IDENTIFICATION- AND SINGULARITY-ROBUST INFERENCE FOR MOMENT CONDITION MODELS

By

Donald W. K. Andrews and Patrik Guggenberger

January 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1978

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

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

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

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

*Andrews and Guggenberger gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-1058376 and SES-1355504, and SES-1021101 and SES-1326827, respectively. The authors thank Isaiah Andrews for providing the details of the simulation designs employed in I. Andrews (2014), which are used in some of the simulations given in this paper. For helpful comments, the authors thank Isaiah Andrews, Xu Cheng, Kirill Evdokimov, Anna Mikusheva, Zhongjun Qu, Jim Stock, and the participants of seminars at the following universities: Boston, Boston College, Brown, Chicago, Columbia, Freiburg, Hanover, Harvard/MIT, Hebrew Jerusalem, Konstanz, Maryland, Michigan, New York, Northwestern, Ohio State, Princeton, Queen's, Strasbourg, and Wisconsin.

Abstract

This paper introduces two new identification- and singularity-robust conditional quasi-likelihood ratio (SR-CQLR) tests and a new identification- and singularity-robust Anderson and Rubin (1949) (SR-AR) test for linear and nonlinear moment condition models. The paper shows that the tests have correct asymptotic size and are asymptotically similar (in a uniform sense) under very weak conditions. For two of the three tests, all that is required is that the moment functions and their derivatives have $2+\gamma$ bounded moments for some $\gamma > 0$ in i.i.d. scenarios. In stationary strong mixing time series cases, the same condition suffices, but the magnitude of γ is related to the magnitude of the strong mixing numbers. For the third test, slightly stronger moment conditions and a (standard, though restrictive) multiplicative structure on the moment functions are imposed. For all three tests, 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 two SR-CQLR tests are 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 two SR-CQLR tests reduce asymptotically to Moreira's CLR test when $p = 1$ in the homoskedastic linear IV model. The first SR-CQLR test, which relies on the multiplicative structure on the moment functions, also does so for $p \geq 2$.

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

JEL Classification Numbers: C10, C12.

1 Introduction

Weak identification and weak instruments (IV's) can arise in a wide variety of empirical applications in economics. Examples include: in macroeconomics and finance, new Keynesian Phillips curve models, dynamic stochastic general equilibrium (DSGE) models, consumption capital asset pricing models (CCAPM), and interest rate dynamics models; in industrial organization, the Berry, Levinsohn, and Pakes (1995) (BLP) model of demand for differentiated products; and in labor economics, returns-to-schooling equations that use IV's, such as quarter of birth or Vietnam draft lottery status, to avoid ability bias.[1] 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.[2] Given this wide range of applications, numerous methods have been developed in the econometrics literature over the last two decades that aim to be identification-robust.

The most important feature of tests and confidence sets (CS's) that aim to be identification-robust is that they control size for a wide range of null distributions regardless of the strength of identification of the parameters. This holds if the tests have correct asymptotic size for a broad class of null distributions. However, the asymptotic size of many tests in the literature that are designed to be identification-robust has not been established. This paper and its companion paper, Andrews and Guggenberger (2014a) (hereafter AG1), help fill this void by establishing the asymptotic size and similarity properties of three new tests and CS's and the influential nonlinear Lagrange multiplier (LM) and conditional likelihood ratio (CLR) tests and CS's of Kleibergen (2005, 2007) and the GMM versions of the tests that appear in Guggenberger and Smith (2005), Otsu (2006), Smith (2007), Newey and Windmeijer (2009), and Guggenberger, Ramalho, and Smith (2012). None of the aforementioned tests and CS's have been shown to have correct asymptotic size for moment condition models (even linear ones) with multiple sources of possible weak identification.

[1] For new Keynesian Phillips curve models, see Dufour, Khalaf, and Kichian (2006), Nason and Smith (2008), and Kleibergen and Mavroeidis (2009). For DSGE models, see Canova and Sala (2009), Iskrev (2010), Qu and Tkachenko (2012), Dufour, Khalaf, and Kichian (2013), Guerron-Quintana, Inoue, and Kilian (2013), I. Andrews and Mikusheva (2014b), Qu (2014), and Schorfheide (2014). For the CCAPM, see Stock and Wright (2000), Neely, Roy, and Whiteman (2001), Yogo (2004), Kleibergen (2005), Carroll, Slacalek, and Sommer (2011), and Gomes and Paz (2013). For interest rate dynamics, see Jegannathan, Skoulakis, and Wang (2002) and Grant (2013). For the BLP model, see Armstrong (2012). For the returns-to-schooling wage equations, see Angrist and Krueger (1991, 1992) and Cruz and Moreira (2005).

[2] For the time series models, see Hannan (1982), Teräsvirta (1994), Nelson and Startz (2007), and Andrews and Cheng (2012, 2013b). For the selection model, see Puhani (2000). For the mixing and regime switching models, see Cho and White (2007), Chen, Ponomareva, and Tamer (2014), and references therein. For the nuisance parameter only under the alternative models, see Davies (1977) and Andrews and Ploberger (1994).

By this we mean that one or more parameters (or transformations of parameters) may be weakly or strongly identified. In addition, the approach and results of the present paper and AG1 should be useful for assessing the asymptotic size of other tests and CS's for moment condition models that allow for multiple sources of weak identification.

The three new tests introduced here include two singularity-robust (SR) conditional quasi-likelihood ratio (SR-CQLR) tests and an SR nonlinear Anderson and Rubin (1949) (SR-AR) test. These tests and CS's are shown to have correct asymptotic size and to be asymptotically similar (in a uniform sense) under very weak conditions. All that is required is that the expected moment functions equal zero at the true parameter value and the moment functions and their derivatives satisfy mild moment conditions. Thus, no identification assumptions of any type are imposed. The results hold for arbitrary fixed $k, p \geq 1$, where k is the number of moment conditions and p is the number of parameters. The case $k \geq p$ is of greatest interest in practice, but the results also hold for $k < p$ and treatment of the $k < p$ case is needed for the SR results. The results allow for any of the p parameters to be weakly or strongly identified, which yields 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 singular (or near singular). This is particularly important in models where lack of identification is accompanied by singularity of the variance matrix of the moments. For example, this occurs in all maximum likelihood scenarios and many quasi-likelihood scenarios. Other examples where it holds are given below. Some finite-sample simulation results, given in the Supplemental Material (SM) to this paper, show that the SR-AR and SR-CQLR tests perform well (in terms of null rejection probabilities) under singular and near singular variance matrices of the moments in the model considered.

In addition, the asymptotic results 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.

The two SR-CQLR tests are 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, as shown in the SM, they reduce to Moreira's (2003) CLR test in the homoskedastic linear IV model with fixed IV's when $p = 1$. This is desirable because the latter test has been shown to have approximate optimal power properties in this model under normality, see Andrews, Moreira, and Stock (2006, 2008) and

Chernozhukov, Hansen, and Jansson (2009).[3] The first SR-CQLR test applies when the moment functions are of the form $u_i(\theta)Z_i$, where $u_i(\theta)$ is a scalar and Z_i is a k vector of IV's, as in Stock and Wright (2000). It reduces to Moreira's CLR test for all $p \geq 1$. The second SR-CQLR test does not require the moment functions to have this form. A drawback of the SR-CQLR tests is that they are not known to have optimality properties under weak identification in other models, see the discussion in Section 2 below. The SR-CQLR tests are easy to compute and their conditional critical values can be simulated easily and very quickly. Constructing CS's by inverting the tests typically is more challenging computationally.

Now, we contrast the aforementioned asymptotic size results with the asymptotic size results of AG1 for Kleibergen's (2005) Lagrange multiplier (LM) and conditional likelihood ratio (CLR) tests. AG1 shows that Kleibergen's LM test has correct asymptotic size for a certain parameter space of null distributions $\mathcal{F}_0$. AG1 shows that this also holds 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 suitable form of a rank statistic, such as the Robin and Smith (2000) rank statistic. Tests of this type have been considered by Newey and Windmeijer (2009) and 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 $p \geq 2$ (i.e., in the case of multiple sources of weak identification). 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 asymptotic independence is needed to show that the asymptotic null rejection probabilities reduce to the nominal size α.[4] 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 (as noted above).[5] Second, the parameter space $\mathcal{F}_0$ restricts

[3] For related results, see Chamberlain (2007), Mikusheva (2010), Montiel Olea (2012), and Ploberger (2012).

[4] Lack of asymptotic independence can occur because the estimation of the variance matrix of the Jacobian of the moments can affect the asymptotic distribution of the Jacobian-variance weighted CLR test statistic under sequences of null distributions that exhibit weak identification of some parameters, or some transformation of the parameters, and strong identification of other parameters, or other transformations of the parameters. Such scenarios occur when $p \geq 2$, but cannot occur when $p = 1$.

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

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.[6] Third, as shown in the SM, Kleibergen's CLR tests with moment-variance-weighting do not reduce to Moreira's CLR test in the homoskedastic normal linear IV model with fixed IV's when $p = 1$. In fact, with the moment-variance-weighting that has been considered in the literature, across different model configurations for which Moreira's conditioning statistic displays the same asymptotic behavior, the magnitude of the conditioning statistic for Kleibergen's CLR tests can be arbitrarily close to zero or infinity (with probability that goes to one). Simulation results given in the SM show that this leads to substantial power loss, in some scenarios of this model, relative to the SR-CQLR tests considered here, Moreira's CLR test, and Kleibergen's CLR test with Jacobian-variance weighting. Fourth, the form of Kleibergen's CLR test statistic for $p \geq 2$ is based on the form of Moreira's test statistic when $p = 1$. In consequence, one needs to make a somewhat arbitrary choice of some rank statistic to reduce the $k \times p$ weighted orthogonalized sample Jacobian to a scalar random variable.[7]

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 two SR-CQLR tests considered in this paper do not have any of these drawbacks.

To establish the asymptotic size and similarity results of the paper, we use the approach in Andrews, Cheng, and Guggenberger (2009) 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. 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.[8]

[6] It is shown in Section 12 in the Appendix to AG1 that this condition is not redundant. Without it, for some models, some sequences of distributions, and some (consistent) choices of variance and covariance estimators, Kleibergen's (2005) LM statistic has a χ^2_k asymptotic distribution, where k is the number of moment conditions. This leads to over-rejection of the null by this LM test when the standard χ^2_p critical value is used, where p is the dimension of the parameter, and the parameter is over-identified (i.e., $k > p$). Kleibergen's CLR tests depend on his LM test statistic, so his CLR tests also rely on the expected outer-product condition.

[7] Several rank statistics in the literature have been suggested, including Cragg and Donald (1996, 1997), Robin and Smith (2000), and Kleibergen and Paap (2006).

[8] 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.

To be more precise, we define these identification categories (when $k \geq p$) here. Let the k vector of moment functions be $g_i(\theta)$ and the $k \times p$ Jacobian matrix be $G_i(\theta) := (\partial/\partial\theta')g_i(\theta)$. The expected Jacobian at the true null value θ_0 is $E_F G_i(\theta_0)$, where F denotes the distribution that generates the observations. The variance matrix of $g_i(\theta_0)$ under F is denoted by $\Omega_F(\theta_0)$. Let $\{s_{jp} : j \leq p\}$ denote the singular values of $\Omega_F^{-1/2}(\theta_0)E_F G_i(\theta_0)$ in nonincreasing order (when $\Omega_F(\theta_0)$ is nonsingular).[9] For a sequence of distributions $\{F_n : n \geq 1\}$, 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 $\{F_n : n \geq 1\}$ 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. Note that 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.

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

To obtain the robustness of the three new tests to the 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 all 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 asymptotic size and similarity results given below do not rely on local or global identification.

[9] 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.

the variance matrix. We show that the resulting tests and CS's have correct asymptotic size. This method of robustifying tests and CS's to singularity of the population variance matrix also can be applied to other tests and CS's in the literature. Hence, it should be a useful addition to the literature with widespread applications. The robustness of the SR-CQLR tests to any form of the expected outer product matrix of the vectorized orthogonalized Jacobian occurs because the SR-CQLR test statistics do not depend on Kleibergen's LM statistic, but rather, on a minimum eigenvalue statistic.

We carry out some asymptotic power comparisons via simulation using eleven linear IV regression models with heteroskedasticity and/or autocorrelation and one right-hand side (rhs) endogenous variable ($p = 1$) and four IV's ($k = 4$). The scenarios considered are the same as in I. Andrews (2014). They are designed to mimic models for the elasticity of inter-temporal substitution estimated by Yogo (2004) for eleven countries using quarterly data from the early 1970's to the late 1990's. The results show that, in an overall sense, the SR-CQLR tests introduced here perform well in the scenarios considered. They have asymptotic power that is competitive with that of the PI-CLC test of I. Andrews (2014) and the MM2-SU test of Moreira and Moreira (2013), have somewhat better overall power than the JVW-CLR and MVW-CLR tests of Kleibergen (2005) and the MM1-SU test of Moreira and Moreira (2013), and have noticeably higher power than Kleibergen's (2005) LM test and the AR test. These results are reported in the SM.

Fast computation of tests is useful when constructing confidence sets by inverting the tests, especially when $p \geq 2$. The SR-CQLR$_2$ test (employed using 5000 critical value repetitions) can be computed $29,411$ times in one minute using a laptop with Intel i7-3667U CPU @2.0GHz in the $(k, p) = (4, 1)$ scenarios described above. The SR-CQLR$_2$ test 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 scenarios considered.[10] The SR-CQLR$_2$ test is found to be noticeably easier to implement than the PI-CLC, MM1-SU, and MM2-SU tests and comparable

[10] These computation times are for the data generating process corresponding to the country Australia, although the choice of country has very little effect on the times. Note that the computation times for the PI-CLC, MM1-SU, and MM2-SU tests depend greatly on the choice of implementation parameters. For the PI-CLC test, these include (i) the number of linear combination coefficients "a" considered in the search over $[0, 1]$, which we take to be 100, (ii) the number of simulation repetitions used to determine the best choice of "a," which we take to be 2000, and (iii) the number of alternative parameter values considered in the search for the best "a," which we take to be 41 for $p = 1$. For the MM1-SU and MM2-SU tests, the implementation parameters include (i) the number of variables in the discretization of the maximization problem, which we take to be 1000, and (ii) the number of points used in the numerical approximations of the integrals $h1$ and $h2$ that appear in the definitions of these tests, which we take to be 1000. The run-times for the PI-CLC, MM1-SU, and MM2-SU tests exclude some items, such as a critical value look up table for the PI-CLC test, that only need to be computed once when carrying out multiple tests. The computations are done in GAUSS using the lmpt application to do the linear programming required by the MM1-SU and MM2-SU tests. Note that the computation time for the SR-CQLR tests could be reduced by using a look up table for the data-dependent critical values, which depend on p singular values. This would be most useful when $p = 2$.

to the JVW-CLR and MVW-CLR tests, in terms of the choice of implementation parameters (see footnote 10) and the robustness of the results to these choices.

The computation time of the SR-CQLR$_2$ test increases relatively slowly with k and p. For example, the times (in minutes) to compute the SR-CQLR test 5000 times (using 5000 critical value repetitions) for $k = 8$ and $p = 1$, 2, 4, 8 are .26, .49, 1.02, 2.46. The times for $p = 1$ and $k = 1$, 2, 4, 8, 16 ,32, 64, 128 are .14, .15, .18, .26, .44, .99, 2.22, 7.76. The times for $(k, p) = (64, 8)$ and $(128, 8)$ are 14.5 and 57.9. Hence, computing tests for large values of (k, p) is quite feasible. These times are for linear IV regression models, but they are the same for any model, linear or nonlinear, when one takes as given the sample moment vector and sample Jacobian matrix.

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 $\theta = (\theta_1, ..., \theta_p)'$, 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 (2013) 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 $p \geq 2$.

Andrews and Guggenberger (2014c) provides SM to this paper. The SM to AG1 is given in Andrews and Guggenberger (2014b).

The paper is organized as follows. Section 2 discusses the related literature. Section 3 introduces the linear IV model and defines Moreira's (2003) CLR test for this model for the case of $p \geq 1$ rhs endogenous variables. Section 4 defines the general moment condition model. Section 5 introduces the SR-AR test. Sections 6 and 7 define the SR-CQLR$_1$ and SR-CQLR$_2$ tests, respectively. Section 8 provides the asymptotic size and similarity results for the tests. Section 9 establishes the asymptotic efficiency in a GMM sense of the SR-CQLR tests under strong and semi-strong identification. An Appendix provides parts of the proofs of the asymptotic size results given in Section 8.

The SM contains the following. Section 12 provides the time series results. Section 13 provides finite-sample null rejection probability simulation results for the SR-AR and SR-CQLR$_2$ tests for cases where the variance matrix of the moment functions is singular and near singular. Section 14 compares the test statistics and conditioning statistics of the SR-CQLR$_1$, SR-CQLR$_2$, and Kleibergen's (2005, 2007) CLR tests to those of Moreira's (2003) LR statistic and conditioning statistic in the homoskedastic linear IV model with fixed (i.e., nonrandom) IV's. Section 15 provides finite-sample simulation results that illustrate that Kleibergen's CLR test with moment-variance weighting can have low power in certain linear IV models with a single rhs endogenous variable, as the theoretical results in Section 14 suggest. Section 16 gives the asymptotic power comparisons based on the estimated models in Yogo (2004). Section 17 establishes some properties of an eigenvalue-adjustment procedure used in the definitions of the two SR-CQLR tests. Section 18 defines a new SR-LM test. The rest of the SM, in conjunction with the Appendix, provides the proofs of the results stated in AG2 and the SM.

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

In this section, we discuss the related literature and, in particular, existing asymptotic results in the literature. Kleibergen (2005) considers standard weak identification and strong identification.[11] This excludes all cases in the nonstandard weak and semi-strong identification categories.

The other papers in the literature that deal with LM and CLR tests for nonlinear moment condition models, including Guggenberger and Smith (2005), Otsu (2006), Smith (2007), Chaudhuri and Zivot (2011), Guggenberger, Ramalho, and Smith (2012), and I. Andrews (2014), rely on Stock and Wright's (2000) Assumption C. (An exception is a recent paper by I. Andrews and Mikusheva (2014a), which considers a different form of CLR test.) Stock and Wright's (2000) Assumption C is an innovative contribution to the literature, but it has some notable drawbacks. For a detailed discussion of Assumption C of Stock and Wright (2000), see Section 2 of AG1. Here we just provide a summary.

First, Assumption C is hard to verify or refute in nonlinear models. As far as we know it has only been verified in the literature for one nonlinear moment condition model, which is a polynomial approximation to the nonlinear CCAPM of interest in Stock and Wright (2000) and Kleibergen (2005). Second, Assumption C is restrictive.[12] It rules out some fairly simple nonlinear models, see AG1. Third, while it covers cases where some parameters are weakly identified and other are strongly identified, it does not cover cases where some transformations of the parameters are weakly identified and other transformations are strongly or semi-strongly identified.

The asymptotic results in this paper and AG1 do not require Assumption C or any related conditions of this type.

[11] The same is true of Andrews and Soares (2007), who consider rank-type CLR tests for linear IV models with multiple endogenous variables. Moreira (2003) considers only standard weak identification asymptotics in the latter model.

[12] The additive separability of the expected moment conditions, which is required by Assumption C, is the condition that leads to the first two drawbacks described here.

Mikusheva (2010) establishes the correct asymptotic size of LM and CLR tests in the linear IV model when there is one rhs endogenous variable ($p = 1$) and the errors are homoskedastic. Guggenberger (2012) establishes the correct asymptotic size of heteroskedasticity-robust LM and CLR tests in a heteroskedastic linear IV model with $p = 1$.

Compared to the standard GMM tests and CS's considered in Hansen (1982), the SR-CQLR and SR-AR tests considered here are robust to weak identification and singularity of the variance matrix of the moments. In particular, the tests considered here 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.[13] Under strong and semi-strong identification, the SR-CQLR procedures considered are asymptotically equivalent under contiguous local alternatives to the procedures considered in Hansen (1982) when the latter are based on asymptotically efficient weighting matrices.

A drawback of the SR-CQLR tests is that they do 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 (2013) provide methods for constructing finite-sample unbiased tests that maximize weighted average power in parametric models. They apply these methods to the heteroskedastic and autocorrelated normal linear IV regression model with $p = 1$. I. Andrews (2014) develops tests that minimize 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.[14] Although these tests are computationally tractable for minimum distance models, they are not for moment condition models. Hence, for moment condition models, I. Andrews proposes plug-in tests that aim to mimic the features of the infeasible optimal tests. (These feasible plug-in tests do not have optimality properties.) He discusses the heteroskedastic normal linear IV regression model with $p = 1$ in detail. Montiel Olea (2012) considers tests that have weighted average power optimality properties in a GMM sense under weak identification in moment condition models when $p = 1$.[15] Elliott, Müller, and Watson (2012) consider tests that maximize weighted average power in a variety of (finite-sample) parametric models where a nuisance parameter appears

[13] Conditions (i)-(iv) appear in Hansen's (1982) assumption (iii) of his Theorem 2.1, Assumption 3.4, assumption that S_w (the asymptotic variance matrix of the sample moments in Hansen's notation) is nonsingular (which is employed in his Theorem 3.2), and Assumption 3.2, respectively.

