Quantitative Economics 10 (2019), 1703–1746 1759-7331/20191703

Identification- and singularity-robust inference for moment condition models

DONALD W. K. ANDREWS
Cowles Foundation, Yale University

PATRIK GUGGENBERGER
Department of Economics, Pennsylvania State University

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

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

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

Donald W. K. Andrews: donald.andrews@yale.edu
Patrik Guggenberger: pxg27@psu.edu
Andrews gratefully acknowledges the research support of the National Science Foundation via grants SES-1058376, SES-1355504, and SES-1656313. Guggenberger gratefully acknowledges the research support of the National Science Foundation via grants SES-1021101 and SES-1326827. The authors thank Isaiah Andrews for providing the details of the simulation designs employed in I. Andrews (2016), which are used in some of the simulations given in this paper. For helpful comments, the authors thank four referees, Isaiah Andrews, Xu Cheng, Kirill Evdokimov, Koohyun Kwon, Anna Mikusheva, Zhongjun Qu, Jim Stock, and the participants of seminars at the following universities: Boston, Boston College, Brown, Chicago, Columbia, Freiburg, Hanover, Harvard/MIT, Hebrew Jerusalem, Konstanz, Maryland, Michigan, New York, Northwestern, Ohio State, Princeton, Queen's, Strasbourg, and Wisconsin.

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

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

JEL classification. C10, C12.

1. Introduction

Weak identification and weak instruments (IVs) can arise in a wide variety of empirical applications in economics. Examples include: in macroeconomics and finance, new Keynesian Phillips curve models, dynamic stochastic general equilibrium (DSGE) models, consumption capital asset pricing models (CCAPM), and interest rate dynamics models; in industrial organization, the Berry, Levinsohn, and Pakes (1995) (BLP) model of demand for differentiated products; and in labor economics, returns-to-schooling equations that use IVs, such as quarter of birth or Vietnam draft lottery status, to avoid ability bias. Other examples include nonlinear regression, autoregressive-moving average, GARCH, and smooth transition autoregressive (STAR) models; parametric selection models estimated by Heckman's two-step method or maximum likelihood; mixture models and regime switching models; and all models where hypothesis testing problems arise where a nuisance parameter appears under the alternative hypothesis, but not under the null.[1]

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

The asymptotic results allow the variance matrix of the moments to be near singular or singular. This is particularly important in models where weak identification (or lack of identification) is necessarily accompanied by near singularity (or exact singularity) of the variance matrix of the moments. This occurs in all maximum likelihood scenarios and many quasi-likelihood scenarios. Furthermore, in models of this type where robustness against lack of identification—not just against weak identification—is important, allowing for singularity of the variance matrix of the moments—not just near

[1] For references, see Section 12 in the Online Supplemental Material (Andrews and Guggenberger (2019)).

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

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

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

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

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

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

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

The SR-CQLR and SR-AR tests are for full vector inference. We develop subvector inference for scenarios in which the nuisance parameters under the null hypothesis are strongly identified. We show that the SR-CQLR subvector test is asymptotically efficient under strong and semi-strong identification. We compare the power of the subvector SR-CQLR and SR-AR tests with the power of the S test in Stock and Wright (2000) and the CLR test in I. Andrews and Mikusheva (2016), which we refer to as the AM test. The model considered is an endogenous probit model with a six- or eight-dimensional nuisance parameter and a scalar parameter of interest. The SR-CQLR and AM tests out perform the SR-AR and S tests in the scenarios considered. The SR-CQLR and AM tests have crisscrossing power functions, which makes a ranking difficult. It takes about 4 minutes to calculate 5000 CQLR tests using an Intel Core 3.4 GHz, 6 MB processor, which is about 59 times faster than for the AM test. The speed difference should be increasing rapidly in the dimension, p, of the parameter specified by the null hypothesis because the AM test requires an optimization over a p dimensional space for each simulation used to compute its conditional critical value, whereas the CQLR test has a closed-form expression. See Section 12 in the SM for references to other subvector inference methods in the literature.

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

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

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

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

The SM, that is, Andrews and Guggenberger (2019), contains the proofs. It also provides (i) time series results, (ii) finite-sample simulations of the null rejection probabilities of the SR-AR and SR-CQLR tests for cases where the variance matrix of the moment functions is singular and near-singular, (iii) analysis of the behavior of the SR-CQLR test and Kleibergen's (2005, 2007) CLR tests in the homoskedastic linear IV model with fixed IVs, (iv) the definition of a new SR-CQLR$_P$ test that reduces asymptotically to Moreira's (2003) CLR test for all $p \geq 1$, but only applies when the moment functions are of a product form, $u_i(\theta)Z_i$, where $u_i(\theta)$ is a scalar and Z_i is a k-vector of instrumental variables, and (v) the definition of a new SR-LM test.

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

2. Discussion of the related literature

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

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

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

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

Although Kleibergen's CLR tests with moment-variance-weighting have correct asymptotic size for $\mathcal{F}_0$, they have some drawbacks. First, the variance matrix of the moment functions must be nonsingular, which can be restrictive. Second, the parameter space $\mathcal{F}_0$ restricts the eigenvalues of the expected outer product of the vectorized orthogonalized sample Jacobian, which can be restrictive and can be difficult to verify in some models. Third, as shown in the SM, Kleibergen's CLR tests with moment-variance-weighting do not reduce to Moreira's CLR test in the homoskedastic normal linear IV model with fixed IVs when $p = 1$. Simulation results in Section 21 of the SM show that this leads to substantial power loss in some scenarios of this model, relative to the SR-CQLR tests considered here, Moreira's CLR test, and Kleibergen's CLR test with Jacobian-variance weighting. Fourth, the form of Kleibergen's CLR test statistic for $p \geq 2$ is based on the form of Moreira's test statistic when $p = 1$. In consequence, one needs to make a somewhat arbitrary choice of some rank statistic to reduce the $k \times p$ weighted orthogonalized sample Jacobian to a scalar random variable.

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

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

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

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

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

3. Moment condition model

3.1 *Moment functions*

The general moment condition model that we consider is

$$E_F g(W_i, \theta) = 0^k, \tag{3.1}$$

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

The Jacobian of the moment functions is

$$G(W_i, \theta) := \frac{\partial}{\partial \theta'} g(W_i, \theta) \in R^{k \times p}.^{2} \tag{3.2}$$

For notational simplicity, we let $g_i(\theta)$ and $G_i(\theta)$ abbreviate $g(W_i, \theta)$ and $G(W_i, \theta)$, respectively. We denote the jth column of $G_i(\theta)$ by $G_{ij}(\theta)$ and $G_{ij} = G_{ij}(\theta_0)$, where θ_0 is the (true) null value of θ, for $j = 1, \ldots, p$. Likewise, we often leave out the argument θ_0 for other functions as well. Thus, we write g_i and G_i, rather than $g_i(\theta_0)$ and $G_i(\theta_0)$. We let I_r denote the r dimensional identity matrix.

We are concerned with tests of the null hypothesis

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0. \tag{3.3}$$

3.2 *Parameter spaces of distributions F*

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

$$\begin{aligned} &\Omega_F(\theta) := E_F\big(g_i(\theta) - E_F g_i(\theta)\big)\big(g_i(\theta) - E_F g_i(\theta)\big)', \\ &r_F(\theta) := \text{rk}\big(\Omega_F(\theta)\big), \quad \text{and} \quad \Omega_F(\theta) := A_F^{\Omega}(\theta) \Pi_F(\theta) A_F^{\Omega}(\theta)', \end{aligned} \tag{3.4}$$

where $\text{rk}(\cdot)$ denotes the rank of a matrix, $\Pi_F(\theta)$ is the $k \times k$ diagonal matrix with the eigenvalues of $\Omega_F(\theta)$ on the diagonal in nonincreasing order, and $A_F^{\Omega}(\theta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\Pi_F(\theta)$. We allow

[2]The asymptotic size results given below do not actually require $G(W_i, \theta)$ to be the derivative matrix of $g(W_i, \theta)$. The matrix $G(W_i, \theta)$ can be any $k \times p$ matrix that satisfies the conditions in $\mathcal{F}^{\text{SR}}$, defined in (3.6) below. For example, $G(W_i, \theta)$ can be the derivative of $g(W_i, \theta)$ almost surely, rather than for all W_i, which allows $g(W_i, \theta)$ to have kinks. The function $G(W_i, \theta)$ also can be a numerical derivative, such as $\big(\big(g(W_i, \theta + \varepsilon e_1) - g(W_i, \theta)\big)/\varepsilon, \ldots, \big(g(W_i, \theta + \varepsilon e_p) - g(W_i, \theta)\big)/\varepsilon\big) \in R^{k \times p}$ for some $\varepsilon > 0$, where e_j is the jth unit vector, for example, $e_1 = (1, 0, \ldots, 0)' \in R^p$.

for the case where $\Omega_F(\theta)$ is singular. We partition $A_F^{\Omega}(\theta)$ according to whether the corresponding eigenvalues are positive or zero:

$$A_F^{\Omega}(\theta) = \left[A_F(\theta), A_F^{\perp}(\theta)\right], \quad \text{where } A_F(\theta) \in R^{k \times r_F(\theta)} \text{ and } A_F^{\perp}(\theta) \in R^{k \times (k - r_F(\theta))}. \tag{3.5}$$

The columns of $A_F(\theta)$ are eigenvectors of $\Omega_F(\theta)$ that correspond to positive eigenvalues of $\Omega_F(\theta)$. Let $\Pi_{1F}(\theta)$ denote the upper left $r_F(\theta) \times r_F(\theta)$ submatrix of $\Pi_F(\theta)$. The matrix $\Pi_{1F}(\theta)$ is diagonal with the positive eigenvalues of $\Omega_F(\theta)$ on its diagonal in nonincreasing order.

The r_F vector $\Pi_{1F}^{-1/2} A_F' g_i$ is a vector of nonredundant linear combinations of the moment functions evaluated at θ_0 rescaled to have variances equal to one: $\mathrm{Var}_F(\Pi_{1F}^{-1/2} A_F' g_i) = \Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2} = I_{r_F}$. The $r_F \times p$ matrix $\Pi_{1F}^{-1/2} A_F' G_i$ is the analogously transformed Jacobian matrix.

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

$$\begin{aligned} &\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}} := \left\{F : E_F g_i = 0^k \text{ and } E_F \left\|\Pi_{1F}^{-1/2} A_F' g_i\right\|^{2+\gamma} \le M\right\} \quad \text{and} \\ &\mathcal{F}^{\mathrm{SR}} := \left\{F \in \mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}} : E_F \left\|\mathrm{vec}\left(\Pi_{1F}^{-1/2} A_F' G_i\right)\right\|^{2+\gamma} \le M\right\}, \end{aligned} \tag{3.6}$$

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

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

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

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

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

In these examples and others like them, $E_F G_i$ is close to having less than full column rank and Ω_F is close to being singular when the null value θ_0 is close to a value which yields reduced column rank of $E_F G_i$ and singularity of Ω_F. Null hypotheses of this type are important for CSs because uniformity over null hypothesis values is necessary for CSs to have correct asymptotic size. Hence, it is important to have procedures available that place no restrictions on either $E_F G_i$ or Ω_F.

The parameter spaces for (F, θ) for the SR-AR and SR-CQLR CSs are

$$\begin{aligned} \mathcal{F}_{\Theta,\mathrm{AR}}^{\mathrm{SR}} &:= \{(F, \theta_0) : F \in \mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}(\theta_0), \theta_0 \in \Theta\} \quad \text{and} \\ \mathcal{F}_{\Theta}^{\mathrm{SR}} &:= \{(F, \theta_0) : F \in \mathcal{F}^{\mathrm{SR}}(\theta_0), \theta_0 \in \Theta\}, \end{aligned} \tag{3.7}$$

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

4. Singularity-robust nonlinear Anderson–Rubin test

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

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

The sample moments and an estimator of the variance matrix of the moments, $\Omega_F(\theta)$, are

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

The usual nonlinear AR statistic is

$$\mathrm{AR}_n(\theta) := n \widehat{g}_n(\theta)' \widehat{\Omega}_n^{-1}(\theta) \widehat{g}_n(\theta). \tag{4.2}$$

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

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

$$\widehat{r}_n(\theta) := \mathrm{rk}\big(\widehat{\Omega}_n(\theta)\big) \quad \text{and} \quad \widehat{\Omega}_n(\theta) := \widehat{A}_n^{\Omega}(\theta)\widehat{\Pi}_n(\theta)\widehat{A}_n^{\Omega}(\theta)', \tag{4.3}$$

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

$$\widehat{A}_n^{\Omega}(\theta) = \big[\widehat{A}_n(\theta), \widehat{A}_n^{\perp}(\theta)\big], \quad \text{where } \widehat{A}_n(\theta) \in R^{k\times\widehat{r}_n(\theta)} \text{ and } \widehat{A}_n^{\perp}(\theta) \in R^{k\times(k-\widehat{r}_n(\theta))}. \tag{4.4}$$

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

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

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

The SR-AR test statistic is defined by

$$\text{SR-AR}_n(\theta) := n\widehat{g}_{An}(\theta)'\widehat{\Omega}_{An}^{-1}(\theta)\widehat{g}_{An}(\theta). \tag{4.6}$$

The SR-AR test rejects the null hypothesis $H_0 : \theta = \theta_0$ if

$$\text{SR-AR}_n(\theta_0) > \chi^2_{\widehat{r}_n(\theta_0),1-\alpha} \quad \text{or} \quad \widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}, \tag{4.7}$$

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

The extra rejection condition in (4.7), $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}$, improves power, but we show it has no effect under H_0 with probability that goes to one (wp $\to$ 1); see Lemma 17.1 in the SM. It improves power because it fully exploits, rather than ignores, the nonstochastic part of the moment conditions associated with the singular part of the variance matrix. For example, if the moment conditions include some identities and the moment variance matrix excluding the identities is nonsingular, then $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0)$ consists of the identities and the SR-AR test rejects H_0 if the identities do not hold when evaluated at θ_0 or if the SR-AR statistic, which ignores the identities, is sufficiently large.

Two other simple examples where the extra rejection condition improves power are given in Section 15.2 in the SM.[3,4]

The SR-AR test statistic can be written equivalently as

$$\text{SR-AR}_n(\theta) = n\widehat{g}_n(\theta)'\widehat{\Omega}_n^+(\theta)\widehat{g}_n(\theta), \tag{4.8}$$

where $\widehat{\Omega}_n^+(\theta)$ is the Moore–Penrose generalized inverse of $\widehat{\Omega}_n(\theta)$; see (69) in the SM.

