Identification- and Singularity-Robust Inference for Moment Condition Models

Donald W. K. Andrews  Patrik Guggenberger
Cowles Foundation  Department of Economics
Yale University  Pennsylvania State University

First Version: March 25, 2011
Revised: April 4, 2019*

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-1058376, SES-1355504, and SES-1656313. Guggenberger gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-1021101 and SES-1326827. The authors thank Isaiah Andrews for providing the details of the simulation designs employed in I. Andrews (2016), which are used in some of the simulations given in this paper. For helpful comments, the authors thank four referees, the co-editor, Isaiah Andrews, Xu Cheng, Kirill Evdokimov, KooHyun Kwon, Anna Mikusheva, Zhongjun Qu, Jim Stock, and the participants of seminars at the following universities: Boston, Boston College, Brown, Chicago, Columbia, Freiburg, Hanover, Harvard/MIT, Hebrew Jerusalem, Konstanz, Maryland, Michigan, New York, Northwestern, Ohio State, Princeton, Queen’s, Strasbourg, and Wisconsin.
Abstract

This paper introduces a new identification- and singularity-robust conditional quasi-likelihood ratio (SR-CQLR) test and a new identification- and singularity-robust Anderson and Rubin (1949) (SR-AR) test for linear and nonlinear moment condition models. Both tests are very fast to compute. The paper shows that the tests have correct asymptotic size and are asymptotically similar (in a uniform sense) under very weak conditions. For example, in i.i.d. scenarios, all that is required is that the moment functions and their derivatives have $2 + \gamma$ bounded moments for some $\gamma > 0$. No conditions are placed on the expected Jacobian of the moment functions, on the eigenvalues of the variance matrix of the moment functions, or on the eigenvalues of the expected outer product of the (vectorized) orthogonalized sample Jacobian of the moment functions.

The SR-CQLR test is shown to be asymptotically efficient in a GMM sense under strong and semi-strong identification (for all $k \geq p$, where $k$ and $p$ are the numbers of moment conditions and parameters, respectively). The SR-CQLR test reduces asymptotically to Moreira’s CLR test when $p = 1$ in the homoskedastic linear IV model. The same is true for $p \geq 2$ in most, but not all, identification scenarios.

We also introduce versions of the SR-CQLR and SR-AR tests for subvector hypotheses and show that they have correct asymptotic size under the assumption that the parameters not under test are strongly identified. The subvector SR-CQLR test is shown to be asymptotically efficient in a GMM sense under strong and semi-strong identification.

Keywords: asymptotics, conditional likelihood ratio test, confidence set, identification, inference, moment conditions, robust, singular variance, subvector test, test, weak identification, weak instruments.

JEL Classification Numbers: C10, C12.
1 Introduction

Weak identification and weak instruments (IV’s) can arise in a wide variety of empirical applications in economics. Examples include: in macroeconomics and finance, new Keynesian Phillips curve models, dynamic stochastic general equilibrium (DSGE) models, consumption capital asset pricing models (CCAPM), and interest rate dynamics models; in industrial organization, the Berry, Levinsohn, and Pakes (1995) (BLP) model of demand for differentiated products; and in labor economics, returns-to-schooling equations that use IV’s, such as quarter of birth or Vietnam draft lottery status, to avoid ability bias. Other examples include nonlinear regression, autoregressive-moving average, GARCH, and smooth transition autoregressive (STAR) models; parametric selection models estimated by Heckman’s two step method or maximum likelihood; mixture models and regime switching models; and all models where hypothesis testing problems arise where a nuisance parameter appears under the alternative hypothesis, but not under the null.

Given this wide range of applications and models, it is useful to have tests and confidence sets (CS’s) that are identification-robust under nearly minimal conditions. This paper introduces two tests (and CS’s) with this feature. The two new tests are a singularity-robust (SR) conditional quasi-likelihood ratio (SR-CQLR) test and an SR nonlinear Anderson and Rubin (1949) (SR-AR) test. These tests and CS’s are shown to have correct asymptotic size and to be asymptotically similar (in a uniform sense) under very weak conditions. All that is required is that the expected moment functions equal zero at the true parameter value and the moment functions and their derivatives satisfy mild moment conditions. Thus, no identification assumptions of any type are imposed. The results hold for arbitrary fixed $k, p \geq 1$, where $k$ is the number of moment conditions and $p$ is the number of parameters. The results allow for any of the $p$ parameters (or any transformations of them) to be weakly or strongly identified, which covers multiple possible sources of weak identification. Results are given for independent identically distributed (i.i.d.) observations as well as stationary strong mixing time series observations.

The asymptotic results allow the variance matrix of the moments to be near singular or singular. This is particularly important in models where weak identification (or lack of identification) is necessarily accompanied by near singularity (or exact singularity) of the variance matrix of the moments. This occurs in all maximum likelihood scenarios and many quasi-likelihood scenarios. Furthermore, in models of this type where robustness against lack of identification—not just against weak identification—is important, allowing for singularity of the variance matrix of the moments—not just near singularity—is necessary. This occurs in likelihood-based models that nest submodels of interest, when the parameters are not identified in the submodel. For examples, this occurs with

---

1 For references, see Section 12 in the Supplemental Material.
(i) factor models with multiple factors, where the submodels of interest have reduced numbers of factors, (ii) mixture models, including regime switching models, where the submodel of interest has only one regime, (iii) asset return models with jumps, where the submodel of interest has no jumps, (iv) random coefficient models with possible correlation between the coefficients, where the submodel has constant coefficients, (v) random coefficient models with possible correlation between a random coefficient and an error term, where the submodel has constant coefficients, (vi) GARCH models and ARCH and GARCH in mean models, where the submodel of interest has no conditional heteroskedasticity, and (vii) ARMA models, where the submodel has i.i.d. (or uncorrelated) observations. In all of these models, ruling out singularity of the variance matrix, rules out the submodel. Note that in these likelihood scenarios (where the moment function is the score function) the SR-AR test is the same as the nonlinear Anderson-Rubin statistic (i.e., the $S$ statistic in Stock and Wright (2000)) and the LM statistic in Andrews and Mikusheva (2015) if the model is identified, but not if it is not identified. Neither Stock and Wright (2000) nor Andrews and Mikusheva (2015) deal with the case where the model is unidentified. Some finite-sample simulation results, given in the Supplemental Material (SM) to this paper, show that the SR-AR and SR-CQLR tests perform well (in terms of null rejection probabilities) under singular and near singular variance matrices of the moments in the model considered.

The asymptotic results also allow the expected outer-product of the vectorized orthogonalized sample Jacobian to be singular. For example, this occurs when some moment conditions do not depend on some parameters. Finally, the asymptotic results allow the true parameter to be on, or near, the boundary of the parameter space.

In sum, the conditions for correct asymptotic size of these tests and CS’s are sufficiently weak and transparent that the practitioner is easily assured of avoiding asymptotic size distortions.

The SR-CQLR test is shown to be 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). Furthermore, it reduces to Moreira’s (2003) CLR test in the homoskedastic linear IV model with fixed IV’s when $p = 1$. This is desirable because the latter test has been shown to have approximate optimal power properties in this model under normality, see Andrews, Moreira, and Stock (2006, 2008), Chernozhukov, Hansen, and Jansson (2009), Mikusheva (2010), and Andrews, Marmer, and Yu (2019). A drawback of the SR-CQLR test is that it is not known to have optimality properties under weak identification in other models. The SR-CQLR test is easy to compute and its conditional critical value can be simulated easily and very quickly.

We recommend the use of the SR-CQLR test over the SR-AR test in over-identified moment
condition models based on power advantages. In exactly-identified models, the SR-CQLR and SR-AR tests are asymptotically equivalent and we recommend the use of the SR-AR test because its critical value is not simulated, whereas that of the SR-CQLR test is simulated.

To establish the asymptotic size and similarity results of the paper, we use the approach in Andrews, Cheng, and Guggenberger (2019) and Andrews and Guggenberger (2010). With this approach, one needs to determine the asymptotic null rejection probabilities of the tests under various drifting sequences of distributions \( \{F_n : n \geq 1\} \). Different sequences can yield different strengths of identification of the unknown parameter \( \theta \). The strength of identification of \( \theta \) depends on the expected Jacobian of the moment functions evaluated at the true parameter, which is a \( k \times p \) matrix. When \( k < p \), the parameter \( \theta \) is unidentified. When \( k \geq p \), the magnitudes of the \( p \) singular values of this matrix determine the strength of identification of \( \theta \). The SR-CQLR statistic has a \( \chi^2_p \) asymptotic null distribution under strong and semi-strong identification and a noticeably more complicated asymptotic null distribution under weak identification.

To obtain the robustness of the two new tests to exact singularity of the variance matrix of the moments, we use the rank of the sample variance matrix of the moments to estimate the rank of the population variance matrix. We use a spectral decomposition of the sample variance matrix to estimate the linear combinations of the moments that are stochastic. We construct the test statistics using these estimated stochastic linear combinations of the moments. When the sample variance matrix is singular, we employ an extra rejection condition that improves power by fully exploiting the nonstochastic part of the moment conditions associated with the singular part of the variance matrix. We show that the resulting tests and CS’s have correct asymptotic size. In contrast, arbitrarily discarding moment conditions when the sample variance matrix is singular can affect the outcome of the test and the power of the test depending on which moment conditions are deleted, see Section 15.2 in the SM for an illustration. In addition, it ignores the information in the extra rejection condition referred to above. The robustness of the SR-CQLR test to any form of the expected outer product matrix of the vectorized orthogonalized Jacobian occurs because the SR-CQLR test statistic does not depend on Kleibergen’s (2005) LM statistic, but rather, on a minimum eigenvalue statistic.

The SR-CQLR and SR-AR tests are for full vector inference. We develop subvector inference for scenarios in which the nuisance parameters under the null hypothesis are strongly identified. We show that the SR-CQLR subvector test is asymptotically efficient under strong and semi-strong identification. We compare the power of the subvector SR-CQLR and SR-AR tests with the power of the \( S \) test in Stock and Wright (2000) and the CLR test in I. Andrews and Mikusheva (2016), which we refer to as the AM test. The model considered is an endogenous probit model with a
six- or eight-dimensional nuisance parameter and a scalar parameter of interest. The SR-CQLR and AM tests out perform the SR-AR and S tests in the scenarios considered. The SR-CQLR and AM tests have crisscrossing power functions, which makes a ranking difficult. It takes about 4 minutes to calculate 5,000 CQLR tests using an Intel Core 3.4GHz, 6MB processor, which is about 59 times faster than for the AM test. The speed difference should be increasing rapidly in the dimension, $p$, of the parameter specified by the null hypothesis because the AM test requires an optimization over a $p$ dimensional space for each simulation used to compute its conditional critical value, whereas the CQLR test has a closed-form expression. See Section 12 in the SM for references to other subvector inference methods in the literature.

We carry out some asymptotic power comparisons of the full-vector versions of the tests via simulation using eleven linear IV regression models with heteroskedasticity and/or autocorrelation and one right-hand side (rhs) endogenous variable ($p = 1$) and four IV’s ($k = 4$). The scenarios considered are the same as in I. Andrews (2016). They are designed to mimic models for the elasticity of inter-temporal substitution estimated by Yogo (2004) for eleven countries using quarterly data from the early 1970’s to the late 1990’s. The results show that, in an overall sense, the SR-CQLR test introduced here performs well in the scenarios considered. It has asymptotic power that is competitive with that of the PI-CLC test of I. Andrews (2016) and the MM2-SU test of Moreira and Moreira (2015), has somewhat better overall power than the JVW-CLR and MVW-CLR tests of Kleibergen (2005) and the MM1-SU test of Moreira and Moreira (2015), and has noticeably higher power than Kleibergen’s (2005) LM test and the AR test.

Fast computation of tests is very useful when constructing confidence sets by inverting the tests. In the model above, the SR-CQLR test (employed using 5000 critical value repetitions) can be computed 29,411 times in one minute using a laptop with Intel i7-3667U CPU @2.0GHz in the $(k,p) = (4,1)$ scenarios described above. This is found to be 115, 292, and 302 times faster than the PI-CLC, MM1-SU, and MM2-SU tests, respectively. For $p \geq 2$, the speed advantage is much larger.

We show how the proposed confidence intervals are implemented by constructing confidence intervals for the elasticity of intertemporal substitution (EIS) and its reciprocal using the models considered in Yogo (2004) and the data from Campbell (2003). The empirical results show no sign of the equity premium puzzle that arises when confidence intervals are constructed using methods that are not robust to weak identification.

The paper is organized as follows. Section 2 discusses the related literature. Section 3 defines the moment condition model. Sections 4 and 5 introduce the SR-AR and SR-CQLR tests, respectively. Section 6 provides the asymptotic size and similarity results for the tests. Section 7 establishes the
asymptotic efficiency of the SR-CQLR test under strong and semi-strong identification. Section 8 provides the empirical application concerning the EIS using the data and models in Yogo (2004). Section 9 provides subvector tests under the assumption that the parameters not under test are strongly identified. Section 9.4 provides the finite-sample results for the subvector tests in the probit model with endogeneity. Section 10 provides the asymptotic power comparisons based on the estimated linear IV models in Yogo (2004).

The SM, i.e., Andrews and Guggenberger (2018), contains the proofs. It also provides (i) time series results, (ii) finite-sample simulations of the null rejection probabilities of the SR-AR and SR-CQLR tests for cases where the variance matrix of the moment functions is singular and near-singular, (iii) analysis of the behavior of the SR-CQLR test and Kleibergen’s (2005, 2007) CLR tests in the homoskedastic linear IV model with fixed IV’s, (iv) the definition of a new SR-CQLR P test that reduces asymptotically to Moreira’s (2003) CLR test for all \( p \geq 1 \), but only applies when the moment functions are of a product form, \( u_i(\theta)Z_i \), where \( u_i(\theta) \) is a scalar and \( Z_i \) is a \( k \)-vector of instrumental variables, and (v) the definition of a new SR-LM test.

All limits below are taken as \( n \to \infty \) and \( A := B \) denotes that \( A \) is defined to equal \( B \).

2 Discussion of the Related Literature

Stock and Wright (2000) consider the nonlinear AR test for nonlinear moment condition models, building on the analysis of Staiger and Stock (1997) for linear IV models with weak identification. Papers in the literature that deal with identification-robust LM and CLR tests for nonlinear moment condition models include Kleibergen (2005, 2007), Guggenberger and Smith (2005), Otsu (2006), Smith (2007), Guggenberger, Ramalho, and Smith (2012), and I. Andrews (2016). None of these papers provide asymptotic size results. Kleibergen (2005) considers standard weak identification and strong identification. This excludes all cases in the nonstandard weak and semi-strong identification categories, see Section 6.2 below. All of the other papers listed obtain asymptotic results under Stock and Wright’s (2000) Assumption C. This assumption is an innovative contribution to the literature, but it has some notable drawbacks. For a detailed discussion, see Section 2 of Andrews and Guggenberger (2017) (AG1). The asymptotic results in this paper do not require Assumption C or any related conditions of this type.

I. Andrews and Mikusheva (2016) consider a different form of CLR test than those above. Their test is asymptotically similar conditional on the entire sample mean process that is orthogonalized to be asymptotically independent of the sample moments evaluated at the null parameter value. They establish correct asymptotic size of this test under an assumption that bounds the minimum
eigenvalue of the variance matrix of the sample moments away from zero. While this condition applies to many models, it rules out likelihood-based models with weak identification.

AG1 analyzes the asymptotic size properties of a class of LM and CLR tests for nonlinear moment condition models. Next, we contrast the asymptotic size results for the SR-AR and SR-CQLR tests with the asymptotic size results of AG1 for variants of Kleibergen’s (2005) CLR tests.

For a certain parameter space of null distributions $\mathcal{F}_0$, AG1 establishes correct asymptotic size for Kleibergen’s CLR tests that are based on (what AG1 calls) moment-variance-weighting (MVW) of the orthogonalized sample Jacobian matrix, combined with a rank statistic, such as the Robin and Smith (2000) rank statistic. Tests of this type have been considered by Guggenberger, Ramalho, and Smith (2012). AG1 also determines a formula for the asymptotic size of Kleibergen’s CLR tests that are based on (what AG1 calls) Jacobian-variance-weighting (JVW) of the orthogonalized sample Jacobian matrix, which is the weighting suggested by Kleibergen. However, AG1 does not show that the latter CLR tests necessarily have correct asymptotic size when $p \geq 2$. The reason is that for some sequences of distributions, the asymptotic versions of the sample moments and the (suitably normalized) rank statistic are not necessarily independent and using asymptotic independence is the only known way of showing that the asymptotic null rejection probabilities reduce to the nominal size $\alpha$. AG1 does show that these tests have correct asymptotic size when $p = 1$, for a certain subset of the parameter space $\mathcal{F}_0$.

Although Kleibergen’s CLR tests with moment-variance-weighting have correct asymptotic size for $\mathcal{F}_0$, they have some drawbacks. First, the variance matrix of the moment functions must be nonsingular, which can be restrictive. Second, the parameter space $\mathcal{F}_0$ restricts the eigenvalues of the expected outer product of the vectorized orthogonalized sample Jacobian, which can be restrictive and can be difficult to verify in some models. Third, as shown in the SM, Kleibergen’s CLR tests with moment-variance-weighting do not reduce to Moreira’s CLR test in the homoskedastic normal linear IV model with fixed IV’s when $p = 1$. Simulation results in the SM show that this leads to substantial power loss in some scenarios of this model, relative to the SR-CQLR tests considered here, Moreira’s CLR test, and Kleibergen’s CLR test with Jacobian-variance weighting. Fourth, the form of Kleibergen’s CLR test statistic for $p \geq 2$ is based on the form of Moreira’s test statistic when $p = 1$. In consequence, one needs to make a somewhat arbitrary choice of some rank statistic to reduce the $k \times p$ weighted orthogonalized sample Jacobian to a scalar random variable.

Kleibergen’s CLR tests with Jacobian-variance weighting also possess drawbacks one, two, and four stated in the previous paragraph, as well as the asymptotic size issue discussed above when $p \geq 2$. In contrast, the SR-CQLR test does not have any of these drawbacks.

Compared to the standard GMM tests considered in Hansen (1982), the SR-CQLR and SR-AR
tests have correct asymptotic size even when any of the following conditions employed in Hansen (1982) fails: (i) the moment functions have a unique zero at the true value, (ii) the expected Jacobian of the moment functions has full column rank, (iii) the variance matrix of the moment functions is nonsingular, and (iv) the true parameter lies on the interior of the parameter space. Under strong and semi-strong identification, the full-vector SR-CQLR test is asymptotically equivalent under contiguous local alternatives to the test in Hansen (1982) that uses an asymptotically efficient weight matrix.

The SR-CQLR and SR-AR tests are shown to be robust to the singularity and near-singularity of the variance matrix of the moments. In somewhat related work, Caner and Yildiz (2012) consider robustness of the continuous updating estimator to near-singularity of the variance matrix of the moments in a many weak IV’s context.

A drawback of the SR-CQLR test is that it does not have any known optimal power properties under weak identification, except in the homoskedastic normal linear IV model with $p = 1$. In contrast, Moreira and Moreira (2015) construct finite-sample unbiased tests that maximize a weighted average power criterion in the heteroskedastic and autocorrelated normal linear IV regression model with $p = 1$. I. Andrews (2016) develops a test that minimizes asymptotic maximum regret among tests that are linear combinations of Kleibergen’s LM and AR tests for linear and nonlinear minimum distance and moment condition models. For moment condition models, this test is not computationally tractable, so he proposes a plug-in test that aims to mimic the features of the infeasible optimal test. This feasible plug-in test does not have optimality properties. I. Andrews (2016) also discusses the relative power performance of the K test in scenarios with Kronecker product and non-Kronecker product variance matrices. Montiel Olea (2012) considers tests that have weighted average power optimality properties in a GMM sense under weak identification in moment condition models when $p = 1$. Whether these tests are asymptotically efficient under strong identification seems to be an open question. None of the previous papers provide asymptotic size results. Elliott, Müller, and Watson (2015) consider tests that maximize weighted average power in a variety of (finite-sample) parametric models where a nuisance parameter appears under the null. The test in I. Andrews and Mikusheva (2016) utilizes information in the entire sample moment process, which other CLR tests do not. But, like the SR-CQLR test, it does not have general asymptotic optimality properties.

Robust inference methods in scenarios where the source of weak identification is known includes Andrews and Cheng (2013), Cox (2017), and Han and McCloskey (2019).
3 Moment Condition Model

3.1 Moment Functions

The general moment condition model that we consider is

\[ E_F g(W_i, \theta) = 0^k, \]  

where the equality holds when \( \theta \in \Theta \subset \mathbb{R}^p \) is the true value, \( 0^k = (0, \ldots, 0)' \in \mathbb{R}^k \), \( \{ W_i \in \mathbb{R}^m : i = 1, \ldots, n \} \) are i.i.d. observations with distribution \( F \), \( g \) is a known (possibly nonlinear) function from \( \mathbb{R}^{m+p} \) to \( \mathbb{R}^k \), \( E_F(\cdot) \) denotes expectation under \( F \), and \( p, k, m \geq 1 \). As noted in the Introduction, we allow for \( k \geq p \) and \( k < p \). In Section 18 in the SM, we consider models with stationary strong mixing observations.

The Jacobian of the moment functions is

\[ G(W_i, \theta) := \frac{\partial}{\partial \theta} g(W_i, \theta) \in \mathbb{R}^{k \times p}. \]  

For notational simplicity, we let \( g_i(\theta) \) and \( G_i(\theta) \) abbreviate \( g(W_i, \theta) \) and \( G(W_i, \theta) \), respectively. We denote the \( j \)th column of \( G_i(\theta) \) by \( G_{ij}(\theta) \) and \( G_{ij} := G_{ij}(\theta_0) \), where \( \theta_0 \) is the (true) null value of \( \theta \), for \( j = 1, \ldots, p \). Likewise, we often leave out the argument \( \theta_0 \) for other functions as well. Thus, we write \( g_i \) and \( G_i \), rather than \( g_i(\theta_0) \) and \( G_i(\theta_0) \). We let \( I_r \) denote the \( r \) dimensional identity matrix.

We are concerned with tests of the null hypothesis

\[ H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \]  

3.2 Parameter Spaces of Distributions \( F \)

The variance matrix of the moments, \( \Omega_F(\theta) \), its rank, and its spectral decomposition are

\[ \Omega_F(\theta) := E_F (g_i(\theta) - E_F g_i(\theta))(g_i(\theta) - E_F g_i(\theta))', \]

\[ r_F(\theta) := rk(\Omega_F(\theta)), \text{ and } \Omega_F(\theta) := A_{\Omega}(\theta) \Pi_F(\theta) A_{\Omega}(\theta)', \]  

where \( rk(\cdot) \) denotes the rank of a matrix, \( \Pi_F(\theta) \) is the \( k \times k \) diagonal matrix with the eigenvalues of \( \Omega_F(\theta) \) on the diagonal in nonincreasing order, and \( A_{\Omega}(\theta) \) is a \( k \times k \) orthogonal matrix of eigenvectors

\footnote{The asymptotic size results given below do not actually require \( G(W_i, \theta) \) to be the derivative matrix of \( g(W_i, \theta) \). The matrix \( G(W_i, \theta) \) can be any \( k \times p \) matrix that satisfies the conditions in \( \mathcal{F}^{\times \mathbb{R}} \), defined in (3.1) below. For example, \( G(W_i, \theta) \) can be the derivative of \( g(W_i, \theta) \) almost surely, rather than for all \( W_i \), which allows \( g(W_i, \theta) \) to have kinks. The function \( G(W_i, \theta) \) also can be a numerical derivative, such as \( ((g(W_i, \theta + \varepsilon e_j) - g(W_i, \theta))/\varepsilon, \ldots, (g(W_i, \theta + \varepsilon e_p) - g(W_i, \theta))/\varepsilon) \in \mathbb{R}^{k \times p} \) for some \( \varepsilon > 0 \), where \( e_j \) is the \( j \)th unit vector, e.g., \( e_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^p \).}
corresponding to the eigenvalues in $\Pi_F(\theta)$. We allow for the case where $\Omega_F(\theta)$ is singular. We partition $A_F^\Omega(\theta)$ according to whether the corresponding eigenvalues are positive or zero:

$$A_F^\Omega(\theta) = [A_F(\theta), A_F^{-}(\theta)],$$

where $A_F(\theta) \in \mathbb{R}^{k \times \rho(\theta)}$ and $A_F^{-}(\theta) \in \mathbb{R}^{k \times (k-\rho(\theta))}$. (3.5)

The columns of $A_F(\theta)$ are eigenvectors of $\Omega_F(\theta)$ that correspond to positive eigenvalues of $\Omega_F(\theta)$. Let $\Pi_1(\theta)$ denote the upper left $r_F(\theta) \times r_F(\theta)$ submatrix of $\Pi_F(\theta)$. The matrix $\Pi_1(\theta)$ is diagonal with the positive eigenvalues of $\Omega_F(\theta)$ on its diagonal in nonincreasing order.

The $r_F$ vector $\Pi_1^{-1/2}A_F g_i$ is a vector of non-redundant linear combinations of the moment functions evaluated at $\theta_0$ rescaled to have variances equal to one: $Var_F(\Pi_1^{-1/2}A_F g_i) = \Pi_1^{-1/2}A_F \Omega_F A_F \Pi_1^{-1/2} = I_{r_F}$. The $r_F \times p$ matrix $\Pi_1^{-1/2}A_F G_i$ is the analogously transformed Jacobian matrix.

For the SR-AR and SR-CQLR tests, we consider the following parameter spaces for the distribution $F$ that generates the data under $H_0 : \theta = \theta_0$:

$$\mathcal{F}^{SR}_{AR} := \{ F : E_F g_i = 0^k \text{ and } E_F \| \Pi_1^{-1/2}A_F g_i\|^{2+\gamma} \leq \mathcal{M} \} \text{ and}$$

$$\mathcal{F}^{SR} := \{ F \in \mathcal{F}^{SR}_{AR} : E_F \| vec(\Pi_1^{-1/2}A_F G_i)\|^{2+\gamma} \leq \mathcal{M} \},$$

respectively, for some $\gamma > 0$ and some $\mathcal{M} < \infty$, where $\| \cdot \|$ denotes the Euclidean norm, and $vec(\cdot)$ denotes the vector obtained from stacking the columns of a matrix.

The first condition in $\mathcal{F}^{SR}_{AR}$ is the defining condition of the model. The second condition in $\mathcal{F}^{SR}_{AR}$ is a mild moment condition on the rescaled non-redundant moment functions $\Pi_1^{-1/2}A_F g_i$. For example, consider the case where $W_i \sim iid N(\theta, \Omega_F)$ for $\theta \in \mathbb{R}^k, \Omega_F \in \mathbb{R}^{k \times k}, g(W_i, \theta) := W_i - \theta, \Omega_F$ has spectral decomposition $A_F \Pi_F A_F'$, and some eigenvalues of $\Omega_F$ may be close to zero or equal to zero. In this case, $\Pi_1^{-1/2}A_F g_i$ is a vector of independent standard normal random variables and the moment conditions in $\mathcal{F}^{SR}_{AR}$ and $\mathcal{F}^{SR}$ hold immediately. The condition in $\mathcal{F}^{SR}$ is a mild moment condition on the analogously transformed derivatives of the moment conditions $\Pi_1^{-1/2}A_F G_i$.

Identification issues arise when $E_F G_i$ has, or is close to having, less than full column rank, which occurs when $k < p$ or $p \geq k$ and one or more of its singular values is zero or close to zero. The sets $\mathcal{F}^{SR}_{AR}$ and $\mathcal{F}^{SR}$ place no restrictions on the column rank or singular values of $E_F G_i$.

The conditions in $\mathcal{F}^{SR}_{AR}$ and $\mathcal{F}^{SR}$ also place no restrictions on the variance matrix $\Omega_F := E_F g_i g_i'$ of $g_i$, such as $\lambda_{\text{min}}(\Omega_F) \geq \delta$ for some $\delta > 0$ or $\lambda_{\text{min}}(\Omega_F) > 0$. This is particularly desirable in cases where identification failure yields singularity of $\Omega_F$ (and weak identification is accompanied by near singularity of $\Omega_F$.) This occurs in all likelihood scenarios. In such scenarios, $g_i(\theta)$ is the score function, the negative expected Jacobian matrix $-E_F G_i$ equals the expected outer product of the
score function $\Omega_F$, i.e., $-E_FG_i = \Omega_F$ (by the information matrix equality), and weak identification occurs when $\Omega_F$ is close to being singular.

Another example where $\Omega_F$ can be singular is in the model for interest rate dynamics in Jagannathan, Skoulakis, and Wang (2002, Sec. 6.2) (JSW). JSW consider five moment conditions for a four dimensional parameter $\theta$. Grant (2013) shows that the variance matrix of the moment functions for this model is singular when one or more of three restrictions on the parameters holds. When any two of these restrictions hold, the parameter also is unidentified, see Section 15.1 in the SM for details.

In these examples and others like them, $E_FG_i$ is close to having less than full column rank and $\Omega_F$ is close to being singular when the null value $\theta_0$ is close to a value which yields reduced column rank of $E_FG_i$ and singularity of $\Omega_F$. Null hypotheses of this type are important for CS’s because uniformity over null hypothesis values is necessary for CS’s to have correct asymptotic size. Hence, it is important to have procedures available that place no restrictions on either $E_FG_i$ or $\Omega_F$.

The parameter spaces for $(F, \theta)$ for the SR-AR and SR-CQLR CS’s are

\[
\mathcal{F}_{\Theta,\text{AR}}^{SR} := \{(F, \theta_0) : F \in \mathcal{F}_{\text{AR}}^{SR}(\theta_0), \theta_0 \in \Theta\} \quad \text{and} \quad \mathcal{F}_{\Theta}^{SR} := \{(F, \theta_0) : F \in \mathcal{F}^{SR}(\theta_0), \theta_0 \in \Theta\},
\]

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

4 Singularity-Robust Nonlinear Anderson-Rubin Test

The nonlinear Anderson-Rubin (AR) test was introduced by Stock and Wright (2000). (They refer to it as an $S$ test.) It is robust to identification failure and weak identification, but it relies on nonsingularity of the variance matrix of the moment functions. In this section, we introduce a singularity-robust nonlinear AR (SR-AR) test that generalizes the $S$ test of Stock and Wright (2000) and allows for a singular variance matrix of the moment functions.

As noted in the Introduction, there are a number of likelihood-based models that nest submodels of interest within which the parameter is not identified. In such models, it is undesirable and unnatural to rule out the case where the true distribution lies in the submodel. In consequence, for such models, the SR-AR test introduced in this section—which allows for lack of identification and singularity of the variance matrix of the moments—has significant advantages over the standard nonlinear AR test—which does not. Seven examples of models of this type are listed in the Introduction. At the end of this section, we provide more detail concerning these models.
The sample moments and an estimator of the variance matrix of the moments, $\Omega_F(\theta)$, are:

$$\hat{g}_n(\theta) := n^{-1} \sum_{i=1}^{n} g_i(\theta) \quad \text{and} \quad \hat{\Omega}_n(\theta) := n^{-1} \sum_{i=1}^{n} g_i(\theta)g_i(\theta)' - \hat{g}_n(\theta)\hat{g}_n(\theta)' .$$  \hfill (4.1)

The usual nonlinear AR statistic is

$$AR_n(\theta) := n\hat{g}_n(\theta)\hat{\Omega}_n^{-1}(\theta)\hat{g}_n(\theta).$$  \hfill (4.2)

The nonlinear AR test rejects $H_0 : \theta = \theta_0$ if $AR_n(\theta_0) > \chi^2_{k,1-\alpha}$, where $\chi^2_{k,1-\alpha}$ is the $1 - \alpha$ quantile of the chi-square distribution with $k$ degrees of freedom.

Now, we introduce sample versions of the population quantities $r_F(\theta)$, $A^\Omega_F(\theta)$, $A_F(\theta)$, $A^\perp_F(\theta)$, and $\Pi_F(\theta)$ in (3.4) and (3.5). The rank and spectral decomposition of $\hat{\Omega}_n(\theta)$ are

$$\hat{r}_n(\theta) := rk(\hat{\Omega}_n(\theta)) \quad \text{and} \quad \hat{\Omega}_n(\theta) := \hat{A}^\Omega_n(\theta)\hat{\Pi}_n(\theta)\hat{A}^\perp_n(\theta)',$$  \hfill (4.3)

where $\hat{\Pi}_n(\theta)$ is the $k \times k$ diagonal matrix with the eigenvalues of $\hat{\Omega}_n(\theta)$ on the diagonal in non-increasing order, and $\hat{A}^\Omega_n(\theta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\hat{\Pi}_n(\theta)$. We partition $\hat{A}^\Omega_n(\theta)$ according to whether the corresponding eigenvalues are positive or zero:

$$\hat{A}^\Omega_n(\theta) = [\hat{A}_n(\theta), \hat{A}^\perp_n(\theta)], \quad \text{where} \quad \hat{A}_n(\theta) \in R^{k \times \hat{r}_n(\theta)} \quad \text{and} \quad \hat{A}^\perp_n(\theta) \in R^{k \times (k - \hat{r}_n(\theta))}. $$  \hfill (4.4)

The columns of $\hat{A}_n(\theta)$ are eigenvectors of $\hat{\Omega}_n(\theta)$ that correspond to positive eigenvalues of $\hat{\Omega}_n(\theta)$. The eigenvectors in $\hat{A}_n(\theta)$ are not uniquely defined, but the eigenspace spanned by these vectors is. The tests and CS’s defined here and below using $\hat{A}_n(\theta)$ are numerically invariant to the particular choice of $\hat{A}_n(\theta)$ (by the invariance results given in Lemma 5.1 below).

Define $\hat{g}_{An}(\theta)$ and $\hat{\Omega}_{An}(\theta)$ as $\hat{g}_n(\theta)$ and $\hat{\Omega}_n(\theta)$ are defined in (4.1), but with $\hat{A}_n(\theta)'g_i(\theta)$ in place of $g_i(\theta)$. That is,

$$\hat{g}_{An}(\theta) := \hat{A}_n(\theta)'\hat{g}_n(\theta) \in R^{\hat{r}_n(\theta)} \quad \text{and} \quad \hat{\Omega}_{An}(\theta) := \hat{A}_n(\theta)'\hat{\Omega}_n(\theta)\hat{A}_n(\theta) \in R^{\hat{r}_n(\theta) \times \hat{r}_n(\theta)}.$$  \hfill (4.5)

The SR-AR test statistic is defined by

$$SR-AR_n(\theta) := n\hat{g}_{An}(\theta)\hat{\Omega}_{An}^{-1}(\theta)\hat{g}_{An}(\theta).$$  \hfill (4.6)
The SR-AR test rejects the null hypothesis \( H_0 : \theta = \theta_0 \) if

\[
SR-AR_n(\theta_0) > \chi^2_{\tilde{r}_n(\theta_0),1-\alpha} \text{ or } \tilde{A}^+_n(\theta_0)'\tilde{g}_n(\theta_0) \neq 0^{k-\tilde{r}_n(\theta_0)}, \tag{4.7}
\]

where by definition the latter condition does not hold if \( \tilde{r}_n(\theta_0) = k \). If \( \tilde{r}_n(\theta_0) = 0 \), then \( SR-AR_n(\theta_0) := 0 \) and \( \chi^2_{\tilde{r}_n(\theta_0),1-\alpha} := 0 \) and the SR-AR test rejects \( H_0 \) if \( \tilde{g}_n(\theta_0) \neq 0^k \).

The extra rejection condition in (4.7), \( \tilde{A}^+_n(\theta_0)'\tilde{g}_n(\theta_0) \neq 0^{k-\tilde{r}_n(\theta_0)} \), improves power, but we show it has no effect under \( H_0 \) with probability that goes to one (wp→1), see Lemma 17.1 in the SM. It improves power because it fully exploits, rather than ignores, the nonstochastic part of the moment conditions associated with the singular part of the variance matrix. For example, if the moment conditions include some identities and the moment variance matrix excluding the identities is nonsingular, then \( \tilde{A}^+_n(\theta_0)'\tilde{g}_n(\theta_0) \) consists of the identities and the SR-AR test rejects \( H_0 \) if the identities do not hold when evaluated at \( \theta_0 \) or if the SR-AR statistic, which ignores the identities, is sufficiently large. Two other simple examples where the extra rejection condition improves power are given in Section 15.2 in the SM.\(^3\)

The SR-AR test statistic can be written equivalently as

\[
SR-AR_n(\theta) = n\tilde{g}_n(\theta)'\tilde{\Omega}^+_n(\theta)\tilde{g}_n(\theta) \tag{4.8}
\]

where \( \tilde{\Omega}^+_n(\theta) \) is the Moore-Penrose generalized inverse of \( \tilde{\Omega}_n(\theta) \), see (14.1) in the SM.

The nominal 100\((1-\alpha)\)% SR-AR CS is

\[
CS_{SR-AR,n} := \{ \theta_0 \in \Theta : SR-AR_n(\theta_0) \leq \chi^2_{\tilde{r}_n(\theta_0),1-\alpha} \text{ and } \tilde{A}^+_n(\theta_0)'\tilde{g}_n(\theta_0) = 0^{k-\tilde{r}_n(\theta_0)} \}. \tag{4.9}
\]

By definition, if \( \tilde{r}_n(\theta_0) = k \), the condition \( \tilde{A}^+_n(\theta_0)'\tilde{g}_n(\theta_0) = 0^{k-\tilde{r}_n(\theta_0)} \) holds. When \( \tilde{r}_n(\theta_0) = k \), \( SR-AR_n(\theta_0) = AR_n(\theta_0) \) because \( \tilde{A}_n(\theta_0) \) is invertible and \( \tilde{\Omega}^{-1}_n(\theta_0) = \tilde{A}_n^{-1}(\theta_0)\tilde{\Omega}_n^{-1}(\theta_0)\tilde{A}_n^{-1}(\theta_0)' \).

Section 20 in the SM provides some finite-sample simulations of the null rejection probabilities of the SR-AR test when the variance matrix of the moments is singular and near singular. The results show that the SR-AR test works very well in the model that is considered in the simulations.

Now we discuss the seven models listed in the Introduction. In each model, the sample moments

\(^3\)In addition, the extra rejection condition has no effect on the finite-sample null rejection probabilities if \( rk(\tilde{\Omega}_n(\theta_0)) = rk(\tilde{\Omega}_F(\theta_0)) \) (a.s., see the proof of Lemma 17.1(b) in the SM. The stochastic part of \( g_i(\theta_0) \) is \( A_F(\theta_0)'g_i(\theta_0) \) and its variance matrix, \( A_F(\theta_0)'\tilde{\Omega}_F(\theta_0)A_F(\theta_0) \), is nonsingular by construction. The previous rank condition holds whenever the sample variance matrix of \( \{ A_F(\theta_0)'g_i(\theta_0) : i \leq n \} \) has full rank \( r_F \) a.s. The latter often holds whenever \( n \geq k+1 \).

\(^4\)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 statistic values 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. See Section 15.2 in the SM for an example that illustrates this.
are the likelihood score. In factor models it is usually the case that the number of factors is uncertain. Hence, in a factor model with \( N_f \) factors, one is usually interested in the case where the actual number of factors is \( J = 0, \ldots, N_f \). However, when the factor loadings are such that only \( J < N_f \) factors enter the model, the variances of the \( N_f - J \) factors that do not enter the model are not identified. Hence, in order to carry out inference that is robust to different numbers of factors in the model, one requires robustness to weak and lack of identification and near and exact singularity of the variance matrix of the moments.

In mixture models and regime switching models, it is usually of interest to consider the submodel in which no mixing (or switching) occurs. But, typically the parameter vector is not identified in this submodel. For example, consider the simple mixture of normals model with mixing distributions \( N(\mu_1, \sigma_1^2) \) and \( N(\mu_2, \sigma_2^2) \) and mixing probability \( p \). In this model, the nested submodel is a \( N(\mu, \sigma^2) \) model and it arises when \( p = 0 \) or \( 1 \) or \( (\mu_1, \sigma_1^2) = (\mu_2, \sigma_2^2) \). In this submodel, the parameter vector \( (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p) \) is not identified and the variance matrix of the moments is singular. Close to this submodel, this parameter vector is weakly identified and the variance matrix is near singular.

A model for asset returns with jumps is another example of a mixture model. The existence or nonexistence of jumps is often an issue of considerable interest. It is common to take the jump component to be of the form \( \sum_{j=0}^{N_J} \xi_j \), where \( \xi_j \sim N(\mu_\xi, \sigma_\xi^2) \) and \( N_J \) has a Poisson distribution with parameter \( \lambda_\xi \), e.g., see Jorion (1988) and Chan and Maheu (2002). When \( \lambda_\xi = 0 \), there are no jumps, the parameters \( (\mu_\xi, \sigma_\xi^2) \) are not identified, and the variance matrix of sample moments is singular.

In a random coefficients model, it is usually of interest to consider the case where the coefficients are nonrandom. In this case, the parameter vector often is not identified and the variance matrix of the sample moments is singular. For example, consider a linear regression model \( Y_i = \mu + X'_i \beta_i + u_i \), where \( \beta_i := \beta + \xi_i \in R^2 \), \( \beta \) is a constant vector, \( \xi_i \sim N(0^2, \xi) \) independent of the error \( u_i \sim N(0, \sigma_u^2) \), and \( V_\xi \) is a \( 2 \times 2 \) variance matrix with variances \( \sigma_{\xi_1}^2 \) and \( \sigma_{\xi_2}^2 \) and correlation \( \rho_\xi \). In the partially or wholly constant coefficient model, we have \( \sigma_{\xi_1}^2 = 0 \) and/or \( \sigma_{\xi_2}^2 = 0 \) and \( \rho_\xi \) is not identified. As another example, suppose \( \beta_i := \beta + \xi_i \) is a scalar random coefficient in the linear regression model above, \( (\xi_i, u_i) \sim N(0^2, \xi_u) \), \( V_u \) is a \( 2 \times 2 \) variance matrix with variances \( \sigma_u^2 \) and \( \sigma_u^2 \) and correlation \( \rho_{\xi_u} \). In the constant coefficient submodel, we have \( \sigma_{\xi}^2 = 0 \), \( \rho_{\xi_u} \) is not identified, and the sample moments have a singular variance matrix.

A GARCH model of conditional heteroskedasticity nests a homoskedastic model, which is often of empirical interest for financial or macroeconomic variables observed at a relatively low frequency, such as a month. For example, the GARCH(1,1) model is of the form: \( Y_i = \sigma_i \varepsilon_i, \sigma_i^2 = \omega + \alpha \varepsilon_{i-1}^2 + \rho \sigma_{i-1}^2, E \varepsilon_i = 0, \) and \( E \varepsilon_i^2 = 1 \). When the GARCH parameter \( \alpha \) equals zero, \( \sigma_i^2 = \omega/(1 - \rho) \), \( (\omega, \rho) \) is
not identified, and the variance matrix of the sample moments is singular. Similarly, an ARCH or GARCH in mean model nests a homoskedastic model with no heteroskedastic mean effect and lack of identification. For example, the ARCH(1) in mean model is of the form: 

\[ Y_i = \mu + \sigma_i^2 \beta + \sigma_i \varepsilon_i, \]

\[ \sigma_i^2 = \omega + \alpha \varepsilon_{i-1}^2, \quad E \varepsilon_i = 0, \quad E \varepsilon_i^2 = 1. \]

When the ARCH parameter \( \alpha \) equals zero, \( \sigma_i^2 = \omega \), the mean of \( Y_i \) becomes \( \mu + \omega \beta \), \( (\mu, \beta) \) is not identified, and the variance matrix of the sample moments is singular.

The ARMA(1,1) model is a workhorse model of time series analysis. It nests the important submodel with no serial correlation. This submodel arises when the AR and MA parameters are equal. The model is of the form:

\[ Y_i = Y_{i-1} + \varepsilon_i - \pi \varepsilon_{i-1}, \quad \text{where} \quad E \varepsilon_i = 0, \quad E \varepsilon_i^2 = \sigma^2, \quad \text{and} \quad \{\varepsilon_i : i \geq 1\} \]

are serially uncorrelated. When \( \rho = \pi \), the model reduces to \( Y_i = \varepsilon_i \), the value of \( \rho = \pi \) is not identified, and the sample moments have a singular variance matrix. Similar “common factor” identification and variance singularity issues also arise in higher-order ARMA(\( p,q \)) models.

5 SR-CQLR Test

This section defines the SR-CQLR test. For expositional clarity and convenience (here and in the proofs), we first define the test in Section 5.1 for the case of nonsingular sample and population moments variance matrices, \( \hat{\Omega}_n(\theta) \) and \( \Omega_F(\theta) \), respectively. Then, we extend the definition in Section 5.2 to the case where these variance matrices may be singular.

5.1 CQLR Test for Nonsingular Moments Variance Matrices

The sample Jacobian is

\[ \hat{G}_n(\theta) := n^{-1} \sum_{i=1}^n G_i(\theta) = (\hat{G}_{1n}(\theta), \ldots, \hat{G}_{pn}(\theta)) \in \mathbb{R}^{k \times p}. \]  

(5.1)

The conditioning matrix \( \hat{D}_n(\theta) \) is defined, as in Kleibergen (2005), to be the sample Jacobian matrix \( \hat{G}_n(\theta) \) adjusted to be asymptotically independent of the sample moments \( \hat{\gamma}_n(\theta) \):

\[ \hat{D}_n(\theta) := (\hat{D}_{1n}(\theta), \ldots, \hat{D}_{pn}(\theta)) \in \mathbb{R}^{k \times p}, \quad \text{where} \]

\[ \hat{D}_{jn}(\theta) := \hat{G}_{jn}(\theta) - \hat{\Gamma}_{jn}(\theta) \hat{\Omega}_n^{-1}(\theta) \hat{\gamma}_n(\theta) \in \mathbb{R}^k \quad \text{for} \quad j = 1, \ldots, p, \quad \text{and} \]

\[ \hat{\Gamma}_{jn}(\theta) := n^{-1} \sum_{i=1}^n (G_{ij}(\theta) - \hat{G}_{jn}(\theta)) g_i(\theta)' \in \mathbb{R}^{k \times k} \quad \text{for} \quad j = 1, \ldots, p. \]  

(5.2)

We call \( \hat{D}_n(\theta) \) the orthogonalized sample Jacobian matrix. This statistic requires that \( \hat{\Omega}_n^{-1}(\theta) \) exists.
Next, we define

\[ \hat{R}_n(\theta) := (B(\theta)' \otimes I_k) \hat{V}_n(\theta) (B(\theta) \otimes I_k) \in R^{(p+1)k \times (p+1)k}, \]

where

\[ \hat{V}_n(\theta) := n^{-1} \sum_{i=1}^{n} \left( f_i(\theta) - \hat{f}_n(\theta) \right) \left( f_i(\theta) - \hat{f}_n(\theta) \right)' \in R^{(p+1)k \times (p+1)k}, \tag{5.3} \]

\[ f_i(\theta) := \begin{pmatrix} g_i(\theta) \\ \text{vec}(G_i(\theta)) \end{pmatrix}, \quad \hat{f}_n(\theta) := \begin{pmatrix} \hat{g}_n(\theta) \\ \text{vec}(\hat{G}_n(\theta)) \end{pmatrix}, \quad \text{and } B(\theta) := \begin{pmatrix} 1 & 0'_p \\ -\theta & -I_p \end{pmatrix}. \]

The estimator \( \hat{R}_n(\theta) \), as well as \( \hat{\Sigma}_n(\theta) \) and \( \hat{L}_n(\theta) \) defined below, are defined so that the CQLR and SR-CQLR tests, which employ them, are asymptotically equivalent to Moreira’s (2003) CLR test in the homoskedastic linear IV model with fixed IV’s with \( p = 1 \) rhs endogenous variable and under standard weak, semi-strong, and strong identification for any \( p \geq 2 \) rhs endogenous variables. See Section 19 in the SM for details. (In the nonstandard weak identification category, see Section 6.2 below, asymptotic non-equivalence is due only to the difference between fixed and random IV’s and, in consequence, it is small.)

We define \( \hat{\Sigma}_n(\theta) \in R^{(p+1) \times (p+1)} \) to be the symmetric positive definite (pd) matrix that minimizes

\[ \left\| (I_{p+1} \otimes \hat{\Omega}_n^{-1/2}(\theta))(\Sigma \otimes \hat{\Omega}_n(\theta) - \hat{R}_n(\theta))(I_{p+1} \otimes \hat{\Omega}_n^{-1/2}(\theta)) \right\| \tag{5.4} \]

over all symmetric pd matrices \( \Sigma \in R^{(p+1) \times (p+1)} \), where \( \| \cdot \| \) denotes the Frobenius norm. This is a weighted minimization problem with the weights given by \( I_{p+1} \otimes \hat{\Omega}_n^{-1/2}(\theta) \). In the homoskedastic linear IV model, the population version of \( \hat{R}_n(\theta) \) has a Kronecker product form and therefore the Kronecker product approximation in (5.4) leads to the asymptotic equivalence of the CQLR test and Moreira’s (2003) CLR test in the homoskedastic linear IV model. We employ the weights above because they lead to a matrix \( \hat{\Sigma}_n(\theta) \) that is invariant to the multiplication of \( g_i(\theta) \) and \( G_i(\theta) \) by any nonsingular matrix \( M \in R^{k \times k} \), see Lemma 5.1 below. Let \( \hat{\Sigma}_{j\ell n}(\theta) \) denote the \((j, \ell)\) element of \( \hat{\Sigma}_n(\theta) \) and \( \hat{R}_{j\ell n}(\theta) \) the \((j, \ell)\) \( k \times k \) submatrix of dimension of \( \hat{R}_n(\theta) \).\footnote{That is, \( \hat{R}_{j\ell n}(\theta) \) contains the elements of \( \hat{R}_n(\theta) \) indexed by rows \((j - 1)k + 1\) to \( jk\) and columns \((\ell - 1)k\) to \( \ell k\).}

By Theorems 3 and 10 of Van Loan and Pitsianis (1993), for \( j, \ell = 1, \ldots, p+1 \), the solution to (5.4) is

\[ \hat{\Sigma}_{j\ell n}(\theta) = \text{tr}(\hat{R}_{j\ell n}(\theta)'\hat{\Omega}_n^{-1}(\theta))/k \tag{5.5} \]

We use an eigenvalue-adjusted version of \( \hat{\Sigma}_n(\theta) \), denoted \( \hat{\Sigma}_n^e(\theta) \), that improves the asymptotic and finite-sample size performance of the CQLR test in some scenarios by making it robust to sin-
gularities and near singularities of the matrix that \( \tilde{\Sigma}_n(\theta) \) estimates. The adjustment affects the test statistic (i.e., \( \hat{\Sigma}_n(\theta) \neq \tilde{\Sigma}_n(\theta) \)) only if the condition number of \( \tilde{\Sigma}_n(\theta) \) (i.e., \( \lambda_{\text{max}}(\tilde{\Sigma}_n(\theta))/\lambda_{\text{min}}(\tilde{\Sigma}_n(\theta)) \)) exceeds \( 1/\varepsilon \). Hence, for a reasonable choice of \( \varepsilon \), it often has no effect even in finite samples. Based on the finite-sample simulations, we recommend using \( \varepsilon = .01 \).

Let \( H \in R^{d_H \times d_H} \) be any non-zero positive semi-definite (psd) matrix with spectral decomposition \( A_H \Lambda_H A_H' \), where \( \Lambda_H = \text{Diag}\{\lambda_{H1}, ..., \lambda_{Hd_H}\} \) is the diagonal matrix of eigenvalues of \( H \) with nonnegative nonincreasing diagonal elements and \( A_H \) is a corresponding orthogonal matrix of eigenvectors of \( H \). For \( \varepsilon > 0 \), the eigenvalue-adjusted matrix \( H^\varepsilon \) is

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

(5.6)

where \( \lambda_{\max}(H) \) denotes the maximum eigenvalue of \( H \). Note that \( H^\varepsilon = H \) whenever the condition number of \( H \) is less than or equal to \( 1/\varepsilon \) (for \( \varepsilon \leq 1 \)). In Lemma 22.1 in the SM, we show that the eigenvalue-adjustment procedure possesses the following desirable properties: (i) \( H^\varepsilon \) is uniquely defined, (ii) \( \lambda_{\min}(H^\varepsilon) \geq \lambda_{\max}(H)\varepsilon \), (iii) \( \lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) \leq \max\{1/\varepsilon, 1\} \), (iv) for all \( c > 0 \), \( (cH)^\varepsilon = cH^\varepsilon \), and (v) \( H_n^\varepsilon \rightarrow H^\varepsilon \) for any sequence of psd matrices \( \{H_n : n \geq 1\} \) with \( H_n \rightarrow H \).

The QLR statistic is

\[
QLR_n(\theta) := AR_n(\theta) - \lambda_{\min}(n\tilde{\Sigma}_n(\theta)),
\]

where

\[
\tilde{\Sigma}_n(\theta) := \left( \tilde{\Theta}_n^{-1/2}(\theta) \tilde{g}_n(\theta), \tilde{D}_n^*(\theta) \right) \left( \tilde{\Theta}_n^{-1/2}(\theta) \tilde{g}_n(\theta), \tilde{D}_n^*(\theta) \right)^t \in R^{(p+1) \times (p+1)},
\]

(5.7)

\[
\tilde{\Theta}_n(\theta) := \tilde{\Theta}_n^{-1/2}(\theta) \tilde{\Theta}_n(\theta) \tilde{\Theta}_n^{-1/2}(\theta) \in R^{k \times p}, \quad \tilde{g}_n(\theta) := (\theta, I_p)(\tilde{\Sigma}_n(\theta))^{-1}(\theta, I_p)^t \in R^{p \times p},
\]

and \( \tilde{\Sigma}_n^\varepsilon(\theta) \) is defined in (5.6) with \( H = \tilde{\Sigma}_n(\theta) \).

The CQLR test uses a conditional critical value that depends on the \( k \times p \) matrix \( n^{1/2}\tilde{D}_n^*(\theta_0) \). For nonrandom \( D \in R^{k \times p} \), let

\[
CLR_{k,p}(D) := Z'Z - \lambda_{\min}(Z, D)'(Z, D)), \quad \text{where} \quad Z \sim N(0^k, I_k).
\]

(5.8)

Define \( c_{k,p}(D, 1 - \alpha) \) to be the \( 1 - \alpha \) quantile of the distribution of \( CLR_{k,p}(D) \). For given \( D \), \( c_{k,p}(D, 1 - \alpha) \) can be computed by simulation very quickly and easily.

For \( \alpha \in (0, 1) \), the nominal \( \alpha \) CQLR test rejects \( H_0 : \theta = \theta_0 \) if

\[
QLR_n(\theta_0) > c_{k,p}(n^{1/2}\tilde{D}_n^*(\theta_0), 1 - \alpha).
\]

(5.9)

The nominal 100(1-\( \alpha \))% CQLR CS is \( CS_{CQLR,n} := \{ \theta_0 \in \Theta : QLR_n(\theta_0) \leq c_{k,p}(n^{1/2}\tilde{D}_n^*(\theta_0), 1 - \alpha) \} \).
Next, we show that the CQLR test is invariant to nonsingular transformations of the moment functions/IV’s. We suppress the dependence on $\theta$ of the statistics in the following lemma.

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

**Comment:** This Lemma is used to obtain the correct asymptotic size of the CQLR test without assuming that $\lambda_{\min}(\Omega_F)$ is bounded away from zero. It suffices that $\Omega_F$ is nonsingular. In the proofs we transform the moments by $g_i \sim M_F g_i$, 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.

### 5.2 Singularity-Robust CQLR Test

Now, we extend the CQLR test to allow for singularity of the population and sample variance matrices of $g_i(\theta)$. First, we adjust $\hat{D}_n(\theta)$ to allow obtaining a conditioning statistic that is robust to the singularity of $\hat{\Omega}_n(\theta)$. For $\hat{\tau}_n(\theta) \geq 1$, where $\hat{\tau}_n(\theta)$ is defined in (4.3), we define $\hat{D}_{An}(\theta)$ as $\hat{D}_n(\theta)$ is defined in (5.2), but with $\hat{A}_n(\theta)'g_i(\theta)$, $\hat{A}_n(\theta)'G_{ij}(\theta)$, and $\hat{\Omega}_{An}(\theta)$ in place of $g_i(\theta)$, $G_{ij}(\theta)$, and $\hat{\Omega}_n(\theta)$, respectively, for $j = 1, ..., p$, where $\hat{A}_n(\theta)$ and $\hat{\Omega}_{An}(\theta)$ are defined in (4.4) and (4.5), respectively:

$$
\hat{D}_{An}(\theta) := (\hat{D}_{A1}(\theta), ..., \hat{D}_{Apn}(\theta)) \in R^{\hat{\tau}_n(\theta) \times p}, \text{ where }
\hat{D}_{Ajn}(\theta) := \hat{G}_{Ajn}(\theta) - \hat{\Gamma}_{Ajn}(\theta)\hat{\Omega}_{An}^{-1}(\theta)\hat{g}_{An}(\theta) \in R^{\hat{\tau}_n(\theta)} \text{ for } j = 1, ..., p,
\hat{G}_{An}(\theta) := \hat{A}_n(\theta)'\hat{G}_n(\theta) = (\hat{G}_{A1}(\theta), ..., \hat{G}_{Apn}(\theta)) \in R^{\hat{\tau}_n(\theta) \times p}, \text{ and }
\hat{\Gamma}_{Ajn}(\theta) := \hat{A}_n(\theta)'\hat{\Gamma}_{jn}(\theta)\hat{A}_n(\theta) \text{ for } j = 1, ..., p. \tag{5.10}
$$

Similarly, we define $\hat{R}_{An}(\theta)$, $\hat{\Sigma}_{An}(\theta)$, $\hat{L}_{An}(\theta)$, and $\hat{D}_{An}^*(\theta)$ just as $\hat{R}_n(\theta)$, $\hat{\Sigma}_n(\theta)$, $\hat{L}_n(\theta)$, and $\hat{D}_n^*(\theta)$ are defined in Section 5.1 but with $\hat{g}_{An}(\theta)$, $\hat{G}_{An}(\theta)$, $\hat{\Omega}_{An}(\theta)$, and $\hat{\tau}_n(\theta)$ in place of $\hat{g}_n(\theta)$, $\hat{G}_n(\theta)$, $\hat{\Omega}_n(\theta)$, and $\hat{\tau}_n(\theta)$, respectively:

$$
\hat{R}_{An}(\theta) := (B(\theta)' \otimes I_{F_n(\theta)})\hat{V}_{An}(\theta) (B(\theta) \otimes I_{F_n(\theta)}) \in R^{(p+1)\hat{\tau}_n(\theta) \times (p+1)\hat{\tau}_n(\theta)}, \text{ where }
\hat{V}_{An}(\theta) := (I_{p+1} \otimes \hat{A}_n(\theta)')(\hat{V}_n(\theta)(I_{p+1} \otimes \hat{A}_n(\theta))) \in R^{(p+1)\hat{\tau}_n(\theta) \times (p+1)\hat{\tau}_n(\theta)},
\hat{\Sigma}_{Aj\ell n}(\theta) := \text{tr}(\hat{R}_{Aj\ell n}(\theta)\hat{\Omega}_{An}^{-1}(\theta))/\hat{\tau}_n(\theta) \text{ for } j, \ell = 1, ..., p + 1,
\hat{L}_{An}(\theta) := (\theta, I_p)(\hat{\Sigma}_{An}(\theta))^{-1}(\theta, I_p)' \in R^{p \times p}, \hat{D}_{An}^*(\theta) := \hat{\Omega}_{An}^{-1/2}(\theta)\hat{D}_{An}(\theta)\hat{L}_{An}^{-1/2}(\theta) \in R^{\hat{\tau}_n(\theta) \times p}, \tag{5.11}
$$

Electronic copy available at: https://ssrn.com/abstract=3366443
\( \hat{\Lambda}_n(\theta) \) is defined in (4.4), \( \hat{\Sigma}_{\text{Aij}}(\theta) \) denotes the \((j, \ell)\) element of \( \hat{\Sigma}_n(\theta) \), and \( \hat{R}_{\text{Aij}}(\theta) \) denotes the \((j, \ell)\) submatrix of dimension \( \hat{r}_n(\theta) \times \hat{r}_n(\theta) \) of \( \hat{R}_n(\theta) \).

For \( \hat{r}_n(\theta) > 0 \), the SR-QLR statistic is defined by

\[
\text{SR-QLR}_n(\theta) := \text{SR-AR}_n(\theta) - \lambda_{\min}(n\hat{Q}_n(\theta)),
\]

where

\[
\hat{Q}_n(\theta) := \begin{pmatrix} \hat{\Omega}^{-1/2}_n(\theta) \hat{g}_n(\theta), \hat{D}^*_n(\theta) \end{pmatrix}' \begin{pmatrix} \hat{\Omega}^{-1/2}_n(\theta) \hat{g}_n(\theta), \hat{D}^*_n(\theta) \end{pmatrix} \in \mathbb{R}^{(p+1) \times (p+1)}.
\]

For \( \alpha \in (0, 1) \), the nominal size \( \alpha \) SR-CQLR test rejects \( H_0 : \theta = \theta_0 \) if

\[
\text{SR-QLR}_n(\theta_0) > c_{\hat{r}_n(\theta_0), p}(n^{1/2}\hat{D}^*_n(\theta_0), 1 - \alpha) \quad \text{or} \quad \hat{A}_n^\perp(\theta_0)'\hat{g}_n(\theta_0) \neq 0^{k - \hat{r}_n(\theta_0)}.
\]

The nominal size 100(1 - \( \alpha \))% SR-CQLR CS is

\[
\text{CS}_{\text{SR-CQLR}_n} := \left\{ \theta_0 \in \Theta : \text{SR-QLR}_n(\theta_0) \leq c_{\hat{r}_n(\theta_0), p}(n^{1/2}\hat{D}^*_n(\theta_0), 1 - \alpha) \text{ and } \hat{A}_n(\theta_0)'\hat{g}_n(\theta_0) = 0^{k - \hat{r}_n(\theta_0)} \right\}.
\]

When \( \hat{r}_n(\theta_0) = k \), \( \hat{A}_n(\theta_0) \) is a nonsingular \( k \times k \) matrix. In consequence, by Lemma 5.1, SR-QLR\(_n(\theta_0) = QLR_n(\theta_0) \) and \( c_{\hat{r}_n(\theta_0), p}(n^{1/2}\hat{D}^*_n(\theta_0), 1 - \alpha) = c_{k, p}(n^{1/2}\hat{D}^*_n(\theta_0), 1 - \alpha) \). That is, the SR-CQLR test is the same as the CQLR test defined in Section 5.1. Of course, when \( \hat{r}_n(\theta) < k \), the CQLR test defined in Section 5.1 is not defined, whereas the SR-CQLR test is. Furthermore, if \( \text{rk}(\Omega_{\text{F}_n}(\theta_0)) = k \) for all \( n \) large, then \( \hat{r}_n(\theta_0) = k \) and \( \text{SR-QLR}_n(\theta_0) = QLR_n(\theta_0) \) wp\( \to \) under \{ \( F_n \in \mathcal{F}^{\text{SR}} : n \geq 1 \) \} (by Lemma 5.1 and Lemma 17.1 in the SM). Note that, if \( \hat{r}_n(\theta_0) \leq p \), then the critical value for the SR-CQLR test is the \( 1 - \alpha \) quantile of \( \chi^2_{\hat{r}_n(\theta_0)} \) (because \( Z'Z - \lambda_{\min}((Z, D)'(Z, D)) = Z'Z \sim \chi^2_{\hat{r}_n(\theta_0)} \) in (5.8) when \( r \leq p \)).

Section 20 in the SM provides finite-sample null rejection probabilities of the SR-CQLR test for singular and near singular variance matrices of the moment functions. The results show that singularity and near singularity of the variance matrix does not lead to distorted null rejection probabilities. The method of robustifying the SR-CQLR test to allow for singular variance matrices, which is introduced above, works quite well in the model that is considered.

### 5.3 Computation

The SR-CQLR test is relatively fast to compute. It is found to be 115, 292, and 302 times faster to compute than the PI-CLC, MM1-SU, and MM2-SU tests, respectively, 1.2 times slower to compute than the JVV-CLR and MVW-CLR tests, and 372 and 495 times slower to compute than the LM and AR tests in the linear IV scenarios described in the Introduction. The SR-CQLR

---

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

8 By definition, if \( \hat{r}_n(\theta_0) = k \), the condition \( \hat{A}_n(\theta_0)'\hat{g}_n(\theta_0) = 0^{k - \hat{r}_n(\theta_0)} \) holds.
test is found to be noticeably easier to implement than the PI-CLC, MM1-SU, and MM2-SU tests and comparable to the JVW-CLR and MVW-CLR tests, in terms of the choice of implementation parameters (see Section 14.2 in the SM for details) and the robustness of the results to these choices.

The computation time of the SR-CQLR test increases relatively slowly with \( k \) and \( p \). For example, the times (in minutes) to compute the SR-CQLR test 5000 times (using 5000 critical value repetitions) for \( k = 8 \) and \( p = 1, 2, 4, 8 \) are .26, .49, 1.02, 2.46. The times for \( p = 1 \) and \( k = 1, 2, 4, 8, 16, 32, 64, 128 \) are .14,.15,.18,.26,.44,.99, 2.22, 7.76. The times for \((k, p) = (64, 8)\) and \((128, 8)\) are 14.5 and 57.9. Hence, computing tests for large values of \((k, p)\) is quite feasible. These times are for linear IV regression models, but they are the same for any model, linear or nonlinear, when one takes as given the sample moment vector and sample Jacobian matrix. Note that most of the computation time for the SR-CQLR test is due to the computation of its conditional critical values.

In contrast, computation of the PI-CLC, MM1-SU, and MM2-SU tests can be expected to increase very rapidly in \( p \). The computation time of the PI-CLC test can be expected to increase in \( p \) proportionally to \( n_\theta^p \), where \( n_\theta \) is the number of points in the grid of alternative parameter values for each component of \( \theta = (\theta_1, \ldots, \theta_p)' \), which are used to assess the minimax regret criterion. We use \( n_\theta = 41 \) in the simulations reported above. Hence, the computation time for \( p = 3 \) should be 1681 times longer than for \( p = 1 \). The MM1-SU and MM2-SU tests are not defined in Moreira and Moreira (2015) for \( p > 1 \), but doing so should be feasible. However, even for \( p = 2 \), one would obtain an infinite number of constraints on the directional derivatives to impose local unbiasedness, in contrast to the \( k \) constraints required when \( p = 1 \). In consequence, computation of the MM1-SU and MM2-SU tests can be expected to be challenging when \( p \geq 2 \).

6 Asymptotic Size

6.1 Definitions of Asymptotic Size and Similarity

Let \( RP_n(\theta_0, F, \alpha) \) denote the null rejection probability of a nominal size \( \alpha \) test with sample size \( n \) when the null distribution of the data is \( F \). The asymptotic size of the test for a null parameter space \( \mathcal{F}(\theta_0) \) is

\[
\text{AsySz} := \limsup_{n \to \infty} \sup_{F \in \mathcal{F}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{6.1}
\]

The test is asymptotically similar (in a uniform sense) to a null parameter space \( \mathcal{F}(\theta_0) \) if

\[
\liminf_{n \to \infty} \inf_{F \in \mathcal{F}(\theta_0)} RP_n(\theta_0, F, \alpha) = \limsup_{n \to \infty} \sup_{F \in \mathcal{F}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{6.2}
\]
The asymptotic size of a CS obtained by inverting tests of $H_0: \theta = \theta_0$ for the parameter space $\mathcal{F}_\Theta := \{(F, \theta_0): F \in \mathcal{F}(\theta_0), \theta_0 \in \Theta\}$ is $\text{AsySz} := \lim \inf_{n \to \infty} \inf_{(F, \theta_0) \in \mathcal{F}_\Theta} (1 - R P_n(\theta_0, F, \alpha))$. The CS is asymptotically similar (in a uniform sense) for $\mathcal{F}_\Theta$ if $\lim \inf_{n \to \infty} \inf_{(F, \theta_0) \in \mathcal{F}_\Theta} (1 - R P_n(\theta_0, F, \alpha)) = \lim \sup_{n \to \infty} \sup_{(F, \theta_0) \in \mathcal{F}_\Theta} (1 - R P_n(\theta_0, F, \alpha))$. Asymptotic size and similarity of a CS require uniformity over the null values $\theta_0 \in \Theta$, as well as uniformity over null distributions $F$ for each null value $\theta_0$.

With the SR-AR and SR-CQLR CS’s, this additional level of uniformity does not cause complications. The same proofs for tests deliver results for CS’s with very minor adjustments.

### 6.2 Identification Categories

To determine the asymptotic size of a test (or CS), one needs to determine the test’s asymptotic null rejection probabilities under sequences that exhibit: (i) standard weak, (ii) nonstandard weak, (iii) semi-strong, and (iv) strong identification, as defined immediately below.\(^9\)

Let $\{s_{jF} : j \leq p\}$ denote the singular values of $\Omega_{F}^{-1/2}(\theta_0)E_F G_i(\theta_0)$ in nonincreasing order (when $\Omega_{F}(\theta_0)$ is nonsingular).\(^10\) For a sequence of distributions $\{F_n : n \geq 1\}$, we say that the parameter $\theta_0$ is: (i) weakly identified in the standard sense if $\lim n^{1/2}s_{1F_n} < \infty$, (ii) weakly identified in the nonstandard sense if $\lim n^{1/2}s_{pF_n} < \infty$ and $\lim n^{1/2}s_{1F_n} = \infty$, (iii) semi-strongly identified if $\lim n^{1/2}s_{pF_n} = \infty$ and $s_{pF_n} = 0$, and (iv) strongly identified if $\lim s_{pF_n} > 0$. For sequences $\{F_n : n \geq 1\}$ for which the previous limits exist (and may equal $\infty$), these categories are mutually exclusive and exhaustive. We say that the parameter $\theta_0$ is weakly identified if $\lim n^{1/2}s_{pF_n} < \infty$, which is the union of the standard and nonstandard weak identification categories. The asymptotics considered in Staiger and Stock (1997) are of the standard weak identification type. The nonstandard weak identification category can be divided into two subcategories: some weak/some strong identification and joint weak identification, see AG1 for details. The asymptotics considered in Stock and Wright (2000) are of the some weak/some strong identification type. For example, joint weak identification occurs in a linear IV model with $p > 1$ when the reduced-form coefficient matrix converges to a matrix of ones.

The SR-CQLR statistic has a $\chi^2_p$ asymptotic null distribution under strong and semi-strong identification and a noticeably more complicated asymptotic null distribution under weak identification. Standard weak identification sequences are relatively easy to analyze asymptotically because all $p$ of the singular values are $O(n^{-1/2})$. Nonstandard weak identification sequences are much more difficult to analyze asymptotically because the $p$ singular values have different orders of

---

\(^9\)As used in this paper, the term “identification” means “local identification.” It is possible for a value $\theta \in \Theta$ to be “strongly identified,” but still be globally unidentified if there exist multiple solutions to the moment functions. The asymptotic size and similarity results given below do not rely on local or global identification.

\(^10\)The definitions of the identification categories when $\Omega_F(\theta_0)$ may be singular, as is allowed in this paper, is somewhat more complicated than the definitions given here.
magnitude. This affects the asymptotic properties of both the test statistics and the conditioning statistics. Contiguous alternatives $\theta$ are at most $O(n^{-1/2})$ from $\theta_0$ when $\theta_0$ is strongly identified, but more distant when $\theta_0$ is semi-strongly or weakly identified. Typically the parameter $\theta$ is not consistently estimable when it is weakly identified.

6.3 Asymptotic Size Results

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

**Theorem 6.1** The asymptotic sizes of the SR-AR and SR-CQLR tests defined in (4.7) and (5.13), respectively, equal their nominal size $\alpha \in (0, 1)$ for the null parameter spaces $\mathcal{F}^{SR}_{AR}$ and $\mathcal{F}^{SR}$, 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 SR-AR and SR-CQLR CS’s for the parameter spaces $\mathcal{F}^{SR}_{\Theta,AR}$ and $\mathcal{F}^{SR}_{\Theta}$.

**Comments:** (i) For distributions $F$ under which $g_i = 0^k$ a.s., the SR-AR and SR-CQLR tests reject 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) SR-LM versions of Kleibergen’s LM test and CS are defined in Section 23 in the SM. However, as discussed there, these procedures are only partially singularity robust.

7 Asymptotic Efficiency of the SR-CQLR Test under Strong and Semi-Strong Identification

Next, we show that the SR-CQLR 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). By this we mean that it is asymptotically equivalent (under the null and contiguous alternatives) to a Wald test constructed using an asymptotically efficient GMM estimator and to the standard GMM LM test, see Newey and West (1987). More specifically, we consider drifting sequences $\{\lambda_{n,h}^*: n \geq 1\}$ of data-generating processes taken from $\mathcal{F}^{SR}$ in (3.6) that correspond to strong or semi-strong identification and establish that the SR-CQLR test statistic equals the standard GMM LM test statistic up to a $o_p(1)$ term and that the conditional critical value of the SR-CQLR test converges in probability to $\chi^2_{p,1-\alpha}$.

Kleibergen’s LM statistic and the standard GMM LM statistic are defined by

$$LM_n := n\tilde{g}_n^{1/2}\tilde{\Omega}_n^{-1/2}P_{\tilde{\delta}_n^{1/2}}\tilde{\delta}_n^{-1/2}\tilde{g}_n$$

and

$$LM_n^{GMM} := n\tilde{g}_n^{1/2}\tilde{\Omega}_n^{-1/2}P_{\tilde{\delta}_n^{1/2}}\tilde{\delta}_n^{-1/2}\tilde{g}_n,$$

(7.1)
respectively, where $\tilde{G}_n$ is the sample Jacobian defined in [4.1] with $\theta = \theta_0$ and $P_A$ denotes the projection matrix onto the column space of the matrix $A$ (i.e., $P_A = A(A' A)^{-1} A'$ when $A$ is full column rank). The critical value for the $LM_n$ and $LM_n^{GMM}$ tests is $\chi^2_{p, 1-\alpha}$, the $1-\alpha$ quantile of the $\chi^2_p$ distribution. The test based on $LM_n^{GMM}$ is asymptotically equivalent to the Wald test based on an asymptotically efficient GMM estimator under (i) strong identification (which requires $k \geq p$), (ii) nonsingular moments-variance matrices (i.e., $\lambda_{\min}(\Omega_{F_n}) \geq \delta > 0$ for all $n \geq 1$), and (iii) a null parameter value that is not on the boundary of the parameter space, see Newey and West (1987).

This also holds true under semi-strong identification (which also requires $k \geq p$). For example, Theorem 5.1 of Andrews and Cheng (2013) shows that the Wald statistic for testing $H_0 : \theta = \theta_0$ based on a GMM estimator with asymptotically efficient weight matrix has a $\chi^2_p$ distribution under semi-strong identification. This Wald statistic can be shown to be asymptotically equivalent to the $LM_n^{GMM}$ statistic under semi-strong identification. (For brevity, we do not do so here.)

Suppose $k \geq p$. The drifting sequences $\{\lambda_{n,h}^* : n \geq 1\}$ referred to above are rather complicated and so, for brevity, we define them at the beginning of Section 28 in the SM. They are defined so that various population quantities that affect the asymptotic distributions of the SR-CQLR test statistic and critical value converge as $n \to \infty$. 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 $n \geq 1$. Most importantly, we have that, along $\{\lambda_{n,h}^* : n \geq 1\}$, $n^{1/2}(s_1 F_n, \ldots, s_p F_n)$ converges to some vector $(h_{1,1}^*, \ldots, h_{1,p}^*)$ whose elements may be finite or infinite, where $(s_1 F_n, \ldots, s_p F_n)$ denote the singular values of the population Jacobian $E_F G_i \in \mathbb{R}^{k \times p}$. Strong or semi-strong identification occurs if the smallest singular value of $E_F G_i$ diverges to infinity after renormalization by $n^{1/2}$, i.e., if $h_{1,p}^* = \infty$.