[14] For $p \geq 2$, the SR-CQLR tests are not in the class of tests considered in I. Andrews (2014).

[15] See Appendix G of Montiel Olea (2012). Whether these tests are asymptotically efficient under strong and semi-strong identification seems to be an open question. Montiel Olea (2012) also considers tests that maximize weighted average power among tests that depend on a score statistic and an identification statistic in the extremum estimator framework of Andrews and Cheng (2012). Only one source of weak identification arises in this framework.

under the null.

None of the previous papers provide asymptotic size results. Moreira and Moreira (2013) only consider finite-sample results. I. Andrews (2014) provides asymptotic results under Stock and Wright's (2000) Assumption C. Montiel Olea (2012) considers standard weak identification asymptotics. The asymptotic framework and results of this paper and AG1 should be useful for determining the asymptotic sizes of the tests considered in these papers. In particular, AG1 shows that the sample moments and the (suitably normalized) Jacobian-variance weighted conditioning statistic are not necessarily asymptotically independent when $p \geq 2$. This may have implications for the asymptotic size properties of moment condition tests that rely on estimation of the variance matrix of the (orthogonalized) sample Jacobian, such as the tests considered in Moreira and Moreira (2013) and I. Andrews (2014), when $p \geq 2$.[16]

A recent paper by I. Andrews and Mikusheva (2014a) considers an identification-robust inference method based on a conditional likelihood ratio approach that differs from those discussed above. The test considered in this paper is asymptotically similar conditional on the entire sample mean process that is orthogonalized to be asymptotically independent of the sample moments evaluated at the null parameter value.

The SR-CQLR and SR-AR tests considered in this paper are for full vector inference. To obtain subvector inference, one needs to employ the Bonferroni method or the Scheffé projection method, see Cavanagh, Elliott, and Stock (1995), Chaudhuri, Richardson, Robins, and Zivot (2010), Chaudhuri and Zivot (2011), and McCloskey (2011) for Bonferroni's method, and Dufour (1989) and Dufour and Jasiak (2001) for the projection method. Both methods are conservative, but Bonferroni's method is found to work quite well by Chaudhuri, Richardson, Robins, and Zivot (2010) and Chaudhuri and Zivot (2011).[17]

Other results in the literature on subvector inference include the following. Subvector inference in which nuisance parameters are profiled out is possible in the linear IV regression model with homoskedastic errors using the AR test, but not the LM or CLR tests, see Guggenberger, Kleibergen, Mavroeidis, and Chen (2012). Andrews and Cheng (2012, 2013a,b) provide subvector tests with correct asymptotic size based on extremum estimator objective functions. These subvector methods depend on the following: (i) one has knowledge of the source of the potential lack of identification (i.e., which subvectors play the roles of β, π, and ζ in their notation), (ii) there is only

[16] Moreira and Moreira (2013) do not explicitly consider tests in linear IV models when $p \geq 2$. However, their approach could be applied in such cases and would require estimation of (what amounts to) the variance matrix of the orthogonalized sample Jacobian when this matrix is unknown (which includes all practical cases of interest), see the appearance of Σ^{-1} in their conditioning statistic T.

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

one source of lack of identification, and (iii) the estimator objective function does not depend on the weakly identified parameters π (in their notation) when $\beta = 0$, which rules out some weak IV's models.[18] Cheng (2014) provides subvector inference in a nonlinear regression model with multiple nonlinear regressors and, hence, multiple potential sources of lack of identification. I. Andrews and Mikusheva (2012) develop subvector inference methods in a minimum distance context based on Anderson-Rubin-type statistics. I. Andrews and Mikusheva (2014b) provide conditions under which subvector inference is possible in exponential family models (but the requisite conditions seem to be quite restrictive).

Phillips (1989) and Choi and Phillips (1992) provide asymptotic and finite-sample results for estimators and classical tests in simultaneous equations models that may be unidentified or partially identified when $p \geq 1$. However, their results do not cover weak identification (of standard or nonstandard form) or identification-robust inference. Hillier (2009) provides exact finite-sample results for CLR tests in the linear model under the assumption of homoskedastic normal errors and known covariance matrix. Antoine and Renault (2009, 2010) consider GMM estimation under semi-strong and strong identification, but do not consider tests or CS's that are robust to weak identification. Armstrong, Hong, and Nekipelov (2012) show that standard Wald tests for multiple restrictions in some nonlinear IV models can exhibit size distortions when some IV's are strongly identified and others are semi-strongly identified—not weakly identified. These results indicate that identification issues can be more severe in nonlinear models than in linear models, which provides further motivation for the development of identification-robust tests for nonlinear models.

3 Linear IV Model with p $\geq$ 1 Endogenous Variables

In this section, we define the CLR test of Moreira (2003) in the homoskedastic Gaussian linear (HGL) IV model with $p \geq 1$ endogenous regressor variables and $k \geq p$ fixed (i.e., nonrandom) IV's. The SR-CQLR$_1$ test introduced below is designed to reduce to Moreira's CLR test in this model asymptotically. The SR-CQLR$_2$ test introduced below reduces to Moreira's CLR test in this model asymptotically when $p = 1$ and in some, but not all, cases when $p \geq 2$ (depending on the behavior of the reduced-form parameters).

[18] Montiel Olea (2012) also provides some subvector analysis in the extremum estimator context of Andrews and Cheng (2012). His efficient conditionally similar tests apply to the subvector (π, ζ) of (β, π, ζ) (in Andrews and Cheng's (2012) notation), where β is a parameter that determines the strength of identification and is known to be strongly identified. The scope of this subvector analysis is analogous to that of Stock and Wright (2000) and Kleibergen (2004).

The linear IV regression model is

$$\begin{aligned} y_{1i} &= Y_{2i}'\theta + u_i \text{ and} \\ Y_{2i} &= \pi' Z_i + V_{2i}, \end{aligned} \tag{3.1}$$

where $y_{1i} \in R$ and $Y_{2i} \in R^p$ are endogenous variables, $Z_i \in R^k$ for $k \geq p$ is a vector of fixed IV's, and $\pi \in R^{k\times p}$ is an unknown unrestricted parameter matrix. In terms of its reduced-form equations, the model is

$$y_{1i} = Z_i'\pi\theta + V_{1i},\ Y_{2i} = \pi' Z_i + V_{2i},\ V_i := (V_{1i}, V_{2i}')',\ V_{1i} = u_i + V_{2i}'\theta, \text{ and } \Sigma_V := EV_iV_i'. \tag{3.2}$$

For simplicity, no exogenous variables are included in the structural equation. The reduced-form errors are $V_i \in R^{p+1}$. In the HGL model, $V_i \sim N(0^{p+1}, \Sigma_V)$ for some positive definite $(p+1)\times(p+1)$ matrix Σ_V.

The IV moment functions and their derivatives with respect to θ are

$$g(W_i, \theta) = Z_i(y_{1i} - Y_{2i}'\theta) \text{ and } G(W_i, \theta) = -Z_i Y_{2i}',\ \text{where } W_i := (y_{1i}, Y_{2i}', Z_i')'. \tag{3.3}$$

Moreira (2003, p. 1033) shows that the LR statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in the HGL model in (3.1)-(3.2) when Σ_V is known is

$$\begin{aligned} LR_{HGL,n} &:= \overline{S}_n'\overline{S}_n - \lambda_{\min}((\overline{S}_n, \overline{T}_n)'(\overline{S}_n, \overline{T}_n)),\ \text{where} \\ \overline{S}_n &:= (Z_{n\times k}'Z_{n\times k})^{-1/2}Z_{n\times k}'Yb_0(b_0'\Sigma_V b_0)^{-1/2} = (n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}\widehat{g}_n(b_0'\Sigma_V b_0)^{-1/2} \in R^k, \\ \overline{T}_n &:= (Z_{n\times k}'Z_{n\times k})^{-1/2}Z_{n\times k}'Y\Sigma_V^{-1}A_0(A_0'\Sigma_V^{-1}A_0)^{-1/2} \\ &= -(n^{-1}Z_{n\times k}'Z_{n\times k})^{-1/2}n^{1/2}(\widehat{G}_n\theta_0 - \widehat{g}_n, \widehat{G}_n)\Sigma_V^{-1}A_0(A_0'\Sigma_V^{-1}A_0)^{-1/2} \in R^{k\times p}, \\ Z_{n\times k} &:= (Z_1, ..., Z_n)' \in R^{n\times k},\ Y := (Y_1, ..., Y_n)' \in R^{n\times(p+1)},\ Y_i := (y_{1i}, Y_{2i}')' \in R^{p+1}, \\ b_0 &:= (1, -\theta_0')' \in R^{p+1},\ \widehat{g}_n := n^{-1}\sum_{i=1}^n g(W_i, \theta_0),\ A_0 := (\theta_0, I_p)' \in R^{(p+1)\times p}, \\ \widehat{G}_n &:= n^{-1}\sum_{i=1}^n G(W_i, \theta_0), \end{aligned} \tag{3.4}$$

$\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix, and the second equality for $\overline{T}_n$ holds by (24.12) in the SM.[19] Note that $(\overline{S}_n, \overline{T}_n)$ is a (conveniently transformed) sufficient statistic for (θ, π) under

[19] We let $Z_{n\times k}$ (rather than Z) denote $(Z_1, ..., Z_n)'$, because we use Z to denote a k vector of standard normals

normality of V_i, known variance matrix Σ_V, and fixed IV's.

Moreira's (2003) CLR test uses the $LR_{HGL,n}$ statistic and a conditional critical value that depends on the $k \times p$ matrix $\overline{T}_n$ through a conditional critical value function $c_{k,p}(D, 1-\alpha)$, which is defined as follows. For nonrandom $D \in R^{k \times p}$, let

$$CLR_{k,p}(D) := Z'Z - \lambda_{\min}((Z,D)'(Z,D)), \text{ where } Z \sim N(0^k, I_k). \tag{3.5}$$

Define $c_{k,p}(D, 1-\alpha)$ to be the $1-\alpha$ quantile of the distribution of $CLR_{k,p}(D)$. For $\alpha \in (0,1)$, Moreira's CLR test with nominal level α rejects H_0 if

$$LR_{HGL,n} > c_{k,p}(\overline{T}_n, 1-\alpha). \tag{3.6}$$

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

Moreira's (2003) CLR test is similar with finite-sample size α in the HGL model with known Σ_V. Intuitively, the strength of the IV's affects the null distribution of the test statistic $LR_{HGL,n}$ and the critical value $c_{k,p}(\overline{T}_n, 1-\alpha)$ adjusts accordingly to yield a test with size α using the dependence of the null distribution of $\overline{T}_n$ on the strength of the IV's. When $p = 1$, this test has been shown to have some (approximate) asymptotic optimality properties, see Andrews, Moreira, and Stock (2006, 2008) and Chernozhukov, Hansen, and Jansson (2009).

For $p \geq 2$, the asymptotic properties of Moreira's CLR test, such as its asymptotic size and similarity, are not available in the literature. The results for the SR-CQLR$_1$ test, specialized to the linear IV model (with or without Gaussianity, homoskedasticity, and/or independence of the errors), fill this gap.

4 Moment Condition Model

4.1 Moment Functions

The general moment condition model that we consider is

$$E_F g(W_i, \theta) = 0^k, \tag{4.1}$$

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

below.

we allow for $k \geq p$ and $k < p$. In Section 12 in the SM, we consider models with stationary strong mixing observations. The parameter space for θ is $\Theta \subset R^p$.

The Jacobian of the moment functions is

$$G(W_i, \theta) := \frac{\partial}{\partial \theta'} g(W_i, \theta) \in R^{k \times p}.^{20} \tag{4.2}$$

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 $j = 1, ..., p$. 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

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \tag{4.3}$$

The SR-CQLR$_1$ test that we introduce in Section 6 below applies when $g_i(\theta)$ has the form

$$g_i(\theta) = u_i(\theta) Z_i, \tag{4.4}$$

where Z_i is a k vector of IV's, $u_i(\theta)$ is a scalar residual, and the (random) function $u_i(\cdot)$ is known. This is the case considered in Stock and Wright (2000). It covers many GMM situations, but can be restrictive. For example, it rules out Hansen and Scheinkman's (1995) moment conditions for continuous-time Markov processes, the moment conditions often used with dynamic panel models, e.g., see Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1995), and moment conditions of the form $g_i(\theta) = u_i(\theta) \otimes Z_i$, where $u_i(\theta)$ is a vector. For the cases ruled out, we introduce a second SR-CQLR test in Section 7 that does not rely on (4.4). The SR-AR test defined in Section 5 also does not require that $g_i(\theta)$ satisfies (4.4).

When (4.4) holds, we define

$$u_{\theta i}(\theta) := \frac{\partial}{\partial \theta} u_i(\theta) \in R^p \text{ and } u_i^*(\theta) := \begin{pmatrix} u_i(\theta) \\ u_{\theta i}(\theta) \end{pmatrix} \in R^{p+1}, \text{ and we have } G_i(\theta) = Z_i u_{\theta i}(\theta)'.^{21} \tag{4.5}$$

[20] 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 $\mathcal{F}_2^{SR}$, defined in (4.9) below. For example, $G(W_i, \theta)$ can be the derivative of $g(W_i, \theta)$ almost surely, rather than for all W_i, which allows $g(W_i, \theta)$ to have kinks. The function $G(W_i, \theta)$ also can be a numerical derivative, such as $((g(W_i, \theta + \varepsilon e_1) - g(W_i, \theta))/\varepsilon, ..., (g(W_i, \theta + \varepsilon e_p) - g(W_i, \theta))/\varepsilon) \in R^{k \times p}$ for some $\varepsilon > 0$, where e_j is the jth unit vector, e.g., $e_1 = (1, 0, ..., 0)' \in R^p$.

[21] As with $G(W_i, \theta)$ defined in (4.2), $u_{\theta i}(\theta)$ need not be a vector of partial derivatives of $u_i(\theta)$ for all sample realizations of the observations. It could be the vector of partial derivatives of $u_i(\theta)$ almost surely, rather than for all W_i, which allows $u_i(\theta)$ to have kinks, or a vector of finite differences of $u_i(\theta)$. For the asymptotic size results for the

4.2 Parameter Spaces of Distributions F

The variance matrix of the moments, $\Omega_F(\theta)$, is defined by

$$\Omega_F(\theta) := E_F(g_i(\theta) - E_F g_i(\theta))(g_i(\theta) - E_F g_i(\theta))'. \tag{4.6}$$

(Under H_0, $\Omega_F(\theta_0) = E_F g_i(\theta_0) g_i(\theta_0)'$.) We allow for the case where $\Omega_F(\theta)$ is singular. The rank and spectral decomposition of $\Omega_F(\theta)$ are denoted by

$$r_F(\theta) := rk(\Omega_F(\theta)) \text{ and } \Omega_F(\theta) := A_F^\dagger(\theta)\Pi_F(\theta)A_F^\dagger(\theta)', \tag{4.7}$$

where $rk(\cdot)$ denotes the rank of a matrix, $\Pi_F(\theta)$ is the $k \times k$ diagonal matrix with the eigenvalues of $\Omega_F(\theta)$ on the diagonal in nonincreasing order, and $A_F^\dagger(\theta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\Pi_F(\theta)$. We partition $A_F^\dagger(\theta)$ according to whether the corresponding eigenvalues are positive or zero:

$$A_F^\dagger(\theta) = [A_F(\theta), A_F^\perp(\theta)], \text{ where } A_F(\theta) \in R^{k \times r_F(\theta)} \text{ and } A_F^\perp(\theta) \in R^{k \times (k - r_F(\theta))}. \tag{4.8}$$

By definition, 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 non-redundant linear combinations of the moment functions evaluated at θ_0 rescaled to have variances equal to one: $Var_F(\Pi_{1F}^{-1/2}A_F' g_i) = \Pi_{1F}^{-1/2}A_F'\Omega_F A_F \Pi_{1F}^{-1/2} = I_{r_F}$. The $r_F \times p$ matrix $\Pi_{1F}^{-1/2}A_F' G_i$ is the analogously transformed Jacobian matrix.

We consider the following parameter spaces for the distribution F that generates the data under $H_0 : \theta = \theta_0$:

$$\begin{aligned}
\mathcal{F}_{AR}^{SR} &:= \{F : E_F g_i = 0^k \text{ and } E_F||\Pi_{1F}^{-1/2}A_F' g_i||^{2+\gamma} \leq M\}, \\
\mathcal{F}_{2}^{SR} &:= \{F \in \mathcal{F}_{AR}^{SR} : E_F||vec(\Pi_{1F}^{-1/2}A_F' G_i)||^{2+\gamma} \leq M\}, \text{ and} \\
\mathcal{F}_{1}^{SR} &:= \{F \in \mathcal{F}_{2}^{SR} : E_F||\Pi_{1F}^{-1/2}A_F' Z_i||^{4+\gamma} \leq M, \ E_F||u_i^*||^{2+\gamma} \leq M, \text{ and} \\
&\qquad E_F||\Pi_{1F}^{-1/2}A_F' Z_i||^2 u_i^2 1(u_i^2 > c) \leq 1/2\}
\end{aligned} \tag{4.9}$$

SR-CQLR$_1$ test given below to hold, $u_{\theta i}(\theta)$ can be any random p vector that satisfies the conditions in $\mathcal{F}_1^{SR}$ (defined in (4.9)).

for some $\gamma > 0$ and some $M, c < \infty$, where $||\cdot||$ denotes the Euclidean norm, and $vec(\cdot)$ denotes the vector obtained from stacking the columns of a matrix. By definition, $\mathcal{F}_1^{SR} \subset \mathcal{F}_2^{SR} \subset \mathcal{F}_{AR}^{SR}$.[22,23]

The null parameter spaces $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ are used for the SR-AR, SR-CQLR$_2$, and SR-CQLR$_1$ tests, respectively. The first condition in $\mathcal{F}_{AR}^{SR}$ is the defining condition of the model. The second condition in $\mathcal{F}_{AR}^{SR}$ is a mild moment condition on the rescaled non-redundant moment functions $\Pi_{1F}^{-1/2} A_F' g_i$. The condition in $\mathcal{F}_2^{SR}$ is a mild moment condition on the analogously transformed derivatives of the moment conditions $\Pi_{1F}^{-1/2} A_F' G_i$. The conditions in $\mathcal{F}_1^{SR}$ are only marginally stronger than those in $\mathcal{F}_2^{SR}$. A sufficient condition for the last condition in $\mathcal{F}_1^{SR}$ to hold for some $c < \infty$ is $E_F u_i^4 \leq M_*$ for some sufficiently large $M_* < \infty$ (using the first condition in $\mathcal{F}_1^{SR}$ and the Cauchy-Bunyakovsky-Schwarz inequality).

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 $k \geq p$ and one or more of its singular values is zero or close to zero. The conditions in $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ place no restrictions on the column rank or singular values of $E_F G_i$.

The conditions in $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ also place no restrictions on the variance matrix $\Omega_F := E_F g_i g_i'$ of g_i, such as $\lambda_{\min}(\Omega_F) \geq \delta$ for some $\delta > 0$ or $\lambda_{\min}(\Omega_F) > 0$. Hence, Ω_F can be singular. 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.) For example, this occurs in all likelihood scenarios, in which case $g_i(\theta)$ is the score function. In such scenarios, the information matrix equality implies that minus the expected Jacobian matrix $E_F G_i$ equals the information matrix, which also equals the expected outer product of the score function Ω_F, i.e., $-E_F G_i = \Omega_F$. In this case, weak identification occurs when Ω_F is close to being singular. Furthermore, identification failure yields singularity of Ω_F in all quasi-likelihood scenarios when the quasi-likelihood does not depend on some element(s) of θ (or some transformation(s) of θ) for θ in a neighborhood of θ_0.[24]

A second example where Ω_F may be singular is the following homoskedastic linear IV model: $y_{1i} = Y_{2i}\beta + U_i$ and $Y_{2i} = Z_i'\pi + V_{1i}$, where all quantities are scalars except $Z_i, \pi \in R^{dz}$, $\theta = (\beta, \pi')' \in R^{\beta + dz}$, $EU_i = EV_{2i} = 0$, $EU_i Z_i = EV_{1i} Z_i = 0^{dz}$, and $E(V_i V_i' | Z_i) = \Sigma_V$ a.s. for

[22] In the results below, we assume that whichever parameter space is being considered is non-empty.

[23] The moment bounds in $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ can be weakened very slightly by, e.g., replacing $E_F||\Pi_{1F}^{-1/2} A_F' g_i||^{2+\gamma} \leq M$ in $\mathcal{F}_{AR}^{SR}$ by $E_F||\Pi_{1F}^{-1/2} A_F' g_i||^2 1(||\Pi_{1F}^{-1/2} A_F' g_i|| > j) \leq \varepsilon_j$ for all integers $j \geq 1$ for some $\varepsilon_j > 0$ (that does not depend on F) for which $\varepsilon_j \to 0$ as $j \to \infty$. The latter conditions are weaker because, for any random variable X and constants $\gamma, j > 0$, $EX^2 1(|X| > j) \leq E|X|^{2+\gamma}/j^\gamma$. The latter conditions allow for the application of Lindeberg's triangular array central limit theorem for independent random variables, e.g., see Billingsley (1979, Thm. 27.2, p. 310), in scenarios where the distribution F depends on n. For simplicity, we define the parameter spaces as is.