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

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

By definition, if $\widehat{r}_n(\theta_0) = k$, the condition $\widehat{A}_n^\perp(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}$ holds. When $\widehat{r}_n(\theta_0) = k$, $\text{SR-AR}_n(\theta_0) = \text{AR}_n(\theta_0)$ because $\widehat{A}_n(\theta_0)$ is invertible and $\widehat{\Omega}_{An}^{-1}(\theta_0) = \widehat{A}_n^{-1}(\theta_0)\widehat{\Omega}_n^{-1}(\theta_0)\widehat{A}_n^{-1}(\theta_0)'$.

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

Now we discuss the seven models listed in the Introduction. In each model, the sample moments are the likelihood score. In factor models, it is usually the case that the number of factors is uncertain. Hence, in a factor model with N_f factors, one is usually interested in the case where the actual number of factors is $J = 0, \dots, N_f$. However, when the factor loadings are such that only $J < N_f$ factors enter the model, the variances of the $N_f - J$ factors that do not enter the model are not identified. Hence, in order to carry out inference that is robust to different numbers of factors in the model, one requires robustness to weak and lack of identification and near and exact singularity of the variance matrix of the moments.

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

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

[4]When the sample variance matrix is singular, an alternative to using the $\text{SR-AR}_n(\theta_0)$ statistic is to arbitrarily delete some moment conditions. However, this typically leads to different test statistic values given the same data and can yield substantially different power properties of the test depending on which moment conditions are deleted, which is highly undesirable. See Section 15.2 in the SM for an example that illustrates this.

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

In a random coefficients model, it is usually of interest to consider the case where the coefficients are nonrandom. In this case, the parameter vector often is not identified and the variance matrix of the sample moments is singular. For example, consider a linear regression model $Y_i = \mu + X_i'\beta_i + u_i$, where $\beta_i := \beta + \xi_i \in R^2$, β is a constant vector, $\xi_i \sim N(0^2, V_\xi)$ independent of the error $u_i \sim N(0, \sigma_u^2)$, and V_ξ is a 2×2 variance matrix with variances $\sigma_{\xi 1}^2$ and $\sigma_{\xi 2}^2$ and correlation ρ_ξ. In the partially or wholly constant coefficient model, we have $\sigma_{\xi_1}^2 = 0$ and/or $\sigma_{\xi_2}^2 = 0$ and ρ_ξ is not identified. As another example, suppose $\beta_i := \beta + \xi_i$ is a scalar random coefficient in the linear regression model above, $(\xi_i, u_i) \sim N(0^2, V_{\xi u})$, $V_{\xi u}$ is a 2×2 variance matrix with variances σ_ξ^2 and σ_u^2 and correlation $\rho_{\xi u}$. In the constant coefficient submodel, we have $\sigma_\xi^2 = 0$, $\rho_{\xi u}$ is not identified, and the sample moments have a singular variance matrix.

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

The ARMA$(1,1)$ model is a workhorse model of time series analysis. It nests the important submodel with no serial correlation. This submodel arises when the AR and MA parameters are equal. The model is of the form: $Y_i = \rho Y_{i-1} + \varepsilon_i - \pi \varepsilon_{i-1}$, where $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma_\varepsilon^2$, and $\{\varepsilon_i : i \geq 1\}$ are serially uncorrelated. When $\rho = \pi$, the model reduces to $Y_i = \varepsilon_i$, the value of $\rho = \pi$ is not identified, and the sample moments have a singular variance matrix. Similar "common factor" identification and variance singularity issues also arise in higher-order ARMA(p, q) models.

5. SR-CQLR test

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

5.1 *CQLR test for nonsingular moments variance matrices*

The sample Jacobian is

$$\widehat{G}_n(\theta) := n^{-1}\sum_{i=1}^{n} G_i(\theta) = \big(\widehat{G}_{1n}(\theta), \dots, \widehat{G}_{pn}(\theta)\big) \in R^{k\times p}. \tag{5.1}$$

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

$$\begin{aligned}
&\widehat{D}_n(\theta) := \big(\widehat{D}_{1n}(\theta), \dots, \widehat{D}_{pn}(\theta)\big) \in R^{k\times p}, \quad \text{where} \\
&\widehat{D}_{jn}(\theta) := \widehat{G}_{jn}(\theta) - \widehat{\Gamma}_{jn}(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \in R^k \quad \text{for } j = 1, \dots, p, \text{ and} \\
&\widehat{\Gamma}_{jn}(\theta) := n^{-1}\sum_{i=1}^{n}\big(G_{ij}(\theta) - \widehat{G}_{jn}(\theta)\big)g_i(\theta)' \in R^{k\times k} \quad \text{for } j = 1, \dots, p.
\end{aligned} \tag{5.2}$$

We call $\widehat{D}_n(\theta)$ the *orthogonalized sample Jacobian* matrix. This statistic requires that $\widehat{\Omega}_n^{-1}(\theta)$ exists.

Next, we define

$$\begin{aligned}
&\widehat{R}_n(\theta) := \big(B(\theta)' \otimes I_k\big)\widehat{V}_n(\theta)\big(B(\theta) \otimes I_k\big) \in R^{(p+1)k\times(p+1)k}, \quad \text{where} \\
&\widehat{V}_n(\theta) := n^{-1}\sum_{i=1}^{n}\big(f_i(\theta) - \widehat{f}_n(\theta)\big)\big(f_i(\theta) - \widehat{f}_n(\theta)\big)' \in R^{(p+1)k\times(p+1)k}, \\
&f_i(\theta) := \begin{pmatrix} g_i(\theta) \\ \operatorname{vec}\big(G_i(\theta)\big)\end{pmatrix}, \qquad \widehat{f}_n(\theta) := \begin{pmatrix} \widehat{g}_n(\theta) \\ \operatorname{vec}\big(\widehat{G}_n(\theta)\big)\end{pmatrix}, \quad \text{and} \quad B(\theta) := \begin{pmatrix} 1 & 0_p' \\ -\theta & -I_p \end{pmatrix}.
\end{aligned} \tag{5.3}$$

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

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

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

over all symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$, where $\|\cdot\|$ denotes the Frobenius norm. This is a weighted minimization problem with the weights given by $I_{p+1} \otimes \widehat{\Omega}_n^{-1/2}(\theta)$. In the homoskedastic linear IV model, the population version of $\widehat{R}_n(\theta)$ has a Kronecker product form and, therefore, the Kronecker product approxmation in (5.4) leads to the asymptotic equivalence of the CQLR test and Moreira's (2003) CLR test in the homoskedastic linear IV model. We employ the weights above because they lead to a matrix $\widehat{\Sigma}_n(\theta)$ that is invariant to the multiplication of $g_i(\theta)$ and $G_i(\theta)$ by any nonsingular matrix $M \in R^{k\times k}$; see Lemma 5.1 below. Let $\widehat{\Sigma}_{j\ell n}(\theta)$ denote the (j,ℓ) element of $\widehat{\Sigma}_n(\theta)$ and $\widehat{R}_{j\ell n}(\theta)$ the (j,ℓ) $k\times k$ submatrix of dimension of $\widehat{R}_n(\theta)$.[5] By Theorems 3 and 10 of Van Loan and Pitsianis (1993), for $j,\ell = 1,\dots,p+1$, the solution to (5.4) is

$$\widehat{\Sigma}_{j\ell n}(\theta) = \operatorname{tr}\bigl(\widehat{R}_{j\ell n}(\theta)'\widehat{\Omega}_n^{-1}(\theta)\bigr)/k.^{6} \tag{5.5}$$

We use an eigenvalue-adjusted version of $\widehat{\Sigma}_n(\theta)$, denoted $\widehat{\Sigma}_n^{\varepsilon}(\theta)$, that improves the asymptotic and finite-sample size performance of the CQLR test in some scenarios by making it robust to singularities and near singularities of the matrix that $\widehat{\Sigma}_n(\theta)$ estimates. The adjustment affects the test statistic (that is, $\widehat{\Sigma}_n^{\varepsilon}(\theta) \neq \widehat{\Sigma}_n(\theta)$) only if the condition number of $\widehat{\Sigma}_n(\theta)$ (that is, $\lambda_{\max}(\widehat{\Sigma}_n(\theta))/\lambda_{\min}(\widehat{\Sigma}_n(\theta))$) exceeds $1/\varepsilon$. Hence, for a reasonable choice of ε, it often has no effect even in finite samples. Based on the finite-sample simulations, we recommend using $\varepsilon = 0.01$.

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

$$\begin{aligned} &H^{\varepsilon} := A_H\Lambda_H^{\varepsilon}A_H', \quad \text{where} \\ &\Lambda_H^{\varepsilon} := \mathrm{Diag}\bigl\{\max\bigl\{\lambda_{H1},\lambda_{\max}(H)\varepsilon\bigr\},\dots,\max\bigl\{\lambda_{Hd_H},\lambda_{\max}(H)\varepsilon\bigr\}\bigr\}, \end{aligned} \tag{5.6}$$

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

The QLR statistic is

$$\begin{aligned} &\mathrm{QLR}_n(\theta) := \mathrm{AR}_n(\theta) - \lambda_{\min}\bigl(n\widehat{Q}_n(\theta)\bigr), \quad \text{where} \\ &\widehat{Q}_n(\theta) := \bigl(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \widehat{D}_n^{*}(\theta)\bigr)'\bigl(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \widehat{D}_n^{*}(\theta)\bigr) \in R^{(p+1)\times(p+1)}, \\ &\widehat{D}_n^{*}(\theta) := \widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)\widehat{L}_n^{1/2}(\theta) \in R^{k\times p}, \qquad \widehat{L}_n(\theta) := (\theta, I_p)\bigl(\widehat{\Sigma}_n^{\varepsilon}(\theta)\bigr)^{-1}(\theta, I_p)' \in R^{p\times p}, \end{aligned} \tag{5.7}$$

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

[5] That is, $\widehat{R}_{j\ell n}(\theta)$ contains the elements of $\widehat{R}_n(\theta)$ indexed by rows $(j-1)k+1$ to jk and columns $(\ell-1)k$ to ℓk.

[6] Moreira and Moreira (forthcoming) utilized the best unweighted Kronecker-product approximation to a matrix, as developed in Van Loan and Pitsianis (1993), but with a different application and purpose than here.

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

$$\mathrm{CLR}_{k,p}(D) := Z'Z - \lambda_{\min}\big((Z,D)'(Z,D)\big), \quad \text{where } Z \sim N\big(0^k, I_k\big). \tag{5.8}$$

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

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

$$\mathrm{QLR}_n(\theta_0) > c_{k,p}\big(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha\big). \tag{5.9}$$

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

$$\mathrm{CS}_{\mathrm{CQLR},n} := \big\{\theta_0 \in \Theta : \mathrm{QLR}_n(\theta_0) \le c_{k,p}\big(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha\big)\big\}.$$

Next, we show that the CQLR test is invariant to nonsingular transformations of the moment functions/IVs. We suppress the dependence on θ of the statistics in the following lemma.

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

Comment. This lemma is used to obtain the correct asymptotic size of the CQLR test without assuming that $\lambda_{\min}(\Omega_F)$ is bounded away from zero. It suffices that Ω_F is nonsingular. In the proofs, we transform the moments by $g_i \rightsquigarrow M_F g_i$, where $M_F\Omega_F M_F' = I_k$, such that the transformed moments have a variance matrix whose eigenvalues are bounded away from zero for some $\delta > 0$ (since $\mathrm{Var}_F(M_F g_i) = I_k$) even if the original moments g_i do not.

5.2 *Singularity-robust CQLR test*

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

$$\begin{aligned}
&\widehat{D}_{An}(\theta) := \big(\widehat{D}_{A1n}(\theta), \dots, \widehat{D}_{Apn}(\theta)\big) \in R^{\widehat{r}_n(\theta) \times p}, \quad \text{where} \\
&\widehat{D}_{Ajn}(\theta) := \widehat{G}_{Ajn}(\theta) - \widehat{\Gamma}_{Ajn}(\theta)\widehat{\Omega}_{An}^{-1}(\theta)\widehat{g}_{An}(\theta) \in R^{\widehat{r}_n(\theta)} \quad \text{for } j = 1, \dots, p, \\
&\widehat{G}_{An}(\theta) := \widehat{A}_n(\theta)'\widehat{G}_n(\theta) = \big(\widehat{G}_{A1n}(\theta), \dots, \widehat{G}_{Apn}(\theta)\big) \in R^{\widehat{r}_n(\theta) \times p}, \quad \text{and} \\
&\widehat{\Gamma}_{Ajn}(\theta) := \widehat{A}_n(\theta)'\widehat{\Gamma}_{jn}(\theta)\widehat{A}_n(\theta) \quad \text{for } j = 1, \dots, p.
\end{aligned} \tag{5.10}$$

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

$$\begin{aligned}
\widehat{R}_{An}(\theta) &:= \big(B(\theta)' \otimes I_{\widehat{r}_n(\theta)}\big)\widehat{V}_{An}(\theta)\big(B(\theta) \otimes I_{\widehat{r}_n(\theta)}\big) \in R^{(p+1)\widehat{r}_n(\theta)\times(p+1)\widehat{r}_n(\theta)}, \quad \text{where} \\
\widehat{V}_{An}(\theta) &:= \big(I_{p+1} \otimes \widehat{A}_n(\theta)'\big)\widehat{V}_n(\theta)\big(I_{p+1} \otimes \widehat{A}_n(\theta)\big) \in R^{(p+1)\widehat{r}_n(\theta)\times(p+1)\widehat{r}_n(\theta)}, \\
\widehat{\Sigma}_{Aj\ell n}(\theta) &:= \operatorname{tr}\big(\widehat{R}_{Aj\ell n}(\theta)'\widehat{\Omega}^{-1}_{An}(\theta)\big)/\widehat{r}_n(\theta) \quad \text{for } j, \ell = 1, \dots, p+1, \\
\widehat{L}_{An}(\theta) &:= (\theta, I_p)\big(\widehat{\Sigma}^{\varepsilon}_{An}(\theta)\big)^{-1}(\theta, I_p)' \in R^{p\times p}, \\
\widehat{D}^*_{An}(\theta) &:= \widehat{\Omega}^{-1/2}_{An}(\theta)\widehat{D}_{An}(\theta)\widehat{L}^{1/2}_{An}(\theta) \in R^{\widehat{r}_n(\theta)\times p},
\end{aligned} \tag{5.11}$$

$\widehat{A}_n(\theta)$ is defined in (4.4), $\widehat{\Sigma}_{Aj\ell n}(\theta)$ denotes the (j,ℓ) element of $\widehat{\Sigma}_{An}(\theta)$, and $\widehat{R}_{Aj\ell n}(\theta)$ denotes the (j,ℓ) submatrix of dimension $\widehat{r}_n(\theta) \times \widehat{r}_n(\theta)$ of $\widehat{R}_{An}(\theta)$.

