Yale University
EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

9-1-2017

Identification-Robust Subvector Inference

Donald W.K. Andrews

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Andrews, Donald W.K., "Identification-Robust Subvector Inference" (2017). *Cowles Foundation Discussion Papers*. 173.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/173

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

IDENTIFICATION-ROBUST SUBVECTOR INFERENCE

By

Donald W. K. Andrews

September 2017
Updated September 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2105



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

http://cowles.yale.edu/

Identification-Robust Subvector Inference

Donald W. K. Andrews
Cowles Foundation
Yale University

First Version: June 12, 2015
Revised: September 21, 2017*

Abstract

This paper introduces identification-robust subvector tests and confidence sets (CS's) that have asymptotic size equal to their nominal size and are asymptotically efficient under strong identification. Hence, inference is as good asymptotically as standard methods under standard regularity conditions, but also is identification robust. The results do not require special structure on the models under consideration, or strong identification of the nuisance parameters, as many existing methods do.

We provide general results under high-level conditions that can be applied to moment condition, likelihood, and minimum distance models, among others. We verify these conditions under primitive conditions for moment condition models. In another paper, we do so for likelihood models.

The results build on the approach of Chaudhuri and Zivot (2011), who introduce a C(α)-type Lagrange multiplier test and employ it in a Bonferroni subvector test. Here we consider two-step tests and CS's that employ a C(α)-type test in the second step. The two-step tests are closely related to Bonferroni tests, but are not asymptotically conservative and achieve asymptotic efficiency under strong identification.

Keywords: Asymptotics, confidence set, identification-robust, inference, instrumental variables, moment condition, robust, test.

JEL Classification Numbers: C10, C12.

*The author gratefully acknowledges the research support of the National Science Foundation via grant SES-1355504. The author thanks Koohyun Kwon for constructing the figures and Jaewon Lee for a very careful proof-reading of the paper. The author thanks Isaiah Andrews, Angelo Melino, Charlie Saunders, and Katsumi Shimotsu for helpful comments.

1 Introduction

Existing identification-robust subvector tests and confidence sets (CS's) have one or more of the following drawbacks: (i) they are asymptotically conservative, such as projection and Bonferroni methods; (ii) they are not asymptotically efficient under strong identification; (iii) they only apply if nuisance parameters are strongly identified; (iv) they only apply to models with special structure, such as knowledge of the source of potential non-identification; (v) they only apply to specific models, such as the homoskedastic linear instrumental variables (IV) model; and/or (vi) they have not been shown to have correct asymptotic size under primitive conditions. In particular, there is no general identification-robust subvector method in the literature that is asymptotically non-conservative, is asymptotically efficient under strong identification, and has been shown to have correct asymptotic size. (Further discussion of the literature is given below.)

This paper aims to fill this gap in the literature. Under a set of high-level conditions, we provide a two-step Bonferroni-like method that is asymptotically non-conservative and asymptotically efficient under strong identification. The method applies to what we call *systems of equations* (SE) models, which include moment condition, likelihood, and minimum distance models, and versions of these models that rely on preliminary $n^{1/2}$-consistent estimators. In this paper, we verify the high-level conditions in moment condition models with independent identically distributed (i.i.d.) and time series observations. In Andrews (2017), we do likewise for likelihood models.

For a parameter $\theta = (\theta_1', \theta_2')' \in R^p$, we consider nominal level α tests of $H_0 : \theta_2 = \theta_{20}$ versus $H_1 : \theta_2 \neq \theta_{20}$, where θ_1 is a nuisance parameter. A two-step test relies on a first-step identification-robust CS, CS_{1n}, for θ_1 of level $1-\alpha_1$ for $\alpha_1 < \alpha$, such as $\alpha_1 = .005$ and $\alpha = .05$, as in a Bonferroni test. This CS is augmented by an estimator set, $\widehat{\Theta}_{1n}$, of θ_1 values that is designed to be such that some element of $CS_{1n}^+ := CS_{1n} \cup \widehat{\Theta}_{1n}$ is necessarily close (within $O_p(n^{-1/2})$) to the true value of θ_1 under locally strongly-identified sequences of distributions. This property is needed to obtain the correct asymptotic level of the two-step test.[1]

The two-step test employs a C(α)-type identification-robust second-step test that takes as given a value of the nuisance parameter θ_1. Chaudhuri and Zivot (2011) introduce a Lagrange multiplier (LM) test of this type for moment condition models. It is based on the (non-identification-robust) C(α) test of Neyman (1959) for likelihood models. I. Andrews (2017) also considers C(α)-type tests. In the moment condition model, we consider C(α)-type identification-robust Anderson-Rubin (AR), LM, and conditional quasi-likelihood ratio (QLR) second-step tests. The C(α)-type

[1] For example, in the moment condition model, if CS_{1n} is the (null-restricted) AR CS, then CS_{1n} is empty with probability bounded away from 0 as $n \to \infty$ when the number of moments k exceeds the dimension of θ_1, which it typically does, and hence, one cannot take $\widehat{\Theta}_{1n} = \varnothing$.

conditional QLR test considered here, which we refer to as C(α)-QLR1, is a C(α) version of a test of Kleibergen (2005) and employs a rank statistic of the form in Robin and Smith (2000).

The second-step test uses a data-dependent significance level, $\widehat{\alpha}_{2n}(\theta_1)$, that lies between $\alpha_2 := \alpha - \alpha_1$ and α, such as $\alpha_2 = .045$ and $\alpha = .05$, and depends on the given value of θ_1. This significance level is designed to equal α_2 under weak identification and transition to α under sufficiently strong identification. It is based on an identification-category-selection statistic.

The two-step test rejects H_0 if the second-step test given θ_1, with significance level $\widehat{\alpha}_{2n}(\theta_1)$, rejects the null hypothesis for all $\theta_1 \in CS_{1n}^{+}$. The two-step CS for θ_2 is obtained by inverting the two-step tests. Computation of the two-step test or CS is essentially the same as that of a Bonferroni test or CS. Thus, in some scenarios, it can be easy to compute, but in other scenarios, it can be difficult to compute.

Different first-step CS's can be employed. For moment condition models, the $H_0 : \theta_2 = \theta_{20}$ null-restricted AR CS is a good choice for power purposes because under the alternative, $H_1 : \theta_2 \neq \theta_{20}$, this CS often is small and has low coverage probability (since it is based on the incorrect null value θ_{20}). For moment condition models, the estimator set $\widehat{\Theta}_{1n}$ can be the set of solutions to the generalized method of moments (GMM) criterion function first-order conditions (FOC's) that minimize, or nearly minimize, the GMM criterion function.

The second-step C(α) tests are based on the sample SE vector, such as the sample moment vector, that has been orthogonalized with respect to (wrt) the sample Jacobian of the SE vector wrt θ_1, which in turn has been transformed to be asymptotically independent of the sample SE vector. Chaudhuri and Zivot (2011) recognize that a C(α)-type test is a good choice for the second-step test in a Bonferroni procedure because it makes the test statistic less sensitive to θ_1 and closer to being asymptotically similar, which is better for power against $\theta_2 \neq \theta_{20}$. For the same reason, C(α)-type tests are good for power in the two-step tests considered here. In fact, the C(α) nature of the second-step tests is needed for the two-step tests to achieve an asymptotic oracle property and asymptotic efficiency under strong identification.

The two-step subvector test does have some potential drawbacks. These include: (i) its asymptotic null rejection probabilities (NRP's) may be less than α under weak identification, (ii) it does not have any asymptotic efficiency properties under weak identification, (iii) it is invariant to scale reparameterizations of θ, but not all reparameterizations, (iv) it requires some tuning parameters, (v) in some scenarios it may be difficult to compute, and (vi) it takes considerable effort to verify the high-level conditions using primitive conditions.

Now we provide a heuristic explanation of the asymptotic properties of the two-step test. First, for locally-strongly-identified sequences of distributions, the two-step test obtains asymptotic NRP's

of α or less by exploiting properties of CS_{1n}^+ and the second-step C(α) test. The true θ_1 value is within $O_p(n^{-1/2})$ of CS_{1n}^+ and the two-step test rejects H_0 only if it rejects for all $\theta_1 \in CS_{1n}^+$. Thus, the test does not reject more often than the level α second-step C(α) test at some point that is $O_p(n^{-1/2})$ from the true θ_1 value. By the properties of the second-step C(α) test, such a test has asymptotic NRP α or less.

Second, for sequences of distributions that are not locally-strongly-identified, the two-step test obtains asymptotic NRP's that are α or less by a Bonferroni argument. Specifically, the augmented first step CS CS_{1n}^+ ($\supseteq CS_{1n}$) has confidence level at least $1-\alpha_1$. By design, the second-step test has significance level $\widehat{\alpha}_{2n}(\theta_1) = \alpha_2$ wp$\to$1 under such sequences when θ_1 is the true value. So, the standard Bonferroni argument gives the asymptotic NRP to be $\alpha_1 + \alpha_2 = \alpha$ or less. To make the transition between sequences of different types seamless, there are some sequences for which one can obtain NRP's of α or less using either the first or the second argument.

Given the asymptotic NRP results for certain sequences, we show that the asymptotic size of the two-step test is less than or equal to α using the subsequence-type argument in Andrews, Cheng, and Guggenberger (2011).

Next, we discuss why the asymptotic size of the two-step test is α, rather than less than α, and why it is asymptotically efficient under strong identification. For globally-strongly-identified sequences, there exists a unique solution to the population system of equations. For such sequences, the true θ_1 value is within $O_p(n^{-1/2})$ of CS_{1n}^+ and all points in CS_{1n}^+ are within $O_p(n^{-1/2})$ of the true θ_1 value. That is, the Hausdorff distance between the singleton set containing the true value θ_1 and CS_{1n}^+ is $O_p(n^{-1/2})$. For such sequences, by design, the data-dependent significance level $\widehat{\alpha}_{2n}(\theta_1)$ satisfies $\widehat{\alpha}_{2n}(\theta_1) = \alpha$ for all θ_1 within $O_p(n^{-1/2})$ of the true θ_1 value wp$\to$1. In this case, by exploiting the properties of the second-step C(α) test, one gets an oracle asymptotic equivalence property. Specifically, the two-step test is asymptotically equivalent to the nominal α oracle second-step test that employs the true value of θ_1. This yields the asymptotic size of the two-step test to be α, not less than α. It also yields asymptotic efficiency of the two-step test under these sequences, if the oracle second-step test is asymptotically efficient. For example, in the moment condition model, this holds for the second-step C(α)-LM and C(α)-QLR1 tests, but not the C(α)-AR test.

This paper considers subvector null hypotheses $H_0 : \theta_2 = \theta_{20}$ and CS's that concern θ_2. But, the results apply to some linear and nonlinear functions of an unknown parameter. Suppose one has a model indexed by $\gamma \in \Gamma \subset R^p$ and the null hypothesis of interest is $H_0 : r(\gamma) = r_0$ for some known function $r(\cdot)$ and vector r_0 of dimension $1 \le d_r \le p$. If there exists a transformation $q(\gamma) \in R^{p-d_r}$ such that $\gamma \to t(\gamma) := (q(\gamma)', r(\gamma)')'$ is a one-to-one function from Γ to $\Theta := \{\theta : \theta = t(\gamma)$ for some $\gamma \in \Gamma\}$, then the results of the present paper can be applied with $\theta = (\theta_1', \theta_2') = (q(\gamma)', r(\gamma)')'$ and

the null hypothesis $H_0 : \theta = \theta_{20}$, where $\theta_{20} = r_0$. For example, if $r(\gamma) = R_2\gamma$ for some full rank $d_r \times p$ matrix R_2, then one can take $t(\gamma) = [R_1' : R_2']'\gamma$ for any $(p - d_r) \times p$ matrix R_1 for which $[R_1' : R_2']' \in R^{p\times p}$ is nonsingular. This transformation method is employed below in a nonlinear IV model that is used in some of the finite-sample simulation results.

A second example, with a nonlinear transformation, arises with a stationary ARMA(1,1) model $Y_i = Y_{i-1}\gamma_1 + \varepsilon_i - \gamma_2\varepsilon_{i-1}$, where the null hypothesis of interest concerns the impulse response at horizon T: $H_0 : \psi_T = r_0$, where $\psi_T := \gamma_1^{T-1}(\gamma_1 - \gamma_2)$. In this case, $\theta := t(\gamma) := (\gamma_1 - \gamma_2, \gamma_1^{T-1}(\gamma_1 - \gamma_2))$ is a one-to-one transformation that yields the transformed hypothesis of interest to be $H_0 : \theta_2 = \theta_{20}$ for $\theta_{20} = r_0$. In this model, lack of identification occurs when $\gamma_1 = \gamma_2$.

The paper provides some finite-sample simulation results in two models. The first model is a heteroskedastic linear IV model with two right-hand side endogenous variables and θ_2 is the coefficient on one of them. The second model is a nonlinear (quadratic) IV model that is parametrized such that θ_2 is the value of the structural function at a point of interest, or reparametrized such that θ_2 is the function's derivative at the point of interest. For both models, we consider the two-step tests based on the first-step AR CS and the second-step C(α)-AR, C(α)-LM, and C(α)-QLR1 tests, which we denote by AR/AR, AR/LM, and AR/QLR1, respectively. We compare the power of these tests with that of the (infeasible) oracle C(α)-QLR1 test, which takes the true value of θ_1 to be known, and the projection (non-C(α)) QLR1 test, which is an existing identification-robust subvector test in the literature. In strong identification scenarios, we also consider the (non-identification-robust) standard 2SLS t test.

In both models, under strong identification, the AR/QLR1, AR/LM, and Oracle C(α)-QLR1 tests have essentially the same power. The 2SLS t test has equal power in the linear IV model to these tests and somewhat higher power in the nonlinear IV model. The AR/AR and Proj-QLR1 tests have noticeably lower power. These results are broadly consistent with the asymptotic theory.

In both models, under weak identification, the AR/QLR1 subvector test performs best in terms of power among the feasible tests, not uniformly, but in an overall sense. It noticeably out-performs the Proj-QLR1 test. The AR/LM test exhibits some quirky power behavior in some scenarios. Not surprisingly, the Oracle C(α)-QLR1 test out-performs the feasible tests in scenarios where θ_1 is weakly identified. However, in the linear IV model with strongly identified θ_1 and weakly identified θ_2, the AR/QLR1 test has equal power to the Oracle C(α)-QLR1 test.

Overall, the AR/QLR1 test is found easily to be the best two-step test in terms of power in the over-identified models considered here and its power is noticeably higher than that of the Proj-QLR1 test. Given this, the remainder of the simulation results focus on the AR/QLR1 test.

The finite-sample NRP's of the AR/QLR1 test are simulated for a range of parameter configu-

rations and sample sizes. In the linear IV model, the maximum NRP's (over the 25 identification scenarios considered) of the AR/QLR1 test are in $[.049, .064]$ for $(n, k) = (100, 4)$, $(250, 4)$, $(500, 4)$, $(100, 8)$, $(250, 8)$, where k is the number of IV's. In the nonlinear IV model, they are in $[.040, .050]$ for the structural function and $[.039, .052]$ for its derivative for the same (n, k) values (with the maximum NRP's being over nine identification scenarios).

We carry out extensive simulations to determine the sensitivity of the AR/QLR1 test to tuning parameters. For some tuning parameters, there are theoretical reasons to expect little or no sensitivity and this is borne out in the simulations. For α_1, we find no sensitivity of the NRP's in both models (and both hypotheses in the nonlinear IV model) and some sensitivity of power. For a constant, K_{rk}, that appears in the rank statistic in the C(α)-QLR1 statistic, we find no sensitivity of the NRP's except some sensitivity in a couple of cases in the linear IV model. For power, we find some sensitivity to K_{rk} in both models, but not a lot. Overall, the base case values of $\alpha_1 = .005$ and $K_{rk} = 1$ (which are used for the power comparisons and the NRP calculations in both models and both hypotheses in the nonlinear model) perform well. These base case values also are used in the simulations for likelihood models in Andrews (2017) and perform well there.

The remainder of the paper is organized as follows. Section 2 discusses subvector methods in the literature. Section 3 introduces SE models, including the moment condition model. Section 4 introduces the two-step tests and CS's for SE models. Section 5 provides asymptotic size and strong-identification asymptotic efficiency results under high-level assumptions. Section 6 proves the asymptotic results in Section 5.

The rest of the paper focuses on the moment condition model. Section 7 introduces the two-step AR/AR, AR/LM, and AR/QLR1 tests and CS's for the moment condition model. Section 8 provides primitive conditions under which these tests have correct asymptotic size and the latter two are asymptotically efficient in a GMM sense under strong identification. The proofs of these results utilize results in Andrews and Guggenberger (2017). Section 9 provides the finite-sample simulation results. The Supplemental Material (SM) to this paper generalizes the results in Section 8 from i.i.d. observations to strictly stationary strong mixing time series observations, proves the results in Section 8, and provides some additional simulation results.

All limits are as $n \to \infty$ unless stated otherwise.

2 Subvector Methods in the Literature

In this section, we discuss existing subvector methods in the literature. Widely used general methods are the Bonferroni and Scheffé projection methods, e.g., see Loh (1985), Berger and Boos

(1994), Cavanagh, Elliott, and Stock (1995), Campbell and Yogo (2006), Chaudhuri, Richardson, Robins, and Zivot (2010), and Chaudhuri and Zivot (2011) for Bonferroni's method, and Dufour (1989) and Dufour and Jasiak (2001) for the projection method. These methods are asymptotically conservative, i.e., their asymptotic size is less than their nominal level. The degree of conservativeness typically is larger for the projection method. It depends on the dimension of the nuisance parameter and the shape of the power function of the joint test that is employed. A refinement of Bonferroni's method that is not conservative, but is much more intensive computationally, is provided by Cavanagh, Elliott, and Stock (1995). McCloskey (2011) also introduces a refinement of Bonferroni's method.

When the nuisance parameters that appear under the null hypothesis are known to be strongly identified, one can obtain identification-robust subvector tests by concentrating out these parameters or replacing them by $n^{1/2}$-consistent asymptotically normal estimators. This method is employed in Stock and Wright (2000), Kleibergen (2004, 2005), Guggenberger and Smith (2005), Otsu (2006), Montiel Olea (2012), Guggenberger, Ramalho, and Smith (2013), I. Andrews and Mikusheva (2015), and Andrews and Guggenberger (2015). This method yields non-conservative inference asymptotically and is asymptotically efficient under strong identification of all of the parameters (for suitable tests). The drawback of this method, however, is that the nuisance parameters cannot be weakly identified.

Andrews and Cheng (2012, 2013a,b), Cheng (2015), Cox (2016), and Han and McCloskey (2016) provide subvector tests with correct asymptotic size based on the asymptotic distributions of standard test statistics under the full range of possible identification scenarios. These subvector methods are not asymptotically conservative and are asymptotically efficient under strong identification of all of the parameters (for suitable tests). However, they require one to have knowledge of the source of the potential lack of identification (e.g., which subvectors play the roles of β, π, and ζ in the Andrews and Cheng (2012) notation) and require special structure of the model considered, such as having a known correspondence between strongly-identified reduced-form parameters and subsets of the structural parameters of interest in the case of Cox (2016).

Elliott, Müller, and Watson (2015) develop nearly optimal subvector tests when a nuisance parameter is present under the null hypothesis, which includes models with weak identification, as exemplified by their example of tests concerning the location of a change point when the magnitude of the change point is moderate. Their tests are nearly optimal in the sense of nearly achieving weighted average power for a given weight function.

Chen, Christensen, O'Hara, and Tamer (2016) provide subvector CS's for the identified set in partially identified models using Monte Carlo Markov chain methods in models where the parame-

ters of interest are functions of strongly-identified reduced-form parameters.

For minimum distance models, I. Andrews and Mikusheva (2016a) provide subvector inference using a geometric approach. This method has asymptotic size equal to its asymptotic nominal level and may or may not be asymptotically efficient under strong identification depending upon the model. For example, in the homoskedastic linear IV model, it does not yield asymptotic efficiency under strong identification, but in other models it does.

I. Andrews (2017) constructs a two-step confidence set for a parameter subvector in a GMM scenario based on identification-robust and standard (non-identifcation-robust) CS's and an identification-category selection method. The two-step CS yields the standard $1-\alpha$ CS with probability that goes to one under strong identification and the identification-robust $1-\alpha-\gamma$ CS otherwise. The asymptotic theory for the method is based on high-level assumptions.

Chaudhuri (2016) extends the subvector Bonferroni test in Chaudhuri and Zivot (2011) to the case of linear restrictions and provides a form of the C(α)-LM test that has some computational advantages.