[24] In this case, the moment functions equal the quasi-score and some element(s) or linear combination(s) of elements of moment functions, equal zero a.s. at θ_0 (because the quasi-score is of the form $g_i(\theta) = (\partial/\partial\theta) \log f(W_i, \theta)$ for some density or conditional density $f(W_i, \theta)$). This yields singularity of the variance matrix of the moment functions and of the expected Jacobian of the moment functions.

$V_i := (V_{1i}, V_{2i})'$ and some 2×2 constant matrix Σ_V. The corresponding reduced-form equations are $y_{1i} = Z_i'\pi\beta + V_{1i}$ and $Y_{2i} = Z_i'\pi + V_{1i}$, where $V_{1i} = U_i + V_{2i}\beta$. The moment conditions for θ are $g_i(\theta) = ((y_{1i} - Z_i'\pi\beta)Z_i', (Y_{2i} - Z_i'\pi)Z_i')' \in R^k$, where $k = 2d_Z$. The variance matrix $\Sigma_V \otimes EZ_iZ_i'$ of $g_i(\theta_0) = (V_{1i}Z_i', V_{2i}Z_i')'$ is singular whenever the covariance between the reduced-form errors V_{1i} and V_{2i} is one (or minus one) or EZ_iZ_i' is singular. In this model, we are interested in joint inference concerning β and π. This is of interest when one wants to see how the magnitude of π affects the range of plausible β values.

A third case where Ω_F can be singular is in the model for interest rate dynamics discussed in Jegannathan, Skoulakis, and Wang (2002, Sec. 6.2) (JSW). JSW consider five moment conditions for a four dimensional parameter θ. Grant (2013) points out 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.[25]

In examples one and three above and others like them, E_FG_i is close to having less than full column rank (i.e., its smallest singular value is small) and Ω_F is close to being singular (i.e., $\lambda_{\min}(\Omega_F)$ is small) when the null value θ_0 is close to a value which yields reduced column rank of E_FG_i and singularity of Ω_F. Null hypotheses of this type are important for the properties of CS's because uniformity over null hypothesis values is necessary for CS's to have correct asymptotic size. Hence, it is important to have procedures available that place no restrictions on either E_FG_i or Ω_F.

In contrast, to obtain the correct asymptotic size of Kleibergen's (2005) LM and moment-variance-weighted CLR tests (and his Jacobian-weighted CLR test when $p = 1$), AG1 imposes the condition $\lambda_{\min}(\Omega_F) > 0$ on all null distributions F, because these tests rely on the inverse of the sample variance matrix $\widehat{\Omega}_n$ being well-defined and well-behaved. AG1 also imposes a second condition that does not appear in the parameter spaces $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$.[26] This second condition can be restrictive and, in some models, difficult to verify. This condition arises because Kleibergen's LM statistic projects onto a p dimensional column space of a weighted version of the $k \times p$ orthogonalized sample Jacobian. To obtain the desired χ_p^2 asymptotic null distribution of this statistic via the continuous mapping theorem, one needs the orthogonalized sample Jacobian to be full column rank p a.s. asymptotically (after suitable renormalization). To obtain this under weak identification, AG1 imposes the condition referred to above.[27] It is shown in Section 12 in

[25] The first four moment functions in JSW are $(a(b-r_i)r_i^{-2\gamma} - \gamma\sigma^2 r_i^{-1}$, $a(b-r_i)r_i^{-2\gamma+1} - (\gamma - 1/2)\sigma^2$, $(b-r_i)r_i^{-a} - (1/2)\sigma^2 r_i^{2\gamma-a-1}$, $a(b-r_i)r_i^{-\sigma} - (1/2)\sigma^3 r_i^{2\gamma-\sigma-1})'$, where $\theta = (a, b, \sigma, \gamma)'$ and r_i is the interest rate. The second and third functions are equivalent if $\gamma = (a+1)/2$; the second and fourth functions are equivalent if $\gamma = (\sigma+1)/2$; and the third and fourth functions are equivalent if $\sigma = a$.

[26] See the definition of $\mathcal{F}_0$ in Section 3 of AG1.

[27] This condition is used in the proof of Lemma 8.3(d) in the Appendix of AG1, which is given in Section 15 in the SM to AG1.

the Appendix to AG1 that this condition is not redundant.

Given the discussion of the previous paragraph, it is clear that the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests introduced below have advantages over Kleibergen's LM and CLR tests in terms of the robustness of their correct asymptotic size properties.

Next, we specify the parameter spaces for (F, θ) that are used with the SR-AR, SR-CQLR$_2$, and SR-CQLR$_1$ CS's. They are denoted by $\mathcal{F}_{\Theta,AR}^{SR}$, $\mathcal{F}_{\Theta,2}^{SR}$, and $\mathcal{F}_{\Theta,1}^{SR}$, respectively. For notational simplicity, the dependence of the parameter spaces $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ in (4.9) on θ_0 is suppressed. When dealing with CS's, rather than tests, we make the dependence explicit and write them as $\mathcal{F}_{AR}^{SR}(\theta_0)$, $\mathcal{F}_2^{SR}(\theta_0)$, and $\mathcal{F}_1^{SR}(\theta_0)$, respectively. We define

$$\begin{aligned}
\mathcal{F}_{\Theta,AR}^{SR} &:= \{(F, \theta_0) : F \in \mathcal{F}_{AR}^{SR}(\theta_0), \theta_0 \in \Theta\}, \\
\mathcal{F}_{\Theta,2}^{SR} &:= \{(F, \theta_0) : F \in \mathcal{F}_2^{SR}(\theta_0), \theta_0 \in \Theta\}, \text{ and} \\
\mathcal{F}_{\Theta,1}^{SR} &:= \{(F, \theta_0) : F \in \mathcal{F}_1^{SR}(\theta_0), \theta_0 \in \Theta\}.
\end{aligned} \tag{4.10}$$

4.3 Definitions of Asymptotic Size and Similarity

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

$$AsySz := \limsup_{n\to\infty} \sup_{F\in\overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{4.11}$$

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

$$\liminf_{n\to\infty} \inf_{F\in\overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha) = \limsup_{n\to\infty} \sup_{F\in\overline{\mathcal{F}}(\theta_0)} RP_n(\theta_0, F, \alpha). \tag{4.12}$$

Below we establish the correct asymptotic size (i.e., asymptotic size equals nominal size) and the asymptotic similarity of the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests for the parameters spaces $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_1^{SR}$, and $\mathcal{F}_2^{SR}$, respectively.

Now we consider a CS that is obtained by inverting tests of $H_0 : \theta = \theta_0$ for all $\theta_0 \in \Theta$. The *asymptotic size* of the CS for the parameter space $\overline{\mathcal{F}}_\Theta := \{(F, \theta_0) : F \in \overline{\mathcal{F}}(\theta_0), \theta_0 \in \Theta\}$ is $AsySz := \liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1 - RP_n(\theta_0, F, \alpha))$. The CS is *asymptotically similar* (in a uniform sense) for the parameter space $\overline{\mathcal{F}}_\Theta$ if $\liminf_{n\to\infty} \inf_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1-RP_n(\theta_0, F, \alpha)) = \limsup_{n\to\infty} \sup_{(F,\theta_0)\in\overline{\mathcal{F}}_\Theta}(1-RP_n(\theta_0, F, \alpha))$. As defined, asymptotic size and similarity of a CS require uniformity over the null values $\theta_0 \in \Theta$, as well as uniformity over null distributions F for each null value θ_0. With the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ CS's considered here, this additional level of uniformity does not cause complications. The same proofs for tests deliver results for CS's with very minor adjustments.

5 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 has correct asymptotic size without any conditions on the variance matrix of the moment functions. The SR-AR test generalizes the S test of Stock and Wright (2000).

When the model is just identified (i.e., the dimension p of θ equals the dimension k of $g_i(\theta)$), the SR-AR test has good power properties. For example, this occurs in likelihood scenarios, in which case the vector of moment functions consists of the score function. However, when the model is over-identified (i.e., $k > p$), the SR-AR test generally sacrifices power because it is a k degrees of freedom test concerning p ($< k$) parameters. Hence, its power is often less than that of the SR-CQLR$_1$ and SR-CQLR$_2$ tests introduced below.

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

$$\widehat{g}_n(\theta) := n^{-1}\sum_{i=1}^{n} g_i(\theta) \text{ and } \widehat{\Omega}_n(\theta) := n^{-1}\sum_{i=1}^{n} g_i(\theta)g_i(\theta)' - \widehat{g}_n(\theta)\widehat{g}_n(\theta)'. \tag{5.1}$$

The usual nonlinear AR statistic is

$$AR_n(\theta) := n\widehat{g}_n(\theta)'\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta). \tag{5.2}$$

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

Now, we introduce a singularity-robust nonlinear AR statistic which applies even if $\Omega_F(\theta)$ is singular. First, we introduce sample versions of the population quantities $r_F(\theta)$, $A_F^\dagger(\theta)$, $A_F(\theta)$, $A_F^\perp(\theta)$, and $\Pi_F(\theta)$, which are defined in (4.7) and (4.8). The rank and spectral decomposition of $\widehat{\Omega}_n(\theta)$ are denoted by

$$\widehat{r}_n(\theta) := rk(\widehat{\Omega}_n(\theta)) \text{ and } \widehat{\Omega}_n(\theta) := \widehat{A}_n^\dagger(\theta)\widehat{\Pi}_n(\theta)\widehat{A}_n^\dagger(\theta)', \tag{5.3}$$

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 $\widehat{A}_n^{\dagger}(\theta)$ is a $k \times k$ orthogonal matrix of eigenvectors corresponding to the eigenvalues in $\widehat{\Pi}_n(\theta)$. We partition $\widehat{A}_n^{\dagger}(\theta)$ according to whether the corresponding eigenvalues are positive or zero:

$$\widehat{A}_n^{\dagger}(\theta) = [\widehat{A}_n(\theta), \widehat{A}_n^{\perp}(\theta)], \text{ where } \widehat{A}_n(\theta) \in R^{k \times \widehat{r}_n(\theta)} \text{ and } \widehat{A}_n^{\perp}(\theta) \in R^{k \times (k - \widehat{r}_n(\theta))}. \tag{5.4}$$

By definition, 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 CS's 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 6.2 below).

Define $\widehat{g}_{An}(\theta)$ and $\widehat{\Omega}_{An}(\theta)$ as $\widehat{g}_n(\theta)$ and $\widehat{\Omega}_n(\theta)$ are defined in (5.1), but with $\widehat{A}_n(\theta)'g_i(\theta)$ in place of $g_i(\theta)$. That is,

$$\widehat{g}_{An}(\theta) := \widehat{A}_n(\theta)'\widehat{g}_n(\theta) \in R^{\widehat{r}_n(\theta)} \text{ and } \widehat{\Omega}_{An}(\theta) := \widehat{A}_n(\theta)'\widehat{\Omega}_n(\theta)\widehat{A}_n(\theta) \in R^{\widehat{r}_n(\theta) \times \widehat{r}_n(\theta)}. \tag{5.5}$$

The SR-AR test statistic is defined by

$$SR\text{-}AR_n(\theta) := n\widehat{g}_{An}(\theta)'\widehat{\Omega}_{An}^{-1}(\theta)\widehat{g}_{An}(\theta). \tag{5.6}$$

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

$$SR\text{-}AR_n(\theta_0) > \chi^2_{\widehat{r}_n(\theta_0),1-\alpha} \text{ or } \widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}, \tag{5.7}$$

where by definition the latter condition does not hold if $\widehat{r}_n(\theta_0) = k$. For completeness of the specification of the SR-AR test, if $\widehat{r}_n(\theta_0) = 0$, then we define $SR\text{-}AR_n(\theta_0) := 0$ and $\chi^2_{\widehat{r}_n(\theta_0),1-\alpha} := 0$. Thus, when $\widehat{r}_n(\theta_0) = 0$, we have $\widehat{A}_n^{\perp}(\theta_0) = I_k$ and the SR-AR test rejects H_0 if $\widehat{g}_n(\theta_0) \neq 0^k$.

The extra rejection condition, $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}$, improves power, but we show it has no effect under H_0 with probability that goes to one (wp→1). 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 the following. First, suppose $(X_{1i}, X_{2i})' \sim$ i.i.d. $N(\theta, \Omega_F)$, where $\theta = (\theta_1, \theta_2)' \in R^2$, Ω_F is a 2×2 matrix of ones, and the moment functions are $g_i(\theta) = (X_{1i} - \theta_1, X_{2i} - \theta_2)'$. In this case, Ω_F is singular, $\widehat{A}_n(\theta_0) =$

$(1,1)'$ a.s., $\widehat{A}_n^{\perp}(\theta_0) = (1,-1)'$ a.s., the SR-AR statistic is a quadratic form in $\widehat{A}_n(\theta_0)'\widehat{g}_n(\theta_0) = \overline{X}_{1n} + \overline{X}_{2n} - (\theta_{10} + \theta_{20})$, where $\overline{X}_{mn} = n^{-1}\sum_{i=1}^n X_{mi}$ for $m = 1,2$, and $A_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = \overline{X}_{1n} - \overline{X}_{2n} - (\theta_{10} - \theta_{20})$ a.s. If one does not use the extra rejection condition, then the SR-AR test has no power against alternatives $\theta = (\theta_1, \theta_2)'$ ($\neq \theta_0$) for which $\theta_1 + \theta_2 = \theta_{10} + \theta_{20}$. However, when the extra rejection condition is utilized, all $\theta \in R^2$ except those on the line $\theta_1 - \theta_2 = \theta_{10} - \theta_{20}$ are rejected with probability one (because $\overline{X}_{1n} - \overline{X}_{2n} = E_F X_{1i} - E_F X_{2i} = \theta_1 - \theta_2$ a.s.) and this includes all of the alternative θ values for which $\theta_1 + \theta_2 = \theta_{10} + \theta_{20}$.

Second, suppose $X_i \sim$ i.i.d. $N(\theta_1, \theta_2)$, $\theta = (\theta_1, \theta_2)' \in R^2$, the moment functions are $g_i(\theta) = (X_i - \theta_1, X_i^2 - \theta_1^2 - \theta_2)'$, and the null hypothesis is $H_0 : \theta = (\theta_{10}, \theta_{20})'$. Consider alternative parameters of the form $\theta = (\theta_1, 0)'$. Under θ, X_i has variance zero, $X_i = \overline{X}_n = \theta_1$ a.s., $X_i^2 = \overline{X_n^2} = \theta_1^2$ a.s., where $\overline{X_n^2} := n^{-1}\sum_{i=1}^n X_i^2$, $\widehat{g}_n(\theta_0) = (\theta_1 - \theta_{10}, \theta_1^2 - \theta_{10}^2 - \theta_{20})'$ a.s., $\widehat{\Omega}_n(\theta_0) = \widehat{g}_n(\theta_0)\widehat{g}_n(\theta_0)' - \widehat{g}_n(\theta_0)\widehat{g}_n(\theta_0)' = 0^{2\times 2}$ a.s. (provided $\widehat{\Omega}_n(\theta_0)$ is defined as in (5.1) with the sample means subtracted off), and $\widehat{r}_n(\theta_0) = 0$ a.s. In consequence, if one does not use the extra rejection condition, then the SR-AR test has no power against alternatives of the form $\theta = (\theta_1, 0)'$ (because by definition the SR-AR test statistic and its critical value equal zero when $\widehat{r}_n(\theta_0) = 0$). However, when the extra rejection condition is utilized, all alternatives of the form $\theta = (\theta_1, 0)'$ are rejected with probability one.[28,29,30,31]

[28] This holds because the extra rejection condition in this case leads one to reject H_0 if $\overline{X}_n \neq \theta_{10}$ or $\overline{X_n^2} - \theta_{10}^2 - \theta_{20} \neq 0$, which is equivalent a.s. to rejecting if $\theta_1 \neq \theta_{10}$ or $\theta_1^2 - \theta_{10}^2 - \theta_{20} \neq 0$ (because $\overline{X}_n = \theta_1$ a.s. and $\overline{X_n^2} = \theta_1^2$ a.s. under θ), which in turn is equivalent to rejecting if $\theta \neq \theta_0$ (because if $\theta_{20} > 0$ one or both of the two conditions is violated when $\theta \neq \theta_0$ and if $\theta_{20} = 0$, then $\theta \neq \theta_0$ only if $\theta_1 \neq \theta_{10}$ since we are considering the case where $\theta_2 = 0$).

[29] In this second example, suppose the null hypothesis is $H_0 : \theta = (\theta_{10}, 0)'$. That is, $\theta_{20} = 0$. Then, the SR-AR test rejects with probability zero under H_0 and the test is not asymptotically similar. This holds because $\widehat{g}_n(\theta_0) = (\overline{X}_n - \theta_{10}, \overline{X_n^2} - \theta_{10}^2)' = (0,0)'$ a.s., $\widehat{r}_n(\theta_0) = 0$ a.s., $SR\text{-}AR_n(\theta_0) = \chi^2_{\widehat{r}_n(\theta_0),1-\alpha} = 0$ a.s. (because $\widehat{r}_n(\theta_0) = 0$ a.s.), and the extra rejection condition leads one to reject H_0 if $\overline{X}_n \neq \theta_{10}$ or $\overline{X_n^2} - \theta_{10}^2 - \theta_{20} \neq 0$, which is equivalent to $\theta_{10} \neq \theta_{10}$ or $\theta_{10}^2 - \theta_{10}^2 - \theta_{20} \neq 0$ (because $X_i = \theta_1$ a.s.), which holds with probability zero.

As shown in Theorem 8.1 below, the SR-AR test is asymptotically similar (in a uniform sense) if one excludes null distributions F for which the $g_i(\theta_0) = 0^k$ a.s. under F, such as in the present example, from the parameter space of null distributions. But, the SR-AR test still has correct asymptotic size without such exclusions.

[30] We thank Kirill Evdokimov for bringing these two examples to our attention.

[31] An alternative definition of the SR-AR test is obtained by altering its definition given here as follows. One omits the extra rejection condition given in (5.7), one defines the SR-AR statistic using a weight matrix that is nonsingular by construction when $\widehat{\Omega}_n(\theta_0)$ is singular, and one determines the critical value by simulation of the appropriate quadratic form in mean zero normal variates when $\widehat{\Omega}_n(\theta_0)$ is singular. For example, such a weight matrix can be constructed by adjusting the eigenvalues of $\widehat{\Omega}_n(\theta_0)$ to be bounded away from zero, and using its inverse. However, this method has two drawbacks. First, it sacrifices power relative to the definition of the SR-AR test in (5.7). The reason is that it does not reject H_0 with probability one when a violation of the nonstochastic part of the moment conditions occurs. This can be seen in the example with identities and the two examples that follow it. Second, it cannot be used with the SR-CQLR tests introduced in Sections 6 and 7 below. The reason is that these tests rely on a statistic $\widehat{D}_n(\theta_0)$, defined in (6.2) below, that employs $\widehat{\Omega}_n^{-1}(\theta_0)$ and if $\widehat{\Omega}_n^{-1}(\theta_0)$ is replaced by a matrix that is nonsingular by construction, such as the eigenvalue-adjusted matrix suggested above, then one does not obtain asymptotic independence of $\widehat{g}_n(\theta_0)$ and $\widehat{D}_n(\theta_0)$ after suitable normalization, which is needed to obtain the correct asymptotic size of the SR-CQLR tests.

The SR-AR test statistic can be written equivalently as

$$SR\text{-}AR_n(\theta) = n\widehat{g}_n(\theta)'\widehat{\Omega}_n^+(\theta)\widehat{g}_n(\theta) = n\widehat{g}_{An}(\theta)'\widehat{\Pi}_{1n}^{-1}(\theta)\widehat{g}_{An}(\theta), \tag{5.8}$$

where $\widehat{\Omega}_n^+(\theta)$ denotes the Moore-Penrose generalized inverse of $\widehat{\Omega}_n(\theta)$, when $\widehat{r}_n(\theta_0) \neq 0$.[32] The expression for the SR-AR statistic given in (5.6) is preferable to the Moore-Penrose expression in (5.8) for the derivation of the asymptotic results. It is not the case that $SR\text{-}AR_n(\theta)$ equals the rhs expression in (5.8) with probability one when $\widehat{\Omega}_n^+(\theta)$ is replaced by an arbitrary generalized inverse of $\widehat{\Omega}_n(\theta)$.

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

$$CS_{SR\text{-}AR,n} := \{\theta_0 \in \Theta : SR\text{-}AR_n(\theta_0) \leq \chi^2_{\widehat{r}_n(\theta_0),1-\alpha} \text{ and } \widehat{A}_n^\perp(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}\}. \tag{5.9}$$

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 $\widehat{r}_n(\theta_0) = k$, the $SR\text{-}AR_n(\theta_0)$ statistic equals $AR_n(\theta_0)$ 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 13 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.