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

$$\begin{aligned}
\text{SR-QLR}_n(\theta) &:= \text{SR-AR}_n(\theta) - \lambda_{\min}\big(n\widehat{Q}_{An}(\theta)\big), \quad \text{where} \\
\widehat{Q}_{An}(\theta) &:= \big(\widehat{\Omega}^{-1/2}_{An}(\theta)\widehat{g}_{An}(\theta), \widehat{D}^*_{An}(\theta)\big)'\big(\widehat{\Omega}^{-1/2}_{An}(\theta)\widehat{g}_{An}(\theta), \widehat{D}^*_{An}(\theta)\big) \\
&\in R^{(p+1)\times(p+1)}.
\end{aligned} \tag{5.12}$$

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

$$\text{SR-QLR}_n(\theta_0) > c_{\widehat{r}_n(\theta_0),p}\big(n^{1/2}\widehat{D}^*_{An}(\theta_0), 1-\alpha\big) \quad \text{or} \quad \widehat{A}^{\perp}_n(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}.\text{[7]} \tag{5.13}$$

The nominal size $100(1-\alpha)\%$ SR-CQLR CS is $CS_{\text{SR-CQLR},n} := \{\theta_0 \in \Theta : \text{SR-QLR}_n(\theta_0) \le c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widehat{D}^*_{An}(\theta_0), 1-\alpha) \text{ and } \widehat{A}^{\perp}_n(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}\}$.[8]

When $\widehat{r}_n(\theta_0) = k$, $\widehat{A}_n(\theta_0)$ is a nonsingular $k \times k$ matrix. In consequence, by Lemma 5.1, $\text{SR-QLR}_n(\theta_0) = \text{QLR}_n(\theta_0)$ and

$$c_{\widehat{r}_n(\theta_0),p}\big(n^{1/2}\widehat{D}^*_{An}(\theta_0), 1-\alpha\big) = c_{k,p}\big(n^{1/2}\widehat{D}^*_n(\theta_0), 1-\alpha\big).$$

That is, the SR-CQLR test is the same as the CQLR test defined in Section 5.1. Of course, when $\widehat{r}_n(\theta) < k$, the CQLR test defined in Section 5.1 is not defined, whereas the SR-CQLR test is. Furthermore, if $\text{rk}(\Omega_{F_n}(\theta_0)) = k$ for all n large, then $\widehat{r}_n(\theta_0) = k$ and $\text{SR-QLR}_n(\theta_0) = \text{QLR}_n(\theta_0)$ wp $\to 1$ under $\{F_n \in \mathcal{F}^{\text{SR}} : n \ge 1\}$ (by Lemma 5.1 and Lemma 17.1 in the SM). Note that, if $\widehat{r}_n(\theta_0) \le p$, then the critical value for the SR-CQLR test is the $1-\alpha$ quantile of $\chi^2_{\widehat{r}_n(\theta_0)}$ (because $Z'Z - \lambda_{\min}((Z,D)'(Z,D)) = Z'Z \sim \chi^2_r$ in (5.8) when $r \le p$).

Section 20 in the SM provides finite-sample null rejection probabilities of the SR-CQLR test for singular and near singular variance matrices of the moment functions. The results show that singularity and near singularity of the variance matrix does not

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

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

lead to distorted null rejection probabilities. The method of robustifying the SR-CQLR test to allow for singular variance matrices, which is introduced above, works quite well in the model that is considered.

5.3 *Computation*

The SR-CQLR test is relatively fast to compute. It is found to be 115, 292, and 302 times faster to compute than the PI-CLC, MM1-SU, and MM2-SU tests, respectively, 1.2 times slower to compute than the JVW-CLR and MVW-CLR tests, and 372 and 495 times slower to compute than the LM and AR tests in the linear IV scenarios described in the Introduction. The SR-CQLR test is found to be noticeably easier to implement than the PI-CLC, MM1-SU, and MM2-SU tests and comparable to the JVW-CLR and MVW-CLR tests, in terms of the choice of implementation parameters (see Section 14.2 in the SM for details) and the robustness of the results to these choices.

The computation time of the SR-CQLR test increases relatively slowly with k and p. For example, the times (in minutes) to compute the SR-CQLR test 5000 times (using 5000 critical value repetitions) for $k = 8$ and $p = 1, 2, 4, 8$ are 0.26, 0.49, 1.02, 2.46. The times for $p = 1$ and $k = 1, 2, 4, 8, 16, 32, 64, 128$ are 0.14, 0.15, 0.18, 0.26, 0.44, 0.99, 2.22, 7.76. The times for $(k, p) = (64, 8)$ and $(128, 8)$ are 14.5 and 57.9. Hence, computing tests for large values of (k, p) is quite feasible. These times are for linear IV regression models, but they are the same for any model, linear or nonlinear, when one takes as given the sample moment vector and sample Jacobian matrix. Note that most of the computation time for the SR-CQLR test is due to the computation of its conditional critical values.

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

6. Asymptotic size

6.1 *Definitions of asymptotic size and similarity*

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

$$\mathrm{AsySz} := \limsup_{n\to\infty} \sup_{F\in\overline{\mathcal{F}}(\theta_0)} \mathrm{RP}_n(\theta_0, F, \alpha). \tag{6.1}$$

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

$$\liminf_{n\to\infty} \inf_{F\in\overline{\mathcal{F}}(\theta_0)} \mathrm{RP}_n(\theta_0, F, \alpha) = \limsup_{n\to\infty} \sup_{F\in\overline{\mathcal{F}}(\theta_0)} \mathrm{RP}_n(\theta_0, F, \alpha). \tag{6.2}$$

The *asymptotic size* of a CS obtained by inverting tests of $H_0 : \theta = \theta_0$ for the parameter space $\overline{\mathcal{F}}_\Theta := \{(F, \theta_0) : F \in \overline{\mathcal{F}}(\theta_0), \theta_0 \in \Theta\}$ is $\mathrm{AsySz} := \liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - \mathrm{RP}_n(\theta_0, F, \alpha))$. The CS is *asymptotically similar* (in a uniform sense) for $\overline{\mathcal{F}}_\Theta$ if $\liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - \mathrm{RP}_n(\theta_0, F, \alpha)) = \limsup_{n\to\infty} \sup_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - \mathrm{RP}_n(\theta_0, F, \alpha))$. Asymptotic size and similarity of a CS require uniformity over the null values $\theta_0 \in \Theta$, as well as uniformity over null distributions F for each null value θ_0. With the SR-AR and SR-CQLR CSs, this additional level of uniformity does not cause complications. The same proofs for tests deliver results for CSs with very minor adjustments.

6.2 *Identification categories*

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

Let $\{s_{jF} : j \le p\}$ denote the singular values of $\Omega_F^{-1/2}(\theta_0)E_F G_i(\theta_0)$ in nonincreasing order (when $\Omega_F(\theta_0)$ is nonsingular).[10] For a sequence of distributions $\{F_n : n \ge 1\}$, we say that the parameter θ_0 is: (i) weakly identified in the standard sense if $\lim n^{1/2} s_{1F_n} < \infty$, (ii) weakly identified in the nonstandard sense if $\lim n^{1/2} s_{pF_n} < \infty$ and $\lim n^{1/2} s_{1F_n} = \infty$, (iii) semi-strongly identified if $\lim n^{1/2} s_{pF_n} = \infty$ and $\lim s_{pF_n} = 0$, and (iv) strongly identified if $\lim s_{pF_n} > 0$. For sequences $\{F_n : n \ge 1\}$ for which the previous limits exist (and may equal ∞), these categories are mutually exclusive and exhaustive. We say that the parameter θ_0 is weakly identified if $\lim n^{1/2} s_{pF_n} < \infty$, which is the union of the standard and nonstandard weak identification categories. The asymptotics considered in Staiger and Stock (1997) are of the standard weak identification type. The nonstandard weak identification category can be divided into two subcategories: some weak/some strong identification and joint weak identification; see AG1 for details. The asymptotics considered in Stock and Wright (2000) are of the some weak/some strong identification type. For example, joint weak identification occurs in a linear IV model with $p > 1$ when the reduced-form coefficient matrix converges to a matrix of ones.

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

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

[10]The definitions of the identification categories when $\Omega_F(\theta_0)$ may be singular, as is allowed in this paper, is somewhat more complicated than the definitions given here.

weak identification sequences are much more difficult to analyze asymptotically because the p singular values have different orders of magnitude. This affects the asymptotic properties of both the test statistics and the conditioning statistics. Contiguous alternatives θ are at most $O(n^{-1/2})$ from θ_0 when θ_0 is strongly identified, but more distant when θ_0 is semi-strongly or weakly identified. Typically, the parameter θ is not consistently estimable when it is weakly identified.

6.3 *Asymptotic size results*

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

THEOREM 6.1. *The asymptotic sizes of the* SR-AR *and* SR-CQLR *tests defined in* (4.7) *and* (5.13), *respectively, equal their nominal size* $\alpha \in (0,1)$ *for the null parameter spaces* $\mathcal{F}_{\mathrm{AR}}^{\mathrm{SR}}$ *and* $\mathcal{F}^{\mathrm{SR}}$, *respectively. These tests are asymptotically similar* (*in a uniform sense*) *for the subsets of these parameter spaces that exclude distributions* F *under which* $g_i = 0^k$ *a.s. Analogous results hold for the corresponding* SR-AR *and* SR-CQLR CSs *for the parameter spaces* $\mathcal{F}_{\Theta,\mathrm{AR}}^{\mathrm{SR}}$ *and* $\mathcal{F}_{\Theta}^{\mathrm{SR}}$.

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

(ii) SR-LM versions of Kleibergen's LM test and CS are defined in Section 23 in the SM. However, as discussed there, these procedures are only partially singularity robust.

7. ASYMPTOTIC EFFICIENCY OF THE SR-CQLR TEST UNDER STRONG AND SEMI-STRONG IDENTIFICATION

Next, we show that the SR-CQLR test is asymptotically efficient in a GMM sense under strong and semi-strong identification (when the variance matrix of the moments is nonsingular and the null parameter value is not on the boundary of the parameter space). By this, we mean that it is asymptotically equivalent (under the null and contiguous alternatives) to a Wald test constructed using an asymptotically efficient GMM estimator and to the standard GMM LM test; see Newey and West (1987). More specifically, we consider drifting sequences $\{\lambda_{n,h}^* : n \geq 1\}$ of data-generating processes taken from $\mathcal{F}^{\mathrm{SR}}$ in (3.6) that correspond to strong or semi-strong identification and establish that the SR-CQLR test statistic equals the standard GMM LM test statistic up to a $o_p(1)$ term and that the conditional critical value of the SR-CQLR test converges in probability to $\chi^2_{p,1-\alpha}$.

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

$$\mathrm{LM}_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \quad \text{and} \quad \mathrm{LM}_n^{\mathrm{GMM}} := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{G}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \quad (7.1)$$

respectively, where $\widehat{G}_n$ is the sample Jacobian defined in (4.1) with $\theta = \theta_0$ and P_A denotes the projection matrix onto the column space of the matrix A (that is, $P_A =$

$A(A'A)^{-1}A'$ when A is full column rank). The critical value for the LM_n and LM_n^{GMM} tests is $\chi^2_{p,1-\alpha}$, the $1-\alpha$ quantile of the χ^2_p distribution. The test based on LM_n^{GMM} is asymptotically equivalent to the Wald test based on an asymptotically efficient GMM estimator under (i) strong identification (which requires $k \geq p$), (ii) nonsingular moments-variance matrices (that is, $\lambda_{\min}(\Omega_{F_n}) \geq \delta > 0$ for all $n \geq 1$), and (iii) a null parameter value that is not on the boundary of the parameter space; see Newey and West (1987). This also holds true under semi-strong identification (which also requires $k \geq p$). For example, Theorem 5.1 of Andrews and Cheng (2013) shows that the Wald statistic for testing $H_0 : \theta = \theta_0$ based on a GMM estimator with asymptotically efficient weight matrix has a χ^2_p distribution under semi-strong identification. This Wald statistic can be shown to be asymptotically equivalent to the LM_n^{GMM} statistic under semi-strong identification. (For brevity, we do not do so here.)

Suppose $k \geq p$. The drifting sequences $\{\lambda^*_{n,h} : n \geq 1\}$ referred to above are rather complicated and so, for brevity, we define them at the beginning of Section 28 in the SM. They are defined so that various population quantities that affect the asymptotic distributions of the SR-CQLR test statistic and critical value converge as $n \to \infty$. We restrict $\{\lambda^*_{n,h} : n \geq 1\}$ to be a sequence for which $\lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$. Most importantly, we have that, along $\{\lambda^*_{n,h} : n \geq 1\}$, $n^{1/2}(s_{1F_n}, \ldots, s_{pF_n})$ converges to some vector $(h^*_{1,1}, \ldots, h^*_{1,p})$ whose elements may be finite or infinite, where $(s_{1F_n}, \ldots, s_{pF_n})$ denote the singular values of the population Jacobian $E_F G_i \in R^{k\times p}$. Strong or semi-strong identification occurs if the smallest singular value of $E_F G_i$ diverges to infinity after renormalization by $n^{1/2}$, that is, if $h^*_{1,p} = \infty$.

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

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

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

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

8. Empirical application

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

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

where τ is a constant, ψ denotes EIS, Δc_{i+1} is consumption growth at time $i+1$, r_{i+1} is the real return on an asset at time $i+1$, and ξ_{i+1} is the error term that is correlated with the regressor. (Note that Yogo (2004) uses a subscript t rather than i.) To identify EIS, we use a vector $Z_i \in R^4$ of IVs consisting of the nominal interest rate, inflation, consumption growth, and log dividend-price ratio, all of which are lagged twice and then satisfy $E(Z_i \xi_{i+1}) = 0^4$. We also consider the reversed form of (8.1):

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

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

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

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

We calculate 95% confidence intervals for ψ and $1/\psi$ (that is, $\alpha = 0.05$) by collecting the values of $\theta_0 = \psi$ for which the null hypothesis in (3.3) is not rejected at the 5% nominal size. To do so, we use a grid of null values with stepsize 0.001 in $[-200, 200]$ and also consider the additional null values ± 500 and ± 1000 (and in some cases larger values).

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

Then define

$$g_i(\theta) = \big(M_{1^n}(\Delta c - \psi r)\big)_i (M_{1^n} Z)_i' \in R^4 \tag{8.3}$$

in the case of regression (8.1) (and analogously $g_i(\theta) = (M_{1^n}(r - (1/\psi)\Delta c))_i (M_{1^n} Z)_i' \in R^4$ in the case of regression (8.2), where $\theta = \psi$ (or $\theta = 1/\psi$). We then obtain

$$G_i(\theta) = G_i = -(M_{1^n} r)_i (M_{1^n} Z)_i' \in R^4 \tag{8.4}$$

for the Jacobian defined in (3.2) (and analogously $G_i = -(M_{1^n}\Delta c)_i (M_{1^n} Z)_i'$ in the case of regression (8.2)).