**Theorem 7.1** Suppose $k \geq p$. For any sequence $\{\lambda_{n,h}^* \in \Lambda^* : n \geq 1\}$ that exhibits strong or semi-strong identification (where the latter and $\Lambda^*$ are defined precisely in Section 28 in the SM), we have

(a) $SR$-$QLR_n = QLR_n + o_p(1) = LM_n + o_p(1) = LM_n^{GMM} + o_p(1)$ and

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

**Comment:** Theorem 7.1 establishes the asymptotic efficiency (in a GMM sense) of the SR-CQLR test under strong and semi-strong identification. Note that Theorem 7.1 provides asymptotic equivalence results under the null hypothesis, but, by the definition of contiguity, these asymptotic equivalence results also hold under contiguous local alternatives.
8 Empirical Application

In this section, we use the AR and CQLR type tests introduced above to do inference on the elasticity of intertemporal substitution (EIS) in consumption. We follow the analysis in Yogo (2004) based on data used in Campbell (2003). Specifically, consider the regression model

\[
\Delta c_{i+1} = \tau + \psi r_{i+1} + \xi_{i+1} \quad \text{for } i = 1, \ldots, n, \tag{8.1}
\]

where \(\tau\) is a constant, \(\psi\) denotes EIS, \(\Delta c_{i+1}\) is consumption growth at time \(i + 1\), \(r_{i+1}\) is the real return on an asset at time \(i + 1\), and \(\xi_{i+1}\) is the error term that is correlated with the regressor. (Note that Yogo (2004) uses a subscript \(t\) rather than \(i\):) To identify EIS, we use a vector \(Z_i \in R^4\) of IV’s consisting of the nominal interest rate, inflation, consumption growth, and log dividend-price ratio, all of which are lagged twice and then satisfy \(E(Z_i \xi_{i+1}) = 0^4\). We also consider the reversed form of (8.1):

\[
r_{i+1} = \mu + (1/\psi) \Delta c_{i+1} + \eta_{i+1} \quad \text{for } i = 1, \ldots, n, \tag{8.2}
\]

where \(\mu\) is a constant and \(\eta_{i+1}\) is the error term, and exploit \(E(Z_i \eta_{i+1}) = 0^4\) to do inference on \(1/\psi\).

Classical inference methods lead to the empirical puzzle that \(\psi\) is found to be significantly less than one but \(1/\psi\) is not found to be significantly different from one. Yogo (2004) addresses this puzzle by applying identification-robust methods. His findings based on the data in Campbell (2003) suggest that \(\psi\) is significantly less than one and not significantly different from zero. The magnitude of \(\psi\) is of economic importance because, as summarized in Yogo (2004), if \(\psi < 1\) (\(\psi > 1\)) then an investor’s optimal consumption-wealth ratio is increasing (decreasing) in expected returns. The analysis based on the AR and CQLR procedures introduced in this paper support the main conclusion in Yogo (2004).

We first replicate the identification-robust inference results in Tables 3, 5, and 6 from Yogo (2004) based on the homoskedastic versions of the AR, LM, and CLR tests (see (25)-(27) in Yogo (2004)) and the heteroskedasticity-robust S-test of Stock and Wright (2000) (see (30) in Yogo (2004)). We then add the new SR-AR and SR-CQLR tests defined in (4.7) and (5.13), respectively, that only impose quite weak restrictions on the parameter space, namely uniform bounds on the moment functions and its derivative, in order to have correct asymptotic size; recall the discussion above regarding the parameter space in (3.6) for the SR-AR and SR-CQLR tests. In particular, heteroskedasticity is allowed. In all of the examples considered here, the estimator of the variance matrix of the moments defined in (4.1) is nonsingular, and therefore, those tests simplify to the ones
defined in (4.2) and (5.9), the first of which is similar to the S-test of Stock and Wright (2000) (see (30) in Yogo (2004)), but differs because we use the recentered estimator of the variance matrix, see (4.1).

We calculate 95% confidence intervals for \( \psi \) and \( 1/\psi \) (i.e. \( \alpha = .05 \)) by collecting the values of \( \theta_0 = \psi \) for which the null hypothesis in (3.3) is not rejected at the 5% nominal size. To do so, we use a grid of null values with stepsize .001 in \([-200, 200]\) and also consider the additional null values \( \pm 500 \) and \( \pm 1000 \) (and in some cases larger values).

To implement our procedures, first premultiply (8.1) and (8.2) by \( M_{1^n} = I_n - P_{1^n} \), where \( 1^n \in \mathbb{R}^n \) denotes a vector of ones, to eliminate the constant term from the regression. Denote by \( Z \in \mathbb{R}^{n \times 4} \) the IV matrix with rows given by \( Z_i' \) for \( i = 1, ..., n \), and define analogously the vectors \( \Delta c \) and \( r \in \mathbb{R}^n \).

Then, define
\[
g_i(\theta) = (M_{1^n}(\Delta c - \psi r))_i(M_{1^n}Z)'_i \in \mathbb{R}^4
\]
in the case of regression (8.1) (and analogously \( g_i(\theta) = (M_{1^n}(r - (1/\psi)\Delta c))_i(M_{1^n}Z)'_i \in \mathbb{R}^4 \) in the case of regression (8.2), where \( \theta = \psi \) (or \( \theta = 1/\psi \)). We then obtain
\[
G_i(\theta) = G_i = -(M_{1^n}r)_i(M_{1^n}Z)'_i \in \mathbb{R}^4
\]
for the Jacobian defined in (3.2) (and analogously \( G_i = -(M_{1^n}\Delta c)_i(M_{1^n}Z)'_i \) in the case of regression (8.2)).

Note that in the regression models considered here the dimension of the parameter of interest equals one, i.e. \( p = 1 \). Next calculate, the quantities \( \hat{g}_n(\theta) \) and \( \Omega_n^{-1}(\theta) \) in (4.1), \( \hat{G}_{1n}, \hat{\Gamma}_{1n}(\theta) \), and \( \hat{D}_{1n}(\theta) \) defined in (5.2), and \( f_i(\theta), \hat{f}_n(\theta), \hat{V}_n(\theta), \) and \( \hat{R}_n(\theta) \) defined in (5.3), with all of these quantities evaluated with \( \theta \) equal to \( \theta_0 = \psi \). Then, calculate \( \hat{\Sigma}_n(\theta) \) and its eigenvalue adjusted version \( \hat{\Sigma}_n^*(\theta) \), see (5.5) and (5.6). For the output below we use \( \varepsilon = .01 \). (We also calculated CI’s for \( \varepsilon = .05 \) and .001, which led to identical results for the case of the real asset being \( r_i = r_{f,i} \), defined below, and comparable results for the case of \( r_i = r_{e,i} \), defined below.) Finally, calculate the quantities \( \hat{L}_n(\theta), \hat{D}_{n}^{*}(\theta) \), and \( \hat{Q}_n(\theta) \), the test statistic \( QLR_n(\theta) \) defined in (5.7), and the test statistic \( AR_n(\theta) \) in (4.2).

The critical value for the AR test is the \( \chi^2_{4,1-\alpha} \) quantile given that there are four instruments. The critical value for the CQLR test is obtained by simulation. Specifically, we generate 10,000 draws from a \( N(0^4, I_4) \) distribution and for each draw we calculate \( CLR_{k,p}(n^{1/2}\hat{D}_{n}^{*}(\theta_0)) \) defined in (5.8). The critical value of the CQLR test is then defined as the \( 1 - \alpha \) sample quantile of these observations, which is denoted by \( c_{k,p}(n^{1/2}\hat{D}_{n}^{*}(\theta_0), 1 - \alpha) \).
The data set from Campbell (2003) employed here consists of quarterly data for the following eleven developed countries: Australia, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland, United Kingdom (U.K.), and the United States (U.S.). The sample period varies across different countries with sample sizes equal to 114, 115, 113, 79, 106, 114, 86, 116, 91, 115, and 114, respectively. For \( r_f \), two candidates for asset returns are used, namely, the real interest rate and the real aggregate stock return, denoted by \( r_{f,i} \) and \( r_{e,i} \), respectively. See Yogo (2004, Section IV. A., p. 803) for details on the data and the precise definition of the variables.

Table I reports the results based on the real interest rate \( r_{f,i} \), whereas Table II reports the results based on the real aggregate stock return \( r_{e,i} \). In each case, we report the CQLR and AR CI’s for \( \psi \) and \( 1/\psi \) based on the regressions (8.1) and (8.2). If a CI contains the right endpoint of the search interval, namely \( \psi \), and analogously for the left endpoint.

We now discuss the findings for \( \psi \) and the implications for the equity premium puzzle obtained from the new inference procedures\(^{11}\). We start with the results based on \( r_{f,i} \). Yogo (2004, p.

\[^{11}\text{Our analysis reveals certain discrepancies with the results reported in Yogo (2004). Namely, the CI’s for } \psi \text{ using } r_{f,i} \text{ based on the LM test (see (26) in Yogo (2004)) are as follows: Australia } [-22.27,5.13,13.74] \text{ by our calculations versus (vs.) } [-22.13,74] \text{ in Table 3 in Yogo (2004), Canada } [-73.02,3.9,14.16] \text{ vs. } [-73.14,15.1, France } [-50.06,36.28,4.7,31] \text{ vs. } [-47.31], \text{ Germany } [-1.21,16.1,11.3] \text{ vs. } [-1.21,26.1], \text{ Italy } [-6.51,3.8,9] \text{ vs. } [-24.11], \text{ Japan } [-24.1,11], \] Sweden } [-59.26,12.1,21.2] \text{ vs. } [-59.26,12.1,21.1], \text{ Netherlands } [-59.26,12.1,21.1] \text{ vs. } [-59.26,12.1,21.2], \text{ Switzerland } [-1.19,0.7,19.9] \text{ vs. } [-1.19,0.7,19.9], \] U.K. } [-1.19,0.7,19.9] \text{ vs. } [-1.19,0.7,19.9], \text{ and U.S. } [-27.86,12.1,21.1] \text{ vs. } [-27.86,12.1,21.2]. \] Finally, the CI’s for \( \psi \) using \( r_{e,i} \) based on the CLR test (see (27) in Yogo (2004)) for the U.S. is } (-\infty,0.01,0.04) \text{ by our calculations vs. } (-\infty,\infty) \] in Yogo (2004).\)
806) concludes from the CI's based on the homogenous CLR test that the “EIS is small and not significantly different from 0 for the eleven developed countries.” This finding is supported also by the new results based on the CQLR test except for the Netherlands (where the CI equals $[-.72, 1.79]$). All eleven CI's contain zero, and nine CI's are bounded from above by .54 (with the exceptions of Germany, where the right endpoint of the CI equals .90, and the Netherlands). The finding is also supported by the CI's from the new AR test for almost all countries, with the exception of Germany (where the right endpoint of the CI is 1.28), the Netherlands (where 0 is not included in the CI), and the U.S. (where the CI is empty). The results based on the CQLR (and also the new AR) CI's are consistent across the regressions (8.1) and (8.2) and the empirical puzzle based on classical (identification non-robust) inference procedures does not occur here. In particular, the left endpoints of the positive portions of the CI’s for $1/\psi$ based on the CQLR test equal 2.9, 4.8, 6.1, 1.1, 9.6, 3.5, .56, 5.0, 5.5, 1.9, and 2.0 for the eleven countries, respectively, which translate into right endpoints of the positive portion of CI's for $\psi$ of .34, .21, .16, .91, .10, .29, 1.79, .20, .18, .53, and .50, respectively. The actual right endpoints of the positive portion of the CI’s for $\psi$ based on the CQLR test equal .34, .21, .16, .90, .10, .29, 1.79, .20, .18, .54, and .49, respectively!

Comparing the CI's based on the new AR and CQLR tests for $\psi$, we find that for Australia, Canada, the Netherlands, the U.K., and the U.S. the former are shorter, while for the other countries the latter are shorter. In fact, for the U.S., the CI from the AR procedure is empty, which points to model misspecification.

Next, we discuss the findings when $r_{e,i}$ is used. Inference on $\psi$ and $1/\psi$ based on $r_{e,i}$ is completely uninformative for both the CQLR and AR CI's for Australia, Germany, Italy, Sweden, and Switzerland with CI's all equal to $(-\infty, \infty)$, see Table II. Inference is also relatively uninformative for all of the other countries, with unbounded CI's for all countries except for Canada, the Netherlands, and the US. And even in the latter three cases, the CI's are too wide to provide information of economic interest. These results are mostly consistent with the findings based on the homoskedastic versions of the AR, LM, and CLR tests that also produce unbounded CI's in almost all cases, see Yogo (2004, Table 5). However, Canada, France, and Japan are three exceptions for which, based on these homoskedastic tests, informative CI's are obtained that imply a small value of $\psi$. It may be the case that the discrepancies between the results based on the new CI’s and those based on the homoskedastic AR, LM, and CLR CI's for these countries are a consequence of undercoverage of the latter CI's because the actual DGP may not satisfy the assumptions necessary for validity of these CI's, such as homoskedasticity.

Note that for the countries where the new CI’s are not equal to $(-\infty, \infty)$, there is complete
Table II. CQLR and AR CI’s for EIS, $\psi$, and its inverse, $1/\psi$, using $r_{e,i}$ with $\varepsilon = .01$

| Country    | CQLR          | AR            |
|------------|---------------|---------------|
| Australia  | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Canada     | $(-\infty, -1.33] \cup [.017, \infty)$ | $(-\infty, -35] \cup [-.01, \infty)$ |
| France     | $(-\infty, .04] \cup [.63, \infty)$ | $(-\infty, .07] \cup [.46, \infty)$ |
| Germany    | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Italy      | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Japan      | $(-\infty, -.336] \cup [-.334, -.333] \cup [-.06, \infty)$ | $(-\infty, -.66] \cup [-.06, \infty)$ |
| Netherlands| $(-\infty, -.002] \cup [.05, \infty)$ | $(-\infty, -.01] \cup [.02, \infty)$ |
| Sweden     | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Switzerland| $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| U.K.       | $(-\infty, -.01] \cup [.048, \infty)$ | $(-\infty, -.01] \cup [.07, \infty)$ |
| U.S.       | $(-\infty, -.01] \cup [.048, \infty)$ | $(-\infty, -.01] \cup [.07, \infty)$ |

| Country    | CQLR          | AR            |
|------------|---------------|---------------|
| Australia  | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Canada     | $[-.75, 60.6]$ | $(-\infty, -182.1] \cup [-2.9, \infty)$ |
| France     | $(-\infty, 1.58] \cup [24.75, \infty)$ | $(-\infty, 2.16] \cup [14.97, \infty)$ |
| Germany    | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Italy      | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Japan      | $(-\infty, -15.8] \cup [-2.994, -2.99] \cup [-2.97, \infty)$ | $(-\infty, -15.7] \cup [-1.5, \infty)$ |
| Netherlands| $[-656.97, -609.34] \cup [-484.1, 20.9]$ | $[-67.27, 51.98]$ |
| Sweden     | $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| Switzerland| $(-\infty, \infty)$ | $(-\infty, \infty)$ |
| U.K.       | $(-\infty, \infty)$ | $(-\infty, 24.4] \cup [509.1, \infty)$ |
| U.S.       | $[-135.01, 21.03]$ | $[-159.57, 13.93]$ |
consistency between the CI’s for $\psi$ and $1/\psi$, analogous to the findings reported in Table I. For example, the CQLR CI for $1/\psi$ has right endpoint equal to 60.6 implying that any positive value of $\psi$ should be contained in $[0.0165, \infty)$. And indeed, the positive portion of the CQLR type CI for $\psi$ is reported as $[0.017, \infty)$. Analogous statements are obtained for the CQLR CI for $1/\psi$, e.g., see the results for France, the Netherlands, and the U.S. In summary, the CI’s based on the new tests reveal that $\psi$ is very weakly identified (or perhaps unidentifiable) when one uses $r_{e,i}$. Unlike CI’s based on a classical inference procedure, such as a $t$-test based CI, the identification-robust results based on the regressions (8.1) and (8.2) are internally consistent.

9 Subvector Inference

We now consider subvector inference based on the AR and CQLR tests under the assumption that the parameters not under test are strongly identified. For brevity, in this section, we assume that the variance matrix of the moment functions evaluated at the true parameters has minimal eigenvalue bounded away from zero. This assumption is eliminated in Section 13 in the SM.

The extension to subvector SR-AR and SR-CQLR tests is analogous to the extension of the full vector tests described in Sections 4 and 5.2. Hence, for brevity, these extensions are given in the SM, see Section 13.

9.1 Model and Hypotheses

The model is

$$E_Fg(W_i, \theta, \beta) = 0^k,$$

(9.1)

where the equality holds when $\eta := (\theta', \beta')' \in \Theta \times B$ is the true value. Here, $\Theta \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^b$ denote the parameter spaces for $\theta$ and $\beta$, respectively, with $p, b \geq 1$ and $k - b \geq 1$. We allow for the possibility that $k - b < p$.

We are concerned with tests of the null hypothesis

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0$$

(9.2)

in the presence of the nuisance parameter $\beta$ and with confidence sets for $\theta$ obtained by inverting the tests.

The first- and second-order partial derivatives of $g(W_i, \eta)$ with respect to $\theta$ and $\beta$ are denoted

Electronic copy available at: https://ssrn.com/abstract=3366443
by
\[ G(W_i, \eta) := \frac{\partial}{\partial \theta} g(W_i, \eta) \in R^{k \times p}, \quad G_\beta(W_i, \eta) := \frac{\partial}{\partial \beta} g(W_i, \eta) \in R^{k \times b}, \]
\[ G_{\theta_j \beta}(W_i, \eta) := \frac{\partial^2}{\partial \theta_j \partial \beta} g(W_i, \eta) \in R^{k \times b} \text{ for } j = 1, \ldots, p, \tag{9.3} \]
and likewise for other expressions, such as \( G_{\beta_j \beta}(W_i, \eta) \). Let \( g_i(\eta) := g(W_i, \eta) \) and \( \hat{g}_n(\eta) := n^{-1} \sum_{i=1}^n g_i(\eta) \) and likewise for other quantities, e.g.,
\[ \hat{G}_{\beta n}(\eta) := n^{-1} \sum_{i=1}^n G_{i \beta}(\eta). \tag{9.4} \]
We use the notation \( G_{ij} := \frac{\partial}{\partial \theta_j} g(W_i, \eta_0) \in R^k \) for \( j = 1, \ldots, p \), where \( \eta_0 := (\theta_0', \beta^*')' \) and \( \beta^* \) denotes the true value of \( \beta \), and likewise for other quantities. For example, \( G_{i0 \beta} := \frac{\partial^2}{\partial \theta_i \partial \beta} g(W_i, \eta_0) \), \( G_{i \beta j \beta} := \frac{\partial^2}{\partial \beta_i \partial \beta_j} g(W_i, \eta_0) \), and \( g_i := g_i(\eta_0) \).

### 9.2 Subvector Tests for Nonsingular Moment Variance Matrices

#### 9.2.1 Definitions of the Subvector Tests

Define
\[ \hat{\Omega}_n(\eta) := n^{-1} \sum_{i=1}^n g_i(\eta)g_i(\eta)' - \hat{\beta}_n(\eta)\hat{\beta}_n(\eta)'. \tag{9.5} \]
Let \( \hat{\beta}_n = \hat{\beta}_n(\theta_0) \) denote the null-restricted two-step GMM estimator of \( \beta \). That is,
\[ \hat{\beta}_n := \arg \min_{\beta \in B} ||\hat{\varphi}_n \hat{g}_n(\theta_0, \beta)||^2, \text{ where } \hat{\varphi}_n \in R^{k \times k}, \hat{\varphi}_n' \hat{\varphi}_n = \hat{\Omega}_n^{-1}(\theta_0, \hat{\beta}_n), \tag{9.6} \]
\( \hat{\beta}_n \) is a solution to \( \text{(9.6)} \) with \( \hat{\varphi}_n \) replaced by \( I_k \). Rather than using the null-restricted two-step GMM estimator \( \hat{\beta}_n \), one could employ the null-restricted continuous-updating estimator of \( \beta \) (e.g., as suggested in Kleibergen (2005)). The same asymptotic results as below would be obtained.

Following Kleibergen’s (2005) approach for the Jacobian, as in (5.2), we now introduce "orthogonalized" estimators of \( E_F g_i g_i' \) and \( E_F G_{ij \beta} \) whose asymptotic distributions are designed to be independent of \( \hat{g}_n^S \), which denotes the asymptotic distribution of \( n^{1/2} \hat{g}_n(\theta_0, \hat{\beta}_n) \), see Lemma 31.5 in the SM. In particular, we do not estimate \( E_F g_i g_i' \) by \( \hat{\Omega}_n(\theta_0, \hat{\beta}_n) \). Rather, we estimate it by
\[ \widetilde{\Omega}_n(\theta_0, \widehat{\beta}_n), \]

where

\[
\tilde{\Omega}_n(\eta) := (\tilde{\Omega}_{1n}(\eta), \ldots, \tilde{\Omega}_{kn}(\eta)) \in R^{k \times k},
\]

\[
\tilde{\Omega}_{jn}(\eta) := n^{-1} \sum_{i=1}^{n} g_i(\eta)g_{ij}(\eta) - \tilde{\Phi}_{jn}(\eta)\tilde{\Omega}_{jn-1}(\eta)\tilde{g}_n(\eta) - \tilde{\gamma}_n(\eta)\tilde{\gamma}_{jn}(\eta) \in R^k, \text{ and}
\]

\[
\tilde{\Phi}_{jn}(\eta) := n^{-1} \sum_{i=1}^{n} \left( g_i(\eta)g_{ij}(\eta) - n^{-1} \sum_{s=1}^{n} (g_s(\eta)g_{sj}(\eta)) \right) g_i(\eta) = \tilde{R}^k \times k \text{ for } j = 1, \ldots, k, \quad (9.7)
\]

where \( \tilde{g}_n(\eta) = (\tilde{g}_{1n}(\eta), \ldots, \tilde{g}_{kn}(\eta))' \). Although it may not be obvious from the expression in (9.7), \( \tilde{\Omega}_n(\eta) \) is symmetric, as desired.

Likewise, we do not estimate \( E_FG_{i\beta} \) by \( \tilde{G}_{\beta n}(\theta_0, \widehat{\beta}_n) \). We estimate it by \( \tilde{G}_{\beta n}(\theta_0, \widehat{\beta}_n) \), where

\[
\tilde{G}_{\beta n}(\eta) := (\tilde{G}_{\beta_{1n}}(\eta), \ldots, \tilde{G}_{\beta_{bn}}(\eta)) \in R^{k \times b},
\]

\[
\tilde{G}_{\beta_{jn}}(\eta) := n^{-1} \sum_{i=1}^{n} G_{i\beta_j}(\eta) - \tilde{F}_{jn}(\eta)\tilde{\Omega}_{jn-1}(\eta)\tilde{g}_n(\eta) \in R^k, \text{ and}
\]

\[
\tilde{F}_{jn}(\eta) := n^{-1} \sum_{i=1}^{n} (G_{i\beta_j}(\eta) - \tilde{G}_{\beta_{jn}}(\eta))g_i(\eta)' \in R^k, \text{ where } \tilde{G}_{\beta_{jn}}(\eta) := n^{-1} \sum_{i=1}^{n} G_{i\beta_j}(\eta), \quad (9.8)
\]

for \( j = 1, \ldots, b \).

We define the following estimator \( \tilde{J}_n(\theta_0, \widehat{\beta}_n) \) of \( (E_Fg_{i\beta})^{-1/2}E_FG_{i\beta} \), which is designed to be asymptotically independent of \( \tilde{\pi}_n^S \). Let

\[
\tilde{J}_n(\eta) := \tilde{\Omega}_n^{-1/2}(\eta)\tilde{G}_{\beta n}(\eta) \in R^{k \times b}. \quad (9.9)
\]

For any matrix \( A \) with \( k \) rows, let \( M_A = I_k - P_A \), where \( P_A \) denotes the projection matrix onto the column space of \( A \).

The subvector AR test statistic is

\[
AR_n^S(\eta) := n\tilde{g}_n(\eta)'\tilde{\Omega}_n^{-1/2}(\eta)M_{\tilde{J}_n}(\eta)\tilde{\Omega}_n^{-1/2}(\eta)\tilde{g}_n(\eta). \quad (9.10)
\]

The superscript \( S \) denotes "subvector". The nominal size \( \alpha \) subvector AR test (without singularity adjustment) rejects \( H_0 \), specified in (9.2), when \( AR_n^S(\theta_0, \widehat{\beta}_n) > \chi_{k-b,1-\alpha}^2 \).

The subvector QLR test statistic \( QLR_n^S(\theta_0, \widehat{\beta}_n) \) is defined as the full vector statistic is defined in (5.7), but with \( (\theta, \widehat{\beta}_n) \) in place of \( \theta, \tilde{\Omega}_n^{-1/2} \) in place of \( \tilde{\Omega}_n^{-1/2} \), and the projection matrix \( M_{\tilde{J}_n}(\theta, \widehat{\beta}_n) \) inserted as a weight matrix. In particular, let \( \tilde{D}_n(\theta, \beta) \in R^{k \times p} \) be defined as \( \tilde{D}_n(\theta) \) is defined in

Electronic copy available at: https://ssrn.com/abstract=3366443
but with \((\theta, \beta)\) in place of \(\theta\). Then, define

\[
QLR_n^S(\theta, \widehat{\beta}_n) := AR_n^S(\theta, \widehat{\beta}_n) - \lambda_{\min}(n\widehat{Q}_n^S(\theta, \widehat{\beta}_n)), \tag{9.11}
\]

where

\[
\widehat{Q}_n^S(\eta) := \left(\widehat{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta), \widehat{D}_n^*(\eta)\right)' M_{\widehat{J}_n}(\eta) \left(\widehat{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta), \widehat{D}_n^*(\eta)\right) \in R^{(p+1)\times(p+1)},
\]

\[
\widehat{D}_n^*(\eta) := \widehat{\Omega}_n^{-1/2}(\eta)\widehat{D}_n(\eta)\widehat{L}_n^{1/2}(\eta) \in R^{k\times p}, \quad \widehat{L}_n(\eta) := (\theta, I_p)(\widehat{\Sigma}_n^e(\eta))^{-1}(\theta, I_p)' \in R^{p\times p},
\]

\(\widehat{\Sigma}_n(\eta) \in R^{(p+1)\times(p+1)}\) is defined as in (5.6) with \(H = \widehat{\Sigma}_n(\eta)\), and \(\widehat{\Sigma}_n(\eta)\) is defined as in (5.3) and (5.5) with \(\eta\) in place of \(\theta\).

Defining \(M_{\widehat{J}_n}(\eta) = I_k\) when \(b = 0\), the definitions of the subvector AR and QLR statistics reduce to the full vector statistics in (5.7), except that they employ \(\widehat{\Omega}_n^{-1/2}\) rather than \(\widehat{\Omega}_n^{-1/2}\)\(^{12}\).

Let \(c_{k,p}(D, J, 1 - \alpha)\) denote the \(1 - \alpha\) quantile of \(CLR_{k,p}(D, J)\), where

\[
CLR_{k,p}(D, J) := Z'M_JZ - \lambda_{\min}((Z, D)'M_J(Z, D)) \text{ and } Z \sim N(0^k, I_k). \tag{9.12}
\]

The conditional critical value of the nominal size \(\alpha\) CQLR test is \(c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0, \widehat{\beta}_n), \widehat{J}_n(\theta_0, \widehat{\beta}_n), 1 - \alpha)\).

The nominal size \(\alpha\) subvector CQLR test rejects the null in (9.2) if

\[
QLR_n^S(\theta_0, \widehat{\beta}_n) > c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0, \widehat{\beta}_n), \widehat{J}_n(\theta_0, \widehat{\beta}_n), 1 - \alpha). \tag{9.13}
\]

### 9.2.2 Asymptotic Size of the Subvector Tests

We make the following assumptions about the function \(g\) and the parameter space \(B\) of \(\beta\). We denote by \(C^j(S)\) the set of \(j\)-times continuously differentiable functions from a set \(S\) into \(R^k\).

**Assumption gB:** (a) For given \(\theta_0\) the domain of \(g\) is \(W \times \{\theta_0\} \times B\), where \(B\) is compact.

(b) \(\forall w \in W, \ g(w, \theta_0, \cdot) \in C^0(B)\).

Note that Assumption gB(a) and gB(b) together imply uniform continuity of \(g(w, \theta_0, \cdot)\) for any given \(w \in W\). We use the latter to prove a uniform law of large numbers via stochastic equicontinuity.

The parameter space \(\mathcal{F}\) in (16.1) needs to be altered from the case of a full vector hypothesis test to the subvector case. Let \(\mu\) denote a probability measure on \(R^m\) for which \(E_\mu \sup_{\beta \in B} ||g_i(\theta, 0, \beta)|| < \infty\), where \(E_\mu\) denotes expectation when \(W_i\) is distributed according to the measure \(\mu\). For \(\theta > 0\) and \(\beta^+ \in R^b\), let

\[
B(\beta^+, \theta) = \{\beta \in R^b : ||\beta^+ - \beta|| < \theta\}.
\]

We abbreviate "absolutely continuous with respect to" by "ac wrt" and "Radon-Nikodym derivative" by "RNd". Next, we define the

\(^{12}\)The reason \(\widehat{\Omega}_n^{-1/2}\) is employed, rather than \(\widehat{\Omega}_n^{-1/2}\), is because \(M_{\widehat{J}_n}(\eta) \neq I_k\) when \(b \geq 1\). When \(b \geq 1\), \(M_{\widehat{J}_n}(\eta)\) has less than full rank and this has consequences for the asymptotic results and their proofs. See the footnote following (31.2) in Section 31 in the SM for details.
null parameter spaces for \((F, \beta^*)\), where \(F\) denotes the distribution of \(W_i\) and \(\beta^*\) denotes the true value of \(\beta\), for the subvector AR and CQLR tests. The following set \(\mathcal{F}^S_{AR,1}\) contains the restrictions needed to guarantee consistency of \(\tilde{\beta}_n\) and \(\tilde{\beta}_n\). Let

\[
\mathcal{F}^S_{AR,1} := \{(F, \beta^*) : E_F g_i = 0^k, \quad F \text{ is ac wrt } \mu \text{ with } \text{RNd } f \text{ satisfying } f \leq M, \\
\inf_{\beta \in B(B(\beta^*, \vartheta))} ||E_F g_i(\theta_0, \beta)||^2 > \delta \zeta, \quad \forall \zeta > 0, \quad E_\mu \sup_{\beta \in B} ||g_i(\theta_0, \beta)|| < \infty, \\
\sup_{\beta \in B} E_F ||g_i(\theta_0, \beta)||^{1+\gamma} \leq M \}
\]  

(9.14)

for constants \(\delta, \gamma > 0\) and \(M < \infty\). Let

\[
\mathcal{F}^S_{AR,2} := \{(F, \beta^*) : B(\beta^*, \vartheta) \subset B, \quad g(w, \theta_0, \cdot) \in C^2(B(\beta^*, \vartheta)) \quad \forall w \in W, \\
E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} ||h_i(\beta)|| \leq M \quad \text{and} \quad \sup_{\beta \in B(\beta^*, \vartheta)} E_F ||h_i(\beta)||^{1+\gamma} \leq M \quad \text{for} \quad h_i(\beta) \in \{||g_i(\theta_0, \beta)||^2, \\
G_{i\beta}(\theta_0, \beta), \quad g_{ij}(\theta_0, \beta)G_{i\beta}(\theta_0, \beta), \quad (\partial^2 / \partial \beta_i \partial \beta^j)g_i(\theta_0, \beta), \quad (\partial^2 / \partial \theta_i \partial \beta^j)g_i(\beta), \\
\lambda_{\min}(E_F g_i g_i') \geq \delta, \quad \tau_{\min}(E_F G_{i\beta}) \geq \delta \}
\]  

(9.15)

for indices \(j = 1, \ldots, k, m = 1, \ldots, b,\) and \(t = 1, \ldots, p,\) and constants \(\vartheta, \delta, \gamma > 0\) and \(M < \infty\), where \(\tau_{\min}(A)\) denotes the smallest singular value of the matrix \(A\).\(^{13}\)

The null parameter space for the subvector AR test is

\[
\mathcal{F}^S_{AR} := \mathcal{F}^S_{AR,1} \cap \mathcal{F}^S_{AR,2}.
\]  

(9.16)

The null parameter space for the subvector CQLR test is

\[
\mathcal{F}^S := \{(F, \beta^*) \in \mathcal{F}^S_{AR} : \max\{E_F ||g_i(\theta_0, \beta^*)||^{4+\gamma}, \quad E_F ||G_{i\beta}(\theta_0, \beta^*)||^{2+\gamma}, \\
E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} ||g_i(\theta_0, \beta)||^3, \quad E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} ||G_i(\theta_0, \beta)||^2, \\
\sup_{\beta \in B(\beta^*, \vartheta)} E_F ||g_i(\theta_0, \beta)||^{3+\gamma}, \quad \sup_{\beta \in B(\beta^*, \vartheta)} E_F ||G_i(\theta_0, \beta)||^{2+\gamma} \leq M \}\}. \quad (9.17)
\]

The parameter spaces \(\mathcal{F}^S_{AR}\) and \(\mathcal{F}^S\) impose correct specification of the model, impose uniform bounds on certain moments (which ensure that laws of large numbers and central limit theorems hold under drifting sequences of distributions), include an identifiability condition for \(\beta^*\) given \(\theta_0\), guarantee invertibility of the covariance matrix of \(g_i\), and impose a minimum singular value condition on the expected Jacobian with respect to \(\beta\) of the moment functions. The condition

\(^{13}\)As with the full vector test, the asymptotic size results given below do not require \(G_i(\eta)\) to be the derivative matrix of \(g_i(\eta)\). The matrix \(G_i(\eta)\) can be any \(k \times p\) matrix that satisfies the moment condition in \(\mathcal{F}^S\).

32
$B(\beta^*, \theta) \subset B$ prevents $\beta^*$ from converging to the boundary of $B$ as $n \to \infty$. The assumption that $g$ is twice continuously differentiable in $\beta$ in a neighborhood of $\beta^*$ is used in the proof of consistency and asymptotic normality of $\hat{\beta}_n$ under drifting sequences of null distributions for $W_i$. The asymptotic results allow $\beta^*$ to change with the sample size.

The asymptotic size and similarity properties of the subvector AR and CQLR tests are given in the following theorem.

**Theorem 9.1** Suppose Assumption gB holds. The subvector AR and CQLR tests (without the SR extensions), defined in and above (9.13), have asymptotic size equal to their nominal size $\alpha \in (0, 1)$ and are asymptotically similar (in a uniform sense) for the parameter spaces $\mathcal{F}_{AR}$ and $\mathcal{F}_{S}$, respectively.

**Comment:** Theorem 9.1 is proved in Section 31 below.

9.3 Asymptotic Efficiency of the Subvector CQLR Test under Strong and Semi-Strong Identification

In Section 7 it is established that the (full vector) SR-CQLR test is asymptotically efficient under strong or semi-strong identification when $\Omega_F$ has eigenvalues that are bounded away from zero and the null value $\theta_0$ is not on the boundary. We next establish the analogous result for the subvector CQLR test. We consider drifting sequences $\{\lambda_{n,h}^S \in \Lambda^S : n \geq 1\}$ of data-generating processes taken from $\mathcal{F}^S$ in (9.17) that correspond to strong or semi-strong identification and establish that the CQLR test statistic equals the subvector LM test statistic up to a $o_p(1)$ term and that the conditional critical value of the subvector CQLR test converges in probability to $2\beta_1 \alpha$. Note that $\mathcal{F}^S$ imposes the minimal eigenvalue restriction $\lambda_{min}(E_F g_i g_i') \geq \delta > 0$. It also imposes the restriction $\tau_{min}(E_F G_{i\beta}) \geq \delta$, which implies strong identification of $\beta$.

As in Newey and West (1987, p.780, third equation in (2.9)), define the subvector LM test statistic as

$$
LM_n^S := n \hat{g}_n(\hat{\eta})' \hat{\Omega}_n^{-1}(\hat{\eta}) \hat{G}_{\eta\eta}(\hat{\eta}) \left( \hat{G}_{\eta\eta}(\hat{\eta})' \hat{\Omega}_n^{-1}(\hat{\eta}) \hat{G}_{\eta\eta}(\hat{\eta}) \right)^{-1} \hat{G}_{\eta\eta}(\hat{\eta})' \hat{\Omega}_n^{-1}(\hat{\eta}) \hat{g}_n(\hat{\eta}),
$$

where

$$
\hat{G}_{\eta\eta}(\hat{\eta}) := [\hat{G}_{\alpha}(\hat{\eta}) : \hat{G}_{\beta\eta}(\hat{\eta})] \in \mathbb{R}^{k \times (p+b)} (9.18)
$$

and $\hat{\eta} := (\theta_0, \hat{\beta}_n)$. The critical value of the subvector LM test of (9.2) is given by $\chi_{p,1-\alpha}^2$.

Suppose $k - b \geq p$. The drifting sequences $\{\lambda_{n,h}^S : n \geq 1\}$ referred to above are rather complicated and so, for brevity, we define them in (31.14) and (31.15) in the SM. They are defined so that various population quantities that affect the asymptotic distributions of the CQLR test statistic

Electronic copy available at: https://ssrn.com/abstract=3366443
and critical value converge as \( n \to \infty \). Most importantly, we have that, along \( \{\lambda_{S,h}^S : n \geq 1\} \), \( n^{1/2}(\tau_{1F,t}^*, ..., \tau_{pF,t}^*) \) converges to some vector \((h_{1,1t}^*, ..., h_{1,pt}^*)\) whose elements may be finite or infinite, where \((\tau_{1F,t}^*, ..., \tau_{pF,t}^*)\) denote the singular values of \( O_{Ft}^*(E_{FGi}g_i^*)^{-1/2}(E_{FGi})U_F \in R^{(k-b) \times p} \). The latter quantity depends on the Jacobian \( E_{FGi} \), the moment variance matrix \( E_{FGi}g_i^* \), the matrix \( U_F \in R^p \times p \), which is the population counterpart of \( \tilde{T}_n^{1/2}(\theta_0, \hat{\beta}_n) \), and the matrix \( O_{Ft}^* \in R^{k \times (k-b)} \), which is defined such that \( O_{Ft}^*O_{Ft}^{1/2} \) is a uniquely-defined population counterpart of the projection weight matrix \( M_{\tilde{f}_n(\theta)} \). \(^{14}\) Strong or semi-strong identification occurs if the smallest singular value of \( O_{Ft}^*(E_{FGi}g_i^*)^{-1/2}(E_{FGi})U_F \) diverges to infinity after renormalization by \( n^{1/2} \), i.e., if \( h_{1,pt}^* = \infty \).

**Theorem 9.2** Suppose Assumption \( gB \) holds and \( k-b \geq p \). For any sequence \( \{\lambda_{S,h}^S : S \geq 1\} \) that exhibits strong or semi-strong identification (where sequences \( \{\lambda_{S,h}^S : S : n \geq 1\} \) are defined precisely following \(^{[31.15]}\) in Section 9.1 in the SM and strong and semi-strong identification are defined precisely in Section 28 in the SM), we have

(a) \( SR-QLR_n^S(\theta_0, \hat{\beta}_A^\prime) = QLR_n^S(\hat{\eta}) + o_p(1) = LM_n^S + o_p(1) \) and

(b) \( c_{r_n(\theta_0, \hat{\beta}_A^\prime)}(n^{1/2} \tilde{D}_n^* (\theta_0, \hat{\beta}_A^\prime), \tilde{\eta}_n(\theta_0, \hat{\beta}_A^\prime), 1-\alpha) = c_{k,p}(n^{1/2} \tilde{D}_n^* (\hat{\eta}), \tilde{\eta}_n(\hat{\eta}), 1-\alpha) + o_p(1) \to_p \chi^2_{k-1-\alpha}^2 \)

9.4 Monte Carlo Study: Probit Model with Endogeneity

In this section we compare the finite-sample rejection probabilities (RP’s), under the null and alternative hypotheses, of the subvector AR and CQLR tests, defined in \([9.10]\) and \([9.13]\), with two tests in the literature. These two tests are the subvector AR-type test in Stock and Wright (2000, Thm. 3), which we refer to as the S test, and the subvector CLR test in Andrews and Mikusheva (2016), which we refer to as the AM test. We consider a probit model with endogeneity:

\[
y_i = 1(y_i^* > 0), \quad y_i^* = \beta_0 + \beta_1 x_{1i} + \theta x_{2i} + u_i, \quad \text{and} \quad x_{2i} = \tilde{Z}_i \pi + v_{2i}, \quad (9.19)
\]

where \( Z_i = (1, x_{1i}, \tilde{Z}_i^\prime) \in R^g \) is a vector of IV’s, \( \theta \) and \( \beta = (\beta_0, \beta_1, \pi^\prime) \) are parameters with \( \theta, \beta_0, \beta_1 \in R \) and \( \pi \in R^{g-2} \), \( x_{1i} \) and \( x_{2i} \) are scalar exogenous and endogenous regressors, respectively, and the observed variables are \( \{(y_i, x_{1i}, x_{2i}, \tilde{Z}_i^\prime) : i = 1, \ldots, n\} \). The reduced form for \( y_i^* \)

\(^{14}\)The indexation of \( O_{Ft}^* \) by \( t^* \) is the result of the need to define a unique matrix \( O_{Ft}^* \) out of the many matrices \( O_{Ft} \in R^{k \times (k-b)} \) for which \( O_{Ft}O_{Ft}^* \) is a population counterpart of \( M_{\tilde{f}_n(\theta)} \). See \([31.13]\) and \([31.15]\) in the SM for details.
is

\[ y_i^* = \beta_0 + \beta_1 x_{1i} + \theta Z_i' \pi + v_{1i}, \text{ where } v_{1i} := \theta v_{2i} + u_i, \]

\[ (v_{1i}, v_{2i})' \sim iid \mathcal{N}(0^2, V), \quad V := \begin{pmatrix} 1 & \rho \sigma \\ \rho \sigma & \sigma^2 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \tag{9.20} \]

for \( \rho \in (-1, 1) \) and \( \sigma^2 > 0 \), and \( (v_{1i}, v_{2i})' \) is independent of \( Z_i \). Also, \( (x_{1i}, Z_i')' \sim iid \mathcal{N}(0^{g-1}, I_{g-1}) \).

The objective is to test \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \) in the presence of the vector of nuisance parameters \( \beta := (\beta_0, \beta_1, \pi)' \in \mathbb{R}^3 \). We have

\[ E(y_i|Z_i) = \Pr(y_i = 1|Z_i) = \Pr(y_i^* > 0|Z_i) = \Pr(\beta_0 + \beta_1 x_{1i} + \theta Z_i' \pi > -v_{1i}|Z_i) = \Phi(\beta_0 + \beta_1 x_{1i} + \theta Z_i' \pi). \tag{9.21} \]

The model implies the moment conditions \( E g_i(\theta, \beta) = 0 \), where

\[ g_i(\theta, \beta) := \frac{1}{(x_{2i} - Z_i' \pi)Z_i} \in \mathbb{R}^g. \tag{9.22} \]

We proceed by estimating the vector of nuisance parameters \( \beta \) under the null by two-step GMM. In the notation employed above, \( k = 2g, b = g, \) and \( p = 1 \).

Given a weighting matrix \( \hat{W}_n \), the GMM criterion function is \( Q_n^{\hat{W}_n}(\theta, \beta) := \hat{g}_n(\theta, \beta) \hat{W}_n \hat{g}_n(\theta, \beta) \). Taking \( \hat{W}_n = I_k \), the first-step GMM estimator \( \hat{\beta}_{n,FS} \) of \( \beta \) minimizes \( Q_n^{I_k}(\theta_0, \beta) \). The second-step GMM estimator \( \hat{\beta}_n \) minimizes \( Q_n^{\hat{W}_n}(\theta_0, \beta) \), where \( \hat{W}_n := n^{-1} \sum_{i=1}^n g_i(\theta_0, \hat{\beta}_{n,FS})g_i(\theta_0, \hat{\beta}_{n,FS})' \).

In the simulation results reported below, the nominal size of the tests is 5%. We take \( \theta_0 = 1 \) (the null value of \( \theta \)), \( \beta_0 = \beta_1 = 1 \), and \( \sigma = 2 \). In addition to the null value, we consider three true values of \( \theta \) on each side of the null such that the resulting RP’s of the subvector CQLR test are roughly equal to 40%, 65%, and 90%. We let \( \pi \in \mathbb{R}^{g-2} \) be a multiple of a vector of ones with a multiplicative constant \( \pi \). The latter determines the strength of identification of \( \theta \). We consider

\[ \text{[15] The other nuisance parameters } \rho \text{ and } \sigma \text{ do not enter the moment function } g_i \text{ defined below.} \]

\[ \text{[16] We use the Newton-Raphson algorithm to find the two-step GMM estimator for } \beta \text{. In both steps we initiate the search from a number of starting points and do ten Newton iterations from each starting point. In particular, for the first step estimator we use } (\hat{\beta}_0, \hat{\pi}) \text{ as one starting value, where } (\hat{\beta}_0, \hat{\beta}_1) \text{ is the OLS estimator of the slope coefficients in a regression of } y - \theta_0 x_2 \text{ on a constant and } x_1 \text{ and } \hat{\pi} \text{ is the OLS estimator in a regression of } x_2 \text{ on } \hat{Z} \text{ and we use } (\hat{\beta}_0, \hat{\pi}) \text{ as another starting value, where } \hat{\pi} \text{ is the true value of the slope coefficients in the third line of } \text{(9.19)} \text{. For the second step, we use the same starting values and also the estimator obtained in the first step. We also experimented using an additional fifteen randomly generated starting points which had little effect on the results. In each Newton iteration, we incorporate a step size control where along the search direction the step is divided in seven equal parts and the next iteration proceeds from the step that yields the smallest criterion function. For numerical stability when inverting matrices, we replace all eigenvalues of the matrices smaller than } 10^{-11} \text{ by } 10^{-11} \text{. We use } \varepsilon = .01 \text{ for the eigenvalue adjustment constant in } \text{(9.11)} \text{. The estimator of } \beta \text{ in each of the two steps is the minimizer of the stochastic criterion function over all candidate vectors for which the criterion function was evaluated in that step.} \]
16 parameter configurations consisting of all of the combinations of \( g = 3,4 \) (which results in \( k = 6,8 \)), \( \rho = 0,.9 \), and \( \pi = 1,.5,.2,.1 \). The sample size is \( n = 1,000 \). The results are based on 5,000 simulation repetitions, and 5,000 simulation repetitions are used to simulate the critical values for each data sample. When calculating the QLR statistic in AM, we use 60 search points to find the infimum over \( \theta \) (see eq. (2) on p. 1575 in AM).

First, we report RP’s under the null hypothesis. Across the 16 parameter configurations, the null RP’s of the CQLR, AR, AM, and S tests fall into the intervals \([5.0\%, 6.7\%]\), \([5.4\%, 6.8\%]\), \([3.6\%, 6.5\%]\), and \([4.7\%, 5.8\%]\), respectively. There is no apparent pattern as to how the RP’s depend on the various parameters \( g, \rho, \) or \( \pi \). Therefore, while there is overrejection under the null for some parameter configurations for all four tests considered, the overrejection is at most slight no matter what the strength or weakness of identification.

Figure 1 reports power for the four tests for \( \rho = .9 \) and \( \pi = 1,.5,.2,.1 \) for \( g = 3 \) (upper row) and \( g = 4 \) (lower row) for three alternatives to the left of the null value of \( \theta \), the null value, and three alternatives to the right of the null value. For clarity, the graphs linearly interpolate the power between the seven \( \theta \) values. Figure SM-1 in the SM provides the corresponding results for \( \rho = 0 \). As expected, the powers of all tests decrease as \( \pi \) decreases. Thus, the CQLR test reaches the 40, 65, and 90% RP’s for alternatives farther from the true value the smaller is \( \pi \), with all other parameters held constant. For example, in the upper panel of Figure 1, which reports power when \( g = 3 \), the sum of the distances to the alternative \( \theta \) values to the left and right of the null value such that the CQLR test has 90% power are roughly 48, .78, 1.82, and 3.59 for \( \pi = 1,.5,.2, \) and \(.1\), respectively. The powers of the tests increase as \( g \) increases from 3 to 4 (with other parameters held constant) with the corresponding sum of the distances being roughly 43, .60, 1.34, and 3.55, see the lower panel of Figure 1. The powers of the tests decrease as \( \rho \) decreases from .9 to 0 with other parameters held constant.

In all scenarios, the AR test has higher power than the S test for alternatives to the left of the null value of \( \theta \). It also has higher power for alternatives to the right of the null value of \( \theta \) except in the most strongly identified case \( \pi = 1 \). The AM test has uniformly higher power than the AR and S tests.

Overall, the CQLR and AM tests are the best two tests among the four tests considered. The CQLR test has higher power than the AM test for all alternatives to the left of the null value in 14 of the 16 parameter configurations with power gains up to 16.5% when \( \pi = 1 \) (see Figure 1 with \( g = 4 \) and \( \pi = 1 \)) and up to 7.5% for \( \pi \leq .5 \) (see Figure 1 with \( g = 3 \) and \( \pi = .2 \)). The AM test has higher power than CQLR for alternatives to the left in the two cases \((g, \pi, \rho) = (4,.1,0)\) and \((4,.1,.9)\), e.g., see Figure 1 with \( g = 4 \) and \( \pi = .1 \). For this parameter configuration the highest
power advantage of the AM test is 23% for $\theta = -.42$.

The CQLR test has comparable or slightly better power than the AM test for all alternatives to the right of the null value except when $\pi = 1$. When $\pi = 1$, the power advantage of the AM test over CQLR is between 1.2% and 2.2% when $(g, \pi, \rho) = (3, 1, 0)$ and it is is between 2.7% and 6.0% when $(g, \pi, \rho) = (4, 1, 0)$ over three three alternatives considered to the right of the null value, see Figure SM-1.

With regard to computation time, it takes about 231 minutes to calculate 5,000 AM tests when $(g, \pi, \rho) = (4, .5, .9)$ under the specifications described above using an Intel Core 3.4GHz, 6MB processor. On the other hand it only takes about 4 minutes to calculate 5,000 CQLR tests, that is, the CQLR test is about 58 times faster to calculate. The difference in computation times is expected to be much larger in cases where $\theta$ is of dimension greater than 1, because the computation time of the AM test increases exponentially in the dimension of $\theta$, whereas the computation time of the CQLR test does not depend on the dimension of $\theta$. Computation time is particularly important when computing a confidence set by inverting a test, because the test has to be computed many times.

10 Power Comparisons in Heteroskedastic/Autocorrelated Linear IV Models

In this section, we present some power comparisons for the AR test, Kleibergen’s (2005) LM, JVW-CLR, and MVW-CLR tests, and the SR-CQLR test introduced above. We also consider the plug-in conditional linear combination (PI-CLC) test introduced in I. Andrews (2016), as well as the MM1-SU and MM2-SU tests introduced in Moreira and Moreira (2015). The PI-CLC test aims to approximate the test that has minimum regret among conditional tests constructed using linear combinations of the LM and AR test statistics (with coefficients that depend on the conditioning statistic), see I. Andrews (2016) for details. The MM1-SU and MM2-SU tests have optimal weighted average power for two different weight functions (over the alternative parameter values $\theta$ and the strength of identification parameter vector $\mu$, given in (10.1) below) among tests that satisfy a sufficient condition for local unbiasedness.

We consider the same designs as in I. Andrews (2016, Sec. 7.2). These designs are for heteroskedastic and autocorrelated data.

---

17 See (4.2), (7.1), and a footnote in Section 21 of the SM for the definitions of the AR test and Kleibergen’s LM, MVW-CLR, and JVW-CLR tests. The AR test is called the S test in Stock and Wright (2000). The LM and JVW-CLR tests are denoted by K and QCLR, respectively, in I. Andrews (2016).

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

19 The weight functions considered depend on the variance parameters $\Sigma_{\theta G}$ and $\Sigma_{GG}$ in (10.1) below.
eroskedastic and/or autocorrelated linear IV models with \( p = 1 \) and \( k = 4 \). The designs are calibrated to mimic the linear IV models for the elasticity of inter-temporal substitution estimated by Yogo (2004) for eleven countries using quarterly data from the early 1970’s to the late 1990’s. The power comparisons are for the limiting experiment under standard weak identification asymptotics. In consequence, for the simulations, the observations are drawn from the following model:

\[
\begin{pmatrix}
\Omega_n^{-1/2} n^{1/2} \tilde{G}_n(\theta_0) \\
\Omega_n^{-1/2} n^{1/2} \tilde{\gamma}_n(\theta_0)
\end{pmatrix}
\sim
N
\left(
\begin{pmatrix}
\mu \theta \\
\mu
\end{pmatrix},
\begin{pmatrix}
I_k & \Sigma_{gG} \\
\Sigma_{gG}' & \Sigma_{GG}
\end{pmatrix}
\right)
\]

(10.1)

for \( \theta \in R, \mu \in R^k, \) and \( \Sigma_{gG}, \Sigma_{GG} \in R^{k \times k} \), where \( \Sigma_{gG} \) and \( \Sigma_{GG} \) are assumed to be known.\(^{20}\)\(^{21}\) The values of \( \mu, \Sigma_{gG}, \) and \( \Sigma_{GG} \) are taken to be equal to the estimated values using the data from Yogo (2004).\(^{22}\) A sample is a single observation from the distribution in (10.1) and the tests are constructed using the known values \( \Sigma_{gG} \) and \( \Sigma_{GG} \).\(^{23}\) The hypotheses are \( H_0 : \theta = 0 \) and \( H_1 : \theta \neq 0 \).

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

Some details concerning the computation and definitions of the SR-CQLR, PI-CLC, MM1-SU, and MM2-SU tests are as follows. The SR-CQLR test uses \( \varepsilon = .01 \), where \( \varepsilon \) appears in the definition of \( \widehat{L}_n(\theta) \) in (5.7).\(^{24}\) For the PI-CLC test, the number of values "a" considered in the search over \([0,1]\) is 100, the number of simulation repetitions used to determine the best choice of "a" is 2000, and the number of alternative parameter values considered in the search for the best "a" is 41. For the MM1-SU and MM2-SU tests, the number of variables in the discretization of maximization problem is 1000, the number of points used in the numerical approximations of the integrals \( h_1 \) and \( h_2 \) that appear in the definitions of these tests is 1000, and when approximating integrals \( h_1 \) and \( h_2 \) by sums of 1000 rectangles these rectangles cover \([-4,4]\).

The asymptotic power functions are given in Figure 2. Each graph is based on 41 equi-spaced

---

\(^{20}\) In linear IV models with i.i.d. observations, the matrix \( \Sigma_{gG} \) is necessarily symmetric. However, with autocorrelation, it need not be. In the eleven countries considered here, it is not.

\(^{21}\) The variance matrix in the limit experiment varies slightly depending on whether one treats the IV’s as fixed or random. For example, the asymptotic variance of \( n^{1/2} \tilde{G}_n(\theta_0) \) under standard weak IV asymptotics varies slightly in these two cases. Power results for the SR-CQLR\( \rho \) test that is introduced in the SM when the limiting variance is computed using fixed IV’s are equivalent to those computed for the SR-CQLR test for the case where the limiting variance is computed using random IV’s. In consequence, we do not separately report power results for the SR-CQLR\( \rho \) test.

\(^{22}\) See I. Andrews (2016, Appendices D.3 and D.4) for details on the calculations of the simulation designs based on Yogo’s (2004) data, as well as for details on the computation of I. Andrews’ PI test, referred to here as PI-CLC, and the two tests of Moreira and Moreira (2013), referred to here and in I. Andrews (2016) as MM1-SU and MM2-SU. The JVW-CLR and LM tests here are the same as the QCLR and K tests, respectively, in I. Andrews (2016).

\(^{23}\) For example, \( \Gamma_{gG}(\theta_0) \) in (5.2) is taken to be known and equal to \( \Sigma_{gG} \), and \( \hat{V}_n(\theta_0) \) in (15.5) is taken to be known.
Table III. Shortfalls in Average-Power (×100)

| Country     | μ/μ' | non-Kron | SR-CQLR | JVV | MVW | PI-CLC | MM1 | MM2 | LM | AR |
|-------------|------|----------|---------|-----|-----|--------|-----|-----|----|----|
| Australia   | 138  | 17       | .0      | .1  | .1  | .2     | 2.4 | .1  | .1 | 6.9|
| Canada      | 48   | 5        | .0      | .0  | .2  | .0     | 1.4 | .5  | .3 | 6.8|
| France      | 79   | 6        | .1      | .2  | .0  | .3     | .7  | .3  | .0 | 8.0|
| Germany     | 10   | 3        | .0      | .1  | .4  | .0     | .2  | .1  | .1 | 2.3|
| Italy       | 84   | 15       | .5      | 1.1 | 2.0 | .2     | 1.1 | .0  | .3 | 8.0|
| Japan       | 17   | 14       | 3.3     | 8.9 | 4.0 | .4     | 2.4 | .0  | 17.4| .6|
| Netherlands | 25   | 3        | 0.2     | .1  | .2  | .9     | .5  | .5  | 1.6| 6.6|
| Sweden      | 174  | 9        | .3      | .2  | .3  | 1.5    | .0  | .3  | 7.5|
| Switzerland | 31   | 4        | .1      | .0  | .0  | .4     | 1.3 | 1.1 | .5 | 7.2|
| U.K.        | 53   | 38       | .7      | 6.0 | 5.4 | .8     | 2.5 | .0  | 7.8| 3.8|
| U.S.        | 81   | 10       | .8      | 2.0 | 2.9 | .0     | 7.3 | .8  | 3.5| 3.2|

Average over Countries .5 1.2 1.8 .2 1.8 .5 3.3 5.7

values on the x axis covering [−6, 6]. The x axis variable is the parameter θ scaled by a fixed value of ||μ|| for a given country, thus θ||μ|| ∈ [−6, 6], where θ is the alternative parameter value (when θ ̸= 0) defined in (10.1) and μ is the mean vector that determines the strength of identification. The y axis variable is power ×100.

Table III provides the shortfall in average-power (×100) of each test for each country relative to the other seven tests considered, where average power is an unweighted average over the 40 alternative parameter values. Table IV provides the maximum power shortfall (×100) of each test for each country relative to the other seven tests considered, where the maximum is taken over the 40 alternative parameter values.25 The shortfall in average-power is an unweighted average power criterion, whereas the maximum power shortfall is a minimax regret criterion.

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

The second and third columns of Table III provide the concentration parameter, μ'μ, which measures of the strength of identification, and a non-Kronecker index, abbreviated by non-Kron, which measures the deviation of the variance matrix in (10.1), call it Ψ, from a Kronecker matrix.

and equal to the variance matrix in (10.1).

24 The numerical results are unchanged when ε = .001 or .05.

25 More precisely, let APt(c) denote the average power of test t for country c, where the average is taken over the 40 parameter values in the alternative hypothesis. By definition, the shortfall in average-power of test t for country c is 

max_{c<θ} APt(c) − APt(0), where the maximum is taken over the eight tests considered.

Let P_{t}(θ) denote the power of test t in country c against the alternative θ. By definition, the power shortfall of test t in country c for alternative θ is max_{c<θ} P_{t}(θ) − P_{t}(θ) and the maximum power shortfall of test t in country c is max_{θ∈Θ_0} (max_{c<θ} P_{t}(θ) − P_{t}(θ)), where Θ_0 contains the 40 alternative parameter values considered.

Note that, as defined, the shortfall in average-power is not equal to the average of the power shortfalls over θ ∈ Θ_0.
Table IV. Maximum Power Shortfalls (×100)

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

Average over Countries | 4.0 | 8.0 | 8.3 | 2.8 | 9.0 | 4.0 | 14.9 | 17.0 |

This deviation is given by the formula $1,000 \times \min_{B,C} \|B \otimes C - \Psi\|$, where the minimum is taken over symmetric pd matrices $B$ and $C$ of dimensions $2 \times 2$ and $4 \times 4$, respectively, $\| \cdot \|$ denotes the Frobenius norm, and the rescaling by 1,000 is for convenience. The non-Kronecker index is computed using the Framework 2 method given in Section 4 of Van Loan and Pitsianis (1993) with symmetry of $C$ imposed by replacing $A_{ij}$ by $(A_{ij} + A_{ji})/2$ in equation (9) of that paper.

Germany, Japan, and the Netherlands exhibit the weakest identification, while Sweden and Australia exhibit the strongest. The U.K., Australia, Italy, and Japan have variance matrices that are farthest from Kronecker-product form, while Germany, the Netherlands, and Switzerland have variance matrices that are closest to Kronecker-product form.

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

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

We conclude that the SR-CQLR test has asymptotic power that is competitive with, or better than, that of other tests in the literature for the particular parameters considered here in the particular model considered here. The SR-CQLR test has advantages compared to the PI-CLC,
MM1-SU, and MM2-SU tests of (i) being applicable in almost any moment condition model, whereas the latter tests are not \(^{27}\) (ii) being easy to implement (i.e., program), and (iii) being fast to compute. \(^{27}\) The PI-CLC test does not apply to moment condition models with possibly singular variance matrices. The MM1-SU and MM2-SU tests apply only to the linear IV model with errors that may be heteroskedastic and/or autocorrelation.
References

Anderson, T. W., and H. Rubin (1949): “Estimation of the Parameters of a Single Equation in a Complete Set of Stochastic Equations,” *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., and X. Cheng (2013): “GMM Estimation and Uniform Subvector Inference with Possible Identification Failure,” *Econometric Theory*, 30, 1–47.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2019): “Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests,” forthcoming in *Journal of Econometrics*.

Andrews, D. W. K., and P. Guggenberger (2010): “Asymptotic Size and a Problem with Subsampling and the $m$ Out of $n$ Bootstrap,” *Econometric Theory*, 26, 426–468.

—— (2017): “Asymptotic Size of Kleibergen’s LM and Conditional LR Tests for Moment Condition Models,” *Econometric Theory*, 33, 1046–1080.

—— (2018): “Supplemental Material to ‘Identification- and Singularity-Robust Inference for Moment Condition Models’,” Cowles Foundation Discussion Paper No. 1978R, Yale University.

Andrews, D. W. K., V. Marmer, and Z. Yu (2019): “On Optimal Inference in the Linear IV Model,” *Quantitative Economics*, 10, forthcoming.

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

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

Andrews, I. (2016): “Conditional Linear Combination Tests for Weakly Identified Models,” *Econometrica*, 84, 2155–2182.

Andrews, I., and A. Mikusheva (2015): “Maximum Likelihood Inference in Weakly Identified DSGE models,” *Quantitative Economics*, 6, 123–152.

—— (2016): “Conditional Inference with a Functional Nuisance Parameter,” *Econometrica*, 84, 1571–1612.

Berry, S., J. Levinsohn, and A. Pakes (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 60, 889–917.
Campbell, J.Y. (2003): “Consumption-Based Asset Pricing,” in G.M. Constantinides, M. Harris, and R.M. Stulz (Eds), Handbook of the Economics of Finance, Amsterdam, Elsevier, vol. 1B, chapter 13, 801–885.

Caner, M. and N. Yildez (2012): “CUE with Many Weak Instruments and Nearly Singular Design,” Journal of Econometrics, 170, 422–441.

Chan, W. H., and J. M. Maheu (2002): “Conditional Jump Dynamics in Stock Market Returns,” Journal of Business and Economic Statistics, 20, 377–389.

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

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.

Elliott, G., U. K. Müller, and M. W. Watson (2015): “Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis,” Econometrica, 83, 771–811.

Grant, N. (2013): “Identification Robust Inference with Singular Variance,” Economics Discussion Paper No. 1315, Department of Economics, University of Manchester.

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

Guggenberger, P., and R. J. Smith (2005): “Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification,” Econometric Theory, 21, 667–709.

Han, S. and A. McCloskey (2019): “Estimation and Inference with a (Nearly) Singular Jacobian,” forthcoming in Quantitative Economics.

Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” Econometrica, 50, 1029–1054.

Jagannathan, R., G. Skoulakis, and Z. Wang (2002): “Generalized Method of Moments: Applications in Finance,” Journal of Business and Economic Statistics, 20, 470–481.

Jorion, P. (1988): “On Jump Processes in the Foreign Exchange and Stock Markets,” Review of Financial Studies, 1, 427–425.

43
Kleibergen, F. (2005): “Testing Parameters in GMM Without Assuming That They Are Identified,” *Econometrica*, 73, 1103–1123.

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

Mikusheva, A. (2010): “Robust Confidence Sets in the Presence of Weak Instruments,” *Journal of Econometrics*, 157, 236–247.

Montiel Olea, J. L. (2012): “Efficient Conditionally Similar-on-the-Boundary Tests,” unpublished manuscript, Department of Economics, New York University.

Moreira, H., and M. J. Moreira (2015): “Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors,” unpublished manuscript, available at arXiv:1505.06644.

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

Newey, W. K., and K. West (1987): “Hypothesis Testing with Efficient Method of Moments Estimation,” *International Economic Review*, 28, 777–787.

Otsu, T. (2006): “Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models Under Weak Identification,” *Econometric Theory*, 22, 513–527.

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

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

Staiger, D., and J. H. Stock (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.

Stock, J. H., and J. H. Wright (2000): “GMM with Weak Identification,” *Econometrica*, 68, 1055–1096.
Van Loan, C. F., and N. Pitsianis (1993): “Approximation with Kronecker Products,” in *Linear Algebra for Large Scale and Real-Time Applications*, ed. by M. S. Moonen, G. H. Golub, and B. L. R. de Moor, 293–314. Amsterdam: Kluwer Academic Publishers.

Yogo, M. (2004): “Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak,” *Review of Economics and Statistics*, 86, 797–810.
FIGURE 1: Power of CQLR, AM, S, and AR as function of $\theta$ for $\rho = .9$ and $\pi = 1, .5, .2, .1$; first/second row $g = 3/4$

Electronic copy available at: https://ssrn.com/abstract=3366443
FIGURE 2: Power of SR-CQLR and other tests as function of $\theta$ for model of Yogo (2004)

2(a) Australia

2(b) Canada

2(c) France

2(d) Germany

Electronic copy available at: https://ssrn.com/abstract=3366443
FIGURE 2: Power of SR-CQLR and other tests as function of $\theta$ for model of Yogo (2004)

2(e) Italy

2(f) Japan

2(g) Netherlands

2(h) Sweden

Electronic copy available at: https://ssrn.com/abstract=3366443
Supplemental Material to
Identification- and Singularity-Robust Inference for Moment Condition Models

By
Donald W. K. Andrews and Patrik Guggenberger

January 2015
Revised April 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 1978R2S

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

http://cowles.yale.edu/

Electronic copy available at: https://ssrn.com/abstract=3366443
Supplemental Material

for

Identification- and Singularity-Robust Inference
for Moment Condition Models

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

Patrik Guggenberger
Department of Economics
Pennsylvania State University

First Version: March 25, 2011
Revised: April 4, 2019
Contents

11 Outline 4

12 Further Discussion of the Related Literature 5

13 Subvector SR Tests for Potentially Singular Moments Variance Matrices 7

14 Miscellanei 10

14.1 Moore-Penrose Expression for the SR-AR Statistic 10

14.2 Computation Implementation 11

15 SR-CQLR_p Test 11

15.1 SR-CQLR_p Parameter Space 12

15.2 Definition of the SR-CQLR_p Test 13

15.3 Asymptotic Size of the SR-CQLR_p Test 18

15.4 Asymptotic Efficiency of the SR-CQLR_p Test under Strong Identification 18

15.5 Summary Comparison of CLR-type Tests in Kleibergen (2005) and AG2 19

16 Tests without the Singularity-Robust Extension 19

16.1 Asymptotic Results for Tests without the SR Extension 20

16.2 Uniformity Framework 21

16.3 General Weight Matrices \( \hat{W}_n \) and \( \hat{U}_n \) 22

16.4 Uniformity Reparametrization 25

16.5 Assumption WU 28

16.6 Asymptotic Distributions 28

17 Singularity-Robust Tests 33

18 Time Series Observations 39

19 SR-CQLR, SR-CQLR_p, and Kleibergen’s Nonlinear CLR Tests in the Homoskedastic Linear IV Model 42

19.1 Normal Linear IV Model with \( p \geq 1 \) Endogenous Variables 43

19.2 Homoskedastic Linear IV Model 45

19.3 SR-CQLR_p Test 46

19.4 SR-CQLR Test 47

19.5 Kleibergen’s Nonlinear CLR Tests 49
31.4 Proof of Theorem 9.2
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 extend 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) IV’s.

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.
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 \( \theta_0 \) for various quantities that depend on the null value \( \theta_0 \).

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 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) establishes the correct asymptotic size of LM and CLR tests in the linear IV model when the errors are homoskedastic. Guggenberger (2012) establishes 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 see 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, Richardson, Robins, and Zivot (2010) and Chaudhuri and Zivot (2011).\textsuperscript{28} Andrews (2017) provides subvector methods that are closely related to

\textsuperscript{28}Cavanagh, Elliott, and Stock (1995) provide a refinement of Bonferroni’s method that is not conservative, but it is much more intensive computationally. McCloskey (2017) also considers a refinement of Bonferroni’s method.
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,b) provide 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 \( \beta, \pi, \) and \( \zeta \) 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 \( \pi \) (in their notation) when \( \beta = 0 \), which rules out some weak IV’s models.\(^{29}\) Cheng (2015) provides 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) develop subvector inference methods in a minimum distance context based on Anderson-Rubin-type statistics. Cox (2017) provides subvector methods in a class of models that allows for multiple sources of weak identification and includes factor models. I. Andrews and Mikusheva (2015) provide conditions under which subvector inference is possible in exponential family models (but the requisite conditions seem to be quite restrictive). I. Andrews (2018) considers 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) provide 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) provides 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) consider GMM estimation under semi-strong and strong identification, but do not consider tests or CS’s that are robust to weak identification. Armstrong, Hong, and Nekipelov (2012) show that standard Wald tests for multiple restrictions in some nonlinear IV models can exhibit size distortions when some IV’s 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

\(^{29}\) Montiel Olea (2012) also provides some subvector analysis in the extremum estimator context of Andrews and Cheng (2012). His efficient conditionally similar tests apply to the subvector \((\pi, \zeta)\) of \((\beta, \pi, \zeta)\) (in Andrews and Cheng’s (2012) notation), where \( \beta \) 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).
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 $\rho = .9$. 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 (4.3) (4.4) with $\eta$ in place of $\theta$. That is, $\bar{r}_n(\theta, \beta) := rk(\bar{\Omega}_n(\theta, \beta))$ and $\bar{\Omega}_n(\theta, \beta) := \hat{A}_n^\Omega(\theta, \beta)\hat{\Pi}_n(\theta, \beta)\hat{A}_n^\Omega(\theta, \beta)'$, where $\hat{\Pi}_n(\theta, \beta)$ is the $k \times k$ diagonal matrix with the eigenvalues of $\bar{\Omega}_n(\theta, \beta)$ on the diagonal in nonincreasing order, and $\hat{A}_n^\Omega(\theta, \beta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\bar{\Omega}_n(\theta, \beta)$. We partition $\hat{A}_n^\Omega(\theta, \beta)$ according to whether the corresponding eigenvalues are positive or zero: $\hat{A}_n^\Omega(\theta, \beta) = [\hat{A}_n(\theta, \beta), \hat{A}_n^+(\theta, \beta)]$, where $\hat{A}_n(\theta, \beta) \in R^{k \times \bar{r}_n(\theta, \beta)}$ and $\hat{A}_n^+(\theta, \beta) \in R^{k \times (k - \bar{r}_n(\theta, \beta))}$. The columns of $\hat{A}_n(\theta, \beta)$ are eigenvectors of $\bar{\Omega}_n(\theta, \beta)$ that correspond to positive eigenvalues of $\bar{\Omega}_n(\theta, \beta)$.

Analogously, consider the spectral decomposition for the population quantity, defined in (3.4) with $\eta$ in place of $\theta$, i.e., $\Omega_F(\theta, \beta) = A_F^\Omega(\theta, \beta)\Pi_F(\theta, \beta)A_F^\Omega(\theta, \beta)'$, and define $r_F(\theta, \beta) := rk(\Omega_F(\theta, \beta))$. We partition $A_F^\Omega(\theta, \beta)$ as

$$A_F^\Omega(\theta, \beta) = [A_F(\theta, \beta), A_F^+(\theta, \beta)], \text{ where } A_F(\theta, \beta) \in R^{k \times r_F(\theta, \beta)}, \ A_F^+(\theta, \beta) \in R^{k \times (k - r_F(\theta, \beta))},$$

(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 $\theta$ and denote by $\bar{\Omega}_n(\beta)$ the matrix $\bar{\Omega}_n(\theta_0, \beta)$ and similarly for other expressions.

