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

By

Donald W. K. Andrews and Patrik Guggenberger

December 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1977

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

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

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

Donald W. K. Andrews
Cowles Foundation
Yale University

Patrik Guggenberger
Department of Economics
Pennsylvania State University

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

*Andrews and Guggenberger gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-1058376 and SES-1355504, and SES-1021101 and SES-1326827, respectively. The authors thank Isaiah Andrews, Xu Cheng, Anna Mikusheva, and Jim Stock and the participants of seminars at the following universities for helpful comments: 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

An influential paper by Kleibergen (2005) introduces Lagrange multiplier (LM) and conditional likelihood ratio-like (CLR) tests for nonlinear moment condition models. These procedures aim to have good size performance even when the parameters are unidentified or poorly identified. However, the asymptotic size and similarity (in a uniform sense) of these procedures has not been determined in the literature. This paper does so.

This paper shows that the LM test has correct asymptotic size and is asymptotically similar for a suitably chosen parameter space of null distributions. It shows that the CLR tests also have these properties when the dimension p of the unknown parameter θ equals 1. When $p \geq 2$, however, the asymptotic size properties are found to depend on how the conditioning statistic, upon which the CLR tests depend, is weighted. Two weighting methods have been suggested in the literature. The paper shows that the CLR tests are guaranteed to have correct asymptotic size when $p \geq 2$ with one weighting method, combined with the Robin and Smith (2000) rank statistic. The paper also determines a formula for the asymptotic size of the CLR test with the other weighting method. However, the results of the paper do not guarantee correct asymptotic size when $p \geq 2$ with the other weighting method, because two key sample quantities are not necessarily asymptotically independent under some identification scenarios.

Analogous results for confidence sets are provided. Even for the special case of a linear instrumental variable regression model with two or more right-hand side endogenous variables, the results of the paper are new to the literature.

Keywords: asymptotics, conditional likelihood ratio test, confidence set, identification, inference, Lagrange multiplier test, moment conditions, robust, test, weak identification, weak instruments.

JEL Classification Numbers: C10, C12.

1 Introduction

We consider the moment condition model

$$E_F g(W_i, \theta) = 0^k, \tag{1.1}$$

where $0^k = (0, ..., 0)' \in R^k$, the equality holds when $\theta \in \Theta \subset R^p$ is the true value, $\{W_i \in R^m : i = 1, ..., n\}$ are stationary and strong mixing observations with distribution F, g is a known (possibly nonlinear) function from R^{m+p} to R^k with $k \geq p$, and $E_F(\cdot)$ denotes expectation under F. This paper is concerned with tests of the null hypothesis

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \tag{1.2}$$

We consider the Lagrange Multiplier (LM) test of Kleibergen (2005) and adaptations of Moreira's (2003) conditional likelihood ratio (CLR) test to the nonlinear moment condition model (1.1), as in Kleibergen (2005, 2007), Smith (2007), Newey and Windmeijer (2009), and Guggenberger, Ramalho, and Smith (2012). The LM and CLR tests are designed to have better overall power than the Anderson and Rubin (1949)-type S-tests of Stock and Wright (2000) when $k > p$.[1]

These tests aim to have good size even when the parameters are unidentified or weakly identified. Weak identification and weak instruments (IV's) can occur in a wide variety of empirical applications in economics with linear and nonlinear models. Examples include: new Keynesian Phillips curve models, dynamic stochastic general equilibrium (DSGE) models, consumption capital asset pricing models (CCAPM), interest rate dynamics models, Berry, Levinsohn, and Pakes (1995) (BLP) models of demand for differentiated products, returns-to-schooling equations, nonlinear regression, autoregressive-moving average models, GARCH models, smooth transition autoregressive (STAR) models, parametric selection models estimated by Heckman's two step method or maximum likelihood, mixture models, regime switching models, and all models where hypotheses testing problems arise in which a nuisance parameter appears under the alternative hypothesis, but not under the null. For references, see (for example) Andrews and Guggenberger (2014a) (hereafter AG2).

The contribution of the paper is to determine the asymptotic sizes of the tests listed above, and the confidence sets (CS's) that correspond to them, for suitably defined parameter spaces of distributions, and to see whether their asymptotic sizes necessarily equal their nominal sizes. We

[1] For the special case of the linear IV model, power comparisons (some theoretical and some simulation based) are given in Kleibergen (2002), Moreira (2003), Andrews, Moreira, and Stock (2006, 2008), Chernozhukov, Hansen, and Jansson (2009), Hillier (2009), Mikusheva (2010), and Ploberger (2012).

also determine whether these tests and CS's are asymptotically similar in a uniform sense. The strength of identification of θ depends on the magnitude of the singular values of the expectation of the Jacobian

$$G(W_i, \theta) := \frac{\partial}{\partial \theta'} g(W_i, \theta) \in R^{k \times p} \tag{1.3}$$

of $g(W_i, \theta)$. The parameter space we consider does not impose any restrictions on the magnitude of these singular values. The results hold for arbitrary fixed k and p with $k \geq p$.

We show that Kleibergen's LM test (and CS) has correct asymptotic size and is uniformly asymptotically similar for a parameter space of null distributions that is fairly general. But, the parameter space does require an eigenvalue condition on the asymptotic variance of a transformation of the conditioning statistic (onto which the normalized sample moments are projected). This condition guarantees that the asymptotic version of the $k \times p$ conditioning statistic (after suitable normalization) is full rank p a.s. This condition is shown not to be redundant in Section 12 in the Appendix to this paper. The parameter space also requires that the variance matrix of the moment functions is nonsingular. This assumption is needed because the inverse of the sample variance matrix is employed to make the conditioning statistic asymptotically independent of the sample moments. This condition can be restrictive because in some models lack of identification is accompanied by singularity of the variance matrix of the moments. For example, this occurs in models in which for some null hypothesis a nuisance parameter appears only under the alternative hypothesis.

The nonlinear CLR tests (and CS's) that we consider depend on a rank statistic, which measures the rank of the expectation of $G(W_i, \theta)$. Following Kleibergen (2005), the rank statistics that have been considered in the literature depend on a weighted orthogonalized version of the sample Jacobian, $n^{-1} \sum_{i=1}^{n} G(W_i, \theta)$, where the orthogonalization is designed to create a conditioning statistic that is asymptotically independent of the sample moments. Two weightings have been considered. The first, proposed by Kleibergen (2005, 2007) and Smith (2007), premultiplies the vectorized orthogonalized sample Jacobian by the negative square root of a consistent estimator of its $kp \times kp$ variance matrix. We call this the *Jacobian-variance weighting.* The second, proposed by Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012), multiplies the $k \times p$ orthogonalized sample Jacobian by the negative square root of a consistent estimator of the $k \times k$ variance matrix of the sample moments. We call this the *moment-variance weighting.*

Given the weighting of the orthogonalized sample Jacobian, several functional forms for the rank statistic have been considered in the literature, including the rank statistics of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006). We provide results for a general form of the rank statistic and verify the conditions imposed on the general form for the Robin and Smith (2000) rank statistic. The latter is a popular choice because it is easy to compute. Note that when $p = 1$, these rank statistics all reduce to the squared Euclidean norm of the weighted orthogonalized sample Jacobian vector.

For the case where $p = 1$, we show that the CLR tests (and CS's) based on either weighting have correct asymptotic size and are asymptotically similar in a uniform sense (for parameter spaces that are the same as those considered for the LM test and CS, or slightly smaller, depending on the method of weighting).

For the case where $p \geq 2$, we show that the CLR test (and CS) based on the Robin and Smith (2000) rank statistic with the moment-variance weighting has correct asymptotic size and is uniformly asymptotically similar for the same parameter spaces of distributions as considered for the LM test (and CS). On the other hand, we cannot show that the CLR test (and CS) based on the Robin and Smith (2000) rank statistic with the Jacobian-variance weighting necessarily has correct asymptotic size. The reason is that the weighted orthogonalized sample Jacobian is not necessarily asymptotically independent of the sample moments under some sequences of null distributions. This occurs because the random variation of the $kp \times kp$ sample variance estimator turns out to affect the asymptotic distribution of the weighted orthogonalized sample Jacobian in some cases. Roughly speaking, this occurs when some parameters are weakly identified and some are strongly identified, or when some transformations of the parameters are weakly identified and some transformations are strongly identified. (Obviously, when $p = 1$ these scenarios cannot occur.) This phenomenon has not been demonstrated previously in the literature.

Simulations in a linear IV regression model with two right-hand side endogenous variables corroborate the existence of the asymptotic correlations discussed in the previous paragraph. However, for the particular model and error distributions considered, these correlations have a small effect on the asymptotic null rejection probabilities of the CLR test with Jacobian-variance weighting. These probabilities are very close to the nominal size of the test.

The results of the paper show that weak identification occurs (i.e., the test statistics have nonstandard asymptotic distributions due to identification deficiency) when $\lim n^{1/2} s_{pF_n} < \infty$, where $\{s_{jF} : j = 1, ..., p\}$ are the singular values of the expected Jacobian, $E_F G(W_i, \theta_0)$, ordered to be nonincreasing in j, F denotes a null distribution, $\{F_n : n \geq 1\}$ denotes a sequence of null distributions for which the previous limit exists, and the limit is taken as $n \to \infty$. Strong or semi-strong identification occurs when $\lim n^{1/2} s_{pF_n} = \infty$. Strong identification occurs when $\lim s_{pF_n} > 0$ and semi-strong identification occurs when $\lim n^{1/2} s_{pF_n} = \infty$ and $\lim s_{pF_n} = 0$. When $p = 1$, $s_{1F} = ||E_F G(W_i, \theta_0)||$ and weak identification occurs when $\lim n^{1/2} ||E_F G(W_i, \theta_0)|| < \infty$. However, when $p \geq 2$, weak identification can take many different forms. Weak identification in the standard sense, i.e., when all parameters are weakly identified, e.g., as in Staiger and Stock (1997), occurs when $\lim n^{1/2} s_{1F_n} < \infty$. This is a relatively easy case to analyze asymptotically. Weak identification also occurs when $\lim n^{1/2} s_{pF_n} < \infty$, but $\lim n^{1/2} s_{1F_n} = \infty$, i.e., different singular values behave differently asymptotically. We refer to this as weak identification in a nonstandard sense. It includes the (some weak/some strong) identification scenario considered in Stock and Wright (2000) based on their Assumption C. The nonstandard weak identification scenario is the scenario in which the weighted orthogonalized sample Jacobian may not be independent of the sample moments when the Jacobian-variance weighting is employed. This case is much more difficult to analyze asymptotically. A subset of this case, which we refer to as *joint weak identification,* is a case in which the previous conditions hold (i.e., $\lim n^{1/2} s_{pF_n} < \infty$ and $\lim n^{1/2} s_{1F_n} = \infty$) and $\lim n^{1/2} ||E_{F_n} G_j(W_i, \theta_0)|| = \infty$ for all $j \leq p$, where $G_j(W_i, \theta_0)$ denotes the jth column of $G(W_i, \theta_0)$. Under joint weak identification, each column of the Jacobian behaves as though the corresponding parameter is strongly or semi-strongly identified, but jointly, weak identification occurs (because $\lim n^{1/2} s_{pF_n} < \infty$). As discussed in Section 2 below, no results in the literature consider all of the cases of weak identification that may occur when $p \geq 2$.[2]

For clarity, the results of the paper are stated and derived first for i.i.d. observations. Then, they are extended to cover time series observations that are stationary and strong mixing. This way of proceeding lets us provide somewhat weaker assumptions in the i.i.d. case than if the i.i.d. case is treated as a special case of the time series results.

All limits below are taken as $n \to \infty$. The expression $A := B$ denotes that A is defined to equal B.

The paper is organized as follows. Section 2 discusses the related literature and the contribution of this paper to the literature. Section 3 defines the moment condition model. Section 4 defines and provides asymptotic results for Kleibergen's (2005) LM test. Section 5 does likewise for Kleibergen's (2005) CLR test with Jacobian-variance weighting. Section 6 does likewise for Kleibergen's CLR test with moment-variance weighting, as in Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012). Section 7 provides results for the tests with time series observations.

An Appendix provides some of the proofs of the results given in the paper. The remaining proofs and some additional results are given in the Supplemental Material to this paper, see Andrews and Guggenberger (2014b).

[2] The definitions of the identification categories given here, which are based on $\{s_{jF_n} : j \leq p, n \geq 1\}$, where s_{jF} is the jth largest singular value of $E_F G(W_i, \theta_0)$, are suitable when $\lambda_{\min}(Var_F(g(W_i, \theta_0)))$ is bounded away from zero over the parameter space of distributions F. When the latter condition does not hold, but $\lambda_{\min}(Var_F(g(W_i, \theta_0))) > 0$ for all distributions F, then s_{jF} should be defined to be the jth largest singular value of the normalized expected Jacobian $Var_F^{-1/2}(g(W_i, \theta_0)) E_F G(W_i, \theta_0)$ in order to obtain the appropriate definitions of the identification categories.

2 Discussion of the Literature

To date in the literature it has only been shown that Kleibergen's LM and CLR tests control the limiting null rejection probability under certain strong instrument and certain weak instrument sequences. For example, concerning the validity of the LM and CLR tests, Kleibergen (2005, proofs of Theorems 1 and 3) deals only with sequences of matrices $E_{F_n} G(W_i, \theta)$ whose limits are a full column rank matrix or a matrix of zeros.[3] Kleibergen (2005) does not consider the cases where

(i) the limit of $E_{F_n} G(W_i, \theta)$ exists and is nonzero, some of its columns are equal to zero, and the remaining columns are linearly independent, and

(ii) the limit of $E_{F_n} G(W_i, \theta)$ exists and is nonzero and some subset of its columns are nonzero but less than full column rank, (2.1)

where $\{F_n : n \geq 1\}$ is a sequence of true null distributions that generates the data. Case (ii) is an example of "joint weak identification" in which several parameters individually satisfy conditions that indicate strong identification, but jointly exhibit weak identification. This paper is the first to investigate joint weak identification. Results for cases (i) and (ii) are needed to establish the asymptotic sizes of the LM and CLR tests.

Example. Consider as a simple example the linear IV regression model

$$\begin{aligned} y_{1i} &= Y_{2i}'\theta + u_i, \\ Y_{2i} &= \pi' Z_i + V_{2i}, \end{aligned} \tag{2.2}$$

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 IV's, and π $(= \pi_F) \in R^{k \times p}$ is an unknown unrestricted parameter matrix.[4] The data $\{W_i = (y_{1i}, Y_{2i}', Z_i')' : i = 1, ..., n\}$ are i.i.d. and $E_F((u_i, V_{2i}')'|Z_i) = 0^{p+1}$ a.s. Here $m = 1 + p + k$ and

$$g(W_i, \theta) = Z_i(y_{1i} - Y_{2i}'\theta) \text{ and } G(W_i, \theta) = -Z_i Y_{2i}'. \tag{2.3}$$

[3] See the first equation of the proof of Kleibergen's (2005) Theorem 1 in which the rate of convergence of his $\widehat{D}_T(\theta_0, Y)$ to its limit is stated to be $T^{-\nu}$ for $\nu = 0$ (which is a typo and should be 1/2) or 1 and $J_\theta(\theta_0)$ (which equals $\lim E_{F_n} G_i(\theta_0)$ in our notation) is assumed to exist and have full column rank when $\nu = 1$.

[4] For simplicity, no exogenous variables are included in the structural equation. See Andrews, Cheng, and Guggenberger (2009) and Mikusheva (2010) for asymptotic size results for the CLR test in linear IV regression models with included exogenous variables, but with only one right-hand side endogenous variable. Due to the latter feature, cases (i) and (ii) in (2.1) and case (iv) in (2.5) below do not arise in the aforementioned papers.

By assumption, $E_F g(W_i,\theta) = E_F Z_i u_i = 0^k$ when θ is the true vector. In addition, we have

$$E_F G(W_i,\theta) = -E_F Z_i Z_i' \pi. \tag{2.4}$$

The latter does not depend on θ but does depend on the reduced-form coefficient matrix π which determines the strength of the IV's. Stock and Wright (2000), Guggenberger and Smith (2005), and Guggenberger, Ramalho, and Smith (2012) consider weak/strong IV sequences $\pi_n = (\pi_{1n}, \pi_{2n}) \in R^{k\times(p_1+p_2)}$, where $\pi_{1n} = n^{-1/2}h_1$ for a fixed h_1 and $\pi_{2n} = \pi_2$ is a fixed matrix (that does not depend on n) with full column rank p_2. Specialized to the linear IV setting, the goal of this paper is to establish that the LM and CLR tests of the hypotheses in (1.2) have asymptotic sizes equal to their nominal sizes for a parameter space that does not impose any restrictions on π.

Case (ii) identification failure in (2.1) occurs in model (2.2) with $p \geq 2$ for sequences π_n where a subset of the columns of π_n converge to nonzero vectors that are linearly dependent. For example, this occurs when $p = 2$, $\pi_n \in R^{k\times 2}$, and the columns of π_n are $(1, ..., 1)'$ and $(1+o(1), ..., 1+o(1))'$. Weak identification of this type has not been dealt with in the literature on LM and CLR tests in linear IV models. We do so in this paper (for both linear and nonlinear models).

We return now to the discussion of the general moment condition model. The missing cases in Kleibergen's (2005) proofs of Theorems 1 and 3 are important because they are likely cases in practice. For example, the case where some parameters are strongly identified and others are weakly identified (likely) occurs in Stock and Wright's (2000) (SW) and Kleibergen's (2005) consumption capital asset pricing model (CCAPM) example.

Guggenberger and Smith (2005), Otsu (2006), Inoue and Rossi (2011), Guggenberger, Ramalho, and Smith (2012), and I. Andrews (2014) deal with a subset of case (i) for generalized empirical likelihood (GEL) and GMM versions of the LM and CLR tests, but rule out case (ii) by assumption.[5,6] Furthermore, their results for case (i) rely on Assumption C of SW.[7,8] This assumption is an innovative contribution to the literature, but it has some significant drawbacks as a general high-level condition.

[5] Case (ii) is ruled out by Assumption ID(iii) in Guggenberger and Smith (2005) and Assumption ID$_\theta$(iii) in Guggenberger, Ramalho, and Smith (2012), which assume that the matrix $M_2(\beta)$ has full column rank, where $M_2(\beta)$ contains the columns of $E_F G(W_i,\theta)$ that correspond to the strongly identified parameters. Case (ii) also is ruled out by Assumption C(ii) in Stock and Wright (2000), which is used to obtain results for GMM estimators.

[6] Guggenberger and Smith (2005), Otsu (2006), and Inoue and Rossi (2011) do not consider CLR tests.

[7] Assumption C of SW requires that the expected moment functions can be written as $n^{-1/2}m_{1n}(\theta) + m_2(\beta)$ for some functions m_{1n} and m_2 and some (α, β) such that $\theta = (\alpha', \beta')'$ and $(\partial/\partial\beta')m_2(\beta_0)$ has full column rank, where β_0 denotes the true value of β. In addition, it requires that $m_{1n}(\theta) \to m_1(\theta)$ uniformly over $\theta \in \Theta$ for some real-valued function m_1, $m_2(\beta_0) = 0^k$, $m_2(\beta) \neq 0^k$ for $\beta \neq \beta_0$, and $(\partial/\partial\beta')m_2(\beta)$ is continuous.

[8] Inoue and Rossi (2011) and I. Andrews (2014, Appendix B) use conditions that are much like Assumption C of SW, but they are not exactly the same. As discussed below, Guggenberger, Ramalho, and Smith (2012) impose high-level conditions on a rank statistic when dealing with a CLR test under Assumption C of SW.

First, while Assumption C is easy to verify or refute in linear IV models, it is hard to verify or refute in many, or most, nonlinear models. As far as we are aware, it has only been verified in the literature for one nonlinear model and that nonlinear model is only a local approximation to the model of interest. The model of interest is the two parameter CCAPM considered in SW and Kleibergen (2005). SW verify Assumption C for a local approximation to this model that is a polynomial in the parameters, see p. 1093 of their Appendix B.[9] It appears to be hard to verify or refute Assumption C in the CCAPM of interest.

Another example where Assumption C is hard to verify or refute is the following simple nonlinear regression model with endogeneity, one weakly-identified parameter, and one strongly-identified parameter: $y_i = f(Y_{1i}\theta_1 + Y_{2i}\theta_2) + u_i$, $Y_{1i} = Z_i'\pi_{1n} + V_{1i}$, $Y_{2i} = Z_i'\pi_2 + V_{2i}$, $\pi_{1n} = Cn^{-1/2}$ for some constant vector $C \in R^k$, $\pi_2 \neq 0^k$, $f(\cdot)$ is a known function, Z_i is a vector of IV's, and $\theta = (\theta_1, \theta_2)'$. The moment functions take the form $(y_i - f(Y_{1i}\theta_1 + Y_{2i}\theta_2))Z_i$. For an arbitrary function f it is difficult to determine whether Assumption C holds or not. If f is a quadratic function, or a polynomial, then it may be possible to verify Assumption C. But, even for such functions, doing so does not seem easy.

Second, Assumption C is restrictive. For example, it fails to hold in a nonlinear regression model with weak identification due to the coefficient on a nonlinear regressor being close to zero. Suppose the model is $y_i = \beta h(X_i, \pi) + u_i$ for $i = 1, ..., n$, where y_i and X_i are observed, u_i is an unobserved mean zero error, and $\theta = (\beta, \pi)'$. The parameter π is weakly identified when $\beta = Cn^{-1/2}$ for some constant C. It is shown in Appendix E of the Supplemental Material to Andrews and Cheng (2012) that Assumption C fails in this case.

Another example where Assumption C fails is a linear IV model with joint estimation of the right-hand side (rhs) endogenous variable parameter, which is weakly identified, and the structural equation error variance, which is strongly identified: $y_{1i} = Y_{2i}\theta_1 + u_i$, $Y_{2i} = Z_i\pi_n + V_{2i}$, $Z_i \in R$ (for simplicity), $\pi_n = Cn^{-1/2}$ for some constant C, $Var(u_i) = \theta_2 > 0$, $\theta = (\theta_1, \theta_2)'$, and $Eu_i = EV_{2i} = EZ_iu_i = EZ_iV_{2i} = 0$. The moment functions are $(y_{1i} - Y_{2i}\theta_1)Z_i$ and $(y_{1i} - Y_{2i}\theta_1)^2 - \theta_2$. Assumption C fails in this model.[10]

[9] The approximate model for which SW verify Assumption C is a local approximation to the model of interest based on a Taylor series expansion about a reference parameter value γ_0, in their notation. This approximation is necessarily accurate only for γ close to γ_0. For other values of γ, the approximate model may be different from the model of interest. Note that Assumption C is a global assumption. So, the fact that it holds for the approximate model local to γ_0 does not imply that it approximately holds for the original model.