Note that in the regression models considered here the dimension of the parameter of interest equals one, that is, $p = 1$. Next, calculate the quantities $\widehat{g}_n(\theta)$ and $\widehat{\Omega}_n^{-1}(\theta)$ in (4.1), $\widehat{G}_{1n}$, $\widehat{\Gamma}_{1n}(\theta)$, and $\widehat{D}_{1n}(\theta)$ defined in (5.2), and $f_i(\theta)$, $\widehat{f}_n(\theta)$, $\widehat{V}_n(\theta)$, and $\widehat{R}_n(\theta)$ defined in (5.3), with all of these quantities evaluated with θ equal to $\theta_0 = \psi$. Then calculate $\widehat{\Sigma}_n(\theta)$ and its eigenvalue adjusted version $\widehat{\Sigma}_n^{\varepsilon}(\theta)$; see (5.5) and (5.6). For the output below, we use $\varepsilon = 0.01$. (We also calculated CIs for $\varepsilon = 0.05$ and 0.001, which led to identical results for the case of the real asset being $r_i = r_{f,i}$, defined below, and comparable results for the case of $r_i = r_{e,i}$, defined below.) Finally, calculate the quantities $\widehat{L}_n(\theta)$, $\widehat{D}_n^*(\theta)$, and $\widehat{Q}_n(\theta)$, the test statistic $\mathrm{QLR}_n(\theta)$ defined in (5.7), and the test statistic $\mathrm{AR}_n(\theta)$ in (4.2).

The critical value for the AR test is the $\chi^2_{4,1-\alpha}$ quantile given that there are four instruments. The critical value for the CQLR test is obtained by simulation. Specifically, we generate 10,000 draws from a $N(0^4, I_4)$ distribution and for each draw we calculate $\mathrm{CLR}_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0))$ defined in (5.8). The critical value of the CQLR test is then defined as the $1-\alpha$ sample quantile of these observations, which is denoted by $c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha)$.

The data set from Campbell (2003) employed here consists of quarterly data for the following eleven developed countries: Australia, Canada, France, Germany, Italy, Japan, Netherlands, Sweden, Switzerland, United Kingdom (U.K.), and the United States (U.S.). The sample period varies across different countries with sample sizes equal to 114, 115, 113, 79, 106, 114, 86, 116, 91, 115, and 114, respectively. For r_i two candidates for asset returns are used, namely, the real interest rate and the real aggregate stock return, denoted by $r_{f,i}$ and $r_{e,i}$, respectively. See Yogo (2004, Section IV. A., p. 803) for details on the data and the precise definition of the variables.

Table 1 reports the results based on the real interest rate $r_{f,i}$, whereas Table 2 reports the results based on the real aggregate stock return $r_{e,i}$. In each case, we report the CQLR and AR CI's for ψ and $1/\psi$ based on the regressions (8.1) and (8.2). If a CI contains the right endpoint of the search interval, namely 200, and the additional points of the search 500 and 1000, we report the right endpoint of the CI as ∞, and analogously for the left endpoint.

We now discuss the findings for ψ and the implications for the equity premium puzzle obtained from the new inference procedures.[11] We start with the results based on $r_{f,i}$. Yogo (2004, p. 806) concludes from the CIs based on the homogenous CLR test that

[11]Our analysis reveals certain discrepancies with the results reported in Yogo (2004). Namely, the CIs for ψ using $r_{f,i}$ based on the LM test (see (26) in Yogo (2004)) are as follows: Australia $[-0.22, 0.27] \cup [5.13, 13.74]$ by our calculations versus (vs.) $[-0.22, 13.74]$ in Table 3 in Yogo (2004), Canada $[-0.73, 0.02] \cup [3.9, 14.16]$ vs. $[-0.73, 14.15]$, France $[-50.06, -36.28] \cup [-0.47, 0.31]$ vs. $[-0.47, 0.31]$, Germany $[-1.21, 0.26] \cup [11.3, 16.02]$ vs. $[-1.21, 0.26]$, Italy $[-6.51, -3.83] \cup [-0.24, 0.11]$ vs. $[-0.24, 0.11]$, Japan $(-\infty, -11.29] \cup [-0.58, 0.47] \cup [6.15, \infty)$ vs. $[-0.24, 0.11]$, Netherlands $(-\infty, -17.21] \cup [-0.76, 0.48] \cup [35.63, \infty)$ vs. $(-\infty, \infty)$, Sweden

Table 1. CQLR and AR CIs for EIS, ψ, and its inverse, $1/\psi$, using $r_{f,i}$.

Country	ψ		$1/\psi$	
	CQLR	AR	CQLR	AR
Australia	$[-0.24, 0.34]$	$[-0.12, 0.27]$	$(-\infty, -4.2] \cup [2.9, \infty)$	$(-\infty, -8.3] \cup [3.8, \infty)$
Canada	$[-0.88, 0.21]$	$[-0.71, 0.05]$	$(-\infty, -1.1] \cup [4.8, \infty)$	$(-\infty, -1.4] \cup [21.8, \infty)$
France	$[-0.39, 0.16]$	$[-0.55, 0.33]$	$(-\infty, -2.6] \cup [6.1, \infty)$	$(-\infty, -1.8] \cup [3.0, \infty)$
Germany	$[-1.5, 0.90]$	$[-1.8, 1.28]$	$(-\infty, -0.66] \cup [1.1, \infty)$	$(-\infty, -0.56] \cup [0.78, \infty)$
Italy	$[-0.25, 0.10]$	$[-0.32, 0.18]$	$(-\infty, -4.0] \cup [9.6, \infty)$	$(-\infty, -3.1] \cup [5.6, \infty)$
Japan	$[-0.78, 0.29]$	$[-0.86, 0.34]$	$(-\infty, -1.3] \cup [3.5, \infty)$	$(-\infty, -1.2] \cup [2.9, \infty)$
Netherlands	$[-0.72, 1.79]$	$[-0.44, -0.11]$	$(-\infty, -1.4] \cup [0.56, \infty)$	$[-9.2, -2.3]$
Sweden	$[-0.20, 0.20]$	$[-0.27, 0.26]$	$(-\infty, -5.1] \cup [5.0, \infty)$	$(-\infty, -3.8] \cup [3.8, \infty)$
Switzerland	$[-1.04, 0.18]$	$[-1.32, 0.41]$	$(-\infty, -0.96] \cup [5.5, \infty)$	$(-\infty, -0.76] \cup [2.4, \infty)$
U.K.	$[-0.97, 0.54]$	$[-0.01, 0.47]$	$(-\infty, -1.0] \cup [1.9, \infty)$	$(-\infty, -68.9] \cup [2.1, \infty)$
U.S.	$[-0.30, 0.49]$	$\emptyset$	$(-\infty, -3.3] \cup [2.0, \infty)$	$\emptyset$

Table 2. CQLR and AR CIs for EIS, ψ, and its inverse, $1/\psi$, using $r_{e,i}$ with $\varepsilon = 0.01$.

Country	ψ	
	CQLR	AR
Australia	$(-\infty, \infty)$	$(-\infty, \infty)$
Canada	$(-\infty, -1.33] \cup [0.017, \infty)$	$(-\infty, -0.35] \cup [-0.01, \infty)$
France	$(-\infty, 0.04] \cup [0.63, \infty)$	$(-\infty, 0.07] \cup [0.46, \infty)$
Germany	$(-\infty, \infty)$	$(-\infty, \infty)$
Italy	$(-\infty, \infty)$	$(-\infty, \infty)$
Japan	$(-\infty, -0.336] \cup [-0.334, -0.333] \cup [-0.06, \infty)$	$(-\infty, -0.66] \cup [-0.06, \infty)$
Netherlands	$(-\infty, -0.002] \cup [0.05, \infty)$	$(-\infty, -0.01] \cup [0.02, \infty)$
Sweden	$(-\infty, \infty)$	$(-\infty, \infty)$
Switzerland	$(-\infty, \infty)$	$(-\infty, \infty)$
U.K.	$(-\infty, \infty)$	$(-\infty, 0.002] \cup [0.04, \infty)$
U.S.	$(-\infty, -0.01] \cup [0.048, \infty)$	$(-\infty, -0.01] \cup [0.07, \infty)$

Country	$1/\psi$	
	CQLR	AR
Australia	$(-\infty, \infty)$	$(-\infty, \infty)$
Canada	$[-0.75, 60.6]$	$(-\infty, -182.1] \cup [-2.9, \infty)$
France	$(-\infty, 1.58] \cup [24.75, \infty)$	$(-\infty, 2.16] \cup [14.97, \infty)$
Germany	$(-\infty, \infty)$	$(-\infty, \infty)$
Italy	$(-\infty, \infty)$	$(-\infty, \infty)$
Japan	$(-\infty, -15.8] \cup [-2.994, -2.99] \cup [-2.97, \infty)$	$(-\infty, -15.7] \cup [-1.5, \infty)$
Netherlands	$[-656.97, -609.34] \cup [-484.1, 20.9]$	$[-67.27, 51.98]$
Sweden	$(-\infty, \infty)$	$(-\infty, \infty)$
Switzerland	$(-\infty, \infty)$	$(-\infty, \infty)$
U.K.	$(-\infty, \infty)$	$(-\infty, 24.4] \cup [509.1, \infty)$
U.S.	$[-135.01, 21.03]$	$[-159.57, 13.93]$

the "EIS is small and not significantly different from 0 for the eleven developed countries." This finding is supported also by the new results based on the CQLR test except for the Netherlands (where the CI equals $[-0.72, 1.79]$). All eleven CIs contain zero, and nine CIs are bounded from above by 0.54 (with the exceptions of Germany, where the right endpoint of the CI equals 0.90, and the Netherlands). The finding is also supported by the CIs from the new AR test for almost all countries, with the exception of Germany (where the right endpoint of the CI is 1.28), the Netherlands (where 0 is not included in the CI), and the U.S. (where the CI is empty). The results based on the CQLR (and also the new AR) CIs are consistent across the regressions (8.1) and (8.2) and the empirical puzzle based on classical (identification nonrobust) inference procedures does not occur here. In particular, the left endpoints of the positive portions of the CIs for $1/\psi$ based on the CQLR test equal 2.9, 4.8, 6.1, 1.1, 9.6, 3.5, 0.56, 5.0, 5.5, 1.9, and 2.0 for the eleven countries, respectively, which translate into right endpoints of the positive portion of CIs for ψ of 0.34, 0.21, 0.16, 0.91, 0.10, 0.29, 1.79, 0.20, 0.18, 0.53, and 0.50, respectively. The actual right endpoints of the positive portion of the CIs for ψ based on the CQLR test equal 0.34, 0.21, 0.16, 0.90, 0.10, 0.29, 1.79, 0.20, 0.18, 0.54, and 0.49, respectively.

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

Next, we discuss the findings when $r_{e,i}$ is used. Inference on ψ and $1/\psi$ based on $r_{e,i}$ is completely uninformative for both the CQLR and AR CIs for Australia, Germany, Italy, Sweden, and Switzerland with CIs all equal to $(-\infty, \infty)$; see Table 2. Inference is also relatively uninformative for all of the other countries, with unbounded CIs for all countries except for Canada, the Netherlands, and the US. And even in the latter three cases, the CIs are too wide to provide information of economic interest. These results are mostly consistent with the findings based on the homoskedastic versions of the AR, LM, and CLR tests that also produce unbounded CIs in almost all cases; see Yogo (2004, Table 5). However, Canada, France, and Japan are three exceptions for which, based on these homoskedastic tests, informative CIs are obtained that imply a small value of ψ. It may be the case that the discrepancies between the results based on the new CIs and those based on the homoskedastic AR, LM, and CLR CIs for these countries are a consequence of undercoverage of the latter CIs because the actual DGP may not satisfy the assumptions necessary for validity of these CIs, such as homoskedasticity.

$(-\infty, -59.26] \cup [-0.21, 0.2] \cup [11.62, \infty]$ vs. $(-\infty, \infty)$, Switzerland $[-1.19, 0.07] \cup [4.9, 7.5]$ vs. $[-1.19, 0.07]$, U.K. $(-\infty, -17.23] \cup [-0.13, 0.45] \cup [7.22, \infty)$ vs. $(-\infty, \infty)$, and U.S. $(-\infty, -27.86] \cup [-0.28, 0.27] \cup [1.41, \infty)$ vs. $(-\infty, \infty)$.

Furthermore, the CIs for ψ using $r_{e,i}$ based on the LM test are for Canada $[-0.11, -0.09] \cup [0.05, 0.35]$ by our calculations vs. $[0.05, 0.35]$ as reported in Table 5 in Yogo (2004), for France $(-\infty, -1.56] \cup [-0.12, 0.07] \cup [0.74, \infty)$ vs. $(-\infty, \infty)$, and for Japan $[-1.01, -0.16] \cup [-0.02, 0.2]$ vs. $[-1.01, 0.20]$.

The CIs for ψ using $r_{e,i}$ based on the AR test (see (25) in Yogo (2004)) are for Australia $(-\infty, -0.21] \cup [-0.04, \infty)$ by our calculations vs. $(-\infty, \infty)$ as reported in Table 5 in Yogo (2004) and for the U.S. $(-\infty, -0.331] \cup [0.048, \infty)$ vs. $(-\infty, \infty)$. Finally, the CI for ψ using $r_{e,i}$ based on the CLR test (see (27) in Yogo (2004)) for the U.S. is $(-\infty, -0.01] \cup [0.048, \infty)$ by our calculations vs. $(-\infty, \infty)$ in Yogo (2004).

Note that for the countries where the new CIs are not equal to $(-\infty, \infty)$, there is complete consistency between the CIs for ψ and $1/\psi$, analogous to the findings reported in Table 1. For example, the CQLR CI for $1/\psi$ has right endpoint equal to 60.6 implying that any positive value of ψ should be contained in $[0.0165, \infty)$. And indeed, the positive portion of the CQLR type CI for ψ is reported as $[0.017, \infty]$. Analogous statements are obtained for the CQLR CI for $1/\psi$, for example, see the results for France, the Netherlands, and the U.S. In summary, the CIs based on the new tests reveal that ψ is very weakly identified (or perhaps unidentified) when one uses $r_{e,i}$. Unlike CIs based on a classical inference procedure, such as a t-test based CI, the identification-robust results based on the regressions (8.1) and (8.2) are internally consistent.

9. Subvector inference

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

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

9.1 *Model and hypotheses*

The model is

$$E_F g(W_i, \theta, \beta) = 0^k, \tag{9.1}$$

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

We are concerned with tests of the null hypothesis

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0 \tag{9.2}$$

in the presence of the nuisance parameter β and with confidence sets for θ obtained by inverting the tests.

