r_1^c}^{cc} & 0^{r_1^c \times (p+1-r_1^c)} \\ 0^{(k-r_1^c)\times r_1^c} & 0^{(k-r_1^c)\times(p+1-r_1^c)} \end{bmatrix},
\end{aligned} \tag{22.34}$$

where $O(d_n)^{s\times s}$ denotes a diagonal $s\times s$ matrix whose elements are $O(d_n)$ for some scalar constants $\{d_n : n\geq 1\}$, $O(d_n)^{s\times 1}$ denotes an s vector whose elements are $O(d_n)$, the equality uses $\tau_{jn}^c/\tau_{r_1 n}^c = \prod_{\ell=1}^{j-1}(\tau_{(\ell+1)n}^c/\tau_{\ell n}^c) = \prod_{\ell=1}^{j-1} h_{6,\ell}^c + o(1)$ for $j = 2, ..., r_1^c$ (which holds by the definition of $h_{6,\ell}^c$) and $\tau_{jn}^c/\tau_{r_1 n}^c = O(\tau_{r_2 n}^c/\tau_{r_1 n}^c)$ for $j = r_2, ..., q$ (because $\{\tau_{jn}^c : j\leq q\}$ are nonincreasing in j), and the convergence uses $\tau_{r_1 n}^c \to \infty$ (by assumption (ii) of the lemma since $r_1\leq q$) and $\tau_{r_2 n}^c/\tau_{r_1 n}^c \to 0$ (by the definition of r_2).

When $k < p$, (22.34) holds but with the rows dimensions of the submatrices in the second line changed by replacing $p-q$ by $k-q$ and $k-p$ by $p-k$ four times each.

Equation (22.34) yields

$$(\tau_{r_1 n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) \to \begin{bmatrix} (h_{6,r_1^c}^{cc})^2 & 0^{r_1^c\times(p+1-r_1^c)} \\ 0^{(p+1-r_1^c)\times r_1^c} & 0^{(p+1-r_1^c)\times(p+1-r_1^c)} \end{bmatrix}. \tag{22.35}$$

The 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 (22.32) and (22.35), the first r_1^c eigenvalues of $(\tau_{r_1 n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)$, i.e., $\{(\tau_{r_1 n}^c)^{-2}\kappa_{jn}^Z : j\leq r_1^c\}$, satisfy

$$((\tau_{r_1 n}^c)^{-2}\kappa_{1n}^Z, ..., (\tau_{r_1 n}^c)^{-2}\kappa_{r_1^c n}^Z) \to_p (1, h_{6,1}^c, h_{6,1}^c h_{6,2}^c, ..., \prod_{\ell=1}^{r_1^c-1} h_{6,\ell}^c) \text{ and so}$$
$$\kappa_{1n}^Z \to \infty\ \forall j = 1, ..., r_1^c \tag{22.36}$$

because $\tau_{r_1 n}^c\to\infty$ (since $r_1\leq q$) and $h_{6,\ell}^c > 0$ for all $\ell\in\{1, ..., r_1^c - 1\}$ (as noted above). By the same argument, the last $p+1-r_1^c$ eigenvalues of $(\tau_{r_1 n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)$, i.e., $\{(\tau_{r_1 n}^c)^{-2}\kappa_{jn}^Z : j = r_1^c + 1, ..., p+1\}$, satisfy

$$(\tau_{r_1 n}^c)^{-2}\kappa_{jn}^Z \to 0\ \forall j = r_1^c + 1, ..., p+1. \tag{22.37}$$

Next, the equality in (22.34) gives

$$(\tau_{r_1 n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) \tag{22.38}$$
$$= \begin{bmatrix} (h_{6,r_1^c}^{cc})^2 + o(1) & 0^{r_1^c\times(q-r_1^c)} & 0^{r_1^c\times(p-q)} & O(1/\tau_{r_1 n}^c)^{r_1^c\times 1} \\ 0^{(q-r_1^c)\times r_1^c} & O((\tau_{r_2 n}^c/\tau_{r_1 n}^c)^2)^{(q-r_1^c)\times(q-r_1^c)} & 0^{(q-r_1^c)\times(p-q)} & O(\tau_{r_2 n}^c/(\tau_{r_1 n}^c)^2)^{(q-r_1^c)\times 1} \\ 0^{(p-q)\times r_1^c} & 0^{(p-q)\times(q-r_1^c)} & O(1/(\tau_{r_1 n}^c)^2)^{(p-q)\times(p-q)} & O(1/(\tau_{r_1 n}^c)^2)^{(p-q)\times 1} \\ O(1/\tau_{r_1 n}^c)^{1\times r_1^c} & O(\tau_{r_2 n}^c/(\tau_{r_1 n}^c)^2)^{1\times(q-r_1^c)} & O(1/(\tau_{r_1 n}^c)^2)^{1\times(p-q)} & O(1/(\tau_{r_1 n}^c)^2)^{1\times 1} \end{bmatrix}.$$

Equation (22.38) holds when $k\geq p$ and $k<p$ (because the column dimensions of the submatrices in the second line of (22.34) are the same when $k\geq p$ and $k<p$).

Define I_{j_1,j_2} to be the $(p+1)\times(j_2-j_1)$ matrix that consists of the $j_1+1,...,j_2$ columns of I_{p+1} for $0\leq j_1<j_2\leq p+1$. We can write

$$I_{p+1}=(I_{0,r_1^c},I_{r_1^c,p+1}),\ \text{ where } I_{0,r_1^c}:=\begin{pmatrix} I_{r_1^c} \\ 0^{(p+1-r_1^c)\times r_1^c}\end{pmatrix}\in R^{(p+1)\times r_1^c} \text{ and}$$
$$I_{r_1^c,p+1}:=\begin{pmatrix} 0^{r_1^c\times(p+1-r_1^c)} \\ I_{p+1-r_1^c}\end{pmatrix}\in R^{(p+1)\times(p+1-r_1^c)}. \tag{22.39}$$

In consequence, we have

$$\begin{aligned}(\Upsilon_n^c,Z) &= ((\Upsilon_n^c,Z)I_{0,r_1^c},(\Upsilon_n^c,Z)I_{r_1^c,p+1}) \text{ and} \\ \varrho_n^c &:= (\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{r_1^c,p+1}=o(\tau_{r_2n}^c/\tau_{r_1n}^c),\end{aligned} \tag{22.40}$$

where the last equality uses the first row of the matrix on the rhs of (22.38) and $O(1/\tau_{r_1n}^c)=o(\tau_{r_2n}^c/\tau_{r_1n}^c)$ (because $\tau_{r_2n}^c\to\infty$).

As in (22.32), $\{(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z : j\leq p+1\}$ solve

$$\begin{aligned}0 &= |(\tau_{r_1n}^c)^{-2}(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)-\kappa I_{p+1}| \\ &= \left|\left[\begin{array}{cc} (\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{0,r_1^c}-\kappa I_{r_1^c} & (\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{r_1^c,p+1} \\ (\tau_{r_1n}^c)^{-2}I_{r_1^c,p+1}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{0,r_1^c} & (\tau_{r_1n}^c)^{-2}I_{r_1^c,p+1}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{r_1^c,p+1}-\kappa I_{p+1-r_1^c}\end{array}\right]\right| \\ &= |(\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{0,r_1^c}-\kappa I_{r_1^c}| \\ &\quad \times|(\tau_{r_1n}^c)^{-2}I_{r_1^c,p+1}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{r_1^c,p+1}-\kappa I_{p+1-r_1^c} \\ &\quad -\varrho_n^{c\prime}((\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{0,r_1^c}-\kappa I_{r_1^c})^{-1}\varrho_n^c|,\end{aligned} \tag{22.41}$$

where the third equality uses the standard formula for the determinant of a partitioned matrix, the definition of ϱ_n^c in (22.40), and the result given in (22.42) below that the matrix which is inverted that appears in the last line of (22.41) is nonsingular for κ equal to any solution $(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z$ to the first equality in (22.41) for $j=r_1^c+1,...,p+1$.

Now we show that, for $j=r_1^c+1,...,p+1$, $(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z$ cannot solve the determinantal equation $|(\tau_{r_1n}^c)^{-2}I_{0,r_1^c}'(\Upsilon_n^c,Z)'(\Upsilon_n^c,Z)I_{0,r_1^c}-\kappa I_{r_1^c}|=0$, where this determinant is the first multiplicand on the rhs of (22.41). Hence, $\{(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z : j=r_1^c+1,...,p+1\}$ must solve the determinantal equation based on the second multiplicand on the rhs of (22.41). For $j = r_1^c + 1, ..., p + 1$, we have

$$(\tau_{r_1 n}^c)^{-2} I'_{0,r_1^c}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) I_{0,r_1^c} - (\tau_{r_1 n}^c)^{-2} \kappa_{jn}^Z I_{r_1^c} = (h_{6,r_1^c}^{cc})^2 + o(1), \quad (22.42)$$

where the equality holds by (22.35) and (22.37). Equation (22.42) and $\lambda_{\min}((h_{6,r_1^c}^{cc})^2) > 0$ (which follows from the definition of $h_{6,r_1^c}^{cc}$ in (22.33) and the fact that $h_{6,j}^c > 0$ for all $j \in \{1, ..., r_1^c - 1\}$) establish the desired result.

For $j = r_1^c + 1, ..., p + 1$, plugging $(\tau_{r_1 n}^c)^{-2} \kappa_{jn}^Z$ into the second multiplicand on the rhs of (22.41) and using (22.40) and (22.42) gives

$$0 = |(\tau_{r_1 n}^c)^{-2} I'_{r_1^c,p+1}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) I_{r_1^c,p+1} + o((\tau_{r_2 F_n}^c / \tau_{r_1 F_n}^c)^2) - (\tau_{r_1 n}^c)^{-2} \kappa_{jn}^Z I_{p+1-r_1^c}|. \quad (22.43)$$

Thus, $\{(\tau_{r_1 n}^c)^{-2} \kappa_{jn}^Z : j = r_1^c + 1, ..., p + 1\}$ solve

$$0 = |(\tau_{r_1 n}^c)^{-2} I'_{r_1^c,p+1}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) I_{r_1^c,p+1} + o((\tau_{r_2 F_n}^c / \tau_{r_1 F_n}^c)^2) - \kappa I_{p+1-r_1^c}|. \quad (22.44)$$

Or equivalently, multiplying through by $(\tau_{r_2 F_n}^c / \tau_{r_1 F_n}^c)^{-2}$, $\{(\tau_{r_2 n}^c)^{-2} \kappa_{jn}^Z : j = r_1^c + 1, ..., p + 1\}$ solve

$$0 = |(\tau_{r_2 n}^c)^{-2} I'_{r_1^c,p+1}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) I_{r_1^c,p+1} + o(1) - \kappa I_{p+1-r_1^c}| \quad (22.45)$$

by the same argument as in (22.31) and (22.32).

Now, we repeat the argument from (22.32) to (22.45) with the expression in (22.45) replacing that in (22.32) and with $I_{p+1-r_1^c}$, $\tau_{r_2 n}^c$, $\tau_{r_3 n}^c$, $r_2^c - r_1^c$, $p + 1 - r_2^c$, and $h_{6,r_2^c}^{cc} = Diag\{1, h_{6,r_1^c+1}^c, h_{6,r_1^c+1}^c h_{6,r_1^c+2}^c, ..., \prod_{\ell=r_1^c+1}^{r_2^c-1} h_{6,\ell}^c\} \in R^{(r_2^c - r_1^c) \times (r_2^c - r_1^c)}$ in place of I_{p+1}, $\tau_{r_1 n}^c$, $\tau_{r_2 n}^c$, r_1^c, $p + 1 - r_1^c$, and $h_{6,r_1^c}^{cc}$, respectively. In addition, I_{0,r_1^c} and $I_{r_1^c,p+1}$ in (22.41) are replaced by the matrices $I_{r_1^c,r_2^c}$ and $I_{r_2^c,p+1}$. This argument gives

$$\kappa_{jn}^Z \to \infty \ \forall j = r_2, ..., r_2^c \text{ and } (\tau_{r_2 n}^c)^{-2} \kappa_{jn}^Z = o(1) \ \forall j = r_2^c + 1, ..., p + 1. \quad (22.46)$$

Repeating the argument $G - 2$ more times yields

$$\kappa_{jn}^Z \to \infty \ \forall j = 1, ..., r_G^c \text{ and } (\tau_{r_g n}^c)^{-2} \kappa_{jn}^Z = o(1) \ \forall j = r_g^c + 1, ..., p + 1, \forall g = 1, ..., G. \quad (22.47)$$

Note that "repeating the argument $G - 2$ more times" is justified by an induction argument that is analogous to that given in the proof of Lemma 16.1 given in Section 16 in the SM of AG1.

Because $r_J^c = q$, the first result in (22.47) proves part (a) of the lemma.

The second result in (22.47) with $g = G$ implies: for all $j = q+1, ..., p+1$,

$$(\tau_{r_G n}^c)^{-2} \kappa_{jn}^Z = o(1) \tag{22.48}$$

because $r_G^c = q$. Either $r_G = r_G^c = q$ or $r_G < r_G^c = q$. In the former case, $(\tau_{qn}^c)^{-2} \kappa_{jn}^Z = o(1)$ for $j = q+1, ..., p+1$ by (22.47). In the latter case, we have

$$\lim \frac{\tau_{qn}^c}{\tau_{r_G n}^c} = \lim \frac{\tau_{r_G^c n}^c}{\tau_{r_G n}^c} = \prod_{j=r_G}^{r_G^c - 1} h_{6,j}^c > 0, \tag{22.49}$$

where the inequality holds because $h_{6,j}^c > 0$ for all $j \in \{r_G, ..., r_G^c - 1\}$, as noted at the beginning of the proof. Hence, in this case too, $(\tau_{qn}^c)^{-2} \kappa_{jn}^Z = o(1)$ for $j = q+1, ..., p+1$ by (22.48) and (22.49). Because $\tau_{jn}^c \geq \tau_{qn}^c$ for all $j \leq q$, this establishes part (b) of the lemma. $\square$

Proof of Lemma 22.6. For $\varepsilon > 0$ such that $q_\infty \pm \varepsilon$ are continuity points of $F^*(x)$, we have

$$\begin{aligned} F_n^*(q_\infty - \varepsilon) &\to F^*(q_\infty - \varepsilon) < 1 - \alpha \text{ and} \\ F_n^*(q_\infty + \varepsilon) &\to F^*(q_\infty + \varepsilon) > 1 - \alpha \end{aligned} \tag{22.50}$$

by assumptions (i) and (ii) of the lemma and $F^*(q_\infty - \varepsilon) < 1 - \alpha$ by the definition of q_∞. The first line of (22.50) implies that $q_n \geq q_\infty - \varepsilon$ for all n large. (If not, there exists an infinite subsequence $\{w_n\}$ of $\{n\}$ for which $q_{w_n} < q_\infty - \varepsilon$ for all $n \geq 1$ and $1 - \alpha \leq F_{w_n}^*(q_{w_n}) \leq F_{w_n}^*(q_\infty - \varepsilon) \to F^*(q_\infty - \varepsilon) < 1 - \alpha$, which is a contradiction). The second line of (22.50) implies that $q_n \leq q_\infty + \varepsilon$ for all n large. There exists a sequence $\{\varepsilon_k > 0 : k \geq 1\}$ for which $\varepsilon_k \to 0$ and $q_\infty \pm \varepsilon_k$ are continuity points of $F^*(x)$ for all $k \geq 1$. Hence, $q_n \to q_\infty$. $\square$

22.4 Proof of Lemma 22.3

Lemma 22.3 is stated in Section 22.1.

Proof of Lemma 22.3. We prove the lemma by proving it separately for four cases: (i) $q \geq 1$, (ii) $k \leq p$, (iii) $\tau_{\min\{k,p\}\infty}^c = 0$, where $\tau_{\min\{k,p\}\infty}^c$ denotes the $\min\{k,p\}$th (and, hence, last and smallest) element of τ_∞^c, and (iv) $q = 0$, $k > p$, and $\tau_{p\infty}^c > 0$. First, suppose $q \geq 1$. Then,

$$\begin{aligned} ACLR_{k,p,q}(\tau_\infty^c) &:= Z'Z - \lambda_{\min}((\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2)) \\ &= Z_1'Z_1 + Z_2'Z_2 - \lambda_{\min}((\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2)) \end{aligned} \tag{22.51}$$

and $ACLR_{k,p,q}(\tau_\infty^c)$ is the convolution of a χ_q^2 distribution (since $Z_1'Z_1 \sim \chi_q^2$) and another distribution. Consider the distribution of $X + Y$, where X is a random variable with an absolutely continuous distribution and X and Y are independent. Let B be a (measurable) subset of R with Lebesgue measure zero. Then,

$$P(X + Y \in B) = \int P(X + y \in B | Y = y) dP_Y(y) = \int P(X \in B - y) dP_Y(y) = 0, \tag{22.52}$$

where P_Y denotes the distribution of Y, the first equality holds by the law of iterated expectations, the second equality holds by the independence of X and Y, and the last equality holds because X is absolutely continuous and the Lebesgue measure of $B - y$ equals zero. Applying (22.52) to (22.51) with $X = Z_1' Z_1$, we conclude that $ACLR_{k,p,q}(\tau_\infty^c)$ is absolutely continuous and, hence, its df is continuous at its $1 - \alpha$ quantile for all $\alpha \in (0, 1)$.