6 SR-CQLR$_1$ Test

This section defines the SR-CQLR$_1$ test. This test applies when the moment functions are of the product form in (4.4). For expositional clarity and convenience (here and in the proofs), we first define the test in Section 6.1 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 6.2 to the case where these variance matrices may be singular.

[32] This holds by the following calculations. For notational simplicity, we suppress the dependence of quantities on θ. We have $SR\text{-}AR_n = n\widehat{g}_n'\widehat{A}_n(\widehat{A}_n'\widehat{\Omega}_n\widehat{A}_n)^{-1}\widehat{A}_n'\widehat{g}_n = n\widehat{g}_n'\widehat{A}_n(\widehat{A}_n'[\widehat{A}_n, \widehat{A}_n^\perp]\widehat{\Pi}_n[\widehat{A}_n, \widehat{A}_n^\perp]'\widehat{A}_n)^{-1}\widehat{A}_n'\widehat{g}_n = n\widehat{g}_n'\widehat{A}_n\widehat{\Pi}_{1n}^{-1}\widehat{A}_n'\widehat{g}_n$ and

$$n\widehat{g}_n'\widehat{\Omega}_n^+\widehat{g}_n = n\widehat{g}_n'[\widehat{A}_n, \widehat{A}_n^\perp]\begin{bmatrix} \widehat{\Pi}_{1n}^{-1} & 0^{\widehat{r}_n \times (k-\widehat{r}_n)} \\ 0^{(k-\widehat{r}_n)\times \widehat{r}_n} & 0^{(k-\widehat{r}_n)(k-\widehat{r}_n)} \end{bmatrix}[\widehat{A}_n, \widehat{A}_n^\perp]'\widehat{g}_n = n\widehat{g}_n'\widehat{A}_n\widehat{\Pi}_{1n}^{-1}\widehat{A}_n'\widehat{g}_n,$$

where the spectral decomposition of $\widehat{\Omega}_n$ given in (4.7) and (5.4) is used once in each equation above.

6.1 CQLR$_1$ Test for Nonsingular Moments Variance Matrices

The sample Jacobian is

$$\widehat{G}_n(\theta) := n^{-1}\sum_{i=1}^{n} G_i(\theta) = (\widehat{G}_{1n}(\theta), ..., \widehat{G}_{pn}(\theta)) \in R^{k\times p}. \tag{6.1}$$

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 $\widehat{g}_n(\theta)$:

$$\begin{aligned}
\widehat{D}_n(\theta) &:= (\widehat{D}_{1n}(\theta), ..., \widehat{D}_{pn}(\theta)) \in R^{k\times p}, \text{ where} \\
\widehat{D}_{jn}(\theta) &:= \widehat{G}_{jn}(\theta) - \widehat{\Gamma}_{jn}(\theta)\widehat{\Omega}_n^{-1}(\theta)\widehat{g}_n(\theta) \in R^k \text{ for } j = 1, ..., p, \text{ and} \\
\widehat{\Gamma}_{jn}(\theta) &:= n^{-1}\sum_{i=1}^{n}(G_{ij}(\theta) - \widehat{G}_{jn}(\theta))g_i(\theta)' \in R^{k\times k} \text{ for } j = 1, ..., p.
\end{aligned} \tag{6.2}$$

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

The statistics $\widehat{g}_n(\theta)$, $\widehat{\Omega}_n(\theta)$, $AR_n(\theta)$, and $\widehat{D}_n(\theta)$ are used by both the (non-SR) CQLR$_1$ test and the (non-SR) CQLR$_2$ test. The CQLR$_1$ test alone uses the following statistics:

$$\begin{aligned}
\widehat{R}_n(\theta) &:= \left(B(\theta)' \otimes I_k\right)\widehat{V}_n(\theta)\left(B(\theta) \otimes I_k\right) \in R^{(p+1)k\times(p+1)k}, \text{ where} \\
\widehat{V}_n(\theta) &:= n^{-1}\sum_{i=1}^{n}\left((u_i^*(\theta) - \widehat{u}_{in}^*(\theta))\left(u_i^*(\theta) - \widehat{u}_{in}^*(\theta)\right)'\right) \otimes \left(Z_iZ_i'\right) \in R^{(p+1)k\times(p+1)k}, \\
\widehat{u}_{in}^*(\theta) &:= \widehat{\Xi}_n(\theta)'Z_i \in R^{p+1}, \\
\widehat{\Xi}_n(\theta) &:= (Z_{n\times k}'Z_{n\times k})^{-1}Z_{n\times k}'U^*(\theta) \in R^{k\times(p+1)}, \\
Z_{n\times k} &:= (Z_1, ..., Z_n)' \in R^{n\times k}, \ U^*(\theta) := (u_1^*(\theta), ..., u_n^*(\theta))' \in R^{n\times(p+1)}, \text{ and} \\
B(\theta) &:= \begin{pmatrix} 1 & 0_p' \\ -\theta & -I_p \end{pmatrix} \in R^{(p+1)\times(p+1)},
\end{aligned} \tag{6.3}$$

where $u_i^*(\theta) := (u_i(\theta), u_{\theta i}(\theta)')'$ is defined in (4.5). Note that (i) $\widehat{V}_n(\theta)$ is an estimator of the variance matrix of the moment function and its vectorized derivatives, (ii) $\widehat{V}_n(\theta)$ exploits the functional form of the moment conditions given in (4.4), (iii) $\widehat{V}_n(\theta)$ typically is not of a Kronecker product form, and (iv) $\widehat{u}_{in}^*(\theta)$ is the best linear predictor of $u_i^*(\theta)$ based on $\{Z_i : n \geq 1\}$. The estimators $\widehat{R}_n(\theta)$, $\widehat{V}_n(\theta)$, and $\widehat{\Sigma}_n(\theta)$ (defined immediately below) are defined so that the SR-CQLR$_1$ test, which employs them, is asymptotically equivalent to Moreira's (2003) CLR test under all strengths of identification in the homoskedastic linear IV model with fixed IV's and p rhs endogenous variables for any $p \geq 1$. See Section 14 in the SM for details.

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

$$\left\| (I_{p+1} \otimes \widehat{\Omega}_n^{-1/2}(\theta))[\Sigma \otimes \widehat{\Omega}_n(\theta) - \widehat{R}_n(\theta)](I_{p+1} \otimes \widehat{\Omega}_n^{-1/2}(\theta)) \right\| \tag{6.4}$$

over all symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$, where $|| \cdot ||$ denotes the Frobenius norm (i.e., the Euclidean norm of the vectorized matrix). This is a weighted minimization problem with the weights given by $I_{p+1} \otimes \widehat{\Omega}_n^{-1/2}(\theta)$. We employ these weights because they lead to a matrix $\widehat{\Sigma}_n(\theta)$ that is invariant to nonsingular transformations of the moment functions. (That is, $\widehat{\Sigma}_n(\theta)$ is invariant to the multiplication of $g_i(\theta)$ and $G_i(\theta)$ by any nonsingular matrix $M \in R^{k\times k}$, wherever $g_i(\theta)$ and $G_i(\theta)$ appear in the definitions of the statistics above, see Lemma 6.2 below.) Equation (6.4) is a least squares minimization problem and, hence, has a closed form solution, which is given as follows. Let $\widehat{\Sigma}_{j\ell n}(\theta)$ denote the (j, ℓ) element of $\widehat{\Sigma}_n(\theta)$. By Theorems 3 and 10 of Van Loan and Pitsianis (1993), for $j, \ell = 1, ..., p + 1$,

$$\widehat{\Sigma}_{j\ell n}(\theta) = tr(\widehat{R}_{j\ell n}(\theta)' \widehat{\Omega}_n^{-1}(\theta))/k, \tag{6.5}$$

where $\widehat{R}_{j\ell n}(\theta)$ denotes the (j, ℓ) submatrix of dimension $k \times k$ of $\widehat{R}_n(\theta)$.[33,34]

The estimator $\widehat{\Sigma}_n(\theta)$ is an estimator of a matrix that could be singular or nearly singular in some cases. For example, in the homoskedastic linear IV model in Section 3, $\widehat{\Sigma}_n(\theta)$ is an estimator of the variance matrix Σ_V of the reduced-form errors when θ is the true parameter, and Σ_V could be singular or nearly singular. In the definition of the $QLR_{1n}(\theta)$ statistic, we use an eigenvalue-adjusted version of $\widehat{\Sigma}_n(\theta)$, denoted by $\widehat{\Sigma}_n^{\varepsilon}(\theta)$, whose condition number (i.e., $\lambda_{\max}(\widehat{\Sigma}_n(\theta))/\lambda_{\min}(\widehat{\Sigma}_n(\theta))$) is bounded above by construction. The reason for making this adjustment is that the inverse of this matrix enters the definition of $QLR_{1n}(\theta)$. The adjustment improves the asymptotic and finite-sample performance of the test by making it robust to singularities and near singularities of the matrix that $\widehat{\Sigma}_n(\theta)$ estimates. The adjustment affects the test statistic (i.e., $\widehat{\Sigma}_n^{\varepsilon}(\theta) \neq \widehat{\Sigma}_n(\theta)$) only if the condition number of $\widehat{\Sigma}_n(\theta)$ exceeds $1/\varepsilon$. Hence, for a reasonable choice of ε, it often has no effect even in finite samples. This differs from many tuning parameters employed in the literature, such as the ones that appear in nonparametric and semiparametric procedures, because their choice often has a substantial effect on the statistic being considered. Based on the finite-sample simulations, we recommend using $\varepsilon = .05$.

The eigenvalue-adjustment procedure is defined as follows for an arbitrary non-zero positive semi-definite (psd) matrix $H \in R^{d_H \times d_H}$ for some positive integer d_H. Let ε be a positive constant.

[33] 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.

[34] Moreira and Moreira (2013) utilize 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.

Let $A_H \Lambda_H A_H'$ be a spectral decomposition of H, where $\Lambda_H = Diag\{\lambda_{H1}, ..., \lambda_{Hd_H}\} \in R^{d_H \times d_H}$ is the diagonal matrix of eigenvalues of H with nonnegative nonincreasing diagonal elements and A_H is a corresponding orthogonal matrix of eigenvectors of H. The eigenvalue-adjusted version of H, denoted $H^\varepsilon \in R^{d_H \times d_H}$, is defined by

$$H^\varepsilon := A_H \Lambda_H^\varepsilon A_H', \text{ where } \Lambda_H^\varepsilon := Diag\{\max\{\lambda_{H1}, \lambda_{\max}(H)\varepsilon\}, ..., \max\{\lambda_{Hd_H}, \lambda_{\max}(H)\varepsilon\}\}, \quad (6.6)$$

where $\lambda_{\max}(H)$ denotes the maximum eigenvalue of H. Note that $\lambda_{\max}(H) = \lambda_{H1}$, and $\lambda_{\max}(H) > 0$ provided the psd matrix H is non-zero. From its definition, it is clear that $H^\varepsilon = H$ whenever the condition number of H is less than or equal to $1/\varepsilon$ (provided $\varepsilon \leq 1$).

In Lemma 17.1 in Section 17 in the SM, we show that the eigenvalue-adjustment procedure possesses the following desirable properties: (i) (uniqueness) H^ε is uniquely defined (i.e., every choice of spectral decomposition of H yields the same matrix H^ε), (ii) (eigenvalue lower bound) $\lambda_{\min}(H^\varepsilon) \geq \lambda_{\max}(H)\varepsilon$, (iii) (condition number upper bound) $\lambda_{\max}(H^\varepsilon)/\lambda_{\min}(H^\varepsilon) \leq \max\{1/\varepsilon, 1\}$, (iv) (scale equivariance) for all $c > 0$, $(cH)^\varepsilon = cH^\varepsilon$, and (v) (continuity) $H_n^\varepsilon \to H^\varepsilon$ for any sequence of psd matrices $\{H_n \in R^{d_H \times d_H} : n \geq 1\}$ that satisfies $H_n \to H$.

The QLR$_1$ statistic, which applies when (4.4) holds, is defined as follows:

$$\begin{aligned} QLR_{1n}(\theta) &:= AR_n(\theta) - \lambda_{\min}(n\widehat{Q}_n(\theta)), \text{ where} \\ \widehat{Q}_n(\theta) &:= \left(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \widehat{D}_n^*(\theta)\right)' \left(\widehat{\Omega}_n^{-1/2}(\theta)\widehat{g}_n(\theta), \widehat{D}_n^*(\theta)\right) \in R^{(p+1)\times(p+1)}, \\ \widehat{D}_n^*(\theta) &:= \widehat{\Omega}_n^{-1/2}(\theta)\widehat{D}_n(\theta)\widehat{L}_n^{1/2}(\theta) \in R^{k\times p}, \text{ and} \\ \widehat{L}_n(\theta) &:= (\theta, I_p)(\widehat{\Sigma}_n^\varepsilon(\theta))^{-1}(\theta, I_p)' \in R^{p\times p}, \end{aligned} \quad (6.7)$$

where $\widehat{\Sigma}_n^\varepsilon(\theta)$ is defined in (6.6) with $H = \widehat{\Sigma}_n(\theta)$.[35] Comparing (3.4) and (6.7), one sees the common structure of the $LR_{HGL,n}$ and $QLR_{1n}(\theta_0)$ statistics, where θ_0 is the null value. The k vector $n^{1/2}\widehat{\Omega}_n^{-1/2}(\theta_0)\widehat{g}_n(\theta_0)$ plays the role of $\overline{S}_n$, and the $k \times p$ matrix $n^{1/2}\widehat{D}_n^*(\theta_0)$ plays the role of $\overline{T}_n$. The matrix $\widehat{L}_n(\theta)$ is defined such that these quantities are asymptotically equivalent in the homoskedastic linear IV regression model with fixed IV's (in scenarios where the eigenvalue adjustment is irrelevant wp→1).

The CQLR$_1$ test uses the QLR$_1$ statistic and a conditional critical value that depends on the $k \times p$ matrix $n^{1/2}\widehat{D}_n^*(\theta_0)$ through the conditional critical value function $c_{k,p}(D, 1-\alpha)$, which is

[35] The asymptotic size result given in Section 8 below for the SR-CQLR$_1$ test still holds if no eigenvalue adjustment is made to $\widehat{\Sigma}_n(\theta)$ provided the parameter space of distributions $\mathcal{F}_1^{SR}$ is restricted so that the population version of $\widehat{\Sigma}_n(\theta)$ has a condition number that is bounded above.

defined in (3.5). For $\alpha \in (0,1)$, the nominal α CQLR$_1$ test rejects $H_0 : \theta = \theta_0$ if

$$QLR_{1n}(\theta_0) > c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha). \tag{6.8}$$

The nominal $100(1-\alpha)\%$ CQLR$_1$ CS is $CS_{CQLR_1,n} := \{\theta_0 \in \Theta : QLR_{1n}(\theta_0) \leq c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha)\}$.

The following lemma shows that the critical value function $c_{k,p}(D, 1-\alpha)$ depends on D only through its singular values.

Lemma 6.1 *Let D be a $k \times p$ matrix with the singular value decomposition $D = C\Upsilon B'$, where C is a $k \times k$ orthogonal matrix of eigenvectors of DD', B is a $p \times p$ orthogonal matrix of eigenvectors of $D'D$, and Υ is the $k \times p$ matrix with the* $\min\{k,p\}$ *singular values $\{\tau_j : j \leq \min\{k,p\}\}$ of D as its first* $\min\{k,p\}$ *diagonal elements and zeros elsewhere, where τ_j is nonincreasing in j. Then, $c_{k,p}(D, 1-\alpha) = c_{k,p}(\Upsilon, 1-\alpha)$.*

Comment: A consequence of Lemma 6.1 is that the critical value $c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha)$ of the CQLR$_1$ test depends on $\widehat{D}_n^*(\theta_0)$ only through $\widehat{D}_n^*(\theta_0)'\widehat{D}_n^*(\theta_0)$ (because, when $k \geq p$, the p singular values of $n^{1/2}\widehat{D}_n^*(\theta_0)$ equal the square roots of the eigenvalues of $n\widehat{D}_n^*(\theta_0)'\widehat{D}_n^*(\theta_0)$ and, when $k < p$, $c_{k,p}(D, 1-\alpha)$ is the $1-\alpha$ quantile of the χ_k^2 distribution which does not depend on D).

The following lemma shows that the CQLR$_1$ test is invariant to nonsingular transformations of the moment functions/IV's. For notational simplicity, we suppress the dependence on θ of the statistics that appear in the lemma.

Lemma 6.2 *The statistics QLR_{1n}, $c_{k,p}(n^{1/2}\widehat{D}_n^*, 1-\alpha)$, $\widehat{D}_n^{*\prime}\widehat{D}_n^*$, AR_n, $\widehat{u}_{in}^*$, $\widehat{\Sigma}_n$, and $\widehat{L}_n$ are invariant to the transformation $(Z_i, u_i^*) \rightsquigarrow (MZ_i, u_i^*)$ for any $k \times k$ nonsingular matrix M. This transformation induces the following transformations: $g_i \rightsquigarrow Mg_i, G_i \rightsquigarrow MG_i, \widehat{g}_n \rightsquigarrow M\widehat{g}_n$, $\widehat{G}_n \rightsquigarrow M\widehat{G}_n$, $\widehat{\Omega}_n \rightsquigarrow M\widehat{\Omega}_nM'$, $\widehat{\Gamma}_{jn} \rightsquigarrow M\widehat{\Gamma}_{jn}M'$, $\widehat{D}_n \rightsquigarrow M\widehat{D}_n$, $Z_{n\times k} \rightsquigarrow Z_{n\times k}M'$, $\widehat{\Xi}_n \rightsquigarrow {M'}^{-1}\widehat{\Xi}_n$, $\widehat{V}_n \rightsquigarrow (I_{p+1} \otimes M)\,\widehat{V}_n\,(I_{p+1} \otimes M')$, and $\widehat{R}_n \rightsquigarrow (I_{p+1} \otimes M)\,\widehat{R}_n\,(I_{p+1} \otimes M')$.*

Comment: This Lemma is important because it implies that one can obtain the correct asymptotic size of the CQLR$_1$ test defined above without assuming that $\lambda_{\min}(\Omega_F)$ is bounded away from zero. It suffices that Ω_F is nonsingular. The reason is that (in the proofs) one can 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 $Var_F(M_F g_i) = I_k$) even if the original moments g_i do not.

6.2 Singularity-Robust CQLR$_1$ Test

Now, we extend the CQLR$_1$ 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 $\widehat{r}_n(\theta) \geq 1$, where $\widehat{r}_n(\theta)$ is defined in (5.3), we define $\widehat{D}_{An}(\theta)$ as $\widehat{D}_n(\theta)$ is defined in (6.2), 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 $j = 1, ..., p$, where $\widehat{A}_n(\theta)$ and $\widehat{\Omega}_{An}$ are defined in (5.4) and (5.5), respectively. That is,

$$
\begin{aligned}
\widehat{D}_{An}(\theta) &:= (\widehat{D}_{A1n}(\theta), ..., \widehat{D}_{Apn}(\theta)) \in R^{\widehat{r}_n(\theta)\times p}, \text{ where} \\
\widehat{D}_{Ajn}(\theta) &:= \widehat{G}_{Ajn}(\theta) - \widehat{\Gamma}_{Ajn}(\theta)\widehat{\Omega}_{An}^{-1}(\theta)\widehat{g}_{An}(\theta) \in R^{\widehat{r}_n(\theta)} \text{ for } j = 1, ..., p, \\
\widehat{G}_{An}(\theta) &:= \widehat{A}_n(\theta)'\widehat{G}_n(\theta) = (\widehat{G}_{A1n}(\theta), ..., \widehat{G}_{Apn}(\theta)) \in R^{\widehat{r}_n(\theta)\times p}, \text{ and} \\
\widehat{\Gamma}_{Ajn}(\theta) &:= \widehat{A}_n(\theta)'\widehat{\Gamma}_{jn}(\theta)\widehat{A}_n(\theta) \text{ for } j = 1, ..., p. \qquad (6.9)
\end{aligned}
$$

Let $Z_{Ai}(\theta) := \widehat{A}_n(\theta)'Z_i \in R^{\widehat{r}_n(\theta)}$ and $Z_{An\times k}(\theta) := Z_{n\times k}\widehat{A}_n(\theta) \in R^{n\times \widehat{r}_n(\theta)}$.