Two recent papers develop methods for subvector inference in moment inequality and/or equality models with partial identification, see Bugni, Canay, and Shi (2016) and Kaido, Molinari, and Stöye (2016). These methods focus on the special difficulties associated with moment inequalities, but can be applied to the moment equality-type models considered in this paper. The proposed methods are non-conservative asymptotically, but do not yield asymptotic efficiency under strong identification.[2]

In the linear IV regression model with homoskedastic errors, subvector inference in which nuisance parameters are profiled out and the χ^2 degrees of freedom are reduced accordingly is possible using the Anderson-Rubin (AR) test, see Guggenberger, Kleibergen, Mavroeidis, and Chen (2012). This method yields asymptotic efficiency under strong identification if the model is exactly identified, but not if the model is over identified. For related results, see Lee (2014). Kleibergen (2015) also provides subvector methods for this model based on the likelihood ratio (LR) test.

3 System of Equations Model

Let $\{W_i \in R^m : i = 1, ..., n\}$ denote the observations with distribution F and let $\theta \in \Theta \subset R^p$ be an unknown parameter. The observations may be independent or temporally dependent. We partition θ as $\theta = (\theta_1', \theta_2')'$ for $\theta_j \in R^{p_j}$ for $j = 1, 2$, where $p_1 + p_2 = p$. This paper is concerned

[2] For models with only moment equalities, the BCS test statistic reduces to the AR statistic or an AR-like statistic based on a diagonal weight matrix. The KMS approach treats each moment equality as two inequalities and employs inf and sup statistics over the different inequalities.

with identification-robust tests of the subvector null hypothesis

$$H_0 : \theta_2 = \theta_{20} \text{ versus } H_1 : \theta_2 \neq \theta_{20}. \tag{3.1}$$

Under H_0 and H_1, θ_1 is a nuisance parameter. The paper also considers CS's for the subvector θ_2.

We consider a general class of models that we call SE models. These models depend on a sample vector $\widehat{g}_n(\theta) \in R^k$ for θ, whose population analogue, $g_F(\theta)$, satisfies

$$g_F(\theta) = 0^k \tag{3.2}$$

when θ is the true parameter value, where $0^k := (0, ..., 0)' \in R^k$. The function $g_F(\theta)$ may or may not equal 0^k for other values of θ depending on whether θ is identified or not. SE models also depend on a consistent estimator $\widehat{\Omega}_n(\theta)$ of the asymptotic variance of $\widehat{g}_n(\theta)$ (after suitable normalization).

Examples of SE models include moment condition models with $\widehat{g}_n(\theta)$ being a sample moment vector:

$$\widehat{g}_n(\theta) := n^{-1} \sum_{i=1}^{n} g_i(\theta), \text{ where } g_i(\theta) := g(W_i, \theta). \tag{3.3}$$

In moment condition models, $g_F(\theta) := E_F g(W_i, \theta) = 0^k$ when θ is the true value and $E_F g(W_i, \theta)$ may or may not equal 0^k otherwise, depending on whether θ is identified.

Likelihood-based models, which we refer to as ML models, are SE models. For ML models, one has a log-likelihood function (divided by n), $\widehat{m}_n(\theta)$, and $\widehat{g}_n(\theta)$ is the score function:

$$\widehat{m}_n(\theta) := n^{-1} \sum_{i=1}^{n} m_i(\theta) \text{ and } \widehat{g}_n(\theta) := \frac{\partial}{\partial \theta} \widehat{m}_n(\theta) = n^{-1} \sum_{i=1}^{n} g_i(\theta), \text{ where } g_i(\theta) := \frac{\partial}{\partial \theta} m_i(\theta), \tag{3.4}$$

$m_i(\theta)$ is the log-likelihood function for the ith observation W_i (conditional on previous observations in time series settings), or $m_i(\theta)$ is the conditional log-likelihood function for Y_i given some covariates X_i when $W_i := (Y_i', X_i')'$. In i.i.d. scenarios, $g_i(\theta) := g(W_i, \theta)$ for some function $g(\cdot, \cdot)$. In ML models, $g_F(\theta) := E_F(\partial/\partial\theta) m_i(\theta)$ and $k = p$. Other models fit into the "sample average" SE framework of (3.4) when $m_i(\theta)$ is a function, such as a least squares or quasi-log-likelihood function, that differs from a log-likelihood function.

Minimum distance models are SE models with $\widehat{g}_n(\theta)$ taking the form

$$\widehat{g}_n(\theta) := \widehat{\pi}_n - g(\theta) \tag{3.5}$$

for some estimator $\widehat{\pi}_n$ of a parameter π and some (known) k-vector of restrictions, $g(\theta)$, on π. The

restrictions on the true values π_F and θ under F are $\pi_F = g(\theta)$. In minimum distance models, $g_F(\theta) := \pi_F - g(\theta)$.

In addition, moment condition, ML, and minimum distance models for which $\widehat{g}_n(\theta)$ depends on a preliminary $n^{1/2}$-consistent estimator, say $\widehat{\gamma}_n$, also are SE models. In these cases, $\widehat{g}_n(\theta) := \widehat{g}_n(\theta, \widehat{\gamma}_n)$.

In the moment condition and ML models, for the case of i.i.d. observations, the estimator $\widehat{\Omega}_n(\theta)$ of the asymptotic variance of $\widehat{g}_n(\theta)$ is given by

$$\widehat{\Omega}_n(\theta) := n^{-1} \sum_{i=1}^{n} (g_i(\theta) - \widehat{g}_n(\theta))(g_i(\theta) - \widehat{g}_n(\theta))' \in R^{k \times k}. \tag{3.6}$$

With time series observations, $\widehat{\Omega}_n(\theta)$ typically needs to be defined differently to account for temporal dependence. For minimum distance models, $\widehat{\Omega}_n(\theta)$ is a consistent estimator of the asymptotic variance of $\widehat{\pi}_n$ (after suitable normalization) and does not depend on θ. In models with preliminary estimators $\widehat{\gamma}_n$, $\widehat{\Omega}_n(\theta)$ needs to be defined to take into account the effect of $\widehat{\gamma}_n$ on the asymptotic variance of $\widehat{g}_n(\theta)$.

The parameter space for θ is $\Theta \subset R^p$. Let Θ_1 denote the null nuisance parameter space:

$$\Theta_1 := \{\theta_1 : \theta = (\theta_1', \theta_{20}')' \in \Theta\}. \tag{3.7}$$

The null parameter space for the pairs (θ_1, F) is denoted by $\mathcal{F}_{SV}$, where SV denotes subvector. When the null hypothesis is true, i.e., $\theta_2 = \theta_{20}$, all such pairs satisfy $g_F(\theta_1, \theta_{20}) = 0^k$ and have $\theta_1 \in \Theta_1$.

When considering CS's for θ_2, the parameter space for (θ, F) is denoted by $\mathcal{F}_{\Theta,SV}$. In this case, we make the dependence of $\mathcal{F}_{SV}$ on the null hypothesis value θ_{20} explicit: $\mathcal{F}_{SV} = \mathcal{F}_{SV}(\theta_{20})$. Let Θ_{2*} denote the set of possible true θ_2 parameter values. We assume that $\Theta_{2*} \subset \Theta_2 := \{\theta_2 : \exists \theta_1$ such that $(\theta_1', \theta_2')' \in \Theta\}$. By definition,

$$\mathcal{F}_{\Theta,SV} := \{(\theta, F) : \theta = (\theta_1', \theta_2')' \in \Theta \text{ such that } (\theta_1, F) \in \mathcal{F}_{SV}(\theta_2) \text{ and } \theta_2 \in \Theta_{2*}\}. \tag{3.8}$$

In SE models, the sample Jacobian is

$$\begin{aligned} \widehat{G}_n(\theta) &:= [\widehat{G}_{1n}(\theta) : \widehat{G}_{2n}(\theta)] \in R^{k \times p}, \text{ where} \\ \widehat{G}_{jn}(\theta) &:= \frac{\partial}{\partial \theta_j'} \widehat{g}_n(\theta) \in R^{k \times p_j} \text{ for } j = 1, 2. \end{aligned} \tag{3.9}$$

Let $\{\theta_{*n} : n \geq 1\}$ be the sequence of true values of θ. We write $\theta_{*n} = (\theta_{1*n}', \theta_{2*n}')'$, where $\theta_{1*n} \in R^{p_1}$ and $\theta_{2*n} \in R^{p_2}$.

For notational simplicity, when considering a test of $H_0 : \theta_2 = \theta_{20}$, we write any function of θ that is evaluated at $\theta_2 = \theta_{20}$ as a function of θ_1 only. For example, $g_i(\theta_1)$ denotes $g_i(\theta_1, \theta_{20})$. When considering a CS for θ_2, uniform asymptotic results require that we consider true values of θ_2 that may depend on n, i.e., $\theta_2 = \theta_{2*n}$. In this case, we write any function of θ that is evaluated at $\theta_2 = \theta_{2*n}$ as a function of θ_1 only.

The high-level results given in Section 5 below apply to the class of SE models. In this paper, we verify the high-level conditions for three two-step subvector tests for moment condition models. In Andrews (2017), we verify them for two two-step subvector tests for ML models.

4 Two-Step Subvector Tests and Confidence Sets

This section provides a general definition of two-step tests of $H_0 : \theta_2 = \theta_{20}$ with nominal level $\alpha \in (0, 1)$ for SE models. Two-step CS's for θ_2 are obtained by inverting the tests. Section 7 below provides detailed descriptions of three two-step tests and CS's in the moment condition model.

The first-step CS CS_{1n}, estimator set $\widehat{\Theta}_{1n}$, and second-step data-dependent significance level $\widehat{\alpha}_{2n}(\theta_1)$ are as described in the Introduction. We define

$$CS_{1n}^{+} = CS_{1n} \cup \widehat{\Theta}_{1n}. \tag{4.1}$$

We denote the second-step nominal level η identification-robust C(α)-test for given θ_1 by $\phi_{2n}(\theta_1, \eta)$, where the test rejects $H_0 : \theta_2 = \theta_{20}$ when $\phi_{2n}(\theta_1, \eta) > 0$. That is, $\phi_{2n}(\theta_1, \eta)$ is the difference between a test statistic and its (possibly data-dependent) critical value. We suppress the dependence of $\phi_{2n}(\theta_1, \eta)$ and $\widehat{\alpha}_{2n}(\theta_1)$ on θ_{20}.

The two-step subvector test with nominal level α is denoted by φ_{2n}^{SV}. It rejects $H_0 : \theta_2 = \theta_{20}$ if $\phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1))$ rejects H_0 for all $\theta_1 \in CS_{1n}^{+}$ and it rejects H_0 if $CS_{1n}^{+} = \varnothing$. That is, the subvector test rejects $H_0 : \theta_2 = \theta_{20}$ if

$$\varphi_{2n}^{SV} := \inf_{\theta_1 \in CS_{1n}^{+}} \phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1)) > 0, \tag{4.2}$$

where the inf over $\theta_1 \in \varnothing$ is defined to equal ∞.

The nominal level α *oracle subvector* test of $H_0 : \theta_2 = \theta_{20}$ is

$$\phi_{2n}(\theta_{1*n}, \alpha), \tag{4.3}$$

where θ_{1*n} is the true value of θ_1. This test is infeasible. Nevertheless, we show that the two-step test

is asymptotically equivalent to the oracle subvector test under most strongly-identified sequences of distributions—both null sequences and sequences that are contiguous to the null. Hence, the two-step test inherits the same asymptotic local power properties as the oracle subvector test for such sequences.

The subvector test described above is similar to the test of Chaudhuri and Zivot (2011), but differs in three ways. First, it employs an estimator set $\widehat{\Theta}_{1n}$ that guarantees that there is an element of CS_{1n}^{+} that is close to the true nuisance parameter θ_{1*n} wp$\to$1 under strongly-identified sequences. Second, it employs a data-dependent second-step significance level $\widehat{\alpha}_{2n}(\theta_1)$ that guarantees that the nominal level of the second-step test $\phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1))$ equals α wp$\to$1 under θ_1-strongly identified sequences. Third, it may differ in its choice of first-step CS and/or second-step test.

To define the two-step CS for θ_2, we make the dependence of the components of the two-step test on the null value θ_{20} explicit and write: $CS_{1n}(\theta_{20})$, $\widehat{\Theta}_{1n}(\theta_{20})$, $CS_{1n}^{+}(\theta_{20})$, $\widehat{\alpha}_{2n}(\theta_1, \theta_{20})$, $\phi_{2n}(\theta_1, \theta_{20}, \eta)$, and $\varphi_{2n}^{SV}(\theta_{20})$ for the quantities defined above. The two-step CS for θ_2 is

$$CS_{2n}^{SV} := \{\theta_2 \in \Theta_2 : \varphi_{2n}^{SV}(\theta_2) \leq 0\}, \text{ where } \Theta_2 := \{\theta_2 : \exists \theta_1 \text{ such that } (\theta_1', \theta_2')' \in \Theta\}. \tag{4.4}$$

5 Asymptotic Results under High-Level Conditions

The results in this section are based on high-level assumptions that are designed to apply to a broad set of SE models. The results can be applied to a variety of first-step CS's CS_{1n}, estimator sets $\widehat{\Theta}_{1n}$, second-step tests $\phi_{2n}(\theta_1, \eta)$, and second-step data-dependent significance levels $\widehat{\alpha}_{2n}(\theta_1)$.

For $\theta_1 \in \Theta_1$, let $B(\theta_1, r)$ denote a closed ball in Θ_1 centered at θ_1 with radius $r > 0$. For $\theta_1 \in R^{p_1}$ and $A_1 \subset R^{p_1}$, let

$$d(\theta_1, A_1) := \inf\{||\theta_a - \theta_1|| : \theta_a \in A_1\} \text{ and } d_H(\theta_1, A_1) := \sup\{||\theta_a - \theta_1|| : \theta_a \in A_1\} \tag{5.1}$$

when $A_1 \neq \varnothing$, and $d(\theta_1, A_1) := d_H(\theta_1, A_1) := \infty$ when $A_1 = \varnothing$. Note that $d_H(\theta_1, A_1)$ is the Hausdorff distance between $\{\theta_1\}$ and A_1.

Let F_n denote the true distribution F when the sample size is n. Let α, α_1, and α_2 be defined as above. That is, $\alpha \in (0, 1)$, $\alpha_1, \alpha_2 > 0$, and $\alpha_1 + \alpha_2 = \alpha$. Let df abbreviate "distribution function."

When testing $H_0 : \theta_2 = \theta_{20}$, let a null sequence be denoted by

$$S := \{(\theta_{*n}, F_n) : (\theta_{1*n}, F_n) \in \mathcal{F}_{SV}, \theta_{2*n} = \theta_{20}, n \geq 1\}, \tag{5.2}$$

where $\mathcal{F}_{SV}$ is the null parameter space for (θ_1, F). Let $\{m_n\}$ denote a subsequence of $\{n\}$. Let

S_m denote the subsequence of S determined by $\{m_n\}$, i.e., $S_m := \{(\theta_{*m_n}, F_{m_n}) : (\theta_{1*m_n}, F_{m_n}) \in \mathcal{F}_{SV}, \theta_{2*m_n} = \theta_{20}, n \geq 1\}$. An alternative sequence $S^A := \{(\theta^A_{*n}, F^A_n) : n \geq 1\}$ is a sequence for which $\theta^A_{*n} = (\theta^{A\prime}_{1*n}, \theta^{A\prime}_{2*n})' \in \Theta$, $\theta^A_{2*n} \neq \theta_{20}$, and (3.2) holds with $(\theta, F) = (\theta^A_{*n}, F^A_n)$, $\forall n \geq 1$.

Given a null sequence S, we define two alternative conditions on the components of the two-step test, i.e., on CS_{1n}, $\widehat{\Theta}_{1n}$, $\phi_{2n}(\theta_1, \eta)$, and $\widehat{\alpha}_{2n}(\theta_1)$. For null sequences S for which Assumption B holds, we bound the asymptotic NRP's of the subvector test by α using a Bonferroni (B) argument. For sequences for which Assumption C holds, we bound the asymptotic NRP's of the subvector test by α using a Neyman C(α)-based (C) argument.

Assumption B. For the null sequence S,

(i) CS_{1n} has asymptotic coverage probability $1 - \alpha_1$ or greater,

(ii) $\phi_{2n}(\theta_{1*n}, \alpha_2)$ has asymptotic NRP α_2 or less, and

(iii) $\widehat{\alpha}_{2n}(\theta_{1*n}) = \alpha_2$ wp$\to 1$.[3]

Assumption C. For the null sequence S,

(i) $d(\theta_{1*n}, CS^+_{1n}) = O_p(n^{-1/2})$,

(ii) $\phi_{2n}(\theta_{1*n}, \alpha)$ has asymptotic NRP equal to α,

(iii) $\phi_{2n}(\theta_{1*n}, \alpha)$ has an asymptotic distribution whose df is continuous at 0,

(iv) $\phi_{2n}(\theta_1, \eta)$ is nondecreasing in η on $[\alpha_2, \alpha]$ $\forall \theta_1 \in \Theta_1$, and

(v) $\sup_{\theta_1 \in B(\theta_{1*n}, K/n^{1/2})} |\phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| = o_p(1)$ $\forall K \in (0, \infty)$.

For a null subsequence S_m, we define Assumptions B and C analogously with m_n in place of n throughout.

Depending on the second-step test $\phi_{2n}(\theta_1, \eta)$, Assumption C is employed in scenarios in which θ_1 is (locally) strongly identified given θ_{20} (e.g., with the second-step C(α)-AR test), or in scenarios in which θ is (locally) strongly identified (e.g., with the second-step C(α)-LM and C(α)-QLR1 tests). Assumption B is employed in other scenarios and in some scenarios in which Assumption C is employed.

Assumption B(i) requires that CS_{1n} is an identification-robust CS for θ_1 given the true value θ_{20}. Assumptions B(ii) and C(ii) require that $\phi_{2n}(\theta_{1*n}, \eta)$ is an identification-robust test of $H_0 : \theta_2 = \theta_{20}$ given the true value θ_{1*n} for $\eta = \alpha_2$ and α. Assumption B(iii) requires that the data-dependent significance level $\widehat{\alpha}_{2n}(\theta_{1*n})$ is small (i.e., equal to α_2) wp$\to$1 in the scenarios for which Assumption B is applied.

Assumption C(i) requires that the true value θ_{1*n} is close to CS^+_{1n} in strongly-identified scenarios. Given the definition of $d(\theta_1, A_1)$, Assumption C(i) requires that CS^+_{1n} is not empty wp$\to$1. Note

[3] More precisely, Assumptions B(i) and B(ii) mean that (i) $\liminf_{n\to\infty} P_{\theta_{*n}, F_n}(\theta_{1*n} \in CS_{1n}) \geq 1 - \alpha_1$ and (ii) $\limsup_{n\to\infty} P_{\theta_{*n}, F_n}(\phi_{2n}(\theta_{1*n}, \alpha_2) > 0) \leq \alpha_2$.

that $d(\theta_{1*n}, \widehat{\Theta}_{1n}) = O_p(n^{-1/2})$ is sufficient for Assumption C(i) and showing this is how Assumption C(i) is verified when CS_{1n} is the AR CS. Assumptions C(iii) and C(iv) are mild conditions. Assumption C(v) typically holds for a test statistic $\phi_{2n}(\theta_1, \eta)$ only if it has been orthogonalized wrt to θ_1 in the Neyman C(α)-type fashion.

The following assumption uses Assumptions B and C. Under this assumption the nominal level α two-step subvector test specified above has correct asymptotic level (CAL), i.e., its asymptotic size equals α or less.

Assumption CAL. For any null sequence S and any subsequence $\{w_n\}$ of $\{n\}$, there exists a subsubsequence $\{m_n\}$ such that S_m satisfies Assumption B or C.

Verifying Assumptions B and C for a selected subsequence $\{m_n\}$, as is required by Assumption CAL, is much easier than verifying it for an arbitrary sequence because one can choose the subsequence to be one for which the limits of various population quantities of interest exist.

The sequential process of specifying CS_{1n}, $\widehat{\Theta}_{1n}$, $\phi_{2n}(\theta_1, \eta)$, and $\widehat{\alpha}_{2n}(\theta_1)$ such that Assumption CAL holds for selected subsequences $\{m_n\}$ is as follows: (i) one selects CS_{1n} and $\phi_{2n}(\theta_1, \eta)$, (ii) given CS_{1n}, one specifies $\widehat{\Theta}_{1n}$ such that Assumption C(i) holds for a broad set of selected subsequences, (iii) given CS_{1n}, $\widehat{\Theta}_{1n}$, and $\phi_{2n}(\theta_1, \eta)$, one determines as large a set of selected subsequences such that Assumption C holds, and (iv) one applies Assumption B to all of the remaining selected subsequences and one specifies $\widehat{\alpha}_{2n}(\theta_1)$ such that Assumption B(iii) holds for each of these subsequences. In step (i), the choice of CS_{1n} does not depend on $\phi_{2n}(\theta_1, \eta)$ and vice versa.

