Supplement to "Identification- and singularity-robust inference for moment condition models"

(*Quantitative Economics*, Vol. 10, No. 4, November 2019, 1703–1746)

Donald W. K. Andrews
Cowles Foundation, Yale University

Patrik Guggenberger
Department of Economics, Pennsylvania State University

Contents

Donald W. K. Andrews: donald.andrews@yale.edu
Patrik Guggenberger: pxg27@psu.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE1219

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 (2017). The SM to AG1 is given in Andrews and Guggenberger (2014).

Section 12 provides further discussion of the literature related to AG2.

Section 13 extends the subvector tests in Section 9 to allow for the possibility that $\Omega_F = E_F g_i g_i'$ is singular.

Section 14 provides some miscellaneous backup material for AG2.

Section 15 introduces the SR-CQLR$_P$ test that applies when the moment functions are of a multiplicative form, $u_i(\theta)Z_i$, where $u_i(\theta)$ is a scalar residual and Z_i is a k-vector of instrumental variables.

Sections 16 and 17 provide parts of the proofs of the asymptotic size results given in Sections 6 and 15.

Section 18 generalizes the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests from i.i.d. observations to strictly stationary strong mixing observations.

Section 19 compares the test statistics and conditioning statistics of the SR-CQLR, SR-CQLR$_P$, 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) IVs.

Section 20 provides finite-sample null rejection probability simulation results for the SR-AR and SR-CQLR tests for cases where the variance matrix of the moment functions is singular and near singular.

Section 21 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 19 suggest.

Section 22 establishes some properties of the eigenvalue-adjustment procedure defined in Section 5.1 and used in the definitions of the SR-CQLR and SR-CQLR$_P$ tests.

Section 23 defines a new SR-LM test.

The remainder of the SM provides the rest of the proofs of the results stated in AG2 and the SM. Section 24 proves Lemmas 16.2, 5.1, and 15.1. Section 25 proves Lemma 16.4 and Proposition 16.5. Section 26 proves Theorem 16.6. Section 27 proves Theorem 16.1 (using Theorem 16.6). Section 28 proves Theorems 7.1 and 15.3. Section 29 proves Lemmas 19.1, 19.2, and 19.3. Section 30 proves Theorem 18.1, which concerns the time series results. Section 31 proves Theorems 9.1, 13.1, and 9.2, which concern the subvector inference results.

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. Further discussion of the related literature

The first paragraph of AG2 lists a number of models in which weak identification may arise. Specific references are as follows. For new Keynesian Phillips curve models, see Dufour, Khalaf, and Kichian (2006), Nason and Smith (2008), and Kleibergen and Mavroeidis (2009). For DSGE models, see Canova and Sala (2009), Iskrev (2010), Qu and Tkachenko (2012), Dufour, Khalaf, and Kichian (2013), Guerron-Quintana, Inoue, and Kilian (2013), Qu (2014), Schorfheide (2014), and I. Andrews and Mikusheva (2015, 2016). For the CCAPM, see Stock and Wright (2000), Neely, Roy, and Whiteman (2001), Yogo (2004), Kleibergen (2005), Carroll, Slacalek, and Sommer (2011), and Gomes and Paz (2013). For interest rate dynamics, see Jegannathan, Skoulakis, and Wang (2002) and Grant (2013). For the BLP model, see Armstrong (2016). For the returns-to-schooling wage equations, see Angrist and Krueger (1991, 1992) and Cruz and Moreira (2005).

For the time series models, see Hannan (1982), Teräsvirta (1994), Nelson and Startz (2007), and Andrews and Cheng (2012, 2013b). For the selection model, see Puhani (2000). For the mixing and regime switching models, see Cho and White (2007), Chen, Ponomareva, and Tamer (2014), and references therein. For the nuisance parameter only under the alternative models, see Davies (1977) and Andrews and Ploberger (1994).

Some asymptotic size results in the linear IV regression model with a single right-hand side endogenous variable (i.e., $p = 1$) include the following. Mikusheva (2010) established the correct asymptotic size of LM and CLR tests in the linear IV model when the errors are homoskedastic. Guggenberger (2012) established the correct asymptotic size of heteroskedasticity-robust LM and CLR tests in a heteroskedastic linear IV model.

Subvector inference via the Bonferroni or Scheffé projection method, is discussed in Cavanagh, Elliott, and Stock (1995), Chaudhuri, Richardson, Robins, and Zivot (2010), Chaudhuri and Zivot (2011), and McCloskey (2017) for Bonferroni's method, and Dufour (1989) and Dufour and Jasiak (2001) for the projection method. Both methods are conservative, but Bonferroni's method is found to work quite well by Chaudhuri et al. (2010) and Chaudhuri and Zivot (2011).[28] Andrews (2017) provided subvector methods that are closely related to the Bonferroni method but are not conservative asymptotically.

Other results in the literature on subvector inference include the following. Subvector inference in which nuisance parameters are profiled out is possible in the linear IV regression model with homoskedastic errors using the AR test, but not the LM or CLR tests; see Guggenberger, Kleibergen, Mavroeidis, and Chen (2012). Andrews and Cheng (2012, 2013a, 2013b) provided subvector tests with correct asymptotic size based on extremum estimator objective functions. These subvector methods depend on the following: (i) one has knowledge of the source of the potential lack of identification (i.e., which subvectors play the roles of β, π, and ζ in their notation), (ii) there is only one source of lack of identification, and (iii) the estimator objective function does not depend on the weakly identified parameters π (in their notation) when $\beta = 0$, which rules out some weak IVs models.[29] Cheng (2015) provided subvector inference in a nonlinear regression model with multiple nonlinear regressors, and hence, multiple potential sources of lack of identification. I. Andrews and Mikusheva (2016) developed subvector inference methods in a minimum distance context based on Anderson–Rubin-type statistics. Cox (2017) provided subvector methods in a class of models that allows for multiple sources of weak identification and includes factor models. I. Andrews and Mikusheva (2015) provided conditions under which subvector inference is possible in exponential family models (but the requisite conditions seem to be quite restrictive). I. Andrews (2018) considered subvector inference in the context of a two-step procedure that determines first whether one should use an identification-robust method or not.

Phillips (1989) and Choi and Phillips (1992) provided asymptotic and finite-sample results for estimators and classical tests in simultaneous equations models that may be unidentified or partially identified when $p \geq 1$. However, their results do not cover weak identification (of standard or nonstandard form) or identification-robust inference. Hillier (2009) provided exact finite-sample results for CLR tests in the linear model under the assumption of homoskedastic normal errors and known covariance matrix. Antoine and Renault (2009, 2010) considered GMM estimation under semi-strong and

[28] Cavanagh, Elliott, and Stock (1995) provided a refinement of Bonferroni's method that is not conservative, but it is much more intensive computationally. McCloskey (2017) also considered a refinement of Bonferroni's method.

[29] Montiel Olea (forthcoming) also provided some subvector analysis in the extremum estimator context of Andrews and Cheng (2012). His efficient conditionally similar tests apply to the subvector (π, ζ) of (β, π, ζ) (in Andrews and Cheng's (2012) notation), where β is a parameter that determines the strength of identification and is known to be strongly identified. The scope of this subvector analysis is analogous to that of Stock and Wright (2000) and Kleibergen (2004).

strong identification, but do not consider tests or CSs that are robust to weak identification. Armstrong, Hong, and Nekipelov (2012) showed that standard Wald tests for multiple restrictions in some nonlinear IV models can exhibit size distortions when some IVs are strongly identified and others are semi-strongly identified—not weakly identified. These results indicate that identification issues can be more severe in nonlinear models than in linear models, which provides further motivation for the development of identification-robust tests for nonlinear models.

13. Subvector SR tests for potentially singular moments variance matrices

Figure SM-1 provides additional power comparisons to those given in Section 9.4 for the subvector null hypothesis in the endogenous probit model. Figure SM-1 provides results for $\rho = 0$, whereas Figure 1 in Section 9.4 provides results for $\rho = 0.9$. See Section 9.4 for a discussion of the results.

In the remainder of this section, we extend the subvector tests in Section 9 to allow for the possibility that $\Omega_F = E_F g_i g_i'$ is singular. We employ the definitions in (10) and (11) with η in place of θ. That is, $\widehat{r}_n(\theta, \beta) := \mathrm{rk}(\widehat{\Omega}_n(\theta, \beta))$ and $\widehat{\Omega}_n(\theta, \beta) := \widehat{A}_n^{\Omega}(\theta, \beta)\widehat{\Pi}_n(\theta, \beta)\widehat{A}_n^{\Omega}(\theta, \beta)'$, where $\widehat{\Pi}_n(\theta, \beta)$ is the $k \times k$ diagonal matrix with the eigenvalues of $\widehat{\Omega}_n(\theta, \beta)$ on the diagonal in nonincreasing order, and $\widehat{A}_n^{\Omega}(\theta, \beta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\widehat{\Pi}_n(\theta, \beta)$. We partition $\widehat{A}_n^{\Omega}(\theta, \beta)$ according to whether the corresponding eigenvalues are positive or zero: $\widehat{A}_n^{\Omega}(\theta, \beta) = [\widehat{A}_n(\theta, \beta), \widehat{A}_n^{\perp}(\theta, \beta)]$, where $\widehat{A}_n(\theta, \beta) \in R^{k \times \widehat{r}_n(\theta, \beta)}$ and $\widehat{A}_n^{\perp}(\theta, \beta) \in R^{k \times (k - \widehat{r}_n(\theta, \beta))}$. The columns of $\widehat{A}_n(\theta, \beta)$ are eigenvectors of $\widehat{\Omega}_n(\theta, \beta)$ that correspond to positive eigenvalues of $\widehat{\Omega}_n(\theta, \beta)$.

Analogously, consider the spectral decomposition for the population quantity, defined in (4) with η in place of θ, that is, $\Omega_F(\theta, \beta) = A_F^{\Omega}(\theta, \beta)\Pi_F(\theta, \beta)A_F^{\Omega}(\theta, \beta)'$, and define $r_F(\theta, \beta) := \mathrm{rk}(\Omega_F(\theta, \beta))$. We partition $A_F^{\Omega}(\theta, \beta)$ as

$$\begin{aligned} &A_F^{\Omega}(\theta, \beta) = \left[A_F(\theta, \beta), A_F^{\perp}(\theta, \beta)\right], \quad \text{where} \\ &A_F(\theta, \beta) \in R^{k \times r_F(\theta, \beta)}, \qquad A_F^{\perp}(\theta, \beta) \in R^{k \times (k - r_F(\theta, \beta))}, \end{aligned} \tag{13.1}$$

and the columns of $A_F(\theta, \beta)$ are eigenvectors of $\Omega_F(\theta, \beta)$ that correspond to positive eigenvalues of $\Omega_F(\theta, \beta)$. Let $\Pi_{1F}(\theta, \beta)$ denote the upper left $r_F(\theta, \beta) \times r_F(\theta, \beta)$ submatrix of $\Pi_F(\theta, \beta)$. The matrix $\Pi_{1F}(\theta, \beta)$ is diagonal with the positive eigenvalues of $\Omega_F(\theta, \beta)$ on its diagonal in nonincreasing order. As above, we sometimes leave out the argument θ and denote by $\widehat{\Omega}_n(\beta)$ the matrix $\widehat{\Omega}_n(\theta_0, \beta)$ and similarly for other expressions.

Recall the definition following (42) of $\widetilde{\beta}_n$, the null-restricted first-stage GMM estimator. Analogously to the full vector SR test, the subvector SR test is defined using the nonredundant moment functions. That is, rather than using the moment function $g_i(\theta, \beta)$, the test of the hypothesis in (38) is based on

$$g_{\widehat{A}i}(\theta, \beta) = \widehat{A}_n(\theta_0, \widetilde{\beta}_n)' g_i(\theta, \beta) \in R^{\widehat{r}_n(\theta_0, \widetilde{\beta}_n)}. \tag{13.2}$$

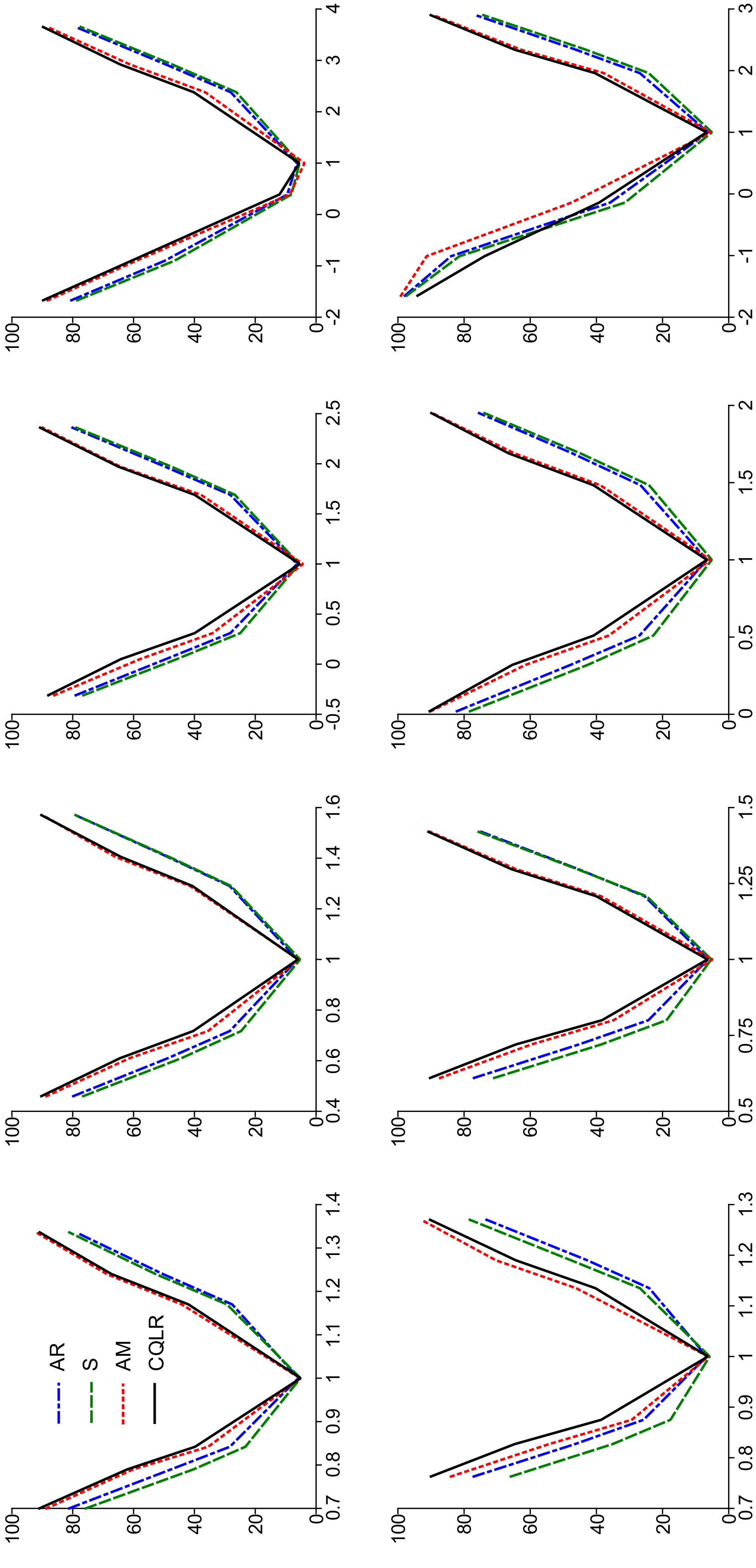

FIGURE SM-1. Power of CQLR, AM, S, and AR as function of θ for $\rho = 0$ and $\pi = 1, 0.5, 0.2, 0.1$; first/second row $g = 3/4$.

From now on, whenever a subindex $\widehat{A}$ appears on an object defined in Section 9.2, it means that it is defined as in Section 9.2 but resulting from a moment condition model defined in terms of the nonredundant moment conditions $g_{\widehat{A}i}(\theta,\beta)$. In particular,

$$\begin{aligned}
\widehat{\Omega}_{\widehat{A}n}(\theta,\beta) &:= n^{-1}\sum_{i=1}^{n} g_{\widehat{A}i}(\theta,\beta)g_{\widehat{A}i}(\theta,\beta)' - \widehat{g}_{\widehat{A}n}(\theta,\beta)\widehat{g}_{\widehat{A}n}(\theta,\beta)' \in R^{\widehat{r}_n(\theta_0,\widetilde{\beta}_n)\times\widehat{r}_n(\theta_0,\widetilde{\beta}_n)},\\
\widehat{g}_{\widehat{A}n}(\theta,\beta) &:= n^{-1}\sum_{i=1}^{n} g_{\widehat{A}i}(\theta,\beta), \quad \text{and} \\
\widehat{\beta}_{\widehat{A}n} &:= \arg\min_{\beta\in B}\left\|\widehat{\varphi}_{\widehat{A}n}\widehat{g}_{\widehat{A}n}(\theta_0,\beta)\right\|^2,
\end{aligned} \tag{13.3}$$

where $\widehat{\varphi}_{\widehat{A}n} \in R^{\widehat{r}_n(\theta_0,\widetilde{\beta}_n)\times\widehat{r}_n(\theta_0,\widetilde{\beta}_n)}$ satisfies

$$\widehat{\varphi}'_{\widehat{A}n}\widehat{\varphi}_{\widehat{A}n} = \widehat{\Omega}^{-1}_{\widehat{A}n}(\theta_0,\widetilde{\beta}_n). \tag{13.4}$$

The subvector SR-AR and SR-CQLR test statistics, denoted by $\text{SR-AR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n})$ and $\text{SR-QLR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n})$, respectively, are defined as the nonrobust tests are defined, but based on the moment functions $g_{\widehat{A}i}(\theta,\beta)$ in place of $g_i(\theta,\beta)$ and using the GMM estimator $\widehat{\beta}_{\widehat{A}n}$ rather than $\widehat{\beta}_n$ to estimate the nuisance parameter β. When $\widehat{r}_n(\theta_0,\widetilde{\beta}_n) > 0$, the subvector SR-AR test at nominal size $\alpha \in (0,1)$ rejects if

$$\text{SR-AR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n}) > \chi^2_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),1-\alpha}. \tag{13.5}$$

The subvector SR-CQLR test at nominal size $\alpha \in (0,1)$ rejects if

$$\text{SR-QLR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n}) > c_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),p}\big(n^{1/2}\widehat{D}^*_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}),\widetilde{J}_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}),1-\alpha\big). \tag{13.6}$$

If $\widehat{r}_n(\theta_0,\widetilde{\beta}_n) = 0$, then $\text{SR-AR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n})$ and $\text{SR-QLR}_n^S(\theta_0,\widehat{\beta}_{\widehat{A}n}) := 0$ and $\chi^2_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),1-\alpha}$ and $c_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),p}(n^{1/2}\widehat{D}^*_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}),\widetilde{J}_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}),1-\alpha) := 0$ and the two tests do not reject H_0.

Next, we define the parameter spaces for the subvector SR-AR and SR-CQLR tests. We denote the column and null spaces of a matrix by $\text{col}(\cdot)$ and $N(\cdot)$, respectively. We impose the conditions in $\mathcal{F}^S_{\text{AR},1}$ defined in (50) which guarantee consistency of the preliminary estimator $\widetilde{\beta}_n$. The parameter space $\mathcal{F}^S_{\text{AR},2}$ defined in (51) is modified in four ways: (i) the condition $\lambda_{\min}(E_F g_i g_i') \geq \delta$ is dropped, (ii) the condition $E_F \sup_{\beta\in B(\beta^*,\zeta)} \|\Pi_{1F}^{-1/2}(\beta)A_F(\beta)'(g_i(\beta) - E_F g_i(\beta))\|^2 \leq M$ is added, (iii) all of the remaining conditions are formulated in terms of the moment functions $\Pi_{1F}^{-1/2}(\theta_0,\beta^*)A_F(\theta_0,\beta^*)'g_i(\theta_0,\beta)$, rather than $g_i(\theta_0,\beta)$, and (iv) the condition, for some $\zeta_\dagger > 0$, $N(\Omega_F(\theta_0,\beta^*)) = N(\Omega_F(\theta_0,\beta))$ for all $\beta \in B(\beta^*,\zeta_\dagger)$, where β^* denotes the true value of β, is added. Call the resulting space $\mathcal{F}^{S,\text{SR}}_{\text{AR},2}$. We define the null parameter space for the subvector SR AR test to be

$$\mathcal{F}^{S,\text{SR}}_{\text{AR}} := \mathcal{F}^S_{\text{AR},1} \cap \mathcal{F}^{S,\text{SR}}_{\text{AR},2}. \tag{13.7}$$

The null parameter space for the subvector SR-CQLR test, denoted by $\mathcal{F}^{S,\mathrm{SR}}$, is defined as $\mathcal{F}^{S}$ is defined in (53) with the following modifications. First, $\mathcal{F}^{S}_{\mathrm{AR}}$ is replaced by $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR}}$, and second, all of the remaining conditions are formulated in terms of the moment functions $\Pi_{1F}^{-1/2}(\theta_0,\beta^*)A_F(\theta_0,\beta^*)'\times g_i(\theta_0,\beta)$, rather than $g_i(\theta_0,\beta)$.

We can also construct confidence regions for θ with correct asymptotic confidence size by inversion of the subvector SR-AR and SR-CQLR tests. The relevant parameter spaces are given by

$$\begin{aligned}\mathcal{F}^{S,\mathrm{SR}}_{\Theta,\mathrm{AR}} &:= \big\{(F,\beta,\theta_0):(F,\beta)\in\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR}}(\theta_0),\ \theta_0\in\Theta\big\} \quad \text{and}\\ \mathcal{F}^{S,\mathrm{SR}}_{\Theta} &:= \big\{(F,\beta,\theta_0):(F,\beta)\in\mathcal{F}^{S,\mathrm{SR}}(\theta_0),\ \theta_0\in\Theta\big\},\end{aligned} \tag{13.8}$$

respectively, where $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR}}(\theta_0)$ and $\mathcal{F}^{S,\mathrm{SR}}(\theta_0)$ denote $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR}}$ and $\mathcal{F}^{S,\mathrm{SR}}$ with the latter sets' dependence on θ_0 made explicit.

Note that condition (iv) of $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR},2}$ can be restrictive. We now discuss a scenario in which it holds. Consider the case where the moment functions are of the form

$$g_i(\theta,\beta)=u_i(\theta,\beta)Z_i, \tag{13.9}$$

where Z_i is a vector of instrument variables, the residual $u_i(\theta,\beta)$ is scalar, $E_F u_i^2(\theta_0,\beta^*)>0$, and $E_F u_i^2(\theta_0,\beta^*)Z_iZ_i'$ factors into $E_F u_i^2(\theta_0,\beta^*)E_F Z_iZ_i'$. (Note that the latter condition is implied by conditional homoskedasticity: $E_F(u_i^2(\theta_0,\beta^*)|Z_i)=\sigma^2$ a.s. for some constant $\sigma^2>0$.) Under these conditions, $\Omega_F(\theta_0,\beta)=E_F u_i^2(\theta_0,\beta)Z_iZ_i'-E_F u_i(\theta_0,\beta)\times Z_iE_F u_i(\theta_0,\beta)Z_i'$, and $\Omega_F(\theta_0,\beta^*)=E_F u_i^2(\theta_0,\beta^*)E_F Z_iZ_i'$. If $A_F\Pi_F A_F'$ denotes a singular value decomposition of $E_F Z_iZ_i'$ with $\Pi_F=\mathrm{Diag}(\Pi_{1F},\Pi_{0F})$, where $\Pi_{1F}\in R^r$ contains the nonzero eigenvalues and $\Pi_{0F}\in R^{k-r}$ contains the zero eigenvalues and $A_F=(A_{1F},A_{0F})$ is a decomposition of the matrix of eigenvectors corresponding to the nonzero/zero eigenvalues, respectively, then $A_{0F}=N(\Omega_F(\theta_0,\beta^*))$. It follows that $A_F'E_F Z_iZ_i'A_F=\mathrm{Diag}(\Pi_{1F},\Pi_{0F})$, and thus, in particular, $E_F(A_F'Z_i)_j^2=0$ for $j=r+1,\dots,k$. Therefore, $(A_F'Z_i)_j=0$ a.s. for $j=r+1,\dots,k$. But then $A_F'\Omega_F(\theta_0,\beta)A_F=E_F u_i^2(\theta_0,\beta)A_F'Z_iZ_i'A_F-E_F u_i(\theta_0,\beta)A_F'Z_i\cdot E_F u_i(\theta_0,\beta)Z_i'A_F$, for any $\beta\in B$, equals a block diagonal matrix with the lower right block equal to $0^{(k-r)\times(k-r)}$. This implies $\Omega_F(\theta_0,\beta)A_{0F}=0^{k\times(k-r)}$, which implies that $N(\Omega_F(\theta_0,\beta^*))\subset N(\Omega_F(\theta_0,\beta))$. Thus, in the setup of (13.9), condition (iv) of $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR},2}$ holds provided $N(\Omega_F(\theta_0,\beta^*))$ is not a strict subset of $N(\Omega_F(\theta_0,\beta))$.

Note that condition (iv) of $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR},2}$ implies that $r_F(\beta)$ is constant for all $\beta\in B(\beta^*,\zeta_\dagger)$. Furthermore, it implies that $\mathrm{col}(\Omega_F(\theta_0,\beta^*))=\mathrm{col}(\Omega_F(\theta_0,\beta))$ for all $\beta\in B(\beta^*,\zeta_\dagger)$, that is, that $\mathrm{col}(A_F(\beta^*))=\mathrm{col}(A_F(\beta))$ for all $\beta\in B(\beta^*,\zeta_\dagger)$. Therefore, without loss of generality, under condition (iv) of $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR},2}$, we can take $A_F(\beta)=A_F(\beta^*)$ for all $\beta\in B(\beta^*,\zeta_\dagger)$, that is, $A_F(\beta)$ does not depend on β for all $\beta\in B(\beta^*,\zeta_\dagger)$.

The asymptotic size and similarity results for the subvector SR-AR and SR-CQLR tests are as follows.

Theorem 13.1. *Suppose Assumption* gB *holds. The asymptotic sizes of the* subvector SR-AR *and* SR-CQLR *tests defined in* (13.5) *and* (13.6)*, respectively, equal their nominal size* $\alpha\in(0,1)$ *for the null parameter spaces* $\mathcal{F}^{S,\mathrm{SR}}_{\mathrm{AR}}$ *and* $\mathcal{F}^{S,\mathrm{SR}}$*, respectively. These tests* *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 corresponding* subvector SR-AR *and* SR-CQLR CSs *for the parameter spaces $\mathcal{F}_{\Theta,\mathrm{AR}}^{S,\mathrm{SR}}$ and $\mathcal{F}_{\Theta}^{S,\mathrm{SR}}$.*

Comment. Theorem 13.1 is proved in Section 31 below.

14. Miscellanei

14.1 *Moore–Penrose expression for the SR-AR statistic*

The expression for the SR-AR statistic given in (15) of AG2 holds by the following calculations. For notational simplicity, we suppress the dependence of quantities on θ. We have

$$\begin{aligned}
\text{SR-AR}_n &= n\widehat{g}_n'\widehat{A}_n\big(\widehat{A}_n'\widehat{\Omega}_n\widehat{A}_n\big)^{-1}\widehat{A}_n'\widehat{g}_n \\
&= n\widehat{g}_n'\widehat{A}_n\big(\widehat{A}_n'[\widehat{A}_n,\widehat{A}_n^{\perp}]\widehat{\Pi}_n[\widehat{A}_n,\widehat{A}_n^{\perp}]'\widehat{A}_n\big)^{-1}\widehat{A}_n'\widehat{g}_n \\
&= n\widehat{g}_n'\widehat{A}_n\widehat{\Pi}_{1n}^{-1}\widehat{A}_n'\widehat{g}_n \quad \text{and} \\
n\widehat{g}_n'\widehat{\Omega}_n^{+}\widehat{g}_n &= n\widehat{g}_n'[\widehat{A}_n,\widehat{A}_n^{\perp}]\begin{bmatrix}\widehat{\Pi}_{1n}^{-1} & 0^{\widehat{r}_n\times(k-\widehat{r}_n)} \\ 0^{(k-\widehat{r}_n)\times\widehat{r}_n} & 0^{(k-\widehat{r}_n)(k-\widehat{r}_n)}\end{bmatrix}[\widehat{A}_n,\widehat{A}_n^{\perp}]'\widehat{g}_n \\
&= n\widehat{g}_n'\widehat{A}_n\widehat{\Pi}_{1n}^{-1}\widehat{A}_n'\widehat{g}_n,
\end{aligned} \tag{14.1}$$

where the spectral decomposition of $\widehat{\Omega}_n$ given in (10) and (11) is used once in each equation above. It is not the case that $\text{SR-AR}_n(\theta)$ equals the rhs expression in (15) with probability one when $\widehat{\Omega}_n^{+}(\theta)$ is replaced by an arbitrary generalized inverse of $\widehat{\Omega}_n(\theta)$.

The expression for the SR-AR statistic given in (13) is preferable to the Moore–Penrose expression in (15) for the derivation of the asymptotic results for the SR-AR test.

14.2 *Computation implementation*

The computation times given in Section 5.3 are for the model in Section 10 for the country Australia, although the choice of country has very little effect on the times. The computation times for the PI-CLC, MM1-SU, and MM2-SU tests depend greatly on the choice of implementation parameters. For the PI-CLC test, these include (i) the number of linear combination coefficients "a" considered in the search over [0, 1], which we take to be 100, (ii) the number of simulation repetitions used to determine the best choice of "a," which we take to be 2000, and (iii) the number of alternative parameter values considered in the search for the best "a," which we take to be 41 for $p = 1$. For the MM1-SU and MM2-SU tests, the implementation parameters include (i) the number of variables in the discretization of the maximization problem, which we take to be 1000, and (ii) the number of points used in the numerical approximations of the integrals $h1$ and $h2$ that appear in the definitions of these tests, which we take to be 1000. The run-times for the PI-CLC, MM1-SU, and MM2-SU tests exclude some items, such as a critical value look up table for the PI-CLC test, that only need to be computed once when carrying out multiple tests. The computations are done in GAUSS using the lmpt application to do the linear programming required by the MM1-SU and MM2-SU tests. Note that the computation time for the SR-CQLR test could be reduced by using a look-up table for the data-dependent critical values, which depend on p singular values. This would be most useful when $p = 2$.

15. SR-CQLR$_P$ TEST

In this section, we define the SR-CQLR$_P$ test, which is quite similar to the SR-CQLR test, but relies on $g_i(\theta)$ having a product form. This form is

$$g_i(\theta) = u_i(\theta) Z_i, \tag{15.1}$$

where Z_i is a k vector of IVs, $u_i(\theta)$ is a scalar residual, and the (random) function $u_i(\cdot)$ is known. This is the case considered in Stock and Wright (2000). It covers many GMM situations, but can be restrictive. For example, it rules out Hansen and Scheinkman's (1995) moment conditions for continuous-time Markov processes, the moment conditions often used with dynamic panel models, for example, see Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1995), and moment conditions of the form $g_i(\theta) = u_i(\theta) \otimes Z_i$, where $u_i(\theta)$ is a vector.

The SR-CQLR$_P$ test reduces asymptotically to Moreira's (2003) CLR test in the homoskedastic linear IV regression model with fixed IVs for sequences of distributions in all identification categories. In contrast, the SR-CQLR test does so only under sequences in the standard weak, semi-strong, and strong identification categories; see Section 6.2 for the definitions of these identification categories.

15.1 SR-CQLR$_P$ *parameter space*

When (15.1) holds, we define

$$\begin{aligned} u_{\theta i}(\theta) &:= \frac{\partial}{\partial \theta} u_i(\theta) \in R^p \quad \text{and} \\ u_i^*(\theta) &:= \begin{pmatrix} u_i(\theta) \\ u_{\theta i}(\theta) \end{pmatrix} \in R^{p+1}, \quad \text{and we have } G_i(\theta) = Z_i u_{\theta i}(\theta)'.^{30} \end{aligned} \tag{15.2}$$

The null parameter space for the SR-CQLR$_P$ test is

$$\begin{aligned} \mathcal{F}_P^{\mathrm{SR}} := \big\{ & F \in \mathcal{F}^{\mathrm{SR}} : E_F \big\| \Pi_{1F}^{-1/2} A_F' Z_i \big\|^{4+\gamma} \le M, E_F \big\| u_i^* \big\|^{2+\gamma} \le M, \text{ and} \\ & E_F \big\| \Pi_{1F}^{-1/2} A_F' Z_i \big\|^2 u_i^2 1\big(u_i^2 > c\big) \le 1/2 \big\}, \end{aligned} \tag{15.3}$$

[30] As with $G(W_i, \theta)$ defined in (2), $u_{\theta i}(\theta)$ need not be a vector of partial derivatives of $u_i(\theta)$ for all sample realizations of the observations. It could be the vector of partial derivatives of $u_i(\theta)$ almost surely, rather than for all W_i, which allows $u_i(\theta)$ to have kinks, or a vector of finite differences of $u_i(\theta)$. For the asymptotic size results for the SR-CQLR$_2$ test given below to hold, $u_{\theta i}(\theta)$ can be any random p vector that satisfies the conditions in $\mathcal{F}_2^{\mathrm{SR}}$ (defined in (15.3)).

for some $\gamma > 0$ and some $M, c < \infty$, where Π_{1F} and A_F are defined in Section 3.2. By definition, $\mathcal{F}_P^{\mathrm{SR}} \subset \mathcal{F}^{\mathrm{SR}} \subset \mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$.

The conditions in $\mathcal{F}_P^{\mathrm{SR}}$ are only marginally stronger than those in $\mathcal{F}^{\mathrm{SR}}$, defined in (6). A sufficient condition for the last condition in $\mathcal{F}_P^{\mathrm{SR}}$ to hold for some $c < \infty$ is $E_F u_i^4 \leq M_*$ for some sufficiently large $M_* < \infty$ (using the first condition in $\mathcal{F}_P^{\mathrm{SR}}$ and the Cauchy–Bunyakovsky–Schwarz inequality).

The conditions in $\mathcal{F}_P^{\mathrm{SR}}$ place no restrictions on the column rank or singular values of $E_F G_i$. The conditions in $\mathcal{F}_P^{\mathrm{SR}}$ also place no restrictions on the variance matrix $\Omega_F := E_F g_i g_i'$ of g_i, such as $\lambda_{\min}(\Omega_F) \geq \delta$ for some $\delta > 0$ or $\lambda_{\min}(\Omega_F) > 0$. Hence, Ω_F can be singular.

In Section 3.2, it is noted that identification failure yields singularity of Ω_F in likelihood scenarios. It also does so in all quasi-likelihood scenarios when the quasi-likelihood does not depend on some element(s) of θ (or some transformation(s) of θ) for θ in a neighborhood of θ_0.[31] Another example where Ω_F may be singular is the following homoskedastic linear IV model: $y_{1i} = Y_{2i}\beta + U_i$ and $Y_{2i} = Z_i'\pi + V_{2i}$, where all quantities are scalars except $Z_i, \pi \in R^{d_Z}$ and $\theta = (\beta, \pi')' \in R^{1+d_Z}$. The corresponding reduced-form equations are $y_{1i} = Z_i'\pi\beta + V_{1i}$ and $Y_{2i} = Z_i'\pi + V_{1i}$, where $V_{1i} = U_i + V_{2i}\beta$. We assume $EU_i = EV_{2i} = 0$, $EU_iZ_i = EV_{2i}Z_i = 0^{d_Z}$, and $E(V_iV_i'|Z_i) = \Sigma_V$ a.s. for $V_i := (V_{1i}, V_{2i})'$ and some 2×2 constant matrix Σ_V. The moment conditions for θ are $g_i(\theta) = ((y_{1i} - Z_i'\pi\beta)Z_i', (Y_{2i} - Z_i'\pi)Z_i')' \in R^k$, where $k = 2d_Z$. The variance matrix $\Sigma_V \otimes EZ_iZ_i'$ of $g_i(\theta_0) = (V_{1i}Z_i', V_{2i}Z_i')'$ is singular whenever the covariance between the reduced-form errors V_{1i} and V_{2i} is one (or minus one) or EZ_iZ_i' is singular. In this model, we are interested in joint inference concerning β and π. This is of interest when one wants to see how the magnitude of π affects the range of plausible β values.

Section 3.2 and Grant (2013) note that Ω_F can be singular in the model for interest rate dynamics in Jegannathan, Skoulakis, and Wang (2002, Section 6.2) (JSW). JSW considered five moment conditions and a four-dimensional parameter θ. The first four moment functions in JSW are $(a(b-r_i)r_i^{-2\gamma} - \gamma\sigma^2 r_i^{-1}, a(b-r_i)r_i^{-2\gamma+1} - (\gamma - 1/2)\sigma^2, (b - r_i)r_i^{-a} - (1/2)\sigma^2 r_i^{2\gamma-a-1}, a(b-r_i)r_i^{-\sigma} - (1/2)\sigma^3 r_i^{2\gamma-\sigma-1})'$, where $\theta = (a, b, \sigma, \gamma)'$ and r_i is the interest rate. The second and third functions are equivalent if $\gamma = (a+1)/2$; the second and fourth functions are equivalent if $\gamma = (\sigma+1)/2$; and the third and fourth functions are equivalent if $\sigma = a$. Hence, the variance matrix of the moment functions is singular when one or more of these three restrictions on the parameters holds. When any two of these restrictions hold, the parameter also is unidentified.

Next, we specify the parameter space for (F, θ) that is used with the SR-CQLR$_P$ CS. It is denoted by $\mathcal{F}_{\Theta,P}^{\mathrm{SR}}$. For notational simplicity, the dependence of the parameter space $\mathcal{F}_P^{\mathrm{SR}}$ in (15.3) on θ_0 is suppressed. When dealing with the SR-CQLR$_P$ CS, rather than test, we make the dependence explicit and write it as $\mathcal{F}_P^{\mathrm{SR}}(\theta_0)$. We define

$$\mathcal{F}_{\Theta,P}^{\mathrm{SR}} := \{(F, \theta_0) : F \in \mathcal{F}_P^{\mathrm{SR}}(\theta_0),\ \theta_0 \in \Theta\}. \tag{15.4}$$

[31] In this case, the moment functions equal the quasi-score and some element(s) or linear combination(s) of elements of moment functions, equal zero a.s. at θ_0 (because the quasi-score is of the form $g_i(\theta) = (\partial/\partial\theta)\log f(W_i, \theta)$ for some density or conditional density $f(W_i, \theta)$). This yields singularity of the variance matrix of the moment functions and of the expected Jacobian of the moment functions.

15.2 *Definition of the* SR-CQLR$_P$ *test*

First, we define the CQLR$_P$ test without the SR extension. It uses the statistics $\widehat{g}_n(\theta)$, $\widehat{\Omega}_n(\theta)$, AR$_n(\theta)$, and $\widehat{D}_n(\theta)$ (defined in (8), (9), and (18)). The CQLR$_P$ test also uses analogues $\widetilde{R}_n(\theta)$ and $\widetilde{V}_n(\theta)$ of $\widehat{R}_n(\theta)$ and $\widehat{V}_n(\theta)$ (defined in (19)), respectively, which are defined as follows:

$$
\begin{aligned}
\widetilde{R}_n(\theta) &:= \big(B(\theta)' \otimes I_k\big)\widetilde{V}_n(\theta)\big(B(\theta) \otimes I_k\big) \in R^{(p+1)k\times(p+1)k}, \quad \text{where} \\
\widetilde{V}_n(\theta) &:= n^{-1}\sum_{i=1}^{n}\Big(\big(u_i^*(\theta) - \widehat{u}_{in}^*(\theta)\big)\big(u_i^*(\theta) - \widehat{u}_{in}^*(\theta)\big)'\Big) \otimes \big(Z_i Z_i'\big) \in R^{(p+1)k\times(p+1)k}, \\
\widehat{u}_{in}^*(\theta) &:= \widetilde{\Xi}_n(\theta)' Z_i \in R^{p+1}, \\
\widetilde{\Xi}_n(\theta) &:= \big(Z_{n\times k}' Z_{n\times k}\big)^{-1} Z_{n\times k}' U^*(\theta) \in R^{k\times(p+1)}, \\
Z_{n\times k} &:= (Z_1, \dots, Z_n)' \in R^{n\times k}, \\
U^*(\theta) &:= \big(u_1^*(\theta), \dots, u_n^*(\theta)\big)' \in R^{n\times(p+1)}, \quad \text{and} \\
B(\theta) &:= \begin{pmatrix} 1 & 0_p' \\ -\theta & -I_p \end{pmatrix} \in R^{(p+1)\times(p+1)},
\end{aligned}
\tag{15.5}
$$

where $u_i^*(\theta) := (u_i(\theta), u_{\theta i}(\theta)')'$ is defined in (15.2). Note that (i) $\widetilde{V}_n(\theta)$ is an estimator of the variance matrix of the moment functions and their vectorized derivatives, (ii) $\widetilde{V}_n(\theta)$ exploits the functional form of the moment conditions given in (15.1), (iii) $\widetilde{V}_n(\theta)$ typically is not of a Kronecker product form (because of the average over $i = 1, \dots, n$), and (iv) $\widehat{u}_{in}^*(\theta)$ is the best linear predictor of $u_i^*(\theta)$ based on $\{Z_i : n \geq 1\}$. The estimators $\widetilde{R}_n(\theta)$, $\widetilde{V}_n(\theta)$, and $\widetilde{\Sigma}_n(\theta)$ (defined immediately below) are defined so that the SR-CQLR$_P$ test, which employs them, is asymptotically equivalent to Moreira's (2003) CLR test under all strengths of identification in the homoskedastic linear IV model with fixed IVs and p rhs endogenous variables for any $p \geq 1$; see Section 19 for details. The SR-CQLR$_P$ test differs from the SR-CQLR test because $\widetilde{V}_n(\theta)$ (and the statistics that depend on it) differs from $\widehat{V}_n(\theta)$ (and the statistics that depend on it).

We define $\widetilde{\Sigma}_n(\theta) \in R^{(p+1)\times(p+1)}$ just as $\widehat{\Sigma}_n(\theta)$ is defined in (20) and (21), but with $\widetilde{R}_n(\theta)$ in place of $\widehat{R}_n(\theta)$. We define $\widetilde{D}_n^*(\theta)$ just as $\widehat{D}_n^*(\theta)$ is defined in (23), but with $\widetilde{\Sigma}_n(\theta)$ in place of $\widehat{\Sigma}_n(\theta)$. That is,

$$
\begin{aligned}
\widetilde{D}_n^*(\theta) &:= \widehat{\Omega}_n(\theta)^{-1/2}\widehat{D}_n(\theta)\widetilde{L}_n^{1/2}(\theta) \in R^{k\times p}, \quad \text{where} \\
\widetilde{L}_n(\theta) &:= (\theta, I_p)\big(\widetilde{\Sigma}_n^{\varepsilon}(\theta)\big)^{-1}(\theta, I_p)'.
\end{aligned}
\tag{15.6}
$$

The estimator $\widetilde{\Sigma}_n(\theta)$ is an estimator of a matrix that could be singular or nearly singular in some cases. For example, in the homoskedastic linear IV model (see Section 19.1 below) $\widetilde{\Sigma}_n(\theta)$ is an estimator of the variance matrix Σ_V of the reduced-form errors when θ is the true parameter, and Σ_V could be singular or nearly singular. In the definition of $\widetilde{L}_n(\theta)$ above, we use an eigenvalue-adjusted version of $\widetilde{\Sigma}_n(\theta)$, denoted by $\widetilde{\Sigma}_n^{\varepsilon}(\theta)$, whose condition number (i.e., $\lambda_{\max}(\widehat{\Sigma}_n(\theta))/\lambda_{\min}(\widehat{\Sigma}_n(\theta))$) is bounded above by construction. Based on the finite-sample simulations, we recommend using $\varepsilon = 0.01$.

The QLR_P statistic without the SR extension, denoted by $\text{QLR}_{Pn}(\theta)$, is defined just as $\text{QLR}_n(\theta)$ is defined in (23), but with $\widetilde{D}_n^*(\theta)$ in place of $\widehat{D}_n^*(\theta)$. For $\alpha \in (0,1)$, the nominal size α CQLR_P test (without the SR extension) rejects $H_0 : \theta = \theta_0$ if

$$\text{QLR}_{Pn}(\theta_0) > c_{k,p}\big(n^{1/2}\widetilde{D}_n^*(\theta_0), 1-\alpha\big), \tag{15.7}$$

where $c_{k,p}(\cdot, 1-\alpha)$ is defined in (24). The nominal size $100(1-\alpha)\%$ CQLR_P CS is $\text{CS}_{\text{CQLR}_P,n} := \{\theta_0 \in \Theta : \text{QLR}_{Pn}(\theta_0) \le c_{k,p}(n^{1/2}\widetilde{D}_n^*(\theta_0), 1-\alpha)\}$.

The CQLR_P test statistic and critical value satisfy the following invariance properties.

Lemma 15.1. *The statistics* QLR_{Pn}, $c_{k,p}(n^{1/2}\widetilde{D}_n^*, 1-\alpha)$, $\widetilde{D}_n^{*\prime}\widetilde{D}_n^*$, AR_n, $\widehat{u}_{in}^*$, $\widetilde{\Sigma}_n$, *and* $\widetilde{L}_n$ *are invariant to the transformation* $(Z_i, u_i^*) \rightsquigarrow (MZ_i, u_i^*)$ $\forall i \le n$ *for any* $k \times k$ *nonsingular matrix* M. *This transformation induces the following transformations*: $g_i \rightsquigarrow Mg_i$ $\forall i \le n$, $G_i \rightsquigarrow MG_i$ $\forall i \le n$, $\widehat{g}_n \rightsquigarrow M\widehat{g}_n$, $\widehat{G}_n \rightsquigarrow M\widehat{G}_n$, $\widehat{\Omega}_n \rightsquigarrow M\widehat{\Omega}_n M'$, $\widehat{\Gamma}_{jn} \rightsquigarrow M\widehat{\Gamma}_{jn}M'$ $\forall j \le p$, $\widehat{D}_n \rightsquigarrow M\widehat{D}_n$, $Z_{n\times k} \rightsquigarrow Z_{n\times k}M'$, $\widetilde{\Xi}_n \rightsquigarrow M'^{-1}\widetilde{\Xi}_n$, $\widetilde{V}_n \rightsquigarrow (I_{p+1}\otimes M)\widetilde{V}_n(I_{p+1}\otimes M')$, *and* $\widetilde{R}_n \rightsquigarrow (I_{p+1}\otimes M)\widetilde{R}_n(I_{p+1}\otimes M')$.

Comment. This lemma is important because it implies that one can obtain the correct asymptotic size of the CQLR_P test defined above without assuming that $\lambda_{\min}(\Omega_F)$ is bounded away from zero. It suffices that Ω_F is nonsingular. The reason is that (in the proofs) one can transform the moments by $g_i \rightsquigarrow M_F g_i$, where $M_F\Omega_F M_F' = I_k$, such that the transformed moments have a variance matrix whose eigenvalues are bounded away from zero for some $\delta > 0$ (since $\text{Var}_F(M_F g_i) = I_k$) even if the original moments g_i do not.

For the CQLR_P test with the SR extension, we define $\widehat{D}_{An}(\theta)$ as in (26). We let $Z_{Ai}(\theta) := \widehat{A}_n(\theta)' Z_i \in R^{\widehat{r}_n(\theta)}$ and $Z_{An\times k}(\theta) := Z_{n\times k}\widehat{A}_n(\theta) \in R^{n\times \widehat{r}_n(\theta)}$. We define

$$\begin{aligned}
\widetilde{V}_{An}(\theta) &:= n^{-1}\sum_{i=1}^n \big(\big(u_i^*(\theta) - \widehat{u}_{Ain}^*(\theta)\big)\big(u_i^*(\theta) - \widehat{u}_{Ain}^*(\theta)\big)'\big) \otimes \big(Z_{Ai}(\theta)Z_{Ai}(\theta)'\big) \\
&\quad \in R^{(p+1)\widehat{r}_n(\theta)\times(p+1)\widehat{r}_n(\theta)}, \quad \text{where} \qquad\qquad (15.8)\\
\widehat{u}_{Ain}^*(\theta) &:= \widetilde{\Xi}_{An}(\theta)' Z_{Ai}(\theta) \in R^{p+1}, \\
\widetilde{\Xi}_{An}(\theta) &:= \big(Z_{An\times k}(\theta)' Z_{An\times k}(\theta)\big)^{-1} Z_{An\times k}(\theta)' U^*(\theta) \in R^{\widehat{r}_n(\theta)\times(p+1)},
\end{aligned}$$

and $\widehat{r}_n(\theta)$ and $\widehat{A}_n(\theta)$ are defined in (10) and (11), respectively. In addition, we define $\widetilde{R}_{An}(\theta)$, $\widetilde{\Sigma}_{An}(\theta)$, $\widetilde{L}_{An}(\theta)$, $\widetilde{D}_{An}^*(\theta)$, and $\widetilde{Q}_{An}(\theta)$ as $\widehat{R}_{An}(\theta)$, $\widehat{\Sigma}_{An}(\theta)$, $\widehat{L}_{An}(\theta)$, $\widehat{D}_{An}^*(\theta)$, and $\widehat{Q}_{An}(\theta)$ are defined, respectively, in (27) and (28), but with $\widetilde{V}_{An}(\theta)$ in place of $\widehat{V}_{An}(\theta)$ in the definition of $\widetilde{R}_{An}(\theta)$, with $\widetilde{R}_{An}(\theta)$ in place of $\widehat{R}_{An}(\theta)$ in the definition of $\widetilde{\Sigma}_{An}(\theta)$, and so on in the definitions of $\widetilde{L}_{An}(\theta)$, $\widetilde{D}_{An}^*(\theta)$, and $\widetilde{Q}_{An}(\theta)$. We define the test statistic $\text{SR-QLR}_{Pn}(\theta)$ as $\text{SR-QLR}_n(\theta)$ is defined in (28), but with $\widetilde{Q}_{An}(\theta)$ in place of $\widehat{Q}_{An}(\theta)$.

Given these definitions, the nominal size α SR-CQLR_P test rejects $H_0 : \theta = \theta_0$ if

$$\text{SR-QLR}_{Pn}(\theta_0) > c_{\widehat{r}_n(\theta_0),p}\big(n^{1/2}\widetilde{D}_{An}^*(\theta_0), 1-\alpha\big) \quad \text{or} \quad \widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}.\text{[32]} \tag{15.9}$$

[32] By definition, $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}$ does not hold if $\widehat{r}_n(\theta_0) = k$. If $\widehat{r}_n(\theta_0) = 0$, then $\text{SR-QLR}_{Pn}(\theta_0) := 0$ and $\chi^2_{\widehat{r}_n(\theta_0),1-\alpha} := 0$. In this case, $\widehat{A}_n^{\perp}(\theta_0) = I_k$ and the SR-CQLR_P test rejects H_0 if $\widehat{g}_n(\theta_0) \neq 0^k$.

The nominal size $100(1-\alpha)\%$ SR-CQLR$_P$ CS is $CS_{\text{SR-CQLR}_P,n} := \{\theta_0 \in \Theta : \text{SR-QLR}_{Pn}(\theta_0) \leq c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widetilde{D}^*_{An}(\theta_0), 1-\alpha)$ and $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}\}$.[33]

Two simple examples where the extra rejection condition in (15.9) for the SR-CQLR$_P$ test (and in (14) and (29) for the SR-AR and SR-CQLR tests, resp.) improves the power of these tests are the following. First, suppose $(X_{1i}, X_{2i})' \sim$ i.i.d. $N(\theta, \Omega_F)$, where $\theta = (\theta_1, \theta_2)' \in R^2$, Ω_F is a 2×2 matrix of ones, and the moment functions are $g_i(\theta) = (X_{1i} - \theta_1, X_{2i} - \theta_2)'$. In this case, Ω_F is singular, $\widehat{A}_n(\theta_0) = (1,1)'$ a.s., $\widehat{A}_n^{\perp}(\theta_0) = (1,-1)'$ a.s., the SR-AR statistic is a quadratic form in $\widehat{A}_n(\theta_0)'\widehat{g}_n(\theta_0) = \overline{X}_{1n} + \overline{X}_{2n} - (\theta_{10} + \theta_{20})$, where $\overline{X}_{mn} = n^{-1}\sum_{i=1}^n X_{mi}$ for $m = 1, 2$, and $A_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = \overline{X}_{1n} - \overline{X}_{2n} - (\theta_{10} - \theta_{20})$ a.s. If one does not use the extra rejection condition, then the SR-AR test has no power against alternatives $\theta = (\theta_1, \theta_2)'$ $(\neq \theta_0)$ for which $\theta_1 + \theta_2 = \theta_{10} + \theta_{20}$. The same is true for the SR-CQLR and SR-CQLR$_P$ tests (because the SR-QLR$_n$ and SR-QLR$_{Pn}$ test statistics depend on the SR-AR$_n$ test statistic). However, when the extra rejection condition is utilized, all $\theta \in R^2$ except those on the line $\theta_1 - \theta_2 = \theta_{10} - \theta_{20}$ are rejected with probability one (because $\overline{X}_{1n} - \overline{X}_{2n} = E_F X_{1i} - E_F X_{2i} = \theta_1 - \theta_2$ a.s.) and this includes all of the alternative θ values for which $\theta_1 + \theta_2 = \theta_{10} + \theta_{20}$.

Second, suppose $X_i \sim$ i.i.d. $N(\theta_1, \theta_2)$, $\theta = (\theta_1, \theta_2)' \in R^2$, the moment functions are $g_i(\theta) = (X_i - \theta_1, X_i^2 - \theta_1^2 - \theta_2)'$, and the null hypothesis is $H_0 : \theta = (\theta_{10}, \theta_{20})'$. Consider alternative parameters of the form $\theta = (\theta_1, 0)'$. Under θ, X_i has variance zero, $X_i = \overline{X}_n = \theta_1$ a.s., $X_i^2 = \overline{X_n^2} = \theta_1^2$ a.s., where $\overline{X_n^2} := n^{-1}\sum_{i=1}^n X_i^2$, $\widehat{g}_n(\theta_0) = (\theta_1 - \theta_{10}, \theta_1^2 - \theta_{10}^2 - \theta_{20})'$ a.s., $\widehat{\Omega}_n(\theta_0) = \widehat{g}_n(\theta_0)\widehat{g}_n(\theta_0)' - \widehat{g}_n(\theta_0)\widehat{g}_n(\theta_0)' = 0^{2\times 2}$ a.s. (provided $\widehat{\Omega}_n(\theta_0)$ is defined as in (8) with the sample means subtracted off), and $\widehat{r}_n(\theta_0) = 0$ a.s. In consequence, if one does not use the extra rejection condition, then the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests have no power against alternatives of the form $\theta = (\theta_1, 0)'$ (because, by definition, the test statistics and critical values equal zero when $\widehat{r}_n(\theta_0) = 0$). However, when the extra rejection condition is utilized, all alternatives of the form $\theta = (\theta_1, 0)'$ are rejected with probability one.[34,35,36,37]

[33] By definition, if $\widehat{r}_n(\theta_0) = k$, the condition $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}$ holds.

[34] This holds because the extra rejection condition in this case leads one to reject H_0 if $\overline{X}_n \neq \theta_{10}$ or $\overline{X_n^2} - \theta_{10}^2 - \theta_{20} \neq 0$, which is equivalent a.s. to rejecting if $\theta_1 \neq \theta_{10}$ or $\theta_1^2 - \theta_{10}^2 - \theta_{20} \neq 0$ (because $\overline{X}_n = \theta_1$ a.s. and $\overline{X_n^2} = \theta_1^2$ a.s. under θ), which in turn is equivalent to rejecting if $\theta \neq \theta_0$ (because if $\theta_{20} > 0$ one or both of the two conditions is violated when $\theta \neq \theta_0$ and if $\theta_{20} = 0$, then $\theta \neq \theta_0$ only if $\theta_1 \neq \theta_{10}$ since we are considering the case where $\theta_2 = 0$).

[35] In this second example, suppose the null hypothesis is $H_0 : \theta = (\theta_{10}, 0)'$. That is, $\theta_{20} = 0$. Then the SR-AR test rejects with probability zero under H_0 and the test is not asymptotically similar. This holds because $\widehat{g}_n(\theta_0) = (\overline{X}_n - \theta_{10}, \overline{X_n^2} - \theta_{10}^2)' = (0,0)'$ a.s., $\widehat{r}_n(\theta_0) = 0$ a.s., SR-AR$_n(\theta_0) = \chi^2_{\widehat{r}_n(\theta_0),1-\alpha} = 0$ a.s. (because $\widehat{r}_n(\theta_0) = 0$ a.s.), and the extra rejection condition leads one to reject H_0 if $\overline{X}_n \neq \theta_{10}$ or $\overline{X_n^2} - \theta_{10}^2 - \theta_{20} \neq 0$, which is equivalent to $\theta_{10} \neq \theta_{10}$ or $\theta_{10}^2 - \theta_{10}^2 - \theta_{20} \neq 0$ (because $X_i = \theta_1$ a.s.), which holds with probability zero.

As shown in Theorem 6.1, the SR-AR test is asymptotically similar (in a uniform sense) if one excludes null distributions F for which the $g_i(\theta_0) = 0^k$ a.s. under F, such as in the present example, from the parameter space of null distributions. But, the SR-AR test still has correct asymptotic size without such exclusions.

[36] We thank Kirill Evdokimov for bringing these two examples to our attention.

[37] An alternative definition of the SR-AR test is obtained by altering its definition given in Section 4 as follows. One omits the extra rejection condition given in (14), one defines the SR-AR statistic using a weight matrix that is nonsingular by construction when $\widehat{\Omega}_n(\theta_0)$ is singular, and one determines the critical value by simulation of the appropriate quadratic form in mean zero normal variates when $\widehat{\Omega}_n(\theta_0)$ is singular. For

When the sample variance matrix is singular, an alternative to using the SR-AR$_n(\theta_0)$ statistic is to arbitrarily delete some moment conditions. However, this typically leads to different test results given the same data and can yield substantially different power properties of the test depending on which moment conditions are deleted, which is highly undesirable. The following simple example illustrates this. Suppose $W_i = (W_{1i}, W_{2i}, W_{3i})'$ has a normal distribution with mean vector $(\theta_1, \theta_2, \theta_2)'$, all variances are equal to one, the covariance between W_{1i} and W_{2i} equals one, (W_{1i}, W_{2i}) and W_{3i} are independent, $g(W_i, \theta) = (W_{1i} - \theta_1, W_{2i} - \theta_2, W_{3i} - \theta_2)'$, and the null hypothesis is $H_0 : \theta = \theta_0$ for some $\theta_0 = (\theta_{01}, \theta_{02})' \in R^2$. The sample variance matrix is singular with probability one. A nonsingular sample variance matrix can be obtained by deleting the first moment condition or the second. If the first moment condition is deleted, the sample moments evaluated at θ_0 are $(\overline{W}_{n2} - \theta_{02}, \overline{W}_{n3} - \theta_{02})'$. If the second moment condition is deleted, they are $(\overline{W}_{n1} - \theta_{01}, \overline{W}_{n3} - \theta_{02})'$. When $\theta_1 - \theta_{10}$ and $\theta_2 - \theta_{20}$ are not equal (where θ_1 and θ_2 denote the true values), these two sets of moment conditions are not the same. Furthermore, it is clear that the power of the two tests based on these two sets of moment conditions is quite different because the first set of sample moments contains no information about θ_1, whereas the second set does.

15.3 *Asymptotic size of the* SR-CQLR$_P$ *test*

The correct asymptotic size and similarity results for the SR-CQLR$_P$ test are as follows.

Theorem 15.2. *The asymptotic size of the* SR-CQLR$_P$ *test defined in* (15.9) *equals its nominal size* $\alpha \in (0, 1)$ *for the null parameter spaces* $\mathcal{F}_P^{\mathrm{SR}}$. *Furthermore, this test is asymptotically similar* (*in a uniform sense*) *for the subset of this parameter space that excludes distributions* F *under which* $g_i = 0^k$ *a.s. Analogous results hold for the corresponding* SR-CQLR$_P$ CS *for the parameter space* $\mathcal{F}_{\Theta,P}^{\mathrm{SR}}$, *defined in* (15.4).

Comment. (i) For distributions F under which $g_i = 0^k$ a.s., the SR-CQLR$_P$ test rejects the null hypothesis with probability zero when the null is true. Hence, asymptotic similarity only holds when these distributions are excluded from the null parameter spaces.

(ii) The proof of Theorem 15.2 is given in Sections 16, 17, and 25–27 below.

---

example, such a weight matrix can be constructed by adjusting the eigenvalues of $\widehat{\Omega}_n(\theta_0)$ to be bounded away from zero, and using its inverse.

However, this method has two drawbacks. First, it sacrifices power relative to the definition of the SR-AR test in (14). The reason is that it does not reject H_0 with probability one when a violation of the nonstochastic part of the moment conditions occurs. This can be seen in the example with identities in Section 4 and the two examples given here.

Second, it cannot be used with the SR-CQLR and SR-CQLR$_2$ tests introduced in Sections 5 and 15. The reason is that these tests rely on the statistic $\widehat{D}_n(\theta_0)$, defined in (18), that employs $\widehat{\Omega}_n^{-1}(\theta_0)$ and if $\widehat{\Omega}_n^{-1}(\theta_0)$ is replaced by a matrix that is nonsingular by construction, such as the eigenvalue-adjusted matrix suggested above, then one does not obtain asymptotic independence of $\widehat{g}_n(\theta_0)$ and $\widehat{D}_n(\theta_0)$ after suitable normalization, which is needed to obtain the correct asymptotic size of the SR-CQLR and SR-CQLR$_2$ tests.

15.4 *Asymptotic efficiency of the* SR-CQLR$_P$ *test under strong identification*

Here, we show that the SR-CQLR$_P$ test is asymptotically efficient in a GMM sense under strong and semi-strong identification (when the variance matrix of the moments is nonsingular and the null parameter value is not on the boundary of the parameter space).

Suppose $k \geq p$. Let A_F and Π_{1F} be defined as in (4) and (5) and the paragraph following these equations with $\theta = \theta_0$. Define λ_F^*, Λ_P^*, and $\{\lambda_{n,h}^* : n \geq 1\}$ as λ_F, $\Lambda_{\mathrm{WU},P}$, and $\{\lambda_{n,h} : n \geq 1\}$, respectively, are defined in (16.16)–(16.18), but with g_i and G_i replaced by $g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i$ and $G_{Fi}^* := \Pi_{1F}^{-1/2} A_F' G_i$, with $\mathcal{F}_P$ replaced by $\mathcal{F}_P^{\mathrm{SR}}$ in the definition of $\mathcal{F}_{\mathrm{WU}}$, and with W_F $(:= W_1(W_{2F}))$ and U_F $(:= U_1(U_{2F}))$ defined as in (16.11) with g_i and G_i replaced by g_{Fi}^* and G_{Fi}^*. In addition, we restrict $\{\lambda_{n,h}^* : n \geq 1\}$ to be a sequence for which $\lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$. By definition, a sequence $\{\lambda_{n,h}^* : n \geq 1\}$ is said to exhibit strong or semi-strong identification if $n^{1/2} s_{pF_n}^* \to \infty$, where s_{pF}^* denotes the smallest singular value of $E_F G_{Fi}^*$.[38]

The LM$_n$ and LM$_n^{\mathrm{GMM}}$ statistics are defined in (32). Let $\chi^2_{p,1-\alpha}$ denote the $1-\alpha$ quantile of the χ^2_p distribution. The critical value for the LM$_n$ and LM$_n^{\mathrm{GMM}}$ tests is $\chi^2_{p,1-\alpha}$.

Theorem 15.3. *Suppose $k \geq p$. For any sequence $\{\lambda_{n,h}^* : n \geq 1\}$ that exhibits strong or semi-strong identification (i.e., for which $n^{1/2} s_{pF_n}^* \to \infty$) and for which $\lambda_{n,h}^* \in \Lambda_P^*$ $\forall n \geq 1$, we have*

(a) SR-QLR$_{Pn}$ $=$ QLR$_{Pn} + o_p(1) =$ LM$_n + o_p(1) =$ LM$_n^{\mathrm{GMM}} + o_p(1)$ *and*

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

Comment. (i) Theorem 15.3 establishes the asymptotic efficiency (in a GMM sense) of the SR-CQLR$_P$ test under strong and semi-strong identification. Theorem 15.3 provides asymptotic equivalence results under the null hypothesis, but by the definition of contiguity, these asymptotic equivalence results also hold under contiguous local alternatives.

(ii) The proof of Theorem 15.3 is given in Section 28.

15.5 *Summary comparison of CLR-type tests in Kleibergen (2005) and AG2*

We briefly summarize some of the results in AG1 and AG2 concerning Kleibergen's (2005) moment-variance-weighted CLR (MVW-CLR) and Jacobian-variance-weighted CLR (JVW-CLR) tests, the SR-CQLR test in AG2, and the SR-CQLR$_P$ test introduced above. (i) The MVW-CLR test has correct asymptotic size for all $p \geq 1$ (for the parameter space in AG1, which imposes nonsingularity of the variance matrix and some other

[38]The singular value s_{pF}^*, defined here, equals s_{pF}, defined in Section 6.2, for all F with $\lambda_{\min}(\Omega_F) > 0$, because in this case $\Omega_F = A_F \Pi_{1F} A_F'$, $\Omega_F^{-1/2} = A_F \Pi_{1F}^{-1/2} A_F'$, $\Omega_F^{-1/2} E_F G_i = A_F \Pi_{1F}^{-1/2} A_F' E_F G_i = A_F E_F G_{Fi}^*$, and A_F is an orthogonal $k \times k$ matrix. Since we consider sequences here with $\lambda_{\min}(\Omega_{F_n}) = \lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$, the definitions of strong and semi-strong identification used here and in Section 6.2 are equivalent.

conditions). (ii) The JVW-CLR test has correct asymptotic size for $p = 1$ (under similar conditions to the MVW-CLR test). (iii) For $p \geq 2$, AG1 provides an expression for the asymptotic size of the JWV-CLR test that depends on a vector of localization parameters. It is unknown whether the asymptotic size exceeds the nominal size. (iv) The MVW-CLR test is not asymptotically equivalent to Moreira's (2003) CLR test in the homoskedastic linear IV (HLIV) model for any $p \geq 1$. (v) The JVW-CLR test is asymptotically equivalent to Moreira's (2003) CLR test in the HLIV model for $p = 1$, but not for $p \geq 2$. (vi) The SR-CQLR test has correct asymptotic size for the parameter space $\mathcal{F}^{\mathrm{SR}}$ in Section 3.2, which is larger than the parameter space in (i) and (ii). (vii) The SR-CQLR$_P$ test has correct asymptotic size for the parameter space $\mathcal{F}_P^{\mathrm{SR}}$ ($\subset \mathcal{F}^{\mathrm{SR}}$). (viii) The SR-CQLR test is asymptotically equivalent to Moreira's (2003) CLR test in the HLIV model for $p = 1$, but not for $p \geq 2$, although the difference for $p \geq 2$ is only due to the difference between treating the IVs as random, rather than fixed. (ix) The SR-CQLR$_P$ test is asymptotically equivalent to Moreira's (2003) CLR test in the HLIV model for all $p \geq 1$.

16. Tests without the singularity-robust extension

The next two sections and Sections 25–27 below are devoted to the proof of Theorems 6.1 and 15.2. The proof proceeds in two steps. First, in this section, we establish the correct asymptotic size and asymptotic similarity of the tests and CSs without the SR extension for parameter spaces of distributions that bound $\lambda_{\min}(\Omega_F)$ away from zero. (These tests are defined in (9), (25), and (15.7).) We provide parts of the proof of this result in this section and other parts in Sections 25–27 below. Second, we extend the proof to the case of the SR tests and CSs. We provide the proof of this extension in Section 17 below.