The SR-CQLR$_1$ test employs statistics $\widehat{R}_{An}(\theta)$, $\widehat{\Sigma}_{An}(\theta)$, $\widehat{L}_{An}(\theta)$, and $\widehat{D}^*_{An}(\theta)$, which are defined just as $\widehat{R}_n(\theta)$, $\widehat{\Sigma}_n(\theta)$, $\widehat{L}_n(\theta)$, and $\widehat{D}^*_n(\theta)$ are defined in Section 6.1, but with $\widehat{g}_{An}(\theta)$, $\widehat{G}_{An}(\theta)$, $\widehat{\Omega}_{An}(\theta)$, $Z_{Ai}(\theta)$, $Z_{An\times k}(\theta)$, and $\widehat{r}_n(\theta)$ in place of $\widehat{g}_n(\theta)$, $\widehat{G}_n(\theta)$, $\widehat{\Omega}_n(\theta)$, Z_i, $Z_{n\times k}$, and k, respectively, using the definitions in (5.3), (5.5) and (6.9). In particular, we have

$$
\begin{aligned}
\widehat{R}_{An}(\theta) &:= \left(B(\theta)' \otimes I_{\widehat{r}_n(\theta)}\right) \widehat{V}_{An}(\theta) \left(B(\theta) \otimes I_{\widehat{r}_n(\theta)}\right) \in R^{(p+1)\widehat{r}_n(\theta)\times(p+1)\widehat{r}_n(\theta)}, \text{ where} \\
\widehat{V}_{An}(\theta) &:= n^{-1}\sum_{i=1}^{n} \left((u_i^*(\theta) - \widehat{u}^*_{Ain}(\theta))(u_i^*(\theta) - \widehat{u}^*_{Ain}(\theta))'\right) \otimes \left(Z_{Ai}(\theta)Z_{Ai}(\theta)'\right) \\
&\in R^{(p+1)\widehat{r}_n(\theta)\times(p+1)\widehat{r}_n(\theta)}, \\
\widehat{u}^*_{Ain}(\theta) &:= \widehat{\Xi}_{An}(\theta)'Z_{Ai}(\theta) \in R^{p+1}, \\
\widehat{\Xi}_{An}(\theta) &:= (Z_{An\times k}(\theta)'Z_{An\times k}(\theta))^{-1}Z_{An\times k}(\theta)'U^*(\theta) \in R^{\widehat{r}_n(\theta)\times(p+1)}, \\
\widehat{\Sigma}_{Aj\ell n}(\theta) &:= tr(\widehat{R}_{Aj\ell n}(\theta)'\widehat{\Omega}_{An}^{-1}(\theta))/\widehat{r}_n(\theta) \text{ for } j, \ell = 1, ..., p+1, \\
\widehat{L}_{An}(\theta) &:= (\theta, I_p)(\widehat{\Sigma}^{\varepsilon}_{An}(\theta))^{-1}(\theta, I_p)' \in R^{p\times p}, \\
\widehat{D}^*_{An}(\theta) &:= \widehat{\Omega}_{An}^{-1/2}(\theta)\widehat{D}_{An}(\theta)\widehat{L}_{An}^{1/2}(\theta) \in R^{\widehat{r}_n(\theta)\times p}, \qquad (6.10)
\end{aligned}
$$

$\widehat{A}_n(\theta)$ is defined in (5.4), $\widehat{\Sigma}_{Aj\ell n}(\theta)$ 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 $\widehat{R}_{An}(\theta)$.

If $\widehat{r}_n(\theta) > 0$, the SR-QLR$_1$ statistic is defined by

$$SR\text{-}QLR_{1n}(\theta) := SR\text{-}AR_n(\theta) - \lambda_{\min}(n\widehat{Q}_{An}(\theta)), \text{ where} \tag{6.11}$$
$$\widehat{Q}_{An}(\theta) := \left(\widehat{\Omega}_{An}^{-1/2}(\theta)\widehat{g}_{An}(\theta), \widehat{D}_{An}^*(\theta)\right)' \left(\widehat{\Omega}_{An}^{-1/2}(\theta)\widehat{g}_{An}(\theta), \widehat{D}_{An}^*(\theta)\right) \in R^{(p+1)\times(p+1)}.$$

For $\alpha \in (0, 1)$, the nominal size α SR-CQLR$_1$ test rejects $H_0 : \theta = \theta_0$ if

$$SR\text{-}QLR_{1n}(\theta_0) > c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widehat{D}_{An}^*(\theta_0), 1-\alpha) \text{ or } \widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}.^{36} \tag{6.12}$$

The nominal size $100(1-\alpha)\%$ SR-CQLR$_1$ CS is $CS_{SR\text{-}CQLR_1,n} := \{\theta_0 \in \Theta : SR\text{-}QLR_{1n}(\theta_0) \leq c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widehat{D}_{An}^*(\theta_0), 1-\alpha)$ and $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}\}$.[37]

Note that if $r \leq p$, then $c_{r,p}(D, 1-\alpha)$ is the $1-\alpha$ quantile of

$$CLR_{r,p}(D) := Z'Z - \lambda_{\min}((Z, D)'(Z, D)) = Z'Z \sim \chi_r^2, \tag{6.13}$$

where $Z \sim N(0^r, I_r)$ and the last equality holds because $(Z, D)'(Z, D)$ is a $(p+1)\times(p+1)$ matrix of rank $r \leq p$, which implies that its smallest eigenvalue is zero. Hence, if $\widehat{r}_n(\theta_0) \leq p$, then the critical value for the SR-CQLR$_1$ test is the $1-\alpha$ quantile of $\chi^2_{\widehat{r}_n(\theta_0)}$, which is denoted by $\chi^2_{\widehat{r}_n(\theta_0),1-\alpha}$.

When $\widehat{r}_n(\theta_0) = k$, $\widehat{A}_n(\theta_0)$ is a nonsingular $k \times k$ matrix. In consequence, by Lemma 6.2, $SR\text{-}QLR_{1n}(\theta_0) = QLR_{1n}(\theta_0)$ and $c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widehat{D}_{An}^*(\theta_0), 1-\alpha) = c_{k,p}(n^{1/2}\widehat{D}_n^*(\theta_0), 1-\alpha)$. That is, the SR-CQLR$_1$ test is the same as the CQLR$_1$ test defined in Section 6.1. Of course, when $\widehat{r}_n(\theta) < k$, the CQLR$_1$ test defined in Section 6.1 is not defined, whereas the SR-CQLR$_1$ test is. Thus, the SR-CQLR$_1$ test defined here is, indeed, an extension of the CQLR$_1$ test defined in Section 6.1 to the case where $\widehat{r}_n(\theta_0) < k$. Furthermore, if $rk(\Omega_{F_n}(\theta_0)) = k$ for all n large, then $\widehat{r}_n(\theta_0) = k$ and $SR\text{-}QLR_{1n}(\theta_0) = QLR_{1n}(\theta_0)$ wp$\to$1 under $\{F_n \in \mathcal{F}_2^{SR} : n \geq 1\}$ (by Lemmas 6.2 and 10.6 below).

7 SR-CQLR$_2$ Test

In this section, we define the SR-CQLR$_2$ test, which is quite similar to the SR-CQLR$_1$ test, but does not rely on $g_i(\theta)$ having the form in (4.4). First, we define the CQLR$_2$ test without the SR

[36] By definition, $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}$ does not hold if $\widehat{r}_n(\theta_0) = k$. If $\widehat{r}_n(\theta_0) = 0$, then $SR\text{-}QLR_{1n}(\theta_0) := 0$ and $\chi^2_{\widehat{r}_n(\theta_0),1-\alpha} := 0$. In this case, $\widehat{A}_n^{\perp}(\theta_0) = I_k$ and the SR-CQLR$_1$ test rejects H_0 if $\widehat{g}_n(\theta_0) \neq 0^k$.

[37] 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.

extension. We define an analogue $\widetilde{R}_n(\theta)$ of $\widehat{R}_n(\theta)$ as follows:

$$\begin{aligned}
\widetilde{R}_n(\theta) &:= \left(B(\theta)' \otimes I_k\right) \widetilde{V}_n(\theta) \left(B(\theta) \otimes I_k\right) \in R^{(p+1)k \times (p+1)k}, \text{ where} \\
\widetilde{V}_n(\theta) &:= n^{-1} \sum_{i=1}^{n} \left(f_i(\theta) - \widehat{f}_n(\theta)\right) \left(f_i(\theta) - \widehat{f}_n(\theta)\right)' \in R^{(p+1)k \times (p+1)k}, \\
f_i(\theta) &:= \begin{pmatrix} g_i(\theta) \\ vec(G_i(\theta)) \end{pmatrix}, \text{ and } \widehat{f}_n(\theta) := \begin{pmatrix} \widehat{g}_n(\theta) \\ vec(\widehat{G}_n(\theta)) \end{pmatrix}.
\end{aligned} \tag{7.1}$$

The SR-CQLR$_2$ test differs from the SR-CQLR$_1$ test because $\widetilde{V}_n(\theta)$ (and the statistics that depend on it) differs from $\widehat{V}_n(\theta)$ (and the statistics that depend on it). The estimator $\widetilde{V}_n(\theta)$ does not depend on the product form of the moment conditions given in (4.4).

We define $\widetilde{\Sigma}_n(\theta) \in R^{(p+1)\times(p+1)}$ just as $\widehat{\Sigma}_n(\theta)$ is defined in (6.4) and (6.5), but with $\widetilde{R}_n(\theta)$ in place of $\widehat{R}_n(\theta)$. We define $\widetilde{D}_n^*(\theta)$ just as $\widehat{D}_n^*(\theta)$ is defined in (6.7), but with $\widetilde{\Sigma}_n(\theta)$ in place of $\widehat{\Sigma}_n(\theta)$. That is,

$$\widetilde{D}_n^*(\theta) := \widehat{\Omega}_n(\theta)^{-1/2} \widehat{D}_n(\theta) \widetilde{L}_n^{1/2}(\theta) \in R^{k \times p}, \text{ where } \widetilde{L}_n(\theta) := (\theta, I_p)(\widetilde{\Sigma}_n^{\varepsilon}(\theta))^{-1}(\theta, I_p)'. \tag{7.2}$$

We use an eigenvalue-adjusted version of $\widetilde{\Sigma}_n(\theta)$ in the definition of $\widetilde{L}_n(\theta)$ because it yields an SR-CQLR test that has correct asymptotic size even if $Var_F(f_i)$ is singular for some F in the parameter space of distributions.

The QLR$_2$ statistic without the SR extension, denoted by $QLR_{2n}(\theta)$, is defined just as $QLR_{1n}(\theta)$ is defined in (6.7), but with $\widetilde{D}_n^*(\theta)$ in place of $\widehat{D}_n^*(\theta)$. For $\alpha \in (0,1)$, the nominal size α CQLR$_2$ test (without the SR extension) rejects $H_0 : \theta = \theta_0$ if

$$QLR_{2n}(\theta_0) > c_{k,p}(n^{1/2} \widetilde{D}_n^*(\theta_0), 1 - \alpha). \tag{7.3}$$

The nominal size $100(1-\alpha)\%$ CQLR$_2$ CS is $CS_{CQLR_2,n} := \{\theta_0 \in \Theta : QLR_{2n}(\theta_0) \leq c_{k,p}(n^{1/2}\widetilde{D}_n^*(\theta_0), 1 - \alpha)\}$.[38]

For the CQLR$_2$ test with the SR extension, we define $\widehat{D}_{An}(\theta)$ as in (6.9). We define

$$\widetilde{V}_{An}(\theta) := (I_{p+1} \otimes \widehat{A}_n(\theta)') \widetilde{V}_n(\theta) (I_{p+1} \otimes \widehat{A}_n(\theta)) \in R^{(p+1)\widehat{r}_n(\theta) \times (p+1)\widehat{r}_n(\theta)}, \tag{7.4}$$

where $\widehat{r}_n(\theta)$ and $\widehat{A}_n(\theta)$ are defined in (5.3) and (5.4), respectively. In addition, we define $\widetilde{R}_{An}(\theta)$,

[38] Analogously to the results of Lemma 6.2, the statistics QLR_{2n}, $c_{k,p}(n^{1/2}\widetilde{D}_n^*, 1-\alpha)$, $\widetilde{D}_n^{*\prime}\widetilde{D}_n^*$, $\widetilde{\Sigma}_n$, and $\widetilde{L}_n$ are invariant to the transformation $(g_i, G_i) \rightsquigarrow (Mg_i, MG_i)$ for any $k \times k$ nonsingular matrix M. This transformation induces the following equivariant transformations: $\widetilde{D}_n^* \rightsquigarrow M\widetilde{D}_n^*$, $\widetilde{V}_n \rightsquigarrow (I_{p+1} \otimes M) \widetilde{V}_n (I_{p+1} \otimes M')$, and $\widetilde{R}_n \rightsquigarrow (I_{p+1} \otimes M) \widetilde{R}_n (I_{p+1} \otimes M')$.

$\widetilde{\Sigma}_{An}(\theta)$, $\widetilde{L}_{An}(\theta)$, $\widetilde{D}^*_{An}(\theta)$, and $\widetilde{Q}_{An}(\theta)$ as $\widehat{R}_{An}(\theta)$, $\widehat{\Sigma}_{An}(\theta)$, $\widehat{L}_{An}(\theta)$, $\widehat{D}^*_{An}(\theta)$, and $\widehat{Q}_{An}(\theta)$ are defined, respectively, in (6.10) and (6.11), but with $\widetilde{V}_{An}(\theta)$ in place of $\widehat{V}_{An}(\theta)$ in the definition of $\widetilde{R}_{An}(\theta)$, with $\widetilde{R}_{An}(\theta)$ in place of $\widehat{R}_{An}(\theta)$ in the definition of $\widetilde{\Sigma}_{An}(\theta)$, and so on in the definitions of $\widetilde{L}_{An}(\theta)$, $\widetilde{D}^*_{An}(\theta)$, and $\widetilde{Q}_{An}(\theta)$. We define the test statistic $SR\text{-}QLR_{2n}(\theta)$ as $SR\text{-}QLR_{1n}(\theta)$ is defined in (6.11), but with $\widetilde{Q}_{An}(\theta)$ in place of $\widehat{Q}_{An}(\theta)$.

Given these definitions, the nominal size α SR-CQLR$_2$ test rejects $H_0 : \theta = \theta_0$ if

$$SR\text{-}QLR_{2n}(\theta_0) > c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widetilde{D}^*_{An}(\theta_0), 1-\alpha) \text{ or } \widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}.^{39} \quad (7.5)$$

The nominal size $100(1-\alpha)\%$ SR-CQLR$_2$ CS is $CS_{SR\text{-}CQLR_2,n} := \{\theta_0 \in \Theta : SR\text{-}QLR_{2n}(\theta_0) \leq c_{\widehat{r}_n(\theta_0),p}(n^{1/2}\widetilde{D}^*_{An}(\theta_0), 1-\alpha)$ and $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) = 0^{k-\widehat{r}_n(\theta_0)}\}$.[40]

Section 13 in the SM provides finite-sample null rejection probabilities of the SR-CQLR$_2$ test for singular and near singular variance matrices of the moment functions.[41] The results show that singularity and near singularity of the variance matrix does not lead to distorted null rejection probabilities. The method of robustifying the SR-CQLR$_2$ test to allow for singular variance matrices, which is introduced above, works quite well in the model that is considered.

8 Asymptotic Size

The correct asymptotic size and similarity results for the SR-AR, SR-CQLR$_1$, and SR-CQLR$_2$ tests are as follows.

Theorem 8.1 *The asymptotic sizes of the* SR-AR, SR-CQLR$_1$, *and* SR-CQLR$_2$ *tests defined in* (5.7), (6.12), *and* (7.5), *respectively, equal their nominal size* $\alpha \in (0,1)$ *for the null parameter spaces* $\mathcal{F}^{SR}_{AR}$, $\mathcal{F}^{SR}_1$, *and* $\mathcal{F}^{SR}_2$, *respectively. Furthermore, 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, SR-CQLR$_1$, *and* SR-CQLR$_2$ CS's *for the parameter spaces* $\mathcal{F}^{SR}_{\Theta,AR}$, $\mathcal{F}^{SR}_{\Theta,1}$, *and* $\mathcal{F}^{SR}_{\Theta,2}$, *respectively, defined in* (4.10).

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

[39] By definition, $\widehat{A}_n^{\perp}(\theta_0)'\widehat{g}_n(\theta_0) \neq 0^{k-\widehat{r}_n(\theta_0)}$ does not hold if $\widehat{r}_n(\theta_0) = k$. If $\widehat{r}_n(\theta_0) = 0$, then $SR\text{-}QLR_{2n}(\theta_0) := 0$ and $\chi^2_{\widehat{r}_n(\theta_0),1-\alpha} := 0$. In this case, $\widehat{A}_n^{\perp}(\theta_0) = I_k$ and the SR-CQLR$_2$ test rejects H_0 if $\widehat{g}_n(\theta_0) \neq 0^k$.

[40] 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.

[41] Analogous results are not given for the SR-CQLR$_1$ test because the moment functions considered are not of the form in (4.4), which is necessary to apply the SR-CQLR$_1$ test.

(ii) SR-LM versions of Kleibergen's LM test and CS can be defined analogously to the SR-AR and SR-CQLR tests and CS's. However, these procedures are only partially singularity robust. See Section 18 in the SM.

(iii) The proof of Theorem 8.1 is given partly in the Appendix and partly in the SM.

9 Asymptotic Efficiency of the SR-CQLR Tests under Strong Identification

Next, we show that the SR-CQLR$_1$ and SR-CQLR$_2$ tests are 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 they are asymptotically equivalent (under the null and contiguous alternatives) to a Wald test constructed using an asymptotically efficient GMM estimator, see Newey and West (1987).

Kleibergen's LM statistic and the standard GMM LM statistic, see Newey and West (1987), are defined by

$$LM_n := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{D}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n \text{ and } LM_n^{GMM} := n\widehat{g}_n'\widehat{\Omega}_n^{-1/2}P_{\widehat{\Omega}_n^{-1/2}\widehat{G}_n}\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \tag{9.1}$$

respectively, where $\widehat{G}_n$ is the sample Jacobian defined in (5.1) with $\theta = \theta_0$. The test based on the standard GMM LM statistic (combined with a χ_p^2 critical value) is asymptotically equivalent to the Wald test based on an asymptotically efficient GMM estimator under (i) strong identification (which requires $k \geq p$), (ii) nonsingular moments-variance matrices (i.e., $\lambda_{\min}(\Omega_{F_n}) \geq \delta > 0$ for all $n \geq 1$), and (iii) a null parameter value that is not on the boundary of the parameter space, see Newey and West (1987). This also holds true under semi-strong identification (which also requires $k \geq p$) . For example, Theorem 5.1 of Andrews and Cheng (2013) shows that the Wald statistic for testing $H_0 : \theta = \theta_0$ based on a GMM estimator with asymptotically efficient weight matrix has a χ_p^2 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 $k \geq p$. Let A_F and Π_{1F} be defined as in (4.7) and (4.8) and the paragraph following these equations with $\theta = \theta_0$. Define λ_F^*, Λ_1^*, Λ_2^*, and $\{\lambda_{n,h}^* : n \geq 1\}$ as λ_F, Λ_1, Λ_2, and $\{\lambda_{n,h} : n \geq 1\}$, respectively, are defined in (10.16)-(10.18) in the Appendix, but with g_i and G_i replaced by $g_{Fi}^* := \Pi_{1F}^{-1/2}A_F'g_i$ and $G_{Fi}^* := \Pi_{1F}^{-1/2}A_F'G_i$, with $\mathcal{F}_1$ replaced by $\mathcal{F}_1^{SR}$, with $\mathcal{F}_2$ replaced by $\mathcal{F}_2^{SR}$ in the definition of $\mathcal{F}_{WU}$, and with W_F $(:= W_1(W_{2F}))$ and U_F $(:= U_1(U_{2F}))$ defined as in (10.8) and (10.11) in the Appendix for the CQLR$_1$ and CQLR$_2$ tests, respectively, with g_i and G_i replaced by g^*_{Fi} and G^*_{Fi}. In addition, 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 $n \geq 1$.[42] By definition, a sequence $\{\lambda^*_{n,h} : n \geq 1\}$ is said to exhibit strong or semi-strong identification if $n^{1/2} s^*_{pF_n} \to \infty$, where s^*_{pF} denotes the smallest singular value of $E_F G^*_{Fi}$.[43]

Let $\chi^2_{p,1-\alpha}$ denote the $1-\alpha$ quantile of the χ^2_p distribution. The critical value for the LM_n and LM_n^{GMM} tests is $\chi^2_{p,1-\alpha}$.

Theorem 9.1 *Suppose $k \geq p$. For any sequence $\{\lambda^*_{n,h} : n \geq 1\}$ that exhibits strong or semi-strong identification (i.e., for which $n^{1/2} s^*_{pF_n} \to \infty$) and for which $\lambda^*_{n,h} \in \Lambda^*_1$ $\forall n \geq 1$ for the* SR-CQLR$_1$ *test statistic and critical value and $\lambda^*_{n,h} \in \Lambda^*_2$ $\forall n \geq 1$ for the* SR-CQLR$_2$ *test statistic and critical value, we have*

(a) $SR\text{-}QLR_{jn} = QLR_{jn} + o_p(1) = LM_n + o_p(1) = LM_n^{GMM} + o_p(1)$ *for* $j = 1, 2$,

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

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

Comments: (i) Theorem 9.1 establishes the asymptotic efficiency (in a GMM sense) of the SR-CQLR$_1$ and SR-CQLR$_2$ tests under strong and semi-strong identification. Note that Theorem 9.1 provides asymptotic equivalence results under the null hypothesis, but, by the definition of contiguity, these asymptotic equivalence results also hold under contiguous local alternatives.