[10] Assumption C of SW fails in the present example because the expected moment functions are $E(y_{1i} - Y_{2i}\theta_1)Z_i = -n^{-1/2}EZ_i^2C(\theta_1 - \theta_{10})$ and $E(y_{1i} - Y_{2i}\theta_1)^2 - \theta_2 = n^{-1}EZ_i^2C^2(\theta_1 - \theta_{10})^2 + a(\theta)$, where $a(\theta) := \sigma_V^2(\theta_1 - \theta_{10})^2 - 2\sigma_{uV}(\theta_1 - \theta_{10}) + \theta_{20} - \theta_2$, $\theta_0 = (\theta_{10}, \theta_{20})'$ denotes the true value of θ, $\sigma_V^2 := Var(V_{2i})$, $\sigma_{uV} := Cov(u_i, V_{2i})$, and σ_V^2 and σ_{uV} do not depend on n. Because $a(\theta)$ does not depend on n, but does depend on both θ_1 and θ_2, one must take $\beta = \theta$ and $m_2(\beta) = (0, a(\theta))'$ in Assumption C (see the footnote above which specifies Assumption C). In this case, $(\partial/\partial\beta')m_2(\beta_0)$ is a 2×2 matrix with less than full rank, because its first row is zero, which violates Assumption C.

The results of this paper do not impose any conditions on the functional form of the expected moment conditions and their derivatives, like Assumption C does. The conditions given are more general than the conditions used in the papers that rely on Assumption C.

We also point out that no papers in the literature deal with cases where $p \geq 2$ and the limit of $E_{F_n} G(W_i, \theta)$ is zero, but $n^{1/2}||E_{F_n} G_j(W_i, \theta)|| \to \infty$ for some $j \leq p$, where, as above, $G_j(W_i, \theta)$ denotes the jth column of $G(W_i, \theta)$. In such situations, analogues of cases (i) and (ii) arise in which suitably rescaled versions of the columns j for which $n^{1/2}||E_{F_n} G_j(W_i, \theta)|| \to \infty$ have limits that are[11]

(iii) nonzero and linearly independent and

(iv) nonzero and linearly dependent. (2.5)

Case (iv) sequences are examples of joint weak identification. Cases (iii) and (iv) sequences need to be considered to establish the correct asymptotic sizes of the LM and CLR tests.

For CLR tests, Guggenberger, Ramalho, and Smith (2012) establish the correct asymptotic null rejection probabilities for GEL versions of the CLR test in a subset of case (i) under Assumption C and the assumption that the conditioning statistic, $rk_n(\theta)$, either diverges to infinity or converges in distribution to a random variable that is random only through its dependence on the limit of the estimated Jacobian. Verifying this condition in cases (i)-(iv) is not easy. We do so in this paper for the Robin and Smith (2000) rank statistic $rk_n(\theta)$ with moment-variance weighting. In sum, Guggenberger, Ramalho, and Smith (2012) do not establish the correct asymptotic null rejection probabilities of the CLR test under Assumption C. They do so only under an additional high level condition on the rank statistic.

Kleibergen's (2005, Thm. 3) results for the CLR test rely on the claim that the conditioning statistic $rk_n(\theta)$ is asymptotically independent of the LM statistic if $rk_n(\theta)$ is a function of a weighting matrix, $\widetilde{V}_{Dn}$ say, and the orthogonalized sample Jacobian, denoted by $\widehat{D}_n(\theta) \in R^{k \times p}$. However, this claim does not hold in general, as shown in Theorem 5.1 below and Section 18 in the Supplemental Material.[12] Newey and Windmeijer (2009) consider the limiting null rejection

[11] For example, suppose $p = 2$. Let $(G_{i1}, G_{i2}) = G(W_i, \theta) \in R^{k \times 2}$. An example of case (iii) occurs when G_{i1} exhibits what might be called "semi-strong identification," i.e., $E_{F_n} G_{i1} = C_{1n} n^{-s}$ for $0 < s < 1/2$ and $C_{1n} \to C_1 \in R^k$, where $C_1 \neq 0^k$, and G_{i2} exhibits the classic features of "weak identification," i.e., $E_{F_n} G_{i2} = C_2 n^{-1/2}$ for some $C_2 \in R^k$. Then, $E_{F_n} G_{i1} \to 0^k$, $E_{F_n} G_{i2} \to 0^k$, $n^{1/2}||E_{F_n} G_{i1}|| \to \infty$, and $n^s E_{F_n} G_{i1} \to C_1 \neq 0^k$.

An example of case (iv) occurs when $E_{F_n} G_{i1}$ is as above and $E_{F_n} G_{i2} = C_{2n} n^{-s_2}$ for $0 < s_2 < 1/2$ and $C_{2n} \to C_2 \in R^k$, where $C_2 \neq 0^k$, and C_1 and C_2 are linearly dependent. If C_1 and C_2 are linearly independent, then this is another example of case (iii).

[12] Under sequences F_n such that $n^{1/2} E_{F_n} G(W_i, \theta)$ converges to a finite matrix, $n^{1/2} \widehat{D}_n(\theta)$ and $n^{1/2} \widehat{g}_n(\theta)$ ($= n^{-1/2} \sum_{i=1}^n g(W_i, \theta)$) are asymptotically independent (see Lemmas 8.2 and 8.3 in Section 8 in the Appendix). Therefore, if $r(\widehat{V}_n, n^{1/2} \widehat{D}_n(\theta))$ is a continuous function of $n^{1/2} \widehat{D}_n(\theta)$ and a weighting matrix $\widehat{V}_n$ (that converges in probability

probability of the CLR test under "many instrument" asymptotics. They do not analyze the effects of weak identification (such as in cases (i)-(iv)). Their Assumption 2 implies global identification of θ.

As a special case of the asymptotic size results of this paper for nonlinear models, this paper provides some new results for the linear IV regression model. Specifically, the results of the present paper establish the correct asymptotic size of LM and CLR tests in the linear IV model with an arbitrary number of rhs endogenous variables, under some maintained assumptions. The results allow for heteroskedasticity of the errors and stationary strong mixing errors and observations.

In contrast, the relevant results available in the literature for the linear IV model are as follows. Kleibergen (2002) shows that his LM test has correct asymptotic null rejection probabilities under fixed full-rank reduced-form matrices, as well as under standard weak IV asymptotics—that is, under the $n^{-1/2}$-local to zero sequences in Staiger and Stock (1997). Also see Moreira (2009). Moreira (2003) proves that the limiting null rejection probability of the CLR test is correct under standard weak IV asymptotics (i.e., of the type considered in Staiger and Stock (1997)). None of these papers considers cases (i)-(iv) above. Mikusheva (2010) establishes the correct asymptotic size of homoskedastic LM and CLR tests and CS's when there is only one endogenous rhs variable, i.e., $p = 1$, and the errors are homoskedastic. Guggenberger (2012) establishes the correct asymptotic size of heteroskedasticity-robust LM and CLR tests in a heteroskedastic model with $p = 1$. I. Andrews (2014) establishes the correct asymptotic size of a class of conditional linear combination (CLC) tests when $p = 1$, which he shows are equivalent to a class of CLR tests. He provides some CLC tests that are designed to have good power under heteroskedasticity and autocorrelation. Moreira and Moreira (2013) introduce some tests that maximize weighted average power in a linear IV model with heteroskedasticity and autocorrelation for the case where $p = 1$. Note that when $p = 1$, i.e., only one rhs endogenous variable appears (and the exogenous variables are projected out), cases (i), (ii), and (iv) above do not arise (because $E_F G(W_i, \theta)$ has a single column). Phillips

to a positive definite matrix), then by the continuous mapping theorem (CMT), $n^{1/2}\widehat{g}_n(\theta)$ and $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ are also asymptotically independent.

However, under sequences for which a component of $n^{1/2}E_{F_n}G(W_i,\theta)$ diverges to plus or minus infinity, the CMT cannot be applied because $n^{1/2}\widehat{D}_n(\theta)$ does not converge in distribution, but rather, some component of it diverges to plus or minus infinity in probability (see Lemma 8.3 in Section 8 in the Appendix when $h_{1,j} = \infty$ for some $j \leq p$). In this case, $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ may not have an asymptotic distribution, and if it does, $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta))$ and $n^{1/2}\widehat{g}_n(\theta)$ are not necessarily asymptotically independent. The following is a simple example of the latter situation when $p = 2$. Let $r(\widehat{V}_n, n^{1/2}\widehat{D}_n(\theta)) = \widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)||$, where $\widehat{V}_{12n}$ is the (1, 2) component of $\widehat{V}_n$ and $\widehat{D}_{1n}(\theta)$ is the first column of $\widehat{D}_n(\theta)$. Assume $\widehat{V}_n - V \to_p 0$ for some matrix V and $n^{1/2}(\widehat{V}_n - V) \to_d \xi$, where ξ is a mean zero normal random matrix. Assume that under F_n the first column $E_{F_n}G_1(W_i,\theta)$ of $E_{F_n}G(W_i,\theta)$ is a fixed nonzero vector, G_1^e say. Assume that the $(1,2)$ element of V, denoted by V_{12}, equals zero under F_n. Then, $\widehat{D}_{1n}(\theta) \to_p G_1^e$ (see Lemma 8.2 in Section 8 in the Appendix) and $\widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)|| = n^{1/2}(\widehat{V}_{12n} - V_{12})||\widehat{D}_{1n}(\theta)|| \to_d \xi_{12}||G_1^e||$. But, in general there is no reason why ξ_{12} and the random limit of $n^{1/2}\widehat{g}_n(\theta)$ are independent. For simplicity, the previous example is somewhat contrived, because rank statistics typically are not of the form $\widehat{V}_{12n}||n^{1/2}\widehat{D}_{1n}(\theta)||$. But, components of rank statistics may be of this form.

(1989) and Choi and Phillips (1992) provide asymptotic and finite sample results for estimators and classical tests in simultaneous equations models with fixed π matrices that may be unidentified or partially identified when $p \geq 1$. Their results do not cover weak identification (of any type). Hillier (2009) provides exact finite sample results for CLR tests in the linear IV model under the assumption of homoskedastic normal errors and known covariance matrix.

We return now to the discussion of a general moment condition model. In this paper, we show that a minimum eigenvalue condition that appears in the parameter space $\mathcal{F}_0$ (defined below) for the null distributions F is necessary in some sense to obtain correct asymptotic size for the LM and CLR tests. For example, in the linear IV regression model, this eigenvalue condition rules out perfect correlation between the structural and reduced-form errors. Without the eigenvalue condition, we show that in some cases the LM statistic equals the AR statistic plus a $o_p(1)$ term. In consequence, the LM test (which uses a χ^2_p critical value) over-rejects the null hypothesis asymptotically when $k > p$. Furthermore, without it, we show that in other cases the LM statistic equals zero a.s. for all $n \geq 1$ and, hence, the LM test rejects the null hypothesis with probability zero for all $n \geq 1$. In such cases, the LM test under-rejects the null asymptotically. These properties of the LM test have not been recognized in the literature, e.g., see Kleibergen (2005, Theorem 1).

We note that the asymptotic framework and results given here should be useful for establishing the asymptotic size of tests (and CS's) for moment condition and linear IV models that differ from the LM and CLR tests (and CS's) considered here, such as the tests in Moreira and Moreira (2013) and I. Andrews (2014). For example, we provide sufficient conditions for a suitably renormalized version of the moment-variance-weighted orthogonalized sample Jacobian to have full rank almost surely asymptotically, which is needed in the latter paper when $p \geq 2$.

AG2 is a sequel to this paper. It introduces two new nonlinear singularity-robust conditional quasi-LR (SR-CQLR) tests and a singularity-robust Anderson-Rubin (SR-AR) test. AG2 shows that these tests (and the corresponding CS's) have correct asymptotic size for all $p \geq 1$ under very weak conditions. For example, in the i.i.d. case, one of the two SR-CQLR tests and the SR-AR test only require the expected moment functions to equal zero at the true parameter and the sample moment functions to have $2+\gamma$ moments uniformly bounded for some $\gamma > 0$. (The other SR-CQLR test imposes somewhat stronger moment conditions.) In particular, none of the tests in AG2 impose any conditions on the expectation of the Jacobian matrix of the moments or any conditions on the variance matrices of the moment functions or the conditioning statistic, which is the meaning of "singularity-robust." The two SR-CQLR tests are shown to be asymptotically efficient in a GMM sense under strong and semi-strong identification.

The new SR-CQLR tests in AG2 have some advantages over the CLR tests considered in this paper. First, they have correct asymptotic size under noticeably weaker conditions. Because they do not require the variance matrix of the moment functions to be nonsingular, they apply to models in which for some null hypothesis a nuisance parameter appears only under the alternative hypothesis and not under the null hypothesis.[13] In addition, they do not place any restrictions on 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.

Second, the tests reduce, or essentially reduce, asymptotically to Moreira's (2003) CLR test in the homoskedastic linear IV model for all $p \geq 1$. In consequence, (a) no arbitrary choice of rank statistic is needed when $p \geq 2$, and (b) the tests have the desirable power properties of Moreira's (2003) CLR test in the homoskedastic normal linear IV model when $p = 1$, which have been established in Andrews, Moreira, and Stock (2006, 2008) and Chernozhukov, Hansen, and Jansson (2009).[14] In contrast, the CLR tests considered here for $p \geq 1$ are all based on the form of Moreira's LR statistic when $p = 1$ and, in consequence, require the specification of some rank statistic. The CLR tests considered here based on the Jacobian-variance weighting reduce to Moreira's CLR test when $p = 1$, but we cannot show that they necessarily have correct asymptotic size when $p \geq 2$. On the other hand, we show that the CLR tests considered here that are based on the moment-variance weighting have correct asymptotic size when $p \geq 1$, but they do not reduce to Moreira's CLR test when $p = 1$ (or $p \geq 2$).

We also mention the recent paper by I. Andrews and Mikusheva (2014a) that introduces a new conditional likelihood ratio test for moment condition models that is robust to weak identification. This 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.

The LM and CLR tests considered in this paper are for full vector inference. To obtain subvector inference, one can employ the Bonferroni method or the Scheffé projection method, see Cavanagh, Elliott, and Stock (1995), Chaudhuri, Richardson, Robins, and Zivot (2010), Chaudhuri and Zivot (2011), and McCloskey (2011) for Bonferroni's method, and Dufour (1989) and Dufour and Jasiak (2001) for the projection method. These methods are conservative, but Bonferroni's method is found to work well by Chaudhuri, Richardson, Robins, and Zivot (2010) and Chaudhuri and Zivot (2011).[15]

[13] Nonsingularity of the variance matrix of the moments is needed for Kleibergen's CLR-type tests, because the inverse of this matrix is used to orthogonalize the sample Jacobian from the sample moments when constructing a conditioning statistic.

[14] For related results, see Chamberlain (2007) and Mikusheva (2010).

[15] A refinement of Bonferroni's method that is not conservative, but is much more intensive computationally, is provided by Cavanagh, Elliott, and Stock (1995). McCloskey (2011) also considers a refinement of Bonferroni's method.

Other methods for 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: (a) one has knowledge of the source of the potential lack of identification (i.e., which subvectors play the roles of β, π, and ζ in their notation), (b) there is only one source of lack of identification, and (c) the estimator objective function does not depend on the weakly identified parameters π (in their notation) when $\beta = 0$, which rules out some weak IV's models. Montiel Olea (2012) provides some subvector analysis in the extremum estimator context of Andrews and Cheng (2012). His efficient conditionally similar tests apply to the subvector (π, ζ) of (β, π, ζ) (in the notation of Andrews and Cheng (2012)), where the parameter β determines the strength of identification and is known to be strongly identified. This subvector analysis is analogous to that of Stock and Wright (2000) and Kleibergen (2004). Cheng (2014) provides subvector inference in a nonlinear regression model with multiple nonlinear regressors and, in consequence, multiple potential sources of lack of identification. I. Andrews and Mikusheva (2012) provide subvector inference methods in a minimum distance context based on Anderson-Rubin-type statistics. I. Andrews and Mikusheva (2014b) provide conditions under which subvector inference is possible in exponential family models (but the requisite conditions seem to be restrictive).

3 Moment Condition Model

3.1 Definition of the Parameter Space for the Distributions F

First we introduce some notation. 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 jth column of $G_i(\theta)$ by $G_{ij}(\theta)$ and $G_{ij} = G_{ij}(\theta_0)$, where θ_0 denotes the (true) null value of θ, for $j = 1, ..., p$. Likewise, we often leave out the argument θ_0 for other functions as well. For example, 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. For a positive semi-definite (psd) matrix A, we let $\lambda_j(A)$ denote the jth largest eigenvalue of A.

For some $\gamma, \delta > 0$ and $M < \infty$, define

$$\mathcal{F} := \{F : E_F g_i = 0^k,\ E_F||(g_i', vec(G_i)')'||^{2+\gamma} \leq M,\ \text{and}\ \lambda_{\min}(E_F g_i g_i') \geq \delta\}, \tag{3.1}$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix, $|| \cdot ||$ denotes the Euclidean norm, and

$vec(\cdot)$ denotes the vector obtained by stacking the columns of a matrix. The first condition in $\mathcal{F}$ is the defining condition of the model. The second condition in $\mathcal{F}$ is a mild moment condition on the moment functions g_i and their derivatives G_i. The last condition in $\mathcal{F}$ rules out singularity and near singularity of the variance matrix of the moments.[16] For example, in the linear IV model it rules out $E_F u_i^2 Z_i Z_i'$ being singular, which usually is not restrictive. Identification issues arise when $E_F G_i$ has, or is close to having, less than full column rank (which occurs when one or more of its singular values is zero or close to zero). The conditions in $\mathcal{F}$ place no restrictions on the singular values or column rank of $E_F G_i$.

The condition $\lambda_{\min}(E_F g_i g_i') \geq \delta$ in $\mathcal{F}$ can be replaced by $\lambda_{\min}(E_F g_i g_i') > 0$ without affecting the asymptotic size and similarity results given in Theorems 4.1 and 6.1 below, provided g_i and G_i are replaced with g_i^* and G_i^*, respectively, in $\mathcal{F}$ and $\mathcal{F}_0$ (defined below), where $g_i^* := (E_F g_i g_i')^{-1/2} g_i$ and $G_i^* := (E_F g_i g_i')^{-1/2} G_i$.[17,18] This allows for the variance matrix of g_i to be arbitrarily close to singular, which occurs in some cases when identification is weak, but rules out singularity.

The parameter spaces for the distribution F that we consider in this paper are subsets of $\mathcal{F}$. The main parameter space that we consider is $\mathcal{F}_0$, which we now define.

For an arbitrary square-integrable (under F) vector a_i, let

$$\Sigma_F^{a_i} := E_F a_i a_i', \ \Gamma_F^{a_i} := E_F a_i g_i', \ \Omega_F := \Sigma_F^{g_i} = E_F g_i g_i', \text{ and } \Psi_F^{a_i} := \Sigma_F^{a_i} - \Gamma_F^{a_i} \Omega_F^{-1} \Gamma_F^{a_i \prime}. \tag{3.2}$$

The matrix $\Psi_F^{a_i}$ is the expected outer product of the vector of residuals from the $L^2(F)$ projections of the components of a_i onto the space spanned by the components of g_i.

Let

$$(\tau_{1F}, ..., \tau_{pF}) \text{ denote the } p \text{ singular values of } \Omega_F^{-1/2} E_F G_i, \tag{3.3}$$

ordered so that τ_{jF} is nonincreasing in j. These singular values are nonnegative and may be zero. Let

$$B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } (E_F G_i)' \Omega_F^{-1} (E_F G_i) \tag{3.4}$$

[16]Note that it is not possible to avoid the assumption $\lambda_{\min}(E_F g_i g_i') \geq \delta$ by replacing an estimator $\widehat{\Omega}_n$ of $E_F g_i g_i'$ by an eigenvalue-adjusted version, e.g., as defined in AG2. The reason is that the eigenvalue adjustment leads to a nonzero asymptotic covariance between the sample moments $\widehat{g}_n$ and the conditioning matrix $\widehat{D}_n$, defined in (4.1) and (4.3) below, which yields a test that does not necessarily have correct asymptotic size. See Comment (ii) to Lemma 8.2 in the Appendix for more details.

[17]This holds because $\lambda_{\min}(E_F g_i^* g_i^{*\prime}) = \lambda_{\min}(I_k) = 1$ and the proofs of the results given below go through with g_i^* and G_i^* in place of g_i and G_i throughout.

[18]The matrix $(E_F g_i g_i')^{-1/2}$ that appears in the definition of g_i^* and G_i^* can be replaced by any nonsingular $k \times k$ matrix, say $K_F(\theta_0)$, that yields $\lambda_{\min}(E_F g_i^* g_i^{*\prime}) \geq \delta > 0$. For example, in somewhat related contexts, Andrews and Cheng (2013b) and I. Andrews and Mikusheva (2014) find it convenient to rescale moment conditions by diagonal matrices.

ordered so that the corresponding eigenvalues $(\kappa_{1F}, ..., \kappa_{pF})$ are nonincreasing. Let

$$C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } \Omega_F^{-1/2}(E_F G_i)(E_F G_i)'\Omega_F^{-1/2} \quad (3.5)$$

ordered so that the corresponding eigenvalues are $(\kappa_{1F}, ..., \kappa_{pF}, 0, ..., 0)' \in R^k$. Note that $\kappa_{jF} = \tau_{jF}^2$ for $j \le p$. With some abuse of notation, for an integer $0 \le j \le p$, let $B_F = (B_{F,j}, B_{F,p-j})$ denote the decomposition of B_F into its first j and last $p-j$ columns, where by definition, when $j = p$, $B_{F,j} = B_F$ and $B_{F,p-j}$ denotes a matrix with no columns and, when $j = 0$, $B_{F,j}$ denotes a matrix with no columns and $B_{F,p-j} = B_F$. Analogously, for an integer $0 \le j \le k$, let $C_F = (C_{F,j}, C_{F,k-j})$ denote the decomposition of C_F into its first j and last $k-j$ columns, where, when $j = 0$ or $j = k$, $C_{F,j}$ and $C_{F,k-j}$ are defined analogously to $B_{F,j}$ and $B_{F,p-j}$.

For $0 \le j \le p-1$ and $\xi \in R^{p-j}$, define

$$\Psi_{jF}(\xi) := \Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_i B_{F,p-j}\xi}. \quad (3.6)$$

For a given $\delta_1 > 0$, we define the parameter space of null distributions to be

$$\begin{aligned} \mathcal{F}_0 &:= \textstyle\bigcup_{j=0}^{p} \mathcal{F}_{0j}, \text{ where} \quad (3.7) \\ \mathcal{F}_{0j} &:= \{F \in \mathcal{F} : \tau_{jF} \ge \delta_1 \text{ and } \lambda_{p-j}(\Psi_{jF}(\xi)) \ge \delta_1 \ \forall \xi \in R^{p-j} \text{ with } ||\xi|| = 1\}, \end{aligned}$$

$\tau_{0F} := \delta_1$, and $\lambda_{p-j}(\Psi_{jF}(\xi)) := \delta_1$ for $j = p$.[19,20] We assume that $\mathcal{F}_0 \neq \varnothing$.