16.1 *Asymptotic results for tests without the SR extension*

For the AR, CQLR, and CQLR$_P$ tests without the SR extension, we consider the following parameter spaces for the distribution F that generates the data under $H_0 : \theta = \theta_0$:

$$\begin{aligned}
\mathcal{F}_{\mathrm{AR}} &:= \left\{F : E_F g_i = 0^k, E_F\|g_i\|^{2+\gamma} \leq M, \text{ and } \lambda_{\min}\left(E_F g_i g_i'\right) \geq \delta\right\}, \\
\mathcal{F} &:= \left\{F \in \mathcal{F}_{\mathrm{AR}} : E_F\left\|\mathrm{vec}(G_i)\right\|^{2+\gamma} \leq M\right\}, \quad \text{and} \\
\mathcal{F}_P &:= \left\{F \in \mathcal{F} : E_F\|Z_i\|^{4+\gamma} \leq M, E_F\left\|u_i^*\right\|^{2+\gamma} \leq M, \lambda_{\min}\left(E_F Z_i Z_i'\right) \geq \delta\right\}
\end{aligned} \tag{16.1}$$

for some $\gamma, \delta > 0$ and $M < \infty$. By definition, $\mathcal{F}_P \subset \mathcal{F} \subset \mathcal{F}_{\mathrm{AR}}$. The parameter spaces $\mathcal{F}_{\mathrm{AR}}$, $\mathcal{F}$, and $\mathcal{F}_P$ are used for the AR, CQLR, and CQLR$_P$ tests, respectively. For the corresponding CSs, we use the parameter spaces: $\mathcal{F}_{\Theta,\mathrm{AR}} := \{(F, \theta_0) : F \in \mathcal{F}_{\mathrm{AR}}(\theta_0), \theta_0 \in \Theta\}$, $\mathcal{F}_{\Theta} := \{(F, \theta_0) : F \in \mathcal{F}(\theta_0), \theta_0 \in \Theta\}$, and $\mathcal{F}_{\Theta,P} := \{(F, \theta_0) : F \in \mathcal{F}_P(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{\mathrm{AR}}(\theta_0)$, $\mathcal{F}(\theta_0)$, and $\mathcal{F}_P(\theta_0)$ equal $\mathcal{F}_{\mathrm{AR}}$, $\mathcal{F}$, and $\mathcal{F}_P$, respectively, with their dependence on θ_0 made explicit.

Theorem 16.1. *The* AR, CQLR, *and* CQLR$_P$ *tests* (*without the* SR *extensions*), *defined in* (9), (25), *and* (15.7), *respectively, have asymptotic size equal to their nominal size* $\alpha \in (0, 1)$ *and are asymptotically similar* (*in a uniform sense*) *for the parameter spaces* $\mathcal{F}_{\mathrm{AR}}$, $\mathcal{F}$, *and*

$\mathcal{F}_P$, respectively. Analogous results hold for the corresponding AR, CQLR, *and* CQLR_P CSs *for the parameter spaces $\mathcal{F}_{\Theta,\mathrm{AR}}$, $\mathcal{F}_{\Theta}$, and $\mathcal{F}_{\Theta,P}$, respectively.*

Comment. (i) The first step of the proof of Theorems 6.1 and 15.2 is to prove Theorem 16.1.

(ii) Theorem 16.1 holds for both $k \geq p$ and $k < p$. Both cases are needed in the proof of Theorems 6.1 and 15.2 (even if $k \geq p$ in Theorems 6.1 and 15.2).

(iii) In Theorem 16.1, as in Theorems 6.1 and 15.2, we assume that the parameter space being considered is nonempty.

(iv) The results of Theorem 6.1 still hold if the moment bounds in $\mathcal{F}_{\mathrm{AR}}$, $\mathcal{F}$, and $\mathcal{F}_P$ are weakened very slightly by, for example, replacing $E_F\|g_i\|^{2+\gamma} \leq M$ in $\mathcal{F}_{\mathrm{AR}}$ by $E_F\|g_i\|^2 1(\|g_i\| > j) \leq \varepsilon_j$ for all integers $j \geq 1$ for some $\varepsilon_j > 0$ (that does not depend on F) for which $\varepsilon_j \to 0$ as $j \to \infty$. The latter conditions are weaker because, for any random variable X and constants $\gamma, j > 0$, $EX^2 1(|X| > j) \leq E|X|^{2+\gamma}/j^\gamma$. The latter conditions allow for the application of Lindeberg's triangular array central limit theorem for independent random variables, for example, see Billingsley (1979, Theorem 27.2, p. 310), in scenarios where the distribution F depends on n. For simplicity, we define the parameter spaces as is.

Analogously, the results in Theorems 6.1 and 15.2 still hold if the moment bounds in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$, $\mathcal{F}^{\mathrm{SR}}$, and $\mathcal{F}_P^{\mathrm{SR}}$ are weakened very slightly by, for example, replacing $E_F\|\varPi_{1F}^{-1/2}A_F'g_i\|^{2+\gamma} \leq M$ in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$ by $E_F\|\varPi_{1F}^{-1/2}A_F'g_i\|^2 1(\|\varPi_{1F}^{-1/2}A_F'g_i\| > j) \leq \varepsilon_j$ for all integers $j \geq 1$ for some $\varepsilon_j > 0$ (that does not depend on F) for which $\varepsilon_j \to 0$ as $j \to \infty$.

The following lemma shows that the critical value function $c_{k,p}(D, 1-\alpha)$ depends on D only through its singular values.

Lemma 16.2. *Let D be a $k \times p$ matrix with the singular value decomposition $D = CYB'$, 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 Y 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 $c_{k,p}(D, 1-\alpha) = c_{k,p}(Y, 1-\alpha)$.*

Comment. A consequence of Lemma 16.2 is that the critical value $c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha)$ of the CQLR test depends on $\widehat{D}_n^*(\theta_0)$ only through $\widehat{D}_n^*(\theta_0)'\widehat{D}_n^*(\theta_0)$ (because when $k \geq p$, the p singular values of $n^{1/2}\widehat{D}_n^*(\theta_0)$ equal the square roots of the eigenvalues of $n\widehat{D}_n^*(\theta_0)'\widehat{D}_n^*(\theta_0)$ and, when $k < p$, $c_{k,p}(D, 1-\alpha)$ is the $1-\alpha$ quantile of the χ_k^2 distribution which does not depend on D).

16.2 *Uniformity framework*

The proofs of Theorems 6.1, 15.2, and 16.1 use Corollary 2.1(c) in Andrews, Cheng, and Guggenberger (2019) (ACG), which provides general sufficient conditions for the correct asymptotic size and (uniform) asymptotic similarity of a sequence of tests.

Now we state Corollary 2.1(c) of ACG. Let $\{\phi_n : n \geq 1\}$ be a sequence of tests of some null hypothesis whose null distributions are indexed by a parameter λ with parameter space Λ. Let $\mathrm{RP}_n(\lambda)$ denote the null rejection probability of ϕ_n under λ. For a finite nonnegative integer J, let $\{h_n(\lambda) = (h_{1n}(\lambda), \ldots, h_{Jn}(\lambda))' \in R^J : n \geq 1\}$ be a sequence of functions on Λ. Define

$$\begin{aligned} H := \big\{h \in \big(R \cup \{\pm\infty\}\big)^J : h_{w_n}(\lambda_{w_n}) \to h \text{ for some subsequence } \{w_n\} \\ \text{of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\big\}. \end{aligned} \tag{16.2}$$

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

Proposition 16.3 (ACG, Corollary 2.1(c)). *Under Assumption* B*, *the tests* $\{\phi_n : n \geq 1\}$ *have asymptotic size* α *and are asymptotically similar* (*in a uniform sense*). *That is,* $\mathrm{AsySz} := \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} \mathrm{RP}_n(\lambda) = \alpha$ *and* $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} \mathrm{RP}_n(\lambda) = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} \mathrm{RP}_n(\lambda)$.

Comment. (i) By Comment 4 to Theorem 2.1 of ACG, Proposition 16.3 provides asymptotic size and similarity results for nominal $1-\alpha$ CSs, rather than tests, by defining λ as one would for a test, but having it depend also on the parameter that is restricted by the null hypothesis, by enlarging the parameter space Λ correspondingly (so it includes all possible values of the parameter that is restricted by the null hypothesis), and by replacing (a) ϕ_n by a CS based on a sample of size n, (b) α by $1-\alpha$, (c) $\mathrm{RP}_n(\lambda)$ by $\mathrm{CP}_n(\lambda)$, where $\mathrm{CP}_n(\lambda)$ denotes the coverage probability of the CS under λ when the sample size is n, and (d) the first $\limsup_{n\to\infty} \sup_{\lambda\in\Lambda}$ that appears by $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda}$. In the present case, where the null hypotheses are of the form $H_0 : \theta = \theta_0$ for some $\theta_0 \in \Theta$, to establish the asymptotic size of CSs, the parameter θ_0 is taken to be a subvector of λ and Λ is specified so that the value of this subvector ranges over Θ.

(ii) In the application of Proposition 16.3 to prove Theorems 6.1, 15.2, and 16.1, one takes Λ to be a one-to-one transformation of $\mathcal{F}_{\mathrm{AR}}$, $\mathcal{F}$, or $\mathcal{F}_P$ for tests, and one takes Λ to be a one-to-one transformation of $\mathcal{F}_{\Theta,\mathrm{AR}}$, $\mathcal{F}_\Theta$, or $\mathcal{F}_{\Theta,P}$ for CSs. With these changes, the proofs for tests and CSs are the same. In consequence, we provide explicit proofs for tests only and obtain the proofs for CSs by analogous applications of Proposition 16.3.

(iii) We prove the test results in Theorems 16.1 and 15.2 using Proposition 16.3 by verifying Assumption B* for a suitable choice of λ, $h_n(\lambda)$, and Λ. The verification of Assumption B* is quite easy for the AR test. It is given in Section 27.6. The verifications of Assumption B* for the CQLR and CQLR_P tests are much more difficult. In the remainder of this Section 16, we provide some key results that are used in doing so. (These results are used only for the CQLR and CQLR_P tests, not the AR test.) The complete verifications for the CQLR and CQLR_P tests are given in Section 27.

16.3 *General weight matrices* $\widehat{W}_n$ *and* $\widehat{U}_n$

As above, for notational simplicity, we suppress the dependence on θ_0 of many quantities, such as g_i, G_i, $u_{\theta i}$, B, and f_i, as well as the quantities V_F, R_F, Ξ_F, $\widetilde{V}_F$, and $\widetilde{R}_F$, that are introduced below. To provide asymptotic results for the CQLR and CQLR$_P$ tests simultaneously, we prove asymptotic results for a QLR test statistic and a conditioning statistic that depend on general random weight matrices $\widehat{W}_n \in R^{k\times k}$ and $\widehat{U}_n \in R^{p\times p}$. In particular, we consider statistics of the form $\widehat{W}_n\widehat{D}_n\widehat{U}_n$ and functions of this statistic, where $\widehat{D}_n$ is defined in (18). Let[39]

$$\begin{aligned} &\mathrm{QLR}_{\mathrm{WU},n} := \mathrm{AR}_n - \lambda_{\min}(n\widehat{Q}_{\mathrm{WU},n}), \quad \text{where} \\ &\widehat{Q}_{\mathrm{WU},n} := \big(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\big)'\big(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\big) \in R^{(p+1)\times(p+1)}. \end{aligned} \tag{16.3}$$

The definitions of the random weight matrices $\widehat{W}_n$ and $\widehat{U}_n$ depend upon the statistic that is of interest. They are taken to be of the form

$$\widehat{W}_n := W_1(\widehat{W}_{2n}) \in R^{k\times k} \quad \text{and} \quad \widehat{U}_n := U_1(\widehat{U}_{2n}) \in R^{p\times p}, \tag{16.4}$$

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

$$W_F := W_1(W_{2F}) \quad \text{and} \quad U_F := U_1(U_{2F}), \tag{16.5}$$

respectively.

Example 1. For the CQLR test,

$$\widehat{W}_n := \widehat{\Omega}_n^{-1/2} \quad \text{and} \quad \widehat{U}_n := \widehat{L}_n^{1/2} := \big((\theta_0, I_p)\big(\widehat{\Sigma}_n^{\varepsilon}\big)^{-1}(\theta_0, I_p)'\big)^{1/2}, \tag{16.6}$$

where $\widehat{\Omega}_n$ is defined in (8) and $\widehat{\Sigma}_n$ is defined in (20) and (21).

The population analogues of $\widehat{V}_n$ and $\widehat{R}_n$, defined in (19), are

$$\begin{aligned} &V_F := E_F(f_i - E_F f_i)(f_i - E_F f_i)' \in R^{(p+1)k\times(p+1)k} \quad \text{and} \\ &R_F := \big(B' \otimes I_k\big)V_F(B \otimes I_k) \in R^{(p+1)k\times(p+1)k}. \end{aligned} \tag{16.7}$$

[39]The definition of $\widehat{Q}_{WUn}$ in (16.3) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$, whereas (23) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$, which has the $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ vector as the first column rather than the last column. The ordering of the columns does not affect the value of the $\lambda_{\min}(\cdot)$ quantity. We use the order $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$ in (23) because it is consistent with the order in Moreira (2003) and Andrews, Moreira, and Stock (2006, 2008). We use the order $(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$ here because it has significant notational advantages in the proof of Theorem 16.6 below, which is given in Section 26.

In this case,

$$\begin{aligned} \widehat{W}_{2n} &:= \widehat{\Omega}_n, \qquad W_{2F} := \Omega_F := E_F g_i g_i', \qquad W_1(W_{2F}) := W_{2F}^{-1/2}, \\ U_1(U_{2F}) &:= \big((\theta_0, I_p)\big(\Sigma^{\varepsilon}(\Omega_F, R_F)\big)^{-1}(\theta_0, I_p)'\big)^{1/2}, \\ \widehat{U}_{2n} &:= (\widehat{\Omega}_n, \widehat{R}_n), \qquad U_{2F} := (\Omega_F, R_F), \quad \text{and} \\ \Sigma_{j\ell}(\Omega_F, R_F) &= \operatorname{tr}\big(R_{j\ell F}' \Omega_F^{-1}\big)/k \end{aligned} \tag{16.8}$$

for $j, \ell = 1, \dots, p+1$, where $\Sigma_{j\ell}(\Omega_F, R_F) \in R^{(p+1)\times(p+1)}$ denotes the (j, ℓ) element $\Sigma(\Omega_F, R_F)$, $\Sigma(\Omega_F, R_F)$ is defined to minimize $\|(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - R_F](I_{p+1} \otimes \Omega_F^{-1/2})\|$ over symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$ (analogously to the definition of $\widehat{\Sigma}_n$ in (20)), the last equality in (16.8) holds by the same argument as used to obtain (21), $\Sigma^{\varepsilon}(\Omega_F, R_F)$ is defined given $\Sigma(\Omega_F, R_F)$ by (22), and $R_{j\ell F}$ denotes the (j, ℓ) $k \times k$ submatrix of R_F.[40]

Example 2. For the CQLR$_P$ test, one takes $\widehat{W}_n$, $\widehat{W}_{2n}$, W_{2F}, $W_1(\cdot)$, and $U_1(\cdot)$ as in Example 1 and

$$\widehat{U}_n := \widetilde{L}_n^{1/2} := \big((\theta_0, I_p)\big(\widetilde{\Sigma}_n^{\varepsilon}\big)^{-1}(\theta_0, I_p)'\big)^{1/2}, \tag{16.9}$$

where $\widetilde{\Sigma}_n = \widetilde{\Sigma}_n(\theta_0)$ is defined just above (15.5) and $\widetilde{\Sigma}_n^{\varepsilon}$ is defined given $\widetilde{\Sigma}_n$ by (22).

The population analogues of $\widetilde{V}_n$ and $\widetilde{R}_n$, defined in (15.5), are

$$\begin{aligned} \widetilde{V}_F &:= E_F f_i f_i' - E_F\big((g_i, G_i)'\Xi_F \otimes Z_i Z_i'\big) - E_F\big(\Xi_F'(g_i, G_i) \otimes Z_i Z_i'\big) \\ &\quad + E_F\big(\Xi_F' Z_i Z_i' \Xi_F \otimes Z_i Z_i'\big) \\ &\in R^{(p+1)k\times(p+1)k} \quad \text{and} \\ \widetilde{R}_F &:= \big(B' \otimes I_k\big)\widetilde{V}_F(B \otimes I_k) \in R^{(p+1)k\times(p+1)k}, \quad \text{where} \\ \Xi_F &:= \big(E_F Z_i Z_i'\big)^{-1} E_F(g_i, G_i) \in R^{k\times(p+1)}, \qquad f_i := \big(g_i', \operatorname{vec}(G_i)'\big)' \in R^{(p+1)k}, \end{aligned} \tag{16.10}$$

and $B = B(\theta_0)$ is defined in (19).

For the CQLR$_P$ test,

$$\begin{aligned} \widehat{U}_{2n} &:= (\widehat{\Omega}_n, \widetilde{R}_n), \qquad U_{2F} := (\Omega_F, \widetilde{R}_F), \quad \text{and} \\ \Sigma_{j\ell}(\Omega_F, \widetilde{R}_F) &= \operatorname{tr}\big(\widetilde{R}_{j\ell F}' \Omega_F^{-1}\big)/k, \end{aligned} \tag{16.11}$$

for $j, \ell = 1, \dots, p+1$, where $\Sigma_{j\ell}(\Omega_F, \widetilde{R}_F) \in R^{(p+1)\times(p+1)}$ denotes the (j, ℓ) element $\Sigma(\Omega_F, \widetilde{R}_F)$, $\Sigma(\Omega_F, \widetilde{R}_F)$ is defined to minimize $\|(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - \widetilde{R}_F](I_{p+1} \otimes \Omega_F^{-1/2})\|$ over symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$ (analogously to the definition of $\widehat{\Sigma}_n(\theta)$ in (20)), the last equality in (16.11) holds by the same argument as used to obtain (21), $\Sigma^{\varepsilon}(\Omega_F, \widetilde{R}_F)$ is defined given $\Sigma(\Omega_F, \widetilde{R}_F)$ by (22), and $\widetilde{R}_{j\ell F}$ denotes the (j, ℓ) $k \times k$ submatrix of $\widetilde{R}_F$.

[40]Note that $W_1(W_{2F})$ and $U_1(U_{2F})$ in (16.8) define the functions $W_1(\cdot)$ and $U_1(\cdot)$ for any conformable arguments, such as $\widehat{W}_{2n}$ and $\widehat{U}_{2n}$, not just for W_{2F} and U_{2F}.

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

$$\mathcal{F}_{\mathrm{WU}} := \left\{F \in \mathcal{F} : \lambda_{\min}(W_F) \geq \delta_1, \lambda_{\min}(U_F) \geq \delta_1, \|W_F\| \leq M_1, \text{ and } \|U_F\| \leq M_1\right\} \tag{16.12}$$

for some constants $\delta_1 > 0$ and $M_1 < \infty$, where $\mathcal{F}$ is defined in (16.1).

For the CQLR test, which uses the definitions in (16.6)–(16.8), we show that $\mathcal{F} \subset \mathcal{F}_{\mathrm{WU}}$ for $\delta_1 > 0$ sufficiently small and $M_1 < \infty$ sufficiently large; see Lemma 27.4(a). Hence, uniform results over $\mathcal{F}_{\mathrm{WU}}$ for this test imply uniform results over $\mathcal{F}$.

For the CQLR$_P$ test, which uses the definitions in (16.9)–(16.11), we show that $\mathcal{F}_P \subset \mathcal{F}_{\mathrm{WU}}$ for $\delta_1 > 0$ sufficiently small and $M_1 < \infty$ sufficiently large, where $\mathcal{F}$ is defined in (16.1); see Lemma 27.4(b) in Section 27.1. Hence, uniform results over $\mathcal{F}_P \cap \mathcal{F}_{\mathrm{WU}}$ for arbitrary $\delta_1 > 0$ and $M_1 < \infty$ for this test imply uniform results over $\mathcal{F}_P$.

16.4 *Uniformity reparametrization*

To apply Proposition 16.3, we reparametrize the null distribution F to a vector λ. The vector λ is chosen such that for a subvector of λ convergence of a drifting subsequence of the subvector (after suitable renormalization) yields convergence in distribution of the test statistic and convergence in distribution of the critical value in the case of the CQLR tests. In this section, we define λ for the CQLR and CQLR$_P$ tests. The same definition is used for both tests. The (much simpler) definition of λ for the AR test is given in Section 27.6 below.

The vector λ depends on the following quantities. Let

$$\begin{aligned}&B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of}\\ &U_F'(E_F G_i)'W_F'W_F(E_F G_i)U_F\end{aligned} \tag{16.13}$$

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

$$\begin{aligned}&C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of}\\ &W_F(E_F G_i)U_F U_F'(E_F G_i)'W_F'.^{41}\end{aligned} \tag{16.14}$$

The corresponding eigenvalues are $(\kappa_{1F}, \ldots, \kappa_{kF}) \in R^k$. Let

$$(\tau_{1F}, \ldots, \tau_{\min\{k,p\}F}) \text{ denote the } \min\{k, p\} \text{ singular values of } W_F(E_F G_i)U_F, \tag{16.15}$$

which are nonnegative, ordered so that τ_{jF} is nonincreasing. (Some of these singular values may be zero.) As is well known, the squares of the $\min\{k, p\}$ singular values of a $k \times p$ matrix A equal the $\min\{k, p\}$ largest eigenvalues of $A'A$ and AA'. In consequence, $\kappa_{jF} = \tau_{jF}^2$ for $j = 1, \ldots, \min\{k, p\}$. In addition, $\kappa_{jF} = 0$ for $j = \min\{k, p\}+1, \ldots, \max\{k, p\}$.

[41] The matrices B_F and C_F are not uniquely defined. We let B_F denote one choice of the matrix of eigenvectors of $U_F'(E_F G_i)'W_F'W_F(E_F G_i)U_F$ and analogously for C_F.

Define the elements of λ to be[42,43]

$$
\begin{aligned}
&\lambda_{1,F} := (\tau_{1F}, \ldots, \tau_{\min\{k,p\}F})' \in R^{\min\{k,p\}},\\
&\lambda_{2,F} := B_F \in R^{p\times p},\\
&\lambda_{3,F} := C_F \in R^{k\times k},\\
&\lambda_{4,F} := E_F G_i \in R^{k\times p},\\
&\lambda_{5,F} := E_F \begin{pmatrix} g_i \\ \mathrm{vec}(G_i)\end{pmatrix}\begin{pmatrix} g_i \\ \mathrm{vec}(G_i)\end{pmatrix}' \in R^{(p+1)k\times(p+1)k},\\
&\lambda_{6,F} = (\lambda_{6,1F}, \ldots, \lambda_{6,(\min\{k,p\}-1)F})' := \left(\frac{\tau_{2F}}{\tau_{1F}}, \ldots, \frac{\tau_{\min\{k,p\}F}}{\tau_{(\min\{k,p\}-1)F}}\right)'\\
&\qquad \in [0,1]^{\min\{k,p\}-1}, \quad \text{where } 0/0 := 0,\\
&\lambda_{7,F} := W_{2F},\\
&\lambda_{8,F} := U_{2F},\\
&\lambda_{9,F} := F, \quad \text{and}\\
&\quad \lambda = \lambda_F := (\lambda_{1,F}, \ldots, \lambda_{9,F}).
\end{aligned} \tag{16.16}
$$

The dimensions of W_{2F} and U_{2F} depend on the choices of $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$. We let $\lambda_{5,gF}$ denote the upper left $k \times k$ submatrix of $\lambda_{5,F.}$ Thus, $\lambda_{5,gF} = E_F g_i g_i' = \Omega_F$. We consider two parameter spaces for λ: Λ_{WU} and $\Lambda_{\mathrm{WU},P}$, which correspond to $\mathcal{F}_{\mathrm{WU}}$ and $\mathcal{F}_{\mathrm{WU}} \cap \mathcal{F}_P$, respectively, where $\mathcal{F}_P$ and $\mathcal{F}_{\mathrm{WU}}$ are defined in (16.1) and (16.12), respectively. The space Λ_{WU} is used for the CQLR test. The space $\Lambda_{\mathrm{WU},P}$ is used for the CQLR$_P$ test.[44] The parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ and the function $h_n(\lambda)$ are defined by

$$
\begin{aligned}
\Lambda_{\mathrm{WU}} &:= \{\lambda : \lambda = (\lambda_{1,F}, \ldots, \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{\mathrm{WU}}\},\\
\Lambda_{\mathrm{WU},P} &:= \{\lambda : \lambda = (\lambda_{1,F}, \ldots, \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{\mathrm{WU}} \cap \mathcal{F}_P\}, \quad \text{and}\\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}).
\end{aligned} \tag{16.17}
$$

By the definition of $\mathcal{F}$, Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ index distributions that satisfy the null hypothesis $H_0 : \theta = \theta_0$. The dimension J of $h_n(\lambda)$ equals the number of elements in $(\lambda_{1,F}, \ldots, \lambda_{8,F})$. Redundant elements in $(\lambda_{1,F}, \ldots \lambda_{8,F})$, such as the redundant off-diagonal elements of the symmetric matrix $\lambda_{5,F}$, are not needed, but do not cause any problem.

We define λ and $h_n(\lambda)$ as in (16.16) and (16.17) because, as shown below, the asymptotic distributions of the test statistics under a sequence $\{F_n : n \geq 1\}$ for which $h_n(\lambda_{F_n}) \to$

[42] For simplicity, when writing $\lambda = (\lambda_{1,F}, \ldots, \lambda_{9,F})$, we allow the elements to be scalars, vectors, matrices, and distributions and likewise in similar expressions.

[43] If $p = 1$, no vector $\lambda_{6,F}$ appears in λ because $\lambda_{1,F}$ only contains a single element.

[44] Note that the parameter λ has different meanings for the CQLR and CQLR$_P$ tests because U_{2F} is different for the two tests.

$h \in H$ depend on the behavior of $\lim n^{1/2}\lambda_{1,F_n}$, as well as $\lim \lambda_{m,F_n}$ for $m = 2, \ldots, 8$. Note that $\lambda_{1,F}$ measures the strength of identification.

For notational convenience,

$$\{\lambda_{n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_n \in \Lambda_{\mathrm{WU}} : n \geq 1\} \text{ for which } h_n(\lambda_n) \to h \in H \tag{16.18}$$

for H defined in (16.2) with Λ equal to Λ_{WU}.[45] By the definitions of Λ_{WU} and $\mathcal{F}_{\mathrm{WU}}$, $\{\lambda_{n,h} : n \geq 1\}$ is a sequence of distributions that satisfies the null hypothesis $H_0 : \theta = \theta_0$.

We decompose h (defined by (16.2), (16.16), and (16.17)) analogously to the decomposition of the first eight components of λ: $h = (h_1, \ldots, h_8)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, \ldots, 8$. We further decompose the vector h_1 as $h_1 = (h_{1,1}, \ldots, h_{1,\min\{k,p\}})'$, where the elements of h_1 could equal ∞. We decompose h_6 as $h_6 = (h_{6,1}, \ldots, h_{6,\min\{k,p\}-1})'$. In addition, we let $h_{5,g}$ denote the upper left $k \times k$ submatrix of h_5. In consequence, under a sequence $\{\lambda_{n,h} : n \geq 1\}$, we have

$$\begin{aligned}
&n^{1/2}\tau_{jF_n} \to h_{1,j} \geq 0 \quad \forall j \leq \min\{k, p\}, \\
&\lambda_{m,F_n} \to h_m \quad \forall m = 2, \ldots, 8, \\
&\lambda_{5,gF_n} = \Omega_{F_n} = E_{F_n} g_i g_i' \to h_{5,g}, \quad \text{and} \\
&\lambda_{6,jF_n} \to h_{6,j} \quad \forall j = 1, \ldots, \min\{k, p\} - 1.
\end{aligned} \tag{16.19}$$

By the conditions in $\mathcal{F}$, defined in (16.1), $h_{5,g}$ is pd.

16.5 *Assumption WU*

We assume that the random weight matrices $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$ defined in (16.4) satisfy the following assumption that depends on a suitably chosen parameter space Λ_* ($\subset \Lambda_{\mathrm{WU}}$), such as Λ_{WU} or $\Lambda_{\mathrm{WU},P}$.

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

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

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

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

In Assumption WU and elsewhere below, "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$" means "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ for any $h \in H$," where H is defined in (16.2) with Λ equal to Λ_{WU}, and likewise with n in place of w_n.

Assumption WU for the parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ is verified in Lemma 27.4 in Section 27 below for the CQLR and CQLR_P tests, respectively.

[45] Analogously, for any subsequence $\{w_n : n \geq 1\}$, $\{\lambda_{w_n,h} : n \geq 1\}$ denotes a sequence $\{\lambda_{w_n} \in \Lambda : n \geq 1\}$ for which $h_{w_n}(\lambda_{w_n}) \to h \in H$.

16.6 *Asymptotic distributions*

This section provides the asymptotic distributions of QLR and QLR_P test statistics and corresponding conditioning statistics. These statistics are used in the proof of Theorem 16.1 to verify Assumption B* of Proposition 16.3.

For any $F \in \mathcal{F}$, define

$$\begin{aligned} \Phi_F^{\mathrm{vec}(G_i)} &:= \mathrm{Var}_F\big(\mathrm{vec}(G_i) - \big(E_F \mathrm{vec}(G_\ell) g_\ell'\big)\Omega_F^{-1} g_i\big) \quad \text{and} \\ \Phi_h^{\mathrm{vec}(G_i)} &:= \lim \Phi_{F_{w_n}}^{\mathrm{vec}(G_i)} \end{aligned} \tag{16.20}$$

whenever the limit exists, where the distributions $\{F_{w_n} : n \geq 1\}$ correspond to $\{\lambda_{w_n,h} : n \geq 1\}$ for any subsequence $\{w_n : n \geq 1\}$. The assumptions allow $\Phi_h^{\mathrm{vec}(G_i)}$ to be singular.

By the CLT and some straightforward calculations, the joint asymptotic distribution of $n^{1/2}(\widehat{g}_n', \mathrm{vec}(\widehat{D}_n - E_{F_n} G_i)')'$ under $\{\lambda_{n,h} : n \geq 1\}$ is given by

$$\begin{pmatrix} \overline{g}_h \\ \mathrm{vec}(\overline{D}_h) \end{pmatrix} \sim N\left(0^{(p+1)k}, \begin{pmatrix} h_{5,g} & 0^{k\times pk} \\ 0^{pk\times k} & \Phi_h^{\mathrm{vec}(G_i)} \end{pmatrix}\right), \tag{16.21}$$

where $\overline{g}_h \in R^k$ and $\overline{D}_h \in R^{k\times p}$ are independent by the definition of $\widehat{D}_n$; see Lemma 16.4 below.[46]

To determine the asymptotic distributions of the QLR_n and QLR_{Pn} statistics (defined in (23) and just below (15.6)) and the conditional critical value of the CQLR and CQLR_P tests (defined in (24), (25), and (15.7)), we need to determine the asymptotic distribution of $W_{F_n}\widehat{D}_n U_{F_n}$ without recentering by $E_{F_n} G_i$. To do so, we post-multiply $W_{F_n}\widehat{D}_n U_{F_n}$ first by B_{F_n} and then by a nonrandom diagonal matrix $S_n \in R^{p\times p}$ (which may depend on F_n and h). The matrix S_n rescales the columns of $W_{F_n}\widehat{D}_n U_{F_n} B_{F_n}$ to ensure that $n^{1/2} W_{F_n}\widehat{D}_n U_{F_n} B_{F_n} S_n$ converges in distribution to a (possibly) random matrix that is finite a.s. and not a.s. zero.

The following is an important definition for the scaling matrix S_n and asymptotic distributions given below. Consider a sequence $\{\lambda_{n,h} : n \geq 1\}$. Let $q = q_h$ ($\in \{0, \ldots, \min\{k, p\}\}$) be such that

$$h_{1,j} = \infty \quad \text{for } 1 \leq j \leq q_h \quad \text{and} \quad h_{1,j} < \infty \quad \text{for } q_h + 1 \leq j \leq \min\{k, p\}, \tag{16.22}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, \ldots, \min\{k, p\}$ by (16.19) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (16.18). This value q exists because $\{h_{1,j} : j \leq \min\{k, p\}\}$ are nonincreasing in j (since $\{\tau_{jF} : j \leq \min\{k, p\}\}$ are nonincreasing in j, as defined in (16.15)). Note that q is the number of singular values of $W_{F_n}(E_{F_n} G_i) U_{F_n}$ that diverge to infinity when multiplied by $n^{1/2}$. Heuristically, q is the

[46] If one eliminates the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}$ and one defines $\widehat{D}_n$ in (18) with $\widehat{\Omega}_n$ replaced by the eigenvalue-adjusted matrix $\widehat{\Omega}_n^{\varepsilon}$ for some $\varepsilon > 0$, then the asymptotic distribution in (16.21) still holds, but without the independence of $\overline{g}_h$ and $\overline{D}_h$. However, this independence is key. Without it, the conditioning argument that is used to establish the correct asymptotic size of the CQLR and CQLR_2 tests does not go through. Thus, we define $\widehat{D}_n$ in (18) using $\widehat{\Omega}_n$, not $\widehat{\Omega}_n^{\varepsilon}$.

maximum number of parameters, or one-to-one transformations of the parameters, that are strongly or semi-strongly identified. (That is, one could partition θ, or a one-to-one transformation of θ, into subvectors of dimension q and $p-q$ such that if the $p-q$ subvector was known, and hence, was no longer part of the parameter, then the q subvector would be strongly or semi-strongly identified in the sense used in this paper.)

Let

$$\begin{aligned} &S_n := \mathrm{Diag}\{(n^{1/2}\tau_{1F_n})^{-1}, \ldots, (n^{1/2}\tau_{qF_n})^{-1}, 1, \ldots, 1\} \in R^{p\times p} \quad \text{and} \\ &T_n := B_{F_n} S_n \in R^{p\times p}, \end{aligned} \tag{16.23}$$

where $q = q_h$ is defined in (16.22). Note that S_n is well-defined for n large, because $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \le q$.

The asymptotic distribution of $\widehat{D}_n$ after suitable rotations and rescaling, but without recentering (by subtracting $E_F G_i$), depends on the following quantities. We partition h_2 and h_3 and define $\overline{\Delta}_h$ as follows:

$$\begin{aligned} h_2 &= (h_{2,q}, h_{2,p-q}), \qquad h_3 = (h_{3,q}, h_{3,k-q}), \\ h^{\diamond}_{1,p-q} &:= \begin{bmatrix} 0^{q\times(p-q)} \\ \mathrm{Diag}\{h_{1,q+1}, \ldots, h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times(p-q)} \quad \text{if } k \ge p, \\ h^{\diamond}_{1,p-q} &:= \begin{bmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ \mathrm{Diag}\{h_{1,q+1}, \ldots, h_{1,k}\} & 0^{(k-q)\times(p-k)} \end{bmatrix} \in R^{k\times(p-q)} \quad \text{if } k < p, \\ \overline{\Delta}_h &= (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k\times p}, \qquad \overline{\Delta}_{h,q} := h_{3,q}, \\ \overline{\Delta}_{h,p-q} &:= h_3 h^{\diamond}_{1,p-q} + h_{71}\overline{D}_h h_{81} h_{2,p-q}, \\ h_{71} &:= W_1(h_7), \quad \text{and} \quad h_{81} := U_1(h_8), \end{aligned} \tag{16.24}$$

where $h_{2,q} \in R^{p\times q}$, $h_{2,p-q} \in R^{p\times(p-q)}$, $h_{3,q} \in R^{k\times q}$, $h_{3,k-q} \in R^{k\times(k-q)}$, $\overline{\Delta}_{h,q} \in R^{k\times q}$, $\overline{\Delta}_{h,p-q} \in R^{k\times(p-q)}$, $h_{71} \in R^{k\times k}$, and $h_{81} \in R^{p\times p}$.[47] Note that when Assumption WU holds $h_{71} = \lim W_{F_n} = \lim W_1(W_{2F_n})$ and $h_{81} = \lim U_{F_n} = \lim U_1(U_{2F_n})$ under $\{\lambda_{n,h} : n \ge 1\}$.

The following lemma allows for $k \ge p$ and $k < p$. For the case where $k \ge p$, it appears in the SM to AG1 as Lemma 10.3.

Lemma 16.4. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h} : n \ge 1\}$ *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* (16.21), (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (16.24), (c) $(\overline{D}_h, \overline{\Delta}_h)$ *and* $\overline{g}_h$ *are independent, and* (d) *under all subsequences*

[47] There is some abuse of notation here. For example, $h_{2,q}$ and $h_{2,p-q}$ denote different matrices even if $p-q$ happens to equal q.

$\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$*, the convergence result above and the results of parts* (a)–(c) *hold with n replaced with* w_n.

COMMENT. (i) Lemma 16.4(c) is a key property that leads to the correct asymptotic size of the CQLR and CQLR_P tests.

(ii) Lemma 10.3 in the SM to AG1 contains a part (part (d)), which does not appear in Lemma 16.4. It states that $\overline{\Delta}_h$ has full column rank a.s. under some additional conditions. For Kleibergen's (2005) LM statistic and Kleibergen's (2005) CLR statistics that employ it, which are considered in AG1, one needs the (possibly) random limit matrix of $n^{1/2}W_{F_n}\widehat{D}_nU_{F_n}B_{F_n}S_n$, namely, $\overline{\Delta}_h$, to have full column rank with probability one, in order to apply the continuous mapping theorem (CMT), which is used to determine the asymptotic distribution of the test statistics. To obtain this full column rank property, AG1 restricts the parameter space for the tests based on aforementioned statistics to be a subset $\mathcal{F}_0$ of $\mathcal{F}$, where $\mathcal{F}_0$ is defined in Section 3 of AG1. In contrast, the QLR_n and QLR_{Pn} statistics considered here do not depend on Kleibergen's LM statistic and do not require the asymptotic distribution of $n^{1/2}W_{F_n}\widehat{D}_nU_{F_n}B_{F_n}S_n$ to have full column rank a.s. In consequence, it is not necessary to restrict the parameter space from $\mathcal{F}$ to $\mathcal{F}_0$ when considering these statistics.

Let

$$\widehat{\kappa}_{jn} \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n, \quad \forall j = 1, \ldots, p, \tag{16.25}$$

ordered to be nonincreasing in j. The jth singular value of $n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n$ equals $\widehat{\kappa}_{jn}^{1/2}$ for $j = 1, \ldots, \min\{k, p\}$.

The following proposition, combined with Lemma 16.2, is used to determine the asymptotic behavior of the data-dependent conditional critical values of the CQLR and CQLR_P tests. The proposition is the same as Theorem 10.4(c)–(f) in the SM to AG1, except that it is extended to cover the case $k < p$, not just $k \geq p$. For brevity, the proof of the proposition given in Section 25 below just describes the changes needed to the proof of Theorem 10.4(c)–(f) in the SM to AG1 in order to cover the case $k < p$. The proof of Theorem 10.4(c)–(f) in the SM to AG1 is similar to, but simpler than, the proof of Theorem 16.6 below, which is given in Section 26.

PROPOSITION 16.5. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

(b) *the* (*ordered*) *vector of the smallest* $p - q$ *eigenvalues of* $n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n$, *that is,* $(\widehat{\kappa}_{(q+1)n}, \ldots, \widehat{\kappa}_{pn})'$, *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* 16.4, *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.

Comment. Proposition 16.5(a) and (b) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = \widehat{L}_n^{1/2}$ is used to determine the asymptotic behavior of the critical value function for the CQLR test, which depends on $n^{1/2}\widehat{D}_n^*$ defined in (23); see the proof of Theorem 27.1 in Section 27.2. Proposition 16.5(a) and (b) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = \widetilde{L}_n^{1/2}$ is used to determine the asymptotic behavior of the critical value function for the CQLR$_P$ test, which depends on $n^{1/2}\widetilde{D}_n^*$ defined in (15.6); see the proof of Theorem 27.1 in Section 27.2.

The next theorem provides the asymptotic distribution of the general QLR$_{\mathrm{WU},n}$ statistic defined in (16.3) and, as special cases, those of the QLR$_n$ and QLR$_{Pn}$ statistics.

Theorem 16.6. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

$$\mathrm{QLR}_{\mathrm{WU},n} \to_d \overline{g}_h' h_{5,g}^{-1} \overline{g}_h - \lambda_{\min}\big((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)\big)$$

and the convergence holds jointly with the convergence in Lemma 16.4 *and Proposition* 16.5. *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.

Comment. (i) Theorem 16.6 gives the asymptotic distributions of the QLR$_n$ and QLR$_{Pn}$ statistics (defined by (23) and (15.6), resp.) once it is verified that the choices of $(\widehat{W}_n, \widehat{U}_n)$ for these statistics satisfy Assumption WU for the parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$, respectively. The latter is done in Lemma 27.4 in Section 27.1.

(ii) When $q = p$, the parameter θ_0 is strongly or semi-strongly identified and Theorem 16.6 shows that the QLR$_{\mathrm{WU},n}$ statistic has a χ_p^2 asymptotic null distribution.

(iii) When $k = p$, Theorem 16.6 shows that the QLR$_{\mathrm{WU},n}$ statistic has a χ_k^2 asymptotic null distribution regardless of the strength of identification.

(iv) When $k < p$, θ is necessarily unidentified and Theorem 16.6 shows that the asymptotic null distribution of QLR$_{\mathrm{WU},n}$ is χ_k^2.

(v) The proof of Theorem 16.6 given in Section 26 also shows that the largest q eigenvalues of $n(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)'(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$ diverge to infinity in probability and the (ordered) vector of the smallest $p+1-q$ eigenvalues of this matrix converges in distribution to the (ordered) vector of the $p+1-q$ eigenvalues of $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)' h_{3,k-q} \times h_{3,k-q}'(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)$.

Propositions 16.3 and 16.5 and Theorem 16.6 are used to prove Theorem 16.1. The proof is given in Section 27 below. Note, however, that the proof is not a straightforward implication of these results. The proof also requires (i) determining the behavior of the conditional critical value function $c_{k,p}(D, 1-\alpha)$, defined in the paragraph containing (24), for sequences of nonrandom $k \times p$ matrices $\{D_n : n \geq 1\}$ whose singular values may converge or diverge to infinity at any rates, (ii) showing that the distribution function of the asymptotic distribution of the $QLR_{\text{WU},n}$ statistic, conditional on the asymptotic version of the conditioning statistic, is continuous and strictly increasing at its $1-\alpha$ quantile for all possible (k, p, q) values and all possible limits of the scaled population singular values $\{n^{1/2}\tau_{jF_n} : n \geq 1\}$ for $j = 1, \ldots, \min\{k, p\}$, and (iii) establishing that Assumption WU holds for the CQLR and CQLR$_P$ tests. These results are established in Lemmas 27.2, 27.3, and 27.4, respectively, in Section 27.

17. Singularity-robust tests

In this section, we prove the main Theorems 6.1 and 15.2 for the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests using Theorem 16.1 for the tests without the SR extension. These tests, defined in (14), (29), and (15.9), depend on the random variable $\widehat{r}_n(\theta)$ and random matrices $\widehat{A}_n(\theta)$ and $\widehat{A}_n^{\perp}(\theta)$, defined in (10) and (11). First, in the following lemma, we show that with probability that goes to one as $n \to \infty$ (wp $\to 1$), the SR test statistics and data-dependent critical values are the same as when the nonrandom and rescaled population quantities $r_F(\theta)$ and $\Pi_{1F}^{-1/2}(\theta)A_F(\theta)'$ are used to define these statistics, rather than $\widehat{r}_n(\theta)$ and $\widehat{A}_n(\theta)'$, where $r_F(\theta)$, $A_F(\theta)$, and $\Pi_{1F}(\theta)$ are defined as in (4) and (5). The lemma also shows that the extra rejection condition in (14), (29), and (15.9) fails to hold wp $\to 1$ under all sequences of null distributions.

In the following lemma, θ_{0n} is the true value that may vary with n (which is needed for the CS results) and col$(\cdot)$ denotes the column space of a matrix.

Lemma 17.1. *For any sequence* $\{(F_n, \theta_{0n}) \in \mathcal{F}_{\Theta,\text{AR}}^{\text{SR}} : n \geq 1\}$, (a) $\widehat{r}_n(\theta_{0n}) = r_{F_n}(\theta_{0n})$ wp $\to 1$, (b) $\text{col}(\widehat{A}_n(\theta_{0n})) = \text{col}(A_{F_n}(\theta_{0n}))$ wp $\to 1$, (c) *the statistics* SR-AR$_n(\theta_{0n})$, SR-QLR$_n(\theta_{0n})$, SR-QLR$_{Pn}(\theta_{0n})$, $c_{\widehat{r}_n(\theta_{0n}),p}(n^{1/2}\widehat{D}_{An}^{*}(\theta_{0n}), 1-\alpha)$, *and* $c_{\widehat{r}_n(\theta_{0n}),p}(n^{1/2}\widetilde{D}_{An}^{*}(\theta_{0n}), 1-\alpha)$ *are invariant* wp $\to 1$ *to the replacement of* $\widehat{r}_n(\theta_{0n})$ *and* $\widehat{A}_n(\theta_{0n})'$ *by* $r_{F_n}(\theta_{0n})$ *and* $\Pi_{1F_n}^{-1/2}(\theta_{0n}) \times A_{F_n}(\theta_{0n})'$, *respectively, and* (d) $\widehat{A}_n^{\perp}(\theta_{0n})'\widehat{g}_n(\theta_{0n}) = 0^{k-\widehat{r}_n(\theta_{0n})}$ wp $\to 1$, *where this equality is defined to hold when* $\widehat{r}_n(\theta_{0n}) = k$.

Comment. 1. We now provide an example that appears to be a counterexample to the claim that $\widehat{r}_n = r$ wp $\to 1$. We show that it is not a counterexample because the distributions considered violate the moment bound in $\mathcal{F}_{\text{AR}}^{\text{SR}}$ in (6). Suppose $k = 1$ and $g_i = 1, -1,$ and 0 with probabilities $p_n/2$, $p_n/2$, and $1 - p_n$, respectively, under F_n, where $p_n = c/n$ for some $0 < c < \infty$. Then $E_{F_n} g_i = 0$, as is required, and $\text{rk}(\Omega_{F_n}) = \text{rk}(E_{F_n} g_i^2) = \text{rk}(p_n) = 1$. We have $\widehat{\Omega}_n = 0$ if $g_i = 0\ \forall i \leq n$. The latter holds with probability $(1 - p_n)^n = (1 - c/n)^n \to e^{-c} > 0$ as $n \to \infty$. In consequence, $P_{F_n}(\text{rk}(\widehat{\Omega}_n) = \text{rk}(\Omega_{F_n})) = P_{F_n}(\text{rk}(\widehat{\Omega}_n) = 1) \leq 1 - P_{F_n}(g_i = 0\ \forall i \leq n) \to 1 - e^{-c} < 1$, which is inconsistent with the claim that $\widehat{r}_n = r$ wp $\to 1$. However, the distributions $\{F_n : n \geq 1\}$ in this example violate the moment bound $E_F\|\Pi_{1F}^{-1/2}A_F'g_i\|^{2+\gamma}\le M$ in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$, so there is no inconsistency with the claim. This holds because for these distributions $E_{F_n}\|\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\|^{2+\gamma}=E_{F_n}|\mathrm{Var}_{F_n}^{-1/2}(g_i)g_i|^{2+\gamma}=p_n^{-(2+\gamma)/2}E_{F_n}|g_i|=p_n^{-\gamma/2}\to\infty$ as $n\to\infty$, where the second equality uses $|g_i|$ equals 0 or 1 and the third equality uses $E_{F_n}|g_i|=p_n$.

2. The example in the previous comment is extreme. A simple version of a more typical example where singularity and near singularity may occur is the case where $W_i\sim$ iid $N(\theta,\Omega_F)$ for $\theta\in R^k$, $\Omega_F\in R^{k\times k}$, $g(W_i,\theta):=W_i-\theta$, Ω_F has spectral decomposition $A_F\Pi_FA_F'$, and some eigenvalues of Ω_F may be close to zero or equal to zero. In this case, $\Pi_F^{-1/2}A_F'g_i$ is a vector of independent standard normal random variables and the moment conditions in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$ and $\mathcal{F}^{\mathrm{SR}}$ hold immediately. In this case, the conditions in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$ and $\mathcal{F}^{\mathrm{SR}}$ are mild moment conditions that allow one to obtain asymptotic results without the normality assumption.

Proof of Lemma 17.1. For notational simplicity, we suppress the dependence of various quantities on θ_{0n}. By considering subsequences, it suffices to consider the case where $r_{F_n}=r$ for all $n\ge1$ for some $r\in\{0,1,\ldots,k\}$.

First, we establish part (a). We have $\widehat{r}_n\le r$ a.s. for all $n\ge1$ because 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. $[F_n]$, where a.s. $[F_n]$ means "with probability one under F_n." This completes the proof of part (a) when $r=0$. Hence, for the rest of the proof of part (a), we assume $r>0$.

We have $\widehat{r}_n:=\mathrm{rk}(\widehat{\Omega}_n)\ge\mathrm{rk}(\Pi_{1F_n}^{-1/2}A_{F_n}'\widehat{\Omega}_nA_{F_n}\Pi_{1F_n}^{-1/2})$ because $\widehat{\Omega}_n$ is $k\times k$, $A_{F_n}\Pi_{1F_n}^{-1/2}$ is $k\times r$, and $1\le r\le k$. In addition, we have

$$\begin{aligned}
&\Pi_{1F_n}^{-1/2}A_{F_n}'\widehat{\Omega}_nA_{F_n}\Pi_{1F_n}^{-1/2}\\
&\quad=n^{-1}\sum_{i=1}^n\big(\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\big)\big(\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\big)'\\
&\qquad-\left(n^{-1}\sum_{i=1}^n\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\right)\left(n^{-1}\sum_{i=1}^n\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\right)',\\
&E_{F_n}\big(\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\big)\big(\Pi_{1F_n}^{-1/2}A_{F_n}'g_i\big)'\\
&\quad=\Pi_{1F_n}^{-1/2}A_{F_n}'\Omega_{F_n}A_{F_n}\Pi_{1F_n}^{-1/2}\\
&\quad=\Pi_{1F_n}^{-1/2}A_{F_n}'A_{F_n}^{\Omega}\Pi_{F_n}A_{F_n}^{\Omega\prime}A_{F_n}\Pi_{1F_n}^{-1/2}=I_r,
\end{aligned}\tag{17.1}$$

and $E_{F_n}\Pi_{1F_n}^{-1/2}A_{F_n}'g_i=0^r$, where the second last equality in (17.1) holds by the spectral decomposition in (4) and the last equality in (17.1) holds by the definitions of A_F^{Ω}, A_F, and Π_{1F} in (4) and (5). By (17.1), the moment conditions in $\mathcal{F}^{\mathrm{SR}}$, and the weak law of large numbers for $L^{1+\gamma/2}$-bounded i.i.d. random variables for $\gamma>0$, we obtain

$\Pi_{1F_n}^{-1/2} A_{F_n}' \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2} \to_p I_r$. In consequence, $\mathrm{rk}(\Pi_{1F_n}^{-1/2} A_{F_n}' \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2}) \geq r$ wp $\to 1$, which concludes the proof that $\widehat{r}_n = r$ wp $\to 1$.[48]

Next, we prove part (b). Let $N(\cdot)$ denote the null space of a matrix. We have

$$\begin{aligned} \lambda \in N(\Omega_{F_n}) \quad &\Longrightarrow \quad \lambda' \Omega_{F_n} \lambda = 0 \quad \Longrightarrow \quad \mathrm{Var}_{F_n}(\lambda' g_i) = 0 \\ &\Longrightarrow \quad \lambda' g_i = 0 \quad \text{a.s. } [F_n] \quad \Longrightarrow \quad \widehat{\Omega}_n \lambda = 0 \quad \text{a.s. } [F_n] \\ &\Longrightarrow \quad \lambda \in N(\widehat{\Omega}_n) \quad \text{a.s. } [F_n]. \end{aligned} \tag{17.2}$$

That is, $N(\Omega_{F_n}) \subset N(\widehat{\Omega}_n)$ a.s. $[F_n]$. This and $\mathrm{rk}(\Omega_{F_n}) = \mathrm{rk}(\widehat{\Omega}_n)$ wp $\to 1$ imply that $N(\Omega_{F_n}) = N(\widehat{\Omega}_n)$ wp $\to 1$ (because if $N(\widehat{\Omega}_n)$ is strictly larger than $N(\Omega_{F_n})$ then the dimension and rank of $\widehat{\Omega}_n$ must exceed the dimension and rank of $N(\Omega_{F_n})$, which is a contradiction). In turn, $N(\Omega_{F_n}) = N(\widehat{\Omega}_n)$ wp $\to 1$ implies that $\mathrm{col}(\widehat{A}_n) = \mathrm{col}(A_{F_n})$ wp $\to 1$, which proves part (b).

To prove part (c), it suffices to consider the case where $r \geq 1$ because the test statistics and their critical values are all equal to zero by definition when $\widehat{r}_n = 0$ and $\widehat{r}_n = 0$ wp $\to 1$ when $r = 0$ by part (a). Part (b) of the lemma implies that there exists a random $r \times r$ nonsingular matrix $\widehat{M}_n$ such that

$$\widehat{A}_n = A_{F_n} \Pi_{1F_n}^{-1/2} \widehat{M}_n \quad \text{wp} \to 1, \tag{17.3}$$

because Π_{1F_n} is nonsingular (since it is a diagonal matrix with the positive eigenvalues of Ω_{F_n} on its diagonal by its definition following (5)). Equation (17.3) and $\widehat{r}_n = r$ wp $\to 1$ imply that the statistics SR-AR$_n$, SR-QLR$_n$, SR-QLR$_{Pn}$, $c_{\widehat{r}_n,p}(n^{1/2}\widehat{D}_{An}^*, 1-\alpha)$, and $c_{\widehat{r}_n,p}(n^{1/2}\widetilde{D}_{An}^*, 1-\alpha)$ are invariant wp $\to 1$ to the replacement of $\widehat{r}_n$ and $\widehat{A}_n'$ by r and $\widehat{M}_n' \Pi_{1F_n}^{-1/2} A_{F_n}'$, respectively. Now we apply the invariance results of Lemmas 5.1 and 15.1 with (k, g_i, G_i) replaced by $(r, \Pi_{1F_n}^{-1/2} A_{F_n}' g_i, \Pi_{1F_n}^{-1/2} A_{F_n}' G_i)$ and with M equal to $\widehat{M}_n'$. These results imply that the previous five statistics when based on r and $\Pi_{1F_n}^{-1/2} A_{F_n}' g_i$ are invariant to the multiplication of the moments $\Pi_{1F_n}^{-1/2} A_{F_n}' g_i$ by the nonsingular matrix $\widehat{M}_n'$. Thus, these five statistics, defined as in Sections 5.2 and 15, are invariant wp $\to 1$ to the replacement of $\widehat{r}_n$ and $\widehat{A}_n'$ by r and $\Pi_{1F_n}^{-1/2} A_{F_n}'$, respectively.

Lastly, we prove part (d). The equality $(\widehat{A}_n^\perp)' \widehat{g}_n = 0^{k-\widehat{r}_n}$ holds by definition when $\widehat{r}_n = k$ (see the statement of Lemma 17.1(d)) and $\widehat{r}_n = r$ wp $\to 1$. Hence, it suffices to consider the case where $r \in \{0, \ldots, k-1\}$. For all $n \geq 1$, we have $E_{F_n}(A_{F_n}^\perp)' \widehat{g}_n = 0^{k-r}$ and

$$n\mathrm{Var}_{F_n}\big((A_{F_n}^\perp)' \widehat{g}_n\big) = (A_{F_n}^\perp)' \Omega_{F_n} A_{F_n}^\perp = (A_{F_n}^\perp)' A_{F_n}^{\Omega} \Pi_{F_n} (A_{F_n}^{\Omega})' A_{F_n}^\perp = 0^{(k-r)\times(k-r)}, \tag{17.4}$$

[48]We now provide an example that appears to be a counterexample to the claim that $\widehat{r}_n = r$ wp $\to 1$. We show that it is not a counterexample because the distributions considered violate the moment bound in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$. Suppose $k = 1$ and $g_i = 1, -1$, and 0 with probabilities $p_n/2$, $p_n/2$, and $1 - p_n$, respectively, under F_n, where $p_n = c/n$ for some $0 < c < \infty$. Then $E_{F_n} g_i = 0$, as is required, and $\mathrm{rk}(\Omega_{F_n}) = \mathrm{rk}(E_{F_n} g_i^2) = \mathrm{rk}(p_n) = 1$. We have $\widehat{\Omega}_n = 0$ if $g_i = 0\ \forall i \leq n$. The latter holds with probability $(1-p_n)^n = (1-c/n)^n \to e^{-c} > 0$ as $n \to \infty$. In consequence, $P_{F_n}(\mathrm{rk}(\widehat{\Omega}_n) = \mathrm{rk}(\Omega_{F_n})) = P_{F_n}(\mathrm{rk}(\widehat{\Omega}_n) = 1) \leq 1 - P_{F_n}(g_i = 0\ \forall i \leq n) \to 1 - e^{-c} < 1$, which is inconsistent with the claim that $\widehat{r}_n = r$ wp $\to 1$. However, the distributions $\{F_n : n \geq 1\}$ in this example violate the moment bound $E_F \|\Pi_{1F}^{-1/2} A_F' g_i\|^{2+\gamma} \leq M$ in $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$, so there is no inconsistency with the claim. This holds because for these distributions $E_{F_n} \|\Pi_{1F_n}^{-1/2} A_{F_n}' g_i\|^{2+\gamma} = E_{F_n} |\mathrm{Var}_{F_n}^{-1/2}(g_i) g_i|^{2+\gamma} = p_n^{-(2+\gamma)/2} E_{F_n}|g_i| = p_n^{-\gamma/2} \to \infty$ as $n \to \infty$, where the second equality uses $|g_i|$ equals 0 or 1 and the third equality uses $E_{F_n}|g_i| = p_n$.

where the second equality uses the spectral decomposition in (4) and the last equality uses $A_n^{\Omega} = [A_F, A_F^{\perp}]$, see (5). In consequence, $(A_{F_n}^{\perp})'\widehat{g}_n = 0^{k-r}$ a.s. This and and the result of part (b) that $\text{col}(\widehat{A}_n^{\perp}) = \text{col}(A_{F_n}^{\perp})$ wp $\to 1$ establish part (d). $\square$

Given Lemma 17.1(d), the extra rejection conditions in the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests and CSs (i.e., the second conditions in (14), (16), (29), (15.9), and in the SR-CQLR and SR-CQLR$_P$ CS definitions following (29) and (15.9)) can be ignored when computing the asymptotic size properties of these tests and CSs (because the condition fails to hold for each test wp $\to 1$ under any sequence of null hypothesis values for any sequence of distributions in the null hypotheses, and the condition holds for each CS wp $\to 1$ under any sequence of true values θ_{0n} for any sequence of distributions for which the moment conditions hold at θ_{0n}).

Given Lemma 17.1(c), the asymptotic size properties of the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests and CSs can be determined by the analogous tests and CSs that are based on $r_{F_n}(\theta_0)$ and $\Pi_{1F_n}^{-1/2}(\theta_0)A_{F_n}(\theta_0)'$ (for fixed θ_0 with tests and for any $\theta_0 \in \Theta$ with CSs). For the tests, we do so by partitioning $\mathcal{F}_{\text{AR}}^{\text{SR}}$, $\mathcal{F}^{\text{SR}}$, and $\mathcal{F}_P^{\text{SR}}$ into k sets based on the value of $\text{rk}(\Omega_F(\theta_0))$ and establishing the correct asymptotic size and asymptotic similarity of the analogous tests separately for each parameter space. That is, we write $\mathcal{F}_{\text{AR}}^{\text{SR}} = \bigcup_{r=0}^{k} \mathcal{F}_{\text{AR}[r]}^{\text{SR}}$, where $\mathcal{F}_{\text{AR}[r]}^{\text{SR}} := \{F \in \mathcal{F}_{\text{AR}}^{\text{SR}} : \text{rk}(\Omega_F(\theta_0)) = r\}$, and establish the desired results for $\mathcal{F}_{\text{AR}[r]}^{\text{SR}}$ separately for each r. Analogously, we write $\mathcal{F}^{\text{SR}} = \bigcup_{r=0}^{k} \mathcal{F}_{[r]}^{\text{SR}}$ and $\mathcal{F}_P^{\text{SR}} = \bigcup_{r=0}^{k} \mathcal{F}_{P[r]}^{\text{SR}}$, where $\mathcal{F}_{[r]}^{\text{SR}} := \mathcal{F}_{\text{AR}[r]}^{\text{SR}} \cap \mathcal{F}^{\text{SR}}$ and $\mathcal{F}_{P[r]}^{\text{SR}} := \mathcal{F}_{\text{AR}[r]}^{\text{SR}} \cap \mathcal{F}_P^{\text{SR}}$. Note that we do not need to consider the parameter space $\mathcal{F}_{\text{AR}[r]}^{\text{SR}}$ for $r = 0$ for the SR-AR test when determining the asymptotic size of the SR-AR test because the test fails to reject H_0 wp $\to 1$ based on the first condition in (14) when $r = 0$ (since the test statistic and critical value equal zero by definition when $\widehat{r}_n = 0$ and $\widehat{r}_n = r = 0$ wp $\to 1$ by Lemma 17.1(a)). In addition, we do not need to consider the parameter space $\mathcal{F}_{\text{AR}[r]}^{\text{SR}}$ for $r = 0$ for the SR-AR test when determining the asymptotic similarity of the test because such distributions are excluded from the parameter space $\mathcal{F}_{\text{AR}}^{\text{SR}}$ by the statement of Theorem 6.1. Analogous arguments regarding the parameter spaces corresponding to $r = 0$ apply to the other tests and CSs. Hence, from here on, we assume $r \in \{1, \dots, k\}$.

For given $r = \text{rk}(\Omega_F(\theta_0))$, the moment conditions and Jacobian are

$$g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i \quad \text{and} \quad G_{Fi}^* := \Pi_{1F}^{-1/2} A_F' G_i, \tag{17.5}$$

where $A_F \in R^{k \times r}$, $\Pi_{1F} \in R^{r \times r}$, and dependence on θ_0 is suppressed for notational simplicity. Given the conditions in $\mathcal{F}^{\text{SR}}$, we have

$$\begin{aligned} E_F \|g_{Fi}^*\|^{2+\gamma} &= E_F \|\Pi_{1F}^{-1/2} A_F' g_i\|^{2+\gamma} \leq M, \\ E_F \|\text{vec}(G_{Fi}^*)\|^{2+\gamma} &= E_F \|\text{vec}(\Pi_{1F}^{-1/2} A_F' G_i)\|^{2+\gamma} \leq M, \\ \lambda_{\min}(E_F g_{Fi}^* g_{Fi}^{*\prime}) &= \lambda_{\min}(\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2}) = \lambda_{\min}(I_r) = 1, \end{aligned} \tag{17.6}$$

and $E_F g_{Fi}^* = 0^r$, where the second equality in the third line of (17.6) holds by the spectral decomposition in (4) and the partition $A_F^{\Omega} = [A_F, A_F^{\perp}]$ in (5). Thus, $F \in \mathcal{F}_{[r]}^{\text{SR}}$ implies that $F \in \mathcal{F}$ with $\delta \leq 1$, when $\mathcal{F}$ is defined with (g^*_{Fi}, G^*_{Fi}) in place of (g_i, G_i), where the definition of $\mathcal{F}$ in (16.1) is extended to allow g_i and G_i to depend on F. Now we apply Theorem 16.1 with (g^*_{Fi}, G^*_{Fi}) and r in place of (g_i, G_i) and k and with $\delta \leq 1$, to obtain the correct asymptotic size and asymptotic similarity of the SR-CQLR test for the parameter space $\mathcal{F}^{\mathrm{SR}}_{[r]}$ for $r = 1, \ldots, k$. This requires that Theorem 16.1 holds for $k < p$, which it does. The fact that g^*_{Fi} and G^*_{Fi} depend on F, whereas g_i and G_i do not, does not cause a problem, because the proof of Theorem 16.1 goes through as is if g_i and G_i depend on F. This establishes the results of Theorem 6.1 for the SR-CQLR test. The proof for the SR-CQLR CS is essentially the same, but with θ_0 taking any value in Θ and with $\mathcal{F}^{\mathrm{SR}}_{\Theta}$ and $\mathcal{F}_{\Theta}$, defined in (7) and just below (16.1), in place of $\mathcal{F}^{\mathrm{SR}}$ and $\mathcal{F}$, respectively.

The proof for the SR-AR test and CS is the same as that for the SR-CQLR test and CS, but with $\mathrm{vec}(G^*_{Fi})$ deleted in (17.6) and with the subscript AR added to the parameter spaces that appear.

Next, we consider the SR-CQLR$_P$ test. When the moment functions satisfy (15.1), that is, $g_i = u_i Z_i$, we define $Z^*_{Fi} := \Pi_{1F}^{-1/2} A'_F Z_i$, $g^*_{Fi} = u_i Z^*_{Fi}$, and $G^*_{Fi} = Z^*_{Fi} u'_{\theta i}$, where $u_{\theta i}$ is defined in (15.2) and the dependence of various quantities on θ_0 is suppressed. In this case, by the conditions in $\mathcal{F}^{\mathrm{SR}}_P$, the IVs Z^*_{Fi} satisfy $E_F\|Z^*_{Fi}\|^{4+\gamma} = E_F\|\Pi_{1F}^{-1/2} A'_F Z_i\|^{4+\gamma} \leq M$ and $E_F\|u^*_i\|^{2+\gamma} \leq M$, where $u^*_i := (u_i, u'_{\theta i})'$. Next, we show that $\lambda_{\min}(E_F Z^*_{Fi} Z^{*\prime}_{Fi})$ is bounded away from zero for $F \in \mathcal{F}^{\mathrm{SR}}_{P[r]}$. We have

$$
\begin{aligned}
&\lambda_{\min}\big(E_F Z^*_{Fi} Z^{*\prime}_{Fi}\big) \\
&\quad = \lambda_{\min}\big(E_F \Pi_{1F}^{-1/2} A'_F Z_i Z'_i A_F \Pi_{1F}^{-1/2}\big) \\
&\quad = \inf_{\lambda \in R^r : \|\lambda\| = 1} \big[E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 1\big(u_i^2 \leq c\big) + E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 1\big(u_i^2 > c\big)\big] \\
&\quad \geq \inf_{\lambda \in R^r : \|\lambda\| = 1} \big[c^{-1} E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 u_i^2 1\big(u_i^2 \leq c\big)\big] \\
&\quad = c^{-1} \inf_{\lambda \in R^r : \|\lambda\| = 1} \big[E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 u_i^2 - E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 u_i^2 1\big(u_i^2 > c\big)\big] \\
&\quad \geq c^{-1}\Big[\lambda_{\min}\big(\Pi_{1F}^{-1/2} A'_F \Omega_F A_F \Pi_{1F}^{-1/2}\big) - \sup_{\lambda \in R^r : \|\lambda\| = 1} E_F\big(\lambda' \Pi_{1F}^{-1/2} A'_F Z_i\big)^2 u_i^2 1\big(u_i^2 > c\big)\Big] \\
&\quad \geq c^{-1}\big[1 - E_F\big\|\Pi_{1F}^{-1/2} A'_F Z_i\big\|^2 u_i^2 1\big(u_i^2 > c\big)\big] \\
&\quad \geq 1/(2c),
\end{aligned}
\tag{17.7}
$$

where the second inequality uses $g_i = Z_i u_i$ and $\Omega_F := E_F g_i g'_i$, the third inequality holds by $\Pi_{1F}^{-1/2} A'_F \Omega_F A_F \Pi_{1F}^{-1/2} = I_r$ (using (4) and (5)) and by the Cauchy–Bunyakovsky–Schwarz inequality applied to $\lambda' \Pi_{1F}^{-1/2} A'_F Z_i$, and the last inequality holds by the condition $E_F\|\Pi_{1F}^{-1/2} A'_F Z_i\|^2 u_i^2 \times 1(u_i^2 > c) \leq 1/2$ in $\mathcal{F}^{\mathrm{SR}}_P$.