(ii) The proof of Theorem 9.1 is given in Section 23 in the SM.

[42] Thus, $A_F = A_F^\dagger$, $\Pi_{1F} = \Pi_F$, $W_F := (\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2})^{-1/2} = I_k$, and by an invariance property, which follows from calculations similar to those used to establish Lemma 6.2, U_F (defined in the Appendix) is the same whether it is defined using g_i and G_i or g^*_{Fi} and G^*_{Fi}.

[43] The singular value s^*_{pF}, defined here, equals s_{pF}, defined in the Introduction, for all F with $\lambda_{\min}(\Omega_F) > 0$, because in this case $\Omega_F = A_F \Pi_{1F} A_F'$, $\Omega_F^{-1/2} = A_F \Pi_{1F}^{-1/2} A_F'$, $\Omega_F^{-1/2} E_F G_i = A_F \Pi_{1F}^{-1/2} A_F' E_F G_i = A_F E_F G^*_{Fi}$, and A_F is an orthogonal $k \times k$ matrix. Since we consider sequences here with $\lambda_{\min}(\Omega_{F_n}) = \lambda_{\min}(E_{F_n} g_i g_i') > 0$ for all $n \geq 1$, the definitions of strong and semi-strong identification used here and in the Introduction are equivalent.

10 Appendix

This Appendix, along with parts of the SM, is devoted to the proof of Theorem 8.1. The proof proceeds in two steps. First, we establish the correct asymptotic size and asymptotic similarity of the tests and CS's without the SR extension for parameter spaces of distributions that bound $\lambda_{\min}(\Omega_F)$ away from zero. (These tests are defined in (5.2), (6.8), and (7.3).) We provide some parts of the proof of this result in Section 10.1 below. The details are given in Section 22 in the SM. Second, we extend the proof to the case of the SR tests and CS's. We provide the proof of this extension in Section 10.2 below.

10.1 Tests without the Singularity-Robust Extension

10.1.1 Asymptotic Results for Tests without the SR Extension

For the AR and CQLR tests without the SR extension, we consider the following parameter spaces for the distribution F that generates the data under $H_0 : \theta = \theta_0$:

$$\begin{aligned} \mathcal{F}_{AR} &:= \{F : E_F g_i = 0^k,\ E_F||g_i||^{2+\gamma} \leq M, \text{ and } \lambda_{\min}(E_F g_i g_i') \geq \delta\}, \\ \mathcal{F}_2 &:= \{F \in \mathcal{F}_{AR} : E_F||vec(G_i)||^{2+\gamma} \leq M\}, \text{ and} \\ \mathcal{F}_1 &:= \{F \in \mathcal{F}_2 : E_F||Z_i||^{4+\gamma} \leq M,\ E_F||u_i^*||^{2+\gamma} \leq M,\ \lambda_{\min}(E_F Z_i Z_i') \geq \delta\} \end{aligned} \tag{10.1}$$

for some $\gamma, \delta > 0$ and $M < \infty$. By definition, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_{AR}$. The parameter spaces $\mathcal{F}_{AR}$, $\mathcal{F}_2$, and $\mathcal{F}_1$, are used for the AR, CQLR$_2$, and CQLR$_1$ tests, respectively. For the corresponding CS's, we use the parameter spaces: $\mathcal{F}_{\Theta,AR} := \{(F, \theta_0) : F \in \mathcal{F}_{AR}(\theta_0), \theta_0 \in \Theta\}$, $\mathcal{F}_{\Theta,2} := \{(F, \theta_0) : F \in \mathcal{F}_2(\theta_0), \theta_0 \in \Theta\}$, and $\mathcal{F}_{\Theta,1} := \{(F, \theta_0) : F \in \mathcal{F}_1(\theta_0), \theta_0 \in \Theta\}$, where $\mathcal{F}_{AR}(\theta_0)$, $\mathcal{F}_2(\theta_0)$, and $\mathcal{F}_1(\theta_0)$ equal $\mathcal{F}_{AR}$, $\mathcal{F}_2$, and $\mathcal{F}_1$, respectively, with their dependence on θ_0 made explicit.

Theorem 10.1 *The* AR, CQLR$_1$, *and* CQLR$_2$ *tests* (*without the* SR *extensions*)*, defined in* (5.2)*,* (6.8)*, and* (7.3)*, respectively, have asymptotic sizes equal to their nominal size* $\alpha \in (0, 1)$ *and are asymptotically similar* (*in a uniform sense*) *for the parameter spaces* $\mathcal{F}_{AR}$*,* $\mathcal{F}_1$*, and* $\mathcal{F}_2$*, respectively. Analogous results hold for the corresponding* AR, CQLR$_1$, *and* CQLR$_2$ CS's *for the parameter spaces* $\mathcal{F}_{\Theta,AR}$*,* $\mathcal{F}_{\Theta,1}$*, and* $\mathcal{F}_{\Theta,2}$*, respectively.*

Comment: (i) The first step of the proof of Theorem 8.1 is to prove Theorem 10.1.

(ii) Theorem 10.1 holds for both $k \geq p$ and $k < p$. Both cases are needed in the proof of Theorem 8.1 (even if $k \geq p$ in Theorem 8.1).

10.1.2 Uniformity Framework

The proof of Theorem 10.1 uses Corollary 2.1(c) in Andrews, Cheng, and Guggenberger (2009) (ACG), which provides general sufficient conditions for the correct asymptotic size and (uniform) asymptotic similarity of a sequence of tests.

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

$$\begin{aligned} H := \{&h \in (R \cup \{\pm\infty\})^J : h_{w_n}(\lambda_{w_n}) \rightarrow h \text{ for some subsequence } \{w_n\} \\ &\text{of } \{n\} \text{ and some sequence } \{\lambda_{w_n} \in \Lambda : n \geq 1\}\}. \end{aligned} \tag{10.2}$$

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

Proposition 10.2 (ACG, Corollary 2.1(c)) *Under Assumption* B*, *the tests* $\{\phi_n : n \geq 1\}$ *have asymptotic size* α *and are asymptotically similar* (*in a uniform sense*). *That is,* $AsySz := \limsup_{n\rightarrow\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda) = \alpha$ *and* $\liminf_{n\rightarrow\infty} \inf_{\lambda\in\Lambda} RP_n(\lambda) = \limsup_{n\rightarrow\infty} \sup_{\lambda\in\Lambda} RP_n(\lambda)$.

Comments: (i) By Comment 4 to Theorem 2.1 of ACG, Proposition 10.2 provides asymptotic size and similarity results for nominal $1-\alpha$ CS's, rather than tests, by defining λ as one would for a test, but having it depend also on the parameter that is restricted by the null hypothesis, by enlarging the parameter space Λ correspondingly (so it includes all possible values of the parameter that is restricted by the null hypothesis), and by replacing (a) ϕ_n by a CS based on a sample of size n, (b) α by $1-\alpha$, (c) $RP_n(\lambda)$ by $CP_n(\lambda)$, where $CP_n(\lambda)$ denotes the coverage probability of the CS under λ when the sample size is n, and (d) the first $\limsup_{n\rightarrow\infty} \sup_{\lambda\in\Lambda}$ that appears by $\liminf_{n\rightarrow\infty} \inf_{\lambda\in\Lambda}$. In the present case, where the null hypotheses are of the form $H_0 : \theta = \theta_0$ for some $\theta_0 \in \Theta$, to establish the asymptotic size of CS's, the parameter θ_0 is taken to be a subvector of λ and Λ is specified so that the value of this subvector ranges over Θ.

(ii) In the application of Proposition 10.2 to prove Theorem 10.1, one takes Λ to be a one-to-one transformation of $\mathcal{F}_{AR}$, $\mathcal{F}_2$, or $\mathcal{F}_1$ for tests, and one takes Λ to be a one-to-one transformation of $\mathcal{F}_{\Theta,AR}$, $\mathcal{F}_{\Theta,2}$, or $\mathcal{F}_{\Theta,1}$ for CS's. With these changes, the proofs for tests and CS's are the same. In consequence, we provide explicit proofs for tests only and obtain the proofs for CS's by analogous applications of Proposition 10.2.

(iii) We prove the test results in Theorem 10.1 using Proposition 10.2 by verifying Assumption

B* for a suitable choice of λ, $h_n(\lambda)$, and Λ. The verification of Assumption B* is quite easy for the AR test. It is given in Section 22.6 in the SM. The verifications of Assumption B* for the CQLR$_1$ and CQLR$_2$ tests are much more difficult. In the remainder of this Section 10.1, we provide some key results that are used in doing so. (These results are used only for the CQLR tests, not the AR test.) The complete verifications for the CQLR$_1$ and CQLR$_2$ tests are given in Section 22 in the SM.

10.1.3 General Weight Matrices $\widehat{\mathbf{W}}_{\mathbf{n}}$ and $\widehat{\mathbf{U}}_{\mathbf{n}}$

As above, for notational simplicity, we suppress the dependence on θ_0 of many quantities, such as g_i, G_i, $u_{\theta i}$, B, and f_i, as well as the quantities V_F, Ξ_F, R_F, $\widetilde{V}_F$, and $\widetilde{R}_F$, that are introduced below. To provide asymptotic results for the CQLR$_1$ and CQLR$_2$ tests simultaneously, we prove asymptotic results for a QLR test statistic and a conditioning statistic that depend on general random weight matrices $\widehat{W}_n \in R^{k \times k}$ and $\widehat{U}_n \in R^{p \times p}$. In particular, we consider statistics of the form $\widehat{W}_n \widehat{D}_n \widehat{U}_n$ and functions of this statistic, where $\widehat{D}_n$ is defined in (6.2). Let[44]

$$
\begin{aligned}
QLR_n &:= AR_n - \lambda_{\min}(n\widehat{Q}_{WU,n}), \text{ where} \\
\widehat{Q}_{WU,n} &:= \left(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\right)' \left(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n\right) \in R^{(p+1)\times(p+1)}.
\end{aligned} \tag{10.3}
$$

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

$$
\widehat{W}_n := W_1(\widehat{W}_{2n}) \in R^{k \times k} \text{ and } \widehat{U}_n := U_1(\widehat{U}_{2n}) \in R^{p \times p}, \tag{10.4}
$$

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

$$
W_F := W_1(W_{2F}) \text{ and } U_F := U_1(U_{2F}), \tag{10.5}
$$

respectively.

[44] The definition of $\widehat{Q}_{WUn}$ in (10.3) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$, whereas (6.7) writes the $\lambda_{\min}(\cdot)$ quantity in terms of $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$, which has the $\widehat{\Omega}_n^{-1/2}\widehat{g}_n$ vector as the first column rather than the last column. The ordering of the columns does not affect the value of the $\lambda_{\min}(\cdot)$ quantity. We use the order $(\widehat{\Omega}_n^{-1/2}\widehat{g}_n, \widehat{D}_n^*)$ in (6.7) because it is consistent with the order in Moreira (2003) and Andrews, Moreira, and Stock (2006, 2008). We use the order $(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2}\widehat{g}_n)$ here because it has significant notational advantages in the proof of Theorem 10.5 below, which is given in Section 21 in the SM.

Example 1: For the CQLR$_1$ test, one takes

$$\widehat{W}_n := \widehat{\Omega}_n^{-1/2} \text{ and } \widehat{U}_n := \widehat{L}_n^{1/2} := ((\theta_0, I_p)(\widehat{\Sigma}_n^{\varepsilon})^{-1}(\theta_0, I_p)')^{1/2}, \tag{10.6}$$

where $\widehat{\Omega}_n$ is defined in (5.1) and $\widehat{\Sigma}_n$ is defined in (6.4) and (6.5).

The population analogues of $\widehat{V}_n$ and $\widehat{R}_n$, defined in (6.3), are

$$\begin{aligned} V_F &:= E_F f_i f_i' - E_F((g_i, G_i)'\Xi_F \otimes Z_i Z_i') - E_F(\Xi_F'(g_i, G_i) \otimes Z_i Z_i') \\ &\quad + E_F(\Xi_F' Z_i Z_i' \Xi_F \otimes Z_i Z_i') \in R^{(p+1)k \times (p+1)k} \text{ and} \\ R_F &:= (B' \otimes I_k) V_F (B \otimes I_k) \in R^{(p+1)k \times (p+1)k}, \text{ where} \\ \Xi_F &:= (E_F Z_i Z_i')^{-1} E_F(g_i, G_i) \in R^{k \times (p+1)}, \ f_i := (g_i', vec(G_i)')' \in R^{(p+1)k}, \end{aligned} \tag{10.7}$$

and $B = B(\theta_0)$ is defined in (6.3).

For the CQLR$_1$ test,

$$\begin{aligned} \widehat{W}_{2n} &:= \widehat{\Omega}_n, \ W_{2F} := \Omega_F := E_F g_i g_i',, \ W_1(W_{2F}) := W_{2F}^{-1/2}, \\ \widehat{U}_{2n} &:= (\widehat{\Omega}_n, \widehat{R}_n), \ U_{2F} := (\Omega_F, R_F), \ U_1(U_{2F}) := ((\theta_0, I_p)(\Sigma^{\varepsilon}(\Omega_F, R_F))^{-1}(\theta_0, I_p)')^{1/2}, \text{ and} \\ \Sigma_{j\ell}(\Omega_F, R_F) &= tr(R_{j\ell F}' \Omega_F^{-1})/k \end{aligned} \tag{10.8}$$

for $j, \ell = 1, ..., p+1$, where $\Sigma_{j\ell}(\Omega_F, R_F) \in R^{(p+1)\times(p+1)}$ denotes the (j, ℓ) element $\Sigma(\Omega_F, R_F)$, $\Sigma(\Omega_F, R_F)$ is defined to minimize $||(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - R_F](I_{p+1} \otimes \Omega_F^{-1/2})||$ over symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$ (analogously to the definition of $\widehat{\Sigma}_n(\theta)$ in (6.4)), the last equality in (10.8) holds by the same argument as used to obtain (6.5), $\Sigma^{\varepsilon}(\Omega_F, R_F)$ is defined given $\Sigma(\Omega_F, R_F)$ by (6.6), and $R_{j\ell F}$ denotes the (j, ℓ) $k \times k$ submatrix of R_F.[45]

Example 2: For the CQLR$_2$ test, one takes $\widehat{W}_n$, $\widehat{W}_{2n}$, W_{2F}, and $W_1(\cdot)$ as in Example 1 and

$$\widehat{U}_n := \widetilde{L}_n^{1/2} := ((\theta_0, I_p)(\widetilde{\Sigma}_n^{\varepsilon})^{-1}(\theta_0, I_p)')^{1/2}, \tag{10.9}$$

where $\widetilde{\Sigma}_n$ is defined in Section 7.

The population analogues of $\widetilde{V}_n$ and $\widetilde{R}_n$, defined in (7.1), are

$$\begin{aligned} \widetilde{V}_F &:= E_F(f_i - E_F f_i)(f_i - E_F f_i)' \in R^{(p+1)k \times (p+1)k} \text{ and} \\ \widetilde{R}_F &:= (B' \otimes I_k)\widetilde{V}_F(B \otimes I_k) \in R^{(p+1)k \times (p+1)k}. \end{aligned} \tag{10.10}$$

[45] Note that $W_1(W_{2F})$ and $U_1(U_{2F})$ in (10.8) define the functions $W_1(\cdot)$ and $U_1(\cdot)$ for any conformable arguments, such as $\widehat{W}_{2n}$ and $\widehat{U}_{2n}$, not just for W_{2F} and U_{2F}.

In this case,

$$\widehat{U}_{2n} := (\widehat{\Omega}_n, \widetilde{R}_n),\ U_{2F} := (\Omega_F, \widetilde{R}_F), \tag{10.11}$$

$W_1(\cdot)$ and $U_1(\cdot)$ are as in (10.8), and $\widetilde{R}_n$ is defined in (7.1). We let $\widetilde{\Sigma}_F$ denote $\Sigma(\Omega_F, \widetilde{R}_F)$, which appears in the definition of $U_1(U_{2F})$ in this case. The matrix $\widetilde{\Sigma}_F$ is defined as Σ_F is defined following (10.8) but with $\widetilde{R}_F$ in place of R_F. As defined, $\widetilde{\Sigma}_F$ minimizes $||(I_{p+1} \otimes \Omega_F^{-1/2})[\Sigma \otimes \Omega_F - \widetilde{R}_F](I_{p+1} \otimes \Omega_F^{-1/2})||$ over symmetric pd matrices $\Sigma \in R^{(p+1)\times(p+1)}$.

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

$$\mathcal{F}_{WU} := \{F \in \mathcal{F}_2 : \lambda_{\min}(W_F) \geq \delta_1, \lambda_{\min}(U_F) \geq \delta_1, ||W_F|| \leq M_1, \text{ and } ||U_F|| \leq M_1\} \tag{10.12}$$

for some constants $\delta_1 > 0$ and $M_1 < \infty$, where $\mathcal{F}_2$ is defined in (10.1).

For the CQLR$_1$ test, which uses the definitions in (10.6)-(10.8), we show that $\mathcal{F}_1 \subset \mathcal{F}_{WU}$ for $\delta_1 > 0$ sufficiently small and $M_1 < \infty$ sufficiently large, where $\mathcal{F}_1$ is defined in (10.1), see Lemma 22.4(a) in Section 22.1 in the SM. Hence, uniform results over $\mathcal{F}_1 \cap \mathcal{F}_{WU}$ for arbitrary $\delta_1 > 0$ and $M_1 < \infty$ for this test imply uniform results over $\mathcal{F}_1$.

For the CQLR$_2$ test, which uses the definitions in (10.9)-(10.11), we show that $\mathcal{F}_2 \subset \mathcal{F}_{WU}$ for $\delta_1 > 0$ sufficiently small and $M_1 < \infty$ sufficiently large, see Lemma 22.4(b). Hence, uniform results over $\mathcal{F}_{WU}$ for this test imply uniform results over $\mathcal{F}_2$.

10.1.4 Uniformity Reparametrization

To apply Proposition 10.2, we reparametrize the null distribution F to a vector λ. The vector λ is chosen such that for a subvector of λ convergence of a drifting subsequence of the subvector (after suitable renormalization) yields convergence in distribution of the test statistic and convergence in distribution of the critical value in the case of the CQLR tests. In this section, we define λ for the CQLR tests. Its (much simpler) definition for the AR test is given in Section 22.6 in the SM.

The vector λ depends on the following quantities. Let

$$B_F \text{ denote a } p \times p \text{ orthogonal matrix of eigenvectors of } U_F'(E_F G_i)' W_F' W_F (E_F G_i) U_F \tag{10.13}$$

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

$$C_F \text{ denote a } k \times k \text{ orthogonal matrix of eigenvectors of } W_F(E_F G_i) U_F U_F'(E_F G_i)' W_F'.^{46} \tag{10.14}$$

[46] The matrices B_F and C_F are not uniquely defined. We let B_F denote one choice of the matrix of eigenvectors of

The corresponding eigenvalues are $(\kappa_{1F}, ..., \kappa_{kF}) \in R^k$. Let

$$(\tau_{1F}, ..., \tau_{\min\{k,p\}F}) \text{ denote the } \min\{k,p\} \text{ singular values of } W_F(E_F G_i)U_F, \tag{10.15}$$

which are nonnegative, ordered so that τ_{jF} is nonincreasing. (Some of these singular values may be zero.) As is well-known, the squares of the $\min\{k,p\}$ singular values of a $k \times p$ matrix A equal the $\min\{k,p\}$ largest eigenvalues of $A'A$ and AA'. In consequence, $\kappa_{jF} = \tau_{jF}^2$ for $j = 1, ..., \min\{k,p\}$. In addition, $\kappa_{jF} = 0$ for $j = \min\{k,p\}, ..., \max\{k,p\}$.