Recall the definition following (9.6) of $\bar{\beta}_n$, 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 (9.2) is based on

$$g_{\bar{A}_n}(\theta, \beta) = \hat{A}_n(\theta_0, \bar{\beta}_n)'g_i(\theta, \beta) \in R^{\bar{r}_n(\theta_0, \bar{\beta}_n)}.$$  

(13.2)
From now on, whenever a subindex \(\hat{A}\) 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 \(g_{\hat{A}i}(\theta, \beta)\). In particular,

\[
\hat{\Omega}_{\hat{A}n}(\theta, \beta) := n^{-1} \sum_{i=1}^{n} g_{\hat{A}i}(\theta, \beta)g_{\hat{A}i}(\theta, \beta)' - \hat{g}_{\hat{A}n}(\theta, \beta)\hat{g}_{\hat{A}n}(\theta, \beta)' \in R^{\hat{r}_n(\theta_0, \beta_n) \times \hat{r}_n(\theta_0, \beta_n)},
\]

\[
\hat{g}_{\hat{A}n}(\theta, \beta) := n^{-1} \sum_{i=1}^{n} g_{\hat{A}i}(\theta, \beta), \text{ and}
\]

\[
\hat{\beta}_{\hat{A}n} := \arg \min_{\beta \in B} \| \hat{\varphi}_{\hat{A}n} \hat{\varphi}_{\hat{A}n}(\theta_0, \beta) \|^2, \tag{13.3}
\]

where \(\hat{\varphi}_{\hat{A}n} \in R^{\hat{r}_n(\theta_0, \beta_n) \times \hat{r}_n(\theta_0, \beta_n)}\) satisfies

\[
\hat{\varphi}_{\hat{A}n} \hat{\varphi}_{\hat{A}n} = \hat{\Omega}_{\hat{A}n}^{-1}(\theta_0, \beta_n). \tag{13.4}
\]

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

\[
SR-AR^S_n(\theta_0, \hat{\beta}_{\hat{A}n}) > \chi^2_{\hat{r}_n(\theta_0, \beta_n), 1-\alpha}. \tag{13.5}
\]

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

\[
SR-QLR^S_n(\theta_0, \hat{\beta}_{\hat{A}n}) > c_{\hat{r}_n(\theta_0, \beta_n), \rho}(n^{1/2} \hat{D}^*_{\hat{A}n}(\theta_0, \hat{\beta}_{\hat{A}n}), \hat{J}_{\hat{A}n}(\theta_0, \hat{\beta}_{\hat{A}n}), 1 - \alpha). \tag{13.6}
\]

If \(\hat{r}_n(\theta_0, \beta_n) = 0\), then \(SR-AR^S_n(\theta_0, \hat{\beta}_{\hat{A}n})\) and \(SR-QLR^S_n(\theta_0, \hat{\beta}_{\hat{A}n}) := 0\) and \(c_{\hat{r}_n(\theta_0, \beta_n), \rho}(n^{1/2} \hat{D}^*_{\hat{A}n}(\theta_0, \hat{\beta}_{\hat{A}n}), \hat{J}_{\hat{A}n}(\theta_0, \hat{\beta}_{\hat{A}n}), 1 - \alpha) := 0\) 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 \(\mathcal{F}_{AR,1}^S\) defined in \([9,14]\) which guarantee consistency of the preliminary estimator \(\hat{\beta}_n\). The parameter space \(\mathcal{F}_{AR,2}^S\) defined in \([9,15]\) is modified in four ways: (i) the condition \(\lambda_{\text{min}}(E_F g_i g_i') \geq \delta\) is dropped, (ii) the condition \(E_F \sup_{\beta \in B(\beta^*, \zeta_1)} \| \Pi_{1F}^{-1/2}(\beta) A_F(\beta)'(g_i(\beta) - E_F g_i(\beta)) \|^2 \leq M\) 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 \(\zeta_1 > 0\), \(N(\Omega_F(\theta_0, \beta^*)) = N(\Omega_F(\theta_0, \beta))\) for all \(\beta \in B(\beta^*, \zeta_1)\), where \(\beta^*\) denotes the true value of \(\beta\), is added. Call the resulting space \(\mathcal{F}_{AR,2}^{S,SR}\). We define the null parameter space for the subvector SR
The null parameter space for the subvector SR-CQLR test, denoted by \( \mathcal{F}_{\Theta,AR}^{S,SR} \), is defined as \( \mathcal{F}^S \) is defined in [9.17] with the following modifications. First, \( \mathcal{F}_{\Theta,AR}^S \) is replaced by \( \mathcal{F}_{\Theta,AR}^{S,SR} \), 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 \( \theta \) with correct asymptotic confidence size by inversion of the subvector SR-AR and SR-CQLR tests. The relevant parameter spaces are given by

\[
\mathcal{F}_{\Theta,AR}^{S,SR} := \{(F, \beta, \theta_0) : (F, \beta) \in \mathcal{F}_{\Theta,AR}^{S,SR}(\theta_0), \theta_0 \in \Theta \} \text{ and } \mathcal{F}_{\Theta}^{S,SR} := \{(F, \beta, \theta_0) : (F, \beta) \in \mathcal{F}^{S,SR}(\theta_0), \theta_0 \in \Theta \},
\]

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

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

\[
g_i(\theta, \beta) = u_i(\theta, \beta)Z_i,
\]

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 \( E_F u_i^2(\theta_0, \beta^*)Z_i Z_i' \) factors into \( E_F u_i^2(\theta_0, \beta^*)E_F Z_i Z_i' \). (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, \( \Omega_F(\theta_0, \beta) = E_F u_i^2(\theta_0, \beta)Z_i Z_i' - E_F u_i(\theta_0, \beta)Z_i E_F u_i(\theta_0, \beta)Z_i' \) and \( \Omega_F(\theta_0, \beta^*) = E_F u_i^2(\theta_0, \beta^*)E_F Z_i Z_i' \). If \( A_F \Pi_F A_F' \) denotes a singular value decomposition of \( E_F Z_i Z_i' \) with \( \Pi_F = Diag(\Pi_{1F}, \Pi_{0F}) \), 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

\[
A_F' E_F Z_i Z_i' A_F = Diag(\Pi_{1F}, \Pi_{0F}) \text{ and thus, in particular, } E_F (A_F' Z_i)^2_j = 0 \text{ for } j = r+1, \ldots, k. \text{ Therefore, } (A_F' Z_i)_j = 0 \text{ a.s. for } j = r+1, \ldots, k. \text{ But then } A_F' \Omega_F(\theta_0, \beta)A_F = E_F u_i^2(\theta_0, \beta)A_F Z_i Z_i' A_F - E_F u_i(\theta_0, \beta)A_F' Z_i \cdot E_F u_i(\theta_0, \beta)Z_i' A_F, \text{ for any } \beta \in B, \text{ equals a block diagonal matrix with lower right block equal to } 0^{(k-r) \times (k-r)}. \text{ This implies } \Omega_F(\theta_0, \beta)A_{0F} = 0^{k \times (k-r)}, \text{ which implies that } N(\Omega_F(\theta_0, \beta^*)) \subset N(\Omega_F(\theta_0, \beta)). \text{ Thus, in the setup of (13.9), condition (iv) of } \mathcal{F}_{AR,2}^{S,SR} \text{ holds provided } N(\Omega_F(\theta_0, \beta^*)) \text{ is not a strict subset of } N(\Omega_F(\theta_0, \beta)). \]
Note that condition (iv) of $\mathcal{F}^{S,S_R}_{AR,2}$ implies that $\tau_F(\beta)$ is constant for all $\beta \in B(\beta^*, \zeta_1)$. Furthermore, it implies that $\text{col}(\Omega_F(\theta_0, \beta^*)) = \text{col}(\Omega_F(\theta_0, \beta))$ for all $\beta \in B(\beta^*, \zeta_1)$, i.e., that $\text{col}(A_F(\beta^*)) = \text{col}(A_F(\beta))$ for all $\beta \in B(\beta^*, \zeta_1)$. Therefore, without loss of generality, under condition (iv) of $\mathcal{F}^{S,S_R}_{AR,2}$, we can take $A_F(\beta) = A_F(\beta^*)$ for all $\beta \in B(\beta^*, \zeta_1)$, i.e., $A_F(\beta)$ does not depend on $\beta$ for all $\beta \in B(\beta^*, \zeta_1)$.

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 $\mathcal{F}^{S,S_R}_{AR}$ and $\mathcal{F}^{S,S_R}$, 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 CS’s for the parameter spaces $\mathcal{F}^{S,S_R}_{\Theta,AR}$ and $\mathcal{F}^{S,S_R}_{\Theta}$.

**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 (4.8) of AG2 holds by the following calculations. For notational simplicity, we suppress the dependence of quantities on $\theta$. We have

$$SR-AR_n = n\tilde{g}_nA_n(\hat{A}_n^\top \hat{\Omega}_n \hat{A}_n)^{-1}A_n^\top \hat{g}_n$$

$$= n\tilde{g}_nA_n(\hat{A}_n^\top \hat{\Omega}_n \hat{A}_n)^{-1}A_n^\top \hat{g}_n = n\tilde{g}_nA_n \hat{\Pi}_n^{-1}A_n^\top \hat{g}_n$$

$$n\tilde{g}_n\hat{\Omega}_n^+\hat{g}_n = n\tilde{g}_nA_n(\hat{A}_n)^\top \left[ \begin{array}{c} \hat{\Pi}_n^{-1} \\ 0_{(k-r_n)\times(k-r_n)} \end{array} \right] \left[ \begin{array}{c} 0_{r_n \times(k-r_n)} \\ 0_{(k-r_n)(k-r_n)} \end{array} \right] [\hat{A}_n, \hat{A}_n^\top] \hat{g}_n = n\tilde{g}_nA_n \hat{\Pi}_n^{-1}A_n^\top \hat{g}_n, (14.1)$$

where the spectral decomposition of $\hat{\Omega}_n$ given in (4.3) and (4.4) is used once in each equation above. It is not the case that $SR-AR_n(\theta)$ equals the rhs expression in (4.8) with probability one when $\hat{\Omega}_n^+(\theta)$ is replaced by an arbitrary generalized inverse of $\hat{\Omega}_n(\theta)$.

The expression for the SR-AR statistic given in (4.6) is preferable to the Moore-Penrose expression in (4.8) 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 \(h_1\) and \(h_2\) 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 IV’s, \(u_i(\theta)\) 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, e.g., 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 IV’s 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\textsubscript{P} Parameter Space

When (15.1) holds, we define

$$u_{\theta i}(\theta) := \frac{\partial}{\partial \theta} u_i(\theta) \in \mathbb{R}^p \text{ and } u_{\theta i}(\theta) := \left( \begin{array}{c} u_i(\theta) \\ u_{\theta i}(\theta) \end{array} \right) \in \mathbb{R}^{p+1}, \text{ and we have } G_i(\theta) = Z_i u_{\theta i}(\theta)'$$  

(15.2)

The null parameter space for the SR-CQLR\textsubscript{P} test is

$$\mathcal{F}_P^{SR} := \{ F \in \mathcal{F}^{SR} : E_F \| \Pi_{1F}^{1/2} A_F Z_i \|^{4+\gamma} \leq M, \ E_F \| u_i^* \|^{2+\gamma} \leq M, \text{ and } E_F \| \Pi_{1F}^{-1/2} A_F Z_i \|^{2} u_i^{2} 1(u_i^{2} > c) \leq 1/2 \}, \quad (15.3)$$

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

The conditions in $\mathcal{F}_P^{SR}$ are only marginally stronger than those in $\mathcal{F}^{SR}$, defined in (3.6). A sufficient condition for the last condition in $\mathcal{F}_P^{SR}$ to hold for some $c < \infty$ is $E_F u_i^4 \leq M_0$ for some sufficiently large $M_0 < \infty$ (using the first condition in $\mathcal{F}_P^{SR}$ and the Cauchy-Bunyakovsky-Schwarz inequality).

The conditions in $\mathcal{F}_P^{SR}$ place no restrictions on the column rank or singular values of $E_F G_i$. The conditions in $\mathcal{F}_P^{SR}$ 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, $\Omega_F$ can be singular.

In Section 3.2 it is noted that identification failure yields singularity of $\Omega_F$ in likelihood scenarios. It also does so in all quasi-likelihood scenarios when the quasi-likelihood does not depend on some element(s) of $\theta$ (or some transformation(s) of $\theta$) for $\theta$ in a neighborhood of $\theta_0$.

Another example where $\Omega_F$ 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 $Z_i, \pi \in \mathbb{R}^{d_2}$ and $\theta = (\beta, \pi)' \in \mathbb{R}^{1+d_2}$.

\textsuperscript{30} As with $G(W_i, \theta)$ defined in (3.2), $u_{\theta i}(\theta)$ 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\textsubscript{2} test given below to hold, $u_{\theta i}(\theta)$ can be any random $p$ vector that satisfies the conditions in $\mathcal{F}_2^{SR}$ (defined in (15.3)).
\textsuperscript{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 $\theta_0$ (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.
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 \( EU_i = EV_{2i} = 0 \), \( EU_i Z_i = EV_{2i} Z_i = 0 \), and \( E(V_i V'_i | Z_i) = \Sigma_V \) a.s. for \( V_i := (V_{1i}, V_{2i})' \) and some \( 2 \times 2 \) constant matrix \( \Sigma_V \). The moment conditions for \( \theta \) are \( g_i(\theta) = ((y_{1i} - Z'_i \pi \beta)Z'_i, (Y_{2i} - Z'_i \pi)Z'_i)' \in \mathbb{R}^k \), where \( k = 2d_Z \). The variance matrix \( \Sigma_V \otimes EZ_i Z'_i \) 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 \( EZ_i Z'_i \) is singular. In this model, we are interested in joint inference concerning \( \beta \) and \( \pi \). This is of interest when one wants to see how the magnitude of \( \pi \) affects the range of plausible \( \beta \) values.

Section 3.2 and Grant (2013) note that \( \Omega_F \) can be singular in the model for interest rate dynamics in Jegannathan, Skoulakis, and Wang (2002, Sec. 6.2) (JSW). JSW consider five moment conditions and a four dimensional parameter \( \theta \). The first four moment functions in JSW are \((a(b - r_i)r_i^{-2\gamma} - \gamma \sigma^2 r_i^{-1}, a(b - r_i)r_i^{-2\gamma+1} - (\gamma - 1/2)\sigma^2, (b - r_i)r_i^{-\sigma} - (1/2)\sigma^2 r_i^{2\gamma - \sigma - 1}, a(b - r_i)r_i^{-\sigma} - (1/2)\sigma^2 r_i^{2\gamma - \sigma - 1})'\), 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 \((F, \theta)\) that is used with the SR-CQLR\(_P\) CS. It is denoted by \( \mathcal{F}_{\Theta,P}^{\text{SR}} \). For notational simplicity, the dependence of the parameter space \( \mathcal{F}_{P}^{\text{SR}} \) in (15.3) on \( \theta_0 \) is suppressed. When dealing with the SR-CQLR\(_P\) CS, rather than test, we make the dependence explicit and write it as \( \mathcal{F}_{P}^{\text{SR}}(\theta_0) \). We define

\[
\mathcal{F}_{\Theta,P}^{\text{SR}} := \{(F, \theta_0) : F \in \mathcal{F}_{P}^{\text{SR}}(\theta_0), \theta_0 \in \Theta\}. \tag{15.4}
\]

15.2 Definition of the SR-CQLR\(_P\) Test

First, we define the CQLR\(_P\) test without the SR extension. It uses the statistics \( \hat{g}_n(\theta), \hat{\Omega}_n(\theta), \) \( AR_n(\theta), \) and \( \hat{D}_n(\theta) \) (defined in (4.1), (4.2), and (5.2)). The CQLR\(_P\) test also uses analogues \( \hat{R}_n(\theta) \)
and $\tilde{V}_n(\theta)$ of $\tilde{R}_n(\theta)$ and $\tilde{V}_n(\theta)$ (defined in (5.3)), respectively, which are defined as follows:

$$\tilde{R}_n(\theta) := (B(\theta)' \otimes I_k) \tilde{V}_n(\theta) (B(\theta) \otimes I_k) \in R^{(p+1)k \times (p+1)k},$$

where

$$\tilde{V}_n(\theta) := n^{-1} \sum_{i=1}^{n} ((u_i^*(\theta) - \tilde{u}_m^*(\theta)) (u_i^*(\theta) - \tilde{u}_m^*(\theta))') \otimes (Z_i Z_i') \in R^{(p+1)k \times (p+1)k},$$

$$\tilde{u}_m^*(\theta) := \tilde{\Xi}_n(\theta)' Z_i \in R^{p+1},$$

$$\tilde{\Xi}_n(\theta) := (Z_n' Z_n)^{-1} Z_n' U^*(\theta) \in R^{k \times (p+1)},$$

$$Z_{n \times k} := (Z_1, ..., Z_n)' \in R^{n \times k}, \quad U^*(\theta) := (u_1^*(\theta), ..., u_n^*(\theta))' \in R^{n \times (p+1)},$$

$$B(\theta) := \begin{pmatrix} 1 & 0_p' \\ -\theta & -I_p \end{pmatrix} \in R^{(p+1) \times (p+1)},$$

(15.5)

where $u_i^*(\theta) := (u_i(\theta), u_{\theta i}(\theta))'$ is defined in (15.2). Note that (i) $\tilde{V}_n(\theta)$ is an estimator of the variance matrix of the moment functions and their vectorized derivatives, (ii) $\tilde{V}_n(\theta)$ exploits the functional form of the moment conditions given in (15.1), (iii) $\tilde{V}_n(\theta)$ typically is not of a Kronecker product form (because of the average over $i = 1, ..., n$), and (iv) $\tilde{u}_m^*(\theta)$ is the best linear predictor of $u_i^*(\theta)$ based on $\{Z_i : n \geq 1\}$. The estimators $\tilde{R}_n(\theta)$, $\tilde{V}_n(\theta)$, and $\tilde{\Sigma}_n(\theta)$ (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 IV’s 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 $\tilde{V}_n(\theta)$ (and the statistics that depend on it) differs from $\tilde{V}_n(\theta)$ (and the statistics that depend on it).

We define $\tilde{\Sigma}_n(\theta) \in R^{(p+1) \times (p+1)}$ just as $\tilde{\Sigma}_n(\theta)$ is defined in (5.4) and (5.5), but with $\tilde{R}_n(\theta)$ in place of $\tilde{R}_n(\theta)$. We define $\tilde{D}_n(\theta)$ just as $\tilde{D}_n(\theta)$ is defined in (5.7), but with $\tilde{\Sigma}_n(\theta)$ in place of $\tilde{\Sigma}_n(\theta)$. That is,

$$\tilde{D}_n(\theta) := \tilde{\Omega}_n(\theta)^{-1/2} \tilde{D}_n(\theta) \tilde{L}_n^{1/2}(\theta) \in R^{k \times p},$$

where $\tilde{L}_n(\theta) := (\theta, I_p)(\tilde{\Sigma}_n^\varepsilon(\theta))^{-1}(\theta, I_p)'$. (15.6)

The estimator $\tilde{\Sigma}_n(\theta)$ 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, $\tilde{\Sigma}_n(\theta)$ is an estimator of the variance matrix $\Sigma_V$ of the reduced-form errors when $\theta$ is the true parameter, and $\Sigma_V$ could be singular or nearly singular. In the definition of $\tilde{L}_n(\theta)$ above, we use an eigenvalue-adjusted version of $\tilde{\Sigma}_n(\theta)$, denoted by $\tilde{\Sigma}_n^\varepsilon(\theta)$, whose condition number (i.e., $\lambda_{\text{max}}(\tilde{\Sigma}_n(\theta))/\lambda_{\text{min}}(\tilde{\Sigma}_n(\theta))$) is bounded above by construction. Based on the finite-sample simulations, we recommend using $\varepsilon = .01$.

The QLR$_P$ statistic without the SR extension, denoted by $QLR_{Pn}(\theta)$, is defined just as $QLR_n(\theta)$
is defined in [5.7], but with \( \hat{D}_n^*(\theta) \) in place of \( \hat{D}_n^*(\theta) \). For \( \alpha \in (0, 1) \), the nominal size \( \alpha \) CQLR\(_P\) test (without the SR extension) rejects \( H_0 : \theta = \theta_0 \) if

\[
QLR_{P,n}(\theta_0) > c_{k,p}(n^{1/2} \hat{D}_n^*(\theta_0), 1 - \alpha),
\]

where \( c_{k,p}(\cdot, 1 - \alpha) \) is defined in [5.8]. The nominal size 100(1 - \( \alpha \))% CQLR\(_P\) CS is \( CS_{CQLR_{P,n}} := \{ \theta_0 \in \Theta : QLR_{P,n}(\theta_0) \leq c_{k,p}(n^{1/2} \hat{D}_n^*(\theta_0), 1 - \alpha) \} \).

The CQLR\(_P\) test statistic and critical value satisfy the following invariance properties.

**Lemma 15.1** The statistics \( QLR_{P,n}, c_{k,p}(n^{1/2} \hat{D}_n^*, 1 - \alpha), \hat{D}_n^* \hat{D}_n^*, AR_n, \hat{u}_n^*, \hat{\Sigma}_n, \) and \( \bar{L}_n \) are invariant to the transformation \((Z_i, u_i^*) \sim (MZ_i, u_i^*) \) \( \forall i \leq n \) for any \( k \times k \) nonsingular matrix \( M \). This transformation induces the following transformations: \( g_i \sim MG_i \) \( \forall i \leq n \), \( G_i \sim MG_i \) \( \forall i \leq n \), \( \bar{g}_n \sim M\bar{g}_n \), \( \bar{G}_n \sim MG_n \), \( \bar{O}_n \sim M\bar{O}_nM' \), \( \bar{\Gamma}_{jn} \sim M\bar{\Gamma}_{jn}M' \) \( \forall j \leq p \), \( \bar{D}_n \sim M\bar{D}_n \), \( Z_{n \times k} \sim Z_{n \times k}M' \), \( \bar{\Xi}_n \sim M^{-1}\bar{\Xi}_n \), \( \bar{V}_n \sim (I_{p+1} \otimes M)\bar{V}_n (I_{p+1} \otimes M') \), and \( \bar{R}_n \sim (I_{p+1} \otimes M)\bar{R}_n (I_{p+1} \otimes M') \).

**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 \( \Omega_F \) is nonsingular. The reason is that (in the proofs) one can transform the moments by \( g_i \sim MF_i g_i \), where \( MF_i \Omega_i F_i' = I_k \), such that the transformed moments have a variance matrix whose eigenvalues are bounded away from zero for some \( \delta > 0 \) (since \( \text{Var}_F(MF_i g_i) = I_k \)) even if the original moments \( g_i \) do not.

For the CQLR\(_P\) test with the SR extension, we define \( \hat{D}_{An}(\theta) \) as in [5.10]. We let \( Z_{Ai}(\theta) := \hat{A}_n(\theta)'Z_i \in R^{\hat{r}_n(\theta)} \) and \( Z_{An \times k}(\theta) := Z_{n \times k} \hat{A}_n(\theta) \in R^{n \times \hat{r}_n(\theta)} \). We define

\[
\bar{V}_{An}(\theta) := n^{-1} \sum_{i=1}^{n} \left( (u_i^*(\theta) - \hat{u}_{Ai}(\theta)) (u_i^*(\theta) - \hat{u}_{Ai}(\theta))^\prime \right) \otimes (Z_{Ai}(\theta)Z_{Ai}(\theta)^\prime)
\in R^{(p+1)\hat{r}_n(\theta) \times (p+1)\hat{r}_n(\theta)},
\]

where

\[
\hat{u}_{Ai}(\theta) := \hat{\Xi}_{An}(\theta)'Z_{Ai}(\theta) \in R^{p+1},
\]

\[
\bar{\Xi}_{An}(\theta) := (Z_{An \times k}(\theta)'Z_{An \times k}(\theta))^{-1}Z_{An \times k}(\theta)'U^*(\theta) \in R^{\hat{r}_n(\theta) \times (p+1)},
\]

and \( \hat{r}_n(\theta) \) and \( \hat{A}_n(\theta) \) are defined in [4.3] and [4.4], respectively. In addition, we define \( \bar{R}_{An}(\theta), \bar{\Sigma}_{An}(\theta), \bar{L}_{An}(\theta), \bar{D}_{An}(\theta), \) and \( \bar{Q}_{An}(\theta) \) as \( \hat{R}_n(\theta), \hat{\Sigma}_n(\theta), \hat{L}_n(\theta), \hat{D}_n(\theta), \) and \( \hat{Q}_n(\theta) \) are defined, respectively, in [5.11] and [5.12], but with \( \bar{V}_{An}(\theta) \) in place of \( \bar{V}_{An}(\theta) \) in the definition of \( \bar{R}_{An}(\theta) \), with \( \bar{R}_{An}(\theta) \) in place of \( \bar{R}_{An}(\theta) \) in the definition of \( \bar{\Sigma}_{An}(\theta) \), and so on in the definitions of \( \bar{L}_{An}(\theta), \bar{D}_{An}(\theta), \) and \( \bar{Q}_{An}(\theta) \). We define the test statistic \( SR-QLR_{P,n}(\theta) \) as \( SR-QLR_{n}(\theta) \) is defined in [5.12], but with \( \bar{Q}_{An}(\theta) \) in place of \( \hat{Q}_n(\theta) \).
Given these definitions, the nominal size $\alpha$ SR-CQLR$_P$ test rejects $H_0 : \theta = \theta_0$ if

$$SR-CQLR_{Pn}(\theta_0) > c_{\alpha}(n, \theta_0, p)(n^{1/2} \tilde{D}_{An}(\theta_0), 1 - \alpha) \text{ or } A_n^+(\theta_0)^Tg_n(\theta_0) \neq 0 \text{ for } n \to \infty \tag{15.9}$$

The nominal size $100(1 - \alpha)$% SR-CQLR$_P$ CS is $CS_{SR-CQLR_{Pn}} := \{\theta_0 \in \Theta : SR-CQLR_{Pn}(\theta_0) \leq c_{\alpha}(n, \theta_0, p)(n^{1/2} \tilde{D}_{An}(\theta_0), 1 - \alpha) \text{ and } A_n^+(\theta_0)^Tg_n(\theta_0) = 0 \}$.

Two simple examples where the extra rejection condition in (15.9) for the SR-CQLR$_P$ test (and in (4.7) and (5.13) for the SR-AR and SR-CQLR tests, respectively) improves the power of these tests are the following. First, suppose $(X_{1i}, X_{2i})' \sim i.i.d. N(\theta, \Omega_F)$, where $\theta = (\theta_1, \theta_2)' \in R^2$, $\Omega_F$ is a $2 \times 2$ matrix of ones, and the moment functions are $g_i(\theta) = (X_{1i} - \theta_1, X_{2i} - \theta_2)'$. In this case, $\Omega_F$ is singular, $\tilde{A}_n(\theta_0) = (1, 1)' \text{ a.s.}, \tilde{A}_n^+(\theta_0) = (1, -1)' \text{ a.s.},$ the SR-AR statistic is a quadratic form in $\tilde{A}_n(\theta_0)^Tg_n(\theta_0) = \sum_{i=1}^n X_{1i} - \sum_{i=1}^n X_{2i} - (\theta_1 \theta_2 + \theta_2 \theta_1)$, where $\sum_{i=1}^n X_{mi}$ for $m = 1, 2$, and $A_n^+(\theta_0)^Tg_n(\theta_0) = \sum_{i=1}^n X_{1i} - \sum_{i=1}^n X_{2i} - (\theta_1 \theta_2 + \theta_2 \theta_1)$ a.s. If one does not use the extra rejection condition, then the SR-AR test has no power against alternatives $\theta = (\theta_1, \theta_2)' (\neq 0)$ for which $\theta_1 + \theta_2 = 0$. The same is true for the SR-CQLR and SR-CQLR$_P$ tests (because the SR-CQLR$_n$ and SR-CQLR$_{Pn}$ 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 $\theta_1 - \theta_2 = \theta_1 + \theta_2$ are rejected with probability one (because $\sum_{i=1}^n X_{1i} - \sum_{i=1}^n X_{2i} = EFX_{1i} - EFX_{2i} = \theta_1 - \theta_2 \text{ a.s.}$) and this includes all of the alternative $\theta$ values for which $\theta_1 + \theta_2 = 0$.

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 $\theta$, $X_i$ has variance zero, $X_i = X_{ni} = \theta_1$ a.s., $X_{2i} = \sum_{i=1}^n X_{2i} = \theta_1^2 \text{ a.s.},$ where $\sum_{i=1}^n X_{2i}^2 := n^{-1} \sum_{i=1}^n X_{2i}^2$, $\tilde{g}_n(\theta_0) = (\theta_1 - \theta_{10}, \theta_1^2 - \theta_{10}^2 - \theta_{20})'$ a.s., $\tilde{\Omega}_n(\theta_0) = \tilde{g}_n(\theta_0)\tilde{g}_n(\theta_0)' - \tilde{g}_n(\theta_0)\tilde{g}_n(\theta_0)' = 0^{2 \times 2}$ a.s. (provided $\tilde{\Omega}_n(\theta_0)$ is defined as in (4.1) with the sample means subtracted off), and $\tilde{\tau}_n(\theta_0) = 0$ 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 $\tilde{\tau}_n(\theta_0) = 0$). However, when the extra rejection condition is utilized, all alternatives of the form

32 By definition, $\tilde{A}_n^+(\theta_0)^Tg_n(\theta_0) \neq 0$ does not hold if $\tilde{\tau}_n(\theta_0) = k$. If $\tilde{\tau}_n(\theta_0) = 0$, then $SR-CQLR_{Pn}(\theta_0) := 0$ and $\chi^2_{\alpha}(\theta_0, 1 - \alpha) := 0$. In this case, $\tilde{A}_n^+(\theta_0) = I_k$ and the SR-CQLR$_P$ test rejects $H_0$ if $\tilde{g}_n(\theta_0) \neq 0$.

33 By definition, if $\tilde{\tau}_n(\theta_0) = k$, the condition $\tilde{A}_n^+(\theta_0)^Tg_n(\theta_0) = 0$ holds.
\( \theta = (\theta_1, 0)' \) are rejected with probability one.

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 \( \theta_0 = (\theta_{01}, \theta_{02})' \in \mathbb{R}^2 \). 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 \( \theta_0 \) 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 \( \theta_1 \) and \( \theta_2 \) 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 \( \theta_1 \), whereas the second set does.

34This 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 \( \theta_2^2 - \theta_{10}^2 - \theta_{20} \neq 0 \) (because \( \overline{X}_n = \theta_1 \) a.s. and \( \overline{X}_n^2 = \theta_2^2 \) a.s. under \( \theta \)), 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 \( \theta_2 = 0 \)).

35In 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 \( \hat{g}_n(\theta_0) = (\overline{X}_n - \theta_{10}, \overline{X}_n^2 - \theta_{10}^2)' = (0, 0)' \) a.s., \( \hat{r}_n(\theta_0) = 0 \) a.s., \( SR-AR_n(\theta_0) = \chi^2_{\hat{r}_n(\theta_0), 1-\alpha} = 0 \) a.s. (because \( \hat{r}_n(\theta_0) = 0 \) 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 \( X_1 = \theta_1 \) 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 \( g_i(\theta_0) = 0 \) 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.

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

37An 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 (4.7), one defines the SR-AR statistic using a weight matrix that is nonsingular by construction when \( \bar{\Omega}_n(\theta_0) \) is singular, and one determines the critical value by simulation of the appropriate quadratic form in mean zero normal variates when \( \bar{\Omega}_n(\theta_0) \) is singular. For example, such a weight matrix can be constructed by adjusting the eigenvalues of \( \bar{\Omega}_n(\theta_0) \) 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 (4.7). 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-CQLR2 tests introduced in Sections 9 and 15. The reason is that these tests rely on the statistic \( \bar{D}_n(\theta_0) \), defined in (5.2), that employs \( \bar{\Omega}^{-1}_n(\theta_0) \) and if \( \bar{\Omega}^{-1}_n(\theta_0) \) 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 \( \hat{g}_n(\theta_0) \) and \( \bar{D}_n(\theta_0) \) after suitable normalization, which is needed to obtain the correct asymptotic size of the SR-CQLR and SR-CQLR2 tests.
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 \( \mathcal{F}^{SR}_P \). 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 \( \mathcal{F}^{SR}_{G,P} \), defined in (15.4).

**Comments:** (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 and 25 below.

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 \( \Pi_F \) be defined as in (3.4) and (3.5) and the paragraph following these equations with \( \theta = \theta_0 \). Define \( \lambda^*_F, \lambda^*_n, \{ \lambda^*_n,h : n \geq 1 \} \) as \( \lambda_F, \lambda_{W,U,F}, \) and \( \{ \lambda^*_n,h : n \geq 1 \} \), respectively, are defined in (16.16)-(16.18), but with \( g_i \) and \( G_i \) replaced by \( g^*_F := \Pi^{-1/2}_F A'_F g_i \) and \( G^*_F := \Pi^{-1/2}_F A'_F G_i \), with \( \mathcal{F}_F \) replaced by \( \mathcal{F}^{SR}_P \) in the definition of \( \mathcal{F}_{W,U} \), and with \( W_F := W(\Pi^{-1}_F) \) and \( U_F := U(\Pi^{-1}_F) \) defined as in (16.11) with \( g_i \) and \( G_i \) replaced by \( g^*_F \) and \( G^*_F \). In addition, we restrict \( \{ \lambda^*_n,h : n \geq 1 \} \) to be a sequence for which \( \lambda_{min}(E_F,g_i g'_i) > 0 \) 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 \( n^{1/2} s^*_p F_n \to \infty \), where \( s^*_p \) denotes the smallest singular value of \( E_F G^*_F \).

The \( LM_n \) and \( LM_n^{GMM} \) statistics are defined in (7.1). Let \( \chi^2_{p,1-\alpha} \) denote the \( 1 - \alpha \) quantile of the \( \chi^2_p \) distribution. The critical value for the \( LM_n \) and \( LM_n^{GMM} \) 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 \( n^{1/2} s^*_p F_n \to \infty \) and for which \( \lambda^*_n,h \in \Lambda_F^* \) \( \forall \) \( n \geq 1 \), we have

---

38 The singular value \( s^*_p \), defined here, equals \( s^*_p \), defined in Section 6.2 for all \( F \) with \( \lambda_{min}(\Omega_F) > 0 \), because in this case \( \Omega_F = A_F \Pi F_A F' \), \( \Omega_F^{1/2} = A_F \Pi F A' F \), \( \Omega_F^{1/2} E F G_i = A_F \Pi F A' F E F G_i = A_F \Pi F A' F E F G_i \), and \( A_F \) is an orthogonal \( k \times k \) matrix. Since we consider sequences here with \( \lambda_{min}(\Omega_{F_i}) = \lambda_{min}(E_F, g_i g'_i) > 0 \) for all \( n \geq 1 \), the definitions of strong and semi-strong identification used here and in Section 6.2 are equivalent.
(a) $SR-QLR_{Pn} = QLR_{Pn} + o_p(1) = LM_n + o_p(1) = LM_{n}^{GMM} + o_p(1)$ and
(b) $c_{k,p}(n^{1/2}D^*_n, 1 - \alpha) \rightarrow_p \chi^2_{p,1-\alpha}.$

Comments: (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
non-singularity of the variance matrix and some other 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 JLV-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 JLV-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 $F_{SR}$ 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 $F_{SR}^P$ ($\subset F_{SR}$). (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 IV’s 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 CS’s without the SR extension for parameter spaces
of distributions that bound $\lambda_{\min}(\Omega_F)$ away from zero. (These tests are defined in (4.2), (5.9), and
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 CS’s. 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 \(H_0: \theta = \theta_0\):

\[
\mathcal{F}_{AR} := \{ F : E_F g_i = 0^k, E_F \|g_i\|^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(E_F g_i g_i^\top) \geq \delta \},
\]

\[
\mathcal{F} := \{ F \in \mathcal{F}_{AR} : E_F \|\text{vec}(G_i)\|^{2+\gamma} \leq M \}, \text{ and}
\]

\[
\mathcal{F}_P := \{ F \in \mathcal{F} : E_F \|Z_i\|^{4+\gamma} \leq M, E_F \|u_i^*\|^{2+\gamma} \leq M, \lambda_{\min}(E_F Z_i Z_i^\top) \geq \delta \} \quad (16.1)
\]

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

**Theorem 16.1** The AR, CQLR, and CQLR\(_P\) tests (without the SR extensions), defined in (4.2), (5.9), 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 \(\mathcal{F}_{AR}, \mathcal{F}, \) and \(\mathcal{F}_P\), respectively. Analogous results hold for the corresponding AR, CQLR, and CQLR\(_P\) CS’s for the parameter spaces \(\mathcal{F}_{\Theta,AR}, \mathcal{F}_{\Theta}, \) and \(\mathcal{F}_{\Theta,P}\), respectively.

**Comments:** (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 non-empty.

(iv) The results of Theorem 6.1 still hold if the moment bounds in \(\mathcal{F}_{AR}, \mathcal{F}, \) and \(\mathcal{F}_P\) are weakened very slightly by, e.g., replacing \(E_F \|g_i\|^{2+\gamma} \leq M\) in \(\mathcal{F}_{AR}\) by \(E_F \|g_i\|^2(\|g_i\| > j) \leq \varepsilon_j\) 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 \(\gamma, j > 0\), \(EX^21(|X| > j) \leq EX^{2+\gamma}/j^\gamma\). The latter conditions allow for the application of Lindeberg’s triangular array central limit theorem for independent random variables, e.g., see Billingsley (1979, 1986).
Thm. 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 $F_{AR}$, $F^{SR}$, and $F_{P}$ are weakened very slightly by, e.g., replacing $E_F||\Pi_F^{-1/2}A_Fg_i||^2+\gamma \leq M$ in $F_{AR}$ by $E_F||\Pi_F^{-1/2}A_Fg_i||^2(||\Pi_F^{-1/2}A_Fg_i|| > j) \leq \varepsilon_j$ 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 $\Upsilon$ 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 $\tau_j$ is nonincreasing in $j$. Then, $c_{k,p}(D,1-\alpha) = c_{k,p}(\Upsilon,1-\alpha)$.

**Comment:** A consequence of Lemma 16.2 is that the critical value $c_{k,p}(n^{1/2}\hat{D}_n^*(\theta_0),1-\alpha)$ of the CQLR test depends on $\hat{D}_n^*(\theta_0)$ only through $\hat{D}_n^*(\theta_0)'\hat{D}_n^*(\theta_0)$ (because, when $k \geq p$, the $p$ singular values of $n^{1/2}\hat{D}_n^*(\theta_0)$ equal the square roots of the eigenvalues of $n\hat{D}_n^*(\theta_0)'\hat{D}_n^*(\theta_0)$ and, when $k < p$, $c_{k,p}(D,1-\alpha)$ is the $1-\alpha$ quantile of the $\chi^2_k$ 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 $\lambda$ with parameter space $\Lambda$. Let $RP_n(\lambda)$ denote the null rejection probability of $\phi_n$ under $\lambda$. For a finite nonnegative integer $J$, let $\{h_n(\lambda) = (h_{1n}(\lambda),...,h_{Jn}(\lambda))' \in \mathbb{R}^J : n \geq 1\}$ be a sequence of functions on $\Lambda$. Define

$$H := \{h \in (\mathbb{R} \cup \{\pm \infty\})^J : h_{wn}(\lambda_{wn}) \to h \text{ for some subsequence } \{wn\}\}$$

of $\{n\}$ and some sequence $\{\lambda_{wn} \in \Lambda : n \geq 1\}$. \hspace{1cm} (16.2)

**Assumption B**: For any subsequence $\{wn\}$ of $\{n\}$ and any sequence $\{\lambda_{wn} \in \Lambda : n \geq 1\}$ for which $h_{wn}(\lambda_{wn}) \to h \in H$, $RP_{wn}(\lambda_{wn}) \to \alpha$ 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 \( \alpha \) and are asymptotically similar (in a uniform sense). That is, \( \text{AsySz} := \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} R_P(n) = \alpha \) and \( \liminf_{n \to \infty} \inf_{\lambda \in \Lambda} R_P(n) = \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} R_P(n) \).

Comments: (i) By Comment 4 to Theorem 2.1 of ACG, Proposition 16.3 provides asymptotic size and similarity results for nominal \( 1 - \alpha \) CS’s, rather than tests, by defining \( \alpha \) 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 \( \Lambda \) correspondingly (so it includes all possible values of the parameter that is restricted by the null hypothesis), and by replacing (a) \( \phi_n \) by a CS based on a sample of size \( n \), (b) \( \alpha \) by \( 1 - \alpha \), (c) \( R_P(n) \) by \( C_P(n) \), where \( C_P(n) \) denotes the coverage probability of the CS under \( \lambda \) 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 CS’s, the parameter \( \theta_0 \) is taken to be a subvector of \( \lambda \) and \( \Lambda \) is specified so that the value of this subvector ranges over \( \Theta \).

(ii) In the application of Proposition 16.3 to prove Theorems 6.1, 15.2, and 16.1, one takes \( \Lambda \) to be a one-to-one transformation of \( F_{AR}, F \), or \( F_P \) for tests, and one takes \( \Lambda \) to be a one-to-one transformation of \( F_{\Theta,AR}, F_{\Theta} \), or \( F_{\Theta,P} \) for CS’s. With these changes, the proofs for tests and CS’s are the same. In consequence, we provide explicit proofs for tests only and obtain the proofs for CS’s by analogous applications of Proposition 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 \( \lambda, h_n(\lambda), \) and \( \Lambda \). 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 \( \hat{W}_n \) and \( \hat{U}_n \)

As above, for notational simplicity, we suppress the dependence on \( \theta_0 \) of many quantities, such as \( g_i, G_i, u_{q_i}, B, \) and \( f_i \), as well as the quantities \( V_F, R_F, \Xi_F, \bar{V}_F, \) and \( \bar{R}_F \), 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 \( \hat{W}_n \in R^{k \times k} \) and \( \hat{U}_n \in R^{p \times p} \). In particular, we consider statistics of the
form $\hat{W}_n \hat{D}_n \hat{U}_n$ and functions of this statistic, where $\hat{D}_n$ is defined in (5.2). Let

$$QLR_{WU,n} := AR_n - \lambda_{\min}(n\hat{Q}_{WU,n}),$$
where

$$\hat{Q}_{WU,n} := \left(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{g}_n\right)' \left(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{g}_n\right) \in R^{(p+1) \times (p+1)}.$$ (16.3)

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

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

where $\hat{W}_{2n}$ and $\hat{U}_{2n}$ 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 $\hat{W}_{2n}$ and $\hat{U}_{2n}$ have corresponding population quantities $W_{2F}$ and $U_{2F}$, respectively. Thus, the population quantities corresponding to $\hat{W}_n$ and $\hat{U}_n$ are

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

respectively.

**Example 1:** For the CQLR test,

$$\hat{W}_n := \hat{\Omega}_n^{-1/2} \text{ and } \hat{U}_n := \hat{I}_n^{1/2} := ((\theta_0, I_p)(\hat{\Sigma}_n^2)^{-1}(\theta_0, I_p)'')^{1/2},$$ (16.6)

where $\hat{\Omega}_n$ is defined in (4.1) and $\hat{\Sigma}_n$ is defined in (5.4) and (5.5).

The population analogues of $\hat{\Omega}_n$ and $\hat{I}_n$, defined in (5.3), are

$$V_F := E_F(f_i - E_F f_i)(f_i - E_F f_i)' \in R^{(p+1)k \times (p+1)k} \text{ and }$$

$$R_F := (B' \otimes I_k)V_F(B \otimes I_k) \in R^{(p+1)k \times (p+1)k}.$$ (16.7)

The definition of $\hat{Q}_{WU,n}$ in (16.3) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{g}_n)$, whereas (5.7) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\hat{\Omega}_n^{-1/2} \hat{g}_n, \hat{D}_n^*)$, which has the $\hat{\Omega}_n^{-1/2} \hat{g}_n$ 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 $(\hat{\Omega}_n^{-1/2} \hat{g}_n, \hat{D}_n^*)$ in (5.7) because it is consistent with the order in Moreira (2003) and Andrews, Moreira, and Stock (2006, 2008). We use the order $(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{g}_n)$ here because it has significant notational advantages in the proof of Theorem [16.6] below, which is given in Section [26]
In this case,

\[
\hat{W}_{2n} := \hat{\Omega}_n, \quad W_{2F} := \Omega_F := \hat{E}_F g_i g_i', \quad W_1(W_{2F}) := W_{2F}^{-1/2},
\]

\[
U_1(U_{2F}) := ((\theta_0, I_p)(\Sigma^e(\Omega_F, R_F))^{-1}(\theta_0, I_p))^{1/2},
\]

\[
\hat{U}_{2n} := (\bar{\Omega}_n, \bar{R}_n), \quad U_{2F} := (\Omega_F, R_F), \quad \text{and}
\]

\[
\Sigma_{j\ell}(\Omega_F, R_F) = tr(R'_{j\ell F}\Omega_F^{-1})/k \tag{16.8}
\]

for \( j, \ell = 1, \ldots, p + 1 \), where \( \Sigma_{j\ell}(\Omega_F, R_F) \in R^{(p+1)\times(p+1)} \) denotes the \((j, \ell)\) 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 \( \hat{\Sigma}_n \) in (5.4)), the last equality in (16.8) holds by the same argument as used to obtain (5.5), \( \Sigma^e(\Omega_F, R_F) \) is defined given \( \Sigma(\Omega_F, R_F) \) by (5.6), and \( R'_{j\ell F} \) denotes the \((j, \ell)\) \( k \times k \) submatrix of \( R_F \).

**Example 2:** For the CQLR\(_P\) test, one takes \( \hat{W}_{2n}, \hat{W}_{2n}, W_{2F}, W_1(\cdot), \) and \( U_1(\cdot) \) as in Example 1 and

\[
\hat{U}_n := \hat{L}_n^{1/2} := ((\theta_0, I_p)(\hat{\Sigma}_n)^{-1}(\theta_0, I_p))^{1/2}, \tag{16.9}
\]

where \( \hat{\Sigma}_n = \hat{\Sigma}_n(\theta_0) \) is defined just above (15.5) and \( \hat{\Sigma}_n^e \) is defined given \( \hat{\Sigma}_n \) by (5.6).

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

\[
\tilde{V}_F := E_F f_i f'_i - E_F((g_i, G_i)\Xi_F \otimes Z_i Z'_i) - E_F(\Xi_F(g_i, G_i) \otimes Z_i Z'_i) + E_F(\Xi'_F Z_i Z'_i \Xi_F \otimes Z_i Z'_i) \in R^{(p+1)k \times (p+1)k}
\]

and

\[
\tilde{R}_F := (B' \otimes I_k)\tilde{V}_F(B \otimes I_k) \in R^{(p+1)k \times (p+1)k}, \quad \text{where}
\]

\[
\Xi_F := (E_F Z_i Z'_i)^{-1} E_F(g_i, G_i) \in R^{k \times (p+1)}, \quad f_i := (g'_i, vec(G_i)'y) \in R^{(p+1)k},
\]

and \( B = B(\theta_0) \) is defined in (5.3).

For the CQLR\(_P\) test,

\[
\hat{U}_{2n} := (\bar{\Omega}_n, \bar{R}_n), \quad U_{2F} := (\Omega_F, \bar{R}_F), \quad \text{and}
\]

\[
\Sigma_{j\ell}(\Omega_F, \bar{R}_F) = tr(R'_{j\ell F}\Omega_F^{-1})/k, \tag{16.11}
\]

for \( j, \ell = 1, \ldots, p + 1 \), where \( \Sigma_{j\ell}(\Omega_F, \bar{R}_F) \in R^{(p+1)\times(p+1)} \) denotes the \((j, \ell)\) element \( \Sigma(\Omega_F, \bar{R}_F) \), \( \Sigma(\Omega_F, \bar{R}_F) \) is defined to minimize \(||(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - \bar{R}_F](I_{p+1} \otimes \Omega_F^{-1/2})||\) over symmetric pd matrices \( \Sigma \in R^{(p+1)\times(p+1)} \) (analogously to the definition of \( \hat{\Sigma}_n(\theta) \) in (5.4)), the last equality in (16.11)

\(^{40}\)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 \( \hat{W}_{2n} \) and \( \hat{U}_{2n} \), not just for \( W_{2F} \) and \( U_{2F} \).
(16.11) holds by the same argument as used to obtain (5.5), \( \Sigma(\Omega_F, \tilde{R}_F) \) is defined given \( \Sigma(\Omega_F, \tilde{R}_F) \) by (5.6), and \( \tilde{R}_{j\ell F} \) denotes the \((j, \ell)\) \(k \times k\) submatrix of \( \tilde{R}_F \).

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

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

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 \( \mathcal{F} \subset \mathcal{F}_{WU} \) for \( \delta_1 > 0 \) sufficiently small and \( M_1 < \infty \) sufficiently large, see Lemma 27.4(a). Hence, uniform results over \( \mathcal{F}_{WU} \) 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 \( \mathcal{F}_P \subset \mathcal{F}_{WU} \) 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 \( \mathcal{F}_P \cap \mathcal{F}_{WU} \) 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 \( \lambda \). The vector \( \lambda \) is chosen such that for a subvector of \( \lambda \) 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 \( \lambda \) for the CQLR and CQLR\(_P\) tests. The same definition is used for both tests. The (much simpler) definition of \( \lambda \) for the AR test is given in Section 27.6 below.

The vector \( \lambda \) depends on the following quantities. Let

\[
B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } U_F'(E_FG_i)'W_F'W_F(E_FG_i)U_F \tag{16.13}
\]

ordered so that the corresponding eigenvalues \((\kappa_1 F, ..., \kappa_p F)\) are nonincreasing. The matrix \( B_F \) is such that the columns of \( W_F(E_FG_i)U_FB_F \) are orthogonal. Let

\[
C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } W_F(E_FG_i)U_FU_F'(E_FG_i)'W_F' \tag{16.14}
\]

\( ^{41} \)The matrices \( B_F \) and \( C_F \) are not uniquely defined. We let \( B_F \) denote one choice of the matrix of eigenvectors of \( U_F'(E_FG_i)'W_F'W_F(E_FG_i)U_F \) and analogously for \( C_F \).
The corresponding eigenvalues are \((\kappa_1, ..., \kappa_k) \in \mathbb{R}^k\). Let

\[
(\tau_1, ..., \tau_{\min\{k,p\}}) \text{ denote the } \min\{k,p\} \text{ singular values of } W_F(E_FG_i)U_F, \tag{16.15}
\]

which are nonnegative, ordered so that \(\tau_j \leq \tau_{j+1}\). (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, \(\kappa_j = \tau^2_j\) for \(j = 1, ..., \min\{k,p\}\). In addition, \(\kappa_j = 0\) for \(j = \min\{k,p\} + 1, ..., \max\{k,p\}\).

Define the elements of \(\lambda\) to be\(^{42,43}\)

\[
\begin{align*}
\lambda_{1,F} &:= (\tau_1, ..., \tau_{\min\{k,p\}})' \in \mathbb{R}^{\min\{k,p\}}, \\
\lambda_{2,F} &:= B_F \in \mathbb{R}^{p \times p}, \\
\lambda_{3,F} &:= C_F \in \mathbb{R}^{k \times k}, \\
\lambda_{4,F} &:= E_FG_i \in \mathbb{R}^{k \times p}, \\
\lambda_{5,F} &:= E_F \left( \begin{array}{c} g_i \\ \text{vec}(G_i) \end{array} \right) \left( \begin{array}{c} g_i \\ \text{vec}(G_i) \end{array} \right)' \in \mathbb{R}^{(p+1)k \times (p+1)k}, \\
\lambda_{6,F} &:= (\lambda_{6,1,F}, ..., \lambda_{6,\min\{k,p\}-1,F})' := \left( \begin{array}{c} \frac{\tau_2}{\tau_1}, ..., \frac{\tau_{\min\{k,p\}}}{\tau_{\min\{k,p\}-1}} \end{array} \right) \in [0, 1]^{\min\{k,p\}-1}, \text{ where } 0/0 := 0, \\
\lambda_{7,F} &:= W_2, \\
\lambda_{8,F} &:= U_2, \\
\lambda_{9,F} &:= F, \text{ and} \\
\lambda = \lambda_F &:= (\lambda_{1,F}, ..., \lambda_{9,F}). \tag{16.16}
\end{align*}
\]

The dimensions of \(W_2\) and \(U_2\) depend on the choices of \(\hat{W}_n = W_1(\hat{W}_{2n})\) and \(\hat{U}_n = U_1(\hat{U}_{2n})\). We let \(\lambda_{5,g,F}\) denote the upper left \(k \times k\) submatrix of \(\lambda_{5,F}\). Thus, \(\lambda_{5,g,F} = E_Fg_i'g_i' = \Omega_F\). We consider two parameter spaces for \(\lambda\): \(\Lambda_{WU}\) and \(\Lambda_{WU,P}\), which correspond to \(\mathcal{F}_{WU}\) and \(\mathcal{F}_{WU} \cap \mathcal{F}_P\), respectively, where \(\mathcal{F}_P\) and \(\mathcal{F}_{WU}\) are defined in \((16.11)\) and \((16.12)\), respectively. The space \(\Lambda_{WU}\) is used for the CQLR test. The space \(\Lambda_{WU,P}\) is used for the CQLR\(_P\) test\(^{44}\). The parameter spaces \(\Lambda_{WU}\) and

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

\(^{43}\)If \(p = 1\), no vector \(\lambda_{6,F}\) appears in \(\lambda\) because \(\lambda_{1,F}\) only contains a single element.

\(^{44}\)Note that the parameter \(\lambda\) has different meanings for the CQLR and CQLR\(_P\) tests because \(U_2\) is different for the two tests.
\( \Lambda_{W,U,P} \) and the function \( h_n(\lambda) \) are defined by

\[
\begin{align*}
\Lambda_{W,U} &:= \{ \lambda : \lambda = (\lambda_{1,F}, \ldots, \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{W,U} \}, \\
\Lambda_{W,U,P} &:= \{ \lambda : \lambda = (\lambda_{1,F}, \ldots, \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{W,U} \cap \mathcal{F}_P \}, \text{ and} \\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}).
\end{align*}
\]

(16.17)

By the definition of \( \mathcal{F} \), \( \Lambda_{W,U} \) and \( \Lambda_{W,U,P} \) 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 \( \lambda \) 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 \( h_n(\lambda_{F_n}) \to h \in H \) 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,

\[
\{\lambda_{n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_n \in \Lambda_{W,U} : n \geq 1\} \text{ for which } h_n(\lambda_n) \to h \in H \quad (16.18)
\]

for \( H \) defined in (16.2) with \( \Lambda \) equal to \( \Lambda_{W,U} \). By the definitions of \( \Lambda_{W,U} \) and \( \mathcal{F}_{W,U} \), \( \{\lambda_{n,h} : n \geq 1\} \) 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 \( \lambda \): \( h = (h_1, \ldots, h_8) \), 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 \( \infty \). 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

\[
n^{1/2} \tau_{jF_n} \to h_{1,j} \geq 0 \ \forall j \leq \min\{k,p\}, \quad \lambda_{m,F_n} \to h_m \ \forall m = 2, \ldots, 8,
\]

\[
\lambda_{5,gF_n} = \Omega_{F_n} = E_{F_n} g_i g_i' \to h_{5,g}, \text{ and } \lambda_{6,jF_n} \to h_{6,j} \ \forall j = 1, \ldots, \min\{k,p\} - 1. \quad (16.19)
\]

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

\( ^{45} \)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 \( h_{w_n}(\lambda_{w_n}) \to h \in H \).
16.5 Assumption WU

We assume that the random weight matrices $\hat{W}_n = W_1(\hat{W}_2n)$ and $\hat{U}_n = U_1(\hat{U}_2n)$ defined in (16.4) satisfy the following assumption that depends on a suitably chosen parameter space $\Lambda_*$ ($\subset \Lambda_{WU}$), such as $\Lambda_{WU}$ or $\Lambda_{WU,P}$.

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

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

(b) $\hat{U}_{2w_n} \rightarrow_p h_8 := \lim U_{2F_{w_n}}$, 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 $\hat{W}_{2w_n}$ wp→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 $\hat{U}_{2w_n}$ wp→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 $\Lambda$ equal to $\Lambda_{WU}$, and likewise with $n$ in place of $w_n$.

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

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

$$\Phi_F^{vec(G_i)} := Var_F(vec(G_i) - (E_F vec(G_i) g_i') \Omega^{-1}_F g_i)$$

and

$$\Phi_h^{vec(G_i)} := \lim \Phi_{F_{w_n}}^{vec(G_i)}$$

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 $\Phi_h^{vec(G_i)}$ to be singular.

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

$$\left(\begin{array}{c}
\bar{\Theta}_h \\
vec(D_h)
\end{array}\right) \sim N \left(0^{(p+1)k}, \begin{pmatrix} h_{5,g} & 0^{k \times pk} \\ 0^{pk \times k} & \Phi_h^{vec(G_i)} \end{pmatrix} \right),$$

(16.21)
where \( \bar{g}_h \in R^k \) and \( \bar{D}_h \in R^{k \times p} \) are independent by the definition of \( \bar{D}_n \), see Lemma \[16.4\] below. 

To determine the asymptotic distributions of the QLR\(_n\) and QLR\(_{P_n}\) statistics (defined in \((5.7)\) and just below \((15.6)\)) and the conditional critical value of the CQLR and CQLR\(_P\) tests (defined in \((5.8)\), \((5.9)\), and \((15.7)\)), we need to determine the asymptotic distribution of \( W_{F_n} \bar{D}_n U_{F_n} \) without recentering by \( E_{F_n} G_i \). To do so, we post-multiply \( W_{F_n} \bar{D}_n U_{F_n} \) 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 \( W_{F_n} \bar{D}_n U_{F_n} B_{F_n} \) to ensure that \( n^{1/2} W_{F_n} \bar{D}_n U_{F_n} B_{F_n} S_n \) 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 \( q = q_h (\in \{0, ..., \min\{k, p\} \}) \) be such that

\[
h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq \min\{k, p\},
\]

where \( h_{1,j} := \lim n^{1/2} \tau_{jF_n} \geq 0 \) for \( j = 1, ..., \min\{k, p\} \) 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 maximum number of parameters, or one-to-one transformations of the parameters, that are strongly or semi-strongly identified. (That is, one could partition \( \theta \), or a one-to-one transformation of \( \theta \), 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

\[
S_n := \text{Diag}\{(n^{1/2} \tau_{1F_n})^{-1}, ..., (n^{1/2} \tau_{qF_n})^{-1}, 1, ..., 1\} \in R^{p \times p} \text{ and } T_n := B_{F_n} S_n \in R^{p \times p},
\]

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 \( j \leq q \).

The asymptotic distribution of \( \bar{D}_n \) after suitable rotations and rescaling, but without recentering (by subtracting \( E_{F_i} G_i \)), depends on the following quantities. We partition \( h_2 \) and \( h_3 \) and define \( \bar{\Delta}_h \)

\[\text{if one eliminates the } \lambda_{\min}(E_{F_i} g_i') \geq \delta \text{ condition in } F \text{ and one defines } \bar{D}_n \text{ in } (5.2) \text{ with } \bar{\Omega}_n \text{ replaced by the eigenvalue-adjusted matrix } \bar{\Omega}_h' \text{ for some } \varepsilon > 0, \text{ then the asymptotic distribution in } (16.21) \text{ still holds, but without } \text{the independence of } \bar{g}_h \text{ and } \bar{D}_h. \text{ However, this independence is key. Without it, the conditioning argument that is used to establish the correct asymptotic size of the CQLR and CQLR}_2 \text{ tests does not go through. Thus, we define } \bar{D}_n \text{ in } (5.2) \text{ using } \bar{\Omega}_n, \text{ not } \bar{\Omega}_h'.\]

\[29\]

Electronic copy available at: https://ssrn.com/abstract=3366443
as follows:

\[
\begin{align*}
    h_2 &= (h_{2,q}, h_{2,p-q}), \quad h_3 = (h_{3,q}, h_{3,k-q}), \\
    h_{1,p-q}^\circ &= \begin{bmatrix} 0_{q \times (p-q)} \\
                \text{Diag}\{h_{1,q+1}, \ldots, h_{1,p}\} \\
                0_{(k-p) \times (p-q)} \end{bmatrix} \in R^{k \times (p-q)} \text{ if } k \geq p, \\
    h_{1,p-q}^\circ &= \begin{bmatrix} 0_{q \times (k-q)} & 0_{q \times (p-k)} \\
                \text{Diag}\{h_{1,q+1}, \ldots, h_{1,k}\} & 0_{(k-q) \times (p-k)} \end{bmatrix} \in R^{k \times (p-q)} \text{ if } k < p, \\
    \overline{\Delta}_h &= (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k \times p}, \quad \overline{\Delta}_{h,q} := h_{3,q}, \quad \overline{\Delta}_{h,p-q} := h_{3} h_{1,p-q}^\circ + h_{71} \overline{D}_h h_{81} h_{2,p-q}, \\
    h_{71} &= W_1(h_7), \quad \text{and } h_{81} := U_1(h_8),
\end{align*}
\]

where \( h_{2,q} \in R^{q \times q}, \ h_{2,p-q} \in R^{q \times (p-q)}, \ h_{3,q} \in R^{k \times q}, \ h_{3,k-q} \in R^{k \times (k-q)}, \ \overline{\Delta}_{h,q} \in R^{k \times q}, \overline{\Delta}_{h,p-q} \in R^{k \times (p-q)}, \ h_{71} \in R^{k \times k}, \) and \( h_{81} \in R^{q \times p}.\) Note that when Assumption WU holds \( h_{71} = \lim W_{F_n} = \lim W_{1(W_2 F_n)} \) and \( h_{81} = \lim U_{F_n} = \lim U_1(U_2 F_n) \) under \( \{\lambda_{n,h} : n \geq 1\}.\)

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

**Lemma 16.4** Suppose Assumption WU holds for some non-empty parameter space \( \Lambda_* \subset \Lambda_W.\)

Under all sequences \( \{\lambda_{n,h} : n \geq 1\} \) with \( \lambda_{n,h} \in \Lambda_*,\)

\[
n^{1/2}(\overline{\gamma}_h, \overline{D}_h - E_{F_n} G_i, W_{F_n} \overline{D}_h U_{F_n} T_n) \rightarrow_d (\overline{\gamma}_h, \overline{D}_h, \overline{\Delta}_h),
\]

where (a) \((\overline{\gamma}_h, \overline{D}_h)\) 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 \( \overline{\gamma}_h \) 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 the results of parts (a)-(c) hold with \( n \) replaced with \( w_n.\)

**Comments:** (i) Lemma 16.4(c) is a key property that leads to the correct asymptotic size of the CQLR and CQLRF 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 \( n^{1/2} W_{F_n} \overline{D}_h U_{F_n} B_{F_n} S_n, \) viz., \( \overline{\Delta}_h,\) to have full column rank with probability one, in order to apply the continuous mapping theorem (CMT), which

\[47\] There is some abuse of notation here. E.g., \( h_{2,q} \) and \( h_{2,p-q} \) denote different matrices even if \( p - q \) happens to equal \( q.\)
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_{P,n} statistics considered here do not depend on Kleibergen’s LM statistic and do not require the asymptotic distribution of \( n^{1/2} W_{F_{n}} D_{n} U_{F_{n}} B_{F_{n}} S_{n} \) 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
\[
\hat{\kappa}_{jn} \text{ denote the } j \text{th eigenvalue of } n \hat{U}_{n} \hat{D}_{n} \hat{W}_{n} \hat{D}_{n} \hat{U}_{n}, \quad \forall j = 1, \ldots, p, \tag{16.25}
\]
ordered to be nonincreasing in \( j \). The \( j \)th singular value of \( n^{1/2} \hat{W}_{n} \hat{D}_{n} \hat{U}_{n} \) equals \( \hat{\kappa}_{jn}^{1/2} \) 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 non-empty parameter space \( \Lambda_{\ast} \subset \Lambda_{WU} \). Under all sequences \( \{\lambda_{n,h} : n \geq 1\} \) with \( \lambda_{n,h} \in \Lambda_{\ast} \),

(a) \( \hat{\kappa}_{jn} \to_{p} \infty \) for all \( j \leq q \),

(b) the (ordered) vector of the smallest \( p-q \) eigenvalues of \( n \hat{U}_{n} \hat{D}_{n} \hat{W}_{n} \hat{D}_{n} \hat{U}_{n} \), i.e., \( \hat{\kappa}_{(q+1)n}, \ldots, \hat{\kappa}_{pn} \)’, converges in distribution to the (ordered) \( p-q \) vector of the eigenvalues of \( \tilde{\Delta}_{h,p-q} h_{3,k-q} \lambda h_{3,k-q} \) \( \times \tilde{\Delta}_{h,p-q} \in \mathbb{R}^{(p-q) \times (p-q)} \),

(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 \( \lambda_{w_{n},h} \in \Lambda_{\ast} \), the results in parts (a)-(c) hold with \( n \) replaced with \( w_{n} \).

**Comment:** Proposition 16.5(a) and (b) with \( \hat{W}_{n} = \hat{\Omega}_{n}^{-1/2} \) and \( \hat{U}_{n} = \hat{L}_{n}^{1/2} \) is used to determine the asymptotic behavior of the critical value function for the CQLR test, which depends on \( n^{1/2} \hat{D}_{n} \) defined in (15.7), see the proof of Theorem 27.1 in Section 27.2. Proposition 16.5(a) and (b) with \( \hat{W}_{n} = \hat{\Omega}_{n}^{-1/2} \) and \( \hat{U}_{n} = \hat{L}_{n}^{1/2} \) is used to determine the asymptotic behavior of the critical value function for the CQLR_{P} test, which depends on \( n^{1/2} \hat{D}_{n} \) 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 $QLR_{W,U,n}$ statistic defined in (16.3) and, as special cases, those of the $QLR_n$ and $QLR_{P,n}$ statistics.

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

$$QLR_{W,U,n} \rightarrow_d \mathcal{G}_h n^{-1/2} - \lambda_{\min}(\langle \Delta_{h,p-q}, n_{5,g}^{-1/2} \mathcal{G}_h \rangle \langle h^3, k-q \rangle h^3, k-q \langle \Delta_{h,p-q}, n_{5,g}^{-1/2} \mathcal{G}_h \rangle)$$

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 $q \leq \min\{k, p\}$), $\Delta_{h,p-q}$ does not appear in the limit random variable and the limit random variable reduces to $(h_{5,g}^{-1/2} \mathcal{G}_h)h_{3,p}h_{5,g}^{-1/2} \mathcal{G}_h \sim \chi^2_p$. 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 $\mathcal{G}_h n_{5,g}^{-1/2} \mathcal{G}_h \sim \chi^2_k$. When $k \leq p$ and $q < k$, the $\lambda_{\min}(\cdot)$ expression equals zero and the limit random variable reduces to $\mathcal{G}_h n_{5,g}^{-1/2} \mathcal{G}_h \sim \chi^2_k$. 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 $w_n$.

**Comments:** (i) Theorem 16.6 gives the asymptotic distributions of the $QLR_n$ and $QLR_{P,n}$ statistics (defined by (5.7) and (15.6), respectively) once it is verified that the choices of $(\hat{W}_n, \hat{U}_n)$ for these statistics satisfy Assumption WU for the parameter spaces $\Lambda_{WU}$ and $\Lambda_{W,U,P}$, respectively. The latter is done in Lemma 27.4 in Section 27.1.

(ii) When $q = p$, the parameter $\theta_0$ is strongly or semi-strongly identified and Theorem 16.6 shows that the $QLR_{W,U,n}$ statistic has a $\chi^2_p$ asymptotic null distribution.

(iii) When $k = p$, Theorem 16.6 shows that the $QLR_{W,U,n}$ statistic has a $\chi^2_k$ asymptotic null distribution regardless of the strength of identification.

(iv) When $k < p$, $\theta$ is necessarily unidentified and Theorem 16.6 shows that the asymptotic null distribution of $QLR_{W,U,n}$ is $\chi^2_k$.

(v) The proof of Theorem 16.6 given in Section 26 also shows that the largest $q$ eigenvalues of $n(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{G}_n)'(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} \hat{G}_n)$ 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 $(\Delta_{h,p-q}, n_{5,g}^{-1/2} \mathcal{G}_h)h_{3,k-q}h^3, k-q \langle \Delta_{h,p-q}, n_{5,g}^{-1/2} \mathcal{G}_h \rangle$.

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 (5.8), for sequences of nonrandom
k \times p \text{ matrices } \{D_n : n \geq 1\} \text{ 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_{WU,n} \text{ statistic, conditional on the asymptotic version of the conditioning statistic, is continuous and strictly increasing at its } 1 - \alpha \text{ quantile for all possible } (k, p, q) \text{ values and all possible limits of the scaled population singular values } \{n^{1/2} \tau_{jF_n} : n \geq 1\} \text{ for } j = 1, ..., \min\{k, p\}, \text{ and (iii) establishing that Assumption WU holds for the CQLR and CQLR}_P \text{ 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 (4.7), (5.13), and (15.9), depend on the random variable \( \widehat{r}_n(\theta) \) and random matrices \( \widehat{A}_n(\theta) \) and \( \widehat{A}^\perp_n(\theta) \), defined in (4.3) and (4.4). First, in the following lemma, we show that with probability that goes to one as } n \to \infty \text{ (wp→1), the SR test statistics and data-dependent critical values are the same as when the non-random and rescaled population quantities } r_F(\theta) \text{ and } \Pi_{1F}^{1/2}(\theta)A_F(\theta)' \text{ are used to define these statistics, rather than } \widehat{r}_n(\theta) \text{ and } \widehat{A}_n(\theta)' \text{, where } r_F(\theta), A_F(\theta), \text{ and } \Pi_{1F}(\theta) \text{ are defined as in (3.4) and (3.5). The lemma also shows that the extra rejection condition in (4.7), (5.13), and (15.9) fails to hold wp→1 under all sequences of null distributions.}

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

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

**Comments.** 1. We now provide an example that appears to be a counter-example to the claim that \( \widehat{r}_n = r \) wp→1. We show that it is not a counter-example because the distributions considered violate the moment bound in \( \mathcal{F}_{\Theta_{AR}}^{SR} \) in (3.6). Suppose } k = 1 \text{ and } g_i = 1, -1, \text{ and 0 with probabilities } p_n/2, p_n/2, \text{ and } 1 - p_n, \text{ respectively, under } F_n, \text{ where } p_n = c/n \text{ for some } 0 < c < \infty. \text{ Then, } E_{F_n}g_i = 0, \text{ as is required, and } rk(\Omega_{F_n}) = rk(E_{F_n}g_i^2) = rk(p_n) = 1. \text{ We have } \Omega_n = 0 \text{ if } g_i = 0 \forall i \leq n. \text{ The latter holds with probability } (1 - p_n)^n = (1 - c/n)^n \to e^{-c} > 0 \text{ as } n \to \infty. \text{ In consequence, } P_{F_n}(rk(\Omega_n) = rk(\Omega_{F_n})) = P_{F_n}(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→1. However, the distributions \( \{F_n : n \geq 1\} \) in this example violate the moment bound \( E_F[\|\Pi_{1F_n}^{1/2}A_F'g_i\|^2 + \gamma] \leq M \) in \( \mathcal{F}^{SR}_{\mathcal{A}R} \), so there is no inconsistency with the claim. This holds because for these distributions \( E_{\mathcal{F}_n}[\|\Pi_{1F_n}^{1/2}A_F'g_i\|^2 + \gamma] = E_{\mathcal{F}_n}[VA_{F_n}^{-1/2}(g_i)g_i]^2 + \gamma = p_n^{-2(\gamma+1)/2}E_{\mathcal{F}_n}[g_i] = p_n^{-\gamma/2} \to \infty \) as \( n \to \infty \), where the second equality uses \( |g_i| \) equals 0 or 1 and the third equality uses \( E_{\mathcal{F}_n}[g_i] = p_n \).

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 \( \theta \in R^k \), \( \Omega_F \in R^{k \times k} \), \( g(W_i, \theta) := W_i - \theta \), \( \Omega_F \) has spectral decomposition \( A_F \Pi_F A_F' \), and some eigenvalues of \( \Omega_F \) 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 \( \mathcal{F}^{SR}_{\mathcal{A}R} \) and \( \mathcal{F}^{SR} \) hold immediately. In this case, the conditions in \( \mathcal{F}^{SR}_{\mathcal{A}R} \) and \( \mathcal{F}^{SR} \) 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 \( \theta_0n \). By considering subsequences, it suffices to consider the case where \( r_{F_n} = r \) for all \( n \geq 1 \) for some \( r \in \{0, 1, \ldots, k\} \).

First, we establish part (a). We have \( \widehat{r}_n \leq r \) a.s. for all \( n \geq 1 \) 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 \( \lambda'\widehat{\Omega}_{F_n}\lambda = n^{-1}\sum_{i=1}^{n}(\lambda'g_i)^2 - (\lambda'\overline{g})^2 = 0 \) 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 \( \widehat{r}_n := rk(\widehat{\Omega}_n) \geq rk(\Pi_{1F_n}^{-1/2}A_F'\widehat{\Omega}_nA_F\Pi_{1F_n}^{-1/2}) \) because \( \widehat{\Omega}_n \) is \( k \times k \), \( A_F\Pi_{1F_n}^{-1/2} \) is \( k \times r \), and \( 1 \leq r \leq k \). In addition, we have

\[
\Pi_{1F_n}^{-1/2}A_F'\widehat{\Omega}_nA_F\Pi_{1F_n}^{-1/2} = n^{-1}\sum_{i=1}^{n}(\Pi_{1F_n}^{-1/2}A_F'g_i)(\Pi_{1F_n}^{-1/2}A_F'g_i)' - (n-1)\sum_{i=1}^{n}\Pi_{1F_n}^{-1/2}A_F'g_i)(n-1)\sum_{i=1}^{n}\Pi_{1F_n}^{-1/2}A_F'g_i)',
\]

\[
E_{\mathcal{F}_n}(\Pi_{1F_n}^{-1/2}A_F'g_i)(\Pi_{1F_n}^{-1/2}A_F'g_i)' = \Pi_{1F_n}^{-1/2}A_F'\Omega_{F_n}A_F\Pi_{1F_n}^{-1/2} = \Pi_{1F_n}^{-1/2}A_F'\Omega_{F_n}A_F\Pi_{1F_n}^{-1/2} = I_r,
\]

and \( E_{\mathcal{F}_n}\Pi_{1F_n}^{-1/2}A_F'g_i = 0^r \), where the second last equality in (17.1) holds by the spectral decomposition in (3.4) and the last equality in (17.1) holds by the definitions of \( A_F^\Omega \), \( A_F \), and \( \Pi_{1F} \) in (3.4) and (3.5). By (17.1), the moment conditions in \( \mathcal{F}^{SR} \), and the weak law of large numbers for \( L^{1+\gamma/2} \)-bounded i.i.d. random variables for \( \gamma > 0 \), we obtain \( \Pi_{1F_n}^{-1/2}A_F'\widehat{\Omega}_nA_F\Pi_{1F_n}^{-1/2} \to_p I_r \). In consequence, \( rk(\Pi_{1F_n}^{-1/2}A_F'\widehat{\Omega}_nA_F\Pi_{1F_n}^{-1/2}) \geq r \) wp→1, which concludes the proof that \( \widehat{r}_n \to r \).
Next, we prove part (b). Let \( N(\cdot) \) denote the null space of a matrix. We have

\[
\lambda \in N(\Omega_{F_n}) \implies \lambda'\Omega_{F_n}\lambda = 0 \implies Var_{F_n}(\lambda'g_i) = 0 \implies \lambda'g_i = 0 \text{ a.s.}[F_n]
\]

\[
\implies \tilde{\Omega}_n\lambda = 0 \text{ a.s.}[F_n] \implies \lambda \in N(\tilde{\Omega}_n) \text{ a.s.}[F_n].
\]

That is, \( N(\Omega_{F_n}) \subset N(\tilde{\Omega}_n) \text{ a.s.}[F_n] \). This and \( rk(\Omega_{F_n}) = rk(\tilde{\Omega}_n) \text{ wp} \to 1 \) imply that \( N(\Omega_{F_n}) = N(\tilde{\Omega}_n) \text{ wp} \to 1 \) (because if \( N(\tilde{\Omega}_n) \) is strictly larger than \( N(\Omega_{F_n}) \) then the dimension and rank of \( \tilde{\Omega}_n \) must exceed the dimension and rank of \( N(\Omega_{F_n}) \), which is a contradiction). In turn, \( N(\Omega_{F_n}) = N(\tilde{\Omega}_n) \text{ wp} \to 1 \) implies that \( col(\tilde{A}_n) = col(A_{F_n}) \text{ wp} \to 1 \), 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 \( \tilde{r}_n = 0 \) and \( \tilde{r}_n = 0 \text{ wp} \to 1 \) when \( r = 0 \) by part (a). Part (b) of the Lemma implies that there exists a random \( r \times r \) nonsingular matrix \( \tilde{M}_n \) such that

\[
\tilde{A}_n = A_{F_n}\Pi_{1F_n}^{-1/2}\tilde{M}_n \text{ wp} \to 1,
\]

because \( \Pi_{1F_n} \) is nonsingular (since it is a diagonal matrix with the positive eigenvalues of \( \Omega_{F_n} \) on its diagonal by its definition following (3.5)). Equation (17.3) and \( \tilde{r}_n = r \text{ wp} \to 1 \) imply that the statistics \( SR-AR_m, SR-QLR_m, SR-QLR_{F_n}, c_{\tilde{r},p}(n^{1/2}\bar{D}_{An}^{*},1-\alpha), \text{ and } c_{\tilde{r},p}(n^{1/2}\bar{D}_{An}^{*},1-\alpha) \) are invariant \( \text{wp} \to 1 \) to the replacement of \( \tilde{r}_n \) and \( \tilde{A}_n' \) by \( r \) and \( \tilde{M}_n'\Pi_{1F_n}^{-1/2}A_{F_n}', \) 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 \( \tilde{M}_n' \). 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 \( \tilde{M}_n' \). Thus, these five statistics, defined as in Sections 5.2 and 15, are invariant \( \text{wp} \to 1 \) to the replacement of \( \tilde{r}_n \) and \( \tilde{A}_n' \) by \( r \) and \( \Pi_{1F_n}^{-1/2}A_{F_n}' \), respectively.