The moment bounds above and (17.7) establish that $F \in \mathcal{F}^{\mathrm{SR}}_{P[r]}$ implies that $F \in \mathcal{F}_P$ for $\delta \leq \min\{1, 1/(2c)\}$, when $\mathcal{F}_P$ is defined with (g^*_{Fi}, G^*_{Fi}) in place of (g_i, G_i), where the definition of $\mathcal{F}_P$ in (16.1) is taken to allow g_i and G_i to depend on F.[49] Now we apply

[49] We require $\delta \leq \min\{1, 1/(2c)\}$, rather than $\delta \leq 1/(2c)$, because $\lambda_{\min}(E_F g^*_{Fi} g^{*\prime}_{Fi}) = 1$ by (17.6) and $\mathcal{F}$ ($\subset \mathcal{F}_{\mathrm{AR}}$) requires $\lambda_{\min}(E_F g^*_{Fi} g^{*\prime}_{Fi}) \geq \delta$.

Theorem 16.1 with (g^*_{Fi}, G^*_{Fi}) and r in place of (g_i, G_i) and k and $\delta \le \min\{1, 1/(2c)\}$ to obtain the correct asymptotic size and asymptotic similarity of the CQLR$_P$ test based on (g^*_{Fi}, G^*_{Fi}) and r for the parameter space $\mathcal{F}^{SR}_{P[r]}$ for $r = 1, \ldots, k$. As noted above, the dependence of g^*_{Fi} and G^*_{Fi} on F does not cause a problem in the application of Theorem 16.1. This establishes the results of Theorem 15.2 for the SR-CQLR$_P$ test by the argument given above.[50] The proof for the SR-CQLR$_P$ CS is essentially the same, but with θ_0 taking any value in Θ and with $\mathcal{F}^{SR}_{\Theta,P}$ and $\mathcal{F}_{\Theta,2}$, defined in (7) and just below (16.1), in place of $\mathcal{F}^{SR}_P$ and $\mathcal{F}_P$, respectively.

This completes the proof of Theorems 6.1 and 15.2 given Theorem 16.1.

18. Time series observations

In this section, we define the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests for observations that are strictly stationary strong mixing. We also generalize the asymptotic size results of Theorems 6.1 and 15.2 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 $\{W_i : i = \ldots, 0, 1, \ldots\}$.[51]

We define

$$V_{F,n}(\theta) := \mathrm{Var}_F\left(n^{-1/2}\sum_{i=1}^n \begin{pmatrix} g_i(\theta) \\ \mathrm{vec}(G_i(\theta)) \end{pmatrix}\right),$$
$$\Omega_{F,n}(\theta) := \mathrm{Var}_F\left(n^{-1/2}\sum_{i=1}^n g_i(\theta)\right), \quad \text{and} \quad r_{F,n}(\theta) := \mathrm{rk}(\Omega_{F,n}(\theta)). \tag{18.1}$$

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), (5), and the paragraph following (5), 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

$$\begin{aligned}\mathcal{F}^{SR}_{TS,AR} := \Big\{&F : \{W_i : i = \ldots, 0, 1, \ldots\} \text{ are stationary and strong mixing under } F \text{ with} \\ &\text{strong mixing numbers } \{\alpha_F(m) : m \ge 1\} \text{ that satisfy } \alpha_F(m) \le Cm^{-d}, \\ &E_F g_i = 0^k, \text{ and } \sup_{n\ge 1} E_F \big\|\Pi^{-1/2}_{1F,n} A'_{F,n} g_i\big\|^{2+\gamma} \le M\Big\}\end{aligned} \tag{18.2}$$

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 CSs, we use the corresponding parameter space

[50] The fact that Z^*_{Fi} depends on θ_0 through $\Pi^{-1/2}_{1F}(\theta_0)A_F(\theta_0)'$ and that $G^*_{Fi}(\theta_0) \ne (\partial/\partial\theta')g^*_{Fi}(\theta_0)$ (because $(\partial/\partial\theta')Z^*_{Fi}$ is ignored in the specification of $G^*_{Fi}(\theta_0)$) does not affect the application of Theorem 16.1. The reason is that the proof of this theorem goes through even if Z_i depends on θ_0 and for any $G_i(\theta_0)$ that satisfies the conditions in $\mathcal{F}_P$, not just for $G_i(\theta_0) := (\partial/\partial\theta')g_i(\theta_0)$.

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

$\mathcal{F}^{SR}_{TS,\Theta,AR} := \{(F, \theta_0) : F \in \mathcal{F}^{SR}_{TS,AR}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}^{SR}_{TS,AR}(\theta_0)$ denotes $\mathcal{F}^{SR}_{TS,AR}$ with its dependence on θ_0 made explicit. The moment conditions in $\mathcal{F}^{SR}_{TS,AR}$ are placed on the normalized moment functions $\Pi^{-1/2}_{1F,n} A'_{F,n} g_i$ that satisfy $\mathrm{Var}_F(n^{-1/2} \sum_{i=1}^n \Pi^{-1/2}_{1F,n} A'_{F,n} g_i) = I_k$ for all $n \geq 1$.

For the SR-CQLR and SR-CQLR$_P$ tests, we use the null parameter spaces $\mathcal{F}^{SR}_{TS}$ and $\mathcal{F}^{SR}_{TS,P}$, respectively, which are defined as $\mathcal{F}^{SR}$ and $\mathcal{F}^{SR}_P$ are defined in (6) and (15.3), but with (i) $\mathcal{F}^{SR}_{TS,AR}$ in place of $\mathcal{F}^{SR}_{AR}$, (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 $\mathcal{F}^{SR}$ and $\mathcal{F}^{SR}_P$ that depend on $A_{F,n}$ and $\Pi_{1F,n}$. For SR-CQLR and SR-CQLR$_P$ CSs, we use the parameter spaces $\mathcal{F}^{SR}_{TS,\Theta}$ and $\mathcal{F}^{SR}_{TS,\Theta,P}$, respectively, which are defined as $\mathcal{F}^{SR}_{TS,\Theta,AR}$ is defined, but with $\mathcal{F}^{SR}_{TS}(\theta_0)$ and $\mathcal{F}^{SR}_{TS,P}(\theta_0)$ in place of $\mathcal{F}^{SR}_{TS,AR}(\theta_0)$, where $\mathcal{F}^{SR}_{TS}(\theta_0)$ and $\mathcal{F}^{SR}_{TS,P}(\theta_0)$ denote $\mathcal{F}^{SR}_{TS}$ and $\mathcal{F}^{SR}_{TS,P}$ with their dependence on θ_0 made explicit.

The SR-CQLR and SR-CQLR$_P$ 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, and SR-CQLR$_P$ tests, these estimators are heteroskedasticity and autocorrelation consistent (HAC) variance matrix estimators based on $\{g_i - \widehat{g}_n : i \leq n\}$, $\{f_i - \widehat{f}_n : i \leq n\}$ (defined in (19)), and $\{(u^*_i - \widehat{u}^*_{in}) \otimes Z_i : i \leq n\}$ (defined in (15.5)), respectively. There are a number of HAC estimators available in the literature, for example, 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 $r = \{1, \ldots, k\}$. 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:

$$\widehat{V}_n = \begin{bmatrix} \widehat{\Omega}_n & \widehat{\Gamma}'_{1n} & \cdots & \widehat{\Gamma}'_{pn} \\ \widehat{\Gamma}_{1n} & \widehat{V}_{G_{11}n} & \cdots & \widehat{V}'_{G_{p1}n} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\Gamma}_{pn} & \widehat{V}_{G_{p1}n} & \cdots & \widehat{V}_{G_{pp}n} \end{bmatrix}. \tag{18.3}$$

We define $\widehat{r}_n$ $(= \widehat{r}_n(\theta_0))$ and $\widehat{A}_n$ $(= \widehat{A}_n(\theta_0))$ as in (10) and (11) with $\theta = \theta_0$, but with $\widehat{\Omega}_n$ defined in (18.3), rather than in (8).

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 test and CS, respectively.

Assumption SR-V. (a) *$[I_{p+1} \otimes (\Pi^{-1/2}_{1F_n,n}(\theta_0) A'_{F_n,n}(\theta_0))][\widehat{V}_n(\theta_0) - V_{F_n,n}(\theta_0)][I_{p+1} \otimes (A_{F_n,n}(\theta_0)\Pi^{-1/2}_{1F_n,n}(\theta_0))] \to_p 0^{(p+1)k\times(p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}^{SR}_{TS} : n \geq 1\}$ 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 $r \in \{1, \ldots, k\}$.*

(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* $F \in \mathcal{F}_{\mathrm{TS}}^{\mathrm{SR}}$.

For SR-CQLR CSs, 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-CS. (a) $[I_{p+1} \otimes (\Pi_{1F_n,n}^{-1/2}(\theta_{0n})A_{F_n,n}'(\theta_{0n}))][\widehat{V}_n(\theta_{0n}) - V_{F_n,n}(\theta_{0n})] \times [I_{p+1} \otimes (A_{F_n,n}(\theta_{0n})\Pi_{1F_n,n}^{-1/2}(\theta_{0n}))] \to_p 0^{(p+1)k\times(p+1)k}$ *under* $\{F_n : n \geq 1\}$ *for any sequence* $\{(F_n, \theta_{0n}) \in \mathcal{F}_{\mathrm{TS},\Theta}^{\mathrm{SR}} : n \geq 1\}$ *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* $r \in \{1, \ldots, k\}$.

(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* $(F, \theta_0) \in \mathcal{F}_{\mathrm{TS},\Theta}^{\mathrm{SR}}$.

Assumptions SR-V(a) and SR-V-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 $\Pi_{1F_n,n}^{-1/2}(\theta_{0n}) \times A_{F_n,n}'(\theta_{0n})G_i(\theta_{0n})$, resp.) to be consistent. This can be verified using standard methods. For typical HAC estimators, equivariance and Assumptions SR-V(c) and SR-V-CS(c) can be shown easily.

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

For the SR-AR test and CS, we use Assumptions SR-Ω and SR-Ω-CS, which are defined as Assumptions SR-V and SR-V-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 $\{F_n \in \mathcal{F}_{\mathrm{TS,AR}}^{\mathrm{SR}} : n \geq 1\}$ 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 $r \in \{1, \ldots, k\}$, (ii) Assumption SR-Ω-CS(a) being as in (i), but with θ_{0n} and $\mathcal{F}_{\mathrm{TS},\Theta,\mathrm{AR}}^{\mathrm{SR}}$ in place of θ_0 and $\mathcal{F}_{\mathrm{TS,AR}}^{\mathrm{SR}}$, (iii) $\widehat{\Omega}_n(\theta_0)$ in place of $\widehat{V}_n(\theta_0)$ in part (b) of each assumption, and (iv) $\mathcal{F}_{\mathrm{TS,AR}}^{\mathrm{SR}}$ in place of $\mathcal{F}_{\mathrm{TS}}^{\mathrm{SR}}$ in part (c) of each assumption.

Now we define the SR-AR, SR-CQLR, and SR-CQLR$_P$ tests in the time series context. The definitions are the same as in the i.i.d. context given in Sections 4, 5, and 15 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 (14) are defined as in (10) and (11), but with $\widehat{\Omega}_n$ defined to satisfy Assumption SR-Ω, rather than being defined in (8). The SR-AR statistic is defined as in Section 4, 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 (10) and (11). 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 statistic is defined as in Section 5, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined to satisfy Assumption SR-V and (18.3) based on $\{f_i - \widehat{f}_n : i \leq n\}$, in place of $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined in (19) and (8), respectively. This affects the definitions of $\widehat{R}_n$, $\widehat{\Sigma}_n$, $\widehat{L}_n$, $\widehat{D}_n^*$, $\widehat{r}_n$, $\widehat{A}_n$, and SR-AR$_n$ (which appears in (23)). Given the previous changes, the definition of the SR-CQLR critical value is unchanged.

In the time series case, the SR-CQLR$_P$ statistic is defined as in Section 15, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined to satisfy Assumption SR-V$_P$ and (18.3) based on $\{(u_i^* - \widehat{u}_{in}^*) \otimes Z_i : i \leq n\}$, rather than in (15.5) and (8), respectively. In turn, this affects the definitions of $\widetilde{R}_n$, $\widetilde{\Sigma}_n$, $\widetilde{L}_n$, $\widetilde{D}_n^*$, $\widetilde{Q}_n$, $\widehat{r}_n$, $\widehat{A}_n$, and SR-AR$_n$. Given the changes described above, the definition of the SR-CQLR$_P$ critical value is unchanged.

In the time series context,

$$
\begin{aligned}
V_F &:= \lim \operatorname{Var}_F\left(n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} g_i \\ \operatorname{vec}(G_i)\end{pmatrix}\right) \\
&= \sum_{m=-\infty}^{\infty} E_F\begin{pmatrix} g_i \\ \operatorname{vec}(G_i - E_F G_i)\end{pmatrix}\begin{pmatrix} g_{i-m} \\ \operatorname{vec}(G_{i-m} - E_F G_{i-m})\end{pmatrix}' \quad \text{and} \qquad (18.4) \\
\Omega_F &:= \sum_{m=-\infty}^{\infty} E_F g_i g_{i-m}',
\end{aligned}
$$

where the dependence of various quantities on the null value θ_0 is suppressed for notational simplicity. The second equality holds for $F \in \mathcal{F}_{\mathrm{TS},P}^{\mathrm{SR}}$.[52]

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

Theorem 18.1. *Suppose the* SR-AR, SR-CQLR, *and* SR-CQLR$_P$ *tests are defined as in this section, the null parameter spaces for F are* $\mathcal{F}_{\mathrm{TS,AR}}^{\mathrm{SR}}$, $\mathcal{F}_{\mathrm{TS}}^{\mathrm{SR}}$, *and* $\mathcal{F}_{\mathrm{TS},P}^{\mathrm{SR}}$, *respectively, and the corresponding Assumption* SR-Ω, SR-V, *or* SR-V$_P$ *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, *and* SR-CQLR$_P$ CSs *for the parameter spaces* $\mathcal{F}_{\mathrm{TS},\Theta,\mathrm{AR}}^{\mathrm{SR}}$, $\mathcal{F}_{\mathrm{TS},\Theta}^{\mathrm{SR}}$, *and* $\mathcal{F}_{\mathrm{TS},\Theta,P}^{\mathrm{SR}}$, *respectively, provided the corresponding Assumption* SR-Ω-CS, SR-V-CS, *or* SR-V$_P$-CS *holds for each* CS, *rather than Assumption* SR-Ω, SR-V, *or* SR-V$_P$.

19. SR-CQLR , SR-CQLR$_P$, 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) IVs 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), and Andrews, Marmer, and Yu (2019).[53]

In this section, we show that the components of the SR-QLR$_P$ statistic and its corresponding conditioning matrix are asymptotically equivalent to those of Moreira's (2003)

[52]This is shown in the proof of Lemma 20.1 in Section 20 in the SM to AG1.

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

LR statistic and its conditioning statistic, respectively, in the homoskedastic linear IV model with $k \geq p$ fixed (i.e., nonrandom) IVs 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 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 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 IVs being fixed, whereas the SR-QLR statistic and its conditioning matrix are designed (essentially) for random IVs.

We also evaluate the behavior of Kleibergen's (2005, 2007) nonlinear CLR tests in the homoskedastic linear IV model with fixed IVs. 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 19.3 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).

Section 21 below provides finite-sample simulation results that illustrate the results of the previous paragraph for Kleibergen's CLR test with moment-variance weighting.

19.1 *Normal linear IV model with $p \geq 1$ endogenous variables*

Here, we define the CLR test of Moreira (2003) in the homoskedastic Gaussian linear (HGL) IV model with $p \geq 1$ endogenous regressor variables and $k \geq p$ fixed (i.e., nonrandom) IVs. The linear IV regression model is

$$\begin{aligned} y_{1i} &= Y_{2i}'\theta + u_i \quad \text{and} \\ Y_{2i} &= \pi' Z_i + V_{2i}, \end{aligned} \tag{19.1}$$

where $y_{1i} \in R$ and $Y_{2i} \in R^p$ are endogenous variables, $Z_i \in R^k$ for $k \geq p$ is a vector of fixed IVs, and $\pi \in R^{k\times p}$ is an unknown unrestricted parameter matrix. In terms of its reduced-form equations, the model is

$$\begin{aligned} &y_{1i} = Z_i'\pi\theta + V_{1i}, \qquad Y_{2i} = \pi' Z_i + V_{2i}, \qquad V_i := (V_{1i}, V_{2i}')', \\ &V_{1i} = u_i + V_{2i}'\theta, \quad \text{and} \quad \Sigma_V := EV_iV_i'. \end{aligned} \tag{19.2}$$

For simplicity, no exogenous variables are included in the structural equation. The reduced-form errors are $V_i \in R^{p+1}$. In the HGL model, $V_i \sim N(0^{p+1}, \Sigma_V)$ for some positive definite $(p+1) \times (p+1)$ matrix Σ_V.

The IV moment functions and their derivatives with respect to θ are

$$\begin{aligned} g(W_i, \theta) &= Z_i\bigl(y_{1i} - Y_{2i}'\theta\bigr) \quad \text{and} \\ G(W_i, \theta) &= -Z_i Y_{2i}', \quad \text{where } W_i := \bigl(y_{1i}, Y_{2i}', Z_i'\bigr)'. \end{aligned} \tag{19.3}$$

Moreira (2003, p. 1033) shows that the LR statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in the HGL model in (19.1)–(19.2) when Σ_V is known is

$$\begin{aligned} \mathrm{LR}_{\mathrm{HGL},n} &:= \overline{S}_n'\overline{S}_n - \lambda_{\min}\bigl((\overline{S}_n, \overline{T}_n)'(\overline{S}_n, \overline{T}_n)\bigr), \quad \text{where} \\ \overline{S}_n &:= \bigl(Z_{n\times k}'Z_{n\times k}\bigr)^{-1/2} Z_{n\times k}' Y b_0 \bigl(b_0'\Sigma_V b_0\bigr)^{-1/2} \\ &= \bigl(n^{-1}Z_{n\times k}'Z_{n\times k}\bigr)^{-1/2} n^{1/2}\widehat{g}_n \bigl(b_0'\Sigma_V b_0\bigr)^{-1/2} \in R^k, \\ \overline{T}_n &:= \bigl(Z_{n\times k}'Z_{n\times k}\bigr)^{-1/2} Z_{n\times k}' Y \Sigma_V^{-1} A_0 \bigl(A_0'\Sigma_V^{-1}A_0\bigr)^{-1/2} \\ &= -\bigl(n^{-1}Z_{n\times k}'Z_{n\times k}\bigr)^{-1/2} n^{1/2}(\widehat{G}_n\theta_0 - \widehat{g}_n, \widehat{G}_n)\Sigma_V^{-1} A_0 \bigl(A_0'\Sigma_V^{-1}A_0\bigr)^{-1/2} \\ &\in R^{k\times p}, \\ Z_{n\times k} &:= (Z_1, \ldots, Z_n)' \in R^{n\times k}, \qquad Y := (Y_1, \ldots, Y_n)' \in R^{n\times(p+1)}, \\ Y_i &:= \bigl(y_{1i}, Y_{2i}'\bigr)' \in R^{p+1}, \qquad b_0 := \bigl(1, -\theta_0'\bigr)' \in R^{p+1}, \\ \widehat{g}_n &:= n^{-1}\sum_{i=1}^n g(W_i, \theta_0), \qquad A_0 := (\theta_0, I_p)' \in R^{(p+1)\times p}, \\ \widehat{G}_n &:= n^{-1}\sum_{i=1}^n G(W_i, \theta_0), \end{aligned} \tag{19.4}$$

$\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix, and the second equality for $\overline{T}_n$ holds by (29.12) in the SM.[54] Note that $(\overline{S}_n, \overline{T}_n)$ is a (conveniently transformed) sufficient statistic for (θ, π) under normality of V_i, known variance matrix Σ_V, and fixed IVs.

Moreira's (2003) CLR test uses the $\mathrm{LR}_{\mathrm{HGL},n}$ statistic and a conditional critical value that depends on the $k \times p$ matrix $\overline{T}_n$ through the conditional critical value function $c_{k,p}(\cdot, 1-\alpha)$ defined in (24). For $\alpha \in (0, 1)$, Moreira's CLR test with nominal level α rejects H_0 if

$$\mathrm{LR}_{\mathrm{HGL},n} > c_{k,p}(\overline{T}_n, 1-\alpha). \tag{19.5}$$

When Σ_V is unknown, Moreira (2003) replaces Σ_V by a consistent estimator.

Moreira's (2003) CLR test is similar with finite-sample size α in the HGL model with known Σ_V. Intuitively, the strength of the IVs affects the null distribution of the test

[54] We let $Z_{n\times k}$ (rather than Z) denote $(Z_1, \ldots, Z_n)'$, because we use Z to denote a k vector of standard normals below.

statistic $\mathrm{LR}_{\mathrm{HGL},n}$ and the critical value $c_{k,p}(\overline{T}_n, 1-\alpha)$ adjusts accordingly to yield a test with size α using the dependence of the null distribution of $\overline{T}_n$ on the strength of the IVs. When $p=1$, this test has been shown to have some (approximate) asymptotic optimality properties; see Andrews, Moreira, and Stock (2006, 2008), Chernozhukov, Hansen, and Jansson (2009), and Andrews, Marmer, and Yu (2019).

For $p \geq 2$, the asymptotic properties of Moreira's CLR test, such as its asymptotic size and similarity, are not available in the literature. The results for the SR-CQLR$_P$ test, specialized to the linear IV model (with or without Gaussianity, homoskedasticity, and/or independence of the errors), fill this gap.

19.2 *Homoskedastic linear IV model*

The model we consider in the remainder of this section is the homoskedastic linear IV model introduced in Section 19.1 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* $E\|V_i\|^4 < \infty$.[55]

(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 *or* SR-QLR$_P$ *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 19.1(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.

[55]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}. Nevertheless, it should be kept in mind that the reduced-form parameters π_n may depend on n.

19.3 SR-CQLR$_P$ *test*

The components of the SR-QLR$_P$ 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 (9) and (15.6)) when $\widehat{r}_n = k$, which holds wp $\to 1$ under Assumption HLIV. Those of Moreira (2003) are $\overline{S}_n$ and $\overline{T}_n$ (see (19.4)). The asymptotic equivalence of these components in the model specified by (19.1)–(19.2) 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 $\widetilde{\Sigma}_n$ of the test statistic SR-QLR$_{Pn}$ and its conditioning statistic.

Lemma 19.1. *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) $\widetilde{R}_n \to_p \Sigma_V \otimes K_Z$,

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

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

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

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

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

Comment. (i) The minus sign in Lemma 19.1(f) is not important because QLR$_{Pn}$ (defined in the paragraph containing (15.7) using the formula in (23)) is unchanged if $\widetilde{D}_n^*$ is replaced by $-\widetilde{D}_n^*$ (and SR-QLR$_{Pn}$ = QLR$_{Pn}$ wp $\to 1$ under Assumption HLIV).[56]

(ii) The results of Lemma 19.1 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 19.1(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 (19.2) 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 IVs are strong, that is, $\liminf_{n\to\infty} \pi_n' n^{-1}\sum_{i=1}^n Z_i Z_i'\pi_n > 0$, then contiguous alternatives have true θ_n values of distance $O(n^{-1/2})$ from the null value θ_0. If the IVs are weak in the standard sense, for example, $\pi_n = \pi n^{-1/2}$ for some fixed matrix π, then all θ values not equal θ_0 yield contiguous alternatives. For semi-strong identification in the standard sense, for example, $\pi_n = \pi n^{-\delta}$ 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 $\pi_n = (\pi_{1n}, \ldots, \pi_{pn}) \in R^{k\times p}$, $n^{1/2}\|\pi_{jn}\| \to \infty$ for all $j \leq p$, $\limsup_{n\to\infty} \lambda_{\min}(n\pi_n'\pi_n) < \infty$, and θ_n is such that $\pi_n(\theta_n - \theta_0) = O(n^{-1/2})$.

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

(iii) The proofs of Lemma 19.1 and Lemmas 19.2 and 19.3 below are given in Section 29 below.

19.4 *SR-CQLR test*

The components of the SR-QLR 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 (8) and (23)) when $\widehat{r}_n = k$, which holds wp $\to 1$ under Assumption HLIV. Here, we show that the conditioning statistic $n^{1/2}\widehat{D}_n^*$ is asymptotically equivalent to Moreira's (2003) conditioning statistic $\overline{T}_n$ (in the homoskedastic linear IV model with fixed IVs) 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 19.2 below.) Nevertheless, under strong and semi-strong IVs, the SR-CQLR test and Moreira's CLR test are asymptotically equivalent.[57] In consequence, when $p = 1$, the SR-CQLR test and Moreira's CLR test are asymptotically equivalent (because standard weak, strong, and semi-strong identification cover all possible cases). When $p \geq 2$, 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 $p \geq 2$, the differences appear to be small because they are due only to the differences between fixed IVs and random IVs (which cause Σ_V to differ somewhat from Σ_{V*} defined below).

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

$$\begin{aligned}\zeta_n(\pi) &:= n^{-1}\sum_{i=1}^{n}(\pi' \otimes Z_i)Z_iZ_i'(\pi \otimes Z_i') \\ &\quad - \left(n^{-1}\sum_{i=1}^{n}(\pi' \otimes Z_i)Z_i\right)\left(n^{-1}\sum_{i=1}^{n}(\pi' \otimes Z_i)Z_i\right)' \\ &\in R^{kp\times kp}. \end{aligned} \tag{19.6}$$

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

$$R(\pi) := \Sigma_V \otimes K_Z + (B' \otimes I_k)\begin{bmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi)\end{bmatrix}(B \otimes I_k) \in R^{k(p+1)\times k(p+1)}, \tag{19.7}$$

where $B = B(\theta_0)$ is defined in (19).

[57] This holds because, under strong and semi-strong IVs, the SR-QLR 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 7.1 for the SR-QLR 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 19.1(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 7.1 for the SR-CQLR critical value.

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

$$\Sigma_{V*j\ell} := \operatorname{tr}\bigl(R_{j\ell}(\pi_*)'K_Z^{-1}\bigr)/k, \tag{19.8}$$

where $R_{j\ell}(\pi_*)$ denotes the (j,ℓ) $k\times k$ submatrix of $R(\pi_*)$ for $j,\ell = 1,\dots,p+1$ 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) $\lim n^{-1}\sum_{i=1}^n \operatorname{vec}(Z_iZ_i')\operatorname{vec}(Z_iZ_i')'$ *exists and is finite,*

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

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

Assumption HLIV2(c) implies that the eigenvalue adjustment to $\widehat{\Sigma}_n$ employed in the SR-QLR statistic has no effect asymptotically. One could analyze the behavior of the SR-CQLR 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 $n^{1/2}\widehat{D}_n^*$ is given in the following lemma. Under Assumption HLIV, $n^{1/2}\widehat{D}_n^*$ equals the SR-CQLR conditioning statistic $n^{1/2}\widehat{D}_{An}^*$ wp $\to 1$ (because $\widehat{r}_n = k$ wp $\to 1$).

Lemma 19.2. *Suppose Assumptions* HLIV *and* HLIV2 *hold. Under the null hypothesis* $H_0: \theta = \theta_0$ *and any* $p \ge 1$, *we have*

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

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

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

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

Comment. (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 19.2(d) yields

$$n^{1/2}\widehat{D}_n^* = -\bigl(I_k + o_p(1)\bigr)\overline{T}_n\bigl(I_p + o_p(1)\bigr) + o_p(1) \tag{19.9}$$

and $n^{1/2}\widehat{D}_n^*$ 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 $n^{1/2}\widehat{D}_n^*$ 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 (19.7). This factor arises because the IVs are fixed in the linear IV model (by assumption), but the variance estimator $\widehat{V}_n$, which appears in $\widehat{R}_n$ (see (19)), and which determines $\widehat{\Sigma}_n$ and Σ_{V*}, treats the IVs as though they are random.

19.5 *Kleibergen's nonlinear CLR tests*

19.5.1 *Definitions of the 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 IVs. 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:

$$\begin{aligned}\mathrm{CLR}_n(\theta) &:= \frac{1}{2}\Big(\mathrm{AR}_n(\theta) - \mathrm{rk}_n(\theta) \\ &\qquad + \sqrt{\big(\mathrm{AR}_n(\theta) - \mathrm{rk}_n(\theta)\big)^2 + 4\mathrm{LM}_n(\theta)\cdot \mathrm{rk}_n(\theta)}\Big), \quad \text{where} \\ \mathrm{LM}_n(\theta) &:= n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1/2}(\theta)P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta)\end{aligned} \tag{19.10}$$

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

The critical value of Kleibergen's CLR test is $c(1-\alpha, \mathrm{rk}_n(\theta))$, where $c(1-\alpha, r)$ is the $1-\alpha$ quantile of the distribution of

$$\mathrm{clr}(r) := \frac{1}{2}\Big(\chi_p^2 + \chi_{k-p}^2 - r + \sqrt{\big(\chi_p^2 + \chi_{k-p}^2 - r\big)^2 + 4\chi_p^2 r}\Big) \tag{19.11}$$

for $0 \leq r < \infty$ and the chi-square random variables χ_p^2 and χ_{k-p}^2 in (19.11) are independent. The CLR test rejects the null hypothesis $H_0 : \theta = \theta_0$ if $\mathrm{CLR}_n > c(1-\alpha, \mathrm{rk}_n)$ (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 $\mathrm{rk}_n(\theta)$. Kleibergen (2005, p. 1114, 2007, equation (37)) and Smith (2007, p. 7, footnote 4) propose to take $\mathrm{rk}_n(\theta)$ to be a function of $\widetilde{V}_{Dn}^{-1/2}(\theta)\,\mathrm{vec}(\widehat{D}_n(\theta))$, where $\widetilde{V}_{Dn}(\theta) \in R^{kp\times kp}$ is a consistent estimator of the covariance matrix of the asymptotic distribution of $\mathrm{vec}(\widehat{D}_n(\theta))$ (after suitable normalization). We refer to $\widetilde{V}_{Dn}^{-1/2}(\theta)\,\mathrm{vec}(\widehat{D}_n(\theta))$ as the orthogonalized sample Jacobian with Jacobian-variance weighting. In the i.i.d. case considered here, we have

$$\begin{aligned}\widetilde{V}_{Dn}(\theta) &:= n^{-1}\sum_{i=1}^{n}\mathrm{vec}\big(G_i(\theta)-\widehat{G}_n(\theta)\big)\,\mathrm{vec}\big(G_i(\theta)-\widehat{G}_n(\theta)\big)' \\ &\quad - \widehat{\varGamma}_n(\theta)\widehat{\varOmega}_n^{-1}(\theta)\widehat{\varGamma}_n(\theta)', \quad \text{where} \\ \widehat{\varGamma}_n(\theta) &:= \big(\widehat{\varGamma}_{1n}(\theta)',\ldots,\widehat{\varGamma}_{pn}(\theta)'\big)' \in R^{pk\times k}\end{aligned} \tag{19.12}$$

and $\widehat{\varGamma}_{1n}(\theta),\ldots,\widehat{\varGamma}_{pn}(\theta)$ are defined in (18).

Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) proposed to take $\mathrm{rk}_n(\theta)$ to be a function of $\widehat{\varOmega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$. We refer to $\widehat{\varOmega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$ 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.

19.5.2 $p=1$ *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 $\widehat{\varOmega}_n(\theta)$. 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

$$\begin{aligned}\mathrm{CLR}_n(\theta) &= \mathrm{AR}_n(\theta) - \lambda_{\min}\big(\big(n^{1/2}\widehat{\varOmega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), r_n(\theta)\big)' \\ &\quad \times \big(n^{1/2}\widehat{\varOmega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), r_n(\theta)\big)\big) \quad \text{provided} \\ \mathrm{rk}_n(\theta) &= r_n(\theta)'r_n(\theta) \quad \text{for some random vector } r_n(\theta)\in R^k.\end{aligned} \tag{19.13}$$

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 $\mathrm{rk}_n(\theta)=r_n(\theta)'r_n(\theta)$ is of the same form as Moreira's (2003) LR statistic with $r_n(\theta_0)$ in place of $\overline{T}_n$ and with $n^{1/2}\widehat{\varOmega}_n^{-1/2}(\theta_0)\widehat{g}_n(\theta_0)$ in place of $\overline{S}_n$, where θ_0 is the null value of θ.[58] The two choices for $\mathrm{rk}_n(\theta)$ that we consider when $p=1$ are

$$\mathrm{rk}_{1n}(\theta) := n\widehat{D}_n(\theta)'\widetilde{V}_{Dn}^{-1}(\theta)\widehat{D}_n(\theta) \quad \text{and} \quad \mathrm{rk}_{2n}(\theta) := n\widehat{D}_n(\theta)'\widehat{\varOmega}_n^{-1}(\theta)\widehat{D}_n(\theta). \tag{19.14}$$

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

[58]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, $\mathrm{rk}_n(\theta)$ equals the squared length of some k vector $r_n(\theta)$.

Let

$$\zeta_n(\pi) := n^{-1}\sum_{i=1}^{n} Z_i Z_i'(Z_i'\pi)^2 - \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\pi\right)\left(n^{-1}\sum_{\ell=1}^{n} Z_\ell Z_\ell'\pi\right)'. \tag{19.15}$$

This definition of $\zeta_n(\pi)$ is the same as in (19.6) when $p=1$.

Lemma 19.3. *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) $\mathrm{rk}_{1n}(\theta_0) = \overline{T}_n'[I_k + L_{V0}K_Z^{-1/2}\zeta_n(\pi_n)K_Z^{-1/2} + o_p(1)]^{-1}\overline{T}_n \cdot (1+o_p(1)) + o_p(1)$,

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

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

Comment. (i) If $\pi_n \to 0$, then $\zeta_n(\pi_n) \to 0$ and Lemma 19.3(a) shows that $\mathrm{rk}_{1n}(\theta_0)$ equals $\overline{T}_n'\overline{T}_n(1+o_p(1)) + o_p(1)$. That is, under weak IVs and semi-strong IVs, $\mathrm{rk}_{1n}(\theta_0)$ reduces asymptotically to Moreira's (2003) conditioning statistic. Under strong IVs, this does not occur. However, under strong IVs, we have $\mathrm{rk}_{1n}(\theta_0) \to_p \infty$, just as $\overline{T}_n'\overline{T}_n \to_p \infty$. In consequence, the test constructed using $\mathrm{rk}_{1n}(\theta_0)$ has the same asymptotic properties as Moreira's (2003) CLR test under the null and contiguous alternative distributions.

(ii) Simple calculations show that $\zeta_n(\pi_n)$ is positive semi-definite (psd). Hence, $\mathrm{rk}_{1n}(\theta_0)$ is smaller than it would be if the second summand in the square brackets in Lemma 19.3(a) was zero.

(iii) Lemma 19.3(b) shows that the rank statistic $\mathrm{rk}_{2n}(\theta_0)$ 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 IVs under weak IVs. This has negative consequences for its power. Under strong or semi-strong IVs, this test does reduce asymptotically to Moreira's (2003) CLR test because $\mathrm{rk}_{1n}(\theta_0) \to_p \infty$, 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 19.3(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 $\mathrm{rk}_{2n}(\theta_0)$ 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 $\mathrm{rk}_{2n}(\theta_0)$ statistic is larger (smaller) than desired and it behaves as though the IVs are stronger (weaker) than they really are, which sacrifices power unless the IVs are quite strong (weak). Note that the inappropriate scale of $\mathrm{rk}_{2n}(\theta_0)$ does not cause asymptotic size problems, only power reductions.

19.5.3 *$p \geq 2$ 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 $\text{rk}_n(\theta_0)$ for several reasons.

First, Moreira's (2003) LR statistic is given in (19.4), whereas Kleibergen's (2005) nonlinear LR statistic is defined in (19.10). By Lemma 19.1(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, $\text{AR}_n = \overline{S}_n'\overline{S}_n + o_p(1)$. Even if rk_n takes the form $r_n'r_n$ for some random k vector r_n, it is not the case that

$$\text{CLR}_n = \text{AR}_n - \lambda_{\min}\big((n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, r_n)'(n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, r_n)\big) \tag{19.16}$$

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, namely, 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 19.1(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 19.3(b) and (c) for the $p = 1$ case.

20. 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 tests when the variance matrix of the moments is singular and near singular.[59] The model we consider is the following homoskedastic linear IV model: $y_{1i} = Y_{2i}\beta + U_i$ and $Y_{2i} = Z_i'\pi + V_{1i}$, where all quantities are scalars except Z_i, $\pi \in R^{d_Z}$, $\theta = (\beta, \pi')' \in R^{\beta + d_Z}$, $EU_i = EV_{2i} = 0$, $EU_iZ_i = EV_{1i}Z_i = 0^{d_Z}$, and $E(V_iV_i'|Z_i) = \Sigma_V$ a.s. for $V_i := (V_{1i}, V_{2i})'$ and some 2×2 constant matrix Σ_V. The corresponding reduced-form equations are $y_{1i} = Z_i'\pi\beta + V_{1i}$ and $Y_{2i} = Z_i'\pi + V_{1i}$, where $V_{1i} = U_i + V_{2i}\beta$. The moment conditions for θ are $g_i(\theta) = ((y_{1i} - Z_i'\pi\beta)Z_i', (Y_{2i} - Z_i'\pi)Z_i')' \in R^k$, where $k = 2d_Z$ and d_Z is the dimension of Z_i. The variance matrix $\Sigma_V \otimes EZ_iZ_i'$ of $g_i(\theta_0) = (V_{1i}Z_i', V_{2i}Z_i')'$ is singular whenever the covariance between the reduced-form

[59] Analogous results for the SR-CQLR$_2$ test are not provided because the moment functions considered are not of the form in (15.1), which is necessary to apply the SR-CQLR$_2$ test.

Table SM-1. Null rejection probabilities (×100) of nominal 5% SR-AR and SR-CQLR tests with singular and near singular variance matrices of the moment functions and $k = 8$.

		SR-AR			SR-CQLR		
n	ρ_V:	0.95	0.999999	1.0	0.95	0.999999	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
1000		5.5	5.5	5.2	5.3	5.3	5.1
2000		5.0	5.0	4.9	4.8	4.8	4.8
4000		5.0	5.0	5.1	4.8	4.8	5.0
8000		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

errors V_{1i} and V_{2i} is one (or minus one) or EZ_iZ_i' is singular. In this model, we are interested in joint inference concerning β and π. This is of interest when one wants to see how the magnitude of π affects the range of plausible β values.

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: $\rho_V = 0.95$, 0.999999, and 1.0; $n = 250$, 500, 1000, 2000, 4000, 8000, 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 $\lambda = \pi' EZ_iZ_i'\pi = 10$ 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 (14) 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 SM-1, SM-2, and SM-3 report results for $k = 8$ (which corresponds to $d_Z = 4$), $k = 4$, and $k = 12$, respectively. Table SM-1 shows that the SR-AR and SR-CQLR 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

Table SM-2. Null rejection probabilities (×100) of nominal 5% SR-AR and SR-CQLR tests with singular and near singular variance matrices of the moment functions and $k = 4$.

		SR-AR			SR-CQLR		
n	ρ_V:	0.95	0.999999	1.0	0.95	0.999999	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
1000		4.9	4.9	5.1	4.8	4.8	4.8
2000		5.1	5.1	5.2	5.0	5.0	5.0
4000		5.1	5.1	5.1	5.0	5.0	4.9
8000		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 SM-3. Null rejection probabilities (×100) of nominal 5% SR-AR and SR-CQLR tests with singular and near singular variance matrices of the moment functions and $k = 12$.

		SR-AR			SR-CQLR		
n	ρ_V:	0.95	0.999999	1.0	0.95	0.999999	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
1000		5.5	5.5	5.3	5.5	5.5	5.3
2000		5.2	5.2	5.1	5.2	5.2	5.1
4000		5.1	5.1	5.1	5.1	5.1	5.1
8000		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

$k = 8$, 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 SM-2 and SM-3 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 $n \leq 500$. The results for $k = 8$ and $C = 0$, 2, $\sqrt{30}$, and 10 are similar. For brevity, these results are not reported.

We conclude that the method introduced in Section 4 to make the SR-AR and SR-CQLR tests robust to singularity works very well in the model that is considered in the simulations.

21. 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 and SR-CQLR$_P$ tests, Kleibergen's CLR test with Jacobian-variance weighting (JVW-CLR), and the CLR test of Moreira (2003) (Mor-CLR).[60] As noted at the beginning of Section 19.5, Lemma 19.3 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.[61] 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 19.3.

The model is the homoskedastic normal linear IV model introduced in Section 19.1 with unknown error variance matrix Σ_V and $p = 1$. The IVs are fixed—they are generated

[60]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 $\mathrm{rk}_n = \lambda_{\min}(n\widehat{D}_n'\widehat{\Omega}_n^{-1/2}\widehat{D}_n)$ and $\mathrm{rk}_n = \lambda_{\min}(n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n)$, respectively, where $\widehat{\Omega}_n$ and $\widehat{D}_n$ are defined in (8) and (18) 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 (19.12). 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.

[61]The constant scale is the constant $(L_{V0}b_0'\Sigma_V b_0)^{-1}$ in Lemma 19.3(b) and (c).

once from a $N(0^k, I_k)$ distribution. The sample size n equals 1000. The hypotheses are $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$. The tests have nominal size 0.05. The power results are based on 40,000 simulation repetitions and 1000 critical value repetitions and are size-corrected (by adding nonnegative constants to the critical values of those tests that overreject 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 IVs is k. The MVW-CLR and JVW-CLR tests employ the Robin and Smith (2000) rank statistic. 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 SM-2 (with $\theta\lambda^{1/2}$ on the horizontal axes with $\lambda^{1/2}$ fixed). Figure SM-2(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, SR-CQLR$_P$, 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, SR-CQLR$_P$, 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 SM-2(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, SR-CQLR$_P$, JVW-CLR, Mor-CLR, and LM tests, which have essentially equal and optimal power.

22. Eigenvalue-adjustment procedure

Eigenvalue adjustments are made to two sample matrices that appear in the SR-CQLR and SR-CQLR$_P$ 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 nonzero 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 $\Lambda_H = \text{Diag}\{\lambda_{H1}, \ldots, \lambda_{Hd_H}\} \in R^{d_H \times d_H}$ 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

$$
\begin{aligned}
&H^\varepsilon := A_H \Lambda_H^\varepsilon A_H', \quad \text{where} \\
&\Lambda_H^\varepsilon := \text{Diag}\big\{\max\big\{\lambda_{H1}, \lambda_{\max}(H)\varepsilon\big\}, \ldots, \max\big\{\lambda_{Hd_H}, \lambda_{\max}(H)\varepsilon\big\}\big\}.
\end{aligned}
\tag{22.1}
$$

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

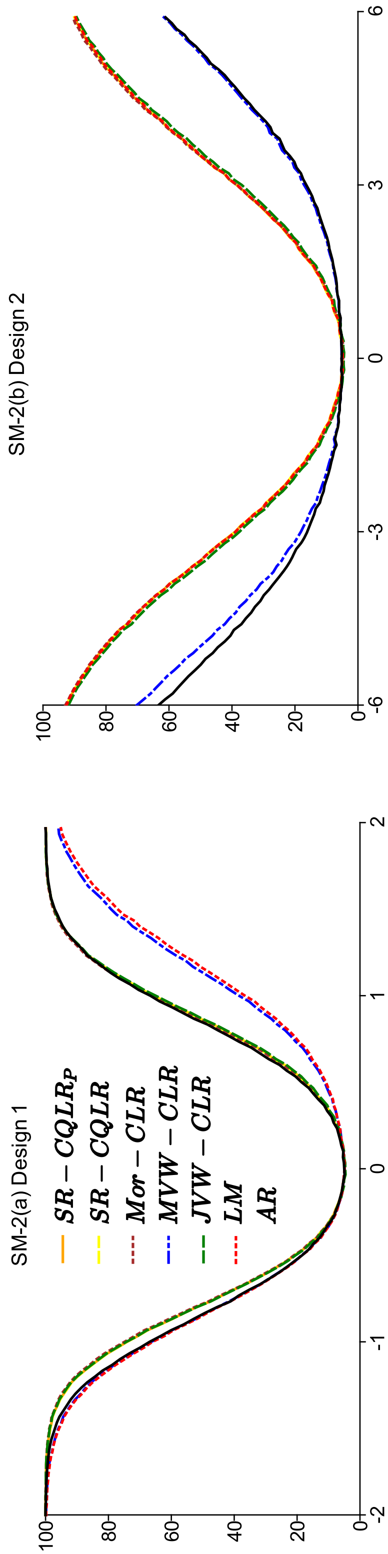

FIGURE SM-2. Power of Kleibergen's MVW-CLR and other tests in linear IV model.

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

LEMMA 22.1. *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 nonzero positive semi-definite nonrandom matrix. Then,*

(a) (*uniqueness*) *H^{ε}, defined in (22.1), is uniquely defined. (i.e., 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 $H_n \to H$.*

COMMENT. (i) The lower bound $\lambda_{\max}(H)\varepsilon$ for $\lambda_{\min}(H^{\varepsilon})$ given in Lemma 22.1(b) is positive provided $H \neq 0^{d_H \times d_H}$.

(ii) Lemma 22.1(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 22.1(d) is an important property. For example, one does not want the choice of measurements in \$ or \$1000 to affect inference.

(iv) Continuity of $(\cdot)^{\varepsilon}$ established in Lemma 22.1(e) is an important property because it implies that for random matrices $\{\widehat{H}_n : n \geq 1\}$ for which $\widehat{H}_n \to_p H$, one has $\widehat{H}_n^{\varepsilon} \to_p H^{\varepsilon}$.

PROOF OF LEMMA 22.1. 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 (22) are unique up to the choice of vectors that span the eigenspace that corresponds to any eigenvalue. Suppose the $j, \ldots, j+d$ 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

$$
\begin{aligned}
A\Lambda^{\varepsilon}A'\xi &= A\Lambda^{\varepsilon}(A_1, A_2, A_3)'(\xi_1 + \xi_2) \\
&= A\Lambda^{\varepsilon}\big(0^{j-1\prime}, (A_2'\xi_1)', 0^{d_H - j - d\prime}\big)' + A\Lambda^{\varepsilon}\big((A_1'\xi_2)', 0^{d+1\prime}, (A_3'\xi_2)'\big)' \\
&= A\lambda_j^{\varepsilon}\big(0^{j-1\prime}, (A_2'\xi_1)', 0^{d_H - j - d\prime}\big)' + (A_1, A_2, A_3)\Lambda^{\varepsilon}\big((A_1'\xi_2)', 0^{d+1\prime}, (A_3'\xi_2)'\big)' \\
&= A_2A_2'\xi_1\lambda_j^{\varepsilon} + (A_1, A_3)\Lambda_{-}^{\varepsilon}\big((A_1'\xi_2)', (A_3'\xi_2)'\big)'
\end{aligned}
$$

$$= A_{2*}A'_{2*}\xi_1\lambda_j^\varepsilon + (A_1, A_3)\Lambda_-^\varepsilon\big(\big(A'_1\xi_2\big)', \big(A'_3\xi_2\big)'\big)'$$
$$= A_*\Lambda^\varepsilon A'_*\xi, \tag{22.2}$$

where $\Lambda_-^\varepsilon \in R^{(d_H-d-1)\times(d_H-d-1)}$ is the diagonal matrix equal to Λ^ε with its $j, \dots, j+d$ 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 $\lambda_j = \cdots = \lambda_{j+d}$ implies that $\lambda_j^\varepsilon = \cdots = \lambda_{j+d}^\varepsilon$, the fourth equality holds using the definition of Λ_-^ε, the fifth equality holds because $A_2A'_2 = A_{2*}A'_{2*}$ (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 (22.2) 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 $\{\max\{\lambda_{Hj}, \lambda_{\max}(H)\varepsilon\} : j = 1, \dots, d_H\}$ 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 nonzero, 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 $\Lambda_{cH}^\varepsilon = c\Lambda_H^\varepsilon$ (using the definition of Λ_H^ε in (22)) and $(cH)^\varepsilon = A_{cH}\Lambda_{cH}^\varepsilon A'_{cH} = cA_H\Lambda_H^\varepsilon A'_H = cH^\varepsilon$, which establishes part (d).

Now we prove part (e). Let $A_n\Lambda_nA'_n$ be a spectral decomposition of H_n for $n \geq 1$. Let $H_n^\varepsilon = A_n\Lambda_n^\varepsilon A'_n$ for $n \geq 1$, where Λ_n^ε is the diagonal matrix with jth diagonal element given by $\lambda_{nj}^\varepsilon = \max\{\lambda_{nj}, \lambda_{\max}(H_n)\varepsilon\}$ and λ_{nj} is the jth largest eigenvalue of H_n. (By part (a) of the lemma, H_n^ε 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, Theorem 3.1, pp. 37–38)). In turn, this gives $\Lambda_n^\varepsilon \to \Lambda^\varepsilon$, where Λ^ε is the diagonal matrix with jth diagonal element given by $\lambda_j^\varepsilon = \max\{\lambda_j, \lambda_{\max}(H)\varepsilon\}$ 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 $H_{n_m} = A_{n_m}\Lambda_{n_m}A'_{n_m} \to A\Lambda A'$, $H = A\Lambda A'$, $H_{n_m}^\varepsilon = A_{n_m}\Lambda_{n_m}^\varepsilon A'_{n_m} \to A\Lambda^\varepsilon A'$, and $A\Lambda^\varepsilon A' = H^\varepsilon$. Because every subsequence $\{n_\ell\}$ of $\{n\}$ has a subsubsequence $\{n_m\}$ for which $H_{n_m}^\varepsilon \to H^\varepsilon$, we obtain $H_n^\varepsilon \to H^\varepsilon$, which completes the proof of part (e). □

23. 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 CSs. 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

$$\text{SR-LM}_n(\theta) := n\widehat{g}_{An}(\theta)' P_{\widehat{\Omega}_{An}^{-1/2}(\theta)\widehat{D}_{An}(\theta)} \widehat{g}_{An}(\theta), \tag{23.1}$$

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

$$\text{SR-LM}_n(\theta_0) > \chi^2_{\min\{\widehat{r}_n(\theta_0),p\},1-\alpha}, \tag{23.2}$$

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 $\delta_1 > 0$,

$$\mathcal{F}^{SR}_{LM} := \bigcup_{j=0}^{\min\{r_F,p\}} \mathcal{F}^{SR}_{LMj}, \quad \text{where}$$

$$\begin{aligned} \mathcal{F}^{SR}_{LMj} &:= \Big\{F \in \mathcal{F}^{SR} : \tau^*_{jF} \geq \delta_1 \text{ and } \lambda_{p-j}\big(\Psi_F^{C^{*\prime}_{F,k-j}G^*_i B^*_{F,p-j}\xi}\big) \geq \delta_1 \\ &\qquad \forall \xi \in R^{p-j} \text{ with } \|\xi\| = 1\Big\}, \\ G^*_i &:= \Pi_{1F}^{-1/2} A'_F G_i \in R^{r_F \times p}, \qquad r_F := \text{rk}(\Omega_F), \qquad g^*_i := \Pi_{1F}^{-1/2} A'_F g_i \in R^{r_F}, \\ \Psi_F^{a_i} &:= E_F a_i a'_i - E_F a_i g^{*\prime}_i \big(E_F g^*_i g^*_i\big)^{-1} E_F g^*_i a'_i \quad \text{for any random vector } a_i, \end{aligned} \tag{23.3}$$

τ^*_{jF} is the jth largest singular value of $E_F G^*_i$ for $j = 1, \ldots, \min\{r_F, p\}$, $\tau^*_{0F} := \delta_1$, 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 $(\kappa^*_{1F}, \ldots, \kappa^*_{pF})$ are nonincreasing, C^*_F is a $r_F \times r_F$ orthogonal matrix of eigenvalues of $(E_F G^*_i)(E_F G^*_i)'$ ordered so that the corresponding eigenvalues $(\kappa^*_{1F}, \ldots, \kappa^*_{r_F F})$ 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 $C^*_{F,k-j} \in R^{r_F \times (r_F - j)}$.[62] ,[63] 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, $a_i - E_F a_i g^{*\prime}_i (E_F g^*_i g^*_i)^{-1} g^*_i$, 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 $\mathcal{F}^{SR}$) 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

[62]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.

[63]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.

(after pre- and post-multiplication by suitable matrices). AG1 shows that these conditions are not redundant. 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 17 below. For brevity, we do not provide the details. Extensions of the asymptotic size and similarity results to SR-LM CSs are analogous to those for the SR-AR and SR-CQLR CSs.

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

24. Proofs of Lemmas 16.2, 5.1, and 15.1

Lemma 16.2 of AG2. *Let D be a $k \times p$ matrix with the singular value decomposition $D = CYB'$, 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 Y 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 $c_{k,p}(D, 1-\alpha) = c_{k,p}(Y, 1-\alpha)$.*

Proof of Lemma 16.2. Define

$$B^+ := \begin{bmatrix} B & 0^p \\ 0^{p'} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}. \tag{24.1}$$

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

$$\begin{aligned}
\left|(D, Z)'(D, Z) - \kappa I_{p+1}\right| &= 0 \quad \text{or} \\
\left|B^{+\prime}(D, Z)'(D, Z)B^+ - \kappa I_{p+1}\right| &= 0 \quad \text{or} \\
\left|(DB, Z)'(DB, Z) - \kappa I_{p+1}\right| &= 0 \quad \text{or} \\
\left|(CY, Z)'CC'(CY, Z) - \kappa I_{p+1}\right| &= 0 \quad \text{or,} \\
\left|\left(Y, Z^*\right)'\left(Y, Z^*\right) - \kappa I_{p+1}\right| &= 0, \quad \text{where } Z^* := C'Z \sim N\left(0^k, I_k\right),
\end{aligned} \tag{24.2}$$

the equivalence of the first and second lines holds because $|A_1 A_2| = |A_1| \cdot |A_2|$, $|B^+| = 1$, 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 $DB = CYB'B = CY$ and $CC' = I_k$, and the equivalence of the last two lines holds by $CC' = I_k$ and the definition of Z^*. Equation (24.2) implies that $\lambda_{\min}((D, Z)'(D, Z))$ equals $\lambda_{\min}((Y, Z^*)'(Y, Z^*))$. In addition, $Z'Z = Z^{*\prime}Z^*$. Hence,[64]

$$\mathrm{CLR}_{k,p}(D) = Z'Z - \lambda_{\min}\big((D, Z)'(D, Z)\big) = Z^{*\prime}Z^* - \lambda_{\min}\big(\left(Y, Z^*\right)'\left(Y, Z^*\right)\big). \tag{24.3}$$

[64]The quantity $\mathrm{CLR}_{k,p}(D)$ is written in terms of (D, Z) in (24.3), whereas it is written in terms of (Z, D) in (24). Both expressions give the same value.

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

LEMMA 5.1 OF AG2. *The statistics QLR_n, $c_{k,p}(n^{1/2}\widehat{D}_n^*, 1-\alpha)$, $\widehat{D}_n^{*\prime}\widehat{D}_n^*$, AR_n, $\widehat{\Sigma}_n$, and $\widehat{L}_n$ are invariant to the transformation $(g_i, G_i) \rightsquigarrow (Mg_i, MG_i)$ $\forall i \le n$ for any $k \times k$ nonsingular matrix M. This transformation induces the following transformations*: $\widehat{g}_n \rightsquigarrow M\widehat{g}_n$, $\widehat{G}_n \rightsquigarrow M\widehat{G}_n$, $\widehat{\Omega}_n \rightsquigarrow M\widehat{\Omega}_n M'$, $\widehat{\Gamma}_{jn} \rightsquigarrow M\widehat{\Gamma}_{jn}M'$ $\forall j \le p$, $\widehat{D}_n \rightsquigarrow M\widehat{D}_n$, $\widehat{V}_n \rightsquigarrow (I_{p+1} \otimes M)\widehat{V}_n(I_{p+1} \otimes M')$, *and* $\widehat{R}_n \rightsquigarrow (I_{p+1} \otimes M)\widehat{R}_n(I_{p+1} \otimes M')$.

PROOF OF LEMMA 5.1. We refer to the results of the lemma for $g_i, G_i, \ldots, \widehat{R}_n$ as equivariance results. The equivariance results are immediate for g_i, G_i, $\widehat{g}_n$, $\widehat{G}_n$, $\widehat{\Omega}_n$, and $\widehat{\Gamma}_{jn}$. For $\widehat{D}_n = (\widehat{D}_{1n}, \ldots, \widehat{D}_{pn})$, we have

$$\widehat{D}_{jn} := \widehat{G}_{jn} - \widehat{\Gamma}_{jn}\widehat{\Omega}_n^{-1}\widehat{g}_n \rightsquigarrow M\widehat{G}_{jn} - M\widehat{\Gamma}_{jn}M'\big(M\widehat{\Omega}_n M'\big)^{-1}M\widehat{g}_n = M\widehat{D}_{jn} \tag{24.4}$$

for $j = 1, \ldots, p$. We have $f_i := (g_i', \mathrm{vec}(G_i)')' \rightsquigarrow ((Mg_i)', \mathrm{vec}(MG_i)')' = (I_{p+1} \otimes M)f_i$. Using this, we obtain $\widehat{V}_n = n^{-1}\sum_{i=1}^n (f_i - \widehat{f}_n)(f_i - \widehat{f}_n)' \rightsquigarrow (I_{p+1} \otimes M)\widehat{V}_n(I_{p+1} \otimes M')$. Next, 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 $\widehat{\Sigma}_{j\ell n} := \mathrm{tr}(\widehat{R}_{j\ell n}'\widehat{\Omega}_n^{-1})/k \rightsquigarrow \mathrm{tr}((M\widehat{R}_{j\ell n}M')' \times (M\widehat{\Omega}_n M')^{-1})/k = \mathrm{tr}(M\widehat{R}_{j\ell n}'M'M'^{-1}\widehat{\Omega}_n^{-1}M^{-1})/k = \widehat{\Sigma}_{j\ell n}$ for $j, \ell = 1, \ldots, p+1$ 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 $\widehat{D}_n^{*\prime}\widehat{D}_n^* := \widehat{L}_n^{1/2}\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n\widehat{L}_n^{1/2} \rightsquigarrow \widehat{L}_n^{1/2}\widehat{D}_n'M'(M\widehat{\Omega}_n M')^{-1}M\widehat{D}_n\widehat{L}_n^{1/2} = \widehat{D}_n^{*\prime}\widehat{D}_n^*$. 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 16.2.

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

$$\begin{aligned}\mathrm{QLR}_n &:= \mathrm{AR}_n - \lambda_{\min}\big(n\big(\widehat{g}_n, \widehat{D}_n\widehat{L}_n^{1/2}\big)'\widehat{\Omega}_n^{-1}\big(\widehat{g}_n, \widehat{D}_n\widehat{L}_n^{1/2}\big)\big)\\ &\rightsquigarrow \mathrm{AR}_n - \lambda_{\min}\big(n\big(M\widehat{g}_n, M\widehat{D}_n\widehat{L}_n^{1/2}\big)'\big(M\widehat{\Omega}_n M'\big)^{-1}\big(M\widehat{g}_n, M\widehat{D}_n\widehat{L}_n^{1/2}\big)\big) = \mathrm{QLR}_n,\end{aligned} \tag{24.5}$$

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

LEMMA 15.1. *The statistics* QLR_{Pn}, $c_{k,p}(n^{1/2}\widetilde{D}_n^*, 1-\alpha)$, $\widetilde{D}_n^{*\prime}\widetilde{D}_n^*$, AR_n, $\widehat{u}_{in}^*$, $\widetilde{\Sigma}_n$, *and* $\widetilde{L}_n$ *are invariant to the transformation $(Z_i, u_i^*) \rightsquigarrow (MZ_i, u_i^*)$ $\forall i \le n$ for any $k \times k$ nonsingular matrix M. This transformation induces the following transformations*: $g_i \rightsquigarrow Mg_i$ $\forall i \le n$, $G_i \rightsquigarrow MG_i$ $\forall i \le n$, $\widehat{g}_n \rightsquigarrow M\widehat{g}_n$, $\widehat{G}_n \rightsquigarrow M\widehat{G}_n$, $\widehat{\Omega}_n \rightsquigarrow M\widehat{\Omega}_n M'$, $\widehat{\Gamma}_{jn} \rightsquigarrow M\widehat{\Gamma}_{jn}M'$ $\forall j \le p$, $\widehat{D}_n \rightsquigarrow M\widehat{D}_n$, $Z_{n\times k} \rightsquigarrow Z_{n\times k}M'$, $\widetilde{\Xi}_n \rightsquigarrow M'^{-1}\widetilde{\Xi}_n$, $\widetilde{V}_n \rightsquigarrow (I_{p+1} \otimes M)\widetilde{V}_n(I_{p+1} \otimes M')$, *and* $\widetilde{R}_n \rightsquigarrow (I_{p+1} \otimes M)\widetilde{R}_n(I_{p+1} \otimes M')$.

PROOF OF LEMMA 15.1. We refer to the results of the lemma for $g_i, G_i, \ldots, \widetilde{R}_n$ 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 $\widehat{D}_n = (\widehat{D}_{1n}, \ldots, \widehat{D}_{pn})$, we have $\widehat{D}_{jn} \rightsquigarrow M\widehat{D}_{jn}$ for $j = 1, \ldots, p$ by (24.4) above. In addition, we have $\widetilde{\Xi}_n := (Z_{n\times k}'Z_{n\times k})^{-1}Z_{n\times k}'U^* \rightsquigarrow (MZ_{n\times k}'Z_{n\times k}M')^{-1}MZ_{n\times k}'U^* =$

$M'^{-1}\widetilde{\Xi}_n$. We have $\widehat{u}^*_{in} := \widetilde{\Xi}'_n Z_i \rightsquigarrow (M'^{-1}\widetilde{\Xi}_n)' M Z_i = \widehat{u}^*_{in}$. We have $\widetilde{V}_n := n^{-1}\sum_{i=1}^n [(u^*_i - \widehat{u}^*_{in})(u^*_i - \widehat{u}^*_{in})' \otimes Z_i Z'_i] \rightsquigarrow n^{-1}\sum_{i=1}^n [(u^*_i - \widehat{u}^*_{in})(u^*_i - \widehat{u}^*_{in})' \otimes M Z_i Z'_i M'] = (I_{p+1} \otimes M) \times \widetilde{V}_n (I_{p+1} \otimes M')$ using the invariance of $\widehat{u}^*_{in}$. We have $\widetilde{R}_n := (B' \otimes I_k)\widetilde{V}_n (B \otimes I_k) \rightsquigarrow (B' \otimes M)\widetilde{V}_n (B \otimes M') = (I_{p+1} \otimes M)\widetilde{R}_n (I_{p+1} \otimes M')$ using the equivariance result for $\widetilde{V}_n$.

We have $\widetilde{\Sigma}_{j\ell n} := \operatorname{tr}(\widetilde{R}'_{j\ell n}\widehat{\Omega}_n^{-1})/k \rightsquigarrow \operatorname{tr}((M\widetilde{R}_{j\ell n}M')'(M\widehat{\Omega}_n M')^{-1})/k = \operatorname{tr}(M\widetilde{R}'_{j\ell n}M' \times M'^{-1}\widehat{\Omega}_n^{-1}M^{-1})/k = \widetilde{\Sigma}_{j\ell n}$ for $j, \ell = 1, \ldots, p+1$ using the equivariance result for $\widetilde{R}_n$. We have $\widetilde{L}_n := (\theta, I_p)(\widetilde{\Sigma}^{\varepsilon}_n)^{-1}(\theta, I_p)' \rightsquigarrow \widetilde{L}_n$ using the invariance result for $\widetilde{\Sigma}_n$. We have $\widetilde{D}^{*\prime}_n \widetilde{D}^*_n := \widetilde{L}_n^{1/2}\widehat{D}'_n\widehat{\Omega}_n^{-1}\widehat{D}_n\widetilde{L}_n^{1/2} \rightsquigarrow \widetilde{L}_n^{1/2}\widehat{D}'_n M'(M\widehat{\Omega}_n M')^{-1}M\widehat{D}_n\widetilde{L}_n^{1/2} = \widetilde{D}^{*\prime}_n\widetilde{D}^*_n$. This implies that $c_{k,p}(n^{1/2}\widetilde{D}^*_n, 1-\alpha) \rightsquigarrow c_{k,p}(n^{1/2}\widetilde{D}^*_n, 1-\alpha)$ because $c_{k,p}(n^{1/2}\widetilde{D}^*_n, 1-\alpha)$ only depends on $\widetilde{D}^*_n$ through $\widetilde{D}^{*\prime}_n\widetilde{D}^*_n$ by the comment to Lemma 16.2.

We have AR_n and QLR_{Pn} are invariant by the argument in the paragraph above that contains (24.5). □

25. Proofs of Lemma 16.4 and Proposition 16.5

Lemma 16.4. *Suppose Assumption* WU *holds for some nonempty parameter space $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. Under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_*$,*

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

where (a) *$(\overline{g}_h, \overline{D}_h)$ are defined in* (16.21), (b) *$\overline{\Delta}_h$ is the nonrandom function of h and $\overline{D}_h$ defined in* (16.24), (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:

$$\begin{aligned} &D_n := E_{F_n}G_i, \qquad B_n := B_{F_n}, \qquad C_n := C_{F_n}, \\ &B_n = (B_{n,q}, B_{n,p-q}), \qquad C_n = (C_{n,q}, C_{n,k-q}), \\ &W_n := W_{F_n}, \qquad W_{2n} := W_{2F_n}, \qquad U_n := U_{F_n}, \quad \text{and} \quad U_{2n} := U_{2F_n}, \end{aligned} \tag{25.1}$$

where $q = q_h$ is defined in (16.22), $B_{n,q} \in R^{p\times q}$, $B_{n,p-q} \in R^{p\times(p-q)}$, $C_{n,q} \in R^{k\times q}$, and $C_{n,k-q} \in R^{k\times(k-q)}$. Let

$$\begin{aligned} &\Upsilon_{n,q} := \operatorname{Diag}\{\tau_{1F_n}, \ldots, \tau_{qF_n}\} \in R^{q\times q}, \\ &\Upsilon_{n,p-q} := \operatorname{Diag}\{\tau_{(q+1)F_n}, \ldots, \tau_{pF_n}\} \in R^{(p-q)\times(p-q)} \quad \text{if } k \geq p, \\ &\Upsilon_{n,k-q} := \operatorname{Diag}\{\tau_{(q+1)F_n}, \ldots, \tau_{kF_n}\} \in R^{(k-q)\times(k-q)} \quad \text{if } k < p, \\ &\Upsilon_n := \begin{bmatrix} \Upsilon_{n,q} & 0^{q\times(p-q)} \\ 0^{(p-q)\times q} & \Upsilon_{n,p-q} \\ 0^{(k-p)\times q} & 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times p} \quad \text{if } k \geq p, \quad \text{and} \\ &\Upsilon_n := \begin{bmatrix} \Upsilon_{n,q} & 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ 0^{(k-q)\times q} & \Upsilon_{n,k-q} & 0^{(k-q)\times(p-k)} \end{bmatrix} \in R^{k\times p} \quad \text{if } k < p. \end{aligned} \tag{25.2}$$

As defined, Y_n is the diagonal matrix of singular values of $W_n D_n U_n$; see (16.15).