Define the elements of λ to be[47,48]

$$\begin{aligned}
&\lambda_{1,F} := (\tau_{1F}, ..., \tau_{\min\{k,p\}F})' \in R^{\min\{k,p\}},\\
&\lambda_{2,F} := B_F \in R^{p\times p},\\
&\lambda_{3,F} := C_F \in R^{k\times k},\\
&\lambda_{4,F} := (E_F G_{i1}, ..., E_F G_{ip}) \in R^{k\times p},\\
&\lambda_{5,F} := E_F \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix} \begin{pmatrix} g_i \\ vec(G_i) \end{pmatrix}' \in R^{(p+1)k\times(p+1)k},\\
&\lambda_{6,F} = (\lambda_{6,1F}, ..., \lambda_{6,(\min\{k,p\}-1)F})' := (\frac{\tau_{2F}}{\tau_{1F}}, ..., \frac{\tau_{\min\{k,p\}F}}{\tau_{(\min\{k,p\}-1)F}})' \in [0,1]^{\min\{k,p\}-1}, \text{ where } 0/0 := 0,\\
&\lambda_{7,F} := W_{2F},\\
&\lambda_{8,F} := U_{2F},\\
&\lambda_{9,F} := F, \text{ and}\\
&\quad \lambda = \lambda_F := (\lambda_{1,F}, ..., \lambda_{9,F}).
\end{aligned} \tag{10.16}$$

The dimensions of W_{2F} and U_{2F} depend on the choices of $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$. We let $\lambda_{5,gF}$ denote the upper left $k \times k$ submatrix of $\lambda_{5,F}$. Thus, $\lambda_{5,gF} = E_F g_i g_i' = \Omega_F$. We consider two parameter spaces for λ: Λ_1 and Λ_2, which correspond to $\mathcal{F}_{WU} \cap \mathcal{F}_1$ and $\mathcal{F}_{WU}$, respectively, where $\mathcal{F}_1$ and $\mathcal{F}_{WU}$ are defined in (10.1) and (10.12), respectively. The space Λ_1 is used for the CQLR_1 test. The space Λ_2 is used for the CQLR_2 test.[49] The parameter spaces Λ_1 and Λ_2 and

$U_F'(E_F G_i)' W_F' W_F (E_F G_i) U_F$ and analogously for C_F.

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

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

[49] Note that the parameter λ has different meanings for the CQLR_1 and CQLR_2 tests because U_{2F} and U_F are different for the two tests.

the function $h_n(\lambda)$ are defined by

$$\begin{aligned}
\Lambda_1 &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU} \cap \mathcal{F}_1\}, \\
\Lambda_2 &:= \{\lambda : \lambda = (\lambda_{1,F}, ..., \lambda_{9,F}) \text{ for some } F \in \mathcal{F}_{WU}\}, \text{ and} \\
h_n(\lambda) &:= (n^{1/2}\lambda_{1,F}, \lambda_{2,F}, \lambda_{3,F}, \lambda_{4,F}, \lambda_{5,F}, \lambda_{6,F}, \lambda_{7,F}, \lambda_{8,F}).
\end{aligned} \tag{10.17}$$

By the definition of $\mathcal{F}_2$, Λ_1 and Λ_2 index distributions that satisfy the null hypothesis $H_0 : \theta = \theta_0$. The dimension J of $h_n(\lambda)$ equals the number of elements in $(\lambda_{1,F}, ..., \lambda_{8,F})$. Redundant elements in $(\lambda_{1,F}, ...\lambda_{8,F})$, such as the redundant off-diagonal elements of the symmetric matrix $\lambda_{5,F}$, are not needed, but do not cause any problem.

We define λ and $h_n(\lambda)$ as in (10.16) and (10.17) because, as shown below, the asymptotic distributions of the test statistics under a sequence $\{F_n : n \geq 1\}$ for which $h_n(\lambda_{F_n}) \to h \in H$ depend on the behavior of $\lim n^{1/2}\lambda_{1,F_n}$, as well as $\lim \lambda_{m,F_n}$ for $m = 2, ..., 8$.

For notational convenience,

$$\{\lambda_{n,h} : n \geq 1\} \text{ denotes a sequence } \{\lambda_n \in \Lambda_2 : n \geq 1\} \text{ for which } h_n(\lambda_n) \to h \in H \tag{10.18}$$

for H defined in (10.2) with Λ equal to Λ_2.[50] By the definitions of Λ_2 and $\mathcal{F}_{WU}$, $\{\lambda_{n,h} : n \geq 1\}$ is a sequence of distributions that satisfies the null hypothesis $H_0 : \theta = \theta_0$.

We decompose h (defined by (10.2), (10.16), and (10.17)) analogously to the decomposition of the first eight components of λ: $h = (h_1, ..., h_8)$, where $\lambda_{m,F}$ and h_m have the same dimensions for $m = 1, ..., 8$. We further decompose the vector h_1 as $h_1 = (h_{1,1}, ..., h_{1,\min\{k,p\}})'$, where the elements of h_1 could equal ∞. We decompose h_6 as $h_6 = (h_{6,1}, ..., h_{6,\min\{k,p\}-1})'$. In addition, we let $h_{5,g}$ denote the upper left $k \times k$ submatrix of h_5. In consequence, under a sequence $\{\lambda_{n,h} : n \geq 1\}$, we have

$$\begin{aligned}
n^{1/2}\tau_{jF_n} &\to h_{1,j} \geq 0 \ \forall j \leq \min\{k,p\}, \ \lambda_{m,F_n} \to h_m \ \forall m = 2, ..., 8, \\
\lambda_{5,gF_n} &= \Omega_{F_n} = E_{F_n} g_i g_i' \to h_{5,g}, \text{ and } \lambda_{6,jF_n} \to h_{6,j} \ \forall j = 1, ..., \min\{k,p\} - 1.
\end{aligned} \tag{10.19}$$

By the conditions in $\mathcal{F}_2$, defined in (10.1), $h_{5,g}$ is pd.

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

10.1.5 Assumption WU

We assume that the random weight matrices $\widehat{W}_n = W_1(\widehat{W}_{2n})$ and $\widehat{U}_n = U_1(\widehat{U}_{2n})$ defined in (10.4) satisfy the following assumption that depends on a suitably chosen parameter space Λ_* ($\subset \Lambda_2$), such as Λ_1 or Λ_2.

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

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

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

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

In Assumption WU and elsewhere below, "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$" means "all sequences $\{\lambda_{w_n,h} : n \geq 1\}$ for any $h \in H$," where H is defined in (10.2) with Λ equal to Λ_2, and likewise with n in place of w_n.

Assumption WU for the parameter spaces Λ_1 and Λ_2 is verified in Lemma 22.4 in Section 22 in the SM for the CQLR_1 and CQLR_2 tests, respectively.

10.1.6 Asymptotic Distributions

This section provides the asymptotic distributions of QLR test statistics and corresponding conditioning statistics that are used in the proof of Theorem 10.1 to verify Assumption B* of Proposition 10.2.

For any $F \in \mathcal{F}_2$, define

$$\Phi_F^{vec(G_i)} := Var_F(vec(G_i) - (E_F vec(G_\ell) g_\ell') \Omega_F^{-1} g_i) \text{ and } \Phi_h^{vec(G_i)} := \lim \Phi_{F_{w_n}}^{vec(G_i)} \qquad (10.20)$$

whenever the limit exists, where the distributions $\{F_{w_n} : n \geq 1\}$ correspond to $\{\lambda_{w_n,h} : n \geq 1\}$ for any subsequence $\{w_n : n \geq 1\}$. The assumptions allow $\Phi_h^{vec(G_i)}$ to be singular.

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

$$\begin{pmatrix} \overline{g}_h \\ vec(\overline{D}_h) \end{pmatrix} \sim N\left(0^{(p+1)k}, \begin{pmatrix} h_{5,g} & 0^{k \times pk} \\ 0^{pk \times k} & \Phi_h^{vec(G_i)} \end{pmatrix}\right), \qquad (10.21)$$

where $\overline{g}_h \in R^k$ and $\overline{D}_h \in R^{k\times p}$ are independent by the definition of $\widehat{D}_n$, see Lemma 10.3 below.[51]

To determine the asymptotic distributions of the QLR_{1n} and QLR_{2n} statistics (defined in (6.7) and just below (7.2)) and the conditional critical value of the CQLR tests (defined in (3.5), (6.8), and (7.3)), we need to determine the asymptotic distribution of $W_{F_n}\widehat{D}_n U_{F_n}$ without recentering by $E_{F_n}G_i$. To do so, we post-multiply $W_{F_n}\widehat{D}_n U_{F_n}$ first by B_{F_n} and then by a nonrandom diagonal matrix $S_n \in R^{p\times p}$ (which may depend on F_n and h). The matrix S_n rescales the columns of $W_{F_n}\widehat{D}_n U_{F_n} B_{F_n}$ to ensure that $n^{1/2}W_{F_n}\widehat{D}_n U_{F_n} B_{F_n} S_n$ converges in distribution to a (possibly) random matrix that is finite a.s. and not a.s. zero.

The following is an important definition for the scaling matrix S_n and asymptotic distributions given below. Consider a sequence $\{\lambda_{n,h} : n \geq 1\}$. Let $q = q_h$ $(\in \{0, ..., \min\{k, p\}\})$ be such that

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq \min\{k, p\}, \tag{10.22}$$

where $h_{1,j} := \lim n^{1/2}\tau_{jF_n} \geq 0$ for $j = 1, ..., \min\{k, p\}$ by (10.19) and the distributions $\{F_n : n \geq 1\}$ correspond to $\{\lambda_{n,h} : n \geq 1\}$ defined in (10.18). This value q exists because $\{h_{1,j} : j \leq \min\{k, p\}\}$ are nonincreasing in j (since $\{\tau_{jF} : j \leq \min\{k, p\}\}$ are nonincreasing in j, as defined in (10.15)). Note that q is the number of singular values of $W_{F_n}(E_{F_n}G_i)U_{F_n}$ that diverge to infinity when multiplied by $n^{1/2}$. Heuristically, q is the maximum number of parameters, or one-to-one transformations of the parameters, that are strongly or semi-strongly identified. (That is, one could partition θ, or a one-to-one transformation of θ, into subvectors of dimension q and $p - q$ such that if the $p - q$ subvector was known and, hence, was no longer part of the parameter, then the q subvector would be strongly or semi-strongly identified in the sense used in this paper.)

Let

$$S_n := Diag\{(n^{1/2}\tau_{1F_n})^{-1}, ..., (n^{1/2}\tau_{qF_n})^{-1}, 1, ..., 1\} \in R^{p\times p} \text{ and } T_n := B_{F_n}S_n \in R^{p\times p}, \tag{10.23}$$

where $q = q_h$ is defined in (10.22). Note that S_n is well defined for n large, because $n^{1/2}\tau_{jF_n} \to \infty$ for all $j \leq q$.

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

[51] If one eliminates the $\lambda_{\min}(E_F g_i g_i') \geq \delta$ condition in $\mathcal{F}_2$ and one defines $\widehat{D}_n$ in (6.2) with $\widehat{\Omega}_n$ replaced by the eigenvalue-adjusted matrix $\widehat{\Omega}_n^{\varepsilon}$ for some $\varepsilon > 0$, then the asymptotic distribution in (10.21) still holds, but without the independence of $\overline{g}_h$ and $\overline{D}_h$. However, this independence is key. Without it, the conditioning argument that is used to establish the correct asymptotic size of the CQLR$_1$ and CQLR$_2$ tests does not go through. Thus, we define $\widehat{D}_n$ in (6.2) using $\widehat{\Omega}_n$, not $\widehat{\Omega}_n^{\varepsilon}$.

as follows:

$$
\begin{aligned}
h_2 &= (h_{2,q}, h_{2,p-q}),\ h_3 = (h_{3,q}, h_{3,k-q}), \\
h^{\diamond}_{1,p-q} &:= \begin{bmatrix} 0^{q\times(p-q)} \\ Diag\{h_{1,q+1}, ..., h_{1,p}\} \\ 0^{(k-p)\times(p-q)} \end{bmatrix} \in R^{k\times(p-q)} \text{ if } k \geq p, \\
h^{\diamond}_{1,p-q} &:= \begin{bmatrix} 0^{q\times(k-q)} & 0^{q\times(p-k)} \\ Diag\{h_{1,q+1}, ..., h_{1,k}\} & 0^{(k-q)\times(p-k)} \end{bmatrix} \in R^{k\times(p-q)} \text{ if } k < p, \\
\overline{\Delta}_h &= (\overline{\Delta}_{h,q}, \overline{\Delta}_{h,p-q}) \in R^{k\times p},\ \overline{\Delta}_{h,q} := h_{3,q},\ \overline{\Delta}_{h,p-q} := h_3 h^{\diamond}_{1,p-q} + h_{71}\overline{D}_h h_{81} h_{2,p-q}, \\
h_{71} &:= W_1(h_7),\ \text{and } h_{81} := U_1(h_8), \qquad (10.24)
\end{aligned}
$$

where $h_{2,q} \in R^{p\times q}$, $h_{2,p-q} \in R^{p\times(p-q)}$, $h_{3,q} \in R^{k\times q}$, $h_{3,k-q} \in R^{k\times(k-q)}$, $\overline{\Delta}_{h,q} \in R^{k\times q}$, $\overline{\Delta}_{h,p-q} \in R^{k\times(p-q)}$, $h_{71} \in R^{k\times k}$, and $h_{81} \in R^{p\times p}$.[52] Note that when Assumption WU holds $h_{71} = \lim W_{F_n} = \lim W_1(W_{2F_n})$ and $h_{81} = \lim U_{F_n} = \lim U_1(U_{2F_n})$ under $\{\lambda_{n,h} : n \geq 1\}$.

The following lemma allows for $k \geq p$ and $k < p$. For the case where $k \geq p$, it appears in the Appendix to AG1 as Lemma 8.3.

Lemma 10.3 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

where (a) $(\overline{g}_h, \overline{D}_h)$ *are defined in* (10.21), (b) $\overline{\Delta}_h$ *is the nonrandom function of* h *and* $\overline{D}_h$ *defined in* (10.24), (c) $(\overline{D}_h, \overline{\Delta}_h)$ *and* $\overline{g}_h$ *are independent, and* (d) *under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$, *the convergence result above and the results of parts* (a)-(c) *hold with* n *replaced with* w_n.

Comments: (i) Lemma 10.3(c) is a key property that leads to the correct asymptotic size of the CQLR_1 and CQLR_2 tests.

(ii) Lemma 8.3 in the Appendix to AG1 contains a part (part (d)), which does not appear in Lemma 10.3. It states that $\overline{\Delta}_h$ has full column rank a.s. under some additional conditions. For Kleibergen's (2005) LM statistic and Kleibergen's (2005) CLR statistics that employ it, which are considered in AG1, one needs the (possibly) random limit matrix of $n^{1/2}W_{F_n}\widehat{D}_n U_{F_n} B_{F_n} S_n$, viz., $\overline{\Delta}_h$, to have full column rank with probability one, in order to apply the continuous mapping theorem

[52] There is some abuse of notation here. E.g., $h_{2,q}$ and $h_{2,p-q}$ denote different matrices even if $p-q$ happens to equal q.

(CMT), which is used to determine the asymptotic distribution of the test statistics. To obtain this full column rank property, AG1 restricts the parameter space for the tests based on aforementioned statistics to be a subset $\mathcal{F}_0$ of $\mathcal{F}_2$, where $\mathcal{F}_0$ is defined in Section 3 of AG1. In contrast, the QLR_{1n} and QLR_{2n} statistics considered here do not depend on Kleibergen's LM statistic and do not require the asymptotic distribution of $n^{1/2}W_{F_n}\widehat{D}_nU_{F_n}B_{F_n}S_n$ to have full column rank a.s. In consequence, it is not necessary to restrict the parameter space from $\mathcal{F}_2$ to $\mathcal{F}_0$ when considering these statistics.

Let

$$\widehat{\kappa}_{jn} \text{ denote the } j\text{th eigenvalue of } n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n,\ \forall j=1,...,p, \tag{10.25}$$

ordered to be nonincreasing in j. The jth singular value of $n^{1/2}\widehat{W}_n\widehat{D}_n\widehat{U}_n$ equals $\widehat{\kappa}_{jn}^{1/2}$ for $j = 1,...,\min\{k,p\}$.

The following proposition, combined with Lemma 6.1, is used to determine the asymptotic behavior of the data-dependent conditional critical values of the CQLR$_1$ and CQLR$_2$ tests. The proposition is the same as Theorem 8.4(c)-(f) in the Appendix to AG1, except that it is extended to cover the case $k < p$, not just $k \geq p$. For brevity, the proof of the proposition given in Section 20 in the SM just describes the changes needed to the proof of Theorem 8.4(c)-(f) of AG1 in order to cover the case $k < p$. The proof of Theorem 8.4(c)-(f) in AG1 is similar to, but simpler than, the proof of Theorem 10.5 below, which is given in Section 21 in the SM.

Proposition 10.4 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

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

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

(c) *the convergence in parts* (a) *and* (b) *holds jointly with the convergence in Lemma* 10.3, *and*

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

Comment: Proposition 10.4(a) and (b) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = \widehat{L}_n^{1/2}$ is used to determine the asymptotic behavior of the critical value function for the CQLR$_1$ test, which depends on $n^{1/2}\widehat{D}_n^*$ defined in (6.7), see the proof of Theorem 22.1 in Section 22.2 in the SM. Proposition 10.4(a) and (b) with $\widehat{W}_n = \widehat{\Omega}_n^{-1/2}$ and $\widehat{U}_n = \widetilde{L}_n^{1/2}$ is used to determine the asymptotic behavior of the critical value function for the CQLR$_2$ test, which depends on $n^{1/2}\widetilde{D}_n^*$ defined in (7.2), see the proof of Theorem 22.1 in Section 22.2 in the SM.

The next theorem provides the asymptotic distribution of the general QLR_n statistic defined in (10.3) and, as special cases, those of the QLR_{1n} and QLR_{2n} statistics.

Theorem 10.5 *Suppose Assumption* WU *holds for some non-empty parameter space* $\Lambda_* \subset \Lambda_2$. *Under all sequences* $\{\lambda_{n,h} : n \geq 1\}$ *with* $\lambda_{n,h} \in \Lambda_*$,

$$QLR_n \to_d \overline{g}_h' h_{5,g}^{-1} \overline{g}_h - \lambda_{\min}((\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h))$$

and the convergence holds jointly with the convergence in Lemma 10.3 *and Proposition* 10.4. *When* $q = p$ *(which can only hold if* $k \geq p$ *because* $q \leq \min\{k, p\}$*),* $\overline{\Delta}_{h,p-q}$ *does not appear in the limit random variable and the limit random variable reduces to* $(h_{5,g}^{-1/2} \overline{g}_h)' h_{3,p} h_{3,p}' h_{5,g}^{-1/2} \overline{g}_h \sim \chi_p^2$. *When* $q = k$ *(which can only hold if* $k \leq p$*), the* $\lambda_{\min}(\cdot)$ *expression does not appear in the limit random variable and the limit random variable reduces to* $\overline{g}_h' h_{5,g}^{-1} \overline{g}_h \sim \chi_k^2$. *When* $k \leq p$ *and* $q < k$*, the* $\lambda_{\min}(\cdot)$ *expression equals zero and the limit random variable reduces to* $\overline{g}_h' h_{5,g}^{-1} \overline{g}_h \sim \chi_k^2$. *Under all subsequences* $\{w_n\}$ *and all sequences* $\{\lambda_{w_n,h} : n \geq 1\}$ *with* $\lambda_{w_n,h} \in \Lambda_*$*, the same results hold with* n *replaced with* w_n.

Comments: (i) Theorem 10.5 gives the asymptotic distributions of the QLR_{1n} and QLR_{2n} statistics (defined by (6.7) and (7.2)) once it is verified that the choices of $(\widehat{W}_n, \widehat{U}_n)$ for these statistics satisfy Assumption WU for the parameter spaces Λ_1 and Λ_2, respectively. The latter is done in Lemma 22.4 in Section 22.1 in the SM.

(ii) When $q = p$, the parameter θ_0 is strongly or semi-strongly identified and Theorem 10.5 shows that the QLR_n statistic has a χ_p^2 asymptotic null distribution.

(iii) When $k = p$, Theorem 10.5 shows that the QLR_n statistic has a χ_k^2 asymptotic null distribution regardless of the strength of identification.

(iv) When $k < p$, θ is necessarily unidentified and Theorem 10.5 shows that the asymptotic null distribution of QLR_n is χ_k^2.

(v) The proof of Theorem 10.5 given in Section 21 in the SM also shows that the largest q eigenvalues of $n(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)'(\widehat{W}_n \widehat{D}_n \widehat{U}_n, \widehat{\Omega}_n^{-1/2} \widehat{g}_n)$ diverge to infinity in probability and the (ordered) vector of the smallest $p+1-q$ eigenvalues of this matrix converges in distribution to the (ordered) vector of the $p+1-q$ eigenvalues of $(\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)' h_{3,k-q} \times h_{3,k-q}' (\overline{\Delta}_{h,p-q}, h_{5,g}^{-1/2} \overline{g}_h)$.

Propositions 10.2 and 10.4 and Theorem 10.5 are used to prove Theorem 10.1. The proof is given in Section 22 in the SM. Note, however, that the proof is not a straightforward implication of these results. The proof also requires (i) determining the behavior of the conditional critical value function $c_{k,p}(D, 1-\alpha)$, defined in the paragraph containing (3.5), for sequences of nonrandom $k \times p$ matrices

$\{D_n : n \geq 1\}$ whose singular values may converge or diverge to infinity at any rates, (ii) showing that the distribution function of the asymptotic distribution of the QLR_n statistic, conditional on the asymptotic version of the conditioning statistic, is continuous and strictly increasing at its $1-\alpha$ quantile for all possible (k, p, q) values and all possible limits of the scaled population singular values $\{n^{1/2}\tau_{jF_n} : n \geq 1\}$ for $j = 1, ..., \min\{k, p\}$, and (iii) establishing that Assumption WU holds for the CQLR$_1$ and CQLR$_2$ tests. These results are established in Lemmas 22.2, 22.3, and 22.4, respectively, in Section 22 in the SM.

