Econometric Theory, **33**, 2017, 1046–1080.
doi:10.1017/S0266466616000347

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

Donald W.K. Andrews
Yale University

Patrik Guggenberger
Pennsylvania State University

An influential paper by Kleibergen (2005, *Econometrica* 73, 1103–1123) 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 have 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 $p \geq 2$ when the weighting is based on an estimator of the variance of the sample moments, i.e., moment-variance weighting, combined with the Robin and Smith (2000, *Econometric Theory* 16, 151–175) rank statistic. The paper also determines a formula for the asymptotic size of the CLR test when the weighting is based on an estimator of the variance of the sample Jacobian. However, the results of the paper do not guarantee correct asymptotic size when $p \geq 2$ with the Jacobian-variance weighting, combined with the Robin and Smith (2000, *Econometric Theory* 16, 151–175) rank statistic, 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.

Andrews and Guggenberger gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-1058376 and SES-1355504, and SES-1021101, SES-1326827, and SES-1462707, respectively. The authors thank four referees, the co-editor, the editor Peter Phillips, 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, Cologne, Columbia, Freiburg, Hanover, Harvard/MIT, Hebrew Jerusalem, Konstanz, Manchester, Maryland, Michigan, Montreal, New York, Northwestern, Ohio State, Pompeu Fabra, Princeton, Queen's, Strasbourg, University College London, Washington, and Wisconsin. Address correspondence to Donald W.K. Andrews, Cowles Foundation, P.O. Box 208281, New Haven, CT 06520-8281; e-mail: donald.andrews@yale.edu.

1. INTRODUCTION

We consider the moment condition model

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

where $0^k = (0,\ldots,0)' \in R^k$, the equality holds when $\theta \in \Theta \subset R^p$ is the true value, $\{W_i \in R^m : i = 1,\ldots,n\}$ 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 $k > p$.

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 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 the conditioning statistic (onto which the 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 14 in the Supplemental Material to this paper, Andrews and Guggenberger (2014b), hereafter SM. 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. Problems of this sort are demonstrated for the Robin and Smith (2000) rank statistic and may or may not occur with other rank statistics.

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., some test statistics have nonstandard asymptotic distributions due to identification deficiency) when $\lim n^{1/2} s_{pF_n} < \infty$, where $\{s_{jF} : j = 1, \ldots, p\}$ 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 semistrong identification occurs when $\lim n^{1/2} s_{pF_n} = \infty$. Strong identification occurs when $\lim s_{pF_n} > 0$ and semistrong 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 $\lim n^{1/2} ||E_F G(W_i, \theta_0)|| < \infty$, where $|| \cdot ||$ denotes the Euclidean norm. 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 semistrongly 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 on identification-robust tests consider all of the cases of weak identification that may occur when $p \geq 2$.[1]

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.

Because of space constraints, the proofs of the results of the paper and some additional results are given in the SM.

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 Thms. 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. 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. When $\lim n^{1/2} s_{pF_n} < \infty$, case (ii) is an example of what we refer to as "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 for identification-robust tests like Kleibergen's LM and CLR tests. Sargan (1959, 1983) provides asymptotic results for estimators in some nonlinear-in-parameters, but linear-in-variables, models that fall into cases (i) and (ii). Phillips (2016) establishes asymptotic results for IV estimators, Wald tests, and tests of over-identifying restrictions in linear IV regression models with nearly singular endogenous variables in the sense that their covariance with the IV's exhibits case (ii) behavior and simultaneously their variance matrix is nearly singular. 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

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

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 $\pi \ (= \pi_F) \in R^{k \times p}$ is an unknown unrestricted parameter matrix. 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.

The data $\{W_i = (y_{1i}, Y_{2i}', Z_i')' : i = 1, \ldots, n\}$ are i.i.d. and $E_F((u_i, V_{2i}')'|Z_i) = 0^{p+1}$ a.s. Here $m = 1 + p + k$ and

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

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 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 $(1, \ldots, 1)'$ and $(1 + o(1), \ldots, 1 + o(1))'$. Identification of this type has not been dealt with in the literature on identification-robust 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.

SW deal with a subset of case (i) for GMM versions of the Anderson–Rubin test. Their conditions rule out case (ii). Guggenberger and Smith (2005), Otsu (2006), Inoue and Rossi (2011), Guggenberger, Ramalho, and Smith (2012), and I. Andrews (2016) deal with a subset of case (i) for generalized empirical likelihood (GEL) and GMM versions of the LM and/or CLR tests, but rule out case (ii) by assumption. Furthermore, their results for case (i) rely on Assumption C of SW. (Inoue and Rossi (2011) and I. Andrews (2016, Appendix C) use conditions that are much like Assumption C of SW, but they are not exactly the same.) 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. This assumption is an innovative contribution to the literature, and it provides a useful guide to the behavior of tests under some types of weak identification. But, it has some significant drawbacks as a general high-level condition.

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.[2] It appears to be hard to verify or refute Assumption C in the CCAPM of interest.