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

$$\begin{aligned} &G(W_i, \eta) := \frac{\partial}{\partial \theta'} g(W_i, \eta) \in R^{k \times p}, \qquad G_\beta(W_i, \eta) := \frac{\partial}{\partial \beta'} g(W_i, \eta) \in R^{k \times b}, \\ &G_{\theta_j \beta}(W_i, \eta) := \frac{\partial^2}{\partial \theta_j \partial \beta'} g(W_i, \eta) \in R^{k \times b} \quad \text{for } j = 1, \ldots, p, \end{aligned} \tag{9.3}$$

and likewise for other expressions, such as $G_{\beta_j\beta}(W_i,\eta)$. Let $g_i(\eta) := g(W_i,\eta)$ and $\widehat{g}_n(\eta) := n^{-1}\sum_{i=1}^n g_i(\eta)$ and likewise for other quantities, for example,

$$\widehat{G}_{\beta n}(\eta) := n^{-1}\sum_{i=1}^n G_{i\beta}(\eta). \tag{9.4}$$

We use the notation $G_{ij} := \frac{\partial}{\partial\theta_j} g(W_i,\eta_0) \in R^k$ for $j = 1,\dots,p$, where $\eta_0 := (\theta_0', \beta^{*\prime})'$ and β^* denotes the true value of β, and likewise for other quantities. For example, $G_{i\theta_j\beta} := \frac{\partial^2}{\partial\theta_j\partial\beta'} g(W_i,\eta_0)$, $G_{i\beta_j\beta} := \frac{\partial^2}{\partial\beta_j\partial\beta'} g(W_i,\eta_0)$, and $g_i := g_i(\eta_0)$.

9.2 *Subvector tests for nonsingular moment variance matrices*

9.2.1 *Definitions of the subvector tests*

Define

$$\widehat{\Omega}_n(\eta) := n^{-1}\sum_{i=1}^n g_i(\eta)g_i(\eta)' - \widehat{g}_n(\eta)\widehat{g}_n(\eta)'. \tag{9.5}$$

Let $\widehat{\beta}_n = \widehat{\beta}_n(\theta_0)$ denote the null-restricted two-step GMM estimator of β. That is,

$$\widehat{\beta}_n := \underset{\beta\in B}{\arg\min}\left\|\widehat{\varphi}_n\widehat{g}_n(\theta_0,\beta)\right\|^2, \quad \text{where } \widehat{\varphi}_n \in R^{k\times k}, \widehat{\varphi}_n'\widehat{\varphi}_n = \widehat{\Omega}_n^{-1}(\theta_0,\widetilde{\beta}_n), \tag{9.6}$$

$\widetilde{\beta}_n$ is a solution to (9.6) with $\widehat{\varphi}_n$ replaced by I_k. Rather than using the null-restricted two-step GMM estimator $\widehat{\beta}_n$, one could employ the null-restricted continuous-updating estimator of β (e.g., as suggested in Kleibergen (2005)). The same asymptotic results as below would be obtained.

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

$$\begin{aligned}
&\widetilde{\Omega}_n(\eta) := \big(\widetilde{\Omega}_{1n}(\eta),\dots,\widetilde{\Omega}_{kn}(\eta)\big) \in R^{k\times k},\\
&\widetilde{\Omega}_{jn}(\eta) := n^{-1}\sum_{i=1}^n g_i(\eta)g_{ij}(\eta) - \widehat{\Phi}_{jn}(\eta)\widehat{\Omega}_n^{-1}(\eta)\widehat{g}_n(\eta) - \widehat{g}_n(\eta)\widehat{g}_{jn}(\eta) \in R^k, \quad \text{and}\\
&\widehat{\Phi}_{jn}(\eta) := n^{-1}\sum_{i=1}^n\left(g_i(\eta)g_{ij}(\eta) - n^{-1}\sum_{s=1}^n\big(g_s(\eta)g_{sj}(\eta)\big)\right)g_i(\eta)'\\
&\qquad \in R^{k\times k} \quad \text{for } j = 1,\dots,k,
\end{aligned} \tag{9.7}$$

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

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

$$
\begin{aligned}
&\widetilde{G}_{\beta n}(\eta) := \big(\widetilde{G}_{\beta_1 n}(\eta), \dots, \widetilde{G}_{\beta_b n}(\eta)\big) \in R^{k\times b}, \\
&\widetilde{G}_{\beta_j n}(\eta) := n^{-1}\sum_{i=1}^n G_{i\beta_j}(\eta) - \widehat{\Gamma}_{jn}(\eta)\widehat{\Omega}_n^{-1}(\eta)\widehat{g}_n(\eta) \in R^k, \quad \text{and} \\
&\widehat{\Gamma}_{jn}(\eta) := n^{-1}\sum_{i=1}^n \big(G_{i\beta_j}(\eta) - \widehat{G}_{\beta_j n}(\eta)\big) g_i(\eta)' \in R^{k\times k}, \quad \text{where} \\
&\widehat{G}_{\beta_j n}(\eta) := n^{-1}\sum_{i=1}^n G_{i\beta_j}(\eta),
\end{aligned}
\tag{9.8}
$$

for $j = 1, \dots, b$.

We define the following estimator $\widetilde{J}_n(\theta_0, \widehat{\beta}_n)$ of $(E_F g_i g_i')^{-1/2} E_F G_{i\beta}$, which is designed to be asymptotically independent of $\overline{g}_h^S$. Let

$$
\widetilde{J}_n(\eta) := \widetilde{\Omega}_n^{-1/2}(\eta)\widetilde{G}_{\beta n}(\eta) \in R^{k\times b}. \tag{9.9}
$$

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

The subvector AR test statistic is

$$
\mathrm{AR}_n^S(\eta) := n\widehat{g}_n(\eta)'\widetilde{\Omega}_n^{-1/2}(\eta) M_{\widetilde{J}_n(\eta)} \widetilde{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta). \tag{9.10}
$$

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

The subvector QLR test statistic $\mathrm{QLR}_n^S(\theta_0, \widehat{\beta}_n)$ is defined as the full vector statistic is defined in (5.7), but with $(\theta, \widehat{\beta}_n)$ in place of θ, $\widetilde{\Omega}_n^{-1/2}$ in place of $\widehat{\Omega}_n^{-1/2}$, and the projection matrix $M_{\widetilde{J}_n(\theta,\widehat{\beta}_n)}$ inserted as a weight matrix. In particular, let $\widehat{D}_n(\theta, \beta) \in R^{k\times p}$ be defined as $\widehat{D}_n(\theta)$ is defined in (5.2), but with (θ, β) in place of θ. Then define

$$
\begin{aligned}
&\mathrm{QLR}_n^S(\theta, \widehat{\beta}_n) := \mathrm{AR}_n^S(\theta, \widehat{\beta}_n) - \lambda_{\min}\big(n\widehat{Q}_n^S(\theta, \widehat{\beta}_n)\big), \quad \text{where} \\
&\quad \widehat{Q}_n^S(\eta) := \big(\widetilde{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta), \widehat{D}_n^*(\eta)\big)' M_{\widetilde{J}_n(\eta)} \big(\widetilde{\Omega}_n^{-1/2}(\eta)\widehat{g}_n(\eta), \widehat{D}_n^*(\eta)\big) \\
&\qquad \in R^{(p+1)\times(p+1)}, \\
&\quad \widehat{D}_n^*(\eta) := \widetilde{\Omega}_n^{-1/2}(\eta)\widehat{D}_n(\eta)\widehat{L}_n^{1/2}(\eta) \in R^{k\times p}, \\
&\quad \widehat{L}_n(\eta) := (\theta, I_p)\big(\widehat{\Sigma}_n^{\varepsilon}(\eta)\big)^{-1}(\theta, I_p)' \in R^{p\times p},
\end{aligned}
\tag{9.11}
$$

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

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

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

$$\mathrm{CLR}_{k,p}(D, J) := Z' M_J Z - \lambda_{\min}\big((Z, D)' M_J (Z, D)\big) \quad \text{and} \quad Z \sim N\big(0^k, I_k\big). \tag{9.12}$$

The conditional critical value of the nominal size α CQLR test is

$$c_{k,p}\big(n^{1/2}\widehat{D}_n^*(\theta_0, \widehat{\beta}_n), \widetilde{J}_n(\theta_0, \widehat{\beta}_n), 1-\alpha\big).$$

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

$$\mathrm{QLR}_n^S(\theta_0, \widehat{\beta}_n) > c_{k,p}\big(n^{1/2}\widehat{D}_n^*(\theta_0, \widehat{\beta}_n), \widetilde{J}_n(\theta_0, \widehat{\beta}_n), 1-\alpha\big). \tag{9.13}$$

9.2.2 *Asymptotic size of the subvector tests* We make the following assumptions about the function g and the parameter space B of β. We denote by $C^j(S)$ the set of j-times continuously differentiable functions from a set S into R^k.

Assumption gB. (a) *For given θ_0 the domain of g is $W \times \{\theta_0\} \times B$, where B is compact.*

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

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

The parameter space $\mathcal{F}$ in (79) needs to be altered from the case of a full vector hypothesis test to the subvector case. Let μ denote a probability measure on R^m for which $E_\mu \sup_{\beta \in B} \|g_i(\theta_0, \beta)\| < \infty$, where E_μ denotes expectation when W_i is distributed according to the measure μ. For $\vartheta > 0$ and $\beta^+ \in R^b$, let $B(\beta^+, \vartheta) = \{\beta \in R^b : \|\beta^+ - \beta\| < \vartheta\}$. We abbreviate "absolutely continuous with respect to" by "ac wrt" and "Radon–Nikodym derivative" by "RNd." Next, we define the null parameter spaces for (F, β^*), where F denotes the distribution of W_i and β^* denotes the true value of β, for the subvector AR and CQLR tests. The following set $\mathcal{F}_{\mathrm{AR},1}^S$ contains the restrictions needed to guarantee consistency of $\widehat{\beta}_n$ and $\widetilde{\beta}_n$. Let

$$\begin{aligned}\mathcal{F}_{\mathrm{AR},1}^S := \Big\{&\big(F, \beta^*\big) : E_F g_i = 0^k, F \text{ is ac wrt } \mu \text{ with RNd } f \text{ satisfying } f \leq M, \\ &\inf_{\beta \in B \setminus B(\beta^*, \zeta)} \big\|E_F g_i(\theta_0, \beta)\big\|^2 \geq \delta_\zeta \ \forall \zeta > 0, E_\mu \sup_{\beta \in B} \big\|g_i(\theta_0, \beta)\big\| < \infty, \\ &\sup_{\beta \in B} E_F \big\|g_i(\theta_0, \beta)\big\|^{1+\gamma} \leq M\Big\}\end{aligned} \tag{9.14}$$

[12]The reason $\widetilde{\Omega}_n^{-1/2}$ is employed, rather than $\widehat{\Omega}_n^{-1/2}$, is because $M_{\widetilde{J}_n(\eta)} \neq I_k$ when $b \geq 1$. When $b \geq 1$, $M_{\widetilde{J}_n(\eta)}$ has less than full rank and this has consequences for the asymptotic results and their proofs. See the footnote following (309) in Section 31 in the SM for details.

for constants $\delta_\zeta, \gamma > 0$ and $M < \infty$. Let

$$\begin{aligned}
\mathcal{F}^S_{\mathrm{AR},2} := \Big\{ &(F, \beta^*) : \text{for } B(\beta^*, \vartheta) \subset B, g(w, \theta_0, \cdot) \in C^2\big(B(\beta^*, \vartheta)\big)\ \forall w \in W, \\
& E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} \|h_i(\beta)\| \leq M \text{ and } \sup_{\beta \in B(\beta^*, \vartheta)} E_F \|h_i(\beta)\|^{1+\gamma} \leq M \\
& \text{for } h_i(\beta) \in \big\{\|g_i(\theta_0, \beta)\|^2, G_{i\beta}(\theta_0, \beta), g_{ij}(\theta_0, \beta) G_{i\beta}(\theta_0, \beta), \\
& \big(\partial^2/\partial\beta_m \partial\beta'\big) g_i(\theta_0, \beta), \big(\partial^2/\partial\theta_t \partial\beta'\big) g_i(\beta)\big\}, \\
& \lambda_{\min}\big(E_F g_i g_i'\big) \geq \delta, \tau_{\min}(E_F G_{i\beta}) \geq \delta \Big\}
\end{aligned} \tag{9.15}$$

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

The null parameter space for the subvector AR test is

$$\mathcal{F}^S_{\mathrm{AR}} := \mathcal{F}^S_{\mathrm{AR},1} \cap \mathcal{F}^S_{\mathrm{AR},2}. \tag{9.16}$$

The null parameter space for the subvector CQLR test is

$$\begin{aligned}
\mathcal{F}^S := \Big\{ &(F, \beta^*) \in \mathcal{F}^S_{\mathrm{AR}} : \max\Big\{ E_F \big\|g_i\big(\theta_0, \beta^*\big)\big\|^{4+\gamma}, E_F \big\|G_{i\beta}\big(\theta_0, \beta^*\big)\big\|^{2+\gamma}, \\
& E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} \|g_i(\theta_0, \beta)\|^3, E_\mu \sup_{\beta \in B(\beta^*, \vartheta)} \|G_i(\theta_0, \beta)\|^2, \\
& \sup_{\beta \in B(\beta^*, \vartheta)} E_F \|g_i(\theta_0, \beta)\|^{3+\gamma}, \sup_{\beta \in B(\beta^*, \vartheta)} E_F \|G_i(\theta_0, \beta)\|^{2+\gamma} \Big\} \leq M \Big\}.
\end{aligned} \tag{9.17}$$

The parameter spaces $\mathcal{F}^S_{\mathrm{AR}}$ and $\mathcal{F}^S$ impose correct specification of the model, impose uniform bounds on certain moments (which ensure that laws of large numbers and central limit theorems hold under drifting sequences of distributions), include an identifiability condition for β^* given θ_0, guarantee invertibility of the covariance matrix of g_i, and impose a minimum singular value condition on the expected Jacobian with respect to β of the moment functions. The condition $B(\beta^*, \vartheta) \subset B$ prevents β^* from converging to the boundary of B as $n \to \infty$. The assumption that g is twice continuously differentiable in β in a neighborhood of β^* is used in the proof of consistency and asymptotic normality of $\widehat{\beta}_n$ under drifting sequences of null distributions for W_i. The asymptotic results allow β^* to change with the sample size.

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

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

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

COMMENT. Theorem 9.1 is proved in Section 31 below.

9.3 *Asymptotic efficiency of the subvector CQLR test under strong and semi-strong identification*

In Section 7, it is established that the (full vector) SR-CQLR test is asymptotically efficient under strong or semi-strong identification when Ω_F has eigenvalues that are bounded away from zero and the null value θ_0 is not on the boundary. We next establish the analogous result for the subvector CQLR test. We consider drifting sequences $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$ of data-generating processes taken from $\mathcal{F}^S$ in (9.17) that correspond to strong or semi-strong identification and establish that the CQLR test statistic equals the subvector LM test statistic up to a $o_p(1)$ term and that the conditional critical value of the subvector CQLR test converges in probability to $\chi^2_{p,1-\alpha}$. Note that $\mathcal{F}^S$ imposes the minimal eigenvalue restriction $\lambda_{\min}(E_F g_i g_i') \geq \delta > 0$. It also imposes the restriction $\tau_{\min}(E_F G_{i\beta}) \geq \delta$, which implies strong identification of β.