Under the next assumption, for a given null subsequence S_m, the subvector test is asymptotically equivalent to the oracle subvector test and has asymptotic NRP equal to α. In consequence, the test has asymptotic size α (not less than α) and is not asymptotically conservative.

Assumption OE. For some null sequence S that satisfies Assumption C,

(i) $d_H(\theta_{1*n}, CS_{1n}^+) = O_p(n^{-1/2})$ and

(ii) $\widehat{\alpha}_{2n}(\theta_1) = \alpha \ \forall \theta_1 \in B(\theta_{1*n}, K/n^{1/2})$ wp$\to$1, $\forall K \in (0, \infty)$

or, for some null subsequence S_m that satisfies Assumption C, the subsequence versions of OE(i) and (ii) hold.

Note that OE abbreviates "oracle equivalence." Assumption OE(i) guarantees that the first-step CS for θ_1 shrinks to θ_{1*n} as $n \to \infty$ and Assumption OE(ii) guarantees that the critical value embodied in the φ_{2n}^{SV} test is α, not less than α, wp$\to$1, for subsequences S_m that satisfy Assumption OE. Whether Assumption OE(i) holds depends on the strength of identification of θ_1, but not θ_2. Assumption OE(i) holds if it holds both with CS_{1n} in place of CS_{1n}^+ and with $\widehat{\Theta}_{1n}$ in place of CS_{1n}^+.

Assumptions B(iii) and OE(ii) are incompatible. Hence, sequences S or subsequences S_m that satisfy one cannot satisfy the other.

Let $AsySz$ denote the asymptotic size of the subvector test φ_{2n}^{SV}. That is,

$$AsySz := \limsup_{n\to\infty} \sup_{(\theta_1,F)\in\mathcal{F}_{SV}} P_{\theta_1,\theta_{20},F}(\varphi_{2n}^{SV} > 0). \tag{5.3}$$

Let $AsyNRP$ denote the asymptotic NRP of the subvector test φ_{2n}^{SV} under a sequence S or subsequence S_m. That is, for a sequence S,

$$AsyNRP := \lim_{n\to\infty} P_{\theta_{1*n},\theta_{20},F_n}(\varphi_{2n}^{SV} > 0), \tag{5.4}$$

where $(\theta_{1*n}, F_n) \in \mathcal{F}_{SV}$, provided this limit exists.

When considering a CS for the subvector θ_2, we define sequences S and subsequences S_m as in (5.2), but with θ_{20} replaced by some $\theta_{2*n} \in \Theta_{2*}$ for $n \geq 1$. Given these definitions, Assumptions B, C, CAL, and OE are defined for CS's just as they are defined for tests. For a CS obtained by inverting a subvector test $\varphi_{2n}^{SV} = \varphi_{2n}^{SV}(\theta_{20})$ (of $H_0 : \theta_2 = \theta_{20}$), asymptotic size is defined by

$$AsySz := 1 - \limsup_{n\to\infty} \sup_{\theta_2\in\Theta_{2*}} \sup_{(\theta_1,F)\in\mathcal{F}_{SV}(\theta_2)} P_{\theta_1,\theta_2,F}(\varphi_{2n}^{SV}(\theta_2) > 0). \tag{5.5}$$

The asymptotic coverage probability of a CS under a sequence S, denoted by $AsyCP$, is

$$AsyCP := 1 - \lim_{n\to\infty} P_{\theta_{1*n},\theta_{2*n},F_n}(\varphi_{2n}^{SV}(\theta_{2*n}) > 0), \tag{5.6}$$

where $(\theta_{1*n}, F_n) \in \mathcal{F}_{SV}(\theta_{2*n})$ and $\theta_{2*n} \in \Theta_{2*}$ for $n \geq 1$, provided this limit exists.

The main result of the paper based on high-level conditions is the following.

Theorem 5.1 *For the parameter space $\mathcal{F}_{SV}$, the nominal level α two-step subvector test φ_{2n}^{SV} satisfies*

(a) *$AsySz \leq \alpha$ under Assumption* CAL,

(b) *$AsySz = \alpha$ under Assumptions* CAL *and* OE,

(c) *$AsyNRP = \alpha$ for all null sequences S for which Assumption* OE *holds,*

(d) *for any null sequence S for which Assumption* OE *holds,*

$\varphi_{2n}^{SV} = \phi_{2n}(\theta_{1*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n},F_n}(\varphi_{2n}^{SV} > 0) = \lim P_{\theta_{*n},F_n}(\phi_{2n}(\theta_{1*n}, \alpha) > 0)$,

(e) *for any alternative sequence S^A that satisfies Assumption* C(iii) *and is contiguous to a null sequence S that satisfies Assumption* OE, $\varphi_{2n}^{SV} = \phi_{2n}(\theta_{1*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}^A,F_n^A}(\varphi_{2n}^{SV} > 0) = \lim P_{\theta_{*n}^A,F_n^A}(\phi_{2n}(\theta_{1*n}, \alpha) > 0)$, *and*

(f) *for the parameter space* $\mathcal{F}_{\Theta,SV}$, *the nominal level* $1-\alpha$ *two-step subvector CS* $\varphi_{2n}^{SV}(\cdot)$ *satisfies* (i) $AsySz \geq 1-\alpha$ *under Assumption* CAL, (ii) $AsySz = 1-\alpha$ *under Assumptions* CAL *and* OE, (iii) $AsyCP = 1-\alpha$ *for all sequences* S *for which Assumption* OE *holds,* (iv) *for any sequence* S *for which Assumption* OE *holds,* $\varphi_{2n}^{SV}(\theta_{2*n}) = \phi_{2n}(\theta_{1*n}, \theta_{2*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n},F_n}(\varphi_{2n}^{SV}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n},F_n}(\phi_{2n}(\theta_{1*n}, \theta_{2*n}, \alpha) > 0)$, *and* (v) *for any alternative sequence* S^A *that satisfies Assumption* C(iii) *and is contiguous to a null sequence* S *that satisfies Assumption* OE, $\varphi_{2n}^{SV}(\theta_{2*n}) = \phi_{2n}(\theta_{1*n}, \theta_{2*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}^A,F_n^A}(\varphi_{2n}^{SV}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n}^A,F_n^A}(\phi_{2n}(\theta_{1*n}, \theta_{2*n}, \alpha) > 0)$.

Comments: (i). In words, Theorem 5.1(a) states that the nominal level α subvector test φ_{2n}^{SV} has correct asymptotic level α (i.e., its asymptotic size is α or less). Theorem 5.1(b) states that it has asymptotic size equal to its nominal level α. Theorem 5.1(c) states that φ_{2n}^{SV} has $AsyNRP$ equal to its nominal level α for certain sequences S. Theorem 5.1(d) and (e) state that φ_{2n}^{SV} is asymptotically equivalent to the oracle subvector test $\phi_{2n}(\theta_{1*n}, \alpha)$ under certain null and contiguous alternative sequences, S and S^A. Theorem 5.1(f) provides analogous results for two-step CS's for θ_2.

(ii). Theorem 5.1(d) and (e) provide an asymptotic efficiency result for the subvector test φ_{2n}^{SV} if the oracle test $\phi_{2n}(\theta_{1*n}, \alpha)$ is asymptotically equivalent to an asymptotically efficient test under the contiguous alternative sequence S^A. More specifically, if θ_2 is strongly identified given $\theta_1 = \theta_{1*n}^A$ under S^A, then the standard LM and Wald tests are asymptotically efficient in a GMM or ML sense (depending on the type of model considered), see Newey and West (1987) for GMM models. Hence, if the oracle test $\phi_{2n}(\theta_{1*n}, \alpha)$ is asymptotically equivalent to these tests under S^A, then it inherits their asymptotic efficiency properties.

(iii). Theorem 5.1(a) is established by showing that the two-step test has asymptotic NRP's equal to α or less for suitable sequences S (and subsequences S_m). To show this for a given sequence S, one uses Assumption B or C depending on the strength of identification local to $(\theta_{1*n}, \theta_{20})$. Depending on the second-step test being considered, the "strength of identification" may refer to the strength of identification of θ_1 (given θ_{20}) or θ.

On the other hand, to verify Assumption OE(i) for some sequence S (or subsequence S_m) one needs global strong identification of θ_1 over Θ_1. By the latter, we mean a global separation between the value of a suitable population criterion function at θ_{1*n} and its value at $\theta_1 \neq \theta_{1*n}$ (when $\theta_2 = \theta_{20}$). Hence, the results of Theorem 5.1(b) and (c) only hold if one has global strong identification of θ_1 over Θ_1 in this sense for some sequence S (or subsequence S_m).

(iv). The results of Theorem 5.1(c)–(e) also apply to subsequences S_m and S_m^A.

(v). The proof of Theorem 5.1(f) is a minor variant of the proof of Theorem 5.1(a)–(e). The only difference is that θ_{20} is replaced by θ_{2*n}, which can depend on n.

6 Proof of Theorem 5.1

Proof of Theorem 5.1. We prove part (a) first. We show below that for any null sequence S and any subsequence $\{w_n\}$ of $\{n\}$, there exists a subsubsequence $\{m_n\}$ such that, under $S_m := \{(\theta_{*m_n}, F_{m_n}) : (\theta_{1*m_n}, F_{m_n}) \in \mathcal{F}_{SV}, \theta_{2*m_n} = \theta_{20}, n \geq 1\}$, $\varphi^{SV}_{2m_n}$ satisfies

$$\limsup_{n\to\infty} P_{\theta_{*m_n},F_{m_n}}(\varphi^{SV}_{2m_n} > 0) \leq \alpha, \tag{6.1}$$

where $\theta_{*m_n} = (\theta'_{1*m_n}, \theta'_{20})'$.

To show $AsySz \leq \alpha$, let S be a null sequence such that $\limsup_{n\to\infty} P_{\theta_{*n}F_n}(\varphi^{SV}_{2n} > 0) = \limsup_{n\to\infty} \sup_{(\theta_1,F)\in\mathcal{F}_{SV}} P_{\theta_1,\theta_{20},F}(\varphi^{SV}_{2n} > 0)$ $(:= AsySz)$, where $\theta_{*n} = (\theta'_{1*n}, \theta'_{20})'$. Such a sequence always exists. Let $\{w_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim P_{\theta_{*w_n}F_{w_n}}(\varphi^{SV}_{2w_n} > 0)$ exists and equals $AsySz$. Such a sequence always exists. By the result stated in the previous paragraph, there exists a subsubsequence $\{m_n\}$ of $\{w_n\}$ such that (6.1) holds. Thus, we have

$$AsySz = \lim P_{\theta_{*w_n},F_{w_n}}(\varphi^{SV}_{2w_n} > 0) = \limsup_{n\to\infty} P_{\theta_{*m_n},F_{m_n}}(\varphi^{SV}_{2m_n} > 0) \leq \alpha, \tag{6.2}$$

where the second equality holds because the limit of any subsequence of a convergent sequence is the same as the limit of the original sequence.

Now we establish (6.1). By Assumption CAL, for any null sequence S and any subsequence $\{w_n\}$ of $\{n\}$, there exists a subsubsequence $\{m_n\}$ such that S_m satisfies Assumption B or C. First, suppose Assumption B holds. With n in place of m_n for notational simplicity, we have

$$\begin{aligned}
& P_{\theta_{*n},F_n}(\varphi^{SV}_{2n} > 0) \\
:= \; & P_{\theta_{*n},F_n}\left(\inf_{\theta_1\in CS^+_{1n}} \phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1)) > 0\right) \\
\leq \; & P_{\theta_{*n},F_n}\left(\inf_{\theta_1\in CS^+_{1n}} \phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1)) > 0,\ \theta_{1*n} \in CS_{1n}\right) + P_{\theta_{*n},F_n}(\theta_{1*n} \notin CS_{1n}) \\
\leq \; & P_{\theta_{*n},F_n}\left(\phi_{2n}(\theta_{1*n}, \widehat{\alpha}_{2n}(\theta_{1*n})) > 0\right) + \alpha_1 + o(1) \\
= \; & P_{\theta_{*n},F_n}\left(\phi_{2n}(\theta_{1*n}, \alpha_2) > 0\right) + \alpha_1 + o(1) \\
\leq \; & \alpha_2 + \alpha_1 + o(1) \\
= \; & \alpha + o(1),
\end{aligned} \tag{6.3}$$

where the second inequality holds using Assumption B(i) and the fact that $CS_{1n} \subset CS^+_{1n}$ by definition, the second last equality holds by Assumption B(iii), the last inequality holds by Assumption B(ii), and the last equality holds by the definition of α_1 and α_2. The inequalities in (6.3) are just

the standard inequalities in the Bonferroni argument. With m_n in place of n, (6.3) establishes (6.1) under Assumption B.

Second, suppose Assumption C holds. That is, for any null sequence S and any subsequence $\{w_n\}$ of $\{n\}$, consider a subsubsequence $\{m_n\}$ such that Assumption C holds under S_m. Let $\widehat{\theta}_{1m_n}$ be an element of $CS^+_{1m_n}$ that satisfies $||\widehat{\theta}_{1m_n} - \theta_{1*m_n}|| = O_p(m_n^{-1/2})$. Such a value $\widehat{\theta}_{1m_n}$ exists wp$\to$1 by Assumption C(i). With n in place of m_n for notational simplicity, we have

$$\begin{aligned} P_{\theta_{*n},F_n}(\varphi^{SV}_{2n} > 0) &:= P_{\theta_{*n},F_n}\left(\inf_{\theta_1 \in CS^+_{1n}} \phi_{2n}(\theta_1, \widehat{\alpha}_{2n}(\theta_1)) > 0\right) \\ &\leq P_{\theta_{*n},F_n}\left(\inf_{\theta_1 \in CS^+_{1n}} \phi_{2n}(\theta_1, \alpha) > 0\right) \\ &\leq P_{\theta_{*n},F_n}\left(\phi_{2n}(\widehat{\theta}_{1n}, \alpha) > 0\right) + o(1), \end{aligned} \tag{6.4}$$

where the first inequality holds by Assumption C(iv) and the second inequality holds because $\widehat{\theta}_{1n} \in CS^+_{1n}$ wp$\to$1.

Next, we show

$$\phi_{2m_n}(\widehat{\theta}_{1m_n}, \alpha) = \phi_{2m_n}(\theta_{1*m_n}, \alpha) + o_p(1). \tag{6.5}$$

Again with n in place of m_n for notational simplicity, we have: for all $\varepsilon, \delta > 0$,

$$\begin{aligned} & P_{\theta_{*n},F_n}\left(|\phi_{2n}(\widehat{\theta}_{1n}, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon\right) \\ \leq\ & P_{\theta_{*n},F_n}\left(|\phi_{2n}(\widehat{\theta}_{1n}, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon,\ n^{1/2}||\widehat{\theta}_{1n} - \theta_{1*n}|| \leq K\right) \\ & + P_{\theta_{*n},F_n}\left(n^{1/2}||\widehat{\theta}_{1n} - \theta_{1*n}|| > K\right) \\ \leq\ & P_{\theta_{*n},F_n}\left(\sup_{\theta_1 \in \Theta_1 : n^{1/2}||\theta_1 - \theta_{1*n}|| \leq K} |\phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon\right) + \delta \\ =\ & o(1) + \delta, \end{aligned} \tag{6.6}$$

where the second inequality holds for K $(= K_\delta)$ sufficiently large and n sufficiently large using the definition of $\widehat{\theta}_{1n}$ and the equality holds by Assumption C(v). Since $\delta > 0$ is arbitrary, this establishes (6.5).

Equation (6.5) and Assumption C(iii) imply that $\phi_{2m_n}(\widehat{\theta}_{1m_n}, \alpha)$ has the same asymptotic distribution as $\phi_{2m_n}(\theta_{1*m_n}, \alpha)$ under S_m, which is absolutely continuous at 0. Hence,

$$\lim P_{\theta_{*m_n},F_{m_n}}(\phi_{2m_n}(\widehat{\theta}_{1m_n}, \alpha) > 0) = \lim P_{\theta_{*m_n},F_{m_n}}(\phi_{2m_n}(\theta_{1*m_n}, \alpha) > 0) = \alpha, \tag{6.7}$$

where the last equality holds by Assumption C(ii). This result and (6.4) (with m_n in place of n)

establish (6.1) under Assumption C. This completes the proof of part (a) of Theorem 5.1.

Now we prove part (b). Given the result of part (a), it suffices to show that there exists a subsequence S_m under which $\lim_{n\to\infty} P_{\theta_{*m_n},F_{m_n}}(\varphi^{SV}_{2m_n} > 0) = \alpha$. We show below that for the subsequence S_m specified in Assumption OE we have

$$\varphi^{SV}_{2m_n} = \phi_{2m_n}(\theta_{1*m_n}, \alpha) + o_p(1). \tag{6.8}$$

This and Assumptions C(ii) and C(iii) give

$$\lim P_{\theta_{*m_n},F_{m_n}}(\varphi^{SV}_{2m_n} > 0) = \lim P_{\theta_{*m_n},F_{m_n}}(\phi_{2m_n}(\theta_{1*m_n}, \alpha) > 0) = \alpha. \tag{6.9}$$

For part (b), it remains to show (6.8). For notational simplicity, we use n in place of m_n from here on. Define $\widehat{\xi}_n := n^{1/2} d_H(\theta_{1*n}, CS^+_{1n})$. We have $\widehat{\xi}_n = O_p(1)$ by Assumption OE(i). Also, $CS^+_{1n} \neq \varnothing$ wp$\to$1 by Assumption OE(i). Given this, there is no loss in generality, and a gain in simplicity of the expressions, in assuming $CS^+_{1n} \neq \varnothing$ in the following calculations. By the definition of d_H, $\theta_1 \in CS^+_{1n}$ implies $||\theta_1 - \theta_{1*n}|| \leq d_H(\theta_{1*n}, CS^+_{1n})$ and $n^{1/2}||\theta_1 - \theta_{1*n}|| \leq \widehat{\xi}_n$. We use this in the following: for all $\varepsilon > 0$,

$$\begin{aligned}
&P_{\theta_{*n},F_n}(|\varphi^{SV}_{2n} - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon) \\
&= P_{\theta_{*n},F_n}\left(\left|\inf_{\theta_1 \in CS^+_{1n}} \phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)\right| > \varepsilon\right) \\
&\leq P_{\theta_{*n},F_n}\left(\sup_{\theta_1 \in CS^+_{1n}} |\phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon\right) \\
&\leq P_{\theta_{*n},F_n}\left(\sup_{\theta_1 \in \Theta_1 : n^{1/2}||\theta_1 - \theta_{1*n}|| \leq \widehat{\xi}_n} |\phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon\right) \\
&\leq P_{\theta_{*n},F_n}\left(\sup_{\theta_1 \in \Theta_1 : n^{1/2}||\theta_1 - \theta_{1*n}|| \leq K} |\phi_{2n}(\theta_1, \alpha) - \phi_{2n}(\theta_{1*n}, \alpha)| > \varepsilon,\ \widehat{\xi}_n \leq K\right) + P_{\theta_{*n},F_n}(\widehat{\xi}_n > K) \\
&\leq o(1) + \delta,
\end{aligned} \tag{6.10}$$

where the equality holds wp$\to$1 because $\widehat{\alpha}_{2n}(\theta_1) = \alpha\ \forall \theta_1 \in CS^+_{1n}$ wp$\to$1 by Assumption OE(ii), the second inequality holds because $\theta_1 \in CS^+_{1n}$ implies $n^{1/2}||\theta_1 - \theta_{1*n}|| \leq \widehat{\xi}_n$, and the last inequality holds for K $(= K_\delta)$ sufficiently large and n sufficiently large by Assumption C(v) and because $\widehat{\xi}_n = O_p(1)$. Since $\delta > 0$ is arbitrary, (6.10) implies (6.8), which completes the proof of part (b).

Next, we prove parts (c) and (d). For any null sequence S that satisfies Assumption OE, (1) the first result of part (d) holds by the proof of (6.8), and (2) part (c) and the second result of part (d) hold by the same argument as for (6.7) using Assumption C(iii).