In addition, I. Andrews and Mikusheva (2016a, Section S8 of their Supplemental Material) verify a generalization of Assumption C, which allows an $O(n^{-1})$ error in the expected moment functions $n^{-1/2}m_{1n}(\theta)+m_2(\beta)$, in a highly-stylized small-scale DSGE model for certain sequences of drifting parameters. Their verification exploits the additive separability in the moment conditions, between the moments and the function of the parameters, which arises in minimum distance models. It utilizes a reparametrization of the structural parameters. I. Andrews and Mikusheva (2016a, Section S8) conclude "even in this highly stylized model deriving the weakly and strongly identified directions in the parameter space is messy, and such derivations will be difficult if not impossible in richer, more empirically relevant models." Note that to verify Assumption C of SW for some sequences of parameters (and/or distributions) involves deriving the weakly and strongly identified directions in the parameter space.

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 $i = 1, \ldots, n$, 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.[3]

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

(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 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 "semistrong 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).

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 19 in the SM.[4] Newey and Windmeijer (2009) consider the limiting null rejection 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, Ramalho, and Smith (2012) establishes the correct asymptotic size of heteroskedasticity-robust LM and CLR tests in a heteroskedastic model with $p = 1$. I. Andrews (2016) 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 (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, Thm. 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 (2016). 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 a certain transformation of 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 semistrong identification. 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). For related results, see Chamberlain (2007) and Mikusheva (2010).

We also mention the recent paper by I. Andrews and Mikusheva (2016b) 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.

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, 2013, 2014) 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 (2013) 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 (2015) 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 (2016a) provide subvector inference methods in a minimum distance context based on Anderson–Rubin-type statistics. I. Andrews and Mikusheva (2015) provide conditions under which subvector inference is possible in exponential family models under an assumption of linearity in any weakly-identified nuisance parameter.

3. MOMENT CONDITION MODEL

3.1. Basic Statistics

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 $j = 1,\ldots,p$. 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 semidefinite (p.s.d.) matrix A, we let $\lambda_j(A)$ denote the jth largest eigenvalue of A.

Let

$$\widehat{g}_n(\theta) := n^{-1}\sum_{i=1}^n g_i(\theta),\ \widehat{G}_n(\theta) := n^{-1}\sum_{i=1}^n G_i(\theta), \text{ and}$$
$$\widehat{\Omega}_n(\theta) := n^{-1}\sum_{i=1}^n g_i(\theta)g_i(\theta)' - \widehat{g}_n(\theta)\widehat{g}_n(\theta)'. \tag{3.1}$$

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 10 in the SM 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 (3.1).

Next, following Kleibergen (2005), we define the *orthogonalized sample Jacobian,* denoted by $\widehat{D}_n(\theta)$, which equals the sample Jacobian $\widehat{G}_n(\theta)$ adjusted to be asymptotically independent of the sample moments $\widehat{g}_n(\theta)$:

$$\widehat{D}_n(\theta) := \left(\widehat{D}_{1n}(\theta),\ldots,\widehat{D}_{pn}(\theta)\right) \in R^{k\times p},$$
$$\widehat{D}_{jn}(\theta) := \widehat{G}_{jn}(\theta) - \widehat{\Gamma}_{jn}(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \in R^k \text{ for } j=1,\ldots,p,$$
$$\widehat{G}_n(\theta) := \left(\widehat{G}_{1n}(\theta),\ldots,\widehat{G}_{pn}(\theta)\right) \in R^{k\times p}, \text{ and}$$
$$\widehat{\Gamma}_{jn}(\theta) := n^{-1}\sum_{i=1}^n \left(G_{ij}(\theta) - \widehat{G}_{jn}(\theta)\right) g_i(\theta)' \in R^{k\times k} \text{ for } j=1,\ldots,p. \tag{3.2}$$

The $\widehat{D}_n(\theta)$ statistic is a basic component of Kleibergen's (2005) LM and CQLR tests.

3.2. Definition of the Parameter Space for the Distributions F

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

$$\mathcal{F} := \Big\{F : \{W_i : i \geq 1\} \text{ are i.i.d. under } F,\ E_F g_i = 0^k,$$
$$E_F\|(g_i', vec(G_i)')'\|^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(E_F g_i g_i') \geq \delta\Big\}, \tag{3.3}$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix and $vec(\cdot)$ denotes the vector obtained by stacking the columns of a matrix. The first condition in $\mathcal{F}$ specifies that the observations are i.i.d. For time series observations, see Section 7. The second condition in $\mathcal{F}$ is the defining condition of the model. The third 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. 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 $G_i^* := (E_F g_i g_i')^{-1/2} G_i$. (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.) 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. Also note that 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 (2015) find it convenient to rescale moment conditions by diagonal matrices.

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.