The conditions in $\mathcal{F}_0$ are used to show that the estimator $\widehat{\Omega}_n^{-1/2}\widehat{D}_n \in R^{k \times p}$, defined below, of the normalized population Jacobian matrix $\Omega_F^{-1/2}E_F G_i$ has full column rank p asymptotically with probability one after suitable normalization (see Lemma 8.3(d) in the Appendix). This almost sure (a.s.) full column rank p property is needed to obtain the desired asymptotic χ_p^2 null distribution of the LM statistic (introduced below), which is used by the LM and CLR tests. The LM statistic is a quadratic form in the sample moments with weight matrix given by the projection matrix onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$.

We obtain the a.s. full column rank property using conditions on both the (asymptotic) mean and variance of $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$. The index j on $\mathcal{F}_{0j}$ denotes the contribution coming from the mean and $p-j$ denotes the contribution coming from the variance. For $j = 0$ (i.e., when the parameters

[19] The matrices B_F and C_F are not necessarily uniquely defined. But, this is not of consequence because the $\lambda_{p-j}(\cdot)$ condition is invariant to the choice of B_F and C_F.

[20] Note that Kleibergen (2005) does not impose any rank restrictions on the variance matrix of the limiting distribution of $n^{-1/2}\sum(g_i', vec(G_i)' - Evec(G_i)')'$. As simple examples show, however, to derive the limiting distribution of the LM test statistic, one needs to impose some restrictions of the type in $\mathcal{F}_0$. For example, the case $g_i(\theta) = 0$ with probability one for all θ vectors is compatible with Kleibergen's (2005) assumptions but violates the nonsingularity claim in the statement of Theorem 1 in Kleibergen (2005).

are weakly identified in the standard sense), the $\tau_{jF} \geq \delta_1$ condition disappears, no restrictions are placed on the mean $\Omega_F^{-1/2} E_F G_i$, and the a.s. full column rank property is obtained using the $\lambda_{p-j}(\cdot)$ condition with $j = 0$. For $j = p$ (i.e., when all parameters are strongly identified), the $\lambda_{p-j}(\cdot)$ condition disappears (because $B_{F,p-j}$ is a matrix with no columns when $j = p$) and the a.s. full rank property is obtained using only the mean condition $\tau_{pF} \geq \delta_1$. For $0 < j < p$ (i.e., when the parameters are weakly identified in the nonstandard sense), the a.s. full rank property is obtained partly via the mean condition $\tau_{jF} \geq \delta_1$ and partly via the $\lambda_{p-j}(\cdot)$ condition.[21,22]

The "variance" (or variability) condition, $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1$, can be interpreted as follows. The $(k-j)\times(p-j)$ matrix $C'_{F,k-j}\Omega_F^{-1/2} G_i B_{F,p-j}$ is a submatrix of the $k \times p$ matrix $C'_F \Omega_F^{-1/2} G_i B_F$, which is just $\Omega_F^{-1/2} G_i$ with its rows and columns rotated. This submatrix $C'_{F,k-j}\Omega_F^{-1/2} G_i B_{F,p-j}$ has the j linear combinations of the rows and columns of $\Omega_F^{-1/2} G_i$ removed for which the mean component of $\widehat{\Omega}_n^{-1/2} \widehat{D}_n$, i.e., $\Omega_F^{-1/2} E_F G_i$, provides a column rank of magnitude j. (More specifically, the mean component of the j linear combinations of the rows and columns of $\Omega_F^{-1/2} G_i$ that are removed equals $C'_{F,j}\Omega_F^{-1/2} E_F G_i B_{F,j} = Diag\{\tau_{1F}, ..., \tau_{jF}\} \in R^{j\times j}$ and the column rank of $Diag\{\tau_{1F}, ..., \tau_{jF}\}$ is j by the definition of $\mathcal{F}_{0j}$.[23]) The $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1$ condition requires that every linear combination ξ (with $||\xi|| = 1$) of the columns of the aforementioned submatrix, i.e., $C'_{F,k-j}\Omega_F^{-1/2} G_i B_{F,p-j}\xi$, has enough variability to provide the requisite additional column rank of magnitude $p - j$. Specifically, the $(p-j)$-th largest eigenvalue of $\Psi_{jF}(\xi)$ $(:= \Psi_F^{C'_{F,k-j}\Omega_F^{-1/2} G_i B_{F,p-j}\xi})$ is bounded away from zero. This allows for the minimal amount of variation that still delivers the incremental $p - j$ column rank that is required. Note that the matrix $\Psi_{jF}(\xi)$ is not actually a variance matrix. It is an expected outer-product matrix, which makes the condition slightly weaker.

We can write

$$\begin{aligned} \Psi_{jF}(\xi) &= (\xi' B'_{F,p-j} \otimes C'_{F,k-j}\Omega_F^{-1/2}) \Psi_F^{vec(G_i)} (B_{F,p-j}\xi \otimes \Omega_F^{-1/2} C_{F,k-j}) \text{ and} \\ \Psi_F^{vec(G_i)} &= E_F G_{Fi} G'_{Fi}, \text{ where } G_{Fi} := vec(G_i) - \Gamma_F^{vec(G_i)} \Omega_F^{-1} g_i \in R^{pk} \end{aligned} \quad (3.8)$$

(using the general formula $vec(ABC) = (C' \otimes A)vec(B)$). The random vector G_{Fi} consists of the residuals from the $L^2(F)$ projections of the components of G_i onto the space spanned by the components of g_i. The matrix $\Psi_F^{vec(G_i)}$ is the expected outer-product of these residuals. Analogously,

[21] Sequences of distributions in the semi-strongly identified category can come from sets $\mathcal{F}_{0j}$ for any $j < p$.

[22] Linking the parameter spaces $\mathcal{F}_{0j}$ for $j = 0, ..., p$ with identification categories, as is done in this paragraph, provides a useful interpretation, but is somewhat heuristic. The reason is that the parameter spaces $\mathcal{F}_{0j}$ place conditions on individual distributions F, whereas the asymptotic identification categories (i.e., strong, semi-strong, and weak in the standard and nonstandard senses) depend on the properties of *sequences* of distributions $\{F_n : n \geq 1\}$.

[23] The stated equality holds because (i) by (3.3)-(3.5) $\Omega_F^{-1/2} E_F G_i = C_F Diag(\tau_F) B'_F$, where $Diag(\tau_F)$ is the $k \times p$ matrix whose (m, m) element equals τ_{mF} for $m = 1, ..., p$ and whose other elements all equal zero, (ii) $C'_F \Omega_F^{-1/2} E_F G_i B_F = Diag(\tau_F)$ by the orthogonality of C_F and B_F, and, hence, (iii) $C'_{F,j}\Omega_F^{-1/2} E_F G_i B_{F,j} = Diag\{\tau_{1F}, ..., \tau_{jF}\}$.

the matrix $\Psi_{jF}(\xi)$ is the expected outer-product of the residuals from the $L^2(F)$ projections of the elements of $C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi$ onto the space spanned by the components of g_i.

If some element of g_i does not depend on some element of θ, then the corresponding element of G_i is identically zero. For example, this occurs with simple mean-variance moment conditions of the form $g_i(\theta) = (Y_i - \theta_1, (Y_i - \theta_1)^2 - \theta_2)'$, where θ_1 is a mean parameter and θ_2 is a variance parameter of the random variable Y_i. In such cases, $\Psi_F^{vec(G_i)}$ is singular. In consequence, it is important to impose the weakest conditions possible on $\Psi_F^{vec(G_i)}$ or $\Psi_F^{vec(\Omega_F^{-1/2}G_i)}$.

In the simple mean-variance model, $k = p = 2$, $E_FG_i = -I_2$, both parameters are strongly identified, and $\mathcal{F}_0$ contains $\mathcal{F}_{0p} = \{F \in \mathcal{F} : \tau_{pF} \geq \delta_1\}$, where τ_{pF} is the smallest singular value of $\Omega_F^{-1/2}$ (because $E_FG_i = -I_2$). In this model, τ_{pF} is bounded away from zero if the fourth moment of Y_i is bounded above, which is implied by the condition in $\mathcal{F}$ that $E_F||g_i||^{2+\gamma} \leq M$.[24] Hence, the condition $\tau_{pF} \geq \delta_1$ is redundant for δ_1 sufficiently small in this model.

If the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_{0j}$ is weakened to $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$ and the variance and covariance matrix estimators $\widehat{\Omega}_n$ and $\widehat{\Gamma}_n$ defined below can be any consistent estimators (under suitable sequences of distributions), then the LM and CLR tests do not necessarily have correct asymptotic size. In particular, we provide an example where the asymptotic distribution of the LM statistic is χ_k^2 in this case, rather than the desired distribution χ_p^2, which leads to over-rejection under the null when $k > p$, see Section 12 in the Appendix.[25] Hence, the restrictions on the parameter space $\mathcal{F}_0$ are not redundant.

In contrast, the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests introduced in AG2 are shown to have correct asymptotic size without any conditions on $\lambda_{p-j}(\Psi_{jF}(\xi))$ or $\lambda_{\min}(E_Fg_ig_i')$. All that is required is the first two conditions in $\mathcal{F}$. Hence, these tests have advantages over the LM and CLR tests considered here in terms of the robustness of their size properties.

Let $\overline{C}_{F,p-j} \in R^{k\times(p-j)}$ denote a matrix that contains $p - j$ columns from the last $k - j$ columns of C_F. Six alternative sufficient conditions for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$, in increasing order of

[24] This holds because $E_FG_i = -I_2$ and Ω_F has elements $[\Omega_F]_{11} = \theta_{20}$, $[\Omega_F]_{12} = [\Omega_F]_{21} = E_FU_i(U_i^2 - \theta_{20})$, and $[\Omega_F]_{22} = E_F(U_i^2 - \theta_{20})^2$, where $\theta_{20} := Var_F(Y_i)$, $U_i := Y_i - \theta_{10}$, $\theta_{10} := E_FY_i$, and $\theta_0 = (\theta_{10}, \theta_{20})'$ denotes the true null value.

[25] This example consists of a standard linear IV regression model with one rhs endogenous variable, IV's that are irrelevant, i.e., $\pi = 0^k$, and a correlation between the structural and reduced-form equation errors that equals one or converges to one as $n \to \infty$. The example also can be extended to cover weak IV cases (where $\pi = \pi_n \neq 0^k$, but $\pi_n \to 0^k$ sufficiently quickly as $n \to \infty$), rather than the irrelevant IV case.

strength, are:

$$
\begin{aligned}
&\text{(i) } \lambda_{\min}\left(\Psi_F^{vec(\overline{C}'_{F,p-j}\Omega_F^{-1/2}G_iB_{F,p-j})}\right) \geq \delta_1 \text{ for some matrix } \overline{C}_{F,p-j},\\
&\text{(ii) } \lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_iB_{F,p-j})}\right) \geq \delta_1,\\
&\text{(iii) } \lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_i)}\right) \geq \delta_1,\\
&\text{(iv) } \lambda_{\min}\left(\Psi_F^{vec(G_i)}\right) \geq \delta_2 := \delta_1 M^{2/(2+\gamma)},\\
&\text{(v) } \lambda_{\min}(\Sigma_F^{f_i}) \geq \delta_2, \text{ where } \Sigma_F^{f_i} := E_F f_i f_i' \text{ and } f_i := \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}, \text{ and}\\
&\text{(vi) } \lambda_{\min}(Var_F(f_i)) \geq \delta_2, \qquad (3.9)
\end{aligned}
$$

where M and γ are as in (3.1) and δ_1 is as in (3.7).[26] See Section 17 in the Supplemental Material for a proof of the sufficiency of these conditions. None of these conditions depend on ξ. Another sufficient condition for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$ is

$$\lambda_p\left(\Psi_F^{\Omega_F^{-1/2}G_iB_{F,p-j}\xi}\right) \geq \delta_1 \ \forall \xi \in R^{p-j} \text{ with } ||\xi|| = 1. \qquad (3.10)$$

For the linear IV model in (2.2), we have $\Omega_F = E_F u_i^2 Z_i Z_i'$, $\Sigma_F^{vec(G_i)} = E_F vec(Z_i Y_{2i}') vec(Z_i Y_{2i}')'$, $\Gamma_F^{vec(G_i)} = -E_F vec(Z_i Y_{2i}') Z_i' u_i$, and $E_F||(g_i', vec(G_i)')'||^{2+\gamma} = E_F||(u_i Z_i', vec(Z_i Y_{2i}')')'||^{2+\gamma}$. Sufficient conditions for condition (vi) in (3.9) (and, hence, for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$) in the linear IV regression model are as follows. We have

$$
\begin{aligned}
\Sigma_F^{f_i} &= E_F((u_i, -Y_{2i}')' \otimes Z_i)((u_i, -Y_{2i}')' \otimes Z_i)'\\
&= E_F(\varepsilon_i \otimes Z_i)(\varepsilon_i \otimes Z_i)' + E_F s_i(\pi) s_i(\pi)' \text{ and}\\
Var_F(f_i) &= E_F(\varepsilon_i \otimes Z_i)(\varepsilon_i \otimes Z_i)' + E_F s_i(\pi) s_i(\pi)' - E_F s_i(\pi) E_F s_i(\pi)'\\
&\geq E_F(\varepsilon_i \varepsilon_i' \otimes Z_i Z_i'), \text{ where}\\
\varepsilon_i &:= (u_i, -V_{2i}')', \ s_i(\pi) := (0^{k\prime}, -(Z_i Z_i' \pi_1)', ..., -(Z_i Z_i' \pi_p)')', \qquad (3.11)
\end{aligned}
$$

$\pi = (\pi_1, ..., \pi_p)$ for $\pi_j \in R^k$ for $j = 1, ..., p$, and the inequality holds in a psd sense. Hence, $\lambda_{\min}(Var_F(f_i)) \geq \delta_2$ holds if $\lambda_{\min}(E_F(\varepsilon_i \varepsilon_i' \otimes Z_i Z_i')) \geq \delta_2$. When ε_i is conditionally homoskedastic, i.e., $\Sigma_{\varepsilon,F} := Var_F(\varepsilon_i) = E_F(\varepsilon_i \varepsilon_i' | Z_i)$ a.s., we have $E_F(\varepsilon_i \varepsilon_i' \otimes Z_i Z_i') = \Sigma_{\varepsilon,F} \otimes E_F Z_i Z_i'$. Hence, for example, $\lambda_{\min}(Var_F(f_i)) \geq \delta_2$ holds if $\Sigma_{\varepsilon,F}$ and $E_F Z_i Z_i'$ have minimum eigenvalues that are

[26] Condition (i) holds if it holds for any $\overline{C}_{F,p-j}$ matrix corresponding to any C_F matrix that satisfies the condition in $\mathcal{F}_{0j}$. Conditions (i) and (ii) are invariant to the choice of the matrix B_F in cases where B_F is not uniquely defined.

bounded away from zero by $\delta_2^{1/2}$.

3.2 Definition of $\mathbf{G(W_i, \boldsymbol{\theta})}$

The $k \times p$ matrix $G(W_i, \theta)$ does not need to equal $(\partial/\partial\theta')g(W_i, \theta)$, as defined in (1.3). Rather, the asymptotic size results given below hold for any matrix $G(W_i, \theta)$ that satisfies the conditions in $\mathcal{F}_0$. 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. Alternatively, the function $G(W_i, \theta)$ can be a numerical derivative, such as $((g(W_i, \theta + \varepsilon e_1) - g(W_i, \theta))/\varepsilon, ..., (g(W_i, \theta + \varepsilon e_p) - g(W_i, \theta))/\varepsilon) \in R^{k \times p}$ for some $\varepsilon > 0$, where e_j is the jth unit vector, e.g., $e_1 = (1, 0, ..., 0)' \in R^p$. This choice of $G(W_i, \theta)$ matrix may be useful for models with quite complicated Jacobian matrices $(\partial/\partial\theta')g(W_i, \theta)$.

3.3 Definitions of Asymptotic Size and Asymptotic Similarity

Now, we define asymptotic size and asymptotic similarity of a test of $H_0 : \theta = \theta_0$ for some given parameter space $\overline{\mathcal{F}}(\theta_0)$ of null distributions F. Let $RP_n(\theta_0, F, \alpha)$ denote the null rejection probability of a nominal size α test with sample size n when the distribution of the data is F. The *asymptotic size* of the test for the parameter space $\overline{\mathcal{F}}(\theta_0)$ is defined by

$$AsySz := \limsup_{n\to\infty} \sup_{F \in \overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{3.12}$$

The test is *asymptotically similar* (in a uniform sense) for the parameter space $\overline{\mathcal{F}}(\theta_0)$ if

$$\liminf_{n\to\infty} \inf_{F \in \overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha) = \limsup_{n\to\infty} \sup_{F \in \overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{3.13}$$

Next, we consider a CS that is obtained by inverting tests of $H_0 : \theta = \theta_0$ for all $\theta_0 \in \Theta$. The *asymptotic size* of the CS for the parameter space $\overline{\mathcal{F}}_\Theta := \{(F, \theta_0) : F \in \overline{\mathcal{F}}(\theta_0), \theta_0 \in \Theta\}$ is $AsySz := \liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - RP_n(\theta_0, F, \alpha))$. The CS is *asymptotically similar* (in a uniform sense) for the parameter space $\overline{\mathcal{F}}_\Theta$ if $\liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - RP_n(\theta_0, F, \alpha)) = \limsup_{n\to\infty} \sup_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - RP_n(\theta_0, F, \alpha))$. As defined, 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 θ_0. This additional level of uniformity does not play a significant role in this paper. The same proofs for tests give results for CS's with only minor changes.

The dependence of the parameter space $\mathcal{F}_0$, defined in (3.7), on θ_0 is suppressed for notational simplicity. When dealing with CS's, rather than tests, we make the dependence explicit and write it as $\mathcal{F}_0(\theta_0)$. The asymptotic size and similarity of CS's is considered for the parameter space $\mathcal{F}_{\Theta,0}$

defined by

$$\mathcal{F}_{\Theta,0} := \{(F, \theta_0) : F \in \mathcal{F}_0(\theta_0), \theta_0 \in \Theta\}. \tag{3.14}$$

4 Kleibergen's Nonlinear LM Test

Here, we define and analyze Kleibergen's (2005) nonlinear LM test for the nonlinear moment condition model in (1.1). Let

$$\widehat{g}_n(\theta) := n^{-1}\sum_{i=1}^{n} g_i(\theta),\ \widehat{G}_n(\theta) := n^{-1}\sum_{i=1}^{n} G_i(\theta),\ \text{and } \widehat{\Omega}_n(\theta) := n^{-1}\sum_{i=1}^{n} g_i(\theta)g_i(\theta)' - \widehat{g}_n(\theta)\widehat{g}_n(\theta)'.^{27} \tag{4.1}$$

For any matrix A with r rows, we define the projection matrices

$$P_A := A(A'A)^{-}A' \text{ and } M_A := I_r - P_A, \tag{4.2}$$

where $(\cdot)^{-}$ denotes any g-inverse.[28] If A has zero columns, we set $M_A = I_r$.

Define the (nonlinear) Anderson and Rubin (1949) (AR) statistic of Stock and Wright (2000), and the Lagrange Multiplier statistic of Kleibergen (2005) as follows:

$$\begin{aligned}
AR_n(\theta) &:= n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \text{ and}\\
LM_n(\theta) &:= n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1/2}(\theta)P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta),\ \text{where}\\
\widehat{D}_n(\theta) &:= (\widehat{D}_{1n}(\theta), ..., \widehat{D}_{pn}(\theta)) \in R^{k\times p},\\
\widehat{D}_{jn}(\theta) &:= \widehat{G}_{jn}(\theta) - \widehat{\Gamma}_{jn}(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \in R^k \text{ for } j = 1, ..., p,\\
\widehat{G}_n(\theta) &:= (\widehat{G}_{1n}(\theta), ..., \widehat{G}_{pn}(\theta)) \in R^{k\times p},\ \text{and}\\
\widehat{\Gamma}_{jn}(\theta) &:= n^{-1}\sum_{i=1}^{n}(G_{ij}(\theta) - \widehat{G}_{jn}(\theta))g_i(\theta)' \in R^{k\times k} \text{ for } j = 1, ..., p.
\end{aligned} \tag{4.3}$$

We refer to $\widehat{D}_n(\theta)$ as the *orthogonalized sample Jacobian* because it equals the sample Jacobian $\widehat{G}_n(\theta)$ adjusted to be asymptotically independent of the sample moments $\widehat{g}_n(\theta)$.

The nominal size α LM test rejects the null hypothesis in (1.2) when $LM_n(\theta_0)$ exceeds the $1-\alpha$ quantile of a χ^2_p distribution, denoted by $\chi^2_{p,1-\alpha}$. The nominal size $1-\alpha$ LM CS is defined by

$$CS_{LM,n} := \{\theta_0 \in \Theta : LM_n(\theta_0) \le \chi^2_{p,1-\alpha}\}. \tag{4.4}$$

[27] Any estimator $\widehat{\Omega}_n(\theta)$ that is consistent for $Eg_i(\theta)g_i(\theta)'$ under the drifting subsequences of distributions considered in Section 8 in the Appendix can be used, such as $n^{-1}\sum_{i=1}^{n} g_i(\theta)g_i(\theta)'$, without changing the asymptotic size results given below. However, we recommend the definition in (4.1).

[28] Projection matrices are invariant to the choice of g-inverse.

The following result establishes the correct asymptotic size and asymptotic similarity of Kleibergen's (2005) LM test and CS for the parameter spaces $\mathcal{F}_0$ and $\mathcal{F}_{\Theta,0}$, respectively.

Theorem 4.1 *The asymptotic size of the* LM *test equals its nominal size* $\alpha \in (0,1)$ *for the parameter space* $\mathcal{F}_0$ (*defined in* (3.7)). *Furthermore, the* LM *test is asymptotically similar* (*in a uniform sense*). *Analogous results hold for the* LM CS *for the parameter space* $\mathcal{F}_{\Theta,0}$, *defined in* (3.14).

Comments: (i) Theorem 4.1 provides a more complete set of asymptotic results under the null hypothesis for the LM statistic than in Kleibergen (2005). See Section 2 for a detailed discussion.