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

$$\begin{aligned}\mathrm{LM}^S_n &:= n\widehat{g}_n(\widehat{\eta})'\widehat{\Omega}_n^{-1}(\widehat{\eta})\widehat{G}_{\eta n}(\widehat{\eta})\big(\widehat{G}_{\eta n}(\widehat{\eta})'\widehat{\Omega}_n^{-1}(\widehat{\eta})\widehat{G}_{\eta n}(\widehat{\eta})\big)^{-1} \\ &\quad \times \widehat{G}_{\eta n}(\widehat{\eta})'\widehat{\Omega}_n^{-1}(\widehat{\eta})\widehat{g}_n(\widehat{\eta}), \quad \text{where} \\ \widehat{G}_{\eta n}(\widehat{\eta}) &:= \big[\widehat{G}_n(\widehat{\eta}) : \widehat{G}_{\beta n}(\widehat{\eta})\big] \in R^{k\times(p+b)}\end{aligned} \tag{9.18}$$

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

Suppose $k - b \geq p$. The drifting sequences $\{\lambda^S_{n,h} : n \geq 1\}$ referred to above are rather complicated and so, for brevity, we define them in (321) and (322) in the SM. They are defined so that various population quantities that affect the asymptotic distributions of the CQLR test statistic and critical value converge as $n \to \infty$. Most importantly, we have that, along $\{\lambda^S_{n,h} : n \geq 1\}$, $n^{1/2}(\tau_{1F_n t^*}, \ldots, \tau_{pF_n t^*})$ converges to some vector $(h_{1,1t^*}, \ldots, h_{1,pt^*})$ whose elements may be finite or infinite, where $(\tau_{1F_n t^*}, \ldots, \tau_{pF_n t^*})$ denote the singular values of $O_{Ft^*}'(E_F g_i g_i')^{-1/2}(E_F G_i)U_F \in R^{(k-b)\times p}$. The latter quantity depends on the Jacobian $E_F G_i$, the moment variance matrix $E_F g_i g_i'$, the matrix $U_F \in R^{k\times p}$, which is the population counterpart of $\widehat{L}_n^{1/2}(\theta_0, \widehat{\beta}_n)$, and the matrix $O_{Ft^*} \in R^{k\times(k-b)}$, which is defined such that $O_{Ft^*}O_{Ft^*}'$ is a uniquely-defined population counterpart of the projection weight matrix $M_{\widetilde{J}_n(\eta)}$.[14] Strong or semi-strong identification occurs if the smallest singular value of $O_{Ft^*}'(E_F g_i g_i')^{-1/2}(E_F G_i)U_F$ diverges to infinity after renormalization by $n^{1/2}$, that is, if $h_{1,pt^*} = \infty$.

THEOREM 9.2. *Suppose Assumption* gB *holds and* $k - b \geq p$. *For any sequence* $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$ *that exhibits strong or semi-strong identification* (*where sequences* $\{\lambda^S_{n,h} \in \Lambda^S : n \geq 1\}$ *are defined precisely following* (322) *in Section* 9.1 *in the* SM *and strong and semi-strong identification are defined precisely in Section* 28 *in the* SM), *we have*

[14] The indexation of O_{Ft^*} by t^* is the result of the need to define a unique matrix O_{Ft^*} out of the many matrices $O_{Ft} \in R^{k\times(k-b)}$ for which $O_{Ft}O_{Ft}'$ is a population counterpart of $M_{\widetilde{J}_n(\eta)}$. See (320) and (322) in the SM for details.

(a) $\text{SR-QLR}_n^S(\theta_0, \widehat{\beta}_{\widehat{A}n}) = \text{QLR}_n^S(\widehat{\eta}) + o_p(1) = \text{LM}_n^S + o_p(1)$ *and*

(b)

$$\begin{aligned}&c_{\widehat{r}_n(\theta_0,\widetilde{\beta}_n),p}\big(n^{1/2}\widehat{D}^*_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}), \widetilde{J}_{\widehat{A}n}(\theta_0,\widehat{\beta}_{\widehat{A}n}), 1-\alpha\big)\\ &\quad = c_{k,p}\big(n^{1/2}\widehat{D}^*_n(\widehat{\eta}), \widetilde{J}_n(\widehat{\eta}), 1-\alpha\big) + o_p(1) \to_p \chi^2_{p,1-\alpha}.\end{aligned}$$

9.4 *Monte Carlo study: Probit model with endogeneity*

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

$$\begin{aligned}y_i &= 1\big(y_i^* > 0\big),\\ y_i^* &= \beta_0 + \beta_1 x_{1i} + \theta x_{2i} + u_i, \quad \text{and}\\ x_{2i} &= \widetilde{Z}_i'\pi + v_{2i},\end{aligned} \tag{9.19}$$

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

$$\begin{aligned}y_i^* &= \beta_0 + \beta_1 x_{1i} + \theta \widetilde{Z}_i'\pi + v_{1i}, \quad \text{where } v_{1i} := \theta v_{2i} + u_i,\\ (v_{1i}, v_{2i})' &\sim \text{iid } N\big(0^2, V\big), \qquad V := \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \in R^{2\times 2}\end{aligned} \tag{9.20}$$

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

The objective is to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ in the presence of the vector of nuisance parameters $\beta := (\beta_0, \beta_1, \pi')' \in R^g$.[15] We have

$$\begin{aligned}E(y_i|Z_i) &= \Pr(y_i = 1|Z_i) = \Pr\big(y_i^* > 0|Z_i\big)\\ &= \Pr\big(\beta_0 + \beta_1 x_{1i} + \theta\widetilde{Z}_i'\pi > -v_{1i}|Z_i\big) = \Phi\big(\beta_0 + \beta_1 x_{1i} + \theta\widetilde{Z}_i'\pi\big).\end{aligned} \tag{9.21}$$

The model implies the moment conditions $Eg_i(\theta, \beta) = 0$, where

$$g_i(\theta, \beta) := \begin{pmatrix} \big(y_i - \Phi(\beta_0 + \beta_1 x_{1i} + \theta\widetilde{Z}_i'\pi)\big)Z_i \\ \big(x_{2i} - \widetilde{Z}_i'\pi\big)Z_i \end{pmatrix} \in R^{2g}. \tag{9.22}$$

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

[15]The other nuisance parameters ρ and σ do not enter the moment function g_i defined below.

Given a weighting matrix $\widehat{W}_n$, the GMM criterion function is $Q_n^{\widehat{W}_n}(\theta,\beta) := \widehat{g}_n(\theta,\beta)' \times \widehat{W}_n \widehat{g}_n(\theta,\beta)$. Taking $\widehat{W}_n = I_k$, the first-step GMM estimator $\widehat{\beta}_{n,FS}$ of β minimizes $Q_n^{I_k}(\theta_0,\beta)$. The second-step GMM estimator $\widehat{\beta}_n$ minimizes $Q_n^{\widehat{W}_n}(\theta_0,\beta)$, where $\widehat{W}_n := n^{-1}\sum_{i=1}^n g_i(\theta_0,\widehat{\beta}_{n,FS}) g_i(\theta_0,\widehat{\beta}_{n,FS})'$.[16]

In the simulation results reported below, the nominal size of the tests is 5%. We take $\theta_0 = 1$ (the null value of θ), $\beta_0 = \beta_1 = 1$, and $\sigma = 2$. In addition to the null value, we consider three true values of θ on each side of the null such that the resulting RPs of the subvector CQLR test are roughly equal to 40%, 65%, and 90%. We let $\pi \in R^{g-2}$ be a multiple of a vector of ones with a multiplicative constant $\overline{\pi}$. The latter determines the strength of identification of θ. We consider 16 parameter configurations consisting of all of the combinations of $g = 3, 4$ (which results in $k = 6, 8$), $\rho = 0, 0.9$, and $\overline{\pi} = 1, 0.5, 0.2, 0.1$. The sample size is $n = 1000$. The results are based on 5000 simulation repetitions, and 5000 simulation repetitions are used to simulate the critical values for each data sample. When calculating the QLR statistic in AM, we use 60 search points to find the infimum over θ (see equation (2) on p. 1575 in AM).

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

Figure 1 reports power for the four tests for $\rho = 0.9$ and $\overline{\pi} = 1, 0.5, 0.2, 0.1$ for $g = 3$ (upper row) and $g = 4$ (lower row) for three alternatives to the left of the null value of θ, the null value, and three alternatives to the right of the null value. For clarity, the graphs linearly interpolate the power between the seven θ values. Figure SM-1 in the SM provides the corresponding results for $\rho = 0$. As expected, the powers of all tests decrease as $\overline{\pi}$ decreases. Thus, the CQLR test reaches the 40, 65, and 90% RPs for alternatives farther from the true value the smaller is $\overline{\pi}$, with all other parameters held constant. For example, in the upper panel of Figure 1, which reports power when $g = 3$, the sum of the distances to the alternative θ values to the left and right of the null

[16]We use the Newton–Raphson algorithm to find the two-step GMM estimator for β. In both steps, we initiate the search from a number of starting points and do ten Newton iterations from each starting point. In particular, for the first step estimator we use $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\pi})$ as one starting value, where $(\widehat{\beta}_0, \widehat{\beta}_1)$ is the OLS estimator of the slope coefficients in a regression of $y - \theta_0 x_2$ on a constant and x_1 and $\widehat{\pi}$ is the OLS estimator in a regression of x_2 on $\widetilde{Z}$ and we use $(\widehat{\beta}_0, \widehat{\beta}_1, \pi)$ as another starting value, where π is the true value of the slope coefficients in the third line of (9.19). For the second step, we use the same starting values and also the estimator obtained in the first step. We also experimented using an additional fifteen randomly generated starting points which had little effect on the results. In each Newton iteration, we incorporate a step size control where along the search direction the step is divided in seven equal parts and the next iteration proceeds from the step that yields the smallest criterion function. For numerical stability when inverting matrices, we replace all eigenvalues of the matrices smaller than 10^{-11} by 10^{-11}. We use $\varepsilon = 0.01$ for the eigenvalue adjustment constant in (9.11). The estimator of β in each of the two steps is the minimizer of the stochastic criterion function over all candidate vectors for which the criterion function was evaluated in that step.

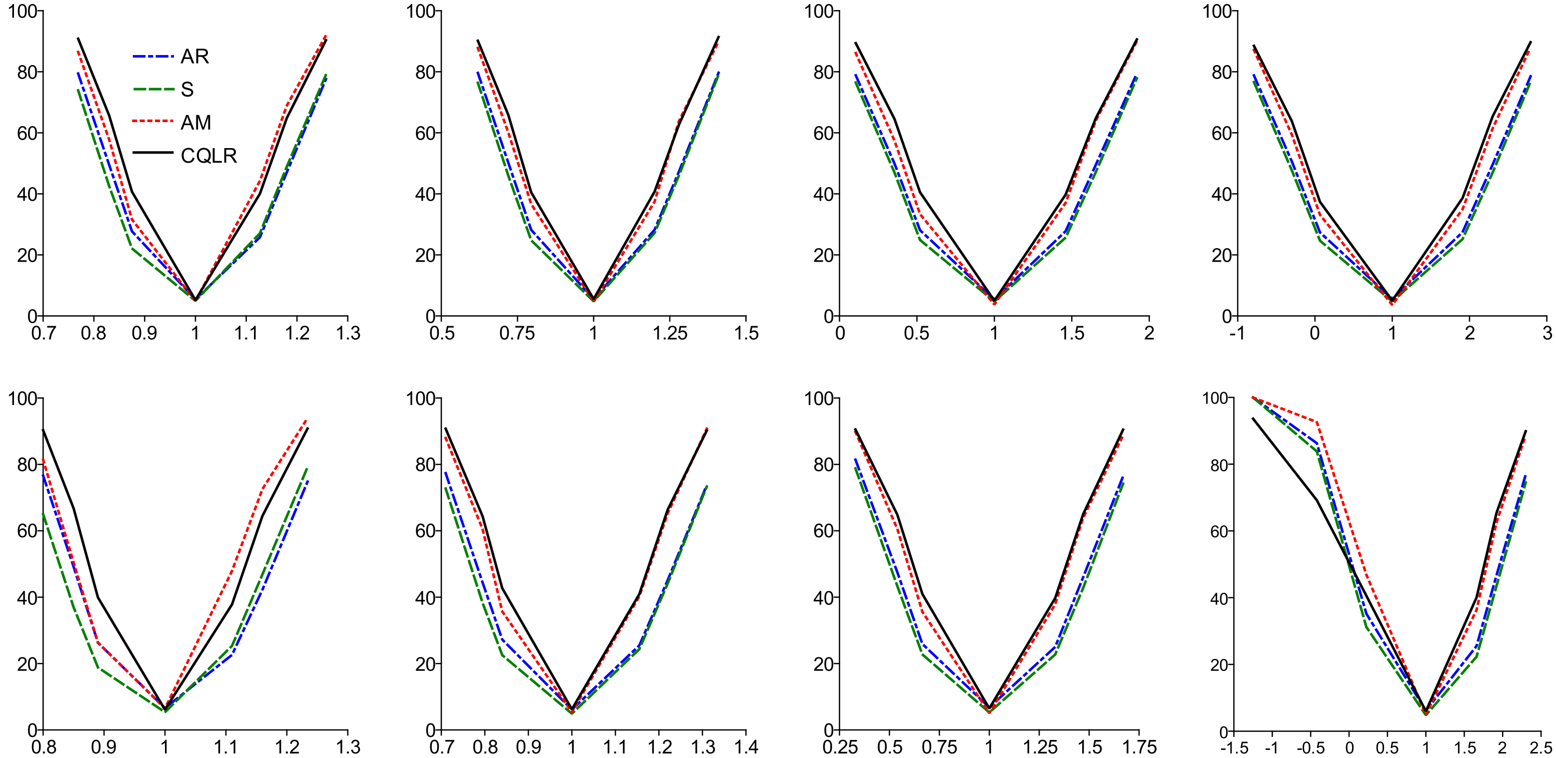

FIGURE 1. Power of CQLR, AM, S, and AR as function of θ for $\rho = 0.9$ and $\pi = 1, 0.5, 0.2, 0.1$; first/second row $g = 3/4$.

value such that the CQLR test has 90% power are roughly 0.48, 0.78, 1.82, and 3.59 for $\overline{\pi} = 1, 0.5, 0.2$, and 0.1, respectively. The powers of the tests increase as g increases from 3 to 4 (with other parameters held constant) with the corresponding sum of the distances being roughly 0.43, 0.60, 1.34, and 3.55; see the lower panel of Figure 1. The powers of the tests decrease as ρ decreases from 0.9 to 0 with other parameters held constant.