Next, we consider the df of $X + Y$, where X has support R_+ and X and Y are independent. Let c denote the $1 - \alpha$ quantile of $X + Y$ for $\alpha \in (0, 1)$, and let c_Y denote the $1 - \alpha$ quantile of Y. Since $X \geq 0$ a.s., $c_Y \leq c$. Hence, for all $\varepsilon > 0$,

$$P(Y < c + \varepsilon) \geq P(Y < c_Y + \varepsilon) \geq 1 - \alpha > 0. \tag{22.53}$$

For $\varepsilon > 0$, we have

$$\begin{aligned} P(X + Y \in [c, c + \varepsilon]) &= \int P(X + y \in [c, c + \varepsilon] | Y = y) dP_Y(y) \\ &= \int P(X \in [c - y, c - y + \varepsilon]) dP_Y(y) > 0, \end{aligned} \tag{22.54}$$

where the first equality holds by the law of iterated expectations, the second equality holds by the independence of X and Y, and the inequality holds because $P(X \in [c - y, c - y + \varepsilon]) > 0$ for all $y < c + \varepsilon$ (because the support of X is R_+) and $P(Y < c + \varepsilon) > 0$ by (22.53). Equation (22.54) implies that the df of $X + Y$ is strictly increasing at its $1 - \alpha$ quantile.

For the case when $q \geq 1$, we apply the result of the previous paragraph with $ACLR_{k,p,q}(\tau_\infty^c) = X + Y$ and $Z_1' Z_1 = X$. This implies that the df of $ACLR_{k,p,q}(\tau_\infty^c)$ is strictly increasing at its $1 - \alpha$ quantile when $q \geq 1$.

Second, suppose $k \leq p$. Then, $(\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2) \in R^{(p-q+1) \times (p-q+1)}$ is singular because $(\Upsilon(\tau_\infty^c), Z_2) \in R^{(k-q) \times (p-q+1)}$ and $k - q < p - q + 1$. Hence, $\lambda_{\min}((\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2)) = 0$, $ACLR_{k,p,q}(\tau_\infty^c) = Z'Z \sim \chi_k^2$, $ACLR_{k,p,q}(\tau_\infty^c)$ is absolutely continuous, and the df of $ACLR_{k,p,q}(\tau_\infty^c)$ is continuous and strictly increasing at its $1 - \alpha$ quantile for all $\alpha \in (0, 1)$.

Third, suppose $\tau_{\min\{k,p\}\infty}^c = 0$. Then, $\lambda_{\min}((\Upsilon(\tau_\infty^c), Z_2)'(\Upsilon(\tau_\infty^c), Z_2)) = 0$, $ACLR_{k,p,q}(\tau_\infty^c) = Z'Z \sim \chi_k^2$, $ACLR_{k,p,q}(\tau_\infty^c)$ is absolutely continuous, and the df of $ACLR_{k,p,q}(\tau_\infty^c)$ is continuous and strictly increasing at its $1-\alpha$ quantile for all $\alpha \in (0,1)$.

Fourth, suppose $q=0$, $k>p$, and $\tau^c_{p\infty}>0$. In this case, $Z_2=Z$ (because $q=0$) and $\Upsilon(\tau^c_\infty)=(D,0^{p\times(k-p)})'$, where $D:=Diag\{\tau^c_\infty\}$ is a pd diagonal $p\times p$ matrix (because $\tau^c_{p\infty}>0$). We write $Z=(Z_a',Z_b')'$ $(\sim N(0^k,I_k))$, where $Z_a\in R^p$ and $Z_b\in R^{k-p}$ and Z_b has a positive number of elements (because $k>p$). Let $ACLR$ abbreviate $ACLR_{k,p,q}(\tau^c_\infty)$. In the present case, we have

$$\begin{aligned} ACLR &= Z'Z-\lambda_{\min}\left(\begin{pmatrix} D & Z_a \\ 0^{(k-p)\times p} & Z_b\end{pmatrix}'\begin{pmatrix} D & Z_a \\ 0^{(k-p)\times p} & Z_b\end{pmatrix}\right) \\ &= Z'Z-\inf_{\xi=(\xi_1',\xi_2)':||\xi||=1}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}'\begin{pmatrix} D^2 & DZ_a \\ Z_a'D & Z'Z\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \\ &= \sup_{\xi=(\xi_1',\xi_2)':||\xi||=1}\left[(1-\xi_2^2)(Z_b'Z_b+Z_a'Z_a)-\xi_1'D^2\xi_1-2\xi_2Z_a'D\xi_1\right], \end{aligned} \tag{22.55}$$

where $\xi_1\in R^p$, $\xi_2\in R$, and $\xi_1'\xi_1+\xi_2^2=1$.

We define the following non-stochastic function

$$ACLR(z_a,\omega):=\sup_{\xi=(\xi_1',\xi_2)':||\xi||=1}\left[(1-\xi_2^2)(\omega+z_a'z_a)-\xi_1'D^2\xi_1-2\xi_2z_a'D\xi_1\right] \tag{22.56}$$

for $z_a\in R^p$ and $\omega\in R_+$. Note that $ACLR=ACLR(Z_a,Z_b'Z_b)$.

We show below that the function $ACLR(z_a,\omega)$ is (i) nonnegative, (ii) strictly increasing in ω on R_+ $\forall z_a\neq 0^p$, and (iii) continuous in (z_a,ω) on $R^p\times R_+$, and $ACLR(z_a,\omega)$ satisfies (iv) $\lim_{\omega\to\infty}ACLR(z_a,\omega)=\infty$. In consequence, $\forall z_a\neq 0^p$, $ACLR(z_a,\omega)$ has a continuous, strictly-increasing inverse function in its second argument with domain $[ACLR(z_a,0),\infty)\subset R_+$, which we denote by $ACLR^{-1}(z_a,x)$.[82] Using this, we have: for all $x\geq ACLR(z_a,0)$ and $z_a\neq 0^p$,

$$ACLR(z_a,\omega)\leq x \text{ iff } \omega\leq ACLR^{-1}(z_a,x), \tag{22.57}$$

where the condition $x\geq ACLR(z_a,0)$ ensures that x is in the domain of $ACLR^{-1}(z_a,\cdot)$.

Now, we show that for all $x_0\in R$ and $z_a\neq 0^p$,

$$\lim_{x\to x_0}P(ACLR(z_a,Z_b'Z_b)\leq x)=P(ACLR(z_a,Z_b'Z_b)\leq x_0). \tag{22.58}$$

[82] Properties (i), (iii), and (iv) determine the domain of $ACLR^{-1}(z_a,x)$ for its second argument.

To prove (22.58), first consider the case $x_0 > ACLR(z_a, 0)$ (≥ 0) and $z_a \neq 0^p$. In this case, we have

$$\begin{aligned} \lim_{x \to x_0} P(ACLR(z_a, Z_b'Z_b) \leq x) &= \lim_{x \to x_0} P(Z_b'Z_b \leq ACLR^{-1}(z_a, x)) \\ &= P(Z_b'Z_b \leq ACLR^{-1}(z_a, x_0)), \end{aligned} \tag{22.59}$$

where the first equality holds by (22.57) and the second equality holds by the continuity of the df of the χ^2_{k-p} random variable $Z_b'Z_b$ and the continuity of $ACLR^{-1}(z_a, x)$ at x_0. Hence, (22.58) holds when $x_0 > ACLR(z_a, 0)$.

Next, consider the case $x_0 < ACLR(z_a, 0)$ and $z_a \neq 0^p$. We have

$$P(ACLR(z_a, Z_b'Z_b) \leq x_0) \leq P(ACLR(z_a, Z_b'Z_b) < ACLR(z_a, 0)) = 0, \tag{22.60}$$

where the equality holds because $ACLR(z_a, x)$ is increasing on by property (ii) and $Z_b'Z_b \geq 0$ a.s. For x sufficiently close to x_0, $x < ACLR(z_a, 0)$ and by the same argument as in (22.60), we obtain $P(ACLR(z_a, Z_b'Z_b) \leq x) = 0$. Thus, (22.58) holds for $x_0 < ACLR(z_a, 0)$.

Finally, consider the case $x_0 = ACLR(z_a, 0)$ and $z_a \neq 0^p$. In this case, (22.58) holds for sequences of values x that strictly decline to x_0 by the same argument as for the first case where $x_0 > ACLR(z_a, 0)$. Next, consider a sequence that strictly increases to x_0. We have $P(ACLR(z_a, Z_b'Z_b) \leq x) = 0$ $\forall x < x_0$ by the same argument as given for the second case where $x_0 < ACLR(z_a, 0)$. In addition, we have

$$P(ACLR(z_a, Z_b'Z_b) \leq x_0) = P(ACLR(z_a, Z_b'Z_b) \leq ACLR(z_a, 0)) \leq P(Z_b'Z_b \leq 0) = 0, \tag{22.61}$$

where the inequality holds because $ACLR(z_a, x)$ is strictly increasing on for $z_a \neq 0^p$ by property (ii). This completes the proof of (22.58).

Using (22.58), we establish the continuity of the df of $ACLR$ on R. For any $x_0 \in R$, we have

$$\begin{aligned} \lim_{x \to x_0} P(ACLR \leq x) &= \lim_{x \to x_0} P(ACLR(Z_a, Z_b'Z_b) \leq x) \\ &= \lim_{x \to x_0} \int P(ACLR(z_a, Z_b'Z_b) \leq x) dF_{Z_a}(z_a) \\ &= \int P\left(ACLR(z_a, Z_b'Z_b) \leq x_0\right) dF_{Z_a}(z_a) \\ &= P(ACLR \leq x_0), \end{aligned} \tag{22.62}$$

where $F_{Z_a}(\cdot)$ denotes the df of Z_a, the first and last equalities hold because $ACLR = ACLR(Z_a, Z_b'Z_b)$, the second equality uses the independence of Z_a and Z_b, and the third equality holds by the bounded convergence theorem using (22.58) and $P(Z_a \neq 0^p) = 1$. Equation (22.62) shows that the df of $ACLR$ is continuous on R.

Next, we show that the df of $ACLR$ is strictly increasing at all $x > 0$. Because the df of $ACLR$ is continuous on R and equals 0 for $x \leq 0$ (because $ACLR \geq 0$ by property (i)), the $1-\alpha$ quantile of $ACLR$ is positive. Hence, the former property implies that the df of $ACLR$ is increasing at its $1-\alpha$ quantile, as stated in the Lemma.

For $x \geq ACLR(z_a, 0)$, $\delta > 0$, and $z_a \neq 0^p$, we have

$$P(ACLR(z_a, Z_b'Z_b) \in [x, x+\delta]) = P\left(Z_b'Z_b \in [ACLR^{-1}(z_a, x), ACLR^{-1}(z_a, x+\delta)]\right) > 0, \quad (22.63)$$

where the equality holds by (22.57) and the inequality holds because $ACLR^{-1}(z_a, x)$ is strictly increasing in x for x in $[ACLR(z_a, 0), \infty)$ when $z_a \neq 0^p$ and $Z_b'Z_b$ has a χ^2_{k-p} distribution, which is absolutely continuous.

The function $ACLR(z_a, 0)$ is continuous at all $z_a \in R^p$ (by property (iii)) and $ACLR(0^p, 0) = 0$ (by a simple calculation using (22.56)). In consequence, for any $x > 0$, there exists a vector $z_a^* \in R^p$ and a constant $\varepsilon > 0$ such that $ACLR(z_a, 0) < x$ for all $z_a \in B(z_a^*, \varepsilon)$, where $B(z_a^*, \varepsilon)$ denotes a ball centered at z_a^* with radius $\varepsilon > 0$. Using this, we have: for any $x > 0$ and $\delta > 0$,

$$\begin{aligned} P(ACLR \in [x, x+\delta]) &= \int P(ACLR(z_a, Z_b'Z_b) \in [x, x+\delta])dF_{Z_a}(z_a) \\ \geq \int_{B(z_a^*, \varepsilon)} & P(ACLR(z_a, Z_b'Z_b) \in [x, x+\delta])dF_{Z_a}(z_a) > 0, \end{aligned} \quad (22.64)$$

where the equality uses the independence of Z_a and Z_b, the first inequality holds because $B(z_a^*, \varepsilon) \subset R$ and the integrand is nonnegative, and the second inequality holds because $P(Z_a \in B(z_a^*, \varepsilon)) > 0$ (since $Z_a \sim N(0^p, I_p)$ and $B(z_a^*, \varepsilon)$ is a ball with positive radius) and the integrand is positive for $z_a \in B(z_a^*, \varepsilon)$ by (22.63) using the fact that $x > ACLR(z_a, 0)$ for all $z_a \in B(z_a^*, \varepsilon)$ by the definition of $B(z_a^*, \varepsilon)$. Equation (22.64) shows that the df of $ACLR$ is strictly increasing at all $x > 0$ and, hence, at its $1-\alpha$ quantile which is positive.

It remains to verify properties (i)-(iv) of the function $ACLR(z_a, \omega)$, which are stated above. The function $ACLR(z_a, \omega)$ is seen to be nonnegative by replacing the supremum in (22.56) by $\xi = (0^{p\prime}, 1)'$. Hence, property (i) holds. The function $ACLR(z_a, \omega)$ can be written as

$$ACLR(z_a, \omega) = \omega + z_a'z_a - \lambda_{\min}\begin{pmatrix} D^2 & Dz_a \\ z_a'D & z_a'z_a + \omega \end{pmatrix} \quad (22.65)$$

by analogous calculations to those in (22.55). The minimum eigenvalue is a continuous function of a matrix is a continuous function of its elements by Elsner's Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38). Hence, $ACLR(z_a, \omega)$ is continuous in $(z_a, \omega) \in R^p \times R_+$ and property (iii) holds.

For any $\xi_{*2}^2 \in [0, 1)$ and $\xi_{*1} \in R^p$ such that $\xi_{*1}'\xi_{*1} = 1 - \xi_{*2}^2$, we have

$$ACLR(z_a, \omega) \geq (1 - \xi_{*2}^2)(\omega + z_a'z_a) - \xi_{*1}'D^2\xi_{*1} - 2\xi_{*2}z_a'D\xi_{*1} \to \infty \text{ as } \omega \to \infty, \qquad (22.66)$$

where the inequality holds by replacing the supremum over ξ in (22.56) by the same expression evaluated at $\xi_* = (\xi_{*1}', \xi_{*2})'$ and the divergence to infinity uses $1 - \xi_{*2}^2 > 0$. Hence, property (iv) holds.

It remains to verify property (ii), which states that $ACLR(z_a, \omega)$ is strictly increasing in ω on R_+ $\forall z_a \neq 0^p$. For $\omega \in R_+$, let $\xi_\omega = (\xi_{\omega 1}', \xi_{\omega 2})'$ (for $\xi_{\omega 1} \in R^p$ and $\xi_{\omega 2} \in R$) be such that $||\xi_\omega|| = 1$ and

$$ACLR(z_a, \omega) = (1 - \xi_{\omega 2}^2)(\omega + z_a'z_a) - \xi_{\omega 1}'D^2\xi_{\omega 1} - 2\xi_{\omega 2}z_a'D\xi_{\omega 1}. \qquad (22.67)$$

Such a vector ξ_ω exists because the supremum in (22.56) is the supremum of a continuous function over a compact set and, hence, the supremum is attained at some vector ξ_ω. (Note that ξ_ω typically depends on z_a as well as ω.) Using (22.67), we obtain: for all $\delta > 0$, if $\xi_{\omega 2}^2 < 1$,

$$\begin{aligned} ACLR(z_a, \omega) &< (1 - \xi_{\omega 2}^2)(\omega + \delta + z_a'z_a) - \xi_{\omega 1}'D^2\xi_{\omega 1} - 2\xi_{\omega 2}z_a'D\xi_{\omega 1} \\ &\leq \sup_{\xi = (\xi_1', \xi_2)' : ||\xi|| = 1} \left[(1 - \xi_2^2)(\omega + \delta + z_a'z_a) - \xi_1'D^2\xi_1 - 2\xi_2 z_a'D\xi_1\right] \\ &= ACLR(z_a, \omega + \delta). \end{aligned} \qquad (22.68)$$

Equation (22.68) shows that $ACLR(z_a, \omega)$ is strictly increasing at ω provided $\xi_{\omega 2}^2 < 1$.

Next, we show that $\xi_{\omega 2}^2 = 1$ only if $z_a = 0^p$. By (22.56) and (22.67), ξ_ω maximizes the rhs expression in (22.56) over $\xi \in R^{p+1}$ subject to $\xi_1'\xi_1 + \xi_2^2 = 1$. The Lagrangian for the optimization problem is

$$(1 - \xi_2^2)(\omega + z_a'z_a) - \xi_1'D^2\xi_1 - 2\xi_2 z_a'D\xi_1 + \gamma(1 - \xi_2^2 - \xi_1'\xi_1), \qquad (22.69)$$

where $\gamma \in R$ is the Lagrange multiplier. The first-order conditions of the Lagrangian with respect to ξ_1, evaluated at the solution $(\xi_{\omega 1}', \xi_{\omega 2})'$ and the corresponding Lagrange multiplier, say γ_ω, are

$$-2D^2\xi_{\omega 1} - 2\xi_{\omega 2}Dz_a - 2\gamma_\omega \xi_{\omega 1} = 0^p. \qquad (22.70)$$

The solution is $\xi_{\omega 1} = 0^p$ (which is an interior point of the set $\{\xi_1 : ||\xi_1|| \leq 1\}$) only if $\xi_{\omega 2} = 0$ or $z_a = 0^p$ (because D is a pd diagonal matrix). Thus, $\xi_{\omega 2}^2 = 1 - \xi_{\omega 1}'\xi_{\omega 1} = 1$ only if $z_a = 0^p$. This concludes the proof of property (iv). $\square$

22.5 Proof of Lemma 22.4

Lemma 22.4 is stated in Section 22.1.