Proof of Lemma 16.4. The asymptotic distribution of $n^{1/2}(\widehat{g}_n, \mathrm{vec}(\widehat{D}_n - D_n))$ given in Lemma 16.4 follows from the Lyapunov triangular-array multivariate CLT (using the moment restrictions in $\mathcal{F}$) and the following:

$$\begin{aligned} n^{1/2}\mathrm{vec}(\widehat{D}_n - D_n) &= n^{-1/2}\sum_{i=1}^{n}\mathrm{vec}(G_i - D_n) - \begin{pmatrix} \widehat{\Gamma}_{1n} \\ \vdots \\ \widehat{\Gamma}_{pn} \end{pmatrix} \widehat{\Omega}_n^{-1} n^{1/2}\widehat{g}_n \\ &= n^{-1/2}\sum_{i=1}^{n}\left[\mathrm{vec}(G_i - D_n) - \begin{pmatrix} E_{F_n}G_{\ell 1}g_\ell' \\ \vdots \\ E_{F_n}G_{\ell p}g_\ell' \end{pmatrix} \Omega_{F_n}^{-1} g_i\right] + o_p(1), \quad (25.3) \end{aligned}$$

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 $j = 1, \ldots, p$, $n^{-1}\sum_{\ell=1}^{n}\mathrm{vec}(G_\ell)$, 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 $n^{1/2}\mathrm{vec}(\widehat{D}_n - D_n)$ 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 $n^{1/2}\mathrm{vec}(\widehat{D}_n - D_n)$ equals $\lim \mathrm{Var}_{F_n}(\mathrm{vec}(G_i) - (E_{F_n}\mathrm{vec}(G_\ell)g_\ell')\Omega_{F_n}^{-1}g_i) = \lim \Phi_{F_n}^{\mathrm{vec}(G_i)} = \Phi_h^{\mathrm{vec}(G_i)}$, see (16.20), and the limit exists because (i) the components of $\Phi_{F_n}^{\mathrm{vec}(G_i)}$ are comprised of λ_{4,F_n} and submatrices of λ_{5,F_n} and (ii) $\lambda_{s,F_n} \to h_s$ for $s = 4, 5$. By the CLT, the limiting variance matrix of $n^{1/2}\widehat{g}_n$ equals $\lim E_{F_n}g_i g_i' = h_{5,g}$.

The asymptotic distribution of $n^{1/2}W_{F_n}\widehat{D}_n U_{F_n}T_n$ is obtained as follows. Using (16.13)–(16.15), the singular value decomposition of $W_n D_n U_n$ is $W_n D_n U_n = C_n Y_n B_n'$. Using this, we get

$$\begin{aligned} W_n D_n U_n B_{n,q} Y_{n,q}^{-1} = C_n Y_n B_n' B_{n,q} Y_{n,q}^{-1} &= C_n Y_n \begin{pmatrix} I_q \\ 0^{(p-q)\times q} \end{pmatrix} Y_{n,q}^{-1} \\ &= C_n \begin{pmatrix} I_q \\ 0^{(k-q)\times q} \end{pmatrix} = C_{n,q}, \quad (25.4) \end{aligned}$$

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

$$\begin{aligned} W_n \widehat{D}_n U_n B_{n,q} Y_{n,q}^{-1} &= W_n D_n U_n B_{n,q} Y_{n,q}^{-1} + W_n n^{1/2}(\widehat{D}_n - D_n) U_n B_{n,q}\left(n^{1/2} Y_{n,q}\right)^{-1} \\ &= C_{n,q} + o_p(1) \to_p h_{3,q} = \overline{\Delta}_{h,q}, \quad (25.5) \end{aligned}$$

where the second equality uses (among other things) $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \le q$ (by the definition of q in (16.22)). The convergence in (25.5) holds by (16.19), (16.24), and (25.1), and the last equality in (25.5) holds by the definition of $\overline{\Delta}_{h,q}$ in (16.24).

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

$$
\begin{aligned}
n^{1/2} W_n D_n U_n B_{n,p-q} &= n^{1/2} C_n \Upsilon_n B_n' B_{n,p-q} = n^{1/2} C_n \Upsilon_n \begin{pmatrix} 0^{q\times(p-q)} \\ I_{p-q} \end{pmatrix} \\
&= C_n \begin{pmatrix} 0^{q\times(p-q)} \\ n^{1/2}\Upsilon_{n,p-q} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \to h_3 \begin{pmatrix} 0^{q\times(p-q)} \\ \operatorname{Diag}\{h_{1,q+1},\ldots,h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \\
&= h_3 h_{1,p-q}^{\diamond}, \qquad (25.6)
\end{aligned}
$$

where the second equality uses $B_n' B_n = I_p$, the third equality and the convergence hold by (16.19) using the definitions in (16.24) and (25.2) with $k \geq p$, and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (16.24) with $k \geq p$. Analogously, if $k < p$, we have

$$
\begin{aligned}
n^{1/2} W_n D_n U_n B_{n,p-q} &= n^{1/2} C_n \Upsilon_n \begin{pmatrix} 0^{q\times(p-q)} \\ I_{p-q} \end{pmatrix} = C_n \begin{pmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ n^{1/2}\Upsilon_{n,k-q} & 0^{(k-q)\times(p-k)} \end{pmatrix} \\
&\to h_3 \begin{pmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ \operatorname{Diag}\{h_{1,q+1},\ldots,h_{1,k}\} & 0^{(k-q)\times(p-k)} \end{pmatrix} = h_3 h_{1,p-q}^{\diamond}, \qquad (25.7)
\end{aligned}
$$

where the third equality holds by (25.2) with $k < p$ and the last equality holds by the definition of $h_{1,p-q}^{\diamond}$ in (16.24) with $k < p$.

Using (25.6), (25.7), and $n^{1/2}(\widehat{g}_n, \widehat{D}_n - D_n) \to_d (\overline{g}_h, \overline{D}_h)$, we get

$$
\begin{aligned}
n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} &= n^{1/2} W_n D_n U_n B_{n,p-q} + W_n n^{1/2}(\widehat{D}_n - D_n) U_n B_{n,p-q} \\
&\to_d h_3 h_{1,p-q}^{\diamond} + h_{71} \overline{D}_h h_{81} h_{2,p-q} = \overline{\Delta}_{h,p-q}, \qquad (25.8)
\end{aligned}
$$

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 (16.24).

Equations (25.5) and (25.8) combine to establish

$$
\begin{aligned}
n^{1/2} W_n \widehat{D}_n U_n T_n &= n^{1/2} W_n \widehat{D}_n U_n B_n S_n = \left(W_n \widehat{D}_n U_n B_{n,q} \Upsilon_{n,q}^{-1}, n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q}\right) \\
&\to_d (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) = \overline{\Delta}_h \qquad (25.9)
\end{aligned}
$$

using the definition of S_n in (16.23). This completes the proof of the convergence result of Lemma 16.4.

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 (16.21)), and part (b) of the lemma. Part (d) is proved by replacing n by w_n in the proofs above. □

Proposition 16.5. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

(b) *the* (*ordered*) *vector of the smallest* $p-q$ *eigenvalues of* $n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n$, *that is,* $(\widehat{\kappa}_{(q+1)n},\ldots,\widehat{\kappa}_{pn})'$, *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* 16.4, *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 16.5. For the case where $k\geq p$, Proposition 16.5 is the same as Theorem 10.4(c)–(f) given in the SM to AG1, which is proved in Section 17 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 10.4(c)–(f) in AG1 is similar to, but simpler than, the proof of Theorem 16.6, which is given in Section 26 below.

In the second line of the proof of Lemma 17.1 in the SM to AG1, p needs to be replaced by $\min\{k,p\}$ three times.

In the fourth line of (17.3) in the SM to AG1, the $k\times p$ matrix that contains six submatrices needs to be replaced by the following matrix when $k<p$:

$$\begin{bmatrix} h^{\diamond}_{6,r_1^{\diamond}}+o(1) & 0^{r_1^{\diamond}\times(k-r_1^{\diamond})} & 0^{r_1^{\diamond}\times(p-k)} \\ 0^{(k-r_1^{\diamond})\times r_1^{\diamond}} & O(\tau_{r_2F_n}/\tau_{r_1F_n})^{(k-r_1^{\diamond})\times(k-r_1^{\diamond})} & 0^{(k-r_1^{\diamond})\times(p-k)} \end{bmatrix}\in R^{k\times p}, \tag{25.10}$$

where $r_1^{\diamond}$ is defined as in the proof of Lemma 17.1 in the SM to AG1.

In the first line of (17.22) 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 $k<p$:

$$\begin{bmatrix} 0^{r_{g-1}^{\diamond}\times(k-r_{g-1}^{\diamond})} & 0^{r_{g-1}^{\diamond}\times(p-k)} \\ \mathrm{Diag}\{\tau_{r_gF_n},\ldots,\tau_{kF_n}\}/\tau_{r_gF_n} & 0^{(k-r_{g-1}^{\diamond})\times(p-k)} \end{bmatrix}\in R^{k\times(p-r_{g-1}^{\diamond})}. \tag{25.11}$$

The limit of this matrix as $n\to\infty$ equals the matrix given in the second line of (17.22) that contains three submatrices. Thus, the limit of the matrix on the first line of (17.22) is the same for the cases where $k\geq p$ and $k<p$.

In the third line of (17.25) 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 (17.22) 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 (17.22) is changed in (25.11), but with $r_g^{\diamond}$ in place of $r_{g-1}^{\diamond}$.

No other changes are needed. □

26. Proof of Theorem 16.6

Theorem 16.6. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_*\subset\Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h}:n\geq 1\}$ *with* $\lambda_{n,h}\in\Lambda_*$,

$$\mathrm{QLR}_{\mathrm{WU},n}\to_d \overline{g}_h'h_{5,g}^{-1}\overline{g}_h-\lambda_{\min}\big((\overline{\Delta}_{h,p-q},h_{5,g}^{-1/2}\overline{g}_h)'h_{3,k-q}h_{3,k-q}'(\overline{\Delta}_{h,p-q},h_{5,g}^{-1/2}\overline{g}_h)\big)$$

and the convergence holds jointly with the convergence in Lemma 16.4 and Proposition 16.5. 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 16.6 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 $n \to \infty$, possibly at different rates. This complicates the proof considerably. For example, the rate of convergence result of Lemma 26.1(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 (25.1) and (25.2) above. The following definitions are used:

$$
\begin{aligned}
&\widehat{D}_n^+ := \left(\widehat{D}_n, \widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n\right) \in R^{k\times(p+1)}, \qquad \widehat{U}_n^+ := \begin{bmatrix} \widehat{U}_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
&U_n^+ := \begin{bmatrix} U_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \qquad h_{81}^+ := \begin{bmatrix} h_{81} & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
&B_n^+ := \begin{bmatrix} B_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \\
&B_n^+ = \left(B_{n,q}^+, B_{n,p+1-q}^+\right) \quad \text{for } B_{n,q}^+ \in R^{(p+1)\times q} \quad \text{and} \quad B_{n,p+1-q}^+ \in R^{(p+1)\times(p+1-q)}, \\
&D_n^+ := \left(D_n, 0^k\right) \in R^{k\times(p+1)}, \qquad Y_n^+ := \left(Y_n, 0^k\right) \in R^{k\times(p+1)}, \\
&S_n^+ := \mathrm{Diag}\left\{\left(n^{1/2}\tau_{1F_n}\right)^{-1}, \ldots, \left(n^{1/2}\tau_{qF_n}\right)^{-1}, 1, \ldots, 1\right\} \\
&\quad = \begin{bmatrix} S_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},
\end{aligned} \tag{26.1}
$$

where $\widehat{g}_n$ and $\widehat{\Omega}_n$ are defined in (8) with $\theta = \theta_0$, $\widehat{D}_n$ is defined in (18) with $\theta = \theta_0$, $\widehat{W}_n$, $\widehat{U}_n$, U_n $(:= U_{F_n})$, and W_n $(:= W_{F_n})$ are defined in (16.4), h_{81} is defined in (16.24), B_n $(:= B_{F_n})$ is defined in (16.13), D_n is defined in (25.1), Y_n is defined in (25.2), and S_n is defined in (16.23).

Let

$$
\widehat{\kappa}_{jn}^+ \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+, \quad \forall j = 1, \ldots, p+1, \tag{26.2}
$$

ordered to be nonincreasing in j. We have[65]

$$\begin{aligned}\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+ &= \left(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\right) \quad \text{and} \\ \lambda_{\min}\left(n\left(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\right)'\left(\widehat{W}_n\widehat{D}_n\widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\right)\right) &= \lambda_{\min}\left(n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+\right) \\ &= \widehat{\kappa}_{(p+1)n}^+.\end{aligned} \tag{26.3}$$

The proof of Theorem 16.6 uses the following rate of convergence lemma, which is analogous to Lemma 17.1 in Section 17 of the SM to AG1.

LEMMA 26.1. *Suppose Assumption* WU *holds for some nonempty parameter space* $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$ *for which* q *defined in* (16.22) *satisfies* $q \geq 1$, *we have* (a) $\widehat{\kappa}_{jn}^+ \to_p \infty$ *for* $j = 1, \ldots, q$ *and* (b) $\widehat{\kappa}_{jn}^+ = o_p((n^{1/2}\tau_{\ell F_n})^2)$ *for all* $\ell \leq q$ *and* $j = q+1, \ldots, p+1$. *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 holds with* n *replaced with* w_n.

PROOF OF THEOREM 16.6. We have $n^{1/2}\widehat{g}_n \to_d \overline{g}_h$ (by Lemma 16.4) 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, $\mathrm{AR}_n \to_d \overline{g}_h' h_{5,g}^{-1}\overline{g}_h$. Given this, the definition of QLR_n in (16.3) and (26.3), to prove the convergence result in Theorem 16.6, it suffices to show that

$$\begin{aligned}&\lambda_{\min}\left(n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+\right) \\ &\quad \to_d \lambda_{\min}\left(\left(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h\right)' h_{3,k-q}h_{3,k-q}'\left(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h\right)\right).\end{aligned} \tag{26.4}$$

Now we establish (26.4). The eigenvalues $\{\widehat{\kappa}_{jn}^+ : j \leq p+1\}$ of $n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+$ are the ordered solutions to the determinantal equation $|n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+ - \kappa I_{p+1}| = 0$. Equivalently, with probability that goes to one (wp $\to$ 1), they are the solutions to

$$\begin{aligned}&\left|Q_n^+(\kappa)\right| = 0, \quad \text{where} \\ &Q_n^+(\kappa) := nS_n^+B_n^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+U_n^+B_n^+S_n^+ \\ &\qquad - \kappa S_n^+B_n^{+\prime}U_n^{+\prime}\left(\widehat{U}_n^+\right)^{-1\prime}\left(\widehat{U}_n^+\right)^{-1}U_n^+B_n^+S_n^+,\end{aligned} \tag{26.5}$$

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

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

$$\begin{aligned}&B_n^+S_n^+ = \left(B_{n,q}^+S_{n,q}, B_{n,p+1-q}^+\right), \quad \text{where} \\ &S_{n,q} := \mathrm{Diag}\left\{\left(n^{1/2}\tau_{1F_n}\right)^{-1}, \ldots, \left(n^{1/2}\tau_{qF_n}\right)^{-1}\right\} \in R^{q\times q}.\end{aligned} \tag{26.6}$$

[65] In (26.3), 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 (23) 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 16.6 in this section.

The convergence result of Lemma 16.4 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

$$\begin{aligned} n^{1/2}W_n\widehat{D}_n^+U_n^+B_{n,q}^+S_{n,q} &= n^{1/2}W_n\widehat{D}_nU_nB_{n,q}S_{n,q} \to_p \overline{\Delta}_{h,q} := h_{3,q} \quad \text{and} \\ n^{1/2}W_n\widehat{D}_n^+U_n^+B_{n,p+1-q}^+ &= n^{1/2}W_n\big(\widehat{D}_n, \widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n\big)U_n^+B_{n,p+1-q}^+ \\ &= n^{1/2}\big(W_n\widehat{D}_nU_nB_{n,p-q}, W_n\widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n\big) \\ &\to_d \big(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h\big), \end{aligned} \tag{26.7}$$

where $\overline{\Delta}_{h,q}$ and $\overline{\Delta}_{h,p-q}$ are defined in (16.24), $B_{n,p-q}$ is defined in (25.1), and the convergence in distribution uses $\widehat{W}_nW_n^{-1} \to_p I_k$ by (26.8).

We have

$$\widehat{W}_nW_n^{-1} \to_p I_k \quad \text{and} \quad \widehat{U}_n^+\big(U_n^+\big)^{-1} \to_p I_{p+1} \tag{26.8}$$

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

By (26.5)–(26.8), we have

$$\begin{aligned} &Q_n^+(\kappa) \\ &= \begin{bmatrix} I_q + o_p(1) & h_{3,q}'n^{1/2}W_n\widehat{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1) \\ n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n'h_{3,q} + o_p(1) & n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n'W_nn^{1/2}\widehat{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1) \end{bmatrix} \\ &\quad - \kappa\begin{bmatrix} S_{n,q}^2 & 0^{q\times(p+1-q)} \\ 0^{(p+1-q)\times q} & I_{p+1-q} \end{bmatrix} - \kappa\begin{bmatrix} S_{n,q}A_{1n}^+S_{n,q} & S_{n,q}A_{2n}^+ \\ A_{2n}^{+\prime}S_{n,q} & A_{3n}^+ \end{bmatrix}, \quad \text{where} \\ &\widehat{A}_n^+ = \begin{bmatrix} A_{1n}^+ & A_{2n}^+ \\ A_{2n}^{+\prime} & A_{3n}^+ \end{bmatrix} := B_n^{+\prime}U_n^{+\prime}\big(\widehat{U}_n^+\big)^{-1\prime}\big(\widehat{U}_n^+\big)^{-1}U_n^+B_n^+ - I_{p+1} = o_p(1) \end{aligned} \tag{26.9}$$

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 (16.14), (16.16), (16.19), and (16.24)). Note that A_{jn}^+ and $\widehat{A}_{jn}^+$ (defined in (26.19) 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 $\widehat{A}_{1n}^+ \in R^{r_1^\diamond\times r_1^\diamond}$, where $r_1^\diamond$ is defined as in the proof of Lemma 17.1 in the SM to AG1.

If $q = 0$, then $B_n^+ = B_{n,p+1-q}^+$ and

$$\begin{aligned} &nB_n^{+\prime}\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n'\widehat{W}_n\widehat{D}_n^+\widehat{U}_n^+B_n^+ \\ &= nB_n^{+\prime}\big(\big(U_n^+\big)^{-1}\widehat{U}_n^+\big)'\big(B_n^+\big)^{-1\prime}B_n^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n'\big(\widehat{W}_nW_n^{-1}\big)' \\ &\quad \times \big(\widehat{W}_nW_n^{-1}\big)\big(W_n\widehat{D}_n^+U_n^+B_n^+\big)\big(B_n^+\big)^{-1}\big(\big(U_n^+\big)^{-1}\widehat{U}_n^+\big)B_n^+ \\ &\to_d \big(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h\big)'\big(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h\big), \end{aligned} \tag{26.10}$$

where the convergence holds by (26.7) and (26.8) and $\overline{\Delta}_{h,p-q}$ is defined as in (16.24) with $q = 0$. The smallest eigenvalue of a matrix is a continuous function of the matrix (by Elsner's theorem, see Stewart (2001, Theorem 3.1, pp. 37–38)). Hence, the smallest eigenvalue of $nB_n^{+\prime}\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{W}_n^{\prime}\widehat{W}_n\widehat{D}_n^{+}\widehat{U}_n^{+}B_n^{+}$ 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 $h_{3,k-q}h_{3,k-q}' = h_3h_3' = I_k$ when $q = 0$), which proves (26.4) when $q = 0$.

In the remainder of the proof of (26.4), we assume $q \geq 1$, which is the remaining case to be considered in the proof of (26.4). The formula for the determinant of a partitioned matrix and (26.9) give

$$
\begin{aligned}
&\left|Q_n^{+}(\kappa)\right| = \left|Q_{1n}^{+}(\kappa)\right| \cdot \left|Q_{2n}^{+}(\kappa)\right|, \quad \text{where} \\
&Q_{1n}^{+}(\kappa) := I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q}A_{1n}^{+}S_{n,q}, \\
&Q_{2n}^{+}(\kappa) := n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n^{\prime}W_n n^{1/2}\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} \\
&\qquad + o_p(1) - \kappa I_{p+1-q} - \kappa A_{3n}^{+} \\
&\qquad - \left[n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n^{\prime}h_{3,q} + o_p(1) - \kappa A_{2n}^{+\prime}S_{n,q}\right] \\
&\qquad \times \left(I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q}A_{1n}^{+}S_{n,q}\right)^{-1} \\
&\qquad \times \left[h_{3,q}^{\prime}n^{1/2}W_n\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} + o_p(1) - \kappa S_{n,q}A_{2n}^{+}\right],
\end{aligned} \tag{26.11}
$$

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 26.1(b) (which applies for $q \geq 1$), for $j = q+1, \ldots, p+1$, and $A_{1n}^{+} = o_p(1)$ (by (26.9)), 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, for $j = q+1, \ldots, p+1$,

$$
Q_{1n}^{+}\left(\widehat{\kappa}_{jn}^{+}\right) = I_q + o_p(1) - \widehat{\kappa}_{jn}^{+}S_{n,q}^2 - \widehat{\kappa}_{jn}^{+}S_{n,q}A_{1n}^{+}S_{n,q} = I_q + o_p(1). \tag{26.12}
$$

By (26.5) and (26.11), $|Q_n^{+}(\widehat{\kappa}_{jn}^{+})| = |Q_{1n}^{+}(\widehat{\kappa}_{jn}^{+})| \cdot |Q_{2n}^{+}(\widehat{\kappa}_{jn}^{+})| = 0$ for $j = 1, \ldots, p+1$. By (26.12), $|Q_{1n}^{+}(\widehat{\kappa}_{jn}^{+})| \neq 0$ for $j = q+1, \ldots, p+1$ wp $\to 1$. Hence, wp $\to 1$,

$$
\left|Q_{2n}^{+}\left(\widehat{\kappa}_{jn}^{+}\right)\right| = 0 \quad \text{for } j = q+1, \ldots, p+1. \tag{26.13}
$$

Now we plug in $\widehat{\kappa}_{jn}^{+}$ for $j = q+1, \ldots, p+1$ into $Q_{2n}^{+}(\kappa)$ in (26.11) and use (26.12). We have

$$
\begin{aligned}
Q_{2n}^{+}\left(\widehat{\kappa}_{jn}^{+}\right) &= nB_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n^{\prime}W_n\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} + o_p(1) \\
&\quad - \left[n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n^{\prime}h_{3,q} + o_p(1)\right]\left(I_q + o_p(1)\right) \\
&\quad \times \left[h_{3,q}^{\prime}n^{1/2}W_n\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} + o_p(1)\right] \\
&\quad - \widehat{\kappa}_{jn}^{+}\left[I_{p+1-q} + A_{3n}^{+}\right. \\
&\quad - \left(n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n^{\prime}h_{3,q} + o_p(1)\right)\left(I_q + o_p(1)\right)S_{n,q}A_{2n}^{+} \\
&\quad - A_{2n}^{+\prime}S_{n,q}\left(I_q + o_p(1)\right)\left(h_{3,q}^{\prime}n^{1/2}W_n\widehat{D}_n^{+}U_n^{+}B_{n,p+1-q}^{+} + o_p(1)\right) \\
&\quad \left. + \widehat{\kappa}_{jn}^{+}A_{2n}^{+\prime}S_{n,q}\left(I_q + o_p(1)\right)S_{n,q}A_{2n}^{+}\right].
\end{aligned} \tag{26.14}
$$

The term in square brackets on the last three lines of (26.14) that multiplies $\widehat{\kappa}_{jn}^{+}$ equals

$$I_{p+1-q} + o_p(1), \tag{26.15}$$

because $A_{3n}^{+} = o_p(1)$ (by (26.9)), $n^{1/2} W_n \widehat{D}_n^{+} U_n^{+} B_{n,p+1-q}^{+} = O_p(1)$ (by (26.7)), $S_{n,q} = o(1)$ (by the definitions of q and $S_{n,q}$ in (16.22) and (26.6), respectively, and $h_{1,j} := \lim n^{1/2} \tau_{jF_n}$), $A_{2n}^{+} = o_p(1)$ (by (26.9)), 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 (26.14) and (26.15) give

$$\begin{aligned} &Q_{2n}^{+}(\widehat{\kappa}_{jn}^{+}) \\ &\quad = n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' [I_k - h_{3,q} h_{3,q}'] n^{1/2} W_n \widehat{D}_n^{+} U_n^{+} B_{n,p+1-q}^{+} \\ &\qquad + o_p(1) - \widehat{\kappa}_{jn}^{+} [I_{p+1-q} + o_p(1)] \\ &\quad = n^{1/2} B_{n,p+1-q}^{+\prime} U_n^{+\prime} \widehat{D}_n^{+\prime} W_n' h_{3,k-q} h_{3,k-q}' n^{1/2} W_n \widehat{D}_n^{+} U_n^{+} B_{n,p+1-q}^{+} \\ &\qquad + o_p(1) - \widehat{\kappa}_{jn}^{+} [I_{p+1-q} + o_p(1)] \\ &\quad := M_{n,p+1-q}^{+} - \widehat{\kappa}_{jn}^{+} [I_{p+1-q} + o_p(1)], \end{aligned} \tag{26.16}$$

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

Equations (26.13) and (26.16) imply that $\{\widehat{\kappa}_{jn}^{+} : j = q+1, \ldots, p+1\}$ are the $p+1-q$ eigenvalues of the matrix

$$M_{n,p+1-q}^{++} := [I_{p+1-q} + o_p(1)]^{-1/2} M_{n,p+1-q}^{+} [I_{p+1-q} + o_p(1)]^{-1/2} \tag{26.17}$$

by pre- and post-multiplying the quantities in (26.16) by the rhs quantity $[I_{p+1-q} + o_p(1)]^{-1/2}$ in (26.16). By (26.7),

$$M_{n,p+1-q}^{++} \to_d (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h). \tag{26.18}$$

The vector of (ordered) eigenvalues of a matrix is a continuous function of the matrix (by Elsner's theorem; see Stewart (2001, Theorem 3.1, pp. 37–38)). By (26.18), 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., $\{\widehat{\kappa}_{jn}^{+} : j = q+1, \ldots, p+1\}$, converges in distribution to the vector of eigenvalues of the limit matrix $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)$. Hence, $\lambda_{\min}(n \widehat{U}_n^{+\prime} \widehat{D}_n^{+\prime} \widehat{W}_n' \widehat{W}_n \widehat{D}_n^{+} \widehat{U}_n^{+})$, which equals the smallest eigenvalue, $\widehat{\kappa}_{(p+1)n}^{+}$, converges in distribution to the smallest eigenvalue of $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)$, which completes the proof of (26.4).

The previous paragraph proves Comment (v) to Theorem 16.6 for the smallest $p+1-q$ eigenvalues of $n(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)'(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)$. In addition, by Lemma 26.1(a), the largest q eigenvalues of this matrix diverge to infinity in probability, which completes the proof of Comment (v) to Theorem 16.6.

When $q = p$, the third and fourth lines in (26.7) 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, that is, $n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q}$ and $\overline{\Delta}_{h,p-q}$ drop out (because $U_n^+B_{n,p+1-q}^+ = (0^{p\prime}, 1)'$ in this case). In consequence, the limit in (26.18) becomes $(h_{5,g}^{-1/2}\overline{g}_h)'h_{3,k-q}h_{3,k-q}'h_{5,g}^{-1/2}\overline{g}_h$, which has a χ^2_{k-p} 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 16.6 holds jointly with that in Lemma 16.4 and Proposition 16.5 because the results in Proposition 16.5 and Theorem 16.6 just rely on the convergence in distribution of $n^{1/2}W_n\widehat{D}_nU_nT_n$, which is part of Lemma 16.4.

When $q = k$, the $\lambda_{\min}(\cdot)$ expression does not appear in the limit random variable in the statement of Theorem 16.6 because, in the second line of (26.16) 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 $M_{n,p+1-q}^{++} = 0^{(p+1-q)\times(p+1-q)} + o_p(1) \rightarrow_p 0^{(p+1-q)\times(p+1-q)}$ in (26.17) and (26.18).

When $k \leq p$ and $q < k$, the $\lambda_{\min}(\cdot)$ expression (in the limit random variable in the statement of Theorem 16.6) 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 16.6 holds for a subsequence $\{w_n : n \geq 1\}$ of $\{n\}$ by the same proof as given above with n replaced by w_n. □

Proof of Lemma 26.1. The proof of Lemma 26.1 is the same as the proof of Lemma 17.1 in Section 17 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}, Y_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}^+, Y_n^+, $B_{n,r_1^\diamond}^+$, and $B_{n,p+1-r_1^\diamond}^+$, respectively, where

$$\widehat{A}_n^+ = \begin{bmatrix} \widehat{A}_{1n}^+ & \widehat{A}_{2n}^+ \\ \widehat{A}_{2n}^{+\prime} & \widehat{A}_{3n}^+ \end{bmatrix} := (B_n^+)'(U_n^+)'(\widehat{U}_n^+)^{-1\prime}(\widehat{U}_n^+)^{-1}U_n^+B_n^+ - I_{p+1}, \tag{26.19}$$

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 17.1 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 (17.2) 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 (17.8) in the SM to AG1), $\widehat{\xi}_{\ell n}(\kappa)$ for $\ell = 1,2,3$ (defined in (17.9) in the SM to AG1), and $\widehat{A}_{j2n}$ (defined in (17.12) 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 (17.3) in the SM to AG1, which states that $W_n D_n U_n B_n = C_n \Upsilon_n$, is a key equality in the proof of Lemma 17.1 in the SM to AG1. The analogue in the proof of the current lemma is

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

$$= (W_n D_n U_n B_n, 0^k) = (C_n \Upsilon_n, 0^k) = C_n \Upsilon_n^+. \tag{26.20}$$

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. □

27. Proofs of the asymptotic size results

In this section, we prove Theorem 16.1, stated in Section 16.

Theorem 16.1 is proved first for the CQLR and CQLR$_P$ tests and CSs. For these test results, we actually prove a more general result that applies to a CQLR test statistic that is defined as the CQLR test statistic is defined in Section 5, 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 $\Lambda_* \subset \Lambda_{\mathrm{WU}}$ (stated in Section 16.5). Then we show that Assumption WU holds for the parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ for the weight matrices employed by the CQLR and CQLR$_P$ tests, respectively, defined in Sections 5 and 15. These results combine to establish the CQLR and CQLR$_P$ test results of Theorem 16.1. The CQLR and CQLR$_P$ CS results of Theorem 16.1 are proved analogously to those for the tests; see the Comment to Proposition 16.3 for details.

In Section 27.6, we prove Theorem 16.1 for the AR test and CS.

27.1 *Statement of results*

A general QLR$_{\mathrm{WU}}$ test statistic for testing $H_0 : \theta = \theta_0$ is defined in (16.3) as

$$\mathrm{QLR}_{\mathrm{WU},n} := \mathrm{AR}_n - \lambda_{\min}(n\widehat{Q}_{\mathrm{WU},n}), \quad \text{where}$$

$$\widehat{Q}_{\mathrm{WU},n} := (\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)' (\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n), \tag{27.1}$$

AR_n is defined in (18), and the dependence of QLR_n, $\widehat{Q}_{\mathrm{WU},n}$, $\widehat{W}_n$, $\widehat{D}_n$, $\widehat{U}_n$, $\widehat{\Omega}_n$, and $\widehat{g}_n$ on θ_0 is suppressed for notational simplicity.

The general CQLR$_{\mathrm{WU}}$ test rejects the null hypothesis if

$$\mathrm{QLR}_{\mathrm{WU},n} > c_{k,p}(n^{1/2} \widehat{W}_n \widehat{D}_n \widehat{U}_n, 1 - \alpha), \tag{27.2}$$

where $c_{k,p}(D, 1-\alpha)$ is defined just below (24).

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

Theorem 27.1. *Suppose Assumption* WU (*defined in Section* 16.5) *holds for some nonempty parameter space $\Lambda_* \subset \Lambda_{\mathrm{WU}}$. Then the asymptotic null rejection probabilities of the nominal size α* CQLR$_{\mathrm{WU}}$ *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 $\lambda_{w_n,h} \in \Lambda_*$.*

COMMENT. (i) Theorem 27.1 and Proposition 16.3 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 27.4 below, we show that the choice of matrices $(\widehat{W}_n, \widehat{U}_n)$ for the CQLR and CQLR$_P$ tests (defined in Sections 5 and 15, resp.) satisfy Assumption WU for the parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ (defined in (16.17)), respectively. In addition, Lemma 27.4 shows that $\mathcal{F} \subset \mathcal{F}_{\mathrm{WU}}$ and $\mathcal{F}_P \subset \mathcal{F}_{\mathrm{WU}}$ when δ_1 and M_1 that appear in the definition of $\mathcal{F}_{\mathrm{WU}}$ are sufficiently small and large, respectively.[66] In consequence, the CQLR and CQLR$_P$ tests have correct asymptotic size α and are asymptotically similar (in a uniform sense) for the parameter spaces $\mathcal{F}$ and $\mathcal{F}_P$, respectively, as stated in Theorem 16.1.

The proof of Theorem 27.1 uses Proposition 16.5 and Theorem 16.6, 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 27.2. *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) $\tau_{jn}^c \to \tau_{j\infty}^c < \infty$ *for* $j = q+1, \ldots, \min\{k,p\}$. *Then*

$$c_{k,p}(D_n^c, 1-\alpha) \to c_{k,p,q}(\tau_\infty^c, 1-\alpha), \quad \textit{where}$$

$$\tau_\infty^c := (\tau_{(q+1)\infty}^c, \ldots, \tau_{\min\{k,p\}\infty}^c)' \in R^{\min\{k,p\}-q},$$

$$Y(\tau_\infty^c) := \begin{pmatrix} \mathrm{Diag}\{\tau_\infty^c\} \\ 0^{(k-p)\times(p-q)} \end{pmatrix} \in R^{(k-q)\times(p-q)} \quad \textit{if } k \geq p,$$

$$Y(\tau_\infty^c) := (\mathrm{Diag}\{\tau_\infty^c\}, 0^{(k-q)\times(p-k)}) \in R^{(k-q)\times(p-q)} \quad \textit{if } k < p,$$

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

$$\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) := Z'Z - \lambda_{\min}((Y(\tau_\infty^c), Z_2)'(Y(\tau_\infty^c), Z_2)), \quad \textit{and}$$

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

COMMENT. (i) The matrix $Y(\tau_\infty^c)$ is the diagonal matrix containing the $\min\{k,p\} - q$ finite limiting eigenvalues of D_n^c. Note that $Y(\tau_\infty^c)$ has only $k-q$ rows, not k rows.

(ii) If $q = p$ (which requires that $k \geq p$), then $Y(\tau_\infty^c)$ has no columns, $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) = Z_1'Z_1 \sim \chi_p^2$, and $c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ equals the $1-\alpha$ quantile of the χ_p^2 distribution.

[66]Note that the set of distributions $\mathcal{F}_{\mathrm{WU}}$ depends on the definitions of (W_F, U_F) (see (16.12)), and (W_F, U_F) are defined differently for the QLR and QLR$_2$ statistics; see (16.6)–(16.8) and (16.9)–(16.11), respectively. Hence, the set of distributions $\mathcal{F}_{\mathrm{WU}}$ differs for the CQLR and CQLR$_2$ tests.

(iii) If $q = k$ (which requires that $k \leq p$), then $Y(\tau_\infty^c)$ and Z_2 have no rows, the $\lambda_{\min}(\cdot)$ expression in $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c)$ disappears, $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) = Z'Z \sim \chi_k^2$, and $c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ is the $1-\alpha$ quantile of the χ_k^2 distribution.

(iv) If $k \leq p$ and $q < k$, then $(Y(\tau_\infty^c), Z_2)$ has fewer rows $(k-q)$ than columns $(p-q+1)$ and, hence, the $\lambda_{\min}(\cdot)$ expression in $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c)$ equals zero, $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) = Z'Z \sim \chi_k^2$, and $c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ is the $1-\alpha$ quantile of the χ_k^2 distribution.

(v) The distribution function (df) of $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c)$ is shown in Lemma 27.3 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 27.2.

The following lemma proves that the df of $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c)$, defined in Lemma 27.2, is continuous and strictly increasing at its $1-\alpha$ quantile. This is a key lemma for showing that the CQLR and CQLR_P tests have correct asymptotic size and are asymptotically similar.

Lemma 27.3. *Let τ_∞^c and $Y(\tau_\infty^c)$ be defined as in Lemma 27.2. 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 nonnegative elements in nonincreasing order, the df of $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) := Z'Z - \lambda_{\min}((Y(\tau_\infty^c), Z_2)'(Y(\tau_\infty^c), Z_2))$ is continuous and strictly increasing at its $1-\alpha$ quantile $c_{k,p,q}(\tau_\infty^c, 1-\alpha)$ 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 and CQLR_P tests. Part (a) of the lemma shows that $\mathcal{F}_{\mathrm{WU}}$, when defined for $(\widehat{W}_n, \widehat{U}_n)$ as in the CQLR test, contains $\mathcal{F}$ for suitable choices of the constants δ_1 and M_1 that appear in the definition of $\mathcal{F}_{\mathrm{WU}}$. Part (b) of the lemma shows that the parameter space $\mathcal{F}_{\mathrm{WU}}$, when defined for $(\widehat{W}_n, \widehat{U}_n)$ as in the CQLR_P test, contains the parameter space $\mathcal{F}_P$ for suitable constants δ_1 and M_1.

Lemma 27.4. (a) *Suppose $(\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widehat{L}_n^{1/2})$, where $\widehat{\Omega}_n$ $(= \widehat{\Omega}_n(\theta_0))$ and $\widehat{L}_n$ $(= \widehat{L}_n(\theta_0))$ are defined in* (8) *and* (23). *Then* (i) *Assumption* WU *holds for the parameter space Λ_{WU} with $(\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widehat{R}_n))$ for $\widehat{R}_n$ defined in* (19), *$W_1(W_2) = W_2^{-1/2}$ for $W_2 \in R^{k\times k}$, $U_1(U_{2F}) = ((\theta_0, I_p)(\Sigma^{\varepsilon}(\Omega_F, R_F))^{-1}(\theta_0, I_p)')^{1/2}$ 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 $\Omega_F := E_F g_i g_i'$, R_F is defined in* (16.7), *$\Sigma(\Omega_F, R_F)$ is defined in* (16.8), *and $\Sigma^{\varepsilon}(\Omega_F, R_F)$ is defined given $\Sigma(\Omega_F, R_F)$ by* (22), *and* (ii) *$\mathcal{F} = \mathcal{F}_{\mathrm{WU}}$ for δ_1 sufficiently small and M_1 sufficiently large in the definition of $\mathcal{F}_{\mathrm{WU}}$, where $\mathcal{F}$ is defined in* (16.1) *and $\mathcal{F}_{\mathrm{WU}}$ is defined in* (16.12).

(b) *Suppose $g_i(\theta) = u_i(\theta)Z_i$, as in* (15.1), *and $(\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widetilde{L}_n^{1/2})$, where $\widehat{\Omega}_n$ $(= \widehat{\Omega}_n(\theta_0))$ and $\widetilde{L}_n$ $(= \widetilde{L}_n(\theta_0))$ are defined in* (8) *and* (15.6), *respectively. Then* (i) *Assumption* WU *holds for the parameter space $\Lambda_{\mathrm{WU},P}$ with $(\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widetilde{R}_n))$ for $\widetilde{R}_n$ defined in* (15.5), *$W_1(W_2) = W_2^{-1/2}$ for $W_2 \in R^{k\times k}$, $U_1(U_{2F}) = ((\theta_0, I_p)(\widetilde{\Sigma}^{\varepsilon}(\Omega_F, \widetilde{R}_F))^{-1}(\theta_0, I_p)')^{1/2}$ for $U_{2F} = (\Omega_F, \widetilde{R}_F)$, $h_7 = \lim W_{2F_{w_n}} := \lim \Omega_{F_{w_n}}$, and $h_8 = \lim U_{2F_{w_n}} := \lim(\Omega_{F_{w_n}}, \widetilde{R}_{F_{w_n}})$, where $\Omega_F := E_F g_i g_i'$, $\widetilde{\Sigma}_F := \Sigma(\Omega_F, \widetilde{R}_F)$ is defined in* (16.11), *$\widetilde{\Sigma}^{\varepsilon}(\Omega_F, \widetilde{R}_F)$ is defined given*

$\Sigma(\Omega_F, \widetilde{R}_F)$ by (22), *and $\widetilde{R}_F$ is defined in* (16.10), *and* (ii) $\mathcal{F}_P \subset \mathcal{F}_{\mathrm{WU}}$ *for δ_1 sufficiently small and M_1 sufficiently large in the definition of $\mathcal{F}_{\mathrm{WU}}$, where $\mathcal{F}_P$ is defined in* (16.1) *and $\mathcal{F}_{\mathrm{WU}}$ is defined in* (16.12).

COMMENT. Theorem 27.1, Lemma 27.4, and Proposition 16.3 combine to prove the CQLR and CQLR$_P$ test results of Theorem 16.1, which state that the CQLR and CQLR$_P$ tests have correct asymptotic size and are asymptotically similar (in a uniform sense) for the parameter spaces $\mathcal{F}$ and $\mathcal{F}_P$, respectively. As stated at the beginning of this section, the proofs of the CQLR and CQLR$_P$ CS results of Theorem 16.1 are analogous to those for the tests; see the Comment to Proposition 16.3, and hence, are not stated explicitly.

27.2 *Proof of Theorem 27.1*

Theorem 27.1 is stated in Section 27.1.

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 27.1. Let

$$\overline{Z}_h = \begin{pmatrix} \overline{Z}_{h1} \\ \overline{Z}_{h2} \end{pmatrix} := \begin{pmatrix} h_{3,q}' h_{5,g}^{-1/2} \overline{g}_h \\ h_{3,k-q}' h_{5,g}^{-1/2} \overline{g}_h \end{pmatrix} = h_3' h_{5,g}^{-1/2} \overline{g}_h \sim N(0^k, I_k), \tag{27.3}$$

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 (16.21)) 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 16.4(c)).

By Theorem 16.6,

$$\begin{aligned} \mathrm{QLR}_{\mathrm{WU},n} &\to_d \overline{g}_h' h_{5,g}^{-1} \overline{g}_h - \lambda_{\min}\big((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2}\overline{g}_h)\big) \\ &= \overline{Z}_h' \overline{Z}_h - \lambda_{\min}\big((h_{3,k-q}'\overline{\Delta}_{h,p-q}, \overline{Z}_{h2})'(h_{3,k-q}'\overline{\Delta}_{h,p-q}, \overline{Z}_{h2})\big) =: \overline{\mathrm{QLR}}_h, \end{aligned} \tag{27.4}$$

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

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

$$\begin{aligned} &\widehat{\tau}_n = (\widehat{\tau}_{[1]n}', \widehat{\tau}_{[2]n}')' \in R^{\min\{k,p\}}, \quad \text{where} \\ &\widehat{\tau}_{[1]n} = (\widehat{\tau}_{1n}, \ldots, \widehat{\tau}_{qn})' \in R^q \quad \text{and} \quad \widehat{\tau}_{[2]n} = (\widehat{\tau}_{(q+1)n}, \ldots, \widehat{\tau}_{\min\{k,p\}n})' \in R^{\min\{k,p\}-q}. \end{aligned} \tag{27.5}$$

By Proposition 16.5(a) and (b), $\widehat{\tau}_{jn} \to_p \infty$ for $j \leq q$ (or, equivalently $\mathrm{Diag}^{-1}\{\widehat{\tau}_{[1]n}\} \to_p 0^{q \times q}$) and

$$\widehat{\tau}_{[2]n} \to_d \overline{\tau}_{[2]h}, \tag{27.6}$$

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, for example, see Pollard (1990, Theorem 9.4, p. 45), there exists a probability space, say $(\Omega^0, \mathcal{F}^0, P^0)$, and random variables $(\mathrm{QLR}_n^0, \widehat{\tau}_n^{0\prime}, \overline{\mathrm{QLR}}_h^0, \overline{\tau}_{[2]h}^{0\prime})'$ defined on it such that $(\mathrm{QLR}_n^0, \widehat{\tau}_n^{0\prime})'$ has the same distribution as $(\mathrm{QLR}_{\mathrm{WU},n}, \widehat{\tau}_n')'$ for all $n \geq 1$, $(\overline{\mathrm{QLR}}_h^0, \overline{\tau}_{[2]h}^{0\prime})'$ has the same distribution as $(\overline{\mathrm{QLR}}_h, \overline{\tau}_{[2]h}')'$, and

$$\begin{pmatrix} \mathrm{QLR}_n^0 \\ \mathrm{Diag}^{-1}\{\widehat{\tau}_{[1]n}^0\} \\ \widehat{\tau}_{[2]n}^0 \end{pmatrix} \to \begin{pmatrix} \overline{\mathrm{QLR}}_h^0 \\ 0^{q\times q} \\ \overline{\tau}_{[2]h}^0 \end{pmatrix} \quad \text{a.s.,} \tag{27.7}$$

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

$$\begin{aligned} &\widehat{Y}_n^0 := \begin{pmatrix} \mathrm{Diag}\{\widehat{\tau}_n^0\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \quad \text{and} \quad \widehat{Y}_n := \begin{pmatrix} \mathrm{Diag}\{\widehat{\tau}_n\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \quad \text{if } k \geq p \quad \text{and} \\ &\widehat{Y}_n^0 := \left(\mathrm{Diag}\{\widehat{\tau}_n^0\}, 0^{k\times(p-k)}\right) \in R^{k\times p} \quad \text{and} \\ &\widehat{Y}_n := \left(\mathrm{Diag}\{\widehat{\tau}_n\}, 0^{k\times(p-k)}\right) \in R^{k\times p} \quad \text{if } k < p. \end{aligned} \tag{27.8}$$

The distributions of $\widehat{Y}_n^0$ and $\widehat{Y}_n$ are the same. The matrix $\widehat{Y}_n^0$ has singular values given by the vector $\widehat{\tau}_n^0$ $(= (\widehat{\tau}_{1n}^0, \ldots, \widehat{\tau}_{\min\{k,p\}n}^0)')$ whose first q elements all diverge to infinity a.s. and whose last $\min\{k, p\} - q$ elements written as the subvector $\widehat{\tau}_{[2]n}^0$ converge to $\overline{\tau}_{[2]h}^0$ a.s. Hence, for some set $C \in \mathcal{F}^0$ with $P^0(\omega \in C) = 1$, we have $\widehat{\tau}_{jn}^0(\omega) \to \infty$ for $j \leq q$ and $\widehat{\tau}_{[2]n}^0(\omega) \to \overline{\tau}_{[2]h}^0(\omega)$, where $\widehat{\tau}_{jn}^0(\omega)$, $\widehat{\tau}_{[2]n}^0(\omega)$, $\overline{\tau}_{[2]h}^0(\omega)$, and $\widehat{Y}_n^0(\omega)$ denote the realizations of the random quantities $\widehat{\tau}_{jn}^0$, $\widehat{\tau}_{[2]n}^0$, $\overline{\tau}_{[2]h}^0$, and $\widehat{Y}_n^0$, respectively, when ω occurs. Thus, using Lemma 27.2 with $D_n^c = \widehat{Y}_n^0(\omega)$ and $\tau_\infty^c = \overline{\tau}_{[2]h}^0(\omega)$, we have

$$c_{k,p}\left(\widehat{Y}_n^0(\omega), 1-\alpha\right) \to c_{k,p,q}\left(\overline{\tau}_{[2]h}^0(\omega), 1-\alpha\right) \quad \text{for all } \omega \in C \text{ with } P^0(\omega \in C) = 1, \tag{27.9}$$

where $c_{k,p,q}(\cdot, 1-\alpha)$ is defined in Lemma 27.2. When $q = \min\{k, p\}$, no vector $\overline{\tau}_{[2]h}^0(\omega)$ appears and by Comments (ii) and (iii) to Lemma 27.2 $c_{k,p,q}(\overline{\tau}_{[2]h}^0(\omega), 1-\alpha)$ equals the $1-\alpha$ quantile of the $\chi^2_{\min\{k,p\}}$ distribution.

Almost sure convergence implies convergence in distribution, so (27.7) and (27.9) 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 $(\mathrm{QLR}_n^0, \widehat{Y}_n^0)$ and $(\mathrm{QLR}_{\mathrm{WU},n}, \widehat{Y}_n)$ for all $n \geq 1$ and of $(\overline{\mathrm{QLR}}_h^0, \overline{\tau}_{[2]h}^{0\prime})'$ and $(\overline{\mathrm{QLR}}_h, \overline{\tau}_{[2]h}')'$, yield the following convergence result:

$$\begin{pmatrix} \mathrm{QLR}_{\mathrm{WU},n} \\ c_{k,p}\left(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha\right) \end{pmatrix} = \begin{pmatrix} \mathrm{QLR}_{\mathrm{WU},n} \\ c_{k,p}(\widehat{Y}_n, 1-\alpha) \end{pmatrix} \to_d \begin{pmatrix} \overline{\mathrm{QLR}}_h \\ c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) \end{pmatrix}, \tag{27.10}$$

where the first equality holds using Lemma 16.2.

Equation (27.10) and the continuous mapping theorem give

$$P\left(\mathrm{QLR}_{\mathrm{WU},n} > c_{k,p}\left(n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n, 1-\alpha\right)\right) \to P\left(\overline{\mathrm{QLR}}_h > c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)\right) \tag{27.11}$$

provided $P(\overline{\text{QLR}}_h = c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)) = 0$. The latter holds because $P(\overline{\text{QLR}}_h = c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)|\overline{D}_h) = 0$ a.s. In turn, the latter holds because, conditional on $\overline{D}_h$, the df of $\overline{\text{QLR}}_h$ is continuous at its $1-\alpha$ quantile (by Lemma 27.3, where $\overline{\text{QLR}}_h$ conditional on $\overline{D}_h$ and $\text{ACLR}_{k,p,q}(\tau^c_\infty)$, which appears in Lemma 27.3, have the same structure with the former being based on $h'_{3,k-q}\overline{\Delta}_{h,p-q}$, which is nonrandom conditional on $\overline{D}_h$, and the latter being based on $\Upsilon(\tau^c_\infty)$, which is nonrandom, and the former only depends on $h'_{3,k-q}\overline{\Delta}_{h,p-q}$ through its singular values (see (24.3)) 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 16.2,

$$c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha) = c_{k,p,q}\big(h'_{3,k-q}\overline{\Delta}_{h,p-q}, 1-\alpha\big), \tag{27.12}$$

where (with some abuse of notation) $c_{k,p,q}(h'_{3,k-q}\overline{\Delta}_{h,p-q}, 1-\alpha)$ denotes the $1-\alpha$ quantile of $Z'Z - \lambda_{\min}((h'_{3,k-q}\overline{\Delta}_{h,p-q}, Z_2)'(h'_{3,k-q}\overline{\Delta}_{h,p-q}, Z_2))$ for Z as in Lemma 27.2, because $\overline{\tau}_{[2]h} \in R^{p-q}$ are the singular values of $h'_{3,k-q}\overline{\Delta}_{h,p-q} \in R^{(k-q)\times(p-q)}$ and $\Upsilon(\overline{\tau}_{[2]h})$ (which appears in $\text{ACLR}_{k,p,q}(\overline{\tau}_{[2]h}) = Z'Z - \lambda_{\min}((\Upsilon(\overline{\tau}_{[2]h}), Z_2)'(\Upsilon(\overline{\tau}_{[2]h}), Z_2)))$ is the $(k-q)\times(p-q)$ matrix with $\overline{\tau}_{[2]h}$ on the main diagonal and zeros elsewhere.

Thus, we have

$$\begin{aligned}
&P\big(\overline{\text{QLR}}_h > c_{k,p,q}(\overline{\tau}_{[2]h}, 1-\alpha)\big)\\
&\quad = P\big(\overline{\text{QLR}}_h > c_{k,p,q}\big(h'_{3,k-q}\overline{\Delta}_{h,p-q}, 1-\alpha\big)\big)\\
&\quad = EP\big(\overline{\text{QLR}}_h > c_{k,p,q}\big(h'_{3,k-q}\overline{\Delta}_{h,p-q}, 1-\alpha\big)|\overline{\Delta}_{h,p-q}\big)\\
&\quad = E\alpha = \alpha,
\end{aligned} \tag{27.13}$$

where the second equality holds by the law of iterated expectations and the third equality holds because, conditional on $\overline{\Delta}_{h,p-q}$, $c_{k,p,q}(h'_{3,k-q}\overline{\Delta}_{h,p-q}, 1-\alpha)$ is the $1-\alpha$ quantile of $\overline{\text{QLR}}_h$ (by the definitions of $c_{k,p,q}(\cdot, 1-\alpha)$ in Lemma 27.2 and $\overline{\text{QLR}}_h$ in (27.4)) and the df of $\overline{\text{QLR}}_h$ is continuous at its $1-\alpha$ quantile (see the explanation following (27.11)). □

27.3 *Proof of Lemma 27.2*

Lemma 27.2 is stated in Section 27.1.

The proof of Lemma 27.2 uses the following two lemmas. Let $\{\tau^c_{jn} : j \le \min\{k,p\}\}$ be the singular values of D^c_n, as in Lemma 27.2. Define

$$\begin{aligned}
\Upsilon^c_n &:= \begin{pmatrix} \text{Diag}\{\tau^c_{1n}, \ldots, \tau^c_{pn}\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p} \quad \text{if } k \ge p \quad \text{and}\\
\Upsilon^c_n &:= \big(\text{Diag}\{\tau^c_{1n}, \ldots, \tau^c_{kn}\}, 0^{k\times(p-k)}\big) \in R^{k\times p} \quad \text{if } k < p.
\end{aligned} \tag{27.14}$$

LEMMA 27.5. *Suppose the scalar constants* $\{\tau^c_{jn} \ge 0 : j \le \min\{k,p\}\}$ *for* $n \ge 1$ *satisfy* (i) $\{\tau^c_{jn} \ge 0 : j \le \min\{k,p\}\}$ *are nonincreasing in* j *for* $n \ge 1$, (ii) $\tau^c_{jn} \to \infty$ *for* $j \le q$ *for some* $1 \le q \le \min\{k,p\}$, (iii) $\tau^c_{jn} \to \tau^c_{j\infty} < \infty$ *for* $j = q+1, \ldots, \min\{k,p\}$, *and* (iv) *when* $p \ge 2$,

$\tau^c_{(j+1)n}/\tau^c_{jn} \to h^c_{6,j}$ *for some* $h^c_{6,j} \in [0,1]$ *for all* $j \le \min\{k,p\}-1$. *Let* Y^c_n *be defined as in* (27.14). *Let* $\{\kappa^Z_{jn} : j \le p+1\}$ *denote the* $p+1$ *eigenvalues of* $(Y^c_n, Z)'(Y^c_n, Z)$, *ordered to be nonincreasing in* j, *where* $Z \sim N(0^k, I_k)$. *Then,*

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

(b) $\kappa^Z_{jn} = o((\tau^c_{\ell n})^2)\ \forall \ell \le q$ *and* $\forall j = q+1, \ldots, p+1$ *for all realizations of* Z.

Comment. Lemma 27.5 only applies when $q \ge 1$, whereas Lemma 27.2 applies when $q \ge 0$.

Lemma 27.6. *Let* $\{F^*_n(x) : n \ge 1\}$ *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* $q_\infty := \inf\{x : F^*(x) \ge 1-\alpha\}$ *is the* $1-\alpha$ *quantile of* $F^*(x)$. *Then the* $1-\alpha$ *quantile of* $F^*_n(x)$, *viz.,* $q_n := \inf\{x : F^*_n(x) \ge 1-\alpha\}$, *satisfies* $q_n \to q_\infty$.

Comment. Condition (ii) of Lemma 27.6 requires that $F^*(x)$ is increasing at its $1-\alpha$ quantile.

Proof of Lemma 27.2. By Lemma 16.2, $c_{k,p}(D^c_n, 1-\alpha) = c_{k,p}(Y^c_n, 1-\alpha)$, where Y^c_n is defined in (27.14). Hence, it suffices to show that $c_{k,p}(Y^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$. To prove the latter, it suffices to show that for any subsequence $\{w_n\}$ of $\{n\}$ there exists a subsubsequence $\{u_n\}$ such that $c_{k,p}(Y^c_{u_n}, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$. When $p \ge 2$, given $\{w_n\}$, we select a subsubsequence $\{u_n\}$ for which $\tau^c_{(j+1)u_n}/\tau^c_{ju_n} \to h^c_{6,j}$ for some constant $h^c_{6,j} \in [0,1]$ for all $j = 1, \ldots, \min\{k,p\}-1$ (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 $p \ge 2$, we prove the full sequence result that $c_{k,p}(Y^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ under the assumption that

$$\tau^c_{(j+1)n}/\tau^c_{jn} \to h^c_{6,j} \quad \text{for all } j \le \min\{k,p\}-1 \tag{27.15}$$

(as well as the other assumptions on the singular values stated in the theorem).[67] 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 $c_{k,p}(Y^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$ without the condition in (27.15) (which is meaningless in this case because there is only one value $\tau^c_{ju_n}$, namely $\tau^c_{1u_n}$, for each n). In this case, too, the same argument holds with n replaced by u_n below, which is the result that is needed to complete the proof. We treat the cases $p \ge 2$ and $p = 1$ simultaneously from here on.

First, we show that

$$\begin{aligned} \mathrm{CLR}_{k,p}(Y^c_n) &:= Z'Z - \lambda_{\min}\big((Y^c_n, Z)'(Y^c_n, Z)\big) \\ &\to Z'Z - \lambda_{\min}\big((Y(\tau^c_\infty), Z_2)'(Y(\tau^c_\infty), Z_2)\big) := \mathrm{ACLR}_{k,p,q}(\tau^c_\infty) \end{aligned} \tag{27.16}$$

[67] The condition in (27.15) is required by Lemma 27.5, which is used in the proof of Lemma 27.2 below.

for all realizations of Z. If $q=0$, then (27.16) holds because $Y_n^c \to Y(\tau_\infty^c)$ (by the definition of Y_n^c in (27.14), the definition of $Y(\tau_\infty^c)$ in the statement of the Lemma 27.2, and assumption (iii) of Lemma 27.2) and the minimum eigenvalue of a matrix is a continuous function of the matrix (by Elsner's theorem, see Stewart (2001, Theorem 3.1, pp. 37–38)).

Now, we establish (27.16) when $q \geq 1$. The (ordered) eigenvalues $\{\kappa_{jn}^Z : j \leq p+1\}$ of $(Y_n^c, Z)'(Y_n^c, Z)$ are solutions to

$$\begin{aligned}
&\left|(Y_n^c, Z)'(Y_n^c, Z) - \kappa I_{p+1}\right| = 0 \quad \text{or} \\
&\qquad \left|Q_n^c(\kappa)\right| = 0, \quad \text{where} \\
&\qquad Q_n^c(\kappa) := S_n^c(Y_n^c, Z)'(Y_n^c, Z)S_n^c - \kappa(S_n^c)^2 \quad \text{and} \\
&\qquad S_n^c := \mathrm{Diag}\{(\tau_{1n}^c)^{-1}, \ldots, (\tau_{qn}^c)^{-1}, 1, \ldots, 1\} \\
&\qquad\quad \in R^{(p+1)\times(p+1)}.
\end{aligned} \tag{27.17}$$

Define

$$S_{n,q}^c := \mathrm{Diag}\{(\tau_{1n}^c)^{-1}, \ldots, (\tau_{qn}^c)^{-1}\} \in R^{q\times q}. \tag{27.18}$$

We have

$$\begin{aligned}
(Y_n^c, Z)S_n^c &= \left((Y_n^c, Z)\begin{pmatrix} I_q \\ 0^{(p+1-q)\times q}\end{pmatrix} S_{n,q}^c, (Y_n^c, Z)\begin{pmatrix} 0^{q\times(p+1-q)} \\ I_{p+1-q}\end{pmatrix}\right) \\
&= (I_{k,q}, Y_{n,p-q}^c, Z) \in R^{k\times(p+1)}, \quad \text{where} \\
I_{k,q} &:= \begin{pmatrix} I_q \\ 0^{(k-q)\times q}\end{pmatrix} \in R^{k\times q}, \\
Y_{n,p-q}^c &:= \begin{pmatrix} 0^{q\times(p-q)} \\ \mathrm{Diag}\{\tau_{(q+1)n}^c, \ldots, \tau_{pn}^c\} \\ 0^{(k-p)\times(p-q)}\end{pmatrix} \in R^{k\times(p-q)} \quad \text{if } k \geq p, \quad \text{and} \\
Y_{n,p-q}^c &:= \begin{pmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ \mathrm{Diag}\{\tau_{(q+1)n}^c, \ldots, \tau_{kn}^c\} & 0^{(k-q)\times(p-k)}\end{pmatrix} \in R^{k\times(p-q)} \quad \text{if } k < p.
\end{aligned} \tag{27.19}$$

By (27.17) and (27.19), we have

$$\begin{aligned}
Q_n^c(\kappa) = &\begin{bmatrix} I_q & I_{k,q}'(Y_{n,p-q}^c, Z) \\ (Y_{n,p-q}^c, Z)'I_{k,q} & (Y_{n,p-q}^c, Z)'(Y_{n,p-q}^c, Z)\end{bmatrix} \\
&- \kappa\begin{bmatrix} (S_{n,q}^c)^2 & 0^{q\times(p+1-q)} \\ 0^{(p+1-q)\times q} & I_{p+1-q}\end{bmatrix}.
\end{aligned} \tag{27.20}$$

By the formula for the determinant of a partitioned inverse,

$$
\begin{aligned}
&\left|Q_n^c(\kappa)\right| = \left|Q_{n,1}^c(\kappa)\right| \cdot \left|Q_{n,2}^c(\kappa)\right|, \quad \text{where} \\
&Q_{n,1}^c(\kappa) := I_q - \kappa\big(S_{n,q}^c\big)^2 \in R^{q\times q} \quad \text{and} \\
&Q_{n,2}^c(\kappa) := \big(Y_{n,p-q}^c, Z\big)'\big(Y_{n,p-q}^c, Z\big) - \kappa I_{p+1-q} \\
&\qquad - \big(Y_{n,p-q}^c, Z\big)' I_{k,q}\big(I_q - \kappa\big(S_{n,q}^c\big)^2\big)^{-1} I_{k,q}'\big(Y_{n,p-q}^c, Z\big) \\
&\qquad \in R^{(p+1-q)\times(p+1-q)}.
\end{aligned} \tag{27.21}
$$

For $j = q+1, \dots, p+1$, we have

$$
Q_{n,1}^c\big(\kappa_{jn}^Z\big) = I_q - \kappa_{jn}^Z\big(S_{n,q}^c\big)^2 = I_q - \text{Diag}\big\{\kappa_{jn}^Z\big(\tau_{1n}^c\big)^{-2}, \dots, \kappa_{jn}^Z\big(\tau_{qn}^c\big)^{-2}\big\} = I_q + o(1) \tag{27.22}
$$

for all realizations of Z, where the last equality holds by Lemma 27.5 (which applies for $q \geq 1$). This implies that $|Q_{n,1}^c(\kappa_{jn}^Z)| \neq 0$ for $j = q+1, \dots, p+1$ for n large. Hence, for n large,

$$
\left|Q_{n,2}^c\big(\kappa_{jn}^Z\big)\right| = 0 \quad \text{for } j = q+1, \dots, p+1. \tag{27.23}
$$

We write

$$
I_k = (I_{k,q}, I_{k,k-q}), \quad \text{where } I_{k,k-q} := \begin{pmatrix} 0^{q\times(k-q)} \\ I_{k-q} \end{pmatrix} \in R^{k\times(k-q)} \tag{27.24}
$$

and $I_{k,q}$ is defined in (27.19).[68]

For $j = q+1, \dots, p+1$, we have

$$
\begin{aligned}
Q_{n,2}^c\big(\kappa_{jn}^Z\big) &= \big(Y_{n,p-q}^c, Z\big)'\big(Y_{n,p-q}^c, Z\big) - \kappa_{jn}^Z I_{p+1-q} \\
&\quad - \big(Y_{n,p-q}^c, Z\big)' I_{k,q}\big(I_q + o(1)\big) I_{k,q}'\big(Y_{n,p-q}^c, Z\big) \\
&= \big(Y_{n,p-q}^c, Z\big)' I_{k,k-q} I_{k,k-q}'\big(Y_{n,p-q}^c, Z\big) + o(1) - \kappa_{jn}^Z I_{p+1-q} \\
&:= M_{n,p+1-q}^c - \kappa_{jn}^Z I_{p+1-q},
\end{aligned} \tag{27.25}
$$

where the first equality holds by (27.22) and the definition of $Q_{n,2}^c(\kappa)$ in (27.21) 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 $Y_{n,p-q}^c = O(1)$ by its definition in (27.19) and the condition (iii) of Lemma 27.2 on $\{\tau_{jn}^c : j = q+1, \dots, \min\{k, p\}\}$ for $n \geq 1$.

Equations (27.23) and (27.25) imply that $\{\kappa_{jn}^Z : j = q+1, \dots, p+1\}$ are the $p+1-q$ eigenvalues of the matrix $M_{n,p+1-q}^c$. By the definition of $Y_{n,p-q}^c$ in (27.19) and the condi-

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

tions of the lemma on $\{\tau^c_{jn} : j = q+1, \ldots, \min\{k, p\}\}$ for $n \geq 1$, we have

$$\begin{aligned} M^c_{n,p+1-q} &\to \left(\begin{pmatrix} 0^{q\times(p-q)} \\ Y(\tau^c_\infty) \end{pmatrix}, Z\right)' I_{k,k-q} I'_{k,k-q} \left(\begin{pmatrix} 0^{q\times(p-q)} \\ Y(\tau^c_\infty) \end{pmatrix}, Z\right) \\ &= \big(Y(\tau^c_\infty), Z_2\big)'\big(Y(\tau^c_\infty), Z_2\big) \end{aligned} \tag{27.26}$$

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

The vector of (ordered) eigenvalues of a matrix is a continuous function of the matrix (by Elsner's theorem, see Stewart (2001, Theorem 3.1, pp. 37–38)). Hence, by (27.26), the eigenvalues $\{\kappa^Z_{jn} : j = q+1, \ldots, p+1\}$ of $M^c_{n,p+1-q}$ converge (for all realizations of Z) to the vector of eigenvalues of $(Y(\tau^c_\infty), Z_2)'(Y(\tau^c_\infty), Z_2)$. In consequence, the smallest eigenvalue $\kappa^Z_{(p+1)n}$ (of both $M^c_{n,p+1-q}$ and $(Y^c_n, Z)'(Y^c_n, Z)$) satisfies

$$\lambda_{\min}\big((Y^c_n, Z)'(Y^c_n, Z)\big) = \kappa^Z_{(p+1)n} \to \lambda_{\min}\big((Y(\tau^c_\infty), Z_2)'(Y(\tau^c_\infty), Z_2)\big), \tag{27.27}$$

where the equality holds by the definition of $\kappa^Z_{(p+1)n}$ in (27.17). This establishes (27.16).