Let

$$\Omega_F := E_F g_i g_i'. \tag{3.4}$$

The following is a singular-value decomposition of $\Omega_F^{-1/2} E_F G_i$:

$$\Omega_F^{-1/2} E_F G_i = C_F \Upsilon_F B_F', \text{ where}$$
$$\Upsilon_F := \begin{pmatrix} Diag\{\tau_{1F}, \ldots, \tau_{pF}\} \\ 0^{(k-p)\times p} \end{pmatrix} \in R^{k\times p},$$

$(\tau_{1F}, \ldots, \tau_{pF})$ denote the p singular values of $\Omega_F^{-1/2} E_F G_i$ in nonincreasing order, and B_F and C_F are $p \times p$ and $k \times k$ orthogonal matrices, respectively.

(3.5)

The singular values $\tau_{1F}, \ldots, \tau_{pF}$ are nonnegative and may be zero. The matrix B_F contains eigenvectors of $(E_F G_i)' \Omega_F^{-1} (E_F G_i)$ ordered so that the corresponding eigenvalues $(\kappa_{1F}, \ldots, \kappa_{pF})$ are nonincreasing. The matrix C_F contains eigenvectors of $\Omega_F^{-1/2}(E_F G_i)(E_F G_i)' \Omega_F^{-1/2}$ ordered so that the corresponding eigenvalues are $(\kappa_{1F}, \ldots, \kappa_{pF}, 0, \ldots, 0)' \in R^k$. Note that $\kappa_{jF} = \tau_{jF}^2$ for $j \leq p$.

With some abuse of notation, for an integer $0 \leq j \leq p$, 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 $0 \leq j \leq k$, 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 an arbitrary square-integrable (under F) vector a_i, let

$$\Sigma_F^{a_i} := E_F a_i a_i', \; \Gamma_F^{a_i} := E_F a_i g_i', \text{ and } \Psi_F^{a_i} := \Sigma_F^{a_i} - \Gamma_F^{a_i} \Omega_F^{-1} \Gamma_F^{a_i\prime}. \tag{3.6}$$

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.

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

$$\Psi_{jF}(\xi) := \Psi_F^{C_{F,k-j}' \Omega_F^{-1/2} G_i B_{F,p-j} \xi}. \tag{3.7}$$

We can write

$$\Psi_{jF}(\xi) = \left(\xi' B_{F,p-j}' \otimes C_{F,k-j}' \Omega_F^{-1/2}\right) \Psi_F^{vec(G_i)} \left(B_{F,p-j}\xi \otimes \Omega_F^{-1/2} C_{F,k-j}\right) \text{ and}$$
$$\Psi_F^{vec(G_i)} = E_F G_{Fi} G_{Fi}', \text{ where } G_{Fi} := vec(G_i) - \Gamma_F^{vec(G_i)} \Omega_F^{-1} g_i \in R^{pk} \tag{3.8}$$

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

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

$$\mathcal{F}_0 := \bigcup_{j=0}^{p} \mathcal{F}_{0j}, \text{ where} \tag{3.9}$$
$$\mathcal{F}_{0j} := \left\{F \in \mathcal{F} : \tau_{jF} \geq \delta_1 \text{ and } \lambda_{p-j}(\Psi_{jF}(\xi)) \geq \delta_1 \; \forall \xi \in R^{p-j} \text{ with } ||\xi|| = 1\right\},$$

$\tau_{0F} := \delta_1$, and $\lambda_{p-j}\left(\Psi_{jF}(\xi)\right) := \delta_1$ for $j = p$.[5] 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. We assume that $\mathcal{F}_0 \neq \varnothing$.

The conditions in $\mathcal{F}_0$ are used to show that the estimator $\widehat{\Omega}_n^{-1/2} \widehat{D}_n \in R^{k \times p}$ of the normalized orthogonalized 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 10.3(d) in the SM). 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 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. Sequences of distributions in the semistrongly identified category can come from sets $\mathcal{F}_{0j}$ for any $j < p$.

Linking the parameter spaces $\mathcal{F}_{0j}$ for $j = 0, \ldots, p$ with identification categories, as is done in the previous 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, semistrong, and weak in the standard and nonstandard senses) depend on the properties of *sequences* of distributions $\{F_n : n \geq 1\}$.

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 $C'_{F,j}\Omega_F^{-1/2} E_F G_i B_{F,j} = Diag\{\tau_{1F}, \ldots, \tau_{jF}\} \in R^{j \times j}$ and the column rank of $Diag\{\tau_{1F}, \ldots, \tau_{jF}\}$ is j by the definition of $\mathcal{F}_{0j}$.)[6] The $\lambda_{p-j}\left(\Psi_{jF}(\xi)\right) \geq \delta_1$ 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.

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, $k = p = 2$, $E_F G_i = -I_2$, 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 $E_F G_i = -I_2$). 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 $E_F||g_i||^{2+\gamma} \leq M$. (This holds because $E_F G_i = -I_2$ and Ω_F has elements $[\Omega_F]_{11} = \theta_{20}$, $[\Omega_F]_{12} = [\Omega_F]_{21} = E_F U_i(U_i^2 - \theta_{20})$, and $[\Omega_F]_{22} = E_F(U_i^2 - \theta_{20})^2$, where $\theta_{20} := Var_F(Y_i)$, $U_i := Y_i - \theta_{10}$, $\theta_{10} := E_F Y_i$, and $\theta_0 = (\theta_{10}, \theta_{20})'$ denotes the true null value.) 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 14 in the SM. Hence, the restrictions on the parameter space $\mathcal{F}_0$ are not redundant. 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.