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 Lemma 22.4. We prove part (a)(i) first. We have

$$\widehat{W}_{2n} = n^{-1} \sum_{i=1}^{n} (g_i g_i' - E_{F_n} g_i g_i') + E_{F_n} g_i g_i' \to_p h_{5,g}, \tag{22.71}$$

where the convergence holds by the WLLN (using the moment conditions in $\mathcal{F}_2$) and $\lambda_{7,F_n} = W_{2F_n} = \Omega_{F_n} := E_{F_n} g_i g_i' \to h_{5,g}$ (by the definition of the sequence $\{\lambda_{n,h} : n \geq 1\}$). Hence, Assumption WU(a) holds for the parameter space Λ_1 with $h_7 = h_{5,g}$.

Next, we verify Assumption WU(b) for the parameter space Λ_1 for $\widehat{U}_{2n} = (\widehat{\Omega}_n, \widehat{R}_n)$. Using the definition of $\widehat{V}_n$ $(= \widehat{V}_n(\theta_0))$ in (6.3), we have

$$\begin{aligned} \widehat{V}_n = {} & n^{-1} \sum_{i=1}^{n} (u_i^* u_i^{*\prime} \otimes Z_i Z_i') - n^{-1} \sum_{i=1}^{n} (\widehat{u}_{in}^* u_i^{*\prime} \otimes Z_i Z_i') - n^{-1} \sum_{i=1}^{n} (u_i^* \widehat{u}_{in}^{*\prime} \otimes Z_i Z_i') \\ & + n^{-1} \sum_{i=1}^{n} (\widehat{u}_{in}^* \widehat{u}_{in}^{*\prime} \otimes Z_i Z_i'). \end{aligned} \tag{22.72}$$

We have

$$\begin{aligned} & n^{-1} \sum_{i=1}^{n} (u_i^* u_i^{*\prime} \otimes Z_i Z_i') = E_{F_n} f_i f_i' + o_p(1), \\ & \widehat{\Xi}_n = (n^{-1} Z_{n\times k}' Z_{n\times k})^{-1} n^{-1} Z_{n\times k}' U^* = (E_{F_n} Z_i Z_i')^{-1} E_{F_n} Z_i u_i^{*\prime} + o_p(1) \\ & \quad = (E_{F_n} Z_i Z_i')^{-1} E_{F_n} (g_i, G_i) + o_p(1) =: \Xi_{F_n} + o_p(1), \qquad (22.73) \\ & n^{-1} \sum_{i=1}^{n} (\widehat{u}_{in}^* u_i^{*\prime} \otimes Z_i Z_i') = n^{-1} \sum_{i=1}^{n} (\widehat{\Xi}_n' Z_i u_i^{*\prime} \otimes Z_i Z_i') = E_{F_n} (\Xi_{F_n}' (g_i, G_i) \otimes Z_i Z_i') + o_p(1), \text{ and} \\ & n^{-1} \sum_{i=1}^{n} (\widehat{u}_{in}^* \widehat{u}_{in}^{*\prime} \otimes Z_i Z_i') = n^{-1} \sum_{i=1}^{n} (\widehat{\Xi}_n' Z_i Z_i' \widehat{\Xi}_n \otimes Z_i Z_i') = E_{F_n} (\Xi_{F_n}' Z_i Z_i' \Xi_{F_n} \otimes Z_i Z_i') + o_p(1), \end{aligned}$$

where the first line holds by the WLLN's (since $u_i^* u_i^{*\prime} \otimes Z_i Z_i' = f_i f_i'$ for f_i defined in (10.7) and using the moment conditions in $\mathcal{F}_2$), the second line holds by the WLLN's (using the conditions in $\mathcal{F}_1$ and $\mathcal{F}_2$), Slutsky's Theorem, and $Z_i u_i^{*\prime} = (g_i, G_i)$, the fourth line holds by the WLLN's (using $E_F((||(g_i, G_i)|| \cdot ||Z_i||^2)^{1+\gamma/4}) \leq (E_F ||(g_i, G_i)||^{2+\gamma/2} E_F ||Z_i||^{4+\gamma})^{1/2} < \infty$ for $\gamma > 0$ by the Cauchy-Bunyakovsky-Schwarz inequality and the moment conditions in $\mathcal{F}_1$ and $\mathcal{F}_2$) and the result of the second and third lines, and the fifth line holds by the WLLN's (using the moment conditions in $\mathcal{F}_1$

and $\mathcal{F}_2$) and the result of the second and third lines.

Equations (10.7) (which defines V_F), (22.72), and (22.73) combine to give

$$\widehat{V}_n - V_{F_n} \to_p 0. \tag{22.74}$$

Using the definitions of $\widehat{R}_n$ and R_F (in (6.3) and (10.7)), (22.71), (22.74), and $h_7 := \lim W_{2F_n} = \lim \Omega_{F_n}$ yield

$$(\widehat{\Omega}_n, \widehat{R}_n) \to_p \lim(\Omega_{F_n}, R_{F_n}) =: h_8. \tag{22.75}$$

This establishes Assumption WU(b) for the parameter space Λ_1 for part (a) of the lemma.

Now we establish Assumption WU(c) for the parameter space Λ_1 for part (a) of the lemma. We take $\mathcal{W}_2$ (which appears in the statement of Assumption WU(c)) to be the space of psd $k \times k$ matrices and $\mathcal{U}_2$ (which also appears in Assumption WU(c)) to be the space of non-zero psd matrices (Ω, R) for $\Omega \in R^{k \times k}$ and $R \in R^{(p+1)k \times (p+1)k}$. By the definition of $\widehat{W}_{2n}$, $\widehat{W}_{2n} \in \mathcal{W}_2$ a.s. We have $W_{2F} \in \mathcal{W}_2 \ \forall F \in \mathcal{F}_{WU}$ because $W_{2F} = E_F g_i g_i'$ is psd. We have $U_{2F} \in \mathcal{U}_2 \ \forall F \in \mathcal{F}_{WU}$ because $U_{2F} = (\Omega_F, R_F)$, $\Omega_F := E_F g_i g_i'$ is psd and non-zero (by the last condition in $\mathcal{F}_2$, even if that condition is weaken to $\lambda_{\max}(E_F g_i g_i') \geq \delta$) and $R_F := (B' \otimes I_k) V_F (B \otimes I_k)$ is psd and non-zero because B (defined in (6.3)) is nonsingular and V_F (defined in (10.7)) is non-zero by the argument given in the paragraph following (22.78) below. By their definitions, $\widehat{\Omega}_n$ and $\widehat{R}_n$ are psd. In addition, they are non-zero wp→1 by (22.75) and the result just established that the two matrices that comprise h_8 are non-zero. Hence, $(\widehat{\Omega}_n, \widehat{R}_n) \in \mathcal{U}_2$ wp→1.

The function $W_1(W_2) = W_2^{-1/2}$ is continuous at $W_2 = h_7$ on $\mathcal{W}_2$ because $\lambda_{\min}(h_7) > 0$ (given that $h_7 = \lim E_{F_n} g_i g_i'$ and $\lambda_{\min}(E_F g_i g_i') \geq \delta$ by the last condition in $\mathcal{F}_2$).

The function $U_1(\cdot)$ defined in (10.8) is well-defined in a neighborhood of h_8 and continuous at h_8 provided all psd matrices $\Omega \in R^{k \times k}$ and $R \in R^{(p+1)k \times (p+1)k}$ with (Ω, R) in a neighborhood of $h_8 := \lim(\Omega_{F_n}, R_{F_n})$ are such that $\Sigma^{\varepsilon}(\Omega, R)$ is nonsingular, where $\Sigma(\Omega, R)$ is defined in the paragraph containing (10.8) with (Ω, R) in place of (Ω_F, R_F) and $\Sigma^{\varepsilon}(\Omega, R)$ is defined given $\Sigma(\Omega, R)$ by (6.6). Lemma 17.1(b) shows that $\Sigma^{\varepsilon}(\Omega, R)$ is nonsingular provided $\lambda_{\max}(\Sigma(\Omega, R)) > 0$. We have

$$\begin{aligned}
&\lambda_{\max}(\Sigma(\Omega, R)) \geq \max_{j \leq p+1} \Sigma_{jj}(\Omega, R) = \max_{j \leq p+1} tr(\Omega^{-1/2} R_{jj} \Omega^{-1/2})/k \\
\geq\ & \max_{j \leq p+1} \lambda_{\max}(\Omega^{-1/2} R_{jj} \Omega^{-1/2})/k = \max_{j \leq p+1} \sup_{\lambda: ||\lambda||=1} \frac{\lambda' \Omega^{-1/2}}{||\Omega^{-1/2}\lambda||} R_{jj} \frac{\Omega^{-1/2}\lambda}{||\Omega^{-1/2}\lambda||} \cdot ||\Omega^{-1/2}\lambda||^2/k \\
\geq\ & \max_{j \leq p+1} \lambda_{\max}(R_{jj}) \lambda_{\min}(\Omega^{-1})/k > 0,
\end{aligned} \tag{22.76}$$

where $\Sigma_{jj}(\Omega, R)$ denotes the (j, j) element of $\Sigma(\Omega, R)$, R_{jj} denotes the (j, j) $k \times k$ submatrix of

R, the first inequality holds by the definition of $\lambda_{\max}(\cdot)$, the first equality holds by (6.5) with (Ω, R) in place of $(\widehat{\Omega}_n(\theta), \widehat{R}_n(\theta))$, the second inequality holds because the trace of a psd matrix equals the sum of its eigenvalues by a spectral decomposition, the third inequality holds by the definition of $\lambda_{\min}(\cdot)$, and the last inequality holds because the conditions in $\mathcal{F}_2$ imply that $\lambda_{\min}(\Omega^{-1}) = 1/\lambda_{\max}(\Omega) > 0$ for Ω in some neighborhood of $\lim \Omega_{F_n}$ (because $\lambda_{\max}(\Omega_F)$ $= \sup_{\lambda \in R^k : ||\lambda||=1} E_F(\lambda' g_i)^2 \leq E_F||g_i||^2 \leq M^{2/(2+\gamma)} < \infty$ for all $F \in \mathcal{F}_2$ using the Cauchy-Bunyakovsky-Schwarz inequality) and $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(R_F) > 0$, which we show below, implies that $\lambda_{\max}(R_{jj}) > 0$ for some $j \leq p+1$.

To establish Assumption WU(c) for part (a) of the lemma, it remains to show that

$$\inf_{F \in \mathcal{F}_2} \lambda_{\max}(R_F) > 0. \tag{22.77}$$

We show that the last condition in $\mathcal{F}_2$, i.e., $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(E_F g_i g_i') > 0$ implies (22.77). In fact, the last condition in $\mathcal{F}_2$ is very much stronger than is needed to get (22.77). (The full strength of the last condition in $\mathcal{F}_2$ is used in the proof of Lemma 10.3, see Section 20, because $\widehat{\Omega}_n^{-1/2}$ enters the definition of $\widehat{D}_n$ and $\widehat{\Omega}_n - \Omega_{F_n} \rightarrow_p 0^{k \times k}$, where $\Omega_F = E_F g_i g_i'$.) We show that (22.77) holds provided $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(E_F g_i g_i') > 0$.

Let $x^* \in R^{(p+1)k}$ be such that $||x^*|| = 1$ and $\lambda_{\max}(V_F) = x^{*\prime} V_F x^*$. Let $x^\dagger = (B \otimes I_k)^{-1} x^*$. Then, we have

$$\begin{aligned}
&\lambda_{\max}(R_F) := \lambda_{\max}((B' \otimes I_k) V_F (B \otimes I_k)) = \sup_{x \in R^{(p+1)k} : ||x||=1} x'(B' \otimes I_k) V_F (B \otimes I_k) x \\
&\geq x^{\dagger\prime}(B' \otimes I_k) V_F (B \otimes I_k) x^\dagger \cdot ||x^\dagger||^{-2} = x^{*\prime} V_F x^* / (x^{*\prime} (B \otimes I_k)^{-1\prime} (B \otimes I_k)^{-1} x^*) \\
&\geq \lambda_{\max}(V_F) / \lambda_{\max}((B \otimes I_k)^{-1\prime} (B \otimes I_k)^{-1}) \geq K \lambda_{\max}(V_F),
\end{aligned} \tag{22.78}$$

where $K := 1/\lambda_{\max}((B \otimes I_k)^{-1\prime}(B \otimes I_k)^{-1})$ is positive and does not depend on F (because B and $B \otimes I_k$ are nonsingular and do not depend on F for $B = B(\theta_0)$ defined in (6.3)).

Next, $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(V_F) \geq \inf_{F \in \mathcal{F}_2} \lambda_{\max}(E_F g_i g_i')$ because V_F can be written as $E_F(u_i^* - \Xi_F' Z_i)(u_i^* - \Xi_F' Z_i)' \otimes Z_i Z_i'$, the first element of $\Xi_F' Z_i$ is zero (because $\Xi_F := (E_F Z_i Z_i')^{-1} E_F(g_i, G_i)$, see (10.7), and $E_F g_i = 0^k$), the first element of $u_i^* - \Xi_F' Z_i = u_i$ (because $u_i^* = (u_i, u_{\theta i}')'$), the upper left $k \times k$ submatrix of V_F equals $E_F u_i^2 Z_i Z_i' = E_F g_i g_i'$, and so, $\lambda_{\max}(V_F) \geq \lambda_{\max}(E_F g_i g_i')$. This result and (22.78) imply that (22.77) holds provided $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(E_F g_i g_i') > 0$. As noted above, the latter is implied by the last condition in $\mathcal{F}_2$. This completes the verification (22.77) and the verification of Assumption WU(c) in part (a) of the lemma.

Now, we prove part (a)(ii) of the lemma. We need to show that the four conditions in the definition of $\mathcal{F}_{WU}$ in (10.12) hold.

(I) We show that $\inf_{F \in \mathcal{F}_1} \lambda_{\min}(W_F) > 0$, where $W_F := W_1(W_{2F}) := \Omega_F^{-1/2} := (E_F g_i g_i')^{-1/2}$ (by (10.5) and the paragraph containing (10.6)). The inequality $E_F||g_i||^{2+\gamma} \leq M$ in $\mathcal{F}_2$ implies $\lambda_{\min}(W_F) \geq \delta_1$ for δ_1 sufficiently small (because the latter holds if $\lambda_{\max}(W_F^{-2}) \leq \delta_1^{-2}$ and $W_F^{-2} = \Omega_F = E_F g_i g_i'$.

(II) We show that $\sup_{F \in \mathcal{F}_2} ||W_F|| < \infty$, where $W_F := W_1(W_{2F}) := \Omega_F^{-1/2} := (E_F g_i g_i')^{-1/2}$ (by (10.5) and the paragraph containing (10.11)). We have $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(\Omega_F) > 0$ (by the last condition in $\mathcal{F}_2$).

(III) We show that $\inf_{F \in \mathcal{F}_1} \lambda_{\min}(U_F) > 0$, where in the present case $U_F := U_1(U_{2F}) := ((\theta_0, I_p)(\Sigma^{\varepsilon}(\Omega_F, R_F))^{-1}(\theta_0, I_p)')^{1/2}$ and $\Sigma(\Omega_F, R_F)$ has (j, ℓ) element equal to $tr(R_{j\ell F}' \Omega_F^{-1})/k$ (by (10.8)). The inequalities $E_F||Z_i||^{4+\gamma} \leq M$, $E_F||(g_i', vec(G_i)')'||^{2+\gamma} \leq M$, and $\lambda_{\min}(E_F Z_i Z_i') \geq \delta$ imply that $\sup_{F \in \mathcal{F}_1}(||\Xi_F|| + ||E_F f_i f_i'|| + ||E_F(\Xi_F' Z_i Z_i' \Xi_F \otimes Z_i Z_i')|| + ||E_F(g_i, G_i)\Xi_F \otimes Z_i Z_i'||) < \infty$, where Ξ_F is defined in (10.7) (using the Cauchy-Bunyakovsky-Schwarz inequality). This, in turn, implies that $\sup_{F \in \mathcal{F}_1} ||V_F|| < \infty$, $\sup_{F \in \mathcal{F}_1} ||R_F|| < \infty$, $\sup_{F \in \mathcal{F}_1} ||\Sigma_F|| < \infty$, $\sup_{F \in \mathcal{F}_1} ||\Sigma_F^{\varepsilon}|| < \infty$, and $\lambda_{\min}(L_F) \geq \delta_2$ for some $\delta_2 > 0$, where V_F and R_F are defined in (10.7), $\Sigma_F := \Sigma(\Omega_F, R_F)$, $L_F := (\theta_0, I_p)(\Sigma_F^{\varepsilon})^{-1}(\theta_0, I_p)'$, and $(\Sigma_F^{\varepsilon})^{-1}$ exists by (IV) below (and $\lambda_{\min}(L_F) \geq \delta_2$ holds because $A := (\theta_0, I_p) \in R^{p \times (p+1)}$ has full row rank p, and $\lambda_{\min}(L_F) = \inf_{\lambda \in R^p : ||\lambda|| = 1} \lambda' A (\Sigma_F^{\varepsilon})^{-1} A' \lambda$ $\geq \inf_{\lambda \in R^p : ||\lambda|| = 1} (A'\lambda)'(\Sigma_F^{\varepsilon})^{-1}(A'\lambda)/||A'\lambda||^2 \times \inf_{\lambda \in R^p : ||\lambda|| = 1} ||A'\lambda||^2 = \lambda_{\min}((\Sigma_F^{\varepsilon})^{-1})\lambda_{\min}(AA') \geq \delta_2$ for some $\delta_2 > 0$ that does not depend on F). Finally, $\lambda_{\min}(L_F) \geq \delta_2$ implies the desired result that $\lambda_{\min}(U_F) \geq \delta_1$ for some $\delta_1 > 0$ (because $U_F := L_F^{1/2}$).