Now, we prove part (e). Let S^A be as in part (e). By the definition of contiguity, any sequence of events whose probabilities converge to zero under S also converge to zero under S^A. Hence, Assumptions OE(i), OE(ii), and C(v) also hold under S^A. (For Assumption OE(i), this uses the fact that $X_n = O_p(1)$ if and only if $P(|X_n| > K_n) \to 0$ for all sequences of finite constants $K_n \to \infty$.) Given this, the first result of part (e) holds by the proof of (6.8) and the second result of part (e) holds by the same argument as for (6.7).

The proof of part (f) is the same as the proof of parts (a)–(e) with some minor changes. Throughout the proof, θ_{20} is replaced by $\theta_{2*n} \in \Theta_{2*}$, the sequences S (and subsequences S_m) considered are null sequences (and subsequences) for null hypotheses that may depend on n (i.e., $H_0 : \theta_2 = \theta_{2*n}$), the quantities $\mathcal{F}_{SV}$, φ_{2n}^{SV}, $\phi_{2n}(\theta_1, \eta)$, $\widehat{\alpha}_{2n}(\theta_1)$, and CS_{1n}^+ are taken to be functions of θ_{2*n} rather than θ_{20}, and the expression $\limsup_{n\to\infty} \sup_{(\theta_1,F)\in\mathcal{F}_{SV}} P_{\theta_1,\theta_{20},F}(\varphi_{2n}^{SV} > 0)$ in the paragraph following (6.1), which is the asymptotic size of the test φ_{2n}^{SV}, see (5.3), is replaced by $\limsup_{n\to\infty} \sup_{\theta_2\in\Theta_{2*}} \sup_{(\theta_1,F)\in\mathcal{F}_{SV}(\theta_2)} P_{\theta_1,\theta_2,F}(\varphi_{2n}^{SV}(\theta_2) > 0)$, which is one minus the asymptotic size of the CS based on $\varphi_{2n}^{SV}(\cdot)$, see (5.5). □

7 Two-Step Tests in the Moment Condition Model

In this section, we describe in detail three two-step tests for the moment condition model. We consider a first-step AR CS for θ_1, an estimator set $\widehat{\Theta}_{1n}$ based on solutions to GMM FOC's, data-dependent significance levels $\widehat{\alpha}_{2n}(\theta_1)$, and second step C($\alpha$)-AR, C($\alpha$)-LM, and C($\alpha$)-QLR1 tests.

Given the definition of two-step CS's for θ_2 in (4.4), this section implicitly also provides detailed descriptions of three two-step CS's for the moment condition model.

7.1 Specification of the First-Step CS

For the first-step CS for θ_1, CS_{1n}, we consider the (null-restricted) AR CS. Other CS's could be used, but the AR CS has power advantages, as noted in the Introduction.

The nominal $1-\eta$ (null-restricted) AR CS for θ_1 is

$$CS_{1n}^{AR} := \{\theta_1 \in \Theta_1 : AR_n(\theta_1, \theta_{20}) \le \chi_k^2(1-\eta)\}, \text{ where } AR_n(\theta) := n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \qquad (7.1)$$

and $\chi_k^2(1-\eta)$ denotes the $1-\eta$ quantile of the χ_k^2 distribution for some $\eta \in (0,1)$.

7.2 Specification of the Estimator Set

Let $\widehat{Q}_n(\theta)$ denote the GMM criterion function

$$\widehat{Q}_n(\theta) := \widehat{g}_n(\theta)'\widehat{W}_{1n}\widehat{g}_n(\theta), \tag{7.2}$$

where $\widehat{W}_{1n}$ is a symmetric, positive semi-definite, possibly data-dependent, $k \times k$ weight matrix that does not depend on θ_1 (but may depend on the null value θ_{20}). When $\widehat{g}_n(\theta)$ is of the form $\widehat{g}_n(\theta) := n^{-1}\sum_{i=1}^n Z_i u_i(\theta)$ for some k vector of instruments and some scalar $u_i(\theta)$, e.g., as in Stock and Wright (2000), one can take $\widehat{W}_{1n} = (n^{-1}\sum_{i=1}^n Z_i Z_i')^{-1}$. This choice yields invariance to nonsingular transformations of Z_i. Or, one can take $\widehat{W}_{1n}$ to be the usual first-step or second-step GMM weight matrix used to compute the two-step GMM estimator. (The usual first-step GMM weight matrix is just $\widehat{W}_{1n} = I_k$.)

The leading choice for the estimator set $\widehat{\Theta}_{1n}$ to be used in the moment condition model is

$$\widehat{\Theta}_{1n} := \{\theta_1 \in \Theta_1 : \widehat{G}_{1n}(\theta_1)'\widehat{W}_{1n}\widehat{g}_n(\theta_1) = 0^{p_1} \ \& \ \widehat{Q}_n(\theta_1) \leq \inf_{\theta_1 \in \Theta_1} \widehat{Q}_n(\theta_1) + c_n\} \tag{7.3}$$

for some positive constants $\{c_n : n \geq 1\}$ for which $c_n \to 0$, where $\widehat{G}_{1n}(\theta_1)$ is defined in (3.9). The choice of the constants $\{c_n\}$ depends on the choice of the criterion function $\widehat{Q}_n(\theta)$. When $\widehat{Q}_n(\theta)$ is a GMM criterion function, we require $nc_n \to \infty$, e.g., $c_n = \log(n)/n$.

We define $\widehat{\Theta}_{1n}$ as in (7.3) because we can show that, under suitable assumptions, there exists a $n^{1/2}$-consistent solution to the FOC's of the GMM criterion function that minimizes the criterion function $\widehat{Q}_n(\theta_1)$ up to c_n. One could omit the minimization condition in (7.3). But, this condition makes $\widehat{\Theta}_{1n}$ smaller, which is desirable for power purposes because it allows one to exclude local minima, local and global maxima, and inflection points from $\widehat{\Theta}_{1n}$.

7.3 Specification of the Second-Step Significance Level

For use with the second-step C(α)-AR test, we employ the following identification-category-selection statistic (ICS):

$$ICS_{1n}(\theta) := \lambda_{\min}^{1/2}\left(\widehat{\Phi}_{1n}(\theta)\widehat{G}_{1n}(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{G}_{1n}(\theta)\widehat{\Phi}_{1n}(\theta)\right), \tag{7.4}$$

where $\widehat{G}_{1n}(\theta)$ is defined in (3.9), $\widehat{\Omega}_n(\theta)$ is defined in (3.6), and

$$\begin{aligned}
&\widehat{\Phi}_{jn}(\theta) := Diag\{\widehat{\sigma}_{j1n}^{-1}(\theta), ..., \widehat{\sigma}_{jp_jn}^{-1}(\theta)\} \in R^{p_j \times p_j}, \\
&\widehat{\sigma}_{jsn}^2(\theta) := n^{-1} \sum_{i=1}^{n} \left(||G_{jsi}(\theta)|| - ||\widehat{G}||_{jsn}(\theta) \right)^2, \\
&G_{jsi}(\theta) := \frac{\partial}{\partial \theta_{js}} g_i(\theta) \in R^k, \ \theta_j = (\theta_{j1}, ..., \theta_{jp_j})', \text{ and } ||\widehat{G}||_{jsn}(\theta) := n^{-1} \sum_{i=1}^{n} ||G_{jsi}(\theta)||
\end{aligned} \tag{7.5}$$

for $s = 1, ..., p_j$ and $j = 1, 2$.

For use with the second-step C(α)-LM and C(α)-QLR1 tests, we employ the following ICS statistic:

$$\begin{aligned}
ICS_n^*(\theta) &:= \lambda_{\min}^{1/2} \left(\widehat{\Phi}_n(\theta) \widehat{G}_n(\theta)' \widehat{\Omega}_n^{-1}(\theta) \widehat{G}_n(\theta) \widehat{\Phi}_n(\theta) \right), \text{ where} \\
\widehat{\Phi}_n(\theta) &:= Diag\{\widehat{\Phi}_{1n}(\theta), \widehat{\Phi}_{2n}(\theta)\} \in R^{p \times p}
\end{aligned} \tag{7.6}$$

and $\widehat{G}_n(\theta)$ is defined in (3.9).

The matrices $\widehat{\Phi}_{1n}(\theta)$ and $\widehat{\Phi}_{2n}(\theta)$ that appear in the definitions of $ICS_{1n}(\theta)$ and $ICS_n^*(\theta)$ ensure that these statistics are invariant to rescaling of the parameters θ_{js} for $s = 1, ..., p_j$ and $j = 1, 2$.

The statistic $ICS_{1n}(\theta)$ is an estimator of the smallest singular value of $\Omega_F^{-1/2}(\theta) E_F G_{1i}(\theta) \Phi_{1F}(\theta)$, where $E_F G_{1i}(\theta)$ is the expected Jacobian of the moment functions wrt θ_1, $\Omega_F(\theta)$ denotes the variance matrix of $n^{1/2} \widehat{g}_n(\theta)$, and $\Phi_{jF}(\theta)$ denotes the diagonal matrix containing the reciprocals of the standard deviations of $||G_{jsi}(\theta)||$ for $s = 1, ..., p_j$ and $j = 1, 2$. Analogously, the statistic $ICS_n^*(\theta)$ is an estimator of the smallest singular value of $\Omega_F^{-1/2}(\theta) E_F G_i(\theta) \Phi_F(\theta)$, where $E_F G_i(\theta)$ is the expected Jacobian of the moment functions wrt θ and $\Phi_F(\theta) := Diag\{\Phi_{1F}(\theta), \Phi_{2F}(\theta)\} \in R^{p \times p}$.

We let[4]

$$ICS_n(\theta_1) := \begin{cases} ICS_{1n}(\theta_1) & \text{for the 2nd-step C}(\alpha)\text{-AR test} \\ ICS_n^*(\theta_1) & \text{for the 2nd-step C}(\alpha)\text{-LM and C}(\alpha)\text{-QLR1 tests.} \end{cases} \tag{7.7}$$

The $ICS_n(\theta)$ statistic is different from, but related to, the ICS statistic employed in Andrews and Cheng (2012, 2013, 2014). The latter is a Wald statistic based on an estimator of a parameter that determines the strength of identification. In the models considered in this paper, no such parameter need exist.

[4]The second-step C(α)-AR test does not rely on $\widehat{G}_{2n}(\theta_1)$, whereas the second-step C(α)-LM and C(α)-QLR1 tests do. In consequence, it turns out that for the latter tests local strong identification of the whole vector θ is required for sequences to satisfy Assumption C. For the second-step C(α)-AR test only local strong identification of θ_1 given the true value of θ_2 is required for sequences to satisfy Assumption C. These differences lead to the different definitions of the $ICS_n(\theta)$ statistic in (7.7) for the different second-step tests.

Given $ICS_n(\theta_1)$, we define the data-dependent significance level $\widehat{\alpha}_{2n}(\theta_1)$ as follows:

$$\widehat{\alpha}_{2n}(\theta_1) := \begin{cases} \alpha_2 & \text{if } ICS_n(\theta_1) \leq K_L \\ \alpha_2 + s\left(\frac{ICS_n(\theta_1)-K_L}{K_U-K_L}\right)\alpha_1 & \text{if } ICS_n(\theta_1) \in (K_L, K_U] \\ \alpha & \text{if } ICS_n(\theta_1) > K_U, \end{cases} \tag{7.8}$$

where $s(\cdot)$ is a strictly increasing continuous function on $[0,1]$ with $s(0) = 0$ and $s(1) = 1$ and $0 < K_L \leq K_U < \infty$. For example, $s(x) = x1(0 < x < 1) + 1(x \geq 1)$.

In some scenarios it may be advantageous to use an ICS statistic that differs from the ones defined in (7.4)–(7.7). For example, for models that fall into the framework considered in Andrews and Cheng (2012, 2013, 2014), one could use the ICS statistics in those papers. One could also consider the ICS statistic in I. Andrews (2017).

7.4 Specification of the Second-Step Test

Next, we specify three second-step C(α)-type tests for moment condition models. They follow the form of Chaudhuri and Zivot's (2011) C(α)-LM test. The latter extends, from likelihood models to moment condition models, the C(α) tests of Neyman (1959), Moran (1970), and Bera and Bilias (2001, eqn. (3.24)). For related results and extensions, see Smith (1987) and I. Andrews (2017).

Following Kleibergen (2005), let $\widehat{D}_{jn}(\theta)$ be the sample Jacobian of the moment functions wrt θ_j adjusted to be asymptotically independent of the sample moments $\widehat{g}_n(\theta)$ for $j = 1, 2$. By definition, for $j = 1, 2$,

$$\begin{aligned} \widehat{D}_{jn}(\theta) &:= [\widehat{D}_{j1n}(\theta) : \cdots : \widehat{D}_{jp_jn}(\theta)], \text{ where, for } s = 1, ..., p_j, \\ \widehat{D}_{jsn}(\theta) &:= \widehat{G}_{jsn}(\theta) - \widehat{\Gamma}_{jsn}(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \in R^k, \\ \widehat{G}_{jsn}(\theta) &:= \frac{\partial}{\partial \theta_{js}}\widehat{g}_n(\theta) \in R^k, \ \theta_j := (\theta_{j1}, ..., \theta_{jp_j})' \in R^{p_j}, \text{ and} \\ \widehat{\Gamma}_{jsn}(\theta) &:= n^{-1}\sum_{i=1}^n \left(\frac{\partial}{\partial \theta_{js}} g_i(\theta) - \widehat{G}_{jsn}(\theta)\right) g_i(\theta)' \in R^{k\times k}. \end{aligned} \tag{7.9}$$

Given a matrix A, let P_A and M_A denote the projection matrices onto the column space of A and the space orthogonal to the column space of A, respectively.

7.4.1 C(α)-AR Test

The second-step C(α)-AR test is a quadratic form in the residuals from the projection of the sample moments onto the space spanned by the random $k \times p_1$ matrix $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)$. This yields a statistic whose power is directed towards violations of $H_0 : \theta_2 = \theta_{20}$. To obtain the desired $\chi^2_{k-p_1}$

asymptotic distribution of this statistic, we need $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)$ to have full rank p_1 a.s. asymptotically. Andrews and Guggenberger (2017) (AG1) provides a fairly general, but complicated, set of conditions under which this holds.

Here we take a different approach that yields a $\chi^2_{k-p_1}$ asymptotic distribution under very simple and general conditions. Rather than projecting onto $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)$, we project onto $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta) + an^{-1/2}\zeta_1$, where ζ_1 is a $k \times p_1$ matrix of independent standard normal random variables that are independent of all of the statistics considered, such as $\widehat{g}_n(\theta)$, $\widehat{G}_n(\theta)$, and $\widehat{\Omega}_n(\theta)$, and a is a small positive constant. This small random perturbation $an^{-1/2}\zeta_1$ guarantees that the space spanned by $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta) + an^{-1/2}\zeta_1$ has dimension p_1 a.s. asymptotically. Under strong and semi-strong identification this perturbation has no effect asymptotically and very little effect in finite samples for a small. Under weak identification it has a small, but non-negligible, asymptotic effect. Note that all of the results given below still hold if one takes $a = 0$ provided one restricts the parameter space for the distributions F as in AG1 (see $\mathcal{F}_0$ in AG1).

For given $\theta_1 \in \Theta_1$, the nominal η second-step C(α)-AR test rejects $H_0 : \theta_2 = \theta_{20}$ when

$$\begin{aligned} \phi_{2n}^{AR}(\theta_1, \eta) &:= AR_{2n}(\theta_1, \theta_{20}) - \chi^2_{k-p_1}(1-\eta) > 0, \text{ where} \\ AR_{2n}(\theta) &:= n\widetilde{g}_n(\theta)'\widehat{M}_{1n}(\theta)\widetilde{g}_n(\theta), \ \widetilde{g}_n(\theta) := \widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \\ \widehat{M}_{1n}(\theta) &:= I_k - P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta) + an^{-1/2}\zeta_1}, \end{aligned} \tag{7.10}$$

$\chi^2_{k-p_1}(1-\eta)$ denotes the $1-\eta$ quantile of the $\chi^2_{k-p_1}$ distribution for some $\eta \in (0,1)$, $\widehat{g}_n(\theta)$ and $\widehat{\Omega}_n(\theta)$ are defined in (3.3) and (3.6), and $\widehat{D}_{1n}(\theta)$ is defined in (7.9).

7.4.2 C(α)-LM Test

The definition of the C(α)-LM test in Chaudhuri and Zivot (2011) involves projection of $\widetilde{g}_n(\theta)$ onto $M_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta)$. To obtain the desired $\chi^2_{p_2}$ asymptotic null distribution of this statistic when θ equals the true value $(\theta_{1*n}, \theta_{20})$, one needs this matrix to have full rank p_2 a.s. asymptotically. This can be violated under weak identification. For example, if $\widehat{D}_{2n}$ $(= \widehat{D}_{2n}(\theta_{1*n}, \theta_{20}))$ has rank less than p_2 with positive probability for all n, then it is violated. Another example occurs when $\widehat{D}_{2n}$ and $\widehat{D}_{1n}$ individually display strong identification, but jointly display weak identification.[5] In cases like these, projection onto $M_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)}\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta)$ does not lead to the desired $\chi^2_{p_2}$ asymptotic distribution of the C(α)-LM statistic.

We introduce a modified C(α)-LM statistic that behaves like a C(α)-LM statistic under strong

[5] That is, $\widehat{D}_{2n}$ and $\widehat{D}_{1n}$ have asymptotic distributions (after suitable normalizations) with positive smallest singular values a.s., but $[\widehat{D}_{1n} : \widehat{D}_{2n}]$ has an asymptotic distribution whose smallest singular value is zero.

identification of θ, but has an asymptotic $\chi^2_{p_2}$ null distribution regardless of the strength of identification of θ. First, we replace $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta)$ by $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta) + an^{-1/2}\zeta_2$, which has a small random perturbation that guarantees that the $k \times p_2$ matrix has full column rank p_2 a.s. asymptotically. Second, we employ $\widehat{M}_{1n}(\theta)$ (defined in (7.10)), rather than $M_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)}$, which utilizes a small random perturbation to $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)$. Third, under weak identification we project onto $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta) + an^{-1/2}\zeta_2$, rather than onto $\widehat{M}_{1n}(\theta)(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta) + an^{-1/2}\zeta_2)$ because this circumvents the potential problem (described in the previous paragraph) that $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta)$ and $\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta)$ might be collinear asymptotically. In particular, we use a data-dependent smoothed indicator function, $WI_n(\theta)$, that equals one under weak identification and equals zero under strong enough identification of θ. We employ $\widehat{M}_{1n}(\theta)$ when $WI_n(\theta) = 1$ and I_k when $WI_n(\theta) = 0$.

We define

$$WI_n(\theta) := 1 - s\left(\frac{ICS_n^*(\theta) - K_L^*}{K_U^* - K_L^*}\right), \tag{7.11}$$

where $ICS_n^*(\theta)$ and $s(\cdot)$ are defined in (7.6) and (7.8), respectively, and K_L^* and K_U^* are tuning parameters that satisfy $0 \leq K_L^* < K_U^* < K_L \leq K_U < \infty$. As defined, $WI_n(\theta) = 1$ if $ICS_n^*(\theta) \leq K_L^*$ and $WI_n(\theta) = 0$ if $ICS_n^*(\theta) \geq K_U^*$.

We project $\widetilde{g}_n(\theta)$ onto the space spanned by

$$\begin{aligned} D_{2n}^\dagger(\theta) &:= (\widehat{M}_{1n}(\theta) + WI_n(\theta)\widehat{P}_{1n}(\theta))(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{2n}(\theta) + an^{-1/2}\zeta_2), \text{ where} \\ \widehat{P}_{1n}(\theta) &:= P_{\widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_{1n}(\theta) + an^{-1/2}\zeta_1} \left(= I_k - \widehat{M}_{1n}(\theta)\right). \end{aligned} \tag{7.12}$$

For given $\theta_1 \in \Theta_1$, the nominal level η second-step C(α)-LM test rejects $H_0 : \theta_2 = \theta_{20}$ when

$$\begin{aligned} \phi_{2n}^{LM}(\theta_1, \eta) &:= LM_{2n}(\theta_1, \theta_{20}) - \chi^2_{p_2}(1-\eta) > 0, \text{ where} \\ LM_{2n}(\theta) &:= n\widetilde{g}_n(\theta)' P_{D_{2n}^\dagger(\theta)} \widetilde{g}_n(\theta), \end{aligned} \tag{7.13}$$

ζ_2 is a $k \times p_2$ matrix of independent standard normal random variables that are independent of all statistics considered, such as $\widehat{g}_n(\theta)$, $\widehat{G}_n(\theta)$, $\widehat{\Omega}_n(\theta)$, and ζ_1, and a is a small positive constant.