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

\[
n Var_{F_n}((A_{F_n}^r)\tilde{g}_n) = (A_{F_n}^r)'\Omega_{F_n}A_{F_n}^r = (A_{F_n}^r)'A_{F_n}^\Omega F_n (A_{F_n}^\Omega)'A_{F_n}^r = 0^{(k-r)\times(k-r)}, \tag{17.4}
\]

where the second equality uses the spectral decomposition in (3.4) and the last equality uses \( A_{n}^\Omega = [A_F, A_F^r], \) see (3.5). In consequence, \((A_{F_n}^r)'\tilde{g}_n = 0^{k-r}\) a.s. This and and the result of part (b) that \(\text{col}(\hat{A}_n^r) = \text{col}(A_{F_n}^r)\) wp\(\rightarrow\)1 establish part (d). \(\blacksquare\)

Given Lemma 17.1(d), the extra rejection conditions in the SR-AR, SR-CQLR, and SR-CQLR\(_P\) tests and CS’s (i.e., the second conditions in (4.7), (4.9), (5.13), (15.9), and in the SR-CQLR and SR-CQLR\(_P\) CS definitions following (5.13) and (15.9)) can be ignored when computing the asymptotic size properties of these tests and CS’s (because the condition fails to hold for each test wp\(\rightarrow\)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\(\rightarrow\)1 under any sequence of true values \(\theta_{0n}\) for any sequence of distributions for which the moment conditions hold at \(\theta_{0n}\)).

Given Lemma 17.1(c), the asymptotic size properties of the SR-AR, SR-CQLR, and SR-CQLR\(_P\) tests and CS’s can be determined by the analogous tests and CS’s that are based on \(r_{F_n}(\theta_0)\) and \(\Pi_{1F_n}^{1/2}(\theta_0)A_{F_n}(\theta_0)'\) (for fixed \(\theta_0\) with tests and for any \(\theta_0 \in \Theta\) with CS’s). For the tests, we do so by partitioning \(\mathcal{F}_{SR\ AR}^F, \mathcal{F}_{SR}^F, \) and \(\mathcal{F}_{SR\ P}^F\) into \(k\) sets based on the value of \(rk(\Omega_F(\theta_0))\) and establishing the correct asymptotic size and asymptotic similarity of the analogous tests separately for each parameter space. That is, we write \(\mathcal{F}_{SR\ AR}^F = \bigcup_{r=0}^k \mathcal{F}_{SR\ AR}^F[r]\), where \(\mathcal{F}_{SR\ AR}^F[r] := \{F \in \mathcal{F}_{SR\ AR}^F : rk(\Omega_F(\theta_0)) = r\}\), and establish the desired results for \(\mathcal{F}_{SR\ AR}^F[r]\) separately for each \(r\). Analogously, we write \(\mathcal{F}_{SR}^F = \bigcup_{r=0}^k \mathcal{F}_{SR}^F[r]\) and \(\mathcal{F}_{SR\ P}^F = \bigcup_{r=0}^k \mathcal{F}_{SR\ P}^F[r]\), where \(\mathcal{F}_{SR}^F[r] := \mathcal{F}_{SR\ AR}^F[r]\cap \mathcal{F}_{SR}^F\) and \(\mathcal{F}_{SR\ P}^F[r] := \mathcal{F}_{SR\ AR}^F[r]\cap \mathcal{F}_{SR\ P}^F\). Note that we do not need to consider the parameter space \(\mathcal{F}_{SR\ AR}^F[r]\) 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\(\rightarrow\)1 based on the first condition in (4.7) when \(r = 0\) (since the test statistic and critical value equal zero by definition when \(\tilde{r}_n = 0\) and \(\tilde{r}_n = r = 0\) wp\(\rightarrow\)1 by Lemma 17.1(a)). In addition, we do not need to consider the parameter space \(\mathcal{F}_{SR\ AR}^F[r]\) 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 \(\mathcal{F}_{SR\ AR}^F\) by the statement of Theorem 6.1. Analogous arguments regarding the parameter spaces corresponding to \(r = 0\) apply to the other tests and CS’s. Hence, from here on, we assume \(r \in \{1, \ldots, k\}\).

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

\[
g_{F_i}^* := \Pi_{1F}^{1/2} A_F' g_i \text{ and } G_{F_i}^* := \Pi_{1F}^{-1/2} A_F' G_i, \tag{17.5}
\]
where $A_F \in \mathbb{R}^{k \times r}$, $\Pi_{1F} \in \mathbb{R}^{r \times r}$, and dependence on $\theta_0$ is suppressed for notational simplicity. Given the conditions in $\mathcal{F}^{SR}$, we have

$$E_F ||g_{F_i}^*||^{2+\gamma} = E_F ||\Pi_{1F}^{-1/2}A_F^t g_i||^{2+\gamma} \leq M,$$

$$E_F ||\text{vec}(G_{F_i}^*)||^{2+\gamma} = E_F ||\text{vec}(\Pi_{1F}^{-1/2}A_F^t G_i)||^{2+\gamma} \leq M,$$

$$\lambda_{\min}(E_F g_{F_i}^* g_{F_i}^{*t}) = \lambda_{\min}(\Pi_{1F}^{-1/2}A_F^t \Omega_F A_F \Pi_{1F}^{-1/2}) = \lambda_{\min}(I_r) = 1,$$ \hspace{1cm} (17.6)

and $E_F g_{F_i}^* = 0^r$, where the second equality in the third line of (17.6) holds by the spectral decomposition in (3.4) and the partition $A_F^\Omega = [A_F, A_F^t]$ in (3.5). Thus, $F \in \mathcal{F}^{SR}$ implies that $F \in \mathcal{F}$ with $\delta \leq 1$, when $\mathcal{F}$ is defined with $(g_{F_i}^*, G_{F_i}^*)$ 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 $(g_{F_i}^*, G_{F_i}^*)$ 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 $\mathcal{F}^{SR}$ for $r = 1, \ldots, k$. This requires that Theorem 16.1 holds for $k < p$, which it does. The fact that $g_{F_i}^*$ and $G_{F_i}^*$ 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 $\theta_0$ taking any value in $\Theta$ and with $\mathcal{F}_{\Theta}^{SR}$ and $\mathcal{F}_{\Theta}$, defined in (3.7) and just below (16.1), in place of $\mathcal{F}^{SR}$ 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 $\text{vec}(G_{F_i}^*)$ 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), i.e., $g_i = u_i Z_i$, we define $Z_{F_i}^* := \Pi_{1F}^{-1/2}A_F^t Z_i$, $g_{F_i}^* = u_i Z_{F_i}^*$, and $G_{F_i}^* = Z_{F_i}^* u_{\theta_i}^t$, where $u_{\theta_i}$ is defined in (15.2) and the dependence of various quantities on $\theta_0$ is suppressed. In this case, by the conditions in $\mathcal{F}^{SR}_P$, the IV’s $Z_{F_i}^*$ satisfy $E_F ||Z_{F_i}^*||^{4+\gamma} = E_F ||\Pi_{1F}^{-1/2}A_F^t Z_i||^{4+\gamma} \leq M$ and $E_F ||u_i^*||^{2+\gamma} \leq M$, where $u_i^* := (u_i, u_{\theta_i}^t)^t$. Next we show that $\lambda_{\min}(E_F Z_{F_i}^* Z_{F_i}^{*t})$ is bounded away from zero for $F \in \mathcal{F}^{SR}_P[r]$. We
have

\[
\lambda_{\min}(E_FZ'_iZ''_i) = \lambda_{\min}(E_F\Pi_{1F}'A_F'Z'_iA_F\Pi_{1F}')
\]
\[
= \inf_{\lambda \in R^d:\|\lambda\|=1} \left[ E_F(\lambda\Pi_{1F}'A_F'Z'_i)^21(u_i^2 > c) + E_F(\lambda\Pi_{1F}'A_F'Z'_i)^21(u_i^2 < c) \right]
\]
\[
\geq \inf_{\lambda \in R^d:\|\lambda\|=1} \left[ c^{-1}E_F(\lambda\Pi_{1F}'A_F'Z'_i)^2u_i^21(u_i^2 > c) \right]
\]
\[
= c^{-1} \inf_{\lambda \in R^d:\|\lambda\|=1} \left[ E_F(\lambda\Pi_{1F}'A_F'Z'_i)^2u_i^2 - E_F(\lambda\Pi_{1F}'A_F'Z'_i)^2u_i^21(u_i^2 > c) \right]
\]
\[
\geq c^{-1} \lambda_{\min}(\Pi_{1F}'A_F'\Omega_AF_i\Pi_{1F}'1/2) - \sup_{\lambda \in R^d:\|\lambda\|=1} E_F(\lambda\Pi_{1F}'A_F'Z'_i)^2u_i^21(u_i^2 > c)
\]
\[
\geq c^{-1}[1 - E_F]\|\Pi_{1F}'A_F'Z'_i\|u_i^21(u_i^2 > c)]
\]
\[
\geq 1/(2c), \tag{17.7}
\]

where the second inequality uses \(g_i = Z_iu_i\) and \(\Omega_F := E_Fg_ig'_i\), the third inequality holds by \(\Pi_{1F}'A_F'\Omega_AF_i\Pi_{1F}'1/2 = I_r\) (using \(3.4\) and \(3.5\)) and by the Cauchy-Bunyakovsky-Schwarz inequality applied to \(\lambda\Pi_{1F}'A_F'Z'_i\), and the last inequality holds by the condition \(E_F\|\Pi_{1F}'A_F'Z'_i\|^2u_i^2 \times 1(u_i^2 > c) \leq 1/2\) in \(F_P^{\text{SR}}\).

The moment bounds above and \(\tag{17.7}\) establish that \(F \in \mathcal{F}_{P[r]}^{\text{SR}}\) implies that \(F \in \mathcal{F}_P\) for \(\delta \leq \min\{1, 1/(2c)\}\), when \(\mathcal{F}_P\) is defined with \((g^*_F, G^*_F)\) in place of \((g_i, G_i)\), where the definition of \(\mathcal{F}_P\) in \(\tag{16.1}\) is taken to allow \(g_i\) and \(G_i\) to depend on \(F\).\footnote{We require \(\delta \leq \min\{1, 1/(2c)\}\), rather than \(\delta \leq 1/(2c)\), because \(\lambda_{\min}(E_Fg^*_Fg'_F) = 1\) by \(\tag{17.6}\) and \(\mathcal{F} \subset \mathcal{F}_{AR}\) requires \(\lambda_{\min}(E_Fg^*_Fg'_F) \geq \delta\).}

Now we apply Theorem \(\tag{16.1}\) with \((g^*_F, G^*_F)\) and \(r\) in place of \((g_i, G_i)\) and \(k\) and \(\delta \leq \min\{1, 1/(2c)\}\) to obtain the correct asymptotic size and asymptotic similarity of the CQLR\(_P\) test based on \((g^*_F, G^*_F)\) and \(r\) for the parameter space \(\mathcal{F}_{P[r]}^{\text{SR}}\) for \(r = 1, \ldots, k\). As noted above, the dependence of \(g^*_F\) and \(G^*_F\) on \(F\) does not cause a problem in the application of Theorem \(\tag{16.1}\). This establishes the results of Theorem \(\tag{15.2}\) for the SR-CQLR\(_P\) test by the argument given above.\footnote{The fact that \(Z'_i\) depends on \(\theta_0\) through \(\Pi_{1F}'/2(\theta_0)AF(\theta_0)'\) and that \(G^*_F(\theta_0) \neq (\partial/\partial\theta')g^*_F(\theta_0)\) (because \((\partial/\partial\theta')Z'_i\) is ignored in the specification of \(G^*_F(\theta_0)\)) does not affect the application of Theorem \(\tag{16.1}\). The reason is that the proof of this Theorem goes through even if \(Z_i\) depends on \(\theta_0\) and for any \(G_i(\theta_0)\) that satisfies the conditions in \(\mathcal{F}_P\), not just for \(G_i(\theta_0) := (\partial/\partial\theta')g_i(\theta_0)\).}

The proof for the SR-CQLR\(_P\) CS is essentially the same, but with \(\theta_0\) taking any value in \(\Theta\) and with \(\mathcal{F}_{\Theta,P}^{\text{SR}}\) and \(\mathcal{F}_{\Theta,2}\), defined in \(\tag{3.7}\) and just below \(\tag{16.1}\), in place of \(\mathcal{F}_{P[r]}^{\text{SR}}\) and \(\mathcal{F}_P\), respectively.

This completes the proof of Theorems \(\tag{6.1}\) and \(\tag{15.2}\) given Theorem \(\tag{16.1}\).
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 $\{W_i : i = \ldots, 0, 1, \ldots\}$.\footnote{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 CS’s. Under such sequences, the observations form a triangular array of row-wise strictly stationary observations.}

We define

$$V_{F,n}(\theta) := \text{Var}_F \left( n^{-1/2} \sum_{i=1}^n \begin{pmatrix} g_i(\theta) \\ \text{vec}(G_i(\theta)) \end{pmatrix} \right),$$

$$\Omega_{F,n}(\theta) := \text{Var}_F (n^{-1/2} \sum_{i=1}^n g_i(\theta)), \quad r_{F,n}(\theta) := rk(\Omega_{F,n}(\theta)).$$

(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 (3.4), (3.5), and the paragraph following (3.3), 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

$$\mathcal{F}^\text{SR}_{TS,AR} := \{ F : \{W_i : i = \ldots, 0, 1, \ldots\} \text{ are stationary and strong mixing under } F \text{ with strong mixing numbers} \{ \alpha_F(m) : m \ge 1 \} \text{ that satisfy} \alpha_F(m) \le Cm^{-d},$$

$$E_F g_i = 0^k, \quad \text{and} \quad \sup_{n \ge 1} E_F \| \Pi_{1F,n}^{-1/2} A_{F,n} g_i \|^2 \le M \}$$

(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 $\theta_0$ is suppressed. For CS’s, we use the corresponding parameter space $\mathcal{F}^\text{SR}_{TS,\Theta,AR} := \{ (F, \theta_0) : F \in \mathcal{F}^\text{SR}_{TS,AR}(\theta_0), \theta_0 \in \Theta \}$, where $\mathcal{F}^\text{SR}_{TS,AR}(\theta_0)$ denotes $\mathcal{F}^\text{SR}_{TS,AR}$ with its dependence on $\theta_0$ made explicit. The moment conditions in $\mathcal{F}^\text{SR}_{TS,AR}$ are placed on the normalized moment functions $\Pi_{1F,n}^{-1/2} A_{F,n} g_i$ that satisfy $\text{Var}_F (n^{-1/2} \sum_{i=1}^n \Pi_{1F,n}^{-1/2} A_{F,n} g_i) = I_k$ for all $n \ge 1$.

For the SR-CQLR and SR-CQLR$P$ tests, we use the null parameter spaces $\mathcal{F}^\text{SR}_{TS}$ and $\mathcal{F}^\text{SR}_{TS,P}$, respectively, which are defined as $\mathcal{F}^\text{SR}$ and $\mathcal{F}^\text{SR}_P$ are defined in (3.6) and (15.3), but with (i) $\mathcal{F}^\text{SR}_{TS,AR}$ in place of $\mathcal{F}^\text{SR}_{AR}$, (ii) $A_F$ and $\Pi_{1F}$ replaced by $A_{F,n}$ and $\Pi_{1F,n}$, respectively, and (iii) $\sup_{n \ge 1}$ added before the quantities $\mathcal{F}^\text{SR}$ and $\mathcal{F}^\text{SR}_P$ that depend on $A_{F,n}$ and $\Pi_{1F,n}$. For SR-CQLR and SR-CQLR$P$ CS’s, we use the parameter spaces $\mathcal{F}^\text{SR}_{TS,\Theta}$ and $\mathcal{F}^\text{SR}_{TS,\Theta,P}$, respectively, which are defined as $\mathcal{F}^\text{SR}_{TS,\Theta,AR}$
is defined, but with $\mathcal{F}^{SR}\theta_0)$ 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 $\theta_0$ made explicit.

The SR-CQLR and SR-CQLR$_P$ test statistics depend on some estimators $\hat{V}_n (= \hat{V}_n(\theta_0))$ of $V_{F,n}$. The SR-AR test statistic only depends on an estimator $\hat{\Omega}_n (= \hat{\Omega}_n(\theta_0))$ 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 \{\(g_i - \hat{g}_n : i \leq n\), \(\{f_i - \hat{f}_n : i \leq n\}\) (defined in (5.3)), and \{\(u_i - \hat{u}_{in}^\sigma \otimes Z_i : i \leq n\)\} (defined in (15.5)), respectively. There are a number of HAC estimators available in the literature, e.g., see Newey and West (1987) and Andrews (1991).

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

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

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

We define $\hat{r}_n (= \hat{r}_n(\theta_0))$ and $\hat{A}_n (= \hat{A}_n(\theta_0))$ as in (4.3) and (4.4) with $\theta = \theta_0$, but with $\hat{\Omega}_n$ defined in (18.3), rather than in (4.1).

The asymptotic size and similarity properties of the tests considered here are the same for any consistent HAC estimator. Hence, for generality, we do not specify a particular estimator $\hat{V}_n$ (or $\hat{\Omega}_n$). Rather, we state results that hold for any estimator $\hat{V}_n$ (or $\hat{\Omega}_n$) that satisfies one the following assumptions when the null value $\theta_0$ 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}^{1/2} F_{F,n}(\theta_0 A_{F,n}(\theta_0)))] [\hat{V}_n(\theta_0) - V_{F,n}(\theta_0)] [-I_{p+1} \otimes (A_{F,n}(\theta_0) \Pi_{-1/2}^{1/2} F_{F,n}(\theta_0))] \rightarrow 0 \uparrow_{\uparrow_n}^{(p+1)k \times (p+1)k} \text{ under } \{F_n : n \geq 1\} \text{ for any sequence } \{F_n \in \mathcal{F}^{SR}_{TS} : n \geq 1\}
\text{ for which } V_{F,n}(\theta_0) \rightarrow V \text{ for some matrix } V \text{ and } r_{F,n}(\theta_0) = r \text{ for all } n \text{ large, for any } r \in \{1, ..., k\}.

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

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

For SR-CQLR CS’s, we use the following assumption that allows both the null parameter $\theta_0$,
as well as the distribution $F_n$, to drift with $n$.

**Assumption SR-V-CS:** $[I_{p+1} \otimes (\Pi_{1F_n,n}(\theta_{0n})A^T_{F_n,n}(\theta_{0n})][\hat{V}_n(\theta_{0n}) - V_{F_n,n}(\theta_{0n})][I_{p+1} \otimes (A^T_{F_n,n}(\theta_{0n})\Pi_{1F_n,n}(\theta_{0n}))] \rightarrow_p 0^{(p+1)k \times (p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{(F_n, \theta_{0n}) \in \mathcal{F}^{SR}_{TS,\Theta}: n \geq 1\}$ for which $V_{F_n,n}(\theta_{0n}) \rightarrow V$ for some matrix $V$ and $r_{F_n,n}(\theta_{0n}) = r$ for all $n$ large, for any $r \in \{1, \ldots, k\}$.

(b) $\hat{V}_n(\theta_{0n})$ is equivariant for all $\theta_0 \in \Theta$.

(c) $\lambda^T g_i(\theta_{0n}) = 0$ a.s.$[F]$ implies that $\lambda^T \hat{\Omega}_n(\theta_{0n}) = 0$ a.s.$[F]$ for all $\lambda \in \mathbb{R}^k$ and $(F, \theta_0) \in \mathcal{F}^{SR}_{TS,\Theta}$. Assumptions SR-V(a) and SR-V-CS(a) require the HAC estimator based on the normalized moments and Jacobian (i.e., $\Pi_{1F_n,n}(\theta_{0n})A^T_{F_n,n}(\theta_{0n})g_i(\theta_{0n})$ and $\Pi_{1F_n,n}(\theta_{0n})A^T_{F_n,n}(\theta_{0n})G_i(\theta_{0n})$, respectively) to be consistent. This can be verified using standard methods. For typical HAC estimators, equivariance and Assumptions SR-V(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-CS$_P$**, which are defined as Assumptions SR-V and SR-V-CS are defined, respectively, but with $\mathcal{F}^{SR}_{TS,P}$ and $\mathcal{F}^{SR}_{TS,\Theta,P}$ in place of $\mathcal{F}^{SR}_{TS}$ and $\mathcal{F}^{SR}_{TS,\Theta}$.

For the SR-AR test and CS, we use **Assumptions SR-$\Omega$ and SR-$\Omega$-CS**, which are defined as Assumptions SR-V and SR-V-CS are defined, respectively, but with (i) Assumption SR-$\Omega$(a) being: $\Pi_{1F_n,n}(\theta_{0n})A^T_{F_n,n}(\theta_{0n})[\hat{\Omega}_n(\theta_{0n}) - \Omega_{F_n,n}(\theta_{0n})]A_{F_n,n}(\theta_{0n})\Pi_{1F_n,n}(\theta_{0n}) \rightarrow_p 0^{k \times k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}^{SR}_{TS,AR}: n \geq 1\}$ for which $\Omega_{F_n,n}(\theta_{0n}) \rightarrow \Omega$ for some matrix $\Omega$ and $r_{F_n,n}(\theta_{0n}) = r$ for all $n$ large, for any $r \in \{1, \ldots, k\}$, (ii) Assumption SR-$\Omega$-CS(a) being as in (i), but with $\theta_{0n}$ and $\mathcal{F}^{SR}_{TS,\Theta,AR}$ in place of $\theta_0$ and $\mathcal{F}^{SR}_{TS,AR}$, (iii) $\hat{\Omega}_n(\theta_{0n})$ in place of $\hat{\Omega}_n(\theta_{0n})$ in part (b) of each assumption, and (iv) $\mathcal{F}^{SR}_{TS,\Theta,AR}$ in place of $\mathcal{F}^{SR}_{TS}$ 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 and 15 with the following changes. For all three tests, $\tilde{r}_n$ and $\hat{A}_n^\perp$ in the condition $\hat{A}_n^\perp \tilde{g}_n \neq 0^{k-\tilde{r}_n}$ in (4.7) are defined as in (4.3) and (4.4), but with $\hat{\Omega}_n$ defined to satisfy Assumption SR-$\Omega$, rather than being defined in (4.1). The SR-AR statistic is defined as in Section 4, but with $\hat{\Omega}_n$ defined to satisfy Assumption SR-$\Omega$. This affects the definitions of $\tilde{r}_n$ and $\hat{A}_n$, given in (4.3) and (4.4). With these changes, the critical value for the SR-AR test in the time series case is defined in the same way as in the i.i.d. case.

In the time series case, the SR-QLR statistic is defined as in Section 5, but with $\hat{V}_n$ and $\hat{\Omega}_n$ defined to satisfy Assumption SR-V and (18.3) based on $\{f_i - \hat{f}_n : i \leq n\}$, in place of $\tilde{V}_n$ and $\hat{\Omega}_n$ defined in (5.3) and (4.1), respectively. This affects the definitions of $\tilde{R}_n$, $\tilde{\Sigma}_n$, $\tilde{L}_n$, $\tilde{D}_n^\ast$, $\tilde{r}_n$, $\hat{A}_n$, and $SR-AR_{\theta_0}$ (which appears in (5.7)). 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 \(\hat{V}_n\) and \(\hat{\Omega}_n\) defined to satisfy Assumption SR-V\(_P\) and (18.3) based on \(\{(u_i^* - \hat{u}_i^n) \otimes Z_j : i \leq n\}\), rather than in (15.5) and (4.1), respectively. In turn, this affects the definitions of \(e_Rn\), \(e_Ln\), \(e_Dn\), \(e_Qn\), \(b_rn\), \(b_An\), 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,

\[
V_F := \lim Var_F \left( n^{-1/2} \sum_{i=1}^n \begin{pmatrix} g_i \\ \text{vec}(G_i) \end{pmatrix} \right)
= \sum_{m=-\infty}^{\infty} E_F \left( \begin{pmatrix} g_i \\ \text{vec}(G_i - E_FG_i) \end{pmatrix} \begin{pmatrix} g_{i-m} \\ \text{vec}(G_{i-m} - E_FG_{i-m}) \end{pmatrix} \right)^t
\]
\[
\Omega_F := \sum_{m=-\infty}^{\infty} E_F g_i g_{i-m}',
\]

(18.4)

where the dependence of various quantities on the null value \(\theta_0\) is suppressed for notational simplicity. The second equality holds for \(F \in \mathcal{F}_{TS,P}^{SR}\).

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 \(\mathcal{F}_{TS,AR}^{SR}\), \(\mathcal{F}_{TS}^{SR}\), and \(\mathcal{F}_{TS,P}^{SR}\), respectively, and the corresponding Assumption SR-\(\Theta\), 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\(_P\) CS’s for the parameter spaces \(\mathcal{F}_{TS,\Theta,AR}^{SR}\), \(\mathcal{F}_{TS,\Theta}^{SR}\), and \(\mathcal{F}_{TS,\Theta,P}^{SR}\), respectively, provided the corresponding Assumption SR-\(\Theta\)-CS, SR-V-CS, or SR-V\(_P\)-CS holds for each CS, rather than Assumption SR-\(\Theta\), 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) IV’s when \(p = 1\), where \(p\) is the number of endogenous rhs variables, which equals the dimension of \(\theta\). The reason is that the latter test has

---

52 This is shown in the proof of Lemma 20.1 in Section 20 in the SM to AG1.
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).\footnote{Whether this also holds for $p \geq 2$ is an open question.}

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) LR statistic and
its conditioning statistic, respectively, in the homoskedastic linear IV model with $k \geq p$ fixed
(i.e., nonrandom) IV’s and nonsingular moments variance matrix (whether or not the errors are
Gaussian). This holds for all values of $p \geq 1$.

We also show that the same is true for the SR-QLR 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 IV’s being fixed, whereas the SR-QLR
statistic and its conditioning matrix are designed (essentially) for random IV’s.

We also evaluate the behavior of Kleibergen’s (2005, 2007) nonlinear CLR tests in the ho-
moskedastic linear IV model with fixed IV’s. Kleibergen’s tests depend on the choice of a weight
matrix for the conditioning statistic (which enters both the CLR test statistic and the critical value
function). We find that when $p = 1$ Kleibergen’s CLR test statistic and conditioning statistic re-
duce 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 arbi-
trarily close to zero or infinity (depending on the value of the reduced-form error matrix $\Sigma_V$ and
null hypothesis value $\theta_0$), see Lemma\footnote{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\footnote{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) IV’s. The linear IV regression model is

$$y_{1i} = Y_{2i}' \theta + u_i$$

and

$$Y_{2i} = \pi' Z_i + V_{2i},$$

(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 IV’s, and $\pi \in R^{k \times p}$ is an unknown unrestricted parameter matrix. In terms of its reduced-form equations, the model is

$$y_{1i} = Z_i' \pi \theta + V_{1i}, \quad Y_{2i} = \pi' Z_i + V_{2i}, \quad V_i := (V_{1i}, V_{2i})', \quad V_{1i} = u_i + V_{2i}' \theta, \quad \text{and} \quad \Sigma_V := EV_i V_i'.$$

(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 $\Sigma_V$.

The IV moment functions and their derivatives with respect to $\theta$ are

$$g(W_i, \theta) = Z_i(y_{1i} - Y_{2i}' \theta) \quad \text{and} \quad G(W_i, \theta) = -Z_i Y_{2i}' \quad \text{where} \quad W_i := (y_{1i}, Y_{2i}, Z_i)' \quad \text{and} \quad \Sigma_V := EV_i V_i'.$$

(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 $\Sigma_V$ is known is

$$LR_{HGL,n} := \overline{S}'_n \overline{S}_n - \lambda_{\min}(\overline{S}'_n, T_n) (\overline{S}'_n, T_n),$$

where

$$\overline{S}_n := (Z_n' \times Z_n \times Yb_0(b_0' \Sigma_V b_0)^{-1/2} = (n^{-1} Z_n' \times Z_n \times Yb_0 b_0' \Sigma_V b_0)^{-1/2} \in R^k,$$

$$T_n := (Z_n' \times Z_n \times Y \Sigma_V^{-1} A_0 (A_0' \Sigma_V^{-1} A_0)^{-1/2}$$

$$= -(n^{-1} Z_n' \times Z_n \times Y \Sigma_V^{-1} A_0 (A_0' \Sigma_V^{-1} A_0)^{-1/2} \in R^{k \times p},$$

$$Z_n \times \times \times Y := (Z_1, \ldots, Z_n)' \in R^{n \times k}, \quad Y := (Y_1, \ldots, Y_n)' \in R^{n \times (p+1)}, \quad Y_i := (y_{1i}, Y_{2i})' \in R^{p+1},$$

$$b_0 := (1, -\theta_0)' \in R^{p+1}, \quad \widehat{\Sigma}_n := n^{-1} \sum_{i=1}^n g(W_i, \theta_0), \quad A_0 := (\theta_0, I_p)' \in R^{(p+1) \times p},$$

$$\widehat{G}_n := n^{-1} \sum_{i=1}^n G(W_i, \theta_0),$$

(19.4)
\( \lambda_{\text{min}}(\cdot) \) denotes the smallest eigenvalue of a matrix, and the second equality for \( T_n \) holds by (29.12) in the SM.\(^{54}\) Note that \((S_n, T_n)\) is a (conveniently transformed) sufficient statistic for \((\theta, \pi)\) under normality of \( V_i \), known variance matrix \( \Sigma_V \), and fixed IV’s.

Moreira’s (2003) CLR test uses the \( LR_{HGL,n} \) statistic and a conditional critical value that depends on the \( k \times p \) matrix \( T_n \) through the conditional critical value function \( c_{k,p}(\cdot, 1 - \alpha) \) defined in (5.8). For \( \alpha \in (0, 1) \), Moreira’s CLR test with nominal level \( \alpha \) rejects \( H_0 \) if

\[
LR_{HGL,n} > c_{k,p}(T_n, 1 - \alpha).
\]

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

Moreira’s (2003) CLR test is similar with finite-sample size \( \alpha \) in the HGL model with known \( \Sigma_V \). Intuitively, the strength of the IV’s affects the null distribution of the test statistic \( LR_{HGL,n} \) and the critical value \( c_{k,p}(T_n, 1 - \alpha) \) adjusts accordingly to yield a test with size \( \alpha \) using the dependence of the null distribution of \( T_n \) on the strength of the IV’s. 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-CQLRP 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) \( \{V_i \in \mathbb{R}^{p+1} : i \geq 1\} \) are i.i.d., \( \{Z_i \in \mathbb{R}^k : i \geq 1\} \) are fixed, not random, and \( k \geq p \).

(b) \( EV_i = 0 \), \( \Sigma_V := E V_i V_i' \) is pd, and \( E||V_i||^4 < \infty \)

(c) \( n^{-1} \sum_{i=1}^{n} Z_i Z_i' \to K_Z \) for some pd matrix \( K_Z \in \mathbb{R}^{k \times k} \), \( n^{-1} \sum_{i=1}^{n} ||Z_i||^6 = o(n) \), and

\[ \sup_{1 \leq i \leq n} (c'Z_i)^2 / \sum_{i=1}^{n} (c'Z_i)^2 \to 0 \forall c \neq 0^k. \]

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

\(^{54}\) 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.

\(^{55}\) 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 \( \pi_n \) may depend on \( n \).
(e) \( \lambda_{\text{max}}(\Sigma_V)/\lambda_{\text{min}}(\Sigma_V) \leq 1/\varepsilon \) 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 \( \tilde{\sigma}_n^2(=\tilde{\sigma}_n^2(\theta_0)) = k \) wp→1 (by Lemma 19.1(b) below) and no SR adjustment of the SR-CQLR tests occurs (wp→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 \( \alpha \), 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.

19.3 SR-CQLR\(_P\) Test

The components of the SR-QLR\(_P\) statistic and its conditioning matrix are \( n^{1/2}\tilde{\Omega}_n^{-1/2}g_n \) and \( n^{1/2}\tilde{D}_n^* \) (see (4.2) and (15.6)) when \( \tilde{\sigma}_n = k \), which holds wp→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 \( \tilde{\Omega}_n \) and \( \tilde{\Sigma}_n \) 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 \( \{\pi_n \in \mathcal{P} : n \geq 1\} \) and any \( p \geq 1 \), we have

(a) \( \tilde{R}_n \overset{\text{p}}{\to} \Sigma_V \otimes K_Z, \)
(b) \( \tilde{\Omega}_n \overset{\text{p}}{\to} (b_0'\Sigma_V b_0)K_Z, \) where \( b_0 := (1, -\theta_0')' \),
(c) \( \tilde{\Sigma}_n \overset{\text{p}}{\to} (b_0'\Sigma_V b_0)^{-1}\Sigma_V, \)
(d) \( \tilde{\Sigma}_n \overset{\text{p}}{\to} (b_0'\Sigma_V b_0)^{-1}\Sigma_V, \)
(e) \( n^{1/2}\tilde{\sigma}_n^{-1/2}\tilde{\gamma}_n = \overline{S}_n + o_p(1), \) and
(f) \( n^{1/2}\tilde{D}_n^* = -(I_k + o_p(1))\overline{T}_n(I_p + o_p(1)) + o_p(1). \)

Comments: (i) The minus sign in Lemma 19.1(f) is not important because QLR\(_Pn\) (defined in the paragraph containing (15.7) using the formula in (5.7)) is unchanged if \( \tilde{D}_n^* \) is replaced by \( -\tilde{D}_n^* \).
(and SR-QLR\textsubscript{Pn} = QLR\textsubscript{Pn} wp→1 under Assumption HLIV)\textsuperscript{56}

(ii) The results of Lemma \textsuperscript{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 \textsuperscript{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 \(\pi_n\). The reduced-form equation (19.2) states that \(y_{1i} = Z_i'\pi_n\theta_n + V_{1i}\) when \(\pi_n\) and \(\theta_n\) are the true values of \(\pi\) and \(\theta\). Contiguous alternatives to the null distributions with parameters \(\pi_n\) and \(\theta_0\) are obtained for parameter values \(\pi_n\) and \(\theta_n (\neq \theta_0)\) that satisfy \(\pi_n\theta_n - \pi_n\theta_0 = \pi_n(\theta_n - \theta_0) = O(n^{-1/2})\). If the IV’s are strong, i.e., \(\liminf_{n→∞} \pi_n'n^{-1}\sum_{i=1}^{n} Z_i'Z_i\pi_n > 0\), then contiguous alternatives have true \(\theta_n\) values of distance \(O(n^{-1/2})\) from the null value \(\theta_0\). If the IV’s are weak in the standard sense, e.g., \(\pi_n = πn^{-1/2}\) for some fixed matrix \(\pi\), then all \(\theta\) values not equal \(\theta_0\) yield contiguous alternatives. For semi-strong identification in the standard sense, e.g., \(\pi_n = \pi n^{-δ}\) for some \(δ \in (0,1/2)\) and some fixed full-column-rank matrix \(\pi\), the contiguous alternatives have \(\theta_n - \theta_0 = O(n^{-(1/2-δ)})\). For joint weak identification, contingency occurs when \(\pi_n = (π_{1n}, \ldots, π_{pn}) ∈ R^{k×p}, n^{1/2}\|\pi_{jn}\| → ∞\) for all \(j ≤ p\), \(\limsup_{n→∞} λ_{min}(np'n\pi_n) < ∞\), and \(\theta_n\) is such that \(\pi_n(\theta_n - \theta_0) = O(n^{-1/2})\).

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

19.4 SR-CQLR Test

The components of the SR-QLR statistic and its conditioning matrix are \(n^{1/2}\hat{Ω}_n^{1/2}g_n\) and \(n^{1/2}\hat{D}_n^*\) (see (4.1) and (5.7)) when \(\hat{n} = k\), which holds wp→1 under Assumption HLIV. Here we show that the conditioning statistic \(n^{1/2}\hat{D}_n^*\) is asymptotically equivalent to Moreira’s (2003) conditioning statistic \(\hat{T}_n\) (in the homoskedastic linear IV model with fixed IV’s) when \(\pi_n → 0^{k×p}\). 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 \textsuperscript{19.2} below.) Nevertheless, under strong and semi-strong IV’s, the SR-CQLR test and Moreira’s CLR test are asymptotically equivalent.\textsuperscript{57} In consequence, when \(p = 1\), the SR-CQLR test and Moreira’s CLR test are asymptotically equivalent.\textsuperscript{57}

\textsuperscript{56}This holds because for \(a_1 ∈ R^k\) and \(A_2 ∈ R^{k×p}\) we have \(\lambda_{min}(a_1, −A_2)'(a_1, −A_2) = \inf_{x_2=(λ_1, x_2)'\|\lambda\|=1}(a_1λ_1 − A_2x_2)'(a_1λ_1 − A_2x_2) = \inf_{x_2=(λ_1, x_2)'\|\lambda\|=1}(a_1λ_1 + A_2λ_2)'(a_1λ_1 + A_2λ_2) = \inf_{x_2=(λ_1, x_2)'\|\lambda\|=1}(a_1λ_1 + A_2λ_2)'(a_1λ_1 + A_2λ_2) = \lambda_{min}(a_1, A_2)'(a_1, A_2)).

\textsuperscript{57}This holds because, under strong and semi-strong IV’s, the SR-QLR statistic and Moreira’s CLR statistic behave asymptotically like LM statistics that project onto \(n^{1/2}\hat{Ω}_n^{1/2}\hat{D}_n\) (or equivalently, \(n^{1/2}\hat{Ω}_n^{1/2}\hat{D}_n\hat{L}_n^{1/2}\)) and \(\hat{T}_n\), respectively, see Theorem \textsuperscript{7.1} for the SR-QLR statistic, and \(n^{1/2}\hat{Ω}_n^{1/2}\hat{D}_n\hat{L}_n^{1/2}\) and \(\hat{T}_n\) are asymptotically equivalent (up to multiplication by \(-1\)) by Lemma \textsuperscript{19.1(f)}. Furthermore, the conditional critical values of the two tests both converge to the same limit.
totically 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 \( \pi_n \rightarrow 0^{k \times p} \), if \( n^{1/2} \) times the smallest singular value of \( \pi_n \) 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 IV’s and random IV’s (which cause \( \Sigma_V \) to differ somewhat from \( \Sigma_{V*} \) defined below).

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

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

(19.6)

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

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

(19.7)

where \( B = B(\theta_0) \) is defined in \((5.3)\).

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

\[
\Sigma_{V* j \ell} := tr(R_{j \ell}(\pi*)' K_Z^{-1})/k,
\]

(19.8)

where \( R_{j \ell}(\pi*) \) denotes the \((j, \ell) \) \( k \times k \) submatrix of \( R(\pi*) \) for \( j, \ell = 1, ..., p+1 \) and \( \pi* = \lim \pi_n \).

Equivalently, \( \Sigma_{V*} \) is the unique minimizer of \( ||[I_{p+1} \otimes ((b_0' \Sigma_V b_0)^{-1/2} K_Z^{-1/2})][\Sigma \otimes K_Z - R(\pi*)] [I_{p+1} \otimes ((b_0' \Sigma_V b_0)^{-1/2} K_Z^{-1/2})]|| \) over all symmetric pd matrices \( \Sigma \in R^{(p+1) \times (p+1)} \). Note that when \( \zeta(\pi*) = 0 \) (as occurs when \( \pi* = 0^{k \times p} \), \( \Sigma_{V*} = \Sigma_V \) (because \( R(\pi*) = \Sigma_V \otimes K_Z \) in this case).

We use the following assumption.

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

(b) \( \pi_n \rightarrow \pi* \) for some \( \pi* \in R^{k \times p} \), and

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

Assumption HLIV2(c) implies that the eigenvalue adjustment to \( \hat{\Sigma} \) 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, in probability to \( \chi^2_{p, 1-\alpha} \) under strong and semi-strong identification, see Theorem 7.1 for the SR-CQLR critical value.
we do not do so here.

The asymptotic behavior of \( n^{1/2} \hat{D}_n^* \) is given in the following lemma. Under Assumption HLIV, \( n^{1/2} \hat{D}_n^* \) equals the SR-CQLR conditioning statistic \( n^{1/2} \hat{D}_{\lambda n}^* \) \( \rightarrow \mathcal{D} \) (because \( \hat{r}_n = k \mathcal{D} \rightarrow 1 \)).

**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) \( \hat{R}_n \rightarrow_p R(\pi_*) \),

(b) \( \hat{\Sigma}_n \rightarrow_p (b_0' \Sigma \sigma_0 b_0)^{-1} \Sigma \),

(c) \( \hat{\Sigma}_n^\varepsilon \rightarrow_p (b_0' \Sigma \sigma_0 b_0)^{-1} \Sigma \), and

(d) \( n^{1/2} \hat{D}_n^* = -(I_k + o_p(1)) \hat{T}_n(L_{V_0}^{-1/2} L_{V_*}^{1/2} + o_p(1)) + o_p(1) \), where \( L_{V_0} := (\theta_0, I_p) \Sigma^{-1} \Sigma V_* (\theta_0, I_p)' \in \mathbb{R}^{p \times p} \) and \( L_{V_*} := (\theta_0, I_p) \Sigma V_*^{-1} (\theta_0, I_p)' \in \mathbb{R}^{p \times p} \).

**Comments:** (i) If \( \pi_* = 0^{k \times p} \), which occurs when all \( \theta \) parameters are either weakly identified in the standard sense or semi-strongly identified, then \( \zeta(\pi_*) = 0^{k p \times k p} \), \( R(\pi_*) = \Sigma V_* \otimes K \), and \( \Sigma V_* = \Sigma V \). In this case, Lemma 19.2(d) yields

\[
n^{1/2} \hat{D}_n^* = -(I_k + o_p(1)) \hat{T}_n(I_p + o_p(1)) + o_p(1) \tag{19.9}
\]

and \( n^{1/2} \hat{D}_n^* \) is asymptotically equivalent to \( T_n \) (up to multiplication by \(-1\)).

(ii) On the other hand, if \( \pi_* \neq 0^{k \times p} \), then \( n^{1/2} \hat{D}_n^* \) is not asymptotically equivalent to \( T_n \) 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 IV’s are fixed in the linear IV model (by assumption), but the variance estimator \( \hat{V}_n \), which appears in \( \hat{R}_n \), see (5.3), and which determines \( \hat{\Sigma}_n \) and \( \hat{\Sigma}_V \), treats the IV’s 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 IV’s. The behavior of Kleibergen’s nonlinear CLR tests is found to depend on the choice of weighting matrix for the conditioning statistic. We find that when \( p = 1 \) (where \( p \) is the dimension of \( \theta \)) 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 $\Sigma_V$ and null hypothesis value $\theta_0$). This has adverse effects on the power of the moment-variance weighted CLR test.

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

Kleibergen’s test statistic takes the form:

$$CLR_n(\theta) := \frac{1}{2} \left( AR_n(\theta) - r k_n(\theta) + \sqrt{(AR_n(\theta) - r k_n(\theta))^2 + 4LM_n(\theta) \cdot r k_n(\theta)} \right),$$

where

$$LM_n(\theta) := n \hat{g}_n(\theta)' \hat{\Omega}_n^{-1/2}(\theta) P_{\hat{\Omega}_n^{-1/2}(\theta) \hat{D}_n(\theta)} \hat{\Omega}_n^{-1/2}(\theta) \hat{g}_n(\theta)$$

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

The critical value of Kleibergen’s CLR test is $c(1 - \alpha, r k_n(\theta))$, where $c(1 - \alpha, r)$ is the $1 - \alpha$ quantile of the distribution of

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

for $0 \leq r < \infty$ and the chi-square random variables $\chi^2_p$ and $\chi^2_{k-p}$ in (19.11) are independent. The CLR test rejects the null hypothesis $H_0 : \theta = \theta_0$ if $CLR_n > c(1 - \alpha, r k_n)$ (where, as elsewhere, the dependence of these statistics on $\theta_0$ is suppressed for simplicity).

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

$$\tilde{V}_{Dn}(\theta) := n^{-1} \sum_{i=1}^{n} \text{vec}(G_i(\theta) - \hat{G}_n(\theta)) \text{vec}(G_i(\theta) - \hat{G}_n(\theta))' - \hat{\Gamma}_n(\theta) \hat{\Omega}_n^{-1}(\theta) \hat{\Gamma}_n(\theta)',$$

where

$$\hat{\Gamma}_n(\theta) := (\hat{\Gamma}_{n1}(\theta)', ..., \hat{\Gamma}_{nk}(\theta))' \in R^{pk \times k}$$

(19.12)
and $\tilde{\Gamma}_1(\theta), \ldots, \tilde{\Gamma}_{pn}(\theta)$ are defined in [5.2].

Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) propose to take $rk_n(\theta)$ to be a function of $\tilde{\Omega}_n^{-1/2}(\theta) \tilde{D}_n(\theta)$. We refer to $\tilde{\Omega}_n^{-1/2}(\theta) \tilde{D}_n(\theta)$ 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 $\tilde{V}_D(\theta)$, but not with a rank statistic based on $\tilde{\Omega}_n(\theta)$. This illustrates that the flexibility in the choice of the rank statistic for Kleibergen’s CLR test can have drawbacks. It may lead to a test that has reduced power.

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

$$CLR_n(\theta) = AR_n(\theta) - \lambda_{\min}((n^{1/2}\tilde{\Omega}_n^{-1/2}(\theta)\tilde{g}_n(\theta), r_n(\theta))'(n^{1/2}\tilde{\Omega}_n^{-1/2}(\theta)\tilde{g}_n(\theta), r_n(\theta))) \text{ provided}$$

$$rk_n(\theta) = r_n(\theta)'r_n(\theta)$$

for some random vector $r_n(\theta) \in \mathbb{R}^k$. (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 $rk_n(\theta) = r_n(\theta)'r_n(\theta)$ is of the same form as Moreira’s (2003) LR statistic with $r_n(\theta_0)$ in place of $\tilde{g}_n$ and with $n^{1/2}\tilde{\Omega}_n^{-1/2}(\theta_0)\tilde{g}_n(\theta_0)$ in place of $\tilde{\Omega}_n$, where $\theta_0$ is the null value of $\theta$. The two choices for $rk_n(\theta)$ that we consider when $p = 1$ are

$$rk_{1n}(\theta) := n\tilde{D}_n(\theta)'\tilde{V}_D^{-1}(\theta)\tilde{D}_n(\theta) \text{ and } rk_{2n}(\theta) := n\tilde{D}_n(\theta)'\tilde{\Omega}_n^{-1}(\theta)\tilde{D}_n(\theta).$$

(19.14)

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

Let

$$\zeta_n(\pi) := n^{-1}\sum_{i=1}^{n} Z_iZ_i'\pi^2 - \left(n^{-1}\sum_{i=1}^{n} Z_iZ_i'\pi\right)\left(n^{-1}\sum_{i=1}^{n} Z_iZ_i'\pi\right)'.$$  

(19.15)

58The functional form of the rank statistics that have been considered in the literature, such as the statistics of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006) all reduce to the same function when $p = 1$. Specifically, $rk_n(\theta)$ equals the squared length of some $k$ vector $r_n(\theta)$. 

58
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) $rk_{1n}(\theta_0) = \mathbf{\bar{F}}_n'[\mathbf{I}_k + L_{V0}K_{Z}^{-1/2}\zeta_n(\pi_n)K_{Z}^{-1/2} + o_p(1)]^{-1}\mathbf{\bar{F}}_n \cdot (1 + o_p(1)) + o_p(1)$,

(b) $rk_{2n}(\theta_0) = \mathbf{\bar{F}}_n'[L_{V0}b_0'S_Vb_0]^{-1} \cdot (1 + o_p(1)) + o_p(1)$, where $L_{V0} := (\theta_0, 1)\Sigma_V^{-1}(\theta_0, 1)' \in R$, and

(c) $L_{V0}b_0'S_Vb_0 = \frac{(1-2\theta_0\rho + \theta_0^2\rho^2)^2}{c^2(1-\rho^2)}$, where $c^2 := Var(V_{2i})/Var(V_{1i}) > 0$ and $\rho = Corr(V_{1i}, V_{2i}) \in (-1, 1)$.

**Comments:** (i) If $\pi_n \to 0$, then $\zeta_n(\pi_n) \to 0$ and Lemma 19.3(a) shows that $rk_{1n}(\theta_0)$ equals $\mathbf{\bar{F}}_n'[\mathbf{I}_k + o_p(1)] + o_p(1)$. That is, under weak IV’s and semi-strong IV’s, $rk_{1n}(\theta_0)$ reduces asymptotically to Moreira’s (2003) conditioning statistic. Under strong IV’s, this does not occur. However, under strong IV’s, we have $rk_{1n}(\theta_0) \rightarrow_p \infty$, just as $\mathbf{\bar{F}}_n'[\mathbf{I}_k] \rightarrow_p \infty$. In consequence, the test constructed using $rk_{1n}(\theta_0)$ has the same asymptotic properties as Moreira’s (2003) CLR test under the null and contiguous alternative distributions.

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

(iii) Lemma 19.3(b) shows that the rank statistic $rk_{2n}(\theta_0)$ differs asymptotically from Moreira’s conditioning statistic $\mathbf{\bar{F}}_n'[\mathbf{I}_k]$ by the scale factor $\left(L_{V0}b_0'S_Vb_0\right)^{-1}$. Thus, the nonlinear CLR test considered by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) does not reduce asymptotically to Moreira’s (2003) CLR test in the homoskedastic linear IV regression model with fixed IV’s under weak IV’s. This has negative consequences for its power. Under strong or semi-strong IV’s, this test does reduce asymptotically to Moreira’s (2003) CLR test because $rk_{1n}(\theta_0) \rightarrow_p \infty$, just as $\mathbf{\bar{F}}_n'[\mathbf{I}_k] \rightarrow_p \infty$, which is sufficient for asymptotic equivalence in these cases.

(iv) For example, if $\rho = 0$ and $c = 1$ in Lemma 19.3(c), then $\left(L_{V0}b_0'S_Vb_0\right)^{-1} = (1+\theta_0^2)^{-2} \leq 1$. In this case, if $|\theta| = 1$, then $\left(L_{V0}b_0'S_Vb_0\right)^{-1} = 1/4$ and $rk_{2n}(\theta_0)$ is $1/4$ as large as $\mathbf{\bar{F}}_n'[\mathbf{I}_k]$ asymptotically. On the other hand, if $\rho = 0$ and $\theta_0 = 0$, then $\left(L_{V0}b_0'S_Vb_0\right)^{-1} = c^2$, which can be arbitrarily close to zero or infinity depending on $c$.

(v) When $\left(L_{V0}b_0'S_Vb_0\right)^{-1}$ is large (small), the $rk_{2n}(\theta_0)$ statistic is larger (smaller) than desired and it behaves as though the IV’s are stronger (weaker) than they really are, which sacrifices power unless the IV’s are quite strong (weak). Note that the inappropriate scale of $rk_{2n}(\theta_0)$ does not cause asymptotic size problems, only power reductions.
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 \( r_{kn}(\theta_0) \) 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), \( n^{1/2}\hat{\Omega}_n^{-1/2}\hat{g}_n = \hat{S}_n + o_p(1) \), where, here and below, we suppress the dependence of various quantities on \( \theta_0 \). Hence, \( AR_n = \hat{S}_n \hat{S}_n + o_p(1) \). Even if \( r_{kn} \) takes the form \( r_n^\prime r_n \) for some random \( k \) vector \( r_n \), it is not the case that

\[
CLR_n = AR_n - \lambda_{\min}\left((n^{1/2}\hat{\Omega}_n^{-1/2}\hat{g}_n, r_n)^\prime(n^{1/2}\hat{\Omega}_n^{-1/2}\hat{g}_n, r_n)\right)
\]  

(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, viz., those of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006), \( r_{kn} \) is not of the form \( r_n^\prime r_n \), when \( p \geq 2 \).

Third, Moreira’s conditioning statistic is the \( k \times p \) matrix \( T_n \). Conditioning on this random matrix is equivalent asymptotically to conditioning on the \( k \times p \) matrix \( n^{1/2}\hat{D}_n^* \) 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 \( \hat{\Omega}_n^{-1} \), 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
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 

\[ Z_i, \pi \in \mathbb{R}^{d_z} \]

\[ \theta = (\beta, \pi)' \in \mathbb{R}^{d_\theta + d_\pi} \]

\[ EU_i = EV_{2i} = 0 \]

\[ EU_i Z_i = EV_{1i} Z_i = 0^{d_z} \]

and 

\[ E(V_i V_i' | Z_i) = \Sigma_V \] a.s. for 

\[ V_i := (V_{1i}, V_{2i})' \]

and some 

\[ 2 \times 2 \]

constant matrix \( \Sigma_V \). 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 \( \theta \) are 

\[ g_i(\theta) = (y_{1i} - Z_i'\pi \beta)(Y_2i - Z_i'\pi)' \in \mathbb{R}^k \]

where 

\[ k = 2d_Z \]

and 

\[ d_Z \]

is the dimension of 

\( Z_i \). The variance matrix 

\[ \Sigma_V \otimes E Z_i Z_i' \]

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 

\[ E Z_i Z_i' \]

is singular. In this model, we are interested in 

joint inference concerning \( \beta \) and \( \pi \). This is of interest when one wants to see how the magnitude of \( \pi \) affects the range of plausible \( \beta \) values.

We take 

\[ (V_{1i}, V_{2i}) \sim N(0^2, \Sigma_V) \]

where 

\( \Sigma_V \) has unit variances and correlation 

\[ \rho_V, Z_i \sim N(0^2, I_{d_Z}), \]

\( (V_{1i}, V_{2i}) \) and 

\( Z_i \) are independent, and the observations are i.i.d. across \( i \). The null hypothesis is 

\[ H_0 : (\beta, \pi) = (\beta_0, \pi_0) \]

We consider the values: 

\[ \rho_V = .95, .999, 999, \] and 

\[ 1.0; n = 250, 500, 1,000, \]

\[ 2,000, 4,000, 8,000, \] and 

\[ 16,000; \pi_0 = (\pi_{10}, 0, 0, 0)' \]

where 

\[ \pi_{10} = \pi_{10n} = C/n^{1/2} \]

and 

\[ C = \sqrt{10} \]

which yields a concentration parameter of 

\[ \lambda = \pi' E Z_i Z_i' \pi = 10 \]

for all \( n \geq 1 \); and \( \beta_0 = 0 \). The variance matrix \( \Omega_F \) 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 

\[ \rho_V \]

is close to one. Under \( H_0 \), with probability one, the extra rejection condition in \( (4.7) \) is: reject \( H_0 \) if 

\[ [I_4,-I_4]g_i(\theta_0) \neq 0^4 \]

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-I, SM-II, and SM-III report results for \( k = 8 \) (which corresponds to \( d_Z = 4 \)), \( k = 4 \), and \( k = 12 \), respectively. Table SM-I shows that the SR-AR and SR-CQLR tests have null rejection

59 Analogous results for the SR-CQLR2 test are not provided because the moment functions considered are not of the form in \( (15.1) \), which is necessary to apply the SR-CQLR2 test.
probabilities that are close to the nominal 5% level for singular and near singular variance matrices as measured by $\rho_V$. As expected, the deviations from 5% decrease with $n$. For all 40,000 simulation repetitions, all values of $n$ considered, and $k = 8$, we obtain $\hat{r}_n(\theta_0) = 8$ when $\rho_V < 1.0$ and $\hat{r}_n(\theta_0) = 4$ when $\rho_V = 1$. The estimator $\hat{r}_n(\theta_0)$ also makes no errors when $k = 4$ and 12. Tables SM-II and SM-III show that the deviations of the null rejection probabilities from 5% are somewhat smaller when $k = 4$ and $n \leq 1000$ than when $k = 8$, and somewhat larger when $k = 12$ and $n \leq 500$. 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-CQLRP tests, Kleibergen’s CLR test with Jacobian-variance weighting (JVW-CLR), and the CLR test of Moreira (2003) (Mor-CLR). 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. 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 $\Sigma_V$ and $p = 1$. The IV’s are fixed—they are generated once from a $N(0^k, I_k)$ distribution. The sample size $n$ equals 1,000. The hypotheses are $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$. The tests have nominal size .05. The power results are based on 40,000 simulation repetitions and 1,000 critical value repetitions and are size-corrected (by adding non-negative constants to the critical values of those tests that over-reject under the null). The reduced-form error variances and correlation are denoted by $\Sigma_{V11}$, $\Sigma_{V22}$, and $\rho$, respectively, and $\lambda := \pi'Z'Z\pi$. The number of IV’s is $k$. The MVW-CLR and JVW-CLR tests employ the Robin and Smith (2000) rank statistic. 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-CQLRP tests, Kleibergen’s CLR test with Jacobian-variance weighting (JVW-CLR), and the CLR test of Moreira (2003) (Mor-CLR). Note that the second formula for $rk_n$ is appropriate only for the case $p = 1$, which is the case considered here. The estimators $\hat{\Omega}_n$ and $\hat{V}_{Dna}$ are estimators of the asymptotic variances of $\hat{D}_n$ (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 $\hat{\Omega}_n$ and $\hat{V}_{Dna}$ are estimators of the asymptotic variances of the sample moments and Jacobian, respectively, which leads to the MVW and JVW terminology.

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 $rk_n = \lambda_{min}(n\hat{D}_n\hat{\Omega}_n^{-1/2}\hat{D}_n)$ and $rk_n = \lambda_{min}(n\hat{D}_n\hat{V}_{Dna}^{-1}\hat{D}_n)$, respectively, where $\hat{\Omega}_n$ and $\hat{D}_n$ are defined in (4.1) and (5.2) with $\theta = \theta_0$ and $\hat{V}_{Dna}$ is an estimator of the asymptotic variance of $\hat{D}_n$ (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 $\hat{\Omega}_n$ and $\hat{V}_{Dna}$ 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 $(L_{V\theta_0^T\Sigma_V\theta_0})^{-1}$ in Lemma 19.3(b) and (c).
SR-CQLR, 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, JVW-CLR, and Mor-CLR tests, for positive $\theta$, and low power, like that of the MVW-CLR and LM tests, for negative $\theta$.

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

The eigenvalue-adjustment procedure can be applied to any non-zero psd matrix $H \in R^{d_H \times d_H}$ for some positive integer $d_H$. Let $\varepsilon$ be a positive constant. Let $A_H \Lambda_H A_H'$ be a spectral decomposition of $H$, where $\Lambda_H = Diag\{\lambda_{H1}, \ldots, \lambda_{Hd_H}\} \in R^{d_H \times d_H}$ 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

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

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

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

**Lemma 22.1** Let $d_H$ be a positive integer, let $\varepsilon$ be a positive constant, and let $H \in R^{d_H \times d_H}$ be a non-zero positive semi-definite non-random matrix. Then,

(a) (uniqueness) $H^\varepsilon$, defined in (22.1), is uniquely defined. (That is, every choice of spectral decomposition of $H$ yields the same matrix $H^\varepsilon$),

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

57

Electronic copy available at: https://ssrn.com/abstract=3366443
Comments: (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 $\varepsilon$ to control the condition number of $H^\varepsilon$. 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 $\$1,000 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 $\{\hat{H}_n : n \geq 1\}$ for which $\hat{H}_n \to_p H$, one has $\hat{H}_n^\varepsilon \to_p H^\varepsilon$.

Proof of Lemma 22.1 For notational simplicity, we drop the $H$ subscript on $A_H$, $\Lambda_H$, and $\Lambda_H^\varepsilon$. We prove part (a) first. The eigenvectors of $H^\varepsilon (= AA^\varepsilon A')$ defined in (5.6) 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, $H^\varepsilon = AA^\varepsilon A'$. To establish part (a), if suffices to show that $H^\varepsilon = A_* \Lambda^\varepsilon A'_*$, or equivalently, $AA^\varepsilon A'_* = A_* \Lambda^\varepsilon A'_*$ for any $\xi \in R^{d_H}$.

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

$$AA^\varepsilon A'_* \xi = A\Lambda^\varepsilon (A_1, A_2, A_3)'(\xi_1 + \xi_2)$$

$$= A\Lambda^\varepsilon (0^{j-1}, (A_2')_2(\xi_2), 0^{d_H - j - d})' + A\Lambda^\varepsilon ((A_1')_2(\xi_2), 0^{d+1}, (A_3')_2)'$$

$$= A\Lambda^\varepsilon_j (0^{j-1}, (A_2')_2(\xi_2), 0^{d_H - j - d})' + (A_1, A_2, A_3)\Lambda^\varepsilon ((A_1')_2(\xi_2), 0^{d+1}, (A_3')_2)'$$

$$= A_2 A_2' \xi_1 + (A_1, A_3)\Lambda^\varepsilon ((A_1')_2(\xi_2), (A_3')_2)'$$

$$= A_* \Lambda^\varepsilon A'_* \xi,$$

where $\Lambda^\varepsilon \in R^{(d_H - d_1) \times (d_H - d_1)}$ is the diagonal matrix equal to $\Lambda^\varepsilon$ with its $j, \ldots, j + d$ rows and columns deleted, $\Lambda^\varepsilon_j = \max\{\lambda_j, \lambda_{\max}(H)\varepsilon\}$, $\lambda_j$ is the $j$th eigenvalue of $\Lambda$, the second equality uses $A_1' \xi_1 = 0^{j-1}$, $A_3' \xi_2 = 0^{d_H - j - d}$, and $A_2' \xi_2 = 0^{d+1}$, the third equality holds because $\lambda_j = \ldots = \lambda_{j+d}$ implies that $\Lambda^\varepsilon_j = \ldots = \Lambda^\varepsilon_{j+d}$, the fourth equality holds using the definition of $\Lambda^\varepsilon$, the fifth equality holds because $A_2 A_2' = A_2 A_2'$ (since both equal the projection matrix onto the column space of
\( A_2 \) (and \( A_{2*} \)), and the last equality holds by reversing the steps in the previous equalities with \( A_*= (A_1, A_{2*}, A_3) \) in place of \( A_*(A_1, A_2, A_3) \). Because \([22.2]\) 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 \( H^\varepsilon \) are \( \{\max\{\lambda_{Hj}, \lambda_{\max}(H)\varepsilon\} : j = 1, ..., d_H\} \) because \( H^\varepsilon = AA^\varepsilon A' \) and \( A \) is an orthogonal matrix. In consequence, \( \lambda_{\min}(H^\varepsilon) \geq \lambda_{\max}(H)\varepsilon \), which establishes part (b). If \( \lambda_{\min}(H) > \lambda_{\max}(H)\varepsilon \), then \( H^\varepsilon = H, \lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H)/\lambda_{\min}(H) < 1/\varepsilon \), and the result of part (c) holds. Alternatively, if \( \lambda_{\min}(H) \leq \lambda_{\max}(H)\varepsilon \), then \( \lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H)\varepsilon \). In addition, we have \( \lambda_{\max}(H^\varepsilon) = \max\{\lambda_{H1}, \lambda_{\max}(H)\varepsilon\} = \lambda_{\max}(H) \times \max\{1, \varepsilon\} \) using \( \lambda_{H1} = \lambda_{\max}(H) \). Combining these two results gives \( \lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) = \lambda_{\max}(H)\max\{1, \varepsilon\}/(\lambda_{\max}(H)\varepsilon) = \max\{1/\varepsilon, 1\} \), where the second equality uses the assumption that \( H \) is non-zero, which implies that \( \lambda_{\max}(H) > 0 \). This gives the result of part (c).

We now prove part (d) and for clarity make the \( H \) subscripts on \( \Lambda_H \) and \( \Lambda_H \) 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 \( \Lambda_H^\varepsilon \) in \([5.6]\)) and \( (cH)^\varepsilon = A_{cH}\Lambda_{cH}^\varepsilon A'_{cH} = cA_H\Lambda_H^\varepsilon A'_{cH} = cH^\varepsilon \), which establishes part (d).

Now we prove part (e). Let \( A_n\Lambda_n A'_n \) be a spectral decomposition of \( H_n \) for \( n \geq 1 \). Let \( H_n^\varepsilon = A_n\Lambda_n^\varepsilon A'_n \) for \( n \geq 1 \), where \( \Lambda_n^\varepsilon \) is the diagonal matrix with \( j \)th diagonal element given by \( \lambda_{nj}^\varepsilon = \max\{\lambda_{nj}, \lambda_{\max}(H_n)\varepsilon\} \) and \( \lambda_{nj} \) is the \( j \)th largest eigenvalue of \( H_n \). (By part (a) of the Lemma, \( H_n^\varepsilon \) 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 subsequence \( \{n_m\} \). (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 \( A_n \to A \), where \( A \) is the diagonal matrix of eigenvalues of \( H \) in nonincreasing order (by Elsner’s Theorem, see Stewart [2001, Thm. 3.1, pp. 37–38]). In turn, this gives \( \Lambda_n^\varepsilon \to \Lambda^\varepsilon \), where \( \Lambda^\varepsilon \) is the diagonal matrix with \( j \)th diagonal element given by \( \lambda_j^\varepsilon = \max\{\lambda_j, \lambda_{\max}(H)\varepsilon\} \) and \( \lambda_j \) is the \( j \)th largest eigenvalue of \( H \), because \( \lambda_{\max}(\cdot) \) is a continuous function (by Elsner’s Theorem again). The previous results imply that \( H_{nm} = A_{nm}\Lambda_{nm} A'_{nm} \to AA' \), \( H = AA' \), \( H_n^\varepsilon = A_{nm}\Lambda_n^\varepsilon A'_{nm} \to AA^\varepsilon A' \), and \( AA^\varepsilon A' = H^\varepsilon \). Because every subsequence \( \{n_\ell\} \) of \( \{n\} \) has a subsubsequence \( \{n_m\} \) for which \( H_{nm}^\varepsilon \to H^\varepsilon \), we obtain \( H_n^\varepsilon \to H^\varepsilon \), which completes the proof of part (e). \( \Box \)
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 CS’s. However, these procedures are only partially singularity robust, see the discussion below. In addition, LM tests have low power in some circumstances under weak identification.

The SR-LM test statistic is

$$SR-LM_n(\theta) := n\tilde{g}_{An}(\theta)^{T} P_{\tilde{\Omega}^{1/2}}(\theta) \tilde{D}_{An}(\theta) \tilde{g}_{An}(\theta),$$

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

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

where $\chi^2_{\min(\tilde{r}_n(\theta_0), p), 1-\alpha}$ denotes the $1 - \alpha$ quantile of a chi-squared distribution with $\min\{\tilde{r}_n(\theta_0), p\}$ 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 $\delta_1 > 0$,

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

where $\mathcal{F}^{SR}_{LMj} := \{F \in \mathcal{F}^{SR} : \tau_{jF} \geq \delta_1 \text{ and } \lambda_{p-j} \left( \Psi_F^{C_{j-k-j}^* G_i^* B_{p-j}^*} \right) \geq \delta_1 \forall \xi \in R^{q-j} \text{ with } ||\xi|| = 1\},

\begin{align*}
G_i^* &:= \Pi_{1F}^{-1/2} A_{F}G_i \in R^{q \times p}, \ \tau_F := rk(\Omega_F), \ g_i^* := \Pi_{1F}^{-1/2} A_{F}g_i \in R^{q_F}, \\
\Psi_F^{g_i} &:= E_F a_i a_i^{T} - E_F a_i g_i^{*} (E_F g_i^{*} g_i^{*})^{-1} E_F g_i^{*} a_i^{T} \text{ for any random vector } a_i,
\end{align*}

$$\tau_{jF}^* := r_{jF}^{*} \text{ is the } j \text{th largest singular value of } E_F G_i^{*} \text{ for } j = 1, \ldots, \min\{r_F, p\}, \ \tau_{0F}^{*} := \delta_1, \ B_{F}^{*} \text{ is a } p \times p \text{ orthogonal matrix of eigenvalues of } (E_F G_i^{*})'(E_F G_i^{*}), \text{ ordered so that the corresponding eigenvalues } (\kappa_{1F}^{*}, \ldots, \kappa_{pF}^{*}) \text{ are nonincreasing, } C_{F}^{*} \text{ is an } r_F \times r_F \text{ orthogonal matrix of eigenvalues of } (E_F G_i^{*})'(E_F G_i^{*})' \text{ ordered so that the corresponding eigenvalues } (\kappa_{1F}^{*}, \ldots, \kappa_{r_F}^{*}) \text{ are nonincreasing, } B_{F}^{*} := (B_{F,j}^{*}, B_{F,p-j}^{*}) \text{ for } B_{F,j}^{*} \in R^{q \times j} \text{ and } B_{F,k-j}^{*} \in R^{q \times (p-j)}, \text{ and } C_{F}^{*} := (C_{F,j}^{*}, C_{F,k-j}^{*}) \text{ for } C_{F,j}^{*} \in R^{q_F \times j} \text{ and } C_{F,k-j}^{*} \in R^{q_F \times (r_F-j)}.$$
parameter space and the quantities upon which it depends. Note that \( \Psi_T^{\text{st}} \) is the expected outer-product matrix of the vector of residuals, \( a_i - E_F a_i g_i^* (E_F g_i^* g_i^*)^{-1} g_i^* \), 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 \( \mathcal{F}_{LM}^{SR} \) (beyond those in \( \mathcal{F}^{SR} \)) are used to guarantee that the conditioning matrix \( \hat{D}_{An} \in \mathbb{R}^{n \times p} \) has full rank \( \min\{n,p\} \) asymptotically with probability one (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 CS’s are analogous to those for the SR-AR and SR-CQLR CS’s.

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

### 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 \( D = C \Sigma B' \), where \( C \) is a \( k \times k \) orthogonal matrix of eigenvectors of \( DD' \), \( B \) is a \( p \times p \) orthogonal matrix of eigenvectors of \( D'D \), and \( \Sigma \) 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 \( \tau_j \) is nonincreasing in \( j \). Then, \( c_{k,p}(D,1-\alpha) = c_{k,p}(\Sigma,1-\alpha) \).

**Proof of Lemma 16.2** Define

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

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

\[
\begin{align*}
|DB)'(DB, Z)' - \kappa I_{p+1}| &= 0 \\
B'^t(D, Z)'(D, Z)B^+ - \kappa I_{p+1}| &= 0 \\
|(C\bar{Y}, Z)'CC'(C\bar{Y}, Z) - \kappa I_{p+1}| &= 0, \quad \text{or,} \\
|(\bar{Y}, Z^*)'(\bar{Y}, Z^*) - \kappa I_{p+1}| &= 0, \text{where } Z^* := C' Z \sim N(0^k, I_k),
\end{align*}
\]

(24.2)

the equivalence of the first and second lines holds because \(|A_1A_2| = |A_1| \cdot |A_2|\), \(|B^{+t}| = 1\), and \(B^tB^+ = I_{p+1}\), the equivalence of the second and third lines holds by matrix algebra, the equivalence of the third and fourth lines holds because \(DB = C\bar{Y}B'B = C\bar{Y}\) 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 \(\lambda_{min}((\bar{Y}, Z^*)'(\bar{Y}, Z^*))\). In addition, \(Z'Z = Z''Z^*\). Hence \(\overline{\text{64}}\)

\[
CLR_{k,p}(D) = Z'Z - \lambda_{min}((D, Z)'(D, Z)) = Z''Z^* - \lambda_{min}((\bar{Y}, Z^*)'(\bar{Y}, Z^*)). \tag{24.3}
\]

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

Lemma 5.1 of AG2. The statistics \(QLR_n, c_{k,p}(n^{1/2}\tilde{D}_n, 1 - \alpha), \tilde{D}_n, AR_n, \tilde{\Sigma}_n, \) and \(\tilde{\Lambda}_n\) are invariant to the transformation \((g_i, G_i) \sim (Mg_i, MG_i) \forall i \leq n\) for any \(k \times k\) nonsingular matrix \(M\). This transformation induces the following transformations: \(\tilde{g}_n \sim M\tilde{g}_n, \tilde{G}_n \sim M\tilde{G}_n, \Omega_n \sim M\Omega_nM', \tilde{\Gamma}_jn \sim M\tilde{\Gamma}_jnM' \forall j \leq p, \tilde{D}_n \sim M\tilde{D}_n, \tilde{V}_n \sim (I_{p+1} \otimes M)\tilde{V}_n (I_{p+1} \otimes M'), \) and \(\tilde{R}_n \sim (I_{p+1} \otimes M)\tilde{R}_n (I_{p+1} \otimes M').\)

Proof of Lemma 5.1. We refer to the results of the Lemma for \(g_i, G_i, ..., \tilde{\Gamma}_jn\) as equivariance results. The equivariance results are immediate for \(g_i, G_i, \tilde{g}_n, \tilde{G}_n, \Omega_n, \) and \(\tilde{\Gamma}_jn\). For \(\tilde{D}_n = (\tilde{D}_{1n}, ..., \tilde{D}_{pn})\), we have

\[
\tilde{D}_jn := \tilde{\Gamma}_jn\tilde{\Omega}^{-1}_jn\tilde{g}_n \sim M\tilde{G}_jn - M\tilde{\Gamma}_jnM'(M\Omega_nM')^{-1}M\tilde{g}_n = M\tilde{D}_jn \tag{24.4}
\]

for \(j = 1, ..., p\). We have \(f_i := (g'_i, vec(G_i))' \sim ((Mg_i)', vec(MG_i))' = (I_{p+1} \otimes M)f_i\). Using this, we obtain \(\tilde{V}_n = n^{-1} \sum_{i=1}^n (f_i - \bar{f}_n)(f_i - \bar{f}_n)' \sim (I_{p+1} \otimes M)\tilde{V}_n (I_{p+1} \otimes M').\) Next, we have

\footnote{The quantity \(CLR_{k,p}(D)\) is written in terms of \((D, Z)\) in (24.3), whereas it is written in terms of \((Z, D)\) in (5.8). Both expressions give the same value.}
\[ \hat{R}_n := (B' \otimes I_k) \hat{V}_n (B \otimes M') \sim (B' \otimes M) \hat{V}_n (B \otimes M') = (I_{p+1} \otimes M) \hat{R}_n (I_{p+1} \otimes M') \] using the equivariance result for \( \hat{V}_n \). We have \( \hat{\Sigma}_{j_1n} := tr(\hat{R}'_{j_1n} \hat{\Omega}^{-1}_n) / k \sim tr((M \hat{R}'_{j_1n} M') (M \hat{\Omega}^{-1}_n M')^{-1}) / k = \widetilde{\hat{\Sigma}}_{j_1n} \) for \( j, \ell = 1, \ldots, p + 1 \) using the equivariance result for \( \hat{R}_n \). We have \( \hat{L}_n := (\theta, I_p)(\hat{\Sigma}^e_n)^{-1}(\theta, I_p)' \sim \hat{L}_n \) using the invariance result for \( \hat{\Sigma}_n \). We have \( \hat{D}'_{n} \hat{D}^*_{n} := \hat{L}_{1/2} \hat{D}'_{n} \hat{\Sigma}_{n}^{-1/2} \hat{D}^*_{n} \hat{L}_{1/2} = \hat{D}'_{n} \hat{D}^*_{n} \). This implies that \( c_{k,p}(n^{1/2} \hat{D}_{n}^*, 1 - \alpha) \sim c_{k,p}(n^{1/2} \hat{D}^*_{n}, 1 - \alpha) \) because \( c_{k,p}(n^{1/2} \hat{D}_n^*, 1 - \alpha) \) only depends on \( \hat{D}_n^* \) through \( \hat{D}'_{n} \hat{D}^*_{n} \) by the Comment to Lemma 16.2.