(ii) In contrast to results in Kleibergen (2005), we impose regularity conditions in the specification of $\mathcal{F}_0$ in order to establish our asymptotic results for the LM test. We show in Section 12 in the Appendix that these regularity conditions are not redundant. Without the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$, we show that, for some models, some sequences of distributions, and some (consistent) choices of variance and covariance estimators, the LM statistic has a χ^2_k asymptotic distribution. This leads to over-rejection of the null when the standard χ^2_p critical value is used and the parameters are over-identified (i.e., $k > p$).

(iii) Kleibergen's LM test is asymptotically efficient in a GMM sense under strong IV's because it is asymptotically equivalent under $n^{-1/2}$ local alternatives to t and/or Wald tests based on asymptotically efficient GMM estimators, e.g., see Newey and West (1987b).

We now provide a brief description of how we obtain the asymptotic distribution of the projection matrix onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$, which appears in the LM statistic, using the conditions in $\mathcal{F}_0$. Projection matrices are invariant to multiplication by scalars, such as $n^{1/2}$, and post-multiplication by nonsingular $p \times p$ matrices. We use this invariance when normalizing $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ to obtain a nondegenerate limit of the projection matrix under a sequence of distributions $\{F_n \in \mathcal{F}_0 : n \geq 1\}$. The appropriate normalization depends on the identification strength under $\{F_n : n \geq 1\}$. For sequences of distributions where all parameters are strongly identified, such as distributions in $\mathcal{F}_{0p}$, no normalization is needed and $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ converges in probability to a nonstochastic matrix that has full column rank p.

For sequences of distributions that are weakly identified in the standard sense (i.e., for which all parameters are weakly identified), such as suitable sequences of distributions in $\mathcal{F}_{00}$, the expected Jacobian $E_{F_n}G_i$ is $O(n^{-1/2})$, we normalize $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ by $n^{1/2}$, the vector $vec(n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n)$ has an asymptotic normal distribution with possibly nonzero mean, and we obtain the desired a.s. full column rank property of the asymptotic version of $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ using the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{00}$ for $j = 0$.

Sequences of distributions $\{F_n : n \geq 1\}$ that are weakly identified in the nonstandard sense are noticeably more complicated to analyze. For such sequences, we multiply $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ by $n^{1/2}$ and post-multiply $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ by a nonstochastic nonsingular $p \times p$ matrix that rotates its columns and then differentially downweights (by suitable functions of n) the q rotated columns that are strongly or semi-strongly identified for $q \in \{1, ..., p\}$, as determined by the magnitude of the singular values $\{\tau_{jF_n} : j \leq p\}$ of $\Omega_{F_n}^{-1/2} E_{F_n} G_i$ for $n \geq 1$. This eliminates the otherwise explosive behavior of these columns. Such sequences of distributions come from $\cup_{j=0}^{q} \mathcal{F}_{0j}$. For such sequences, the asymptotic version of the normalized $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ matrix has full column rank a.s. because, for all $j \leq q$, (i) the first j nonstochastic (rotated) columns have full column rank by the choice of rotation and (ii) the expected outer-product matrix of every linear combination of the remaining $p - j$ asymptotically normal (rotated) rows and columns, i.e., $\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_i B_{F,p-j}\xi}$, satisfies the $\lambda_{p-j}(\cdot)$ lower bound condition in $\mathcal{F}_{0j}$.

5 Kleibergen's CLR Test with Jacobian-Variance Weighting

In this section, we consider Kleibergen's (2005, Sec. 5.1) nonlinear CLR test that employs the Jacobian-variance weighting. This test utilizes a rank statistic, $rk_n(\theta)$, that is suitable for testing the hypothesis $rank[E_F G_i] \leq p - 1$ against $rank[E_F G_i] = p$. For example, the rank statistics of Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006) have been suggested for this purpose. Given $rk_n(\theta)$ and any $p \geq 1$, Kleibergen (2005) defines the nonlinear CLR test statistic as

$$CLR_n(\theta) := \frac{1}{2}\left(AR_n(\theta) - rk_n(\theta) + \sqrt{(AR_n(\theta) - rk_n(\theta))^2 + 4LM_n(\theta) \cdot rk_n(\theta)}\right). \quad (5.1)$$

This definition mimics the definition of the likelihood ratio (LR) statistic in the homoskedastic normal linear IV regression model with fixed regressors when $p = 1$, see Moreira (2003, eqn. (3)). However, it differs from the LR statistic in the latter model when $p \geq 2$. Smith (2007), Newey and Windmeijer (2009), and Guggenberger, Ramalho, and Smith (2012) consider GEL versions of the CLR statistic in (5.1).

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

$$clr(r) := \frac{1}{2}\left(\chi_p^2 + \chi_{k-p}^2 - r + \sqrt{(\chi_p^2 + \chi_{k-p}^2 - r)^2 + 4\chi_p^2 r}\right) \quad (5.2)$$

for $0 \leq r < \infty$ and the chi-square random variables χ_p^2 and χ_{k-p}^2 in (5.2) are independent. The

CLR test rejects the null hypothesis $H_0 : \theta = \theta_0$ if $CLR_n(\theta_0) > c(1-\alpha, rk_n(\theta_0))$.

Kleibergen (2005, p. 1114) recommends using a rank statistic that is a function of $\widehat{D}_n(\theta)$ and a consistent estimator of the covariance matrix of the asymptotic distribution of $vec(\widehat{D}_n(\theta))$ (after suitable normalization), denoted $\widetilde{V}_{Dn}(\theta) \in R^{kp \times kp}$. (Also, see (37) of Kleibergen (2007).) In the i.i.d. case considered here, $\widetilde{V}_{Dn}(\theta)$ is defined by

$$\widetilde{V}_{Dn}(\theta) := n^{-1}\sum_{i=1}^{n} vec(G_i(\theta) - \widehat{G}_n(\theta))vec(G_i(\theta) - \widehat{G}_n(\theta))' - \widehat{\Gamma}_n(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{\Gamma}_n(\theta)', \text{ where}$$
$$\widehat{\Gamma}_n(\theta) := (\widehat{\Gamma}_{1n}(\theta)', ..., \widehat{\Gamma}_{pn}(\theta)')' \in R^{pk \times k}. \tag{5.3}$$

The Jacobian-variance weighted version of $\widehat{D}_n(\theta)$ upon which the rank statistic depends is

$$\widehat{D}_n^{\dagger}(\theta) := vec_{k,p}^{-1}(\widetilde{V}_{Dn}^{-1/2}(\theta)vec(\widehat{D}_n(\theta))) = \sum_{j=1}^{p}(\widetilde{M}_{1jn}(\theta)\widehat{D}_{jn}(\theta), ..., \widetilde{M}_{pjn}(\theta)\widehat{D}_{jn}(\theta)), \text{ where}$$
$$\widetilde{M}_n(\theta) = \begin{bmatrix} \widetilde{M}_{11n}(\theta) & \cdots & \widetilde{M}_{1pn}(\theta) \\ \vdots & \ddots & \vdots \\ \widetilde{M}_{p1n}(\theta) & \cdots & \widetilde{M}_{ppn}(\theta) \end{bmatrix} := \widetilde{V}_{Dn}^{-1/2}(\theta) \in R^{kp \times kp} \text{ and } \widetilde{M}_{j\ell n}(\theta) \in R^{k \times k} \text{ for } j, \ell \leq p. \tag{5.4}$$

The function $vec_{k,p}^{-1}(\cdot)$ is the inverse of the $vec(\cdot)$ function for $k \times p$ matrices.[29] Similarly, Smith's (2007) nonlinear CLR test relies on a rank statistic that is a function of $\widehat{D}_n^{\dagger}(\theta)$. We refer to $\widehat{D}_n^{\dagger}(\theta)$ as the *Jacobian-variance-weighted* orthogonalized sample Jacobian.

For example, Kleibergen's (2005, 2007) rank statistic based on the Robin and Smith (2000) statistic is

$$rk_n(\theta) := \lambda_{\min}(n(\widehat{D}_n^{\dagger}(\theta))'\widehat{D}_n^{\dagger}(\theta)). \tag{5.5}$$

The asymptotic null distribution of $n^{1/2}\widehat{D}_n^{\dagger}T_n^{\dagger}$ is given in the following theorem.[30] Here $T_n^{\dagger}$ is a nonstochastic $p \times p$ matrix that rotates $\widehat{D}_n^{\dagger}$ by an orthogonal matrix and then rescales the resulting columns so that $n^{1/2}\widehat{D}_n^{\dagger}T_n^{\dagger}$ has a non-degenerate asymptotic distribution. We let $\{\lambda_{n,h} : n \geq 1\}$ index a sequence of distributions $\{F_n : n \geq 1\}$ that has certain properties, including convergence of

$$E_{F_n}G_i \text{ and } Var_{F_n}\begin{pmatrix} f_i^* \\ vech(f_i^* f_i^{*\prime}) \end{pmatrix}, \text{ where } f_i^* := \begin{pmatrix} g_i \\ vec(G_i - E_{F_n}G_i) \end{pmatrix}, \tag{5.6}$$

[29] Thus, the domain of $vec_{k,p}^{-1}(\cdot)$ consists of kp-vectors and its range consists of $k \times p$ matrices.

[30] As mentioned above, for notational simplicity, we often drop the dependence on θ_0 for statistics that are computed under the null hypothesis value $\theta = \theta_0$. Thus, $\widehat{D}_n^{\dagger}$ and $T_n^{\dagger}$ denote $\widehat{D}_n^{\dagger}(\theta_0)$ and $T_n^{\dagger}(\theta_0)$, respectively.

and convergence (possibly to infinity) of certain functions of $n^{1/2}E_{F_n}G_i$. In (5.6), $vech(\cdot)$ denotes the half vectorization operator that vectorizes the elements in the columns of a symmetric matrix that are on and below the main diagonal. We define $T_n^\dagger$ and $\{\lambda_{n,h} : n \geq 1\}$ precisely in Section 18 in the Supplemental Material, see (18.9) and (18.28), rather than here. The reason is that it takes several pages to define these quantities precisely, and the exact form of these quantities is not important. What is important is the general form of the asymptotic distribution of $n^{1/2}\widehat{D}_n^\dagger T_n^\dagger$, which can be specified without these definitions.

The following theorem is a key ingredient in determining the asymptotic size of Kleibergen's CLR test with Jacobian-variance weighting when $p \geq 2$. For this CLR test based on the Robin and Smith (2000) rank statistic (defined in (5.5)), the asymptotic size is determined and a formula for it is stated in Section 18 in the Supplemental Material. The formula for asymptotic size is given by the supremum of the asymptotic null rejection probabilities over sequences of distributions with different identification strengths. For some sequences, the asymptotic versions of the sample moments and the (suitably normalized) Jacobian-variance weighted orthogonalized sample Jacobian are independent, and the asymptotic null rejection probabilities are necessarily equal to the nominal size α.

However, when $p \geq 2$, for some sequences, these asymptotic quantities are not necessarily independent, and the asymptotic null rejection probabilities are not necessarily equal to the nominal size α. (The problematic sequences of distributions are of the nonstandard weak identification type, which requires $p \geq 2$.) The asymptotic null rejection probabilities could be larger or smaller than α (or both) depending on the model. If they are larger (or larger and smaller), the test does not have correct asymptotic size and is not asymptotically similar. If they are smaller, the test has correct asymptotic size, but is not asymptotically similar. The outcome that obtains depends on the specific model and moment conditions. Hence, when $p \geq 2$, we cannot say that, under general conditions, the Jacobian-variance weighted CLR test has correct asymptotic size.

Although the asymptotic size formula for the Jacobian-variance weighted CLR test is an important result of this paper, it is stated in the Supplemental Material because the notation and definitions needed to state it are extremely lengthy. Instead, we state the following result here, which shows why we cannot show that this CLR test necessarily has correct asymptotic size when $p \geq 2$.

Theorem 5.1 *Under the null hypothesis* $H_0 : \theta = \theta_0$ *and under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_{KCLR}\ \forall n \geq 1$ *(as defined in Section* 18.2 *in the Supplemental Material),* $n^{1/2}(\widehat{g}_n, \widehat{D}_n^\dagger T_n^\dagger) \to_d (\overline{g}_h, \overline{\Delta}_h^\dagger + \overline{M}_h^\dagger)$, *where* $(\overline{g}_h, \overline{\Delta}_h^\dagger, \overline{M}_h^\dagger)$ *has a multivariate normal distribution whose mean and variance matrix depend on* $\lim Var_{F_n}((f_i^{*\prime}, vech\,(f_i^* f_i^{*\prime})')')$ *and on the limits of certain functions of* $E_{F_n}G_i$

and $\overline{g}_h$ *and* $\overline{\Delta}_h^{\dagger}$ *are independent.*

Comments: **(i)** The quantities $\overline{g}_h$, $\overline{\Delta}_h^{\dagger}$, and $\overline{M}_h^{\dagger}$, which appear in Theorem 5.1, are complicated nonrandom linear functions of a mean zero multivariate normal random vector $\overline{L}_h$ whose variance matrix equals the limit of the variance that appears in (5.6). These linear functions are given explicitly in (18.13), (18.15), and (18.19) in Section 18 in the Supplemental Material.

(ii) When trying to show that Kleibergen's (2005, 2007) and Smith's (2007) CLR tests have correct asymptotic size, one needs the conditional asymptotic distributions of the LM statistic and the statistic $J_n(\theta_0) := AR_n(\theta_0) - LM_n(\theta_0)$ given the asymptotic rank statistic, which is a nonrandom function of $\overline{\Delta}_h^{\dagger} + \overline{M}_h^{\dagger}$, to be χ^2_p and χ^2_{k-p} distributions, respectively.[31] The asymptotic distributions of $LM_n(\theta_0)$ and $J_n(\theta_0)$ are quadratic forms in $\overline{g}_h$ with random idempotent weight matrices that depend on $\overline{\Delta}_h^{\dagger} + \overline{M}_h^{\dagger}$. If $\overline{M}_h^{\dagger} = 0^{k\times p}$ a.s., then conditional on $\overline{\Delta}_h^{\dagger}$, these asymptotic distributions are χ^2_p and χ^2_{k-p} distributions, as desired, because $\overline{g}_h$ and $\overline{\Delta}_h^{\dagger}$ are independent. Alternatively, if $(\overline{M}_h^{\dagger}, \overline{\Delta}_h^{\dagger})$ is independent of $\overline{g}_h$, one obtains the desired conditional asymptotic distributions given $(\overline{M}_h^{\dagger}, \overline{\Delta}_h^{\dagger})$. However, when $\overline{M}_h^{\dagger} \neq 0^{k\times p}$ with positive probability, one typically does not get the desired conditional asymptotic distributions, because $\overline{M}_h^{\dagger}$ and $\overline{g}_h$ typically are correlated in this case.

(iii) In some scenarios, $\overline{M}_h^{\dagger} = 0^{k\times p}$ a.s. This always occurs if $p = 1$.[32] If $p \geq 2$, it occurs if $E_{F_n} G_i \to 0^{k\times p}$, which covers the cases where all of the parameters are weakly identified in the standard sense or semi-strongly identified. If $p \geq 2$, it also occurs if the smallest singular value of $n^{1/2} E_{F_n} G_i$ diverges to infinity, which covers the case where all of the parameters are strongly or semi-strongly identified.

In addition, $(\overline{M}_h^{\dagger}, \overline{\Delta}_h^{\dagger})$ is independent of $\overline{g}_h$, if g_i and $f_i^* f_i^{*\prime}$ are uncorrelated (for all F in the parameter space of interest), which holds in some special cases. For example, in a homoskedastic linear IV model with p rhs endogenous variables and fixed IV's, it holds if (i) the reduced-form equation error vector V_{2i} is of the form $V_{2i} = K_1 u_i + K_2 \xi_i$, where u_i is the structural equation error, K_1 is some constant p vector, K_2 is some constant $p \times p$ matrix, and ξ_i is some mean zero random p vector, (ii) u_i is independent of ξ_i, and (iii) u_i is symmetrically distributed about zero with three moments finite. These conditions hold if $(u_i, V_{2i}')'$ has a multivariate normal distribution, but fail for most joint distributions of $(u_i, V_{2i}')'$.[33,34]

---

[31] See the proof of Theorem 10.1 for details.

[32] The proof of this is given in Comment (ii) to Theorem 18.3 in the Supplemental Material.

[33] The correlation between g_i and $f_i^* f_i^{*\prime}$ is zero in this case by the following: $y_{1i} = Y_{2i}'\theta + u_i$, $Y_{2i} = Z_i'\pi + V_{2i}$, $g_i = Z_i u_i$, $G_i = -Z_i Y_{2i}$, and $f_i^* = (u_i, -V_{2i}')' \otimes Z_i$. In consequence, the product of any element of g_i and any element of $f_i^* f_i^{*\prime}$ is of the form of a constant times $Z_{is} Z_{it} Z_{i\ell}$ times a linear combination (with constant coefficients) of u_i^3, $u_i \xi_{ij}^2$, $u_i \xi_{ij} \xi_{im}$, and $u_i^2 \xi_{ij}$ for some $s, t, \ell, j, m \geq 1$, where Z_{is} and ξ_{ij} denote the sth element of Z_i and the jth element of ξ_i, respectively. The expectations of these terms are all zero under conditions (i)-(iii).

[34] In addition, lack of correlation between g_i and $f_i^* f_i^{*\prime}$ typically does not hold if the IV's are random and independent

Typically, $\overline{M}_h^\dagger$ is non-zero (with positive probability) and correlated with $\overline{g}_h$ whenever some parameters are strongly identified and others are weakly identified in either the standard sense or in a jointly weakly-identified sense. In consequence, in general, when $p \geq 2$, one cannot verify that Kleibergen's (2005, 2007) and Smith's (2007) CLR tests have correct asymptotic size using the standard proof. Depending upon the particular sequence of distributions considered and the particular moment functions considered, the correlation between $\overline{g}_h$ and $\overline{\Delta}_h^\dagger + \overline{M}_h^\dagger$ could increase or decrease the asymptotic null rejection probability from the nominal probability α.

(iv) Numerical simulations of a linear IV model (with $p = 2$, one parameter strongly identified, one parameter weakly identified, and a particular distribution of the errors) corroborate the finding that $\overline{M}_h^\dagger$ and $\overline{g}_h$ can be correlated asymptotically, see Section 18.3 in the Supplemental Material for details. In the model considered, the simulated asymptotic null rejection probabilities are found to be in $[4.95, 5.01]$, which are very close to the test's nominal size of 5.00. Whether this occurs for a wide range of error distributions and for other moment condition models is an open question. It appears that this question needs to be answered on a case by case basis.

(v) If the random weight matrix $\widetilde{V}_{Dn}^{-1/2}(\theta)$ is replaced in the definition of $\widehat{D}_n^\dagger(\theta)$ by the non-random quantity that it is estimating, call it $V_{Dn}^{-1/2}(\theta)$, then the asymptotic distribution of the quantities in Theorem 5.1 is given by $(\overline{g}_h, \overline{\Delta}_h^\dagger)$, where $\overline{g}_h$ and $\overline{\Delta}_h^\dagger$ are independent. Thus, the appearance of $\overline{M}_h^\dagger$ in Theorem 5.1 is due to the estimation of the weight matrix. If $V_{Dn}^{-1/2}(\theta)$ is known (which almost never occurs in practice) and is used to define $\widehat{D}_n^\dagger(\theta)$, then the Kleibergen (2005, 2007) and Smith (2007) CLR tests can be shown to have correct asymptotic size even when $p \geq 2$.

(vi) The reason that the estimator $\widetilde{V}_{Dn}^{-1/2}$ affects the limit distribution of $n^{1/2}\widehat{D}_n^\dagger T_n^\dagger$ is because it weights the columns of $\widehat{D}_n$ differently. If one bases the rank statistic on $\widetilde{W}_n \widehat{D}_n$, where $\widetilde{W}_n$ $(= \widetilde{W}_n(\theta_0))$ is some random $k \times k$ matrix that converges in probability to a nonsingular matrix, then the nondegenerate asymptotic distribution of $\widetilde{W}_n$ (after suitable normalization) does not affect the asymptotic distribution of $\widetilde{W}_n \widehat{D}_n$, only the plim of $\widetilde{W}_n$ does (and the corresponding CLR test has correct asymptotic size). The proof is given in Section 18.5 in the Supplemental Material.

(vii) In Section 18.1 in the Supplemental Material, we provide an example that illustrates the results of Theorem 5.1 and Comments (iv) and (v) to Theorem 5.1.

(viii) Given the result of Theorem 5.1, we do not recommend using a rank statistic that depends on an estimator of the asymptotic variance matrix of $vec(\widehat{D}_n(\theta))$ (after suitable normalization) when $p \geq 2$.

(ix) The CLR test with Jacobian-variance weighting (in the rank statistic) is asymptotically efficient in a GMM sense under strong IV's provided $rk_n(\theta) \rightarrow_p \infty$ under strong IV's, which is the

of $(u_i, V_{2i}')'$. This is a consequence of the definition of $E_F G_i$ being different between the fixed and random IV cases.

case for all of the rank tests considered in the literature.[35]

As indicated in Comment (iii) to Theorem 5.1, when $p = 1$, $\overline{M}_h^\dagger = 0^{k \times p}$ a.s. In consequence, Kleibergen's (2005) CLR test has correct asymptotic size when $p = 1$ for a suitable parameter space of distributions F and a suitable rank statistic, such as that in (5.5). We consider the parameter space

$$\mathcal{F}_{JVW,p=1} := \{F \in \mathcal{F} : \lambda_{\min}(\Psi_F^{G_i} - E_F G_i E_F G_i') \geq \delta_3\} \tag{5.7}$$

for some $\delta_3 > 0$. For the corresponding CS, we consider the parameter space $\mathcal{F}_{\Theta,JVW,p=1} := \{(F, \theta_0) : F \in \mathcal{F}_{JVW,p=1}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{JVW,p=1}(\theta_0)$ denotes the set $\mathcal{F}_{JVW,p=1}$ defined in (5.7) with its dependence on θ_0 made explicit.

We have $\mathcal{F}_{JVW,p=1} \subset \mathcal{F}_{00}$ ($\subset \mathcal{F}_0$) when $\delta_3 = \delta_2$ (by (3.7) and condition (iv) in (3.9)), where $\mathcal{F}_{00} = \mathcal{F}_{0j}$ with $j = 0$ (for $\mathcal{F}_{0j}$ defined in (3.7)) and $\mathcal{F}_0$ is the parameter space for which the moment-variance weighted CLR test has correct asymptotic size, see Theorem 6.1 below. When $p = 1$, $\mathcal{F}_0 = \mathcal{F}_{00} \cup \mathcal{F}_{01}$ and the set $\mathcal{F}_{01}$ places no restrictions on the variance matrix or outer-product matrix of the orthogonalized sample Jacobian (i.e., $\Psi_{1F}(\xi)$). The parameter space $\mathcal{F}_{JVW,p=1}$ cannot be enlarged to include a set like $\mathcal{F}_{01}$, because the condition on the variance matrix of the orthogonalized sample Jacobian $\Psi_F^{G_i} - E_F G_i E_F G_i'$ in $\mathcal{F}_{JVW,p=1}$ is needed to obtain the nonsingularity of the probability limit of the weight matrix $\widetilde{V}_{Dn}$.