Now we use (27.16) to establish that $c_{k,p}(Y^c_n, 1-\alpha) \to c_{k,p,q}(\tau^c_\infty, 1-\alpha)$, which proves the lemma. Let

$$F_{k,p,q,\tau^c_\infty}(x) = P\big(\mathrm{ACLR}_{k,p,q}(\tau^c_\infty) \leq x\big). \tag{27.28}$$

By (27.16), for any $x \in R$ that is a continuity point of $F_{k,p,q,\tau^c_\infty}(x)$, we have

$$1\big(\mathrm{CLR}_{k,p}(Y^c_n) \leq x\big) \to 1\big(\mathrm{ACLR}_{k,p,q}(\tau^c_\infty) \leq x\big) \quad \text{a.s.} \tag{27.29}$$

Equation (27.29) and the bounded convergence theorem give

$$P\big(\mathrm{CLR}_{k,p}(Y^c_n) \leq x\big) \to P\big(\mathrm{ACLR}_{k,p,q}(\tau^c_\infty) \leq x\big) = F_{k,p,q,\tau^c_\infty}(x). \tag{27.30}$$

Now Lemma 27.6 gives the desired result, because (27.30) verifies assumption (i) of Lemma 27.6 and the df of $\mathrm{ACLR}_{k,p,q}(\tau^c_\infty)$ is strictly increasing at its $1-\alpha$ quantile (by Lemma 27.3), which verifies assumption (ii) of Lemma 27.6. □

Proof of Lemma 27.5. The proof is similar to the proof of Lemma 17.1 given in Section 17 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 27.5, $h^c_{6,q} = 0$ if $q < \min\{k, p\}$. If $q = \min\{k, p\}$, then $h^c_{6,q}$ is not defined in the statement of Lemma 27.5 and we define it here to equal zero. If $h^c_{6,j} > 0$, then $\{\tau^c_{jn} : n \geq 1\}$ and $\{\tau^c_{(j+1)n} : n \geq 1\}$ are of the same order of magnitude, that is, $0 < \lim \tau^c_{(j+1)n}/\tau^c_{jn} \leq 1$. We group the first q values of τ^c_{jn} into groups that have the same order of magnitude within each group. Let G $(\in \{1, \ldots, q\})$ denote the number of groups. Note that G equals the number of values in $\{h^c_{6,1}, \ldots, h^c_{6,q}\}$ that equal zero. Let r_g and r^c_g denote the indices of the first and last values in the gth group, respectively, for $g = 1, \ldots, G$. Thus, $r_1 = 1$, $r^c_g = r_{g+1} - 1$, where by definition $r_{G+1} = q+1$, and $r^c_G = q$. By definition, the τ^c_{jn} values in the gth group, which have the gth largest order of magnitude, are $\{\tau^c_{r_g n}: n \geq 1\}, \dots, \{\tau^c_{r^c_g n}: n \geq 1\}$. By construction, $h^c_{6,j} > 0$ for all $j \in \{r_g, \dots, r^c_g - 1\}$ for $g = 1, \dots, G$. (The reason is: if $h^c_{6,j}$ is equal to zero for some $j \leq r^c_g - 1$, then $\{\tau^c_{r^c_g n}: n \geq 1\}$ is of smaller order of magnitude than $\{\tau^c_{r_g n}: n \geq 1\}$, which contradicts the definition of r^c_g.) Also by construction, $\lim \tau^c_{j'n}/\tau^c_{jn} = 0$ for any (j, j') in groups (g, g'), respectively, with $g < g'$.

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

$$\left|\left(\tau^c_{r_1 n}\right)^{-2}\left(Y^c_n, Z\right)'\left(Y^c_n, Z\right) - \left(\tau^c_{r_1 n}\right)^{-2}\kappa I_{p+1}\right| = 0. \tag{27.31}$$

Thus, $\{(\tau^c_{r_1 n})^{-2}\kappa^Z_{jn}: j \leq p+1\}$ solve

$$\left|\left(\tau^c_{r_1 n}\right)^{-2}\left(Y^c_n, Z\right)'\left(Y^c_n, Z\right) - \kappa I_{p+1}\right| = 0. \tag{27.32}$$

Let

$$h^{cc}_{6,r^c_1} := \mathrm{Diag}\left\{1, h^c_{6,1}, h^c_{6,1}h^c_{6,2}, \dots, \prod_{\ell=1}^{r^c_1 - 1} h^c_{6,\ell}\right\} \in R^{r^c_1 \times r^c_1}. \tag{27.33}$$

When $k \geq p$, we have

$$\begin{aligned}
&\left(\tau^c_{r_1 n}\right)^{-1}\left(Y^c_n, Z\right) \\
&= \begin{bmatrix}
h^{cc}_{6,r^c_1} + o(1) & 0^{r^c_1 \times (q - r^c_1)} & 0^{r^c_1 \times (p-q)} & O\left(1/\tau^c_{r_1 n}\right)^{r^c_1 \times 1} \\
0^{(q - r^c_1) \times r^c_1} & O\left(\tau^c_{r_2 n}/\tau^c_{r_1 n}\right)^{(q - r^c_1) \times (q - r^c_1)} & 0^{(q - r^c_1) \times (p-q)} & O\left(1/\tau^c_{r_1 n}\right)^{(q - r^c_1) \times 1} \\
0^{(p-q) \times r^c_1} & 0^{(p-q) \times (q - r^c_1)} & O\left(1/\tau^c_{r_1 n}\right)^{(p-q) \times (p-q)} & O\left(1/\tau^c_{r_1 n}\right)^{(p-q) \times 1} \\
0^{(k-p) \times r^c_1} & 0^{(k-p) \times (q - r^c_1)} & 0^{(k-p) \times (p-q)} & O\left(1/\tau^c_{r_1 n}\right)^{(k-p) \times 1}
\end{bmatrix} \\
&\to \begin{bmatrix}
h^{cc}_{6,r^c_1} & 0^{r^c_1 \times (p+1-r^c_1)} \\
0^{(k - r^c_1) \times r^c_1} & 0^{(k - r^c_1) \times (p+1-r^c_1)}
\end{bmatrix},
\end{aligned} \tag{27.34}$$

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 a s vector whose elements are $O(d_n)$, the equality uses $\tau^c_{jn}/\tau^c_{r_1 n} = \prod_{\ell=1}^{j-1}(\tau^c_{(\ell+1)n}/\tau^c_{\ell n}) = \prod_{\ell=1}^{j-1} h^c_{6,\ell} + o(1)$ for $j = 2, \dots, r^c_1$ (which holds by the definition of $h^c_{6,\ell}$) and $\tau^c_{jn}/\tau^c_{r_1 n} = O(\tau^c_{r_2 n}/\tau^c_{r_1 n})$ for $j = r_2, \dots, q$ (because $\{\tau^c_{jn}: j \leq q\}$ are nonincreasing in j), and the convergence uses $\tau^c_{r_1 n} \to \infty$ (by assumption (ii) of the lemma since $r_1 \leq q$) and $\tau^c_{r_2 n}/\tau^c_{r_1 n} \to 0$ (by the definition of r_2).

When $k < p$, (27.34) 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 (27.34) yields

$$\left(\tau^c_{r_1 n}\right)^{-2}\left(Y^c_n, Z\right)'\left(Y^c_n, Z\right) \to \begin{bmatrix}
\left(h^{cc}_{6,r^c_1}\right)^2 & 0^{r^c_1 \times (p+1-r^c_1)} \\
0^{(p+1-r^c_1) \times r^c_1} & 0^{(p+1-r^c_1) \times (p+1-r^c_1)}
\end{bmatrix}. \tag{27.35}$$

The vector of eigenvalues of a matrix is a continuous function of the matrix (by Elsner's theorem, see Stewart (2001, Theorem 3.1, pp. 37–38)). Hence, by (27.32) and

(27.35), the first r_1^c eigenvalues of $(\tau_{r_1n}^c)^{-2}(Y_n^c, Z)'(Y_n^c, Z)$, that is, $\{(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z : j \leq r_1^c\}$, satisfy

$$\left((\tau_{r_1n}^c)^{-2}\kappa_{1n}^Z, \ldots, (\tau_{r_1n}^c)^{-2}\kappa_{r_1^cn}^Z\right) \to_p \left(1, h_{6,1}^c, h_{6,1}^c h_{6,2}^c, \ldots, \prod_{\ell=1}^{r_1^c-1} h_{6,\ell}^c\right) \quad \text{and so}$$
$$\kappa_{1n}^Z \to \infty \quad \forall j = 1, \ldots, r_1^c \tag{27.36}$$

because $\tau_{r_1n}^c \to \infty$ (since $r_1 \leq q$) and $h_{6,\ell}^c > 0$ for all $\ell \in \{1, \ldots, r_1^c - 1\}$ (as noted above). By the same argument, the last $p + 1 - r_1^c$ eigenvalues of $(\tau_{r_1n}^c)^{-2}(Y_n^c, Z)'(Y_n^c, Z)$, that is, $\{(\tau_{r_1n}^c)^{-2}\kappa_{jn}^Z : j = r_1^c + 1, \ldots, p + 1\}$, satisfy

$$\left(\tau_{r_1n}^c\right)^{-2}\kappa_{jn}^Z \to 0 \quad \forall j = r_1^c + 1, \ldots, p + 1. \tag{27.37}$$

Next, the equality in (27.34) gives

$$\left(\tau_{r_1n}^c\right)^{-2}\left(Y_n^c, Z\right)'\left(Y_n^c, Z\right)$$
$$= \begin{bmatrix} \left(h_{6,r_1^c}^{cc}\right)^2 + o(1) & 0^{r_1^c\times(q-r_1^c)} & 0^{r_1^c\times(p-q)} & O\left(1/\tau_{r_1n}^c\right)^{r_1^c\times 1} \\ 0^{(q-r_1^c)\times r_1^c} & O\left(\left(\tau_{r_2n}^c/\tau_{r_1n}^c\right)^2\right)^{(q-r_1^c)\times(q-r_1^c)} & 0^{(q-r_1^c)\times(p-q)} & O\left(\tau_{r_2n}^c/\left(\tau_{r_1n}^c\right)^2\right)^{(q-r_1^c)\times 1} \\ 0^{(p-q)\times r_1^c} & 0^{(p-q)\times(q-r_1^c)} & O\left(1/\left(\tau_{r_1n}^c\right)^2\right)^{(p-q)\times(p-q)} & O\left(1/\left(\tau_{r_1n}^c\right)^2\right)^{(p-q)\times 1} \\ O\left(1/\tau_{r_1n}^c\right)^{1\times r_1^c} & O\left(\tau_{r_2n}^c/\left(\tau_{r_1n}^c\right)^2\right)^{1\times(q-r_1^c)} & O\left(1/\left(\tau_{r_1n}^c\right)^2\right)^{1\times(p-q)} & O\left(1/\left(\tau_{r_1n}^c\right)^2\right)^{1\times 1} \end{bmatrix}. \tag{27.38}$$

Equation (27.38) holds when $k \geq p$ and $k < p$ (because the column dimensions of the submatrices in the second line of (27.34) 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 $j_1 + 1, \ldots, j_2$ columns of I_{p+1} for $0 \leq j_1 < j_2 \leq p + 1$. We can write

$$I_{p+1} = (I_{0,r_1^c}, I_{r_1^c,p+1}), \quad \text{where } I_{0,r_1^c} := \begin{pmatrix} I_{r_1^c} \\ 0^{(p+1-r_1^c)\times r_1^c} \end{pmatrix} \in R^{(p+1)\times r_1^c} \quad \text{and}$$
$$I_{r_1^c,p+1} := \begin{pmatrix} 0^{r_1^c\times(p+1-r_1^c)} \\ I_{p+1-r_1^c} \end{pmatrix} \in R^{(p+1)\times(p+1-r_1^c)}. \tag{27.39}$$

In consequence, we have

$$\left(Y_n^c, Z\right) = \left((Y_n^c, Z)I_{0,r_1^c}, (Y_n^c, Z)I_{r_1^c,p+1}\right) \quad \text{and}$$
$$\varrho_n^c := \left(\tau_{r_1n}^c\right)^{-2} I_{0,r_1^c}'\left(Y_n^c, Z\right)'\left(Y_n^c, Z\right)I_{r_1^c,p+1} = o\left(\tau_{r_2n}^c/\tau_{r_1n}^c\right), \tag{27.40}$$

where the last equality uses the expressions in the first row of the matrix on the rhs of (27.38) 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 (27.32), $\{(\tau^c_{r_1n})^{-2}\kappa^Z_{jn} : j \le p+1\}$ solve

$$
\begin{aligned}
0 &= \big|(\tau^c_{r_1n})^{-2}(Y^c_n, Z)'(Y^c_n, Z) - \kappa I_{p+1}\big| \\
&= \left|\begin{bmatrix} (\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} - \kappa I_{r^c_1} \vdots & (\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} \\ (\tau^c_{r_1n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} & (\tau^c_{r_1n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} - \kappa I_{p+1-r^c_1} \end{bmatrix}\right| \\
&= \big|(\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} - \kappa I_{r^c_1}\big| \\
&\quad \times \big|(\tau^c_{r_1n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} - \kappa I_{p+1-r^c_1} \\
&\quad - \varrho^{c\prime}_n\big((\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} - \kappa I_{r^c_1}\big)^{-1}\varrho^c_n\big|,
\end{aligned} \tag{27.41}
$$

where the third equality uses the standard formula for the determinant of a partitioned matrix, the definition of ϱ^c_n in (27.40), and the result given in (27.42) below that the matrix which is inverted that appears in the last line of (27.41) is nonsingular for κ equal to any solution $(\tau^c_{r_1n})^{-2}\kappa^Z_{jn}$ to the first equality in (27.41) for $j = r^c_1 + 1, \dots, p+1$.

Now we show that, for $j = r^c_1 + 1, \dots, p+1$, $(\tau^c_{r_1n})^{-2}\kappa^Z_{jn}$ cannot solve the determinantal equation $|(\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} - \kappa I_{r^c_1}| = 0$ for n sufficiently large, where this determinant is the first multiplicand on the rhs of (27.41). Hence, $\{(\tau^c_{r_1n})^{-2}\kappa^Z_{jn} : j = r^c_1 + 1, \dots, p+1\}$ must solve the determinantal equation based on the second multiplicand on the rhs of (27.41) for n sufficiently large. For $j = r^c_1 + 1, \dots, p+1$, we have

$$
(\tau^c_{r_1n})^{-2}I'_{0,r^c_1}(Y^c_n, Z)'(Y^c_n, Z)I_{0,r^c_1} - (\tau^c_{r_1n})^{-2}\kappa^Z_{jn}I_{r^c_1} = (h^{cc}_{6,r^c_1})^2 + o(1), \tag{27.42}
$$

where the equality holds by (27.35) and (27.37). Equation (27.42) and $\lambda_{\min}((h^{cc}_{6,r^c_1})^2) > 0$ (which follows from the definition of h^{cc}_{6,r^c_1} in (27.33) and the fact that $h^c_{6,j} > 0$ for all $j \in \{1, \dots, r^c_1 - 1\}$) establish the desired result.

For $j = r^c_1 + 1, \dots, p+1$, plugging $(\tau^c_{r_1n})^{-2}\kappa^Z_{jn}$ into the second multiplicand on the rhs of (27.41) and using (27.40) and (27.42) gives

$$
0 = \big|(\tau^c_{r_1n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} + o\big((\tau^c_{r_2n}/\tau^c_{r_1n})^2\big) - (\tau^c_{r_1n})^{-2}\kappa^Z_{jn}I_{p+1-r^c_1}\big|. \tag{27.43}
$$

Thus, $\{(\tau^c_{r_1n})^{-2}\kappa^Z_{jn} : j = r^c_1 + 1, \dots, p+1\}$ solve

$$
0 = \big|(\tau^c_{r_1n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} + o\big((\tau^c_{r_2n}/\tau^c_{r_1n})^2\big) - \kappa I_{p+1-r^c_1}\big|. \tag{27.44}
$$

Or equivalently, multiplying through by $(\tau^c_{r_2n}/\tau^c_{r_1n})^{-2}$, $\{(\tau^c_{r_2n})^{-2}\kappa^Z_{jn} : j = r^c_1 + 1, \dots, p+1\}$ solve

$$
0 = \big|(\tau^c_{r_2n})^{-2}I'_{r^c_1,p+1}(Y^c_n, Z)'(Y^c_n, Z)I_{r^c_1,p+1} + o(1) - \kappa I_{p+1-r^c_1}\big| \tag{27.45}
$$

by the same argument as in (27.31) and (27.32).

Now, we repeat the argument from (27.32) to (27.45) with the expression in (27.45) replacing that in (27.32) and with $I_{p+1-r_1^c}$, $\tau_{r_2n}^c$, $\tau_{r_3n}^c$, $r_2^c - r_1^c$, $p+1-r_2^c$, and $h_{6,r_2^c}^{cc} = \mathrm{Diag}\{1, h_{6,r_1^c+1}^c, h_{6,r_1^c+1}^c h_{6,r_1^c+2}^c, \ldots, \prod_{\ell=r_1^c+1}^{r_2^c-1} h_{6,\ell}^c\} \in R^{(r_2^c-r_1^c)\times(r_2^c-r_1^c)}$ in place of I_{p+1}, $\tau_{r_1n}^c$, $\tau_{r_2n}^c$, 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 (27.41) are replaced by the matrices $I_{r_1^c,r_2^c}$ and $I_{r_2^c,p+1}$. This argument gives

$$\kappa_{jn}^Z \to \infty \quad \forall j = r_2, \ldots, r_2^c \quad \text{and} \quad \left(\tau_{r_2n}^c\right)^{-2} \kappa_{jn}^Z = o(1) \quad \forall j = r_2^c+1, \ldots, p+1. \tag{27.46}$$

Repeating the argument $G-2$ more times yields

$$\begin{aligned} &\kappa_{jn}^Z \to \infty \quad \forall j = 1, \ldots, r_G^c \quad \text{and} \\ &\left(\tau_{r_gn}^c\right)^{-2} \kappa_{jn}^Z = o(1) \quad \forall j = r_g^c+1, \ldots, p+1, \forall g = 1, \ldots, G. \end{aligned} \tag{27.47}$$

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 17.1 given in Section 17 in the SM of AG1.

Because $r_J^c = q$, the first result in (27.47) proves part (a) of the lemma.

The second result in (27.47) with $g = G$ implies: for all $j = q+1, \ldots, p+1$,

$$\left(\tau_{r_Gn}^c\right)^{-2} \kappa_{jn}^Z = o(1) \tag{27.48}$$

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 $j = q+1, \ldots, p+1$ by (27.47). In the latter case, we have

$$\lim \frac{\tau_{qn}^c}{\tau_{r_Gn}^c} = \lim \frac{\tau_{r_G^cn}^c}{\tau_{r_Gn}^c} = \prod_{j=r_G}^{r_G^c-1} h_{6,j}^c > 0, \tag{27.49}$$

where the inequality holds because $h_{6,j}^c > 0$ for all $j \in \{r_G, \ldots, r_G^c - 1\}$, as noted at the beginning of the proof. Hence, in this case too, $(\tau_{qn}^c)^{-2}\kappa_{jn}^Z = o(1)$ for $j = q+1, \ldots, p+1$ by (27.48) and (27.49). Because $\tau_{jn}^c \geq \tau_{qn}^c$ for all $j \leq q$, this establishes part (b) of the lemma. □

Proof of Lemma 27.6. For $\varepsilon > 0$, such that $q_\infty \pm \varepsilon$ are continuity points of $F^*(x)$, we have

$$\begin{aligned} &F_n^*(q_\infty - \varepsilon) \to F^*(q_\infty - \varepsilon) < 1 - \alpha \quad \text{and} \\ &F_n^*(q_\infty + \varepsilon) \to F^*(q_\infty + \varepsilon) > 1 - \alpha \end{aligned} \tag{27.50}$$

by assumptions (i) and (ii) of the lemma and $F^*(q_\infty - \varepsilon) < 1 - \alpha$ by the definition of q_∞. The first line of (27.50) 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 (27.50) 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, $q_n \to q_\infty$. □

27.4 *Proof of Lemma 27.3*

Lemma 27.3 is stated in Section 27.1.

Proof of Lemma 27.3. We prove the lemma by proving it separately for four cases: (i) $q \geq 1$, (ii) $k \leq p$, (iii) $\tau^c_{\min\{k,p\}\infty} = 0$, where $\tau^c_{\min\{k,p\}\infty}$ denotes the $\min\{k, p\}$th (and hence, last and smallest) element of τ^c_∞, and (iv) $q = 0$, $k > p$, and $\tau^c_{p\infty} > 0$. First, suppose $q \geq 1$. Then

$$\begin{aligned}\mathrm{ACLR}_{k,p,q}(\tau^c_\infty) &:= Z'Z - \lambda_{\min}\big((Y(\tau^c_\infty), Z_2)'(Y(\tau^c_\infty), Z_2)\big)\\ &= Z_1'Z_1 + Z_2'Z_2 - \lambda_{\min}\big((Y(\tau^c_\infty), Z_2)'(Y(\tau^c_\infty), Z_2)\big) \qquad (27.51)\end{aligned}$$

and $\mathrm{ACLR}_{k,p,q}(\tau^c_\infty)$ is the convolution of a χ^2_q distribution (since $Z_1'Z_1 \sim \chi^2_q$) 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

$$P(X + Y \in B) = \int P(X + y \in B|Y = y)\, dP_Y(y) = \int P(X \in B - y)\, dP_Y(y) = 0, \qquad (27.52)$$

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 (27.52) to (27.51) with $X = Z_1'Z_1$, we conclude that $\mathrm{ACLR}_{k,p,q}(\tau^c_\infty)$ 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 $\varepsilon > 0$,

$$P(Y < c + \varepsilon) \geq P(Y < c_Y + \varepsilon) \geq 1 - \alpha > 0. \qquad (27.53)$$

For $\varepsilon > 0$, we have

$$\begin{aligned}P\big(X + Y \in [c, c + \varepsilon]\big) &= \int P\big(X + y \in [c, c + \varepsilon]|Y = y\big)\, dP_Y(y)\\ &= \int P\big(X \in [c - y, c - y + \varepsilon]\big)\, dP_Y(y) > 0, \qquad (27.54)\end{aligned}$$

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 (27.53). Equation (27.54) 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 $\mathrm{ACLR}_{k,p,q}(\tau^c_\infty) = X + Y$ and $Z_1'Z_1 = X$. This implies that the df of $\mathrm{ACLR}_{k,p,q}(\tau^c_\infty)$ is strictly increasing at its $1 - \alpha$ quantile when $q \geq 1$.

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

Fourth, suppose $q = 0$, $k > p$, and $\tau_{p\infty}^c > 0$. In this case, $Z_2 = Z$ (because $q = 0$) and $Y(\tau_\infty^c) = (D, 0^{p\times(k-p)})'$, where $D := \mathrm{Diag}\{\tau_\infty^c\}$ is a pd diagonal $p \times p$ matrix (because $\tau_{p\infty}^c > 0$). 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 $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c)$. In the present case, we have

$$\begin{aligned}
\mathrm{ACLR} &= Z'Z - \lambda_{\min}\left(\begin{pmatrix} D & Z_a \\ 0^{(k-p)\times p} & Z_b \end{pmatrix}' \begin{pmatrix} D & Z_a \\ 0^{(k-p)\times p} & Z_b \end{pmatrix}\right) \\
&= Z'Z - \inf_{\xi=(\xi_1',\xi_2)':\|\xi\|=1} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}' \begin{pmatrix} D^2 & DZ_a \\ Z_a'D & Z'Z \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \\
&= \sup_{\xi=(\xi_1',\xi_2)':\|\xi\|=1} \left[(1-\xi_2^2)(Z_b'Z_b + Z_a'Z_a) - \xi_1'D^2\xi_1 - 2\xi_2 Z_a'D\xi_1\right], \qquad (27.55)
\end{aligned}$$

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

We define the following nonstochastic function:

$$\mathrm{ACLR}(z_a, \omega) := \sup_{\xi=(\xi_1',\xi_2)':\|\xi\|=1} \left[(1-\xi_2^2)(\omega + z_a'z_a) - \xi_1'D^2\xi_1 - 2\xi_2 z_a'D\xi_1\right] \qquad (27.56)$$

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

We show below that the function $\mathrm{ACLR}(z_a, \omega)$ 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 $\mathrm{ACLR}(z_a, \omega)$ satisfies (iv) $\lim_{\omega\to\infty} \mathrm{ACLR}(z_a, \omega) = \infty$. In consequence, $\forall z_a \neq 0^p$, $\mathrm{ACLR}(z_a, \omega)$ has a continuous, strictly-increasing inverse function in its second argument with domain $[\mathrm{ACLR}(z_a, 0), \infty) \subset R_+$, which we denote by $\mathrm{ACLR}^{-1}(z_a, x)$.[69] Using this, we have: for all $x \geq \mathrm{ACLR}(z_a, 0)$ and $z_a \neq 0^p$,

$$\mathrm{ACLR}(z_a, \omega) \leq x \quad \text{iff} \quad \omega \leq \mathrm{ACLR}^{-1}(z_a, x), \qquad (27.57)$$

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

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

$$\lim_{x\to x_0} P(\mathrm{ACLR}(z_a, Z_b'Z_b) \leq x) = P(\mathrm{ACLR}(z_a, Z_b'Z_b) \leq x_0). \qquad (27.58)$$

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

To prove (27.58), first consider the case $x_0 > \text{ACLR}(z_a, 0)$ (≥ 0) and $z_a \neq 0^p$. In this case, we have

$$\begin{aligned}\lim_{x \to x_0} P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq x\big) &= \lim_{x \to x_0} P\big(Z_b' Z_b \leq ACLR^{-1}(z_a, x)\big) \\ &= P\big(Z_b' Z_b \leq \text{ACLR}^{-1}(z_a, x_0)\big), \qquad (27.59)\end{aligned}$$

where the first equality holds by (27.57) 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 $\text{ACLR}^{-1}(z_a, x)$ at x_0. Hence, (27.58) holds when $x_0 > \text{ACLR}(z_a, 0)$.

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

$$P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq x_0\big) \leq P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) < \text{ACLR}(z_a, 0)\big) = 0, \qquad (27.60)$$

where the equality holds because $\text{ACLR}(z_a, x)$ is increasing in x on R_+ by property (ii) and $Z_b' Z_b \geq 0$ a.s. For x sufficiently close to x_0, $x < \text{ACLR}(z_a, 0)$ and by the same argument as in (27.60), we obtain $P(\text{ACLR}(z_a, Z_b' Z_b) \leq x) = 0$. Thus, (27.58) holds for $x_0 < \text{ACLR}(z_a, 0)$.

Finally, consider the case $x_0 = \text{ACLR}(z_a, 0)$ and $z_a \neq 0^p$. In this case, (27.58) holds for sequences of values x that strictly decline to x_0 by the same argument as for the first case where $x_0 > \text{ACLR}(z_a, 0)$. Next, consider a sequence that strictly increases to x_0. We have $P(\text{ACLR}(z_a, Z_b' Z_b) \leq x) = 0\ \forall x < x_0$ by the same argument as given for the second case where $x_0 < \text{ACLR}(z_a, 0)$. In addition, we have

$$\begin{aligned}P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq x_0\big) &= P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq \text{ACLR}(z_a, 0)\big) \\ &\leq P\big(Z_b' Z_b \leq 0\big) = 0, \qquad (27.61)\end{aligned}$$

where the inequality holds because $\text{ACLR}(z_a, x)$ is strictly increasing on for $z_a \neq 0^p$ by property (ii). This completes the proof of (27.58).

Using (27.58), we establish the continuity of the df of ACLR on R. For any $x_0 \in R$, we have

$$\begin{aligned}\lim_{x \to x_0} P(\text{ACLR} \leq x) &= \lim_{x \to x_0} P\big(\text{ACLR}\big(Z_a, Z_b' Z_b\big) \leq x\big) \\ &= \lim_{x \to x_0} \int P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq x\big)\, dF_{Z_a}(z_a) \\ &= \int P\big(\text{ACLR}\big(z_a, Z_b' Z_b\big) \leq x_0\big)\, dF_{Z_a}(z_a) \\ &= P(\text{ACLR} \leq x_0), \qquad (27.62)\end{aligned}$$

where $F_{Z_a}(\cdot)$ denotes the df of Z_a, the first and last equalities hold because $\text{ACLR} = \text{ACLR}(Z_a, Z_b' Z_b)$, the second equality uses the independence of Z_a and Z_b, and the third equality holds by the bounded convergence theorem using (27.58) and $P(Z_a \neq 0^p) = 1$. Equation (27.62) 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 $\text{ACLR} \geq 0$ by property (i)), the $1-\alpha$ quantile of ACLR is positive. Hence, the former property implies that the df of ACLR is strictly increasing at its $1-\alpha$ quantile, as stated in the lemma.

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

$$\begin{aligned} &P\big(\mathrm{ACLR}\big(z_a, Z_b' Z_b\big) \in [x, x+\delta]\big) \\ &\quad = P\big(Z_b' Z_b \in \big[\mathrm{ACLR}^{-1}(z_a, x), \mathrm{ACLR}^{-1}(z_a, x+\delta)\big]\big) > 0, \end{aligned} \tag{27.63}$$

where the equality holds by (27.57) and the inequality holds because $\mathrm{ACLR}^{-1}(z_a, x)$ is strictly increasing in x for x in $[\mathrm{ACLR}(z_a, 0), \infty)$ when $z_a \neq 0^p$ and $Z_b' Z_b$ has a χ^2_{k-p} distribution, which is absolutely continuous.

The function $\mathrm{ACLR}(z_a, 0)$ is continuous at all $z_a \in R^p$ (by property (iii)) and $\mathrm{ACLR}(0^p, 0) = 0$ (by a simple calculation using (27.56)). In consequence, for any $x > 0$, there exists a vector $z_a^* \in R^p$ and a constant $\varepsilon > 0$ such that $\mathrm{ACLR}(z_a, 0) < x$ 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 $\delta > 0$,

$$\begin{aligned} P\big(\mathrm{ACLR} \in [x, x+\delta]\big) &= \int P\big(\mathrm{ACLR}\big(z_a, Z_b' Z_b\big) \in [x, x+\delta]\big)\, dF_{Z_a}(z_a) \\ &\geq \int_{B(z_a^*, \varepsilon)} P\big(\mathrm{ACLR}\big(z_a, Z_b' Z_b\big) \in [x, x+\delta]\big)\, dF_{Z_a}(z_a) > 0, \end{aligned} \tag{27.64}$$

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 (27.63) using the fact that $x > \mathrm{ACLR}(z_a, 0)$ for all $z_a \in B(z_a^*, \varepsilon)$ by the definition of $B(z_a^*, \varepsilon)$. Equation (27.64) 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 $\mathrm{ACLR}(z_a, \omega)$, which are stated above. The function $\mathrm{ACLR}(z_a, \omega)$ is seen to be nonnegative by replacing the supremum in (27.56) by $\xi = (0^{p\prime}, 1)'$. Hence, property (i) holds. The function $\mathrm{ACLR}(z_a, \omega)$ can be written as

$$\mathrm{ACLR}(z_a, \omega) = \omega + z_a' z_a - \lambda_{\min} \begin{pmatrix} D^2 & D z_a \\ z_a' D & z_a' z_a + \omega \end{pmatrix} \tag{27.65}$$

by analogous calculations to those in (27.55). The minimum eigenvalue is a continuous function of a matrix is a continuous function of its elements by Elsner's theorem; see Stewart (2001, Theorem 3.1, pp. 37–38). Hence, $\mathrm{ACLR}(z_a, \omega)$ 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

$$\mathrm{ACLR}(z_a, \omega) \geq \big(1 - \xi_{*2}^2\big)\big(\omega + z_a' z_a\big) - \xi_{*1}' D^2 \xi_{*1} - 2\xi_{*2} z_a' D \xi_{*1} \to \infty \quad \text{as } \omega \to \infty, \tag{27.66}$$

where the inequality holds by replacing the supremum over ξ in (27.56) 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 $\text{ACLR}(z_a,\omega)$ 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

$$\text{ACLR}(z_a,\omega) = \big(1-\xi_{\omega 2}^2\big)\big(\omega + z_a' z_a\big) - \xi_{\omega 1}' D^2 \xi_{\omega 1} - 2\xi_{\omega 2} z_a' D \xi_{\omega 1}. \tag{27.67}$$

Such a vector ξ_ω exists because the supremum in (27.56) 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 (27.67), we obtain: for all $\delta > 0$, if $\xi_{\omega 2}^2 < 1$,

$$\begin{aligned}
\text{ACLR}(z_a,\omega) &< \big(1-\xi_{\omega 2}^2\big)\big(\omega + \delta + z_a' z_a\big) - \xi_{\omega 1}' D^2 \xi_{\omega 1} - 2\xi_{\omega 2} z_a' D \xi_{\omega 1} \\
&\le \sup_{\xi=(\xi_1',\xi_2)':\|\xi\|=1} \big[\big(1-\xi_2^2\big)\big(\omega+\delta+z_a'z_a\big) - \xi_1' D^2 \xi_1 - 2\xi_2 z_a' D\xi_1\big] \\
&= \text{ACLR}(z_a,\omega+\delta).
\end{aligned} \tag{27.68}$$

Equation (27.68) shows that $\text{ACLR}(z_a,\omega)$ 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 (27.56) and (27.67), ξ_ω maximizes the rhs expression in (27.56) over $\xi \in R^{p+1}$ subject to $\xi_1'\xi_1 + \xi_2^2 = 1$. The Lagrangian for the optimization problem is

$$\big(1-\xi_2^2\big)\big(\omega + z_a' z_a\big) - \xi_1' D^2 \xi_1 - 2\xi_2 z_a' D \xi_1 + \gamma\big(1 - \xi_2^2 - \xi_1'\xi_1\big), \tag{27.69}$$

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

$$-2D^2\xi_{\omega 1} - 2\xi_{\omega 2} D z_a - 2\gamma_\omega \xi_{\omega 1} = 0^p. \tag{27.70}$$

The solution is $\xi_{\omega 1} = 0^p$ (which is an interior point of the set $\{\xi_1 : \|\xi_1\| \le 1\}$) 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). □

27.5 *Proof of Lemma 27.4*

Lemma 27.4 is stated in Section 27.1.

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

Proof of Lemma 27.4. We prove part (a)(i) first. We have

$$\widehat{W}_{2n} = n^{-1}\sum_{i=1}^n \big(g_i g_i' - E_{F_n} g_i g_i'\big) - \widehat{g}_n \widehat{g}_n' + E_{F_n} g_i g_i' \to_p h_{5,g}, \tag{27.71}$$

where the convergence holds by the WLLN (using the moment conditions in $\mathcal{F}$), $E_{F_n} g_i = 0^k$, 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 $\{\lambda_{n,h} : n \ge 1\}$). Hence, Assumption WU(a) holds for the parameter space Λ_{WU} with $h_7 = h_{5,g}$.

Next, we establish Assumption WU(b) for the parameter space Λ_{WU}. Using the definition of $\widehat{V}_n$ $(= \widehat{V}_n(\theta_0))$ in (19), we have

$$\widehat{V}_n = n^{-1}\sum_{i=1}^{n} f_i f_i' - \widehat{f}_n \widehat{f}_n' = E_{F_n} f_i f_i' - (E_{F_n} f_i)(E_{F_n} f_i)' + o_p(1) \tag{27.72}$$

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

$$\begin{aligned}\widehat{R}_n &= (B' \otimes I_k)\big(E_{F_n} f_i f_i' - (E_{F_n} f_i)\big(E_{F_n} f_i'\big)\big)(B \otimes I_k) + o_p(1) \\ &\rightarrow_p R_h := (B' \otimes I_k)\big[h_5 - \mathrm{vec}\big((0^k, h_4)\big)\mathrm{vec}\big((0^k, h_4)\big)'\big](B \otimes I_k),\end{aligned} \tag{27.73}$$

where $B = B(\theta_0)$ is defined in (19), the convergence uses the definitions of $\lambda_{4,F}$ and $\lambda_{5,F}$ in (16.16), and the definition of $\{\lambda_{n,h} : n \geq 1\}$ in (16.18).

This yields

$$\widehat{U}_{2n} = (\widehat{\Omega}_n, \widehat{R}_n) \rightarrow_p (h_{5,g}, R_h) = h_8, \tag{27.74}$$

which verifies Assumption WU(b) for the parameter space Λ_{WU} for part (a) of the lemma.

Now we establish Assumption WU(c) for the parameter space Λ_{WU} 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 nonzero 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 $W_{2F} \in \mathcal{W}_2$ $\forall F \in \mathcal{F}_{\mathrm{WU}}$ because $W_{2F} = E_F g_i g_i'$ is psd. We have $U_{2F} \in \mathcal{U}_2$ $\forall F \in \mathcal{F}_{\mathrm{WU}}$ because $U_{2F} = (\Omega_F, R_F)$, $\Omega_F := E_F g_i g_i'$ is psd and nonzero (by the last condition in $\mathcal{F}$, 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 nonzero because B is nonsingular and V_F (defined in (16.7)) is nonzero by the argument given in the paragraph containing (27.77) below. By their definitions, $\widehat{\Omega}_n$ and $\widehat{R}_n$ are psd. In addition, they are nonzero wp $\rightarrow 1$ by (27.74) and the result just established that the two matrices that comprise h_8 are nonzero. Hence, $(\widehat{\Omega}_n, \widehat{R}_n) \in \mathcal{U}_2$ wp $\rightarrow 1$.

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 $\mathcal{F}$).

The function $U_1(\cdot)$ defined in (16.8) 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 (16.8) with (Ω, R) in place of (Ω_F, R_F) and $\Sigma^{\varepsilon}(\Omega, R)$ is defined given $\Sigma(\Omega, R)$ by (22). Lemma 22.1(b) shows that $\Sigma^{\varepsilon}(\Omega, R)$ is nonsingular provided $\lambda_{\max}(\Sigma(\Omega, R)) > 0$. We have

$$\begin{aligned}\lambda_{\max}\big(\Sigma(\Omega, R)\big) &\geq \max_{j \leq p+1} \Sigma_{jj}(\Omega, R) = \max_{j \leq p+1} \mathrm{tr}\big(\Omega^{-1/2} R_{jj} \Omega^{-1/2}\big)/k \\ &\geq \max_{j \leq p+1} \lambda_{\max}\big(\Omega^{-1/2} R_{jj} \Omega^{-1/2}\big)/k\end{aligned}$$

$$= \max_{j \le p+1} \sup_{\lambda:\|\lambda\|=1} \frac{\lambda' \Omega^{-1/2}}{\|\Omega^{-1/2}\lambda\|} R_{jj} \frac{\Omega^{-1/2}\lambda}{\|\Omega^{-1/2}\lambda\|} \cdot \|\Omega^{-1/2}\lambda\|^2 / k$$

$$\ge \max_{j \le p+1} \lambda_{\max}(R_{jj}) \lambda_{\min}(\Omega^{-1})/k > 0, \tag{27.75}$$

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 (21) 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 $\mathcal{F}$ imply that $\lambda_{\min}(\Omega^{-1}) = 1/\lambda_{\max}(\Omega) > 0$ for Ω in some neighborhood of $\lim \Omega_{F_n}$ (because $\lambda_{\max}(\Omega_F) = \sup_{\lambda \in R^k:\|\lambda\|=1} E_F(\lambda' g_i)^2 \le E_F\|g_i\|^2 \le M^{2/(2+\gamma)} < \infty$ for all $F \in \mathcal{F}$ using the Cauchy–Bunyakovsky–Schwarz inequality) and $\inf_{F\in\mathcal{F}} \lambda_{\max}(R_F) > 0$, which we show below, implies that $\lambda_{\max}(R_{jj}) > 0$ for some $j \le p+1$.

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

$$\inf_{F\in\mathcal{F}} \lambda_{\max}(R_F) > 0. \tag{27.76}$$

We show that the last condition in $\mathcal{F}$, that is, $\inf_{F\in\mathcal{F}} \lambda_{\min}(E_F g_i g_i') > 0$ implies (27.76). In fact, the last condition in $\mathcal{F}$ is very much stronger than is needed to get (27.76). (The full strength of the last condition in $\mathcal{F}$ is used in the proof of Lemma 16.4 (see Section 25) because $\widehat{\Omega}_n^{-1/2}$ enters the definition of $\widehat{D}_n$ and $\widehat{\Omega}_n - \Omega_{F_n} \to_p 0^{k\times k}$, where $\Omega_F = E_F g_i g_i'$.) We show that (27.76) holds provided $\inf_{F\in\mathcal{F}} \lambda_{\max}(E_F g_i g_i') > 0$.

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

$$\begin{aligned}
\lambda_{\max}(R_F) &:= \lambda_{\max}\big((B' \otimes I_k) V_F (B \otimes I_k)\big) = \sup_{x \in R^{(p+1)k}:\|x\|=1} x'(B' \otimes I_k) V_F (B \otimes I_k) x \\
&\ge x^{\dagger\prime}(B' \otimes I_k) V_F (B \otimes I_k) x^\dagger \cdot \|x^\dagger\|^{-2} = x^{*\prime} V_F x^* / \big(x^{*\prime}(B \otimes I_k)^{-1\prime}(B \otimes I_k)^{-1} x^*\big) \\
&\ge \lambda_{\max}(V_F)/\lambda_{\max}\big((B \otimes I_k)^{-1\prime}(B \otimes I_k)^{-1}\big) = K \lambda_{\max}(V_F),
\end{aligned} \tag{27.77}$$

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 (19)). Next, $\inf_{F\in\mathcal{F}} \lambda_{\max}(V_F) \ge \inf_{F\in\mathcal{F}} \lambda_{\max}(E_F g_i g_i') \ge \delta$ because $E_F g_i g_i'$ is the upper left $p \times p$ submatrix of V_F, which implies that $\lambda_{\max}(V_F) \ge \lambda_{\max}(E_F g_i g_i')$, and $\lambda_{\max}(E_F g_i g_i') \ge \delta$ by the last condition in $\mathcal{F}$. This completes the verification (27.76) and the verification of Assumption WU(c) in part (a) of the lemma.

Now we prove part (a)(ii). It suffices to show that $\mathcal{F} \subset \mathcal{F}_{\mathrm{WU}}$ for δ_1 sufficiently small and M_1 sufficiently large because $\mathcal{F}_{\mathrm{WU}} \subset \mathcal{F}$ by the definition of $\mathcal{F}_{\mathrm{WU}}$. We need to show that the four conditions in the definition of $\mathcal{F}_{\mathrm{WU}}$ in (16.12) hold.

(I) We show that $\inf_{F\in\mathcal{F}}\lambda_{\min}(W_F)>0$, where $W_F:=W_1(W_{2F}):=\Omega_F^{-1/2}:=(E_Fg_ig_i')^{-1/2}$ (by (16.5), (16.8), and (16.11)). The inequality $E_F\|g_i\|^{2+\gamma}\leq M$ in $\mathcal{F}$ 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 $W_F^{-2}=\Omega_F=E_Fg_ig_i'$.

(II) We show that $\sup_{F\in\mathcal{F}}\|W_F\|<\infty$, where $W_F:=W_1(W_{2F}):=\Omega_F^{-1/2}:=(E_Fg_ig_i')^{-1/2}$ (by (16.5) and (16.8)). We have $\inf_{F\in\mathcal{F}}\lambda_{\min}(\Omega_F)>0$ (by the last condition in $\mathcal{F}$).

(III) We show that $\inf_{F\in\mathcal{F}}\lambda_{\min}(U_F)>0$, where in the present case $U_F:=U_1(U_{2F}):=((\theta_0,I_p)(\Sigma_F^{\varepsilon})^{-1}(\theta_0,I_p)')^{1/2}$ and $\Sigma_F:=\Sigma(\Omega_F,R_F)$ has (j,ℓ) element equal to $\mathrm{tr}(R_{j\ell F}'\Omega_F^{-1})/k$ (by (16.8)). We have $\sup_{F\in\mathcal{F}}\|R_F\|=\sup_{F\in\mathcal{F}}\|(B'\otimes I_k)\mathrm{Var}_F(f_i)(B\otimes I_k)\|<\infty$ (where the inequality uses the condition $E_F\|(g_i',\mathrm{vec}(G_i)')'\|^{2+\gamma}\leq M$ in $\mathcal{F}$). In addition, $\inf_{F\in\mathcal{F}}\lambda_{\min}(\Omega_F)>0$ (by the last condition in $\mathcal{F}$). The latter results imply that $\sup_{F\in\mathcal{F}}\|\Sigma_F\|<\infty$ (because Σ_F minimizes $\|(I_{p+1}\otimes\Omega_F^{-1/2})[\Sigma\otimes\Omega_F-R_F](I_{p+1}\otimes\Omega_F^{-1/2})\|$, see the paragraph containing (16.8)). This implies that $\sup_{F\in\mathcal{F}}\|\Sigma_F^{\varepsilon}\|<\infty$. In addition, Σ_F is nonsingular $\forall F\in\mathcal{F}$ (because $\inf_{F\in\mathcal{F}}\lambda_{\min}(\Sigma_F)>0$ by the proof of result (IV) below). The last two results imply the desired result $\inf_{F\in\mathcal{F}}\lambda_{\min}(U_F)=\inf_{F\in\mathcal{F}}\lambda_{\min}((\theta_0,I_p)(\Sigma_F^{\varepsilon})^{-1}(\theta_0,I_p)')^{1/2})>0$ (because $A:=(\theta_0,I_p)\in R^{p\times(p+1)}$ has full row rank p and $\lambda_{\min}(U_F)=\inf_{\lambda\in R^p:\|\lambda\|=1}\lambda'A(\Sigma_F^{\varepsilon})^{-1}A'\lambda\geq\inf_{\lambda\in R^p:\|\lambda\|=1}(A'\lambda)'(\Sigma_F^{\varepsilon})^{-1}\times(A'\lambda)/\|A'\lambda\|^2\times\inf_{\lambda\in R^p:\|\lambda\|=1}\|A'\lambda\|^2=\lambda_{\min}((\Sigma_F^{\varepsilon})^{-1})\lambda_{\min}(AA')\geq\delta_2$ for some $\delta_2>0$ that does not depend on F).

(IV) We show that $\sup_{F\in\mathcal{F}}\|U_F\|<\infty$, where U_F is defined in (III) immediately above. By the same calculations as in (27.75) (which use (27.76)) with Σ_F and (Ω_F,R_F) in place of $\Sigma(\Omega,R)$ and (Ω,R), respectively, we have $\inf_{F\in\mathcal{F}_P}\lambda_{\max}(\Sigma_F)>0$. The latter implies $\inf_{F\in\mathcal{F}_P}\lambda_{\min}(\Sigma_F^{\varepsilon})>0$ by Lemma 22.1(b). In turn, the latter implies the desired result $\sup_{F\in\mathcal{F}_P}\|U_F\|=\sup_{F\in\mathcal{F}_P}\|((\theta_0,I_p)\times(\Sigma_F^{\varepsilon})^{-1}(\theta_0,I_p)')^{1/2}\|<\infty$.

This completes the proof of part (a)(ii).

Now, we prove part (b)(i) of the lemma. Assumption WU(a) holds for the parameter space $\Lambda_{\mathrm{WU},P}$ with $h_7=h_{5,g}$ by the same argument as for part (a)(i).

Next, we verify Assumption WU(b) for the parameter space $\Lambda_{\mathrm{WU},P}$ for $\widehat{U}_{2n}=(\widehat{\Omega}_n,\widetilde{R}_n)$. Using the definition of $\widetilde{V}_n$ $(=\widetilde{V}_n(\theta_0))$ in (15.5), we have

$$\widetilde{V}_n=n^{-1}\sum_{i=1}^{n}(u_i^*u_i^{*\prime}\otimes Z_iZ_i')-n^{-1}\sum_{i=1}^{n}(\widehat{u}_{in}^*u_i^{*\prime}\otimes Z_iZ_i')-n^{-1}\sum_{i=1}^{n}(u_i^*\widehat{u}_{in}^{*\prime}\otimes Z_iZ_i')$$
$$+n^{-1}\sum_{i=1}^{n}(\widehat{u}_{in}^*\widehat{u}_{in}^{*\prime}\otimes Z_iZ_i'). \tag{27.78}$$

We have

$$\begin{aligned}
n^{-1}\sum_{i=1}^{n}\big(u_i^{*}u_i^{*\prime}\otimes Z_iZ_i'\big) &= E_{F_n}f_if_i' + o_p(1),\\
\widetilde{\Xi}_n &= \big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1}n^{-1}Z_{n\times k}'U^{*}\\
&= \big(E_{F_n}Z_iZ_i'\big)^{-1}E_{F_n}Z_iu_i^{*\prime} + o_p(1)\\
&= \big(E_{F_n}Z_iZ_i'\big)^{-1}E_{F_n}(g_i, G_i) + o_p(1) =: \Xi_{F_n} + o_p(1),\\
n^{-1}\sum_{i=1}^{n}\big(\widehat{u}_{in}^{*}u_i^{*\prime}\otimes Z_iZ_i'\big) &= n^{-1}\sum_{i=1}^{n}\big(\widetilde{\Xi}_n'Z_iu_i^{*\prime}\otimes Z_iZ_i'\big)\\
&= E_{F_n}\big(\Xi_{F_n}'(g_i, G_i)\otimes Z_iZ_i'\big) + o_p(1), \quad \text{and}\\
n^{-1}\sum_{i=1}^{n}\big(\widehat{u}_{in}^{*}\widehat{u}_{in}^{*\prime}\otimes Z_iZ_i'\big) &= n^{-1}\sum_{i=1}^{n}\big(\widetilde{\Xi}_n'Z_iZ_i'\widetilde{\Xi}_n\otimes Z_iZ_i'\big)\\
&= E_{F_n}\big(\Xi_{F_n}'Z_iZ_i'\Xi_{F_n}\otimes Z_iZ_i'\big) + o_p(1),
\end{aligned} \tag{27.79}$$

where the first line holds by the WLLN's (since $u_i^{*}u_i^{*\prime}\otimes Z_iZ_i' = f_if_i'$ for f_i defined in (16.10) and using the moment conditions in $\mathcal{F}$), the second line holds by the WLLN's (using the conditions in $\mathcal{F}$ and $\mathcal{F}_P$), Slutsky's theorem, and $Z_iu_i^{*\prime} = (g_i, G_i)$, the fourth line holds by the WLLN's (using $E_F((\|(g_i, G_i)\|\cdot\|Z_i\|^2)^{1+\gamma/4}) \le (E_F\|(g_i, G_i)\|^{2+\gamma/2}E_F\|Z_i\|^{4+\gamma})^{1/2} < \infty$ for $\gamma > 0$ by the Cauchy–Bunyakovsky–Schwarz inequality and the moment conditions in $\mathcal{F}$ and $\mathcal{F}_P$) and the result of the second and third lines, and the fifth line holds by the WLLN's (using the moment conditions in $\mathcal{F}$ and $\mathcal{F}_P$) and the result of the second and third lines.

Equations (16.10) (which defines $\widetilde{V}_F$) with $F = F_n$, (27.78), and (27.79) combine to give

$$\widetilde{V}_n - \widetilde{V}_{F_n} \to_p 0. \tag{27.80}$$

Using the definitions of $\widetilde{R}_n$ and $\widetilde{R}_F$ (in (15.5) and (16.10)), (27.71), (27.80), and $h_7 := \lim W_{2F_n} = \lim \Omega_{F_n}$ yield

$$(\widehat{\Omega}_n, \widetilde{R}_n) \to_p \lim(\Omega_{F_n}, \widetilde{R}_{F_n}) =: h_8. \tag{27.81}$$

This establishes Assumption WU(b) for the parameter space $\Lambda_{\mathrm{WU},P}$ for part (b) of the lemma.

Assumption WU(c) holds for the parameter space $\Lambda_{\mathrm{WU},P}$, 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 $\inf_{F\in\mathcal{F}}\lambda_{\max}(\widetilde{R}_F) > 0$. The latter holds by the same argument as used above to show $\inf_{F\in\mathcal{F}}\lambda_{\max}(R_F) > 0$ (which is given in the paragraph containing (27.77) and the paragraph following it), but with (i) $\widetilde{R}_F$ in place of R_F and (ii) $\inf_{F\in\mathcal{F}}\lambda_{\max}(\widetilde{V}_F) > 0$, rather than $\inf_{F\in\mathcal{F}}\lambda_{\max}(V_F) > 0$, holding. Condition (ii) holds because $\inf_{F\in\mathcal{F}}\lambda_{\max}(\widetilde{V}_F) \ge \inf_{F\in\mathcal{F}}\lambda_{\max}(E_Fg_ig_i') > 0$ because $\widetilde{V}_F$ 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 (16.10), 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 $\widetilde{V}_F$ equals $E_F u_i^2 Z_i Z_i' = E_F g_i g_i'$, and so, $\lambda_{\max}(V_F) \geq \lambda_{\max}(E_F g_i g_i')$, and $\inf_{F \in \mathcal{F}} \lambda_{\max}(E_F g_i g_i') > 0$ is implied by the last condition in $\mathcal{F}$. This completes the verification of Assumption WU(c) in part (b) of the lemma.

Now, we prove part (b)(ii) of the lemma. We need to show that the four conditions in the definition of $\mathcal{F}_{\mathrm{WU}}$ in (16.12) hold for all $F \in \mathcal{F}_P$, for some δ_1 sufficiently small and some M_1 sufficiently large.

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

(III) We show that $\inf_{F \in \mathcal{F}_P} \lambda_{\min}(U_F) > 0$, where in the present case $U_F := U_1(U_{2F}) := ((\theta_0, I_p)(\Sigma^{\varepsilon}(\Omega_F, \widetilde{R}_F))^{-1}(\theta_0, I_p)')^{1/2}$ and $\Sigma(\Omega_F, \widetilde{R}_F)$ has (j, ℓ) element equal to $\mathrm{tr}(\widetilde{R}_{j\ell F}' \Omega_F^{-1})/k$ (by (16.11)). The inequalities $E_F \|Z_i\|^{4+\gamma} \leq M$, $E_F \|(g_i', \mathrm{vec}(G_i)')'\|^{2+\gamma} \leq M$, and $\lambda_{\min}(E_F Z_i Z_i') \geq \delta$ imply that $\sup_{F \in \mathcal{F}_P}(\|\Xi_F\| + \|E_F f_i f_i'\| + \|E_F(\Xi_F' Z_i Z_i' \Xi_F \otimes Z_i Z_i')\| + \|E_F(g_i, G_i)\Xi_F \otimes Z_i Z_i'\|) < \infty$, where Ξ_F is defined in (16.10) (using the Cauchy–Bunyakovsky–Schwarz inequality). This, in turn, implies that $\sup_{F \in \mathcal{F}_P} \|\widetilde{V}_F\| < \infty$, $\sup_{F \in \mathcal{F}_P} \|\widetilde{R}_F\| < \infty$, $\sup_{F \in \mathcal{F}_P} \|\widetilde{\Sigma}_F\| < \infty$, $\sup_{F \in \mathcal{F}_P} \|\widetilde{\Sigma}_F^{\varepsilon}\| < \infty$, and $\lambda_{\min}(\widetilde{L}_F) \geq \delta_2$ for some $\delta_2 > 0$, where $\widetilde{V}_F$ and $\widetilde{R}_F$ are defined in (16.10), $\widetilde{\Sigma}_F := \Sigma(\Omega_F, \widetilde{R}_F)$, $\widetilde{L}_F := (\theta_0, I_p)(\widetilde{\Sigma}_F^{\varepsilon})^{-1}(\theta_0, I_p)'$, and $(\widetilde{\Sigma}_F^{\varepsilon})^{-1}$ exists by (IV) below (and $\lambda_{\min}(\widetilde{L}_F) \geq \delta_2$ holds because $A := (\theta_0, I_p) \in R^{p \times (p+1)}$ has full row rank p and $\lambda_{\min}(\widetilde{L}_F) = \inf_{\lambda \in R^p : \|\lambda\|=1} \lambda' A(\widetilde{\Sigma}_F^{\varepsilon})^{-1} A' \lambda \geq \inf_{\lambda \in R^p : \|\lambda\|=1} (A'\lambda)'(\widetilde{\Sigma}_F^{\varepsilon})^{-1}(A'\lambda)/\|A'\lambda\|^2 \times \inf_{\lambda \in R^p : \|\lambda\|=1} \|A'\lambda\|^2 = \lambda_{\min}((\widetilde{\Sigma}_F^{\varepsilon})^{-1})\lambda_{\min}(AA') \geq \delta_2$ for some $\delta_2 > 0$ that does not depend on F). Finally, $\lambda_{\min}(\widetilde{L}_F) \geq \delta_2$ implies the desired result that $\lambda_{\min}(U_F) \geq \delta_1$ for some $\delta_1 > 0$ (because $U_F := \widetilde{L}_F^{1/2}$).

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

Results (I)–(IV) establish the result of part (b)(ii) of the lemma. □

27.6 *Proof of Theorem 16.1 for the Anderson–Rubin test and CS*

Proof of Theorem 16.1 for AR test and CS. We prove the AR test results of Theorem 16.1 by applying Proposition 16.3 with

$$\lambda = \lambda_F := E_F g_i g_i', \; h_n(\lambda) := \lambda, \quad \text{and} \quad \Lambda := \{\lambda : \lambda = \lambda_F \text{ for some } F \in \mathcal{F}_{\mathrm{AR}}\}. \tag{27.82}$$

We define the parameter space H as in (16.2). For notational simplicity, we verify Assumption B* used in Proposition 16.3 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

$$\lambda_{F_n} \to h := \lim E_{F_n} g_i g_i'. \tag{27.83}$$

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 $\mathcal{F}_{\mathrm{AR}}$) 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

$$n^{1/2} \widehat{g}_n \to_d h^{1/2} Z, \quad \text{where } Z \sim N\big(0^k, I_k\big). \tag{27.84}$$

We have

$$\widehat{\Omega}_n = n^{-1} \sum_{i=1}^{n} \big(g_i g_i' - E_{F_n} g_i g_i'\big) - \widehat{g}_n \widehat{g}_n' + E_{F_n} g_i g_i' \to_p h \quad \text{and} \quad \widehat{\Omega}_n^{-1} \to_p h^{-1}, \tag{27.85}$$

where the equality holds by definition of $\widehat{\Omega}_n$ in (8), the first convergence result uses (27.83), (27.84), 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 (27.84) and (27.85) give

$$\mathrm{AR}_n := n \widehat{g}_n' \widehat{\Omega}_n^{-1} \widehat{g}_n \to_d Z' h^{1/2} h^{-1} h^{1/2} Z = Z'Z \sim \chi_k^2. \tag{27.86}$$

In turn, (27.86) gives

$$P_{F_n}\big(\mathrm{AR}_n > \chi^2_{k,1-\alpha}\big) \to P\big(Z'Z > \chi^2_{k,1-\alpha}\big) = \alpha. \tag{27.87}$$

where the equality holds because $\chi^2_{k,1-\alpha}$ is the $1-\alpha$ quantile of $Z'Z$. Equation (27.87) verifies Assumption B* and the proof of the AR test results of Theorem 16.1 is complete.

The proof of the AR CS results of Theorem 16.1 is analogous to those for the tests; see the Comment to Proposition 16.3. □

28. Proofs of Theorems 7.1 and 15.3

Suppose $k \geq p$. Let A_F and Π_{1F} be defined as in (4) and (5) and the paragraph following these equations with $\theta = \theta_0$. Define λ_F^*, Λ^*, and $\{\lambda_{n,h}^* : n \geq 1\}$ as λ_F, Λ_{WU}, and $\{\lambda_{n,h} : n \geq 1\}$, respectively, are defined in (16.16)–(16.18), but with g_i and G_i replaced by $g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i$ and $G_{Fi}^* := \Pi_{1F}^{-1/2} A_F' G_i$, with $\mathcal{F}$ replaced by $\mathcal{F}^{\mathrm{SR}}$, and with W_F $(:= W_1(W_{2F}))$ and U_F $(:= U_1(U_{2F}))$ defined as in (16.8) with g_i and G_i replaced by g_{Fi}^* and G_{Fi}^*. In addition, we restrict $\{\lambda_{n,h}^* : n \geq 1\}$ to be a sequence for which $\lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$. Let $(s_{1F_n}^*, \ldots, s_{pF_n}^*)$ denote the singular values of $E_F G_{Fi}^*$. Under these conditions, $A_{F_n} = A_{F_n}^{\Omega}$, $\Pi_{1F_n} = \Pi_{F_n}$, $W_{F_n} := (\Pi_{1F_n}^{-1/2} A_{F_n}' \Omega_{F_n} A_{F_n} \Pi_{1F_n}^{-1/2})^{-1/2} = I_k$, and $n^{1/2} s_{pF_n}^* \to \infty$ iff $n^{1/2} s_{pF_n} \to \infty$.

Theorem 7.1 of AG2. *Suppose $k \geq p$. For any sequence $\{\lambda_{n,h}^* : n \geq 1\}$ that exhibits strong or semi-strong identification (i.e., for which $n^{1/2} s_{pF_n}^* \to \infty$) and for which $\lambda_{n,h}^* \in \Lambda^*$ $\forall n \geq 1$ for the* SR-CQLR *test statistic and critical value, we have*

(a) $\text{SR-QLR}_n = \text{QLR}_n + o_p(1) = \text{LM}_n + o_p(1) = LM_n^{\text{GMM}} + o_p(1)$ *and*

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

THEOREM 15.3. *Suppose $k \geq p$. For any sequence $\{\lambda^*_{n,h} : n \geq 1\}$ that exhibits strong or semi-strong identification (i.e., for which $n^{1/2}s^*_{pF_n} \to \infty$) and for which $\lambda^*_{n,h} \in \Lambda^*_P\ \forall n \geq 1$, we have*

(a) $\text{SR-QLR}_{Pn} = \text{QLR}_{Pn} + o_p(1) = \text{LM}_n + o_p(1) = \text{LM}_n^{\text{GMM}} + o_p(1)$ *and*

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

The proofs of Theorems 7.1 and 15.3 use the following lemma that concerns the $\text{QLR}_{\text{WU},n}$ statistic, which is based on general weight matrices $\widehat{W}_n$ and $\widehat{U}_n$ (see (16.3)), and considers sequences of distributions F in $\mathcal{F}$ or $\mathcal{F}_P$, rather than sequences in $\mathcal{F}^{\text{SR}}$ or $\mathcal{F}_P^{\text{SR}}$. Given the result of this lemma, we obtain the results of Theorems 7.1 and 15.3 using an argument that is similar to that employed in Section 17, combined with the verification of Assumption WU for the parameter spaces Λ_{WU} and $\Lambda_{\text{WU},P}$ for the CQLR and CQLR_P tests, respectively, that is given in Lemma 27.4 in Section 27.

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

$$\begin{aligned} \text{LM}_n(\widehat{W}_n) &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{W}_n\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \quad \text{and} \\ \text{LM}_n^{\text{GMM}}(\widehat{W}_n) &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{W}_n\widehat{G}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \end{aligned} \tag{28.1}$$

respectively, where $\widehat{G}_n$ is the sample Jacobian defined in (8) with $\theta = \theta_0$. In Lemma 28.1, we show that when $n^{1/2}\tau_{pF_n} \to \infty$, the $\text{QLR}_{\text{WU},n}$ statistic is asymptotically equivalent to the $\text{LM}_n(\widehat{W}_n)$ and $\text{LM}_n^{\text{GMM}}(\widehat{W}_n)$ 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 $F \in \mathcal{F}_{\text{WU}}$, the smallest and largest singular values of $W_F(E_FG_i)U_F$ (i.e., $\tau_{\min\{k,p\}F}$ and τ_{1F}) are related to those of $\Omega_F^{-1/2}E_FG_i$, denoted (as in Section 6.2 of AG2) by $s_{\min\{k,p\}F}$ and s_{1F}, via $c_1s_{jF} \leq \tau_{jF} \leq c_2s_{jF}$ for $j = \min\{k, p\}$ and $j = 1$ for some constants $0 < c_1 < c_2 < \infty$. This result uses the condition $\lambda_{\min}(\Omega_F) \geq \delta > 0$ in $\mathcal{F}_{\text{WU}}$. (See Section 10.3 in the SM to AG1 for the argument used to prove this result.) In consequence, when $k \geq p$, the standard weak, nonstandard weak, semi-strong, and strong identification categories defined in Section 6.2 are unchanged if s_{jF_n} is replaced by τ_{jF_n} in their definitions for $j = 1, p$.

LEMMA 28.1. *Suppose $k \geq p$ and Assumption* WU *holds for some nonempty parameter space $\Lambda_* \subset \Lambda_{\text{WU}}$. Under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_*$ for which $n^{1/2}\tau_{pF_n} \to \infty$, we have*

(a) $\text{QLR}_{\text{WU},n} = \text{LM}_n(\widehat{W}_n) + o_p(1) = \text{LM}_n^{\text{GMM}}(\widehat{W}_n) + o_p(1)$ *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 $\mathrm{QLR}_{\mathrm{WU},n}$ statistic, defined in (16.3), does not affect the asymptotic distribution of $\mathrm{QLR}_{\mathrm{WU},n}$ statistic under strong or semi-strong identification. This holds because $\mathrm{QLR}_{\mathrm{WU},n}$ 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 28.1 (and are defined in (28.1)) do not depend on $\widehat{U}_n$.

Proofs of Theorem 7.1 of AG2 and Theorem 15.3. By the second last paragraph of Section 5.2, $\text{SR-QLR}_n(\theta_0) = QLR_n(\theta_0)$ wp $\to 1$ under any sequence $\{F_n \in \mathcal{F}^{\mathrm{SR}} : n \geq 1\}$ with $r_{F_n}(\theta_0) = k$ for n large. By the same argument as given there, $\text{SR-QLR}_{Pn}(\theta_0) = \mathrm{QLR}_{Pn}(\theta_0)$ wp $\to 1$ under any sequence $\{F_n \in \mathcal{F}_P^{\mathrm{SR}} : n \geq 1\}$ with $r_{F_n}(\theta_0) = k$ for n large. This establishes the first equality in part (a) of Theorems 7.1 and 15.3 because by assumption $\lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$ (see the paragraphs preceding Theorems 7.1 and 15.3).

Assumption WU for the parameter spaces Λ_{WU} and $\Lambda_{\mathrm{WU},P}$ is verified in Lemma 27.4 in Section 27 for the CQLR and CQLR_P tests, respectively. Hence, Lemma 28.1 implies that under sequences $\{\lambda_{n,h} : n \geq 1\}$ we have $\mathrm{QLR}_n = \mathrm{LM}_n(\widehat{\Omega}_n^{-1/2}) + o_p(1) = \mathrm{LM}_n^{\mathrm{GMM}}(\widehat{\Omega}_n^{-1/2}) + o_p(1)$ and likewise for QLR_{Pn}, where QLR_n and QLR_{Pn} are defined in (23) and in the paragraph containing (15.7), respectively, and $\mathrm{LM}_n(\widehat{\Omega}_n^{-1/2})$ and $\mathrm{LM}_n^{\mathrm{GMM}}(\widehat{\Omega}_n^{-1/2})$ are defined in (28.1) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$. In addition, Lemma 28.1 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 $\mathcal{F}$ or $\mathcal{F}_P$, not $\mathcal{F}^{\mathrm{SR}}$ or $\mathcal{F}_P^{\mathrm{SR}}$.