In contrast, the SR-AR and SR-CQLR type tests introduced in AG2 are shown to have correct asymptotic size without any conditions on $\lambda_{p-j}(\Psi_{jF}(\xi))$ or $\lambda_{\min}(E_F g_i g_i')$. 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 strength, are:

$$\text{(i) } \lambda_{\min}\left(\Psi_F^{vec(\overline{C}_{F,p-j}'\Omega_F^{-1/2}G_i B_{F,p-j})}\right) \geq \delta_1 \text{ for some matrix } \overline{C}_{F,p-j},$$

$$\text{(ii) } \lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_i B_{F,p-j})}\right) \geq \delta_1,$$

$$\text{(iii) } \lambda_{\min}\left(\Psi_F^{vec(\Omega_F^{-1/2}G_i)}\right) \geq \delta_1,$$

$$\text{(iv) } \lambda_{\min}\left(\Psi_F^{vec(G_i)}\right) \geq \delta_2 := \delta_1 M^{2/(2+\gamma)},$$

$$\text{(v) } \lambda_{\min}(\Sigma_F^{f_i}) \geq \delta_2, \text{ where } \Sigma_F^{f_i} := E_F f_i f_i' \text{ and } f_i := \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}, \text{ and}$$

$$\text{(vi) } \lambda_{\min}(Var_F(f_i)) \geq \delta_2, \qquad \textbf{(3.10)}$$

where M and γ are as in (3.3) and δ_1 is as in (3.9). 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. See Section 18 in the SM 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

$$\lambda_p\left(\Psi_F^{\Omega_F^{-1/2}G_i B_{F,p-j}\xi}\right) \geq \delta_1 \ \forall \xi \in R^{p-j} \text{ with } ||\xi|| = 1. \tag{3.11}$$

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 (3.10) (and, hence, for the $\lambda_{p-j}(\cdot)$ condition in $\mathcal{F}_{0j}$) in the linear IV regression model are as follows. We have

$$\begin{aligned}
\Sigma_F^{f_i} &= E_F\left((u_i, -Y_{2i}')' \otimes Z_i\right)\left((u_i, -Y_{2i}')' \otimes Z_i\right)' \\
&= E_F(\varepsilon_i \otimes Z_i)(\varepsilon_i \otimes Z_i)' + E_F s_i(\pi) s_i(\pi)' \text{ and} \\
Var_F(f_i) &= E_F(\varepsilon_i \otimes Z_i)(\varepsilon_i \otimes Z_i)' + E_F s_i(\pi) s_i(\pi)' - E_F s_i(\pi) E_F s_i(\pi)' \\
&\geq E_F(\varepsilon_i \varepsilon_i' \otimes Z_i Z_i'), \text{ where} \\
&\varepsilon_i := (u_i, -V_{2i}')', \ s_i(\pi) := (0^{k\prime}, -(Z_i Z_i' \pi_1)', \ldots, -(Z_i Z_i' \pi_p)')',
\end{aligned} \tag{3.12}$$

$\pi = (\pi_1, \ldots, \pi_p)$ for $\pi_j \in R^k$ for $j = 1, \ldots, p$, and the inequality holds in a p.s.d. 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 bounded away from zero by $\delta_2^{1/2}$.

3.3. Definition of $\mathbf{G(W_i, \theta)}$

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 $((g(W_i, \theta + \varepsilon e_1) - g(W_i, \theta))/\varepsilon, \ldots, (g(W_i, \theta + \varepsilon e_p) - g(W_i, \theta))/\varepsilon) \in R^{k \times p}$ for some $\varepsilon > 0$, where e_j is the jth unit vector, e.g., $e_1 = (1, 0, \ldots, 0)' \in R^p$. This choice of $G(W_i, \theta)$ matrix may be useful for models with quite complicated Jacobian matrices $(\partial/\partial\theta')g(W_i, \theta)$.

3.4. 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

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

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

$$\liminf_{n\to\infty} \inf_{F\in\overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha) = \limsup_{n\to\infty} \sup_{F\in\overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{3.14}$$

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 reason is that to determine the asymptotic coverage probabilities of CS's under all identification scenarios one needs to consider all possible asymptotic behavior of $\Omega_{F_n}^{-1/2} E_{F_n} G_i$. It does not matter whether $\Omega_{F_n}^{-1/2} E_{F_n} G_i$ varies due to variation of F_n, which is what occurs in the analysis of tests with a fixed null hypothesis, or due to variation of both F_n and the (true) null parameter upon which $\Omega_{F_n}^{-1/2} E_{F_n} G_i$ depends, which is what occurs for CS's. The analysis for tests already covers all possible variations, so nothing new is needed for the CS analysis beyond minor notational adjustments.