When $p = 1$, the Robin and Smith (2000) rank statistic given in (5.5) (with $\theta = \theta_0$), which is based on Kleibergen's (2005, 2007) recommended Jacobian-variance weight matrix $\widetilde{V}_{Dn}^{-1/2}$, reduces to

$$rk_n := n\widehat{D}_n{}'\widetilde{V}_{Dn}^{-1}\widehat{D}_n. \tag{5.8}$$

Theorem 5.2 *Suppose* $p = 1$. *The asymptotic size of the* CLR *test with Jacobian-variance weighting, defined by* (5.1), (5.2), *and* (5.8), *equals its nominal size* $\alpha \in (0, 1)$ *for the parameter space* $\mathcal{F}_{JVW,p=1}$. *Furthermore, this* CLR *test is asymptotically similar* (*in a uniform sense*) *for this parameter space. Analogous results hold for the* CLR CS *with Jacobian-variance weighting for the parameter space* $\mathcal{F}_{\Theta,JVW,p=1}$.

Comment: Correct asymptotic size holds for Kleibergen's CLR test with Jacobian-variance weighting when $p = 1$ because $\widehat{D}_n$ has only one column in this case, so it is impossible to have unequal column weights.

[35] This holds because all CLR tests of the form in (5.1) and (5.2) are asymptotically equivalent to the LM test in (4.3) under the null and $n^{-1/2}$ local alternatives under strong IV's, by (10.3) and (10.4) in the proof of Theorem 10.1 in Section 10 in the Appendix, and, as noted above, the LM test is asymptotically efficient in a GMM sense under strong IV's. Note that, by definition in (4.3), the LM statistic uses moment-variance weighting of $\widehat{D}_n(\theta)$ in its projection matrix.

6 Kleibergen's CLR Test with Moment-Variance Weighting

Newey and Windmeijer (2009) and Guggenberger, Ramalho, and Smith (2012) consider a version of Kleibergen's (2005) CLR test that uses a rank statistic that depends on

$$\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta), \tag{6.1}$$

rather than $\widehat{D}_n^{\dagger}(\theta)$. We refer to $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$ as the *moment-variance-weighted* orthogonalized sample Jacobian. This choice gives equal weight to each of the columns of $\widehat{D}_n$. In this section, we show that this choice combined with the Robin and Smith (2000) rank statistic yields a nonlinear CLR test that has correct asymptotic size for the parameter space $\mathcal{F}_0$. In this case, the rank statistic is

$$rk_n(\theta) := \lambda_{\min}(n\widehat{D}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{D}_n(\theta)). \tag{6.2}$$

Theorem 6.1 *The asymptotic size of the* CLR *test with moment-variance weighting, defined by* (5.1), (5.2), *and* (6.2), *equals its nominal size* $\alpha \in (0,1)$ *for the parameter space* $\mathcal{F}_0$ (*defined in* (3.7)). *Furthermore, this* CLR *test is asymptotically similar* (*in a uniform sense*) *for this parameter space. Analogous results hold for the* CLR CS *with moment-variance weighting for the parameter space* $\mathcal{F}_{\Theta,0}$, *defined in* (3.14).

Comments: (i) Neither Newey and Windmeijer (2009) nor Guggenberger, Ramalho, and Smith (2012) provide an asymptotic size result like that in Theorem 6.1. Guggenberger, Ramalho, and Smith (2012) provide asymptotic null rejection probabilities only under Stock and Wright's (2000) Assumption C, plus a high-level condition that involves the asymptotic behavior of the rank statistic. Verifying this high-level assumption under parameter sequences that satisfy Assumption C turns out to be very challenging. We do so in this paper, also see Comment (ii). But note that the proof of Theorem 6.1, given in Section 10 in the Appendix, involves much more than this. It is complicated because it needs to consider a broad array of different types of identification ranging from standard weak identification, to joint weak identification, to semi-strong and strong identification.

(ii) The proof of Theorem 6.1 actually allows for the use of any rank statistic that satisfies an assumption called Assumption R, which is stated in Section 10, not just the rank statistic $rk_n(\theta)$ in (6.2). Assumption R is verified using Theorem 8.4 below for the rank statistic in (6.2). With some changes, Assumption R can be verified using Theorem 8.4 when the rank statistic is of an "equally-weighted" Robin-Smith form, but with a different weight matrix than in (6.2). That is, Assumption R can be verified when $rk_n(\theta)$ is as in (6.2) but with $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)$ replaced by $\widetilde{W}_n(\theta)\widehat{D}_n(\theta)$ for some $k \times k$ weight matrix $\widetilde{W}_n(\theta)$ that is positive definite (pd) asymptotically. (This is what we mean by equally-weighted.) This is done in Section 18.5 in the Supplemental Material. In contrast, by Theorem 5.1, when $p \geq 2$, Assumption R typically does not hold for any rank statistic that depends on the Jacobian-variance weighted statistic $\widehat{D}_n^{\dagger}(\theta)$.

(iii) The CLR test considered in Theorem 6.1 is asymptotically efficient in a GMM sense under strong IV's provided $rk_n(\theta) \to_p \infty$ under strong IV's, see Comment (iii) to Theorem 4.1 for more details.

(iv) Assumption R likely holds for the Cragg and Donald (1996, 1997) and Kleibergen and Paap (2006) rank statistics when they are based on an equally-weighted function of $\widehat{D}_n(\theta)$. However, showing this is not easy. We do not do so here.

Although the rank statistic in (6.2) yields a test with correct asymptotic size, it has some drawbacks. The use of the pre-multiplication weight matrix $\widehat{\Omega}_n^{-1/2}(\theta)$ and no post-multiplication weight matrix for $\widehat{D}_n(\theta)$ is arbitrary. The choice of these weight matrices is important for power purposes because it is a major determinant of the magnitude of $rk_n(\theta)$ and the latter enters both the test statistic and the data-dependent critical value function. We show in Section 14 in the Supplemental Material to AG2 that the rank statistic in (6.2) does not reduce to the rank statistic in Moreira's (2003) CLR test in the homoskedastic normal linear IV regression model with fixed regressors even when $p = 1$. Specifically, the $rk_n(\theta)$ statistic in (6.2) differs asymptotically from the rank statistic in Moreira's CLR test by a scale factor that can range between 0 and ∞ depending on the scenario considered. This is undesirable because Moreira's CLR test has been shown to have some approximate optimal power properties in the aforementioned model when $p = 1$.

In addition, the CLR test with moment-variance weighting, which is considered in this section, has correct asymptotic size for the parameter space $\mathcal{F}_0$, but not necessarily for the larger parameter space $\mathcal{F}$.

These disadvantages motivate interest in the SR-CQLR$_1$ and SR-CQLR$_2$ tests considered in AG2.

7 Time Series Observations

In this section, we generalize the results of Theorems 4.1, 5.2, and 6.1 from i.i.d. observations to strictly stationary strong mixing observations. In the time series case, F denotes the distribution of the stationary infinite sequence $\{W_i : i = ..., 0, 1, ...\}$.[36] Let a_i be a random vector that depends on W_i, such as $vec(G_i)$ or $C_{F,k-j}' \Omega_F^{-1/2} G_i B_{F,p-j} \xi$. In the time series case, we define Ω_F and $\Psi_F^{a_i}$

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

differently from their definitions in (3.2) for the i.i.d. case. For the time series case, we define $\Sigma_F^{a_i}$, $\Gamma_F^{a_i}$, Ω_F, and $\Psi_F^{a_i}$ as follows:[37]

$$\Sigma_F^{a_i} := \sum_{m=-\infty}^{\infty} E_F(a_i - E_F a_i)(a_{i-m} - E_F a_{i-m})', \ \Gamma_F^{a_i} := \sum_{m=-\infty}^{\infty} E_F a_i g_{i-m}',$$
$$\Omega_F := \sum_{m=-\infty}^{\infty} E_F g_i g_{i-m}', \text{ and } \Psi_F^{a_i} := \Sigma_F^{a_i} - \Gamma_F^{a_i} \Omega_F^{-1} \Gamma_F^{a_i \prime}. \quad (7.1)$$

Note that $\Psi_F^{a_i} = \lim Var_F(n^{-1/2} \sum_{i=1}^n (a_i - \Gamma_F^{a_i} \Omega_F^{-1} g_i))$.[38]

The time series analogue $\mathcal{F}_{TS}$ of the space of distributions $\mathcal{F}$, defined in (3.1), is

$$\begin{aligned} \mathcal{F}_{TS} := \{F : \{W_i : i = ..., 0, 1, ...\} &\text{ are stationary and strong mixing under } F \text{ with} \\ &\text{strong mixing numbers } \{\alpha_F(m) : m \geq 1\} \text{ that satisfy } \alpha_F(m) \leq Cm^{-d}, \\ &E_F g_i = 0^k, \ E_F ||(g_i', vec(G_i)')'||^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(\Omega_F) \geq \delta\} \end{aligned} \quad (7.2)$$

for some $\gamma, \delta > 0$, $d > (2+\gamma)/\gamma$, and $C, M < \infty$, where Ω_F is defined in (7.1).

We define the time series parameter spaces of distributions $\mathcal{F}_{TS,0}$ and $\{\mathcal{F}_{TS,0j} : 0 \leq j \leq p\}$ as $\mathcal{F}_0$ and $\{\mathcal{F}_{0j} : 0 \leq j \leq p\}$ are defined in (3.7), but with $\mathcal{F}_{TS}$ in place of $\mathcal{F}$, with $\Psi_F^{a_i}$ defined as in (7.1), and with the definitions of $(\tau_{1F}, ..., \tau_{pF})$, B_F, and C_F in (3.3)-(3.5) employing the definition of Ω_F in (7.1). We define the time series parameter space of distributions $\mathcal{F}_{TS,JVW,p=1}$ as $\mathcal{F}_{JVW,p=1}$ is defined in (5.7), but with $\mathcal{F}_{TS}$ in place of $\mathcal{F}$, with $\Psi_F^{G_i}$ defined as in (7.1), and with $E_F G_i E_F G_i'$ deleted (because $\Psi_F^{G_i} := \Sigma_F^{G_i} - \Gamma_F^{G_i} \Omega_F^{-1} \Gamma_F^{G_i \prime}$ and $\Sigma_F^{G_i}$ is defined to be $E_F(G_i - E_F G_i)(G_i - E_F G_i)'$ in the time series case, rather than $E_F G_i G_i'$). That is, $\mathcal{F}_{TS,JVW,p=1} := \{F \in \mathcal{F}_{TS} : \lambda_{\min}(\Psi_F^{G_i}) \geq \delta_3\}$ for some $\delta_3 > 0$. For CS's, we use the parameter spaces $\mathcal{F}_{\Theta,TS,0} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,0}(\theta_0), \theta_0 \in \Theta\}$ and $\mathcal{F}_{\Theta,TS,JVW,p=1} := \{(F, \theta_0) : F \in \mathcal{F}_{TS,JVW,p=1}(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{TS,0}(\theta_0)$ and $\mathcal{F}_{TS,JVW,p=1}(\theta_0)$ denote $\mathcal{F}_{TS,0}$ and $\mathcal{F}_{TS,JVW,p=1}$ with their dependence on θ_0 made explicit.

The sufficient conditions for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$ provided in (3.9) and (3.10) also hold in the time series setting with $\Psi_F^{a_i}$ and $\Sigma_F^{a_i}$ defined as in (7.1).

[37] Note that the definition of $\Sigma_F^{a_i}$ in (7.1) differs from its definition in (3.2) in two ways. First, there are the lag $m \neq 0$ terms. Second, there is the re-centering of a_i by its mean $E_F a_i$. Re-centering is needed in the time series context to ensure that $\Sigma_F^{a_i}$ is a convergent sum. In the i.i.d. case, we avoid re-centering because without it the restriction in $\mathcal{F}_0$, defined in (3.7), is weaker.

[38] This follows by calculations analogous to those in (19.3) and (19.4) in the proof of Theorem 7.1 below.

Now, we define the LM and CLR test statistics in the time series context. To do so, we let

$$
\begin{aligned}
V_F &:= \lim Var_F\left(n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} g_i \\ vec(G_i)\end{pmatrix}\right) \\
&= \sum_{m=-\infty}^{\infty} E_F\begin{pmatrix} g_i \\ vec(G_i - E_F G_i)\end{pmatrix}\begin{pmatrix} g_{i-m} \\ vec(G_{i-m} - E_F G_{i-m})\end{pmatrix}'. \qquad (7.3)
\end{aligned}
$$

The second equality holds for all $F \in \mathcal{F}_{TS}$ (as shown in the proof of Lemma 19.1 in Section 19 in the Supplemental Material).

The test statistics depend on an estimator $\widehat{V}_n(\theta_0)$ of V_F. This estimator is (typically) a heteroskedasticity and autocorrelation consistent (HAC) variance estimator based on the observations $\{f_i - \widehat{f}_n : i \leq n\}$, where $f_i := (g_i', vec(G_i)')'$ and $\widehat{f}_n(\theta) := (\widehat{g}_n', vec(\widehat{G}_n)')'$. There are a number of HAC estimators available in the literature, e.g., see Newey and West (1987a) and Andrews (1991). The asymptotic size and similarity properties of the tests are the same for any consistent HAC estimator. Hence, for generality, we do not specify a particular estimator $\widehat{V}_n(\theta_0)$. Rather, we state results that hold for any estimator $\widehat{V}_n(\theta_0)$ that satisfies the following consistency condition when the null value θ_0 is the true value.

Assumption V: $\widehat{V}_n(\theta_0) - V_{F_n} \rightarrow_p 0^{(p+1)k\times(p+1)k}$ under $\{F_n : n \geq 1\}$ for any sequence $\{F_n \in \mathcal{F}_{TS} : n \geq 1\}$ for which $V_{F_n} \rightarrow V$ for some pd matrix V.

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

$$
\widehat{V}_n(\theta) = \begin{bmatrix} \widehat{\Omega}_n(\theta) & \widehat{\Gamma}_{1n}(\theta)' & \cdots & \widehat{\Gamma}_{pn}(\theta)' \\ \widehat{\Gamma}_{1n}(\theta) & \widehat{V}_{G_{11}n}(\theta) & \cdots & \widehat{V}_{G_{p1}n}'(\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\Gamma}_{pn}(\theta) & \widehat{V}_{G_{p1}n}(\theta) & \cdots & \widehat{V}_{G_{pp}n}(\theta) \end{bmatrix}. \qquad (7.4)
$$

Under Assumption V, $\widehat{\Omega}_n(\theta_0) \rightarrow_p \Omega_F$ under F and $\widehat{\Gamma}_n(\theta_0) = (\widehat{\Gamma}_{1n}(\theta_0)', ..., \widehat{\Gamma}_{pn}(\theta_0)')' \rightarrow_p \Gamma_F^{vec(G_i)}$ under F.

In the time series case, for the LM test, the CLR test with moment-variance weighting, and when $p = 1$ the CLR test with Jacobian-variance weighting, the definitions of the statistics $\widehat{g}_n(\theta)$, $\widehat{G}_n(\theta)$, $AR_n(\theta)$, $LM_n(\theta)$, $\widehat{D}_n(\theta)$, $CLR_n(\theta)$, and $rk_n(\theta)$ are the same as in (4.1)-(5.1), but with $\widehat{\Omega}_n(\theta)$ and $\widehat{\Gamma}_{jn}(\theta)$ for $j = 1, ..., p$ defined as in Assumption V and (7.4) rather than as in Sections 4 and 5. In addition, when $p = 1$, for the CLR test with Jacobian-variance weighting, in the definition of $\widetilde{V}_{Dn}$ in (5.3), the matrix $n^{-1}\sum_{i=1}^{n} vec(G_i(\theta) - \widehat{G}_n(\theta))vec(G_i(\theta) - \widehat{G}_n(\theta))'$ is replaced by the lower right $pk \times pk$ submatrix of $\widehat{V}_n(\theta)$ in (7.4) (and $\widehat{\Omega}_n(\theta)$ and $\widehat{\Gamma}_{jn}(\theta)$ for $j = 1, ..., p$ are defined as in (7.4)). With these changes, the critical values for the time series case are defined in the same way as in the i.i.d. case.

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

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

Appendix

This Appendix provides proofs of some of the results stated in the paper and shows that the eigenvalue condition in $\mathcal{F}_0$ is not redundant. For brevity, other proofs are provided in the Supplemental Material to this paper given in Andrews and Guggenberger (2014b). Section 8 in this Appendix states some basic results that are used in all of the proofs. For brevity, these results are proved in Sections 14-16 in the Supplemental Material. These results also are used in Andrews and Guggenberger (2014a) and should be useful for establishing the asymptotic sizes of other tests for moment condition models when strong identification is not assumed. Given the results in Section 8, Section 9 proves Theorem 4.1, Section 10 proves Theorem 6.1, and Section 11 proves Theorem 5.2. Theorem 5.1 is proved in Section 18 in the Supplemental Material. Section 12 shows that the eigenvalue condition in $\mathcal{F}_0$, defined in (3.7), is not redundant in Theorems 4.1, 5.2, and 6.1.

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

8 Basic Framework and Results for the Proofs

8.1 Uniformity

The proofs of Theorems 4.1, 5.2, and 6.1 use Corollary 2.1(c) in Andrews, Cheng, and Guggenberger (2009) (ACG). The latter result provides general sufficient conditions for the correct asymptotic size and (uniform) asymptotic similarity of a sequence of tests.

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

$$H := \{h \in (R \cup \{\pm\infty\})^J : h_{w_n}(\lambda_{w_n}) \to h \text{ for some subsequence } \{w_n\} \text{ of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\}. \quad (8.1)$$

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

Proposition 8.1 (ACG, Corollary 2.1(c)) *Under Assumption* B*, *the tests* $\{\phi_n : n \geq 1\}$ *have asymptotic size* α *and are asymptotically similar* (*in a uniform sense*). *That is,* $AsySz := \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda) = \alpha$ *and* $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda} RP_n(\lambda) = \limsup_{n\to\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda)$.

Comments: **(i)** By Comment 4 to Theorem 2.1 of ACG, Proposition 8.1 provides asymptotic size and similarity results for nominal $1-\alpha$ CS's, rather than tests, by defining λ as one would for a test, but having it depend also on the parameter that is restricted by the null hypothesis, by enlarging the parameter space Λ correspondingly (so it includes all possible values of the parameter that is restricted by the null hypothesis), and by replacing (i) ϕ_n by a CS based on a sample of size n, (ii) α by $1-\alpha$, (iii) $RP_n(\lambda)$ by $CP_n(\lambda)$, where $CP_n(\lambda)$ denotes the coverage probability of the CS under λ when the sample size is n, and (iv) the first $\limsup_{n\to\infty} \sup_{\lambda\in\Lambda}$ that appears by $\liminf_{n\to\infty} \inf_{\lambda\in\Lambda}$. In the present case, where the null hypotheses are of the form $H_0 : \theta = \theta_0$ for $\theta \in \Theta$, for CS's, θ_0 is taken to be a subvector of λ and Λ is specified so that the value of this subvector ranges over Θ.

(ii) In the application of Proposition 8.1 to prove Theorems 4.1 and 6.1, one takes Λ to be a one-to-one transformation of $\mathcal{F}_0$ for tests, and one takes Λ to be a one-to-one transformation of $\mathcal{F}_{\Theta,0}$ for CS's. With these changes, the proofs for tests and CS's are the same. In consequence, we provide explicit proofs for tests only and obtain the proofs for CS's by analogous applications of Proposition 8.1. In the application of Proposition 8.1 to prove Theorem 5.2, the same is done but with $\mathcal{F}_{JVW,p=1}$ in place of $\mathcal{F}_0$.

(iii) We prove the test results in Theorems 4.1, 5.2, and 6.1 using Proposition 8.1 by verifying Assumption B* for suitable choices of λ and $h_n(\lambda)$.

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

We prove results for statistics that depend on random weight matrices $\widehat{W}_n \in R^{k\times k}$ and $\widehat{U}_n \in R^{p\times p}$. In particular, we consider statistics of the form $\widehat{W}_n\widehat{D}_n\widehat{U}_n$ and functions of this statistic, where $\widehat{D}_n$ is defined in (4.3). The definitions of the random weight matrices $\widehat{W}_n$ and $\widehat{U}_n$ depend upon the statistic that is of interest. They are taken to be of the form

$$\widehat{W}_n := W_1(\widehat{W}_{2n}) \in R^{k\times k} \text{ and } \widehat{U}_n := U_1(\widehat{U}_{2n}) \in R^{p\times p}, \tag{8.2}$$

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

$$W_F := W_1(W_{2F}) \text{ and } U_F := U_1(U_{2F}), \tag{8.3}$$

respectively.

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

$$\widehat{W}_n = \widehat{\Omega}_n^{-1/2} \text{ and } \widehat{U}_n = I_p. \tag{8.4}$$

In this case, the functions $W_1(\cdot)$ and $U_1(\cdot)$ are the identity functions, and the corresponding population quantities are $W_F = W_{2F} = \Omega_F^{-1/2}$, where $\Omega_F := E_F g_i g_i'$, see (3.2), and $U_F = U_{2F} = I_p$.

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

Example 3: With Kleibergen's (2005) CLR test with Jacobian-variance weighting and $p = 1$, which is considered in Section 5, we determine the asymptotic distribution of the rank statistic in (5.8) by taking $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$ and $\widehat{U}_n = I_p$. In this case, the functions $W_1(\cdot)$ and $U_1(\cdot)$ are as in Example 1, and the corresponding population quantities are $W_F = W_{2F} = (Var_F(vec(G_i)) - \Gamma_F^{vec(G_i)}\Omega_F^{-1}\Gamma_F^{vec(G_i)\prime})^{-1/2} = (\Psi_F^{vec(G_i)} - E_F G_i E_F G_i')^{-1/2}$, and $U_F = U_{2F} = I_p$. For this test, we need the asymptotic distribution of the LM statistic. In consequence, for this test, we also establish some asymptotic results with $\widehat{W}_n$ and $\widehat{U}_n$ defined as in Example 1.