The second-step C(α)-LM test reduces to the C(α)-LM test in Chaudhuri and Zivot (2011) when $WI_n(\theta) := 0$ and $a = 0$. We provide correct asymptotic size results both for the case where $WI_n(\theta)$ is defined in (7.11) and for the case where $WI_n(\theta) := 0$. But, the latter case requires a more restrictive parameter space, see (8.12) below.

7.4.3 C(α)-QLR1 Test

Next, we consider a C(α) version of Kleibergen's (2005, Sec. 5.1) nonlinear CLR test. This test employs a rank statistic, $rk_{2n}(\theta)$, that is suitable for testing the hypothesis $rank(E_F G_{2i}) \leq p_2 - 1$ against $rank(E_F G_{2i}) = p_2$, where $rank(A)$ denotes the rank of a matrix A. For this we use the rank statistic of Robin and Smith (2000). The second-step C(α)-QLR1 test statistic is

$$QLR1_{2n}(\theta) := \frac{1}{2}\left(AR^{\dagger}_{2n}(\theta) - rk_{2n}(\theta) + \sqrt{(AR^{\dagger}_{2n}(\theta) - rk_{2n}(\theta))^2 + 4LM_{2n}(\theta)\cdot rk_{2n}(\theta)}\right), \text{ where}$$
$$AR^{\dagger}_{2n}(\theta) := n\widetilde{g}_n(\theta)'\left(\widehat{M}_{1n}(\theta) + WI_n(\theta)\widehat{P}_{1n}(\theta)\right)\widetilde{g}_n(\theta), \tag{7.14}$$

$LM_{2n}(\theta)$ is the second-step C(α)-LM test statistic defined in (7.13), and $\widehat{P}_{1n}(\theta)$ is defined in (7.12). Given the C(α) definition of $LM_{2n}(\theta)$, the C(α)-QLR1 statistic, $QLR1_{2n}(\theta)$, is defined with the modified C(α)-AR statistic, $AR^{\dagger}_{2n}(\theta)$, rather than the $AR_{2n}(\theta)$ statistic defined in (7.10).

The Robin and Smith (2000)-type rank statistic that we consider is

$$rk_{2n}(\theta) := \lambda_{\min}(K_{rk} n\widehat{\Phi}_{2n}(\theta)\widehat{D}_{2n}(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{D}_{2n}(\theta)\widehat{\Phi}_{2n}(\theta)), \tag{7.15}$$

where $\widehat{\Phi}_{2n}(\theta)$ is defined in (7.5) and $K_{rk} > 0$ is a constant. The matrix $\widehat{\Phi}_{2n}(\theta)$ that appears in the definition of $rk_{2n}(\theta)$ ensures that $rk_{2n}(\theta)$ is invariant to rescaling of the parameters θ_{2s} for $s = 1, ..., p_2$. This is a useful feature because one does not want this statistic to vary when one changes the unit of measurement of a parameter.

The C(α)-QLR1 test uses a conditional critical value that depends on the rank statistic and $WI_n(\theta)$. For fixed $0 \leq r < \infty$ and $w \in \{0, 1\}$, let $c^{QLR1}(1-\eta, r, w)$ denote the $1-\eta$ quantile of the distribution of

$$QLR1(r, w) := \frac{1}{2}\left(\chi^2_{p_2} + \chi^2_{k-p+wp_1} - r + \sqrt{(\chi^2_{p_2} + \chi^2_{k-p+wp_1} - r)^2 + 4\chi^2_{p_2} r}\right), \tag{7.16}$$

where $\chi^2_{p_2}$ and $\chi^2_{k-p+wp_1}$ are independent chi-square random variables with p_2 and $k - p + wp_1$ degrees of freedom, respectively. For a nominal level η test, the C(α)-QLR1 critical value is

$$c^{QLR1}(1-\eta, rk_{2n}(\theta), WI^{\dagger}_n(\theta)), \text{ where } WI^{\dagger}_n(\theta) := 1(WI_n(\theta) > 0). \tag{7.17}$$

As defined, the critical value depends on a $\chi^2_{k-p_2}$ distribution when $WI_n(\theta) > 0$ (i.e., $ICS^*_n(\theta) < K^*_U$) and a χ^2_{k-p} distribution when $WI_n(\theta) = 0$. It can be shown that when $rk_{2n}(\theta) \to_p \infty$ (which holds under strong identification of θ_2), $c^{QLR1}(1-\eta, rk_{2n}(\theta), WI^{\dagger}_n(\theta)) \to_p \chi^2_{p_2}(1-\eta)$ whether $WI^{\dagger}_n(\theta) = 0$ or 1 for any $n \geq 1$. Thus, the value of $WI^{\dagger}_n(\theta)$ is asymptotically irrelevant in this case.

Given $\theta_1 \in \Theta_1$, the nominal level η second-step C(α)-QLR1 test rejects $H_0 : \theta_2 = \theta_{20}$ when

$$\phi_{2n}^{QLR1}(\theta_1, \eta) := QLR1_{2n}(\theta_1, \theta_{20}) - c^{QLR1}(1 - \eta, rk_{2n}(\theta_1, \theta_{20}), WI_n^{\dagger}(\theta_1, \theta_{20})) > 0. \tag{7.18}$$

8 Asymptotic Results under Primitive Conditions in the Moment Condition Model

In this section, we provide asymptotic results under primitive conditions for three subvector tests AR/AR, AR/LM, and AR/CQLR1 (and corresponding CS's) for the moment condition model in (3.3). All three tests use the first-step AR CS defined in (7.1) and the estimator set $\widehat{\Theta}_{1n}$ defined in (7.3). The results are obtained by verifying the assumptions of Theorem 5.1. Here we consider the case where the observations $\{W_i : i = 1, 2, ...\}$ are i.i.d. under any distribution F.

8.1 Parameter Space Definitions

8.1.1 Notation

The moment functions $g_i(\theta)$ are defined on Θ. The parameter space Θ is assumed to be an open subset of R^p. Given Θ, Θ_1 is defined in (3.7). The parameter space Θ_1 is employed in the definitions of CS_{1n} and $\widehat{\Theta}_{1n}$. Given that Θ is open, Θ_1 is open. The true value of the nuisance parameter θ_1 is assumed to lie in a set Θ_{1*} that satisfies Θ_{1*} is bounded and $B(\Theta_{1*}, \varepsilon) \subset \Theta_1$ for some $\varepsilon > 0$, where $B(\Theta_{1*}, \varepsilon)$ denotes the union of closed balls in R^{p_1} with radius ε centered at points in Θ_{1*}. This implies that the true value of θ_1 cannot be on the boundary of the optimization set Θ_1.[6]

When considering CS's for θ_2, we assume that Θ is open and the true parameter θ lies in a set Θ_* that is bounded and satisfies $B(\Theta_*, \varepsilon) \subset \Theta$ for some $\varepsilon > 0$. In the CS case, we define $\Theta_{1*} := \{\theta_1 : \exists \theta_2 \text{ such that } (\theta_1', \theta_2')' \in \Theta_*\}$ and $\Theta_{2*} := \{\theta_2 : \exists \theta_1 \text{ such that } (\theta_1', \theta_2')' \in \Theta_*\}$.

The variance matrix of the moments is denoted by

$$\Omega_F(\theta) := E_F(g_i(\theta) - E_F g_i(\theta))(g_i(\theta) - E_F g_i(\theta))'. \tag{8.1}$$

Let $\theta_j = (\theta_{j1}, ..., \theta_{jp_j})'$ for $j = 1, 2$. When the following derivatives exist, we define

$$\begin{aligned} \xi_{ji} &:= \max_{s,u \le p_j} \sup_{\theta_1 \in \Theta_1} \left\| \frac{\partial^2}{\partial \theta_{js} \partial \theta_{ju}} g_i(\theta_1, \theta_{20}) \right\| \text{ for } j = 1, 2 \text{ and} \\ \xi_{12i} &:= \max_{s \le p_1, u \le p_2} \sup_{\theta_1 \in \Theta_1} \left\| \frac{\partial^2}{\partial \theta_{1s} \partial \theta_{2u}} g_i(\theta_1, \theta_{20}) \right\|. \end{aligned} \tag{8.2}$$

[6] If this condition is violated, the possible effect is small. Specifically, asymptotic NRP's are still $\alpha + \alpha_1$ or less and, hence, the distortion is at most α_1, such as $\alpha_1 = .005$.

We let $\tau_{jn}^{\Phi} = \tau_{jn}^{\Phi}(\theta_{*n})$ for $\theta_{*n} = (\theta_{1*n}', \theta_{20}')'$. Here, $\tau_{jn}^{\Phi}(\theta)$ denotes the smallest singular value of

$$\begin{aligned}&\Omega_{F_n}^{-1/2}(\theta) E_{F_n} \widehat{G}_{jn}(\theta) \Phi_{jF_n}(\theta), \text{ where} \\ &\Phi_{jF}(\theta) := Diag\{Var_F^{-1/2}(||G_{j1i}(\theta)||), ..., Var_F^{-1/2}(||G_{jp_ji}(\theta)||)\}\end{aligned} \tag{8.3}$$

and $G_{ji}(\theta) = (G_{j1i}(\theta), ..., G_{jp_ji}(\theta)) \in R^{k \times p_j}$ is defined in (7.5) for $j = 1, 2$. We let $\tau_n^{\Phi} = \tau_n^{\Phi}(\theta_{*n})$, where $\tau_n^{\Phi}(\theta)$ denotes the smallest singular value of

$$\Omega_{F_n}^{-1/2}(\theta) E_{F_n} \widehat{G}_n(\theta) \Phi_{F_n}(\theta), \text{ and } \Phi_F(\theta) := Diag\{\Phi_{1F}(\theta), \Phi_{2F}(\theta)\} \in R^{p \times p}. \tag{8.4}$$

Define

$$r_{jF}(\theta) := rank(\Omega_F^{-1/2}(\theta) E_F \widehat{G}_{jn}(\theta)) \text{ for } j = 1, 2. \tag{8.5}$$

A compact singular value decomposition (SVD) of $\Omega_F^{-1/2}(\theta) E_F \widehat{G}_{jn}(\theta)$ is

$$\Omega_F^{-1/2}(\theta) E_F \widehat{G}_{jn}(\theta) = C_{*jF}(\theta) \Upsilon_{*jF}(\theta) B_{*jF}(\theta)' \text{ for } j = 1, 2, \tag{8.6}$$

where $C_{*jF}(\theta) \in R^{k \times r_{jF}(\theta)}$, $\Upsilon_{*jF}(\theta) \in R^{r_{jF}(\theta) \times r_{jF}(\theta)}$, $B_{*jF}(\theta) \in R^{p_j \times r_{jF}(\theta)}$, the columns of $C_{*jF}(\theta)$ are orthonormal, the columns of $B_{*jF}(\theta)$ are orthonormal, and $\Upsilon_{*jF}(\theta)$ is the diagonal matrix with the positive singular values of $\Omega_F^{-1/2}(\theta) E_F \widehat{G}_{jn}(\theta)$ on its diagonal in non-increasing order.[7] Define

$$C_{*F}(\theta) := [C_{*1F}(\theta) : C_{*2F}(\theta)] \in R^{k \times (r_{1F}(\theta) + r_{2F}(\theta))}. \tag{8.7}$$

8.1.2 AR/AR Subvector Test

For the AR/AR subvector test, we assume that $g_i(\theta_1)$ is twice continuously differentiable in θ_1 on Θ_1 for all sample realizations. For this test, the null parameter space for the true (θ_1, F) is

$$\begin{aligned}\mathcal{F}_{AR/AR} := \{&(\theta_1, F) : E_F g_i(\theta_1) = 0^k,\ \theta_1 \in \Theta_{1*},\ \{W_i : i \geq 1\} \text{ are i.i.d. under } F, \\ &E_F ||g_i(\theta_1)||^{2+\gamma} \leq M,\ E_F ||vec(G_{1i}(\theta_1))||^{2+\gamma} \leq M,\ E_F \xi_{1i}^2 \leq M, \\ &\lambda_{\min}(\Omega_F(\theta_1)) \geq \delta,\ Var_F(||G_{1si}(\theta_1)||) \geq \delta\ \forall s = 1, ..., p_1\}\end{aligned} \tag{8.8}$$

for some $\gamma, \delta > 0$ and $M < \infty$.

The second last condition in $\mathcal{F}_{AR/AR}$ bounds $\lambda_{\min}(\Omega_F(\theta_1))$ away from zero. This is not restrictive in most moment condition models, but it is restrictive in likelihood scenarios because under

[7] A compact SVD can be obtained from any SVD by deleting the non-essential rows and columns of the matrices in the SVD as in (8.6), e.g., see Demmel (2000).

weak identification the Jacobian is close to being singular and this implies that the variance matrix $\Omega_F(\theta_1)$ also is close to being singular (by the information matrix equality).

The last condition in $\mathcal{F}_{AR/AR}$ is not restrictive. For example, in the linear IV model with multiple right-hand side (rhs) endogenous variables, $G_{1si} = Z_i X_{1si}$, where Z_i is an IV vector and X_{1si} is the sth rhs endogenous variable whose coefficient is not specified by the null hypothesis. In this case, this condition is quite mild.

If $\widehat{W}_{1n} \neq I_k$, then some conditions that control the behavior of $\widehat{W}_{1n}$ typically need to be added to the definition of $\mathcal{F}_{AR/AR}$ in order to verify the condition on $\widehat{W}_{1n}$ in Theorem 8.1 below. For example, if $\widehat{W}_{1n} = (n^{-1}\sum_{i=1}^n Z_i Z_i')^{-1}$, then the following conditions are added to the definition of $\mathcal{F}_{AR/AR}$: $\lambda_{\min}(E_F Z_i Z_i') \geq \delta$ and $E_F||Z_i||^{2+\gamma} \leq M$.

For the AR/AR CS, the parameter space for the true (θ, F) is

$$\mathcal{F}_{\Theta,AR/AR} := \{(\theta, F) : \theta = (\theta_1', \theta_2')' \in \Theta \text{ such that } (\theta_1, F) \in \mathcal{F}_{AR/AR}(\theta_2) \text{ and } \theta_2 \in \Theta_{2*}\}, \tag{8.9}$$

where $\mathcal{F}_{AR/AR}(\theta_{20})$ denotes $\mathcal{F}_{AR/AR}$ with its dependence on the null value θ_{20} made explicit.

8.1.3 AR/LM and AR/QLR1 Subvector Tests

For the AR/LM and AR/QLR1 subvector tests, we assume that $g_i(\theta_1)$ is twice continuously differentiable in θ_1 on Θ_1, $g_i(\theta_1, \theta_2)$ is differentiable in θ_2 at θ_{20} $\forall \theta_1 \in \Theta_1$, and $(\partial/\partial\theta_2')g_i(\theta_1, \theta_{20})$ is differentiable in θ_1 $\forall \theta_1 \in \Theta_1$ for all sample realizations. A sufficient condition for these conditions is $g_i(\theta)$ is twice continuously differentiable in θ at $(\theta_1', \theta_{20}')'$ $\forall \theta_1 \in \Theta_1$ for all sample realizations.

For the AR/LM and AR/QLR1 subvector tests, the null parameter space for the true (θ_1, F) is

$$\begin{aligned}\mathcal{F}_{AR/LM,QLR1} := \{(\theta_1, F) \in \mathcal{F}_{AR/AR} : E_F||vec(G_{2i}(\theta_1))||^{2+\gamma} \leq M,\ E_F\xi_{2i}^2 \leq M,\ E_F\xi_{12i} \leq M,\\ Var_F(||G_{2si}(\theta_1)||) \geq \delta\ \forall s = 1, ..., p_2\}\end{aligned} \tag{8.10}$$

for $\gamma, \delta > 0$ and $M < \infty$ as in the definition of $\mathcal{F}_{AR/AR}$.

For the AR/LM and AR/QLR1 CS's, the parameter space for the true (θ, F) is

$$\mathcal{F}_{\Theta,AR/LM,QLR1} := \{(\theta, F) : \theta = (\theta_1', \theta_2')' \in \Theta \text{ such that } (\theta_1, F) \in \mathcal{F}_{AR/LM,QLR1}(\theta_2) \text{ and } \theta_2 \in \Theta_{2*}\}, \tag{8.11}$$

where $\mathcal{F}_{AR/LM,QLR1}(\theta_{20})$ denotes $\mathcal{F}_{AR/LM,QLR1}$ with its dependence on the null value θ_{20} made explicit.

Next, we define the null parameter space for the AR/LM and AR/QLR1 subvector tests if one defines the $LM_{2n}(\theta)$ and $AR^{\dagger}_{2n}(\theta)$ statistics and C(α)-QLR1 critical value with $WI_n(\theta) := 0$, which yields pure C(α)-LM and C(α)-QLR1 tests. In this case, $\mathcal{F}_{AR/LM,QLR1}$ needs to include the additional condition

$$\lambda_{\min}(C_{*F}(\theta_1)'C_{*F}(\theta_1)) \geq \delta \tag{8.12}$$

for $\delta > 0$ as above in $\mathcal{F}_{AR/LM,QLR1}$.[8] This condition is used to guarantee that the asymptotic distribution of the matrix in the projection in the $LM_{2n}(\theta)$ statistic (see (7.13)) has full column rank p_2 a.s. It allows the rank of $\Omega_F^{-1/2}(\theta_1)E_F\widehat{G}_{jn}(\theta_1)$ to take any value in $\{0, ..., p_j\}$ for $j = 1, 2$. But, it precludes the column spaces of $\Omega_F^{-1/2}(\theta_1)E_F\widehat{G}_{1n}(\theta_1)$ and $\Omega_F^{-1/2}(\theta_1)E_F\widehat{G}_{2n}(\theta_1)$ from being too similar, which is restrictive. The condition in (8.12) is not redundant.[9]

8.2 Asymptotic Results

8.2.1 AR/AR Subvector Test

Next, we provide asymptotic size results for the two-step AR/AR subvector test, denoted by $\varphi_{2n}^{AR/AR}$, and the corresponding two-step AR/AR CS. Here, null sequences S are defined as in (5.2), but with the generic parameter space $\mathcal{F}_{SV}$ replaced by $\mathcal{F}_{AR/AR}$, defined in (8.8).

For null sequences S that satisfy the following strong identification (SI) assumption, the $\varphi_{2n}^{AR/AR}$ test has asymptotic NRP equal to α. For other sequences, its asymptotic NRP's may be less than α. The (smallest singular) value τ_{1n}^{Φ} is defined in (8.3) above.

Assumption SI. For the null sequence S and some $r > 2$, (i) $\liminf_{n\to\infty} \inf_{\theta_1 \notin B(\theta_{1*n},\varepsilon)} ||E_{F_n} g_i(\theta_1)|| > 0$ for all $\varepsilon > 0$, (ii) $\liminf_{n\to\infty} \tau_{1n}^{\Phi} > K_U$ (for $K_U > 0$ as in the definition of $\widehat{\alpha}_{2n}(\theta_1)$ in (7.8)), (iii) $\limsup_{n\to\infty} E_{F_n} \sup_{\theta_1\in\Theta_1} ||g_i(\theta_1)||^r < \infty$, (iv) $\limsup_{n\to\infty} E_{F_n} \sup_{\theta_1\in\Theta_1} ||G_{1i}(\theta_1)||^r < \infty$, (v) Θ_1 is convex and bounded, and (vi) $\liminf_{n\to\infty} \inf_{\theta_1\in\Theta_1} \lambda_{\min}(\Omega_{F_n}(\theta_1)) > 0$.

Assumptions SI(i) and SI(ii) are global and local strong-identification assumptions, respectively, on θ_1 at $\{\theta_{1*n} : n \geq 1\}$ given θ_{20}. Assumptions SI(iv) and (v) can be replaced by the Lipschitz condition: $||g_i(\theta_a) - g_i(\theta_b)|| \leq B_{1i}||\theta_a - \theta_b|| \; \forall \theta_a, \theta_b \in \Theta_1$ for some random variable B_{1i} that satisfies $\limsup_{n\to\infty} E_{F_n} B_{1i}^r < \infty$ for some $r > 2$ and Θ_1 is bounded.

We use the following condition on $\widehat{W}_{1n}$, which appears in (7.2) and (7.3).

Assumption W. For the null sequence S, (i) $\widehat{W}_{1n}$ is symmetric and positive semidefinite (psd) and (ii) $\widehat{W}_{1n} \to_p W_{1\infty}$ for some nonrandom nonsingular $k \times k$ matrix $W_{1\infty}$.

[8] This condition does not depend on the particular choice of matrix $C_{*F}(\theta_1)$ (which is not uniquely defined).