The dependence of the parameter space $\mathcal{F}_0$, defined in (3.9), 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

$$\mathcal{F}_{\Theta,0} := \{(F,\theta_0) : F \in \mathcal{F}_0(\theta_0), \theta_0 \in \Theta\}. \tag{3.15}$$

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).

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

$$P_A := A\left(A'A\right)^{-} A' \text{ and } M_A := I_r - P_A, \tag{4.1}$$

where $(\cdot)^-$ denotes any g-inverse. 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:

$$AR_n(\theta) := n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \text{ and}$$
$$LM_n(\theta) := n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1/2}(\theta)P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \quad \textbf{(4.2)}$$

where $\widehat{g}_n(\theta)$, $\widehat{\Omega}_n(\theta)$, and $\widehat{D}_n(\theta)$ are defined in (3.1) and (3.2).

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

$$CS_{LM,n} := \left\{\theta_0 \in \Theta : LM_n(\theta_0) \le \chi^2_{p,1-\alpha}\right\}. \quad \textbf{(4.3)}$$

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* (3.9)). *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* (3.15).

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 14 in the SM 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 postmultiplication 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 postmultiply $n^{1/2}\widehat{\Omega}_n^{-1/2}\widehat{D}_n$ 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 semistrongly identified for $q \in \{1, \ldots, p\}$, 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., $\Psi_F^{C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi}$, 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, Section 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 $rank[E_FG_i] \leq p - 1$ against $rank[E_FG_i] = p$. 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

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

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, equation (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

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

for $0 \leq r < \infty$, where χ_p^2 and χ_{k-p}^2 are independent chi-square random variables with p and $k-p$ degrees of freedom, respectively. 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

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

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

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

The function $vec_{k,p}^{-1}(\cdot)$ is the inverse of the $vec(\cdot)$ function for $k \times p$ matrices. Thus, the domain of $vec_{k,p}^{-1}(\cdot)$ consists of kp-vectors and its range consists of $k \times p$ matrices. 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. (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.) 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 nondegenerate 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

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

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 19 in the SM, see (19.9) and (19.26), 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 19 in the SM. 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 with the Robin and Smith (2000) rank statistic has correct asymptotic size.

Although the asymptotic size formula for the Jacobian-variance weighted CLR test with the Robin and Smith (2000) rank statistic is an important result of this paper, it is stated in the SM 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 19.2 in the SM), $n^{1/2}(\widehat{g}_n, \widehat{D}_n^{\dagger} T_n^{\dagger}) \rightarrow_d (\overline{g}_h, \overline{\Delta}_h^{\dagger} + \overline{M}_h^{\dagger})$, where $(\overline{g}_h, \overline{\Delta}_h^{\dagger}, \overline{M}_h^{\dagger})$ has a multivariate normal distribution whose mean and variance matrix depend on $\lim Var_{F_n}((f_i^{*\prime}, vech(f_i^* f_i^{*\prime})^{\prime})^{\prime})$ 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 (19.13), (19.15), and (19.19) in Section 19 in the SM.

(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. See the proof of Theorem 12.1 in the SM for details. 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 $p = 1$. The proof of this is given in Comment (ii) to Theorem 19.3 in the SM. If $p \geq 2$, it occurs if $E_{F_n} G_i \rightarrow 0^{k \times p}$, which covers the cases where all of the parameters are weakly identified in the standard sense or semistrongly 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 semistrongly 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 $\left(u_i, V_{2i}'\right)'$ has a multivariate normal distribution, but fail for most joint distributions of $\left(u_i, V_{2i}'\right)'$.[7] 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 of $\left(u_i, V_{2i}'\right)'$. This is a consequence of $E_F G_i$ being different between the fixed and random IV cases.

Typically, $\overline{M}_h^\dagger$ is nonzero (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 with the Robin and Smith (2000) rank statistic 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 19.3 in the SM for details. In the model considered, the simulated asymptotic null rejection probabilities are found to be in $[4.99, 5.11]$, which are quite 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 nonrandom 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 $\left(\overline{g}_h, \overline{\Delta}_h^\dagger\right)$, 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 19.5 in the SM.

(vii) In Section 19.1 in the SM, 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 the CLR test with the Robin and Smith (2000) rank statistic based on an estimator of the asymptotic variance matrix of $vec\left(\widehat{D}_n(\theta)\right)$ (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 case for all of the rank tests considered in the literature.[8]

Next, we provide a result that gives an upper bound on the possible asymptotic size distortion of any CLR test, including the Jacobian-variance weighted CLR test. Consider the test that rejects $H_0 : \theta = \theta_0$ when

$$\sup_{r\in[0,\infty]} \left[\frac{1}{2} \left(AR_n(\theta_0) - r + \sqrt{(AR_n(\theta_0) - r)^2 + 4LM_n(\theta_0) \cdot r} \right) - c(1-\alpha, r) \right] \quad \textbf{(5.7)}$$

is positive, where $c(1-\alpha, r)$ is defined in (5.2). This test rejects whenever the CLR test defined in (5.1)–(5.2) rejects, no matter how the rank statistic $rk_n(\theta_0) \in R_+$ is defined. Therefore, the asymptotic size of the test in (5.7) provides an upper bound on the asymptotic size of the CLR test for any rank statistic $rk_n(\theta_0) \in R_+$.[9]