(IV) We show that $\sup_{F \in \mathcal{F}_1} ||U_F|| < \infty$, where U_F is as in (III) immediately above. By the same calculations as in (22.76) (which use (22.77)) with Σ_F and (Ω_F, R_F) in place of $\Sigma(\Omega, R)$ and (Ω, R), respectively, we have $\inf_{F \in \mathcal{F}_1} \lambda_{\max}(\Sigma_F) > 0$. The latter implies $\inf_{F \in \mathcal{F}_1} \lambda_{\min}(\Sigma_F^{\varepsilon}) > 0$ by Lemma 17.1(b). In turn, the latter implies the desired result $\sup_{F \in \mathcal{F}_1} ||U_F|| = \sup_{F \in \mathcal{F}_1} ||((\theta_0, I_p)(\Sigma_F^{\varepsilon})^{-1} \times (\theta_0, I_p)')^{1/2}|| < \infty$.

Results (I)-(IV) establish the result of part (a)(ii).

Now, we prove part (b)(i) of the lemma. Assumption WU(a) holds for the parameter space Λ_2 with $h_7 = h_{5,g}$ by the same argument as for part (a)(i). Next, we establish Assumption WU(b) for the parameter space Λ_2. Using the definition of $\widetilde{V}_n$ $(= \widetilde{V}_n(\theta_0))$ in (7.1), we have

$$\widetilde{V}_n = n^{-1} \sum_{i=1}^{n} f_i f_i' - \widehat{f}_n \widehat{f}_n' = E_{F_n} f_i f_i' - (E_{F_n} f_i)(E_{F_n} f_i)' + o_p(1) \tag{22.79}$$

by the WLLN's (using the moment conditions in $\mathcal{F}_2$). In consequence, we have

$$\begin{aligned}\widetilde{R}_n &= \left(B' \otimes I_k\right) \left(E_{F_n} f_i f_i' - (E_{F_n} f_i)(E_{F_n} f_i')\right) \left(B \otimes I_k\right) + o_p(1) \\ &\rightarrow_p \widetilde{R}_h := \left(B' \otimes I_k\right) \left[h_5 - vec((0^k, h_4))vec((0^k, h_4))'\right] \left(B \otimes I_k\right), \end{aligned} \qquad (22.80)$$

where $B = B(\theta)$ is defined in (6.3), the convergence uses the definitions of $\lambda_{4,F}$ and $\lambda_{5,F}$ in (10.16), and the definition of $\{\lambda_{n,h} : n \geq 1\}$ in (10.18).

This yields

$$\widehat{U}_{2n} = (\widehat{\Omega}_n, \widetilde{R}_n) \rightarrow_p (h_{5,g}, \widetilde{R}_h) = h_8, \qquad (22.81)$$

which verifies Assumption WU(b) for the parameter space Λ_2 for part (b) of the lemma.

Assumption WU(c) holds for the parameter space Λ_2, with $\mathcal{W}_2$ and $\mathcal{U}_2$ defined as above, by the argument given above to verify Assumption WU(c) in part (a) of the lemma plus the inequality $\lambda_{\max}(\widetilde{R}_h) > 0$, which is established as follows. The inequality $\lambda_{\max}(\widetilde{R}_h) > 0$ is implied by $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(\widetilde{R}_F) > 0$. The latter holds by the same argument as used above to show $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(R_F) > 0$ (which is given in the paragraph containing (22.78) and the paragraph following it), but with (i) $\widetilde{R}_F$ in place of R_F and (ii) $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(\widetilde{V}_F) > 0$, rather than $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(V_F) > 0$, holding because $E_F g_i g_i'$ is the upper left $p \times p$ submatrix of $\widetilde{V}_F$, which implies that $\lambda_{\max}(\widetilde{V}_F) \geq \lambda_{\max}(E_F g_i g_i')$, and $\lambda_{\max}(E_F g_i g_i') \geq \delta$ by the last condition in $\mathcal{F}_2$.

Now we prove part (b)(ii). It suffices to show that $\mathcal{F}_2 \subset \mathcal{F}_{WU}$ for δ_1 sufficiently small and M_1 sufficiently large because $\mathcal{F}_{WU} \subset \mathcal{F}_2$ by the definition of $\mathcal{F}_{WU}$. We need to show that the four conditions in the definition of $\mathcal{F}_{WU}$ in (10.12) hold.

(I) & (II) We have $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(W_F) > 0$ and $\sup_{F \in \mathcal{F}_2} ||W_F|| < \infty$ by the proofs of (I) and (II) for part (a)(ii) of the lemma.

(III) We show that $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(U_F) > 0$, where in the present case $U_F := U_1(U_{2F}) := ((\theta_0, I_p)(\widetilde{\Sigma}^{\varepsilon}_F)^{-1}(\theta_0, I_p)')^{1/2}$ and $\widetilde{\Sigma}_F := \Sigma(\Omega_F, \widetilde{R}_F)$ has (j, ℓ) element equal to $tr(\widetilde{R}_{j\ell F}' \Omega_F^{-1})/k$ (by the paragraph containing (10.11)). We have $\sup_{F \in \mathcal{F}_2} ||\widetilde{R}_F|| = \sup_{F \in \mathcal{F}_2} ||\left(B' \otimes I_k\right) \times Var_F(f_i) \left(B \otimes I_k\right)|| < \infty$ (where the inequality uses the condition $E_F||(g_i', vec(G_i)')'||^{2+\gamma} \leq M$ in $\mathcal{F}_2$). In addition, $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(\Omega_F) > 0$ (by the last condition in $\mathcal{F}_2$). The latter results imply that $\sup_{F \in \mathcal{F}_2} ||\widetilde{\Sigma}_F|| < \infty$ (because $\widetilde{\Sigma}_F$ minimizes $||(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - \widetilde{R}_F](I_{p+1} \otimes \Omega_F^{-1/2})||$, see the paragraph containing (10.11)). This implies that $\sup_{F \in \mathcal{F}_2} ||\widetilde{\Sigma}^{\varepsilon}_F|| < \infty$. In addition, $\widetilde{\Sigma}_F$ is nonsingular $\forall F \in \mathcal{F}_2$ (because $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(\widetilde{\Sigma}_F) > 0$ by the proof of result (IV) below). The last two results imply the desired result $\inf_{F \in \mathcal{F}_2} \lambda_{\min}(U_F) = \inf_{F \in \mathcal{F}_2} \lambda_{\min}((\theta_0, I_p)(\widetilde{\Sigma}^{\varepsilon}_F)^{-1}(\theta_0, I_p)')^{1/2}) > 0$ (because $(\theta_0, I_p) \in R^{p \times (p+1)}$ has full row rank p).

(IV) We show that $\sup_{F \in \mathcal{F}_2} ||U_F|| < \infty$, where U_F is defined in (III) immediately above. The proof is the same as the proof of (IV) for part (a) of the lemma given above, but with $\widetilde{R}_F$ in place of R_F and with the verification that $\inf_{F \in \mathcal{F}_2} \lambda_{\max}(\widetilde{R}_F) > 0$ given in the the verification of Assumption WU(c)) above.

This completes the proof of part (b)(ii). $\square$

22.6 Proof of Theorem 10.1 for the Anderson-Rubin Test and CS

Theorem 10.1 is stated in Section 8 of AG2 and, for convenience, is restated at the beginning of this section, i.e., Section 22.

Proof of Theorem 10.1 for AR Test and CS. We prove the AR test results of Theorem 10.1 by applying Proposition 10.2 with

$$\lambda = \lambda_F := E_F g_i g_i',\ h_n(\lambda) := \lambda,\ \text{and } \Lambda := \{\lambda : \lambda = \lambda_F \text{ for some } F \in \mathcal{F}_{AR}\}. \tag{22.82}$$

We define the parameter space H as in (10.2). For notational simplicity, we verify Assumption B* used in Proposition 10.2 for a sequence $\{\lambda_n \in \Lambda : n \geq 1\}$ for which $h_n(\lambda_n) \to h \in H$, rather than a subsequence $\{\lambda_{w_n} \in \Lambda : n \geq 1\}$ for some subsequence $\{w_n\}$ of $\{n\}$. The same argument as given below applies with a subsequence $\{\lambda_{w_n} : n \geq 1\}$. For the sequence $\{\lambda_n \in \Lambda : n \geq 1\}$, we have

$$\lambda_{F_n} \to h := \lim E_{F_n} g_i g_i'. \tag{22.83}$$

The $k \times k$ matrix h is pd because $\lambda_{\min}(E_{F_n} g_i g_i') \geq \delta > 0$ for all $n \geq 1$ (by the last condition in $\mathcal{F}_{AR}$) and $\lim \lambda_{\min}(E_{F_n} g_i g_i') = \lambda_{\min}(h)$ (because the minimum eigenvalue of a matrix is a continuous function of the matrix).

By the multivariate central limit theorem for triangular arrays of row-wise i.i.d. random vectors with mean 0^k, variance λ_{F_n} that satisfies $\lambda_{F_n} \to h$, and uniformly bounded $2+\gamma$ moments, we have

$$n^{1/2}\widehat{g}_n \to_d h^{1/2} Z,\ \text{where } Z \sim N(0^k, I_k). \tag{22.84}$$

We have

$$\widehat{\Omega}_n = n^{-1} \sum_{i=1}^{n} (g_i g_i' - E_{F_n} g_i g_i') - \widehat{g}_n \widehat{g}_n' + E_{F_n} g_i g_i' \to_p h \text{ and } \widehat{\Omega}_n^{-1} \to_p h^{-1}, \tag{22.85}$$

where the equality holds by definition of $\widehat{\Omega}_n$ in (5.1), the first convergence result uses (22.83), (22.84), and the WLLN's for triangular arrays of row-wise i.i.d. random vectors with expectation that converges to h, and uniformly bounded $1 + \gamma/2$ moments, and the second convergence result holds by Slutsky's Theorem because h is pd.

Equations (22.84) and (22.85) give

$$AR_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \rightarrow_d Z'h^{1/2}h^{-1}h^{1/2}Z = Z'Z \sim \chi_k^2. \tag{22.86}$$

In turn, (22.86) gives

$$P_{F_n}(AR_n > \chi^2_{k,1-\alpha}) \rightarrow P(Z'Z > \chi^2_{k,1-\alpha}) = \alpha. \tag{22.87}$$

where the equality holds because $\chi^2_{k,1-\alpha}$ is the $1-\alpha$ quantile of $Z'Z$. Equation (22.87) verifies Assumption B* and the proof of the AR test results of Theorem 10.1 is complete.

The proof of the AR CS results of Theorem 10.1 is analogous to those for the tests, see the Comment to Proposition 10.2. $\square$

23 Proof of Theorem 9.1

Theorem 9.1 of AG2. *Suppose $k \geq p$. For any sequence $\{\lambda^*_{n,h} : n \geq 1\}$ that exhibits strong or semi-strong identification and for which $\lambda^*_{n,h} \in \Lambda_1^*$ $\forall n \geq 1$ for the* SR-CQLR$_1$ *test statistic and critical value and $\lambda^*_{n,h} \in \Lambda_2^*$ $\forall n \geq 1$ for the* SR-CQLR$_2$ *test statistic and critical value, we have*

(a) SR-$QLR_{jn} = QLR_{jn} + o_p(1) = LM_n + o_p(1) = LM_n^{GMM} + o_p(1)$ for $j = 1, 2$,

(b) $c_{k,p}(n^{1/2}\widehat{D}_n^*, 1-\alpha) \rightarrow_p \chi^2_{p,1-\alpha}$, and

(c) $c_{k,p}(n^{1/2}\widetilde{D}_n^*, 1-\alpha) \rightarrow_p \chi^2_{p,1-\alpha}$.

The proof of Theorem 9.1 uses the following Lemma that concerns the QLR_n statistic, which is based on general weight matrices $\widehat{W}_n$ and $\widehat{U}_n$, see (10.3), and considers sequences of distributions F in $\mathcal{F}_1$ or $\mathcal{F}_2$, rather than sequences in $\mathcal{F}_1^{SR}$ or $\mathcal{F}_2^{SR}$. Given the result of this Lemma, we obtain the results of Theorem 9.1 using an argument that is similar to that employed in Section 10.2, combined with the verification of Assumption WU for the parameter spaces Λ_1 and Λ_2 for the CQLR$_1$ and CQLR$_2$ tests, respectively, that is given in Lemma 22.4 in Section 22.

For the weight matrix $\widehat{W}_n \in R^{k \times k}$, Kleibergen's LM statistic and the standard GMM LM statistic are defined by

$$LM_n(\widehat{W}_n) := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{W}_n\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \text{ and } LM_n^{GMM}(\widehat{W}_n) := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{W}_n\widehat{G}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \tag{23.1}$$

respectively, where $\widehat{G}_n$ is the sample Jacobian defined in (5.1) with $\theta = \theta_0$. In Lemma 23.1, we show that when $n^{1/2}\tau_{pF_n} \rightarrow \infty$, the QLR_n statistic is asymptotically equivalent to the $LM_n(\widehat{W}_n)$ and $LM_n^{GMM}(\widehat{W}_n)$ statistics.

The condition $n^{1/2}\tau_{pF_n} \to \infty$ corresponds to strong or semi-strong identification in the present context. This holds because, for $F \in \mathcal{F}_{WU}$, the smallest and largest singular values of $W_F(E_F G_i)U_F$ (i.e., $\tau_{\min\{k,p\}F}$ and τ_{1F}) are related to those of $\Omega_F^{-1/2} E_F G_i$, denoted (as in the Introduction) by $s_{\min\{k,p\}F}$ and s_{1F}, via $c_1 s_{jF} \le \tau_{jF} \le c_2 s_{jF}$ for $j = \min\{k,p\}$ and $j = 1$ for some constants $0 < c_1 < c_2 < \infty$. This result uses the condition $\lambda_{\min}(\Omega_F) \ge \delta > 0$ in $\mathcal{F}_{WU}$. (See Section 8.3 in the Appendix of AG1 for the argument used to prove this result.) In consequence, when $k \ge p$, the standard weak, nonstandard weak, semi-strong, and strong identification categories defined in the Introduction are unchanged if s_{jF_n} is replaced by τ_{jF_n} in their definitions for $j = 1, p$.

Lemma 23.1 *Suppose $k \ge p$ and Assumption* WU *holds for some non-empty parameter space $\Lambda_* \subset \Lambda_2$. Under all sequences $\{\lambda_{n,h} : n \ge 1\}$ with $\lambda_{n,h} \in \Lambda_*$ for which $n^{1/2}\tau_{pF_n} \to \infty$, we have*

(a) $QLR_n = LM_n(\widehat{W}_n) + o_p(1) = LM_n^{GMM}(\widehat{W}_n) + o_p(1)$ and

(b) $c_{k,p}(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha) \to_p \chi^2_{p,1-\alpha}$.

Comment: The choice of the weight matrix $\widehat{U}_n$ that appears in the definition of the QLR_n statistic, defined in (10.3), does not affect the asymptotic distribution of QLR_n statistic under strong or semi-strong identification. This holds because QLR_n is within $o_p(1)$ of LM statistics that project onto the matrices $\widehat{W}_n\widehat{D}_n\widehat{U}_n$ and $\widehat{W}_n\widehat{G}_n\widehat{U}_n$, but such statistics do not depend on $\widehat{U}_n$ because $P_{\widehat{W}_n\widehat{D}_n\widehat{U}_n} = P_{\widehat{W}_n\widehat{D}_n}$ and $P_{\widehat{W}_n\widehat{G}_n\widehat{U}_n} = P_{\widehat{W}_n\widehat{G}_n}$ when $\widehat{U}_n$ is a nonsingular $p \times p$ matrix. In consequence, the LM statistics that appear in Lemma 23.1 (and are defined in (23.1)) do not depend on $\widehat{U}_n$.

Proof of Theorem 9.1 of AG2. By the last paragraph of Section 6.2, for $j = 1$, $SR\text{-}QLR_{jn}(\theta_0) = QLR_{jn}(\theta_0)$ wp$\to$1 under any sequence $\{F_n \in \mathcal{F}_2^{SR} : n \ge 1\}$ with $r_{F_n}(\theta_0) = k$ for n large. By the same argument as given there, the same result holds for $j = 2$. This establishes the first equality in part (a) of Theorem 9.1 because by assumption $\lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \ge 1$ (see the paragraph preceding Theorem 9.1).