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

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

$$\mathcal{F}_{WU} := \{F \in \mathcal{F} : \lambda_{\min}(W_F) \geq \delta_{WU}, \lambda_{\min}(U_F) \geq \delta_{WU}, ||W_F|| \leq M_{WU}, \text{ and } ||U_F|| \leq M_{WU}\} \tag{8.5}$$

for some constants $\delta_{WU} > 0$ and $M_{WU} < \infty$, where $\mathcal{F}$ is defined in (3.1). The set $\mathcal{F}_{WU} \cap \mathcal{F}_0$ is used to establish results for Kleibergen's LM and the CLR test with moment-variance weighting, considered in Section 6, using the fact that $\mathcal{F}_0 = \mathcal{F}_{WU} \cap \mathcal{F}_0$ for $\delta_{WU} > 0$ sufficiently small and $M_{WU} < \infty$ sufficiently large. This holds because for all $F \in \mathcal{F}_0$, $\lambda_{\min}(W_F) = \lambda_{\min}(\Omega_F^{-1/2}) = \lambda_{\max}^{-1/2}(\Omega_F) \geq ||\Omega_F||^{-1/2} \geq M_*^{-1/2}$ for some $M_* < \infty$ (because $||\Omega_F|| = ||E_F g_i g_i'|| \leq M_*$ for some $M_* < \infty$ by the moment conditions in $\mathcal{F}$), $||W_F|| = ||\Omega_F^{-1/2}|| \leq \lambda_{\min}^{-1/2}(\Omega_F) \leq \delta^{-1/2}$ (using the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}$), where $\delta > 0$, $\lambda_{\min}(U_F) = \lambda_{\min}(I_p) = 1$, and $||U_F|| = ||I_p|| = p$.

8.3 Reparametrization

To apply Proposition 8.1, we reparametrize the null distribution F to a vector λ. The vector λ is chosen such that for a subvector of λ convergence of a drifting subsequence of the subvector (after suitable renormalization) yields convergence in distribution of the test statistic and convergence in distribution of the critical value in the case of the CLR tests.

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

The vector λ depends on the following quantities. Let

$$B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } U_F'(E_F G_i)' W_F' W_F (E_F G_i) U_F \tag{8.6}$$

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

$$C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } W_F(E_F G_i) U_F U_F' (E_F G_i)' W_F' \text{[39]} \tag{8.7}$$

ordered so that the corresponding eigenvalues are $(\kappa_{1F}, ..., \kappa_{pF}, 0, ..., 0) \in R^k$. Let

$$(\tau_{1F}, ..., \tau_{pF}) \text{ denote the } p \text{ singular values of } W_F(E_F G_i) U_F, \tag{8.8}$$

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

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

Define the elements of λ to be[40,41,42]

$$\begin{aligned}
\lambda_{1,F} &:= (\tau_{1F}, ..., \tau_{pF})' \in R^p, \\
\lambda_{2,F} &:= B_F \in R^{p \times p}, \\
\lambda_{3,F} &:= C_F \in R^{k \times k}, \\
\lambda_{4,F} &:= (E_F G_{i1}, ..., E_F G_{ip}) \in R^{k \times p}, \\
\lambda_{5,F} &:= E_F \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix} \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}' \in R^{(p+1)k \times (p+1)k}, \\
\lambda_{6,F} &= (\lambda_{6,1F}, ..., \lambda_{6,(p-1)F})' := (\frac{\tau_{2F}}{\tau_{1F}}, ..., \frac{\tau_{pF}}{\tau_{(p-1)F}})' \in R^{p-1}, \text{ where } 0/0 := 0, \\
\lambda_{7,F} &:= W_{2F}, \\
\lambda_{8,F} &:= U_{2F}, \\
\lambda_{9,F} &:= F, \text{ and} \\
\lambda &= \lambda_F := (\lambda_{1,F}, ..., \lambda_{9,F}). \qquad (8.9)
\end{aligned}$$

The dimensions of W_{2F} and U_{2F} depend on the choices of $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$. We let $\lambda_{5,gF}$ denote the upper left $k \times k$ submatrix of $\lambda_{5,F}$. Thus, $\lambda_{5,gF} = E_F g_i g_i' = \Omega_F$.

We consider the parameter space Λ_0 for λ, which corresponds to $\mathcal{F}_{WU} \cap \mathcal{F}_0$, where $\mathcal{F}_{WU}$ and $\mathcal{F}_0$ are defined in (8.5) and (3.7), respectively. The parameter space Λ_0 and the function $h_n(\lambda)$ are defined by

$$\begin{aligned}
\Lambda_0 &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU} \cap \mathcal{F}_0\} \text{ and} \\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}). \qquad (8.10)
\end{aligned}$$

By the definition of $\mathcal{F}$, Λ_0 indexes distributions that satisfy the null hypothesis $H_0 : \theta = \theta_0$. The dimension J of $h_n(\lambda)$ equals the number of elements in $(\lambda_{1,F}, ..., \lambda_{8,F})$. Redundant elements in $(\lambda_{1,F}, ..., \lambda_{8,F})$, such as the redundant off-diagonal elements of the symmetric matrix $\lambda_{5,F}$, are not needed, but do not cause any problem. Note that two parameter spaces denoted by Λ_1 and Λ_2, which are larger than Λ_0, are considered for the two SR-CQLR tests analyzed in AG2. (We also use Λ_2 in this paper, see (8.11) below.)

We define λ and $h_n(\lambda)$ as in (8.9) and (8.10) because, as shown below, the asymptotic dis-

[40] 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.

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

[42] The vector $\lambda_{6,F}$ is only used in the proofs for CLR tests. It could be deleted when considering only an LM test.

tributions of the test statistics under a sequence $\{F_n : n \geq 1\}$ for which $h_n(\lambda_{F_n}) \rightarrow h \in H$ depend on the behavior of $\lim n^{1/2}\lambda_{1,F_n}$, as well as $\lim \lambda_{m,F_n}$ for $m = 2, ..., 8$. For example, the LM statistic in (4.3) depends on $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$, or equivalently, on $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n B_{F_n} S_n$ (because projections are invariant to rescaling and rhs transformations by nonsingular matrices), where S_n is a pd diagonal matrix that is designed to make this quantity $O_p(1)$ and not $o_p(1)$. We show that this quantity is asymptotically equivalent to $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{D}_n B_{F_n} S_n$. In turn, the latter quantity depends on $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{G}_n B_{F_n} = n^{1/2}\Omega_{F_n}^{-1/2}(\widehat{G}_n B_{F_n} - E_{F_n}G_i B_{F_n}) + n^{1/2}\Omega_{F_n}^{-1/2}E_{F_n}G_i B_{F_n}$. The quantity $vec(n^{1/2}\Omega_{F_n}^{-1/2}(\widehat{G}_n B_{F_n} - E_{F_n}G_i B_{F_n}))$ has a nondegenerate asymptotic normal distribution by the central limit theorem (CLT), using the behavior of $\lim \lambda_{s,F_n}$ for $s = 2, 4, 5$, the fact that B_{F_n} is an orthogonal matrix, and the restriction in $\mathcal{F}_0$. Hence, the asymptotic behavior of $vec(n^{1/2}\Omega_{F_n}^{-1/2}\widehat{G}_n B_{F_n})$ depends on that of $n^{1/2}\Omega_{F_n}^{-1/2}E_{F_n}G_i B_{F_n}$. Using the SVD of $\Omega_{F_n}^{-1/2}E_{F_n}G_i$, the latter is shown below to equal $\lambda_{3,F_n}Diag\{n^{1/2}\lambda_{1,F_n}\}$, where $Diag\{n^{1/2}\lambda_{1,F_n}\}$ denotes the $k \times p$ matrix with $n^{1/2}\lambda_{1,F_n}$ on the main diagonal and zeros elsewhere.

In Example 1 of Section 8.2 applied to the linear model (2.2), we have $W_F = \Omega_F^{-1/2}$ and τ_{jF} is the jth singular value of $-\Omega_F^{-1/2}E_F Z_i Y_{2i}' = -\Omega_F^{-1/2}E_F Z_i Z_i'\pi$, where $\Omega_F = E_F u_i^2 Z_i Z_i'$ for $j = 1, ..., p$. As is well known, if π is close to zero, weak instrument problems occur. But, as we show, matrices π that are close to being singular, without their columns being close to zero, also lead to weak IV problems. This is captured in the present set-up by τ_{pF} being close to zero in the sense that $\lim n^{1/2}\tau_{pF_n} < \infty$. If this occurs, then weak identification problems arise.

For notational convenience,

$$\{\lambda_{n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_n \in \Lambda_2 : n \geq 1\} \text{ for which } h_n(\lambda_n) \rightarrow h \in H, \text{ where}$$
$$\Lambda_2 := \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU}\} \tag{8.11}$$

and H is defined in (8.1) with Λ replaced by Λ_2.[43] By definition, $\Lambda_0 \subset \Lambda_2$. We use the parameter space Λ_2 in many places in the paper, rather than Λ_0, for two reasons. First, this makes it clear where the conditions specified in $\mathcal{F}_0$ (and Λ_0) are really needed. Second, some of the results given here are used in AG2, which does not employ the smaller set Λ_0, but does use Λ_2. By the definitions of Λ_2 and $\mathcal{F}_{WU}$, $\{\lambda_{n,h} : n \geq 1\}$ is a sequence of distributions that satisfies the null hypothesis $H_0 : \theta = \theta_0$.

We decompose h (defined by (8.1), (8.9), and (8.10)) analogously to the decomposition of the first eight components of λ: $h = (h_1, ..., h_8)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, ..., 8$. We further decompose the vector h_1 as $h_1 = (h_{1,1}, ..., h_{1,p})'$, where the elements of

[43] Analogously, for any subsequence $\{w_n : n \geq 1\}$, $\{\lambda_{w_n,h} : n \geq 1\}$ denotes a sequence $\{\lambda_{w_n} \in \Lambda_2 : n \geq 1\}$ for which $h_{w_n}(\lambda_{w_n}) \rightarrow h \in H$.

h_1 could equal ∞. We decompose h_6 as $h_6 = (h_{6,1}, ..., h_{6,p-1})'$. In addition, we let $h_{5,g}$ denote the upper left $k \times k$ submatrix of h_5. In consequence, under a sequence $\{\lambda_{n,h} : n \geq 1\}$, we have

$$\begin{aligned} n^{1/2}\tau_{jF_n} &\rightarrow h_{1,j} \geq 0 \; \forall j \leq p, \; \lambda_{m,F_n} \rightarrow h_m \; \forall m = 2, ..., 8, \\ \lambda_{5,gF_n} &= \Omega_{F_n} = E_{F_n} g_i g_i' \rightarrow h_{5,g}, \text{ and } \lambda_{6,jF_n} \rightarrow h_{6,j} \; \forall j = 1, ..., p-1. \end{aligned} \tag{8.12}$$

By the conditions in $\mathcal{F}$, defined in (3.1), $h_{5,g}$ is pd.

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

$$c_1 s_{jF} \leq \tau_{jF} \leq c_2 s_{jF} \text{ for } j = 1 \text{ and } j = p \text{ for some constants } 0 < c_1 < c_2 < \infty \tag{8.13}$$

that do not depend on F. As shown below, the parameter θ is strongly or semi-strongly identified under $\{\lambda_{n,h} : n \geq 1\}$ if $\lim n^{1/2}\tau_{pF_n} = \infty$. In consequence of (8.13), this holds iff $\lim n^{1/2} s_{pF_n} = \infty$. The parameters are weakly identified in the standard sense if $\lim n^{1/2}\tau_{jF_n} < \infty \; \forall j \leq p$ or, equivalently, if $\lim n^{1/2}\tau_{1F_n} < \infty$, which holds by (8.13) iff $\lim n^{1/2} s_{1F_n} < \infty$. The parameters are weakly identified in the non-standard sense if $\lim n^{1/2}\tau_{1F_n} = \infty$ and $\lim n^{1/2}\tau_{pF_n} < \infty$, which holds by (8.13) iff $\lim n^{1/2} s_{1F_n} = \infty$ and $\lim n^{1/2} s_{pF_n} < \infty$.

The proof of (8.13) is as follows. For notational simplicity, we drop the subscript F in some of the calculations. We have

$$\begin{aligned} & \lambda_{\min}(U'EG_i'W'WEG_iU) \\ = & \min_{\lambda:||\lambda||=1} (U\lambda/||U\lambda||)'EG_i'W'WEG_i(U\lambda/||U\lambda||) \cdot ||U\lambda||^2 \\ \leq & \min_{\lambda:||\lambda||=1} \lambda'EG_i'W'WEG_i\lambda \cdot \lambda_{\max}(U'U) \\ = & \min_{\lambda:||\lambda||=1} (EG_i\lambda/||EG_i\lambda||)'W'W(EG_i\lambda/||EG_i\lambda||) \cdot ||EG_i\lambda||^2 \cdot \lambda_{\max}(U'U) \\ \leq & \lambda_{\max}(W'W)\lambda_{\min}(EG_i'EG_i)\lambda_{\max}(U'U) \\ \leq & c_2^2 \lambda_{\min}(EG_i'EG_i), \text{ where} \\ c_2 := & \sup_{F \in \mathcal{F}_{WU}} [\lambda_{\max}(W_F'W_F)\lambda_{\max}(U_F'U_F)]^{1/2} < \infty \end{aligned} \tag{8.14}$$

and the last inequality holds by the conditions in $\mathcal{F}_{WU}$ (defined in (8.5)). Because the smallest eigenvalues of $U'EG_i'W'WEG_iU$ and $EG_i'EG_i$ equal the squares of the smallest singular values of WEG_iU and EG_i, respectively, (8.14) establishes the second inequality in (8.13) for $j = p$. Analogous calculations establish the lower bound in (8.14) for $j = p$ and the bounds for $j = 1$ by replacing min and $\leq$ by max and $\geq$, respectively, in the appropriate places and taking $c_1 :=$

$\inf_{F \in \mathcal{F}_{WU}} [\lambda_{\min}(W_F' W_F) \lambda_{\min}(U_F' U_F)]^{1/2} > 0.$

8.4 Assumption WU

We assume that the random weight matrices $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$ defined in (8.2) satisfy the following assumption that depends on a suitably chosen parameter space Λ_* ($\subset \Lambda_2$), such as Λ_2, Λ_0, or Λ_1.

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

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

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

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

In Assumption WU and elsewhere below, "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$" means "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ for any $h \in H$" and likewise with n in place of w_n. Note that, by definition, a sequence $\{\lambda_{w_n,h} : n \geq 1\}$ determines a sequence of distributions $\{F_{w_n} : n \geq 1\}$, see (8.9).

Assumption WU for the parameter space Λ_0 is verified in Comment (ii) to Theorem 10.1 given below for the CLR test with moment-variance weighting, which is considered in Section 6. It also holds for Kleibergen's LM test (for the same parameter space Λ_0) by the same argument (because $\widehat{W}_{2n}$, $\widehat{U}_{2n}$, $W_1(\cdot)$, and $U_1(\cdot)$ are the same for these two tests, see (8.4)).

8.5 Basic Results

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

$$\Phi_F^{a_i} := Var_F(a_i - (E_F a_\ell g_\ell')\Omega_F^{-1} g_i) \text{ and } \Phi_h^{a_i} := \lim \Phi_{F_{w_n}}^{a_i} \tag{8.15}$$

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

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

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

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

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

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

(ii) Suppose one eliminates the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}$ and one defines $\widehat{D}_n$ in (4.3) with $\widehat{\Omega}_n$ replaced by an eigenvalue-adjusted matrix, denoted by $\widehat{\Omega}_n^{\varepsilon}$, which is constructed to have its smallest eigenvalue greater than or equal to $\varepsilon > 0$ multiplied by its largest eigenvalue, see AG2 for the details of such a construction. In this case, the result of Lemma 8.2 still holds and all of the other asymptotic results following from Lemma 8.2 still hold, except the independence of $\overline{g}_h$ and $\overline{D}_h$. However, this independence is key because it is used in the conditioning argument that establishes the correct asymptotic size of all of the tests that are shown to have correct asymptotic size. Without it, these tests do not necessarily have correct asymptotic size. In consequence, we define $\widehat{D}_n$ in (4.3) using $\widehat{\Omega}_n$, not $\widehat{\Omega}_n^{\varepsilon}$.

The reason that independence does not necessarily hold when $\widehat{D}_n$ is defined using $\widehat{\Omega}_n^{\varepsilon}$, rather than $\widehat{\Omega}_n$, is that the covariance term $E_{F_n}[G_{ij} - E_{F_n}G_{ij} - (E_{F_n}G_{\ell j}g_\ell')(\Omega_{F_n}^{\varepsilon})^{-1}g_i]g_i'$ typically does not equal $0^{k\times k}$ when $\Omega_{F_n}^{\varepsilon} \neq \Omega_{F_n}$, whereas $E_{F_n}[G_{ij} - E_{F_n}G_{ij} - (E_{F_n}G_{\ell j}g_\ell')\Omega_{F_n}^{-1}g_i]g_i'$ necessarily equals $0^{k\times k}$, see the proof of Lemma 8.2 in Section 14 in the Supplementary Material for more details.

(iii) The proofs of Lemma 8.2 and other results in this section are given in Sections 14-16 in the Supplemental Appendix.

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

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq p, \tag{8.16}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, ..., p$ by (8.12) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (8.11). Such a q exists because $\{h_{1,j} : j \leq p\}$ are nonincreasing in j (since $\{\tau_{jF} : j \leq p\}$ are the ordered singular values of $W_F(E_F G_i)U_F$, as defined in (8.8)). As defined, q is the number of singular values of $W_{F_n}(E_{F_n}G_i)U_{F_n}$ that diverge to infinity when multiplied by $n^{1/2}$. Roughly speaking, q is the number of parameters, or one-to-one transformations of the parameters, that are strongly or semi-strongly identified.

The following quantities appear in Lemma 8.3 below, which gives the asymptotic distribution of $\widehat{D}_n$ after suitable rotations and rescaling, but without the recentering (by subtracting $E_{F_n} G_i$) that appears in Lemma 8.2. We partition h_2 and h_3 and define $\overline{\Delta}_h$ as follows:

$$\begin{aligned} h_2 &= (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}),\ h_{1,p-q}^{\diamond} := \begin{bmatrix} 0^{q \times (p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p) \times (p-q)} \end{bmatrix} \in R^{k \times (p-q)}, \\ \overline{\Delta}_h &= (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k \times p},\ \overline{\Delta}_{h,q} := h_{3,q},\ \overline{\Delta}_{h,p-q} := h_3 h_{1,p-q}^{\diamond} + h_{71} \overline{D}_h h_{81} h_{2,p-q}, \\ h_{71} &:= W_1(h_7),\ \text{and } h_{81} := U_1(h_8), \end{aligned} \tag{8.17}$$

where $h_{2,q} \in R^{p \times q}$, $h_{2,p-q} \in R^{p \times (p-q)}$, $h_{3,q} \in R^{k \times q}$, $h_{3,k-q} \in R^{k \times (k-q)}$, $\overline{\Delta}_{h,q} \in R^{k \times q}$, $\overline{\Delta}_{h,p-q} \in R^{k \times (p-q)}$, $h_{71} \in R^{k \times k}$, $h_{81} \in R^{p \times p}$, and $\overline{D}_h$ is defined in Lemma 8.2.[44] Note that when Assumption WU holds $h_{71} = \lim W_{F_n} = \lim W_1(W_{2F_n})$ and $h_{81} = \lim U_{F_n} = \lim U_1(U_{2F_n})$ under $\{\lambda_{n,h} : n \geq 1\}$.

The case where $q = p$ (i.e., $n^{1/2} \tau_{jF_n} \to \infty$ for all $j \leq p$) is the strong or semi-strong identification case. In this case, no $h_{2,p-q}$, $h_{1,p-q}^{\diamond}$, and $\overline{\Delta}_{h,p-q}$ matrices appear in (8.17), $\overline{\Delta}_h = h_{3,q} = h_{3,p}$, and $\overline{\Delta}_h$ is non-random. In consequence, the limit in distribution (or probability) of the normalized matrix $n^{1/2} W_{F_n} \widehat{D}_n U_{F_n} T_n$, where $T_n \in R^{p \times p}$ is defined below, is non-random, see Lemma 8.3 below. When $q < p$, identification is weak and the limit of this matrix is random.

Now we provide some motivation for Lemma 8.3, which is stated below. To show that the LM statistic has a χ_p^2 asymptotic distribution we need to determine the asymptotic behavior of $\widehat{D}_n$ without the recentering by $E_{F_n} G_i$ that occurs in Lemma 8.2. In addition, to determine the asymptotic distribution of the rk_n statistic in (6.2), we need to determine the asymptotic distribution of $W_{F_n} \widehat{D}_n U_{F_n}$ without recentering by $E_{F_n} G_i$.[45] To do so, we post-multiply $W_{F_n} \widehat{D}_n U_{F_n}$ first by B_{F_n} and then by a nonrandom diagonal matrix $S_n \in R^{p \times p}$ (which may depend on F_n and h). The matrix S_n rescales the columns of $W_{F_n} \widehat{D}_n U_{F_n} B_{F_n}$ to ensure that $n^{1/2} W_{F_n} \widehat{D}_n U_{F_n} B_{F_n} S_n$ converges in distribution to a (possibly) random matrix that is finite a.s. and not almost surely zero. For $F \in \mathcal{F}_{WU} \cap \mathcal{F}_0$, it ensures that the (possibly) random limit matrix has full column rank with probability one. For example, in the case of the LM statistic, these transformations are applied with $W_{F_n} = \Omega_{F_n}^{-1/2}$ and $U_{F_n} = I_p$.

For the LM statistic and the CLR statistics that employ it, we need the full column rank property of the limit random matrix in order to apply the continuous mapping theorem (CMT). For the LM statistic, the full rank property ensures that the quantity $\widehat{D}_n' \widehat{\Omega}_n^{-1} \widehat{D}_n$ (whose inverse

[44] For simplicity, there is some abuse of notation here, e.g., $h_{2,q}$ and $h_{2,p-q}$ denote different matrices even if $p - q$ happens to equal q.

[45] Furthermore, to determine the asymptotic distributions of the two SR-CQLR test statistics and conditional critical values considered in AG2, we need to determine the asymptotic distribution of $W_{F_n} \widehat{D}_n U_{F_n}$ without recentering by $E_{F_n} G_i$.

appears in the expression for LM_n, see (4.3)), is nonsingular asymptotically with probability one after $\widehat{D}_n$ has been transformed and rescaled to yield $n^{1/2}\Omega_{F_n}^{-1/2}\widehat{D}_n B_{F_n} S_n$. Note that $P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}$, which appears in the definition of LM_n in (4.3), can be written as

$$\begin{aligned} P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} &:= \widehat{\Omega}_n^{-1/2}\widehat{D}_n(\widehat{D}_n'\widehat{\Omega}_n^{-1}\widehat{D}_n)^{-1}\widehat{D}_n'\widehat{\Omega}_n^{-1/2} \\ &= (\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})(n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n)\Big[(n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n)'(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2})'(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2}) \\ &\quad \times (n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n)\Big]^{-1}(n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n)'(\Omega_n^{1/2}\widehat{\Omega}_n^{-1/2}), \text{ where} \\ T_n &:= B_{F_n} S_n \in R^{p\times p} \text{ and } \Omega_n := \Omega_{F_n} \ (= E_{F_n} g_i g_i'), \end{aligned} \tag{8.18}$$

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

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

Define

$$S_n := Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}, 1, ..., 1\} \in R^{p\times p}, \tag{8.19}$$

where $q = q_h$ is defined in (8.16).[46]

The proof of Theorem 9.1 for the LM test, the proofs of Theorems 8.4 and 10.1 for the CLR test with moment-variance weighting, and the proofs for the two SR-CQLR tests in AG2 use the following lemma. The $p \times p$ matrix T_n is defined in (8.18).