LEMMA 5.2. *The asymptotic size of the test in* (5.7) *for the parameter space* $\mathcal{F}_0$ (*defined in* (3.9)) *equals the probability that*

$$\sup_{r\in[0,\infty]} \left[\frac{1}{2} \left(\chi_p^2 + \chi_{k-p}^2 - r + \sqrt{(\chi_p^2 + \chi_{k-p}^2 - r)^2 + 4\chi_p^2 \cdot r} \right) - c(1-\alpha, r) \right] > 0, \quad \textbf{(5.8)}$$

where χ_p^2 *and* χ_{k-p}^2 *are independent random variables with chi-square distributions with* p *and* $k-p$ *degrees of freedom, respectively.*

Comments. (i) The probability of the event in (5.8) only depends on p, k, and α and can be computed easily by simulation. Using 10^6 simulation repetitions, a grid for r given by $\{0, .5, 1, 1.5, \ldots, 100, \infty\}$, and $\alpha = 0.05$, we find that for $p = 2$ and $k = 3, 5, 10, 15, 20, \ldots, 50, 75, 100$ the probability of the event in (5.8) times 100 equals 7.00, 8.11, 9.00, 9.36, 9.57, 9.71, 9.81, 9.88, 9.95, 10.00, 10.04, 10.14, and 10.01, respectively. When $p = 5$ and $k = 6, 10, 15, 20, \ldots, 50, 75, 100$, the probability times 100 equals 6.38, 7.85, 8.54, 8.90, 9.17, 9.31, 9.45, 9.56, 9.63, 9.68, 9.91, and 9.91, respectively.

(ii) The simulation results in (i) imply that the potential asymptotic size distortion (times 100) of the nominal size 5% CLR test with Jacobian-variance weighting is bounded above by the quantities in (i) minus 5.00 for the various choices of p and k.

(iii) The proof of Lemma 5.2 is given in Section 19.7 in the SM. It uses the decomposition $AR_n(\theta) = LM_n(\theta) + J_n(\theta)$ and shows that asymptotic versions of $LM_n(\theta_0)$ and $J_n(\theta_0)$ are independent and distributed as χ^2_p and χ^2_{k-p}, respectively, conditional on a certain random matrix. Because their conditional distribution does not depend on the conditioning matrix, their unconditional and conditional distributions are the same.

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) Jacobian-variance weighted 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

$$\mathcal{F}_{JVW,p=1} := \{F \in \mathcal{F} : \lambda_{\min}(\Psi_F^{G_i} - E_F G_i E_F G_i') \geq \delta_3\} \tag{5.9}$$

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 (5.9) 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 (3.9) and condition (iv) in (3.10)), where $\mathcal{F}_{00} = \mathcal{F}_{0j}$ with $j = 0$ (for $\mathcal{F}_{0j}$ defined in (3.9)) 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

$$rk_n := n\widehat{D}_n{}'\widetilde{V}_{Dn}^{-1}\widehat{D}_n. \tag{5.10}$$

THEOREM 5.3. *Suppose* $p = 1$. *The asymptotic size of the* CLR *test with Jacobian-variance weighting, defined by* (5.1), (5.2), *and* (5.10), *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.

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* (3.9)). *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* (3.15).

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 12 in the SM to this paper, 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 semistrong 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 12 in the SM, not just the rank statistic $rk_n(\theta)$ in (6.2). Assumption R is verified using Theorem 10.4 in the SM for the rank statistic in (6.2). With some changes, Assumption R can be verified using Theorem 10.4 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 19.5 in the SM. 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, this is just a conjecture. Proving or disproving it seems to be difficult.

Although the rank statistic in (6.2) yields a test with correct asymptotic size, it has some drawbacks. The use of the premultiplication weight matrix $\widehat{\Omega}_n^{-1/2}(\theta)$ and no postmultiplication 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 16 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, see Lemma 16.3 in the Supplemental Material to AG2. 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 type tests considered in AG2.

7. TIME SERIES OBSERVATIONS

In this section, we generalize the results of Theorems 4.1, 5.3, 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 $\{W_i : i = \ldots, 0, 1, \ldots\}$. 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. Let a_i be a random vector that depends on W_i, such as $vec(G_i)$ or $C'_{F,k-j}\Omega_F^{-1/2}G_iB_{F,p-j}\xi$. In the time series case, we define Ω_F and $\Psi_F^{a_i}$ differently from their definitions in (3.6) 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:

$$\Sigma_F^{a_i} := \sum_{m=-\infty}^{\infty} E_F(a_i - E_F a_i)(a_{i-m} - E_F a_{i-m})', \ \Gamma_F^{a_i} := \sum_{m=-\infty}^{\infty} E_F a_i g_{i-m}',$$

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

Note that $\Psi_F^{a_i} = \lim Var_F(n^{-1/2}\sum_{i=1}^n (a_i - \Gamma_F^{a_i}\Omega_F^{-1} g_i))$.[10] The definition of $\Sigma_F^{a_i}$ in (7.1) differs from its definition in (3.6) 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 (3.9), is weaker.