Assumption WU for the parameter spaces Λ_1 and Λ_2 is verified in Lemma 22.4 in Section 22 for the CQLR$_1$ and CQLR$_2$ tests, respectively. Hence, Lemma 23.1 implies that under sequences $\{\lambda_{n,h} : n \ge 1\}$ we have $QLR_{jn} = LM_n(\widehat{\Omega}_n^{-1/2}) + o_p(1) = LM_n^{GMM}(\widehat{\Omega}_n^{-1/2}) + o_p(1)$ for $j = 1, 2$, where QLR_{1n} and QLR_{2n} are defined in (6.7) and in the paragraph containing (7.3), respectively, and $LM_n(\widehat{\Omega}_n^{-1/2})$ and $LM_n^{GMM}(\widehat{\Omega}_n^{-1/2})$ are defined in (23.1) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$. In addition, Lemma 23.1 implies that $c_{k,p}(n^{1/2}\widehat{D}_n^*, 1-\alpha) \to_p \chi^2_{p,1-\alpha}$ and $c_{k,p}(n^{1/2}\widetilde{D}_n^*, 1-\alpha) \to_p \chi^2_{p,1-\alpha}$. Note that all of these results are for sequences of distributions F in $\mathcal{F}_1$ or $\mathcal{F}_2$, not $\mathcal{F}_1^{SR}$ or $\mathcal{F}_2^{SR}$.

Next, we employ a similar argument to that in (10.30)-(10.32) of Section 10.2. Specifically, we apply the version of Lemma 23.1 described in the previous paragraph with $g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i$

and $G^*_{Fi} := \Pi^{-1/2}_{1F} A'_F G_i$ in place of g_i and G_i to the QLR_{jn} test statistics and their corresponding critical values for $j = 1, 2$. We have $n^{1/2} s^*_{pF_n} \to \infty$ iff $n^{1/2}\tau^*_{pF_n} \to \infty$, where s^*_{pF} denotes the smallest singular value of $E_F G^*_{Fi}$ and τ^*_{pF} is defined to be the smallest singular value of $(E_F g^*_{Fi} g^{*\prime}_{Fi})^{-1/2}(E_F G^*_{Fi})U_F = (\Pi^{-1/2}_{1F} A'_F \Omega_F A_F \Pi^{-1/2}_{1F})^{-1/2}(E_F G^*_i)U_F = (E_F G^*_i)U_F$. In consequence, the condition $n^{1/2}\tau_{pF_n} \to \infty$ of Lemma 23.1 holds for the transformed variables $g^*_{F_n i}$ and $G^*_{F_n i}$, i.e., $n^{1/2}\tau^*_{pF_n} \to \infty$. In the present case, $\{\Pi^{-1/2}_{1F_n} A'_{F_n} : n \geq 1\}$ are nonsingular $k \times k$ matrices by the assumption that $\lambda_{\min}(E_{F_n} g_i g'_i) > 0$ for all $n \geq 1$ (as specified in the paragraph preceding Theorem 9.1). In consequence, by Lemma 6.2 (and a footnote in Section 7, which extends the results of Lemma 6.2 to the QLR_{2n} statistic and its critical value), the QLR_{1n} and QLR_{2n} test statistics and their corresponding critical values are exactly the same when based on g^*_{Fi} and G^*_{Fi} as when based on g_i and G_i. By the definitions of $\mathcal{F}^{SR}_1$ and $\mathcal{F}^{SR}_2$, the transformed variables g^*_{Fi} and G^*_{Fi} satisfy the conditions in $\mathcal{F}_1$ and $\mathcal{F}_2$, see (10.31) and (10.32). In particular, $E_F g^*_{Fi} g^{*\prime}_{Fi} = I_k$ and $\lambda_{\min}(E_F Z^*_{Fi} Z^{*\prime}_{Fi}) \geq 1/(2c) > 0$, where $Z^*_{Fi} := \Pi^{-1/2}_{1F} A'_F Z_i$ and c is as in the definition of $\mathcal{F}^{SR}_1$ in (4.9). In addition, the LM_n and LM^{GMM}_n statistics are exactly the same when based on g^*_{Fi} and G^*_{Fi} as when based on g_i and G_i. (This holds because, for any $k \times k$ nonsingular matrix M, such as $M = \Pi^{-1/2}_{1F} A'_F$, we have $LM_n := n\widehat{g}'_n \widehat{\Omega}^{-1}_n \widehat{D}_n [\widehat{D}'_n \widehat{\Omega}^{-1}_n \widehat{D}_n]^{-1} \widehat{D}'_n \widehat{\Omega}^{-1}_n \widehat{g}_n = n\widehat{g}'_n M'(M\widehat{\Omega}_n M')^{-1} M\widehat{D}_n$ $[\widehat{D}'_n M'(M\widehat{\Omega}_n M')^{-1} M\widehat{D}_n]^{-1} \widehat{D}'_n M'(M\widehat{\Omega}_n M')^{-1}\widehat{g}_n$ and likewise for LM^{GMM}_n.) Using these results, the version of Lemma 23.1 described in the previous paragraph applied to the transformed variables g^*_{Fi} and G^*_{Fi} establishes the second and third equalities of part (a) and parts (b) and (c) of Theorem 9.1. $\square$

Proof of Lemma 23.1. We start by proving the first result of part (a) of the lemma. We have $n^{1/2}\tau_{pF_n} \to \infty$ iff $q = p$ (by the definition of q in (10.22)). Hence, by assumption, $q = p$. Given this, $Q^+_{2n}(\kappa)$ (defined in (21.11) in the proof of Theorem 10.5) is a scalar. In consequence, (21.13) and (21.16) with $j = p + 1$ give

$$\begin{aligned} 0 &= |Q^+_{2n}(\widehat{\kappa}^+_{(p+1)n})| = |M^+_{n,p+1-q} - \widehat{\kappa}^+_{(p+1)n}(1 + o_p(1))| \text{ and, hence,} \\ \widehat{\kappa}^+_{(p+1)n} &= M^+_{n,p+1-q}(1 + o_p(1)) \\ &= (n^{1/2} B^{+\prime}_{n,p+1-q} U^{+\prime}_n \widehat{D}^{+\prime}_n W'_n) h_{3,k-q} h'_{3,k-q} (n^{1/2} W_n \widehat{D}^+_n U^+_n B^+_{n,p+1-q})(1 + o_p(1)) + o_p(1) \\ &= (n^{1/2} \widehat{g}'_n \widehat{\Omega}^{-1/2}_n \widehat{W}^{-1\prime}_n W'_n) h_{3,k-q} h'_{3,k-q} (n^{1/2} W_n \widehat{W}^{-1}_n \widehat{\Omega}^{-1/2}_n \widehat{g}_n)(1 + o_p(1)) + o_p(1) \\ &= n\widehat{g}'_n \widehat{\Omega}^{-1/2}_n h_{3,k-q} h'_{3,k-q} \widehat{\Omega}^{-1/2}_n \widehat{g}_n + o_p(1), \end{aligned} \tag{23.2}$$

where $\widehat{\kappa}^+_{(p+1)n}$ is defined in (21.2), the equality on the third line holds by the definition of $M^+_{n,p+1-q}$ in (21.16), the equality on the fourth line holds by lines two and three of (21.7) because when $q = p$

the third line of (21.7) becomes $n^{1/2}W_n\widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n$, i.e., $n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q}$ drops out, as noted near the end of the proof of Theorem 10.5, and the last equality holds because $W_n\widehat{W}_n^{-1} = I_k + o_p(1)$ by Assumption WU and $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = O_p(1)$.

Next, we have

$$\begin{aligned} QLR_n &:= AR_n - \lambda_{\min}(n\widehat{Q}_{WU,n}) \\ &= AR_n - \widehat{\kappa}^+_{(p+1)n} \\ &= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}(I_k - h_{3,k-q}h_{3,k-q}')\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1) \\ &= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}h_{3,q}h_{3,q}'\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1), \end{aligned} \tag{23.3}$$

where the first equality holds by the definition of QLR_n in (10.3), the second equality holds by the definition of $\widehat{\kappa}^+_{(p+1)n}$ in (21.2), the third equality holds by (23.2) and the definition $AR_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n$ in (5.2), and the last equality holds because $h_3 = (h_{3,q}, h_{3,k-q})$ is a $k \times k$ orthogonal matrix.

When $q = p$, by Lemma 10.3, we have

$$\begin{aligned} & n^{1/2}W_n\widehat{D}_nU_nT_n \to_d \overline{\Delta}_h = h_{3,q} \text{ and so} \\ & n^{1/2}\widehat{W}_n\widehat{D}_nU_nT_n \to_p h_{3,q}, \end{aligned} \tag{23.4}$$

where the equality holds by the definition of $\overline{\Delta}_h$ in (10.24) when $q = p$ and the second convergence uses $W_n\widehat{W}_n^{-1} = I_k + o_p(1)$ by Assumption WU. In consequence,

$$\begin{aligned} P_{\widehat{W}_n\widehat{D}_n} &= P_{n^{1/2}\widehat{W}_n\widehat{D}_nU_nT_n} = P_{h_{3,q}} + o_p(1) = h_{3,q}h_{3,q}' + o_p(1) \text{ and} \\ QLR_n &= LM_n(\widehat{W}_n) + o_p(1), \end{aligned} \tag{23.5}$$

where the first equality holds because $n^{1/2}U_nT_n$ is nonsingular wp→1 by Assumption WU and post-multiplication by a nonsingular matrix does not affect the resulting projection matrix, the second equality holds by (23.4), the third equality holds because $h_{3,q}'h_{3,q} = I_q$ (since $h_3 = (h_{3,q}, h_{3,k-q})$ is an orthogonal matrix), and the second line holds by the first line, (23.3), $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = O_p(1)$, and the definition of $LM_n(\widehat{W}_n)$ in (23.1).

As in (20.5) in Section 20 with $\widehat{G}_n$ in place of $\widehat{D}_n$, we have

$$\begin{aligned} W_n\widehat{G}_nU_nB_{n,q}\Upsilon_{n,q}^{-1} &= W_nD_nU_nB_{n,q}\Upsilon_{n,q}^{-1} + W_nn^{1/2}(\widehat{G}_n - D_n)U_nB_{n,q}(n^{1/2}\Upsilon_{n,q})^{-1} \\ &= C_{n,q} + o_p(1) \to_p h_{3,q}, \end{aligned} \tag{23.6}$$

where $D_n := E_{F_n}G_i$, the second equality uses (among other things) $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$

(by the definition of q in (10.22)). The convergence in (23.6) holds by (10.19), (10.24), and (20.1). Using (23.6) in place of the first line of (23.4), the proof of $QLR_n = LM_n^{GMM}(\widehat{W}_n) + o_p(1)$ is the same as that given for $QLR_n = LM_n(\widehat{W}_n) + o_p(1)$. This completes the proof of part (a) of Lemma 23.1.

By (22.10) in the proof of Theorem 22.1, we have

$$\begin{aligned} c_{k,p}(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha) &\rightarrow_d c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) \text{ and} \\ c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) &= \chi^2_{p,1-\alpha} \text{ when } q = p, \end{aligned} \tag{23.7}$$

where the second line of (23.7) holds by the sentence following (22.9). This proves part (b) of Lemma 23.1 because convergence in distribution to a constant is equivalent to convergence in probability to the same constant. □

24 Proofs of Lemmas 14.1, 14.2, and 14.3

24.1 Proof of Lemma 14.1

In this section, we suppress the dependence of various quantities on θ_0 for notational simplicity. Thus, $g_i := g_i(\theta_0)$, $G_i := G_i(\theta_0) = (G_{i1}, ..., G_{ip}) \in R^{k \times p}$, and similarly for $\widehat{g}_n$, $\widehat{G}_n$, f_i, B, $\widehat{R}_n$, $\widehat{D}_n^*$, $\widehat{D}_n$, $\widehat{L}_n$, $\widehat{\Gamma}_{jn}$, and $\widehat{\Omega}_n$.

The proof of Lemma 14.1 uses the following lemmas. Define

$$\begin{aligned} A_0^* &:= \Sigma_V B \begin{pmatrix} b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0 \\ I_p \end{pmatrix} \in R^{(p+1)\times p},\ B := \begin{pmatrix} 1 & 0_p' \\ -\theta_0 & -I_p \end{pmatrix} \in R^{(p+1)\times(p+1)}, \\ c_0 &:= (b_0'\Sigma_V b_0)^{-1}, b_0 := (1, -\theta_0')',\ (\Sigma_{V1}, ..., \Sigma_{Vp+1}) := \Sigma_V \in R^{(p+1)\times(p+1)},\ \text{and} \\ L_{V0} &:= (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' \in R^{p\times p}. \end{aligned} \tag{24.1}$$

As defined in (3.4), $A_0 := (\theta_0, I_p)' \in R^{(p+1)\times p}$.

Lemma 24.1 $A_0^* L_{V0} = -A_0$.

Comment: Some calculations show that the columns of A_0^* and A_0 are all orthogonal to b_0. Also, A_0^* and A_0 both have full column rank p. Hence, the columns of A_0^* and A_0 span the same space in R^{p+1}. It is for this reason that there exists a $p \times p$ positive definite matrix $L = L_{V0}$ that solves $A_0^* L = -A_0$.

Lemma 24.2 *Suppose Assumption* HLIV *holds. Under* H_0, *we have* (a) $n^{1/2}\widehat{g}_n \rightarrow_d N(0^k, b_0'\Sigma_V b_0 \cdot K_Z)$, (b) $n^{-1}\sum_{i=1}^n (G_{ij}g_i' - EG_{ij}g_i') = o_p(1)\ \forall j \leq p$, (c) $\widehat{G}_n = O_p(1)$, (d) $n^{-1}\sum_{i=1}^n (g_i g_i' - E g_i g_i') = o_p(1)$, *and* (e) $\widehat{G}_n - n^{-1}\sum_{i=1}^n EG_i = O_p(n^{-1/2})$.

Proof of Lemma 14.1. To prove part (a), we determine the probability limit of $\widehat{V}_n$ defined in (6.3). By (6.3) and (3.1)-(3.3), in the linear IV regression model with reduced-form parameter π_n, we have

$$\begin{aligned}
& u_i := u_i(\theta_0) = y_{1i} - Y_{2i}'\theta_0,\ Eu_i = 0,\ u_{\theta i} = -Y_{2i} = -\pi_n' Z_i - V_{2i},\ Eu_{\theta i} = -\pi_n' Z_i, \\
& u_i^* := \begin{pmatrix} u_i \\ u_{\theta i} \end{pmatrix} = \begin{pmatrix} u_i \\ -Y_{2i} \end{pmatrix} = \Xi_n' Z_i + \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix},\ \text{where } \Xi_n = (0^k, -\pi_n) \in R^{k\times(p+1)}, \\
& Eu_i^* = \Xi_n' Z_i,\ u_i^* - Eu_i^* = \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix} = B'V_i,\ \widehat{u}_{in}^* - Eu_i^* = (\widehat{\Xi}_n - \Xi_n)' Z_i,\ \text{and} \\
& U^* := (u_1^*, ..., u_n^*)' = Z_{n\times k}\Xi_n + VB,\ \text{where } V := (V_1, ..., V_n)' \in R^{n\times(p+1)} \qquad (24.2)
\end{aligned}$$

and $B := B(\theta_0)$ is defined in (6.3).

Next, we have

$$\widehat{\Xi}_n - \Xi_n = (Z_{n\times k}' Z_{n\times k})^{-1} Z_{n\times k}' U^* - \Xi_n = (n^{-1} Z_{n\times k}' Z_{n\times k})^{-1} n^{-1} Z_{n\times k}' VB = O_p(n^{-1/2}), \qquad (24.3)$$

where the first equality holds by the definition of $\widehat{\Xi}_n$ in (6.3), the second equality uses the last line of (24.2), and the third equality holds by Assumption HLIV(c) (specifically, $n^{-1}Z_{n\times k}' Z_{n\times k} \rightarrow K_Z$ and K_Z is pd) and by $n^{-1/2} Z_{n\times k}' V = O_p(1)$ (which holds because $EZ_{n\times k}' V = 0$ and the variance of the (j, ℓ) element of $n^{-1/2} Z_{n\times k}' V$ is $n^{-1}\sum_{i=1}^n Z_{ij}^2 EV_{i\ell}^2 \rightarrow K_{Zjj} EV_{i\ell}^2 < \infty$ using Assumption HLIV(c), where K_{Zjj} denotes the (j, j) element of K_Z, for all $j \leq k, \ell \leq p+1$).

By the definition of $\widehat{V}_n$ in (6.3) and simple algebra, we have

$$\begin{aligned}
\widehat{V}_n &:= n^{-1}\sum_{i=1}^n \left[(u_i^* - \widehat{u}_{in}^*)(u_i^* - \widehat{u}_{in}^*)' \otimes Z_i Z_i'\right] \qquad (24.4) \\
&= n^{-1}\sum_{i=1}^n \left[(u_i^* - Eu_i^*)(u_i^* - Eu_i^*)' \otimes Z_i Z_i'\right] - n^{-1}\sum_{i=1}^n \left[(\widehat{u}_{in}^* - Eu_i^*)(u_i^* - Eu_i^*)' \otimes Z_i Z_i'\right] \\
&\quad - n^{-1}\sum_{i=1}^n \left[(u_i^* - Eu_i^*)(\widehat{u}_{in}^* - Eu_i^*)' \otimes Z_i Z_i'\right] + n^{-1}\sum_{i=1}^n \left[(\widehat{u}_{in}^* - Eu_i^*)(\widehat{u}_{in}^* - Eu_i^*)' \otimes Z_i Z_i'\right].
\end{aligned}$$

Using the third line of (24.2), the fourth summand on the rhs of (24.4) equals

$$n^{-1}\sum_{i=1}^{n}\left[(\widehat{\Xi}_n-\Xi_n)'Z_iZ_i'(\widehat{\Xi}_n-\Xi_n)\otimes Z_iZ_i'\right]. \tag{24.5}$$

The elements of the fourth summand on the rhs of (24.4) are each $o_p(1)$ because each is bounded by $O_p(n^{-1})n^{-1}\sum_{i=1}^n ||Z_i||^4$ using (24.3) and $n^{-1}\sum_{i=1}^n ||Z_i||^4 \le n^{-1}\sum_{i=1}^n ||Z_i||^4 1(||Z_i||>1)+1 \le n^{-1}\sum_{i=1}^n ||Z_i||^6+1=o(n)$ by Assumption HLIV(c).