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

\[ QLR_n := AR_n - \lambda_{\min}\left(n \left( \hat{g}_n, \hat{D}_n \hat{L}_{n/2} \right)' \hat{\Sigma}_n^{-1} \left( \hat{g}_n, \hat{D}_n \hat{L}_{n/2} \right) \right) \sim AR_n - \lambda_{\min}\left(n \left( M \hat{g}_n, M \hat{D}_n \hat{L}_{n/2} \right)' (M \hat{\Omega}_n M')^{-1} \left( M \hat{g}_n, M \hat{D}_n \hat{L}_{n/2} \right) \right) = QLR_n, \quad (24.5) \]

using the invariance of \( AR_n \) and \( \hat{L}_n \) and the equivariance of the other statistics that appear. □

**Lemma 15.1.** The statistics \( QLR_Pn, c_{k,p}(n^{1/2} \hat{D}_n^*, 1 - \alpha) \), \( \hat{D}'_{n} \hat{D}^*_{n} \), \( AR_n \), \( \hat{u}^*_n, \hat{\Sigma}_n \), and \( \hat{\Sigma}_n \) are invariant to the transformation \(( Z_i, u^*_i) \sim (M Z_i, u^*_i) \) \( \forall i \) for any \( k \times k \) nonsingular matrix \( M \).

This transformation induces the following transformations: \( g_i \sim M g_i \) \( \forall i \leq n \), \( G_i \sim M G_i \) \( \forall i \leq n \), \( \hat{g}_n \sim M \hat{g}_n \), \( \hat{G}_n \sim M \hat{G}_n \), \( \hat{\Sigma}_n \sim M \hat{\Sigma}_n \). In addition, we have \( \hat{D}'_{n} \hat{D}^*_{n} := \hat{L}_{1/2} \hat{D}'_{n} \hat{\Omega}^{-1}_{n} \hat{D}^*_{n} \hat{L}_{1/2} = \hat{D}'_{n} \hat{D}^*_{n} \). This implies that \( c_{k,p}(n^{1/2} \hat{D}_n^*, 1 - \alpha) \sim c_{k,p}(n^{1/2} \hat{D}^*_{n}, 1 - \alpha) \) because \( c_{k,p}(n^{1/2} \hat{D}_n^*, 1 - \alpha) \) only depends on \( \hat{D}_n^* \) through \( \hat{D}'_{n} \hat{D}^*_{n} \) 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
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 \( n^{1/2}(\hat{g}_n, \text{vec}(\hat{D}_n - D_n)) \) given in Lemma 16.4 follows from the Lyapunov triangular-array multivariate CLT (using the moment restrictions

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

$$n^{1/2} \text{vec}(\widehat{D}_n - D_n) = n^{-1/2} \sum_{i=1}^{n} \text{vec}(G_i - D_n) - \left( \begin{array}{c} \widehat{\Gamma}_{1n} \\
\vdots \\
\widehat{\Gamma}_{pn} \end{array} \right) \Omega_n^{-1} n^{1/2} \widehat{g}_n$$

(25.3)

$$= n^{-1/2} \sum_{i=1}^{n} \left[ \text{vec}(G_i - D_n) - \begin{pmatrix} E_{F_n} G_{i\ell} g_{i\ell}' \\
\vdots \\
E_{F_n} G_{i\ell_0} g_{i\ell_0}' \end{pmatrix} \Omega_n^{-1} g_i \right] + o_p(1),$$

where the second equality holds by (i) the weak law of large numbers (WLLN) applied to $n^{-1} \sum_{\ell=1}^{n} G_{ij} g_{ij}'$ for $j = 1, \ldots, p$, $n^{-1} \sum_{\ell=1}^{n} \text{vec}(G_\ell)$, and $n^{-1} \sum_{\ell=1}^{n} g_{ij} g_{ij}'$, (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 $n^{1/2} \text{vec}(\widehat{D}_n - D_n)$ and $n^{1/2} \widehat{g}_n$ is a zero matrix because $E_{F_n} [G_{ij} - E_{F_n} G_{ij}] - (E_{F_n} G_{ij} g_{ij}') \Omega_n^{-1} g_i = 0^{k \times k}$, where $G_{ij}$ denotes the $j$th column of $G_i$. By the CLT, the limiting variance matrix of $n^{1/2} \text{vec}(\widehat{D}_n - D_n)$ equals $\lim \text{Var}_{F_n}(\text{vec}(G_i) - (E_{F_n} \text{vec}(G_\ell) g_{\ell}') \Omega_n^{-1} g_i) = \lim \Phi_{F_n}^{\text{vec}(G_i)} = \Phi_{h}^{\text{vec}(G_i)}$, see (16.20), and the limit exists because (i) the components of $\Phi_{F_n}^{\text{vec}(G_i)}$ are comprised of $\lambda_{4,F_n}$ and submatrices of $\lambda_{5,F_n}$ and (ii) $\lambda_{s,F_n} \rightarrow h_s$ for $s = 4, 5$. By the CLT, the limiting variance matrix of $n^{1/2} \widehat{g}_n$ equals $\lim E_{F_n} g_i g_i' = h_{5,g}$.

The asymptotic distribution of $n^{1/2} W_{F_n} \widehat{D}_n U_n \mathcal{T}_n$ 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

$$W_n \widehat{D}_n U_n B_{n,q} Y_{n,q}^{-1} = C_n Y_n B_{n,q} B_{n,q} Y_{n,q}^{-1} = C_n Y_n \begin{pmatrix} I_q \\
0_{(p-q) \times q} \end{pmatrix} Y_{n,q}^{-1} = C_n \begin{pmatrix} I_q \\
0_{(k-q) \times q} \end{pmatrix} = C_{n,q},$$

(25.4)

where the second equality uses $B_{n,q} B_{n,q} = I_p$. Hence, we obtain

$$W_n \widehat{D}_n U_n B_{n,q} Y_{n,q}^{-1} = W_n D_n U_n B_{n,q} Y_{n,q}^{-1} + W_n n^{1/2} (\widehat{D}_n - D_n) U_n B_{n,q} (n^{1/2} Y_{n,q})^{-1}$$

$$= C_{n,q} + o_p(1) \rightarrow_p h_{3,q} = \Delta_{h,q},$$

(25.5)

where the second equality uses (among other things) $n^{1/2} \tau_{j,F_n} \rightarrow \infty$ for all $j \leq q$ (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 $\Delta_{h,q}$ in (16.24).
Using the singular value decomposition of \( W_n D_n U_n \) again, we obtain: if \( k \geq p \),

\[
n^{1/2} W_n D_n U_n B_{n,p-q} = n^{1/2} C_n \Psi_n B'_n B_{n,p-q} = n^{1/2} C_n \Psi_n \begin{pmatrix} 0^{q \times (p-q)} \\ I_{p-q} \end{pmatrix}
\]

\[
= C_n \begin{pmatrix} n^{1/2} \Psi_n (p-q) \\ 0^{(k-p) \times (p-q)} \end{pmatrix} \rightarrow h_3 \begin{pmatrix} 0^{q \times (p-k)} \\ 0^{q \times (k-p)} \\ \text{Diag}\{h_{1,q+1}, \ldots, h_{1,p}\} \end{pmatrix} = h_3 h^\circ_{1,p-q}, \tag{25.6}
\]

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 \( h^\circ_{1,p-q} \) in (16.24) with \( k \geq p \). Analogously, if \( k < p \), we have

\[
n^{1/2} W_n D_n U_n B_{n,p-q} = n^{1/2} C_n \Psi_n \begin{pmatrix} 0^{q \times (p-q)} \\ I_{p-q} \end{pmatrix} = C_n \begin{pmatrix} 0^{q \times (k-q)} \\ 0^{q \times (p-k)} \\ \text{Diag}\{h_{1,q+1}, \ldots, h_{1,k}\} \end{pmatrix} = h_3 h^\circ_{1,p-q}, \tag{25.7}
\]

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

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

\[
n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q} = n^{1/2} W_n D_n U_n B_{n,p-q} + W_n n^{1/2}(\widehat{D}_n - D_n) U_n B_{n,p-q}
\]

\[
\rightarrow_d h_3 h^\circ_{1,p-q} + h_7 \overline{D}_h h_{81} h_{2,p-q} = \overline{\Lambda}_{h,p-q}, \tag{25.8}
\]

where \( B_{n,p-q} \rightarrow h_{2,p-q} \), \( W_n \rightarrow h_7 \), and \( U_n \rightarrow h_{81} \), and the last equality holds by the definition of \( \overline{\Lambda}_{h,p-q} \) in (16.24).

Equations (25.5) and (25.8) combine to establish

\[
n^{1/2} W_n \widehat{D}_n U_n T_n = n^{1/2} W_n \widehat{D}_n U_n B_n S_n = (W_n \widehat{D}_n U_n B_{n,q}^{-1} T_{n,q}) \cdot n^{1/2} W_n \widehat{D}_n U_n B_{n,p-q})
\]

\[
\rightarrow_d (\overline{\Lambda}_{h,q}, \overline{\Lambda}_{h,p-q}) = \overline{\Lambda}_h \tag{25.9}
\]

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

Parts (a) and (b) of the lemma hold by the definitions of \( (\overline{g}_h, \overline{D}_h) \) and \( \overline{\Lambda}_h \). The independence of \( (\overline{D}_h, \overline{\Lambda}_h) \) and \( \overline{g}_h \), stated in part (c) of the lemma, holds by the independence of \( \overline{g}_h \) and \( D_h \) (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 non-empty parameter space $\Lambda_* \subset \Lambda_{WU}$. Under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_*$,

(a) $\tilde{\kappa}_{jn} \rightarrow_p \infty$ for all $j \leq q$,

(b) the (ordered) vector of the smallest $p-q$ eigenvalues of $n\tilde{U}'_n \tilde{D}'_n \tilde{W}_n \tilde{D}_n \tilde{U}_n$, i.e., $(\tilde{\kappa}_{(q+1)n}, \ldots, \tilde{\kappa}_{pn})'$, converges in distribution to the (ordered) $p-q$ vector of the eigenvalues of $\Delta'_h p-q h_{3,k-q} h_{3,k-q}^t \times \Delta_{h,p-q} \in R^{(p-q) \times (p-q)}$,

(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 $k < p$:

$$
\begin{bmatrix}
  h_{\tilde{r}_1} + o(1) & 0 \times (k-r_1^*p) & 0 \times (p-k) \\
  0 \times (k-r_1^*p) & O(\tau_{r_2 F_n} / \tau_{r_1 F_n}) \times (k-r_1^*p) & 0 \times (p-k) \\
  0 \times (k-r_1^*p) & 0 \times (p-k) & 0 \times (p-k)
\end{bmatrix} \in R^{k \times p},
$$

(25.10)

where $r_1^*$ 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 $k \times (p - r_{g-1}^*)$ matrix that contains three submatrices needs to be replaced by the following matrix when $k < p$:

$$
\begin{bmatrix}
  0 \times (k-r_{g-1}^*) & 0 \times (p-k) \\
  Diag\{\tau_{r_{g-1} F_n}, \ldots, \tau_{k F_n}\} / \tau_{r_{g-1} F_n} & 0 \times (k-r_{g-1}^*) \times (p-k)
\end{bmatrix} \in R^{k \times (p-r_{g-1}^*)},
$$

(25.11)

The limit of this matrix as $n \rightarrow \infty$ 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 \( k \times (p - r_g^2) \) matrix) is the same as the matrix in the first line of (17.22) in the SM to AG1, but with \( r_g^2 \) in place of \( r_{g-1}^2 \) (using \( r_{g+1} = r_g^2 + 1 \) and \( r_g = r_{g-1}^2 + 1 \)). 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 \( r_g^2 \) in place of \( r_{g-1}^2 \).

No other changes are needed. \( \square \)

26 Proof of Theorem 16.6

**Theorem 16.6.** Suppose Assumption WU holds for some non-empty parameter space \( \Lambda_s \subset \Lambda_W \). Under all sequences \( \{\lambda_{n,h} : n \geq 1\} \) with \( \lambda_{n,h} \in \Lambda_s \),

\[
QLR_{WU,n} \to_d \Phi_n h_{5,g}^{-1}g_h - \lambda_{\min}(\left(\sum_{h,p-q} h_{5,g}^{-1/2}g_h h_{3,k-q} h'_{3,k-q} (\sum_{h,p-q} h_{5,g}^{-1/2}g_h)\right))
\]

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 \( q \leq \min\{k, p\} \)), \( \sum_{h,p-q} \) does not appear in the limit random variable and the limit random variable reduces to \( (h_{5,g}^{-1/2}g_h) h_{3,p} h'_{3,p} h_{5,g}^{-1/2}g_h \sim \chi^2_p \). 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 \( \Phi_n h_{5,g}^{-1}g_h \sim \chi^2_k \). When \( k \leq p \) and \( q < k \), the \( \lambda_{\min}(\cdot) \) expression equals zero and the limit random variable reduces to \( \Phi_n h_{5,g}^{-1}g_h \sim \chi^2_k \). Under all subsequences \( \{w_n\} \) and all sequences \( \{\lambda_{w_n,h} : n \geq 1\} \) with \( \lambda_{w_n,h} \in \Lambda_s \), 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:

\[
\hat{D}_n^+ := (\hat{D}_n, \hat{W}_n, \hat{\Omega}_n^{-1/2} g_n) \in R^{k \times (p+1)} \quad \hat{U}_n^+ := \begin{bmatrix} \hat{U}_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
\]

\[
U_n^+ := \begin{bmatrix} U_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)} \quad h_{81}^+ := \begin{bmatrix} h_{81} & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
\]

\[
B_n^+ := \begin{bmatrix} B_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
\]

\[
B_n^+ = (B_{n,q}^+, B_{n,p+1-q}^+) \text{ for } B_{n,q}^+ \in R^{(p+1) \times q} \text{ and } B_{n,p+1-q}^+ \in R^{(p+1) \times (p+1-q)},
\]

\[
D_n^+ := (D_n, 0^k) \in R^{k \times (p+1)}, \quad \Upsilon_n^+ := (\Upsilon_n, 0^k) \in R^{k \times (p+1)},
\]

\[
S_n^+ := \text{Diag}\{ (n^{1/2} \tau_{1,F_n})^{-1}, \ldots, (n^{1/2} \tau_{q,F_n})^{-1}, 1, \ldots, 1 \} = \begin{bmatrix} S_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
\]

where \( \hat{g}_n \) and \( \hat{\Omega}_n \) are defined in (4.1) with \( \theta = \theta_0 \), \( \hat{D}_n \) is defined in (5.2) with \( \theta = \theta_0 \), \( \hat{W}_n, \hat{U}_n, U_n \) (: = \( U_{F_n} \)), 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), \( \Upsilon_n \) is defined in (25.2), and \( S_n \) is defined in (16.23).

Let \( \hat{\kappa}^+_j \) denote the \( j \)th eigenvalue of \( n\hat{U}_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+\hat{U}_n^+ \), \( \forall j = 1, \ldots, p+1 \), (26.2) ordered to be nonincreasing in \( j \). We have\(^65\)

\[
\hat{W}_n \hat{D}_n^+ \hat{U}_n^+ = (\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} g_n) \quad \text{and}
\]

\[
\lambda_{\min}(n(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} g_n)(\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} g_n)) = \lambda_{\min}(n\hat{U}_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+\hat{U}_n^+) = \hat{\kappa}^+_j \in (p+1)n^\prime.
\]

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 non-empty parameter space \( \Lambda_* \subset \Lambda_{WU} \). 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) \( \hat{\kappa}^+_j \rightarrow_p \infty \) for \( j = 1, \ldots, g \) and (b) \( \hat{\kappa}^+_j = o_p((n^{1/2} \tau_{F_n})^2) \) 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

\(^65\) In (26.3), we write \( (\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} g_n) \), whereas we write its analogue \( (\hat{\Omega}_n^{-1/2} g_n, \hat{D}_n^*) \) in (5.7) with its columns in the reverse order. Both ways give the same value for the minimum eigenvalue of the inner product of the matrix with itself, which is the statistic of interest. We use the order \( (\hat{\Omega}_n^{-1/2} g_n, \hat{D}_n^*) \) in AG2 because it is consistent with the order in Moreira (2003) and Andrews, Moreira, and Stock (2006). We use the order \( (\hat{W}_n \hat{D}_n \hat{U}_n, \hat{\Omega}_n^{-1/2} g_n) \) 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.
holds with \( n \) replaced with \( w_n \).

**Proof of Theorem 16.6.** We have \( n^{1/2}g_n \to_d \overline{g}_h \) (by Lemma 16.4) and \( \hat{\Omega}_n^{-1/2} \to_p h_{5,g}^{-1/2} \) (because \( \hat{\Omega}_n - \Omega_{F_n} \to_p k \times k \) by the WLLN, \( \Omega_{F_n} \to h_{5,g} \), and \( h_{5,g} \) is pd). In consequence, \( AR_n \to_d \overline{g}_h h_{5,g}^{-1} \overline{g}_h \).

Given this, the definition of \( QLR_n \) in (16.3), and (26.3), to prove the convergence result in Theorem 16.6 it suffices to show that

\[
\lambda_{\min}(nU_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+\hat{U}_n^+ - \kappa I_{p+1}) \to_d \lambda_{\min}(\Delta_{h,p-q}, h_{5,g}^{-1/2}g_h)h_{3,k-q}h_{3,k-q}'(\Delta_{h,p-q}, h_{5,g}^{-1/2}g_h)).
\] (26.4)

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

\[
|Q_n^+(\kappa)| = 0, \quad Q_n^+(\kappa) := nS_nB_n^+U_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+U_n^+B_n^+S_n - \kappa S_nB_n^+U_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+U_n^+ - \kappa S_nB_n^+U_n^+\hat{D}_n^+\hat{W}_n^+\hat{D}_n^+U_n^+,
\]

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

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

\[
B_n^+S_n^+ = (B_{n,q}^+S_{n,q}, B_{n,p+1-q}^+), \quad \text{where} \quad S_{n,q} := \text{Diag}\{(n^{1/2}r_{1F_n})^{-1}, \ldots, (n^{1/2}r_{qF_n})^{-1}\} \in \mathbb{R}^{q \times q}. \] (26.6)

The convergence result of Lemma 16.4 for \( n^{1/2}W_n\hat{D}_nU_nT_n \) \( (= n^{1/2}W_n\hat{D}_nU_nB_nS_n) \) can be written as

\[
n^{1/2}W_n\hat{D}_nU_nB_{n,q}S_{n,q} \to_p \Delta_{h,q} := h_{3,q} \quad \text{and} \quad n^{1/2}W_n\hat{D}_nU_nB_{n,p+1-q} \to_d \Delta_{h,p-q}. \] (26.7)

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

We have

\[
\hat{W}_nW_n^{-1} \to_p I_k \quad \text{and} \quad \hat{U}_n^+(U_n^+)^{-1} \to_p I_{p+1}, \] (26.8)

70
because \( \tilde{W}_n \to_p h_{71} := \lim W_n \) (by Assumption WU(a) and (c)), \( \tilde{U}_n^+ \to_p h_{81}^+ := \lim U_n^+ \) (by Assumption WU(b) and (c)), and \( h_{71} \) and \( h_{81}^+ \) are pd (by the conditions in \( \mathcal{F}_{WU} \)).

By (26.5)-(26.8), we have

\[
Q_n^+(\kappa) = \begin{bmatrix}
I_q + o_p(1) & h_{3,q}^t n^{1/2} W_n \hat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1) \\
n^{1/2} B_{n,p+1-q}^t U_n^+ \hat{D}_n^+ W_n h_{3,q} + o_p(1) & n^{1/2} B_{n,p+1-q}^t U_n^+ \hat{D}_n^+ W_n n^{1/2} \hat{D}_n^+ U_n^+ B_{n,p+1-q}^+ + o_p(1)
\end{bmatrix}
\]

\[
-\kappa \begin{bmatrix}
S_{n,q}^2 & 0^q \times (p+1-q) \\
0^{p+1-q} \times q & I_{p+1-q}
\end{bmatrix} - \kappa \begin{bmatrix}
S_{n,q} A_{1n}^t S_{n,q} & S_{n,q} A_{2n}^t \\
A_{2n}^t S_{n,q} & A_{3n}
\end{bmatrix}, \text{ where }
\]

(26.9)

for \( A_{1n}^+ \in \mathbb{R}^{q \times q} \), \( A_{2n}^+ \in \mathbb{R}^{q \times (p+1-q)} \), and \( A_{3n}^+ \in \mathbb{R}^{(p+1-q) \times (p+1-q)} \), and the first equality uses \( \Delta_{h,q} := h_{3,q}^t h_{3,q} = h_{3,q}^t h_{3,q} = \lim C'_{n,q} C_{n,q} = I_q \) (by (16.14), (16.16), (16.19), and (16.24)). Note that \( A_{jn}^+ \) and \( \tilde{A}_{jn}^+ \) (defined in (26.19) below) are not the same in general for \( j = 1, 2, 3 \) because their dimensions differ. For example, \( A_{1n}^+ \in \mathbb{R}^{q \times q} \), whereas \( \tilde{A}_{1n}^+ \in \mathbb{R}^{q \times \tilde{r}_q} \), where \( \tilde{r}_q \) 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

\[
n B_n^t \tilde{U}_n^t \hat{D}_n^+ W_n \tilde{W}_n \hat{D}_n^+ \tilde{U}_n^+ B_n^+ \\
= n B_n^t \left((U_n^+)^{-1} \tilde{U}_n^+\right)' \left(B_n^+\right)^{-1} B_n^t \tilde{U}_n^t \hat{D}_n^+ W_n \tilde{W}_n W_n^{-1}^{-1} B_n^+ \\
\times \left(W_n W_n^{-1}\right) \left(W_n \hat{D}_n^+ U_n^+ B_n^+\right) \left(B_n^+\right)^{-1} \left((U_n^+)^{-1} \tilde{U}_n^+\right) B_n^+ \\
\to_d \left(\Delta_{h,p-q}, h_{5,q}^{-1/2} g_h\right)' \left(\Delta_{h,p-q}, h_{5,q}^{-1/2} g_h\right),
\]

(26.10)

where the convergence holds by (26.7) and (26.8) and \( \Delta_{h,p-q} \) 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, Thm. 3.1, pp. 37–38)). Hence, the smallest eigenvalue of \( n B_n^t \tilde{U}_n^t \hat{D}_n^+ W_n \tilde{W}_n \hat{D}_n^+ \tilde{U}_n^+ B_n^+ \) converges in distribution to the smallest eigenvalue of \( \left(\Delta_{h,p-q}, h_{5,q}^{-1/2} g_h\right)' h_{3,k-q} h_{3,k-q} \left(\Delta_{h,p-q}, h_{5,q}^{-1/2} g_h\right) \) (using \( h_{3,k-q} h_{3,k-q} = h_{3,k-q} I_k \) when \( q = 0 \)), which proves (26.4) when \( q = 0 \).
\[ |Q^+_n(\kappa)| = |Q^+_{1n}(\kappa)| \cdot |Q^+_{2n}(\kappa)|, \]

where

\[ Q^+_{1n}(\kappa) := I_q + o_p(1) - \kappa S_{n,q}^2 - \kappa S_{n,q}A^+_{1n}S_{n,q}, \]

\[ Q^+_{2n}(\kappa) := n^{1/2}B_{n,p+1-q}^+U_n^+\tilde{D}_n^+W_n^+I_n^{1/2}\tilde{D}_n^+B_{n,p+1-q}^+ + o_p(1) - \kappa I_{p+1-q} - \kappa A^+_{3n} \]

\[ -[n^{1/2}B_{n,p+1-q}^+U_n^+\tilde{D}_n^+W_n^+h_{3,q} + o_p(1)](I_q + o_p(1))[h_{3,q}n^{1/2}W_n^+I_n^{1/2}\tilde{D}_n^+B_{n,p+1-q}^+ + o_p(1)] \]

\[ \times [h_{3,q}n^{1/2}W_n^+\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1) - \kappa S_{n,q}A^+_{2n}], \]  

(26.11)

none of the \( o_p(1) \) terms depend on \( \kappa \), 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 \( j = q + 1, ..., p + 1 \), and \( A^+_{1n} = o_p(1) \) (by 26.9), we have \( \tilde{\kappa}_{jn}^+ S_{n,q} = o_p(1) \) and \( \tilde{\kappa}_{jn}^+ S_{n,q}A^+_{1n}S_{n,q} = o_p(1) \). Thus, for \( j = q + 1, ..., p + 1 \),

\[ Q^+_{1n}(\tilde{\kappa}_{jn}^+) = I_q + o_p(1) - \tilde{\kappa}_{jn}^+ S_{n,q}^2 - \tilde{\kappa}_{jn}^+ S_{n,q}A^+_{1n}S_{n,q} = I_q + o_p(1). \]  

(26.12)

By (26.5) and (26.11), \( |Q^+_{n}(\tilde{\kappa}_{jn}^+)| = |Q^+_{1n}(\tilde{\kappa}_{jn}^+)| \cdot |Q^+_{2n}(\tilde{\kappa}_{jn}^+)| = 0 \) for \( j = 1, ..., p + 1 \). By (26.12), \( |Q^+_{1n}(\tilde{\kappa}_{jn}^+)| \neq 0 \) for \( j = q + 1, ..., p + 1 \) \( \text{wp-}1 \). Hence, \( \text{wp-}1 \),

\[ |Q^+_{2n}(\tilde{\kappa}_{jn}^+)| = 0 \] for \( j = q + 1, ..., p + 1 \).  

(26.13)

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

\[ Q^+_{2n}(\tilde{\kappa}_{jn}^+) = nB_{n,p+1-q}^+U_n^+\tilde{D}_n^+W_n^+I_n^{1/2}\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1) \]

\[ -[n^{1/2}B_{n,p+1-q}^+U_n^+\tilde{D}_n^+W_n^+h_{3,q} + o_p(1)](I_q + o_p(1))[h_{3,q}n^{1/2}W_n^+I_n^{1/2}\tilde{D}_n^+B_{n,p+1-q}^+ + o_p(1)] \]

\[ \times [h_{3,q}n^{1/2}W_n^+\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1)](I_q + o_p(1))[h_{3,q}n^{1/2}W_n^+I_n^{1/2}\tilde{D}_n^+B_{n,p+1-q}^+ + o_p(1)] \]

\[ \times [h_{3,q}n^{1/2}W_n^+\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1)] \]

\[ \times [h_{3,q}n^{1/2}W_n^+\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1)] \]

\[ \times [h_{3,q}n^{1/2}W_n^+\tilde{D}_n^+U_n^+B_{n,p+1-q}^+ + o_p(1)] \]

(26.14)

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

\[ I_{p+1-q} + o_p(1), \]

(26.15)

because \( A^+_{3n} = o_p(1) \) (by 26.9), \( n^{1/2}W_n^+I_n^{1/2}\tilde{D}_n^+B_{n,p+1-q}^+ = O_p(1) \) (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 \( h_{1,j} := \lim n^{1/2}\tau_{j,F_n} \), \( A^+_{2n} = o_p(1) \) (by 26.9), and \( \tilde{\kappa}_{jn}^+ A^+_{2n}S_{n,q}(I_q + o_p(1))S_{n,q}A^+_{2n} = A^+_{2n}\tilde{\kappa}_{jn}^+ S_{n,q}^2A^+_{2n} + A^+_{2n}\tilde{\kappa}_{jn}^+ S_{n,q}o_p(1)S_{n,q}A^+_{2n} = o_p(1) \)

72
(using $\tilde{\kappa}_{jn}^+ S_{n,q}^2 = o_p(1)$ and $A_{2n}^+ = o_p(1)$).

Equations (26.14) and (26.15) give

$$Q_{2n}^+ (\tilde{\kappa}_{jn}^+)$$

$$= n^{1/2} B_{n,p+1\rightarrow q} U_{n}^T \tilde{D}_{n,q}^+ W_n[U_{n} - h_{3,q} h_{3,q}']n^{1/2}W_n \tilde{D}_{n,q}^+ U_{n}^T B_{n,p+1\rightarrow q} + o_p(1) - \tilde{\kappa}_{jn}^+[I_{p+1-q} + o_p(1)]$$

$$= n^{1/2} B_{n,p+1\rightarrow q} U_{n}^T \tilde{D}_{n,q}^+ W_n h_{3,k-q} h_{3,k-q}' n^{1/2}W_n \tilde{D}_{n,q}^+ U_{n}^T B_{n,p+1\rightarrow q} + o_p(1) - \tilde{\kappa}_{jn}^+[I_{p+1-q} + o_p(1)]$$

$$:= M_{n,p+1\rightarrow q}^+ - \tilde{\kappa}_{jn}^+[I_{p+1-q} + o_p(1)],$$

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

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

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),

$$M_{n,p+1\rightarrow q}^+ \rightarrow_d (\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h') h_{3,k-q} h_{3,k-q}' (\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h).$$

The vector of (ordered) eigenvalues of a matrix is a continuous function of the matrix (by Elsner’s Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). By (26.18), the matrix $M_{n,p+1\rightarrow q}^+$ converges in distribution. In consequence, by the CMT, the vector of eigenvalues of $M_{n,p+1\rightarrow q}^+$, viz., \{\tilde{\kappa}_{jn}^+ : j = q + 1, \ldots, p + 1\}, converges in distribution to the vector of eigenvalues of the limit matrix $(\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h') h_{3,k-q} h_{3,k-q}' (\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h)$. Hence, $\lambda_{\min}(nU_{n}^T \tilde{D}_{n,q}^+ W_n \tilde{D}_{n,q}^+ U_{n})$ converges in distribution to the smallest eigenvalue of $(\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h') h_{3,k-q} h_{3,k-q}' (\Delta_{h,p-q, h_{5,g}}^{-1/2} \tilde{g}_h)$, 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 $n(\tilde{W}_n \tilde{D}_n \tilde{U}_n, \tilde{\Omega}_n^{-1/2} \tilde{g}_n)'(\tilde{W}_n \tilde{D}_n \tilde{U}_n, \tilde{\Omega}_n^{-1/2} \tilde{g}_n)$. 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 $n^{1/2}W_n \tilde{W}_n^{-1} \tilde{D}_n^{-1/2} \tilde{g}_n$ and $h_{5,g}^{-1/2} \tilde{g}_h$, respectively, i.e., $n^{1/2}W_n \tilde{D}_n U_n B_{n,p-q}$ and $\Delta_{h,p-q}$ drop out (because $U_{n}^T B_{n,p+1\rightarrow q}^+ = (0^p, 1)'$ in this case). In consequence, the limit in (26.18) becomes $(h_{5,g}^{-1/2} \tilde{g}_h') h_{3,k-q} h_{3,k-q}' h_{5,g}^{-1/2} \tilde{g}_h$, which has a $\chi^2_{k-p}$ distribution (because $h_{5,g}^{-1/2} \tilde{g}_h \sim N(0^k, I_k)$, $h_3 = (h_{3,q}, h_{3,k-q}) \in R^{k \times k}$ 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 \( n^{1/2}W_n\bar{D}_nU_nT_n \), 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 \( I_k - h_{3,q}h_{3,q}^t \) equals zero because the quantities on which they depend differ between the two proofs. Similarly, the quantities \( \tau_{(p+1)} F_n \), \( \tau_p F_n = 0 \) (using 0/0 := 0), and with \( \bar{D}_n, \tilde{U}_n, B_n, \bar{k}_jn, \tilde{\lambda}_n, D_n, U_n, h_81, \chi_n, B_n, r_{i} \), and \( B_n, r_{i} \) replaced by \( \tilde{\lambda}_n, U_n, B_n, \bar{k}_jn, \tilde{\lambda}_n, D_n, U_n, h_81, \chi_n, B_n, r_{i} \), respectively, where

\[
\tilde{\lambda}_n = \begin{bmatrix} \tilde{\lambda}_{1n}^+ & \tilde{\lambda}_{2n}^+ \\ \tilde{\lambda}_{2n}^+ & \tilde{\lambda}_{3n}^+ \end{bmatrix} := (B_n^+)'(U_n^+)'(\tilde{U}_n^+)^{-1}(\tilde{U}_n^+)^{-1}U_n^+D_n^+ - I_{p+1}, \tag{26.19}
\]

where \( \tilde{\lambda}_{1n}^+ \in R^{(p+1) \times r_{i}}, \tilde{\lambda}_{2n}^+ \in R^{r_{i} \times (p+1 - r_{i})}, \tilde{\lambda}_{3n}^+ \in R^{(p+1 - r_{i}) \times (p+1 - r_{i})} \), and \( r_{i} \) is defined as in the proof of Lemma 16.1 in the SM to AG1. Note that the quantities \( \tilde{\lambda}_{tn} \) for \( \ell = 1, 2, 3 \), which depend on \( \tilde{\lambda}_n \) (see (17.2) in the SM to AG1), differ between the two proofs (because \( \tilde{\lambda}_n \) differs from \( \tilde{\lambda}_n^+ \)). Similarly, the quantities \( \tilde{\lambda}_n \) (defined in (17.8) in the SM to AG1), \( \tilde{\lambda}_{2n}(\kappa) \) for \( \ell = 1, 2, 3 \) (defined in (17.9) in the SM to AG1), and \( \tilde{\lambda}_{2n} \) (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: \( \{\tau_j F_n : j \leq p\}, q, \{h_{6,j} : j \leq p-1\}, G_h, \{r_j : j \leq G_h\}, \{r_{i}^\circ : j \leq G_h\}, h_{6, r_{i}^\circ}, \bar{W}_n, h_{71}, C_n, \) and \( h_3 \). Note that the first \( p \) singular values of \( W_nD_nU_n \) (i.e., \( \{\tau_j F_n : j \leq p\} \)) 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 \( \kappa_{jF_n} \) is the \( j \)th eigenvalue of \( W_nD_nU_nU_n^D_n^{}W_n^{} \), \( W_nD_n^+U_n^+ = W_n(D_n, 0^{\circ})U_n^+ = (W_nD_nU_n^{}, 0^{\circ}) \), and hence, \( W_nD_n^+U_n^+U_n^D_n^{}W_n^{} = W_nD_nU_nU_n^D_n^{}W_n^{} \).

The second equality in (17.3) in the SM to AG1, which states that \( W_nD_n^{}U_n^{}B_n = C_n \chi_n \), is a key equality in the proof of Lemma 16.1 in the SM to AG1. The analogue in the proof of the
current lemma is
\[ W_n D_n^+ U_n^+ B_n^+ = (W_n D_n, 0^k) \begin{bmatrix} U_n B_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} = (W_n D_n U_n B_n, 0^k) = (C_n \gamma_n, 0^k) = C_n \gamma_n^+. \] (26.20)

Hence, this part of the proof goes through when \( D_n, U_n, B_n, \) and \( \gamma_n \) are replaced by \( D_n^+, U_n^+, B_n^+, \) and \( \gamma_n^+ \), 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 CS’s. 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 \( (\hat{\Omega}_n^{-1/2}, \hat{L}_n^{1/2}) \) replaced by any matrices \( (\tilde{W}_n, \tilde{U}_n) \) that satisfy Assumption WU for some parameter space \( \Lambda_* \subset \Lambda_{WU} \) (stated in Section 16.5). Then, we show that Assumption WU holds for the parameter spaces \( \Lambda_{WU} \) and \( \Lambda_{WU,P} \) 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\(_{WU}\) test statistic for testing \( H_0 : \theta = \theta_0 \) is defined in (16.3) as
\[
QLR_{WU,n} := AR_n - \lambda_{\min}(n\tilde{Q}_{WU,n}), \quad \text{where} \quad \tilde{Q}_{WU,n} := \left(\tilde{W}_n \tilde{D}_n \tilde{U}_n, \tilde{\Omega}_n^{-1/2} \tilde{g}_n\right) \left(\tilde{W}_n \tilde{D}_n \tilde{U}_n, \tilde{\Omega}_n^{-1/2} \tilde{g}_n\right),
\] (27.1)

\( AR_n \) is defined in (5.2), and the dependence of \( QLR_n, \tilde{Q}_{WU,n}, \tilde{W}_n, \tilde{D}_n, \tilde{U}_n, \tilde{\Omega}_n, \) and \( \tilde{g}_n \) on \( \theta_0 \) is suppressed for notational simplicity.

The general CQLR\(_{WU}\) test rejects the null hypothesis if
\[
QLR_{WU,n} > c_{k,p}(n^{1/2} \tilde{W}_n \tilde{D}_n \tilde{U}_n, 1 - \alpha),
\] (27.2)

where \( c_{k,p}(D, 1 - \alpha) \) is defined just below (5.8).
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 non-empty parameter space \( \Lambda_s \subset \Lambda_{WU} \). Then, the asymptotic null rejection probabilities of the nominal size \( \alpha \) CQLR\(_{WU} \) test based on \( (\hat{W}_n, \hat{U}_n) \) equal \( \alpha \) under all subsequences \( \{w_n\} \) and all sequences \( \{\lambda_{w_n,h} : n \geq 1\} \) with \( \lambda_{w_n,h} \in \Lambda_s \).

**Comments:** (i) Theorem 27.1 and Proposition 16.3 imply that any nominal size \( \alpha \) CQLR test based on matrices \( (\hat{W}_n, \hat{U}_n) \) that satisfy Assumption WU for some parameter space \( \Lambda_s \) has correct asymptotic size \( \alpha \) and is asymptotically similar (in a uniform sense) for the parameter space \( \Lambda_s \).

(ii) In Lemma 27.4 below, we show that the choice of matrices \( (\hat{W}_n, \hat{U}_n) \) for the CQLR and CQLR\(_P \) tests (defined in Sections 3 and 15, respectively) satisfy Assumption WU for the parameter spaces \( \Lambda_{WU} \) and \( \Lambda_{WU,P} \) (defined in (16.17)), respectively. In addition, Lemma 27.4 shows that \( \mathcal{F} \subset \mathcal{F}_{WU} \) and \( \mathcal{F}_P \subset \mathcal{F}_{WU} \) when \( \delta_1 \) and \( M_1 \) that appear in the definition of \( \mathcal{F}_{WU} \) are sufficiently small and large, respectively.\(^{66}\) In consequence, the CQLR and CQLR\(_P \) tests have correct asymptotic size \( \alpha \) 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 \( \{D^c_n : n \geq 1\} \) be a sequence of constant (i.e., nonrandom) \( k \times p \) matrices. Here, we determine the limit as \( n \to \infty \) of \( c_{k,p}(D^c_n, 1 - \alpha) \) under certain assumptions on the singular values of \( D^c_n \).

**Lemma 27.2** Suppose \( \{D^c_n : n \geq 1\} \) is a sequence of constant (i.e., nonrandom) \( k \times p \) matrices with singular values \( \{\tau^c_{jn} \geq 0 : j \leq \min\{k, p\}\} \) for \( n \geq 1 \) that satisfy (i) \( \{\tau^c_{jn} \geq 0 : j \leq \min\{k, p\}\} \) are nonincreasing in \( j \) for \( n \geq 1 \), (ii) \( \tau^c_{jn} \to \infty \) for \( j \leq q \) for some \( 0 \leq q \leq \min\{k, p\} \) and (iii)\(^{66}\) Note that the set of distributions \( \mathcal{F}_{WU} \) 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 \( \mathcal{F}_{WU} \) differs for the CQLR and CQLR\(_2 \) tests.
\[ \tau_{jn}^c \to \tau_{j\infty}^c < \infty \text{ for } j = q + 1, \ldots, \min\{k,p\}. \] Then,

\[ c_{k,p}(D_n^c, 1-\alpha) \to c_{k,p,q}(\tau_{\infty}^c, 1-\alpha), \text{ where } \tau_{\infty}^c := (\tau_{(q+1)\infty}^c, \ldots, \tau_{\min\{k,p\}\infty}^c)' \in \mathbb{R}^{\min\{k,p\}-q}, \]

\[ \Upsilon(\tau_{\infty}^c) := \begin{pmatrix} \text{Diag}\{\tau_{\infty}^c\} \\ (0^{(k-p)\times(p-q)}) \end{pmatrix} \in \mathbb{R}^{(k-q)\times(p-q)} \text{ if } k \geq p, \]

\[ \Upsilon(\tau_{\infty}^c) := \begin{pmatrix} \text{Diag}\{\tau_{\infty}^c\}, (0^{(q-k)\times(p-k)}) \end{pmatrix} \in \mathbb{R}^{(k-q)\times(p-q)} \text{ if } k < p, \]

\[ c_{k,p,q}(\tau_{\infty}^c, 1-\alpha) \text{ denotes the } 1-\alpha \text{ quantile of } ACLR_{k,p,q}(\tau_{\infty}^c) := Z'Z - \lambda_{\min}(\Upsilon(\tau_{\infty}^c), Z_2)'(\Upsilon(\tau_{\infty}^c), Z_2), \text{ and } \]

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

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

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

(iii) If \( q = k \) (which requires that \( k \leq p \)), then \( \Upsilon(\tau_{\infty}^c) \) and \( Z_2 \) have no rows, the \( \lambda_{\min}(\cdot) \) expression in \( ACLR_{k,p,q}(\tau_{\infty}^c) \) disappears, \( ACLR_{k,p,q}(\tau_{\infty}^c) = Z_1'Z_1 \sim \chi^2_k \), and \( c_{k,p,q}(\tau_{\infty}^c, 1-\alpha) \) is the \( 1-\alpha \) quantile of the \( \chi^2_k \) distribution.

(iv) If \( k \leq p \) and \( q < k \), then \( \Upsilon(\tau_{\infty}^c), Z_2 \) has fewer rows \( (k-q) \) than columns \( (p-q+1) \) and, hence, the \( \lambda_{\min}(\cdot) \) expression in \( ACLR_{k,p,q}(\tau_{\infty}^c) \) equals zero, \( ACLR_{k,p,q}(\tau_{\infty}^c) = Z_1'Z_1 \sim \chi^2_k \), and \( c_{k,p,q}(\tau_{\infty}^c, 1-\alpha) \) is the \( 1-\alpha \) quantile of the \( \chi^2_k \) distribution.

(v) The distribution function (df) of \( ACLR_{k,p,q}(\tau_{\infty}^c) \) is shown in Lemma 27.3 below to be continuous and strictly increasing at its \( 1-\alpha \) quantile for all possible \( (k,p,q,\tau_{\infty}^c) \) values, which is required in the proof of Lemma 27.2.

The following lemma proves that the df of \( ACLR_{k,p,q}(\tau_{\infty}^c) \), 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 \( \tau_{\infty}^c \) and \( \Upsilon(\tau_{\infty}^c) \) 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 \( \min\{k,p\} - q \) \( (\geq 0) \) vectors \( \tau_{\infty}^c \) with non-negative elements in non-increasing order, the df of \( ACLR_{k,p,q}(\tau_{\infty}^c) := Z'Z - \lambda_{\min}(\Upsilon(\tau_{\infty}^c), Z_2)'(\Upsilon(\tau_{\infty}^c), Z_2) \) 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 \( Z_1 \in \mathbb{R}^q \) and \( Z_2 \in \mathbb{R}^{k-q} \).
the CQLR and CQLR<sub>P</sub> tests. Part (a) of the lemma shows that \( F_{WU} \), when defined for \((\widehat{W}_n, \widehat{U}_n)\) as in the CQLR test, contains \( F \) for suitable choices of the constants \( \delta_1 \) and \( M_1 \) that appear in the definition of \( F_{WU} \). Part (b) of the lemma shows that the parameter space \( F_{WU} \), when defined for \((\widehat{W}_n, \widehat{U}_n)\) as in the CQLR<sub>P</sub> test, contains the parameter space \( F_P \) for suitable constants \( \delta_1 \) and \( M_1 \).

**Lemma 27.4** (a) Suppose \((\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widehat{\Omega}_n^{1/2})\), where \( \widehat{\Omega}_n = \widehat{\Omega}_n(\theta_0) \) and \( \widehat{L}_n = \widehat{L}_n(\theta_0) \) are defined in \((4.1)\) and \((5.7)\). Then, (i) Assumption WU holds for the parameter space \( \Lambda_{WU} \) with \((\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widehat{R}_n))\) for \( \widehat{R}_n \) defined in \((5.3)\), \( W_1(W_2) = W_2^{-1/2} \) for \( W_2 \in \mathbb{R}^{k \times k} \), \( U_1(U_2F) = ((\theta_0, I_p)(\Sigma^e(\Omega_F, R_F)))^{-1}(\theta_0, I_p)'/2 \) for \( U_2F = (\Omega_F, R_F) \), \( h_7 = \lim W_{2Fw_n} := \lim \Omega_{Fw_n} \), and \( h_8 = \lim U_{2Fw_n} := \lim(\Omega_{Fw_n}, R_{Fw_n}) \), where \( \Omega_F := E_Fg_i g'_i, R_F \) is defined in \((16.7)\), \( \Sigma(\Omega_F, R_F) \) is defined in \((16.8)\), and \( \Sigma^e(\Omega_F, R_F) \) is defined given \( \Sigma(\Omega_F, R_F) \) by \((5.6)\), and (ii) \( F = F_{WU} \) for \( \delta_1 \) sufficiently small and \( M_1 \) sufficiently large in the definition of \( F_{WU} \), where \( F \) is defined in \((16.1)\) and \( F_{WU} \) is defined in \((16.12)\).

(b) Suppose \( g_i(\theta) = u_i(\theta)Z_i \), as in \((15.1)\), and \((\widehat{W}_n, \widehat{U}_n) = (\widehat{\Omega}_n^{-1/2}, \widehat{L}_n^{1/2})\), where \( \widehat{\Omega}_n = \widehat{\Omega}_n(\theta_0) \) and \( \widehat{L}_n = \widehat{L}_n(\theta_0) \) are defined in \((4.1)\) and \((15.6)\), respectively. Then, (i) Assumption WU holds for the parameter space \( \Lambda_{WU,P} \) with \((\widehat{W}_{2n}, \widehat{U}_{2n}) = (\widehat{\Omega}_n, (\widehat{\Omega}_n, \widehat{R}_n))\) for \( \widehat{R}_n \) defined in \((15.5)\), \( W_1(W_2) = W_2^{-1/2} \) for \( W_2 \in \mathbb{R}^{k \times k} \), \( U_1(U_2F) = ((\theta_0, I_p)(\Sigma^e(\Omega_F, \widehat{R}_F)))^{-1}(\theta_0, I_p)'/2 \) for \( U_2F = (\Omega_F, \widehat{R}_F) \), \( h_7 = \lim W_{2Fw_n} := \lim \Omega_{Fw_n} \), and \( h_8 = \lim U_{2Fw_n} := \lim(\Omega_{Fw_n}, \widehat{R}_{Fw_n}) \), where \( \Omega_F := E_Fg_i g'_i, \Sigma_F := \Sigma(\Omega_F, \widehat{R}_F) \) is defined in \((16.11)\), \( \Sigma^e(\Omega_F, \widehat{R}_F) \) is defined given \( \Sigma(\Omega_F, \widehat{R}_F) \) by \((5.6)\), and \( \widehat{R}_F \) is defined in \((16.10)\), and (ii) \( F_P \subset F_{WU} \) for \( \delta_1 \) sufficiently small and \( M_1 \) sufficiently large in the definition of \( F_{WU} \), where \( F_P \) is defined in \((16.1)\) and \( F_{WU} \) is defined in \((16.12)\).

**Comment:** Theorem 27.1, Lemma 27.4 and Proposition 16.3 combine to prove the CQLR and CQLR<sub>P</sub> test results of Theorem 16.1, which state that the CQLR and CQLR<sub>P</sub> tests have correct asymptotic size and are asymptotically similar (in a uniform sense) for the parameter spaces \( F \) and \( F_P \), respectively. As stated at the beginning of this section, the proofs of the CQLR and CQLR<sub>P</sub> 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

\[ \mathcal{Z}_h = \begin{pmatrix} \mathcal{Z}_{h1} \\ \mathcal{Z}_{h2} \end{pmatrix} := \begin{pmatrix} h_{3,q}^{-1/2} g_h \\ h_{3,k-q}^{-1/2} g_h \end{pmatrix} = h_3^{1/2} g_h \sim N(0^k, I_k), \tag{27.3} \]

where \( \mathcal{Z}_{h1} \in R^q \) and \( \mathcal{Z}_{h2} \in R^{k-q} \) and the distributional result holds because \( g_h \sim N(0^k, h_{5,q}) \) (by (16.21)) and \( h_3^1 = \lim C_n C_n = I_k \). Note that \( \mathcal{Z}_h \) and \( (\mathcal{D}_h, \Sigma_h) \) are independent because \( g_h \) and \( (\mathcal{D}_h, \Sigma_h) \) are independent (by Lemma 16.4(c)).

By Theorem 16.6

\[ QLR_{WU,n} \rightarrow_d \mathcal{Z}_h h_{5,g}^{-1} g_h - \lambda_{\min}(\Sigma_{h,p-q}, h_{5,g}^{-1/2} g_h) h_{3,k-q}^{-1} h_{3,k-q}(\Sigma_{h,p-q}, h_{5,g}^{-1/2} g_h) = \mathcal{Z}' \mathcal{Z}_h - \lambda_{\min}(h_{3,k-q}^{-1} \Sigma_{h,p-q}, \Sigma_{h2})(h_{3,k-q}^{-1} \Sigma_{h,p-q}, \Sigma_{h2}) =: QLR_h, \tag{27.4} \]

where the equality uses \( h_3 h_3' = \lim C_n C_n = I_k \). When \( q = p \), the term \( \Sigma_{h,p-q} \) does not appear and \( QLR_h := \mathcal{Z}_h \mathcal{Z}_h - \mathcal{Z}' \mathcal{Z}_h = -\mathcal{Z}' \mathcal{Z}_h. \)

Let \( \{\tau_{jn} : j = \min\{k, p\}\} \) denote the min\(\{k, p\}\) singular values of \( n^{1/2} \hat{W} \hat{D} \hat{U} \) in nonincreasing order. They equal the vector of square roots of the first \(\min\{k, p\}\) eigenvalues of \( n \hat{U} n \hat{D} \hat{W} \hat{U} \hat{D} \hat{U} \) in nonincreasing order. Define

\[ \tau_n = (\tau_1^{[1]n}, \tau_2^{[2]n})' \in R^{\min\{k, p\}}, \text{ where} \]

\[ \tau_1^{[1]n} = (\tau_1, ..., \tau_{q_1})' \in R^q \text{ and } \tau_2^{[2]n} = (\tau_{q_2}, ..., \tau_{\min\{k, p\}n})' \in R^{\min\{k, p\}-q}. \tag{27.5} \]

By Proposition 16.5(a) and (b), \( \tau_{jn} \rightarrow_p \infty \) for \( j \leq q \) (or, equivalently \( \text{Diag}^{-1}\{\tau_{1n}\} \rightarrow_p 0^{q \times q} \)) and

\[ \tau_2^{[2]n} \rightarrow_d \tau_{[2]h}. \tag{27.6} \]

where \( \tau_j \rightarrow \tau_j^{1/2} \) for \( j \leq q \) and \( \tau_{[2]}h \) is the vector of square roots of the first \(\min\{k, p\}-q\) eigenvalues of \( \Sigma_{h,p-q} h_{3,k-q}^{-1} h_{3,k-q}^{-1} \Sigma_{h,p-q} \in R^{q-q} \times (p-q) \) in nonincreasing order. (When \( q = \min\{k, p\} \), no vector \( \tau_{[2]}h \) appears.) By an almost sure representation argument, e.g., see Pollard (1990, Thm. 9.4, p. 45), there exists a probability space, say \((\Omega_0, \mathcal{F}_0, P_0)\), and random variables \( (QLR_{0, n}, \tau_{0n}, QLR_{0,h}, \tau_{0,h}^{[2]n})' \) defined on it such that \((QLR_{0, n}, \tau_{0n})'\) has the same distribution as \((QLR_{WU,n}, \tau_{jn})'\) for all \( n \geq 1 \), \((QLR_{0,h}, \tau_{0,h}^{[2]n})'\) has the same distribution as \((QLR_{h}, \tau_{[2]h})'\), and

\[
\begin{pmatrix}
QLR_{0,n} \\
\text{Diag}^{-1}\{\tau_{0}^{[1]n}\} \\
\tau_{0}^{[2]n}
\end{pmatrix}
\rightarrow
\begin{pmatrix}
QLR_{h} \\
0^{q \times q} \\
\tau_{[2]h}
\end{pmatrix}
a.s., \tag{27.7}
\]
where \( \tau_{[2]h}^0 \in R^{\min\{k,p\} - q} \). Let

\[
\tilde{\tau}_n^0 := \left( \begin{array}{c} \text{Diag}\{\tau_n^0\} \\ 0^{(k-p)\times p} \end{array} \right) \in R^{k\times p} \quad \text{and} \quad \tilde{\tau}_n := \left( \begin{array}{c} \text{Diag}\{\tau_n^0\} \\ 0^{(k-p)\times p} \end{array} \right) \in R^{k\times p} \quad \text{if} \quad k \geq p \quad \text{and} \quad (27.8)
\]

\[
\tilde{\tau}_n^0 := \left( \begin{array}{c} \text{Diag}\{\tau_n^0\}, 0 \times (p-k) \end{array} \right) \in R^{k\times p} \quad \text{and} \quad \tilde{\tau}_n := \left( \begin{array}{c} \text{Diag}\{\tau_n^0\}, 0 \times (p-k) \end{array} \right) \in R^{k\times p} \quad \text{if} \quad k < p.
\]

The distributions of \( \tilde{\tau}_n^0 \) and \( \tilde{\tau}_n \) are the same. The matrix \( \tilde{\tau}_n^0 \) has singular values given by the vector \( \tau_n^0 := (\tau_{1n}^0, \ldots, \tau_{\min\{k,p\}n}^0)' \) whose first \( q \) elements all diverge to infinity a.s. and whose last \( \min\{k,p\} - q \) elements written as the subvector \( \tau_{[2]n}^0 \) converge to \( \tau_{[2]h}^0 \) a.s. Hence, for some set \( C \in \mathcal{F}_0 \) with \( P^0(\omega \in C) = 1 \), we have \( \tau_{jn}(\omega) \rightarrow \infty \) for \( j \leq q \) and \( \tau_{[2]n}(\omega) \rightarrow \tau_{[2]h}^0(\omega) \), where \( \tau_{jn}^0(\omega), \tau_{[2]n}^0(\omega), \tau_{[2]h}^0(\omega), \) and \( \tilde{\tau}_n^0(\omega) \) denote the realizations of the random quantities \( \tilde{\tau}_{jn}^0, \tilde{\tau}_{[2]n}^0, \tilde{\tau}_{[2]h}^0, \) and \( \tilde{\tau}_n^0 \), respectively, when \( \omega \) occurs. Thus, using Lemma 27.2 with \( D_{c}^0 = \tilde{\tau}_n^0(\omega) \) and \( \tau_{c}^0 = \tilde{\tau}_{[2]h}^0(\omega) \), we have

\[
c_{k,p}(\tilde{\tau}_n^0(\omega), 1 - \alpha) \rightarrow c_{k,p,q}(\tau_{[2]h}^0(\omega), 1 - \alpha) \quad \text{for all} \quad \omega \in C \quad \text{with} \quad P^0(\omega \in C) = 1,
\]

where \( c_{k,p,q}(. , 1 - \alpha) \) is defined in Lemma 27.2. When \( q = \min\{k,p\} \), no vector \( \tau_{[2]h}^0(\omega) \) appears and by Comments (ii) and (iii) to Lemma 27.2, \( c_{k,p,q}(\tau_{[2]h}^0(\omega), 1 - \alpha) \) 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 \( (QLR_{n,0}^0, \tilde{\tau}_n^0) \) and \( (QLR_{W,U,n}, \tilde{\tau}_n) \) for all \( n \geq 1 \) and of \( (QLR_{h,n}^0, \tau_{[2]h}^0)' \) and \( (QLR_{h,n}^0, \tau_{[2]h}^0)' \), yield the following convergence result:

\[
\left( \begin{array}{c} QLR_{W,U,n} \\ c_{k,p}(n^{1/2} \tilde{W}_n \tilde{D}_n \tilde{U}_n, 1 - \alpha) \end{array} \right) \rightarrow_d \left( \begin{array}{c} QLR_{W,U,n} \\ c_{k,p}(\tilde{\tau}_n, 1 - \alpha) \end{array} \right),
\]

where the first equality holds using Lemma 16.2.

Equation (27.10) and the continuous mapping theorem give

\[
P(QLR_{W,U,n} > c_{k,p}(n^{1/2} \tilde{W}_n \tilde{D}_n \tilde{U}_n, 1 - \alpha)) \rightarrow P(QLR_{h} > c_{k,p,q}(\tau_{[2]h}, 1 - \alpha))
\]

provided \( P(QLR_{h} = c_{k,p,q}(\tau_{[2]h}, 1 - \alpha)) = 0 \). The latter holds because \( P(QLR_{h} = c_{k,p,q}(\tau_{[2]h}, 1 - \alpha) | \tilde{D}_h) = 0 \) a.s. In turn, the latter holds because, conditional on \( \tilde{D}_h \), the df of \( QLR_{h} \) is continuous at its \( 1 - \alpha \) quantile (by Lemma 27.3) where \( QLR_{h} \) conditional on \( \tilde{D}_h \) and ACLR_{k,p,q}(\tau_{c}^0), which
appears in Lemma \[27.3\] have the same structure with the former being based on \( h'_{3,k-q} \Sigma_{h,p-q} \), 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 \( h'_{3,k-q} \Sigma_{h,p-q} \) through its singular values, see \[24.3\] and \( c_{k,p,q}(\tau^c_{[2]h}, 1 - \alpha) \) is a constant (because \( \tau^c_{[2]h} \) is random only through \( \overline{D}_h \)).

By the same argument as in the proof of Lemma \[16.2\]

\[
c_{k,p,q}(\tau^c_{[2]h}, 1 - \alpha) = c_{k,p,q}(h'_{3,k-q} \Sigma_{h,p-q}, 1 - \alpha),
\]

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

Thus, we have

\[
P(\overline{QLR}_h > c_{k,p,q}(\tau^c_{[2]h}, 1 - \alpha))
= P(\overline{QLR}_h > c_{k,p,q}(h'_{3,k-q} \Sigma_{h,p-q}, 1 - \alpha))
= EP(\overline{QLR}_h > c_{k,p,q}(h'_{3,k-q} \Sigma_{h,p-q}, 1 - \alpha))
= E\alpha = \alpha,
\]

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

### 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 \( \{\tau^c_{jn} : j \leq \min\{k, p\}\} \) be the singular values of \( D^c_{nj} \), as in Lemma \[27.2\]. Define

\[
\Upsilon^c_n := \begin{pmatrix}
\text{Diag} \{\tau^c_{ijn}, \ldots, \tau^c_{ijn} \} \\
0_{(k-p)\times p}
\end{pmatrix} \in \mathbb{R}^{k\times p} \text{ if } k \geq p
\]

and

\[
\Upsilon^c_n := \begin{pmatrix}
\text{Diag} \{\tau^c_{ijn}, \ldots, \tau^c_{i kn} \}, 0^{k\times (p-k)}
\end{pmatrix} \in \mathbb{R}^{k\times p} \text{ if } k < p.
\]

(27.14)
Lemma 27.5 Suppose the scalar constants \( \{ \tau^c_{jn} \geq 0 : j \leq \min\{k, p\} \} \) for \( n \geq 1 \) satisfy (i) \( \{ \tau^c_{jn} \geq 0 : j \leq \min\{k, p\} \} \) are nonincreasing in \( j \) for \( n \geq 1 \), (ii) \( \tau^c_{jn} \to \infty \) for \( j \leq q \) for some \( 1 \leq q \leq \min\{k, p\} \), (iii) \( \tau^c_{jn} \to \tau^c_{jf} < \infty \) for \( j = q+1, \ldots, \min\{k, p\} \), and (iv) when \( p \geq 2 \), \( \tau^c_{jn+1} / \tau^c_{jn} \to h^c_{0,j} \) for some \( h^c_{0,j} \in [0, 1] \) for all \( j \leq \min\{k, p\} - 1 \). Let \( \Upsilon^c_n \) be defined as in (27.14). Let \( \{ \kappa^Z_{jn} : j \leq p + 1 \} \) denote the \( p + 1 \) eigenvalues of \( (\Upsilon^c_n, Z)^T(\Upsilon^c_n, Z) \), ordered to be nonincreasing in \( j \), where \( Z \sim N(0^k, I_k) \). Then,

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

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

Comment: Lemma 27.5 only applies when \( q \geq 1 \), whereas Lemma 27.2 applies when \( q \geq 0 \).

Lemma 27.6 Let \( \{ F_n^*(x) : n \geq 1 \} \) and \( F^*(x) \) be df’s on \( R \) and let \( \alpha \in (0, 1) \) be given. Suppose (i) \( F_n^*(x) \to F^*(x) \) for all continuity points \( x \) of \( F^*(x) \) and (ii) \( F^*(q_\infty + \varepsilon) > 1 - \alpha \) for all \( \varepsilon > 0 \), where \( q_\infty := \inf\{ x : F^*(x) \geq 1 - \alpha \} \) is the \( 1 - \alpha \) quantile of \( F^*(x) \). Then, the \( 1 - \alpha \) quantile of \( F_n^*(x) \), viz., \( q_n := \inf\{ x : F_n^*(x) \geq 1 - \alpha \} \), satisfies \( q_n \to q_\infty \).

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

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

\[
\tau^c_{(j+1)n} / \tau^c_{jn} \to h^c_{0,j} \text{ for all } j \leq \min\{k, p\} - 1 \tag{27.15}
\]

(as well as the other assumptions on the singular values stated in the theorem). The same argument holds with \( n \) replaced by \( u_n \) below, which is the result that is needed to complete the proof. When \( p = 1 \), we prove the full sequence result that \( c_{k,p}(\Upsilon^c_{u_n}, 1 - \alpha) \to c_{k,p,q}(\tau^c_{\infty}, 1 - \alpha) \) without the condition in (27.15) (which is meaningless in this case because there is only one value \( \tau^c_{jn} \), namely \( \tau^c_{jn} \) for each \( n \)). In this case too, the same argument holds with \( n \) replaced by \( u_n \).

---

67 The condition in (27.15) is required by Lemma 27.5, which is used in the proof of Lemma 27.2 below.
below, which is the result that is needed to complete the proof. We treat the cases \( p \geq 2 \) and \( p = 1 \) simultaneously from here on.

First, we show that

\[
CLR_{k,p}(\mathcal{Y}_n^c) := Z'Z - \lambda_{\min}(\mathcal{Y}_n^c, Z) \rightarrow Z'Z - \lambda_{\min}(\mathcal{Y}(\tau^c_{\infty}), Z_2) := ACLR_{k,p,q}(\tau^c_{\infty})
\]

(27.16)

for all realizations of \( Z \). If \( q = 0 \), then (27.16) holds because \( \mathcal{Y}_n^c \rightarrow \mathcal{Y}(\tau^c_{\infty}) \) (by the definition of \( \mathcal{Y}_n^c \) in (27.14), the definition of \( \mathcal{Y}(\tau^c_{\infty}) \) 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, Thm. 3.1, pp. 37–38)).

Now, we establish (27.16) when \( q \geq 1 \). The (ordered) eigenvalues \( \{\kappa_j^Z : j \leq p + 1\} \) of \( (\mathcal{Y}_n^c, Z)'(\mathcal{Y}_n^c, Z) \) are solutions to

\[
|((\mathcal{Y}_n^c, Z)'(\mathcal{Y}_n^c, Z) - \kappa I_{p+1}| = 0 \text{ or }
|Q_n^c(\kappa)| = 0, \text{ where } Q_n^c(\kappa) := S_n^c((\mathcal{Y}_n^c, Z)'(\mathcal{Y}_n^c, Z)S_n^c - \kappa(S_n^c)^2 \text{ and }
S_n^c := \text{Diag}((\tau^c_1)^{-1}, ..., (\tau^c_n)^{-1}, 1, ..., 1) \in R^{(p+1)\times(p+1)}.
\]

(27.17)

Define

\[
S_{n,q}^c := \text{Diag}((\tau^c_1)^{-1}, ..., (\tau^c_n)^{-1}) \in R^{q\times q}.
\]

(27.18)

We have

\[
(\mathcal{Y}_n^c, Z)S_n^c = \begin{pmatrix} I_q & 0_{(p+1-q)\times q} \end{pmatrix} S_{n,q}^c \begin{pmatrix} I_q & 0_{q\times(p+1-q)} \end{pmatrix} = (I_{k,q}, \mathcal{Y}_{n,p-q}, Z) \in R^{k\times(p+1)}, \text{ where }
\]

\[
I_{k,q} := \begin{pmatrix} I_q & 0_{q\times(p-q)} \end{pmatrix} \in R^{k\times q},
\]

(27.19)

\[
\mathcal{Y}_{n,p-q}^c := \begin{pmatrix} \text{Diag}\{\tau^c_{(q+1)n}, ..., \tau^c_{pn}\} & 0_{(k-p)\times(p-q)} \end{pmatrix} \in R^{k\times(p-q)} \text{ if } k \geq p, \text{ and }
\]

\[
\mathcal{Y}_{n,p-q}^c := \begin{pmatrix} \text{Diag}\{\tau^c_{(q+1)n}, ..., \tau^c_{kn}\} & 0_{(k-q)\times(p-k)} \end{pmatrix} \in R^{k\times(p-q)} \text{ if } k < p.
\]
By (27.17) and (27.19), we have

\[ Q^c_n(\kappa) = \begin{bmatrix} I_q & I'_{k,q}(\Upsilon^c_{n,p-q}, Z) \\ (\Upsilon^c_{n,p-q}, Z)'I_{k,q} & (\Upsilon^c_{n,p-q}, Z)'(\Upsilon^c_{n,p-q}, Z) \end{bmatrix} - \kappa \begin{bmatrix} (S^c_{n,q})^2 & 0^{q \times (p+1-q)} \\ 0^{(p+1-q) \times q} & I_{p+1-q} \end{bmatrix}. \]  

By the formula for the determinant of a partitioned inverse,

\[ |Q^c_n(\kappa)| = |Q^c_{n,1}(\kappa)| \cdot |Q^c_{n,2}(\kappa)|, \quad \text{where} \]

\[ Q^c_{n,1}(\kappa) := I_q - \kappa(S^c_{n,q})^2 \in \mathbb{R}^{q \times q} \quad \text{and} \]

\[ Q^c_{n,2}(\kappa) := (\Upsilon^c_{n,p-q}, Z)'(\Upsilon^c_{n,p-q}, Z) - \kappa I_{p+1-q} - (\Upsilon^c_{n,p-q}, Z)'I_{k,q}(I_q - \kappa(S^c_{n,q})^2)^{-1}I'_{k,q}(\Upsilon^c_{n,p-q}, Z) \in \mathbb{R}^{(p+1-q) \times (p+1-q)}. \]  

(27.21)

For \( j = q + 1, \ldots, p + 1 \), we have

\[ Q^c_{n,1}(\kappa^Z_{jn}) = I_q - \kappa^Z_{jn}(S^c_{n,q})^2 = I_q - \text{Diag}\{\kappa^Z_{jn}(\tau^c_{1n})^{-2}, \ldots, \kappa^Z_{jn}(\tau^c_{qn})^{-2}\} = I_q + o(1) \]  

(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 \( |Q^c_{n,1}(\kappa^Z_{jn})| \neq 0 \) for \( j = q + 1, \ldots, p + 1 \) for \( n \) large. Hence, for \( n \) large,

\[ |Q^c_{n,2}(\kappa^Z_{jn})| = 0 \quad \text{for} \quad j = q + 1, \ldots, p + 1. \]  

(27.23)

We write

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

(27.24)

and \( I_{k,q} \) is defined in (27.19) \(^{68}\)

For \( j = q + 1, \ldots, p + 1 \), we have

\[ Q^c_{n,2}(\kappa^Z_{jn}) = (\Upsilon^c_{n,p-q}, Z)'(\Upsilon^c_{n,p-q}, Z) - \kappa^Z_{jn}I_{p+1-q} - (\Upsilon^c_{n,p-q}, Z)'I_{k,q}(I_q + o(1))I'_{k,q}(\Upsilon^c_{n,p-q}, Z) \]

\[ = (\Upsilon^c_{n,p-q}, Z)'I_{k,k-q}I_{k,k-q}(\Upsilon^c_{n,p-q}, Z) + o(1) - \kappa^Z_{jn}I_{p+1-q} \]

\[ := M'_{n,p+1-q} - \kappa^Z_{jn}I_{p+1-q}, \]  

(27.25)

where the first equality holds by (27.22) and the definition of \( Q^c_{n,2}(\kappa) \) 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 \( \Upsilon^c_{n,p-q} = O(1) \) by its definition in (27.19) and the condition (iii) of Lemma 27.2 on \( \{\tau^c_{jn} : j = q + 1, \ldots, \min\{k,p\}\} \)

\(^{68}\)There is some abuse of notation here because \( I_{k,q} \) does not equal \( I_{k,k-q} \) even if \( q \) equals \( k - q \).
for \( n \geq 1 \).

Equations (27.23) and (27.25) imply that \( \kappa_{j_n}^Z : j = q + 1, \ldots, p + 1 \) are the \( p + 1 - q \) eigenvalues of the matrix \( M_{n,p+1-q}^c \). By the definition of \( Y_n^c, p-q \) in (27.19) and the conditions of the lemma on \( \{r_{jn} : j = q + 1, \ldots, \min\{k,p\} \} \) for \( n \geq 1 \), we have

\[
M_{n,p+1-q}^c \rightarrow \left( \begin{pmatrix} 0 & (p-q) \\ Y(\tau_{\infty}^c) \end{pmatrix} \right) I_k k_{p-q} I_k' \left( \begin{pmatrix} 0 & (p-q) \\ Y(\tau_{\infty}^c) \end{pmatrix} \right) Z
\]

for all realizations of \( Z \), where the equality uses the definitions of \( Y(\tau_{\infty}^c) \) 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, Thm. 3.1, pp. 37–38)). Hence, by (27.26), the eigenvalues \( \{\kappa_{j_n}^Z : j = q + 1, \ldots, p + 1\} \) of \( M_{n,p+1-q}^c \) converge (for all realizations of \( Z \)) to the vector of eigenvalues of \( (Y(\tau_{\infty}^c), Z_2)'(Y(\tau_{\infty}^c), Z_2) \). In consequence, the smallest eigenvalue \( \kappa_{(p+1)n}^Z \) (of both \( M_{n,p+1-q}^c \) and \( (Y_n^c, Z)'(Y_n^c, Z) \)) satisfies

\[
\lambda_{\min}(\langle Y_n^c, Z \rangle'(Y_n^c, Z)) = \kappa_{(p+1)n}^Z \rightarrow \lambda_{\min}(\langle Y(\tau_{\infty}^c), Z_2 \rangle'(Y(\tau_{\infty}^c), Z_2)),
\]

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

Now we use (27.16) to establish that \( c_{k,p}(Y_n^c, 1-\alpha) \rightarrow c_{k,p,q}(\tau_{\infty}^c, 1-\alpha) \), which proves the lemma. Let

\[
F_{k,p,q,\tau_{\infty}^c}(x) = P(ACLRL_{k,p,q}(\tau_{\infty}^c) \leq x).
\]

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

\[
1(CLR_{k,p}(Y_n^c) \leq x) \rightarrow 1(ACLRL_{k,p,q}(\tau_{\infty}^c) \leq x) \text{ a.s.}
\]

Equation (27.29) and the bounded convergence theorem give

\[
P(CLR_{k,p}(Y_n^c) \leq x) \rightarrow P(ACLRL_{k,p,q}(\tau_{\infty}^c) \leq x) = F_{k,p,q,\tau_{\infty}^c}(x).
\]

Now Lemma 27.6 gives the desired result, because (27.30) verifies assumption \((i)\) of Lemma 27.6 and the df of \( ACLRL_{k,p,q}(\tau_{\infty}^c) \) 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 \( q \geq 1 \) in the statement of Lemma \textbf{27.5}, \( h_{6,q}^c = 0 \) if \( q < \min\{k,p\} \). If \( q = \min\{k,p\} \), then \( h_{6,q}^c \) is not defined in the statement of Lemma \textbf{27.5} and we define it here to equal zero. If \( h_{6,j}^c \), then \( \{\tau_{j,n}^c : n \geq 1\} \) and \( \{\tau_{(j+1)n}^c : n \geq 1\} \) are of the same order of magnitude, i.e., \( 0 < \lim_{n \to \infty} \tau_{(j+1)n}^c / \tau_{j,n}^c \leq 1 \). We group the first \( q \) values of \( \tau_{j,n}^c \) 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 \( \{h_{6,1}^c, \ldots, h_{6,q}^c\} \) that equal zero. Let \( r_g \) and \( r_g^c \) denote the indices of the first and last values in the \( g \)th group, respectively, for \( g = 1, \ldots, G \). Thus, \( r_1 = 1 \), \( r_g = r_{g+1} - 1 \), where by definition \( r_{G+1} = q + 1 \), and \( r_G^c = q \). By definition, the \( \tau_{j,n}^c \) values in the \( g \)th group, which have the \( g \)th largest order of magnitude, are \( \{\tau_{r_g,n}^c : n \geq 1\}, \ldots, \{\tau_{r_g^c,n}^c : n \geq 1\} \). By construction, \( h_{6,j}^c \) and \( h_{6,j}^c \) have the same order of magnitude, are \( \{\tau_{r_g,n}^c : n \geq 1\}, \ldots, \{\tau_{r_g^c,n}^c : n \geq 1\} \). By construction, \( h_{6,j}^c > 0 \) for all \( j \in \{r_g, \ldots, r_g^c - 1\} \) for \( g = 1, \ldots, G \). (The reason is: if \( h_{6,j}^c \) is equal to zero for some \( j \leq r_g^c - 1 \), then \( \{\tau_{r_g^c,n}^c : n \geq 1\} \) is of smaller order of magnitude than \( \{\tau_{r_g,n}^c : n \geq 1\} \), which contradicts the definition of \( r_g^c \).) Also by construction, \( \lim_{n \to \infty} \tau_{j,n}^c / \tau_{j,n}^c = 0 \) for any \( (j, j') \) in groups \( (g, g') \), respectively, with \( g < g' \).