The time series analogue $\mathcal{F}_{TS}$ of the space of distributions $\mathcal{F}$, defined in (3.3), is

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

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 (3.9), 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 $(\tau_{1F}, \ldots, \tau_{pF})$, B_F, and C_F in (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 (5.9), 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 (3.10) and (3.11) also hold in the time series setting with $\Psi_F^{a_i}$ and $\Sigma_F^{a_i}$ defined as in (7.1).

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

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

The second equality holds for all $F \in \mathcal{F}_{TS}$ (as shown in the proof of Lemma 20.1 in Section 20 in the SM).

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 $\left\{f_i - \widehat{f}_n : i \leq n\right\}$, where $f_i := (g_i', vec(G_i)')'$ and $\widehat{f}_n(\theta) := \left(\widehat{g}_n', vec\left(\widehat{G}_n\right)'\right)'$. 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:

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

Under Assumption V, $\widehat{\Omega}_n(\theta_0) \rightarrow_p \Omega_F$ under F and $\widehat{\Gamma}_n(\theta_0) = \left(\widehat{\Gamma}_{1n}(\theta_0)', \ldots, \widehat{\Gamma}_{pn}(\theta_0)'\right)' \rightarrow_p \Gamma_F^{vec(G_i)}$ 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 (3.1)–(5.1), but with $\widehat{\Omega}_n(\theta)$ and $\widehat{\Gamma}_{jn}(\theta)$ for $j = 1, \ldots, p$ 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 $n^{-1} \sum_{i=1}^{n} vec\left(G_i(\theta) - \widehat{G}_n(\theta)\right) vec\left(G_i(\theta) - \widehat{G}_n(\theta)\right)'$ 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 $j = 1, \ldots, p$ 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* $\mathcal{F}_{\Theta,TS,JVW,p=1}$.

Comment. Theorem 7.1 shows that the results of Theorems 4.1, 5.3, and 6.1 for i.i.d. observations generalize to strictly stationary strong mixing observations, provided the space of distributions $\mathcal{F}$ is adjusted suitably and the variance estimator $\widehat{V}_n(\theta_0)$ of V_F is defined appropriately.

8. CONCLUSION

This paper analyzes the asymptotic properties of the LM and CLR tests and confidence sets introduced in Kleibergen (2005) for nonlinear moment condition models. These procedures aim to be identification robust. This paper determines the asymptotic size and similarity (in a uniform sense) of these procedures for suitably specified parameter spaces of null distributions of the observations.

The LM test is found to have correct asymptotic size and to be asymptotically similar for a suitably chosen parameter space of null distributions. The same is true for the CLR tests when $p = 1$, where p is the dimension of the unknown parameter θ. However, when $p \geq 2$, the asymptotic sizes of the CLR tests are found to depend on how the conditioning statistic employed by CLR tests is weighted. When the weighting is based on an estimator of the variance of the sample moments, i.e., moment-variance weighting, combined with the Robin and Smith (2000) rank statistic, the paper finds that the CLR tests are guaranteed to have correct asymptotic size when $p \geq 2$. When the weighting is based on an estimator of the variance of the sample Jacobian, the paper determines a formula for the asymptotic size of the CLR test. However, the results of the paper do not guarantee correct asymptotic size when $p \geq 2$ with the Jacobian-variance weighting, combined with the Robin and Smith (2000) rank statistic, because two key sample quantities are not necessarily asymptotically independent under some identification scenarios.

The results of this paper are employed in AG2, which develops some new identification-robust tests and confidence sets that allow for a broader specification of the parameter space of null distributions that generate the data than the procedures considered in this paper.

NOTES

1. 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. 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.

3. 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. 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.

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

However, under sequences for which a component of $n^{1/2}E_{F_n}G(W_i,\theta)$ diverges to plus or minus infinity, the CMT cannot be applied because $n^{1/2}\widehat{D}_n(\theta)$ does not converge in distribution, but rather, some component of it diverges to plus or minus infinity in probability (see Lemma 10.3 in Section 10 in the SM when $h_{1,j}=\infty$ for some $j\le p$). 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. A simple example is given at the beginning of Section 19 in the SM.

5. 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).