[9] For example, it is violated (in the unlikely case) when $\widehat{G}_{1n}(\theta_1) = \widehat{G}_{2n}(\theta_1)$ because $C_{*1F}(\theta_1) = C_{*2F}(\theta_1)$ and $\lambda_{\min}(C_{*F}(\theta_1)'C_{*F}(\theta_1)) = 0$.

Theorem 8.1 *Suppose* $\widehat{g}_n(\theta_1)$ *are the moment functions defined in* (3.3), $\widehat{D}_{1n}(\theta)$ *is defined in* (7.9), *and* $\widehat{M}_{1n}(\theta_1)$ *is defined in* (7.10) *with* $a > 0$. *Suppose* CS_{1n} *is the first-step* AR CS CS_{1n}^{AR}, $\phi_{2n}(\theta_1, \eta)$ *is the second-step* C(α)-AR *test* $\phi_{2n}^{AR}(\theta_1, \eta)$, *and* $\widehat{\alpha}_{2n}(\theta_1)$ *is defined in* (7.4)–(7.8). *Suppose* $g_i(\theta)$ *is a function on* Θ *for all* $i \geq 1$, Θ *is an open subset of* R^p, $g_i(\theta_1)$ $(:= g_i(\theta_1, \theta_{20}))$ *is twice continuously differentiable in* θ_1 *on* Θ_1 *for all sample realizations for* Θ_1 *defined in* (3.7), Θ_{1*} *in* $\mathcal{F}_{AR/AR}$ *is bounded,* $B(\Theta_{1*}, \varepsilon) \subset \Theta_1$ *for some* $\varepsilon > 0$, $p_1 < k$, *and the positive constants* $\{c_n : n \geq 1\}$ *in* (7.3) *satisfy* $c_n \to 0$ *and* $nc_n \to \infty$. *Suppose for every subsequence* $\{w_n\}$ *of* $\{n\}$ *there exists a subsubsequence* $\{m_n\}$ *such that the null subsequence* S_m *in* $\mathcal{F}_{AR/AR}$ *satisfies Assumption* W. *Then, the two-step* AR/AR *subvector test,* $\varphi_{2n}^{AR/AR}$, *satisfies*

(a) $AsySz \leq \alpha$ *for the null parameter space* $\mathcal{F}_{AR/AR}$,

(b) $AsyNRP = \alpha$ *for all null sequences* S *in* $\mathcal{F}_{AR/AR}$ *that satisfy Assumption* SI,

(c) $AsySz = \alpha$ *provided some null sequence* S *in* $\mathcal{F}_{AR/AR}$ *satisfies Assumption* SI,

(d) *for any null sequence* S *in* $\mathcal{F}_{AR/AR}$ *that satisfies Assumption* SI, $\varphi_{2n}^{AR/AR} = \phi_{2n}^{AR}(\theta_{*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}, F_n}(\varphi_{2n}^{AR/AR} > 0) = \lim P_{\theta_{*n}, F_n}(\phi_{2n}^{AR}(\theta_{*n}, \alpha) > 0)$,

(e) *for any alternative sequence* $S^A = \{(\theta_{*n}^A, F_n^A) : n \geq 1\}$ *that satisfies Assumption* C(iii) *and is contiguous to a null sequence* S *that satisfies Assumption* OE, $\varphi_{2n}^{AR/AR} = \phi_{2n}^{AR}(\theta_{*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}^A, F_n^A}(\varphi_{2n}^{AR/AR} > 0) = \lim P_{\theta_{*n}^A, F_n^A}(\phi_{2n}^{AR}(\theta_{*n}, \alpha) > 0)$, *and*

(f) *under the assumptions stated in the Theorem before part* (a), *plus* Θ_* *is bounded and satisfies* $B(\Theta_*, \varepsilon) \subset \Theta$ *for some* $\varepsilon > 0$, *the two-step* AR/AR CS *satisfies* (i) $AsySz \geq 1 - \alpha$ *for the parameter space* $\mathcal{F}_{\Theta, AR/AR}$, (ii) $AsyCP = 1 - \alpha$ *for all sequences* S *in* $\mathcal{F}_{\Theta, AR/AR}$ *that satisfy Assumption* SI, (iii) $AsySz = 1 - \alpha$ *provided some sequence* S *in* $\mathcal{F}_{\Theta, AR/AR}$ *satisfies Assumption* SI, (iv) *for any sequence* S *in* $\mathcal{F}_{\Theta, AR/AR}$ *that satisfies Assumption* SI, $\varphi_{2n}^{AR/AR}(\theta_{2*n}) = \phi_{2n}^{AR}(\theta_{*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}, F_n}(\varphi_{2n}^{AR/AR}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n}, F_n}(\phi_{2n}^{AR}(\theta_{*n}, \alpha) > 0)$, *and* (v) *for any alternative sequence* S^A *that satisfies Assumption* C(iii) *and is contiguous to a null sequence* S *that satisfies Assumption* OE, $\varphi_{2n}^{AR/AR}(\theta_{2*n}) = \phi_{2n}^{AR}(\theta_{*n}, \alpha) + o_p(1)$ *and* $\lim P_{\theta_{*n}^A, F_n^A}(\varphi_{2n}^{AR/AR}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n}^A, F_n^A}(\phi_{2n}^{AR}(\theta_{*n}, \alpha) > 0)$.

Comments: (i). In Theorem 8.1(c), the existence of a null sequence S that satisfies Assumption SI is not restrictive because the latter imposes standard strong-identification regularity conditions.

(ii). Theorem 8.1(d) and (e) show that, under global strong identification, $\varphi_{2n}^{AR/AR}$ is asymptotically equivalent to the oracle second-step C(α)-AR test $\phi_{2n}^{AR}(\theta_{1*n}, \alpha)$ under the null hypothesis and contiguous local alternatives. When there are no over-identifying restrictions, i.e., $k = p$, the latter test is asymptotically efficient in a GMM sense, e.g., as defined in Newey and West (1987), under global strong identification. Hence, the two-step $\varphi_{2n}^{AR/AR}$ test is as well (when $k = p$).

(iii). The proof of Theorem 8.1 in the SM employs Theorem 5.1. In the proof, we show that

sequences S for which $\lim_{n\to\infty} \tau_{1n}^{\Phi} < K_L$ (and some other conditions hold) satisfy Assumption B of Section 5. We show that sequences S for which $\lim_{n\to\infty} \tau_{1n}^{\Phi} > 0$ (and some other conditions hold) satisfy Assumption C of Section 5. In addition, we show that sequences S that satisfy Assumption SI (and some other conditions) satisfy Assumption OE of Section 5.

8.2.2 AR/LM and AR/QLR1 Subvector Tests

Next, we provide asymptotic size results for the two-step AR/LM and AR/QLR1 subvector tests, denoted by $\varphi_{2n}^{AR/LM}$ and $\varphi_{2n}^{AR/QLR1}$, respectively, with the parameter space $\mathcal{F}_{AR/LM,QLR1}$, and the corresponding CS's. For these two-step tests, asymptotic NRP's that necessarily equal α, not α or less, are achieved for sequences S for which $\liminf_{n\to\infty} \tau_n^{\Phi} > K_U$ (as well as global strong identification of θ_1 given θ_{20}). This condition requires local strong identification of θ, rather than local strong identification of θ_1 given θ_{20}. It is needed because these tests rely on the projection of $\widehat{\Omega}_n^{-1/2}\widehat{D}_{2n}$ onto the space orthogonal to $\widehat{\Omega}_n^{-1/2}\widehat{D}_{1n}$, which yields a matrix with full column rank p_2 wp$\to$1 only if $\widehat{\Omega}_n^{-1/2}[\widehat{D}_{1n} : \widehat{D}_{2n}]$ has full column rank p wp$\to$1.

Assumption SI2. Assumption SI holds with τ_n^{Φ} in place of τ_{1n}^{Φ} in part (ii).

Theorem 8.2 *Suppose the statistics and conditions are as in Theorem* 8.1 *except that* $\phi_{2n}(\theta_1, \eta)$ *is the second-step* C(α)-LM *test* $\phi_{2n}^{LM}(\theta_1, \eta)$ *or* C(α)-QLR1 *test* $\phi_{2n}^{QLR1}(\theta_1, \eta)$ *with* $WI_n(\theta)$ *defined as in* (7.11), $\widehat{\alpha}_{2n}(\theta_1)$ *is defined accordingly in* (7.6)–(7.8), *the parameter space* $\mathcal{F}_{AR/AR}$ *is replaced by the parameter space* $\mathcal{F}_{AR/LM,QLR1}$, *and the condition* $p_1 < k$ *is replaced by* $p_2 \geq 1$ *for the* C(α)-LM *test and by* $p_2 \geq 1$ *and* $p \leq k$ *for the* C(α)-QLR1 *test. In addition, suppose* $g_i(\theta_1, \theta_2)$ *is differentiable in* θ_2 *at* θ_{20} $\forall \theta_1 \in \Theta_1$ *and* $(\partial/\partial\theta_2')g_i(\theta_1, \theta_{20})$ *is differentiable in* θ_1 $\forall \theta_1 \in \Theta_1$ *for all sample realizations. Then, the two-step* AR/LM *and* AR/QLR1 *subvector tests satisfy*

(a) $AsySz \leq \alpha$ *for the null parameter space* $\mathcal{F}_{AR/LM,QLR1}$,

(b) $AsyNRP = \alpha$ *for all null sequences* S *in* $\mathcal{F}_{AR/LM,QLR1}$ *that satisfy Assumption* SI2,

(c) $AsySz = \alpha$ *provided some null sequence* S *in* $\mathcal{F}_{AR/LM,QLR1}$ *satisfies Assumption* SI2,

(d) *for any null sequence* S *in* $\mathcal{F}_{AR/LM,QLR1}$ *that satisfies Assumption* SI2, $\varphi_{2n}^{AR/LM} = \phi_{2n}^{LM}(\theta_{*n}, \alpha) + o_p(1)$, $\lim P_{\theta_{*n},F_n}(\varphi_{2n}^{AR/LM} > 0) = \lim P_{\theta_{*n},F_n}(\phi_{2n}^{LM}(\theta_{*n}, \alpha) > 0)$, *and analogous results hold for* $\varphi_{2n}^{AR/QLR1}$ *and* $\phi_{2n}^{QLR1}(\theta_{1*n}, \alpha)$,

(e) *for any alternative sequence* $S^A = \{(\theta_{*n}^A, F_n^A) : n \geq 1\}$ *that satisfies Assumption* C(iii) *and is contiguous to a null sequence* S *that satisfies Assumption* OE, $\varphi_{2n}^{AR/LM} = \phi_{2n}^{LM}(\theta_{*n}, \alpha) + o_p(1)$, $\lim P_{\theta_{*n}^A,F_n^A}(\varphi_{2n}^{AR/LM} > 0) = \lim P_{\theta_{*n}^A,F_n^A}(\phi_{2n}^{LM}(\theta_{*n}, \alpha) > 0)$ *and analogous results hold for* $\varphi_{2n}^{AR/QLR1}$ *and* $\phi_{2n}^{QLR1}(\theta_{1*n}, \alpha)$, *and*

(f) *under the assumptions stated before part* (a) *of the Theorem, plus* Θ_* *is bounded and satisfies*

$B(\Theta_*, \varepsilon) \subset \Theta$ *for some* $\varepsilon > 0$, *the two-step* AR/LM *and* AR/QLR1 CS's *satisfy* (i) $AsySz \geq 1-\alpha$ *for the parameter space* $\mathcal{F}_{\Theta,AR/LM,QLR1}$, (ii) $AsyCP = 1-\alpha$ *for all sequences* S *in* $\mathcal{F}_{\Theta,AR/LM,QLR1}$ *that satisfy Assumption* SI2, (iii) $AsySz = 1-\alpha$ *provided some sequence* S *in* $\mathcal{F}_{\Theta,AR/LM,QLR1}$ *satisfies Assumption* SI2, (iv) *for any sequence* S *in* $\mathcal{F}_{\Theta,AR/LM,QLR1}$ *that satisfies Assumption* SI2, $\varphi_{2n}^{AR/LM}(\theta_{2*n}) = \phi_{2n}^{LM}(\theta_{*n}, \alpha) + o_p(1)$, $\lim P_{\theta_{*n},F_n}(\varphi_{2n}^{AR/LM}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n},F_n}(\phi_{2n}^{LM}(\theta_{*n}, \alpha) > 0)$, *and analogous results for* $\varphi_{2n}^{AR/QLR1}(\theta_{2*n})$ *and* $\phi_{2n}^{QLR1}(\theta_{*n}, \alpha)$, *and* (v) *for any alternative sequence* S^A *that satisfies Assumption* C(iii) *and is contiguous to a null sequence* S *that satisfies Assumption* OE, $\varphi_{2n}^{AR/LM}(\theta_{2*n}) = \phi_{2n}^{LM}(\theta_{*n}, \alpha) + o_p(1)$, $\lim P_{\theta_{*n}^A,F_n^A}(\varphi_{2n}^{AR/LM}(\theta_{2*n}) > 0) = \lim P_{\theta_{*n}^A,F_n^A}(\phi_{2n}^{LM}(\theta_{*n}, \alpha) > 0)$, *and analogous results hold for* $\varphi_{2n}^{AR/QLR1}(\theta_{2*n})$ *and* $\phi_{2n}^{QLR1}(\theta_{*n}, \alpha)$.

Comments: (i). Theorem 8.2(d) and (e) show that, under global strong identification, $\varphi_{2n}^{AR/LM}$ and $\varphi_{2n}^{AR/QLR1}$ are asymptotically equivalent to the oracle second-step C(α)-LM test $\phi_{2n}^{LM}(\theta_{1*n}, \alpha)$ and the oracle second-step C(α)-QLR1 test $\phi_{2n}^{QLR1}(\theta_{1*n}, \alpha)$, respectively, under the null hypothesis and contiguous local alternatives. The latter tests are asymptotically efficient in a GMM sense, e.g., as defined in Newey and West (1987), under global strong identification when $k \geq p$. Hence, the two-step $\varphi_{2n}^{AR/LM}$ and $\varphi_{2n}^{AR/QLR1}$ tests are as well.

(ii). The proof of Theorem 8.2 in the SM employs Theorem 5.1. In the proof, we show that sequences S for which $\lim_{n\to\infty} \tau_n^\Phi < K_L$ (and some other conditions hold) satisfy Assumption B of Section 5. We show that sequences S for which $\lim_{n\to\infty} \tau_n^\Phi > K_U^*$ (and some other conditions hold) satisfy Assumption C of Section 5, where $K_U^* < K_L$ by assumption. We also show that sequences S that satisfy Assumption SI2 (and some other conditions) satisfy Assumption OE of Section 5.

(iii). The results of Theorem 8.2 also hold when $WI_n(\theta) := 0$, which yields pure C(α)-LM and C(α)-QLR1 tests, provided $\mathcal{F}_{AR/LM,QLR1}$ in (8.10) is defined to include the condition in (8.12). (This result is proved in the SM.)

(iv) Time series versions of Theorems 8.1 and 8.2 are given in the SM.

9 Finite-Sample Simulations

9.1 Heteroskedastic Linear IV Model

9.1.1 Simulation Set-up

In this section, we consider a heteroskedastic linear IV model with two rhs endogenous variables. We consider tests concerning the coefficient on the second rhs endogenous variable Y_{2i}. The coefficient on the first rhs endogenous variable is a nuisance parameter. The model and sample

moment vector are

$$\begin{aligned} Y_i &= Y_{1i}\theta_1 + Y_{2i}\theta_2 + U_i, \\ Y_{ji} &= Z_i'(\pi_j/n^{1/2}) + V_{ji} \text{ for } j = 1, 2, \text{ and} \\ g_i(\theta) &:= (Y_i - Y_{1i}\theta_1 - Y_{2i}\theta_2)Z_i, \end{aligned} \tag{9.1}$$

where $(U_i, V_{1i}, V_{2i})' = ((||Z_i||/k^{1/2})\varepsilon_{Ui},\ (||Z_i||/k^{1/2})\varepsilon_{1i},\ (||Z_i||/k^{1/2})\varepsilon_{2i})'$, $(\varepsilon_{Ui},\ \varepsilon_{1i},\ \varepsilon_{2i})' \sim$ i.i.d. $N(0^k, V)$ for $V \in R^{3\times 3}$ with $V_{jj} = 1\ \forall j \leq 3$, $V_{1j} = .8$ for $j = 2, 3$, and $V_{23} = .3$, $Z_i \sim$ i.i.d. $N(0^k, I_k)$ independent of $(\varepsilon_{Ui}, \varepsilon_{1i}, \varepsilon_{2i})'$, $\pi_1 = ||\pi_1||1^k/k^{1/2}$ for some $||\pi_1||$, $\pi_2 = ||\pi_2||\pi_2^*/k^{1/2}$ for $\pi_2^* = (1^{k/2\prime}, -1^{k/2\prime})'$ and some $||\pi_2||$, and k is an even number. The coefficient vectors $\pi_j/n^{1/2}$ on Z_i in the reduced-form equations are scaled by $n^{-1/2}$. This is innocuous to the finite-sample results. It is done only to facilitate the assessment of the effect of n on power. If the asymptotic results are accurate, power should not be sensitive to n with this rescaling. Similarly, the π_j vectors are scaled by $k^{-1/2}$ to ensure that the expected concentration parameter $E\pi_j'Z_iZ_i'\pi_j/n = ||\pi_j||^2/n$ does not depend on k, which facilitates the assessment of the effect of k on power.

The hypotheses are $H_0 : \theta_2 = \theta_{20}$ and $H_1 : \theta_2 \neq \theta_{20}$. The NRP's and power of the tests considered are invariant wrt θ_1 and equivariant wrt θ_2. In consequence, without loss of generality, we take $\theta_1 = 0$ and $\theta_{20} = 0$.

The tests considered include the two-step AR/AR, AR/LM, and AR/QLR1 tests defined in Section 7. We also consider (i) the Oracle C(α)-QLR1 test, which is the infeasible C(α)-QLR1 test $\phi_{2n}^{QLR1}(\theta_1, \alpha)$ (defined in (7.18)) evaluated at the true value of θ_1, and (ii) the projection (non-C(α)) conditional QLR1 test, which is denoted by Proj-QLR1.[10] The Oracle C(α)-QLR1 test is used to assess the effect of not knowing θ_1 on the power of the two-step AR/QLR1 test. The (non-C(α)) Proj-QLR1 test is considered because it is the existing test in the literature that is closest to the AR/QLR1 two-step test. We do not report results for the Oracle C(α)-AR, Oracle C(α)-LM, Proj-AR, or Proj-LM tests because they have lower power than the corresponding QLR1 tests. For the case of strong identification, we also consider the two-stage least-squares (2SLS) test.[11] The nominal size of the tests is .05.

For NRP's and power, we consider four identification cases: (i) $||\pi_1|| = ||\pi_2|| = 40$ (strong

[10] The (non-C(α)) QLR1 test statistic is $QLR1_{2n}(\theta)$ (defined in (7.14)) with $AR_n(\theta)$ (defined in (7.1)) in place of $AR_{2n}^{\dagger}(\theta)$, with $LM_{2n}(\theta)$ (defined in (7.13)) defined using the weight matrix $P_{\widehat{\Omega}_n^{-1/2}(\theta)(\widehat{D}_{1n}(\theta):\widehat{D}_{2n}(\theta))}$ in place of $P_{D_{2n}^{\dagger}(\theta)}$, and with $rk_{2n}(\theta)$ (defined in (7.15)) defined with $[\widehat{D}_{1n}(\theta) : \widehat{D}_{2n}(\theta)]$ and $[\widehat{\Phi}_{1n}(\theta) : \widehat{\Phi}_{2n}(\theta)]$ in place of $\widehat{D}_{2n}(\theta)$ and $\widehat{\Phi}_{2n}(\theta)$, respectively. Its conditional critical value is given by the $1-\alpha$ quantile of $QLR1(r, 0)$ defined in (7.16)) with p in place of p_2 and evaluated at $r = rk_{2n}(\theta)$. The (non-C(α)) Proj-QLR1 test rejects H_0 only if it rejects H_0 when evaluated at $\theta = (\theta_1', \theta_{20}')'$ for all $\theta_1 \in R$.

[11] The 2SLS test is not considered in the other cases, because it is not identification robust and, hence, over-rejects in these cases.

identification of θ_1 and θ_2), (ii) $||\pi_1|| = ||\pi_2|| = 4$ (weak identification of θ_1 and θ_2), (iii) $||\pi_1|| = 4$ and $||\pi_2|| = 40$, and (iv) $||\pi_1|| = 40$ and $||\pi_2|| = 4$. For each case, we consider power for $\theta_2 \in [-B, B]$ for B chosen suitably.