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

Overall, the CQLR and AM tests are the best two tests among the four tests considered. The CQLR test has higher power than the AM test *for all alternatives to the left of the null* value in 14 of the 16 parameter configurations with power gains up to 16.5% when $\overline{\pi} = 1$ (see Figure 1 with $g = 4$ and $\overline{\pi} = 1$) and up to 7.5% for $\overline{\pi} \leq 0.5$ (see Figure 1 with $g = 3$ and $\overline{\pi} = 0.2$). The AM test has higher power than CQLR for alternatives to the left in the two cases $(g, \overline{\pi}, \rho) = (4, 0.1, 0)$ and $(4, 0.1, 0.9)$, for example, see Figure 1 with $g = 4$ and $\overline{\pi} = 0.1$. For this parameter configuration, the highest power advantage of the AM test is 23% for $\theta = -0.42$.

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

With regard to computation time, it takes about 231 minutes to calculate 5000 AM tests when $(g, \overline{\pi}, \rho) = (4, 0.5, 0.9)$ under the specifications described above using an Intel Core 3.4 GHz, 6 MB processor. On the other hand, it only takes about 4 minutes to calculate 5000 CQLR tests, that is, the CQLR test is about 58 times faster to calculate. The difference in computation times is expected to be much larger in cases where θ is of dimension greater than 1, because the computation time of the AM test increases exponentially in the dimension of θ, whereas the computation time of the CQLR test does not depend on the dimension of θ. Computation time is particularly important when computing a confidence set by inverting a test, because the test has to be computed many times.

10. Power comparisons in heteroskedastic/autocorrelated linear IV models

In this section, we present some power comparisons for the AR test, Kleibergen's (2005) LM, JVW-CLR, and MVW-CLR tests, and the SR-CQLR test introduced above.[17] We also consider the plug-in conditional linear combination (PI-CLC) test introduced in I. Andrews (2016), as well as the MM1-SU and MM2-SU tests introduced in Moreira and Moreira (forthcoming). The PI-CLC test aims to approximate the test that has minimum

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

regret among conditional tests constructed using linear combinations of the LM and AR test statistics (with coefficients that depend on the conditioning statistic); see I. Andrews (2016) for details.[18] The MM1-SU and MM2-SU tests have optimal weighted average power for two different weight functions (over the alternative parameter values θ and the strength of identification parameter vector μ, given in (10.1) below) among tests that satisfy a sufficient condition for local unbiasedness.[19]

We consider the same designs as in I. Andrews (2016, Section 7.2). These designs are for heteroskedastic and/or autocorrelated linear IV models with $p=1$ and $k=4$. The designs are calibrated to mimic the linear IV models for the elasticity of intertemporal substitution estimated by Yogo (2004) for eleven countries using quarterly data from the early 1970s to the late 1990s. The power comparisons are for the limiting experiment under standard weak identification asymptotics. In consequence, for the simulations, the observations are drawn from the following model:

$$\begin{pmatrix} \widehat{\Omega}_n^{-1/2} n^{1/2} \widehat{g}_n(\theta_0) \\ \widehat{\Omega}_n^{-1/2} n^{1/2} \widehat{G}_n(\theta_0) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu\theta \\ \mu \end{pmatrix}, \begin{pmatrix} I_k & \Sigma_{gG} \\ \Sigma_{gG}' & \Sigma_{GG} \end{pmatrix} \right) \tag{10.1}$$

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

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

Some details concerning the computation and definitions of the SR-CQLR, PI-CLC, MM1-SU, and MM2-SU tests are as follows. The SR-CQLR test uses $\varepsilon = 0.01$, where ε

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

[19]The weight functions considered depend on the variance parameters Σ_{gG} and Σ_{GG} in (10.1) below.

[20]In linear IV models with i.i.d. observations, the matrix Σ_{gG} is necessarily symmetric. However, with autocorrelation, it need not be. In the eleven countries considered here, it is not.

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

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

[23]For example, $\widehat{\Gamma}_{jn}(\theta_0)$ in (5.2) is taken to be known and equal to Σ_{gG}', and $\widetilde{V}_n(\theta_0)$ in (74) is taken to be known and equal to the variance matrix in (10.1).

appears in the definition of $\widehat{L}_n(\theta)$ in (5.7).[24] For the PI-CLC test, the number of values "a" considered in the search over [0, 1] is 100, the number of simulation repetitions used to determine the best choice of "a" is 2000, and the number of alternative parameter values considered in the search for the best "a" is 41. For the MM1-SU and MM2-SU tests, the number of variables in the discretization of maximization problem is 1000, the number of points used in the numerical approximations of the integrals $h1$ and $h2$ that appear in the definitions of these tests is 1000, and when approximating integrals $h1$ and $h2$ by sums of 1000 rectangles these rectangles cover $[-4, 4]$.

The asymptotic power functions are given in Figure 2. Each graph is based on 41 equispaced values on the x axis covering $[-6, 6]$. The x axis variable is the parameter θ scaled by a fixed value of $\|\mu\|$ for a given country, thus $\theta\|\mu\| \in [-6, 6]$, where θ is the alternative parameter value (when $\theta \neq 0$) defined in (10.1) and μ is the mean vector that determines the strength of identification. The y axis variable is power $\times 100$.

Table 3 provides the *shortfall in average-power* ($\times 100$) of each test for each country relative to the other seven tests considered, where average power is an unweighted average over the 40 alternative parameter values. Table 4 provides the *maximum power shortfall* ($\times 100$) of each test for each country relative to the other seven tests considered, where the maximum is taken over the 40 alternative parameter values.[25] The shortfall in average-power is an unweighted average power criterion, whereas the maximum power shortfall is a minimax regret criterion.

The last row of Table 3 shows the average (across countries) of the shortfall in average-power ($\times 100$) of each test. This provides a summary measure. Similarly, the last row of Table 4 shows the average (across countries) of the maximum power shortfall ($\times 100$) of each test.

The second and third columns of Table 3 provide the concentration parameter, $\mu'\mu$, which measures the strength of identification, and a non-Kronecker index, abbreviated by non-Kron, which measures the deviation of the variance matrix in (10.1), call it Ψ, from a Kronecker matrix. This deviation is given by the formula $1000 \times \min_{B,C} \|B \otimes C - \Psi\|$, where the minimum is taken over symmetric pd matrices B and C of dimensions 2×2 and 4×4, respectively, $\|\cdot\|$ denotes the Frobenius norm, and the rescaling by 1000 is for convenience.[26] Germany, Japan, and the Netherlands exhibit the weakest identification, while Sweden and Australia exhibit the strongest. The U.K., Australia, Italy, and Japan

[24]The numerical results are unchanged when $\varepsilon = 0.001$ or 0.05.

[25]More precisely, let AP_{tc} denote the average power of test t for country c, where the average is taken over the 40 parameter values in the alternative hypothesis. By definition, the *shortfall in average-power* of test t for country c is $\max_{s\leq 8} \mathrm{AP}_{sc} - \mathrm{AP}_{tc}$, where the maximum is taken over the eight tests considered.

Let $P_{tc}(\theta)$ denote the power of test t in country c against the alternative θ. By definition, the power shortfall of test t in country c for alternative θ is $\max_{s\leq 8} P_{sc}(\theta) - P_{tc}(\theta)$ and the *maximum power shortfall* of test t in country c is $\max_{\theta\in\Theta_{40}}(\max_{s\leq 8} P_{sc}(\theta) - P_{tc}(\theta))$, where Θ_{40} contains the 40 alternative parameter values considered.

Note that, as defined, the shortfall in average-power is not equal to the average of the power shortfalls over $\theta \in \Theta_{40}$.

[26]The non-Kronecker index is computed using the Framework 2 method given in Section 4 of Van Loan and Pitsianis (1993) with symmetry of C imposed by replacing $\widehat{A}_{ij}$ by $(\widehat{A}_{ij} + \widehat{A}_{ji})/2$ in equation (9) of that paper.

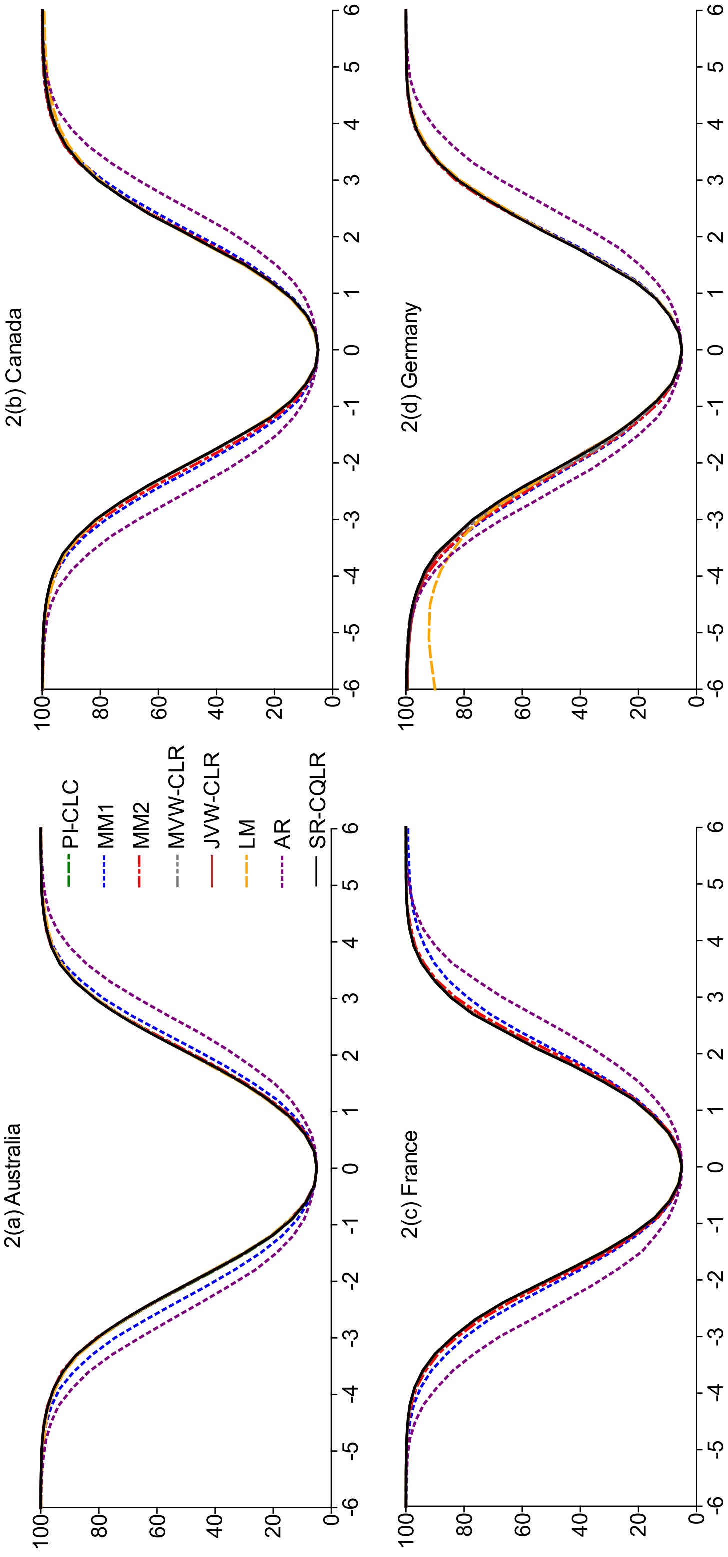

Figure 2. Power of SR-CQLR and other tests as function of θ for model of Yogo (2004).