Next, we employ a similar argument to that in (17.5)–(17.7) of Section 17. Specifically, we apply the version of Lemma 28.1 described in the previous paragraph with $g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i$ and $G_{Fi}^* := \Pi_{1F}^{-1/2} A_F' G_i$ in place of g_i and G_i to the QLR_n and QLR_{Pn} test statistics and their corresponding critical values. 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 $(E_F g_{Fi}^* g_{Fi}^{*\prime})^{-1/2}(E_F G_{Fi}^*)U_F = (\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2})^{-1/2} \times (E_F G_i^*)U_F = (E_F G_i^*)U_F$. In consequence, the condition $n^{1/2}\tau_{pF_n} \to \infty$ of Lemma 28.1 holds for the transformed variables $g_{F_n i}^*$ and $G_{F_n i}^*$, that is, $n^{1/2}\tau_{pF_n}^* \to \infty$. In the present case, $\{\Pi_{1F_n}^{-1/2} A_{F_n}' : n \geq 1\}$ 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 paragraphs preceding Theorems 7.1 and 15.3). In consequence, by Lemmas 5.1 and 15.1, the QLR_n and QLR_{Pn} 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 $\mathcal{F}^{\mathrm{SR}}$ and $\mathcal{F}_P^{\mathrm{SR}}$, the transformed variables g_{Fi}^* and G_{Fi}^* satisfy the conditions in $\mathcal{F}$ and $\mathcal{F}_P$; see (17.6) and (17.7). In particular, $E_F g_{Fi}^* g_{Fi}^{*\prime} = I_k$ and $\lambda_{\min}(E_F Z_{Fi}^* Z_{Fi}^{*\prime}) \geq 1/(2c) > 0$, where $Z_{Fi}^* := \Pi_{1F}^{-1/2} A_F' Z_i$ and c is as in the definition of $\mathcal{F}_P^{\mathrm{SR}}$ in (15.3). In addition, the LM_n and $\mathrm{LM}_n^{\mathrm{GMM}}$ 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 $M = \Pi_{1F}^{-1/2} A_F'$, we have $\mathrm{LM}_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n[\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n]^{-1}\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n$ $= n\widehat{g}_n'M'(M\widehat{\Omega}_nM')^{-1}M\widehat{D}_n[\widehat{D}_n'M'(M\widehat{\Omega}_nM')^{-1}M\widehat{D}_n]^{-1}\widehat{D}_n'M'(M\widehat{\Omega}_nM')^{-1}\widehat{g}_n$ and likewise for $\mathrm{LM}_n^{\mathrm{GMM}}$.) Using these results, the version of Lemma 28.1 described in the previous paragraph applied to the transformed variables g_{Fi}^* and G_{Fi}^* establishes the second and third equalities of part (a) of Theorems 7.1 and 15.3 and part (b) of Theorems 7.1 and 15.3. □

Proof of Lemma 28.1. 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 (16.22)). Hence, by assumption, $q = p$. Given this, $Q_{2n}^+(\kappa)$ (defined in (26.11) in the proof of Theorem 16.6) is a scalar. In consequence, (26.13) and (26.16) with $j = p+1$ give

$$
\begin{aligned}
0 = \left|Q_{2n}^+(\widehat{\kappa}_{(p+1)n}^+)\right| = \left|M_{n,p+1-q}^+ - \widehat{\kappa}_{(p+1)n}^+\big(1+o_p(1)\big)\right| \quad \text{and, hence,}&\\
\widehat{\kappa}_{(p+1)n}^+ = M_{n,p+1-q}^+\big(1+o_p(1)\big)&\\
= \big(n^{1/2}B_{n,p+1-q}^{+\prime}U_n^{+\prime}\widehat{D}_n^{+\prime}W_n'\big)&\\
\times h_{3,k-q}h_{3,k-q}'\big(n^{1/2}W_n\widehat{D}_n^+U_n^+B_{n,p+1-q}^+\big)\big(1+o_p(1)\big)+o_p(1)&\\
= \big(n^{1/2}\widehat{g}_n'\widehat{\Omega}_n^{-1/2}\widehat{W}_n^{-1\prime}W_n'\big)&\\
\times h_{3,k-q}h_{3,k-q}'\big(n^{1/2}W_n\widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n\big)\big(1+o_p(1)\big)+o_p(1)&\\
= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}h_{3,k-q}h_{3,k-q}'\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1),&
\end{aligned}
\tag{28.2}
$$

where $\widehat{\kappa}_{(p+1)n}^+$ is defined in (26.2), the equality on the third line holds by the definition of $M_{n,p+1-q}^+$ in (26.16), the equality on the fourth line holds by lines two and three of (26.7) because when $q = p$ the third line of (26.7) becomes $n^{1/2}W_n\widehat{W}_n^{-1}\widehat{\Omega}_n^{-1/2}\widehat{g}_n$, that is, $n^{1/2}W_n\widehat{D}_nU_nB_{n,p-q}$ drops out, as noted near the end of the proof of Theorem 16.6, 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

$$
\begin{aligned}
\mathrm{QLR}_{\mathrm{WU},n} &:= \mathrm{AR}_n - \lambda_{\min}(n\widehat{Q}_{\mathrm{WU},n})\\
&= \mathrm{AR}_n - \widehat{\kappa}_{(p+1)n}^+\\
&= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}\big(I_k - h_{3,k-q}h_{3,k-q}'\big)\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1)\\
&= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}h_{3,q}h_{3,q}'\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1),
\end{aligned}
\tag{28.3}
$$

where the first equality holds by the definition of $\mathrm{QLR}_{\mathrm{WU},n}$ in (16.3), the second equality holds by the definition of $\widehat{\kappa}_{(p+1)n}^+$ in (26.2), the third equality holds by (28.2) and the definition $\mathrm{AR}_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n$ in (9), 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 16.4, we have

$$
\begin{aligned}
& n^{1/2} W_n \widehat{D}_n U_n T_n \to_d \overline{\Delta}_h = h_{3,q} \quad \text{and so} \\
& n^{1/2} \widehat{W}_n \widehat{D}_n U_n T_n \to_p h_{3,q},
\end{aligned} \tag{28.4}
$$

where the equality holds by the definition of $\overline{\Delta}_h$ in (16.24) when $q = p$ and the second convergence uses $W_n \widehat{W}_n^{-1} = I_k + o_p(1)$ by Assumption WU. In consequence,

$$
\begin{aligned}
& P_{\widehat{W}_n \widehat{D}_n} = P_{n^{1/2} \widehat{W}_n \widehat{D}_n U_n T_n} = P_{h_{3,q}} + o_p(1) = h_{3,q} h_{3,q}' + o_p(1) \quad \text{and} \\
& \mathrm{QLR}_{\mathrm{WU},n} = \mathrm{LM}_n(\widehat{W}_n) + o_p(1),
\end{aligned} \tag{28.5}
$$

where the first equality holds because $n^{1/2} U_n T_n$ is nonsingular wp $\to 1$ by Assumption WU and post-multiplication by a nonsingular matrix does not affect the resulting projection matrix, the second equality holds by (28.4), 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, (28.3), $n^{1/2} \widehat{\Omega}_n^{-1/2} \widehat{g}_n = O_p(1)$, and the definition of $\mathrm{LM}_n(\widehat{W}_n)$ in (28.1).

As in (25.5) in Section 25 with $\widehat{G}_n$ in place of $\widehat{D}_n$, we have

$$
\begin{aligned}
W_n \widehat{G}_n U_n B_{n,q} Y_{n,q}^{-1} &= W_n D_n U_n B_{n,q} Y_{n,q}^{-1} + W_n n^{1/2} (\widehat{G}_n - D_n) U_n B_{n,q} \big(n^{1/2} Y_{n,q} \big)^{-1} \\
&= C_{n,q} + o_p(1) \to_p h_{3,q},
\end{aligned} \tag{28.6}
$$

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 $j \le q$ (by the definition of q in (16.22)). The convergence in (28.6) holds by (16.19), (16.24), and (25.1). Using (28.6) in place of the first line of (28.4), the proof of $\mathrm{QLR}_{\mathrm{WU},n} = \mathrm{LM}_n^{\mathrm{GMM}}(\widehat{W}_n) + o_p(1)$ is the same as that given for $\mathrm{QLR}_{\mathrm{WU},n} = \mathrm{LM}_n(\widehat{W}_n) + o_p(1)$. This completes the proof of part (a) of Lemma 28.1.

By (27.10) in the proof of Theorem 27.1, we have

$$
\begin{aligned}
& c_{k,p} \big(n^{1/2} \widehat{W}_n \widehat{D}_n \widehat{U}_n, 1 - \alpha \big) \to_d c_{k,p,q}(\overline{\tau}_{[2]h}, 1 - \alpha) \quad \text{and} \\
& c_{k,p,q}(\overline{\tau}_{[2]h}, 1 - \alpha) = \chi^2_{p,1-\alpha} \quad \text{when } q = p,
\end{aligned} \tag{28.7}
$$

where the second line of (28.7) holds by the sentence following (27.9). This proves part (b) of Lemma 28.1 because convergence in distribution to a constant is equivalent to convergence in probability to the same constant. □

29. Proofs of Lemmas 19.1, 19.2, and 19.3

29.1 *Proof of Lemma 19.1*

Lemma 19.1. *Suppose Assumption* HLIV *holds. Under the null hypothesis* $H_0 : \theta = \theta_0$, *for any sequence of reduced-form parameters* $\{\pi_n \in \Pi : n \ge 1\}$ *and any* $p \ge 1$, *we have*: (a) $\widetilde{R}_n \to_p \Sigma_V \otimes K_Z$, (b) $\widehat{\Omega}_n \to_p (b_0' \Sigma_V b_0) K_Z$, *where* $b_0 := (1, -\theta_0')'$, (c) $\widetilde{\Sigma}_n \to_p (b_0' \Sigma_V b_0)^{-1} \Sigma_V$, (d) $\widetilde{\Sigma}_n^{\varepsilon} \to_p (b_0' \Sigma_V b_0)^{-1} \Sigma_V$, (e) $n^{1/2} \widehat{\Omega}_n^{-1/2} \widehat{g}_n = \overline{S}_n + o_p(1)$, *and* (f) $n^{1/2} \widetilde{D}_n^* = -(I_k + o_p(1)) \overline{T}_n (I_p + o_p(1)) + o_p(1)$.

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

The proof of Lemma 19.1 uses the following lemmas. Define

$$
\begin{aligned}
A_0^* &:= \Sigma_V B \begin{pmatrix} b_0'\Sigma_{V2}c_0, \ldots, b_0'\Sigma_{Vp+1}c_0 \\ I_p \end{pmatrix} \in R^{(p+1)\times p}, \\
B &:= \begin{pmatrix} 1 & 0_p' \\ -\theta_0 & -I_p \end{pmatrix} \in R^{(p+1)\times(p+1)}, \\
c_0 &:= \left(b_0'\Sigma_V b_0\right)^{-1}, \qquad b_0 := \left(1, -\theta_0'\right)', \qquad (29.1)\\
(\Sigma_{V1}, \ldots, \Sigma_{Vp+1}) &:= \Sigma_V \in R^{(p+1)\times(p+1)}, \quad \text{and} \\
L_{V0} &:= (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' \in R^{p\times p}.
\end{aligned}
$$

As defined in (19.4), $A_0 := (\theta_0, I_p)' \in R^{(p+1)\times p}$.

Lemma 29.1. $A_0^* L_{V0} = -A_0$.

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

Proof of Lemma 19.1. To prove part (a), we determine the probability limit of $\widetilde{V}_n$ defined in (15.5). By (15.5) and (19.1)–(19.3), in the linear IV regression model with reduced-form parameter π_n, we have

$$
\begin{aligned}
& u_i := u_i(\theta_0) = y_{1i} - Y_{2i}'\theta_0, \qquad Eu_i = 0, \\
& u_{\theta i} = -Y_{2i} = -\pi_n' Z_i - V_{2i}, \qquad Eu_{\theta i} = -\pi_n' Z_i, \\
& u_i^* := \begin{pmatrix} u_i \\ u_{\theta i} \end{pmatrix} = \begin{pmatrix} u_i \\ -Y_{2i} \end{pmatrix} = \Xi_n' Z_i + \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix}, \quad \text{where } \Xi_n = \left(0^k, -\pi_n\right) \in R^{k\times(p+1)}, \qquad (29.2)\\
& Eu_i^* = \Xi_n' Z_i,\ u_i^* - Eu_i^* = \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix} = B' V_i, \qquad \widehat{u}_{in}^* - Eu_i^* = (\widetilde{\Xi}_n - \Xi_n)' Z_i, \quad \text{and} \\
& U^* := \left(u_1^*, \ldots, u_n^*\right)' = Z_{n\times k}\Xi_n + VB, \quad \text{where } V := (V_1, \ldots, V_n)' \in R^{n\times(p+1)}
\end{aligned}
$$

and $B := B(\theta_0)$ is defined in (15.5).

Next, we have

$$\begin{aligned}\widetilde{\Xi}_n - \Xi_n &= \left(Z_{n\times k}' Z_{n\times k}\right)^{-1} Z_{n\times k}' U^* - \Xi_n \\ &= \left(n^{-1} Z_{n\times k}' Z_{n\times k}\right)^{-1} n^{-1} Z_{n\times k}' V B = O_p\left(n^{-1/2}\right), \end{aligned} \tag{29.3}$$

where the first equality holds by the definition of $\widetilde{\Xi}_n$ in (15.5), the second equality uses the last line of (29.2), and the third equality holds by Assumption HLIV(c) (specifically, $n^{-1} Z_{n\times k}' Z_{n\times k} \to K_Z$ 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 $n^{-1}\sum_{i=1}^n Z_{ij}^2 EV_{i\ell}^2 \to K_{Zjj} EV_{i\ell}^2 < \infty$ 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 $\widetilde{V}_n$ in (15.5) and simple algebra, we have

$$\begin{aligned}\widetilde{V}_n &:= n^{-1}\sum_{i=1}^n \left[\left(u_i^* - \widehat{u}_{in}^*\right)\left(u_i^* - \widehat{u}_{in}^*\right)' \otimes Z_i Z_i'\right] \\ &= n^{-1}\sum_{i=1}^n \left[\left(u_i^* - Eu_i^*\right)\left(u_i^* - Eu_i^*\right)' \otimes Z_i Z_i'\right] \\ &\quad - n^{-1}\sum_{i=1}^n \left[\left(\widehat{u}_{in}^* - Eu_i^*\right)\left(u_i^* - Eu_i^*\right)' \otimes Z_i Z_i'\right] \\ &\quad - n^{-1}\sum_{i=1}^n \left[\left(u_i^* - Eu_i^*\right)\left(\widehat{u}_{in}^* - Eu_i^*\right)' \otimes Z_i Z_i'\right] \\ &\quad + n^{-1}\sum_{i=1}^n \left[\left(\widehat{u}_{in}^* - Eu_i^*\right)\left(\widehat{u}_{in}^* - Eu_i^*\right)' \otimes Z_i Z_i'\right]. \end{aligned} \tag{29.4}$$

Using the third line of (29.2), the fourth summand on the rhs of (29.4) equals

$$n^{-1}\sum_{i=1}^n \left[(\widetilde{\Xi}_n - \Xi_n)' Z_i Z_i' (\widetilde{\Xi}_n - \Xi_n) \otimes Z_i Z_i'\right]. \tag{29.5}$$

The elements of the fourth summand on the rhs of (29.4) are each $o_p(1)$ because each is bounded by $O_p(n^{-1}) n^{-1}\sum_{i=1}^n \|Z_i\|^4$ using (29.3) and $n^{-1}\sum_{i=1}^n \|Z_i\|^4 \leq n^{-1}\sum_{i=1}^n \|Z_i\|^4 \times 1(\|Z_i\| > 1) + 1 \leq n^{-1}\sum_{i=1}^n \|Z_i\|^6 + 1 = o(n)$ by Assumption HLIV(c).

Using the third line of (29.2), the second summand on the rhs of (29.4) (excluding the minus sign) equals

$$n^{-1}\sum_{i=1}^n \left[(\widetilde{\Xi}_n - \Xi_n)' Z_i V_i' B \otimes Z_i Z_i'\right]. \tag{29.6}$$

The elements of the second summand on the rhs of (29.4) are each $o_p(1)$ because $\widetilde{\Xi}_n - \Xi_n = O_p(n^{-1/2})$ by (29.3) and for any $j_1, j_2, j_3 \leq k$ and $\ell \leq p$ we have $n^{-1}\sum_{i=1}^n Z_{ij_1} Z_{ij_2} \times Z_{ij_3} V_{i\ell} = o_p(n^{1/2})$ because its mean is zero and its variance is $EV_{i\ell}^2 n^{-1}\sum_{i=1}^n Z_{ij_1}^2 Z_{ij_2}^2 Z_{ij_3}^2 =$

$o(n)$ by Assumption HLIV(c). By the same argument, the elements of the third summand on the rhs of (29.4) are each $o_p(1)$.

In consequence, we have

$$\begin{aligned}
\widetilde{V}_n &= n^{-1}\sum_{i=1}^{n}\left[B'V_iV_i'B \otimes Z_iZ_i'\right] + o_p(1) \\
&= n^{-1}\sum_{i=1}^{n}\left[\left(B'V_iV_i'B - B'\Sigma_V B\right)\otimes Z_iZ_i'\right] + \left[B'\Sigma_V B \otimes n^{-1}\sum_{i=1}^{n} Z_iZ_i'\right] + o_p(1) \\
&\to_p B'\Sigma_V B \otimes K_Z, \qquad (29.7)
\end{aligned}$$

where the first equality holds using (29.4), the argument in the two paragraphs following (29.4), and the third line of (29.2), 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 (29.5)).

Equation (29.7) gives

$$\widetilde{R}_n := \left(B' \otimes I_k\right)\widetilde{V}_n(B \otimes I_k) \to_p \Sigma_V \otimes K_Z \qquad (29.8)$$

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

To prove part (b), we have

$$\begin{aligned}
\widehat{\Omega}_n &:= n^{-1}\sum_{i=1}^{n} g_ig_i' - \widehat{g}_n\widehat{g}_n' = n^{-1}\sum_{i=1}^{n} Eg_ig_i' + n^{-1}\sum_{i=1}^{n}\left(g_ig_i' - Eg_ig_i'\right) + O_p\left(n^{-1}\right) \\
&= n^{-1}\sum_{i=1}^{n} Z_iZ_i'Eu_i^2 + o_p(1) \to_p \left(b_0'\Sigma_V b_0\right)K_Z, \qquad (29.9)
\end{aligned}$$

where the first equality holds by the definition in (8), second equality uses $n^{1/2}\widehat{g}_n = O_p(1)$ by Lemma 29.2(a), the third equality holds by Lemma 29.2(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

$$\widetilde{\Sigma}_{j\ell n} = \mathrm{tr}\left(\widetilde{R}_{j\ell n}\widehat{\Omega}_n^{-1}\right)/k \to_p \mathrm{tr}\left(\Sigma_{Vj\ell}K_Z\left(b_0'\Sigma_V b_0\right)^{-1}K_Z^{-1}\right)/k = \Sigma_{Vj\ell}\left(b_0'\Sigma_V b_0\right)^{-1}, \qquad (29.10)$$

where $\widetilde{\Sigma}_{j\ell n}$ and $\Sigma_{Vj\ell}$ denote the (j,ℓ) elements of $\widetilde{\Sigma}_n$ and Σ_V, respectively, $\widetilde{R}_{j\ell n}$ denotes the (j,ℓ) submatrix of $\widetilde{R}_n$ of dimension $k \times k$, and the convergence holds because $\widetilde{R}_{j\ell n} \to_p \Sigma_{Vj\ell}K_Z$ for $j,\ell = 1,\ldots,p+1$ 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 $\widetilde{\Sigma}_n^{\varepsilon} \to_p ((b_0'\Sigma_V b_0)^{-1}\Sigma_V)^{\varepsilon}$ by part (c) of the lemma and Lemma 22.1(e), $((b_0'\Sigma_V b_0)^{-1}\Sigma_V)^{\varepsilon} = (b_0'\Sigma_V b_0)^{-1}\Sigma_V^{\varepsilon}$ by Lemma 22.1(d), and $\Sigma_V^{\varepsilon} = \Sigma_V$ by Assumption HLIV(e) and Comment (ii) to Lemma 22.1.

We prove part (f) next. We have

$$
\begin{aligned}
n^{-1}Z_{n\times k}'Y &= \left(n^{-1}\sum_{i=1}^{n} Z_i\big(y_{1i}-Y_{2i}'\theta_0\big)+n^{-1}\sum_{i=1}^{n} Z_iY_{2i}'\theta_0, n^{-1}\sum_{i=1}^{n} Z_iY_{2i}\right)\\
&= (\widehat{g}_n-\widehat{G}_n\theta_0,-\widehat{G}_n)=(\widehat{g}_n,\widehat{G}_n)\begin{pmatrix}1 & 0_p'\\ -\theta_0 & -I_p\end{pmatrix}=(\widehat{g}_n,\widehat{G}_n)B, \qquad (29.11)
\end{aligned}
$$

where the expressions for $\widehat{g}_n$ and $\widehat{G}_n$ use (19.3). Using (29.11) and the definition of L_{V0} in (29.1), the statistic $\overline{T}_n$ defined in (19.4) can be written as

$$
\begin{aligned}
\overline{T}_n &:= \big(Z_{n\times k}'Z_{n\times k}\big)^{-1/2}Z_{n\times k}'Y\Sigma_V^{-1}A_0\big(A_0'\Sigma_V^{-1}A_0\big)^{-1/2}\\
&= n^{1/2}\big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1/2}. \qquad (29.12)
\end{aligned}
$$

Note that, using the definitions of B and L_{V0} in (29.1) and A_0 in (19.4), the rhs expression for $\overline{T}_n$ equals the expression in (19.4).

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

$$
\begin{aligned}
\widehat{\Gamma}_{jn} &:= n^{-1}\sum_{i=1}^{n}(G_{ij}-\widehat{G}_{jn})g_i' = n^{-1}\sum_{i=1}^{n}EG_{ij}g_i' + n^{-1}\sum_{i=1}^{n}\big(G_{ij}g_i'-EG_{ij}g_i'\big)-\widehat{G}_{jn}\widehat{g}_n'\\
&= n^{-1}\sum_{i=1}^{n}EG_{ij}g_i'+o_p(1) = -n^{-1}\sum_{i=1}^{n}EZ_iY_{2ij}Z_i'u_i+o_p(1)\\
&= -n^{-1}\sum_{i=1}^{n}Z_iZ_i'EV_{2ij}V_i'b_0+n^{-1}\sum_{i=1}^{n}Z_iZ_i'\big(Z_i'\pi_{jn}\big)Eu_i+o_p(1)\\
&= -n^{-1}\sum_{i=1}^{n}Z_iZ_i'\Sigma_{Vj+1}'b_0+o_p(1), \qquad (29.13)
\end{aligned}
$$

where $g_i = Z_i(y_{1i}-Y_{2i}'\theta_0) = Z_iu_i$ by (19.3), the third equality holds by Lemma 29.2(a)–(c), the fourth equality holds by (19.3) 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 $\Sigma_V := (\Sigma_{V1},\ldots,\Sigma_{Vp+1}) := EV_iV_i'$.

Equations (29.9) and (29.13) give

$$
\begin{aligned}
\widehat{D}_{jn} &:= \widehat{G}_{jn}-\widehat{\Gamma}_{jn}\widehat{\Omega}_n^{-1}\widehat{g}_n = \widehat{G}_{jn}+\Sigma_{Vj+1}'b_0\big(b_0'\Sigma_Vb_0\big)^{-1}\widehat{g}_n+o_p\big(n^{-1/2}\big) \quad \text{and}\\
\widehat{D}_n &:= (\widehat{D}_{1n},\ldots,\widehat{D}_{pn}) = (\widehat{g}_n,\widehat{G}_n)\begin{pmatrix}\Sigma_{V2}'b_0c_0,\ldots,\Sigma_{Vp+1}'b_0c_0\\ I_p\end{pmatrix}+o_p\big(n^{-1/2}\big)\\
&= (\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}\left(\Sigma_VB\begin{pmatrix}\Sigma_{V2}'b_0c_0,\ldots,\Sigma_{Vp+1}'b_0c_0\\ I_p\end{pmatrix}\right)+o_p\big(n^{-1/2}\big)\\
&= (\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^*+o_p\big(n^{-1/2}\big),
\end{aligned} \qquad (29.14)
$$

where the second equality on the first line uses $\widehat{g}_n = O_p(n^{-1/2})$ by Lemma 29.2(a), the second line uses $c_0 = (b_0'\Sigma_V b_0)^{-1}$, the second last equality holds because $B^{-1} = B$, and the last equality holds by the definition of A_0^* in (29.1).

Now, we have

$$\begin{aligned} n^{1/2}\widetilde{D}_n^* &:= n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widehat{L}_n^{1/2} \\ &= \big(b_0'\Sigma_V b_0\big)^{-1/2}\big(I_k + o_p(1)\big)\big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1/2}n^{1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0^* \\ &\quad \times \big(b_0'\Sigma_V b_0\big)^{1/2}L_{V0}^{1/2}\big(I_p + o_p(1)\big) + o_p(1) \\ &= -\big(I_k + o_p(1)\big)\big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1/2}n^{1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1/2}\big(I_p + o_p(1)\big) \\ &\quad + o_p(1) \\ &= -\big(I_k + o_p(1)\big)\overline{T}_n\big(I_p + o_p(1)\big) + o_p(1), \end{aligned} \tag{29.15}$$

where the first equality holds by the definition of $\widetilde{D}_n^*$ in (23), the second equality holds by (29.14), $\widehat{\Omega}_n \to_p (b_0'\Sigma_V b_0)K_Z$ (which holds by part (b) of the lemma), and $\widetilde{L}_n := (\theta_0, I_p)(\widetilde{\Sigma}_n^{\varepsilon})^{-1}(\theta_0, I_p)' \to_p (b_0'\Sigma_V b_0)L_{V0}$ (which holds because $\widetilde{\Sigma}_n^{\varepsilon} \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_V$ by part (d) of the lemma), for $L_{V0} := (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)'$ defined in (29.1), the third equality holds by Lemma 29.1, and the last equality holds by (29.12). This completes the proof of part (f).

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

$$\begin{aligned} \overline{S}_n &:= \big(Z_{n\times k}'Z_{n\times k}\big)^{-1/2}Z_{n\times k}'Yb_0\big(b_0'\Sigma_V b_0\big)^{-1/2} \\ &= n^{1/2}\left(n^{-1}\sum_{i=1}^n Z_iZ_i'\right)^{-1/2}\widehat{g}_n\big(b_0'\Sigma_V b_0\big)^{-1/2} \\ &= n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1), \end{aligned} \tag{29.16}$$

where the first equality holds by the definition of $\overline{S}_n$ in (19.4), the second equality holds because $Y_i'b_0 = u_i$, and the third equality holds by (29.9) and $n^{1/2}\widehat{g}_n = O_p(1)$ by Lemma 29.2(a). This proves part (e). □

Proof of Lemma 29.1. By premultiplying by $B\Sigma_V^{-1}$, the equation $A_0^*L_{V0} = -A_0$ is seen to be equivalent to

$$\begin{pmatrix} b_0'\Sigma_{V2}c_0, \ldots, b_0'\Sigma_{Vp+1}c_0 \\ I_p \end{pmatrix} L_{V0} = -B\Sigma_V^{-1}\begin{pmatrix} \theta_0' \\ I_p \end{pmatrix} = \begin{pmatrix} -1 & 0^{p\prime} \\ \theta_0 & I_p \end{pmatrix}\Sigma_V^{-1}\begin{pmatrix} \theta_0' \\ I_p \end{pmatrix}. \tag{29.17}$$

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

$$L_{V0} = (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)', \tag{29.18}$$

which hold by the definition of L_{V0} in (29.1).

Substituting in the definition of L_{V0}, the first row of the equations in (29.17) is

$$\big(b_0'\Sigma_{V2}c_0, \dots, b_0'\Sigma_{Vp+1}c_0\big)(\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' = \big(-1, 0^{p\prime}\big)\Sigma_V^{-1}(\theta_0, I_p)'. \tag{29.19}$$

Equation (29.19) 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 (29.19) equals

$$\begin{aligned}
&\big(b_0'\Sigma_{V2}^*\theta_0 c_0, b_0'\Sigma_{V2}c_0, \dots, b_0'\Sigma_{Vp+1}c_0\big)\Sigma_V^{-1}(\theta_0, I_p)' \\
&\quad = \big(\big(-b_0'\Sigma_V b_0 + b_0'\Sigma_{V1}\big)c_0, b_0'\Sigma_{V2}c_0, \dots, b_0'\Sigma_{Vp+1}c_0\big)\Sigma_V^{-1}(\theta_0, I_p)' \\
&\quad = \big(-1 + b_0'\Sigma_{V1}c_0, b_0'\Sigma_{V2}c_0, \dots, b_0'\Sigma_{Vp+1}c_0\big)\Sigma_V^{-1}(\theta_0, I_p)',
\end{aligned} \tag{29.20}$$

where the second equality uses the definition of c_0 in (29.1).

Hence, the difference between the left-hand side (lhs) and the rhs of (29.19) equals

$$\big(b_0'\Sigma_{V1}c_0, \dots, b_0'\Sigma_{Vp+1}c_0\big)\Sigma_V^{-1}(\theta_0, I_p)' = c_0 b_0'\Sigma_V\Sigma_V^{-1}\begin{pmatrix}\theta_0' \\ I_p\end{pmatrix} = 0_p' \tag{29.21}$$

using $b_0' := (1, -\theta_0')$. Thus, (29.19) holds, which completes the proof. □

Proof of Lemma 29.2. Part (a) holds by the CLT of Eicker (1963, Theorem 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

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n \big(G_{ij}g_i' - EG_{ij}g_i'\big) \\
&\quad = -n^{-1}\sum_{i=1}^n Z_i Z_i'(Y_{2ij}u_i - EY_{2ij}u_i) \\
&\quad = -n^{-1}\sum_{i=1}^n Z_i Z_i'\big(Z_i'\pi_{jn}\big)u_i - n^{-1}\sum_{i=1}^n Z_i Z_i'\big(V_{2ij}u_i - \Sigma_{Vj+1}'b_0\big),
\end{aligned} \tag{29.22}$$

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 $n^{-2}\sum_{i=1}^n (Z_{i\ell_1}Z_{i\ell_2}Z_i'\pi_{jn})^2 \times \mathrm{Var}(u_i)$, which converges to zero for all $\ell_1, \ell_2 \le k$ because $n^{-1}\sum_{i=1}^n \|Z_i\|^6 = o(n)$, $\mathrm{Var}(u_i) = b_0'\Sigma_V b_0 < \infty$, 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 $n^{-2}\sum_{i=1}^n Z_{i\ell_1}^2 Z_{i\ell_2}^2 \mathrm{Var}(V_{2ij}u_i)$, 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 $\mathrm{Var}(V_{2ij}u_i) \le E(V_{2ij}V_i'b_0)^2 \le b_0'b_0 E\|V_i\|^4 < \infty$ by Assumptions HLIV(b)–(c). This establishes part (b).

For part (c), we have

$$\widehat{G}_n = -n^{-1}\sum_{i=1}^{n} Z_i Y_{2i}' = -n^{-1}\sum_{i=1}^{n} Z_i Z_i'\pi_n - n^{-1}\sum_{i=1}^{n} Z_i V_{2i}'. \tag{29.23}$$

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 $\ell \leq k$ and $j \leq p$ 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

$$n^{-1}\sum_{i=1}^{n}\big(g_i g_i' - E g_i g_i'\big) = n^{-1}\sum_{i=1}^{n} Z_i Z_i'\big(u_i^2 - Eu_i^2\big) \to_p 0, \tag{29.24}$$

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 $n^{-1}\sum_{i=1}^n (Z_{i\ell_1}^2 Z_{i\ell_2}^2 \mathrm{Var}((V_i' b_0)^2) \leq n^{-1}\sum_{i=1}^n \|Z_i\|^4 E\|V_i\|^4 \|b_0\|^4 < \infty$ by Assumption HLIV(b)–(c) for all $\ell_1, \ell_2 \leq k$. This proves part (d).

Part (e) holds by the following argument:

$$\widehat{G}_n - n^{-1}\sum_{i=1}^{n} EG_i = -n^{-1}\sum_{i=1}^{n} Z_i(Y_{2i} - EY_{2i})' = -n^{-1}\sum_{i=1}^{n} Z_i V_{2i}' = O_p\big(n^{-1/2}\big), \tag{29.25}$$

where the last equality holds by the argument following (29.23). □

29.2 *Proof of Lemma 19.2*

LEMMA 19.2. *Suppose Assumptions* HLIV *and* HLIV2 *hold. Under the null hypothesis* $H_0 : \theta = \theta_0$ *and any* $p \geq 1$, *we have*: (a) $\widehat{R}_n \to_p R(\pi_*)$, (b) $\widehat{\Sigma}_n \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$, (c) $\widehat{\Sigma}_n^{\varepsilon} \to_p (b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$, *and* (d) $n^{1/2}\widehat{D}_n^* = -(I_k + o_p(1))\overline{T}_n(L_{V0}^{-1/2}L_{V*}^{1/2} + o_p(1)) + o_p(1)$, *where* $L_{V0} := (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' \in R^{p\times p}$ *and* $L_{V*} := (\theta_0, I_p)\Sigma_{V*}^{-1}(\theta_0, I_p)' \in R^{p\times p}$.

PROOF OF LEMMA 19.2. To prove part (a), we determine the probability limit of $\widehat{V}_n$ defined in (19), where $f_i = (Z_i'u_i, -\mathrm{vec}(Z_i Y_{2i}')')'$ by (19.1) and (19.3). For $\zeta_n(\pi)$ defined in (19.6), we can write

$$\zeta_n(\pi_n) = n^{-1}\sum_{i=1}^{n} Z_{ni}^* Z_{ni}^{*\prime}, \quad \text{where}$$

$$Z_{ni}^* := \mathrm{vec}\left(Z_i Z_i'\pi_n - n^{-1}\sum_{\ell=1}^{n} Z_\ell Z_\ell'\pi_n\right) \tag{29.26}$$

$$= \big(\pi_n' \otimes Z_i\big)Z_i - n^{-1}\sum_{\ell=1}^{n}\big(\pi_n' \otimes Z_\ell\big)Z_\ell \in R^{kp}$$

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

We have

$$\begin{aligned}
\widehat{V}_n &:= n^{-1}\sum_{i=1}^{n}\left(f_i - n^{-1}\sum_{\ell=1}^{n} Ef_\ell\right)\left(f_i - n^{-1}\sum_{\ell=1}^{n} Ef_\ell\right)' \\
&\quad - \left(\widehat{f}_n - n^{-1}\sum_{\ell=1}^{n} Ef_\ell\right)\left(\widehat{f}_n - n^{-1}\sum_{\ell=1}^{n} Ef_\ell\right)' \\
&= n^{-1}\sum_{i=1}^{n}\begin{pmatrix} Z_i u_i \\ -\mathrm{vec}\big(Z_i V_{2i}'\big) - Z_{ni}^* \end{pmatrix}\begin{pmatrix} Z_i u_i \\ -\mathrm{vec}\big(Z_i V_{2i}'\big) - Z_{ni}^* \end{pmatrix}' + o_p(1) \\
&= n^{-1}\sum_{i=1}^{n}\left(\begin{pmatrix} u_i \\ -V_{2i}\end{pmatrix}\begin{pmatrix} u_i \\ -V_{2i}\end{pmatrix}' \otimes Z_i Z_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta_n(\pi_n)\end{pmatrix} \\
&\quad + n^{-1}\sum_{i=1}^{n}\begin{pmatrix} Z_i u_i \\ -\mathrm{vec}\big(Z_i V_{2i}'\big)\end{pmatrix}\begin{pmatrix} 0^k \\ -Z_{ni}^*\end{pmatrix}' + n^{-1}\sum_{i=1}^{n}\begin{pmatrix} 0^k \\ -Z_{ni}^*\end{pmatrix}\begin{pmatrix} Z_i u_i \\ -\mathrm{vec}\big(Z_i V_{2i}'\big)\end{pmatrix}' + o_p(1) \\
&= \left(\begin{pmatrix} 1 & -\theta_0' \\ 0^p & -I_p\end{pmatrix}\Sigma_V\begin{pmatrix} 1 & -\theta_0' \\ 0^p & -I_p\end{pmatrix}'\right) \otimes \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi_*)\end{pmatrix} + o_p(1) \\
&= \big(B'\Sigma_V B\big) \otimes \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right) + \begin{pmatrix} 0^{k\times k} & 0^{k\times kp} \\ 0^{kp\times k} & \zeta(\pi_*)\end{pmatrix} + o_p(1), \qquad (29.27)
\end{aligned}$$

where the second equality holds using $Eu_i = 0$, $EV_{2i} = 0^p$, $Y_{2i} = \pi_n' Z_i + V_{2i}$, $\mathrm{vec}(Z_i Y_{2i}' - n^{-1}\sum_{\ell=1}^n EZ_\ell Y_{2\ell}') = \mathrm{vec}(Z_i V_{2i}') + Z_{ni}^*$, and Lemma 29.2(a) and (e) because $\widehat{f}_n - n^{-1}\sum_{\ell=1}^n Ef_\ell = (\widehat{g}_n', \mathrm{vec}(\widehat{G}_n - n^{-1}\sum_{\ell=1}^n EG_\ell)')'$, the third equality holds by (29.26) 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 $n^{-1}\sum_{i=1}^n \|Z_i\|^4 \le n^{-1}\sum_{i=1}^n \|Z_i\|^4 1(\|Z_i\| > 1) + 1 \le n^{-1}\sum_{i=1}^n \|Z_i\|^6 + 1 = o(n)$ 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 $n^{-2}\sum_{i=1}^n \|Z_{ni}^*\|^2\|Z_i\|^2 \le \|\pi_n\|^2(n^{-2}\sum_{i=1}^n \|Z_i\|^6 + 2n^{-2}\sum_{i=1}^n \|Z_i\|^4 n^{-1}\sum_{\ell=1}^n \|Z_\ell\|^2 + n^{-2}\sum_{i=1}^n \|Z_i\|^2(n^{-1}\sum_{\ell=1}^n \|Z_\ell\|^2)^2) = o(1)$, using $\|Z_{ni}^*\| \le \|\pi_n\|(\|Z_i\|^2 + n^{-1}\sum_{\ell=1}^n \|Z_\ell\|^2)$, $\sup_{\pi\in\Pi}\|\pi_n\| < \infty$, and $E\|V_i\|^2 < \infty$ by Assumption HLIV(b)–(d)), and the fifth equality holds by the definition of B in (19).

Using the definitions of $\widehat{R}_n$ in (19) and $R(\pi_*)$ in (19.7), part (a) of the lemma follows from (29.27).

Next, we prove part (b). We have

$$\widehat{\Sigma}_{j\ell n} = \mathrm{tr}\big(\widehat{R}_{j\ell n}'\widehat{\Omega}_n^{-1}\big)/k \to_p \mathrm{tr}\big(R_{j\ell}(\pi_*)'\big(b_0'\Sigma_V b_0\big)^{-1} K_Z^{-1}\big)/k =: \big(b_0'\Sigma_V b_0\big)^{-1}\Sigma_{V*j\ell}, \qquad (29.28)$$

where $\widehat{\Sigma}_{j\ell n}$ and $\Sigma_{V*j\ell}$ denote the (j,ℓ) elements of $\widehat{\Sigma}_n$ and Σ_{V*}, respectively, $\widehat{R}'_{j\ell n}$ and $R_{j\ell}(\pi_*)$ denote the (j,ℓ) submatrices of dimension $k\times k$ of $\widehat{R}'_n$ and $R(\pi_*)$, respectively, the convergence holds by part (a) of the lemma and Lemma 19.1(b), and the last equality holds by the definition of $\Sigma_{V*j\ell}$ in (19.8). Equation (29.28) establishes part (b).

Part (c) holds because part (b) of the lemma and Lemma 22.1(e) imply that $\widehat{\Sigma}_n^{\varepsilon}\to_p ((b_0'\Sigma_V b_0)^{-1}\Sigma_{V*})^{\varepsilon}$, Lemma 22.1(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

$$\begin{aligned}
&n^{1/2}\widehat{D}_n^*\\
&\quad:=n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\widehat{L}_n^{1/2}\\
&\quad=\left((b_0'\Sigma_V b_0K_Z)^{-1/2}K_Z^{1/2}+o_p(1)\right)\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}n^{1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^*L_{V0}^{1/2}\\
&\qquad\times\left(L_{V0}^{-1/2}(b_0'\Sigma_V b_0L_{V*})^{1/2}+o_p(1)\right)+o_p(1)\\
&\quad=-\left(I_k+o_p(1)\right)\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}n^{1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1/2}\\
&\qquad\times\left(L_{V0}^{-1/2}L_{V*}^{1/2}+o_p(1)\right)+o_p(1)\\
&\quad=-\left(I_k+o_p(1)\right)\overline{T}_n\left(L_{V0}^{-1/2}L_{V*}^{1/2}+o_p(1)\right)+o_p(1),
\end{aligned}\tag{29.29}$$

where the first equality holds by the definition of $\widehat{D}_n^*$ in (23), the second equality holds by (i) (29.14), (ii) the result of part (c) of the lemma that $\widehat{\Sigma}_n^{\varepsilon}\to_p(b_0'\Sigma_V b_0)^{-1}\Sigma_{V*}$, (iii) the result of Lemma 19.1(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) $\widehat{L}_n:=(\theta_0,I_p)(\widehat{\Sigma}_n^{\varepsilon})^{-1}(\theta_0,I_p)'$ as defined in (23) with $\theta=\theta_0$, and (vi) $\widehat{L}_n\to_p b_0'\Sigma_V b_0L_{V*}$ for L_{V*} defined in part (d) of the lemma, the third equality holds by Lemma 29.1, and the last equality holds by (29.12). This completes the proof of part (d). □

29.3 *Proof of Lemma 19.3*

Lemma 19.3. *Suppose Assumption* HLIV *holds and* $p=1$. *Under the null hypothesis* $H_0:\theta=\theta_0$, *for any sequence of reduced-form parameters* $\{\pi_n\in\Pi:n\geq1\}$, *we have*: (a) $\mathrm{rk}_{1n}(\theta_0)=\overline{T}_n'[I_k+L_{V0}K_Z^{-1/2}\zeta_n(\pi_n)K_Z^{-1/2}+o_p(1)]^{-1}\overline{T}_n\cdot(1+o_p(1))+o_p(1)$, (b) $\mathrm{rk}_{2n}(\theta_0)=\overline{T}_n'\overline{T}_n(L_{V0}b_0'\Sigma_V b_0)^{-1}\cdot(1+o_p(1))+o_p(1)$, *where* $L_{V0}:=(\theta_0,1)\Sigma_V^{-1}(\theta_0,1)'\in R$, *and* (c) $L_{V0}b_0'\Sigma_V b_0=\frac{(1-2\theta_0\rho c+\theta_0^2c^2)^2}{c^2(1-\rho^2)}$, *where* $c^2:=\mathrm{Var}(V_{2i})/\mathrm{Var}(V_{1i})>0$ *and* $\rho=\mathrm{Corr}(V_{1i},V_{2i})\in(-1,1)$.

When $p=1$, we write

$$\Sigma_V:=EV_iV_i':=(\Sigma_{V1},\Sigma_{V2}):=\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}\in R^{2\times2}\tag{29.30}$$

for $\Sigma_{V1},\Sigma_{V2}\in R^2$, using the definition in (19.2).

The proof of Lemma 19.3 uses the following lemma.

Lemma 29.3. *Under the conditions of Lemma* 19.3, (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 19.3. We prove part (b) first. By (29.9) and (29.14),

$$\begin{aligned}
&n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n\\
&\quad=n^{1/2}\big(I_k+o_p(1)\big)\big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0^*\big(b_0'\Sigma_V b_0\big)^{-1/2}+o_p(1)\\
&\quad=-n^{1/2}\big(I_k+o_p(1)\big)\big(n^{-1}Z_{n\times k}'Z_{n\times k}\big)^{-1/2}(\widehat{g}_n,\widehat{G}_n)B\Sigma_V^{-1}A_0L_{V0}^{-1}\big(b_0'\Sigma_V b_0\big)^{-1/2}\\
&\qquad+o_p(1)\\
&\quad=-\big(I_k+o_p(1)\big)\overline{T}_n\big(L_{V0}b_0'\Sigma_V b_0\big)^{-1/2}+o_p(1),
\end{aligned}\tag{29.31}$$

where the second equality holds by Lemma 29.1 and the third equality holds by (29.12). 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

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n(G_i-\widehat{G}_n)(G_i-\widehat{G}_n)'\\
&\quad=n^{-1}\sum_{i=1}^n\Bigg(G_i-n^{-1}\sum_{\ell=1}^nEG_\ell\Bigg)\Bigg(G_i-n^{-1}\sum_{\ell=1}^nEG_\ell\Bigg)'\\
&\qquad-\Bigg(\widehat{G}_n-n^{-1}\sum_{i=1}^nEG_i\Bigg)\Bigg(\widehat{G}_n-n^{-1}\sum_{i=1}^nEG_i\Bigg)'\\
&\quad=n^{-1}\sum_{i=1}^n\Bigg(-Z_iZ_i'\pi_n-Z_iV_{2i}+n^{-1}\sum_{\ell=1}^nZ_\ell Z_\ell'\pi_n\Bigg)\\
&\qquad\times\Bigg(-Z_iZ_i'\pi_n-Z_iV_{2i}+n^{-1}\sum_{\ell=1}^nZ_\ell Z_\ell'\pi_n\Bigg)'+o_p(1)\\
&\quad=n^{-1}\sum_{i=1}^n(Z_iV_{2i})(Z_iV_{2i})'+2n^{-1}\sum_{i=1}^n\big(Z_iZ_i'\pi_n\big)(Z_iV_{2i})'\\
&\qquad-2\Bigg(n^{-1}\sum_{\ell=1}^nZ_\ell Z_\ell'\pi_n\Bigg)\Bigg(n^{-1}\sum_{\ell=1}^nZ_iV_{2i}\Bigg)'\\
&\qquad+\zeta_n(\pi_n)+o_p(1)\\
&\quad=n^{-1}Z_{n\times k}'Z_{n\times k}\sigma_2^2+\zeta_n(\pi_n)+o_p(1),
\end{aligned}\tag{29.32}$$

where the first equality holds by algebra, the second equality holds by Lemma 29.2(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

(19.15), 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 29.2(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 $4\sigma_2^2 n^{-2}\sum_{i=1}^n \|Z_i\|^6 \sup_{\pi\in\Pi}\|\pi\|^2$, 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 $n^{-1}\sum_{\ell=1}^n Z_i V_{2i} = o_p(1)$ by the argument following (29.23).

Combining (29.9), (29.13), (29.32) and the definition of $\widetilde{V}_{Dn}$ in (19.14), we obtain

$$\begin{aligned}\widetilde{V}_{Dn} &= n^{-1}\sum_{i=1}^n Z_i Z_i'\left(\sigma_2^2 - \left(b_0'\Sigma_{V2}\right)^2\left(b_0'\Sigma_V b_0\right)^{-1}\right) + \zeta_n(\pi_n) + o_p(1)\\ &= K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1), \qquad (29.33)\end{aligned}$$

where the second equality holds by Lemma 29.3(c) and Assumption HLIV(c).

Next, we have

$$\begin{aligned}&n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}\widehat{D}_n L_{V0}^{1/2}\\ &= n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0^* L_{V0}^{1/2} + o_p(1)\\ &= -n^{1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{-1/2}(\widehat{g}_n, \widehat{G}_n)B\Sigma_V^{-1}A_0 L_{V0}^{-1/2} + o_p(1) = -\overline{T}_n + o_p(1), \qquad (29.34)\end{aligned}$$

where the first equality holds by (29.14), the second equality holds by Lemma 29.1, and the third equality holds by (29.12).

Using (29.33), we obtain

$$\begin{aligned}n^{1/2}\widetilde{V}_{Dn}^{-1/2}\widehat{D}_n &= \left[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)\right]^{-1/2} n^{1/2}\widehat{D}_n\\ &= -\left[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)\right]^{-1/2}\left(n^{-1}Z_{n\times k}'Z_{n\times k}\right)^{1/2}\overline{T}_n L_{V0}^{-1/2} + o_p(1)\\ &= -\left[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)\right]^{-1/2} K_Z^{1/2}\overline{T}_n L_{V0}^{-1/2}\left(1 + o_p(1)\right) + o_p(1), \qquad (29.35)\end{aligned}$$

where the second equality holds using (29.34) and Assumption HLIV(c), the third equality holds by Assumption HLIV(c) and some calculations. Using this, we obtain

$$\begin{aligned}\mathrm{rk}_{1n} &:= n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n\\ &= \overline{T}_n' K_Z^{1/2}\left[K_Z L_{V0}^{-1} + \zeta_n(\pi_n) + o_p(1)\right]^{-1} K_Z^{1/2}\overline{T}_n L_{V0}^{-1}\left(1 + o_p(1)\right) + o_p(1)\\ &= \overline{T}_n'\left[I_k + L_{V0}K_Z^{-1/2}\zeta_n(\pi_n)K_Z^{-1/2} + o_p(1)\right]^{-1}\overline{T}_n\left(1 + o_p(1)\right) + o_p(1), \qquad (29.36)\end{aligned}$$

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

Part (c) of the lemma follows from Lemma 29.3(a) and (b) by substituting in $\sigma_2^2 = c^2\sigma_1^2$. $\square$

PROOF OF LEMMA 29.3. Part (a) holds by the following calculations:

$$\begin{aligned} L_{V0} &:= (\theta_0, 1)\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}^{-1}\begin{pmatrix} \theta_0 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}(\theta_0, 1)\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}\begin{pmatrix} \theta_0 \\ 1 \end{pmatrix} \\ &= \frac{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2}{\sigma_1^2\sigma_2^2(1-\rho^2)}. \end{aligned} \tag{29.37}$$

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:

$$\begin{aligned} & b_0'\Sigma_V b_0 := (1, -\theta_0)\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\begin{pmatrix} 1 \\ -\theta_0 \end{pmatrix} = \sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2 \quad \text{and} \\ & b_0'\Sigma_{V2} := (1, -\theta_0)\big(\rho\sigma_1\sigma_2, \sigma_2^2\big)' = \rho\sigma_1\sigma_2 - \theta_0\sigma_2^2. \end{aligned} \tag{29.38}$$

Using (29.38), we obtain

$$\begin{aligned} \sigma_2^2 - \big(b_0'\Sigma_{V2}\big)^2\big(b_0'\Sigma_V b_0\big)^{-1} &= \sigma_2^2 - \frac{\big(\rho\sigma_1\sigma_2 - \theta_0\sigma_2^2\big)^2}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} \\ &= \frac{\sigma_1^2\sigma_2^2 - 2\theta_0\rho\sigma_1\sigma_2^3 + \theta_0^2\sigma_2^4 - \big(\rho\sigma_1\sigma_2 - \theta_0\sigma_2^2\big)^2}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} \\ &= \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 - 2\theta_0\rho\sigma_1\sigma_2 + \theta_0^2\sigma_2^2} = L_{V0}^{-1}, \end{aligned} \tag{29.39}$$

which proves part (c). □

30. PROOF OF THEOREM 18.1

In Sections 16 and 17, we establish Theorems 6.1 and 15.2 by first establishing Theorem 16.1, which concerns non-SR versions of the AR, CQLR, and CQLR$_P$ tests and employs the parameter spaces $\mathcal{F}_{\text{AR}}$, $\mathcal{F}$, and $\mathcal{F}_P$, rather than $\mathcal{F}_{\text{AR}}^{\text{SR}}$, $\mathcal{F}^{\text{SR}}$, and $\mathcal{F}_P^{\text{SR}}$. We prove Theorem 18.1 here using the same two-step approach.

In the time series context, the non-SR version of the AR statistic is defined as in (9) based on $\{f_i - \widehat{f}_n : i \leq n\}$, but with $\widehat{\Omega}_n$ defined in (18.3) and Assumption Ω below, rather than in (8), and the critical value is $\chi^2_{k,1-\alpha}$. The non-SR QLR time series test statistic and conditional critical value are defined as in Section 5.1, but with $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined in (18.3) and Assumption V below based on $\{f_i - \widehat{f}_n : i \leq n\}$, in place of $\widehat{V}_n$ and $\widehat{\Omega}_n$ defined in (19) and (8), respectively. The non-SR QLR$_P$ time series test statistic and conditional critical value are defined as in Section 15, but with $\widetilde{V}_n$ and $\widehat{\Omega}_n$ defined in (18.3) and Assumption V$_P$ below based on $\{(u_i^* - \widehat{u}_{in}^*) \otimes Z_i : i \leq n\}$, rather than in (15.5) and (8), respectively.

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

$$\begin{aligned}\mathcal{F}_{\text{TS,AR}} := \Big\{F : \{W_i : i = \cdots, 0, 1, \ldots\} &\text{ are stationary and strong mixing under } F \text{ with}\\ &\text{strong mixing numbers } \big\{\alpha_F(m) : m \geq 1\big\} \text{ that satisfy } \alpha_F(m) \leq Cm^{-d},\\ &E_F g_i = 0^k, E_F\|g_i\|^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(\Omega_F) \geq \delta\Big\}\end{aligned} \tag{30.1}$$

for some $\gamma, \delta > 0$, $d > (2+\gamma)/\gamma$, and $C, M < \infty$, where Ω_F is defined in (18.4). We define $\mathcal{F}_{\text{TS}}$ and $\mathcal{F}_{\text{TS},P}$ as $\mathcal{F}$ and $\mathcal{F}_P$ are defined in (16.1), respectively, but with $\mathcal{F}_{\text{TS,AR}}$ in place of $\mathcal{F}_{\text{AR}}$. For CSs, we use the corresponding parameter spaces $\mathcal{F}_{\text{TS},\Theta,\text{AR}} := \{(F, \theta_0) : F \in \mathcal{F}_{\text{TS,AR}}(\theta_0), \theta_0 \in \Theta\}$, $\mathcal{F}_{\text{TS},\Theta} := \{(F, \theta_0) : F \in \mathcal{F}_{\text{TS}}(\theta_0), \theta_0 \in \Theta\}$, and $\mathcal{F}_{\text{TS},\Theta,P} := \{(F, \theta_0) : F \in \mathcal{F}_{\text{TS},P}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{\text{TS,AR}}(\theta_0)$, $\mathcal{F}_{\text{TS}}(\theta_0)$, and $\mathcal{F}_{\text{TS},P}(\theta_0)$ denote $\mathcal{F}_{\text{TS,AR}}$, $\mathcal{F}_{\text{TS}}$, and $\mathcal{F}_{\text{TS},P}$, respectively, with their dependence on θ_0 made explicit.

For the (non-SR) CQLR 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 $\{F_n \in \mathcal{F}_{\text{TS},P} : n \geq 1\}$ 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 $\{(F_n, \theta_{0n}) \in \mathcal{F}_{\text{TS},\Theta,P} : n \geq 1\}$ 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$_P$ test and CS, we use Assumptions V$_P$ and V$_P$-CS, which are defined to be the same as Assumptions V and V-CS, respectively, but with $\mathcal{F}_{\text{TS},P}$ and $\mathcal{F}_{\text{TS},\Theta,P}$ in place of $\mathcal{F}_{\text{TS}}$ and $\mathcal{F}_{\text{TS},\Theta}$.

For the (non-SR) AR test and CS, we use Assumptions Ω and Ω-CS, which are defined as follows.

Assumption Ω. *$\widehat{\Omega}_n(\theta_0) - \Omega_{F_n,n}(\theta_0) \to_p 0^{k\times k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{\text{TS,AR}} : n \geq 1\}$ 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 $r \in \{1, \ldots, k\}$.*

Assumption Ω-CS is the same as Assumption Ω, but with θ_{0n} and $\mathcal{F}_{\text{TS},\Theta,\text{AR}}$ in place of θ_0 and $\mathcal{F}_{\text{TS,AR}}$.

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

Theorem 30.1. *Suppose the* AR, CQLR, *and* CQLR$_P$ *tests are defined as above, the parameter spaces for F are $\mathcal{F}_{\text{TS,AR}}$, $\mathcal{F}_{\text{TS}}$, and $\mathcal{F}_{\text{TS},P}$, respectively (defined in the paragraph containing (30.1)), and the corresponding Assumption Ω,* V, *or* V$_P$ *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, *and* CQLR$_P$

CSs *for the parameter spaces* $\mathcal{F}_{\mathrm{TS},\Theta,\mathrm{AR}}$, $\mathcal{F}_{\mathrm{TS},\Theta}$, *and* $\mathcal{F}_{\mathrm{TS},\Theta,P}$, *respectively, provided the corresponding Assumption* Ω-CS, V-CS, *or* V_P-CS *holds for each* CS, *rather than Assumption* Ω, V, *or* V_P.

The proof of Theorem 18.1 uses Theorem 30.1 and the following lemma.

LEMMA 30.2. *Suppose* $\{X_i : i = \cdots, 0, 1, \ldots\}$ *is a strictly stationary sequence of mean zero, square integrable, strong mixing random variables. Then* $\mathrm{Var}(\overline{X}_n) = 0$ *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 18.1. The proof of Theorem 18.1 using Theorem 30.1 is essentially the same as the proof (given in Section 17) of Theorems 6.1 and 15.2 using Theorem 16.1 and Lemma 17.1. Thus, we need an analogue of Lemma 17.1 to hold in the time series case. The proof of Lemma 17.1 (given in Section 17) 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 30.2 (with $X_i = \lambda' g_i$) and in consequence $\lambda'\widehat{\Omega}_n\lambda = 0$ a.s. $[F_n]$ by Assumption SR-V(c), SR-V-CS(c), SR-V_P(c), SR-V_P-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(a) or SR-V-CS(a), rather than by the definition of $\widehat{\Omega}_n$ combined with a WLLN for i.i.d. random variables,

(iii) in (17.2), the second implication holds by Lemma 30.2 (with $X_i = \lambda' g_i$) and the fourth implication holds by Assumption SR-V(c), SR-V-CS(c), SR-V_P(c), SR-V_P-CS(c), SR-Ω(c), or SR-Ω-CS(c), and

(iv) the results of Lemmas 5.1 and 15.1, which are used in the proof of Lemma 17.1, holds using the equivariance condition in Assumption SR-V(b), SR-V-CS(b), SR-V_P(b), SR-V_P-CS(b), SR-Ω(b), or SR-Ω-CS(b). □

PROOF OF THEOREM 30.1. The proof is essentially the same as the proof of Theorem 16.1 (given in Section 27) and the proofs of Lemma 16.4 and Proposition 16.5 (given in Section 25 above and Section 17 in the SM of AG1, resp.) 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 20 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_P, V_P-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_P, or V_P-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 20.1 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 20 in the SM to AG1. These modifications affect the proof of Lemma 16.4. No modifications are needed elsewhere. □

PROOF OF LEMMA 30.2. Suppose $\mathrm{Var}(\overline{X}_n) = 0$. 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

$$\begin{aligned}\xi_s &:= \left|P(X_1 \in A, X_{1+sn} \in A) - P(X_1 \in A)P(X_{1+sn} \in A)\right| \\ &\leq \alpha_X(sn) \to 0 \quad \text{as } s \to \infty,\end{aligned} \tag{30.2}$$

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

$$\xi_s = \left|P(X_1 \in A) - P(X_1 \in A)^2\right| = P(X_1 \in A)\big(1 - P(X_1 \in A)\big), \tag{30.3}$$

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 (30.2) and ξ_s does not depend on s by (30.3), we have $\xi_s = 0$. That is, $P(X_1 \in A)$ equals zero or one (using (30.3)) for all Borel sets A, and hence, X_i equals a constant a.s. Because $EX_i = 0$, the constant equals zero. □

31. PROOF OF THEOREMS 9.1, 13.1, AND 9.2

31.1 *Proof of Theorem 9.1*

To prove Theorem 9.1, we use the same proof structure as for the full vector test. Like the proof for the full vector test, the proof of Theorem 9.1 is based on a number of intermediate lemmas, propositions, and theorems. A key change is that the role of $E_F G_i \in R^{k\times p}$ in the full vector case is played by $O_F'(E_F g_i g_i')^{-1/2} E_F G_i \in R^{(k-b)\times p}$ in the subvector case, where $O_F \in R^{k\times(k-b)}$, defined below, is such that $M_{(E_F g_i g_i')^{-1/2} E_F G_{i\beta}} = O_F O_F'$. In this sense, the role of k is replaced by $k - b$.

The proof of the full vector case is given for a general CQLR test that employs weighting matrices $\widehat{W}_n$ and $\widehat{U}_n$ that satisfy a certain high level condition Assumption WU. In particular, $\widehat{W}_n$ and $\widehat{U}_n$ converge to certain matrices W_{F_n} and U_{F_n}, respectively. We follow that structure and prove the result of the theorem for a general CQLR test. However, for the subvector test, the weighting matrices $\widehat{W}_n$ and W_{F_n} are set equal to the identity matrix and, therefore, do not appear in the high level Assumption WU^S, which adapts Assumption WU from the full vector test. We verify Assumption WU^S for the specific choice of weighting matrix $\widehat{U}_n$ employed in the subvector CQLR test (47), which is $\widehat{U}_n = \widehat{L}_n^{1/2}(\theta_0, \widehat{\beta}_n)$, in Lemma 31.9 below.

A general QLR_{WU} subvector test statistic is defined as

$$\begin{aligned}\text{QLR}^S_{\text{WU},n} &:= \text{AR}^S_n(\theta_0, \widehat{\beta}_n) - \lambda_{\min}\big(n\widehat{Q}^S_{\text{WU},n}\big), \quad \text{where} \\ \widehat{Q}^S_{\text{WU},n} &:= \big(\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})\widehat{U}_n, \widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})\big)' \\ &\quad \times M_{\widetilde{J}_n(\widehat{\eta})}\big(\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})\widehat{U}_n, \widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})\big)\end{aligned} \tag{31.1}$$

for $\widehat{\eta} := (\theta_0', \widehat{\beta}_n')'$, and $\widehat{U}_n := U_1(\widehat{U}_{2n})$ is defined as in (16.4). Here, we keep the WU notation from the full vector test, even though no W-type matrix affects the statistic. The population counterpart $U_F := U_1(U_{2F})$ of $\widehat{U}_n$ is defined as in (16.5). The general $\text{CQLR}^S_{\text{WU}}$ test rejects the null hypothesis if

$$\text{QLR}^S_{\text{WU},n} > c_{k,p}\big(n^{1/2}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})\widehat{U}_n, \widetilde{J}_n(\widehat{\eta}), 1-\alpha\big), \tag{31.2}$$

where $c_{k,p}(D, J, 1-\alpha)$ is defined in (48).[70]

The proof for the subvector test result is based on working out the asymptotic null rejection probabilities along certain drifting sequences of parameters $\{\lambda^S_{n,h} : n \geq 1\}$ that we introduce below (31.15). The notation involving λ and h in (16.16) and (16.19) for the full vector case has to be adapted to the subvector case. The argument θ_0 in the notation for expressions for full vector inference is replaced throughout by the argument (θ_0, β^*). For example, in $\lambda^S_{4,F} = E_F G_i$, G_i abbreviates $G_i(\theta_0, \beta^*)$, rather than $G_i(\theta_0)$ as in the full vector case. In addition, relative to $\lambda_{n,h}$ for the full vector case, $\lambda^S_{n,h}$ contains several additional components, such as $\lambda^S_{4,\theta_j\beta,F} := E_F G_{i\theta_j\beta}$ for $j = 1, \ldots, p$ and $\lambda^S_{4,\beta_j\beta,F} := E_F G_{i\beta_j\beta}$ for $j = 1, \ldots, b$.

Construction of bases O_{F_n} and $\widetilde{O}_{F_n}$ for the spaces spanned by the eigenvectors corresponding to the eigenvalue 1 *of two projection matrices* For a projection matrix, the eigenvalues are 0 or 1. When deriving the asymptotic distribution of $\widehat{Q}^S_n(\theta_0, \widehat{\beta}_n)$ in (47), which is part of the test statistic $\text{QLR}^S_n(\theta_0, \widehat{\beta}_n)$, it is helpful to factor $M_{\widetilde{J}_n(\eta)}$ into a product $\widetilde{O}_{F_n}\widetilde{O}'_{F_n}$ where $\widetilde{O}_{F_n} \in R^{k\times(k-b)}$ contains a basis for the space of eigenvectors spanned by the eigenvalue 1 of the projection matrix $M_{\widetilde{J}_n(\eta)}$. Given this factorization, we consider the quantities $(\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta), \widetilde{O}'_{F_n}\widehat{D}^*_n(\eta))$, which puts us into the framework used in the proof for the full vector test. Note that, in general, eigenvectors are not continuous functions of a matrix. However, in the case of a projection matrix, the eigenvalues are well separated and eigenvectors that are continuous can be explicitly constructed.

We now outline this construction. First, given a sequence of nonstochastic matrices $\{J_n \in R^{k\times b} : n \geq 1\}$ that satisfy $J_n \to J$ with J of full column rank b, we construct matrices

[70]The reason $\widetilde{\Omega}_n^{-1/2}$ is used in the definitions of $\text{QLR}^S_n(\theta, \widehat{\beta}_n)$ in (47) and $\text{QLR}^S_{\text{WU},n}$, rather than $\widehat{\Omega}_n^{-1/2}$, is that we prove the subvector results using the proof of the full vector result with $\widehat{W}_n$, W_{F_n}, and $\widehat{D}_n \in R^{k\times p}$ replaced by I_k, I_k, and $\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n) \in R^{(k-b)\times p}$, respectively, where $\widetilde{O}'_{F_n} \in R^{(k-b)\times k}$, defined below, is such that $\widetilde{O}_{F_n}\widetilde{O}'_{F_n} = M_{\widetilde{J}_n(\theta_0, \widehat{\beta}_n)}$. For the full vector results, the difference between $\widehat{W}_n$ and W_{F_n} can be handled easily because $\widehat{W}_n W_{F_n}^{-1} \to_p I_p$ (as in (26.8)). But, in the subvector case, the same strategy cannot be applied to $\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)$ and $(E_{F_n} g_i g_i')^{-1/2}$, because of the factor $\widetilde{O}'_{F_n}$ that precedes $\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)$ in the definition of $\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)$, which is the subvector equivalent to $\widehat{D}_n$.

O_n and $O \in R^{k\times(k-b)}$ such that $M_{J_n} = O_n O_n'$, $M_J = OO'$, and $O_n \to O$. To do so, note first that for any $O' \in R^{(k-b)\times k}$ having rows that contain an orthonormal basis of eigenvectors of the eigenvalue 1, we have $M_J = OO'$. A basis of eigenvectors of the eigenvalue 0 is given by the columns of J. Therefore, the space of eigenvectors corresponding to the eigenvalue 1 is given by $\mathrm{span}(J)^\perp$, the orthogonal complement of $\mathrm{span}(J)$. We have $\mathrm{span}(J)^\perp = N(J')$.

There are $T := \binom{k}{b}$ different sets of b rows from the set of k rows of $J \in R^{k\times b}$. Given that J has full column rank, there is at least one choice of b rows of J that form a basis of R^b. For notational simplicity, assume that the first b columns of J' form a basis of R^b.[71] Decompose $J' = (J_1', J_2')$ with $J_1' \in R^{b\times b}$ and $J_2' = (j_1, \ldots, j_{k-b}) \in R^{b\times(k-b)}$, $j_s \in R^b$ for $s = 1, \ldots, k-b$. It follows that a basis of $N(J')$ is given by the vectors $(-j_s' J_1^{-1}, e_s')' \in R^k$ for $s = 1, \ldots, k-b$, where e_s denotes the sth coordinate vector in R^{k-b}. This holds because

$$J' \begin{pmatrix} -(J_1^{-1})' j_s \\ e_s \end{pmatrix} = (J_1', J_2') \begin{pmatrix} -(J_1^{-1})' j_s \\ e_s \end{pmatrix} = 0^b \quad \text{for } s = 1, \ldots, k-b. \tag{31.3}$$

Let $Q' \in R^{(k-b)\times k}$ be a matrix whose sth row is given by

$$(-j_s' J_1^{-1}, e_s') \tag{31.4}$$

for $s = 1, \ldots, k-b$. Define

$$O' = O(J)' := \left(Q'Q\right)^{-1/2} Q'. \tag{31.5}$$

The matrix OO' is symmetric and idempotent, and hence, is a projection matrix. Since the rows of Q' are orthogonal to the rows of J', OO' projects onto the space orthogonal to the columns of J. That is, $OO' = M_J$. When we want to emphasize which choice of the $t = 1, \ldots, T$ sets of b columns from the set of k columns of J' is used in the above construction of $O' = O(J)'$ we add an additional subindex and write

$$O_t' = O_t(J)' \tag{31.6}$$

instead.