The results are for sample size $n = 250$ and $k = 4$ IV's, except in Table III. For the two-step tests, we use $\alpha_1 = .005$ for first-step CS, $K_L = K_U = .05$ for second-step significance level, and $K_{rk} = 1$ for the QLR1 rank statistic. These are referred to as the base case values. A sensitivity analysis of the results to these choices is provided in Table II. The data-dependent critical values are taken from a look-up table that was simulated using $500,000$ simulation repetitions. The number of simulations employed for the rejection probabilities is $10,000$, except in Table I, which employs $25,000$ repetitions for the NRP's. The grid used for the first-step CI values of θ_1 is $[-3, 3]$ with a grid width of .1. In the tables, the base case values of n, k, and the tuning parameters is indicated by bold face. In the figures, the power of the Oracle C(α)-QLR1 and 2SLS tests are NRP-corrected because they over-reject somewhat in finite samples. The power for the other tests are not NRP-corrected because they do not over-reject.

9.1.2 Simulation Results

Figures 1 and 2 provide finite-sample power curves for identification cases (i)–(iv). In Figure 1 with strong identification (top), the power curves for the AR/LM (yellow), AR/QLR1 (blue), and Oracle C(α)-QLR1 (red) tests are high and are on top of each other. The 2SLS power curve (circles) is quite similar, but with somewhat lower power for negative θ_2 values and somewhat higher power for positive θ_2 values. The Proj-QLR1 (black) and AR/AR (green) tests have noticeably lower power than the other tests.

In Figure 1 with weak identification (bottom), the AR/AR and AR/QLR1 tests have equal power and have the highest power of the feasible tests. The AR/LM test has the lowest power of all of the tests for negative θ_2 values, while the Proj-QLR1 has the lowest power for positive θ_2 values. The Oracle C(α)-QLR1 test has noticeably higher power than any of the feasible tests. This is not surprising, because weak identification of θ_1 implies that knowledge of the true value of θ_1 is quite valuable. Note that the scales of the θ_2 axes in the two graphs in Figure 1 are quite different. This reflects the differing amounts of information available about θ_2 in these two cases.

In Figure 2 top, the AR/QLR1 and Oracle C(α)-QLR1 tests have equal power—due to the strong identification of θ_1. The AR/AR test has similar power, but its power is lower for negative θ_2 values where the power curves are steep. The AR/LM test has poor (quirky) power for negative θ_2 values, but the highest power of all of the tests for positive θ_2 values. The Proj-QLR1 test has the lowest power of all of the tests except for the AR/LM test for some of the negative θ_2 values.

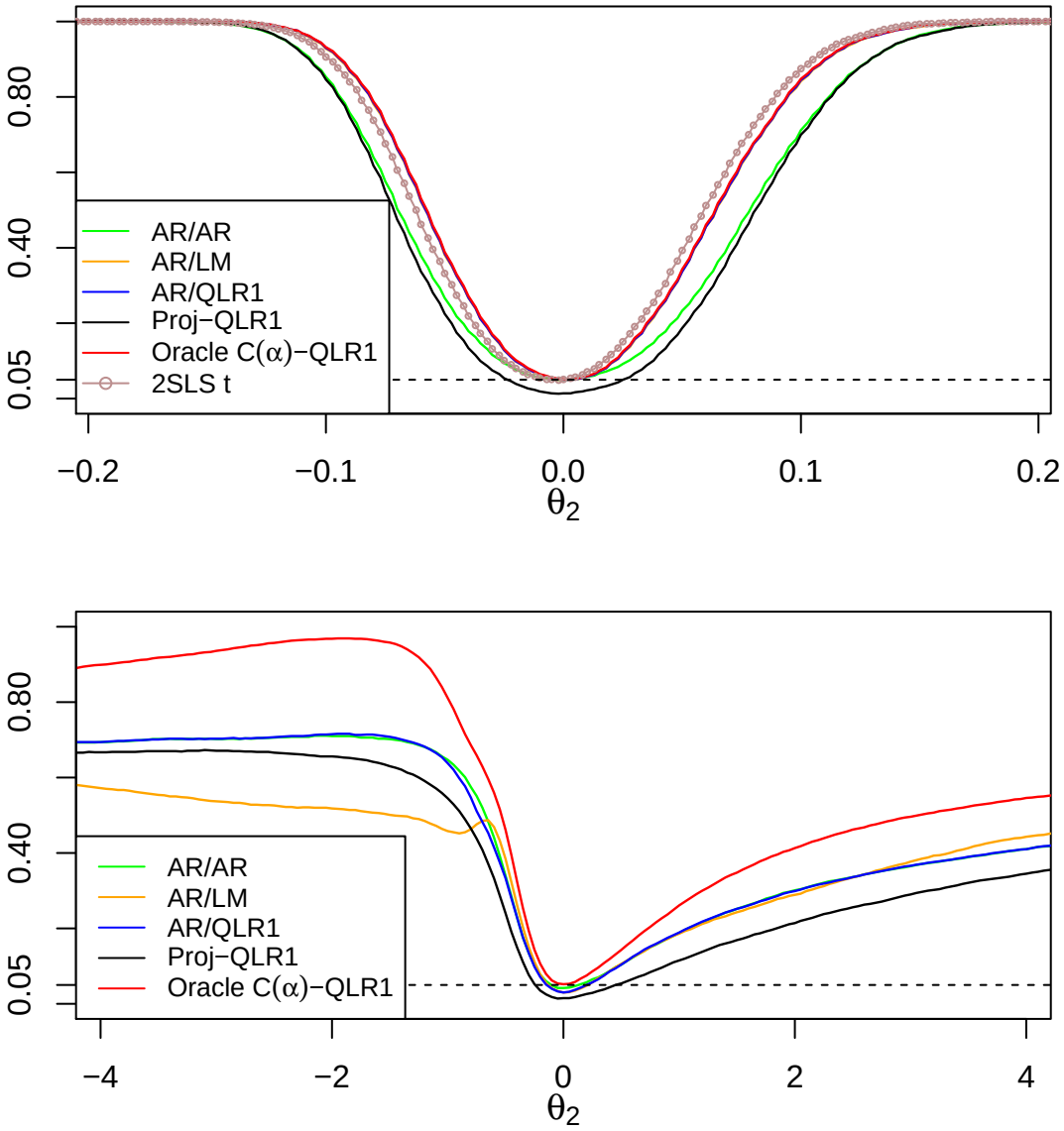


Figure 1. Heteroskedastic Linear IV: Power for n=250, k=4 with $(||\pi_1||,||\pi_2||)=(40,40)$ (top) and $(||\pi_1||,||\pi_2||)=(4,4)$ (bottom)

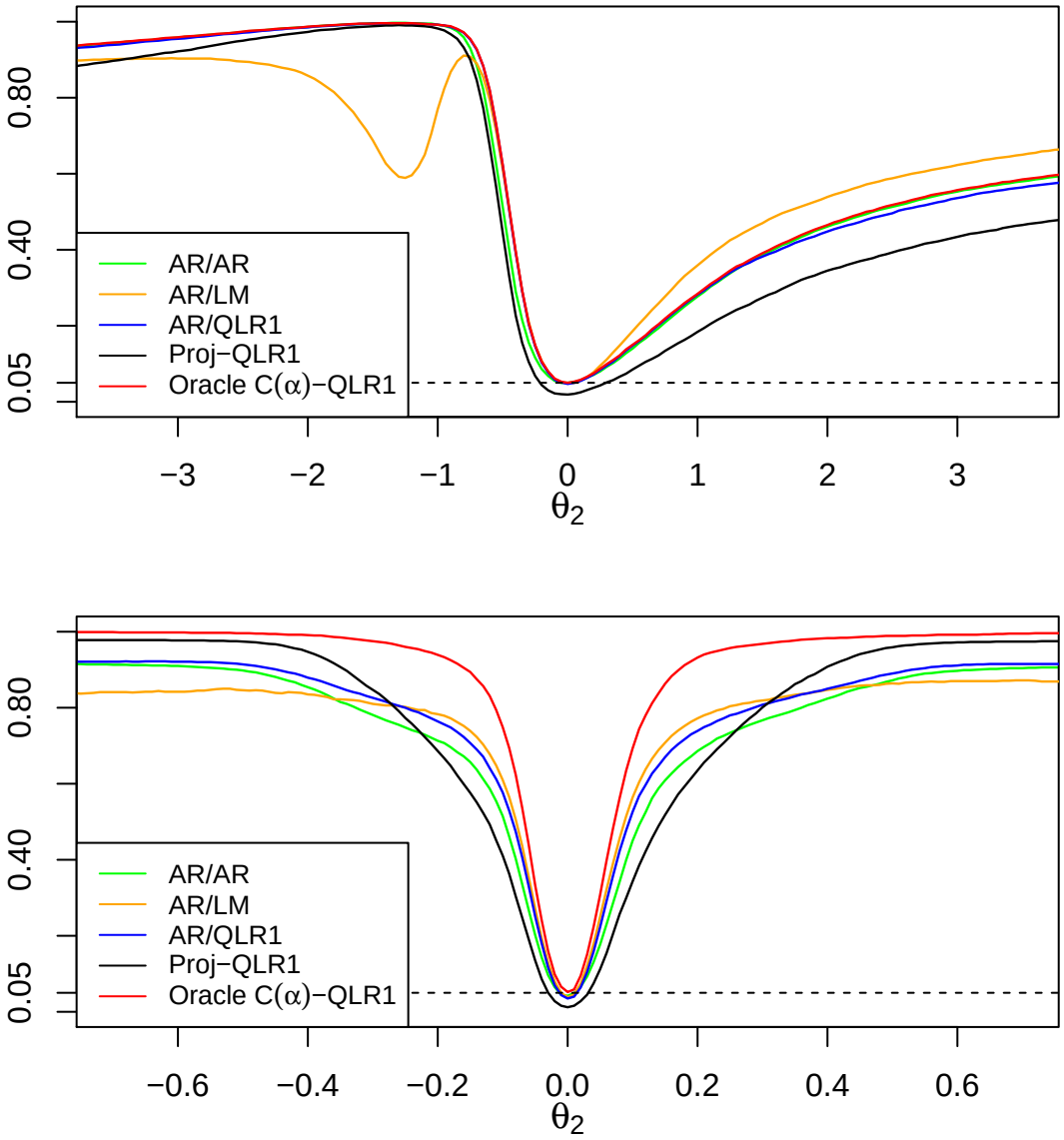


Figure 2. Heteroskedastic Linear IV: Power for n=250, k=4 with $(||\pi_1||,||\pi_2||)=(40,4)$ (top) and $(||\pi_1||,||\pi_2||)=(4,40)$ (bottom)

In Figure 2 bottom, the Oracle C(α)-QLR1 test has the highest power by a substantial margin. This is because θ_1 is weakly identified in this case. The ranking of the other tests' power has the interesting feature that it is reversed between θ_2 values where power is $\leq .80$ and $\geq .80$. In the former case, the ranking from highest to lowest power is AR/LM, AR/QLR1, AR/AR, and Proj-QLR1. In the latter case, it is the reverse.

In conclusion, the two-step AR/QLR1 subvector test performs the best in terms of power among the feasible tests in Figures 1 and 2. It noticeably out-performs the Proj-QLR1 test. We now investigate its NRP's and the sensitivity of its NRP's and power to the tuning parameters.

Table I provides NRP's for the nominal .05 AR/QLR1 test for $n = 100, 250$ and $||\pi_1||, ||\pi_2|| \in \{40, 20, 12, 4, 0\}$. The results show that the NRP's vary between .000 and .052 over these cases. The NRP's are in [.043, .052] for $||\pi_1|| \geq 12$ and all $||\pi_2||$ values. They are in [.000, .039] for $||\pi_1|| \leq 4$ and all $||\pi_2||$ values. Hence, the finite-sample size of the AR/QLR1 test is close to its nominal size and it under-rejects the null noticeably only for $||\pi_1|| \leq 4$.

TABLE I. NRP's of the Nominal .05 AR/QLR1 Test for $k = 4$, $N = 100$ and 250, and Base Case Tuning Parameters in the Heteroskedastic Linear Instrumental Variables Model

		$n = 100$					$n = 250$				
	$\|\|\pi_2\|\|$:	40	20	12	4	0	40	20	12	4	0
	40	.046	.045	.046	.048	.052	.049	.049	.049	.047	.046
	20	.045	.044	.044	.046	.050	.049	.049	.049	.046	.045
$\|\|\pi_1\|\|$	12	.044	.043	.043	.044	.049	.048	.048	.048	.044	.044
	4	.025	.025	.025	.029	.039	.033	.032	.030	.030	.037
	0	.000	.001	.001	.001	.001	.000	.000	.000	.001	.001

Table II investigates the sensitivity of the NRP and power of the nominal .05 AR/QLR1 test to the tuning parameters α_1, K_L ($= K_U$), K_{rk}, K_L^*, and a for identification cases (i)–(iv) and five values of θ_2 including the null value zero, two negative values, and two positive values, which are chosen (differently in different scenarios) to yield power around .80 and .50 (when the identification strength is sufficient to yield such power).

In Table II, for changes in α_1 (where $1-\alpha_1$ is the first-step CI nominal level), there is very little sensitivity of the NRP's. There is some sensitivity of power for some θ_2 values in cases (ii)–(iv) with power decreasing as α_1 is increased from its base case value of .005 and power being relatively insensitive to reductions of α_1 from its base case value. The base case value works well in an overall sense. For K_L, K_L^*, and a, there is very little or no sensitivity of NRP's or power. For K_{rk}, in case (i), there is no sensitivity of NRP's or power; for case (ii), there is a little sensitivity of NRP's, a

noticeable drop in power for $\theta_2 = -1.05$ as K_{rk} increased from its base case value to the largest K_{rk} value, and little sensitivity in power for other θ_2 values. For K_{rk}, in case (iii), there is no sensitivity of NRP's or power except for $\theta_2 = -.45$, where power drops noticeably for the smallest K_{rk} value; and in case (iv), there is no sensitivity of NRP's, but sensitivity of power with power generally increasing in K_{rk} and power at the base case K_{rk} value being in the middle of the range.

TABLE II. Sensitivity of NRP and Power of the Nominal .05 AR/QLR1 Test to the Tuning Parameters α_1, K_L, K_{rk}, K_L^*, and a for $(||\pi_1||, ||\pi_2||) = (40, 40)$, $(4, 4)$, $(40, 4)$, and $(4, 40)$ and for Five Values of θ_2 in the Heteroskedastic Linear Instrumental Variables Model

Tuning			$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (40, 40)$					$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (4, 4)$				
Parameter		θ_2 :	.00	-.082	-.059	.064	.094	.00	-1.05	-.575	2.00	3.75
	.0010		.050	.804	.510	.501	.799	.031	.663	.406	.314	.417
	.0025		.050	.804	.510	.501	.799	.031	.660	.402	.309	.410
α_1	**.0050**		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	.0100		.050	.804	.511	.502	.799	.029	.638	.385	.280	.380
	.0150		.050	.805	.512	.502	.800	.028	.619	.372	.258	.360
	.01		.050	.804	.511	.501	.799	.032	.666	.409	.317	.420
K_L	**.05**		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	.10		.050	.804	.511	.501	.799	.029	.651	.397	.299	.401
	.25		.051	.799	.503	.483	.780	.039	.657	.412	.299	.400
	.50		.050	.805	.510	.498	.797	.036	.659	.405	.299	.401
K_{rk}	**1.0**		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	2.0		.050	.803	.510	.500	.799	.027	.633	.393	.301	.403
	4.0		.050	.803	.510	.501	.798	.028	.593	.397	.305	.404
	.001		.050	.804	.511	.501	.799	.031	.654	.398	.299	.401
K_L^*	**.005**		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	.010		.050	.804	.511	.501	.799	.030	.653	.397	.299	.401
	.00		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
a	$\mathbf{10^{-6}}$		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	.01		.050	.804	.511	.501	.799	.030	.652	.393	.300	.401

TABLE II. (cont.)

Tuning			$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (40, 4)$					$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (4, 40)$				
Parameter		θ_2 :	.00	-.59	-.45	1.65	4.00	.00	-.250	-.085	.095	.290
	.0010		.047	.803	.504	.412	.599	.033	.804	.497	.493	.806
	.0025		.047	.803	.505	.408	.595	.035	.802	.502	.500	.806
α_1	**.0050**		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	.0100		.047	.804	.505	.392	.567	.034	.791	.494	.489	.793
	.0150		.047	.804	.505	.382	.547	.033	.781	.481	.476	.783
	.01		.047	.803	.505	.413	.602	.037	.811	.517	.517	.814
K_L	**.05**		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	.10		.045	.803	.504	.394	.586	.034	.799	.500	.500	.802
	.25		.050	.769	.458	.401	.586	.033	.766	.471	.469	.774
	.50		.047	.794	.489	.402	.586	.034	.780	.485	.480	.784
K_{rk}	**1.0**		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	2.0		.049	.799	.512	.405	.587	.036	.811	.528	.530	.818
	4.0		.048	.793	.510	.410	.588	.038	.815	.541	.543	.828
	.001		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
K_L^*	**.005**		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	.010		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	0		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
a	$\mathbf{10^{-6}}$		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	.01		.047	.803	.505	.403	.586	.035	.798	.502	.500	.801

Overall, the base case value of K_{rk} performs well.

Table III investigates the sensitivity of the NRP and power of the nominal .05 AR/QLR1 test to the sample size $n \in \{50, 100, 250, 500, 1000\}$ and the number of IV's $k \in \{4, 8, 12\}$. In Table III, the NRP's are insensitive to n for $n \geq 100$ and slightly lower for $n = 50$. The NRP's are close to .05 in cases (i) and (iii), which both have $||\pi_1|| = 40$, and less than .05 in the other two cases, uniformly across n. In Table III, power increases from $n = 50$ to 100 and in some cases to 200 (even with the $n^{-1/2}$ scaling of the coefficients on Z_i). Power is stable for larger values of n.

In Table III, the NRP's vary with k, but there are no clear patterns. NRP's increase with k in

case (iii), with some over-rejection, .059, for $k = 12$, but the NRP's decrease with k in case (iv). In Table III, power is strongly decreasing in k in cases (ii)–(iv), but not in case (i).

In conclusion, the simulations show that the AR/QLR1 test performs best in terms of power of the feasible tests considered across all four identification scenarios. Its power is essentially equivalent to that of the Oracle C(α)-QLR1 and 2SLS tests under strong identification. The NRP's of the AR/QLR1 test are close to its nominal level for $||\pi_1|| \geq 12$ and over-rejection of the null as large as .059 is detected only in one case, when $k = 12$. The AR/QLR1 test exhibits some sensitivity to

TABLE III. Sensitivity of NRP and Power of the Nominal .05 AR/QLR1 Test to the Sample Size, n, and Number of Instruments, k, for $(||\pi_1||, ||\pi_2||) = (40, 40)$, $(4, 4)$, $(40, 4)$, and $(4, 40)$ and for Five Values of θ_2 in the Heteroskedastic Linear Instrumental Variables Model

			$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (40, 40)$					$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (4, 4)$				
		θ_2 :	.00	-.082	-.059	.064	.094	.00	-1.05	-.575	2.00	3.75
	50		.042	.672	.417	.399	.671	.023	.578	.341	.257	.353
n	100		.048	.765	.478	.480	.772	.032	.648	.400	.292	.396
	250		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
	500		.050	.804	.513	.514	.812	.028	.657	.388	.296	.404
	1000		.052	.811	.516	512	.807	.033	.657	.394	.292	.398
	4		.050	.804	.511	.501	.799	.030	.654	.398	.299	.401
k	8		.044	.846	.553	.536	.841	.036	.589	.333	.226	.330
	12		.045	.844	.549	.527	.826	.043	.524	.291	.177	.269
			$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (40, 4)$					$(\|\|\pi_1\|\|, \|\|\pi_2\|\|) = (4, 40)$				
		θ_2 :	.00	-.59	-.45	1.65	4.00	.00	-.250	-.085	.095	.290
	50		.045	.676	.422	.385	.546	.019	.743	.365	.362	.739
n	100		.051	.773	.489	.423	.589	.025	.784	.456	.450	.774
	250		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
	500		.047	.809	.514	.396	.595	.037	.807	.521	.524	.809
	1000		.047	.817	.521	.394	.596	.040	.810	.548	.538	.816
	4		.047	.803	.505	.403	.586	.035	.800	.503	.502	.802
k	8		.053	.793	.480	.341	.522	.017	.744	.429	.432	.743
	12		.059	.730	.412	.301	.468	.012	.671	.363	.363	.667

the choice of K_{rk}, a little sensitivity to the choice of α_1, but little or no sensitivity to the choices of K_L, K_L^*, and a. Even for sample sizes as small as 50, the AR/QLR1 test has NRP's less than its nominal level. However, its power at this small a sample size is below what the asymptotic results suggest. Overall, the AR/QLR1 test seems to perform quite well in this model for the parameter scenarios considered.