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

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

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

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

Let

\[
h_{6,r_1}^{cc} := \text{Diag}\{1, h_{6,1}^c, h_{6,2}^c, \ldots, \prod_{\ell=1}^{r_1-1} h_{6,\ell}^c\} \in \mathbb{R}^{r_1 \times r_1}.\]

When \( k \geq p \), we have

\[
(\tau_{r_1,n}^c)^{-1}(\Upsilon_n^c, Z) = \\
\begin{bmatrix}
\begin{array}{cccc}
h_{6,r_1}^{cc} + o(1) & 0^{r_1 \times (q-r_1^c)} & 0^{r_1 \times (p-q)} & O(1/\tau_{r_1,n}^c) 0^{r_1 \times (p+1-r_1^c)} \\
0^{(q-r_1^c) \times r_1} & O(\tau_{r_2,n}^c \tau_{r_1,n}^c)^{(q-r_1^c) \times (q-r_1^c)} & 0^{(q-r_1^c) \times (p-q)} & O(1/\tau_{r_1,n}^c) 0^{(q-r_1^c) \times (p+1-q)} \\
0^{(p-q) \times r_1^c} & 0^{(p-q) \times (q-r_1^c)} & O(1/\tau_{r_1,n}^c) 0^{(p-q) \times (p-q)} & O(1/\tau_{r_1,n}^c) 0^{(p-q) \times (p+1-q)} \\
0^{(k-p) \times r_1^c} & 0^{(k-p) \times (q-r_1^c)} & 0^{(k-p) \times (p-q)} & O(1/\tau_{r_1,n}^c) 0^{(k-p) \times (p+1-q)} \\
h_{6,r_1}^{cc} & 0^{r_1 \times (p+1-r_1^c)} & 0^{(k-p) \times (p+1-r_1^c)} & 0^{(k-p) \times (p+1-r_1^c)}
\end{array}
\end{bmatrix}.
\]

(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 an \( s \) vector whose elements are \( O(d_n) \), the equality uses \( \tau_{jn}^c / \tau_{r1n}^c = \prod_{\ell=1}^{j-1} (\tau_{(\ell+1)n}^c / \tau_{\ell n}^c) = \prod_{\ell=1}^{j-1} h_{6,\ell}^c + o(1) \) for \( j = 2, \ldots, r_1^c \) (which holds by the definition of \( h_{6,\ell}^c \)) and \( \tau_{jn}^c / \tau_{r1n}^c = O(\tau_{r2n}^c / \tau_{r1n}^c) \) for \( j = r_2, \ldots, q \) (because \( \{\tau_{jn}^c: j \leq q\} \) are nonincreasing in \( j \)), and the convergence uses \( \tau_{r1n}^c \to \infty \) (by assumption (ii) of the lemma since \( r_1 \leq q \)) and \( \tau_{r2n}^c / \tau_{r1n}^c \to 0 \) (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

\[
(\tau_{r1n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) \to \begin{bmatrix}
(h_{6,1}^c)^2 & 0_{r_1^c \times (p+1-r_1^c)} \\
0_{(p+1-r_1^c) \times r_1^c} & 0_{(p+1-r_1^c) \times (p+1-r_1^c)}
\end{bmatrix}.
\tag{27.35}
\]

The vector of eigenvalues of a matrix is a continuous function of the matrix (by Elsner’s Theorem, see Stewart (2001, Thm. 3.1, pp. 37–38)). Hence, by (27.32) and (27.35), the first \( r_1^c \) eigenvalues of \((\tau_{r1n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)\), i.e., \{\((\tau_{r1n}^c)^{-2}\kappa_{1n}^Z: j \leq r_1^c\}\), satisfy

\[
((\tau_{r1n}^c)^{-2}\kappa_{1n}^Z, \ldots, (\tau_{r1n}^c)^{-2}\kappa_{r1^c n}^Z) \to_p (1, h_{6,1}^c, h_{6,2}^c, \ldots, \prod_{\ell=1}^{r_1^c-1} h_{6,\ell}^c) \quad \text{and so}
\]

\[
\kappa_{1n}^Z \to \infty \quad \forall j = 1, \ldots, r_1^c
\tag{27.36}
\]

because \( \tau_{r1n}^c \to \infty \) (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 \((\tau_{r1n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z)\), i.e., \{\((\tau_{r1n}^c)^{-2}\kappa_{r1^c n}^Z: j = r_1^c + 1, \ldots, p + 1\}\), satisfy

\[
((\tau_{r1n}^c)^{-2}\kappa_{r1^c n}^Z \to 0 \quad \forall j = r_1^c + 1, \ldots, p + 1.
\tag{27.37}
\]

Next, the equality in (27.34) gives

\[
(\tau_{r1n}^c)^{-2}(\Upsilon_n^c, Z)'(\Upsilon_n^c, Z) = \begin{bmatrix}
(h_{6,1}^c r_1^c)^2 + o(1) & 0_{r_1^c \times (q-r_1^c)} & 0_{r_1^c \times (p-q)} & O(1/\tau_{r1n}^c) r_1^c \times 1 \\
0_{(q-r_1^c) \times r_1^c} & O((\tau_{r2n}^c / \tau_{r1n}^c)^2(q-r_1^c) \times (q-r_1^c)) & 0_{(q-r_1^c) \times (p-q)} & O(\tau_{r2n}^c / \tau_{r1n}^c)^2(q-r_1^c) \times 1 \\
0_{(p-q) \times r_1^c} & 0_{(q-r_1^c) \times (p-q)} & O(1/\tau_{r1n}^c)^2(p-q) \times (p-q) & O(1/\tau_{r1n}^c)^2(q-r_1^c) \times 1 \\
O(1/\tau_{r1n}^c)^{1 \times r_1^c} & O(\tau_{r2n}^c / \tau_{r1n}^c)^2(1 \times (q-r_1^c)) & O(1/\tau_{r1n}^c)^{2 \times (p-q)} & O(1/\tau_{r1n}^c)^2(1 \times 1)
\end{bmatrix}.
\tag{27.38}
\]

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

\[
I_{p+1} = (I_{0, r_{j_1}}, I_{r_{j_1}, p+1}), \quad \text{where} \quad I_{0, r_{j_1}} := \begin{pmatrix} I_{r_{j_1}} \\ 0_{(p+1-r_{j_1}) \times r_{j_1}} \end{pmatrix} \in R^{(p+1) \times r_{j_1}} \text{ and}
\]

\[
I_{r_{j_1}, p+1} := \begin{pmatrix} 0_{r_{j_1} \times (p+1-r_{j_1})} \\ I_{p+1-r_{j_1}} \end{pmatrix} \in R^{(p+1) \times (p+1-r_{j_1})}. \quad (27.39)
\]

In consequence, we have

\[
(Y_n^c, Z) = ((Y_n^c, Z)I_{0, r_{j_1}}, (Y_n^c, Z)I_{r_{j_1}, p+1}) \quad \text{and} \quad
\]

\[
\varrho_n^c := (\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}, p+1} = o(\tau_{r_{j_2}}^c/\tau_{r_{j_1}}^c), \quad (27.40)
\]

where the last equality uses the expressions in the first row of the matrix on the rhs of \((27.38)\) and \(O(1/\tau_{r_{j_1}}^c) = o(\tau_{r_{j_2}}^c/\tau_{r_{j_1}}^c)\) (because \(\tau_{r_{j_2}}^c \to \infty\)).

As in \((27.32)\), \(\{(\tau_{r_{j_n}}^c)^{-2}\kappa_{j_n}^Z : j \leq p+1\}\) solve

\[
0 = |(\tau_{r_{j_1}}^c)^{-2}(Y_n^c, Z)'(Y_n^c, Z) - \kappa I_{p+1}|
\]

\[
= \left| \begin{array}{c}
(\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}} - \kappa I_{r_{j_1}} \\
(\tau_{r_{j_1}}^c)^{-2}I_{r_{j_1}, p+1}(Y_n^c, Z)'(Y_n^c, Z)I_{0, r_{j_1}} \\
(\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}, p+1} - \kappa I_{p+1-r_{j_1}} \\
(\tau_{r_{j_1}}^c)^{-2}I_{r_{j_1}, p+1}(Y_n^c, Z)'(Y_n^c, Z)I_{0, r_{j_1}} - \kappa I_{r_{j_1}} \\
\end{array} \right|
\]

\[
= |(\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}} - \kappa I_{r_{j_1}}|
\]

\[
\times |(\tau_{r_{j_1}}^c)^{-2}I_{r_{j_1}, p+1}(Y_n^c, Z)'(Y_n^c, Z)I_{0, r_{j_1}} - \kappa I_{r_{j_1}}|
\]

\[
\times |\varrho_n^c((\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}} - \kappa I_{r_{j_1}})^{-1}\varrho_n^c|, \quad (27.41)
\]

where the third equality uses the standard formula for the determinant of a partitioned matrix, the definition of \(\varrho_n^c\) 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 \(\kappa\) equal to any solution \((\tau_{r_{j_1}}^c)^{-2}\kappa_{j_n}^Z\) to the first equality in \((27.41)\) for \(j = r_{j_1}^c + 1, \ldots, p+1\).

Now we show that, for \(j = r_{j_1}^c + 1, \ldots, p+1\), \((\tau_{r_{j_1}}^c)^{-2}\kappa_{j_n}^Z\) cannot solve the determinantal equation \(|(\tau_{r_{j_1}}^c)^{-2}I_{0, r_{j_1}}(Y_n^c, Z)'(Y_n^c, Z)I_{r_{j_1}} - \kappa I_{r_{j_1}}| = 0\) for \(n\) sufficiently large, where this determinant is the first multiplicand on the rhs of \((27.41)\). Hence, \(\{(\tau_{r_{j_1}}^c)^{-2}\kappa_{j_n}^Z : j = r_{j_1}^c + 1, \ldots, p+1\}\) must solve the determinantal equation based on the second multiplicand on the rhs of \((27.41)\) for \(n\) sufficiently
large. For \( j = r_1^c + 1, ..., p + 1 \), we have

\[
(\tau_{r_1^c})^{-2} I'_{0,r_1^c} \left( Y_n^c, Z \right)' (Y_n^c, Z) I_{0,r_1^c} - (\tau_{r_1^c})^{-2} \kappa_{Jn}^Z I_{r_1^c} = (h_{6,r_1^c}^c)^2 + o(1),
\]

where the equality holds by (27.35) and (27.37). Equation (27.42) and \( \lambda_{\min}((h_{6,r_1^c}^c)^2) > 0 \) (which follows from the definition of \( h_{6,r_1^c}^c \) in (27.33) and the fact that \( h_{6,j}^c > 0 \) for all \( j \in \{1, ..., r_1^c - 1\} \)) establish the desired result.

For \( j = r_1^c + 1, ..., p + 1 \), plugging \( (\tau_{r_1^c})^{-2} \kappa_{Jn}^Z \) into the second multiplicand on the rhs of (27.41) and using (27.40) and (27.42) gives

\[
0 = |(\tau_{r_1^c})^{-2} I'_{r_1^c,p+1} (Y_n^c, Z)' (Y_n^c, Z) I_{r_1^c,p+1} + o((\tau_{r_2^c}/\tau_{r_1^c})^2) - (\tau_{r_1^c})^{-2} \kappa_{Jn}^Z I_{p+1-r_1^c}|. \tag{27.43}
\]

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

\[
0 = |(\tau_{r_2^c})^{-2} I'_{r_1^c,p+1} (Y_n^c, Z)' (Y_n^c, Z) I_{r_1^c,p+1} + o(1) - \kappa_{p+1-r_1^c}|. \tag{27.44}
\]

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

\[
0 = |(\tau_{r_2^c})^{-2} I'_{r_1^c,p+1} (Y_n^c, Z)' (Y_n^c, Z) I_{r_1^c,p+1} + o(1) - \kappa_{p+1-r_1^c}|. \tag{27.45}
\]

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 \( I_{p+1-r_1^c}, \tau_{r_2^c}, \tau_{r_3^c}, r_2^c-r_1^c, p+1-r_2^c, \) and \( h_{6,r_2^c}^c = \text{Diag}\{1, h_{6,r_1^c+1}^c, \ldots, h_{6,r_2^c+1}^c \} \in R^{(r_2^c-r_1^c) \times (r_2^c-r_1^c)} \) in place of \( I_{p+1}, \tau_{r_1^c}, \tau_{r_2^c}, r_1^c, p+1-r_1^c, \) and \( h_{6,r_1^c}^c \), 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

\[
\kappa_{Jn}^Z \to \infty \forall j = r_2^c, ..., r_2^c \text{ and } (\tau_{r_2^c})^{-2} \kappa_{Jn}^Z = o(1) \forall j = r_2^c + 1, ..., p + 1. \tag{27.46}
\]

Repeating the argument \( G - 2 \) more times yields

\[
\kappa_{Jn}^Z \to \infty \forall j = 1, ..., r_G^c \text{ and } (\tau_{r_G^c})^{-2} \kappa_{Jn}^Z = o(1) \forall j = r_G^c + 1, ..., p + 1, \forall g = 1, ..., G. \tag{27.47}
\]

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 \( r_j^G = q \), the first result in (27.47) proves part (a) of the lemma.
The second result in (27.47) with \( g = G \) implies: for all \( j = q+1, \ldots, p+1 \),

\[
(\tau_{rGn}^\epsilon)^{-2} \kappa_{jn}^Z = o(1) \tag{27.48}
\]

because \( r_{Gj}^\epsilon = q \). Either \( r_G = r_{Gj}^\epsilon = q \) or \( r_G < r_{Gj}^\epsilon = q \). In the former case, \( (\tau_{qGn}^\epsilon)^{-2} \kappa_{jn}^Z = o(1) \) for \( j = q+1, \ldots, p+1 \) by (27.47). In the latter case, we have

\[
\lim \frac{\tau_{qGn}^\epsilon}{\tau_{rGn}^\epsilon} = \lim \frac{\tau_{rGn}^\epsilon}{\tau_{rGn}^q} = \prod_{j=r_G}^{r_{Gj}^\epsilon-1} h_{0,j}^c > 0, \tag{27.49}
\]

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

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

\[
F_n^*(q_\infty - \epsilon) \to F^*(q_\infty - \epsilon) < 1 - \alpha \quad \text{and} \quad F_n^*(q_\infty + \epsilon) \to F^*(q_\infty + \epsilon) > 1 - \alpha \tag{27.50}
\]

by assumptions (i) and (ii) of the lemma and \( F^*(q_\infty - \epsilon) < 1 - \alpha \) by the definition of \( q_\infty \). The first line of (27.50) implies that \( q_n \geq q_\infty - \epsilon \) for all \( n \) large. (If not, there exists an infinite subsequence \( \{w_n\} \) of \( \{n\} \) for which \( q_{wn} < q_\infty - \epsilon \) for all \( n \geq 1 \) and \( 1 - \alpha \leq F_{wn}^*(q_{wn}) \leq F_{wn}^*(q_\infty - \epsilon) \to F^*(q_\infty - \epsilon) < 1 - \alpha \), which is a contradiction). The second line of (27.50) implies that \( q_n \leq q_\infty + \epsilon \) for all \( n \) large. There exists a sequence \( \{\varepsilon_k > 0 : k \geq 1\} \) for which \( \varepsilon_k \to 0 \) and \( q_\infty \pm \varepsilon_k \) are continuity points of \( F^*(x) \) for all \( k \geq 1 \). Hence, \( q_n \to q_\infty \). \( \square \)

### 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) \( \tau_{\min(k,p)}^\epsilon \triangleq 0 \), where \( \tau_{\min(k,p)}^\epsilon \) denotes the \( \min\{k,p\} \)th (and, hence, last and smallest) element of \( \tau_{\cdot}^\epsilon \), and (iv) \( q = 0 \), \( k > p \), and \( \tau_{p\infty}^\epsilon > 0 \). First, suppose \( q \geq 1 \). Then,

\[
ACLR_{k,p,q}(\tau_{\infty}^\epsilon) = Z'Z - \lambda_{\min}((\Upsilon(\tau_{\infty}^\epsilon), Z_2)'(\Upsilon(\tau_{\infty}^\epsilon), Z_2))
\]

\[
= Z_1^T Z_1 + Z_2^T Z_2 - \lambda_{\min}((\Upsilon(\tau_{\infty}^\epsilon), Z_2)'(\Upsilon(\tau_{\infty}^\epsilon), Z_2)) \tag{27.51}
\]

and \( ACLR_{k,p,q}(\tau_{\infty}^\epsilon) \) is the convolution of a \( \chi_q^2 \) distribution (since \( Z_1^T Z_1 \sim \chi_q^2 \)) and another dis-
tribution. 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 \( \mathbb{R} \) with Lebesgue measure zero. Then,

\[
P(X + Y \in B) = \int P(X + y \in B | Y = y) dP_Y(y) = \int P(X \in B - y) dP_Y(y) = 0,
\]

(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 \( ACLR_{k,p,q}(\tau_c^\infty) \) 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 \( \varepsilon > 0 \),

\[
P(Y < c + \varepsilon) \geq P(Y < c_Y + \varepsilon) \geq 1 - \alpha > 0.
\]

(27.53)

For \( \varepsilon > 0 \), we have

\[
P(X + Y \in [c, c + \varepsilon]) = \int P(X + y \in [c, c + \varepsilon] | Y = y) dP_Y(y)
= \int P(X \in [c - y, c - y + \varepsilon]) dP_Y(y) > 0,
\]

(27.54)

where the first equality holds by the law of iterated expectations, the second equality holds by the independence of \( X \) and \( Y \), and the inequality holds because \( P(X \in [c - y, c - y + \varepsilon]) > 0 \) for all \( y < c + \varepsilon \) (because the support of \( X \) is \( R_+ \)) and \( P(Y < c + \varepsilon) > 0 \) by (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 \( ACLR_{k,p,q}(\tau_c^\infty) = X + Y \) and \( Z_1'Z_1 = X \). This implies that the df of \( ACLR_{k,p,q}(\tau_c^\infty) \) is strictly increasing at its \( 1 - \alpha \) quantile when \( q \geq 1 \).

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

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

Fourth, suppose $q = 0$, $k > p$, and $\tau_{\rho \infty}^c > 0$. In this case, $Z_2 = Z$ (because $q = 0$) and $\Upsilon(\tau_{\infty}^c) = (D, 0^{p \times (k-p)})'$, where $D := \text{Diag}\{\tau_{\infty}^c\}$ is a pd diagonal $p \times p$ matrix (because $\tau_{\rho \infty}^c > 0$).

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 $ACLR_{k,p,q}(\tau_{\infty}^c)$. In the present case, we have

$$ACLR = Z'Z - \lambda_{\min} \left( \begin{pmatrix} D & Z_a \\ 0^{(k-p) \times p} & Z_b \end{pmatrix} \right)' \left( \begin{pmatrix} D & Z_a \\ 0^{(k-p) \times p} & Z_b \end{pmatrix} \right) = Z'Z - \inf_{\xi = (\xi_1, \xi_2)'} \left( \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \right)' \left( \begin{pmatrix} D^2 & DZ_a \\ Z_a'D & Z'Z \end{pmatrix} \right) \left( \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \right)$$

$$= \sup_{\xi = (\xi_1, \xi_2)'} \left[ (1 - \xi_1^2z_a)(Z_b'Z_b + Z_a'Z_a) - \xi_1'D^2\xi_1 - 2\xi_2Z_a'D\xi_1 \right],$$

(27.55)

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

We define the following non-stochastic function

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

(27.56)

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

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

Using this, we have: for all $x \geq ACLR(z_a, 0)$ and $z_a \neq 0^p$,

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

(27.57)

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

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

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

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

$$
\lim_{x \to x_0} P(ACLR(z_a, Z'_b Z_b) \leq x) = \lim_{x \to x_0} P(Z'_b Z_b \leq ACLR^{-1}(z_a, x)) = P(Z'_b Z_b \leq ACLR^{-1}(z_a, x_0)), \tag{27.59}
$$

where the first equality holds by (27.57) and the second equality holds by the continuity of the df of the $\chi^2_{k-p}$ random variable $Z'_b Z_b$ and the continuity of $ACLR^{-1}(z_a, x)$ at $x_0$. Hence, (27.58) holds when $x_0 > ACLR(z_a, 0)$.

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

$$
P(ACLR(z_a, Z'_b Z_b) \leq x_0) \leq P(ACLR(z_a, Z'_b Z_b) < ACLR(z_a, 0)) = 0, \tag{27.60}
$$

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

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

$$
P(ACLR(z_a, Z'_b Z_b) \leq x_0) = P(ACLR(z_a, Z'_b Z_b) \leq ACLR(z_a, 0)) \leq P(Z'_b Z_b \leq 0) = 0, \tag{27.61}
$$

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

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

$$
\lim_{x \to x_0} P(ACLR \leq x) = \lim_{x \to x_0} P(ACLR(Z_a, Z'_b Z_b) \leq x) \\
= \lim_{x \to x_0} \int P(ACLR(z_a, Z'_b Z_b) \leq x) dF_{Z_a}(z_a) \\
= \int P(ACLR(z_a, Z'_b Z_b) \leq x_0) dF_{Z_a}(z_a) \\
= P(ACLR \leq x_0), \tag{27.62}
$$

where $F_{Z_a}(\cdot)$ denotes the df of $Z_a$, the first and last equalities hold because $ACLR = ACLR(Z_a, Z'_b Z_b)$, the second equality uses the independence of $Z_a$ and $Z_b$, and the third equality holds by the
bounded convergence theorem using (27.58) and \( P(Z_0 \neq 0^p) = 1 \). 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 \( ACLR \geq 0 \) 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 \( x \geq ACLR(z_a, 0), \delta > 0, \) and \( z_a \neq 0^p \), we have

\[
P(ACLR(z_a, Z'_bZ_b) \in [x, x + \delta]) = P(Z'_bZ_b \in [ACLR^{-1}(z_a, x), ACLR^{-1}(z_a, x + \delta)]) > 0, \tag{27.63}
\]

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

The function \( ACLR(z_a, 0) \) is continuous at all \( z_a \in R^p \) (by property (iii)) and \( ACLR(0^p, 0) = 0 \) (by a simple calculation using (27.56)). In consequence, for any \( x > 0 \), there exists a vector \( z_a^* \in R^p \) and a constant \( \varepsilon > 0 \) such that \( ACLR(z_a, 0) < x \) for all \( z_a \in B(z_a^*, \varepsilon) \), where \( B(z_a^*, \varepsilon) \) denotes a ball centered at \( z_a^* \) with radius \( \varepsilon > 0 \). Using this, we have: for any \( x > 0 \) and \( \delta > 0 \),

\[
P(ACLR \in [x, x + \delta]) \geq \int_{B(z_a^*, \varepsilon)} P(ACLR(z_a, Z'_bZ_b) \in [x, x + \delta])dF_{Z_a}(z_a) > 0, \tag{27.64}
\]

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 \( x > ACLR(z_a, 0) \) 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 \( ACLR(z_a, \omega) \), which are stated above. The function \( ACLR(z_a, \omega) \) is seen to be nonnegative by replacing the supremum in (27.56) by \( \xi = (0^p, 1)' \). Hence, property (i) holds. The function \( ACLR(z_a, \omega) \) can be written as

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

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, Thm. 3.1, pp. 37–38). Hence, $ACLR(z_a, \omega)$ is continuous in $(z_a, \omega) \in R^p \times R_+$ and property (iii) holds.

For any $\xi_{s_2}^2 \in [0,1)$ and $\xi_{s_1} \in R^p$ such that $\xi'_{s_1} \xi_{s_1} = 1 - \xi_{s_2}^2$, we have

$$ACLR(z_a, \omega) \geq (1 - \xi_{s_2}^2)(\omega + z'_a z_a) - \xi'_{s_1} D^2 \xi_{s_1} - 2 \xi_{s_2} z'_a D \xi_{s_1} \to \infty \text{ as } \omega \to \infty,$$  \hspace{1cm} (27.66)

where the inequality holds by replacing the supremum over $\xi$ in (27.56) by the same expression evaluated at $\xi = (\xi'_{s_1}, \xi_{s_2})'$ and the divergence to infinity uses $1 - \xi_{s_2}^2 > 0$. Hence, property (iv) holds.

It remains to verify property (ii), which states that $ACLR(z_a, \omega)$ is strictly increasing in $\omega$ on $R_+ \forall z_a \neq 0^p$. For $\omega \in R_+$, let $\xi_{\omega} = (\xi'_{\omega_1}, \xi_{\omega_2})'$ (for $\xi_{\omega_1} \in R^p$ and $\xi_{\omega_2} \in R$) be such that $||\xi_{\omega}|| = 1$ and

$$ACLR(z_a, \omega) = (1 - \xi_{\omega_2}^2)(\omega + z'_a z_a) - \xi'_{\omega_1} D^2 \xi_{\omega_1} - 2 \xi_{\omega_2} z'_a D \xi_{\omega_1}.$$  \hspace{1cm} (27.67)

Such a vector $\xi_{\omega}$ 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 $\xi_{\omega}$. (Note that $\xi_{\omega}$ typically depends on $z_a$ as well as $\omega$.) Using (27.67), we obtain: for all $\delta > 0$, if $\xi_{\omega_2} < 1$,

$$ACLR(z_a, \omega) < (1 - \xi_{\omega_2}^2)(\omega + \delta + z'_a z_a) - \xi'_{\omega_1} D^2 \xi_{\omega_1} - 2 \xi_{\omega_2} z'_a D \xi_{\omega_1}$$

$$\leq \sup_{\xi = (\xi'_{\omega_1}, \xi_{\omega_2})' ||\xi||=1} \left[(1 - \xi_{\omega_2}^2)(\omega + \delta + z'_a z_a) - \xi'_{\omega_1} D^2 \xi_{\omega_1} - 2 \xi_{\omega_2} z'_a D \xi_{\omega_1}\right]$$

$$= ACLR(z_a, \omega + \delta).$$  \hspace{1cm} (27.68)

Equation (27.68) shows that $ACLR(z_a, \omega)$ is strictly increasing at $\omega$ provided $\xi_{\omega_2} < 1$.

Next, we show that $\xi_{\omega_2} = 1$ only if $z_a = 0^p$. By (27.56) and (27.67), $\xi_{\omega}$ 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

$$(1 - \xi_{2}^2)(\omega + z'_a z_a) - \xi'_{1} D^2 \xi_{1} - 2 \xi_{2} z'_a D \xi_{1} + \gamma (1 - \xi_{2}^2 - \xi_{1}^2),$$  \hspace{1cm} (27.69)

where $\gamma \in R$ is the Lagrange multiplier. The first-order conditions of the Lagrangian with respect to $\xi_1$, evaluated at the solution $(\xi'_{\omega_1}, \xi_{\omega_2})'$ and the corresponding Lagrange multiplier, say $\gamma_{\omega}$, are

$$- 2 D^2 \xi_{\omega_1} - 2 \xi_{\omega_2} D z_a - 2 \gamma_{\omega} \xi_{\omega_1} = 0^p.$$  \hspace{1cm} (27.70)

The solution is $\xi_{\omega_1} = 0^p$ (which is an interior point of the set $\{\xi_1 : ||\xi_1|| \leq 1\}$) only if $\xi_{\omega_2} = 0$ or $z_a = 0^p$ (because $D$ is a pd diagonal matrix). Thus, $\xi_{\omega_2} = 1 - \xi_{\omega_1} \xi_{\omega_1} = 1$ only if $z_a = 0^p$. This concludes the proof of property (iv). \quad \Box
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 \( \{w_n : n \geq 1\} \). The same proof holds for any subsequence \( \{w_n : n \geq 1\} \).

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

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

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 \( \{\lambda_{n,h} : n \geq 1\} \)). Hence, Assumption WU(a) holds for the parameter space \( \Lambda_{WU} \) with \( h_7 = h_{5,g} \).

Next, we establish Assumption WU(b) for the parameter space \( \Lambda_{WU} \). Using the definition of \( \hat{V}_n (= \hat{V}_n(\theta_0)) \) in (5.3), we have

\[
\hat{V}_n = n^{-1} \sum_{i=1}^{n} f_i f'_i - \hat{f}_n \hat{f}'_n = E_{F_n} f_i f'_i - (E_{F_n} f_i)(E_{F_n} f'_i) + o_p(1) \tag{27.72}
\]

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

\[
\hat{R}_n = (B' \otimes I_k) (E_{F_n} f_i f'_i - (E_{F_n} f_i)(E_{F_n} f'_i)) (B \otimes I_k) + o_p(1)
\to_p R_h := (B' \otimes I_k) [h_5 - vec((0^k, h_4))vec((0^k, h_4))'] (B \otimes I_k), \tag{27.73}
\]

where \( B = B(\theta_0) \) is defined in (5.3), 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

\[
\hat{U}_{2n} = (\hat{\Omega}_n, \hat{R}_n) \to_p (h_{5,g}, R_h) = h_8, \tag{27.74}
\]

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

Now we establish Assumption WU(c) for the parameter space \( \Lambda_{WU} \) for part (a) of the lemma. We take \( W_2 \) (which appears in the statement of Assumption WU(c)) to be the space of psd \( k \times k \) matrices and \( U_2 \) (which also appears in Assumption WU(c)) to be the space of non-zero psd matrices \( (\Omega, R) \) for \( \Omega \in R^{k \times k} \) and \( R \in R^{(p+1)k \times (p+1)k} \). By the definition of \( \hat{W}_{2n}, \hat{W}_{2n} \in W_2 \) a.s. We have \( W_{2F} \in W_2 \forall F \in \mathcal{F}_{WU} \) because \( W_{2F} = E_{F} g_i g'_i \) is psd. We have \( U_{2F} \in U_2 \forall F \in \mathcal{F}_{WU} \) because \( U_{2F} = (\Omega_F, R_F) \), \( \Omega_F := E_{F} g_i g'_i \) is psd and non-zero (by the last condition in \( \mathcal{F} \), even if that condition is weaken to \( \lambda_{\text{max}}(E_{F} g_i g'_i) \geq \delta \)) and \( R_F := (B' \otimes I_k)V_F(B \otimes I_k) \) is psd and non-zero.
because $B$ is nonsingular and $V_F$ (defined in (16.7)) is non-zero by the argument given in the paragraph containing (27.77) below. By their definitions, $\hat{\Omega}_n$ and $\hat{R}_n$ are psd. In addition, they are non-zero wp→1 by (27.74) and the result just established that the two matrices that comprise $h_8$ are non-zero. Hence, $(\hat{\Omega}_n, \hat{R}_n) \in \mathcal{U}_2$ wp→1.

The function $W_1(W_2) = W_2^{-1/2}$ is continuous at $W_2 = h_7$ on $\mathcal{W}_2$ because $\lambda_{\min}(h_7) > 0$ (given that $h_7 = \lim E_{F_n} g_i g_i'$ and $\lambda_{\min}(E_{F_n} g_i g_i') \geq \delta$ 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 $\Omega \in \mathbb{R}^{k \times k}$ and $R \in \mathbb{R}^{(p+1)k \times (p+1)k}$ with $(\Omega, R)$ 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 $(\Omega, R)$ in place of $(\Omega_{F_n}, R_{F_n})$ and $\Sigma^{\varepsilon}(\Omega, R)$ is defined given $\Sigma(\Omega, R)$ by (5.6). Lemma 22.1(b) shows that $\Sigma^{\varepsilon}(\Omega, R)$ is nonsingular provided $\lambda_{\max}(\Sigma(\Omega, R)) > 0$. We have

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

$$\geq \max_{j \leq p+1} \lambda_{\max}(\Omega^{-1/2} R_{jj} \Omega^{-1/2}) / k = \max_{j \leq p+1} \lambda_{\max}(\Omega^{-1/2}) \sup_{\lambda \in \mathbb{R}^{|X|}} \frac{|\Omega^{-1/2} R_{jj}| \Omega^{-1/2} \lambda \cdot \|\Omega^{-1/2} \lambda\|^2 / k}{\|\Omega^{-1/2} \lambda\|^2}$$

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

(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 (5.5) with $(\hat{\Omega}_n(\theta), \hat{R}_n(\theta))$, 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 $\Omega$ in some neighborhood of $\lim \Omega_{F_n}$ (because $\lambda_{\max}(\Omega_F) = \sup_{\lambda \in \mathbb{R}^{|X|}} |\lambda| = 1 \frac{\lambda g_f}{E_{F}(\lambda g_f)^2 \leq E_{F}|g_f|^2 \leq M^2/(2+\gamma) < \infty$ 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 $j \leq p + 1$.

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

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

(27.76)

We show that the last condition in $\mathcal{F}$, i.e., $\inf_{F \in \mathcal{F}} \lambda_{\min}(E_F g_i g_i') > 0$ 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 $\hat{\Omega}_{n}^{-1/2}$ enters the definition of $\hat{D}_n$ and $\hat{\Omega}_n - \Omega_{F_n} \rightarrow_p 0^{k \times k}$, where $\Omega_F = E_F g_i g_i'$.) We show that (27.76) holds provided $\inf_{F \in \mathcal{F}} \lambda_{\max}(E_F g_i g_i') > 0$. 

97
Let \( x^* \in R^{(p+1)k} \) be such that \( ||x^*|| = 1 \) and \( \lambda_{\text{max}}(V_F) = x^* V_F x^* \). Let \( x^+ = (B \otimes I_k)^{-1} x^* \). Then, we have

\[
\lambda_{\text{max}}(R_F) := \lambda_{\text{max}}((B' \otimes I_k) V_F (B \otimes I_k)) = \sup_{x \in R^{(p+1)k}} x' (B' \otimes I_k) V_F (B \otimes I_k) x
\]

\[
\geq x'' (B' \otimes I_k) V_F (B \otimes I_k) x^+ \cdot ||x||^{-2} = x'' V_F x^* / (x'' (B \otimes I_k)^{-1} (B \otimes I_k)^{-1} x^*)
\]

\[
\geq \lambda_{\text{max}}(V_F) / \lambda_{\text{max}}((B \otimes I_k)^{-1} (B \otimes I_k)^{-1}) = K \lambda_{\text{max}}(V_F),
\]

where \( K := 1 / \lambda_{\text{max}}((B \otimes I_k)^{-1} (B \otimes I_k)^{-1}) \) 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 (5.3)). Next, \( \inf_{F \in \mathcal{F}} \lambda_{\text{max}}(V_F) \geq \inf_{F \in \mathcal{F}} \lambda_{\text{max}}(E_F g_i g_i') \geq \delta \) because \( E_F g_i g_i' \) is the upper left \( p \times p \) submatrix of \( V_F \), which implies that \( \lambda_{\text{max}}(V_F) \geq \lambda_{\text{max}}(E_F g_i g_i') \), and \( \lambda_{\text{max}}(E_F g_i g_i') \geq \delta \) 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 \( \mathcal{F} \subset \mathcal{F}_{WU} \) for \( \delta_1 \) sufficiently small and \( M_1 \) sufficiently large because \( \mathcal{F}_{WU} \subset \mathcal{F} \) by the definition of \( \mathcal{F}_{WU} \). We need to show that the four conditions in the definition of \( \mathcal{F}_{WU} \) in (16.12) hold.

(I) We show that \( \inf_{F \in \mathcal{F}} \lambda_{\text{min}}(W_F) > 0 \), where \( W_F := W_1(\Omega_F) := \Omega_F^{-1/2} := (E_F g_i g_i')^{-1/2} \) (by (16.5), (16.8), and (16.11)). The inequality \( E_F ||g_i||^{2+\gamma} \leq M \) in \( \mathcal{F} \) implies \( \lambda_{\text{min}}(W_F) \geq \delta_1 \) for \( \delta_1 \) sufficiently small (because the latter holds if \( \lambda_{\text{max}}(W_F^{-2}) \leq \delta_1^{-2} \) and \( W_F^{-2} = \Omega_F = E_F g_i g_i' \).)

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

(III) We show that \( \inf_{F \in \mathcal{F}} \lambda_{\text{min}}(U_F) > 0 \), where in the present case \( U_F := U_1(\Omega_F) := ((\theta_0, I_p') (\Sigma_{F}^{\epsilon})^{-1} (\theta_0, I_p'))^{1/2} \) and \( \Sigma_{F} := \Sigma(\Omega_F, R_F) \) has \( (j, \ell) \) element equal to \( \text{tr}(R_{ij\ell} \Omega_F^{-1/2}) / k \) (by (16.8)). We have \( \sup_{F \in \mathcal{F}} ||R_F|| = \sup_{F \in \mathcal{F}} ||(B' \otimes I_k) \text{Var}_F(f_i) (B \otimes I_k)|| < \infty \) (where the inequality uses the condition \( E_F ||(g_i', \text{vec}(G_i'))^{(2+\gamma)} \leq M \) in \( \mathcal{F} \)). In addition, \( \inf_{F \in \mathcal{F}} \lambda_{\text{min}}(\Omega_F) > 0 \) (by the last condition in \( \mathcal{F} \)). The latter results imply that \( \sup_{F \in \mathcal{F}} ||\Sigma_{F}|| < \infty \) (because \( \Sigma_{F} \) 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 \( \sup_{F \in \mathcal{F}} ||\Sigma_{F}|| < \infty \). In addition, \( \Sigma_{F} \) is nonsingular \( \forall F \in \mathcal{F} \) (because \( \inf_{F \in \mathcal{F}} \lambda_{\text{min}}(\Sigma_{F}) > 0 \) by the proof of result (IV) below). The last two results imply the desired result \( \inf_{F \in \mathcal{F}} \lambda_{\text{min}}(U_F) = \lambda_{\text{min}}(\Sigma_{F}^{\epsilon})^{-1} \lambda_{\text{min}}(\Sigma_{F}^{\epsilon})^{-1} (\theta_0, I_p')^{1/2} > 0 \) (because \( A := (\theta_0, I_p) \) in \( R^{p \times (p+1)} \) has full row rank \( p \) and \( \lambda_{\text{min}}(U_F) = \inf_{A \in R^p \cdot ||A||=1} \lambda' A (\Sigma_{F}^{\epsilon})^{-1} A' \lambda \geq \inf_{A \in R^p \cdot ||A||=1} \lambda (A' \lambda)' (\Sigma_{F}^{\epsilon})^{-1} (A' \lambda)' ||A' \lambda||^2 \times \inf_{A \in R^p \cdot ||A||=1} ||A' \lambda||^2 = \lambda_{\text{min}}(\Sigma_{F}^{\epsilon})^{-1} \lambda_{\text{min}}(A A') \geq \delta_2 \) 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 \( \Sigma_F \) and \((\Omega_F, R_F)\) in place of \((\Omega, R)\) and \((\Omega, R)\), respectively, we have \( \inf_{F \in \mathcal{F}_P} \lambda_{\max}(\Sigma_F) > 0 \). The latter implies \( \inf_{F \in \mathcal{F}_P} \lambda_{\min}(\Sigma_F) > 0 \) by Lemma \( \overline{22.1} \)(b). In turn, the latter implies the desired result \( \sup_{F \in \mathcal{F}_P} ||U_F|| = \sup_{F \in \mathcal{F}_P} ||(\theta_0, I_p) \times (\Sigma_F)^{-1}(\theta_0, I_p)'||^{1/2} < \infty \).

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 \( \Lambda_{WU,P} \) with \( h_t = h_{5,g} \) by the same argument as for part (a)(i).

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

\[
\tilde{V}_n = n^{-1} \sum_{i=1}^{n} (u_i^* u_i^{*\prime} \otimes Z_i Z_i') - n^{-1} \sum_{i=1}^{n} (\tilde{u}_i^* \tilde{u}_i^{*\prime} \otimes Z_i Z_i') - n^{-1} \sum_{i=1}^{n} (u_i^* \tilde{u}_i^* \otimes Z_i Z_i') + n^{-1} \sum_{i=1}^{n} (\tilde{u}_i^* \tilde{u}_i^{*\prime} \otimes Z_i Z_i').
\]

(27.78)

We have

\[
\begin{align*}
n^{-1} \sum_{i=1}^{n} (u_i^* u_i^{*\prime} \otimes Z_i Z_i') &= E_{F_n} f_i f_i' + o_p(1), \\
\tilde{\Xi}_n &= (n^{-1} Z_{n \times k} Z_{n \times k})^{-1} n^{-1} Z_{n \times k} U^* = (E_{F_n} Z_i Z_i')^{-1} E_{F_n} Z_i u_i^{*\prime} + o_p(1) \\
&= (E_{F_n} Z_i Z_i')^{-1} E_{F_n} (g_i, G_i) + o_p(1) =: \Xi_{F_n} + o_p(1), \\
n^{-1} \sum_{i=1}^{n} (\tilde{u}_i^* \tilde{u}_i^{*\prime} \otimes Z_i Z_i') &= n^{-1} \sum_{i=1}^{n} (\tilde{\Xi}_n Z_i u_i^{*\prime} \otimes Z_i Z_i') = E_{F_n} (\Xi_{F_n}(g_i, G_i) \otimes Z_i Z_i') + o_p(1), \text{ and} \\
n^{-1} \sum_{i=1}^{n} (\tilde{u}_i^* \tilde{u}_i^{*\prime} \otimes Z_i Z_i') &= n^{-1} \sum_{i=1}^{n} (\tilde{\Xi}_n Z_i Z_i' \tilde{\Xi}_n \otimes Z_i Z_i') = E_{F_n} (\Xi_{F_n}(g_i, G_i) \otimes Z_i Z_i') + o_p(1),
\end{align*}
\]

where the first line holds by the WLLN’s (since \( u_i^* u_i^{*\prime} \otimes Z_i Z_i' = f_i f_i' \) for \( f_i \) defined in (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 \( Z_i u_i^{*\prime} = (g_i, G_i) \), the fourth line holds by the WLLN’s (using \( E_F(||(g_i, G_i)|| \cdot ||Z_i||^{2+\gamma/4}) \leq (E_F||g_i, G_i||)^2+\gamma/2 E_F||Z_i||^{1+\gamma} \)) 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 \( \tilde{V}_F \)) with \( F = F_n \), (27.78), and (27.79) combine to give

\[
\tilde{V}_n - \tilde{V}_{F_n} \rightarrow_{p} 0.
\]
Using the definitions of $\tilde{R}_n$ and $\tilde{R}_F$ (in (15.5) and (16.10), (27.71), (27.80), and $h_T := \lim W_{2F_n} = \lim \Omega_{F_n}$ yield

$$
(\tilde{\Omega}_n, \tilde{R}_n) \to_p \lim(\Omega_{F_n}, \tilde{R}_{F_n}) := h_8. 
$$

(27.81)

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

Assumption WU(c) holds for the parameter space $\Lambda_{\mathcal{U}, P}$, 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 $\inf F \in \mathcal{F} \lambda_{\max}(\tilde{R}_F) > 0$. 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 (27.7) and the paragraph following it), but with (i) $\tilde{R}_F$ in place of $R_F$ and (ii) $\inf F \in \mathcal{F} \lambda_{\max}(\tilde{V}_F) > 0$, rather than $\inf F \in \mathcal{F} \lambda_{\max}(V_F) > 0$, holding. Condition (ii) holds because $\inf F \in \mathcal{F} \lambda_{\max}(\tilde{V}_F) \geq \inf F \in \mathcal{F} \lambda_{\max}(E_F g_i g_i') > 0$ because $\tilde{V}_F$ can be written as $E_F(u_i^* - \Xi_i^* Z_i^*)(u_i^* - \Xi_i^* Z_i^*)' \otimes Z_i Z_i^*$, the first element of $Z_i^* Z_i$ is zero (because $\Xi := (E_F Z_i Z_i^*)^{-1} E_F (g_i, G_i)$, see (16.10), and $E_F g_i = 0$), the first element of $u_i^* - \Xi_i^* Z_i = u_i$ (because $u_i^* = (u_i, u_i')'$), the upper left $k \times k$ submatrix of $\tilde{V}_F$ equals $E_F u_i^* Z_i Z_i^* = E_F g_i g_i'$, 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 $\mathcal{F}_{\mathcal{W}, \mathcal{U}}$ in (16.12) hold for all $F \in \mathcal{F}_P$, for some $\delta_1$ 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 $U_F := U_1(U_2 F) := ((\theta_0, I_p)(\Sigma^x(\Omega_F, \tilde{R}_F))^{-1}(\theta_0, I_p)' \otimes (\theta_0, I_p)' \otimes (\theta_0, I_p)' / k$ (by (16.11)). The inequalities $E_F \|Z_i\|^{4+\gamma} \leq M$, $E_F \|g_i, vec(G_i)\|^{2+\gamma} \leq M$, and $\lambda_{\min}(E_F Z_i Z_i^*) \geq \delta$ imply that $\sup F \in \mathcal{F}_P \|\Xi F\| + \|E_F f_i f_i'\| + \|E_F(\Xi F Z_i Z_i^* \Xi F \otimes Z_i Z_i^*)\| + \|E_F(g_i, G_i)\Xi F \otimes Z_i Z_i^*)\| < \infty$, where $\Xi_F$ is defined in (16.10) (using the Cauchy-Bunyakovsky-Schwarz inequality). This, in turn, implies that $\sup F \in \mathcal{F}_P \|\bar{V}_F\| < \infty$, $\sup F \in \mathcal{F}_P \|\tilde{R}_F\| < \infty$, $\sup F \in \mathcal{F}_P \|\tilde{S}_F\| < \infty$, $\sup F \in \mathcal{F}_P \|\tilde{S}_P\| < \infty$, and $\lambda_{\min}(\tilde{L}_F) \geq \delta_2$ for some $\delta_2 > 0$, where $\tilde{V}_F$ and $\tilde{R}_F$ are defined in (16.10), $\tilde{S}_F := \Sigma(\Omega_F, \tilde{R}_F)$, $\tilde{L}_F := (\theta_0, I_p)(\tilde{S}_F)^{-1}(\theta_0, I_p)'$, and $(\tilde{S}_F)^{-1}$ exists by (IV) below (and $\lambda_{\min}(\tilde{L}_F) \geq \delta_2$ holds because $A := (\theta_0, I_p) \in \mathbb{R}^{p \times (p+1)}$ has full rank $p$ and $\lambda_{\min}(\tilde{L}_F) := \inf_{\lambda \in \mathcal{R}}(\|A\|_1 = 1, A' A(\tilde{S}_F)^{-1} A' A \geq \inf_{\lambda \in \mathcal{R}}(\|A' A\|_1 = 1, A' A \|_{L_2}^2 \|A' A\|_2 \times \inf_{\lambda \in \mathcal{R}}(\|A' A\|_1 = 1, A' A \|_{L_2}^2 \|A' A\|_2 = \lambda_{\min}(\tilde{S}_F)^{-1}) \lambda_{\min}(A A') \geq \delta_2$ for some $\delta_2 > 0$ that does not depend on $F$). Finally, $\lambda_{\min}(\tilde{L}_F) \geq \delta_2$ implies the desired result that $\lambda_{\min}(U_F) \geq \delta_1$ for some $\delta_1 > 0$ (because $U_F := \tilde{L}_F^{1/2}$).

(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 $\tilde{R}_F$ in place of...
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

$$\lambda = \lambda_F := E_F g_i g_i^T, \ h_n(\lambda) := \lambda, \text{ and } \Lambda := \{\lambda : \lambda = \lambda_F \text{ for some } F \in F_{AR}\}. \quad (27.82)$$

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

$$\lambda_{F_n} \to h := \lim E_{F_n} g_i g_i^T. \quad (27.83)$$

The $k \times k$ matrix $h$ is pd because $\lambda_{\min}(E_{F_n} g_i g_i^T) \geq \delta > 0$ for all $n \geq 1$ (by the last condition in $F_{AR}$) and $\lim \lambda_{\min}(E_{F_n} g_i g_i^T) = \lambda_{\min}(h)$ (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 $\lambda_{F_n}$ that satisfies $\lambda_{F_n} \to h$, and uniformly bounded $2 + \gamma$ moments, we have

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

We have

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

where the equality holds by definition of $\tilde{\Omega}_n$ in \[4.1\], 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
\[ AR_n := n g_n^2 \tilde{\sigma}_n^{-1} g_n \rightarrow_d Z' h^{1/2} h^{-1} Z = Z' Z \sim \chi^2_k. \] (27.86)

In turn, (27.86) gives
\[ P_{F_n}(AR_n > \chi^2_{k,1-\alpha}) \rightarrow P(Z' Z > \chi^2_{k,1-\alpha}) = \alpha. \] (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 \( \square \)

### 28 Proofs of Theorems 7.1 and 15.3

Suppose \( k \geq p \). Let \( A_F \) and \( \Pi_1F \) be defined as in (3.4) and (3.5) and the paragraph following these equations with \( \theta = \theta_0 \). Define \( \lambda^*_F, \Lambda^* \), and \( \{ \lambda^*_{nh} : n \geq 1 \} \) as \( \lambda_F, \Lambda_{WU} \), and \( \{ \lambda_{nh} : n \geq 1 \} \), respectively, are defined in (16.16)-(16.18), but with \( g_i \) and \( G_i \) replaced by \( g_i^* \) and \( G_i^* := \Pi_1F^{-1/2} A'_FG_i \), with \( F \) replaced by \( F^{SR} \), 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 \( g_i^* \) and \( G_i^* \). In addition, we restrict \( \{ \lambda^*_{nh} : n \geq 1 \} \) to be a sequence for which \( \lambda_{\min}(E_{Fn}g_i^*g_i^*) > 0 \) for all \( n \geq 1 \). Let \( (s^*_{1F_n}, \ldots, s^*_{pF_n}) \) denote the singular values of \( E_FG_i^* \). Under these conditions, \( A_{F_n} = A_{F_n}^\Omega, \Pi_1F_n = \Pi_{F_n}, W_{F_n} := (\Pi_1F_n^{-1/2} A'_{F_n} \Omega_{F_n} A_{F_n} \Pi_1F_n^{-1/2})^{-1} = I_k, \) and \( n^{1/2} s^*_{pF_n} \rightarrow \infty \) iff \( n^{1/2} s^*_{pF_n} \rightarrow \infty \).

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

(a) \( SR-QLRn = QLRn + o_p(1) = LM_n + o_p(1) = LM_n^{GM2} + o_p(1) \) and

(b) \( c_{k,p}(n^{1/2} \tilde{D}_n, 1-\alpha) \rightarrow_p \chi^2_{k,1-\alpha} \).

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

(a) \( SR-QLRp_n = QLRp_n + o_p(1) = LM_n + o_p(1) = LM_n^{GM2} + o_p(1) \) and

(b) \( c_{k,p}(n^{1/2} \tilde{D}_n, 1-\alpha) \rightarrow_p \chi^2_{k,1-\alpha} \).

The proofs of Theorems 7.1 and 15.3 use the following Lemma that concerns the \( QLR_{WU,n} \) statistic, which is based on general weight matrices \( \tilde{W}_n \) and \( \tilde{U}_n \), see (16.3), and considers sequences of distributions \( F \) in \( F \) or \( F_P \), rather than sequences in \( F^{SR} \) or \( F^{SR}_P \). 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 \( \Lambda_{WU} \) and \( \Lambda_{WU,P} \) for the CQLR and CQLRP tests, respectively, that is given in Lemma 27.4 in Section 27.

For the weight matrix \( \hat{W}_n \in R^{k \times k} \), Kleibergen’s LM statistic and the standard GMM LM statistic are defined by

\[
LM_n(\hat{W}_n) := n\hat{g}_n^{*} \hat{\Omega}_n^{-1/2} P_{\hat{\omega}_n, \hat{\omega}_n} \hat{\Omega}_n^{-1/2} \hat{g}_n \quad \text{and} \quad LM_n^{GMM}(\hat{W}_n) := n\hat{g}_n^{*} \hat{\Omega}_n^{-1/2} P_{\hat{\omega}_n, \hat{\omega}_n} \hat{\Omega}_n^{-1/2} \hat{g}_n, \tag{28.1}
\]

respectively, where \( \hat{G}_n \) is the sample Jacobian defined in (4.1) with \( \theta = \theta_0 \). In Lemma 28.1, we show that when \( n^{1/2} T_{pF_n} \to \infty \), the QLR_{WU,n} statistic is asymptotically equivalent to the \( LM_n(\hat{W}_n) \) and \( LM_n^{GMM}(\hat{W}_n) \) statistics.

The condition \( n^{1/2} T_{pF_n} \to \infty \) corresponds to strong or semi-strong identification in the present context. This holds because, for \( F \in \mathcal{F}_{WU} \), the smallest and largest singular values of \( W_F(E_F G_i) U_F \) (i.e., \( \tau_{\min(k,p)} \) and \( \tau_{1F} \)) are related to those of \( \Omega_{F}^{-1/2} E_F G_i \), denoted (as in Section 6.2 of AG2) by \( s_{\min(k,p)} \) and \( s_{1F} \), via \( c_1 s_{jF} \leq \tau_{jF} \leq c_2 s_{jF} \) for \( j = \min\{k,p\} \) and \( j = 1 \) for some constants \( 0 < c_1 < c_2 < \infty \). This result uses the condition \( \lambda_{\min}(\Omega_F) \geq \delta > 0 \) in \( \mathcal{F}_{WU} \). (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 \( \tau_{jF_n} \) in their definitions for \( j = 1, p \).

**Lemma 28.1** Suppose \( k \geq p \) and Assumption WU holds for some non-empty parameter space \( \Lambda_\ast \subset \Lambda_{WU} \). Under all sequences \( \{\lambda_{n,h} : n \geq 1\} \) with \( \lambda_{n,h} \in \Lambda_\ast \) for which \( n^{1/2} T_{pF_n} \to \infty \), we have

(a) \( QLR_{WU,n} = LM_n(\hat{W}_n) + o_p(1) = LM_n^{GMM}(\hat{W}_n) + o_p(1) \) and

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

**Comment:** The choice of the weight matrix \( \hat{U}_n \), that appears in the definition of the \( QLR_{WU,n} \) statistic, defined in (16.3), does not affect the asymptotic distribution of \( QLR_{WU,n} \) statistic under strong or semi-strong identification. This holds because \( QLR_{WU,n} \) is within \( o_p(1) \) of LM statistics that project onto the matrices \( \hat{W}_n \hat{D}_n \hat{U}_n \) and \( \hat{W}_n \hat{G}_n \hat{U}_n \), but such statistics do not depend on \( \hat{U}_n \) because \( P_{\hat{W}_n, \hat{D}_n, \hat{U}_n} = P_{\hat{W}_n, \hat{D}_n} \) and \( P_{\hat{W}_n, \hat{G}_n, \hat{U}_n} = P_{\hat{W}_n, \hat{G}_n} \) when \( \hat{U}_n \) 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 \( \hat{U}_n \).

**Proofs of Theorem 7.1 of AG2 and Theorem 15.3** By the second last paragraph of Section 5.2, \( SR-QLR_n(\theta_0) = QLR_n(\theta_0) \) wp→1 under any sequence \( \{F_n \in \mathcal{F}_{SR} : n \geq 1\} \) with \( r_{F_n}(\theta_0) = k \) for \( n \) large. By the same argument as given there, \( SR-QLRP_n(\theta_0) = QLRP_n(\theta_0) \) wp→1 under any
sequence \( \{ F_n \in \mathcal{F}^S_R : n \geq 1 \} \) 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 \( \lambda_{\min}(E_{F_n}g_i^g_t') > 0 \) for all \( n \geq 1 \) (see the paragraphs preceding Theorems 7.1 and 15.3).

Assumption WU for the parameter spaces \( \Lambda_{WU} \) and \( \Lambda_{WU,P} \) 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 \( QL_{Rn} = LM_n(\hat{\Omega}_n^{-1/2}) + \alpha_p(1) = LM_n^{GMM}(\hat{\Omega}_n^{-1/2}) + \alpha_p(1) \) and likewise for \( QL_{RnP} \), where \( QL_{Rn} \) and \( QL_{RnP} \) are defined in (5.7) and in the paragraph containing (15.7), respectively, and \( LM_n(\hat{\Omega}_n^{-1/2}) \) and \( LM_n^{GMM}(\hat{\Omega}_n^{-1/2}) \) are defined in (28.1) with \( \hat{W}_n = \hat{\Omega}_n^{-1/2} \). In addition, Lemma 28.1 implies that \( c_k(\sigma_{1/2}\hat{D}_n^*1 - \alpha) \rightarrow_p \chi^2_{p,1-\alpha} \) and \( c_k(\sigma_{1/2}\hat{D}_n^*1 - \alpha) \rightarrow_p \chi^2_{p,1-\alpha} \).

Note that all of these results are for sequences of distributions \( F \) in \( \mathcal{F} \) or \( \mathcal{F}_P \), not \( \mathcal{F}^{SR} \) or \( \mathcal{F}^{SR}_P \).

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 \( g_{F_i} := \Pi_{1/2}A'_Fg_i \) and \( G_{F_i} := \Pi_{1/2}A'_Fg_i \) in place of \( g_i \) and \( G_i \) to the \( QL_{Rn} \) and \( QL_{RnP} \) test statistics and their corresponding critical values. We have \( n^{1/2}s^*_{pF_n} \rightarrow \infty \iff n^{1/2}\tau^*_{pF_n} \rightarrow \infty \), where \( s^*_{pF} \) denotes the smallest singular value of \( E_FG^*_F \), and \( \tau^*_{pF} \) is defined to be the smallest singular value of \( (E_Fg_{F_i}^*g_{F_i}^*)_F^{-1/2}(E_FG^*_F)_F \). In consequence, the condition \( n^{1/2}\tau^*_{pF_n} \rightarrow \infty \) holds by Lemma 28.1 for the transformed variables \( g_{F_n,i} \) and \( G_{F_n,i} \), i.e., \( n^{1/2}\tau^*_{pF_n} \rightarrow \infty \). In the present case, \( \{ \Pi_{1/2}A'_F : n \geq 1 \} \) are nonsingular \( k \times k \) matrices by the assumption that \( \lambda_{\min}(E_{F_i}g_i^g_t') > 0 \) 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 \( QL_{Rn} \) and \( QL_{RnP} \) test statistics and their corresponding critical values are exactly the same when based on \( g_{F_i} \) and \( G_{F_i} \) as when based on \( g_i \) and \( G_i \). By the definitions of \( \mathcal{F}^{SR} \) and \( \mathcal{F}^{SR}_P \), the transformed variables \( g_{F_i}^* \) and \( G_{F_i}^* \) satisfy the conditions in \( \mathcal{F} \) and \( \mathcal{F}_P \), see (16.6) and (17.7). In particular, \( E_Fg_{F_i}^*g_{F_i}^* = \Pi_k \) and \( \lambda_{\min}(E_FZ_{F_i}^*Z_{F_i}^*) \geq 1/(2c) > 0 \), where \( Z_{F_i} := \Pi_{1/2}A'_FZ_i \), and \( c \) is as in the definition of \( \mathcal{F}^{SR}_P \) in (15.3). In addition, the \( LM_n \) and \( LM_n^{GMM} \) statistics are exactly the same when based on \( g_{F_i}^* \) and \( G_{F_i}^* \) as when based on \( g_i \) and \( G_i \). (This holds because, for any \( k \times k \) nonsingular matrix \( M \), such as \( M = \Pi_{1/2}A'_F \), we have \( LM_n := (\hat{O}_n^{-1})^{-1}\hat{D}n[\hat{D}'(\hat{O}_n^{-1})\hat{D}n]^{-1}(\hat{D}n^{-1})^{-1} = n_{\hat{g}_n}'M'(\hat{O}_nM')^{-1}1\hat{D}n[\hat{D}'M'(\hat{O}_nM')^{-1}M\hat{D}n]^{-1}\hat{D}'M'(\hat{O}_nM')^{-1}\hat{g}_n \) and likewise for \( LM_n^{GMM} \).) Using these results, the version of Lemma 28.1 described in the previous paragraph applied to the transformed variables \( g_{F_i}^* \) and \( G_{F_i}^* \) 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 \( \Box \)

Proof of Lemma 28.1. We start by proving the first result of part (a) of the lemma. We have \( n^{1/2}\tau^*_{pF_n} \rightarrow \infty \iff q = p \) (by the definition of \( q \) in (16.22)). Hence, by assumption, \( q = p \). Given this, \( Q_{2n}^2(\kappa) \) (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

\[
0 = |Q_{2n}^- \hat{\nu}^{(p+1)}_{n+1}| = |M_{n,p+1-q}^+ \hat{\nu}^{(p+1)}_{n} (1 + o_p(1))| \quad \text{and, hence,}
\]

\[
\hat{\nu}^{(p+1)}_{n+1} = M_{n,p+1-q}^+ (1 + o_p(1))
\]

\[
= (n^{1/2} D_{n,p+1-q}^1 U_n^1 D_n^1 W_n^1 h_{3,k-q} h_{3,k-q}^t (n^{1/2} W_n D_n^1 U_n^1 D_n^1 W_n^1) (1 + o_p(1)) + o_p(1)
\]

\[
= (n^{1/2} g_n^t \hat{\nu} \hat{\nu}^t h_{3,k-q} h_{3,k-q} (n^{1/2} W_n \hat{\nu} \hat{\nu}^t W_n^1) (1 + o_p(1)) + o_p(1)
\]

\[
= n g_n^t \hat{\nu} \hat{\nu}^t h_{3,k-q} h_{3,k-q}^t \hat{\nu} \hat{\nu}^t + o_p(1),
\]

(28.2)

where \( \hat{\nu}^{(p+1)}_{n+1} \) 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 \( n^{1/2} W_n \hat{\nu} \hat{\nu}^t h_{3,k-q} h_{3,k-q}^t (n^{1/2} W_n D_n^1 U_n^1 D_n^1 W_n^1) (1 + o_p(1)) + o_p(1) \), i.e., \( n^{1/2} W_n D_n^1 U_n^1 D_n^1 W_n^1 B_{n,p-q} \) drops out, as noted near the end of the proof of Theorem 16.6 and the last equality holds because \( W_n \hat{\nu} \hat{\nu}^t W_n^1 = I_k + o_p(1) \) by Assumption WU and \( n^{1/2} \hat{\nu} \hat{\nu}^t \hat{\nu} \hat{\nu}^t = O_p(1) \).

Next, we have

\[
QLR_{WU,n} := AR_n - \lambda_{\min}(n \hat{\nu} \hat{\nu}^t W_n)
\]

\[
= AR_n - \hat{\nu}^{(p+1)}_{n+1}
\]

\[
= n g_n^t \hat{\nu} \hat{\nu}^t (I_k - h_{3,k-q} h_{3,k-q}^t) \hat{\nu} \hat{\nu}^t + o_p(1)
\]

\[
= n g_n^t \hat{\nu} \hat{\nu}^t h_{3,q} h_{3,k-q}^t \hat{\nu} \hat{\nu}^t + o_p(1),
\]

(28.3)

where the first equality holds by the definition of \( QLR_{WU,n} \) in (16.3), the second equality holds by the definition of \( \hat{\nu}^{(p+1)}_{n+1} \) in (26.2), the third equality holds by (28.2) and the definition \( AR_n := n g_n^t \hat{\nu} \hat{\nu}^t \hat{\nu} \hat{\nu}^t \) in (4.2), and the last equality holds because \( h_3 = (h_{3,q}, h_{3,k-q}) \) is a \( k \times k \) orthogonal matrix.

When \( q = p \), by Lemma 16.4, we have

\[
n^{1/2} W_n \hat{\nu} \hat{\nu}^t D_n^1 U_n^1 T_n \rightarrow_d \Sigma_h = h_{3,d} \quad \text{and so}
\]

\[
n^{1/2} \hat{\nu} \hat{\nu}^t D_n^1 U_n^1 T_n \rightarrow_p h_{3,q},
\]

(28.4)

where the equality holds by the definition of \( \Sigma_h \) in (16.24) when \( q = p \) and the second convergence uses \( W_n \hat{\nu} \hat{\nu}^t W_n^1 = I_k + o_p(1) \) by Assumption WU. In consequence,

\[
P_{W_n} n_d = P_n \rightarrow_p h_{3,q} + o_p(1) = h_{3,q} h_{3,q}^t + o_p(1)
\]

\[
QLR_{WU,n} = LM_n(\hat{\nu} \hat{\nu}^t) + o_p(1),
\]

(28.5)

Electronic copy available at: https://ssrn.com/abstract=3366443
where the first equality holds because \( n^{1/2}U_n T_n \) is nonsingular \( \text{wp} \to 1 \) 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), \( n^{1/2} \hat{\Omega}_n^{-1/2} g_n = O_p(1) \), and
the definition of \( LM_n(\hat{W}_n) \) in (28.1).

As in (25.5) in Section 25 with \( \hat{G}_n \) in place of \( \hat{D}_n \), we have

\[
W_n \hat{G}_n U_n B_{n,q} \Sigma_{n,q}^{-1} = W_n D_n U_n B_{n,q} \Sigma_{n,q}^{-1} + W_n n^{1/2}(\hat{G}_n - D_n) U_n B_{n,q} (n^{1/2} \Sigma_{n,q})^{-1}
\]

\[
= C_{n,q} + o_p(1) \rightarrow_p h_{3,q},
\]

(28.6)

where \( D_n := E_{F_n} G_i \), the second equality uses (among other things) \( n^{1/2} \tau_j F_n \to \infty \) for all \( j \leq q \)
(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 \( QLR_{W,n} = LM_n^{GM}(\hat{W}_n) + o_p(1) \) is
the same as that given for \( QLR_{W,n} = LM_n(\hat{W}_n) + o_p(1) \). This completes the proof of part (a) of
Lemma 28.1.

By (27.10) in the proof of Theorem 27.1 we have

\[
c_{k,p}(n^{1/2} \hat{W}_n, \hat{D}_n, \hat{U}_n, 1 - \alpha) \rightarrow_d c_{k,p,q}(\tau_q, 1 - \alpha) \text{ and}
\]

\[
c_{k,p,q}(\tau_q, 1 - \alpha) = \chi^2_{p,1-\alpha} \text{ when } q = p,
\]

(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. \( \square \)

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 \( \{ \pi_n \in \Pi : n \geq 1 \} \) and any \( p \geq 1 \), we have: (a) \( \hat{R}_n \rightarrow_p \Sigma_V \otimes K_Z \), (b) \( \hat{\Omega}_n \rightarrow_p (b_0' \Sigma_V b_0) K_Z \), where \( b_0 := (1, -\theta_0)' \), (c) \( \hat{\Sigma}_n \rightarrow_p (b_0' \Sigma_V b_0)^{-1} \Sigma_V \), (d) \( \hat{\Sigma}^e_n \rightarrow_p (b_0' \Sigma_V b_0)^{-1} \Sigma_V \), (e) \( n^{1/2} \hat{\Omega}_n^{-1/2} \hat{g}_n = \hat{S}_n + o_p(1) \), and (f) \( n^{1/2} \hat{D}_n^* = -(I_k + o_p(1)) T_n (I_p + o_p(1)) + o_p(1) \).

In this section, we suppress the dependence of various quantities on \( \theta_0 \) for notational simplicity.
Thus, \( g_i := g_i(\theta_0) \), \( G_i := G_i(\theta_0) = (G_{i1}, \ldots, G_{ip}) \in \mathbb{R}^{k \times p} \), and similarly for \( \hat{g}_n, \hat{G}_n, f_i, B, \hat{D}_n, \hat{\Gamma}_{jn}, \hat{\Omega}_n, \hat{R}_n, \hat{D}_n^* \), and \( \hat{L}_n \).
The proof of Lemma 19.1 uses the following lemmas. Define

\[ A_0^* := \Sigma_V B \left( b_0^0 \Sigma_V \alpha_0, \ldots, b_0^p \Sigma_V \varphi_{p+1} \alpha_0 \right) \in R^{(p+1) \times p}, \quad B := \begin{pmatrix} 1 & 0_p' \\ -\theta_0 & -I_p \end{pmatrix} \in R^{(p+1) \times (p+1)}, \]

\[ c_0 := (b_0^0 \Sigma_V b_0)^{-1}, \quad b_0 := (1, -\theta_0)', \quad (\Sigma_V, \ldots, \Sigma_{V_{p+1}}) := \Sigma_V \in R^{(p+1) \times (p+1)}, \quad \text{and} \]

\[ L_{V_0} := (\theta_0, I_p) \Sigma_{V_{p+1}}^{-1}(\theta_0, I_p)' \in R^{p \times p}. \]

(29.1)

As defined in (19.4), \( A_0 := (\theta_0, I_p)' \in R^{(p+1) \times p}. \)

**Lemma 29.1** \( A_0^* L_{V_0} = -A_0. \)

**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 \( L = L_{V_0} \) that solves \( A_0^* L = -A_0 \).

**Lemma 29.2** Suppose Assumption HLIV holds. Under \( H_0 \), we have (a) \( n^{1/2} \tilde{g}_n \to_d N(0, b_0^0 \Sigma_V b_0 \cdot K_Z) \), (b) \( n^{-1} \sum_{i=1}^n (G_{ij} g_i - E G_{ij} g_i') = o_p(1) \) \( \forall j \leq p \), (c) \( \tilde{G}_n = O_p(1) \), (d) \( n^{-1} \sum_{i=1}^n (g_i g_i' - E g_i g_i') = o_p(1) \), and (e) \( \tilde{G}_n - n^{-1} \sum_{i=1}^n E G_i = O_p(n^{-1/2}) \).

**Proof of Lemma 19.1** To prove part (a), we determine the probability limit of \( \tilde{V}_n \) defined in (15.5). By (15.5) and (19.1)-(19.3), in the linear IV regression model with reduced-form parameter \( \pi_n \), we have

\[ u_i := u_i(\theta_0) = y_{i1} - Y_{2i} \theta_0, \quad E u_i = 0, \quad u_{\theta_i} = -Y_{2i} = -\pi_i' Z_i - V_{2i}, \quad E u_{\theta_i} = -\pi_i' Z_i, \]

\[ u_i^* := \begin{pmatrix} u_i \\ u_{\theta_i} \end{pmatrix} = \begin{pmatrix} u_i \\ -Y_{2i} \end{pmatrix} = \Xi'_n Z_i + \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix}, \quad \text{where} \quad \Xi_n = (0^k, -\pi_n) \in R^{k \times (p+1)}, \]

\[ E u_i^* = \Xi'_n Z_i, \quad u_i^* - E u_i^* = \begin{pmatrix} u_i \\ -V_{2i} \end{pmatrix} = B' \tilde{V}_i, \quad \tilde{u}_n^* - E u_i^* = (\tilde{\Xi}_n - \Xi_n)' Z_i, \quad \text{and} \]

\[ U^* := (u_1^*, \ldots, u_n^*)' = Z_{n \times k} \Xi_n + VB, \quad \text{where} \quad V := (V_1, \ldots, V_n)' \in R^{n \times (p+1)} \]

(29.2) and \( B := B(\theta_0) \) is defined in (15.5).

Next, we have

\[ \tilde{\Xi}_n - \Xi_n = (Z_{n \times k} Z_{n \times k})^{-1} Z_{n \times k} U^* - \Xi_n = (n^{-1} Z_{n \times k} Z_{n \times k})^{-1} n^{-1} Z_{n \times k} V B = O_p(n^{-1/2}), \]

(29.3)
where the first equality holds by the definition of $\tilde{\Xi}_n$ in (15.5), the second equality uses the last line of (29.2), and the third equality holds by Assumption HLIV(c) (specifically, $n^{-1}Z_{n \times k}^\prime Z_{n \times k} \to K_Z$ and $K_Z$ is pd) and by $n^{-1/2}Z_{n \times k}^\prime V = O_p(1)$ (which holds because $EZ_{n \times k}^\prime V = 0$ and the variance of the $(j, \ell)$ element of $n^{-1/2}Z_{n \times k}^\prime V$ 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 $\tilde{V}_n$ in (15.5) and simple algebra, we have

$$
\tilde{V}_n := n^{-1} \sum_{i=1}^n \left[ (u_i^* - \hat{u}_{in}^*) (u_i^* - \hat{u}_{in}^*)^\prime \otimes Z_i Z_i^\prime \right] \quad (29.4)
$$

$$
= n^{-1} \sum_{i=1}^n \left[ (u_i^* - Eu_i^*) (u_i^* - Eu_i^*)^\prime \otimes Z_i Z_i^\prime \right] - n^{-1} \sum_{i=1}^n \left[ (\hat{u}_{in}^* - Eu_i^*) (u_i^* - Eu_i^*)^\prime \otimes Z_i Z_i^\prime \right]
- n^{-1} \sum_{i=1}^n \left[ (u_i^* - Eu_i^*) (\hat{u}_{in}^* - Eu_i^*)^\prime \otimes Z_i Z_i^\prime \right] + n^{-1} \sum_{i=1}^n \left[ (\hat{u}_{in}^* - Eu_i^*) (\hat{u}_{in}^* - Eu_i^*)^\prime \otimes Z_i Z_i^\prime \right].
$$

Using the third line of (29.2), the fourth summand on the rhs of (29.4) equals

$$
n^{-1} \sum_{i=1}^n \left[ (\tilde{\Xi}_n - \Xi_n)^\prime Z_i Z_i^\prime (\tilde{\Xi}_n - \Xi_n) \otimes Z_i Z_i^\prime \right]. \quad (29.5)
$$

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 $n^{-1} \sum_{i=1}^n ||Z_i||^4 \leq n^{-1} \sum_{i=1}^n ||Z_i||^4 1(||Z_i|| > 1) + 1 \leq n^{-1} \sum_{i=1}^n ||Z_i||^6 + 1 = o(n)$ 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

$$
n^{-1} \sum_{i=1}^n \left[ (\tilde{\Xi}_n - \Xi_n)^\prime Z_i V_i^\prime B \otimes Z_i Z_i^\prime \right]. \quad (29.6)
$$

The elements of the second summand on the rhs of (29.4) are each $o_p(1)$ because $\tilde{\Xi}_n - \Xi_n = O_p(n^{-1/2})$ by (29.3) and for any $j_1, j_2, j_3 \leq k$ and $\ell \leq p$ we have $n^{-1} \sum_{i=1}^n Z_{ij_1} Z_{ij_2} Z_{ij_3} V_{i\ell} = o_p(n^{1/2})$ 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

\[
\tilde{V}_n = n^{-1} \sum_{i=1}^{n} [B'V_i V_i' B \otimes Z_i'] + o_p(1)
\]

\[
= n^{-1} \sum_{i=1}^{n} [(B'V_i V_i' B - B'\Sigma_V B) \otimes Z_i'] + \left[ B'\Sigma_V B \otimes n^{-1} \sum_{i=1}^{n} Z_i Z_i' \right] + o_p(1)
\]

\[
\to_p B'\Sigma_V B \otimes K_Z, \tag{29.7}
\]

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, the third line of (29.2), the second equality holds by adding and subtracting the same quantity, and where the first equality holds using (29.4), the argument in the two paragraphs following (29.4), and holds by the calculations following (29.5)).

Equation (29.7) gives

\[
\tilde{R}_n := (B' \otimes I_k) \tilde{V}_n (B \otimes I_k) \to_p \Sigma_V \otimes K_Z \tag{29.8}
\]

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

To prove part (b), we have

\[
\hat{\Omega}_n := n^{-1} \sum_{i=1}^{n} g_i g_i' - \hat{g}_n \hat{g}_n' = n^{-1} \sum_{i=1}^{n} E g_i g_i' + n^{-1} \sum_{i=1}^{n} (g_i g_i' - E g_i g_i') + O_p(n^{-1})
\]

\[
= n^{-1} \sum_{i=1}^{n} Z_i Z_i' E u_i^2 + o_p(1) \to_p (b_0' \Sigma_V b_0) K_Z, \tag{29.9}
\]

where the first equality holds by the definition in (4.1), second equality uses \(n^{1/2} \hat{g}_n = O_p(1) \) by Lemma 29.2(a), the third equality holds by Lemma 29.2(d), and the convergence holds by Assumption HLIV(c) and because \(E u_i^2 = E (V_i' b_0)^2 = b_0' \Sigma_V b_0 \) by Assumption HLIV(b).

Part (c) holds because

\[
\tilde{\Sigma}_{jfn} = \text{tr}(\tilde{R}_{jfn} \hat{\Omega}_n^{-1})/k \to_p \text{tr}(\Sigma_{V,jf} K_Z (b_0' \Sigma_V b_0)^{-1} K_Z^{-1})/k = \Sigma_{V,jf} (b_0' \Sigma_V b_0)^{-1}, \tag{29.10}
\]

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

Part (d) holds because \(\tilde{\Sigma}_n \to_p ((b_0' \Sigma_V b_0)^{-1} \Sigma_V)^{\varepsilon} \) by part (c) of the lemma and Lemma 22.1(e),

109
\[ ((b_0^\prime \Sigma V b_0)^{-1} \Sigma V)^\varepsilon = (b_0^\prime \Sigma V b_0)^{-1} \Sigma V^\varepsilon \] 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

\[
n^{-1} Z'_{n \times k} Y = \left( n^{-1} \sum_{i=1}^{n} Z_i(y_{1i} - Y_{2i}^0 \theta_0) + n^{-1} \sum_{i=1}^{n} Z_i Y_{2i}^0 \theta_0, n^{-1} \sum_{i=1}^{n} Z_i Y_{2i} \right) \]
\[
= (\hat{g}_n - \tilde{G}_n \theta_0, -\tilde{G}_n) = (\hat{g}_n, \tilde{G}_n) \begin{pmatrix} 1 & 0' \\ -\theta_0 & -I_p \end{pmatrix} = (\hat{g}_n, \tilde{G}_n) B, \quad (29.11)
\]

where the expressions for \( \hat{g}_n \) and \( \tilde{G}_n \) use (19.3). Using (29.11) and the definition of \( L_{V0} \) in (29.1), the statistic \( T_n \) defined in (19.4) can be written as

\[
T_n := (Z'_{n \times k} Z_{n \times k})^{-1/2} Z'_{n \times k} Y \Sigma V^{-1} A_0 (A_0' \Sigma V^{-1} A_0)^{-1/2} A_0 L_{V0}^{-1/2}. \quad (29.12)
\]

Note that, using the definitions of \( B \) and \( L_{V0} \) in (29.1) and \( A_0 \) in (19.4), the rhs expression for \( T_n \) equals the expression in (19.4).