Use analogous notation for $J_n' = (J_{n1}', J_{n2}')$, $J_{n2}' = (j_{n1}, \ldots, j_{n(k-b)})$, the matrix $Q_n' \in R^{(k-b)\times k}$, whose sth row is given by $(-j_{ns}' J_{n1}^{-1}, e_s')$, and $O_n' = O(J_n)' := (Q_n' Q_n)^{-1/2} Q_n'$. Then $O_n O_n' = M_{J_n}$, $OO' = M_J$, and $O_n' \to O'$ as desired, where the convergence follows directly from $J_n \to J$. Again, when we want to emphasize which set of b columns of J_n' is used in the construction, we write

$$O_{nt}' = O_t(J_n)' \tag{31.7}$$

instead.

Under sequences $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$ (defined below), this construction is applied to

$$J_n = \left(E_{F_n} g_i g_i'\right)^{-1/2} E_{F_n} G_{i\beta} \tag{31.8}$$

[71] If that is not the case and the first b columns do not form a basis, simply adapt the notation in what follows so that the b columns of J' that are referred to, do indeed form a basis of R^b.

and the matrix O_n just constructed also is sometimes denoted by O_{F_n}. Under the sequence $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$, it follows that J_n converges to the matrix $J_h := (h_{5,g})^{-1/2} h_{4,\beta}$ defined below.

As in (31.5), for given $F \in \mathcal{F}^S$,

$$O_F' = O_{Ft}' = O\big((E_F g_i g_i')^{-1/2} E_F G_{i\beta}\big)' \tag{31.9}$$

denotes a basis of the space of eigenvectors for the eigenvalue 1 for $M_{(E_F g_i g_i')^{-1/2} E_F G_{i\beta}}$ using the construction outlined above for any choice $t = 1, \ldots, T$ of any b columns of $((E_F g_i g_i')^{-1/2} E_F G_{i\beta})'$ that form a basis of R^b.

Under sequences $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$, Lemma 31.5 below implies that $\widetilde{J}_n(\theta_0, \widehat{\beta}_n) - J_n \in R^{k\times b}$ converges in probability to zero and $J_n = (E_{F_n} g_i g_i')^{-1/2} E_{F_n} G_{i\beta} \to J_h := (h_{5,g})^{-1/2} h_{4,\beta}$. In addition, J_n has full column rank b for all n sufficiently large, under the restrictions in $\mathcal{F}^S$. Therefore, $\widetilde{J}_n(\theta_0, \widehat{\beta}_n)$ has full column rank b wp $\to 1$. For any b columns indexed by $t = 1, \ldots, T$ of J_h' that form a basis of R^b and apply the above construction with this choice of columns to both $\widetilde{J}_n(\theta_0, \widehat{\beta}_n)'$ and J_n' to obtain

$$\widetilde{O}_{F_n}' = O\big(\widetilde{J}_n(\theta_0, \widehat{\beta}_n)\big)' \in R^{(k-b)\times k} \quad \text{and} \quad O_{F_n}' = O(J_n)' \tag{31.10}$$

using the notation in (31.5). Given that $\widetilde{J}_n(\theta_0, \widehat{\beta}_n) - J_n \to_p 0^{k\times b}$, it follows that $\widetilde{O}_{F_n}' - O_{F_n}' \to_p 0^{(k-b)\times k}$.

Definition of $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$ As described above, each $t = 1, \ldots, T$ indexes a set of b columns of $((E_F g_i g_i')^{-1/2} E_F G_{i\beta})'$. For any $t = 1, \ldots, T$ for which the b columns of $((E_F g_i g_i')^{-1/2} E_F G_{i\beta})'$ form a basis of R^b, consider a singular value decomposition of $O_{Ft}'(E_F g_i g_i')^{-1/2}(E_F G_i) U_F \in R^{(k-b)\times p}$. More precisely, let $B_F = B_{Ft}$ denote a $p \times p$ orthogonal matrix of eigenvectors of

$$U_F'(E_F G_i)'\big(E_F g_i g_i'\big)^{-1/2} O_{Ft} O_{Ft}'\big(E_F g_i g_i'\big)^{-1/2} E_F G_i U_F \tag{31.11}$$

ordered so that the corresponding eigenvalues $(\kappa_{1Ft}, \ldots, \kappa_{pFt})$ are nonincreasing. Let $C_F = C_{Ft}$ denote a $(k-b) \times (k-b)$ orthogonal matrix of eigenvectors of

$$O_{Ft}'\big(E_F g_i g_i'\big)^{-1/2}(E_F G_i) U_F U_F'(E_F G_i)'\big(E_F g_i g_i'\big)^{-1/2} O_{Ft}. \tag{31.12}$$

The corresponding eigenvalues are $(\kappa_{1Ft}, \ldots, \kappa_{k-bFt})$.

Let $(\tau_{1Ft}, \ldots, \tau_{\min\{k-b,p\}Ft})$ denote the $\min\{k-b, p\}$ singular values of

$$O_{Ft}'\big(E_F g_i g_i'\big)^{-1/2}(E_F G_i) U_F, \tag{31.13}$$

which are nonnegative and ordered so that τ_{jFt} is nonincreasing in j. For all other $t = 1, \ldots, T$ (for which the b columns of $((E_F g_i g_i')^{-1/2} E_F G_{i\beta})'$ indexed by t do not form a basis of R^b), define $(\tau_{1Ft}, \ldots, \tau_{\min\{k-b,p\}Ft})$ to be a vector of minus ones and B_{Ft} and

C_{Ft} to be identity matrices in $R^{p\times p}$ and $R^{(k-b)\times(k-b)}$, respectively. (This definition is arbitrary and could be replaced by other choices.)

Define the elements of λ^S to be

$$
\begin{aligned}
\lambda^S_{1,F} &:= (\tau_{1F1},\ldots,\tau_{\min\{k-b,p\}F1},\ldots,\tau_{1FT},\ldots,\tau_{\min\{k-b,p\}FT})' \\
&\in R^{T\min\{k-b,p\}}, \\
\lambda^S_{2,F} &:= (B_{F1},\ldots,B_{FT}) \in R^{p\times Tp}, \\
\lambda^S_{3,F} &:= (C_{F1},\ldots,C_{FT}) \in R^{(k-b)\times T(k-b)}, \\
\lambda^S_{4,F} &:= E_F G_i \in R^{k\times p}, \\
\lambda^S_{4,\beta,F} &:= E_F G_{i\beta} \in R^{k\times b}, \\
\lambda^S_{4,\theta_j\beta,F} &:= E_F G_{i\theta_j\beta} \in R^{k\times b} \quad \text{for } j=1,\ldots,p, \\
\lambda^S_{4,\beta_j\beta,F} &:= E_F G_{i\beta_j\beta} \in R^{k\times b} \quad \text{for } j=1,\ldots,b, \\
\lambda^S_{5,F} &:= E_F \begin{pmatrix} g_i \\ \mathrm{vec}(G_i) \end{pmatrix} \begin{pmatrix} g_i \\ \mathrm{vec}(G_i) \end{pmatrix}' \in R^{(p+1)k\times(p+1)k}, \\
\lambda^S_{5,\beta_j,F} &:= E_F G_{i\beta_j} g_i' \in R^{k\times k} \quad \text{for } j=1,\ldots,b, \\
\lambda^S_{5,3,j,F} &:= E_F g_i g_{ij} g_i' \in R^{k\times k} \quad \text{for } j=1,\ldots,k, \\
\lambda^S_{6,F} &:= (\lambda_{6,1F1},\ldots,\lambda_{6,(\min\{k-b,p\}-1)F1},\ldots,\lambda_{6,1FT},\ldots,\lambda_{6,(\min\{k-b,p\}-1)FT})' \\
&:= \left(\frac{\tau_{2F1}}{\tau_{1F1}},\ldots,\frac{\tau_{\min\{k-b,p\}F1}}{\tau_{(\min\{k-b,p\}-1)F1}},\ldots,\frac{\tau_{2FT}}{\tau_{1FT}},\ldots,\frac{\tau_{\min\{k-b,p\}FT}}{\tau_{(\min\{k-b,p\}-1)FT}}\right)' \\
&\in [0,1]^{T(\min\{k-b,p\}-1)}, \\
\lambda^S_{8,F} &:= U_{2F}, \\
\lambda^S_{9,F} &:= F, \\
\lambda^S_{10,F} &:= \mathrm{Var}_F\big(g_i', \mathrm{vec}(G_i)', \mathrm{vec}\big(g_i g_i'\big)', \mathrm{vec}(G_{\beta i})'\big)', \\
\lambda^S &:= \lambda^S_F := \big(\lambda^S_{1,F},\ldots,\lambda^S_{10,F}\big),
\end{aligned} \tag{31.14}
$$

where $0/0 := 0$ for the components of $\lambda^S_{6,F}$, and λ^S is the vector that collects all the above terms in one vector. As mentioned above, there is no weighting matrix $\widehat{W}_n$ for the subvector test and therefore, no $\lambda^S_{7,F}$ component appears. For $j=1,\ldots,b$, we denote the jth column of $\lambda^S_{4,\beta,F} \in R^{k\times b}$ by $\lambda^S_{4,\beta_j,F} \in R^k$. Let

$$
\begin{aligned}
\Lambda^S &:= \big\{\lambda^S_F : F \in \mathcal{F}^S\big\}, \quad \text{and} \\
h_n(\lambda^S) &:= \big(n^{1/2}\lambda^S_{1,F}, \lambda^S_{2,F}, \lambda^S_{3,F}, \lambda^S_{4,F},\ldots,\lambda^S_{6,F}, \lambda^S_{8,F}, \lambda^S_{10,F}\big).
\end{aligned} \tag{31.15}
$$

Let $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$ denote a sequence $\{\lambda^S_n \in \Lambda^S : n \geq 1\}$ for which $h_n(\lambda^S_n) \to h \in H$, for H as in (16.2).[72] Denote by $h_{4,\beta}$, $h_{4,\theta_j\beta}$, $h_{4,\beta_j\beta}$, $h_{5,Gg}$, and $h_{5,G}$, the limits of λ^S_{4,β,F_n}, $\lambda^S_{4,\theta_j\beta,F_n}$, $\lambda^S_{4,\beta_j\beta,F_n}$, λ^S_{5,GgF_n}, and λ^S_{5,GF_n} under the sequence $\{\lambda^S_{n,h} : n \geq 1\}$, respectively, and analogously for other expressions, where by $\lambda^S_{5,GgF}$ and $\lambda^S_{5,GF}$ we denote the lower left and lower right submatrices of $\lambda^S_{5,F}$ of dimensions $R^{pk\times k}$ and $R^{pk\times pk}$.

Consider a sequence $\{\lambda^S_{n,h} : n \geq 1\}$ and let the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda^S_{n,h} : n \geq 1\}$. Because under $\{\lambda^S_{n,h} : n \geq 1\}$, $(E_{F_n} g_i g_i')^{-1/2} E_{F_n} G_{i\beta}$ converges to a full column rank matrix, there exists a smallest index $t^* \in \{1, \ldots, T\}$ such that for all n sufficiently large the b columns of $((E_{F_n} g_i g_i')^{-1/2} E_{F_n} G_{i\beta})'$ indexed by t^* form a basis of R^b, and by definition of $\{\lambda^S_{n,h} : n \geq 1\}$, $n^{1/2}(\tau_{1F_nt^*}, \ldots, \tau_{\min\{k-b,p\}F_nt^*}) \to (h_{1,1t^*}, \ldots, h_{1,\min\{k-b,p\}t^*})$. Note that t^* depends on the sequence $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$. We include τ_{1Ft}, B_{Ft}, and C_{Ft} for all $t = 1, \ldots, T$ in the definition of $\lambda^S_{1,F}$, $\lambda^S_{2,F}$, and $\lambda^S_{3,F}$ in (31.14) because this ensures the convergence of $n^{1/2}\tau_{1F_nt^*}$, $B_{F_nt^*}$, and $C_{F_nt^*}$ for the value t^* just defined.

In what follows, with slight abuse of notation, we leave out the index t^* from the notation.

As in (16.22), let $q^S = q^S_h$ $(\in \{0, \ldots, \min\{k-b, p\}\})$ be such that

$$h_{1,j} = \infty \quad \text{for } 1 \leq j \leq q^S_h \quad \text{and} \quad h_{1,j} < \infty \quad \text{for } q^S_h + 1 \leq j \leq \min\{k-b, p\}, \tag{31.16}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, \ldots, \min\{k-b, p\}$.

Define $\mathcal{F}^S_{\mathrm{WU}}$ as $\mathcal{F}_{\mathrm{WU}}$ in (16.12) with $\mathcal{F}$ replaced by $\mathcal{F}^S$ and W replaced by I_k. Define Λ^S_{WU} as Λ_{WU} in (16.17) with $\mathcal{F}_{\mathrm{WU}}$ replaced by $\mathcal{F}^S_{\mathrm{WU}}$.

Assumption WU^S for the parameter space $\Lambda^S_* \subset \Lambda^S_{\mathrm{WU}}$. *Under all subsequences $\{w_n\}$ and all sequences $\{\lambda^S_{w_n,h} : n \geq 1\}$ with $\lambda^S_{w_n,h} \in \Lambda^S_*$,*

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

(b) $U_1(\cdot)$ *is a continuous function at h_8 on some set $\mathcal{U}_2$ that contains* $\{\lambda^S_{8,F} (= U_{2F}) : \lambda^S \in \Lambda^S_*\}$ *and contains* $\widehat{U}_{2w_n}$ wp $\to 1$.

As in (16.23), let (and recall again that we leave out the index t^* from the notation)

$$\begin{aligned} S_n &:= \mathrm{Diag}\{(n^{1/2}\tau_{1F_n})^{-1}, \ldots, (n^{1/2}\tau_{q^SF_n})^{-1}, 1, \ldots, 1\} \in R^{p\times p} \quad \text{and} \\ T_n &:= B_{F_n} S_n \in R^{p\times p}. \end{aligned} \tag{31.17}$$

The random function $\mathrm{CLR}_{k,p}(D, J)$ in (48) that generates the conditional critical value of the CLR subvector test can be expressed as follows. Suppose $M_J = OO'$, for O

[72]Regarding the notation, it would be more consistent to put a superscript S on all of the expressions involving h. However, this would introduce too much clutter, so we do not do so.

defined in (31.5). Then we can write

$$
\begin{aligned}
\mathrm{CLR}_{k,p}(D,J) &:= Z'M_J Z - \lambda_{\min}\big((Z,D)'M_J(Z,D)\big) \\
&= \big(O'Z\big)'O'Z - \lambda_{\min}\big(\big(O'Z,O'D\big)'\big(O'Z,O'D\big)\big) \\
&= \overline{Z}'\overline{Z} - \lambda_{\min}\big(\big(\overline{Z},O'D\big)'\big(\overline{Z},O'D\big)\big), \\
&\sim \mathrm{CLR}_{k-b,p}\big(O'D,0^{(k-b)\times 0}\big) = \mathrm{CLR}_{k-b,p}\big(O'D\big),
\end{aligned} \tag{31.18}
$$

where $Z \sim N(0^k, I_k)$, $\overline{Z} := O'Z \sim N(0^{k-b}, I_{k-b})$, "$\sim$" denotes "has the same distribution as," and $\mathrm{CLR}_{k-b,p}(O'D)$ is the expression from the full vector test defined in (24).

We now state the intermediate lemmas, propositions, and theorems upon which the proof of Theorem 9.1 is based. Using them, the proof of Theorem 9.1 follows the same lines as the proof of Theorem 16.1 for the full vector case.

By Lemma 16.2, the $1-\alpha$ quantile $c_{k-b,p}(O'D,1-\alpha)$ of $\mathrm{CLR}_{k-b,p}(O'D)$ depends on $O'D$ only through the singular values of $O'D$. By (31.18), that immediately implies the following analogue to Lemma 16.2.

LEMMA 31.1. *Let D and J be $k\times p$ and $k\times b$ matrices, respectively, where J has full column rank b. Let CYB' denote a singular value decomposition of $O'D \in R^{(k-b)\times p}$, where Y contains the singular values in nonincreasing order and $O' = O(J)'$ is defined in* (31.5). *Then $c_{k,p}(D,J,1-\alpha)$ depends on D and J only through Y and*

$$
c_{k,p}(D,J,1-\alpha) = c_{k-b,p}\big(Y,0^{(k-b)\times 0},1-\alpha\big) = c_{k-b,p}(Y,1-\alpha).
$$

Just like the full vector test in Lemma 5.1, the subvector CQLR test is invariant to nonsingular transformations of the moment functions. We suppress the dependence on θ_0 of the statistics in the following lemma.

LEMMA 31.2. *Given the preliminary estimator $\widetilde{\beta}_n$ of β_n^*, the statistics $\mathrm{AR}_n^S(\widehat{\beta}_n)$, $\mathrm{QLR}_n^S(\widehat{\beta}_n)$, $\widehat{\beta}_n$, $c_{k,p}(n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n),\widetilde{J}_n(\widehat{\beta}_n),1-\alpha)$, $\widehat{D}_n^{*\prime}(\widehat{\beta}_n)M_{\widetilde{J}_n(\widehat{\beta}_n)}\widehat{D}_n^*(\widehat{\beta}_n)$, $\widehat{g}_n(\widehat{\beta}_n)'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)M_{\widetilde{J}_n(\widehat{\beta}_n)}\times$ $\widehat{D}_n^*(\widehat{\beta}_n)$, $\widehat{\Sigma}_n(\widehat{\beta}_n)$, and $\widehat{L}_n(\widehat{\beta}_n)$ are invariant to the transformation $(g_i(\beta),G_i(\beta)) \rightsquigarrow (Mg_i(\beta),MG_i(\beta))$ $\forall i\le n$ for any $k\times k$ nonsingular matrix M. This transformation induces the following transformations: $\widehat{g}_n(\widehat{\beta}_n)\rightsquigarrow M\widehat{g}_n(\widehat{\beta}_n)$, $\widehat{G}_n(\widehat{\beta}_n)\rightsquigarrow M\widehat{G}_n(\widehat{\beta}_n)$, $\widetilde{G}_{\beta n}(\widehat{\beta}_n)\rightsquigarrow M\widetilde{G}_{\beta n}(\widehat{\beta}_n)$, $\widehat{\Gamma}_{jn}(\widehat{\beta}_n)\rightsquigarrow M\widehat{\Gamma}_{jn}(\widehat{\beta}_n)M'$ $\forall j\le p$, $\widehat{D}_n(\widehat{\beta}_n)\rightsquigarrow M\widehat{D}_n(\widehat{\beta}_n)$, $\widehat{\Omega}_n(\widehat{\beta}_n)\rightsquigarrow M\widehat{\Omega}_n(\widehat{\beta}_n)M'$, $\widetilde{\Omega}_n(\widehat{\beta}_n)\rightsquigarrow M\widetilde{\Omega}_n(\widehat{\beta}_n)M'$, $\widehat{V}_n(\widehat{\beta}_n)\rightsquigarrow (I_{p+1}\otimes M)\widehat{V}_n(\widehat{\beta}_n)\times(I_{p+1}\otimes M')$, and $\widehat{R}_n(\widehat{\beta}_n)\rightsquigarrow (I_{p+1}\otimes M)\widehat{R}_n(\widehat{\beta}_n)(I_{p+1}\otimes M')$.*

The proof of the lemma is straightforward for all quantities except $c_{k,p}(n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n),$ $\widetilde{J}_n(\widehat{\beta}_n),1-\alpha)$. Using Lemma 31.1, this quantity depends on $n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n)$ and $\widetilde{J}_n(\widehat{\beta}_n)$ only through the nonzero singular values of $O(\widetilde{J}_n(\widehat{\beta}_n))'n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n)$, which equal the square roots of the nonzero eigenvalues of $n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n)'M_{\widetilde{J}_n(\widehat{\beta}_n)}n^{1/2}\widehat{D}_n^*(\widehat{\beta}_n)$. But, the latter quantity is invariant to the transformation $(g_i(\beta),G_i(\beta))\rightsquigarrow(Mg_i(\beta),MG_i(\beta))$.

The derivation in (31.18) immediately implies an analogue of the result in Lemma 27.2. Let $c_{k-b,p,q}(\tau^c_\infty, 1-\alpha)$ denote the $1-\alpha$ quantile of

$$\begin{aligned}
\mathrm{ACLR}_{k-b,p,q}(\tau^c_\infty) &:= \overline{Z}'\overline{Z} - \lambda_{\min}\big(\big(Y(\tau^c_\infty), \overline{Z}_2\big)'\big(Y(\tau^c_\infty), \overline{Z}_2\big)\big), \quad \text{where} \\
\overline{Z} &:= \begin{pmatrix} \overline{Z}_1 \\ \overline{Z}_2 \end{pmatrix} \sim N\big(0^{k-b}, I_{k-b}\big) \quad \text{for } \overline{Z}_1 \in R^q \text{ and } \overline{Z}_2 \in R^{k-b-q}, \\
\tau^c_\infty &:= \big(\tau^c_{(q+1)\infty}, \ldots, \tau^c_{\min\{k-b,p\}\infty}\big)' \in R^{\min\{k-b,p\}-q}, \\
Y(\tau^c_\infty) &:= \begin{pmatrix} \mathrm{Diag}\{\tau^c_\infty\} \\ 0^{(k-b-p)\times(p-q)} \end{pmatrix} \in R^{(k-b-q)\times(p-q)} \quad \text{if } k-b \geq p, \quad \text{and} \\
Y(\tau^c_\infty) &:= \big(\mathrm{Diag}\{\tau^c_\infty\}, 0^{(k-b-q)\times(p-k-b)}\big) \\
&\quad\ \in R^{(k-b-q)\times(p-q)} \quad \text{if } k-b < p.
\end{aligned} \tag{31.19}$$

Lemma 31.3. *Suppose* $\{(D^c_n, J^c_n) : n \geq 1\}$ *is a sequence of constant* (*i.e., nonrandom*) $k \times p$ *and* $k \times b$ *matrices, respectively, such that* $O^{c\prime}_n D^c_n$ (*for* $O^c_n O^{c\prime}_n = M_{J^c_n}$ *and* O^c_n *defined in* (31.5)) *has singular values* $\{\tau^c_{jn} \geq 0 : j \leq \min\{k-b, p\}\}$ *for* $n \geq 1$ *that satisfy* (i) $\{\tau^c_{jn} \geq 0 : j \leq \min\{k-b, p\}\}$ *are nonincreasing in* j *for* $n \geq 1$, (ii) $\tau^c_{jn} \to \infty$ *for* $j \leq q$ *for some* $0 \leq q \leq \min\{k-b, p\}$ *and* (iii) $\tau^c_{jn} \to \tau^c_{j\infty} < \infty$ *for* $j = q+1, \ldots, \min\{k-b, p\}$. *Then*

$$c_{k,p}\big(D^c_n, J^c_n, 1-\alpha\big) \to c_{k-b,p,q}\big(\tau^c_\infty, 1-\alpha\big).$$

The next lemma is a restatement of Lemma 27.3 with k replaced by $k-b$.

Lemma 31.4. *For all admissible integers* $(k-b, p, q)$ (*i.e.,* $k-b \geq 1$, $p \geq 1$, *and* $0 \leq q \leq \min\{k-b, p\}$) *and all* $\min\{k-b, p\} - q$ (≥ 0) *vectors* τ^c_∞ *with nonnegative elements in nonincreasing order, the df of* $\mathrm{ACLR}_{k-b,p,q}(\tau^c_\infty) := \overline{Z}'\overline{Z} - \lambda_{\min}((Y(\tau^c_\infty), \overline{Z}_2)'(Y(\tau^c_\infty), \overline{Z}_2))$ *is continuous and strictly increasing at its* $1-\alpha$ *quantile* $c_{k-b,p,q}(\tau^c_\infty, 1-\alpha)$ *for all* $\alpha \in (0,1)$, *where* $\overline{Z} := (\overline{Z}'_1, \overline{Z}'_2)' \sim N(0^{k-b}, I_{k-b})$ *for* $\overline{Z}_1 \in R^q$ *and* $\overline{Z}_2 \in R^{k-b-q}$ *and* τ^c_∞ *and* $Y(\tau^c_\infty)$ *are defined in* (31.19).

The next lemma is an important ingredient in the proof of Theorem 9.1 because it provides the asymptotic distributions of key quantities. It is the analogue and extension of Lemma 16.4 for the subvector test. We now introduce some notation that is used in the lemma.

By the Lyapunov CLT, under sequences $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$, we have

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^n \begin{pmatrix} g_i \\ \mathrm{vec}(G_i) \\ \mathrm{vec}\big(g_i g_i' - \Omega_n\big) \\ \mathrm{vec}(G_{\beta i} - E_n G_{\beta i}) \end{pmatrix} \\
&\quad \to_d \overline{L}_h \sim N\big(0^{d^*}, h_{10}\big), \quad \text{where} \\
&\overline{L}_h := \big(\overline{g}'_h, \overline{L}'_{h,2}, \overline{L}'_{h,3}, \overline{L}'_{h,4}\big)' \quad \text{for } \overline{g}_h \in R^k, \overline{L}_{h,2} \in R^{kp}, \overline{L}_{h,3} \in R^{k^2}, \overline{L}_{h,4} \in R^{kb}, \\
&\overline{G}_h := \mathrm{vec}^{-1}_{k,p}(\overline{L}_h) \in R^{k\times p},
\end{aligned} \tag{31.20}$$

$d^* = k + kp + k^2 + kb$, and the function $\mathrm{vec}_{k,p}^{-1}(\cdot)$ is the inverse of the $\mathrm{vec}(\cdot)$ function for $k \times p$ matrices. (Thus, the domain of $\mathrm{vec}_{k,p}^{-1}(\cdot)$ consists of kp-vectors and its range consists of $k \times p$ matrices.) As defined in (31.20), $\overline{g}_h$ is the same as in (16.21) for the full vector case.

The asymptotic distributions of (i) $n^{1/2}(\widehat{\beta}_n - \beta_n^*)$, (ii) $n^{1/2}\widehat{g}_n(\widehat{\beta}_n)$, (iii) $n^{1/2} \times \mathrm{vec}(\widehat{D}_n(\widehat{\beta}_n) - D_n)$, where $D_n := E_n G_i$, (iv) $n^{1/2}(\widetilde{\Omega}_n(\widehat{\beta}_n) - \Omega_n)$, and (v) $n^{1/2}(\widetilde{G}_{\beta_j n}(\widehat{\beta}_n) - E_n G_{i\beta_j})$ are given by

$$
\begin{aligned}
&\text{(i)} \quad \overline{\beta}_h := \left[(h_{5,g}^{-1/2} h_{4,\beta})'(h_{5,g}^{-1/2} h_{4,\beta})\right]^{-1} (h_{5,g}^{-1/2} h_{4,\beta})' h_{5,g}^{-1/2} \overline{g}_h,\\
&\text{(ii)} \quad \overline{g}_h^S := h_{5,g}^{1/2} M_{h_{5,g}^{-1/2} h_{4,\beta}} h_{5,g}^{-1/2} \overline{g}_h,\\
&\text{(iii)} \quad \mathrm{vec}(\overline{D}_h^S) := \left(\mathrm{vec}(\overline{G}_h) - h_{5,Gg} h_{5,g}^{-1} \overline{g}_h\right) + \mathrm{vec}(h_{4,\theta_1\beta}\overline{\beta}_h, \ldots, h_{4,\theta_p\beta}\overline{\beta}_h)\\
&\qquad\qquad - h_{5,Gg} h_{5,g}^{-1} h_{4,\beta}\overline{\beta}_h,\\
&\text{(iv)} \quad \overline{\varkappa}_h^S := (\overline{\varkappa}_{h,1}^S, \ldots, \overline{\varkappa}_{h,k}^S), \quad \text{and} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (31.21)\\
&\text{(v)} \quad \overline{\varrho}_h^S := (\overline{\varrho}_{h,1}^S, \ldots, \overline{\varrho}_{h,b}^S) \quad \text{where}\\
&\qquad \overline{\varkappa}_{h,j}^S := \overline{L}_{j,h,3} - h_{5,3.j} h_{5,g}^{-1} \overline{g}_h\\
&\qquad\qquad + \left[(h_{5,\beta_1,j}, \ldots, h_{5,\beta_b,j}) + \left((h_{5,\beta_1})_j', \ldots, (h_{5,\beta_b})_j'\right) - h_{5,3,j} h_{5,g}^{-1} h_{4,\beta}\right]\overline{\beta}_h\\
&\qquad \text{for } j = 1, \ldots, k,\\
&\qquad \overline{\varrho}_{h,j}^S := \overline{L}_{j,h,4} - h_{5,\beta_j} h_{5,g}^{-1} \overline{g}_h + \left(h_{4,\beta_j\beta} - h_{5,\beta_j} h_{5,g}^{-1} h_{4,\beta}\right)\overline{\beta}_h \quad \text{for } j = 1, \ldots, b,
\end{aligned}
$$

$\overline{L}_{j,h,3}, \overline{L}_{j,h,4} \in R^k$ denote the $(j-1)k+1, \ldots, jk$ components of $\overline{L}_{h,3}$ and $\overline{L}_{h,4}$, respectively, and $(h_{5,\beta_s})_j' \in R^k$ denotes the jth column of $(h_{5,\beta_s})' \in R^{k \times k}$ for $s = 1, \ldots, b$.[73] If no preliminary estimator appears, that is, $\widehat{\beta}_n = \beta_n^*$, then the quantities in (31.21) reduce to those in the full vector case. In particular, $\overline{\beta}_h = 0^b$, $\overline{g}_h^S = \overline{g}_h$, and $\mathrm{vec}(\overline{D}_h^S) = \mathrm{vec}(\overline{G}_h) - h_{5,Gg} h_{5,g}^{-1} \overline{g}_h = \mathrm{vec}(\overline{D}_h)$.

Consider the function that maps $\mathrm{vec}(\varphi)$ onto $\mathrm{vec}(\varphi^{-1/2})$, where $\varphi \in R^{k \times k}$ is positive definite. Let $\overline{\varphi}_h \in R^{k^2 \times k^2}$ denote the matrix of partial derivatives of that mapping evaluated at $\mathrm{vec}(h_{5,g})$. Consider the function that maps $\mathrm{vec}(J)$ for $J \in R^{k \times b}$ onto $\mathrm{vec}((-j_1'(J_1)^{-1}, e_1'), \ldots, (-j_{(k-b)}'(J_1)^{-1}, e_{k-b}')) \in R^{k(k-b)}$, as defined in (31.4) and (31.5). Denote by $\overline{B}_h \in R^{k(k-b) \times kb}$ the matrix of partial derivatives of that mapping evaluated at $\mathrm{vec}(h_{5,g}^{-1/2} h_{4,\beta})$.

The asymptotic distributions of (vi) $n^{1/2}(\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) - \Omega_n^{-1/2})$, (vii) $n^{1/2} \times (\widetilde{J}_n(\widehat{\beta}_n) - \Omega_n^{-1/2} E_n G_{\beta i})$, (viii) $n^{1/2}(\widetilde{O}_n - O_n)$, (ix) $n^{1/2}(\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n' \Omega_n^{-1/2} D_n)$,

[73] See (31.45)–(31.46) for (i), (31.48) for (ii), (31.52) for (iii), (31.54) for (iv), and (31.55) for (v).

(x) $n^{1/2}\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) \times U_n B_n S_n$ are given by

$$
\begin{aligned}
&\text{(vi)} \quad \operatorname{vec}_{k,k}^{-1}\big(\overline{\varphi}_h \operatorname{vec}(\overline{\varkappa}_h^S)\big),\\
&\text{(vii)} \quad \overline{\omega}_h^S := h_{5,g}^{-1/2}\overline{\varrho}_h^S + \operatorname{vec}_{k,k}^{-1}\big(\overline{\varphi}_h \operatorname{vec}(\overline{\varkappa}_h^S)\big)h_{4,\beta},\\
&\text{(viii)} \quad \operatorname{vec}_{k,k-b}^{-1}\big(\overline{B}_h \operatorname{vec}(\overline{\omega}_h^S)\big),\\
&\text{(ix)} \quad \chi_h := \operatorname{vec}_{k,k-b}^{-1}\big(\overline{B}_h \operatorname{vec}(\overline{\omega}_h^S)\big)' h_{5,g}^{-1/2} h_4\\
&\qquad + O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' \operatorname{vec}_{k,k}^{-1}\big(\overline{\varphi}_h \operatorname{vec}(\overline{\varkappa}_h^S)\big)h_4\\
&\qquad + O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' h_{5,g}^{-1/2}\overline{D}_h,\\
&\text{(x)} \quad \overline{\Delta}_h^S := \big(\overline{\Delta}_{h,q^S}^S, \overline{\Delta}_{h,p-q^S}^S\big), \quad \text{where } \overline{\Delta}_{h,q^S}^S := h_{3,q^S} \in R^{(k-b)\times q^S},\\
&\qquad \overline{\Delta}_{h,p-q^S}^S := h_3 h_{1,p-q^S}^{\diamond} + \chi_h h_{81} h_{2,p-q^S} \in R^{(k-b)\times(p-q^S)},
\end{aligned}
\tag{31.22}
$$

and $h_{1,p-q^S}^{\diamond} \in R^{(k-b)\times(p-q^S)}$ is defined as in (16.24) with $k-b$ and q^S in place of k and q, respectively.[74]

LEMMA 31.5. *Suppose Assumptions* gB *and* WU^S *hold for some nonempty parameter space $\Lambda_*^S \subset \Lambda_{\text{WU}}^S$. Under all sequences $\{\lambda_{n,h}^S \in \Lambda_*^S : n \geq 1\}$,*

(a) $n^{1/2}(\widehat{\beta}_n - \beta_n^*) \to_d \overline{\beta}_h$,

(b) $\widetilde{J}_n(\widehat{\beta}_n) \to_p h_{5,g}^{-1/2}h_{4,\beta}$,

(c)

$$
n^{1/2}\begin{pmatrix} \widehat{g}_n(\theta_0, \widehat{\beta}_n) \\ \widehat{D}_n(\theta_0, \widehat{\beta}_n) - E_{F_n}G_i \\ \widetilde{\Omega}_n(\theta_0, \widehat{\beta}_n) - E_{F_n}g_i g_i' \\ \widetilde{G}_{\beta n}(\theta_0, \widehat{\beta}_n) - E_{F_n}G_{i\beta} \end{pmatrix} \to_d \begin{pmatrix} \overline{g}_h^S \\ \overline{D}_h^S \\ \overline{\varkappa}_h^S \\ \overline{\varrho}_h^S \end{pmatrix},
$$

where $(\overline{\beta}_h, \overline{D}_h^S, \overline{\varkappa}_h^S, \overline{\varrho}_h^S)$ and $\overline{g}_h^S$ are independent,

(d) *for $\widetilde{O}_{F_n}'$ defined in* (31.10),

$$
n^{1/2}\widetilde{O}_{F_n}'\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)U_{F_n}T_n \to_d \overline{\Delta}_h^S \in R^{(k-b)\times p},
$$

where $(\overline{\beta}_h, \overline{D}_h^S, \overline{\varkappa}_h^S, \overline{\varrho}_h^S, \overline{\Delta}_h^S)$ and $\overline{g}_h^S$ are independent, and

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

[74] See (31.56) for (vi), (31.57) for (vii), (31.59) for (viii), (31.64) for (ix), and (31.60), (31.61), and (31.65) for (x). Recall again that we leave out a subindex t^* from certain expressions.

Lemma 31.5 is proved in Section 31.2 below. Note that in order to obtain consistency of the first step estimator $\widetilde{\beta}_n$ we only need to impose the conditions in $\mathcal{F}^S_{\mathrm{AR},1}$. In particular, for consistency of $\widetilde{\beta}_n$, the variance matrix Ω_{F_n} is allowed to be rank deficient. Lemma 31.5(b) and (c) implies Theorem 9.1 for the subvector AR test. This holds because $\mathrm{AR}^S_n(\theta_0, \widehat{\beta}_n)$ is a quadratic form in $M_{\widetilde{J}_n(\theta_0,\widehat{\beta}_n)}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)n^{1/2}\widehat{g}_n(\theta_0, \widehat{\beta}_n)$ which converges in distribution to $M_{h_{5,g}^{-1/2}h_{4,\beta}} h_{5,g}^{-1/2}\overline{g}_h$. Because $h_{5,g}^{-1/2}h_{4,\beta}$ has full column rank b, the desired result follows.

An analogue of Proposition 16.5 holds where $\widehat{W}_n$, W_{F_n}, and $\widehat{D}_n \in R^{k\times p}$ are replaced by I_k, I_k, and $\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n) \in R^{(k-b)\times p}$, respectively. In particular, $\widehat{\kappa}_{jn}$ is defined as the jth eigenvalue of

$$n\big(\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)\widehat{U}_n\big)'\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)\widehat{U}_n. \tag{31.23}$$

Recall the following notation as for the full vector test, $B_{F_n} = (B_{F_n,q^S}, B_{F_n,p-q^S})$, $C_{F_n} = (C_{F_n,q^S}, C_{F_n,k-b-q^S})$, with $B_{F_n,q^S} \in R^{p\times q^S}$, $B_{F_n,p-q^S} \in R^{p\times(p-q^S)}$, $C_{F_n,q^S} \in R^{(k-b)\times q^S}$, and $C_{F_n,k-b-q^S} \in R^{(k-b)\times(k-b-q^S)}$ and corresponding decompositions for the limiting matrices $h_2 = (h_{2,q^S}, h_{2,p-q^S})$ and $h_3 = (h_{3,q^S}, h_{3,k-b-q^S})$. Recall that we leave out a subindex t^* from certain expressions.

Proposition 31.6. *Suppose Assumption* WU^S *holds for some nonempty parameter space $\Lambda^S_* \subset \Lambda^S_{\mathrm{WU}}$. Under all sequences $\{\lambda^S_{n,h} : n \geq 1\}$ with $\lambda^S_{n,h} \in \Lambda^S_*$,*

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

(b) *$(\widehat{\kappa}_{(q^S+1)n}, \ldots, \widehat{\kappa}_{pn})'$ converges in distribution to the (ordered) $p - q^S$ vector of the eigenvalues of $\overline{\Delta}^{S\prime}_{h,p-q^S} h_{3,k-b-q^S} h'_{3,k-b-q^S}\overline{\Delta}^S_{h,p-q^S} \in R^{(p-q^S)\times(p-q^S)}$,*

(c) *the convergence in parts (a) and (b) holds jointly with the convergence in Lemma 31.5, and*

(d) *under all subsequences $\{w_n\}$ and all sequences $\{\lambda^S_{w_n,h} : n \geq 1\}$ with $\lambda^S_{w_n,h} \in \Lambda^S_*$, the results in parts (a)–(c) hold with n replaced with w_n.*

An analogue of Theorem 16.6 holds for $\mathrm{QLR}^S_{\mathrm{WU},n} = \mathrm{AR}^S_n(\theta_0, \widehat{\beta}_n) - \lambda_{\min}(n\widehat{Q}^S_{\mathrm{WU},n})$, defined in (31.1). For $\widehat{\eta} := (\theta_0, \widehat{\beta}_n)$, wp $\to 1$, we can write

$$\begin{aligned}\widehat{Q}^S_{\mathrm{WU},n} = {} & \big(\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})\widehat{U}_n, \widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})\big)' \\ & \times \big(\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})\widehat{U}_n, \widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})\big)\end{aligned} \tag{31.24}$$

by again replacing $\widehat{W}_n$, W_{F_n}, $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$, and $\widehat{D}_n \in R^{k\times p}$ by I_k, I_k, $\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})$, and $\widetilde{O}'_{F_n}\widetilde{\Omega}_n(\theta_0, \widehat{\beta}_n)^{-1/2}\widehat{D}_n(\theta_0, \widehat{\beta}_n) \in R^{(k-b)\times p}$, respectively. This implies that the role of k is played by $k - b$. Note that by Lemma 31.5(b) and (c) and (31.59) below, which implies

$\widetilde{O}'_{F_n} = O(\widetilde{J}_n(\theta_0, \widehat{\beta}_n))' \to_p O(h_{5,g}^{-1/2} h_{4,\beta})'$, we have

$$\begin{aligned} n^{1/2}\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta}) \to_d\ & O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' h_{5,g}^{-1/2}\overline{g}_h^S \\ =\ & O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' M_{h_{5,g}^{-1/2}h_{4,\beta}} h_{5,g}^{-1/2}\overline{g}_h \\ =\ & O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' h_{5,g}^{-1/2}\overline{g}_h \sim N\big(0^{k-b}, I_{k-b}\big), \end{aligned} \tag{31.25}$$

using $\overline{g}_h^S := h_{5,g}^{1/2} M_{h_{5,g}^{-1/2}h_{4,\beta}} h_{5,g}^{-1/2}\overline{g}_h$, $OO' = M_{h_{5,g}^{-1/2}h_{4,\beta}}$, and $O'O = I_{k-b}$.

Theorem 31.7. *Suppose Assumption* WU^S *holds for some nonempty parameter space* $\Lambda_*^S \subset \Lambda_{\mathrm{WU}}^S$. *Under all sequences* $\{\lambda_{n,h}^S : n \geq 1\}$ *with* $\lambda_{n,h}^S \in \Lambda_*^S$,

$$\begin{aligned} \mathrm{QLR}_{\mathrm{WU},n}^S \to_d\ & l_h' l_h - \lambda_{\min}\big(\big(\overline{\Delta}_{h,p-q^S}^S, l_h\big)' h_{3,k-b-q^S} h_{3,k-b-q^S}' \big(\overline{\Delta}_{h,p-q^S}^S, l_h\big)\big), \quad \textit{where} \\ l_h :=\ & O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' h_{5,g}^{-1/2}\overline{g}_h, \end{aligned}$$

$\overline{\Delta}_{h,p-q^S}^S$ *is defined in* (31.22), *and the convergence holds jointly with the convergence in Lemma* 31.5 *and Proposition* 31.6. *When* $q^S = p$ (*which can only hold if* $k-b \geq p$ *because* $q^S \leq \min\{k-b, p\}$), $\overline{\Delta}_{h,p-q^S}^S$ *does not appear in the limit random variable and the limit random variable reduces to*

$$l_h' h_{3,p} h_{3,p}' l_h \sim \chi_p^2.$$

When $q^S = k-b$ (*which can only hold if* $k-b \leq p$), *the* $\lambda_{\min}(\cdot)$ *expression does not appear in the limit random variable and the limit random variable reduces to*

$$l_h' l_h \sim \chi_{k-b}^2. \tag{31.26}$$

When $k-b \leq p$ *and* $q^S < k-b$, *the* $\lambda_{\min}(\cdot)$ *expression equals zero and the limit random variable reduces to the one in* (31.26). *Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h}^S : n \geq 1\}$ *with* $\lambda_{w_n,h}^S \in \Lambda_*^S$, *the same results hold with* n *replaced with* w_n.

The following lemma, which the proof of Theorem 31.7 relies on, adapts Lemma 26.1 from the full vector test and Lemma 17.1 in AG1. Define

$$\begin{aligned} Y_n :=\ & \begin{bmatrix} Y_{n,q^S} & 0^{q^S \times (p-q^S)} \\ 0^{(p-q^S)\times q^S} & Y_{n,p-q^S} \\ 0^{(k-b-p)\times q^S} & 0^{(k-b-p)\times(p-q^S)} \end{bmatrix} \in R^{(k-b)\times p} \quad \text{if } k-b \geq p, \quad \text{and} \\ Y_n :=\ & \begin{bmatrix} Y_{n,q^S} & 0^{q^S\times(k-b-q^S)} & 0^{q^S\times(p-(k-b))} \\ 0^{(k-b-q^S)\times q^S} & Y_{n,k-b-q^S} & 0^{(k-b-q^S)\times(p-(k-b))} \end{bmatrix} \\ & \in R^{(k-b)\times p} \quad \text{if } k-b < p, \end{aligned} \tag{31.27}$$

as in (25.2), but with $\tau_{1F_nt^*}, \ldots, \tau_{pF_nt^*}$ and q^S in place of $\tau_{1F_n}, \ldots, \tau_{pF_n}$ and q, respectively. Define

$$
\begin{aligned}
&\widehat{D}_n^+ := \big(\widetilde{O}_{F_n}'\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n), \widetilde{O}_{F_n}'\widetilde{\Omega}_n^{-1/2}\widehat{g}_n\big) \in R^{(k-b)\times(p+1)},\\
&\widehat{U}_n^+ := \begin{bmatrix} \widehat{U}_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \qquad U_n^+ = \begin{bmatrix} U_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},\\
&h_{81}^+ := \begin{bmatrix} h_{81} & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)}, \qquad B_n^+ = \begin{bmatrix} B_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},\\
&B_n^+ := \big(B_{n,q^S}^+, B_{n,p+1-q^S}^+\big) \quad \text{for } B_{n,q^S}^+ \in R^{(p+1)\times q^S} \text{ and } B_{n,p+1-q^S}^+ \in R^{(p+1)\times(p+1-q^S)}, \qquad (31.28)\\
&D_n^+ := \big(O(J_n)'\Omega_n^{-1/2}D_n, 0^k\big) \in R^{(k-b)\times(p+1)}, \qquad Y_n^+ := \big(Y_n, 0^{k-b}\big) \in R^{(k-b)\times(p+1)},\\
&S_n^+ := \mathrm{Diag}\big\{\big(n^{1/2}\tau_{1F_n}\big)^{-1}, \ldots, \big(n^{1/2}\tau_{q^SF_n}\big)^{-1}, 1, \ldots, 1\big\}\\
&\qquad = \begin{bmatrix} S_n & 0^{p\times 1} \\ 0^{1\times p} & 1 \end{bmatrix} \in R^{(p+1)\times(p+1)},
\end{aligned}
$$

with J_n defined in (31.8). Let

$$\widehat{\kappa}_{jn}^+ \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n^{+\prime}\widehat{D}_n^{+\prime}\widehat{D}_n^+\widehat{U}_n^+, \quad \forall j = 1, \ldots, p+1, \tag{31.29}$$

ordered to be nonincreasing in j.

LEMMA 31.8. *Suppose Assumption* WU^S *holds for some nonempty parameter space* $\Lambda_*^S \subset \Lambda_{\mathrm{WU}}^S$. *Under all sequences* $\{\lambda_{n,h}^S : n \geq 1\}$ *with* $\lambda_{n,h}^S \in \Lambda_*^S$ *for which* q^S *satisfies* $q^S \geq 1$, *we have* (a) $\widehat{\kappa}_{jn}^+ \to_p \infty$ *for* $j = 1, \ldots, q^S$ *and* (b) $\widehat{\kappa}_{jn}^+ = o_p((n^{1/2}\tau_{\ell F_n})^2)$ *for all* $\ell \leq q^S$ *and* $j = q^S + 1, \ldots, p+1$. *Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h}^S : n \geq 1\}$ *with* $\lambda_{w_n,h}^S \in \Lambda_*^S$, *the same result holds with n replaced with* w_n.

The proof of Lemma 17.1, with analogous modifications that were made in order to prove Lemma 26.1, applies to prove Lemma 31.8. For example, the equivalent of (17.3) of AG1 is

$$
\begin{aligned}
\tau_{r_1F_n}^{-1}\widehat{D}_n^+U_n^+B_n^+ &= \tau_{r_1F_n}^{-1}D_n^+U_n^+B_n^+ + \big(n^{1/2}\tau_{r_1F_n}\big)^{-1}n^{1/2}\big(\widehat{D}_n^+ - D_n^+\big)U_n^+B_n^+\\
&= \tau_{r_1F_n}^{-1}\big(O(J_n)'\Omega_n^{-1/2}D_nU_nB_n, 0^{k-b}\big) + O_p\big(\big(n^{1/2}\tau_{r_1F_n}\big)^{-1}\big)\\
&= \tau_{r_1F_n}^{-1}C_nY_n^+ + O_p\big(\big(n^{1/2}\tau_{r_1F_n}\big)^{-1}\big)\\
&\to_p h_3\begin{bmatrix} h_{6,r_1^\diamond}^\diamond & 0^{r_1^\diamond\times(p+1-r_1^\diamond)} \\ 0^{(k-b-r_1^\diamond)\times r_1^\diamond} & 0^{(k-b-r_1^\diamond)\times(p+1-r_1^\diamond)} \end{bmatrix}, \quad \text{where}
\end{aligned} \tag{31.30}
$$

$$h_{6,r_1^\diamond}^\diamond := \mathrm{Diag}\left\{1, h_{6,1}, h_{6,1}h_{6,2}, \ldots, \prod_{\ell=1}^{r_1^\diamond-1} h_{6,\ell}\right\}$$

and the second equality uses $n^{1/2}(\widehat{D}_n^+ - D_n^+) = O_p(1)$, which holds by (31.62) below and Lemma 31.5(b). Note that here, unlike in the fourth line of (17.3) of AG1, no $o_p(1)$ term arises. Also recall again that we leave out the subindex t^* from the notation, for example, in $h_{6,j}$ for $j = 1, \ldots, r_1^\diamond - 1$.

As mentioned above, the proof of Theorem 9.1 now follows the same lines as the proof of Theorem 16.1 for the full vector case. The roles of k, $h_{5,g}^{-1/2}\overline{g}_h$, $n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n$, and $\overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q}$ in the proof of Theorem 16.1 are played by $k-b$, l_h (defined in Theorem 31.7), $n^{1/2}\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n) \times \widehat{D}_n(\theta_0, \widehat{\beta}_n)\widehat{U}_n$, and $\overline{\Delta}^{S\prime}_{h,p-q^S}h_{3,k-b-q^S}h'_{3,k-b-q^S} \times \overline{\Delta}^{S}_{h,p-q^S}$, respectively. By Lemma 31.1, the almost sure representation argument used in the proof of the full vector result, and Lemma 31.3, we have

$$
\begin{aligned}
&c_{k,p}\big(n^{1/2}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)\widehat{U}_n, \widetilde{J}_n(\theta_0, \widehat{\beta}_n), 1-\alpha\big)\\
&\quad = c_{k-b,p}\big(\widehat{\Upsilon}_n, 0^{(k-b)\times 0}, 1-\alpha\big)\\
&\quad = c_{k-b,p}(\widehat{\Upsilon}_n, 1-\alpha)\\
&\quad \to_d c_{k-b,p,q^S}\big(h'_{3,k-b-q^S}\overline{\Delta}^{S}_{h,p-q^S}, 1-\alpha\big),
\end{aligned}
\tag{31.31}
$$

where $\widehat{\Upsilon}_n$ denotes the matrix of singular values of $n^{1/2}\widetilde{O}'_{F_n}\widetilde{\Omega}_n^{-1/2}(\theta_0, \widehat{\beta}_n)\widehat{D}_n(\theta_0, \widehat{\beta}_n)\widehat{U}_n$, defined as in (27.8), $c_{k-b,p,q^S}(\cdot, 1-\alpha)$ is defined in (31.19) (and $c_{k-b,p,q^S}(h'_{3,k-b-q^S}\overline{\Delta}^{S}_{h,p-q^S}, 1-\alpha)$ uses the notation in (27.12)), and the convergence in (31.31) is joint with the convergence in Theorem 31.7.

To conclude the proof of Theorem 9.1, we state the equivalent of Lemma 27.4 for the subvector case, which verifies that Assumption WU^S holds when $\widehat{U}_n$ is defined as $\widehat{L}_n^{1/2}$, where $\widehat{L}_n := (\theta_0, I_p)(\widehat{\Sigma}_n^{\varepsilon}(\theta_0, \widehat{\beta}_n))^{-1}(\theta_0, I_p)' \in R^{p\times p}$ is defined in (47). Furthermore, the following lemma shows that $\mathcal{F}^S = \mathcal{F}^S_{\mathrm{WU}}$, where $\mathcal{F}^S$ is defined in (53) and $\mathcal{F}^S_{\mathrm{WU}}$ is defined just below (31.16). Recall the definition $\Sigma_{j\ell}(\Omega_F, R_F) := \mathrm{tr}(R'_{j\ell F}\Omega_F^{-1})/k$ for the (j,ℓ)-th component of Σ, where $\Omega_F := E_F g_i g_i'$, $V_F := E_F(f_i - E_F f_i)(f_i - E_F f_i)' \in R^{(p+1)k\times(p+1)k}$, $R_F := (B' \otimes I_k)V_F(B \otimes I_k) \in R^{(p+1)k\times(p+1)k}$ in (16.7). Also, recall the definition of $\widehat{R}_n(\theta_0, \widehat{\beta}_n) := (B' \otimes I_k)\widehat{V}_n(\theta_0, \widehat{\beta}_n)(B \otimes I_k)$, which is given by (19) with $(\theta_0, \widehat{\beta}_n)$ in place of θ.

Lemma 31.9. (a) *Assumption* WU^S *holds with* $\widehat{U}_{2n} = (\widehat{\Omega}_n(\theta_0, \widehat{\beta}_n), \widehat{R}_n(\theta_0, \widehat{\beta}_n))$, $U_1(U_{2F}) = U_1(\Omega_F, R_F) = ((\theta_0, I_p)(\Sigma^{\varepsilon}(\Omega_F, R_F))^{-1}(\theta_0, I_p)')^{1/2}$ *defined in* (16.8), *and* $h_8 = \lim U_{2F_{wn}} = \lim(\Omega_{F_{wn}}, R_{F_{wn}})$, *under any sequence* $\{\lambda_{w_n,h} \in \Lambda_*^S : n \geq 1\}$, *and*

(b) $\mathcal{F}^S = \mathcal{F}^S_{\mathrm{WU}}$ *for* δ_1 *sufficiently small and* M_1 *sufficiently large in the definition of* $\mathcal{F}^S_{\mathrm{WU}}$.

The proof of Lemma 31.9 follows the same lines as the proof of Lemma 27.4. As in (27.73), we have

$$
\widehat{V}_n(\theta_0, \widehat{\beta}_n) = E_{F_n} f_i f_i' - (E_{F_n} f_i)(E_{F_n} f_i)' + o_p(1) \tag{31.32}
$$

and

$$\begin{aligned}\widehat{R}_n(\theta_0, \widehat{\beta}_n) &= \big(B' \otimes I_k\big)\big(E_{F_n} f_i f_i' - (E_{F_n} f_i)\big(E_{F_n} f_i'\big)\big)(B \otimes I_k) + o_p(1) \\ &\to_p R_h := \big(B' \otimes I_k\big)\big[h_5 - \mathrm{vec}\big((0^k, h_4)\big)\,\mathrm{vec}\big((0^k, h_4)\big)'\big](B \otimes I_k),\end{aligned} \tag{31.33}$$

where the convergence holds by results stated (or proved exactly as) in the proof of Lemma 31.5(b) below. This implies that Assumption WU^S(a) holds, namely, $\widehat{U}_{2w_n} - U_{2F_{w_n}} = (\widehat{\Omega}_{w_n}(\theta_0, \widehat{\beta}_n), \widehat{R}_{w_n}(\theta_0, \widehat{\beta}_n)) - (\Omega_{F_{w_n}}, R_{F_{w_n}}) = o_p(1)$. Assumption WU^S(b) holds by the same argument as the one for the full vector case that starts in the paragraph containing (27.75). This establishes Lemma 31.9(a).

Lemma 31.9(b) holds by the same argument as the one for the full vector case that starts after the paragraph that contains (27.77).

31.2 *Proof of Lemma 31.5*

Throughout the proof, we use the shorthand notation $g_i(\beta) = g_i(\theta_0, \beta)$ and $\widehat{g}_n(\beta) = n^{-1}\sum_{i=1}^n g_i(\theta_0, \beta)$ and write g_i for $g_i(\beta^*)$, where β^* is the true value of β, and analogously for other expressions, for example, we write $\widehat{D}_n(\beta)$ for $\widehat{D}_n(\theta_0, \beta)$ and G_i for $G_i(\theta_0, \beta^*)$. Furthermore, to simplify notation, we replace subscripts F_n by n, for example, we write E_n, rather than E_{F_n}.

Proof of Lemma 31.5(a). Given $\{\lambda^S_{n,h} : n \geq 1\}$, let F_n and β^*_n denote the distribution of W_i and the true parameter β when the sample size is n. Let $\widehat{Q}_n(\beta) = \|\widehat{g}_n(\beta)\|^2$ and $Q_n(\beta) = \|E_n g_i(\beta)\|^2$, where a subscript n on E or P denotes expectation or probability under F_n, respectively. The following proof adapts the standard proof for consistency of extremum estimators to the case of drifting DGP's $\{\lambda^S_{n,h} : n \geq 1\}$.

(a1). We first show consistency of the first-step estimator, that is, $\widetilde{\beta}_n - \beta^*_n \to_p 0^b$ under $\{\lambda^S_{n,h} : n \geq 1\}$. Let $\varepsilon > 0$. By the identifiability condition in $\mathcal{F}^S_{\mathrm{AR},1}$ in (50), there exists $\delta_\varepsilon > 0$ such that $\beta \in B \backslash B(\beta^*_n, \varepsilon e)$ implies $Q_n(\beta) \geq \delta_\varepsilon$. Thus,

$$\begin{aligned} P_n\big(\|\widetilde{\beta}_n - \beta^*_n\| > \varepsilon\big) &= P_n\big(\widetilde{\beta}_n \in B\backslash B\big(\beta^*_n, \varepsilon\big)\big) \\ &\leq P_n\big(Q_n(\widetilde{\beta}_n) - \widehat{Q}_n(\widetilde{\beta}_n) + \widehat{Q}_n(\widetilde{\beta}_n) \geq \delta_\varepsilon\big) \\ &\leq P_n\big(Q_n(\widetilde{\beta}_n) - \widehat{Q}_n(\widetilde{\beta}_n) + \widehat{Q}_n\big(\beta^*_n\big) \geq \delta_\varepsilon\big) \\ &\leq P_n\Big(2\sup_{\beta \in B}\big|Q_n(\beta) - \widehat{Q}_n(\beta)\big| \geq \delta_\varepsilon\Big) \\ &\to 0, \end{aligned}$$

where the second inequality holds because $\widehat{Q}_n(\cdot)$ is minimized by $\widetilde{\beta}_n$, the third inequality holds because $Q_n(\beta^*_n) = 0$, and the convergence result holds, because, as we show now, $\sup_{\beta \in B} |\widehat{Q}_n(\beta) - Q_n(\beta)| \to_p 0$.

For $\delta > 0$, define

$$Y_{i\delta} := \sup_{\beta \in B}\ \sup_{\beta' \in B(\beta, \delta)} \big\|g_i\big(\beta'\big) - g_i(\beta)\big\|, \tag{31.34}$$

whose distribution depends on F_n. By Assumption gB, $g_i(\cdot)$ is uniformly continuous on B and therefore $Y_{i\delta} \to 0$ a.s. $[\mu]$ as $\delta \to 0$. Furthermore, $E_\mu Y_{i\delta} \le 2E_\mu \sup_{\beta\in B} \|g_i(\beta)\| < \infty$, where the latter inequality holds by the conditions in $\mathcal{F}^S_{\mathrm{AR},1}$. Therefore, by the dominated convergence theorem (DCT) it follows that $E_\mu Y_{i\delta} \to 0$ as $\delta \to 0$. Let f_n denote the Radon–Nikodym derivative of F_n wrt μ and note that by assumption $f_n \le M$. We have $\sup_n E_n Y_{i\delta} = \sup_n E_\mu f_n Y_{i\delta} \le E_\mu M Y_{i\delta} \to 0$ as $\delta \to 0$.

By Assumption gB, B is compact. Therefore, for $\delta > 0$ there is a finite cover of B by balls of radius δ centered at some points β_j, $j = 1, \dots, J_\delta$, that is, $B \subset \bigcup_{j=1}^{J_\delta} B(\beta_j, \delta)$. Let

$$H_n(\beta) = \widehat{g}_n(\beta) - E_n g_i(\beta). \tag{31.35}$$

Because $\mathcal{F}^S_{\mathrm{AR},1}$ imposes $\sup_{\beta\in B} E_F\|g_i(\beta)\|^{1+\gamma} \le M$, a Lyapunov-type WLLN implies that for any fixed $\beta \in B$ we have $H_n(\beta) \to_p 0^k$ as $n \to \infty$. It then follows that for $\varepsilon > 0$ we have

$$\begin{aligned}
&P_n\Big(\sup_{\beta\in B}\big\|H_n(\beta)\big\| > 2\varepsilon\Big) \\
&\quad\le P_n\Big(\max_{j=1,\dots,J_\delta}\sup_{\beta\in B(\beta_j,\delta)}\big\|H_n(\beta) - H_n(\beta_j)\big\| + \big\|H_n(\beta_j)\big\| > 2\varepsilon\Big) \\
&\quad\le P_n\Big(\sup_{\beta\in B}\sup_{\beta'\in B(\beta,\delta)}\big\|H_n(\beta') - H_n(\beta)\big\| > \varepsilon\Big) + P_n\Big(\max_{j=1,\dots,J_\delta}\big\|H_n(\beta_j)\big\| > \varepsilon\Big),
\end{aligned} \tag{31.36}$$

where the first inequality holds by the triangle inequality.

For the first summand in (31.36), we have the following bound:

$$\begin{aligned}
&P_n\Big(\sup_{\beta\in B}\sup_{\beta'\in B(\beta,\delta)}\big\|H_n(\beta') - H_n(\beta)\big\| > \varepsilon\Big) \\
&\quad\le P_n\left(\frac{1}{n}\sum_{i=1}^n (Y_{i\delta} + E_n Y_{i\delta}) > \varepsilon\right) \\
&\quad\le E_n\frac{1}{n}\sum_{i=1}^n (Y_{i\delta} + E_n Y_{i\delta})/\varepsilon \\
&\quad= 2E_n Y_{i\delta}/\varepsilon,
\end{aligned} \tag{31.37}$$

where the first inequality holds by the triangle inequality and the second inequality holds by Markov's inequality. Because, as shown above, $\sup_n E_n Y_{i\delta} \to 0$ as $\delta \to 0$, for given $\nu > 0$ there is $\delta_\nu > 0$ such that $2E_n Y_{i\delta}/\varepsilon < \nu/2$ for all n and for all $\delta \le \delta_\nu$. Because $H_n(\beta) \to_p 0$, we can find a finite $n_{\delta_\nu} \in N$ such that for all $n \ge n_{\delta_\nu}$ we have $P_n(\max_{j=1,\dots,J_{\delta_\nu}} \|H_n(\beta_j)\| > \varepsilon) < \nu/2$. This proves

$$P_n\Big(\sup_{\beta\in B}\big\|H_n(\beta)\big\| > 2\varepsilon\Big) \to 0 \tag{31.38}$$

as $n \to \infty$. By the reverse triangle inequality, we then obtain the desired $\sup_{\beta\in B} |\widehat{Q}_n(\beta) - Q_n(\beta)| \to_p 0$ as $n \to \infty$.

(a2). Next, we show consistency of $\widehat{\beta}_n$. Let $\{\beta_n - \beta_n^\dagger : n \geq 1\}$ be any nonstochastic sequence that converges to 0^b. We can write $E_n g_i(\beta_n) - E_n g_i(\beta_n^\dagger) = E_\mu h_n$ for $h_n = (g_i(\beta_n) - g_i(\beta_n^\dagger)) f_n$. Because $f_n \leq M$ and $g_i(\cdot)$ is uniformly continuous on B by Assumption gB, it follows that $h_n \to 0^k$ a.s. $[\mu]$. Furthermore, $E_\mu h_n \leq 2M E_\mu \sup_{\beta \in B} \|g_i(\beta)\| < \infty$ by the conditions in $\mathcal{F}_{\mathrm{AR}}^S$. Therefore, by the DCT, $E_\mu h_n \to 0^k$.

Define $E_n g_i(\widetilde{\beta}_n) = E_n g_i(\beta)|_{\beta = \widetilde{\beta}_n}$. That is, the expectation is taken first treating β as nonrandom, and then the resulting expression is evaluated at the random vector $\widetilde{\beta}_n$. For any given $\varepsilon > 0$,

$$\begin{aligned} \left\|\widehat{g}_n(\widetilde{\beta}_n)\right\| &\leq \left\|\widehat{g}_n(\widetilde{\beta}_n) - E_n g_i(\widetilde{\beta}_n)\right\| + \left\|E_n g_i(\widetilde{\beta}_n) - E_n g_i\big(\beta_n^*\big)\right\| \\ &\leq \sup_{\beta \in B(\beta_n^*, \varepsilon)} \left\|\widehat{g}_n(\beta) - E_n g_i(\beta)\right\| + o_p(1) \\ &= o_p(1), \end{aligned} \tag{31.39}$$

where the first inequality holds by the triangle inequality, the second inequality holds wp $\to 1$ because $\widetilde{\beta}_n - \beta_n^* \to_p 0^b$ and $E_\mu h_n \to 0^k$, and the equality holds by (31.38).

Furthermore, $n^{-1} \sum_{i=1}^n g_i(\widetilde{\beta}_n) g_i(\widetilde{\beta}_n)' - E_n g_i g_i' \to_p 0^{k \times k}$. This result is proved as in (31.39), by establishing a UWLLN on $B(\beta_n^*, \varepsilon)$ for $n^{-1} \sum_{i=1}^n g_i(\cdot) g_i(\cdot)'$ and by showing that $E_\mu h_n \to 0^{k \times k}$ for $h_n = (g_i(\beta_n) g_i(\beta_n)' - g_i(\beta_n^\dagger) g_i(\beta_n^\dagger)') f_n$ when $\beta_n - \beta_n^\dagger$ converges to 0^b. The latter follows as above from the DCT using $E_\mu \sup_{\beta \in B(\beta_n^*, \varepsilon)} \|g_i(\beta)\|^2 < \infty$ by the conditions in $\mathcal{F}_{\mathrm{AR}}^S$. The former follows using the same proof as for (31.38) noting that by the conditions in $\mathcal{F}_{\mathrm{AR}}^S$ we have $E_\mu \sup_{\beta \in B(\beta_n^*, \varepsilon)} \|g_i(\beta)\|^2 < \infty$ and $\sup_{\beta \in B(\beta_n^*, \varepsilon)} E_F \|g_i(\beta)\|^{2+\gamma}$ is uniformly bounded. We have therefore shown that $(\widehat{\varphi}_n' \widehat{\varphi}_n)^{-1} - E_n g_i g_i' \to_p 0^{k \times k}$, where $\widehat{\varphi}_n' \widehat{\varphi}_n$ is defined in (42). Because $\mathcal{F}_{\mathrm{AR}}^S$ imposes $\lambda_{\min}(E_F g_i g_i') \geq \delta$, it follows that

$$\widehat{\varphi}_n' \widehat{\varphi}_n - \big(E_n g_i g_i'\big)^{-1} \to_p 0^{k \times k}. \tag{31.40}$$

The remainder of the consistency proof is analogous to the proof in part (a1), but with $\widehat{Q}_n(\beta) := \|\widehat{\varphi}_n \widehat{g}_n(\beta)\|^2$ and $Q_n(\beta) := \|(E_n g_i g_i')^{-1/2} E_n g_i(\beta)\|^2$. To establish a UWLLN for $\widehat{Q}_n(\beta)$, note that

$$\begin{aligned} &\sup_{\beta \in B(\beta_n^*, \varepsilon)} \left\|\widehat{\varphi}_n \widehat{g}_n(\beta) - \big(E_n g_i g_i'\big)^{-1/2} E_n g_i(\beta)\right\| \\ &\leq \|\widehat{\varphi}_n\| \sup_{\beta \in B(\beta_n^*, \varepsilon)} \left\|\widehat{g}_n(\beta) - E_n g_i(\beta)\right\| + \left\|\widehat{\varphi}_n - \big(E_n g_i g_i'\big)^{-1/2}\right\| \sup_{\beta \in B(\beta_n^*, \varepsilon)} \left\|E_n g_i(\beta)\right\| \\ &= o_p(1), \end{aligned} \tag{31.41}$$

where the inequality uses the triangle inequality and the equality uses (31.38), (31.40), the assumption that $\sup_{\beta \in B(\beta_n^*, \varepsilon)} \|E_n g_i(\beta)\|$ is uniformly bounded by a finite number, and that $\|\widehat{\varphi}_n\| = O(1)$ because $\lambda_{\min}(E_n g_i g_i') \geq \delta$. Equation (31.41) implies that $\sup_{\beta \in B(\beta_n^*, \varepsilon)} |\widehat{Q}_n(\beta) - Q_n(\beta)| = o_p(1)$.