9.2 Nonlinear IV Model

9.2.1 Simulation Set-up

Next, we consider an IV model with one rhs endogenous variable that enters nonlinearly:

$$\begin{aligned} Y_i &= Y_{1i}\gamma_1 + Y_{1i}^2\gamma_2 + U_i, \\ Y_{1i} &= Z_{1i}'(\pi/n^{1/2}) + V_i, \end{aligned} \tag{9.2}$$

where $(U_i, V_i) \sim$ i.i.d. $N(0, V)$ for $V \in R^{2\times 2}$ with $V_{jj} = 1$ for $j = 1, 2$ and $V_{12} = .8$, $Z_i = (Z_{1i}', Z_{2i}')' \in R^k$, $Z_{1i} = (Z_{11i}, ..., Z_{1(k/2)i})' \sim$ i.i.d. $N(0^{k/2}, I_{k/2})$ independent of (U_i, V_i), $Z_{2i} = (Z_{11i}^2, ..., Z_{1(k/2)i}^2)' \in R^{k/2}$, and $\pi = ||\pi||1^k/k^{1/2}$. The errors are homoskedastic. The coefficient vector $\pi/n^{1/2}$ on Z_i in the reduced-form equation is scaled by $n^{-1/2}$ and π is scaled by $k^{-1/2}$ for the same reasons as in Section 9.1.

We consider hypotheses concerning the value and derivative of the quadratic structural function $y_1\gamma_1 + y_1^2\gamma_2$ at the point $y_1 = 2$. For the function value, we set $\theta_1 := y_1\gamma_1$ and $\theta_2 := y_1\gamma_1 + y_1^2\gamma_2$ and test $H_0 : \theta_2 = \theta_{20}$. That is, we transform the parameters from (γ_1, γ_2) to (θ_1, θ_2) and the structural equation to $Y_i = Y_{1i}^*\theta_1 + Y_{2i}^*\theta_2 + U_i$, where $Y_{1i}^* := Y_{1i}/y_1 - Y_{1i}^2/y_1^2$ and $Y_{2i}^* := Y_{1i}^2/y_1^2$. For the function derivative, we set $\theta_1 := \gamma_1$ and $\theta_2 := \gamma_1 + 2y_1\gamma_2$, test $H_0 : \theta_2 = \theta_{20}$, and the transformed structural equation has $Y_{1i}^* := Y_{1i} - Y_{1i}^2/(2y_1)$ and $Y_{2i}^* := Y_{1i}^2/(2y_1)$. In both cases, the moment vector is

$$g_i(\theta) := (Y_i - Y_{1i}^*\theta_1 - Y_{2i}^*\theta_2)Z_i, \tag{9.3}$$

but with different definitions of (Y_{1i}^*, Y_{2i}^*). The NRP's and power of the tests considered are invariant wrt θ_1 and equivariant wrt θ_2. In consequence, without loss of generality, we take the true value of θ_1 to be zero, the null value θ_{20} to be zero, and test the hypotheses $H_0 : \theta_2 = 0$ versus $H_1 : \theta_2 \neq 0$.[12]

The same tests, base case tuning parameters, and simulation repetition numbers are used as for the linear IV model. As above, $\alpha = .05$. Figures 3 and 4 are for $n = 500$ and $k = 4$.

For NRP's and power, we consider two identification cases: (i) $||\pi|| = 50$ (strong identification)

[12] The hypothesis $H_0 : \theta_2 = 0$ is obtained by replacing Y_i by $Y_i - Y_{2i}^*\theta_{20}$ and θ_2 by $\theta_2 - \theta_{20}$.

and (ii) $||\pi|| = 4$ (weak identification). For each case, we consider power for $\theta_2 \in [-B, B]$ for B chosen suitably.

9.2.2 Simulation Results

Figures 3 and 4 provide finite-sample power curves for identification cases (i) and (ii) for the hypotheses that concern the value and derivative of the structural function, respectively.

In Figure 3 for strong identification (top), the AR/LM, AR/QLR1, and 2SLS tests have equal and highest power for negative θ_2 values. The power of the Oracle C(α)-QLR1 test is similar, but slightly lower for some negative θ_2 values. For positive θ_2 values, the 2SLS test clearly has the highest power, while the AR/LM, AR/QLR1, and Oracle C(α)-QLR1 have equal, but lower power than 2SLS. Note that the power curves of the tests are not symmetric about $\theta_2 = 0$ (including 2SLS, but to a lesser extent than the other tests). This indicates that the values of n and $||\pi||$ are not sufficiently large for the strong-identification normal approximation to be highly accurate (although this does not cause over-rejection under H_0). The power curves of the AR/AR and Proj-QLR1 tests are noticeably below those of the other tests, as is expected in this case.

In Figure 3 for weak identification (bottom), the Oracle C(α)-QLR1 test has the highest power for all θ_2 values. The AR/QLR1 test has equal power to it for negative θ_2 values, but noticeably lower power for positive θ_2 values. The AR/LM test has quirky, low power for some negative θ_2 values, but relatively high power for positive θ_2 values. The AR/AR test has somewhat lower power than the AR/QLR1 test. The Proj-QLR1 has noticeably lower power than the AR/QLR1 test for all θ_2 values.

Figure 4 for the derivative of the structural function is quite similar to Figure 3. This is due to the similarity of the transformed parameters θ_1 and θ_2 in these two cases.

Table IV provides NRP's for the nominal .05 AR/QLR1 test for a range of values of $||\pi||$, n, and k with homoskedastic errors, and in one case heteroskedastic errors (with the same form of heteroskedasticity as in (9.1) for the linear IV model). The table shows that the NRP's vary between .007 and .050 over these cases. The lowest NRP's occur for $||\pi|| = 0$. In the base case scenario, $n = 500$ and $k = 4$, the NRP's are in $[.038, .042]$ for $||\pi|| \geq 4$.

The SM provides tables that are analogous to Tables II and III, which concern sensitivity of the AR/QLR1 test to the tuning parameters, as well as to n and k, but for the nonlinear IV model with hypotheses concerning the structural function and its derivative. Broadly speaking, the results are similar to those in Tables II and III.

Overall, the AR/QLR1 test performs well in terms of NRP's and power in the nonlinear IV model for the parameter scenarios considered.

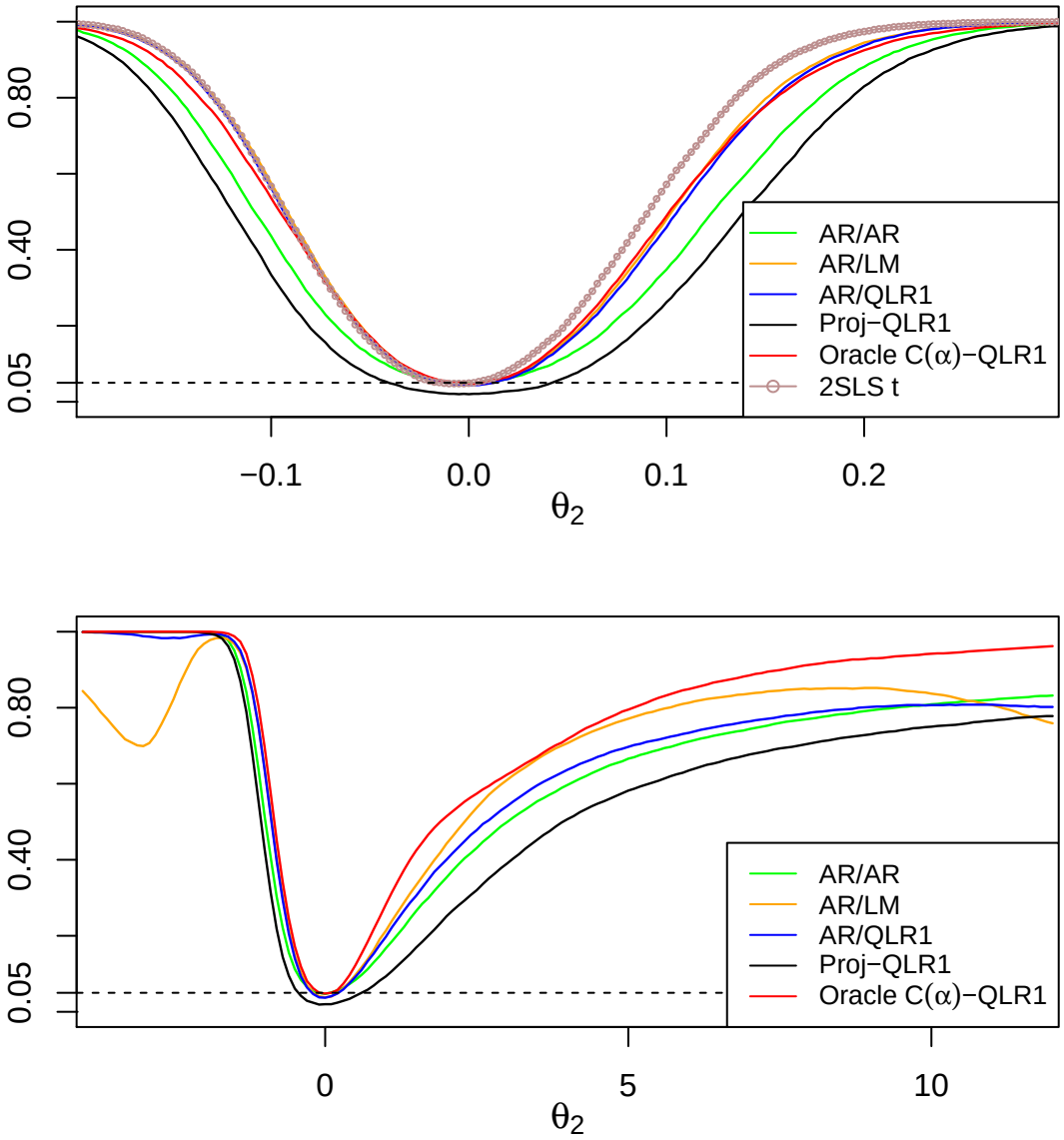


Figure 3. Nonlinear IV Model, Structural Function: Power for n=500, k=4, $||\pi||$=50 (top) and $||\pi||$=4 (bottom)

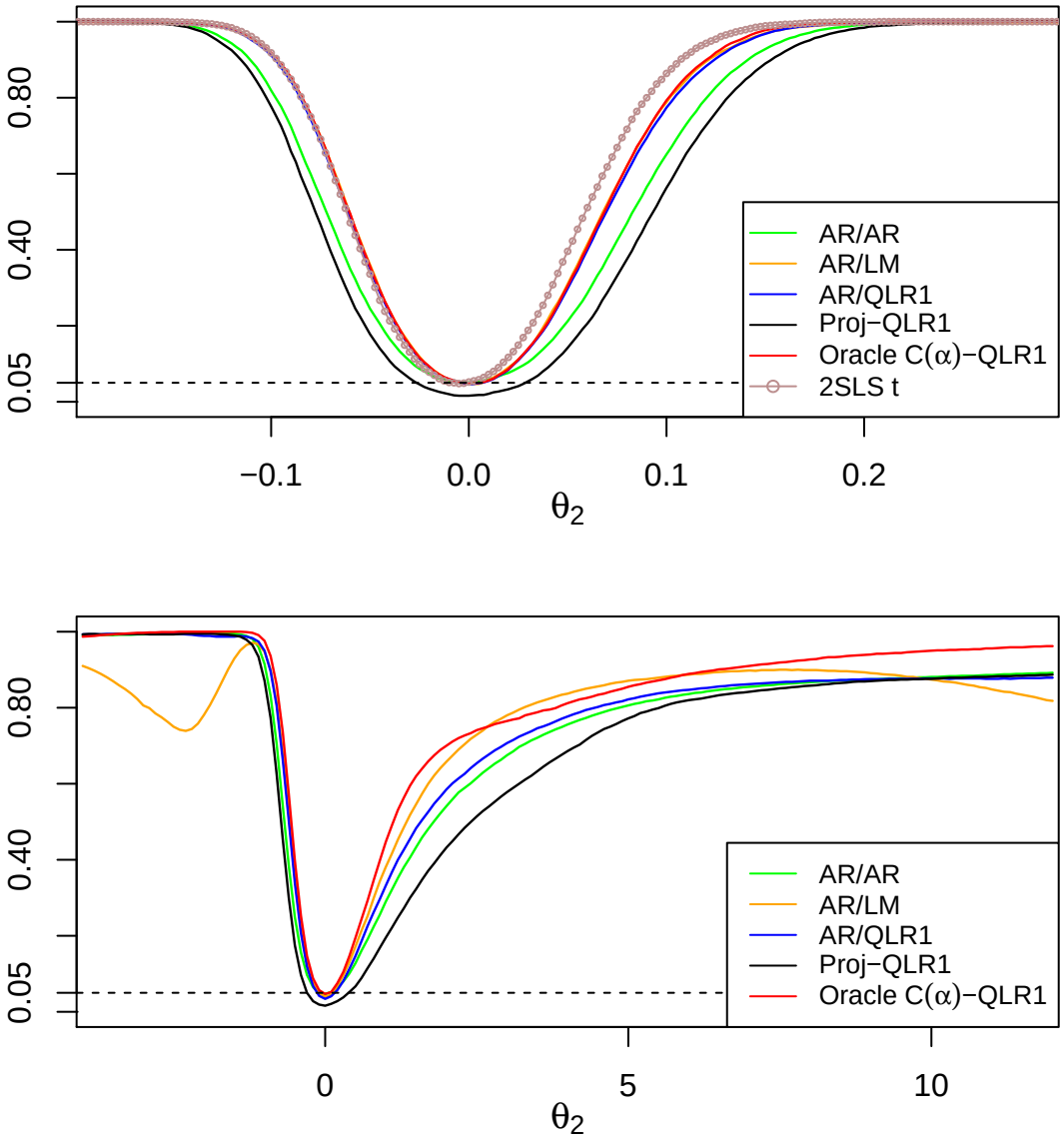


Figure 4. Nonlinear IV Model, Derivative of Structural Function: Power for n=500, k=4, ||π||=50 (top) and ||π||=4 (bottom)

TABLE IV. NRP's of the Nominal .05 AR/QLR1 Test for Base Case Tuning Parameters for Inference on the Structural Function at $y_1 = 2$ in the Nonlinear Instrumental Variables Model

k	n	Errors	$\|\|\pi\|\|$:	100	75	50	35	20	14	8	4	0
4	50	Homoskedastic		.032	.031	.026	.021	.018	.019	.019	.018	.009
4	100	Homoskedastic		.040	.039	.036	.033	.033	.033	.032	.027	.017
4	250	Homoskedastic		.041	.041	.041	.040	.039	.039	.038	.035	.024
4	**500**	**Homoskedastic**		.042	.043	.045	.044	.043	.042	.039	.038	.026
8	100	Homoskedastic		.050	.050	.046	.043	.044	.043	.041	.035	.025
8	250	Homoskedastic		.043	.043	.044	.044	.045	.044	.042	.039	.035
4	250	Heteroskedastic		.032	.030	.027	.025	.021	.018	.013	.009	.007

References

Andrews, D. W. K. (2017): "Identification-Robust Subvector Inference for Likelihood-Based Models," manuscript in preparation, Cowles Foundation, Yale University.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica,* 80, 2153–2211. Supplemental Material is available at *Econometrica Supplemental Material,* 80, http://www.econometricsociety.org/ecta/Supmat/9456_miscellaneous.pdf.

——— (2013): "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," *Journal of Econometrics,* 173, 36–56. Supplemental Material is available with Cowles Foundation Discussion Paper No. 1824R, 2011, Yale University.

——— (2014): "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," *Econometric Theory,* 30, 287–333.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper No. 1813, Yale University.

Andrews, D. W. K., and P. Guggenberger (2015): "Identification- and Singularity-Robust Inference for Moment Condition Models," Cowles Foundation Discussion Paper No. 1978, Yale University, revised 2017.

——— (2017): "Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models," *Econometric Theory,* 33, forthcoming. Earlier version available as Cowles Foundation Discussion Paper No. 1977, Yale University.

Andrews, I. (2017): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics,* forthcoming.

Andrews, I., and A. Mikusheva (2015): "Maximum Likelihood Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics,* 6, 123–152.

——— (2016a): "A Geometric Approach to Nonlinear Econometric Models," *Econometrica,* 84, 1249–1264.

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

Bera, A. K., and Y. Bilias (2001): "Rao's Score, Neyman's C(α) and Silvey's LM Tests: An Essay on Historical Developments and Some New Results," *Journal of Statistical Inference and Planning,* 97, 9–44.

Berger, R. L., and D. D. Boos (1994): "P Values Maximized over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association,* 89, 1012–1016.

Bugni, F. A., I. A. Canay, and X. Shi (2016): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," unpublished manuscript, Department of Economics, University of Wisconsin, Madison.

Campbell, J. Y., and M. Yogo (2006): "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics,* 81, 27–60.

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

Chaudhuri, S. (2016): "Improved Projection GMM-LM Tests for Linear Restrictions," unpublished manuscript, Department of Economics, University of Montreal.

Chaudhuri, S., T. Richardson, J. Robins, and E. Zivot (2010): "A New Projection-Type Split-Sample Score Test in Linear Instrumental Variables Regression," *Econometric Theory,* 26, 1820–1837.

Chaudhuri, S., and E. Zivot (2011): "A New Method of Projection-Based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics,* 164, 239–251.

Chen, X., T. M. Christensen, K. O'Hara, and E. Tamer (2016): "Monte Carlo Markov Chain Confidence Sets for Identified Sets," Cowles Foundation Discussion Paper No. 2017R, Yale University.

Cheng, X. (2015): "Robust Inference in Nonlinear Models with Mixed Identification Strength," *Journal of Econometrics,* 189, 207–228.

Cox, G. (2016): "Robust Inference in a Class of Generically Identified Models with an Application to Factor Models," unpublished manuscript, Department of Economics, Yale University.

Demmel, J. (1980): "Singular Value Decomposition," in Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst (eds.) *Templates for the Solution of Algebraic Eigenvalue Problems*: *A Practical Guide.* Philadelphia: SIAM.

Dufour, J.-M. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica,* 57, 335–355.

Dufour, J.-M., and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Structural Models with Generated Regressors," *International Economic Review,* 42, 815–843.

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

Guggenberger, P., F. Kleibergen, S. Mavroeidis, and L. Chen (2012): "On the Asymptotic Sizes of Subset Anderson-Rubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression," *Econometrica,* 80, 2649–2666.

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

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

Han, S., and A. McCloskey (2016): "Estimation and Inference with a (Nearly) Singular Jacobian," unpublished manuscript, Department of Economics, University of Texas, Austin.

Kaido, H., F. Molinari, and J. Stöye (2016): "Confidence Intervals for Projections of Partially Identified Parameters," unpublished manuscript, Department of Economics, Boston University.

Kleibergen, F. (2004): "Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model," *Review of Economics and Statistics,* 86, 418–423.

——— (2005): "Testing Parameters in GMM Without Assuming That They Are Identified," *Econometrica,* 73, 1103–1123.

——— (2015): "Efficient Size Correct Subset Inference in Linear Instrumental Variables Regression," unpublished manuscript, Amsterdam School of Economics, University of Amsterdam.

Lee, J. (2014): "Asymptotic Sizes of Subset Anderson-Rubin Tests with Weakly Identified Nuisance Parameters and General Covariance Structure," unpublished manuscript, Department of Economics, M.I.T.

Loh, W.-Y. (1985): "A New Method for Testing Separate Families of Hypotheses," *Journal of the American Statistical Association,* 80, 362–368.

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

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

Moran, P. A. P. (1979): "On Asymptotically Optimal Tests of Composite Hypotheses," *Biometrika,* 57, 47–55.

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

Neyman, J. (1959): "Optimal Asymptotic Test of Composite Statistical Hypothesis," in U. Grenander (ed.) *Probability and Statistics, The Harold Cramér Volume.* Uppsala: Almqvist and Wiksell, pp. 313–334.

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

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

Smith, R. J. (1987): "Alternative Asymptotically Optimal Tests and Their Application to Dynamic Specification," *Review of Economic Studies,* 54, 665–680.

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