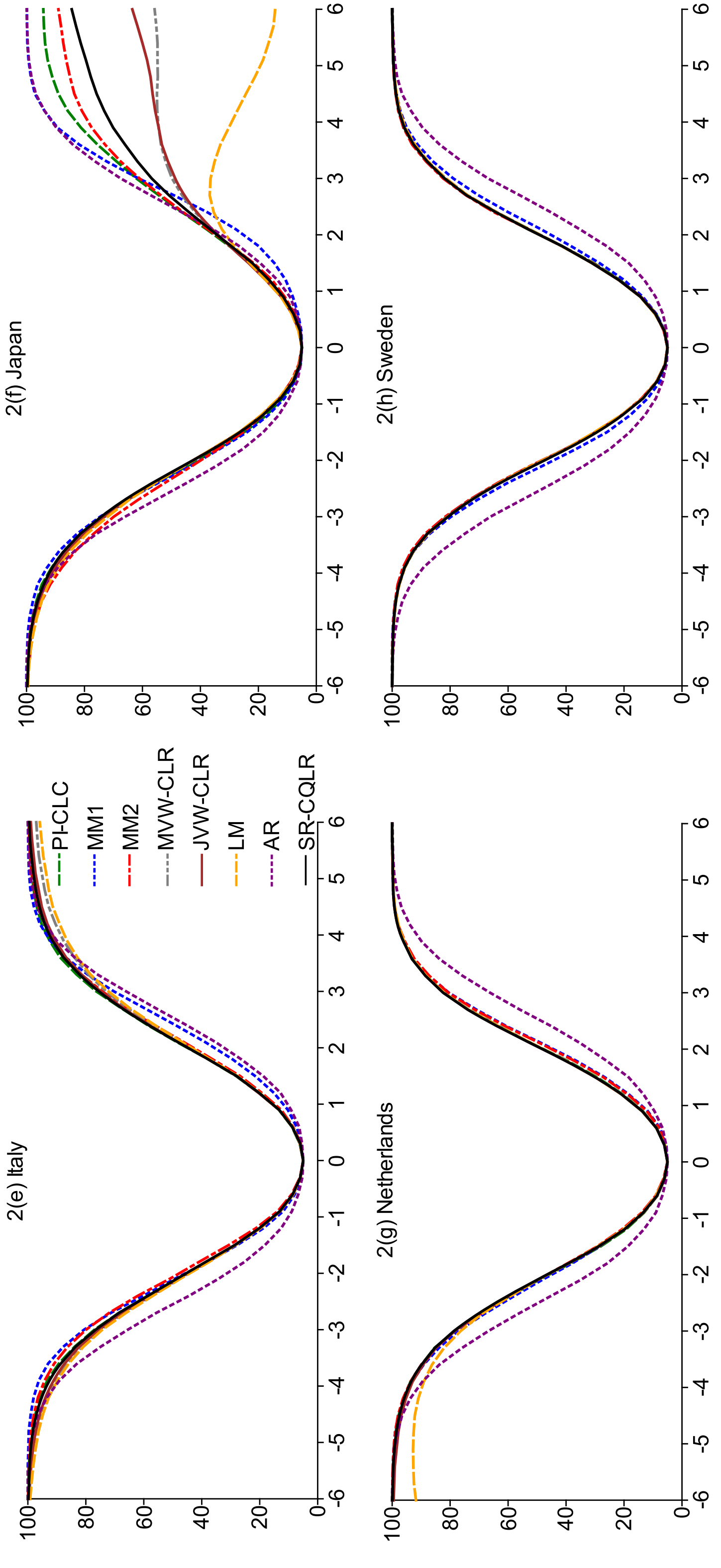

FIGURE 2. *Continued.*

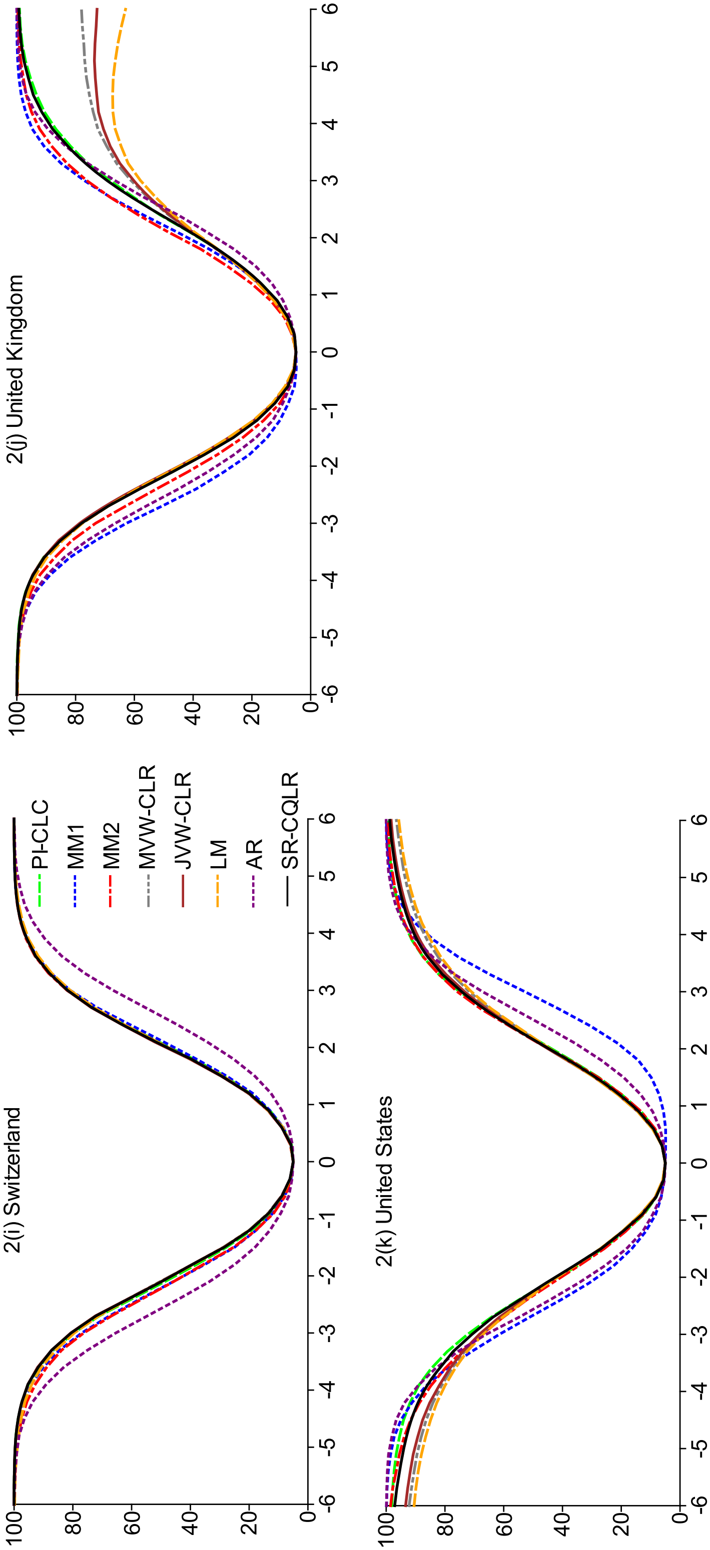

FIGURE 2. *Continued.*

Table 3. Shortfalls in average-power (×100).

Country	$\mu'\mu$	non-Kron	SR-CQLR	JVW	MVW	PI-CLC	MM1	MM2	LM	AR
Australia	138	17	0.0	0.1	0.1	0.2	2.4	0.1	0.1	6.9
Canada	48	5	0.0	0.0	0.2	0.0	1.4	0.5	0.3	6.8
France	79	6	0.1	0.2	0.0	0.3	0.7	0.3	0.0	8.0
Germany	10	3	0.0	0.1	0.4	0.0	0.2	0.1	2.3	6.5
Italy	84	15	0.5	1.1	2.0	0.2	1.1	0.0	2.6	5.5
Japan	17	14	3.3	3.2	8.9	0.4	0.0	2.4	17.4	0.6
Netherlands	25	3	0.0	0.2	0.1	0.2	0.9	0.5	1.6	6.6
Sweden	174	9	0.3	0.2	0.3	0.2	1.5	0.0	0.3	7.5
Switzerland	31	4	0.1	0.0	0.0	0.4	1.3	1.1	0.5	7.2
U. K.	53	38	0.7	6.0	5.4	0.8	2.5	0.0	7.8	3.8
U.S.	81	10	0.8	2.0	2.9	0.0	7.3	0.8	3.5	3.2
Average over countries			0.5	1.2	1.8	0.2	1.8	0.5	3.3	5.7

have variance matrices that are farthest from Kronecker-product form, while Germany, the Netherlands, and Switzerland have variance matrices that are closest to Kronecker-product form.

The test that performs best in Tables 1 and 2 is the PI-CLC test, followed closely by the SR-CQLR and MM2-SU tests. The difference between these tests is not large. For example, the difference in the average (across countries) shortfall in average-power (not rescaled by multiplication by 100 in contrast to the results in Table 3) of the PI-CLC test and the SR-CQLR and MM2-SU tests is 0.003. This small power advantage is almost entirely due to the relative performances for Japan, which exhibits very weak identification and moderately large non-Kronecker index.

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

Table 4. Maximum power shortfalls (×100).

Country	$\mu'\mu$	non-Kron	SR-CQLR	JVW	MVW	PI-CLC	MM1	MM2	LM	AR
Australia	138	17	0.5	0.6	0.8	1.0	8.2	1.3	0.9	17.2
Canada	48	5	0.6	0.5	0.9	0.7	5.4	3.0	1.7	17.7
France	79	6	0.7	0.8	0.5	1.0	3.0	1.6	0.4	19.9
Germany	10	3	0.8	0.8	2.2	0.6	1.0	0.8	10.6	18.4
Italy	84	15	4.4	5.7	6.5	3.9	9.7	2.3	7.1	17.7
Japan	17	14	21.3	41.4	44.9	8.6	10.1	13.6	85.8	11.9
Netherlands	25	3	0.9	1.1	0.9	1.4	3.9	3.3	8.2	18.6
Sweden	174	9	1.0	0.6	1.0	0.7	4.9	0.4	1.1	19.6
Switzerland	31	4	0.5	0.3	0.5	1.6	4.8	5.5	1.4	18.8
U. K.	53	38	8.4	27.3	23.2	9.0	20.6	7.1	37.0	14.7
U.S.	81	10	5.2	9.0	10.2	2.6	27.7	5.1	11.7	12.4
Average over countries			4.0	8.0	8.3	2.8	9.0	4.0	14.9	17.0

noticeably lower power than the other tests in an overall sense, and the AR test has noticeably lower power than the LM test.

We conclude that the SR-CQLR test has asymptotic power that is competitive with, or better than, that of other tests in the literature for the particular parameters considered here in the particular model considered here. The SR-CQLR test has advantages compared to the PI-CLC, MM1-SU, and MM2-SU tests of (i) being applicable in almost any moment condition model, whereas the latter tests are not,[27] (ii) being easy to implement (that is, program), and (iii) being fast to compute.

References

Anderson, T. W. and H. Rubin (1949), "Estimation of the parameters of a single equation in a complete set of stochastic equations." *Annals of Mathematical Statistics*, 20, 46–63. [1703, 1704]

Andrews, D. W. K. and X. Cheng (2013), "GMM estimation and uniform subvector inference with possible identification failure." *Econometric Theory*, 30, 1–47. [1710, 1723]

Andrews, D. W. K., X. Cheng, and P. Guggenberger (forthcoming), "Generic results for establishing the asymptotic size of confidence sets and tests." *Journal of Econometrics*. [1706]

Andrews, D. W. K. and P. Guggenberger (2010), "Asymptotic size and a problem with subsampling and the *m* out of *n* bootstrap." *Econometric Theory*, 26, 426–468. [1706]

Andrews, D. W. K. and P. Guggenberger (2017), "Asymptotic size of Kleibergen's LM and conditional LR tests for moment condition models." *Econometric Theory*, 33, 1046–1080. [1708]

Andrews, D. W. K. and P. Guggenberger (2019), "Supplement to 'Identification- and singularity-robust inference for moment condition models'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE1219. [1704, 1707]

Andrews, D. W. K., V. Marmer, and Z. Yu (2019), "On optimal inference in the linear IV model." *Quantitative Eonomics*, 10, 457–485. [1705]

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006), "Optimal two-sided invariant similar tests for instrumental variables regression." *Econometrica*, 74, 715–752. [1705]

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2008), "Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments." *Journal of Econometrics*, 146, 241–254. [1705]

Andrews, I. (2016), "Conditional linear combination tests for weakly identified models." *Econometrica*, 84, 2155–2182. [1703, 1707, 1708, 1709, 1737, 1738]

[27]The PI-CLC test does not apply to moment condition models with possibly singular variance matrices. The MM1-SU and MM2-SU tests apply only to the linear IV model with errors that may be heteroskedastic and/or autocorrelation.

Andrews, I. and A. Mikusheva (2015), "Maximum likelihood inference in weakly identified DSGE models." *Quantitative Economics*, 6, 123–152. [1705]

Andrews, I. and A. Mikusheva (2016), "Conditional inference with a functional nuisance parameter." *Econometrica*, 84, 1571–1612. [1706, 1708, 1709, 1734]

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile prices in market equilibrium." *Econometrica*, 60, 889–917. [1704]

Campbell, J. Y. (2003), "Consumption-based asset pricing." In *Handbook of the Economics of Finance*, Vol. 1B (G. M. Constantinides, M. Harris, and R. M. Stulz, eds.), 801–885, Elsevier, Amsterdam. Chapter 13. [1707, 1723, 1724, 1725]

Caner, M. and N. Yildez (2012), "CUE with many weak instruments and nearly singular design." *Journal of Econometrics*, 170, 422–441. [1709]

Chan, W. H. and J. M. Maheu (2002), "Conditional jump dynamics in stock market returns." *Journal of Business and Economic Statistics*, 20, 377–389. [1715]

Chernozhukov, V., C. Hansen, and M. Jansson (2009), "Admissible invariant similar tests for instrumental variables regression." *Econometric Theory*, 25, 806–818. [1705]

Cox, G. (2017), "Weak identification in a class of generically identified models with an application to factor models." Unpublished working paper, Department of Economics, Columbia University. [1710]

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. [1709]

Grant, N. (2013), "Identification robust inference with singular variance." Economics Discussion Paper No. 1315, Department of Economics, University of Manchester. [1711]

Guggenberger, P., J. J. S. Ramalho, and R. J. Smith (2012), "GEL statistics under weak identification." *Journal of Econometrics*, 170, 331–349. [1708]

Guggenberger, P. and R. J. Smith (2005), "Generalized empirical likelihood estimators and tests under partial, weak and strong identification." *Econometric Theory*, 21, 667–709. [1708]

Han, S. and A. McCloskey (2019), "Estimation and inference with a (nearly) singular Jacobian." *Quantitative Economics*, 10, 1019–1068. [1710]

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators." *Econometrica*, 50, 1029–1054. [1709]

Jagannathan, R., G. Skoulakis, and Z. Wang (2002), "Generalized method of moments: Applications in finance." *Journal of Business and Economic Statistics*, 20, 470–481. [1711]

Jorion, P. (1988), "On jump processes in the foreign exchange and stock markets." *Review of Financial Studies*, 1, 427–445. [1715]

Kleibergen, F. (2005), "Testing parameters in GMM without assuming that they are identified." *Econometrica*, 73, 1103–1123. [1706, 1707, 1708, 1716, 1729, 1737]

Kleibergen, F. (2007), "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics." *Journal of Econometrics*, 139, 181–216. [1707, 1708]

Mikusheva, A. (2010), "Robust confidence sets in the presence of weak instruments." *Journal of Econometrics*, 157, 236–247. [1705]

Montiel Olea, J. L. (forthcoming), "Efficient conditionally similar-on-the-boundary tests." *Econometric Theory*. [1709]

Moreira, H. and M. J. Moreira (forthcoming), "Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors." *Journal of Econometrics*. [1707, 1709, 1717, 1720, 1737, 1738]

Moreira, M. J. (2003), "A conditional likelihood ratio test for structural models." *Econometrica*, 71, 1027–1048. [1705, 1707, 1716, 1717]

Newey, W. K. and K. West (1987), "Hypothesis testing with efficient method of moments estimation." *International Economic Review*, 28, 777–787. [1722, 1723, 1733]

Otsu, T. (2006), "Generalized empirical likelihood inference for nonlinear and time series models under weak identification." *Econometric Theory*, 22, 513–527. [1708]

Robin, J.-M. and R. J. Smith (2000), "Tests of rank." *Econometric Theory*, 16, 151–175. [1708]

Smith, R. J. (2007), "Weak instruments and empirical likelihood: A discussion of the papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura." In *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Vol. III* (R. Blundell, W. K. Newey, and T. Persson, eds.), Cambridge University Press, Cambridge, UK. Also available as CEMMAP Working Paper No. 13/05, UCL. [1708]

Staiger, D. and J. H. Stock (1997), "Instrumental variables regression with weak instruments." *Econometrica*, 65, 557–586. [1708, 1721]

Stock, J. H. and J. H. Wright (2000), "GMM with weak identification." *Econometrica*, 68, 1055–1096. [1705, 1706, 1708, 1712, 1721, 1724, 1734, 1737]

Van Loan, C. F. and N. Pitsianis (1993), "Approximation with Kronecker products." In *Linear Algebra for Large Scale and Real-Time Applications* (M. S. Moonen, G. H. Golub, and B. L. R. de Moor, eds.), 293–314, Kluwer Academic Publishers, Amsterdam. [1717, 1739]

Yogo, M. (2004), "Estimating the elasticity of intertemporal substitution when instruments are weak." *Review of Economics and Statistics*, 86, 797–810. [1707, 1723, 1724, 1725, 1727, 1738, 1740]

Co-editor Andres Santos handled this manuscript.

Manuscript received 8 October, 2018; final version accepted 5 April, 2019; available online 23 April, 2019.