6. The stated equality holds because (i) by (3.5) $\Omega_F^{-1/2}E_FG_i=C_FDiag(\tau_F)B_F'$, where $Diag(\tau_F)$ is the $k\times p$ matrix whose (m,m) element equals τ_{mF} for $m=1,\ldots,p$ and whose other elements all equal zero, (ii) $C_F'\Omega_F^{-1/2}E_FG_iB_F=Diag(\tau_F)$ by the orthogonality of C_F and B_F, and, hence, (iii) $C_{F,j}'\Omega_F^{-1/2}E_FG_iB_{F,j}=Diag\{\tau_{1F},\ldots,\tau_{jF}\}$.

7. The correlation between g_i and $f_i^*f_i^{*\prime}$ is zero in this case by the following: $y_{1i}=Y_{2i}'\theta+u_i$, $Y_{2i}=Z_i'\pi+V_{2i}$, $g_i=Z_iu_i$, $G_i=-Z_iY_{2i}$, 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 $s,t,\ell,j,m\ge 1$, 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).

8. This holds because all CLR tests of the form in (5.1) and (5.2) are asymptotically equivalent to the LM test in (4.2) under the null and $n^{-1/2}$ local alternatives under strong IV's, by (12.3) and (12.4) in the proof of Theorem 12.1 in Section 12 in the SM, and, as noted above, the LM test is asymptotically efficient in a GMM sense under strong IV's. Note that, by definition in (4.2), the LM statistic uses moment-variance weighting of $\widehat{D}_n(\theta)$ in its projection matrix.

9. We thank a referee for suggesting an asymptotic size upper bound based on the test in (5.7).

10. This follows by calculations analogous to those in (20.3) and (20.4) in the proof of Theorem 7.1 in the SM.

REFERENCES

Anderson, T.W. & 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. & 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.

Andrews, D.W.K. & X. Cheng (2013) 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 (2014) GMM estimation and uniform subvector inference with possible identification failure. *Econometric Theory* 30, 1–47.

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

Andrews, D.W.K. & P. Guggenberger (2014b) Supplement to 'Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models'. *Econometric Theory*, Supplementary Material available at http://dx.doi.org/10.1017/S0266466616000347.

Andrews, D.W.K., M.J. Moreira, & J.H. Stock (2006) Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Andrews, D.W.K., M.J. Moreira, & J.H. Stock (2008) Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments. *Journal of Econometrics* 146, 241–254.

Andrews, I. (2016) Conditional linear combination tests for weakly identified models. *Econometrica,* first published online 29 October 2013.

Andrews, I. & A. Mikusheva (2015) Maximum likelihood inference in weakly identified DSGE models. *Quantitative Economics* 6, 123–152.

Andrews, I. & A. Mikusheva (2016a) A geometric approach to nonlinear econometric models. *Econometrica* 84, 1249–1264.

Andrews, I. & A. Mikusheva (2016b) Conditional inference with a functional nuisance parameter. *Econometrica* 84, 1571–1612.

Berry, S., J. Levinsohn, & A. Pakes (1995) Automobile prices in market equilibrium. *Econometrica* 63, 841–890.

Cavanagh, C.L., G. Elliott, & 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, & E. Zivot (2010) A new projection-type split-sample score test in linear instrumental variables regression. *Econometric Theory* 26, 1820–1837.

Chaudhuri, S. & 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. (2015) Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics* 189, 207–228.

Chernozhukov, V., C. Hansen, & M. Jansson (2009) Admissible invariant similar tests for instrumental variables regression. *Econometric Theory* 25, 806–818.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics* 51, 113–150.

Cragg, J.C. & 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.

Cragg, J.C. & S.G. Donald (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. & 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, & 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, & R.J. Smith (2012) GEL statistics under weak identification. *Journal of Econometrics* 170, 331–349.

Guggenberger, P. & 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. & 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.

Kleibergen, F. (2004) Testing subsets of structural parameters in the instrumental variables regression model. *Review of Economics and Statistics* 86, 418–423.

Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified. *Econometrica* 73, 1103–1123.

Kleibergen, F. (2007) Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* 139, 181–216.

Kleibergen, F. & 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. (2013) Admissible, Similar Tests: A characterization. Unpublished manuscript, Department of Economics, New York University.

Moreira, M.J. (2003) A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.

Moreira, M.J. (2009) Tests with correct size when instruments can be arbitrarily weak. *Journal of Econometrics* 152, 131–140.

Moreira, H. & M.J. Moreira (2013) Contributions to the theory of similar tests. Unpublished manuscript, FGV/EPGE, Rio de Janeiro, Brazil.

Newey, W.K. & K. West (1987a) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Newey, W.K. & K. West (1987b) Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777–787.

Newey, W.K. & 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.

Phillips, P.C.B. (2016) Inference in near-singular regression. *Essays in Honor of Aman Ullah, Advances in Econometrics* 36, 461–486.

Robin, J.-M. & R.J. Smith (2000) Tests of rank. *Econometric Theory* 16, 151–175.

Sargan, J.D. (1959) The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society, Series B* 21, 91–105.

Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* 51, 1605–1633.

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 R. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications*: *Ninth World Congress of the Econometric Society*, chapter 8, vol. III, Econometric Society Monograph Series, ESM 43, pp. 238–260, Cambridge University Press.

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.