Using the third line of (24.2), the second summand on the rhs of (24.4) (excluding the minus sign) equals

$$n^{-1}\sum_{i=1}^{n}\left[(\widehat{\Xi}_n-\Xi_n)'Z_iV_i'B\otimes Z_iZ_i'\right]. \tag{24.6}$$

The elements of the second summand on the rhs of (24.4) are each $o_p(1)$ because $\widehat{\Xi}_n-\Xi_n = O_p(n^{-1/2})$ by (24.3) and for any $j_1, j_2, j_3 \le k$ and $\ell \le p$ we have $n^{-1}\sum_{i=1}^n Z_{ij_1}Z_{ij_2}Z_{ij_3}V_{i\ell}=o_p(n^{1/2})$ because its mean is zero and its variance is $EV_{i\ell}^2 n^{-1}\sum_{i=1}^n Z_{ij_1}^2Z_{ij_2}^2Z_{ij_3}^2 = o(n)$ by Assumption HLIV(c). By the same argument, the elements of the third summand on the rhs of (24.4) are each $o_p(1)$.

In consequence, we have

$$\begin{aligned}
\widehat{V}_n &= n^{-1}\sum_{i=1}^{n}\left[B'V_iV_i'B\otimes Z_iZ_i'\right]+o_p(1)\\
&= n^{-1}\sum_{i=1}^{n}\left[(B'V_iV_i'B-B'\Sigma_V B)\otimes Z_iZ_i'\right]+\left[B'\Sigma_V B\otimes n^{-1}\sum_{i=1}^{n}Z_iZ_i'\right]+o_p(1)\\
&\to_p B'\Sigma_V B\otimes K_Z,
\end{aligned} \tag{24.7}$$

where the first equality holds using (24.4), the argument in the two paragraphs following (24.4), and the third line of (24.2), the second equality holds by adding and subtracting the same quantity, and the convergence holds by Assumption HLIV(c) (specifically, $n^{-1}\sum_{i=1}^n Z_iZ_i' \to K_Z$) and because the first summand on the second line is $o_p(1)$ (which holds because it has mean zero and each of its elements has variance that is bounded by $O(n^{-2}\sum_{i=1}^n ||Z_i||^4)=o(1)$, where the latter equality holds by the calculations following (24.5)).

Equation (24.7) gives

$$\widehat{R}_n := \left(B'\otimes I_k\right)\widehat{V}_n\left(B\otimes I_k\right)\to_p \Sigma_V\otimes K_Z \tag{24.8}$$

because $B'B'=BB=I_{p+1}$. Hence, part (a) holds.

To prove part (b), we have

$$\widehat{\Omega}_n := n^{-1}\sum_{i=1}^n g_i g_i' - \widehat{g}_n\widehat{g}_n' = n^{-1}\sum_{i=1}^n Eg_i g_i' + n^{-1}\sum_{i=1}^n (g_i g_i' - Eg_i g_i') + O_p(n^{-1})$$
$$= n^{-1}\sum_{i=1}^n Z_i Z_i' E u_i^2 + o_p(1) \to_p (b_0'\Sigma_V b_0)K_Z, \quad (24.9)$$

where the first equality holds by the definition in (5.1), second equality uses $n^{1/2}\widehat{g}_n = O_p(1)$ by Lemma 24.2(a), the third equality holds by Lemma 24.2(d), and the convergence holds by Assumption HLIV(c) and because $Eu_i^2 = E(V_i'b_0)^2 = b_0'\Sigma_V b_0$ by Assumption HLIV(b).

Part (c) holds because

$$\widehat{\Sigma}_{j\ell n} = tr(\widehat{R}_{j\ell n}\widehat{\Omega}_n^{-1})/k \to_p tr(\Sigma_{Vj\ell}K_Z(b_0'\Sigma_V b_0)^{-1}K_Z^{-1})/k = \Sigma_{Vj\ell}(b_0'\Sigma_V b_0)^{-1}, \quad (24.10)$$

where $\widehat{\Sigma}_{j\ell n}$ and $\Sigma_{Vj\ell}$ denote the (j,ℓ) elements of $\widehat{\Sigma}_n$ and Σ_V, respectively, $\widehat{R}_{j\ell n}$ denotes the (j,ℓ) submatrix of $\widehat{R}_n$ of dimension $k \times k$, and the convergence holds because $\widehat{R}_{j\ell n} \to_p \Sigma_{Vj\ell}K_Z$ for $j,\ell = 1,...,p+1$ and $\widehat{\Omega}_n \to_p (b_0'\Sigma_V b_0)K_Z$ by parts (a) and (b) of the lemma.

Part (d) holds because $\widehat{\Sigma}_n^{\varepsilon} \to_p ((b_0'\Sigma_V b_0)^{-1}\Sigma_V)^{\varepsilon}$ by part (c) of the lemma and Lemma 17.1(e), $((b_0'\Sigma_V b_0)^{-1}\Sigma_V)^{\varepsilon} = (b_0'\Sigma_V b_0)^{-1}\Sigma_V^{\varepsilon}$ by Lemma 17.1(d), and $\Sigma_V^{\varepsilon} = \Sigma_V$ by Assumption HLIV(e) and Comment (ii) to Lemma 17.1).

We prove part (f) next. We have

$$n^{-1}Z_{n\times k}'Y = \left(n^{-1}\sum_{i=1}^n Z_i(y_{1i} - Y_{2i}'\theta_0) + n^{-1}\sum_{i=1}^n Z_i Y_{2i}'\theta_0, n^{-1}\sum_{i=1}^n Z_i Y_{2i}\right)$$
$$= (\widehat{g}_n - \widehat{G}_n\theta_0, -\widehat{G}_n) = (\widehat{g}_n, \widehat{G}_n)\begin{pmatrix} 1 & 0_p' \\ -\theta_0 & -I_p \end{pmatrix} = (\widehat{g}_n, \widehat{G}_n)B, \quad (24.11)$$

where the expressions for $\widehat{g}_n$ and $\widehat{G}_n$ use (3.3). Using (24.11) and the definition of L_{V0} in (24.1), the statistic $\overline{T}_n$ defined in (3.4) can be written as

$$\overline{T}_n := (Z_{n\times k}'Z_{n\times k})^{-1/2}Z_{n\times k}'Y\Sigma_V^{-1}A_0(A_0'\Sigma_V^{-1}A_0)^{-1/2}$$
$$= n^{1/2}(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1/2}. \quad (24.12)$$

Note that, using the definitions of B and L_{V0} in (24.1) and A_0 in (3.4), the rhs expression for $\overline{T}_n$ equals the expression in (3.4).

Now we simplify the statistic $\widehat{D}_n := (\widehat{D}_{1n}, ..., \widehat{D}_{pn})$, where $\widehat{D}_{jn} := \widehat{G}_{jn} - \widehat{\Gamma}_{jn}\widehat{\Omega}_n^{-1}\widehat{g}_n$ for $j = 1,...,p$, by replacing $\widehat{\Gamma}_{jn}$ and $\widehat{\Omega}_n$ by their probability limits plus $o_p(1)$ terms. Let $\pi_n := (\pi_{1n}, ..., \pi_{pn}) \in$

$R^{k\times p}$. For $j=1,...,p$, we have

$$\begin{aligned}\widehat{\Gamma}_{jn} &:= n^{-1}\sum_{i=1}^{n}(G_{ij}-\widehat{G}_{jn})g_i' = n^{-1}\sum_{i=1}^{n}EG_{ij}g_i' + n^{-1}\sum_{i=1}^{n}(G_{ij}g_i' - EG_{ij}g_i') - \widehat{G}_{jn}\widehat{g}_n' \\ &= n^{-1}\sum_{i=1}^{n}EG_{ij}g_i' + o_p(1) = -n^{-1}\sum_{i=1}^{n}EZ_iY_{2ij}Z_i'u_i + o_p(1) \\ &= -n^{-1}\sum_{i=1}^{n}Z_iZ_i'EV_{2ij}V_i'b_0 + n^{-1}\sum_{i=1}^{n}Z_iZ_i'(Z_i'\pi_{jn})Eu_i + o_p(1) \\ &= -n^{-1}\sum_{i=1}^{n}Z_iZ_i'\Sigma_{Vj+1}'b_0 + o_p(1), \end{aligned} \tag{24.13}$$

where $g_i = Z_i(y_{1i} - Y_{2i}'\theta_0) = Z_iu_i$ by (3.3), the third equality holds by Lemma 24.2(a)-(c), the fourth equality holds by (3.3) with $\theta=\theta_0$, the fifth equality uses $Y_{2ij} = Z_i'\pi_{jn} + V_{2ij}$ and $u_i = V_i'b_0$, and the sixth equality holds because $EV_i = 0$ by Assumption HLIV(b), $u_i = V_i'b_0$, and $\Sigma_V := (\Sigma_{V1},...,\Sigma_{Vp+1}) := EV_iV_i'$.

Equations (24.9) and (24.13) give

$$\begin{aligned}\widehat{D}_{jn} &:= \widehat{G}_{jn} - \widehat{\Gamma}_{jn}\widehat{\Omega}_n^{-1}\widehat{g}_n = \widehat{G}_{jn} + \Sigma_{Vj+1}'b_0(b_0'\Sigma_Vb_0)^{-1}\widehat{g}_n + o_p(n^{-1/2}) \text{ and} \\ \widehat{D}_n &:= (\widehat{D}_{1n},...,\widehat{D}_{pn}) = (\widehat{g}_n,\widehat{G}_n)\begin{pmatrix}\Sigma_{V2}'b_0c_0,...,\Sigma_{Vp+1}'b_0c_0 \\ I_p\end{pmatrix} + o_p(n^{-1/2}) \\ &= (\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}\left(\Sigma_VB\begin{pmatrix}\Sigma_{V2}'b_0c_0,...,\Sigma_{Vp+1}'b_0c_0 \\ I_p\end{pmatrix}\right) + o_p(n^{-1/2}) \\ &= (\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^* + o_p(n^{-1/2}), \end{aligned} \tag{24.14}$$

where the second equality on the first line uses $\widehat{g}_n = O_p(n^{-1/2})$ by Lemma 24.2(a), the second line uses $c_0 = (b_0'\Sigma_Vb_0)^{-1}$, the second last equality holds because $B^{-1}=B$, and the last equality holds by the definition of A_0^* in (24.1).

Now, we have

$$\begin{aligned} n^{1/2}\widehat{D}_n^* &:= n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widehat{L}_n^{1/2} \\ &= (b_0'\Sigma_Vb_0)^{-1/2}(I_k+o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^* \\ &\quad \times(b_0'\Sigma_Vb_0)^{1/2}L_{V0}^{1/2}(I_p+o_p(1)) + o_p(1) \\ &= -(I_k+o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1/2}(I_p+o_p(1)) + o_p(1) \\ &= -(I_k+o_p(1))\overline{T}_n(I_p+o_p(1)) + o_p(1), \end{aligned} \tag{24.15}$$

where the first equality holds by the definition of $\widehat{D}_n^*$ in (6.7), the second equality holds by (24.14),

$\widehat{\Omega}_n \to_p (b_0'\Sigma_V b_0)K_Z$ (which holds by part (b) of the lemma), and $\widehat{L}_n := (\theta_0, I_p)(\widehat{\Sigma}_n^{\varepsilon})^{-1}(\theta_0, I_p)' \to_p (b_0'\Sigma_V b_0)L_{V0}$ (which holds because $\widehat{\Sigma}_n^{\varepsilon} \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_V$ by part (d) of the lemma), for $L_{V0} := (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)'$ defined in (24.1), the third equality holds by Lemma 24.1, and the last equality holds by (24.12). This completes the proof of part (f).

Lastly, we prove part (e). The statistic $\overline{S}_n$ satisfies

$$\begin{aligned}\overline{S}_n &:= (Z_{n\times k}'Z_{n\times k})^{-1/2}Z_{n\times k}'Yb_0(b_0'\Sigma_V b_0)^{-1/2}\\ &= n^{1/2}(n^{-1}\sum_{i=1}^{n} Z_iZ_i')^{-1/2}\widehat{g}_n(b_0'\Sigma_V b_0)^{-1/2}\\ &= n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1),\end{aligned} \tag{24.16}$$

where the first equality holds by the definition of $\overline{S}_n$ in (3.4), the second equality holds because $Y_i'b_0 = u_i$, and the third equality holds by (24.9) and $n^{1/2}\widehat{g}_n = O_p(1)$ by Lemma 24.2(a). This proves part (e). $\square$

Proof of Lemma 24.1. By pre-multiplying by $B\Sigma_V^{-1}$, the equation $A_0^*L_{V0} = -A_0$ is seen to be equivalent to

$$\begin{pmatrix} b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0 \\ I_p \end{pmatrix} L_{V0} = -B\Sigma_V^{-1}\begin{pmatrix} \theta_0' \\ I_p \end{pmatrix} = \begin{pmatrix} -1 & 0^{p\prime} \\ \theta_0 & I_p \end{pmatrix}\Sigma_V^{-1}\begin{pmatrix} \theta_0' \\ I_p \end{pmatrix}. \tag{24.17}$$

The last p rows of these $p+1$ equations are

$$L_{V0} = (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)', \tag{24.18}$$

which hold by the definition of L_{V0} in (24.1).

Substituting in the definition of L_{V0}, the first row of the equations in (24.17) is

$$(b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0)(\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' = (-1, 0^{p\prime})\Sigma_V^{-1}(\theta_0, I_p)'. \tag{24.19}$$

Equation (24.19) holds by the following argument. Write $\Sigma_V := (\Sigma_{V1}, \Sigma_{V2}^*)$ for $\Sigma_{V2}^* \in R^{(p+1)\times p}$. Then, $b_0'\Sigma_{V2}^*\theta_0 = -b_0'\Sigma_V b_0 + b_0'\Sigma_{V1}$, since $b_0 := (1, -\theta_0')'$. The left-hand side of (24.19) equals

$$\begin{aligned}&(b_0'\Sigma_{V2}^*\theta_0 c_0, b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0)\Sigma_V^{-1}(\theta_0, I_p)'\\ &= ((-b_0'\Sigma_V b_0 + b_0'\Sigma_{V1})c_0, b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0)\Sigma_V^{-1}(\theta_0, I_p)'\\ &= (-1 + b_0'\Sigma_{V1}c_0, b_0'\Sigma_{V2}c_0, ..., b_0'\Sigma_{Vp+1}c_0)\Sigma_V^{-1}(\theta_0, I_p)',\end{aligned} \tag{24.20}$$

where the second equality uses the definition of c_0 in (24.1).

Hence, the difference between the left-hand side (lhs) and the rhs of (24.19) equals

$$(b_0'\Sigma_{V1}c_0, ..., b_0'\Sigma_{Vp+1}c_0)\Sigma_V^{-1}(\theta_0, I_p)' = c_0 b_0'\Sigma_V\Sigma_V^{-1}\begin{pmatrix}\theta_0' \\ I_p\end{pmatrix} = 0_p' \tag{24.21}$$

using $b_0' := (1, -\theta_0')$. Thus, (24.19) holds, which completes the proof. $\square$

Proof of Lemma 24.2. Part (a) holds by the CLT of Eicker (1963, Thm. 3) and the Cramér-Wold device under Assumptions HLIV(a)-(c) because $n^{1/2}\widehat{g}_n = n^{-1}\sum_{i=1}^n Z_i u_i$ is an average of i.i.d. mean-zero finite-variance random variables u_i with nonrandom weights Z_i.

To show part (b), we write

$$\begin{aligned} n^{-1}\sum_{i=1}^n (G_{ij}g_i' - EG_{ij}g_i') &= -n^{-1}\sum_{i=1}^n Z_i Z_i'(Y_{2ij}u_i - EY_{2ij}u_i) \\ &= -n^{-1}\sum_{i=1}^n Z_i Z_i'(Z_i'\pi_{jn})u_i - n^{-1}\sum_{i=1}^n Z_i Z_i'(V_{2ij}u_i - \Sigma_{Vj+1}'b_0), \end{aligned} \tag{24.22}$$

where the first equality holds because $g_i = Z_i u_i$ and $G_{ij} = -Z_i Y_{2ij}$, the second equality holds because $Y_{2ij} = Z_i'\pi_{jn} + V_{2ij}$ and $EV_{2ij}u_i = EV_{2ij}V_i'b_0 = \Sigma_{Vj+1}'b_0$. Both summands on the rhs have mean zero. The (ℓ_1, ℓ_2) element of the first summand has variance equal to $n^{-2}\sum_{i=1}^n (Z_{i\ell_1}Z_{i\ell_2}Z_i'\pi_{jn})^2$ $\times Var(u_i)$, which converges to zero for all $\ell_1, \ell_2 \le k$ because $n^{-1}\sum_{i=1}^n ||Z_i||^6 = o(n)$, $Var(u_i) = b_0'\Sigma_V b_0 < \infty$, and $\sup_{j\le p, n\ge 1}||\pi_{jn}|| < \infty$ by Assumption HLIV(b)-(d). The (ℓ_1, ℓ_2) element of the second summand has variance equal to $n^{-2}\sum_{i=1}^n Z_{i\ell_1}^2 Z_{i\ell_2}^2 Var(V_{2ij}u_i)$, which converges to zero for all $\ell_1, \ell_2 \le k$ because $n^{-1}\sum_{i=1}^n ||Z_i||^6 = o(n)$ and $Var(V_{2ij}u_i) \le E(V_{2ij}V_i'b_0)^2 \le b_0'b_0 E||V_i||^4 < \infty$ by Assumptions HLIV(b)-(c). This establishes part (b).