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

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

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

Comments: (i) Lemma 8.3(c)-(d) are key properties of the asymptotic distribution of $n^{1/2}(\widehat{g}_n,$

[46] Note that $\tau_{jF_n} > 0$ for n large for $j \leq q$ and, hence, S_n is well defined for n large, because $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$.

$W_{F_n}\widehat{D}_n U_{F_n} T_n$) that lead to the LM statistic having a χ^2_p asymptotic distribution and the CLR test with moment-variance weighting having correct asymptotic size. Lemma 8.3(c) is a key property that leads to the correct asymptotic size of the two SR-CQLR tests in AG2. Lemma 8.3(d) is not needed for these tests because they do not rely on an LM statistic.

(ii) The conditions in $\mathcal{F}_0$ are used in the proofs to obtain the result of Lemma 8.3(d) and are not used elsewhere in the proofs, except where Lemma 8.3(d) is used.

The following theorems are used only for the CLR tests. For the proof of Theorem 4.1 concerning Kleibergen's (2005) LM test, one can go from here to Section 9.

Let

$$\widehat{\kappa}_{jn} \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n,\ \forall j = 1, ..., p, \tag{8.20}$$

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

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

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

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

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

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

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

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

Comments: (i) The statistic $\widehat{\kappa}_{pn} = \lambda_{\min}(n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n)$ in Theorem 8.4(a) and (b) is a Robin and Smith (2000)-type rank statistic.

(ii) Theorem 8.4(a) and (b) is used to determine the asymptotic behavior of the statistic rk_n defined in (6.2) (which is employed by the CLR test with moment-variance weighting that is considered in Section 6). More specifically, Theorem 8.4(a) and (b) is used to verify Assumption R in Section 10 below.

(iii) Theorem 8.4(c) and (d) is used to determine the asymptotic behavior of the critical value functions for the two SR-CQLR tests considered in AG2 (with $\widehat{W}_n$ and $\widehat{U}_n$ defined suitably). Because Theorem 8.4(c) and (d) are immediate by-products of the proofs of Theorem 8.4(a) and (b), they are stated and proved here, rather than in AG2.

(iv) The statement of Theorem 3 in Kleibergen (2005) is difficult to interpret because the expression given for the conditional asymptotic distribution of the CLR statistic involves Kleibergen's (2005) statistic $\mathrm{rk}(\theta_0)$, which is a finite-sample object. Based on Theorem 8.4, (10.7) below provides the asymptotic distribution of a class of CLR statistics in terms of an asymptotic version of the rank statistic employed, which is necessary for a precise statement of the asymptotic distribution. The class of CLR statistics considered are those defined in (5.1) and based on the rank statistic in Theorem 8.4 for some choices of $\widehat{W}_n$ and $\widehat{U}_n$, which is a Robin and Smith (2000)-type rank statistic. In particular, taking $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$ gives the rank statistic defined in (6.2).

9 Asymptotic Size of the Nonlinear LM Test

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

We state a theorem that verifies Assumption B* of ACG (stated in Section 8) for the LM test. The following theorem applies with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $W_F = \Omega_F^{-1/2}$, and $\widehat{U}_n = U_F = I_p$. (These definitions affect the definition of $\lambda_{n,h}$, which appears in the theorem).

Theorem 9.1 *The asymptotic null rejection probabilities of the nominal size $\alpha \in (0,1)$ LM test equal α under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$.*

Comments: (i) The requirement that $\lambda_{w_n,h} \in \Lambda_0$ (defined in (8.10)) implies that the parameter space for F is $\mathcal{F}_0$ (defined in (3.7)) for the results given in Theorems 4.1 and 9.1 (because the restrictions in $\mathcal{F}_{WU}$ are not binding, see the discussion in the paragraph containing (8.5)).

(ii) Proposition 8.1 and Theorem 9.1 prove Theorem 4.1 for the LM test. The proof of Theorem 4.1 for the LM CS is analogous, see Comments (i) and (ii) to Proposition 8.1.

For notational simplicity, we prove Theorem 9.1 for the sequence $\{n\}$, rather than a subsequence $\{w_n : n \geq 1\}$. We note here that the same proof holds for any subsequence $\{w_n : n \geq 1\}$.

Proof of Theorem 9.1. Let $\Omega_n := \Omega_{F_n}$. We derive the limiting distribution of the statistic LM_n using the CMT applied to $\Omega_n^{-1/2} n^{1/2} \widehat{g}_n$, $\widehat{\Omega}_n^{-1/2} \Omega_n^{1/2}$, and $n^{1/2} \Omega_n^{-1/2} \widehat{D}_n T_n$, where the latter two quantities appear in the expression on the rhs of (8.18). Note that $\widehat{\Omega}_n \to_p h_{5,g}$ by the WLLN, $\Omega_n \to h_{5,g}$, and $h_{5,g}$ is pd. Thus, $\widehat{\Omega}_n^{-1/2} \Omega_n^{1/2} \to_p I_k$. By Lemma 8.3 applied with $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$ (which results from taking $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$), we get $(\Omega_n^{-1/2} n^{1/2} \widehat{g}_n, n^{1/2} \Omega_n^{-1/2} \widehat{D}_n T_n) \to_d (h_{5,g}^{-1/2} \overline{g}_h, \overline{\Delta}_h)$. For the CMT to apply, it is enough to show that the function $f : R^{k \times p} \to R^{k \times k}$ defined by $f(D) := D(D'D)^{-1} D'$ for $D \in R^{k \times p}$ is continuous on a set $C \subset R^{k \times p}$ with $P(\overline{\Delta}_h \in$

$C) = 1$.[47] Note that f is continuous at each D that has full column rank. And, by Lemma 8.3(d), $\overline{\Delta}_h$ has full column rank a.s. because $\lambda_{n,h} \in \Lambda_0$, $F_n \in \mathcal{F}_0$, $W_F = \Omega_F^{-1/2}$, and $U_F = I_p$. Hence, f is continuous a.s. By $\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2} \to_p I_k$, the convergence result in Lemma 8.3, and the CMT, we have

$$P_{D_n^\diamond}\widehat{\Omega}_n^{-1/2}n^{1/2}\widehat{g}_n = D_n^\diamond(D_n^{\diamond\prime}D_n^\diamond)^{-1}D_n^{\diamond\prime}\widehat{\Omega}_n^{-1/2}n^{1/2}\widehat{g}_n \to_d \overline{v}_h := P_{\overline{\Delta}_h}h_{5,g}^{-1/2}\overline{g}_h, \tag{9.1}$$

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

Conditional on $\overline{\Delta}_h$, $\overline{v}_h'\overline{v}_h$ is distributed as χ_p^2 because (i) $\overline{\Delta}_h$ and $\overline{g}_h$ are independent by property (c) in Lemma 8.3, (ii) $h_{5,g}^{-1/2}\overline{g}_h$ is conditionally distributed as $N(0^k, I_k)$ by $\overline{g}_h \sim N(0^k, h_{5,g})$ and (i), and (iii) $P_{\overline{\Delta}_h}$ is fixed given $\overline{\Delta}_h$ and projects onto a space of dimension p a.s. by property (d) in Lemma 8.3. Because the χ_p^2 distribution does not depend on $\overline{\Delta}_h$, $\overline{v}_h'\overline{v}_h$ is unconditionally distributed as χ_p^2 as well. In consequence, using the CMT again, we have

$$LM_n \to_d \overline{LM}_h := \overline{v}_h'\overline{v}_h \sim \chi_p^2. \tag{9.2}$$

Given this result and the use of the $\chi_{p,1-\alpha}^2$ critical value by the LM test, we obtain the conclusion of Theorem 9.1 for the LM test: $\lim P_{F_n}(LM_n > \chi_{p,1-\alpha}^2) = \alpha$. $\square$

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

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

Section 18.5 in the Supplemental Material provides additional asymptotic size results for equally-weighted CLR tests (and CS's), which are CLR tests that are based on rk_n statistics that depend on $\widehat{D}_n$ only through $\widetilde{W}_n\widehat{D}_n$ for some $k \times k$ weighting matrix $\widetilde{W}_n$. These results show that equally-weighted CLR tests (and CS's) based on the Robin and Smith (2000) rank statistic with a general weight matrix $\widetilde{W}_n$ $(\in R^{k\times k})$ have correct asymptotic size under suitable conditions on $\widetilde{W}_n$. One can

[47] This holds because the function $f_2(D, L) := LD((LD)'(LD))^{-1}D'L'$ for a nonsingular $k\times k$ matrix L is continuous at (D, I_k) if $f(D)$ is continuous at D.

view these results as verifying Assumption R for a broad class of rk_n statistics. In contrast, the results in the present section establish the correct asymptotic size of CLR tests (and CS's) under the high-level condition Assumption R and for the Robin and Smith (2000) rank statistic when $\widetilde{W}_n$ is the moment-variance weighting matrix $\widehat{\Omega}_n^{-1/2}$, see Comment (ii) to Theorem 10.1 below.

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

Assumption R: For any subsequence $\{w_n\}$ and any sequence $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$ either (a) $rk_{w_n} \rightarrow_p r_h = \infty$ or (b) $rk_{w_n} \rightarrow_d r_h(\overline{D}_h)$ for some nonrandom function $r_h : R^{k\times p} \rightarrow R$, where $\overline{D}_h$ is defined in Lemma 8.2, and the convergence is joint with that in Lemma 8.2.[48]

The following theorem applies when the LM statistic is defined as in (4.3) with projection onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$. In consequence, the quantities in (8.2) in the present case are $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $W_F = \Omega_F^{-1/2}$, and $\widehat{U}_n = U_F = I_p$. (These definitions affect the definition of $\lambda_{n,h}$, which appears in the theorem).

Theorem 10.1 *For any statistic rk_n that satisfies Assumption* R, *the asymptotic null rejection probabilities of the nominal size $\alpha \in (0,1)$* CLR *test defined in* (4.3)-(5.2) *based on rk_n equal α under all subsequences $\{w_n\}$ and all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$.*

Comments: (i) Theorem 10.1 and Proposition 8.1 imply that a nominal size α CLR test based on any rank statistic that satisfies Assumption R has asymptotic size α and is asymptotically similar. Analogous CS results (to the test results stated in Theorem 10.1) hold for a parameter space $\Lambda_{\Theta,0}$ that is a reparametrization of $\mathcal{F}_{\Theta,0}$ and is defined as Λ_0 is defined, but with the adjustments outlined in Comments (i) and (ii) to Proposition 8.1.

(ii) Theorems 8.4 and 10.1 and Proposition 8.1 establish the test results of Theorem 6.1. This holds because Theorem 8.4(a), (b), (e), and (f) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$ imply that Assumption R holds for the CLR test with moment-variance weighting, that is considered in Section 6, which uses the Robin and Smith (2000) rk_n statistic defined in (6.2). (In the present context, Theorem 8.4 requires that Assumption WU holds for the parameter space Λ_0. It holds with $\widehat{W}_n = \widehat{W}_{2n}$, $W_1(w) = w$ for $w \in R^{k\times k}$, $\mathcal{W}_2 = R^{k\times k}$, $\widehat{U}_n = \widehat{U}_{2n}$, $U_1(u) = u$ for $u \in R^{p\times p}$, and $\mathcal{U}_2 = R^{p\times p}$, because $\widehat{W}_n = \widehat{\Omega}_n^{-1/2} \rightarrow_p h_{5,g}^{-1/2}$ under all sequences $\{\lambda_{n,h} : n \geq 1\}$ with $\lambda_{n,h} \in \Lambda_0$ and $\widehat{U}_n = I_p$ for all $n \geq 1$.) In particular, Assumption R holds with $r_h = \infty$ if $q = p$ and with $r_h(\overline{D}_h)$ equal to the smallest eigenvalue of $\overline{\Delta}'_{h,p-q}h_{3,k-q}h'_{3,k-q}\overline{\Delta}_{h,p-q}$ if $q < p$ (where $\overline{\Delta}_{h,p-q}$ and $h_{3,k-q}$ are defined in (8.17) based on $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$). The CS results of Theorem 6.1 hold by Theorem 8.4, Comment (i) to Theorem 10.1, and Comment (i) to Proposition 8.1.

[48]By $rk_{\omega_n} \rightarrow_p \infty$, we mean that for every $K < \infty$ we have $P_{\theta_0,\lambda_{\omega_n}}(rk_{\omega_n} > K) \rightarrow 1$, where $P_{\theta_0,\lambda_{\omega_n}}(\cdot)$ denotes probability under λ_{ω_n} when the true parameter vector equals θ_0.

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

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

Proof of Theorem 10.1. Let

$$J_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}M_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n. \tag{10.1}$$

It follows from (4.3) that

$$AR_n = LM_n + J_n. \tag{10.2}$$

We now distinguish two cases. First, suppose Assumption R(a) holds: $rk_n \to_p \infty$. By (10.2) and some algebra, we have $(AR_n - rk_n)^2 + 4LM_n \cdot rk_n = (LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n$. Therefore,

$$CLR_n = \frac{1}{2}\left(LM_n + J_n - rk_n + \sqrt{(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n}\right). \tag{10.3}$$

Using a mean-value expansion of the square-root expression in (10.3) about $(LM_n - J_n + rk_n)^2$, we have

$$\sqrt{(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n} = LM_n - J_n + rk_n + (2\sqrt{\zeta_n})^{-1}4LM_n \cdot J_n \tag{10.4}$$

for an intermediate value ζ_n between $(LM_n - J_n + rk_n)^2$ and $(LM_n - J_n + rk_n)^2 + 4LM_n \cdot J_n$. It follows that $CLR_n = LM_n + o_p(1) \to_d \chi_p^2$ using (9.2) and $(\sqrt{\zeta_n})^{-1} = o_p(1)$ (which holds because $rk_n \to_p \infty$, $LM_n = O_p(1)$, and $J_n = O_p(1)$ by (10.6) below). Analogously, it can be shown that the critical value $c(1-\alpha, rk_n)$, defined above (5.2), of the CLR test converges in probability to $\chi^2_{p,1-\alpha}$. The result of Theorem 10.1 then follows by the definition of convergence in distribution.

Second, suppose Assumption R(b) holds. Then, using Lemma 8.2, we have $(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i), rk_n) \to_d (\overline{g}_h, \overline{D}_h, r_h(\overline{D}_h))$. By the proof of Lemma 8.3 applied with $W_F = \Omega_F^{-1/2}$ and $U_F = I_p$ (which correspond to $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = I_p$), using the former result in place of $(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i)) \to_d (\overline{g}_h, \overline{D}_h)$ gives

$$(n^{1/2}\widehat{g}_n, n^{1/2}(\widehat{D}_n - E_{F_n}G_i), n^{1/2}\Omega_n^{-1/2}\widehat{D}_nT_n, rk_n) \to_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h, r_h(\overline{D}_h)), \tag{10.5}$$

where $\Omega_n := \Omega_{F_n}$, $(\overline{D}_h, \overline{\Delta}_h)$ and $\overline{g}_h$ are independent, and $\overline{\Delta}_h$ has full column rank p with probability one by Lemma 8.3(d) (because we are considering sequences $\{\lambda_{w_n,h} : n \geq 1\}$ with $\lambda_{w_n,h} \in \Lambda_0$ $\forall n \geq 1$, $W_F = \Omega_F^{-1/2}$, and $U_F = I_p$). In addition, $\widehat{\Omega}_n \rightarrow_p h_{5,g}$, $h_{5,g}$ is pd, and $M_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n} = M_{\left(\widehat{\Omega}_n^{-1/2}\Omega_n^{1/2}\right)n^{1/2}\Omega_n^{-1/2}\widehat{D}_n T_n}$ because T_n (defined in (8.18)) and $\Omega_n^{-1/2}$ are nonsingular. These results and the CMT imply that

$$J_n \rightarrow_d \overline{J}_h := \overline{g}_h' h_{5,g}^{-1/2} M_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h. \tag{10.6}$$

The convergence results in (9.2) and (10.6) and $rk_n \rightarrow_d r_h(\overline{D}_h)$ hold jointly by (10.5) and the definitions of LM_n and J_n in (4.3) and (10.1).

Note that $\overline{LM}_h = \overline{g}_h' h_{5,g}^{-1/2} P_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ by (9.1) and (9.2). Conditional on $\overline{\Delta}_h$, $P_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ and $M_{\overline{\Delta}_h} h_{5,g}^{-1/2} \overline{g}_h$ have a joint normal distribution with zero covariance (because $Var(h_{5,g}^{-1/2} \overline{g}_h) = I_k$ and $P_{\overline{\Delta}_h} M_{\overline{\Delta}_h} = 0^{k \times k}$) and, hence, are independent. The same holds true conditional on $\overline{D}_h$, because $\overline{\Delta}_h$ is a nonrandom function of $\overline{D}_h$ and $\overline{D}_h$ is independent of $\overline{g}_h$. In consequence, conditional on $\overline{D}_h$, $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2, respectively.

Using the convergence results in (10.5) and (10.6), the definition of CLR_n in (5.1) with $AR_n = LM_n + J_n$ substituted in, and the CMT, we obtain

$$CLR_n \rightarrow_d \overline{CLR}_h := \frac{1}{2}\left(\overline{LM}_h + \overline{J}_h - \overline{r}_h + \sqrt{(\overline{LM}_h + \overline{J}_h - \overline{r}_h)^2 + 4\overline{LM}\overline{r}_h}\right), \tag{10.7}$$

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

The function $c(1-\alpha, r)$ (defined in (5.2)) is continuous in r on R_+ by the absolute continuity of the distributions of χ_p^2 and χ_{k-p}^2, which appear in $clr(r)$ (also defined in (5.2)), and the continuity of $clr(r)$ in r a.s. This, $rk_n \rightarrow_d \overline{r}_h$, and (10.7) yield

$$CLR_n - c(1-\alpha, rk_n) \rightarrow_d \overline{CLR}_h - c(1-\alpha, \overline{r}_h). \tag{10.8}$$

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

$$P_{\theta_0,\lambda_n}(CLR_n > c(1-\alpha, rk_n)) \rightarrow P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)) \tag{10.9}$$

provided $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)) = 0$, which holds because $P(\overline{CLR}_h = c(1-\alpha, \overline{r}_h)|\overline{D}_h) = 0$ a.s. The latter holds because conditional on $\overline{D}_h$, $\overline{CLR}_h$ is absolutely continuous (by (10.7) since $\overline{LM}_h$ and $\overline{J}_h$ are independent and distributed as χ_p^2 and χ_{k-p}^2 and $\overline{r}_h$ is a nonrandom function of $\overline{D}_h$) and $c(1-\alpha, \overline{r}_h)$ is a constant.

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

$$P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h)|\overline{D}_h) = \alpha \text{ a.s.} \tag{10.10}$$

Because the left-hand side conditional probability equals α a.s. and α does not depend on $\overline{D}_h$, the unconditional probability $P(\overline{CLR}_h > c(1-\alpha, \overline{r}_h))$ equals α as well. Combined with (10.9), this gives the desired result. $\square$

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

In this section, we prove the test results of Theorem 5.2, which concerns Kleibergen's CLR test (and CS) with Jacobian-variance weighting when $p = 1$. The CS results of Theorem 5.2 hold by an analogous argument, see Comments (i) and (ii) to Proposition 8.1.

Proof of Theorem 5.2. We prove the test results of Theorem 5.2 using Proposition 8.1 and results (or variants of results) in Lemma 8.3 and Theorems 8.4, 9.1, and 10.1. The proof is made more complicated by the fact that we need to use two different definitions of $\widehat{W}_n$. To obtain the asymptotic distribution of the LM statistic (which is a component of the CLR statistic), we need to take $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = 1$, because the LM statistic (defined in (4.3)) depends on $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$. But, to obtain the asymptotic distribution of the rank statistic $rk_n := n\widehat{D}_n{}'\widetilde{V}_{Dn}^{-1}\widehat{D}_n$ (defined in (5.8)), we need to take $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$ and $\widehat{U}_n = 1$, because rk_n depends on $\widetilde{V}_{Dn}^{-1/2}\widehat{D}_n$.

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

We proceed as follows. First, we apply Lemma 8.3 exactly as in the proof of Theorem 9.1 with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$, $\widehat{U}_n = 1$, $W_F = \Omega_F^{-1/2}$, and $U_F = 1$. This yields $n^{1/2}(\widehat{g}_n, \widehat{D}_n - E_{F_n}G_i, W_{F_n}\widehat{D}_nU_{F_n}T_n) \to_d (\overline{g}_h, \overline{D}_h, \overline{\Delta}_h)$ for sequences $\{\lambda_{n,h} : n \geq 1\}$ that correspond to distributions F in $\mathcal{F}_{WU} \cap \mathcal{F}_0$ based on these definitions of W_F and U_F. As discussed in the paragraph containing (8.5), $\mathcal{F}_0 = \mathcal{F}_{WU} \cap \mathcal{F}_0$ for δ_{WU} sufficiently small and M_{WU} sufficiently large. We employ constants δ_{WU} and M_{WU} for which this holds. The joint convergence result above yields the asymptotic distributions of the AR_n, LM_n, and J_n statistics via the calculations in (9.1), (9.2), (10.1), (10.2), and (10.6).