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

\[
\hat{\Gamma}_n := n^{-1} \sum_{i=1}^{n} (G_{ij} - \hat{G}_{jn}) g^i = n^{-1} \sum_{i=1}^{n} E G_{ij} g^i + n^{-1} \sum_{i=1}^{n} (G_{ij} g^i - E G_{ij} g^i) - \hat{G}_{jn} \hat{g}_n
\]
\[
= n^{-1} \sum_{i=1}^{n} E G_{ij} g^i + o_p(1) = -n^{-1} \sum_{i=1}^{n} E Z_i Y_{2ij} Z'_i u_i + o_p(1)
\]
\[
= -n^{-1} \sum_{i=1}^{n} Z_i Z'_i E V_{2ij} V'_i b_0 + n^{-1} \sum_{i=1}^{n} Z_i Z'_i (Z'_i \pi_{jn}) E u_i + o_p(1)
\]
\[
= -n^{-1} \sum_{i=1}^{n} Z_i Z'_i \Sigma_{V,j+1} b_0 + o_p(1), \quad (29.13)
\]

where \( g_i = Z_i(y_{1i} - Y_{2i}^0 \theta_0) = Z_i u_i \) 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 \( \Sigma V := (\Sigma V_1, ..., \Sigma V_{p+1}) := EV_i V'_i \).  

110

Electronic copy available at: https://ssrn.com/abstract=3366443
Equations (29.9) and (29.13) give

\[
\hat{D}_{jn} := \hat{G}_{jn} - \bar{\Gamma}_{jn} \hat{\Omega}_n^{-1} \hat{g}_n = \hat{G}_{jn} + \Sigma_{V,j+1} b_0 (b'_0 \Sigma_V b_0)^{-1} \hat{g}_n + o_p(n^{-1/2}) \quad \text{and}
\]

\[
\hat{D}_n := (\hat{D}_{1n}, \ldots, \hat{D}_{pn}) = (\hat{g}_n, \hat{G}_n) \left( \begin{array}{c} \Sigma_{V,2} b_0 c_0, \ldots, \Sigma_{V,p+1} b_0 c_0 \\ I_p \end{array} \right) + o_p(n^{-1/2})
\]

\[
= (\hat{g}_n, \hat{G}_n) B \Sigma_V^{-1} \left( \Sigma_V B \left( \begin{array}{c} \Sigma_{V,2} b_0 c_0, \ldots, \Sigma_{V,p+1} b_0 c_0 \\ I_p \end{array} \right) \right) + o_p(n^{-1/2})
\]

\[
= (\hat{g}_n, \hat{G}_n) B \Sigma_V^{-1} A_0^* + o_p(n^{-1/2}),
\]

(29.14)

where the second equality on the first line uses \( \hat{g}_n = O_p(n^{-1/2}) \) 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

\[
n^{1/2} \tilde{D}_n^* := n^{1/2} \tilde{\Omega}_n^{-1/2} \tilde{D}_n \tilde{\Omega}_n^{1/2} = (b'_0 \Sigma_V b_0)^{-1/2} (I_k + o_p(1))(n^{-1} Z_{n \times k}^o Z_{n \times k})^{-1/2} n^{1/2} (\hat{g}_n, \hat{G}_n) B \Sigma_V^{-1} A_0^* 
\]

\[
\times (b'_0 \Sigma_V b_0)^{1/2} L_{V_0}^{-1/2} (I_p + o_p(1)) + o_p(1)
\]

\[
= - (I_k + o_p(1))(n^{-1} Z_{n \times k}^o Z_{n \times k})^{-1/2} n^{1/2} (\hat{g}_n, \hat{G}_n) B \Sigma_V^{-1} A_0 L_{V_0}^{-1/2} (I_p + o_p(1)) + o_p(1)
\]

\[
= - (I_k + o_p(1)) \bar{T}_n (I_p + o_p(1)) + o_p(1),
\]

(29.15)

where the first equality holds by the definition of \( \tilde{D}_n^* \) in (5.7), the second equality holds by (29.14), \( \tilde{\Omega}_n \to_p (b'_0 \Sigma_V b_0) K \) (which holds by part (b) of the lemma), and \( \tilde{L}_n := (\theta_0, I_p) \tilde{\Sigma}_n^{-1}(\theta_0, I_p)' \to_p (b'_0 \Sigma_V b_0)L_{V_0} \) (which holds because \( \tilde{\Sigma}_n \to_p (b'_0 \Sigma_V b_0)^{-1} \Sigma_V \) by part (d) of the lemma), for \( L_{V_0} := (\theta_0, I_p) \tilde{\Sigma}_n^{-1}(\theta_0, I_p)' \) 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 \( \tilde{S}_n \) satisfies

\[
\tilde{S}_n := (Z_{n \times k}^o Z_{n \times k})^{-1/2} Z_{n \times k}^o V b_0 (b'_0 \Sigma_V b_0)^{-1/2} = \Gamma_{jn}^{-1/2} \sum_{i=1}^n Z_i Z_i' \hat{g}_n (b'_0 \Sigma_V b_0)^{-1/2}
\]

\[
= n^{1/2} \tilde{\Omega}_n^{-1/2} \hat{g}_n + o_p(1),
\]

(29.16)

where the first equality holds by the definition of \( \tilde{S}_n \) in (19.4), the second equality holds because \( Y_i' b_0 = u_i \), and the third equality holds by (29.9) and \( n^{1/2} \tilde{g}_n = O_p(1) \) by Lemma 29.2(a). This proves part (e). □
**Proof of Lemma 29.1** By pre-multiplying by $B\Sigma_V^{-1}$, the equation $A_0^* L_{V0} = -A_0$ is seen to be equivalent to

$$
\begin{pmatrix}
  b_0' \Sigma_V c_0, \ldots, b_0' \Sigma_V^{-1} c_0 \\
  I_p
\end{pmatrix} L_{V0} = -B \Sigma_V^{-1} \begin{pmatrix}
  \theta_0' \\
  I_p
\end{pmatrix} = \begin{pmatrix}
  -1 & 0^p' \\
  \theta_0 & I_p
\end{pmatrix} \Sigma_V^{-1} \begin{pmatrix}
  \theta_0' \\
  I_p
\end{pmatrix} . \tag{29.17}
$$

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

$$(b_0' \Sigma_V c_0, \ldots, b_0' \Sigma_V^{-1} c_0)(\theta_0, I_p) \Sigma_V^{-1} (\theta_0, I_p)' = (-1, 0^p') \Sigma_V^{-1} (\theta_0, I_p)' . \tag{29.19}$$

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

$$
(b_0' \Sigma_{V2} \theta_0 c_0, b_0' \Sigma_{V2} c_0, \ldots, b_0' \Sigma_{Vp+1} c_0) \Sigma_V^{-1} (\theta_0, I_p)' \\
= ((-b_0' \Sigma_{V} b_0 + b_0' \Sigma_{V1}) c_0, b_0' \Sigma_{V2} c_0, \ldots, b_0' \Sigma_{Vp+1} c_0) \Sigma_V^{-1} (\theta_0, I_p)' \\
= (-1 + b_0' \Sigma_{V1} c_0, b_0' \Sigma_{V2} c_0, \ldots, b_0' \Sigma_{Vp+1} c_0) \Sigma_V^{-1} (\theta_0, I_p)' , \tag{29.20}
$$

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

$$
(b_0' \Sigma_{V1} c_0, \ldots, b_0' \Sigma_{Vp+1} c_0) \Sigma_V^{-1} (\theta_0, I_p)' = c_0' b_0' \Sigma_{V2} \Sigma_V^{-1} \begin{pmatrix}
  \theta_0' \\
  I_p
\end{pmatrix} = 0' , \tag{29.21}
$$

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, Thm. 3) and the Cramér-Wold device under Assumptions IIIIV(a)-(c) because $\frac{1}{2} \hat{g}_n = n^{-1} \sum_{i=1}^{n} Z_i u_i$ is an average of i.i.d. mean-zero finite-variance random variables $u_i$ with nonrandom weights $Z_i$.  

112
To show part (b), we write

\[ n^{-1} \sum_{i=1}^{n} (G_{ij}g_i' - EG_{ij}g_i') = n^{-1} \sum_{i=1}^{n} Z_i Z_i'(Y_{2ij}u_i - EY_{2ij}u_i) \]

\[ = -n^{-1} \sum_{i=1}^{n} Z_i Z_i'(Z_i \pi_{j,n})u_i - n^{-1} \sum_{i=1}^{n} Z_i Z_i'(V_{2ij}u_i - \Sigma'_{V_{j+1}b_0}), \]

where the first equality holds because \( g_i = Z_i u_i \) and \( G_{ij} = -Z_i Y_{2ij} \), the second equality holds because \( Y_{2ij} = Z_i \pi_{j,n} + V_{2ij} \) and \( EV_{2ij}u_i = EV_{2ij}V_i'b_0 = \Sigma'_{V_{j+1}b_0} \). Both summands on the rhs have mean zero. The \((\ell, \ell_2)\) element of the first summand has variance equal to

\[ n^{-2} \sum_{i=1}^{n} (Z_\ell_i Z_{\ell_i2} Z_i' \pi_{j,n})^2 \times Var(u_i), \]

which converges to zero for all \( \ell, \ell_2 \leq k \) because \( n^{-1} \sum_{i=1}^{n} ||Z_i||^6 = o(n) \), \( Var(u_i) = b_0' \Sigma V b_0 < \infty \), and \( \sup_{j,p,n} ||\pi_{j,n}|| < \infty \) by Assumption HLIV(b)-(d). The \((\ell, \ell_2)\) element of the second summand has variance equal to

\[ n^{-2} \sum_{i=1}^{n} Z_\ell^2 Z_{\ell_2}^2 Var(V_{2ij}u_i), \]

which converges to zero for all \( \ell, \ell_2 \leq k \) because \( n^{-1} \sum_{i=1}^{n} ||Z_i||^6 = o(n) \) and \( Var(V_{2ij}u_i) \leq E(V_{2ij}V_i'b_0)^2 \leq b_0' b_0 E||V_i||^4 < \infty \) by Assumptions HLIV(b)-(c). This establishes part (b).

For part (c), we have

\[ \hat{G}_n = -n^{-1} \sum_{i=1}^{n} Z_i Y_{2i} = -n^{-1} \sum_{i=1}^{n} Z_i Z_i' \pi_n - n^{-1} \sum_{i=1}^{n} Z_i V_{2i}. \]

(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 \((\ell, j)\) element for \( \ell \leq k \) and \( j \leq p \) has variance \( n^{-2} \sum_{i=1}^{n} Z_{\ell}^2 \Sigma_{V_{j^{*}j^{*}}} \) where \( \Sigma_{V_{j^{*}j^{*}}} < \infty \) is the \((j^{*}, j^{*})\) element of \( \Sigma_{V} \) and \( j^{*} = j + 1 \), and

\[ n^{-1} \sum_{i=1}^{n} Z_{\ell_2}^2 \Sigma_{V_{j^{*}j^{*}}} \rightarrow K_{Z\ell \ell'} \Sigma_{V_{j^{*}j^{*}}}, \]

where \( K_{Z\ell \ell} < \infty \) is the \((\ell, \ell)\) element of \( K_Z \). Hence, the rhs is \( O_p(1) \), which establishes part (c).

To prove part (d), we have

\[ n^{-1} \sum_{i=1}^{n} (g_i g_i' - Eg_i g_i') = n^{-1} \sum_{i=1}^{n} Z_i Z_i'(u_i^2 - Eu_i^2) \rightarrow_p 0, \]

(29.24)

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

Part (e) holds by the following argument:

\[ \hat{G}_n - n^{-1} \sum_{i=1}^{n} EG_i = -n^{-1} \sum_{i=1}^{n} Z_i (Y_{2i} - EY_{2i}) = -n^{-1} \sum_{i=1}^{n} Z_i V_{2i} = O_p(n^{-1/2}), \]

(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) $\hat{R}_n \rightarrow_p R(\pi_*)$, (b) $\hat{\Sigma}_n \rightarrow_p (b_0^2 \Sigma_V b_0)^{-1} \Sigma_{V*}$, (c) $\hat{\Sigma}_n \rightarrow_p (b_0^2 \Sigma_V b_0)^{-1} \Sigma_{V*}$, and (d) $n^{1/2} \hat{D}_n = -(I_k + o_p(1))T_n(LV_0^{-1/2}LV_0^{1/2} + o_p(1)) + o_p(1)$, where $LV_0 := (\theta_0, I_p)\Sigma_V^{-1}(\theta_0, I_p)' \in R^{p \times p}$ and $LV_*: := (\theta_0, I_p)\Sigma_{V*}^{-1}(\theta_0, I_p)' \in R^{p \times p}$.

Proof of Lemma 19.2. To prove part (a), we determine the probability limit of $\hat{V}_n$ defined in (5.3), where $f_i = (Z_i' u_i, -vec(Z_i Y_{2i}'))'$ by (19.1) and (19.3). For $\zeta_n(\pi)$ defined in (19.6), we can write

$$\zeta_n(\pi_n) = n^{-1} \sum_{i=1}^{n} Z_{ni}^* Z_{ni}', \text{ where}$$

$$Z_{ni}^* := vec \left( Z_i Z_i' \pi_n - n^{-1} \sum_{\ell=1}^{n} Z_i Z_i' \pi_n \right) = (\pi_n' \otimes Z_i) Z_i - n^{-1} \sum_{\ell=1}^{n} (\pi_n' \otimes Z_\ell) Z_\ell \in R^{kp}$$

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

We have

$$\hat{V}_n := n^{-1} \sum_{i=1}^{n} \left( f_i - n^{-1} \sum_{\ell=1}^{n} Ef_\ell \right) \left( f_i - n^{-1} \sum_{\ell=1}^{n} Ef_\ell \right)' - \left( \hat{f}_n - n^{-1} \sum_{\ell=1}^{n} Ef_\ell \right) \left( \hat{f}_n - n^{-1} \sum_{\ell=1}^{n} Ef_\ell \right)'$$

$$= n^{-1} \sum_{i=1}^{n} \left( Z_i u_i \right) \left( -vec(Z_i V_{2i}') - Z_{ni}^* \right) \left( Z_i u_i \right)' + o_p(1)$$

$$= n^{-1} \sum_{i=1}^{n} \left( u_i \otimes Z_i Z_i' \right) + \left( 0_{kp \times kp} \right) + o_p(1)$$

$$+ n^{-1} \sum_{i=1}^{n} \left( Z_i u_i \right) \left( -vec(Z_i V_{2i}') \right) \left( 0_{kp} \right)' + o_p(1)$$

$$= \left( \left( 1 - \theta_0' \right) \Sigma_V \left( 1 - \theta_0' \right) \right)' \otimes \left( n^{-1} \sum_{i=1}^{n} Z_i Z_i' \right) + \left( 0_{kp \times kp} \right) + o_p(1)$$

$$= (B'\Sigma_V B) \otimes \left( n^{-1} \sum_{i=1}^{n} Z_i Z_i' \right) + \left( 0_{kp \times kp} \right) + o_p(1), \quad (29.26)$$

where the second equality holds using $E u_i = 0$, $EV_{2i} = 0_p$, $Y_{2i} = \pi_n' Z_i + V_{2i}$, $vec(Z_i Y_{2i}') = vec(Z_i V_{2i}') + Z_{ni}$, and Lemma 29.2 (a) and (e) because $\hat{f}_n - n^{-1} \sum_{\ell=1}^{n} Ef_\ell = (\hat{G}_n, vec(\hat{G}_n - n^{-1} \sum_{\ell=1}^{n} EG_\ell'))'$, 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 \( u_i = (1, -\theta'_{0j})V_i \), (ii) the variance of each element of the lhs matrix is \( o(1) \) because \( E||V_i||^4 < \infty \) and \( n^{-1}\sum_{i=1}^{n}||Z_i||^4 = o(n) \) by Assumption HLIV(b)-(c) (because \( n^{-1}\sum_{i=1}^{n}||Z_i||^4 \leq n^{-1}\sum_{i=1}^{n}||Z_i||^1(||Z_i|| > 1) + 1 \leq n^{-1}\sum_{i=1}^{n}||Z_i||^6 + 1 = o(n) \) using Assumption HLIV(c)), (iii) \( \zeta_n(\pi_n) \to \zeta(\pi_s) \) 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 summand on the lhs of the fourth equality using \( \text{...} \).

Using the definitions of \( \tilde{R}_n \) in (5.3) and \( \rho(\pi_s) \) in (19.7), part (a) of the lemma follows from (29.27).

Next we prove part (b). We have

\[
\hat{\Sigma}_{j\ell n} = \text{tr}(\hat{R}^j_{j\ell n} \hat{\Sigma}_n^{-1})/k \to_p \text{tr}(R_{j\ell}(\pi_s)(b'_0 \Sigma V b_0)^{-1}K^{-1}_Z)/k =: (b'_0 \Sigma V b_0)^{-1} \Sigma_{V, j\ell}, \tag{29.28}
\]

where \( \hat{\Sigma}_{j\ell n} \) and \( \Sigma_{V, j\ell} \) denote the \((j, \ell)\) elements of \( \hat{\Sigma}_n \) and \( \Sigma_{V, *} \), respectively, \( \hat{R}^j_{j\ell n} \) and \( R_{j\ell}(\pi_s) \) denote the \((j, \ell)\) submatrices of dimension \( k \times k \) of \( \hat{R}_n \) and \( R(\pi_s) \), 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 \( \hat{\Sigma}^\varepsilon_n \to_p ((b'_0 \Sigma V b_0)^{-1} \Sigma_{V, *})^\varepsilon \), Lemma 22.1(d) implies that \( ((b'_0 \Sigma V b_0)^{-1} \Sigma_{V, *})^\varepsilon = (b'_0 \Sigma V b_0)^{-1} \Sigma_{V, *}, \) and Assumption HLIV2(c) implies that \( \Sigma_{V, *}^\varepsilon = \Sigma_{V, *} \).

To prove part (d), we have

\[
n^{1/2}\hat{D}^*_n := n^{1/2}\hat{\Omega}_n^{-1/2} \hat{D}_n \hat{L}^{1/2}_n
\]

where the first equality holds by the definition of \( \hat{D}^*_n \) in (5.7), the second equality holds by (i) (29.14), (ii) the result of part (c) of the lemma that \( \hat{\Sigma}^\varepsilon_n \to_p (b'_0 \Sigma V b_0)^{-1} \Sigma_{V, *}, \) (iii) the result of Lemma 19.1(b) that \( \hat{\Omega}_n \to_p (b'_0 \Sigma V b_0)K_Z, \) (iv) \( n^{-1}Z'_{n \times k}Z_{n \times k} \to K_Z \) by Assumption HLIV(c), (v)
\[ \hat{L}_n := (\theta_0, I_p)(\hat{\Sigma}_n^{-1})^{-1}(\theta_0, I_p)' \] as defined in (5.7) with \( \theta = \theta_0 \), and (vi) \( \hat{L}_n \rightarrow_p b_0'\Sigma_Vb_0L_V^* \) 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). \( \square \)

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 \( \{\pi_n \in \Pi : n \geq 1\} \), we have: (a) \( r_k \Pi(\theta_0) = T_n'[I_k + L_V^{-1/2} \zeta(\pi_n)K_Z^{-1/2} + o_p(1)]^{-1}T_n \cdot (1 + o_p(1)) + o_p(1) \), (b) \( r_k \Pi(\theta_0) = T_n'[T_n(L_Vb_0'\Sigma_Vb_0)^{-1} \cdot (1 + o_p(1)) + o_p(1) \), where \( L_V := (\theta_0, 1)'\Sigma_V^{-1}(\theta_0, 1)' \in R \), and (c) \( L_Vb_0'\Sigma_Vb_0 = \frac{(1-2\theta_0\rho+c_2)^2}{c^2(1-\rho^2)} \), where \( c^2 := \text{Var}(V_{2i})/\text{Var}(V_{1i}) > 0 \) and \( \rho = \text{Corr}(V_{1i}, V_{2i}) \in (-1, 1) \).

When \( p = 1 \), we write

\[ \Sigma_V := EV_{V_1'} := (\Sigma_{V_1}, \Sigma_{V_2}) := \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \in R^{2 \times 2} \quad (29.30) \]

for \( \Sigma_{V_1}, \Sigma_{V_2} \in R^2 \), 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) \( L_V = \frac{\sigma_1^2 - 2\theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} > 0 \), (b) \( b_0'\Sigma_Vb_0 = \sigma_1^2 - 2\theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2 \), and (c) \( L_Vb_0'\Sigma_Vb_0 - (b_0'\Sigma_Vb_0)^2(b_0'\Sigma_Vb_0)^{-1} = 1 \).

Proof of Lemma 19.3 We prove part (b) first. By (29.9) and (29.14),

\[
\begin{align*}
n^{1/2} \tilde{\Omega}_n^{-1/2} \tilde{D}_n & = n^{1/2}(I_k + o_p(1))(n^{-1} Z_{n \times k} Z_{n \times k})^{-1/2}(\tilde{g}_n, \tilde{G}_n)B \Sigma_V^{-1} A_0(b_0' \Sigma_Vb_0)^{-1/2} + o_p(1) \\
& = -n^{1/2}(I_k + o_p(1))(n^{-1} Z_{n \times k} Z_{n \times k})^{-1/2}(\tilde{g}_n, \tilde{G}_n)B \Sigma_V^{-1} A_0 L_V^{-1}(b_0' \Sigma_Vb_0)^{-1/2} + o_p(1) \\
& = - (I_k + o_p(1)) T_n (L_Vb_0' \Sigma_Vb_0)^{-1/2} + o_p(1),
\end{align*}
\]

where the second equality holds by Lemma 29.1 and the third equality holds by (29.12). Because \( \tilde{T}_n'(I_k + o_p(1)) \tilde{T}_n = \tilde{T}_n' \tilde{T}_n + o_p(1)||\tilde{T}_n||^2 \), the result of part (b) follows.
Next, we prove part (a). We have

\[ n^{-1} \sum_{i=1}^{n} (G_i - \hat{G}_n)(G_i - \hat{G}_n)' = n^{-1} \sum_{i=1}^{n} \left( G_i - n^{-1} \sum_{\ell=1}^{n} EG_{\ell} \right) \left( G_i - n^{-1} \sum_{\ell=1}^{n} EG_{\ell} \right)' - \left( \hat{G}_n - n^{-1} \sum_{i=1}^{n} EG_i \right) \left( \hat{G}_n - n^{-1} \sum_{i=1}^{n} EG_i \right)' \]

\[ = n^{-1} \sum_{i=1}^{n} \left( -Z_i'Z_i\pi_n - Z_iV_{2i} + n^{-1} \sum_{\ell=1}^{n} Z_{\ell}\pi_n \right) \left( -Z_i'Z_i\pi_n - Z_iV_{2i} + n^{-1} \sum_{\ell=1}^{n} Z_{\ell}\pi_n \right)' + o_p(1) \]

\[ = n^{-1} \sum_{i=1}^{n} (Z_iV_{2i})'(Z_iV_{2i})' + 2n^{-1} \sum_{i=1}^{n} (Z_iZ_i'\pi_n)(Z_iV_{2i})' - 2 \left( n^{-1} \sum_{\ell=1}^{n} Z_{\ell}\pi_n \right) \left( n^{-1} \sum_{\ell=1}^{n} Z_{\ell}V_{2i} \right) \]

\[ + \zeta_n(\pi_n) + o_p(1) \]

\[ = n^{-1} Z_n'Z_n\sigma_2^2 + \zeta_n(\pi_n) + o_p(1), \tag{29.32} \]

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 \( Y_{2i} - EV_{2i} = V_{2i}, \) 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 \( 4\sigma_2^2n^{-2} \sum_{i=1}^{n} ||Z_i||^6 \sup_{\pi \in \Pi} ||\pi||^2, \) which is \( o(1) \) by Assumption HLIV(c)-(d), and the third summand on the lhs of the fourth equality is \( o_p(1) \) because \( n^{-1} \sum_{\ell=1}^{n} Z_{\ell}Z_{\ell}'\pi_n = O(1) \) by Assumption HLIV(c) and (d) and \( n^{-1} \sum_{\ell=1}^{n} Z_{\ell}V_{2i} = o_p(1) \) by the argument following (29.23).

Combining (29.9), (29.13), (29.32) and the definition of \( \bar{V}_{Dn} \) in (19.14), we obtain

\[ \bar{V}_{Dn} = n^{-1} \sum_{i=1}^{n} Z_iZ_i'(\sigma_2^2 - (b'_0\Sigma V_2)^2(b'_0\Sigma V_0b_0)^{-1}) + \zeta_n(\pi_n) + o_p(1) \]

\[ = K_ZL_{V_0}^{-1} + \zeta_n(\pi_n) + o_p(1), \tag{29.33} \]

where the second equality holds by Lemma 29.3(c) and Assumption HLIV(c).

Next, we have

\[ n^{1/2} \left( n^{-1} Z_n'Z_n \right)^{-1/2} \bar{D}_nL_{V_0}^{-1/2} = n^{1/2} \left( n^{-1} Z_n'Z_n \right)^{-1/2} \left( \bar{g}_n, \bar{G}_n \right)B\Sigma_V^{-1}A_0^*L_{V_0}^{-1/2} + o_p(1) \]

\[ = -n^{1/2} \left( n^{-1} Z_n'Z_n \right)^{-1/2} \left( \bar{g}_n, \bar{G}_n \right)B\Sigma_V^{-1}A_0L_{V_0}^{-1/2} + o_p(1) = -\bar{T}_n + o_p(1), \tag{29.34} \]

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

\[ n^{1/2} \hat{D}_n - 1/2 \hat{D}_n = [K_Z L_{Y0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} n^{1/2} \hat{D}_n \]

\[ = -[K_Z L_{Y0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} (n^{-1} Z_{n \times k} Z_{n \times k})^{1/2} T_n L_{Y0}^{-1/2} + o_p(1) \]

\[ = -[K_Z L_{Y0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} K_Z^{-1/2} T_n L_{Y0}^{-1/2} (1 + o_p(1)) + o_p(1), \quad (29.35) \]

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

\[ r_{k1} := n^{1/2} \hat{D}_n - 1/2 \hat{D}_n = T_n K_Z^{-1/2} [K_Z L_{Y0}^{-1} + \zeta_n(\pi_n) + o_p(1)]^{-1/2} K_Z^{-1/2} T_n L_{Y0}^{-1/2} (1 + o_p(1)) + o_p(1) \]

\[ = T_n [I_k + L_{Y0} K_Z^{-1/2} \zeta_n(\pi_n) K_Z^{-1/2} + o_p(1)]^{-1} T_n (1 + o_p(1)) + o_p(1), \quad (29.36) \]

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 \( \sigma^2 = c^2 \sigma_1^2 \). □

**Proof of Lemma 29.3** Part (a) holds by the following calculations:

\[ L_{Y0} := (\theta_0, 1) \left( \begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right)^{-1} \left( \begin{array}{c} \theta_0 \\ 1 \end{array} \right) \quad (29.37) \]

\[ = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} (\theta_0, 1) \left( \begin{array}{cc} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{array} \right) \left( \begin{array}{c} \theta_0 \\ 1 \end{array} \right) = \frac{\sigma_1^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}. \]

We have \( L_{Y0} > 0 \) because \( \Sigma_V \) is pd by Assumption HLIV(b) and \( (\theta_0, 1) \neq 0_2 \).

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

\[ b_0' \Sigma_V b_0 := (1, -\theta_0) \left( \begin{array}{cc} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{array} \right) \left( \begin{array}{c} 1 \\ -\theta_0 \end{array} \right) = \sigma_1^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2 \]

\[ b_0' \Sigma_V 2 := (1, -\theta_0)(\rho \sigma_1 \sigma_2, \sigma_2') = \rho \sigma_1 \sigma_2 - \theta_0 \sigma_2^2. \quad (29.38) \]

Using (29.38), we obtain

\[ \sigma_2^2 - (b_0' \Sigma_V 2)^2 (b_0' \Sigma_V b_0)^{-1} = \sigma_2^2 - \frac{(\rho \sigma_1 \sigma_2 - \theta_0 \sigma_2^2)^2}{\sigma_1^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2} \quad (29.39) \]

\[ = \frac{\sigma_1^2 \sigma_2^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2 - (\rho \sigma_1 \sigma_2 - \theta_0 \sigma_2^2)^2}{\sigma_1^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2} = \frac{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}{\sigma_1^2 - 2 \theta_0 \rho \sigma_1 \sigma_2 + \theta_0^2 \sigma_2^2} = L_{Y0}^{-1}. \]

118
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 \(\mathcal{F}_{AR}, \mathcal{F},\) and \(\mathcal{F}_P\), rather than \(\mathcal{F}_{AR}^{SR}, \mathcal{F}^{SR},\) and \(\mathcal{F}_P^{SR}\). 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 (4.2) based on \(\{f_i - \hat{f}_n : i \leq n\}\), but with \(\hat{\Omega}_n\) defined in (18.3) and Assumption \(\Omega\) below, rather than in (4.1), and the critical value is \(\chi^2_{1-\alpha}\). The non-SR QLR time series test statistic and conditional critical value are defined as in Section 5.1, but with \(\tilde{V}_n\) and \(\tilde{\Omega}_n\) defined in (18.3) and Assumption V below based on \(\{f_i - \hat{f}_n : i \leq n\}\), in place of \(\hat{V}_n\) and \(\hat{\Omega}_n\) defined in (5.3) and (4.1), respectively. The non-SR QLR\(_P\) time series test statistic and conditional critical value are defined as in Section 15, but with \(\tilde{V}_n\) and \(\tilde{\Omega}_n\) defined in (18.3) and Assumption \(\text{V}_P\) below based on \(\{(u_i^* - \hat{u}_n^*) \otimes Z_i : i \leq n\}\), rather than in (15.5) and (4.1), 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

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

(30.1)

for some \(\gamma, \delta > 0, d > (2 + \gamma)/\gamma,\) and \(C,M < \infty,\) where \(\Omega_F\) is defined in (18.4). We define \(\mathcal{F}_{TS}\) and \(\mathcal{F}_{TS,P}\) as \(\mathcal{F}\) and \(\mathcal{F}_P\) are defined in (16.1), respectively, but with \(\mathcal{F}_{TS,AR}\) in place of \(\mathcal{F}_{AR}\). For CS’s, we use the corresponding parameter spaces \(\mathcal{F}_{TS,\Theta,AR} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,AR}(\theta_0), \theta_0 \in \Theta\}, \mathcal{F}_{TS,\Theta} := \{(F, \theta_0) : F \in \mathcal{F}_{TS}(\theta_0), \theta_0 \in \Theta\}, \text{ and } \mathcal{F}_{TS,\Theta,P} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,P}(\theta_0), \theta_0 \in \Theta\}, \)

where \(\mathcal{F}_{TS,AR}(\theta_0), \mathcal{F}_{TS}(\theta_0),\) and \(\mathcal{F}_{TS,P}(\theta_0)\) denote \(\mathcal{F}_{TS,AR}, \mathcal{F}_{TS},\) and \(\mathcal{F}_{TS,P},\) respectively, with their dependence on \(\theta_0\) made explicit.

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

**Assumption V:** \(\tilde{V}_n(\theta_0) - V_{F_n}(\theta_0) \rightarrow_p 0^{(p+1)k \times (p+1)k}\) under \(\{F_n : n \geq 1\}\) for any sequence \(\{F_n \in \mathcal{F}_{TS,P} : n \geq 1\}\) for which \(V_{F_n}(\theta_0) \rightarrow V\) for some matrix \(V\) whose upper left \(k \times k\) submatrix \(\Omega\) is pd.

**Assumption V-CS:** \(\tilde{V}_n(\theta_{0n}) - V_{F_n}(\theta_{0n}) \rightarrow_p 0^{(p+1)k \times (p+1)k}\) under \(\{(F_n, \theta_{0n}) : n \geq 1\}\) for any
sequence \( \{(F_n, \theta_{0n}) \in \mathcal{F}_{TS, \Theta, P} : n \geq 1\} \) for which \( V_{F_n}(\theta_{0n}) \rightarrow V \) for some matrix \( V \) whose upper left \( k \times k \) submatrix \( \Omega \) 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 \( \mathcal{F}_{TS, P} \) and \( \mathcal{F}_{TS, \Theta, P} \) in place of \( \mathcal{F}_{TS} \) and \( \mathcal{F}_{TS, \Theta} \).

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

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

**Theorem 30.1** Suppose the AR, CQLR, and CQLR\(_P\) tests are defined as above, the parameter spaces for \( F \) are \( \mathcal{F}_{TS, AR} \), \( \mathcal{F}_{TS} \), and \( \mathcal{F}_{TS, P} \), respectively (defined in the paragraph containing (30.1)), and the corresponding Assumption \( \Omega \), \( V \), 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\(_P\) CS’s for the parameter spaces \( \mathcal{F}_{TS, \Theta, AR} \), \( \mathcal{F}_{TS, \Theta} \), and \( \mathcal{F}_{TS, \Theta, P} \), respectively, provided the corresponding Assumption \( \Omega-CS \), \( V-CS \), or \( V_P-CS \) holds for each CS, rather than Assumption \( \Omega \), \( V \), or \( V_P \).

The proof of Theorem [18.1] uses Theorem [30.1] and the following lemma.

**Lemma 30.2** Suppose \( \{X_i : i = \ldots, 0, 1, \ldots\} \) is a strictly stationary sequence of mean zero, square integrable, strong mixing random variables. Then, \( \text{Var}(\overline{X}_n) = 0 \) 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 \( \hat{r}_n \leq r \) (\( = r_{F_n} \)) 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 \( \lambda' \tilde{\Omega}_n \lambda = n^{-1} \sum_{i=1}^n (\lambda' g_i)^2 - (\lambda' \tilde{g}_n)^2 = 0 \) 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 \( \lambda' \tilde{\Omega}_n \lambda = 0 \) a.s.\( [F_n] \) by Assumption SR-V(c), SR-V-CS(c), SR-V\(_P\)(c), SR-V\(_P\)-CS(c), SR-\( \Omega \)(c), or SR-\( \Omega-CS \)(c).”
(ii) in the proof of \( \widehat{\sigma}_n \geq r \) a.s. \( \forall n \geq 1 \) we have \( \Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2} \rightarrow_p I_r \), with \( \Pi_{1F_n} \) 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-\(\Omega\)(c), or SR-\(\Omega\)-CS(c), and

(iv) the results of Lemmas [5.1] and [15.1] which are used in the proof of Lemma [17.1] hold using the equivariance condition in Assumption SR-V(b), SR-V-CS(b), SR-V\(_P\)(b), SR-V\(_P\)-CS(b), SR-\(\Omega\)(b), or SR-\(\Omega\)-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, respectively) for the i.i.d. case, but with some modifications. The modifications are the first, second, third, and fifth modifications stated in the proof of Theorem 7.1 in AG1, which is given in Section 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, \(\Omega\), or \(\Omega\)-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 \( \text{Var}(X_n) = 0 \). Then, \( X_n \) 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, \( \sum_{i=1}^{n} X_{i+s} = 0 \) a.s. and \( \sum_{i=2}^{n+1} X_{i+s} = 0 \) a.s. for all integers \( s \geq 0 \). Taking differences yields \( X_{1+n+s} = X_{1+n+s} \) for all \( s \geq 0 \). That is, \( X_1 = X_{1+n} \) for all \( s \geq 1 \).

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

\[
\xi_s := |P(X_1 \in A, X_{1+n+s} \in A) - P(X_1 \in A)P(X_{1+n+s} \in A)| \leq \alpha_X(sn) \rightarrow 0 \text{ as } s \rightarrow \infty, \quad (30.2)
\]

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

\[
\xi_s = |P(X_1 \in A) - P(X_1 \in A)|^2 = P(X_1 \in A)(1 - P(X_1 \in A)), \quad (30.3)
\]
where the first equality holds because $X_1 = X_{1+m}$ a.s. and by strict stationarity. Because $\xi_s \to 0$ as $s \to \infty$ by (30.2) and $\xi_s$ 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 $O_F^e (E_F g_i d_i^t)^{-1/2} E_F G_i \in R^{(k-b) \times p}$ in the subvector case, where $O_F \in R^{k \times (k-b)}$, defined below, is such that $M_{(E_F g_i d_i^t)^{-1/2} E_F G_i} = O_F O_F^e$. 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 $W_n$ and $U_n$ that satisfy a certain high level condition Assumption WU. In particular, $W_n$ and $U_n$ 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 $W_n$ and $W_{F_n}$ are set equal to the identity matrix and therefore do not appear in the high level Assumption WUS, which adapts Assumption WU from the full vector test. We verify Assumption WUS for the specific choice of weighting matrix $U_n$ employed in the subvector CQLR test (9.11), which is $\tilde{U}_n = \tilde{L}_n^{1/2} (\theta_0, \tilde{\beta}_n)$, in Lemma 31.9 below.

A general QLRWU subvector test statistic is defined as

$$QLR_{WU,n}^S := AR_{n}^S(\theta_0, \tilde{\beta}_n) - \lambda_{\min}(\hat{Q}_{WU,n}^S),$$

$$\hat{Q}_{WU,n}^S := (\hat{\Omega}_{n}^{-1/2}(\tilde{\eta}) \hat{D}_{n}(\tilde{\eta}) \hat{U}_n, \hat{\Omega}_{n}^{-1/2}(\tilde{\eta}) \hat{g}_n(\tilde{\eta}))' M_{\hat{J}_n(\tilde{\eta})} (\hat{\Omega}_{n}^{-1/2}(\tilde{\eta}) \hat{D}_{n}(\tilde{\eta}) \hat{U}_n, \hat{\Omega}_{n}^{-1/2}(\tilde{\eta}) \hat{g}_n(\tilde{\eta}))$$

for $\tilde{\eta} := (\theta'_0, \tilde{\beta}'_n)'$, and $\hat{U}_n := U_1(\hat{U}_2n)$ 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 $\hat{U}_n$ is defined as in (16.5). The general CQLRWU test rejects the null hypothesis if

$$QLR_{WU,n}^S > c_{k,p}(n^{1/2} \hat{\Omega}_{n}^{-1/2}(\tilde{\eta}) \hat{D}_{n}(\tilde{\eta}) \hat{U}_n, \hat{J}_n(\tilde{\eta}), 1 - \alpha),$$

where $c_{k,p}(D, J, 1 - \alpha)$ is defined in (9.12).}

70The reason $\hat{\Omega}_{n}^{-1/2}$ is used in the definitions of $QLR_{n}^S(\theta, \tilde{\beta}_n)$ in (9.11) and $QLR_{WU,n}^S$, rather than $\tilde{\Omega}_{n}^{-1/2}$, is that
The proof for the subvector test result is based on working out the asymptotic null rejection probabilities along certain drifting sequences of parameters \(\{\lambda_{n,h}^S : n \geq 1\}\) that we introduce below (31.15). The notation involving \(\lambda\) and \(h\) in (16.16) and (16.19) for the full vector case has to be adapted to the subvector case. The argument \(\theta_0\) in the notation for expressions for full vector inference is replaced throughout by the argument \((\theta_0, \beta^*)\). For example, in \(\lambda_{4,F}^S = E_FG_i, G_i\) 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, \(\lambda_{n,h}^S\) contains several additional components, such as \(\lambda_{4,\theta_j,\beta,F}^S := E_FG_i\theta_j\beta\) for \(j = 1, \ldots, p\) and \(\lambda_{4,\beta_j,\beta,F}^S := E_FG_i\beta_j\beta\) for \(j = 1, \ldots, b\).

**Construction of bases \(O_{F_n}\) and \(\tilde{O}_{F_n}\) 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 \(\tilde{Q}_n(\theta_0, \beta_n)\) in (9.11), which is part of the test statistic \(QLR_n^S(\theta_0, \beta_n)\), it is helpful to factor \(M^{\tilde{X}_n(\eta)}\) into a product \(\tilde{O}_{F_n}\tilde{O}'_{F_n}\) where \(\tilde{O}_{F_n} \in R^{k \times (k-b)}\) contains a basis for the space of eigenvectors spanned by the eigenvalue 1 of the projection matrix \(M^{\tilde{X}_n(\eta)}\). Given this factorization, we consider the quantities \((\tilde{O}_{F_n}\tilde{\Omega}_n^{-1/2}(\eta)\tilde{g}_n(\eta), \tilde{O}'_{F_n}\tilde{D}_n^*(\eta))\), 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 \(O_n\) and \(O \in R^{k \times (k-b)}\) such that \(M^{J_n} = O_nO_n', M = OO'\), 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 \(span(J)^\perp\), the orthogonal complement of \(span(J)\). We have \(span(J)^\perp = N(J')\).

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 \(R^b\). Decompose \(J' = (J'_1, J'_2)\) with

\[
\begin{align*}
\text{we prove the subvector results using the proof of the full vector result with } & W_n, W_{F_n}, \text{ and } D_n \in R^{k \times p} \text{ replaced by } I_b, I_b, \text{ and } \tilde{O}'_{F_n}\tilde{\Omega}_n^{-1/2}(\theta_0, \beta_n)\tilde{D}_n(\theta_0, \beta_n) \in R^{(k-b) \times p}, \text{ respectively, where } \tilde{O}'_{F_n} \in R^{(k-b) \times k}, \text{ defined below, is such that } \\
\tilde{O}'_{F_n}\tilde{O}_{F_n} = M^{J_n(\theta_0, \beta_n)} \text{ for the full vector results, the difference between } \tilde{W}_n \text{ and } W_{F_n} \text{ can be handled easily because } \\
\tilde{W}_nW_{F_n}^{-1} = I_p \text{ as in (26.8). But, in the subvector case, the same strategy cannot be applied to } \tilde{\Omega}_n^{-1/2}(\theta_0, \beta_n) \text{ and } (E\tilde{F}_n, g, g')^{-1/2}, \text{ because of the factor } \tilde{O}'_{F_n} \text{ that precedes } \tilde{\Omega}_n^{-1/2}(\theta_0, \beta_n) \text{ in the definition of } \\
\tilde{O}'_{F_n}\tilde{\Omega}_n^{-1/2}(\theta_0, \beta_n)\tilde{D}_n(\theta_0, \beta_n), \text{ which is the subvector equivalent of } D_n. \\
\text{If that is not the case and the first } b \text{ columns do not form a basis, simply adapt the notation in what follows so that the } b \text{ columns of } J' \text{ that are referred to, do indeed form a basis of } R^b.
\end{align*}
\]
$J_1' \in R^{b\times b}$ 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 $(-j_s'^{-1}J_1, e_s') \in R^k$ for $s = 1, \ldots, k - b$, where $e_s$ denotes the $s$-th 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 $s$-th row is given by

$$(-j_s'^{-1}J_1, e_s') \tag{31.4}$$

for $s = 1, \ldots, k - b$. Define

$$O' = O(J)' := (Q'Q)^{-1/2}Q'. \tag{31.5}$$

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 $J_n' = (j_n', J_{n2})$, $J_n' = (j_{n1}, \ldots, j_{n(k-b)})$, the matrix $Q_n' \in R^{(k-b)\times k}$, whose $s$-th row is given by $(-j_n'^{-1}J_{n1}, e_s')$, and $O_n' = O(J_n)' := (Q_n'Q_n)^{-1/2}Q_n'$. Then, $O_nO_n' = M_{J_n}$, $OO' = M_J$, and $O_n' \rightarrow O'$ as desired, where the convergence follows directly from $J_n \rightarrow J$. 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

$$J_n = (E_{F_nG_n})^{-1/2}E_{F_n}G_{i\beta} \tag{31.8}$$

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 $J_h := (h_{5,4})^{-1/2}h_{4,\beta}$ defined below.
As in (31.5), for given $F \in \mathcal{F}^S$,

$$O'_F = O'_{Ft} = O((E_F g_i g'_i)^{-1/2} E_F G_{i\beta})'$$

(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 $\{\lambda_{n,h}^S \in \mathcal{A}^S : n \geq 1\}$, Lemma 31.5 below implies that $\tilde{J}_n(\theta_0, \hat{\beta}_n) - J_n \in R^{k \times b}$ converges in probability to zero and $J_n = (E_{F_n} g_i g'_i)^{-1/2} E_{F_n} G_{i\beta} \rightarrow J_n := (h_{5, g})^{-1/2} h_{4, \beta}$. In addition, $J_n$ has full column rank $b$ for all $n$ sufficiently large, under the restrictions in $\mathcal{F}^S$. Therefore, $\tilde{J}_n(\theta_0, \hat{\beta}_n)$ has full column rank $b$ wp→ 1. For any $b$ columns indexed by $t = 1, \ldots, T$ of $J'_n$ that form a basis of $R^b$ and apply the above construction with this choice of columns to both $\tilde{J}_n(\theta_0, \hat{\beta}_n)'$ and $J'_n$ to obtain

$$\tilde{O}'_{Ft} = O(\tilde{J}_n(\theta_0, \hat{\beta}_n))' \in R^{(k-b) \times k} \quad \text{and} \quad O'_{Ft} = O(J_n)'$$

(31.10)

using the notation in (31.5). Given that $\tilde{J}_n(\theta_0, \hat{\beta}_n) - J_n \rightarrow_p 0^{k \times b}$, it follows that $\tilde{O}'_{Ft} - O'_{Ft} \rightarrow_p 0^{(k-b) \times k}$.

**Definition of $\{\lambda_{n,h}^S \in \mathcal{A}^S : n \geq 1\}$**. 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 $O'_{Ft}(E_F g_i g'_i)^{-1/2}(E_F G_i)_{UF} \in R^{(k-b) \times p}$. More precisely, let $B_F = B_{Ft}$ denote a $p \times p$ orthogonal matrix of eigenvectors of

$$U_F'(E_F G_i)'(E_F g_i g'_i)^{-1/2} O_{Ft}(E_F g_i g'_i)^{-1/2} E_F G_i U_F$$

(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

$$O_{Ft}(E_F g_i g'_i)^{-1/2}(E_F G_i)_{UF} U_F'(E_F G_i)'(E_F g_i g'_i)^{-1/2} O_{Ft}.$$  

(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

$$O_{Ft}(E_F g_i g'_i)^{-1/2}(E_F G_i)_{UF},$$

(31.13)

which are nonnegative and ordered so that $\tau_{jFt}$ 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
\((\tau_{1F}, \ldots, \tau_{\min(k-b,p)F})\) to be a vector of minus ones and \(B_F\) and \(C_F\) 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 \(\lambda^S\) to be

\[
\lambda^S_{i,F} := (\tau_{1F}, \ldots, \tau_{\min(k-b,p)F})^T \in R^{\min(k-b,p)} , \\
\lambda^S_{2,F} := (B_F, \ldots, B_{FT})^T \in R^{p \times Tp} , \\
\lambda^S_{3,F} := (C_F, \ldots, C_{FT})^T \in R^{(k-b) \times T(k-b)} , \\
\lambda^S_{4,F} := E_F G_i \in R^{k \times p} , \\
\lambda^S_{3,3,F} := E_F G_i \beta \in R^{k \times b} , \\
\lambda^S_{4,\beta_j,F} := E_F G_i \beta_j \in R^{k \times b} \text{ for } j = 1, \ldots, p , \\
\lambda^S_{4,\beta_j,F} := E_F G_i \beta_j \in R^{k \times b} \text{ for } j = 1, \ldots, b , \\
\lambda^S_{5,\beta_j,F} := E_F \binom{g_i}{vec(G_i)} \left( \begin{array}{c} g_i \\ vec(G_i) \end{array} \right)^T \in R^{(p+1)k \times (p+1)k} , \\
\lambda^S_{5,\beta_j,F} := E_F \binom{g_i g_i^T}{vec(G_i) vec(G_i)} \frac{1}{vec(G_i)} \in R^{k \times k} \text{ for } j = 1, \ldots, b , \\
\lambda^S_{6,F} := \varphi_F (g_i, vec(G_i))^T , \\
\lambda^S_{10,F} := \varphi_F (g_i, vec(G_i))^T , \\
\lambda^S := \lambda^S_F := (\lambda^S_{1,F}, \ldots, \lambda^S_{10,F}) , \tag{31.14}
\]

where \(0/0 := 0\) for the components of \(\lambda^S_{6,F}\), and \(\lambda^S\) 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 \(j\)-th column of \(\lambda^S_{4,\beta_j,F} \in R^{k \times b}\) by \(\lambda^S_{4,\beta_j,F} \in R^k\). Let

\[
\Lambda^S := \{\lambda^S_F : F \in \mathcal{F}^S\} , \quad \text{and} \\
h_n(\lambda^S) := (n^{1/2} \lambda^S_{1,F} , \lambda^S_{2,F} , \lambda^S_{3,F} , \lambda^S_{4,F} , \lambda^S_{5,F} , \lambda^S_{6,F} , \lambda^S_{8,F} , \lambda^S_{10,F}) . \tag{31.15}
\]

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)\footnote{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.}. Denote by $h_{4,3}$, $h_{4,3}g_{3}$, $h_{4,3}g_{3}g_{3}$, and $h_{5,5}$, the limits of $\lambda_{4,3,F_{n}}^{S}$, $\lambda_{4,3}^{S}$, $\lambda_{4,3}^{S}G_{F_{n}}$, and $\lambda_{4,3}^{S}G_{F_{n}}$ under the sequence $\{\lambda_{n,h}^{S} : n \geq 1\}$, respectively, and analogously for other expressions, where by $\lambda_{4,3}^{S}G_{F_{n}}$ and $\lambda_{4,3}^{S}G_{F_{n}}$ we denote the lower left and lower right submatrices of $\lambda_{4,3}^{S}$ of dimensions $R^{b \times k}$ and $R^{p \times k}$.

Consider a sequence $\{\lambda_{n,h}^{S} : n \geq 1\}$ and let the distributions $\{F_{n} : n \geq 1\}$ correspond to $\{\lambda_{n,h}^{S} : n \geq 1\}$. Because under $\{\lambda_{n,h}^{S} : n \geq 1\}$, $(E_{F_{n}}g_{n}g_{n}^{t})^{-1/2}E_{F_{n}}G_{i\beta}^{t}$ converges to a full column rank matrix, there exists a smallest index $t^{*} \in \{1, ..., T\}$ such that for all $n$ sufficiently large the $b$ columns of $((E_{F_{n}}g_{n}g_{n}^{t})^{-1/2}E_{F_{n}}G_{i\beta})^{t}$ indexed by $t^{*}$ form a basis of $R^{b}$, and by definition of $\{\lambda_{n,h}^{S} : n \geq 1\}$, $n^{1/2}(t_{1,F_{n,t^{*}}}, ..., t_{\min\{k-b,p\},F_{n,t^{*}}}) \rightarrow (h_{1,1}, ..., h_{1,\min\{k-b,p\}}, t^{*})$. Note that $t^{*}$ depends on the sequence $\{\lambda_{n,h}^{S} \in \Lambda^{S} : n \geq 1\}$. We include $\tau_{1,F_{t}}$, $B_{F_{t}}$, and $C_{F_{t}}$ for all $t = 1, ..., T$ in the definition of $\lambda_{1,F_{t}}^{S}$, $\lambda_{2,F_{t}}^{S}$, and $\lambda_{3,F_{t}}^{S}$ in (31.14) because this ensures the convergence of $n^{1/2}\tau_{1,F_{n,t^{*}}}$, $B_{F_{n,t^{*}}}$, and $C_{F_{n,t^{*}}}$ 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 $q_{n}^{S} = q_{n}^{S} \in \{0, ..., \min\{k-b,p\}\}$ be such that

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_{n}^{S} \text{ and } h_{1,j} < \infty \text{ for } q_{n}^{S} + 1 \leq j \leq \min\{k-b,p\},$$

(31.16)

where $h_{1,j} := \lim n^{1/2}t_{j,F_{n}} \geq 0$ for $j = 1, ..., \min\{k-b,p\}$.

Define $F_{W_{U}}^{S}$ as $F_{W_{U}}$ in (16.12) with $F$ replaced by $F^{S}$ and $W$ replaced by $I_{k}$. Define $\Lambda_{W_{U}}^{S}$ as $\Lambda_{W_{U}}$ in (16.17) with $F_{W_{U}}$ replaced by $F_{W_{U}}^{S}$.

**Assumption WUS for the parameter space $\Lambda_{W_{U}}^{S}$:** Under all subsequences $\{w_{n}\}$ and all sequences $\{\lambda_{w_{n},h}^{S} : n \geq 1\}$ with $\lambda_{w_{n},h}^{S} \in \Lambda_{W_{U}}^{S}$,

(a) $\hat{U}_{2w_{n}} \rightarrow_{p} h_{8} := \lim U_{2F_{w_{n}}}$( and

(b) $U_{1}(\cdot)$ is a continuous function at $h_{8}$ on some set $\mathcal{U}_{2}$ that contains $\{\lambda_{8,F}^{S} (= U_{2F}) : \lambda^{S} \in \Lambda_{W_{U}}^{S}\}$ and contains $\hat{U}_{2w_{n}}$ wp→1.

As in (16.23), let (and recall again that we leave out the index $t^{*}$ from the notation)

$$S_{n} := \text{Diag}\{n^{1/2}\tau_{1,F_{n}}, ..., (n^{1/2}\tau_{q_{n}^{S},F_{n}})^{-1}, 1, ..., 1\} \in R^{p \times p} \text{ and } T_{n} := B_{F_{n}}S_{n} \in R^{p \times p}. \quad (31.17)$$

The random function $CLR_{k,p}(D,J)$ in (9.12) that generates the conditional critical value of the CLR subvector test can be expressed as follows. Suppose $M_{J} = OO'$, for $O$ defined in (31.5).
Then, we can write

$$CLR_{k,p}(D, J) := Z' M J Z - \lambda_{\min}((Z, D)' M J (Z, D))$$

$$= (O' Z)' O' Z - \lambda_{\min}((O' Z, O' D)'(O' Z, O' D))$$

$$= Z' Z - \lambda_{\min}((Z, O' D)'(Z, O' D)),$$

$$\sim CLR_{k-b,p}(O' D, 0^{(k-b)\times0}) = CLR_{k-b,p}(O' D), \quad (31.18)$$

where $Z \sim N(0^k, I_k)$, $Z := O' Z \sim N(0^{k-b}, I_{k-b})$, “~” denotes “has the same distribution as,” and $CLR_{k-b,p}(O' D)$ is the expression from the full vector test defined in (5.5).

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 $CLR_{k-b,p}(O' D)$ 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

$$c_{k,p}(D, J, 1 - \alpha) = c_{k-b,p}(Y, 0^{(k-b)\times0}, 1 - \alpha) = c_{k-b,p}(Y, 1 - \alpha).$$

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 $\theta$ of the statistics in the following lemma.

**Lemma 31.2** Given the preliminary estimator $\hat{\beta}_n$ of $\beta^*_n$, the statistics $AP_n^S(\hat{\beta}_n)$, $QLR_n^S(\hat{\beta}_n)$, $\tilde{n}_n$, $c_{k,p}(n^{1/2} \tilde{D}_n(\hat{\beta}_n), \tilde{n}_n(\hat{\beta}_n), 1 - \alpha)$, $\tilde{D}_n^*(\hat{\beta}_n) M_{\tilde{n}_n(\hat{\beta}_n)} \tilde{D}_n(\hat{\beta}_n)$, $\tilde{g}_n(\hat{\beta}_n)' \tilde{\Omega}_n^{-1/2}(\hat{\beta}_n) M_{\tilde{n}_n(\hat{\beta}_n)} \tilde{D}_n(\hat{\beta}_n)$, $\tilde{\Sigma}_n(\hat{\beta}_n)$, and $\tilde{L}_n(\hat{\beta}_n)$ are invariant to the transformation $(g_i(\beta), G_i(\beta)) \sim (MG_i(\beta), MG_i(\beta)) \forall i \leq n$ for any $k \times k$ nonsingular matrix $M$. This transformation induces the following transformations: $\tilde{g}_n(\hat{\beta}_n) \sim M\tilde{g}_n(\hat{\beta}_n), \tilde{G}_n(\hat{\beta}_n) \sim M\tilde{G}_n(\hat{\beta}_n), \tilde{G}_n(\hat{\beta}_n) \sim M\tilde{G}_n(\hat{\beta}_n), \tilde{\Omega}_n(\hat{\beta}_n) \sim M\tilde{\Omega}_n(\hat{\beta}_n) M', \tilde{\Sigma}_n(\hat{\beta}_n) \sim M\tilde{\Sigma}_n(\hat{\beta}_n) M', \tilde{V}_n(\hat{\beta}_n) \sim (I_{p+1} \otimes M) \tilde{V}_n(\hat{\beta}_n) \times (I_{p+1} \otimes M'), and \tilde{R}_n(\hat{\beta}_n) \sim (I_{p+1} \otimes M) \tilde{R}_n(\hat{\beta}_n) (I_{p+1} \otimes M').
The proof of the lemma is straightforward for all quantities except \(c_{k,p}(n^{1/2}\hat{D}_n^*(\beta_n), \hat{J}_n(\beta_n), 1 - \alpha)\). Using Lemma 31.1 this quantity depends on \(n^{1/2}\hat{D}_n(\beta_n)\) and \(J_n(\beta_n)\) only through the nonzero singular values of \(O(\hat{J}_n(\beta_n))n^{1/2}\hat{D}_n(\beta_n)\), which equal the square roots of the nonzero eigenvalues of \(n^{1/2}\hat{D}_n(\beta_n)'M_{J_n(\beta_n)}n^{1/2}\hat{D}_n(\beta_n)\). But, the latter quantity is invariant to the transformation \((g_i(\beta), G_i(\beta)) \sim (Mg_i(\beta), MG_i(\beta))\).

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

\[
ACLR_{k-b,p,q}(\tau^c_{\infty}) := \mathbf{Z}'\mathbf{Z} - \lambda_{\min}((\Upsilon(\tau^c_{\infty}), \mathbf{Z}_2)'(\Upsilon(\tau^c_{\infty}), \mathbf{Z}_2)) \text{ where}
\]

\[
\mathbf{Z} := \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \sim N(0^{k-b}, I_{k-b}) \text{ for } \mathbf{Z}_1 \in \mathbb{R}^d \text{ and } \mathbf{Z}_2 \in \mathbb{R}^{k-b-q},
\]

\[
\tau^c_{\infty} := (\tau^c_{(q+1)\infty}, \ldots, \tau^c_{\min(k-b,p)\infty})' \in \mathbb{R}^{\min(k-b,p)-q},
\]

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

\[
\Upsilon(\tau^c_{\infty}) := \begin{pmatrix} \text{Diag}(\tau^c_{\infty}), 0^{(k-b-q)\times(p-k-q)} \end{pmatrix} \in \mathbb{R}^{(k-b-q)\times(p-q)} \text{ if } k-b < p. \tag{31.19}
\]

**Lemma 31.3** Suppose \(\{(\mathbf{D}^n, J^c_n) : n \geq 1\}\) is a sequence of constant (i.e., nonrandom) \(k \times p\) and \(k \times b\) matrices, respectively, such that \(O^D_n\mathbf{D}^c_n\) (for \(O^D_nO^c_n = M_{J^c_n}\) 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) \(\tau^c_{jn} \rightarrow \infty\) for \(j \leq q\) for some \(0 \leq q \leq \min\{k-b,p\}\) and (iii) \(\tau^c_{jn} \rightarrow \tau^c_{j\infty} < \infty\) for \(j = q+1, \ldots, \min\{k-b,p\}\). Then,

\[
c_{k,p}(\mathbf{D}^c_n, J^c_n, 1 - \alpha) \rightarrow c_{k-b,p,q}(\tau^c_{\infty}, 1 - \alpha).
\]

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 \(\min\{k-b,p\} - q \geq 0\) vectors \(\tau^c_{\infty}\) with nonnegative elements in nonincreasing order, the df of \(ACLR_{k-b,p,q}(\tau^c_{\infty}) := \mathbf{Z}'\mathbf{Z} - \lambda_{\min}((\Upsilon(\tau^c_{\infty}), \mathbf{Z}_2)'(\Upsilon(\tau^c_{\infty}), \mathbf{Z}_2))\) 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 \(\mathbf{Z} := (\mathbf{Z}_1, \mathbf{Z}_2)' \sim N(0^{k-b}, I_{k-b})\) for \(\mathbf{Z}_1 \in \mathbb{R}^d\) and \(\mathbf{Z}_2 \in \mathbb{R}^{k-b-q}\) and \(\tau^c_{\infty}\) and \(\Upsilon(\tau^c_{\infty})\) 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.

129
By the Lyapunov CLT, under sequences \( \{ \lambda_{n,h}^S \in \Lambda^S : n \geq 1 \} \), we have

\[
n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} g_i \\ vec(G_i) \\ vec(g_i g_i' - \Omega_n) \\ vec(G_{j1} - E_n G_{j1}) \end{pmatrix} \to_d T_h \sim N(0d^*, h_{10}), \text{ where}
\]

\[
T_h := (\vec{g}_h, T_{h,2}, T_{h,3}, \vec{T}'_h) \text{ for } \vec{g}_h \in R^k, \ T_{h,2} \in R^{kp}, \ T_{h,3} \in R^{k^2}, \ T_{h,4} \in R^{kb},
\]

\[
G_h := vec_{k,p}^{-1}(T_h) \in R^{k \times p},
\]

(31.20)

d^* = k + kp + k^2 + kb, and the function \( vec_{k,p}^{-1}(\cdot) \) is the inverse of the \( vec(\cdot) \) function for \( k \times p \) matrices. (Thus, the domain of \( vec_{k,p}^{-1}(\cdot) \) consists of \( kp \)-vectors and its range consists of \( k \times p \) matrices.) As defined in (31.20), \( \vec{g}_h \) is the same as in (16.21) for the full vector case.

The asymptotic distributions of (i) \( n^{1/2}(\vec{\beta}_n - \beta^*_n) \), (ii) \( n^{1/2}\vec{g}_n(\vec{\beta}_n) \), (iii) \( n^{1/2}vec(\vec{D}_n(\vec{\beta}_n) - D_n) \), where \( D_n := E_n G_i \), (iv) \( n^{1/2}(\vec{\beta}_n - \vec{\Omega}_n - \vec{\theta}_n) \), and (v) \( n^{1/2}(\vec{G}_{\theta,\beta}(\vec{\beta}_n) - E_n G_{i\beta}) \) are given by

(i) \( \vec{\beta}_h := \left[ (h_{5,g}^{-1/2} h_{4,\beta}')(h_{5,g}^{-1/2} h_{4,\beta}) \right]^{-1} (h_{5,g}^{-1/2} h_{4,\beta})' h_{5,g}^{-1/2} \vec{g}_h \),

(ii) \( \vec{g}_h^S := h_{5,g}^{1/2} M_{h_{5,g}^{-1/2} h_{4,\beta}}^{-1/2} \vec{g}_h \),

(iii) \( vec(\vec{D}_h^S) := vec(\vec{G}_h) - h_{5,G} h_{5,g}^{-1} \vec{g}_h + vec(h_{4,\theta,3} \vec{\beta}_h, ..., h_{4,\theta,3} \vec{\beta}_h) - h_{5,G} h_{5,g}^{-1} h_{4,\beta} \vec{\beta}_h \),

(iv) \( \vec{\varphi}_h^S := (\vec{\varphi}_h^S, ..., \vec{\varphi}_h^{S,b}) \), and

(v) \( \vec{\varphi}_h^S := (\vec{\varphi}_h^S, ..., \vec{\varphi}_h^{S,b}) \)

where

\[
\vec{\varphi}_h^S := \vec{\varphi}_h^S - h_{5,\theta} h_{5,g}^{-1} \vec{g}_h + (h_{4,\theta,3} \vec{\beta}_h - h_{5,\theta} h_{5,g}^{-1} h_{4,\beta}) \vec{\beta}_h \text{ for } j = 1, ..., b,
\]

\[
T_{h,3}, T_{h,4} \in R^k \text{ denote the } (j - 1)k + 1, ..., jk \text{ components of } T_{h,3} \text{ and } T_{h,4}, \text{ respectively, and}
\]

\( (h_{5,\theta} s_j)' \in R^k \) denotes the \( j \)-th column of \( (h_{5,\theta} s_j)' \in R^{k \times k} \) for \( s = 1, ..., b \).\(^73\) If no preliminary estimator appears, i.e., \( \vec{\beta}_n = \beta^*_n \), then the quantities in (31.21) reduce to those in the full vector case. In particular, \( \vec{\beta}_h = 0^b, \vec{g}_h^S = \vec{g}_h \), and \( vec(\vec{D}_h^S) = vec(\vec{G}_h) - h_{5,G} h_{5,g}^{-1} \vec{g}_h = vec(\vec{D}_h) \).

Consider the function that maps \( vec(\varphi) \) onto \( vec(\varphi^{-1/2}) \), where \( \varphi \in R^{k \times k} \) is positive definite. Let \( \vec{\varphi}_h \in R^{k^2 \times k^2} \) denote the matrix of partial derivatives of that mapping evaluated at \( vec(h_{5,g}) \). Consider the function that maps \( vec(J) \) for \( J \in R^{k \times b} \) onto \( vec((-J'(J)^{-1}, e_1'), ..., (-J'(k-b)(J_1)^{-1}, e_{k-b}')) \in R^{k(k-b)} \), as defined in (31.14) and (31.5). Denote by \( \vec{B}_h \in R^{k(k-b) \times kb} \)

\(^73\) See (31.45)-(31.46) for (i), (31.48) for (ii), (31.52) for (iii), (31.54) for (iv), and (31.55) for (v).
the matrix of partial derivatives of that mapping evaluated at $vec(h_{5g}^{-1/2}h_{4\beta})$.

The asymptotic distributions of (vi) $n^{1/2}(\Omega^{-1/2}(\hat{\beta}_n) - \Omega_n^{-1/2})$, (vii) $n^{1/2}(\Omega_n^{-1/2}(\hat{\beta}_n) - \Omega^{-1/2}E_nG_{\beta})$, (viii) $n^{1/2}(\Omega - O_n)$, (ix) $n^{1/2}(\Omega_n^{-1/2}(\hat{\beta}_n) - O_n'\Omega_n^{-1/2}D_n)$, (x) $n^{1/2}\Omega_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n) \times U_nB_nS_n$ are given by

$$(vi) \ vec_{k,k}^{-1}(\Sigma_h vec(\Sigma_h^S)),
(vii) \ \Sigma_h^S := h_{5g}^{-1/2}\Sigma_h^S + vec_{k,k}^{-1}(\Sigma_h vec(\Sigma_h^S))h_{4\beta},
(viii) \ vec_{k,k-b}(\Sigma_h vec(\Sigma_h^S)),
(ix) \ \chi_n := vec_{k,k-b}(\Sigma_h vec(\Sigma_h^S))'h_{5g}^{-1/2}h_4 + O(h_{5g}^{-1/2}h_{4\beta})'vec_{k,k}(\Sigma_h vec(\Sigma_h^S))h_4 + O(h_{5g}^{-1/2}h_{4\beta})'h_{5g}^{-1/2}\Sigma_h,
(x) \ \Sigma_h := (\Sigma_{h,q}\Sigma_h^S, \Sigma_{h,p-q}\Sigma_h^S), \ where \ \Sigma_{h,q}\Sigma_h^S := h_{3,q}\Sigma_h \in \mathbb{R}^{(k-b)\times q^S}, \ \Sigma_{h,p-q}\Sigma_h^S := h_{3,p-q}\Sigma_h + \chi_n h_{81}h_{2,p-q}\Sigma_h \in \mathbb{R}^{(k-b)\times (p-q^S)}, \ (31.22)$$

and $h_{1,p-q^S} \in \mathbb{R}^{(k-b)\times (p-q^S)}$ is defined as in $[16.24]$ with $k - b$ and $q^S$ in place of $k$ and $q$, respectively.\(^4\)

**Lemma 31.5** Suppose Assumptions gB and WUS hold for some non-empty parameter space $\Lambda^S_W \subset \Lambda_{WU}$. Under all sequences $\{\lambda^S_{n,h} : n \geq 1\}$,
(a) $n^{1/2}(\hat{\beta}_n - \beta_n^*) \rightarrow_d \Sigma_h,
(b) \tilde{J}_n(\hat{\beta}_n) \rightarrow_p h_{5g}^{-1/2}h_{4\beta},
(c) \ n^{1/2} \begin{pmatrix} \tilde{g}_n(\theta_0, \hat{\beta}_n) \\ \tilde{D}_n(\theta_0, \hat{\beta}_n) - EF_{n}G_{i} \\ \tilde{\Omega}_n(\theta_0, \hat{\beta}_n) - EF_{n}g_{i}g_{i}' \\ \tilde{G}_{\beta n}(\theta_0, \hat{\beta}_n) - EF_{n}G_{i\beta} \end{pmatrix} \rightarrow_d \begin{pmatrix} \overline{\Sigma}_h^S \\ \overline{\Sigma}_h \\ \overline{\Sigma}_h \\ \overline{\Sigma}_h \end{pmatrix},$

where $(\overline{\Sigma}_h, \overline{\Sigma}_h^S, \overline{\Sigma}_h^S, \overline{\Sigma}_h^S)$ and $\overline{\Sigma}_h^S$ are independent,
(d) for $\overline{\Omega}_{F_{n}}$ defined in $[31.10]$, $n^{1/2}\overline{\Omega}_{F_{n}}\Omega_n^{-1/2}(\theta_0, \hat{\beta}_n)\tilde{D}_n(\theta_0, \hat{\beta}_n)U_{F_{n}}T_{n} \rightarrow_d \overline{\Sigma}_h \in \mathbb{R}^{(k-b)\times p},$

where $(\overline{\Sigma}_h, \overline{\Sigma}_h^S, \overline{\Sigma}_h^S, \overline{\Sigma}_h^S)$ and $\overline{\Sigma}_h^S$ are independent, and
(e) under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_{n,h}} : n \geq 1\}$ with $\lambda_{w_{n,h}} \in \Lambda^S_w$, the convergence results in parts (a)-(d) hold with $n$ replaced with $w_n$.\(^4\)

\(^4\)See $[31.56]$ for (vi), $[31.57]$ for (vii), $[31.59]$ for (viii), $[31.64]$ for (ix), and $[31.66], [31.61]$, 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 \( \hat{\beta}_n \) we only need to impose the conditions in \( S_{AR,1}^S \). In particular, for consistency of \( \tilde{\beta}_n \), the variance matrix \( \Omega_{Fn} \) is allowed to be rank deficient. Lemma 31.5(b) and (c) implies Theorem 9.1 for the subvector AR test. This holds because \( AR_n^S(\theta_0, \tilde{\beta}_n) \) is a quadratic form in \( M_{\tilde{\eta}_n}(\theta_0, \tilde{\beta}_n) \) which converges in distribution to \( M_{h_5^{-1/2}, h_4, \beta}^{-1/2} \mathcal{F}_0 \). Because \( h_5^{-1/2}/h_{4, \beta} \) has full column rank \( b \), the desired result follows.

An analogue of Proposition 16.5 holds where \( \tilde{W}_n, W_{Fn} \) and \( \tilde{D}_n \) are replaced by \( I_k, I_k, \) and \( \tilde{O}_n^{-1/2}(\theta_0, \tilde{\beta}_n)\tilde{D}_n(\theta_0, \tilde{\beta}_n) \in R^{(k-b) \times p} \), respectively. In particular, \( \tilde{\kappa}_{jn} \) is defined as the \( j \)-th eigenvalue of

\[
n(\tilde{O}_n^{-1/2}(\theta_0, \tilde{\beta}_n)\tilde{D}_n(\theta_0, \tilde{\beta}_n)\tilde{U}_n)'(\tilde{O}_n^{-1/2}(\theta_0, \tilde{\beta}_n)\tilde{D}_n(\theta_0, \tilde{\beta}_n)\tilde{U}_n).
\]

Recall the following notation as for the full vector test, \( B_{Fn} = (B_{Fn,q}, B_{Fn,p-q}) \), \( C_{Fn} = (C_{Fn,q}, C_{Fn,k-b-q}) \), with \( B_{Fn,q} \in R^{p \times q}, B_{Fn,p-q} \in R^{p \times (p-q)}, \) and \( C_{Fn,q} \in R^{(k-b) \times q} \). Proposition 31.6

**Proposition 31.6** Suppose Assumption WU^S holds for some non-empty parameter space \( \Lambda_{WU}^S \subset \Lambda_{WU}^S \). Under all sequences \( \{\lambda_{n,h}^S : n \geq 1\} \) with \( \lambda_{n,h}^S \in \Lambda_{WU}^S \),

(a) \( \tilde{\kappa}_{jn} \rightarrow_p \infty \) for all \( j \leq q^S \),

(b) \( (\tilde{\kappa}_{(q^S+1)n}, \ldots, \tilde{\kappa}_{pn})' \) converges in distribution to the (ordered) \( p-q^S \)-vector of the eigenvalues of \( \tilde{\Sigma}_{h,p-q^S} \),

(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 \( \{\lambda_{w_n,h}^S : n \geq 1\} \) with \( \lambda_{w_n,h}^S \in \Lambda_{WU}^S \), the results in parts (a)-(c) hold with \( n \) replaced with \( w_n \).

An analogue of Theorem 16.6 holds for \( QLS_{WU,n} = AR_n^S(\hat{\theta}_0, \hat{\beta}_n) - \lambda_{\min}(n\hat{Q}_{WU,n}) \), defined in (31.1). For \( \tilde{\eta} = (\theta_0, \tilde{\beta}_n) \), \( \tilde{w} \rightarrow \tilde{w} \), we can write

\[
Q_{WU,n}^S = (\tilde{O}_n^{-1/2}(\tilde{\eta})\tilde{D}_n(\tilde{\eta})\tilde{U}_n, \tilde{O}_n^{-1/2}(\tilde{\eta})\tilde{g}_n(\tilde{\eta}))' \left( \tilde{O}_n^{-1/2}(\tilde{\eta})\tilde{D}_n(\tilde{\eta})\tilde{U}_n, \tilde{O}_n^{-1/2}(\tilde{\eta})\tilde{g}_n(\tilde{\eta}) \right)
\]

by again replacing \( \tilde{W}_n, W_{Fn}, \tilde{O}_n^{-1/2}g_n, \) and \( \tilde{D}_n \in R^{k \times p} \) by \( I_k, I_k, \tilde{O}_n^{-1/2}(\tilde{\eta})\tilde{g}_n(\tilde{\eta}) \), respectively. This implies that the role of \( k \) is played by \( k-b \). Note that by Lemma 31.5(b) and (c) and (31.5) below, which implies \( \tilde{O}_n = O(\tilde{\eta}(\theta_0, \tilde{\beta}_n))' \rightarrow_p \)
\( O(h_{5,g}^{-1/2} h_{4,\beta})' \), we have

\[
n^{1/2} \tilde{E}_n \tilde{G}_n^{-1/2}(\tilde{\eta}) \tilde{g}_n(\tilde{\eta}) \rightarrow_d O(h_{5,g}^{-1/2} h_{4,\beta})' h_{5,g}^{-1/2} \tilde{g}_h = O(h_{5,g}^{-1/2} h_{4,\beta})' h_{5,g}^{-1/2} \tilde{g}_h \sim N(0^{k-b}, I_{k-b}), \tag{31.25}
\]

using \( \tilde{g}_h^S := h_{5,g}^{-1/2} M_{h_{5,g}^{-1/2} h_{4,\beta}^{-1/2} \tilde{g}_h}, OO' = M_{h_{5,g}^{-1/2} h_{4,\beta}^{-1/2}}, \) and \( O'O = I_{k-b} \).