For part (c), we have

$$\widehat{G}_n = -n^{-1}\sum_{i=1}^n Z_i Y_{2i}' = -n^{-1}\sum_{i=1}^n Z_i Z_i'\pi_n - n^{-1}\sum_{i=1}^n Z_i V_{2i}'. \tag{24.23}$$

The first term on the rhs is $O(1)$ by Assumption HLIV(c)-(d). The second term on the rhs is $O_p(n^{-1/2})$ $(= o_p(1))$ because it has mean zero and its (ℓ, j) element for $\ell \le k$ and $j \le p$ has variance $n^{-2}\sum_{i=1}^n Z_{i\ell}^2\Sigma_{Vj^*j^*}$, where $\Sigma_{Vj^*j^*} < \infty$ is the (j^*, j^*) element of Σ_V and $j^* = j+1$, and $n^{-1}\sum_{i=1}^n Z_{i\ell}^2\Sigma_{Vj^*j^*} \to K_{Z\ell\ell}\Sigma_{Vj^*j^*}$, where $K_{Z\ell\ell} < \infty$ is the (ℓ, ℓ) element of K_Z. Hence, the rhs is $O_p(1)$, which establishes part (c).

To prove part (d), we have

$$n^{-1}\sum_{i=1}^{n}(g_i g_i' - E g_i g_i') = n^{-1}\sum_{i=1}^{n} Z_i Z_i'(u_i^2 - E u_i^2) \to_p 0, \tag{24.24}$$

where the convergence holds because the rhs of the equality has mean zero and its (ℓ_1, ℓ_2) element has variance equal to n^{-1} times $n^{-1}\sum_{i=1}^{n}(Z_{i\ell_1}^2 Z_{i\ell_2}^2 Var((V_i' b_0)^2) \leq n^{-1}\sum_{i=1}^{n}||Z_i||^4 E||V_i||^4||b_0||^4 < \infty$ by Assumption HLIV(b)-(c) for all $\ell_1, \ell_2 \leq k$. This proves part (d).

Part (e) holds by the following argument:

$$\widehat{G}_n - n^{-1}\sum_{i=1}^{n} EG_i = -n^{-1}\sum_{i=1}^{n} Z_i(Y_{2i} - EY_{2i})' = -n^{-1}\sum_{i=1}^{n} Z_i V_{2i}' = O_p(n^{-1/2}), \tag{24.25}$$

where the last equality holds by the argument following (24.23). $\square$

24.2 Proof of Lemma 14.2

Proof of Lemma 14.2. To prove part (a), we determine the probability limit of $\widetilde{V}_n$ defined in (7.1), where $f_i = (Z_i' u_i, -vec(Z_i Y_{2i}')')'$ by (3.1) and (3.3). For $\zeta_n(\pi)$ defined in (14.1), we can write

$$\begin{aligned} &\zeta_n(\pi_n) = n^{-1}\sum_{i=1}^{n} Z_{ni}^* Z_{ni}^{*\prime}, \text{ where} && (24.26) \\ &Z_{ni}^* := vec\left(Z_i Z_i' \pi_n - n^{-1}\sum_{\ell=1}^{n} Z_\ell Z_\ell' \pi_n\right) = (\pi_n' \otimes Z_i)Z_i - n^{-1}\sum_{\ell=1}^{n}(\pi_n' \otimes Z_\ell)Z_\ell \in R^{kp} \end{aligned}$$

and the second equality in the second line follows from $vec(ABC) = (C' \otimes A)vec(B)$.

We have

$$\widetilde{V}_n := n^{-1}\sum_{i=1}^n\left(f_i - n^{-1}\sum_{\ell=1}^n Ef_\ell\right)\left(f_i - n^{-1}\sum_{\ell=1}^n Ef_\ell\right)' - \left(\widehat{f}_n - n^{-1}\sum_{\ell=1}^n Ef_\ell\right)\left(\widehat{f}_n - n^{-1}\sum_{\ell=1}^n Ef_\ell\right)'$$
$$= n^{-1}\sum_{i=1}^n\begin{pmatrix} Z_iu_i \\ -vec(Z_iV_{2i}') - Z_{ni}^* \end{pmatrix}\begin{pmatrix} Z_iu_i \\ -vec(Z_iV_{2i}') - Z_{ni}^* \end{pmatrix}' + o_p(1)$$
$$= n^{-1}\sum_{i=1}^n\left(\begin{pmatrix} u_i \\ -V_{2i}\end{pmatrix}\begin{pmatrix} u_i \\ -V_{2i}\end{pmatrix}' \otimes Z_iZ_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta_n(\pi_n)\end{pmatrix}$$
$$+n^{-1}\sum_{i=1}^n\begin{pmatrix} Z_iu_i \\ -vec(Z_iV_{2i}')\end{pmatrix}\begin{pmatrix} 0^k \\ -Z_{ni}^*\end{pmatrix}' + n^{-1}\sum_{i=1}^n\begin{pmatrix} 0^k \\ -Z_{ni}^*\end{pmatrix}\begin{pmatrix} Z_iu_i \\ -vec(Z_iV_{2i}')\end{pmatrix}' + o_p(1)$$
$$= \left(\begin{pmatrix} 1 & -\theta_0' \\ 0^p & -I_p\end{pmatrix}\Sigma_V\begin{pmatrix} 1 & -\theta_0' \\ 0^p & -I_p\end{pmatrix}'\right)\otimes\left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi_*)\end{pmatrix} + o_p(1)$$
$$= \left(B'\Sigma_V B\right)\otimes\left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi_*)\end{pmatrix} + o_p(1), \qquad (24.27)$$

where the second equality holds using $Eu_i = 0$, $EV_{2i} = 0^p$, $Y_{2i} = \pi_n'Z_i + V_{2i}$, $vec(Z_iY_{2i}' - n^{-1}\sum_{\ell=1}^n EZ_\ell Y_{2\ell}') = vec(Z_iV_{2i}') + Z_{ni}^*$, and Lemma 24.2(a) and (e) because $\widehat{f}_n - n^{-1}\sum_{\ell=1}^n Ef_\ell = (\widehat{g}_n', vec(\widehat{G}_n - n^{-1}\sum_{\ell=1}^n EG_\ell)')'$, the third equality holds by (24.26) and simple rearrangement, the fourth equality holds because (i) the first summand on the rhs of the fourth equality is the mean of the first summand on the lhs of the fourth equality using $u_i = (1, -\theta_0')V_i$, (ii) the variance of each element of the lhs matrix is $o(1)$ because $E||V_i||^4 < \infty$ and $n^{-1}\sum_{i=1}^n ||Z_i||^4 = o(n)$ by Assumption HLIV(b)-(c) (because $n^{-1}\sum_{i=1}^n ||Z_i||^4 \leq n^{-1}\sum_{i=1}^n ||Z_i||^4 1(||Z_i|| > 1) + 1 \leq n^{-1}\sum_{i=1}^n ||Z_i||^6 + 1 = o(n)$ using Assumption HLIV(c)), (iii) $\zeta_n(\pi_n) \to \zeta(\pi_*)$ by Assumption HLIV2(a)-(b), and (iv) the third and fourth summands on the lhs of the fourth equality have zero means and the variance of each element of these summands is $o(1)$ (because each variance is bounded by $n^{-2}\sum_{i=1}^n ||Z_{ni}^*||^2||Z_i||^2 \leq ||\pi_n||^2(n^{-2}\sum_{i=1}^n ||Z_i||^6 + 2n^{-2}\sum_{i=1}^n ||Z_i||^4 n^{-1}\sum_{\ell=1}^n ||Z_\ell||^2 + n^{-2}\sum_{i=1}^n ||Z_i||^2(n^{-1}\sum_{\ell=1}^n ||Z_\ell||^2)^2) = o(1)$, using $||Z_{ni}^*|| \leq ||\pi_n||(||Z_i||^2 + n^{-1}\sum_{\ell=1}^n ||Z_\ell||^2)$, $\sup_{\pi\in\Pi} ||\pi_n|| < \infty$, and $E||V_i||^2 < \infty$ by Assumption HLIV(b)-(d)), and the fifth equality holds by the definition of B in (6.3).

Using the definitions of $\widetilde{R}_n$ in (7.1) and $R(\pi_*)$ in (14.2), part (a) of the lemma follows from (24.27).

Next we prove part (b). We have

$$\widetilde{\Sigma}_{j\ell n} = tr(\widetilde{R}_{j\ell n}'\widehat{\Omega}_n^{-1})/k \to_p tr(R_{j\ell}(\pi_*)'(b_0'\Sigma_V b_0)^{-1}K_Z^{-1})/k =: (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*j\ell}, \qquad (24.28)$$

where $\widetilde{\Sigma}_{j\ell n}$ and $\Sigma_{V*j\ell}$ denote the (j,ℓ) elements of $\widetilde{\Sigma}_n$ and Σ_{V*}, respectively, $\widetilde{R}'_{j\ell n}$ and $R_{j\ell}(\pi_*)$ denote the (j,ℓ) submatrices of dimension $k \times k$ of $\widetilde{R}'_n$ and $R(\pi_*)$, respectively, the convergence holds by part (a) of the lemma and Lemma 14.1(b), and the last equality holds by the definition of $\Sigma_{V*j\ell}$ in (14.3). Equation (24.28) establishes part (b).

Part (c) holds because part (b) of the lemma and Lemma 17.1(e) imply that $\widetilde{\Sigma}_n^{\varepsilon} \to_p ((b_0'\Sigma_V b_0)^{-1}\Sigma_{V*})^{\varepsilon}$, Lemma 17.1(d) implies that $((b_0'\Sigma_V b_0)^{-1}\Sigma_{V*})^{\varepsilon} = (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}^{\varepsilon}$, and Assumption HLIV2(c) implies that $\Sigma_{V*}^{\varepsilon} = \Sigma_{V*}$.

To prove part (d), we have

$$\begin{aligned}
& n^{1/2}\widetilde{D}_n^* \\
:= {} & n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widetilde{L}_n^{1/2} \\
= {} & ((b_0'\Sigma_V b_0 K_Z)^{-1/2}K_Z^{1/2} + o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0^* L_{V0}^{1/2} \\
& \times(L_{V0}^{-1/2}(b_0'\Sigma_V b_0 L_{V*})^{1/2} + o_p(1)) + o_p(1) \\
= {} & -(I_k + o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0 L_{V0}^{-1/2}(L_{V0}^{-1/2}L_{V*}^{1/2} + o_p(1)) + o_p(1) \\
= {} & -(I_k + o_p(1))\overline{T}_n(L_{V0}^{-1/2}L_{V*}^{1/2} + o_p(1)) + o_p(1), \qquad (24.29)
\end{aligned}$$

where the first equality holds by the definition of $\widetilde{D}_n^*$ in (7.2), the second equality holds by (i) (24.14), (ii) the result of part (c) of the lemma that $\widetilde{\Sigma}_n^{\varepsilon} \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$, (iii) the result of Lemma 14.1(b) that $\widehat{\Omega}_n \to_p (b_0'\Sigma_V b_0)K_Z$, (iv) $n^{-1}Z_{n\times k}'Z_{n\times k} \to K_Z$ by Assumption HLIV(c), (v) $\widetilde{L}_n := (\theta_0, I_p)(\widetilde{\Sigma}_n^{\varepsilon})^{-1}(\theta_0, I_p)'$ as defined in (7.2) with $\theta = \theta_0$, and (vi) $\widetilde{L}_n \to_p b_0'\Sigma_V b_0 L_{V*}$ for L_{V*} defined in part (d) of the lemma, the third equality holds by Lemma 24.1, and the last equality holds by (24.12). This completes the proof of part (d). □

24.3 Proof of Lemma 14.3

When $p = 1$, we write

$$\Sigma_V := EV_iV_i' := (\Sigma_{V1}, \Sigma_{V2}) := \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \in R^{2\times 2} \qquad (24.30)$$

for $\Sigma_{V1}, \Sigma_{V2} \in R^2$, using the definition in (3.2).

The proof of Lemma 14.3 uses the following lemma.

Lemma 24.3 *Under the conditions of Lemma* 14.3, (a) $L_{V0} = \frac{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2}{\sigma_1^2\sigma_2^2(1-\rho^2)} > 0$, (b) $b_0'\Sigma_V b_0 = \sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2$, *and* (c) $L_{V0}(\sigma_2^2 - (b_0'\Sigma_{V2})^2(b_0'\Sigma_V b_0)^{-1}) = 1$.

Proof of Lemma 14.3. We prove part (b) first. By (24.9) and (24.14),

$$\begin{aligned} n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n &= n^{1/2}(I_k+o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^*(b_0'\Sigma_V b_0)^{-1/2}+o_p(1) \\ &= -n^{1/2}(I_k+o_p(1))(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1}(b_0'\Sigma_V b_0)^{-1/2}+o_p(1) \\ &= -(I_k+o_p(1))\overline{T}_n(L_{V0}b_0'\Sigma_V b_0)^{-1/2}+o_p(1), \qquad (24.31) \end{aligned}$$

where the second equality holds by Lemma 24.1 and the third equality holds by (24.12). Because $\overline{T}_n'(I_k+o_p(1))\overline{T}_n=\overline{T}_n'\overline{T}_n+o_p(1)||\overline{T}_n||^2$, the result of part (b) follows.

Next, we prove part (a). We have

$$\begin{aligned} & n^{-1}\sum_{i=1}^n (G_i-\widehat{G}_n)(G_i-\widehat{G}_n)' \\ &= n^{-1}\sum_{i=1}^n \left(G_i-n^{-1}\sum_{\ell=1}^n EG_\ell\right)\left(G_i-n^{-1}\sum_{\ell=1}^n EG_\ell\right)'-\left(\widehat{G}_n-n^{-1}\sum_{i=1}^n EG_i\right)\left(\widehat{G}_n-n^{-1}\sum_{i=1}^n EG_i\right)' \\ &= n^{-1}\sum_{i=1}^n\left(-Z_iZ_i'\pi_n-Z_iV_{2i}+n^{-1}\sum_{\ell=1}^n Z_\ell Z_\ell'\pi_n\right)\left(-Z_iZ_i'\pi_n-Z_iV_{2i}+n^{-1}\sum_{\ell=1}^n Z_\ell Z_\ell'\pi_n\right)'+o_p(1) \\ &= n^{-1}\sum_{i=1}^n (Z_iV_{2i})(Z_iV_{2i})'+2n^{-1}\sum_{i=1}^n (Z_iZ_i'\pi_n)(Z_iV_{2i})'-2\left(n^{-1}\sum_{\ell=1}^n Z_\ell Z_\ell'\pi_n\right)\left(n^{-1}\sum_{\ell=1}^n Z_iV_{2i}\right)' \\ &\quad +\zeta_n(\pi_n)+o_p(1) \\ &= n^{-1}Z_{n\times k}'Z_{n\times k}\sigma_2^2+\zeta_n(\pi_n)+o_p(1), \qquad (24.32) \end{aligned}$$

where the first equality holds by algebra, the second equality holds by Lemma 24.2(e), $G_i=-Z_iY_{2i}$, $Y_{2i}=Z_i'\pi_n+V_{2i}$, and so $Y_{2i}-EY_{2i}=V_{2i}$, the third equality holds by multiplying out the terms on the lhs of the third equality and using the definition of $\zeta_n(\pi)$ in (14.10), the first summand on the lhs of the fourth equality equals the first summand on the rhs of the fourth equality plus $o_p(1)$ by the same argument as for Lemma 24.2(d) with V_{2i}^2 in place of u_i^2 and $\sigma_2^2 := EV_{2i}^2$ in place of Eu_i^2, the second summand on the lhs of the fourth equality is $o_p(1)$ because it has mean zero and its elements have variances that are bounded by $4\sigma_2^2n^{-2}\sum_{i=1}^n||Z_i||^6\sup_{\pi\in\Pi}||\pi||^2$, which is $o(1)$ by Assumption HLIV(c)-(d), and the third summand on the lhs of the fourth equality is $o_p(1)$ because $n^{-1}\sum_{\ell=1}^n Z_\ell Z_\ell'\pi_n=O(1)$ by Assumption HLIV(c) and (d) and $n^{-1}\sum_{\ell=1}^n Z_iV_{2i}=o_p(1)$ by the argument following (24.23).

Combining (24.13), (24.9), (24.32) and the definition of $\widetilde{V}_{Dn}$ in (14.9), we obtain

$$\begin{aligned}\widetilde{V}_{Dn} &= n^{-1}\sum_{i=1}^{n} Z_i Z_i'(\sigma_2^2 - (b_0'\Sigma_{V2})^2(b_0'\Sigma_V b_0)^{-1}) + \zeta_n(\pi_n) + o_p(1)\\ &= K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1), \end{aligned} \tag{24.33}$$

where the second equality holds by Lemma 24.3(c) and Assumption HLIV(c).