(a3). Now, we derive the limiting distribution of $\widehat{\beta}_n$ under $\{\lambda_{n,h}^S : n \geq 1\}$. As above, $\widehat{Q}_n(\beta) := \|\widehat{\varphi}_n \widehat{g}_n(\beta)\|^2$. Because β_n^* is bounded away from the boundary of B, $\widehat{\beta}_n - \beta_n^* \to_p$

0^b, and $g_i(\cdot) \in C^2(B(\beta_n^*, \vartheta))$, the following FOC holds wp $\to 1$ and element-by-element mean-value expansions of $\frac{\partial}{\partial\beta} Q_n(\widehat{\beta}_n)$ exist:

$$0^b = \frac{\partial}{\partial\beta} Q_n(\widehat{\beta}_n) = \frac{\partial}{\partial\beta} Q_n\big(\beta_n^*\big) + \frac{\partial^2}{\partial\beta\partial\beta'} Q_n\big(\beta_n^+\big)\big(\widehat{\beta}_n - \beta_n^*\big), \tag{31.42}$$

where the mean-value β_n^+ lies on the segment joining $\widehat{\beta}_n$ and β_n^* (and hence satisfies $\widehat{\beta}_n - \beta_n^* \to_p 0^b$).

For $m, j = 1, \dots, b$, we have

$$\begin{aligned} \frac{\partial}{\partial\beta} Q_n(\beta) &= \left[n^{-1} \sum_{i=1}^n \frac{\partial}{\partial\beta'} g_i(\beta) \right]' \widehat{\varphi}_n' \widehat{\varphi}_n n^{-1} \sum_{i=1}^n g_i(\beta) \quad \text{and} \\ \left[\frac{\partial^2}{\partial\beta\partial\beta'} Q_n(\beta) \right]_{mj} &= n^{-1} \sum_{i=1}^n \frac{\partial}{\partial\beta_m} g_i(\beta)' \widehat{\varphi}_n' \widehat{\varphi}_n n^{-1} \sum_{i=1}^n \frac{\partial}{\partial\beta_j} g_i(\beta) \\ &\quad + n^{-1} \sum_{i=1}^n \frac{\partial^2}{\partial\beta_m\partial\beta_j} g_i(\beta)' \widehat{\varphi}_n' \widehat{\varphi}_n n^{-1} \sum_{i=1}^n g_i(\beta). \end{aligned} \tag{31.43}$$

By the argument in (31.39), $n^{-1} \sum_{i=1}^n g_i(\beta_n^+) \to_p 0^k$ under $\{\lambda_{n,h}^S : n \geq 1\}$. Furthermore,

$$n^{-1} \sum_{i=1}^n \frac{\partial}{\partial\beta_j} g_i\big(\beta_n^+\big) - E_n \frac{\partial}{\partial\beta_j} g_i \to_p 0^k \tag{31.44}$$

under $\{\lambda_{n,h}^S : n \geq 1\}$. The latter holds by the argument in (31.39) with $g_i(\widetilde{\beta}_n)$ replaced by $\frac{\partial}{\partial\beta_j} g_i(\beta_n^+)$ and using the assumptions $\sup_{\beta \in B(\beta_n^*, \vartheta)} E_n \|G_{i\beta}(\beta)\|^{1+\gamma}$ and $E_\mu \sup_{\beta \in B(\beta_n^*, \vartheta)} \|G_{i\beta}(\beta)\|$ are uniformly bounded in $\mathcal{F}^S$. In addition, $n^{-1} \sum_{i=1}^n \frac{\partial^2}{\partial\beta_m\partial\beta_j} \times g_i(\beta_n^+) = O_p(1)$ (again by an argument as in (31.39) with $g_i(\widetilde{\beta}_n)$ replaced by $\frac{\partial^2}{\partial\beta_m\partial\beta_j} g_i(\beta_n^+)$ and using the fact that $\sup_{\beta \in B(\beta_n^*, \vartheta)} E_n \| \frac{\partial^2}{\partial\beta_m\partial\beta_j} g_i(\beta) \|^{1+\gamma}$ and $E_\mu \sup_{\beta \in B(\beta_n^*, \vartheta)} \| \frac{\partial^2}{\partial\beta_m\partial\beta_j} \times g_i(\beta) \|$ are uniformly bounded by the conditions in $\mathcal{F}^S$). It follows that $\frac{\partial^2}{\partial\beta\partial\beta'} Q_n(\beta_n^+) - B_n^* \to_p 0^{b \times b}$ under $\{\lambda_{n,h}^S : n \geq 1\}$, where $B_n^* := E_n G_{i\beta}' (E_n g_i g_i')^{-1} E_n G_{i\beta}$.

Because $\lambda_{\min}(B_n^*)$ is bounded away from zero (since $\tau_{\min}(E_n G_{i\beta}) \geq \delta$ for $F_n \in \mathcal{F}^S$), it follows that $\frac{\partial^2}{\partial\beta\partial\beta'} Q_n(\beta_n^+)$ is invertible wp $\to 1$. This and (31.42) give

$$n^{1/2}\big(\widehat{\beta}_n - \beta_n^*\big) = -\big(B_n^* + o_p(1)\big)^{-1} \sqrt{n} \frac{\partial}{\partial\beta} Q_n\big(\beta_n^*\big). \tag{31.45}$$

From above, we have

$$n^{1/2} \frac{\partial}{\partial\beta} Q_n\big(\beta_n^*\big) = (E_n G_{i\beta})' \big(E_n g_i g_i'\big)^{-1} n^{-1/2} \sum_{i=1}^n g_i\big(\beta_n^*\big) + o_p(1). \tag{31.46}$$

By the CLT result in (31.20), $n^{-1/2} \sum_{i=1}^n g_i(\beta_n^*) \to_d \overline{g}_h$. Combining the previous results and using the definition of the vector h, we obtain the result of Lemma 31.5(a). $\square$

Proof of Lemma 31.5(b). Using Lemma 31.5(a) or the same argument employed multiple times in the proof of Lemma 31.5(a), we have: $n^{-1}\sum_{i=1}^n g_i(\widehat{\beta}_n) \rightarrow_p 0^k$, $n^{-1} \times \sum_{i=1}^n g_i(\widehat{\beta}_n) g_{ij}(\widehat{\beta}_n) - E_n g_i g_{ij} \rightarrow_p 0^k$, $n^{-1}\sum_{i=1}^n g_i(\widehat{\beta}_n) g_{ij}(\widehat{\beta}_n) g_i(\widehat{\beta}_n)' - \lambda^S_{5,3,j,n} \rightarrow_p 0^{k\times k}$, $\widehat{\Omega}_n^{-1}(\widehat{\beta}_n) - (E_n g_i g_i)^{-1} \rightarrow_p 0^{k\times k}$, $n^{-1}\sum_{i=1}^n G_{i\beta_j}(\widehat{\beta}_n) - \lambda^S_{4,\beta_j,n} \rightarrow_p 0^k$, and $n^{-1} \times \sum_{i=1}^n G_{i\beta_j}(\widehat{\beta}_n) g_i(\widehat{\beta}_n)' - \lambda^S_{5,\beta_j,n} \rightarrow_p 0^{k\times k}$. Therefore, $\widetilde{\Omega}_n(\widehat{\beta}_n) \rightarrow_p h_{5,g}$, $\widetilde{G}_{\beta n}(\widehat{\beta}_n) \rightarrow_p h_{4,\beta}$, and $\widetilde{J}_n(\widehat{\beta}_n) \rightarrow_p h_{5,g}^{-1/2} h_{4,\beta}$. □

Proof of Lemma 31.5(c). We derive the limit distributions of (i) $\widehat{g}_n(\widehat{\beta}_n)$, (ii) $\widehat{D}_n(\widehat{\beta}_n) - E_n G_i$, (iii) $\widetilde{\Omega}_n(\theta_0, \widehat{\beta}_n) - E_n g_i g_i'$, and (iv) $\widetilde{G}_{\beta n}(\theta_0, \widehat{\beta}_n) - E_n G_{i\beta}$ under $\{\lambda^S_{n,h} : n \geq 1\}$ in (c1)–(c4) below, respectively.

(c1). We have

$$n^{1/2}\widehat{g}_n(\widehat{\beta}_n) = n^{1/2}\widehat{g}_n(\beta_n^*) + \widehat{G}_{\beta n}(\beta_n^+) n^{1/2}(\widehat{\beta}_n - \beta_n^*)$$
$$= \left(I_k - (E_n G_{i\beta}) B_n^{*-1} (E_n G_{i\beta})' \left(E_n g_i g_i'\right)^{-1}\right) n^{1/2}\widehat{g}_n(\beta_n^*) + o_p(1), \quad (31.47)$$

where the first equality uses a mean-value expansion with β_n^+ on the segment joining $\widehat{\beta}_n$ and β_n^* and the second equality holds by (31.45) and (31.46). Therefore,

$$n^{1/2}\widehat{g}_n(\widehat{\beta}_n) \rightarrow_d \overline{g}^S_h := h_{5,g}^{1/2} M_{h_{5,g}^{-1/2} h_{4,\beta}} h_{5,g}^{-1/2} \overline{g}_h. \quad (31.48)$$

Note that the assumption of strong identification of β, namely $\tau_{\min}(E_F G_{i\beta}) \geq \delta$ in $\mathcal{F}^S$, implies that $h_{4,\beta}$ has full column rank b.

(c2). Recall the definition of $\widehat{\Gamma}_{jn}(\cdot)$ for $j = 1, \ldots, p$ in (18). Under sequences $\{\lambda^S_{n,h} : n \geq 1\}$, we have

$$\left(\widehat{\Gamma}'_{1n}(\widehat{\beta}_n), \ldots, \widehat{\Gamma}'_{pn}(\widehat{\beta}_n)\right)' \widehat{\Omega}_n^{-1}(\widehat{\beta}_n) - \left(\left(E_n G_{i1} g_i'\right)', \ldots, \left(E_n G_{ip} g_i'\right)'\right)' \Omega_n^{-1} \rightarrow_p 0, \quad (31.49)$$

which is established analogously to the results in (31.40) and (31.44), using the uniform finite bounds on $\sup_{\beta \in B(\beta_n^*, \vartheta)} E_n \|(\frac{\partial}{\partial \beta'} g_i(\beta)) g_{ij}(\beta)\|^{1+\gamma}$ and $E_\mu \sup_{\beta \in B(\beta_n^*, \vartheta)} \|(\frac{\partial}{\partial \beta'} g_i(\beta)) \times g_{ij}(\beta)\|$ for $j = 1, \ldots, k$ in $\mathcal{F}^S$.

Using $g_i(\cdot) \in C^2(B(\beta^*, \vartheta))$, by a second-order Taylor expansion of $\widehat{g}_n(\widehat{\beta}_n)$ about β_n^* and a mean-value expansion of $G_i(\widehat{\beta}_n)$ (as in (31.47)), we obtain

$$n^{1/2}\operatorname{vec}\left(\widehat{D}_n(\widehat{\beta}_n) - D_n\right)$$
$$= n^{-1/2}\sum_{i=1}^n \left[\operatorname{vec}(G_i - E_n G_i) - \begin{pmatrix} E_n G_{\ell 1} g_\ell' \\ \vdots \\ E_n G_{\ell p} g_\ell' \end{pmatrix} \Omega_n^{-1} g_i\right]$$
$$+ \operatorname{vec}\left((E_n G_{i\theta_1\beta}) n^{1/2}(\widehat{\beta}_n - \beta_n^*), \ldots, (E_n G_{i\theta_p\beta}) n^{1/2}(\widehat{\beta}_n - \beta_n^*)\right)$$
$$- \left[\begin{pmatrix} E_n G_{\ell 1} g_\ell' \\ \vdots \\ E_n G_{\ell p} g_\ell' \end{pmatrix} \Omega_n^{-1}\right] (E_n G_{i\beta}) n^{1/2}(\widehat{\beta}_n - \beta_n^*) + o_p(1), \quad (31.50)$$

where $E_n G_{\ell j} g_\ell' = E_n G_{ij} g_i'$ for any observation indices $\ell, i \geq 1$ by stationarity. The terms on the rhs of the first line of (31.50) consist of the term $D_n = E_n G_i$ and the first term of the expansions of $G_i(\widehat{\beta}_n)$ and $g_i(\widehat{\beta}_n)$, respectively, replacing sample averages by expectations as in (31.49). The term in the second line comes from the second term of the expansion of $G_i(\widehat{\beta}_n)$. For this, we use

$$\widehat{G}_{\theta_j \beta n}(\beta_n^+) - E_n G_{i\theta_j\beta} \rightarrow_p 0^{k\times b} \quad \text{for } j = 1, \ldots, p \tag{31.51}$$

for any sequence β_n^+ such that $\beta_n^+ - \beta_n^* \rightarrow_p 0$. The latter is established (as in several places above) using the assumptions that $\sup_{\beta\in B(\beta_n^*,\vartheta)} E_n \| \frac{\partial^2}{\partial\theta_\iota \partial\beta'} g_i(\beta)\|^{1+\gamma}$ and $E_\mu \sup_{\beta\in B(\beta_n^*,\vartheta)} \|\frac{\partial^2}{\partial\theta_\iota\partial\beta'} g_i(\beta)\|$ are uniformly bounded in $\mathcal{F}^S$. The first term of the third line comes from the second term of the expansion of $g_i(\widehat{\beta}_n)$ and using (31.44) and (31.49). The $o_p(1)$ term contains the errors caused by the approximations in (31.49) and (31.51) and from the third term of the expansion of $g_i(\widehat{\beta}_n)$ (which is indeed $o_p(1)$ given the moment bounds in $\mathcal{F}^S$ on $\frac{\partial^2}{\partial\theta_\iota\partial\beta'} g_i(\beta)$).

Equation (31.50), combined with Lemma 31.5(a), (31.20), (31.14), and the paragraph containing (31.16), give

$$\begin{aligned} & n^{1/2}\operatorname{vec}\big(\widehat{D}_n(\widehat{\beta}_n) - D_n\big) \\ & \quad \rightarrow_d \operatorname{vec}\big(\overline{D}_h^S\big) \\ & \quad := \big(\operatorname{vec}(\overline{G}_h) - h_{5,Gg} h_{5,g}^{-1} \overline{g}_h\big) + \operatorname{vec}(h_{4,\theta_1\beta}\overline{\beta}_h, \ldots, h_{4,\theta_p\beta}\overline{\beta}_h) \\ & \qquad - h_{5,Gg} h_{5,g}^{-1} h_{4,\beta}\overline{\beta}_h. \end{aligned} \tag{31.52}$$

Note that $\overline{g}_h^S$ and $\overline{\beta}_h$ are independent because

$$\begin{aligned} & \operatorname{cov}\big(M_{h_{5,g}^{-1/2} h_{4,\beta}} h_{5,g}^{-1/2} \overline{g}_h, \overline{\beta}_h\big) \\ & = M_{h_{5,g}^{-1/2} h_{4,\beta}} \big(h_{5,g}^{-1/2} h_{4,\beta}\big)\big[\big(h_{5,g}^{-1/2} h_{4,\beta}\big)'\big(h_{5,g}^{-1/2} h_{4,\beta}\big)\big]^{-1} = 0^{k\times b}. \end{aligned} \tag{31.53}$$

Next, we establish that $\overline{g}_h^S$ and $\overline{D}_h^S$ (defined in (31.21)) are independent. The last two summands that make up $\overline{D}_h^S$ are independent of $\overline{g}_h^S$ because $\overline{g}_h^S$ and $\overline{\beta}_h$ are independent. Regarding the first summand, recall that from (31.20) we know that $\operatorname{vec}(\overline{G}_h)$ and $\overline{g}_h$ are jointly normally distributed and because $\operatorname{cov}(\overline{g}_h, \operatorname{vec}(\overline{G}_h) - h_{5,Gg} h_{5,g}^{-1}\overline{g}_h) = 0^{k\times pk}$, it follows that $\operatorname{vec}(\overline{G}_h) - h_{5,Gg} h_{5,g}^{-1}\overline{g}_h$ and $\overline{g}_h^S$ are independent.

(c3). Next, we derive the asymptotic distribution of $\widetilde{\Omega}_n(\widehat{\beta}_n)$. Let $j \in \{1, \ldots, k\}$. By a mean-value expansion, for some vectors β_n^+ and $\beta_n^\dagger$ on the line segment joining $\widehat{\beta}_n$ and β_n^*, under $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$, we have

$$\begin{aligned} & n^{1/2}\big[\widetilde{\Omega}_{jn}(\widehat{\beta}_n) - \Omega_{jn}\big] \\ & \quad = n^{1/2}\left[n^{-1}\sum_{i=1}^n g_i g_{ij} - \Omega_{jn}\right] \end{aligned}$$

$$
\begin{aligned}
&+ n^{-1}\sum_{i=1}^{n}\left[G_{i\beta}(\beta_n^{+})g_{ij}(\beta_n^{+}) + g_i(\beta_n^{+})\left(\frac{\partial g_{ij}(\beta_n^{+})}{\partial\beta}\right)'\right]n^{1/2}(\widehat{\beta}_n - \beta_n^{*})\\
&- \widehat{\Phi}_{jn}(\widehat{\beta}_n)\widehat{\Omega}_n^{-1}(\widehat{\beta}_n)\left[n^{1/2}\widehat{g}_n + G_{i\beta}(\beta_n^{\dagger})n^{1/2}(\widehat{\beta}_n - \beta_n^{*})\right] + o_p(1)\\
\to_d\ & \overline{\varkappa}_{h,j}^{S}\\
:=\ & \overline{L}_{j,h,3} - h_{5,3,j}h_{5,g}^{-1}\overline{g}_h\\
&+ \left[(h_{5,\beta_1,j},\ldots,h_{5,\beta_b,j}) + \left((h_{5,\beta_1})_j',\ldots,(h_{5,\beta_b})_j'\right) - h_{5,3,j}h_{5,g}^{-1}h_{4,\beta}\right]\overline{\beta}_h, \quad (31.54)
\end{aligned}
$$

where $\overline{L}_{j,h,3} \in R^k$ denotes the $(j-1)k+1,\ldots,jk$ components of $\overline{L}_{h,3}$, $(h_{5,\beta_l})_j' \in R^k$ denotes the jth column of $(h_{5,\beta_l})' \in R^{k\times k}$ for $l = 1,\ldots,b$, and the convergence result holds by the moment restrictions in the parameter space, WLLN's, (31.20), and part (a) of the lemma. Equation (31.54) yields $n^{1/2}(\widetilde{\Omega}_n(\widehat{\beta}_n) - \Omega_n) \to_d \overline{\varkappa}_h^S := (\overline{\varkappa}_{h,1}^S,\ldots,\overline{\varkappa}_{h,k}^S)$.

By definition, $\overline{\varkappa}_{h,j}^S$ is a nonrandom function of $\overline{L}_{j,h,3} - h_{5,3,j}h_{5,g}^{-1}\overline{g}_h$ and $\overline{\beta}_h$. By (31.53), $\overline{\beta}_h$ and $\overline{g}_h^S$ are independent. In addition, $\overline{L}_{j,h,3} - h_{5,j}h_{5,g}^{-1}\overline{g}_h$ and $\overline{g}_h$ are independent, because they are jointly normal with a zero covariance matrix. Therefore, $\overline{L}_{j,h,3} - h_{5,j}h_{5,g}^{-1}\overline{g}_h$ and $\overline{g}_h^S$ $(:= h_{5,g}^{1/2}M_{h_{5,g}^{-1/2}h_{4,\beta}}h_{5,g}^{-1/2}\overline{g}_h)$ are independent. This shows that $\overline{\varkappa}_h^S := (\overline{\varkappa}_{h,1}^S,\ldots,\overline{\varkappa}_{h,k}^S)$ and $\overline{g}_h^S$ are independent.

Equation (31.54) yields $n^{1/2}(\widetilde{\Omega}_{jn}(\widehat{\beta}_n) - \Omega_{jn}) \to_d \overline{\varkappa}_h^S$, as desired.

(c4). As in (31.54), for $j \in \{1,\ldots,b\}$, under $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$, we have

$$
\begin{aligned}
&n^{1/2}\left[\widetilde{G}_{\beta_j n}(\widehat{\beta}_n) - E_nG_{i\beta_j}\right]\\
&= n^{1/2}\left[n^{-1}\sum_{i=1}^{n}G_{i\beta_j n} - E_nG_{i\beta_j}\right] + n^{-1}\sum_{i=1}^{n}G_{i\beta_j\beta}(\beta_n^{+})n^{1/2}(\widehat{\beta}_n - \beta_n^{*})\\
&\quad - \widehat{F}_{jn}(\widehat{\beta}_n)\widehat{\Omega}_n(\widehat{\beta}_n)^{-1}\left[n^{1/2}\widehat{g}_n + G_{i\beta}(\beta_n^{\perp})n^{1/2}(\widehat{\beta}_n - \beta_n^{*})\right]\\
&\to_d \overline{\varrho}_{h,j}^S := \overline{L}_{j,h,4} - h_{5,\beta_j}h_{5,g}^{-1}\overline{g}_h + \left(h_{4,\beta_j\beta} - h_{5,\beta_j}h_{5,g}^{-1}h_{4,\beta}\right)\overline{\beta}_h, \quad (31.55)
\end{aligned}
$$

where $\overline{L}_{j,h,4} \in R^k$ denotes the $(j-1)k+1,\ldots,jk$ components of $\overline{L}_{h,4}$. Equation (31.55) and $\overline{\varrho}_h^S := (\overline{\varrho}_{h,1}^S,\ldots,\overline{\varrho}_{h,b}^S)$ yield $n^{1/2}(\widetilde{G}_{\beta n}(\widehat{\beta}_n) - E_nG_{i\beta_j}) \to_d \overline{\varrho}_h^S$, as desired.

By the same argument as for $\overline{\varkappa}_h^S$ above, $\overline{\varrho}_h^S := (\overline{\varrho}_{h,1}^S,\ldots,\overline{\varrho}_{h,b}^S)$ and $\overline{g}_h^S$ are independent. □

Proof of Lemma 31.5(d). First, we obtain the asymptotic distributions of $\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)$, $\widetilde{J}_n(\widehat{\beta}_n)$, and $\widetilde{O}_n = \widetilde{O}_n(\widetilde{J}_n(\widehat{\beta}_n))$.

Consider the function that maps $\text{vec}(\varphi)$ onto $\text{vec}(\varphi^{-1/2})$, where $\varphi \in R^{k\times k}$ is positive definite. Denote by $\overline{\varphi}_h \in R^{k^2\times k^2}$ the matrix of partial derivatives of that mapping evaluated at $\text{vec}(h_{5,g})$. By $n^{1/2}(\widetilde{\Omega}_{jn}(\widehat{\beta}_n) - \Omega_{jn}) \to_d \overline{\varkappa}_h^S$, which holds by part (c) of the lemma (and is proved in (31.54)), and the delta method, we have

$$
n^{1/2}\left[\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) - \Omega_n^{-1/2}\right] \to_d \text{vec}_{k,k}^{-1}\left(\overline{\varphi}_h\,\text{vec}(\overline{\varkappa}_h^S)\right). \quad (31.56)
$$

The asymptotic distribution $\widetilde{J}_n(\widehat{\beta}_n) := \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widetilde{G}_{\beta n}(\widehat{\beta}_n)$ is obtained as follows:

$$
\begin{aligned}
& n^{1/2}\big[\widetilde{J}_n(\widehat{\beta}_n) - \Omega_n^{-1/2} E_n G_{\beta i}\big] \\
&\quad = \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) n^{1/2}\big[\widehat{G}_{\beta n}(\widehat{\beta}_n) - E_n G_{\beta i}\big] + n^{1/2}\big[\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) - \Omega_n^{-1/2}\big] E_n G_{\beta i} \\
&\quad \to_d \overline{\omega}_h^S := h_{5,g}^{-1/2}\overline{\varrho}_h^S + \operatorname{vec}_{k,k}^{-1}\big(\overline{\varphi}_h \operatorname{vec}(\overline{\varkappa}_h^S)\big) h_{4,\beta},
\end{aligned}
\tag{31.57}
$$

where the convergence uses (31.56) and $n^{1/2}(\widetilde{G}_{\beta n}(\widehat{\beta}_n) - E_n G_{i\beta_j}) \to_d \overline{\varrho}_h^S$, which holds by part (c) of the lemma (and is proved in (31.55)).

Assume wlog that the first b columns of $(h_{5,g}^{-1/2} h_{4,\beta})'$ are linearly independent.[75] Then, by (31.4) and (31.5), we have

$$
\begin{aligned}
O_n &= O(J_n) = \big(\big(-j_{n1}'(J_{n1})^{-1}, e_1'\big)', \ldots, \big(-j_{n(k-b)}'(J_{n1})^{-1}, e_{k-b}'\big)'\big) \quad \text{and} \\
\widetilde{O}_n &= O(\widetilde{J}_n) = \big(\big(-\widetilde{j}_{n1}'(\widetilde{J}_{n1})^{-1}, e_1'\big)', \ldots, \big(-\widetilde{j}_{n(k-b)}'(\widetilde{J}_{n1})^{-1}, e_{k-b}'\big)'\big),
\end{aligned}
\tag{31.58}
$$

where again $J_n' = (J_{n1}', J_{n2}')$ with $J_{n1}' \in R^{b\times b}$ and $J_{n2}' = (j_{n1}, \ldots, j_{nk-b}) \in R^{b\times(k-b)}$, $j_{nl} \in R^b$ for $l = 1, \ldots, k-b$, and analogously for $\widetilde{J}_n'$. Consider the function that maps $\operatorname{vec}(J)$ for $J \in R^{k\times b}$ onto $\operatorname{vec}(O(J)) \in R^{k(k-b)}$, where $O(J)$ is defined by (31.4) and (31.5). Denote by $\overline{B}_h \in R^{k(k-b)\times kb}$ the matrix of partial derivatives of that mapping evaluated at $\operatorname{vec}(h_{5,g}^{-1/2} h_{4,\beta})$. Then, by the delta method,

$$
n^{1/2}(\widetilde{O}_n - O_n) \to_d \operatorname{vec}_{k,k-b}^{-1}\big(\overline{B}_h \operatorname{vec}(\overline{\omega}_h^S)\big)
\tag{31.59}
$$

and the asymptotic distribution is independent of $\overline{g}_h^S$.

Given the asymptotic distributions of $\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)$ and $\widetilde{O}_n$, the asymptotic distribution of $n^{1/2}\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_n T_n$ is obtained as follows. We write this matrix in terms of two submatrices:

$$
\begin{aligned}
& n^{1/2}\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_n B_n S_n \\
&\quad = \big(\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_n B_{n,q^S} Y_{n,q^S}^{-1}, n^{1/2}\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_n B_{n,p-q^S}\big).
\end{aligned}
\tag{31.60}
$$

Consider the first component on the rhs of (31.60). By definition, the singular value decomposition of $O_n'\Omega_n^{-1/2}D_n U_n$ is $C_n Y_n B_n'$ and, as in the proof of the full vector case preceding (25.5), we have $O_n'\Omega_n^{-1/2}D_n U_n B_{n,q^S} Y_{n,q^S}^{-1} = C_{n,q^S}$. Hence, we obtain

$$
\begin{aligned}
& \widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_n B_{n,q^S} Y_{n,q^S}^{-1} \\
&\quad = O_n'\Omega_n^{-1/2}D_n U_n B_{n,q^S} Y_{n,q^S}^{-1} \\
&\qquad + n^{1/2}\big(\widetilde{O}_n'\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n'\Omega_n^{-1/2}D_n\big)U_n B_{n,q^S}\big(n^{1/2}Y_{n,q^S}\big)^{-1}
\end{aligned}
$$

[75] If the first b columns of $(h_{5,g}^{-1/2} h_{4,\beta})'$ are linearly dependent, O_n and $\widetilde{O}_n$ are given by analogous formulas involving $R_n' = E_n G_{i\beta}'\Omega_n^{-1/2} = (R_{n1}', R_{n2}')$ and $\widetilde{R}_n(\widehat{\beta}_n)' = (\widetilde{R}_{n1}', \widetilde{R}_{n2}')$ just based on a different set of b columns of $(h_{5,g}^{-1/2} h_{4,\beta})'$.

$$
\begin{aligned}
&= C_{n,q^S} + o_p(1) \\
&\to_p \overline{\Delta}^S_{h,q^S} := h_{3,q^S} \in R^{(k-b)\times q^S},
\end{aligned} \tag{31.61}
$$

where the second equality uses $n^{1/2}(\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n'\Omega_n^{-1/2}D_n) = O_p(1)$ and $n^{1/2}\tau_{jn} \to \infty$ for all $j \le q^S$ (by the definition of q^S in (31.16)). The convergence in (31.61) holds by (31.14) and (31.15), and the last equality in (31.61) holds by definition. To see that the $O_p(1)$ result holds, we write

$$
\begin{aligned}
&n^{1/2}\big(\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n'\Omega_n^{-1/2}E_nG_i\big) \\
&\quad = \widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) n^{1/2}\big(\widehat{D}_n(\widehat{\beta}_n) - E_nG_i\big) + \widetilde{O}_n' n^{1/2}\big(\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) - \Omega_n^{-1/2}\big)E_nG_i \\
&\qquad + n^{1/2}\big(\widetilde{O}_n' - O_n'\big)\Omega_n^{-1/2}E_nG_i.
\end{aligned} \tag{31.62}
$$

The $O_p(1)$ result then holds by (31.52), (31.56), (31.59), $O_n = O(1)$, $\Omega_n^{-1/2} = O(1)$, and $D_n = O(1)$.

Next, consider with the second component on the rhs of (31.60). As in (25.6) and (25.7), we have

$$
n^{1/2}O_n'\Omega_n^{-1/2}D_nU_nB_{n,p-q^S} \to h_3h^{\diamond}_{1,p-q^S}. \tag{31.63}
$$

By (31.52), (31.56), and (31.59), we have

$$
\begin{aligned}
&n^{1/2}\big(\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n'\Omega_n^{-1/2}D_n\big) \\
&\quad = n^{1/2}(\widetilde{O}_n - O_n)' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) + O_n' n^{1/2}\big(\widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n) - \Omega_n^{-1/2}\big)\widehat{D}_n(\widehat{\beta}_n) \\
&\qquad + O_n'\Omega_n^{-1/2} n^{1/2}\big(\widehat{D}_n(\widehat{\beta}_n) - D_n\big) \\
&\quad \to_d \chi_h \\
&\quad := \mathrm{vec}^{-1}_{k,k-b}\big(\overline{B}_h \,\mathrm{vec}(\overline{\omega}^S_h)\big)' h_{5,g}^{-1/2} h_4 \\
&\qquad + O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' \mathrm{vec}^{-1}_{k,k}\big(\overline{\varphi}_h \,\mathrm{vec}(\overline{\varkappa}^S_h)\big) h_4 + O\big(h_{5,g}^{-1/2}h_{4,\beta}\big)' h_{5,g}^{-1/2}\overline{D}_h.
\end{aligned} \tag{31.64}
$$

Using (31.63) and (31.64), we obtain

$$
\begin{aligned}
&n^{1/2}\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n)U_nB_{n,p-q^S} \\
&\quad = n^{1/2}O_n'\Omega_n^{-1/2}D_nU_nB_{n,p-q^S} \\
&\qquad + n^{1/2}\big(\widetilde{O}_n' \widetilde{\Omega}_n^{-1/2}(\widehat{\beta}_n)\widehat{D}_n(\widehat{\beta}_n) - O_n'\Omega_n^{-1/2}D_n\big)U_nB_{n,p-q^S} \\
&\quad \to_d \overline{\Delta}^S_{h,p-q^S} := h_3h^{\diamond}_{1,p-q^S} + \chi_h h_{81} h_{2,p-q^S} \in R^{(k-b)\times(p-q^S)},
\end{aligned} \tag{31.65}
$$

where $B_{n,p-q^S} \to h_{2,p-q^S}$, $U_{2n} \to h_8$, and $U_n = U_1(U_{2n}) \to U_1(h_8) =: h_{81}$, using the definitions in (16.4), (16.5), and (16.24). Combining (31.60), (31.61), and (31.65) gives the desired asymptotic result because $\overline{\Delta}^S_h := (\overline{\Delta}^S_{h,q^S}, \overline{\Delta}^S_{h,p-q^S})$ by (31.22).

We have $(\overline{\beta}_h, \overline{D}_h^S, \overline{\varkappa}_h^S, \overline{\varrho}_h^S, \overline{\Delta}_h^S)$ is independent of $\overline{g}_h^S$ because $\overline{\Delta}_h^S$ is a nonrandom function of h and $(\overline{D}_h^S, \overline{\varkappa}_h^S, \overline{\varrho}_h^S)$, see (31.22), and $(\overline{\beta}_h, \overline{D}_h^S, \overline{\varkappa}_h^S, \overline{\varrho}_h^S)$ is independent of $\overline{g}_h^S$ by Lemma 31.5(c). □

Proof of Lemma 31.5(e). The proofs of parts (a)–(d) of the lemma go through when n is replaced by w_n. □

31.3 *Proof of Theorem 13.1*

The proof of Theorem 13.1 is a combination of the following lemma and the correct asymptotic size results for the subvector AR and CQLR tests given in Theorem 9.1.

In the following lemma, θ_{0n} is the true value that may vary with n. For notational simplicity, we suppress the dependence of various quantities on θ_{0n}.

Lemma 31.10. *Suppose Assumption* gB *holds. Then, for any sequence* $\{(F_n, \beta_n^*, \theta_{0n}) \in \mathcal{F}_{\Theta,\mathrm{AR}}^{S,\mathrm{SR}} : n \geq 1\}$, (a) $\widehat{r}_n(\widetilde{\beta}_n) = r_{F_n}(\widetilde{\beta}_n) = r_{F_n}(\beta_n^*)$ wp $\to 1$, (b) $\mathrm{col}(\widehat{A}_n(\widetilde{\beta}_n)) = \mathrm{col}(A_{F_n}(\widetilde{\beta}_n)) = \mathrm{col}(A_{F_n}(\beta_n^*))$ wp $\to 1$, *and* (c) *given the first-stage estimator* $\widetilde{\beta}_n$, *the statistics* $\mathrm{SR\text{-}AR}_n^S(\widehat{\beta}_{\widehat{A}n})$, $\mathrm{SR\text{-}QLR}_n^S(\widehat{\beta}_{\widehat{A}n})$, $\chi^2_{\widehat{r}_n(\widetilde{\beta}_n),1-\alpha}$, $c_{\widehat{r}_n(\widetilde{\beta}_n),p}(n^{1/2}\widehat{D}^*_{\widehat{A}n}(\widehat{\beta}_{\widehat{A}n}), \widetilde{J}_{\widehat{A}n}(\widehat{\beta}_{\widehat{A}n}), 1-\alpha)$ *are invariant* wp $\to 1$ *to the replacement of* $\widehat{r}_n(\widetilde{\beta}_n)$ *and* $\widehat{A}_n(\widetilde{\beta}_n)'$ *by* $r_{F_n}(\beta_n^*)$ *and* $\Pi_{1F_n}^{-1/2}(\beta_n^*) \times A_{F_n}(\beta_n^*)'$, *respectively.*

Proof of Lemma 31.10. First, we establish part (a). For any $\overline{\beta} \in B(\beta_n^*, \varepsilon)$,

$$
\begin{aligned}
\lambda \in N\big(\Omega_{F_n}(\overline{\beta})\big) \quad &\Longrightarrow \quad \lambda \in \bigcap_{\beta \in B(\beta_n^*, \varepsilon)} N\big(\Omega_{F_n}(\beta)\big) \\
&\Longrightarrow \quad \sup_{\beta \in B(\beta_n^*, \varepsilon)} \lambda' \Omega_{F_n}(\beta)\lambda = 0 \\
&\Longrightarrow \quad \sup_{\beta \in B(\beta_n^*, \varepsilon)} \mathrm{Var}_{F_n}\big(\lambda' g_i(\beta)\big) = 0 \\
&\Longrightarrow \quad \sup_{\beta \in B(\beta_n^*, \varepsilon)} \big|\lambda' g_i(\beta) - E_{F_n}\lambda' g_i(\beta)\big| = 0 \quad \text{a.s. } [F_n] \\
&\Longrightarrow \quad \sup_{\beta \in B(\beta_n^*, \varepsilon)} \lambda' \widehat{\Omega}_n(\beta)\lambda = 0 \quad \text{a.s. } [F_n] \\
&\Longrightarrow \quad \forall \beta \in B\big(\beta_n^*, \varepsilon\big), \widehat{\Omega}_n(\beta)\lambda = 0 \text{ a.s. } [F_n] \\
&\Longrightarrow \quad \lambda \in \bigcap_{\beta \in B(\beta_n^*, \varepsilon)} N\big(\widehat{\Omega}_n(\beta)\big) \text{ a.s. } [F_n],
\end{aligned}
\tag{31.66}
$$

where the first implication holds by condition (iv) of $\mathcal{F}_{\mathrm{AR},2}^{S,\mathrm{SR}}$. From the proof of Lemma 31.5, under sequences $\{(F_n, \beta_n^*, \theta_{0n}) \in \mathcal{F}_{\Theta,\mathrm{AR}}^{S,\mathrm{SR}} : n \geq 1\}$, we have that $\widetilde{\beta}_n - \beta_n^* \to_p 0^b$. Thus, wp $\to 1$ it follows that $\widetilde{\beta}_n \in B(\beta_n^*, \varepsilon)$. Thus, from (31.66), $N(\Omega_{F_n}(\widetilde{\beta}_n)) \subset N(\widehat{\Omega}_n(\widetilde{\beta}_n))$ wp $\to 1$ and $\widehat{r}_n(\widetilde{\beta}_n) \leq r_{F_n}(\widetilde{\beta}_n)$ wp $\to 1$.

Next, we prove $\widehat{r}_n(\widetilde{\beta}_n) \geq r_{F_n}(\widetilde{\beta}_n)$ wp $\to 1$. By considering subsequences, it suffices to consider the case where $r_{F_n}(\beta_n^*) = r$ for all $n \geq 1$ for some $r \in \{0, 1, \ldots, k\}$. We have

$$\widehat{r}_n(\widetilde{\beta}_n) = \mathrm{rk}\big(\widehat{\Omega}_n(\widetilde{\beta}_n)\big) \geq \mathrm{rk}\big(\Pi_{1F_n}^{-1/2}(\widetilde{\beta}_n) A_{F_n}(\widetilde{\beta}_n)' \widehat{\Omega}_n(\widetilde{\beta}_n) A_{F_n}(\widetilde{\beta}_n) \Pi_{1F_n}^{-1/2}(\widetilde{\beta}_n)\big) \tag{31.67}$$

because $\widehat{\Omega}_n(\widetilde{\beta}_n)$ is $k \times k$, the matrix $A_{F_n}(\widetilde{\beta}_n) \Pi_{1F_n}^{-1/2}(\widetilde{\beta}_n)$ is $k \times r$ wp $\to 1$ by condition (iv) of $\mathcal{F}_{\mathrm{AR},2}^{S,\mathrm{SR}}$ and consistency of $\widetilde{\beta}_n$, and wlog $1 \leq r \leq k$. (If $r = 0$, then the desired inequality $\widehat{r}_n(\widetilde{\beta}_n) \geq 0 = r_{F_n}(\widetilde{\beta}_n)$ holds trivially wp $\to 1$, where the equality holds by condition (iv) of $\mathcal{F}_{\mathrm{AR},2}^{S,\mathrm{SR}}$ and consistency of $\widetilde{\beta}_n$).) From condition (iv) of $\mathcal{F}_{\mathrm{AR},2}^{S,\mathrm{SR}}$, it follows that $A_{F_n}(\beta) = A_{F_n}(\beta_n^*)$ and therefore $A_{F_n}(\beta)$ does not depend on β for all $\beta \in B(\beta_n^*, \varepsilon)$. For $\beta \in B(\beta_n^*, \varepsilon)$, we therefore write A_{F_n} for $A_{F_n}(\beta)$ to simplify notation. Furthermore,

$$\begin{aligned}
&\Pi_{1F_n}^{-1/2}(\beta) A_{F_n}' \widehat{\Omega}_n(\beta) A_{F_n} \Pi_{1F_n}^{-1/2}(\beta) \\
&= n^{-1} \sum_{i=1}^n \Pi_{1F_n}^{-1/2}(\beta) A_{F_n}' \big(g_i(\beta) - E_{F_n} g_i(\beta)\big)\big(g_i(\beta) - E_{F_n} g_i(\beta)\big)' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta) \\
&\quad - \left[n^{-1} \sum_{i=1}^n \Pi_{1F_n}^{-1/2}(\beta) A_{F_n}' \big(g_i(\beta) - E_{F_n} g_i(\beta)\big) \right] \\
&\quad \times \left[n^{-1} \sum_{i=1}^n \big(g_i(\beta) - E_{F_n} g_i(\beta)\big)' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta) \right].
\end{aligned} \tag{31.68}$$

By construction and using condition (iv) of $\mathcal{F}_{\mathrm{AR},2}^{S,\mathrm{SR}}$, we have, for all $\beta \in B(\beta_n^*, \varepsilon)$,

$$E_{F_n} \Pi_{1F_n}^{-1/2}(\beta) A_{F_n}' \big(g_i(\beta) - E_{F_n} g_i(\beta)\big)\big(g_i(\beta) - E_{F_n} g_i(\beta)\big)' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta) = I_r.$$

By the uniform moment bound in $\mathcal{F}_{\Theta,\mathrm{AR}}^{S,\mathrm{SR}}$, namely, $E_F \sup_{\beta \in B(\beta^*, \varepsilon)} \|\Pi_{1F}^{-1/2}(\beta) A_F(\beta)' \times (g_i(\beta) - E_F g_i(\beta))\|^2 \leq M$ and continuity of $(g_i(\beta) - E_{F_n} g_i(\beta))' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta)$ as a function of β (which holds by Elsner's theorem and Assumption gB), it follows from a uniform weak law of large numbers for $L^{1+\gamma/2}$-bounded i.i.d. random variables for $\gamma > 0$ that the expressions in the second and third lines of (31.68) converge in probability to I_r and $0^{r \times r}$, respectively, uniformly over $\beta \in B(\beta^*, \varepsilon)$. This implies that

$$\Pi_{1F_n}^{-1/2}(\widetilde{\beta}_n) A_{F_n}(\widetilde{\beta}_n)' \widehat{\Omega}_n(\widetilde{\beta}_n) A_{F_n}(\widetilde{\beta}_n) \Pi_{1F_n}^{-1/2}(\widetilde{\beta}_n) \to_p I_r.$$

This establishes that $\widehat{r}_n(\widetilde{\beta}_n) \geq r$ wp $\to 1$ and, therefore, $\widehat{r}_n(\widetilde{\beta}_n) = r$ and $N(\Omega_{F_n}(\widetilde{\beta}_n)) = N(\widehat{\Omega}_n(\widetilde{\beta}_n))$ wp $\to 1$, which proves (a). In turn, the latter implies that $\mathrm{col}(A_{F_n}(\widetilde{\beta}_n)) = \mathrm{col}(\widehat{A}_n(\widetilde{\beta}_n))$ wp $\to 1$, which also proves part (b).

To prove part (c), it suffices to consider the case where $r \geq 1$ because the test statistics and their critical values are all equal to zero by definition when $\widehat{r}_n(\widetilde{\beta}_n) = 0$ and $\widehat{r}_n(\widetilde{\beta}_n) = 0$ wp $\to 1$ when $r = 0$ by part (a). Part (b) of the lemma implies that there exists a random $r \times r$ nonsingular matrix $\widehat{M}_n$ such that

$$\widehat{A}_n(\widetilde{\beta}_n) = A_{F_n}(\beta_n^*) \Pi_{1F_n}^{-1/2}(\beta_n^*) \widehat{M}_n \quad \text{wp} \to 1, \tag{31.69}$$

because $\varPi_{1F_n}^{-1/2}(\beta_n^*)$ is nonsingular (since by its definition it is a diagonal matrix with the positive eigenvalues of $\varOmega_{F_n}(\beta_n^*)$ on its diagonal.) Equation (31.69) and $\widehat{r}_n(\widetilde{\beta}_n) = r$ wp $\to$ 1 imply that the statistics SR-AR$_n^S(\widehat{\beta}_{\widehat{A}n})$, SR-QLR$_n^S(\widehat{\beta}_{\widehat{A}n})$, $\chi^2_{\widehat{r}_n(\widetilde{\beta}_n),1-\alpha}$, $c_{\widehat{r}_n(\widetilde{\beta}_n),p}(n^{1/2} \times \widehat{D}^*_{\widehat{A}n}(\widehat{\beta}_{\widehat{A}n}), \widetilde{J}_{\widehat{A}n}(\widehat{\beta}_{\widehat{A}n}), 1-\alpha)$, are invariant wp $\to$ 1 to the replacement of $\widehat{r}_n(\widetilde{\beta}_n)$ and $\widehat{A}_n(\widetilde{\beta}_n)'$ by r and $A_{F_n}(\beta_n^*)\varPi_{1F_n}^{-1/2}(\beta_n^*)\widehat{M}_n$, respectively. Now we apply the invariance results of Lemma 31.2 with $(k, g_i(\beta), G_i(\beta))$ replaced by $(r, \varPi_{1F_n}^{-1/2}(\beta_n^*)A_{F_n}(\beta_n^*)'g_i(\beta),$ $\varPi_{1F_n}^{-1/2}(\beta_n^*)A_{F_n}(\beta_n^*)'G_i(\beta))$ and with M equal to $\widehat{M}_n'$. These results imply that the previous four statistics when based on r and $\varPi_{1F_n}^{-1/2}(\beta_n^*)A_{F_n}(\beta_n^*)'g_i(\beta)$ are invariant to the multiplication of the moments $\varPi_{1F_n}^{-1/2}(\beta_n^*)A_{F_n}(\beta_n^*)'g_i(\beta)$ by the nonsingular matrix $\widehat{M}_n'$. Thus, the statistics, defined as in Section 5.2, are invariant wp $\to$ 1 to the replacement of $\widehat{r}_n(\widetilde{\beta}_n)$ and $\widehat{A}_n'(\widetilde{\beta}_n)$ by r and $\varPi_{1F_n}^{-1/2}(\beta_n^*)A_{F_n}(\beta_n^*)'$, respectively, which proves part (c). □

31.4 *Proof of Theorem 9.2*

Proof of Theorem 9.2. By Lemma 31.10(a) and (b) and $\mathcal{F}^S \subset \mathcal{F}_{\mathrm{AR}}^{S,\mathrm{SR}}$ (because $\mathcal{F}^S$ imposes $\lambda_{\min}(E_F g_i g_i') \geq \delta$, where $\mathcal{F}_{\mathrm{AR}}^{S,\mathrm{SR}}$ is defined in (52)), we have $\widehat{r}_n(\widetilde{\beta}_n) = r_{F_n}(\widetilde{\beta}_n) = r_{F_n}(\beta_n^*)$ and $\mathrm{col}(\widehat{A}_n(\widetilde{\beta}_n)) = \mathrm{col}(A_{F_n}(\widetilde{\beta}_n)) = \mathrm{col}(A_{F_n}(\beta_n^*))$ wp $\to$ 1. Also, given $\lambda_{\min}(E_{F_n} g_i g_i') \geq \delta$, it follows that the orthogonal matrix $A_{F_n}(\beta_n^*)$ is in $R^{k\times k}$. Given that the statistics QLR$_n^S(\widetilde{\beta}_n)$ and $c_{k,p}(n^{1/2}\widehat{D}_n^*(\widetilde{\beta}_n), \widetilde{J}_n(\widetilde{\beta}_n), 1-\alpha)$ are invariant to nonsingular transformations by Lemma 31.2, the definition of the subvector SR test in (13.2), combined with the previous two statements imply that SR-QLR$_n^S(\theta_0, \widehat{\beta}_{\widehat{A}n}) =$ QLR$_n^S(\widehat{\eta}) + o_p(1)$. Because $\widehat{r}_n(\widetilde{\beta}_n) = k$ wp $\to$ 1, it follows that $c_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),p}(n^{1/2}\widehat{D}^*_{\widehat{A}n}(\theta_0, \widehat{\beta}_{\widehat{A}n}),$ $\widetilde{J}_{\widehat{A}n}(\theta_0, \widehat{\beta}_{\widehat{A}n}), 1-\alpha) = c_{k,p}(n^{1/2}\widehat{D}_n^*(\widehat{\eta}), \widetilde{J}_n(\widehat{\eta}), 1-\alpha)$ wp $\to$ 1, where the latter critical value is the one for the subvector CQLR test without singularity robustness; see (48). This proves the first equalities in parts (a) and (b).

We now proceed as in the proof of Theorem 7.1. We replace $\widehat{W}_n$, W_{F_n}, $\widehat{\varOmega}_n^{-1/2}\widehat{g}_n$, and $\widehat{D}_n \in R^{k\times p}$ by the corresponding quantities I_k, I_k, $\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta})$, and $\widetilde{O}_{F_n}'\widetilde{\varOmega}_n(\widehat{\eta})^{-1/2}\widehat{D}_n(\widehat{\eta}) \in R^{(k-b)\times p}$, respectively. Note that $q^S = p$ under $\{\lambda_{n,h}^S : n \geq 1\}$. The analogue to (28.2) with $q^S = p$ therefore states that

$$\widehat{\kappa}^+_{(p+1)n} = n\widehat{g}_n'(\widehat{\eta})\widehat{\varOmega}_n^{-1/2}\widetilde{O}_{F_n}h_{3,k-p}h_{3,k-p}'\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta}) + o_p(1). \tag{31.70}$$

In addition, the analogue to (28.3) with $q^S = p$ states that

$$\mathrm{QLR}^S_{\mathrm{WU},n} = n\widehat{g}_n'(\widehat{\eta})\widehat{\varOmega}_n^{-1/2}\widetilde{O}_{F_n}h_{3,p}h_{3,p}'\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta}) + o_p(1), \tag{31.71}$$

where $\mathrm{QLR}^S_{\mathrm{WU},n} := \mathrm{AR}_n^S(\widehat{\eta}) - \lambda_{\min}(n\widehat{Q}^S_{\mathrm{WU},n})$ is defined below Proposition 31.6 and equals QLR$_n^S(\widehat{\eta})$ when $\widehat{U}_n$ is taken to be $\widetilde{L}_n^{1/2}(\widehat{\eta})$. Equation (31.61) implies that $h_{3,p} = \widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})U_nB_{n,p}Y_{n,p}^{-1} + o_p(1)$. Because $U_nB_{n,p}Y_{n,p}^{-1}$ is an invertible matrix, it follows that $P_{h_{3,p}} = P_{\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})} + o_p(1)$. Therefore, using $h_{3,p}'h_{3,p} = I_p$, it follows that

$$\mathrm{QLR}^S_{\mathrm{WU},n} = n\widehat{g}_n'(\widehat{\eta})\widehat{\varOmega}_n^{-1/2}\widetilde{O}_{F_n}P_{\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta})}\widetilde{O}_{F_n}'\widetilde{\varOmega}_n^{-1/2}(\widehat{\eta})\widehat{g}_n(\widehat{\eta}) + o_p(1). \tag{31.72}$$

By (31.48), $\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})n^{1/2}\widehat{g}_n(\widehat{\eta}) = M_{J_h}h_{5,g}^{-1/2}n^{1/2}\widehat{g}_n(\theta_0, \beta_n^*) + o_p(1)$, where $J_h = h_{5,g}^{-1/2}h_{4,\beta}$. Also, $\widetilde{O}_{F_n} = O(J_h) + o_p(1)$ and $\widetilde{\Omega}_n^{-1/2}(\widehat{\eta})\widehat{D}_n(\widehat{\eta}) = J_{\theta h} + o_p(1)$, where $J_{\theta h} := h_{5,g}^{-1/2}h_4$ and $J_{\theta h}$ has full column rank p. Thus, we obtain

$$\begin{aligned}\mathrm{QLR}_n^S(\widehat{\eta}) &= n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big)'h_{5,g}^{-1/2}M_{J_h}O(J_h)P_{O(J_h)'J_{\theta h}}O(J_h)'M_{J_h}h_{5,g}^{-1/2}n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big) + o_p(1)\\ &= n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big)'h_{5,g}^{-1/2}P_{M_{J_h}J_{\theta h}}h_{5,g}^{-1/2}n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big) + o_p(1), \qquad (31.73)\end{aligned}$$

where the second equality uses $O(J_h)O(J_h)' = M_{J_h} = M_{J_h}'M_{J_h}$.

From the above, it also follows that

$$\mathrm{LM}_n^S = n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big)'h_{5,g}^{-1/2}(M_{J_h}P_{[J_{\theta h}:J_h]}M_{J_h})h_{5,g}^{-1/2}n^{1/2}\widehat{g}_n\big(\theta_0, \beta_n^*\big) + o_p(1) \qquad (31.74)$$

using $h_{5,g}^{-1/2}\widehat{G}_{\eta n}(\widehat{\eta}) \to_p [J_{\theta h} : J_h] := [h_{5,g}^{-1/2}h_4 : h_{5,g}^{-1/2}h_{4,\beta}]$ and $[J_{\theta h} : J_h]$ has full column rank $p + b$.

Next, we have

$$M_{J_h}P_{[J_{\theta h}:J_h]}M_{J_h} = M_{J_h}P_{[M_{J_h}J_{\theta h}:J_h]}M_{J_h} = M_{J_h}(P_{M_{J_h}J_{\theta h}} + P_{J_h})M_{J_h} = P_{M_{J_h}J_{\theta h}}, \qquad (31.75)$$

where the first equality holds because $[J_{\theta h} : J_h]$ and $[M_{J_h}J_{\theta h} : J_h]$ span the same space, the second equality holds because $M_{J_h}J_{\theta h}$ and J_h are orthogonal, and the last equality holds because $P_{J_h}M_{J_h} = 0^{k\times k}$ and $P_{M_{J_h}J_{\theta h}}M_{J_h} = P_{M_{J_h}J_{\theta h}}$. Equations (31.73)–(31.75) combine to show that $\mathrm{QLR}_n^S(\widehat{\eta}) = \mathrm{LM}_n^S + o_p(1)$, which establishes the second equality of part (a).

By (31.31), $c_{k,p}(n^{1/2}\widehat{D}_n^*(\widehat{\eta}), \widetilde{J}_n(\widehat{\eta}), 1-\alpha) + o_p(1) \to_p c_{k-b,p,q^S}(h_{3,k-b-q^S}'\overline{\Delta}_{h,p-q^S}^S, 1-\alpha)$, where $c_{k-b,p,q^S}(\cdot, 1-\alpha)$ is defined in (31.19) (and uses the notation in (27.12)). In the present case, $q^S = p$, which implies that $\overline{\Delta}_{h,p-q^S}^S$ has no columns, $\mathrm{ACLR}_{k,p,q}(\tau_\infty^c) = Z_1'Z_1 \sim \chi_p^2$, and $c_{k,p,q}(h_{3,k-b-q^S}' \times \overline{\Delta}_{h,p-q^S}^S, 1-\alpha)$ equals the $1-\alpha$ quantile of the χ_p^2 distribution. Hence, the convergence result in part (b) holds. □

References

Ahn, S. C. and P. Schmidt (1995), "Efficient estimation of models for dynamic panel data." *Journal of Econometrics*, 68, 5–27. [10]

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59, 817–858. [35]

Andrews, D. W. K. (2017), "Identification-robust subvector inference." Cowles Foundation Discussion Paper No. 3003, 2017. [4]

Andrews, D. W. K. and X. Cheng (2012), "Estimation and inference with weak, semi-strong, and strong identification." *Econometrica*, 80, 2153–2211. Supplemental Material is available at *Econometrica Supplemental Material*, 80, http://www.econometricsociety.org/ecta/Supmat/9456_miscellaneous.pdf. [3, 4]

Andrews, D. W. K. and X. Cheng (2013a), "GMM estimation and uniform subvector inference with possible identification failure." *Econometric Theory*, 30, 1–47. [4]

Andrews, D. W. K. and X. Cheng (2013b), "Maximum likelihood estimation and uniform inference with sporadic identification failure." *Journal of Econometrics*, 173, 36–56. Supplemental Material is available with Cowles Foundation Discussion Paper No. 1824R, 2011, Yale University. [3, 4]

Andrews, D. W. K., X. Cheng, and P. Guggenberger (forthcoming), "Generic results for establishing the asymptotic size of confidence sets and tests." *Journal of Econometrics*. [18]

Andrews, D. W. K. and P. Guggenberger (2014), "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. [2]

Andrews, D. W. K. and P. Guggenberger (2017), "Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models." *Econometric Theory*, 33, 1046–1080. [2]

Andrews, D. W. K., V. Marmer, and Z. Yu (2019), "On optimal inference in the linear IV model." *Quantitative Eonomics*, 10, 457–485. [37, 40]

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. [20, 37, 40, 62]

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008), "Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments." *Journal of Econometrics*, 146, 241–254. [20, 37, 40]

Andrews, D. W. K. and W. Ploberger (1994), "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62, 1383–1414. [3]

Andrews, I. (2018), "Valid two-step identification-robust confidence sets for GMM." *Review of Economics and Statistics*, 100, 337–348. [4]

Andrews, I. and A. Mikusheva (2015), "Maximum likelihood inference in weakly identified models." *Quantitative Economics*, 6, 123–152. [3, 4]

Andrews, I. and A. Mikusheva (2016), "A geometric approach to nonlinear econometric models." *Econometrica*, 84, 1249–1264. [3, 4]

Angrist, J. D. and A. B. Krueger (1991), "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106, 979–1014. [3]

Angrist, J. D. and A. B. Krueger (1992), "Estimating the payoff to schooling using the Vietnam-era draft lottery." NBER Working Paper No. 4067. [3]

Antoine, B. and E. Renault (2009), "Efficient GMM with nearly weak instruments." *Econometrics Journal*, 12, S135–S171. [4]

Antoine, B. and E. Renault (2010), "Efficient inference with poor instruments, a general framework." In *Handbook of Empirical Economics and Finance* (D. Giles and A. Ullah, eds.). Taylor and Francis, Oxford. [4]

Arellano, M. and O. Bover (1995), "Another look at instrumental variable estimation of error component models." *Journal of Econometrics*, 68, 29–51. [10]

Armstrong, T. B. (2016), "Large market asymptotics for differentiated product demand estimators with economic models of supply." *Econometrica*, 84, 1961–1980. [3]

Armstrong, T. B., H. Hong, and D. Nekipelov (2012), "How strong must identification be for conventional asymptotics in nonlinear models?" Unpublished manuscript, Cowles Foundation, Yale University. [5]

Billingsley, P. (1979), *Probability and Measure*. John Wiley and Sons, New York, NY. [18]

Blundell, R. and S. Bond (1995), *Initial Conditions and Moment Restrictions in Dynamic Panel Data Models*. The Institute for Fiscal Studies, London. Working Paper No. W95/17. [10]

Canova, F. and L. Sala (2009), "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics*, 56, 431–449. [3]

Carroll, C. D., J. Slacalek, and M. Sommer (2011), "International evidence on sticky consumption growth." *Review of Economics and Statistics*, 93, 1135–1145. [3]

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995), "Inference in models with nearly integrated regressors." *Econometric Theory*, 11, 1131–1147. [4]

Chaudhuri, S., T. Richardson, J. Robins, and E. Zivot (2010), "A new projection-type split-sample score test in linear instrumental variables regression." *Econometric Theory*, 26, 1820–1837. [4]

Chaudhuri, S. and E. Zivot (2011), "A new method of projection-based inference in GMM with weakly identified nuisance parameters." *Journal of Econometrics*, 164, 239–251. [4]

Chen, X., M. Ponomareva, and E. Tamer (2014), "Likelihood inference in some finite mixture models." *Journal of Econometrics*, 182, 87–99. [3]

Cheng, X. (2015), "Uniform inference in nonlinear models with mixed identification strength." *Journal of Econometrics*, 189, 207–228. [4]

Chernozhukov, V., C. Hansen, and M. Jansson (2009), "Admissible invariant similar tests for instrumental variables regression." *Econometric Theory*, 25, 806–818. [37, 40]

Cho, J. S. and H. White (2007), "Testing for regime switching." *Econometrica*, 75, 1671–1720. [3]

Choi, I. and P. C. B. Phillips (1992), "Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations." *Journal of Econometrics*, 51, 113–150. [4]

Cox, G. (2017), "Weak identification in a class of generically identified models with an application to factor models." Unpublished working paper, Department of Economics, Columbia University. [4]

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. [45, 47]

Cragg, J. C. and S. G. Donald (1997), "Inferring the rank of a matrix." *Journal of Econometrics*, 76, 223–250. [45, 47]

Cruz, L. M. and M. J. Moreira (2005), "On the validity of econometric techniques with weak instruments: Inference on returns to education using compulsory school attendance laws." *Journal of Human Resources*, 40, 393–410. [3]

Davies, R. B. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative." *Biometrika*, 64, 247–254. [3]

Dufour, J.-M. (1989), "Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions." *Econometrica*, 57, 335–355. [4]

Dufour, J.-M. and J. Jasiak (2001), "Finite sample limited information inference methods for structural equations and structural models with generated regressors." *International Economic Review*, 42, 815–843. [4]

Dufour, J.-M., L. Khalaf, and M. Kichian (2006), "Inflation dynamics and the new Keynesian Phillips curve: An identification robust econometric analysis." *Journal of Economic Dynamics and Control*, 30, 1707–1727. [3]

Dufour, J.-M., L. Khalaf, and M. Kichian (2013), "Identification-robust analysis of DSGE and structural macroeconomic models." *Journal of Monetary Economics*, 60, 340–350. [3]

Eicker, F. (1963), "Central limit theorems for families of sequences of random variables." *Annals of Mathematical Statistics*, 34, 439–446. [101]

Gomes, F. A. R. and L. S. Paz (2013), "Estimating the elasticity of intertemporal substitution: Is the aggregate financial return free from the weak instrument problem." *Journal of Macroeconomics*, 36, 63–75. [3]

Grant, N. (2013), "Identification robust inference with singular variance." Economics Discussion Paper No. 1315, Department of Economics, University of Manchester. [3, 11]

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013), "Frequentist inference in weakly identified dynamic stochastic general equilibrium models." *Quantitative Economics*, 4, 197–229. [3]

Guggenberger, P. (2012), "On the asymptotic size distortion of tests when instruments locally violate the exogeneity condition." *Econometric Theory*, 28, 387–421. [4]

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012), "On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression." *Econometrica*, 80, 2649–2666. [4]

Guggenberger, P., J. J. S. Ramalho, and R. J. Smith (2012), "GEL statistics under weak identification." *Journal of Econometrics*, 170, 331–349. [38, 44, 45, 46, 47]

Hannan, E. J. (1982), "Testing for autocorrelation and Akaike's criterion." In *Essays in Statistical Science* (J. Gani and E. J. Hannan, eds.). *Journal of Applied Probability*, Vol. Special Volume 19A. [3]

Hansen, L. P. and J. A. Scheinkman (1995), "Back to the future: Generating moment implications for continuous-time Markov processes." *Econometrica*, 63, 767–804. [10]

Hillier, G. (2009), "Exact properties of the conditional likelihood ratio test in an IV regression model." *Econometric Theory*, 25, 915–957. [4]

Iskrev, N. (2010), "Local identification in DSGE." *Journal of Monetary Economics*, 57, 189–202. [3]

Jegannathan, R., G. Skoulakis, and Z. Wang (2002), "Generalized method of moments: Applications in finance." *Journal of Business and Economic Statistics*, 20, 470–481. [3, 11]

Johansen, S. (1991), "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models." *Econometrica*, 59, 1551–1580. [61]

Kleibergen, F. (2004), "Testing subsets of structural parameters in the instrumental variables regression model." *Review of Economics and Statistics*, 86, 418–423. [4]

Kleibergen, F. (2005), "Testing parameters in GMM without assuming that they are identified." *Econometrica*, 73, 1103–1123. [3, 16, 27, 38, 44, 45, 47, 49]

Kleibergen, F. (2007), "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics." *Journal of Econometrics*, 139, 181–216. [3, 38, 44, 45]

Kleibergen, F. and S. Mavroeidis (2009), "Weak instrument robust tests and the new Keynesian Phillips curve." *Journal of Business and Economic Statistics*, 27, 293–311. [3]

Kleibergen, F. and R. Paap (2006), "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133, 97–126. [45, 47]

McCloskey, A. (2017), "Bonferroni-based size-correction for nonstandard testing problems." *Journal of Econometrics*, 200, 17–35. [4]

Mikusheva, A. (2010), "Robust confidence sets in the presence of weak instruments." *Journal of Econometrics*, 157, 236–247. [3]

Montiel Olea, J. L. (forthcoming), "Efficient conditionally similar-on-the-boundary tests." *Econometric Theory*. [4]

Moreira, M. J. (2003), "A conditional likelihood ratio test for structural models." *Econometrica*, 71, 1027–1048. [3, 10, 12, 17, 20, 37, 38, 39, 41, 42, 44, 45, 46, 47, 49, 62]

Nason, J. M. and G. W. Smith (2008), "Identifying the new Keynesian Phillips curve." *Journal of Applied Econometrics*, 23, 525–551. [3]

Neely, C. J., A. Roy, and C. H. Whiteman (2001), "Risk aversion versus intertemporal substitution: A case study of identification failure in the intertemporal consumption capital asset pricing model." *Journal of Business and Economic Statistics*, 19, 395–403. [3]

Nelson, C. R. and R. Startz (2007), "The zero-information-limit condition and spurious inference in weakly identified models." *Journal of Econometrics*, 138, 47–62. [3]

Newey, W. K. and K. West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, 55, 703–708. [35]

Newey, W. K. and F. Windmeijer (2009), "Generalized method of moments with many weak moment conditions." *Econometrica*, 77, 687–719. [38, 44, 45, 46, 47]

Phillips, P. C. B. (1989), "Partially identified econometric models." *Econometric Theory*, 5, 181–240. [4]

Pollard, D. (1990), *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Institute of Mathematical Statistics, Hayward, CA. [71]

Puhani, P. A. (2000), "The Heckman correction for sample selection and its critique." *Journal of Economic Surveys*, 14, 53–68. [3]

Qu, Z. (2014), "Inference in DSGE models with possible weak identification." *Quantitative Economics*, 5, 457–494. [3]

Qu, Z. and D. Tkachenko (2012), "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models." *Quantitative Economics*, 3, 95–132. [3]

Robin, J.-M. and R. J. Smith (2000), "Tests of rank." *Econometric Theory*, 16, 151–175. [45, 47, 49, 50, 61]

Schorfheide, F. (2014), "Estimation and evaluation of DSGE models: Progress and challenges." In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Vol. III* (D. Acemoglu, M. Arellano, and E. Dekel, eds.), 184–230, Cambridge University Press, Cambridge, UK. [3]

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* (R. Blundell, W. K. Newey, and T. Persson, eds.). Cambridge University Press, Cambridge, UK. Also available as CEMMAP Working Paper No. 13/05, UCL. [38, 44, 45, 47]

Stewart, G. W. (2001), *Matrix Algorithms Volume II: Eigensystems*. SIAM, Philadelphia. [53, 64, 65, 74, 76, 77, 84]

Stock, J. H. and J. H. Wright (2000), "GMM with weak identification." *Econometrica*, 68, 1055–1096. [3, 4, 10]

Teräsvirta, T. (1994), "Specification, estimation, and evaluation of smooth transition autoregressive models." *Journal of the American Statistical Association*, 89, 208–218. [3]

Yogo, M. (2004), "Estimating the elasticity of intertemporal substitution when instruments are weak." *Review of Economics and Statistics*, 86, 797–810. [3]

Co-editor Andres Santos handled this manuscript.

Manuscript received 8 October, 2018; final version accepted 5 April, 2019; available online 23 April, 2019.