**Theorem 31.7** Suppose Assumption WUS holds for some non-empty parameter space \( \Lambda^S_s \subset \Lambda^S_W \). Under all sequences \( \{\lambda^S_{n,h} : n \geq 1\} \) with \( \lambda^S_{n,h} \in \Lambda^S_s \),

\[
QLR^S_{W,U,n} \rightarrow_d l'_h l_h - \lambda_{\min}(\Delta^S_{h,p-q^S}, l_h)' h_{3,k-b-q^S} h_{3,k-b-q^S}' (\Delta^S_{h,p-q^S}, l_h),
\]

where \( l_h := O(h_{5,g}^{-1/2} h_{4,\beta})' h_{5,g}^{-1/2} \tilde{g}_h \), \( \Delta^S_{h,p-q^S} \) 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 \( q^S \leq \min\{k - b, p\} \)), \( \Delta^S_{h,p-q^S} \) does not appear in the limit random variable and the limit random variable reduces to

\[
l'_h h_{3,p} h'_{3,p} l_h \sim \chi^2_p.
\]

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

\[
l'_h l_h \sim \chi^2_{k-b}. \tag{31.26}
\]

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 \( \{\lambda^S_{w_n,h} : n \geq 1\} \) with \( \lambda^S_{w_n,h} \in \Lambda^S_s \), 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

\[
\Upsilon_n := \begin{bmatrix}
\Upsilon_{n,q^S} & 0^{q^S \times (p-q^S)} \\
0^{(p-q^S) \times q^S} & \Upsilon_{n,p-q^S} \\
0^{(k-b-p) \times q^S} & 0^{(k-b-p) \times (p-q^S)}
\end{bmatrix} \in R^{(k-b) \times p} \text{ if } k - b \geq p, \text{ and} \tag{31.27}
\]

\[
\Upsilon_n := \begin{bmatrix}
\Upsilon_{n,q^S} & 0^{q^S \times (k-b-q^S)} & 0^{q^S \times (p-(k-b))} \\
0^{(k-b-q^S) \times q^S} & \Upsilon_{n,k-b-q^S} & 0^{(k-b-q^S) \times (p-(k-b))}
\end{bmatrix} \in R^{(k-b) \times p} \text{ if } k - b < p,
\]

133
as in (25.2), but with $\tau_{1,F_n}, \ldots, \tau_{p,F_n}$ and $q^S$ in place of $\tau_{1,F_n}, \ldots, \tau_{p,F_n}$ and $q$, respectively. Define

$$
\tilde{D}_n^+ := (\hat{O}_n^T \hat{\Omega}_n^{1/2}(\theta_0, \beta_n) \hat{\tilde{D}}_n(\theta_0, \beta_n), \hat{O}_n^T \hat{\Omega}_n^{1/2} \hat{\tilde{g}}_n) \in R^{(k-b) \times (p+1)},
$$

$$
\hat{U}_n^+ := \begin{bmatrix} \hat{U}_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)}, U_n^+ = \begin{bmatrix} U_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
$$

$$
B_n^+ := (B_{n,q_1}^+, B_{n,p+1-q_1}^+) \quad \text{for} \quad B_{n,q_1}^+ \in R^{(p+1) \times q_1} \text{ and } B_{n,p+1-q_1}^+ \in R^{(p+1) \times (p+1-q_1)}, \quad (31.28)
$$

$$
D_n^+ := (O(J_n)^T \hat{\Omega}_n^{1/2} D_n, 0^k) \in R^{(k-b) \times (p+1)}, \quad \gamma_n^+ := (\gamma_n, 0^{k-b}) \in R^{(k-b) \times (p+1)},
$$

$$
S_n^+ := Diag\{ (n^{1/2} \tau_{1,F_n})^{-1}, \ldots, (n^{1/2} \tau_{q^S,F_n})^{-1}, 1, \ldots, 1 \} = \begin{bmatrix} S_n & 0^{p \times 1} \\ 0^{1 \times p} & 1 \end{bmatrix} \in R^{(p+1) \times (p+1)},
$$

with $J_n$ defined in (31.8). Let

$$
\hat{\kappa}_{jn}^+ \quad \text{denote the } j\text{th eigenvalue of } n\hat{U}_n^+ \hat{D}_n^+ \hat{D}_n^+ \hat{U}_n^+, \quad \forall j = 1, \ldots, p + 1, \quad (31.29)
$$

ordered to be nonincreasing in $j$.

**Lemma 31.8** Suppose Assumption WUS holds for some non-empty parameter space $\Lambda^S \subset \Lambda^S_{WUS}$. 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) $\hat{\kappa}_{jn}^+ \rightarrow_p \infty$ for $j = 1, \ldots, q^S$ and (b) $\hat{\kappa}_{jn}^+ = \lambda_{p,h}^S(n^{1/2} \tau_{1,F_n})^2$ 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

$$
\tau_{1,F_n}^{-1} \hat{D}_n^+ U_n^+ B_n^+ = \tau_{1,F_n}^{-1} D_n^+ U_n^+ B_n^+ + (n^{1/2} \tau_{1,F_n})^{-1} n^{1/2} (\hat{D}_n^+ - D_n^+) U_n^+ B_n^+ = \tau_{1,F_n}^{-1} O(J_n)^T \hat{\Omega}_n^{1/2} D_n U_n B_n, 0^{k-b} + O_p(n^{1/2} \tau_{1,F_n}^{-1}) = \tau_{1,F_n}^{-1} C_n \gamma_n^+ + O_p((n^{1/2} \tau_{1,F_n}^{-1})^{-1}) \rightarrow_p h_3 \begin{bmatrix} h_6^{\omega} & 0^{(p+1-q_1) \times (p+1-q_1)} \\ (k-b-q_1)^{\times (p+1-q_1)} & 0^{(k-b-q_1) \times (p+1-q_1)} \end{bmatrix}, \quad \text{where } h_6^{\omega} := Diag\{ h_{6,1}, h_{6,1} h_{6,2}, \ldots, \prod_{\ell=1}^{r_{\gamma}^1-1} h_{6,\ell} \}(31.30)
$$

and the second equality uses $n^{1/2}(\hat{D}_n^+ - D_n^+) = O_p(1)$, which holds by (31.62) below and Lemma

Electronic copy available at: https://ssrn.com/abstract=3366443
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, e.g. in \( h_{6,j} \) for \( j = 1, \ldots, r^*_1 - 1 \).

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 \( k, h_{5,q}^{1/2} \tilde{y}_n, n^{1/2} \tilde{W}_n \tilde{D}_n \tilde{U}_n \) and \( \sum_{h,p-q} h_{3,k-q} h'_{3,k-q} \sum_{h,p-q} \) in the proof of Theorem 16.1 are played by \( k-b, l_b \) (defined in Theorem 31.7), \( n^{1/2} \tilde{O}'_{F_n} \tilde{\Omega}_n^{-1/2} (\theta_0, \beta_n) \times \tilde{D}_n(\theta_0, \beta_n) \tilde{U}_n \), and \( \sum_{h,p-q} h_{3,k-b-q} h'_{3,k-b-q} \sum_{h,p-q} \), 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

\[
\begin{align*}
   &c_{k,p}(n^{1/2} \tilde{\Omega}_n^{-1/2} (\theta_0, \beta_n) \tilde{D}_n(\theta_0, \beta_n) \tilde{U}_n) \\
  &\quad \rightarrow_{d} c_{k-b,p,q,} (h'_{3,k-b-q} \sum_{h,p-q} \sum_{h,p-q}, 1 - \alpha), \\
  &\hspace{1cm} (31.31)
\end{align*}
\]

where \( \tilde{\Omega}_n \) denotes the matrix of singular values of \( n^{1/2} \tilde{O}'_{F_n} \tilde{\Omega}_n^{-1/2} (\theta_0, \beta_n) \tilde{D}_n(\theta_0, \beta_n) \tilde{U}_n \), defined as in (27.8), \( c_{k-b,p,q,} (\cdot, 1 - \alpha) \) is defined in (31.19) (and \( c_{k-b,p,q,} (h'_{3,k-b-q} \sum_{h,p-q} \sum_{h,p-q}, 1 - \alpha) \) 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 WUS holds when \( \tilde{U}_n \) is defined as \( \tilde{L}_n^{1/2} \), where \( \tilde{L}_n := \langle \theta_0, I_p \rangle (\sum_{n}(\theta_0, \beta_n))^{-1} (\theta_0, I_p) \in \mathbb{R}^{p \times p} \) is defined in (9.11). Furthermore, the following lemma shows that \( \mathcal{F}^S = \mathcal{F}^S_{WU} \), where \( \mathcal{F}^S \) is defined in (9.17) and \( \mathcal{F}^S_{WU} \) is defined just below (31.16). Recall the definition \( \Sigma_{j\ell}(\Omega_F, R_F) := tr(R_{j\ell}^{1/2} \Omega_F^{-1})/k \) for the \((j, \ell)\)-th component of \( \Sigma \), where \( \Omega_F := E_F g_i g_i' \), \( V_F := E_F (f_i - E_F f_i) (f_i - E_F f_i)' \in \mathbb{R}^{(p+1)k \times (p+1)k} \), \( R_F := (B' \otimes I_k) V_F (B \otimes I_k) \in \mathbb{R}^{(p+1)k \times (p+1)k} \) in (16.7). Also, recall the definition of \( \tilde{R}_n(\theta_0, \beta_n) := (B' \otimes I_k) \tilde{V}_n(\theta_0, \beta_n)(B \otimes I_k) \), which is given by (5.3) with \( (\theta_0, \beta_n) \) in place of \( \theta \).

**Lemma 31.9** (a) Assumption WUS holds with \( \tilde{U}_2n = (\tilde{\Omega}_n(\theta_0, \beta_n), \tilde{R}_n(\theta_0, \beta_n)) \), \( U_1(U_2F) = U_1(\Omega_F, R_F) = ((\theta_0, I_p) (\sum_{n}(\Omega_F, R_F))^{-1} (\theta_0, I_p))^{1/2} \) defined in (16.8), and \( h_8 = \lim U_2F_{\infty} = \lim(\Omega_{F \infty}, R_{F \infty}) \), under any sequence \( \{\lambda_{w, n} : n \geq 1\} \) and \( \delta_1 \) sufficiently small and \( M_1 \) sufficiently large in the definition of \( \mathcal{F}^S_{WU} \).

(b) \( \mathcal{F}^S = \mathcal{F}^S_{WU} \), As in (27.73), we have

\[
\tilde{V}_n(\theta_0, \beta_n) = E_F f_i f_i' - (E_F f_i) (E_F f_i)' + o_p(1) \quad (31.32)
\]
and
\[
\hat{R}_n(\theta_0, \hat{\beta}_n) = (B' \otimes I_k) (E_{F_n}f_1f_1' - (E_{F_n}f_1)(E_{F_n}f_1')) (B \otimes I_k) + o_p(1)
\]
\[
\rightarrow_p R_h := (B' \otimes I_k) [h_3 - vec((0^k, h_4))vec((0^k, h_4))'] (B \otimes I_k),
\] (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, \( \hat{U}_{2W_n} - U_{2F_{w_n}} = (\hat{\Omega}_{W_n}(\theta_0, \hat{\beta}_n), \hat{R}_{W_n}(\theta_0, \hat{\beta}_n)) - (\Omega_{F_{w_n}}, R_{F_{w_n}}) = o_p(1) \). 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(b).

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 \( \hat{g}_n(\beta) = n^{-1} \sum_{i=1}^{n} g_i(\theta_0, \beta) \) and write \( g_i \) for \( g_i(\beta^*) \), where \( \beta^* \) is the true value of \( \beta \), and analogously for other expressions, e.g., we write \( \hat{D}_n(\beta) \) for \( \hat{D}_n(\theta_0, \beta) \) and \( G_i \) for \( G_i(\theta_0, \beta^*) \). Furthermore, to simplify notation, we replace subscripts \( F_n \) by \( n \), e.g., we write \( E_n \), rather than \( E_{F_n} \).

**Proof of Lemma 31.5(a).** Given \( \{\lambda_{n,h}^S : n \geq 1\} \), let \( F_n \) and \( \beta_n^* \) denote the distribution of \( W_i \) and the true parameter \( \beta \) when the sample size is \( n \). Let \( \hat{Q}_n(\beta) = ||\hat{g}_n(\beta)||^2 \) 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 \( \{\lambda_{n,h}^S : n \geq 1\} \).

(a1). We first show consistency of the first-step estimator, i.e., \( \overline{\beta}_n - \beta_n^* \rightarrow_p 0^p \) under \( \{\lambda_{n,h}^S : n \geq 1\} \). Let \( \varepsilon > 0 \). By the identifiability condition in \( \mathcal{F}_{AR,1}^S \) in (9.14), there exists \( \delta_\varepsilon > 0 \) such that \( \beta \in B \setminus B(\beta_n^*, \varepsilon) \) implies \( Q_n(\beta) \gtrless \delta_\varepsilon \). Thus,

\[
P_n(||\overline{\beta}_n - \beta_n^*|| > \varepsilon) = P_n(\overline{\beta}_n \in B \setminus B(\beta_n^*, \varepsilon))
\leq P_n(Q_n(\overline{\beta}_n) - \hat{Q}_n(\overline{\beta}_n) + \hat{Q}_n(\beta_n^*) \geq \delta_\varepsilon)
\leq P_n(Q_n(\overline{\beta}_n) - \hat{Q}_n(\beta_n^*) + \hat{Q}_n(\beta_n^*) \geq \delta_\varepsilon)
\leq P_n(2 \sup_{\beta \in B} |Q_n(\beta) - \hat{Q}_n(\beta)| \geq \delta_\varepsilon)
\rightarrow 0,
\]
where the second inequality holds because \( \hat{Q}_n(\cdot) \) is minimized by \( \hat{\beta}_n \), the third inequality holds because \( Q_n(\beta^*_n) = 0 \), and the convergence result holds, because, as we show now, \( \sup_{\beta \in B} |\hat{Q}_n(\beta) - Q_n(\beta)| \to_p 0 \).

For \( \delta > 0 \), define

\[
Y_{i\delta} := \sup_{\beta \in B} \sup_{\beta' \in B(\beta, \delta)} ||g_i(\beta') - g_i(\beta)||, \tag{31.34}
\]

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, \( E_{\mu}Y_{i\delta} \leq 2E_{\mu}\sup_{\beta \in B} ||g_i(\beta)|| < \infty \), where the latter inequality holds by the conditions in \( \mathcal{F}_{AR,1}^S \). 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 \( \mu \) and note that by assumption \( f_n \leq M \). We have \( \sup_n E_nY_{i\delta} = \sup_n E_{\mu}f_nY_{i\delta} \leq E_{\mu}MY_{i\delta} \to 0 \) 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 \( \delta \) centered at some points \( \beta_j, j = 1, \ldots, J_\delta \), i.e., \( B \subset \bigcup_{j=1}^{J_\delta} B(\beta_j, \delta) \). Let

\[
H_n(\beta) = \hat{g}_n(\beta) - E_ng_i(\beta). \tag{31.35}
\]

Because \( \mathcal{F}_{AR,1}^S \) imposes \( \sup_{\beta \in B} E_F ||g_i(\beta)||^{1+\gamma} \leq M \), 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

\[
P_n(\sup_{\beta \in B} ||H_n(\beta)|| > 2\varepsilon) \\
\leq P_n(\max_{j=1,\ldots,J_\delta} \sup_{\beta \in B(\beta_j, \delta)} ||H_n(\beta) - H_n(\beta_j)|| + ||H_n(\beta_j)|| > 2\varepsilon) \\
\leq P_n(\sup_{\beta \in B} \sup_{\beta' \in B(\beta, \delta)} ||H_n(\beta') - H_n(\beta)|| > \varepsilon) + P_n(\max_{j=1,\ldots,J_\delta} ||H_n(\beta_j)|| > \varepsilon), \tag{31.36}
\]

where the first inequality holds by the triangle inequality.

For the first summand in (31.36), we have the following bound

\[
P_n(\sup_{\beta \in B} \sup_{\beta' \in B(\beta, \delta)} ||H_n(\beta') - H_n(\beta)|| > \varepsilon) \\
\leq P_n(\frac{1}{n} \sum_{i=1}^{n} (Y_{i\delta} + E_nY_{i\delta}) > \varepsilon) \\
\leq E_n \left( \frac{1}{n} \sum_{i=1}^{n} (Y_{i\delta} + E_nY_{i\delta}) \right)/\varepsilon \\
= 2E_nY_{i\delta}/\varepsilon, \tag{31.37}
\]

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 \( \delta \leq \delta_\nu \). Because \( H_n(\beta) \to_p 0 \) we can find a finite \( n_{\delta_\nu} \in N \) such that for all \( n \geq n_{\delta_\nu} \) we have \( P_n(\max_{j=1, \ldots, j_{\delta_\nu}} ||H_n(\beta_j)|| > \varepsilon) < \nu/2 \). This proves

\[
P_n(\sup_{\beta \in B} ||H_n(\beta)|| > 2\varepsilon) \to 0 \quad (31.38)
\]
as \( n \to \infty \). By the reverse triangle inequality, we then obtain the desired \( \sup_{\beta \in B} |\hat{\varphi}_n(\beta) - Q_n(\beta)| \to_p 0 \) as \( n \to \infty \).

(a2). Next, we show consistency of \( \hat{\beta}_n \). Let \( \{\beta_n - \beta_n^*: n \geq 1\} \) be any nonstochastic sequence that converges to \( 0^b \). We can write \( E_n g_i(\beta_n) - E_n g_i(\beta_n^*) = E_n h_n \) for \( h_n = (g_i(\beta_n) - g_i(\beta_n^*)) f_n \). 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, \( E_n h_n \leq 2M E_n \sup_{\beta \in B} ||g_i(\beta)|| < \infty \) by the conditions in \( \mathcal{F}_{\mathcal{A},R}^S \). Therefore, by the DCT, \( E_n h_n \to 0^k \).

Define \( E_n g_i(\beta_n) = E_n g_i(\beta)|_{\beta = \beta_n} \). That is, the expectation is taken first treating \( \beta \) as nonrandom, and then the resulting expression is evaluated at the random vector \( \beta_n \). For any given \( \varepsilon > 0 \),

\[
||\hat{g}_n(\beta_n)|| \leq ||\hat{g}_n(\beta_n) - E_n g_i(\beta_n)|| + ||E_n g_i(\beta_n) - E_n g_i(\beta_n^*)||
\leq \sup_{\beta \in B(\beta_n^*, \varepsilon)} ||\hat{g}_n(\beta) - E_n g_i(\beta)|| + o_p(1)
= o_p(1),
\]
where the first inequality holds by the triangle inequality, the second inequality holds \( \text{wp} \to 1 \) because \( \beta_n - \beta_n^* \to_p 0^b \) and \( E_n h_n \to 0^k \), and the equality holds by \( 31.38 \).

Furthermore, \( n^{-1} \sum_{i=1}^n g_i(\beta_n)g_i(\beta_n') - E_n g_i g_i' \to_p 0^{k \times k} \). 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 \( E_n h_n \to 0^{k \times k} \) for \( h_n = (g_i(\beta_n) - g_i(\beta_n^*)) g_i(\beta_n' - \beta_n^*) f_n \) when \( \beta_n - \beta_n^* \) converges to \( 0^b \). The latter follows as above from the DCT using \( E_n \sup_{\beta \in B(\beta_n^*, \varepsilon)} ||g_i(\beta)||^2 < \infty \) by the conditions in \( \mathcal{F}_{\mathcal{A},R}^S \). The former follows using the same proof as for \( (31.38) \) noting that by the conditions in \( \mathcal{F}_{\mathcal{A},R}^S \) we have \( E_n \sup_{\beta \in B(\beta_n^*, \varepsilon)} ||g_i(\beta)||^2 < \infty \) and \( \sup_{\beta \in B(\beta_n^*, \varepsilon)} E_F ||g_i(\beta)||^{2+\gamma} \) is uniformly bounded. We have therefore shown that \( (\hat{\varphi}_n^n)_{i,j} - E_n g_i g_j' \to_p 0^{k \times k} \), where \( \hat{\varphi}_n^n \) is defined in \( [9,6] \). Because \( \mathcal{F}_{\mathcal{A},R}^S \) imposes \( \lambda_{\min}(E_F g_i g_j') \geq \delta \), it follows that

\[
(\hat{\varphi}_n^n)^{-1} - E_n g_i g_i' \to_p 0^{k \times k} \quad (31.40)
\]
The remainder of the consistency proof is analogous to the proof in part (a1), but with \( \hat{\varphi}_n^n :=

Electronic copy available at: https://ssrn.com/abstract=3366443
The inequality uses the triangle inequality and the equality uses (31.38), the assumption that \( \sup_{\beta \in (\beta_{cun}^*, \epsilon)} |E_n g_i(\beta)| \) is uniformly bounded by a finite number, and that \( |\varphi_n| = O(1) \) because \( \lambda_{\text{min}}(E_n g_i(\beta)) \geq \delta \). Equation (31.41) implies that \( \sup_{\beta \in (\beta_{cun}^*, \epsilon)} |\hat{Q}_n(\beta) - Q_n(\beta)| = o_p(1) \).

(a3). Now, we derive the limiting distribution of \( \hat{\beta}_n \) under \( \{\lambda_{n, i}^S : n \geq 1\} \). As above, \( \hat{Q}_n(\beta) := ||\varphi_n g_n(\beta)||^2 \). Because \( \beta_{cun}^* \) is bounded away from the boundary of \( B \), \( \hat{\beta}_n - \beta_{cun}^* \to_p 0^b \), and \( g_i(\cdot) \in C^2(B(\beta_{cun}^*, \vartheta)) \), the following FOC holds \( wp \to 1 \) and element-by-element mean-value expansions of \( \frac{\partial \hat{Q}_n(\hat{\beta}_n)}{\partial \beta} \) exist:

\[
0^b = \frac{\partial \hat{Q}_n(\hat{\beta}_n)}{\partial \beta} = \frac{\partial \hat{Q}_n(\beta_{cun}^*)}{\partial \beta} + \frac{\partial^2 \hat{Q}_n(\beta_{cun}^*)(\hat{\beta}_n - \beta_{cun}^*)}{\partial \beta_2 \beta_3},
\]

where the mean-value \( \beta_{cun}^* \) lies on the segment joining \( \hat{\beta}_n \) and \( \beta_{cun}^* \) (and hence satisfies \( \hat{\beta}_n - \beta_{cun}^* \to_p 0^b \)).

For \( m, j = 1, \ldots, b \), we have

\[
\frac{\partial^2 \hat{Q}_n(\beta_{cun}^*)}{\partial \beta_m \partial \beta_j} = n^{-1} \sum_{i=1}^n \frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta_{cun}^*) + \frac{\partial^2}{\partial \beta_m \partial \beta_j} \hat{Q}_n(\beta_{cun}^*) n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \beta_j} g_i(\beta_{cun}^*) + n^{-1} \sum_{i=1}^n \frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta_{cun}^*) \hat{Q}_n(\beta_{cun}^*) n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \beta_j} g_i(\beta_{cun}^*).
\]

By the argument in (31.39), \( n^{-1} \sum_{i=1}^n g_i(\beta_{cun}^*) \to_p 0^k \) under \( \{\lambda_{n, i}^S : n \geq 1\} \). Furthermore,

\[
n^{-1} \sum_{i=1}^n \frac{\partial}{\partial \beta_j} g_i(\beta_{cun}^*) - E_n \frac{\partial}{\partial \beta_j} g_i \to_p 0^k
\]

under \( \{\lambda_{n, i}^S : n \geq 1\} \). The latter holds by the argument in (31.39) with \( g_i(\hat{\beta}_n) \) replaced by \( \frac{\partial}{\partial \beta_j} g_i(\beta_{cun}^*) \) and using the assumptions \( \sup_{\beta \in (\beta_{cun}^*, \vartheta)} E_n ||G_{ij}(\beta)||^{1+\gamma} \) and \( E \sup_{\beta \in (\beta_{cun}^*, \vartheta)} ||G_{ij}(\beta)|| \) are uniformly bounded in \( F^S \). In addition, \( n^{-1} \sum_{i=1}^n \frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta_{cun}^*) = O_p(1) \) (again by an argument as in (31.39) with \( g_i(\hat{\beta}_n) \) replaced by \( \frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta_{cun}^*) \) and using the fact that \( \sup_{\beta \in (\beta_{cun}^*, \vartheta)} E_n ||\frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta)||^{1+\gamma} \) and \( E \sup_{\beta \in (\beta_{cun}^*, \vartheta)} ||\frac{\partial^2}{\partial \beta_m \partial \beta_j} g_i(\beta)|| \) are uniformly bounded by the conditions in \( F^S \)). It follows that \( \frac{\partial^2}{\partial \beta_m \partial \beta_j} \hat{Q}_n(\beta_{cun}^*) = B^* \to_p 0^{b \times b} \) under \( \{\lambda_{n, i}^S : n \geq 1\} \), where \( B^* := \ldots \)
\[ E_n G_{i\beta} (E_n g_i g_i')^{-1} E_n G_{i\beta}. \]

Because \( \lambda_{\min}(B_n^*\beta) \) is bounded away from zero (since \( \tau_{\min}(E_n G_{i\beta}) \geq \delta \) for \( F_n \in F^S \)), it follows that \( \frac{\partial^2}{\partial \beta \partial \beta'} Q_n(\beta_n^+) \) is invertible wp\( \rightarrow 1 \). This and (31.42) give

\[ n^{1/2}(\hat{\beta}_n - \beta_n^*) = -(B_n^* + o_p(1))^{-1} \sqrt{n} \frac{\partial}{\partial \beta} Q_n(\beta_n^*). \quad (31.45) \]

From above, we have

\[ n^{1/2} \frac{\partial}{\partial \beta} Q_n(\beta_n^*) = (E_n G_{i\beta})'(E_n g_i g_i')^{-1} n^{-1/2} \sum_{i=1}^n g_i(\beta_n^*) + o_p(1). \quad (31.46) \]

By the CLT result in (31.20), \( n^{-1/2} \sum_{i=1}^n g_i(\beta_n^*) \rightarrow_d \mathcal{G}_h \). 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: \( n^{-1} \sum_{i=1}^n g_i(\hat{\beta}_n) \rightarrow_p 0^k \), \( n^{-1} \sum_{i=1}^n g_i(\hat{\beta}_n) g_{ij}(\hat{\beta}_n) - E_n g_i g_{ij} \rightarrow_p 0^k \), \( n^{-1} \sum_{i=1}^n g_i(\hat{\beta}_n) g_{ij}(\hat{\beta}_n) g_i(\hat{\beta}_n)' - \lambda_{5,3,j,n}^S \rightarrow_p 0^{k \times k} \), \( \tilde{\Omega}_n^{-1}(\hat{\beta}_n) - (E_n g_i g_i)^{-1} \rightarrow_p 0^{k \times k} \), \( n^{-1} \sum_{i=1}^n G_{i\beta_j}(\hat{\beta}_n) - \lambda_{5,\beta_j,n}^S \rightarrow_p 0^k \), and \( n^{-1} \sum_{i=1}^n G_{i\beta_j}(\hat{\beta}_n) g_i(\hat{\beta}_n)' - \lambda_{5,\beta_j,n}^S \rightarrow_p 0^{k \times k} \). Therefore, \( \tilde{\Omega}_n(\hat{\beta}_n) \rightarrow_p h_{5,g}, \tilde{G}_{\beta n}(\hat{\beta}_n) \rightarrow_p h_{4,\beta}, \) and \( \tilde{J}_n(\hat{\beta}_n) \rightarrow_p h_{5,g}^{1/2} h_{4,\beta} \). \( \square \)

**Proof of Lemma 31.5(c).** We derive the limit distributions of (i) \( \tilde{g}_n(\beta_n) \), (ii) \( \tilde{D}_n(\hat{\beta}_n) - E_n G_{i\beta} \), (iii) \( \tilde{\Omega}_n(\theta_n, \hat{\beta}_n) - E_n g_i g_i' \), and (iv) \( \tilde{G}_{\beta n}(\hat{\beta}_n, \hat{\beta}_n) - E_n G_{i\beta} \) under \( \{\lambda_{n,h}^S : n \geq 1\} \) in (c1)-(c4) below, respectively.

\( (c1). \) We have

\[ n^{1/2} \tilde{g}_n(\hat{\beta}_n) = n^{1/2} \tilde{g}_n(\beta_n^*) + \tilde{G}_{\beta n}(\beta_n^+) n^{1/2}(\hat{\beta}_n - \beta_n^*) = (I_k - (E_n G_{i\beta})B_n^{-1}(E_n G_{i\beta})'(E_n g_i g_i')^{-1}) n^{1/2} \tilde{g}_n(\beta_n^*) + o_p(1), \quad (31.47) \]

where the first equality uses a mean-value expansion with \( \beta_n^+ \) on the segment joining \( \hat{\beta}_n \) and \( \beta_n^* \) and the second equality holds by (31.45) and (31.46). Therefore,

\[ n^{1/2} \tilde{g}_n(\hat{\beta}_n) \rightarrow_d \tilde{g}_n^S := h_{5,g}^{1/2} M_{h_{5,g}^{1/2} h_{4,\beta} h_{5,g}^{-1/2} \tilde{g}_h}. \quad (31.48) \]

Note that the assumption of strong identification of \( \beta \), namely \( \tau_{\min}(E_F G_{i\beta}) \geq \delta \) in \( F^S \), implies that \( h_{4,\beta} \) has full column rank \( b \).

\( (c2). \) Recall the definition of \( \tilde{\Gamma}_{jn}(\cdot) \) for \( j = 1, \ldots, p \) in (5.2). Under sequences \( \{\lambda_{n,h}^S : n \geq 1\} \), we
have
\[
\hat{\Gamma}_{ln}^{*} (\hat{\beta}_n), \ldots, \hat{\Gamma}_{pm}^{*} (\hat{\beta}_n) \hat{\Omega}_n^{-1} (\hat{\beta}_n) - ((E_n G_{i1} g_i')', \ldots, (E_n G_{ip} g_i')') \Omega_n^{-1} \rightarrow_p 0,
\]
(31.49)
which is established analogously to the results in (31.40) and (31.44), using the uniform finite bounds on \( \sup_{\beta \in B(\beta_n^*, \nu)} E_n \left\| \left( \frac{\partial}{\partial \beta} g_i (\beta) \right) g_i (\beta) \right\|^{1+\gamma} \) and \( E_\mu \sup_{\beta \in B(\beta_n^*, \nu)} \left\| \left( \frac{\partial^2}{\partial \beta, \partial \beta} g_i (\beta) \right) g_{ij} (\beta) \right\| \) for \( j = 1, \ldots, k \) in \( \mathcal{F}^S \).

Using \( g_i (\cdot) \in C^2 (B (\beta^*, \nu)) \), by a second-order Taylor expansion of \( \hat{g}_n (\hat{\beta}_n) \) about \( \beta_n^* \) and a mean-value expansion of \( G_i (\hat{\beta}_n) \) (as in (31.47)), we obtain

\[
n^{1/2} \text{vec}(\hat{D}_n (\hat{\beta}_n) - D_n) = n^{-1/2} \sum_{i=1}^n \left[ \text{vec}(G_i - E_n G_i) - \begin{pmatrix} E_n G_{i1} g_i' \\
\vdots \\
E_n G_{ip} g_i' \end{pmatrix} \Omega_n^{-1} g_i \right] 
+ \text{vec}((E_n G_{i1} g_i') n^{1/2} (\hat{\beta}_n - \beta_n^*), \ldots, (E_n G_{ip} g_i') n^{1/2} (\hat{\beta}_n - \beta_n^*)) 
- \begin{pmatrix} E_n G_{i1} g_i' \\
\vdots \\
E_n G_{ip} g_i' \end{pmatrix} \Omega_n^{-1} (E_n G_{ij} a) n^{1/2} (\hat{\beta}_n - \beta_n^*) + o_p (1),
\]
(31.50)
where \( E_n G_{ij} g_i' = E_n G_{ij} g_i' \) 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 \( G_i (\hat{\beta}_n) \) and \( g_i (\hat{\beta}_n) \), 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 \( G_i (\hat{\beta}_n) \). For this, we use

\[
\hat{G}_{ij, \beta_n} (\beta_n^*) - E_n G_{ij, \beta} \rightarrow_p 0^{k \times b} \text{ for } j = 1, \ldots, p
\]
(31.51)
for any sequence \( \beta_n^* \) such that \( \beta_n^* - \beta_n^* \rightarrow_p 0 \). The latter is established (as in several places above) using the assumptions that \( \sup_{\beta \in B(\beta_n^*, \nu)} E_n \left\| \left( \frac{\partial^2}{\partial \beta, \partial \beta} g_i (\beta) \right) g_i (\beta) \right\|^{1+\gamma} \) and \( E_\mu \sup_{\beta \in B(\beta_n^*, \nu)} \left\| \left( \frac{\partial^2}{\partial \beta, \partial \beta} g_i (\beta) \right) g_{ij} (\beta) \right\| \) are uniformly bounded in \( \mathcal{F}^S \). The first term of the third line comes from the second term of the expansion of \( g_i (\hat{\beta}_n) \) 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 \( g_i (\hat{\beta}_n) \) (which is indeed \( o_p (1) \) given the moment bounds in \( \mathcal{F}^S \) on \( \frac{\partial^2}{\partial \beta, \partial \beta} g_i (\beta) \)).

Equation (31.50), combined with Lemma (31.5a), (31.20), (31.14), and the paragraph containing (31.16), give

\[
n^{1/2} \text{vec}(\hat{D}_n (\hat{\beta}_n) - D_n) \rightarrow_d \text{vec}(\hat{D}_h)
:= (\text{vec}(\mathcal{G}_h) - h_{5,G} h_{5,g} \mathcal{G}_h) + \text{vec}(h_{4,\beta, \beta, \beta} \mathcal{G}_h, \ldots, h_{4,\beta, \beta, \beta} \mathcal{G}_h) - h_{5,G} h_{5,g} h_{4,\beta, \beta, \beta} \mathcal{G}_h.
\]
(31.52)
Note that \( \bar{\gamma}_h^S \) and \( \bar{\Omega}_h \) are independent because

\[
\text{cov}(M_{h_{5,g}^{-1/2}h_{4,\beta}} h_{5,g}^{-1/2} \bar{\gamma}_h, \bar{\Omega}_h) = M_{h_{5,g}^{-1/2}h_{4,\beta}} (h_{5,g}^{-1/2} \bar{\Omega}_h) [\eta_{5,g}^{-1/2} h_{4,\beta}]^{-1} = 0^{k \times h}. \tag{31.53}
\]

Next, we establish that \( \bar{\gamma}_h^S \) and \( \bar{D}_h^S \) (defined in \((31.21)\)) are independent. The last two summands that make up \( \bar{D}_h^S \) are independent of \( \bar{\gamma}_h^S \) because \( \bar{\gamma}_h^S \) and \( \bar{\Omega}_h \) are independent. Regarding the first summand, recall that from \((31.20)\) we know that \( \text{vec}(\bar{\gamma}_h) \) and \( \bar{\gamma}_h \) are jointly normally distributed and because \( \text{cov}(\bar{\gamma}_h, \text{vec}(\bar{\Omega}_h) - h_{5,G} h_{5,g}^{-1} \bar{\gamma}_h) = 0^{k \times pk} \), it follows that \( \text{vec}(\bar{\gamma}_h) - h_{5,G} h_{5,g}^{-1} \bar{\gamma}_h \) and \( \bar{\gamma}_h^S \) are independent.

\((c3)\). Next we derive the asymptotic distribution of \( \bar{\Omega}_n(\hat{\beta}_n) \). Let \( j \in \{1, \ldots, k\} \). By a mean-value expansion, for some vectors \( \beta_n^\circ \) and \( \beta_n^\dagger \) on the line segment joining \( \hat{\beta}_n \) and \( \beta_n^* \), under \( \{\lambda_{n,h}^S : n \geq 1\} \), we have

\[
n^{1/2} \left[ \bar{\Omega}_n(\hat{\beta}_n) - \Omega_{jn} \right] = n^{1/2} \left[ n^{-1} \sum_{i=1}^n g_i g_{ij} - \Omega_{jn} \right] + n^{-1} \sum_{i=1}^n \left[ G_{ij}(\beta_n^\circ) g_{ij}(\beta_n^\circ + g_i(\beta_n^\circ)^\dagger \frac{\partial g_{ij}(\beta_n^\circ)^\dagger}{\partial \beta} \right] n^{1/2}(\hat{\beta}_n - \beta_n^*) \tag{31.54}
\]

where \( \bar{L}_{j,h} \in \mathbb{R}^k \) denotes the \( (j - 1)k + 1, \ldots, jk \) components of \( \bar{L}_{h,3} \), \( (h_{5,b})_{j} \in \mathbb{R}^k \) denotes the \( j \)-th column of \( (h_{5,b})_l \in \mathbb{R}^{k \times k} \) 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 \( n^{1/2}(\bar{\Omega}_n(\hat{\beta}_n) - \Omega_{jn}) \rightarrow_d \bar{\gamma}_h^S := (\bar{\gamma}_h^S_{1,:}, \ldots, \bar{\gamma}_h^S_{h,:}) \).

By definition, \( \bar{\gamma}_h^S_{h,:} \) is a nonrandom function of \( \bar{L}_{j,h,3} - h_{5,3,j} h_{5,g}^{-1} \bar{\gamma}_h \) and \( \bar{\Omega}_h \). By \((31.53)\), \( \bar{\Omega}_h \) and \( \bar{\gamma}_h^S \) are independent. In addition, \( \bar{L}_{j,h,3} - h_{5,3,j} h_{5,g}^{-1} \bar{\gamma}_h \) and \( \bar{\gamma}_h \) are independent, because they are jointly normal with a zero covariance matrix. Therefore, \( \bar{L}_{j,h,3} - h_{5,3,j} h_{5,g}^{-1} \bar{\gamma}_h \) and \( \bar{\gamma}_h^S := h_{5,3,j}^{-1/2} \bar{\gamma}_h \) are independent. This shows that \( \bar{\gamma}_h^S := (\bar{\gamma}_h^S_{1,:}, \ldots, \bar{\gamma}_h^S_{h,:}) \) and \( \bar{\gamma}_h^S \) are independent. Equation \((31.54)\) yields \( n^{1/2}(\bar{\Omega}_n(\hat{\beta}_n) - \Omega_{jn}) \rightarrow_d \bar{\gamma}_h^S \), 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

\[
n^{1/2} \left[ \bar{G}_{j,n}(\hat{\beta}_{n}) - E_{n}G_{ij} \right]
= n^{1/2} \left[ -n^{-1} \sum_{i=1}^{n} G_{i:j,n} - E_{n}G_{ij} \right] + n^{-1} \sum_{i=1}^{n} G_{i:j}^{+}(\beta_{n}) n^{1/2}(\hat{\beta}_{n} - \beta_{n}^{*}) - \bar{f}_{j,n}(\beta_{n})^{-1/2} \left[ 2\bar{g}_{n} + G_{i:j}(\beta_{n}) n^{1/2}(\hat{\beta}_{n} - \beta_{n}^{*}) \right]
- d \bar{G}_{h:j} = T_{j,h,4} - h_{5,j}n^{1/2}(\hat{\beta}_{n}) + (h_{4,j} - h_{5,j}n^{1/2}(\hat{\beta}_{n})\bar{\beta}_{h}), \tag{31.55}
\]

where \( T_{j,h,4} \in R^{k} \) denotes the \((j - 1)k + 1, \ldots, jk \) components of \( \bar{T}_{h,4} \). Equation (31.55) and \( \bar{\beta}_{h} = (\bar{\beta}_{h,1}, \ldots, \bar{\beta}_{h,b}) \) yield \( n^{1/2}(\bar{G}_{\beta:n}(\hat{\beta}_{n}) - E_{n}G_{ij}) \rightarrow_{d} \bar{\beta}_{h}^{S} \), as desired.

By the same argument as for \( \bar{\tau}_{h}^{S} \) above, \( \bar{\tau}_{h} = (\bar{\tau}_{h,1}, \ldots, \bar{\tau}_{h,b}) \) and \( \bar{\tau}_{h}^{S} \) are independent. □

**Proof of Lemma 31.5(d).** First, we obtain the asymptotic distributions of \( \bar{\Omega}_{n}^{-1/2}(\hat{\beta}_{n}), \bar{J}_{n}(\hat{\beta}_{n}), \) and \( \bar{O}_{n} = \bar{O}_{n}(\bar{J}_{n}(\hat{\beta}_{n})) \).

Consider the function that maps \( vec(\varphi) \) onto \( vec(\varphi^{-1/2}) \), where \( \varphi \in R^{k \times k} \) is positive definite. Denote by \( \bar{\tau}_{h} \in R^{k \times k} \) the matrix of partial derivatives of that mapping evaluated at \( vec(h_{5,g}) \).

By \( n^{1/2}(\bar{\Omega}_{j,n}(\hat{\beta}_{n}) - \Omega_{j,n}) \rightarrow_{d} \bar{\tau}_{h}^{S} \), which holds by part (c) of the lemma (and is proved in (31.54)), and the delta method, we have

\[
n^{1/2} \left[ \bar{\Omega}_{n}^{-1/2}(\hat{\beta}_{n}) - \Omega_{n}^{-1/2} \right] \rightarrow_{d} vec^{-1}_{k,k}(\bar{\tau}_{h}vec(\bar{\tau}_{h}^{S})). \tag{31.56}
\]

The asymptotic distribution \( \bar{J}_{n}(\hat{\beta}_{n}) = \bar{\Omega}_{n}^{-1/2}(\hat{\beta}_{n})\bar{G}_{\beta:n}(\hat{\beta}_{n}) \) is obtained as follows:

\[
n^{1/2} \left[ \bar{J}_{n}(\hat{\beta}_{n}) - \Omega_{n}^{-1/2}E_{n}G_{i:j} \right]
= \bar{\Omega}_{n}^{-1/2}(\hat{\beta}_{n})n^{1/2} \left[ \bar{G}_{\beta:n}(\hat{\beta}_{n}) - E_{n}G_{i:j} \right] + n^{1/2} \left[ \bar{\Omega}_{n}^{-1/2}(\hat{\beta}_{n}) - \Omega_{n}^{-1/2} \right] E_{n}G_{i:j}
- d \bar{\tau}_{h}^{S} = h_{5,g}^{-1}\bar{\beta}_{h} + vec^{-1}_{k,k}(\bar{\tau}_{h}vec(\bar{\tau}_{h}^{S}))h_{4,\beta}, \tag{31.57}
\]

where the convergence uses (31.56) and \( n^{1/2}(\bar{G}_{\beta:n}(\hat{\beta}_{n}) - E_{n}G_{i:j}) \rightarrow_{d} \bar{\tau}_{h}^{S} \), 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

\[
O_{n} = O(J_{n}) = ((-j_{n1}'(J_{n1})^{-1}, e_{1}', \ldots, -j_{n(k-b)}'(J_{n(k-b)})^{-1}, e_{k-b}')') \quad \text{and} \quad \bar{O}_{n} = O(J_{n}) = ((-\bar{j}_{n1}'(\bar{J}_{n1})^{-1}, e_{1}', \ldots, -\bar{j}_{n(k-b)}'(\bar{J}_{n(k-b)})^{-1}, e_{k-b}')'), \tag{31.58}
\]

\(^{75}\)If the first \( b \) columns of \( (h_{5,g}^{-1/2}h_{4,\beta})' \) are linearly dependent, \( O_{n} \) and \( \bar{O}_{n} \) are given by analogous formulas involving \( R_{n} = E_{n}G_{i:j}^{+} \Omega_{n}^{-1/2} = (R_{n1}, R_{n2}) \) and \( \bar{R}_{n}(\beta_{n})' = (\bar{R}_{n1}, \bar{R}_{n2}) \) just based on a different set of \( b \) columns of \( (h_{5,g}^{-1/2}h_{4,\beta})' \).
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 \( J'_n \). Consider the function that maps \( vec(J) \) for \( J \in R^{k \times b} \) onto \( vec(O(J)) \in R^{b(k-b)} \), where \( O(J) \) is defined by (31.4) and (31.5). Denote by \( \overline{B}_h \in R^{b(k-b) \times kb} \) the matrix of partial derivatives of that mapping evaluated at \( vec(h_{n,g}^{-1/2} h_{4,\beta}) \). Then, by the delta method,

\[
n^{1/2}(\tilde{O}_n - O_n) \to_d vec^{-1}_{k,b-k-b}(\overline{B}_h vec(\overline{w}_h^S))
\]  

(31.59)

and the asymptotic distribution is independent of \( \overline{w}_h^S \).

Given the asymptotic distributions of \( \tilde{O}_n^{-1/2}(\hat{\beta}_n) \) and \( \tilde{O}_n \), the asymptotic distribution of \( n^{1/2}\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n)U_nT_n \) is obtained as follows. We write this matrix in terms of two submatrices:

\[
n^{1/2}\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n)U_nB_nS_n = (\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n)U_nB_{n,q^S} S_n^{-1} Y_{n,q^S}, n^{1/2}\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n)U_{n,p-q^S}) \]  

(31.60)

Consider the first component on the rhs of (31.60). By definition, the singular value decomposition of \( O'_n\Omega_n^{-1/2}D_nU_n \) is \( C_n Y_n B'_n \) and, as in the proof of the full vector case preceding (25.5), we have \( O'_n\Omega_n^{-1/2}D_nU_nB_{n,q^S} Y_{n,q^S} = C_{n,q^S} \). Hence, we obtain

\[
\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n)U_nB_{n,q^S} Y_{n,q^S}^{-1} = O'_n\Omega_n^{-1/2}D_nU_nB_{n,q^S} Y_{n,q^S}^{-1} + n^{1/2}(\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n) - O'_n\Omega_n^{-1/2}D_n)U_nB_{n,q^S}(n^{1/2}Y_{n,q^S})^{-1} \]

\[
= C_{n,q^S} + o_p(1)
\]

\[
\to_p \sum_{h,q^S}^S := h_{3,q^S} \in R^{(k-b) \times q^S},
\]  

(31.61)

where the second equality uses \( n^{1/2}(\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n) - O'_n\Omega_n^{-1/2}D_n) = O_p(1) \) and \( n^{1/2} \tau_{ijn} \to \infty \) for all \( j \leq q^S \) (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

\[
n^{1/2}(\tilde{O}_n^{-1/2}(\hat{\beta}_n)\tilde{D}_n(\hat{\beta}_n) - O'_n\Omega_n^{-1/2}E_nG_i)
\]

\[
= \tilde{O}_n^{-1/2}(\hat{\beta}_n)n^{1/2}(\tilde{D}_n(\hat{\beta}_n) - E_nG_i) + O'_n n^{1/2}(\tilde{O}_n^{-1/2}(\hat{\beta}_n) - \Omega_n^{-1/2})E_nG_i
\]

\[
+ n^{1/2}(\tilde{O}_n^{-1/2}(\hat{\beta}_n) - O'_n\Omega_n^{-1/2})E_nG_i.
\]  

(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
\[ n^{1/2}O'_n \omega^{-1/2}D_n U_n B_{n,p-q} \to h_3 h^{q}_{1,p-q} \].

By (31.52), (31.56), and (31.59), we have
\[ n^{1/2}(\tilde{\Omega}_n^{1/2}(\tilde{\beta}_n) \tilde{D}_n(\tilde{\beta}_n) - O_n^{1/2}D_n) \]
\[ = n^{1/2}(\tilde{\Omega}_n - O_n)\tilde{\Omega}_n^{-1/2}(\tilde{\beta}_n) \tilde{D}_n(\tilde{\beta}_n) + O_n n^{1/2}(\tilde{\Omega}_n^{-1/2}(\tilde{\beta}_n) - \Omega_n^{-1/2}) \tilde{D}_n(\tilde{\beta}_n) \]
\[ + O_n \Omega_n^{-1/2}n^{1/2}(\tilde{D}_n(\tilde{\beta}_n) - D_n) \]
\[ \to_d \chi_n := \text{vec}_{k,k-b}(\tilde{\Omega}_h \text{vec}(\tilde{\Omega}_h^S))'h_{5,g}^{-1/2}h_4 + O(h_{5,g}^{-1/2}h_{4,b})' \text{vec}_{k,k-b}(\tilde{\Omega}_h \text{vec}(\tilde{\Omega}_h^S))h_4 + O(h_{5,g}^{-1/2}h_{4,b})'h_{5,g}^{-1/2}D_h. \] (31.64)

Using (31.63) and (31.64), we obtain
\[ n^{1/2}O'_n \omega^{-1/2}(\tilde{\beta}_n) \tilde{D}_n(\tilde{\beta}_n)U_n B_{n,p-q} \]
\[ = n^{1/2}O_n \omega^{-1/2}D_n U_n B_{n,p-q} + n^{1/2}(\tilde{\Omega}_n^{1/2}(\tilde{\beta}_n) \tilde{D}_n(\tilde{\beta}_n) - O_n^{1/2}D_n)U_n B_{n,p-q} \]
\[ \to_d \Delta^S_{h,p-q} := h_3 h^{q}_{1,p-q} + \chi_n h_8 h_{2,p-q} \in R^{(k-b) \times (p-q)}, \] (31.65)

where \( B_{n,p-q} \to h_{2,p-q}, U_{2n} \to h_8, \) 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 \( \Delta^S_{h} := (\Delta^S_{h,q}, \Delta^S_{h,p-q}) \) by (31.22).

We have \( (\tilde{\beta}_h, \tilde{D}_h, \tilde{\Omega}_h^S, \tilde{\beta}_h^S, \tilde{\Omega}_h^S) \) is independent of \( \tilde{\gamma}_h^S \) because \( \tilde{\Omega}_h^S \) is a nonrandom function of \( h \) and \( (\tilde{\Omega}_h^S, \tilde{\beta}_h^S, \tilde{\Omega}_h^S) \), see (31.22), and \( (\tilde{\beta}_h, \tilde{D}_h, \tilde{\Omega}_h^S, \tilde{\Omega}_h^S) \) is independent of \( \tilde{\gamma}_h^S \) 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, \( \theta_{0n} \) is the true value that may vary with \( n \). For notational simplicity, we suppress the dependence of various quantities on \( \theta_{0n} \).

**Lemma 31.10** Suppose Assumption gB holds. Then, for any sequence \( \{ (F_n, \beta^*_n, \theta_{0n}) \in \mathcal{F}^{S,S,R}_{\Theta,AR} : n \geq 1 \} \), (a) \( \widehat{\gamma}_n(\beta_n) = r_{F_n}(\beta_n) = r_{F_n}(\beta_n^*) \) wp→1, (b) \( \text{col}(\widehat{\Lambda}_n(\beta_n)) = \text{col}(A_{F_n}(\beta_n)) = \text{col}(A_{F_n}(\beta_n^*)) \) wp→1, and (c) given the first-stage estimator \( \widetilde{\beta}_n \), the statistics \( \text{SR-AR}_n^S(\widetilde{\beta}_n) \), \( \text{SR-QLR}_n^S(\widetilde{\beta}_n) \),
\[ \chi^2_{\tau_n(\beta_n),1-\alpha} = c_{\tau_n(\beta_n),p}(n^{1/2} \tilde{D}^*_{\Lambda_n}(\tilde{\beta}_n), \tilde{J}_{\Lambda_n}(\tilde{\beta}_n)), 1 - \alpha \] are invariant \( \text{wp} \rightarrow 1 \) to the replacement of \( \tilde{\tau}_n(\beta_n) \) and \( \Lambda_n(\beta_n)' \) by \( r_{F_n}(\beta_n) \) and \( \Pi_{1F_n}^{-1/2}(\beta_n) A_{F_n}(\beta_n)' \), respectively.

**Proof of Lemma 31.10**

First, we establish part (a). For any \( \beta \in B(\beta_n^*, \varepsilon) \),

\[ \lambda \in N(\Omega_{F_n}(\beta)) \implies \lambda \in \cap_{\beta \in B(\beta_n^*, \varepsilon)} N(\Omega_{F_n}(\beta)) \]

\[ \implies \sup_{\beta \in B(\beta_n^*, \varepsilon)} \lambda' \Omega_{F_n}(\beta) \lambda = 0 \implies \sup_{\beta \in B(\beta_n^*, \varepsilon)} \text{Var}_{F_n}(\lambda' g_i(\beta)) = 0 \]

\[ \implies \sup_{\beta \in B(\beta_n^*, \varepsilon)} |\lambda' g_i(\beta) - E_{F_n} \lambda' g_i(\beta)| = 0 \text{ a.s.}[F_n] \]

\[ \implies \lambda \in \cap_{\beta \in B(\beta_n^*, \varepsilon)} N(\tilde{\Omega}_n(\beta)) \text{ a.s.}[F_n], \]

(31.66)

where the first implication holds by condition (iv) of \( \mathcal{F}^{S,SR}_{\mathcal{A}R,2} \). From the proof of Lemma 31.5 under sequences \( \{(F_n, \beta_n^*, \theta_0_n) \in \mathcal{F}^{S,SR}_{\mathcal{O}AR} : n \geq 1\} \), we have that \( \beta_n - \beta_n^* \rightarrow_p 0 \). Thus, \( \text{wp} \rightarrow 1 \) if follows that \( \tilde{\beta}_n \in B(\beta_n^*, \varepsilon) \). Thus, from (31.66), \( N(\Omega_{F_n}(\tilde{\beta}_n)) \subseteq N(\tilde{\Omega}_n(\tilde{\beta}_n)) \) \( \text{wp} \rightarrow 1 \) and \( \tilde{\tau}_n(\tilde{\beta}_n) \leq r_{F_n}(\tilde{\beta}_n) \) \( \text{wp} \rightarrow 1 \).

Next we prove \( \tilde{\tau}_n(\tilde{\beta}_n) \geq r_{F_n}(\tilde{\beta}_n) \) \( \text{wp} \rightarrow 1 \). By considering subsequences, it suffices to consider the case where \( r_{F_n}(\beta_n^*) = r \) for all \( n \geq 1 \) for some \( r \in \{0, 1, ..., k\} \). We have

\[ \tilde{\tau}_n(\tilde{\beta}_n) = rk(\tilde{\Omega}_n(\tilde{\beta}_n)) \geq rk(\Pi_{1F_n}^{-1/2}(\beta_n) A_{F_n}(\beta_n)' \tilde{\Omega}_n(\beta_n) A_{F_n}(\beta_n) \Pi_{1F_n}^{-1/2}(\beta_n)) \]

(31.67)

because \( \tilde{\Omega}_n(\tilde{\beta}_n) \) is \( k \times k \), the matrix \( A_{F_n}(\beta_n) \Pi_{1F_n}^{-1/2}(\beta_n) \) is \( k \times r \) \( \text{wp} \rightarrow 1 \) by condition (iv) of \( \mathcal{F}^{S,SR}_{\mathcal{A}R,2} \) and consistency of \( \beta_n^* \), and wlog \( 1 \leq r \leq k \). (If \( r = 0 \), then the desired inequality \( \tilde{\tau}_n(\beta_n) \geq 0 = r_{F_n}(\beta_n) \) holds trivially \( \text{wp} \rightarrow 1 \), where the equality holds by condition (iv) of \( \mathcal{F}^{S,SR}_{\mathcal{A}R,2} \) and consistency of \( \beta_n^* \).)

From condition (iv) of \( \mathcal{F}^{S,SR}_{\mathcal{A}R,2} \), it follows that \( A_{F_n}(\beta_n) = A_{F_n}(\beta_n^*) \) and therefore \( A_{F_n}(\beta) \) does not depend on \( \beta \) 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,

\[ \Pi_{1F_n}^{-1/2}(\beta_n) A_{F_n}(\beta_n) = n^{-1} \sum_{i=1}^{n} \Pi_{1F_n}^{-1/2}(\beta_n) A_{F_n}(g_i(\beta) - E_{F_n} g_i(\beta)) (g_i(\beta) - E_{F_n} g_i(\beta))' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta_n) - \]

\[ n^{-1} \sum_{i=1}^{n} \Pi_{1F_n}^{-1/2}(\beta_n) A_{F_n}(g_i(\beta) - E_{F_n} g_i(\beta))' \left[ n^{-1} \sum_{i=1}^{n} (g_i(\beta) - E_{F_n} g_i(\beta))' A_{F_n} \Pi_{1F_n}^{-1/2}(\beta_n) \right]. \]

(31.68)
By construction and using condition (iv) of $F_{AR,2}^{S,SR}$, we have, for all $\beta \in B(\beta_n^*, \varepsilon)$,

$$E_F \Pi_{1}^{-1/2}(\beta)A_F(g(\beta) - E_F g(\beta)) (g(\beta) - E_F g(\beta))^\prime A_F \Pi_{1}^{-1/2}(\beta) = I_r.$$ 

By the uniform moment bound in $F_{\Theta,AR}^{S,SR}$, namely, $E_F \sup_{\beta \in B(\beta^*, \varepsilon)} \| \Pi_{1}^{-1/2}(\beta)A_F(\beta) (g(\beta) - E_F g(\beta)) \|^2 \leq M$ and continuity of $(g(\beta) - E_F g(\beta))^\prime A_F \Pi_{1}^{-1/2}(\beta)$ as a function of $\beta$ (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

$$\Pi_{1}^{-1/2}(\beta_n)A_F(\beta_n)\hat{\Omega}_n(\beta_n)A_F(\beta_n)\Pi_{1}^{-1/2}(\beta_n) \rightarrow_p I_r.$$ 

This establishes that $\hat{\tau}_n(\beta_n) \geq r$ wp→ 1 and therefore $\hat{\tau}_n(\beta_n) = r$ and $N(\Omega_F(\beta_n)) = N(\hat{\Omega}_n(\beta_n))$ wp→ 1, which proves (a). In turn, the latter implies that $\text{col}(A_F(\beta_n)) = \text{col}(\hat{\Omega}_n(\beta_n))$ wp→ 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 $\hat{\tau}_n(\beta_n) = 0$ and $\hat{\tau}_n(\beta_n) \equiv 0$ wp→ 1 when $r = 0$ by part (a). Part (b) of the Lemma implies that there exists a random $r \times r$ nonsingular matrix $\hat{M}_n$ such that

$$\hat{A}_n(\beta_n) = A_F(\beta_n^*) \Pi_{1}^{-1/2}(\beta_n^*) \hat{M}_n \text{ wp→ 1}, \quad (31.69)$$

because $\Pi_{1}^{-1/2}(\beta_n^*)$ is nonsingular (since by its definition it is a diagonal matrix with the positive eigenvalues of $\Omega_F(\beta_n^*)$ on its diagonal.) Equation (31.69) and $\hat{\tau}_n(\beta_n) = r$ wp→ 1 imply that the statistics $SR-AR^nS(\beta_n^*, \beta_n^*)$, $SR-QLR^nS(\beta_n^*, \beta_n^*)$, $\chi^2_{\hat{\tau}_n(\beta_n),1-\alpha}$, $c_{\hat{\tau}_n(\beta_n),p}(n^{1/2}d_{\hat{A}_n(\beta_n^*)})$, $I_{\hat{A}_n(\beta_n^*)}(1-\alpha)$, are invariant wp→ 1 to the replacement of $\hat{\tau}_n(\beta_n)$ and $\hat{A}_n(\beta_n^*)$ by $r$ and $A_F(\beta_n^*) \Pi_{1}^{-1/2}(\beta_n^*) \hat{M}_n$, respectively. Now we apply the invariance results of Lemma 31.2 with $(k, g(\beta), G_1(\beta))$ replaced by $(r, \Pi_{1}^{-1/2}(\beta_n^*)A_F(\beta_n^*) g(\beta), \Pi_{1}^{-1/2}(\beta_n^*)A_F(\beta_n^*) G_1(\beta))$ and with $M$ equal to $\hat{M}_n$. These results imply that the previous four statistics when based on $r$ and $\Pi_{1}^{-1/2}(\beta_n^*)A_F(\beta_n^*) g(\beta)$ are invariant to the multiplication of the moments $\Pi_{1}^{-1/2}(\beta_n^*)A_F(\beta_n^*) g(\beta)$ by the nonsingular matrix $\hat{M}_n$. Thus, the statistics, defined as in Section 5.2, are invariant wp→ 1 to the replacement of $\hat{\tau}_n(\beta_n)$ and $\hat{A}_n(\beta_n)$ by $r$ and $\Pi_{1}^{-1/2}(\beta_n^*)A_F(\beta_n^*)$, 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 \( F_S \subset F_{AR}^{SR} \) (because \( F_S \) imposes \( \lambda_{\min}(E_F g_i g_i') \geq \delta \), where \( F_{AR}^{SR} \) is defined in (9.16)), we have \( \tilde{r}_n(\beta_n) = r_{F_n}(\beta_n) = r_{F_n}(\beta_n^*) \) and \( \text{col}(\tilde{A}_n(\beta_n)) = \text{col}(A_{F_n}(\beta_n)) = \text{col}(A_{F_n}(\beta_n^*)) \) wp→1. Also, given \( \lambda_{\min}(E_F g_i g_i') \geq \delta \), it follows that the orthogonal matrix \( A_{F_n}(\beta_n^*) \) is in \( R^{k \times k} \). Given that the statistics \( QLR_n^2(\beta_n) \) and \( c_k,p(n^{1/2} \tilde{D}_n^*(\beta_n), \tilde{J}_n(\beta_n), 1 - \alpha) \) 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 \( SR-QLR_n^S(\theta_0, \beta_{\tilde{A}_n}) = QLR_n^S(\eta) + o_p(1) \). Because \( \tilde{r}_n(\beta_n) = k \) wp→1, it follows that \( c_{\tilde{r}_n(\theta_0, \beta_{\tilde{A}_n})}(n^{1/2} \tilde{D}_{\tilde{A}_n}^*(\theta_0, \beta_{\tilde{A}_n}), \tilde{J}_{\tilde{A}_n}(\theta_0, \beta_{\tilde{A}_n}), 1 - \alpha) = c_{k,p}(n^{1/2} \tilde{D}_n^*(\eta), \tilde{J}_n(\eta), 1 - \alpha) \) wp→1, where the latter critical value is the one for the subvector CQLR test without singularity robustness, see (9.12). This proves the first equalities in parts (a) and (b).

We now proceed as in the proof of Theorem 7.1. We replace \( \widehat{W}_n, W_{F_n}, \widehat{\Omega}^{-1/2}_n, \) and \( \bar{D}_n \in R^{k \times p} \) by the corresponding quantities \( I_k, I_k, \bar{O}'_F \Omega_{\tilde{A}_n}^{-1/2}(\eta) \tilde{g}_n(\eta), \) and \( \bar{O}'_F \Omega_{\tilde{A}_n}^{-1/2}(\eta) \tilde{D}_n(\eta) \in R^{(k-b) \times p} \), respectively. Note that \( q^S = p \) under \( \{\lambda^S_{n,h} : n \geq 1\} \). The analogue to (28.2) with \( q^S = p \) therefore states that

\[
\tilde{r}_n^{(p+1)} = n \tilde{g}'_n(\eta) \Omega_n^{-1/2} \bar{O}'_F h_{3,k-p} h_{3,k-p}' \bar{O}'_F \Omega_n^{-1/2}(\eta) \tilde{g}_n(\eta) + o_p(1). \tag{31.70}
\]

In addition, the analogue to (28.3) with \( q^S = p \) states that

\[
QLR_{W,n} = n \tilde{g}'_n(\eta) \Omega_n^{-1/2} \bar{O}'_F h_{3,p} h_{3,p}' \bar{O}'_F \Omega_n^{-1/2}(\eta) \tilde{g}_n(\eta) + o_p(1), \tag{31.71}
\]

where \( QLR_{W,n} := AR_n^S(\eta) - \lambda_{\min}(n \tilde{Q}_{W,n}) \) is defined below Proposition 31.6 and equals \( QLR_n^S(\eta) \) when \( \tilde{U}_n \) is taken to be \( \tilde{L}_n^{1/2}(\eta) \). Equation (31.61) implies that \( h_{3,p} = \bar{O}'_F \Omega_{\tilde{A}_n}^{-1/2}(\eta) \tilde{D}_n(\eta) U_n B_{n,p} \Upsilon_{n,p}^{-1} + o_p(1) \). Because \( U_n B_{n,p} \Upsilon_{n,p}^{-1} \) is an invertible matrix, it follows that \( P_{h_{3,p}} = P_{\bar{O}'_F} \Omega_n^{-1/2}(\eta) \tilde{D}_n(\eta) + o_p(1) \). Therefore, using \( h_{3,p}' = I_p \), it follows that

\[
QLR_{W,n} = n \tilde{g}'_n(\eta) \Omega_n^{-1/2} \bar{O}'_F P_{\bar{O}'_F} \Omega_n^{-1/2}(\eta) \tilde{g}_n(\eta) + o_p(1). \tag{31.72}
\]

By (31.48), \( \Omega_n^{-1/2}(\eta) n^{1/2} \tilde{g}_n(\eta) = M_{J_h} h_{5,g}^{-1/2} n^{1/2} \tilde{g}_n(\theta_0, \beta_n^*) + o_p(1) \), where \( J_h = h_{5,g}^{-1/2} h_{4,\beta} \). Also, \( \bar{O}_F = O(J_h) + o_p(1) \) and \( \tilde{\Omega}_n^{-1/2}(\eta) \tilde{D}_n(\eta) = J_{gh} + o_p(1) \), where \( J_{gh} := h_{5,g}^{-1/2} h_4 \) and \( J_{gh} \) has full column rank \( p \). Thus, we obtain

\[
QLR_n^S(\eta) = n^{1/2} \tilde{g}_n(\theta_0, \beta_n^*) h_{5,g}^{-1/2} M_{J_h} O(J_h) P_{O(J_h)' J_{gh}} O(J_h)' M_{J_h} h_{5,g}^{-1/2} n^{1/2} \tilde{g}_n(\theta_0, \beta_n^*) + o_p(1) \\
= n^{1/2} \tilde{g}_n(\theta_0, \beta_n^*) h_{5,g}^{-1/2} P_{M_{J_h} J_{gh} h_{5,g}^{-1/2}} n^{1/2} \tilde{g}_n(\theta_0, \beta_n^*) + o_p(1), \tag{31.73}
\]
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

$$LM_n^S = n^{1/2}g_n(\theta_0, \beta^*_n)h_{5,g}^{-1/2}(M_{J_h}P_{[J_{\theta h} : J_h]}M_{J_h})h_{5,g}^{-1/2}g_n(\theta_0, \beta^*_n) + o_p(1) \quad (31.74)$$

using $h_{5,g}^{-1/2}\tilde{G}_{\eta(n)}(\tilde{\eta}) \rightarrow_p [J_{\theta h} : J_h] := [h_{5,g}^{-1/2}h_4 : h_{5,g}^{-1/2}h_4, \beta]$ and $[J_{\theta h} : J_h]$ has full column rank $p + b$.

Next, we have

$$M_{J_h}P_{[J_{\theta h} : J_h]}M_{J_h} = M_{J_h}P_{[M_{J_h}, J_{\theta h} : J_h]}M_{J_h} = M_{J_h} \left( P_{M_{J_h}J_{\theta h}} + P_{J_h} \right) M_{J_h} = P_{M_{J_h}J_{\theta h}}, \quad (31.75)$$

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 $QLR_n^S(\tilde{\eta}) = LM_n^S + o_p(1)$, which establishes the second equality of part (a).

By (31.31), $c_{k,p}(n^{1/2}\tilde{D}_n^* (\tilde{\eta}), \tilde{J}_n(\tilde{\eta}), 1 - \alpha) + o_p(1) \rightarrow_p c_{k - b, p, q, S}(h'_{3,k - b - q, S}A^S_{h, p - q, S}, 1 - \alpha)$, 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 $A^S_{h, p - q, S}$ has no columns, $ACLR_{k,p,q}(\tau_c) = Z[\chi^2_p, Z_1 \sim \chi^2_p$, and $c_{k,p,q}(h'_{3,k - b - q, S} \times \tilde{\Delta}_{h, p - q, S}, 1 - \alpha)$ equals the $1 - \alpha$ quantile of the $\chi^2_p$ distribution. Hence, the convergence result in part (b) holds. \(\square\)
References

Ahn, S. C., and P. Schmidt (1995): “Efficient Estimation of Models for Dynamic Panel Data,” *Journal of Econometrics*, 68, 5–27.

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

——— (2017): “Identification-Robust Subvector Inference,” Cowles Foundation Discussion Paper No. 3003, 2017.

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.econometricssociety.org/ecta/Supmat/9456_miscellaneous.pdf.

——— (2013a): “GMM Estimation and Uniform Subvector Inference with Possible Identification Failure,” *Econometric Theory*, 30, 1–47.

——— (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.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2019): “Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests,” forthcoming in *Journal of Econometrics*.

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.

——— (2017): “Asymptotic Size of Kleibergen’s LM and Conditional LR Tests for Moment Condition Models,” *Econometric Theory*, 33, 1046-1080.

Andrews, D. W. K., V. Marmer, and Z. Yu (2019): “On Optimal Inference in the Linear IV Model,” *Quantitative Economics*, 10, forthcoming.

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

——— (2008): “Efficient Two-sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments,” *Journal of Econometrics*, 146, 241–254.
Andrews, D. W. K., and W. Ploberger (1994): “Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative,” *Econometrica*, 62, 1383–1414.

Andrews, I. (2016): “Conditional Linear Combination Tests for Weakly Identified Models,” *Econometrica*, 84, 2155–2182.

—— (2018): “Valid Two-Step Identification-Robust Confidence Sets for GMM,” *Review of Economics and Statistics*, 100, 337–348.

Andrews, I., and A. Mikusheva (2015): “Maximum Likelihood Inference in Weakly Identified Models,” *Quantitative Economics*, 6, 123–152.

—— (2016): “A Geometric Approach to Nonlinear Econometric Models,” *Econometrica*, 84, 1249–1264.

Angrist, J. D., and A. B. Krueger (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, 106, 979–1014.

—— (1992): “Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery,” NBER Working Paper No. 4067.

Antoine, B., and E. Renault (2009): “Efficient GMM with Nearly Weak Instruments,” *Econometrics Journal*, 12, S135–S171.

—— (2010): “Efficient Inference with Poor Instruments, a General Framework,” in *Handbook of Empirical Economics and Finance*, ed. by D. Giles and A. Ullah. Oxford: Taylor and Francis.

Arellano, M., and O. Bover (1995): “Another Look at Instrumental Variable Estimation of Error Component Models,” *Journal of Econometrics*, 68, 29–51.

Armstrong, T. B. (2016): “Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply,” *Econometrica*, 84, 1961–1980.

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.

Billingsley, P. (1979): *Probability and Measure*. New York: John Wiley and Sons.

Blundell, R., and S. Bond (1995): “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,” Working Paper No. W95/17, The Institute for Fiscal Studies, London.
Canova, F., and L. Sala (2009): “Back to Square One: Identification Issues in DSGE Models,” *Journal of Monetary Economics*, 56, 431–449.

Carroll, C. D., J. Slacalek, and M. Sommer (2011): “International Evidence on Sticky Consumption Growth,” *Review of Economics and Statistics*, 93, 1135–1145.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): “Inference in Models with Nearly Integrated Regressors,” *Econometric Theory*, 11, 1131–1147.

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.

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.

Chen, X., M. Ponomareva, and E. Tamer (2014): “Likelihood Inference in Some Finite Mixture Models,” *Journal of Econometrics*, 182, 87–99.

Cheng, X. (2015): “Uniform Inference in Nonlinear Models with Mixed Identification Strength,” *Journal of Econometrics*, 189, 207–228.

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

Cho, J. S., and H. White (2007): “Testing for Regime Switching,” *Econometrica*, 75, 1671–1720.

Choi, I., and Phillips, P. C. B. (1992): “Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations,” *Journal of Econometrics*, 51, 113–150.

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.

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

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

Davies, R. B. (1977): “Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative,” *Biometrika*, 64, 247–254.

Dufour, J.-M. (1989): “Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions,” *Econometrica*, 57, 335–355.

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.

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.

—— (2013): “Identification-Robust Analysis of DSGE and Structural Macroeconomic Models,” *Journal of Monetary Economics*, 60, 340–350.

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

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.

Grant, N. (2013): “Identification Robust Inference with Singular Variance,” Economics Discussion Paper No. 1315, Department of Economics, University of Manchester.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): “Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models, *Quantitative Economics*, 4, 197–229.

Guggenberger, P. (2012): “On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Condition,” *Econometric Theory*, 28, 387–421.

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

Hannan, E. J. (1982): “Testing for Autocorrelation and Akaike’s Criterion,” in Essays in Statistical Science, ed. by J. Gani and E. J. Hannan, *Journal of Applied Probability*, Special Volume 19A.

Hansen, L. P. and J. A. Scheinkman (1995): “Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes,” *Econometrica*, 63, 767–804.

Hillier, G. (2009): “Exact Properties of the Conditional Likelihood Ratio Test in an IV Regression Model,” *Econometric Theory*, 25, 915–957.

Iskrev, N. (2010): “Local Identification in DSGE,” *Journal of Monetary Economics*, 57, 189–202.

Jegannathan, R., G. Skoulakis, and Z. Wang (2002): “Generalized Method of Moments: Applications in Finance,” *Journal of Business and Economic Statistics*, 20, 470–481.

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

Kleibergen, F. (2004): “Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model,” *Review of Economics and Statistics*, 86, 418–423.

— — — (2005): “Testing Parameters in GMM Without Assuming That They Are Identified,” *Econometrica*, 73, 1103–1123.

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

Kleibergen, F., and S. Mavroeidis (2009): “Weak Instrument Robust Tests and the New Keynesian Phillips Curve,” *Journal of Business and Economic Statistics*, 27, 293–311.

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

McCloskey, A. (2017): “Bonferroni-Based Size-Correction for Nonstandard Testing Problems,” *Journal of Econometrics*, 200, 17–35.

Mikusheva, A. (2010): “Robust Confidence Sets in the Presence of Weak Instruments,” *Journal of Econometrics*, 157, 236–247.
Montiel Olea, J. L. (2012): “Efficient Conditionally Similar-on-the-Boundary Tests,” unpublished manuscript, Department of Economics, New York University.

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

Nason, J. M., and G. W. Smith (2008): “Identifying the New Keynesian Phillips Curve,” *Journal of Applied Econometrics*, 23, 525–551.

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.

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.

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

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

Phillips, P. C. B. (1989): “Partially Identified Econometric Models,” *Econometric Theory*, 5, 181–240.

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

Puhani, P. A. (2000): “The Heckman Correction for Sample Selection and its Critique,” *Journal of Economic Surveys*, 14, 53–68.

Qu, Z. (2014): “Inference in DSGE Models with Possible Weak Identification,” *Quantitative Economics*, 5, 457–494.

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.

Robin, J.-M., and R. J. Smith (2000): “Tests of Rank,” *Econometric Theory*, 16, 151–175.
Schorfeide, F. (2014): “Estimation and Evaluation of DSGE Models: Progress and Challenges,” in Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, ed. by D. Acemoglu, M. Arellano, and E. Dekel, Vol. III, 184–230. Cambridge, UK: Cambridge University Press.

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

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

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

Teräsvirta, T. (1994): “Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models,” Journal of the American Statistical Association, 89, 208–218.

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

Yogo, M. (2004): “Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak,” Review of Economics and Statistics, 86, 797–810.
FIGURE SM-1: Power of CQLR, AM, S, and AR as function of $\theta$ for $\rho = 0$ and $\pi = 1, .5, .2, .1$; first/second row $g = 3/4$
FIGURE SM-2: Power for Kleibergen's MVW-CLR and Other Tests in Linear IV Model

SM-2(a) Design 1

- SR – CQLRP
- SR – CQLR
- Mor – CLR
- MVW – CLR
- JVW – CLR
- LM
- AR

SM-2(b) Design 2

Electronic copy available at: https://ssrn.com/abstract=3366443