Next, we take $\widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$, $\widehat{U}_n = 1$, $W_F = W_{2F} = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$, where $\Gamma_F^{G_i}$ and Ω_F are defined in (3.2), $W_1(\cdot)$ equals the identity function on $\mathcal{W}_2 := R^{k\times k}$, $U_F = U_{2F} = 1$, and $U_1(\cdot)$ equals the identity function on $\mathcal{U}_2 := R$. We consider distributions in $\mathcal{F}_{JVW,p=1}$ (which is a subset of $\mathcal{F}_0$ when $\delta_3 = \delta_2$ by the paragraph following (5.7)). We obtain the asymptotic distribution of rk_n

under the corresponding sequences $\{\lambda_{n,h} : n \geq 1\}$ (which differ from the sequences $\{\lambda_{n,h} : n \geq 1\}$ in the previous paragraph due to the difference between the two definitions of W_F). More specifically, we verify the convergence results in Assumption R for $rk_n := n\widehat{D}_n'\widetilde{V}_{Dn}^{-1}\widehat{D}_n$ (defined in (5.8)) for the $\{\lambda_{n,h} : n \geq 1\}$ sequences of this paragraph. The result of Theorem 8.4(a), (b), (e), and (f) verifies the convergence results in Assumption R for sequences $\{\lambda_{n,h} : n \geq 1\}$ for which $F_n \in \mathcal{F}_{JVW,p=1}$ $\forall n \geq 1$ provided Assumption WU holds for such sequences with $\widehat{W}_{2n} = \widehat{W}_n = \widetilde{V}_{Dn}^{-1/2}$, $W_1(\cdot)$ equal to the identity function, $\widehat{U}_{2n} = \widehat{U}_n = 1$, $U_1(\cdot)$ equal to the identity function, and the parameter space Λ_* being equal to $\Lambda_{JVW,p=1} := \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F})$ for some $F \in \mathcal{F}_{WU} \cap \mathcal{F}_{JVW,p=1}\}$. Here $\mathcal{F}_{WU}$ is defined in (8.5) with $W_F = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$ and $U_F = 1$. Note that $\mathcal{F}_{JVW,p=1} = \mathcal{F}_{WU} \cap \mathcal{F}_{JVW,p=1}$ for $\delta_{WU} > 0$ sufficiently small and $M_{WU} < \infty$ sufficiently large (and we employ constants δ_{WU} and M_{WU} that satisfy these conditions). This holds because for all $F \in \mathcal{F}_{JVW,p=1}$, $\lambda_{\min}(W_F) = \lambda_{\min}((Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}) = \lambda_{\max}^{-1/2}(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})$ $\geq \lambda_{\max}^{-1/2}(E_F G_i G_i') \geq M_+^{-1/2}$ for some $M_+ < \infty$ (because $E_F G_i G_i' - (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}) = E_F G_i E_F G_i' + \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}$ is psd and $||E_F G_i G_i'|| \leq M_+$ for some $M_+ < \infty$ by the moment conditions in $\mathcal{F}$), $||W_F|| = ||(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}|| \leq \lambda_{\min}^{-1/2}(Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime}) \leq \delta_3^{-1/2}$ (using the condition in $\mathcal{F}_{JVW,p=1}$ and the fact that $Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime} = \Psi_F^{G_i} - E_F G_i E_F G_i'$ using the definition of $\Psi_F^{G_i}$ in (3.2)), where $\delta_3 > 0$, and $||U_F|| = \lambda_{\min}(U_F) = 1$.

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

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

$$\begin{aligned}
\widetilde{V}_{Dn} &:= n^{-1}\sum_{i=1}^{n}(G_i - \widehat{G}_n)(G_i - \widehat{G}_n)' - \widehat{\Gamma}_n\widehat{\Omega}_n^{-1}\widehat{\Gamma}_n' \\
&= E_{F_n}(G_i - E_{F_n}G_i)(G_i - E_{F_n}G_i)' - \Gamma_{F_n}^{G_i}\Omega_{F_n}^{-1}\Gamma_{F_n}^{G_i\prime} + o_p(1) \\
&= W_{2F_n}^{-2} + o_p(1) \\
&\rightarrow_p h_7^{-2}, \qquad (11.1)
\end{aligned}$$

where the first equality holds by (5.3), the second equality holds by the WLLN's applied multiple times and Slutsky's Theorem using the conditions in $\mathcal{F}$, the third equality holds by the definition of W_{2F}, and the convergence holds because $W_{2F_n} = \lambda_{7,F_n} \rightarrow h_7$ by the definition of the sequence $\{\lambda_{n,h} : n \geq 1\}$ and h_7 is pd (since $h_7 = \lim W_{2F_n}$ and the eigenvalues of W_{2F}^{-2} are bounded above for $F \in \mathcal{F}$). Equation (11.1) and Slutsky's Theorem give $\widetilde{V}_{Dn}^{-1/2} \rightarrow_p h_7$ because h_7^{-2} is pd using the condition in $\mathcal{F}_{JVW,p=1}$ that $\lambda_{\min}(\Psi_F^{G_i} - E_F G_i E_F G_i') \geq \delta$. In consequence, Assumption WU(a) holds.

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

Now we consider sequences $\{\lambda_{n,h} : n \geq 1\}$ that satisfy the conditions on $\{\lambda_{n,h} : n \geq 1\}$ given in both the third and fourth paragraphs of this proof. These sequences correspond to distributions F in $\mathcal{F}_{JVW,p=1}$. These sequences satisfy the convergence conditions in (8.11) using the definitions in (8.9) and (8.10) with τ_{jF}, B_F, C_F, and W_{2F} defined based on $W_F = \Omega_F^{-1/2}$ and with these quantities based on $W_F = (Var_F(G_i) - \Gamma_F^{G_i}\Omega_F^{-1}\Gamma_F^{G_i\prime})^{-1/2}$. In consequence, for these sequences of distributions $\{\lambda_{n,h} : n \geq 1\}$, the results above establish the asymptotic distributions of the AR_n, LM_n, J_n, and rk_n statistics and the convergence is joint because all of the convergence results are based on the underlying CLT result in Lemma 8.2. Given this joint convergence, by the same arguments as given in the proof of Theorem 10.1, we obtain that the CLR test with Jacobian-variance weighting has asymptotic null rejection probabilities equal to α under all such sequences $\{\lambda_{n,h} : n \geq 1\}$ (and all subsequences of such sequences).

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

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

In this section, we show that the restriction $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_{0j}$, defined in (3.7), is not redundant. If this restriction is weakened to $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$, we show that, for some models, some sequences of distributions, and some (consistent) choices of variance and covariance estimators, the LM statistic in (4.3) has a χ^2_k asymptotic distribution. This leads to over-rejection of the null when the standard χ^2_p critical value is used and the parameters are over-identified (i.e., $k > p$). On the other hand, we show that the LM statistic equals zero a.s. for some models and some distributions F if the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ is removed entirely. This implies that the LM test also under-rejects the null hypothesis and is nonsimilar in both finite samples and asymptotically for some F.

All of the CLR tests considered in Sections 5 and 6, except that of Smith (2007), are functions of the LM statistic in (4.3) (and other statistics). In consequence, the aberrant behavior of the LM statistic and test demonstrated in this section, when the restriction $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_0$ is weakened or eliminated, carries over to the CLR statistics and tests in Sections 5 and 6.[49]

12.1 Eigenvalue Condition Counter-Examples

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

Lemma 12.1 (a) *Suppose $\mathcal{F}_0$ is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$ in place of $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ in $\mathcal{F}_{0j}$ for all $j \in \{0, ..., p\}$, where $p = 1$. Suppose $\widehat{\Omega}_n(\theta)$ is defined in* (4.1) *and $\widehat{\Gamma}_{1n}(\theta) = n^{-1}\sum_{i=1}^n G_i(\theta)g_i(\theta)'$ (which differs from its definition in* (4.3)*). Then, there exist moment functions $g(W_i, \theta)$ and a sequence of null distributions $\{F_n \in \mathcal{F}_0 : n \geq 1\}$ for which $\widehat{\Omega}_n = \widehat{\Omega}_n(\theta_0)$ and $\widehat{\Gamma}_{1n} = \widehat{\Gamma}_{1n}(\theta_0)$ are well-behaved (in the sense that $\widehat{\Omega}_n - E_{F_n} g_i g_i' \to_p 0^{k\times k}$ and $\widehat{\Gamma}_{1n} - E_{F_n} G_i g_i' \to_p 0^{k\times k}$) and $LM_n(\theta_0) = AR_n(\theta_0) + o_p(1) \to_d \chi^2_k$.*

(b) *Suppose $\mathcal{F}_0$ is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ deleted in $\mathcal{F}_{0j}$ for all $j \in \{0, ..., p\}$, where $p = 1$. Suppose $\widehat{\Omega}_n(\theta)$ and $\widehat{\Gamma}_{1n}(\theta)$ are defined in* (4.1) *and* (4.3), *respectively. Then, there exists moment functions and a null distribution $F \in \mathcal{F}_0$ for which $LM_n(\theta_0) = 0$ a.s. for all $n \geq 1$.*

Comments: (i) The model we use to prove Lemma 12.1(a) is the linear IV regression model with one endogenous rhs variable and (for simplicity) no exogenous variables. Specifically, the model is

$$y_{1i} = y_{2i}\theta + u_i \text{ and } y_{2i} = Z_i'\pi + v_{2i}, \tag{12.1}$$

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

$$\begin{aligned}\mathcal{F}^* := \{F : &(12.1) \text{ holds with } \theta = \theta_0,\ \pi = \pi_F \in R^k,\ \rho = \rho_F \in (-1, 1),\\ &Z_i,\ u_i, \text{ and } \xi_i \text{ are mutually independent, } E_F u_i = E_F \xi_i = 0,\\ &E_F u_i^2 = E_F \xi_i^2 = 1,\ E_F||(u_i, \xi_i, Z_i'Z_i)||^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(E_F Z_i Z_i') \geq \delta\}\end{aligned} \tag{12.2}$$

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

[49] Smith's (2007) CLR test is a function of the LM statistic in (4.3) but with $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ replaced by $\widehat{D}_n^{\dagger}$.

The moment functions are $g(W_i, \theta) = Z_i(y_{1i} - y_{2i}\theta)$. When the null value θ_0 is the true value, this gives $g_i = g_i(\theta_0) = Z_i u_i$ and $G_i = G_i(\theta_0) = -Z_i y_{2i}$. The set $\mathcal{F}^*$ is a subset of $\mathcal{F}_0$ when the latter is defined with the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) > 0$ in place of $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$. This holds because (i) for all $F \in \mathcal{F}^*$, $\lambda_{\min}(\Psi_F^{vec(G_i)}) > 0$ (by the argument in the paragraph that contains (3.11) because $\lambda_{\min}(E_F Z_i Z_i') > 0$ and $\lambda_{\min}(E_F \varepsilon_i \varepsilon_i') > 0$, where $\varepsilon_i = (u_i, -\rho u_i - \delta \xi_i)'$ for $\rho \in (-1, 1)$), (ii) $\lambda_{\min}(E_F g_i g_i') = E_F u_i^2 \lambda_{\min}(E_F Z_i Z_i') \geq \delta > 0$, and (iii) $\lambda_{p-j}(\Psi_F^{C_{F,k-j}' \Omega_F^{-1/2} G_i B_{F,p-j}\xi}) \geq \lambda_{\min}(\Psi_F^{vec(G_i)}) M^{-2/(2+\gamma)}$ for all $\xi \in R^{p-j}$ with $||\xi|| = 1$ and all $j \in \{0, ..., p\}$ (by the results and arguments in the paragraphs that contain (17.1)-(17.3), which verify that condition (iv), stated in (3.9), is a sufficient condition for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$). The quantity $\lambda_{\min}(\Psi_F^{vec(G_i)})$ is arbitrarily close to zero for ρ arbitrarily close to one.

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

$$G_i = -\rho_n g_i + \delta_n G_i^*, \text{ where } G_i^* := -Z_i \xi_i \text{ and } \delta_n := (1 - \rho_n^2)^{1/2}. \tag{12.3}$$

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

(ii) The model we consider in Lemma 12.1(b) is the same as that in part (a) except that $\mathcal{F}^*$ allows for $\rho = \rho_F \in (-1, 1]$ and we consider a single distribution F with $\pi = 0^k$ and $\rho = 1$, rather than a drifting sequence of distributions. For this distribution, $\lambda_{\min}(\Psi_F^{vec(G_i)}) = 0$.

(iii) The intuition for the results in Lemma 12.1(a) and (b) is as follows. As (12.3) shows, G_i is close to being proportional to g_i when $\pi_{F_n} = 0^k$ and ρ_n is close to one. And, when $\pi_{F_n} = 0^k$ and $\rho_n = 1$, they are exactly proportional. By averaging over $i = 1, ..., n$ and by taking expectations, the same properties are seen to hold for $\widehat{G}_n$ and $\widehat{g}_n$ and their population counterparts. In consequence, $\widehat{D}_n$ $(:= \widehat{G}_n - \widehat{\Gamma}_n \widehat{\Omega}_n^{-1} \widehat{g}_n$ when $p = 1)$ is close to 0^k (because it is a sample version of the $L^2(F)$ projection of G_i on g_i) and the same is true of the population counterpart of $\widehat{D}_n$ (because it is the $L^2(F)$ projection of G_i on g_i). The latter implies that the direction of the k-vector $\widehat{D}_n$ is primarily random. In consequence, this direction turns out to be sensitive to the specification of the sample matrices $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ even within the class of consistent estimators of their population counterparts.

One consistent choice of $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ (used in Lemma 12.1(a)) yields $\widehat{D}_n$ to be very close to being proportional to $\widehat{g}_n$. In this case, the projection of $\widehat{\Omega}_n^{-1/2} \widehat{g}_n$ onto $\widehat{\Omega}_n^{-1/2} \widehat{D}_n$ is asymptotically equivalent to $\widehat{\Omega}_n^{-1/2} \widehat{g}_n$ itself. The LM statistic is a quadratic form in this projection k-vector (i.e., $P_{\widehat{\Omega}_n^{-1/2} \widehat{D}_n} \widehat{\Omega}_n^{-1/2} \widehat{g}_n$) multiplied by n. Hence, it behaves asymptotically like a quadratic form in $\widehat{\Omega}_n^{-1/2} \widehat{g}_n$ multiplied by n, which is just the AR statistic. This explains the result in Lemma 12.1(a).

On the other hand, when $\rho_n = 1$ (which implies that $\widehat{G}_n = -\widehat{g}_n$ by (12.3)), another consistent choice of $\widehat{\Gamma}_n$ and $\widehat{\Omega}_n$ (used in Lemma 12.1(b)) yields $\widehat{D}_n = 0^k$ a.s. In this case, the projection of $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ onto $\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ equals 0^k a.s. Hence, the LM statistic (which is a quadratic form in this projection times n) equals zero a.s. This explains the result in Lemma 12.1(b).

(iv) The result of Lemma 12.1(a) also holds for the model described in Comment (ii). Hence, drifting sequences of distributions are not required to show the result of Lemma 12.1(a) if one removes the condition $\lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 > 0$ entirely from $\mathcal{F}_{0j}$. Furthermore, the result of Lemma 12.1(a) can be extended to cover weak IV cases (in which $\pi = \pi_n \neq 0^k$, but $\pi_n \to 0^k$ sufficiently quickly as $n \to \infty$), rather than the irrelevant IV case (in which $\pi = 0^k$).

(v) Finite sample simulations corroborate the asymptotic result given in Lemma 12.1(a). For the model and LM test described in Comment (i) with $k = 5$, $\pi = 0^k$, $\rho = 1$, $Z_i \sim N(0^5, I_5)$, $(u_i, \xi_i) \sim N(0^2, I_2)$, and Z_i independent of (u_i, ξ_i), the null rejection rate of the nominal 5% LM test is 59.4% when $n = 200$ and 57.6% when $n = 1000$. However, when ρ deviates from 1 even by a small amount, the magnitude of over-rejection drops very quickly. The null rejection rate of this nominal 5% LM test is 10.1% when $\rho = 0.99$ and $n = 200$ and 12.9% when $\rho = 0.998$ and $n = 1000$. (These simulation results are based on 50,000 simulation repetitions.)

(vi) The conditions of Lemma 12.1(a) and (b) are consistent with those of Theorem 1 of Kleibergen (2005). This implies that the χ^2_p asymptotic distribution of the LM statistic obtained in the latter only holds under additional conditions, such as those in $\mathcal{F}_0$.

12.2 Proof of Lemma 12.1

Proof of Lemma 12.1. To prove part (a), we use the model defined in (12.1)-(12.3). We have

$$\begin{aligned}
\widehat{G}_n &= -\rho_n \widehat{g}_n + \delta_n \widehat{G}_n^*, \text{ where } \widehat{G}_n^* := n^{-1}\sum_{i=1}^n G_i^*, \text{ and} \\
\widehat{\Gamma}_{1n} &= n^{-1}\sum_{i=1}^n G_i g_i' = n^{-1}\sum_{i=1}^n (-\rho_n g_i + \delta_n G_i^*) g_i' = -\rho_n \widehat{\Omega}_n - \rho_n \widehat{g}_n \widehat{g}_n' + \delta_n \widehat{\Gamma}_{1n}^*, \text{ where} \\
\widehat{\Gamma}_{1n}^* &:= n^{-1}\sum_{i=1}^n G_i^* g_i'. \qquad (12.4)
\end{aligned}$$

We choose $\{\rho_n : n \geq 1\}$ to converge to one sufficiently fast that $n\delta_n \to 0$, where $\delta_n = (1-\rho_n^2)^{1/2}$ by (12.3). For example, we can take $\rho_n = (1 - n^{-3})^{1/2}$. Using the results above, we obtain

$$\begin{aligned}
\widehat{D}_n &= \widehat{G}_n - \widehat{\Gamma}_{1n}\widehat{\Omega}_n^{-1}\widehat{g}_n \\
&= -\rho_n \widehat{g}_n + \delta_n \widehat{G}_n^* - [-\rho_n \widehat{\Omega}_n - \rho_n \widehat{g}_n \widehat{g}_n' + \delta_n \widehat{\Gamma}_{1n}^*]\widehat{\Omega}_n^{-1}\widehat{g}_n \\
&= \rho_n(\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)\widehat{g}_n + \delta_n(\widehat{G}_n^* - \widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1}\widehat{g}_n). \qquad (12.5)
\end{aligned}$$

This gives

$$\widetilde{g}_n := \widehat{g}_n + n\delta_n\zeta_n = \widehat{D}_n/(\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n), \text{ where}$$
$$\zeta_n := (\widehat{G}_n^* - \widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1}\widehat{g}_n)/(\rho_n n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n) = O_p(n^{-1/2}) \text{ and } \widetilde{g}_n = \widehat{g}_n + o_p(n^{-1/2}), \quad (12.6)$$

where $\zeta_n = O_p(n^{-1/2})$ because $\rho_n \to 1$, $\widehat{G}_n^* = O_p(n^{-1/2})$ by the CLT since $E_{F_n}G_i^* = -E_{F_n}Z_i \cdot E_{F_n}\xi_i = 0^k$, $\widehat{\Gamma}_{1n}^*\widehat{\Omega}_n^{-1} = O_p(1)$ by the WLLN applied twice and $\lambda_{\min}(E_{F_n}g_ig_i') = \lambda_{\min}(E_{F_n}Z_iZ_i') \geq \delta > 0$, $\widehat{g}_n = O_p(n^{-1/2})$ by the CLT, and $(n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)^{-1} = O_p(1)$, which holds by the CMT because $AR_n = n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \to_d \chi_k^2$ (by the CLT, WLLN, and CMT) and $\chi_k^2 > 0$ a.s., and lastly the result for $\widetilde{g}_n$ in the second line of (12.6) holds by $\zeta_n = O_p(n^{-1/2})$ and $n\delta_n = o(1)$.

Projections are invariant to nonzero scalar multiplications of the matrix that defines the projection. That is, $P_A = P_{cA}$ for any matrix A and any scalar $c \neq 0$. We have $\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n \neq 0$ wp$\to$1 because $(n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n)^{-1} = O_p(1)$ and $\rho_n \to 1$. So, the LM statistic is unchanged wp$\to$1 when $\widehat{D}_n$ is replaced by $\widehat{D}_n/(\rho_n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n) = \widetilde{g}_n = \widehat{g}_n + o_p(n^{-1/2})$ using (12.6). Thus, we have

$$\begin{aligned} LM_n &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \\ &= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widetilde{g}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n + o_p(1) \\ &= n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widetilde{g}_n(\widetilde{g}_n'\widehat{\Omega}_n^{-1}\widetilde{g}_n)^{-1}\widetilde{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n + o_p(1) \\ &= n\widehat{g}_n'\widehat{\Omega}_n^{-1}\widehat{g}_n + o_p(1) = AR_n + o_p(1) \to_d \chi_k^2, \end{aligned} \quad (12.7)$$

which completes the proof of part (a).

Next, we prove part (b). In this case, we use the model in (12.1)-(12.3) with $\rho_n = 1$ and $\delta_n = 0$ for all $n \geq 1$. In consequence, $G_i = -g_i$ and $\widehat{G}_n = -\widehat{g}_n$. Given the definitions of $\widehat{\Omega}_n$ and $\widehat{\Gamma}_{1n}$ in (4.1) and (4.3), this yields

$$\begin{aligned} \widehat{\Gamma}_{1n} &= n^{-1}\sum_{i=1}^n G_ig_i' - \widehat{G}_n\widehat{g}_n' = -n^{-1}\sum_{i=1}^n g_ig_i' + \widehat{g}_n\widehat{g}_n' = -\widehat{\Omega}_n, \\ \widehat{D}_n &= \widehat{G}_n - \widehat{\Gamma}_{1n}\widehat{\Omega}_n^{-1}\widehat{g}_n = 0^k, \text{ and} \\ LM_n &:= n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{0^k}\widehat{\Omega}_n^{-1/2}\widehat{g}_n = 0 \end{aligned} \quad (12.8)$$

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

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

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica,* 80, 2153–2211. Supplemental Material available at *Econometrica Supplemental Material,* 80, http://www.econometricsociety.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 available with Cowles Foundation Discussion Paper No. 1824R, 2011, Yale University.

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

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

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

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

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

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

Andrews, I., and A. Mikusheva (2012): "A Geometric Approach to Weakly Identified Econometric Models," unpublished manuscript, Department of Economics, MIT.

——— (2014a): "Conditional Inference with a Functional Nuisance Parameter," unpublished manuscript, Department of Economics, MIT.

——— (2014b): "Maximum Likelihood Inference in Weakly Identified Models," *Quantitative Economics,* forthcoming.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica,* 63, 841–890.

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory,* 11, 1131–1147.

Chamberlain, G. (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica,* 75, 609–652.

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.

Cheng, X. (2014): "Uniform Inference in Nonlinear Models with Mixed Identification Strength," unpublished manuscript, Department of Economics, University of Pennsylvania.

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

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.

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.

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.

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.

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

Hillier, G. (2009): "Exact Properties of the Conditional Likelihood Ratio Test in an IV Regression Model," *Econometric Theory,* 25, 915–957.

Inoue, A., and B. Rossi (2011): "Testing for Weak Identification in Possibly Nonlinear Models," *Journal of Econometrics,* 161, 246–261.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica,* 70, 1781–1803.

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

McCloskey, A. (2011): "Bonferroni-Based Size-Correction for Nonstandard Testing Problems," unpublished manuscript, Department of Economics, Brown University.

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.

——— (2009): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," *Journal of Econometrics,* 152, 131–140.

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

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

——— (1987b): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review,* 28, 777–787.

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

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

Phillips, P. C. B. (1989): "Partially Identified Econometric Models," *Econometric Theory,* 5, 181–240.

Ploberger, W. (2012): "Optimal Tests for Models with Weak Instruments," unpublished manuscript, Department of Economics, Washington University in St. Louis.

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.