Next, we have

$$\begin{aligned} n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}\widehat{D}_n L_{V0}^{1/2} &= n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^* L_{V0}^{1/2} + o_p(1)\\ = -n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0 L_{V0}^{-1/2} &+ o_p(1) = -\overline{T}_n + o_p(1), \end{aligned} \tag{24.34}$$

where the first equality holds by (24.14), the second equality holds by Lemma 24.1, and the third equality holds by (24.12).

Using (24.33), we obtain

$$\begin{aligned} n^{1/2}\widetilde{V}_{Dn}^{-1/2}\widehat{D}_n &= [K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} n^{1/2}\widehat{D}_n\\ &= -[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{1/2}\overline{T}_n L_{V0}^{-1/2} + o_p(1)\\ &= -[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} K_Z^{1/2}\overline{T}_n L_{V0}^{-1/2}(1 + o_p(1)) + o_p(1), \end{aligned} \tag{24.35}$$

where the second equality holds using (24.34) and Assumption HLIV(c), the third equality holds by Assumption HLIV(c) and some calculations. Using this, we obtain

$$\begin{aligned} rk_{1n} := n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n &= \overline{T}_n' K_Z^{1/2}[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1} K_Z^{1/2}\overline{T}_n L_{V0}^{-1}(1 + o_p(1)) + o_p(1)\\ &= \overline{T}_n'[I_k + L_{V0}K_Z^{-1/2}\zeta_n(\pi_n)K_Z^{-1/2} + o_p(1)]^{-1}\overline{T}_n(1 + o_p(1)) + o_p(1), \end{aligned} \tag{24.36}$$

where the last equality holds by some algebra. This proves part (a) of the lemma.

Part (c) of the lemma follows from Lemma 24.3(a) and (b) by substituting in $\sigma_2^2 = c^2\sigma_1^2$. □

Proof of Lemma 24.3. Part (a) holds by the following calculations:

$$\begin{aligned} L_{V0} &:= (\theta_0, 1)\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1}\begin{pmatrix} \theta_0 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}(\theta_0, 1)\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}\begin{pmatrix} \theta_0 \\ 1 \end{pmatrix} = \frac{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}. \end{aligned} \tag{24.37}$$

We have $L_{V0} > 0$ because Σ_V is pd by Assumption HLIV(b) and $(\theta_0, 1) \neq 0_2$.

Part (b) holds by the first of the following two calculations:

$$b_0'\Sigma_V b_0 := (1, -\theta_0) \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} 1 \\ -\theta_0 \end{pmatrix} = \sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2 \text{ and}$$
$$b_0'\Sigma_{V2} := (1, -\theta_0)(\rho\sigma_1\sigma_2, \sigma_2^2)' = \rho\sigma_1\sigma_2 - \theta_0\sigma_2^2. \tag{24.38}$$

Using (24.38), we obtain

$$\sigma_2^2 - (b_0'\Sigma_{V2})^2(b_0'\Sigma_V b_0)^{-1} = \sigma_2^2 - \frac{(\rho\sigma_1\sigma_2 - \theta_0\sigma_2^2)^2}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} \tag{24.39}$$
$$= \frac{\sigma_1^2\sigma_2^2 - 2\theta_0\rho\sigma_1\sigma_2^3 + \theta_0^2\sigma_2^4 - (\rho\sigma_1\sigma_2 - \theta_0\sigma_2^2)^2}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} = L_{V0}^{-1},$$

which proves part (c). □

25 Proof of Theorem 12.1

In Section 8, we establish Theorem 8.1 by first establishing Theorem 10.1, which concerns non-SR versions of the AR, CQLR_1, and CQLR_2 tests and employs the parameter spaces $\mathcal{F}_{AR}$, $\mathcal{F}_1$, and $\mathcal{F}_2$, rather than $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_1^{SR}$, and $\mathcal{F}_2^{SR}$. We prove Theorem 12.1 here using the same two-step approach.

In the time series context, the non-SR version of the AR statistic is defined as in (5.2) based on $\{f_i - \widehat{f}_n : i \leq n\}$, but with $\widehat{\Omega}_n$ defined in (12.3) and Assumption Ω below, rather than in (5.1), and the critical value is $\chi^2_{k,1-\alpha}$. The non-SR QLR_1 time series test statistic and conditional critical value are defined as in Section 6.1, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined in (12.3) and Assumption V_1 below based on $\{(u_i^* - \widehat{u}_{in}^*) \otimes Z_i : i \leq n\}$, rather than in (6.3) and (5.1), respectively. The non-SR QLR_2 time series test statistic and conditional critical value are defined as in Section 7, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined in (12.3) and Assumption V below based on $\{f_i - \widehat{f}_n : i \leq n\}$, in place of $\widetilde{V}_n$ and $\widehat{\Omega}_n$ defined in (7.1) and (5.1), respectively.

For the non-SR AR and non-SR CQLR tests in the time series context, we use the following parameter spaces. We define

$$\mathcal{F}_{TS,AR} := \{F : \{W_i : i = ..., 0, 1, ...\} \text{ are stationary and strong mixing under } F \text{ with}$$
$$\text{strong mixing numbers } \{\alpha_F(m) : m \geq 1\} \text{ that satisfy } \alpha_F(m) \leq Cm^{-d},$$
$$E_F g_i = 0^k,\ E_F||g_i||^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(\Omega_F) \geq \delta\} \tag{25.1}$$

for some $\gamma, \delta > 0$, $d > (2+\gamma)/\gamma$, and $C, M < \infty$, where Ω_F is defined in (12.4). We define $\mathcal{F}_{TS,2}$ and $\mathcal{F}_{TS,1}$ as $\mathcal{F}_2$ and $\mathcal{F}_1$ are defined in (10.1), respectively, but with $\mathcal{F}_{TS,AR}$ in place of $\mathcal{F}_{AR}$. For CS's, we use the corresponding parameter spaces $\mathcal{F}_{TS,\Theta,AR} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,AR}(\theta_0), \theta_0 \in \Theta\}$, $\mathcal{F}_{TS,\Theta,2} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,2}(\theta_0), \theta_0 \in \Theta\}$, and $\mathcal{F}_{TS,\Theta,1} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,1}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{TS,AR}(\theta_0)$, $\mathcal{F}_{TS,2}(\theta_0)$, and $\mathcal{F}_{TS,1}(\theta_0)$ denote $\mathcal{F}_{TS,AR}$, $\mathcal{F}_{TS,2}$, and $\mathcal{F}_{TS,1}$, respectively, with their dependence on θ_0 made explicit.

For the (non-SR) CQLR$_2$ test and CS in the time series context, we use the following assumptions.

Assumption V: $\widehat{V}_n(\theta_0) - V_{F_n}(\theta_0) \to_p 0^{(p+1)k \times (p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{TS,2} : n \geq 1\}$ for which $V_{F_n}(\theta_0) \to V$ for some matrix V whose upper left $k \times k$ submatrix Ω is pd.

Assumption V-CS: $\widehat{V}_n(\theta_{0n}) - V_{F_n}(\theta_{0n}) \to_p 0^{(p+1)k \times (p+1)k}$ under $\{(F_n, \theta_{0n}) : n \geq 1\}$ for any sequence $\{(F_n, \theta_{0n}) \in \mathcal{F}_{TS,\Theta,2} : n \geq 1\}$ for which $V_{F_n}(\theta_{0n}) \to V$ for some matrix V whose upper left $k \times k$ submatrix Ω is pd.

For the (non-SR) CQLR$_1$ test and CS, we use **Assumptions $\mathbf{V_1}$** and **$\mathbf{V_1}$-CS**, which are defined to be the same as Assumptions V and V-CS, respectively, but with $\mathcal{F}_{TS,1}$ and $\mathcal{F}_{TS,\Theta,1}$ in place of $\mathcal{F}_{TS,2}$ and $\mathcal{F}_{TS,\Theta,2}$.

For the (non-SR) AR test and CS, we use Assumptions Ω and Ω-CS, which are defined as follows. **Assumption $\mathbf{\Omega}$**: $\widehat{\Omega}_n(\theta_0) - \Omega_{F_n,n}(\theta_0) \to_p 0^{k \times k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{TS,AR} : n \geq 1\}$ for which $\Omega_{F_n,n}(\theta_0) \to \Omega$ for some pd matrix Ω and $r_{F_n,n}(\theta_0) = r$ for all n large, for any $r \in \{1, ..., k\}$. **Assumption $\mathbf{\Omega}$-CS** is the same as Assumption Ω, but with θ_{0n} and $\mathcal{F}_{TS,\Theta,AR}$ in place of θ_0 and $\mathcal{F}_{TS,AR}$.

For the time series case, the asymptotic size and similarity results for the non-SR tests and CS's are as follows.

Theorem 25.1 *Suppose the* AR, CQLR$_1$, *and* CQLR$_2$ *tests are defined as above, the parameter spaces for F are $\mathcal{F}_{TS,AR}$, $\mathcal{F}_{TS,1}$, and $\mathcal{F}_{TS,2}$, respectively* (*defined in the paragraph containing* (25.1))*, and the corresponding Assumption* Ω, V$_1$, *or* V *holds for each test. 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* AR, CQLR$_1$, *and* CQLR$_2$ CS's *for the parameter spaces $\mathcal{F}_{TS,\Theta,AR}$, $\mathcal{F}_{TS,\Theta,1}$, and $\mathcal{F}_{TS,\Theta,2}$, respectively, provided the corresponding Assumption* Ω-CS, V$_1$-CS, *or* V-CS *holds for each* CS*, rather than Assumption* Ω, V$_1$, *or* V.

The proof of Theorem 12.1 uses Theorem 25.1 and the following lemma.

Lemma 25.2 *Suppose $\{X_i : i = ..., 0, 1, ...\}$ is a strictly stationary sequence of mean zero, square integrable, strong mixing random variables. Then, $Var(\overline{X}_n) = 0$ for any $n \geq 1$ implies that $X_i = 0$ a.s., where $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$.*

Proof of Theorem 12.1. The proof of Theorem 12.1 using Theorem 25.1 is essentially the same as the proof (given in Section 10.2) of Theorem 8.1 using Theorem 10.1 and Lemma 10.6. Thus, we need an analogue of Lemma 10.6 to hold in the time series case. The proof of Lemma 10.6 (given in Section 10.2) goes through in the time series case, except for the following:

(i) in the proof of $\widehat{r}_n \leq r\ (= r_{F_n})$ a.s. $\forall n \geq 1$ we replace the statement "for any constant vector $\lambda \in R^k$ for which $\lambda'\Omega_{F_n}\lambda = 0$, we have $\lambda' g_i = 0$ a.s.$[F_n]$ and $\lambda'\widehat{\Omega}_n\lambda = n^{-1}\sum_{i=1}^n(\lambda' g_i)^2 - (\lambda'\widehat{g}_n)^2 = 0$ a.s.$[F_n]$" by the statement "for any constant vector $\lambda \in R^k$ for which $\lambda'\Omega_{F_n}\lambda = 0$, we have $\lambda' g_i = 0$ a.s.$[F_n]$ by Lemma 25.2 (with $X_i = \lambda' g_i$) and in consequence $\lambda'\widehat{\Omega}_n\lambda = 0$ a.s.$[F_n]$ by Assumption SR-V$_2$(c), SR-V$_2$-CS(c), SR-V$_1$(c), SR-V$_1$-CS(c), SR-Ω(c), or SR-Ω-CS(c)."

(ii) in the proof of $\widehat{r}_n \geq r$ a.s. $\forall n \geq 1$ we have $\Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2} \to_p I_r$, with Π_{1F_n} and A_{F_n} replaced by $\Pi_{1F_n,n}$ and $A_{F_n,n}$, respectively, by Assumption SR-V$_2$(a) or SR-V$_2$-CS(a), rather than by the definition of $\widehat{\Omega}_n$ combined with a WLLN for i.i.d. random variables,

(iii) in (10.27), the second implication holds by Lemma 25.2 (with $X_i = \lambda' g_i$) and the fourth implication holds by Assumption SR-V$_2$(c), SR-V$_2$-CS(c), SR-V$_1$(c), SR-V$_1$-CS(c), SR-Ω(c), or SR-Ω-CS(c), and

(iv) the result of Lemma 6.2, which is used in the proof of Lemma 10.6, holds using the equivariance condition in Assumption SR-V$_2$(b), SR-V$_2$-CS(b), SR-V$_1$(b), SR-V$_1$-CS(b), SR-Ω(b), or SR-Ω-CS(b). $\square$

Proof of Theorem 25.1. The proof is essentially the same as the proof of Theorem 10.1 (given in Section 22) and the proofs of Lemma 10.3 and Proposition 10.4 (given in Section 20 above and Section 16 in the SM of AG1, respectively) for the i.i.d. case, but with some modifications. The modifications are the first, second, third, and fifth modifications stated in the proof of Theorem 7.1 in AG1, which is given in Section 19 in the SM to AG1. Briefly, these modifications involve: (i) the definition of $\lambda_{5,F}$, (ii) justifying the convergence in probability of $\widehat{\Omega}_n$ and the positive definiteness of its limit by Assumption V, V-CS, V$_1$, V$_1$-CS, Ω, or Ω-CS, rather than by the WLLN for i.i.d. random variables, (iii) justifying the convergence in probability of $\widehat{\Gamma}_{jn}\ (= \widehat{\Gamma}_{jn}(\theta_0))$ by Assumption V, V-CS, V$_1$, or V$_1$-CS, rather than by the WLLN for i.i.d. random variables, and (iv) using the WLLN and CLT for triangular arrays of strong mixing random vectors given in Lemma 16.1 in the SM of AG1, rather than the WLLN and CLT for i.i.d. random vectors. For more details on the modifications, see Section 19 in the SM to AG1. These modifications affect the proof of Lemma

10.3. No modifications are needed elsewhere. □

Proof of Lemma 25.2. Suppose $Var(\overline{X}_n) = 0$. Then, $\overline{X}_n$ equals a constant a.s. Because $E\overline{X}_n = 0$, the constant equals zero. Thus, $\sum_{i=1}^{n} X_i = 0$ a.s. By strict stationarity, $\sum_{i=1}^{n} X_{i+sn} = 0$ a.s. and $\sum_{i=2}^{n+1} X_{i+sn} = 0$ a.s. for all integers $s \geq 0$. Taking differences yields $X_{1+sn} = X_{1+n+sn}$ for all $s \geq 0$. That is, $X_1 = X_{1+sn}$ for all $s \geq 1$.

Let A be any Borel set in R. By the strong mixing property, we have

$$\xi_s := |P(X_1 \in A, X_{1+sn} \in A) - P(X_1 \in A)P(X_{1+sn} \in A)| \leq \alpha_X(sn) \to 0 \text{ as } s \to \infty, \tag{25.2}$$

where $\alpha_X(m)$ denotes the strong mixing number of $\{X_i : i = ..., 0, 1, ...\}$ for time period separations of size $m \geq 1$. We have

$$\xi_s = |P(X_1 \in A) - P(X_1 \in A)^2| = P(X_1 \in A)(1 - P(X_1 \in A)), \tag{25.3}$$

where the first equality holds because $X_1 = X_{1+sn}$ a.s. and by strict stationarity. Because $\xi_s \to 0$ as $s \to \infty$ by (25.2) and ξ_s does not depend on s by (25.3), we have $\xi_s = 0$. That is, $P(X_1 \in A)$ equals zero or one (using (25.3)) for all Borel sets A and, hence, X_i equals a constant a.s. Because $EX_i = 0$, the constant equals zero. □

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica,* 59, 817–858.

Andrews, D. W. K., and P. Guggenberger (2014a): "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," Cowles Foundation Discussion Paper No. 1977, Yale University.

——— (2014b): "Supplemental Material to 'Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models'," Cowles Foundation Discussion Paper No. 1977, Yale University.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica, 74,* 715–752.

——— (2008): "Efficient Two-sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments," *Journal of Econometrics,* 146, 241–254.

Andrews, I. (2014): "Conditional Linear Combination Tests for Weakly Identified Models," unpublished manuscript, Department of Economics, MIT.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Admissible Invariant Similar Tests for Instrumental Variables Regression," *Econometric Theory,* 25, 806–818.

Cragg, J. C., and S. G. Donald (1996): "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association,* 91, 1301–1309.

——— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics,* 76, 223–250.

Eicker, F. (1963): "Central Limit Theorems for Families of Sequences of Random Variables," *Annals of Mathematical Statistics,* 34, 439–446.

Guggenberger, P., J. J. S. Ramalho, and R. J. Smith (2012): "GEL Statistics Under Weak Identification," *Journal of Econometrics,* 170, 331–349.

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.

——— (2007): "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics,* 139, 181–216.

Kleibergen, F., and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics,* 133, 97–126.

Moreira, H., and M. J. Moreira (2013): "Contributions to the Theory of Similar Tests," unpublished manuscript, FGV/EPGE, Rio de Janeiro, Brazil.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica,* 71, 1027–1048.

Newey, W. K., and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55, 703–708.

Newey, W. K., and F. Windmeijer (2009): "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica,* 77, 687–719.

Pollard, D. (1990): *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

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.

Stock, J. H., and J. H. Wright (2000): "GMM with Weak Identification," *Econometrica,* 68, 1055–1096.

Van Loan, C. F., and N. Pitsianis (1993): "Approximation with Kronecker Products," in *Linear Algebra for Large Scale and Real-Time Applications,* ed. by M. S. Moonen, G. H. Golub, and B. L. R. de Moor, 293–314. Amsterdam: Kluwer Academic Publishers.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics,* 86, 797–810.