10.2 Singularity-Robust Tests

In this section, we prove the main Theorem 8.1 for the SR tests using Theorem 10.1 for the tests without the SR extension. The SR-AR and SR-CQLR tests, defined in (5.7), (6.12), and (7.5), depend on the random variable $\widehat{r}_n(\theta)$ and random matrices $\widehat{A}_n(\theta)$ and $\widehat{A}_n^{\perp}(\theta)$, defined in (5.3) and (5.4). First, in the following lemma, we show that with probability that goes to one as $n \to \infty$ (wp$\to$1), the SR test statistics and data-dependent critical values are the same as when the non-random and rescaled population quantities $r_F(\theta)$ and $\Pi_{1F}^{-1/2}(\theta)A_F(\theta)'$ are used to define these statistics, rather than $\widehat{r}_n(\theta)$ and $\widehat{A}_n(\theta)'$, where $r_F(\theta)$, $A_F(\theta)$, and $\Pi_{1F}(\theta)$ are defined as in (4.7) and (4.8). The lemma also shows that the extra rejection condition in (5.7), (6.12), and (7.5) fails to hold wp$\to$ 1 under all sequences of null distributions.

In the following lemma, θ_{0n} is the true value that may vary with n (which is needed for the CS results) and col$(\cdot)$ denotes the column space of a matrix.

Lemma 10.6 *For any sequence* $\{(F_n, \theta_{0n}) \in \mathcal{F}_{\Theta,AR}^{SR} : n \geq 1\}$, (a) $\widehat{r}_n(\theta_{0n}) = r_{F_n}(\theta_{0n})$ wp$\to$1, (b) col$(\widehat{A}_n(\theta_{0n})) =$ col$(A_{F_n}(\theta_{0n}))$ wp$\to$1, (c) *the statistics* $SR\text{-}AR_n(\theta_{0n})$, $SR\text{-}QLR_{1n}(\theta_{0n})$, $SR\text{-}QLR_{2n}(\theta_{0n})$, $c_{\widehat{r}_n(\theta_{0n}),p}(n^{1/2}\widehat{D}_{An}^{*}(\theta_{0n}), 1-\alpha)$, *and* $c_{\widehat{r}_n(\theta_{0n}),p}(n^{1/2}\widetilde{D}_{An}^{*}(\theta_{0n}), 1-\alpha)$ *are invariant* wp$\to$1 *to the replacement of* $\widehat{r}_n(\theta_{0n})$ *and* $\widehat{A}_n(\theta_{0n})'$ *by* $r_{F_n}(\theta_{0n})$ *and* $\Pi_{1F_n}^{-1/2}(\theta_{0n})A_{F_n}(\theta_{0n})'$, *respectively, and* (d) $\widehat{A}_n^{\perp}(\theta_{0n})'\widehat{g}_n(\theta_{0n}) = 0^{k-\widehat{r}_n(\theta_{0n})}$ wp$\to$1, *where this equality is defined to hold when* $\widehat{r}_n(\theta_{0n}) = k$.

Proof of Lemma 10.6. For notational simplicity, we suppress the dependence of various quantities on θ_{0n}. By considering subsequences, it suffices to consider the case where $r_{F_n} = r$ for all $n \geq 1$ for some $r \in \{0, 1, ..., k\}$.

First, we establish part (a). We have $\widehat{r}_n \leq r$ a.s. for all $n \geq 1$ because for any constant vector $\lambda \in R^k$ for which $\lambda'\Omega_{F_n}\lambda = 0$, we have $\lambda' g_i = 0$ a.s.$[F_n]$ and $\lambda'\widehat{\Omega}_n\lambda = n^{-1}\sum_{i=1}^{n}(\lambda' g_i)^2 - (\lambda'\widehat{g}_n)^2 = 0$ a.s.$[F_n]$, where a.s.$[F_n]$ means "with probability one under F_n." This completes the proof of part (a) when $r = 0$. Hence, for the rest of the proof of part (a), we assume $r > 0$.

We have $\widehat{r}_n := rk(\widehat{\Omega}_n) \geq rk(\Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2})$ because $\widehat{\Omega}_n$ is $k \times k$, $A_{F_n} \Pi_{1F_n}^{-1/2}$ is $k \times r$, and $1 \leq r \leq k$. In addition, we have

$$\begin{aligned}
\Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2} &= n^{-1} \sum_{i=1}^n (\Pi_{1F_n}^{-1/2} A'_{F_n} g_i)(\Pi_{1F_n}^{-1/2} A'_{F_n} g_i)' \\
&\quad - (n^{-1} \sum_{i=1}^n \Pi_{1F_n}^{-1/2} A'_{F_n} g_i)(n^{-1} \sum_{i=1}^n \Pi_{1F_n}^{-1/2} A'_{F_n} g_i)', \\
E_{F_n} (\Pi_{1F_n}^{-1/2} A'_{F_n} g_i)(\Pi_{1F_n}^{-1/2} A'_{F_n} g_i)' &= \Pi_{1F_n}^{-1/2} A'_{F_n} \Omega_{F_n} A_{F_n} \Pi_{1F_n}^{-1/2} \\
&= \Pi_{1F_n}^{-1/2} A'_{F_n} A^{\dagger}_{F_n} \Pi_{F_n} A^{\dagger\prime}_{F_n} A_{F_n} \Pi_{1F_n}^{-1/2} = I_r, \qquad (10.26)
\end{aligned}$$

and $E_{F_n} \Pi_{1F_n}^{-1/2} A'_{F_n} g_i = 0^r$, where the second last equality in (10.26) holds by the spectral decomposition in (4.7) and the last equality in (10.26) holds by the definitions of $A_F^{\dagger}$, A_F, and Π_{1F} in (4.7) and (4.8). By (10.26), the moment conditions in $\mathcal{F}_2^{SR}$, and the weak law of large numbers for $L^{1+\gamma/2}$-bounded i.i.d. random variables for $\gamma > 0$, we obtain $\Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2} \to_p I_r$. In consequence, $rk(\Pi_{1F_n}^{-1/2} A'_{F_n} \widehat{\Omega}_n A_{F_n} \Pi_{1F_n}^{-1/2}) \geq r$ wp$\to$1, which concludes the proof that $\widehat{r}_n = r$ wp$\to$1.[53]

Next, we prove part (b). Let $N(\cdot)$ denotes the null space of a matrix. We have

$$\begin{aligned}
\lambda \in N(\Omega_{F_n}) &\Longrightarrow \lambda' \Omega_{F_n} \lambda = 0 \Longrightarrow Var_{F_n}(\lambda' g_i) = 0 \Longrightarrow \lambda' g_i = 0 \text{ a.s.}[F_n] \\
&\Longrightarrow \widehat{\Omega}_n \lambda = 0 \text{ a.s.}[F_n] \Longrightarrow \lambda \in N(\widehat{\Omega}_n) \text{ a.s.}[F_n]. \qquad (10.27)
\end{aligned}$$

That is, $N(\Omega_{F_n}) \subset N(\widehat{\Omega}_n)$ a.s.$[F_n]$. This and $rk(\Omega_{F_n}) = rk(\widehat{\Omega}_n)$ wp$\to$1 imply that $N(\Omega_{F_n}) = N(\widehat{\Omega}_n)$ wp$\to$1 (because if $N(\widehat{\Omega}_n)$ is strictly larger than $N(\Omega_{F_n})$ then the dimension and rank of $\widehat{\Omega}_n$ must exceed the dimension and rank of $N(\Omega_{F_n})$, which is a contradiction). In turn, $N(\Omega_{F_n}) = N(\widehat{\Omega}_n)$ wp$\to$1 implies that $\mathrm{col}(\widehat{A}_n) = \mathrm{col}(A_{F_n})$ wp$\to$1, which proves part (b).

To prove part (c), it suffices to consider the case where $r \geq 1$ because the test statistics and their critical values are all equal to zero by definition when $\widehat{r}_n = 0$ and $\widehat{r}_n = 0$ wp$\to$1 when $r = 0$ by part (a). Part (b) of the Lemma implies that there exists a random $r \times r$ nonsingular matrix

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

$\widehat{M}_n$ such that

$$\widehat{A}_n = A_{F_n}\Pi_{1F_n}^{-1/2}\widehat{M}_n \text{ wp} \to 1, \tag{10.28}$$

because Π_{1F_n} is nonsingular (since it is a diagonal matrix with the positive eigenvalues of Ω_{F_n} on its diagonal by its definition following (4.8)). Equation (10.28) and $\widehat{r}_n = r$ wp→1 imply that the statistics $SR\text{-}AR_n$, $SR\text{-}QLR_{1n}$, $SR\text{-}QLR_{2n}$, $c_{\widehat{r}_n,p}(n^{1/2}\widehat{D}^*_{An}, 1-\alpha)$, and $c_{\widehat{r}_n,p}(n^{1/2}\widetilde{D}^*_{An}, 1-\alpha)$ are invariant wp→1 to the replacement of $\widehat{r}_n$ and $\widehat{A}_n'$ by r and $\widehat{M}_n'\Pi_{1F_n}^{-1/2}A_{F_n}'$, respectively. Now we apply the invariance result of Lemma 6.2 with (k, g_i, G_i) replaced by $(r, \Pi_{1F_n}^{-1/2}A_{F_n}'g_i, \Pi_{1F_n}^{-1/2}A_{F_n}'G_i)$ and with M equal to $\widehat{M}_n'$. (The extension of Lemma 6.2 to cover the statistics employed by the CQLR$_2$ test is stated in a footnote in Section 7.) This result implies that the previous five statistics when based on r and $\Pi_{1F_n}^{-1/2}A_{F_n}'g_i$ are invariant to the multiplication of the moments $\Pi_{1F_n}^{-1/2}A_{F_n}'g_i$ by the nonsingular matrix $\widehat{M}_n'$. Thus, these five statistics, defined as in Sections 6.2 and 7, are invariant wp→1 to the replacement of $\widehat{r}_n$ and $\widehat{A}_n'$ by r and $\Pi_{1F_n}^{-1/2}A_{F_n}'$, respectively.

Lastly, we prove part (d). The equality $(\widehat{A}_n^\perp)'\widehat{g}_n = 0^{k-\widehat{r}_n}$ holds by definition when $\widehat{r}_n = k$ (see the statement of Lemma 10.6(d)) and $\widehat{r}_n = r$ wp→1. Hence, it suffices to consider the case where $r \in \{0, ..., k-1\}$. For all $n \geq 1$, we have $E_{F_n}(A_{F_n}^\perp)'\widehat{g}_n = 0^{k-r}$ and

$$nVar_{F_n}((A_{F_n}^\perp)'\widehat{g}_n) = (A_{F_n}^\perp)'\Omega_{F_n}A_{F_n}^\perp = (A_{F_n}^\perp)'A_{F_n}^\dagger\Pi_{F_n}(A_{F_n}^\dagger)'A_{F_n}^\perp = 0^{(k-r)\times(k-r)}, \tag{10.29}$$

where the second equality uses the spectral decomposition in (4.7) and the last equality uses $A_n^\dagger = [A_F, A_F^\perp]$, see (4.8). In consequence, $(A_{F_n}^\perp)'\widehat{g}_n = 0^{k-r}$ a.s. This and and the result of part (b) that $\text{col}(\widehat{A}_n^\perp) = \text{col}(A_{F_n}^\perp)$ wp→1 establish part (d). □

Given Lemma 10.6(d), the extra rejection conditions in the SR-AR and SR-CQLR tests and CS's (i.e., the second conditions in (5.7), (5.9), (6.12), (7.5), and in the SR-CQLR CS definitions following (6.12) and (7.5)) can be ignored when computing the asymptotic size properties of these tests and CS's (because the condition fails to hold for each test wp→1 under any sequence of null hypothesis values for any sequence of distributions in the null hypotheses, and the condition holds for each CS wp→1 under any sequence of true values θ_{0n} for any sequence of distributions for which the moment conditions hold at θ_{0n}).

Given Lemma 10.6(c), the asymptotic size properties of the SR-AR and SR-CQLR tests and CS's can be determined by the analogous tests and CS's that are based on $r_{F_n}(\theta_0)$ and $\Pi_{1F_n}^{-1/2}(\theta_0)A_{F_n}(\theta_0)'$ (for fixed θ_0 with tests and for any $\theta_0 \in \Theta$ with CS's). For the tests, we do so by partitioning $\mathcal{F}_{AR}^{SR}$, $\mathcal{F}_2^{SR}$, and $\mathcal{F}_1^{SR}$ into k sets based on the value of $rk(\Omega_F(\theta_0))$ and establishing the correct asymptotic size and asymptotic similarity of the analogous tests separately for each parameter space. That is, we write $\mathcal{F}_{AR}^{SR} = \cup_{r=0}^k \mathcal{F}_{AR[r]}^{SR}$, where $\mathcal{F}_{AR[r]}^{SR} := \{F \in \mathcal{F}_{AR}^{SR} : rk(\Omega_F(\theta_0)) = r\}$, and establish the desired results for $\mathcal{F}_{AR[r]}^{SR}$ separately for each r. Analogously, we write $\mathcal{F}_2^{SR} = \cup_{r=0}^{k} \mathcal{F}_{2[r]}^{SR}$ and $\mathcal{F}_1^{SR} = \cup_{r=0}^{k} \mathcal{F}_{1[r]}^{SR}$, where $\mathcal{F}_{2[r]}^{SR} := \mathcal{F}_{AR[r]}^{SR} \cap \mathcal{F}_2^{SR}$ and $\mathcal{F}_{1[r]}^{SR} := \mathcal{F}_{AR[r]}^{SR} \cap \mathcal{F}_1^{SR}$. Note that we do not need to consider the parameter space $\mathcal{F}_{AR[r]}^{SR}$ for $r = 0$ for the SR-AR test when determining the asymptotic size of the SR-AR test because the test fails to reject H_0 wp$\rightarrow$1 based on the first condition in (5.7) when $r = 0$ (since the test statistic and critical value equal zero by definition when $\widehat{r}_n = 0$ and $\widehat{r}_n = r = 0$ wp$\rightarrow$1 by Lemma 10.6(a)). In addition, we do not need consider the parameter space $\mathcal{F}_{AR[r]}^{SR}$ for $r = 0$ for the SR-AR test when determining the asymptotic similarity of the test because such distributions are excluded from the parameter space $\mathcal{F}_{AR}^{SR}$ by the statement of Theorem 8.1. Analogous arguments regarding the parameter spaces corresponding to $r = 0$ apply to the other tests and CS's. Hence, from here on, we assume $r \in \{1, ..., k\}$.

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

$$g_{Fi}^* := \Pi_{1F}^{-1/2} A_F' g_i \text{ and } G_{Fi}^* := \Pi_{1F}^{-1/2} A_F' G_i, \tag{10.30}$$

where $A_F \in R^{k \times r}$, $\Pi_{1F} \in R^{r \times r}$, and dependence on θ_0 is suppressed for notational simplicity. Given the conditions in $\mathcal{F}_2^{SR}$, we have

$$\begin{aligned} E_F||g_{Fi}^*||^{2+\gamma} &= E_F||\Pi_{1F}^{-1/2} A_F' g_i||^{2+\gamma} \leq M, \\ E_F||vec(G_{Fi}^*)||^{2+\gamma} &= E_F||vec(\Pi_{1F}^{-1/2} A_F' G_i)||^{2+\gamma} \leq M, \\ \lambda_{\min}(E_F g_{Fi}^* g_{Fi}^{*\prime}) &= \lambda_{\min}(\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2}) = \lambda_{\min}(I_r) = 1, \end{aligned} \tag{10.31}$$

and $E_F g_{Fi}^* = 0^r$, where the second equality in the third line of (10.31) holds by the spectral decomposition in (4.7) and the partition $A_F^{\dagger} = [A_F, A_F^{\perp}]$ in (4.8). Thus, $F \in \mathcal{F}_{2[r]}^{SR}$ for (g_i, G_i) implies that $F \in \mathcal{F}_2$ with $\delta \leq 1$ for (g_{Fi}^*, G_{Fi}^*), where the definition of $\mathcal{F}_2$ in (10.1) is extended to allow g_i and G_i to depend on F. Now we apply Theorem 10.1 with (g_{Fi}^*, G_{Fi}^*) and r in place of (g_i, G_i) and k and with $\delta \leq 1$, to obtain the correct asymptotic size and asymptotic similarity of the SR-CQLR$_2$ test for the parameter space $\mathcal{F}_{2[r]}^{SR}$ for $r = 1, ..., k$. This requires that Theorem 10.1 holds for $k < p$, which it does. The fact that g_{Fi}^* and G_{Fi}^* depend on F, whereas g_i and G_i do not, does not cause a problem, because the proof of Theorem 10.1 goes through as is if g_i and G_i depend on F. This establishes the results of Theorem 8.1 for the SR-CQLR$_2$ test. The proof for the SR-CQLR$_2$ CS is essentially the same, but with θ_0 taking any value in Θ and with $\mathcal{F}_{\Theta,2}^{SR}$ and $\mathcal{F}_{\Theta,2}$, defined in (4.10) and just below (10.1), in place of $\mathcal{F}_2^{SR}$ and $\mathcal{F}_2$, respectively.

The proof for the SR-AR test and CS is the same as that for the SR-CQLR$_2$ test and CS, but with $vec(G_{Fi}^*)$ deleted in (10.31) and with the subscript 2 replaced by AR on the parameter spaces that appear.

Next, we consider the SR-CQLR$_1$ test. When the moment functions satisfy (4.4), i.e., $g_i = u_i Z_i$, we define $Z^*_{Fi} := \Pi_{1F}^{-1/2} A_F' Z_i$, $g^*_{Fi} = u_i Z^*_{Fi}$, and $G^*_{Fi} = Z^*_{Fi} u'_{\theta i}$, where $u_{\theta i}$ is defined in (4.5) and the dependence of various quantities on θ_0 is suppressed. In this case, by the conditions in $\mathcal{F}_1^{SR}$, the IV's Z^*_{Fi} satisfy $E_F||Z^*_{Fi}||^{4+\gamma} = E_F||\Pi_{1F}^{-1/2} A_F' Z_i||^{4+\gamma} \leq M$ and $E_F||u^*_i||^{2+\gamma} \leq M$, where $u^*_i := (u_i, u'_{\theta i})'$. Next we show that $\lambda_{\min}(E_F Z^*_{Fi} Z^{*\prime}_{Fi})$ is bounded away from zero for $F \in \mathcal{F}^{SR}_{1[r]}$. We have

$$
\begin{aligned}
\lambda_{\min}(E_F Z^*_{Fi} Z^{*\prime}_{Fi}) &= \lambda_{\min}(E_F \Pi_{1F}^{-1/2} A_F' Z_i Z_i' A_F \Pi_{1F}^{-1/2}) \\
&= \inf_{\lambda \in R^r: ||\lambda||=1} [E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 1(u_i^2 \leq c) + E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 1(u_i^2 > c)] \\
&\geq \inf_{\lambda \in R^r: ||\lambda||=1} [c^{-1} E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 u_i^2 1(u_i^2 \leq c)] \\
&= c^{-1} \inf_{\lambda \in R^r: ||\lambda||=1} [E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 u_i^2 - E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 u_i^2 1(u_i^2 > c)] \\
&\geq c^{-1} [\lambda_{\min}(\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2}) - \sup_{\lambda \in R^r: ||\lambda||=1} E_F(\lambda' \Pi_{1F}^{-1/2} A_F' Z_i)^2 u_i^2 1(u_i^2 > c)] \\
&\geq c^{-1} [1 - E_F||\Pi_{1F}^{-1/2} A_F' Z_i||^2 u_i^2 1(u_i^2 > c)] \\
&\geq 1/(2c), \qquad (10.32)
\end{aligned}
$$

where the second inequality uses $g_i = Z_i u_i$ and $\Omega_F := E_F g_i g_i'$, the third inequality holds by $\Pi_{1F}^{-1/2} A_F' \Omega_F A_F \Pi_{1F}^{-1/2} = I_r$ (using (4.7) and (4.8)) and by the Cauchy-Bunyakovsky-Schwarz inequality applied to $\lambda' \Pi_{1F}^{-1/2} A_F' Z_i$, and the last inequality holds by the condition $E_F||\Pi_{1F}^{-1/2} A_F' Z_i||^2 u_i^2 \times 1(u_i^2 > c) \leq 1/2$ in $\mathcal{F}_1^{SR}$.

The moment bounds above and (10.32) establish that $F \in \mathcal{F}^{SR}_{1[r]}$ for (g_i, G_i) implies that $F \in \mathcal{F}_1$ for (g^*_{Fi}, G^*_{Fi}) for $\delta \leq \min\{1, 1/(2c)\}$, where the definition of $\mathcal{F}_1$ in (10.1) is taken to allow g_i and G_i to depend on F.[54] Now we apply Theorem 10.1 with (g^*_{Fi}, G^*_{Fi}) and r in place of (g_i, G_i) and k and $\delta \leq \min\{1, 1/(2c)\}$ to obtain the correct asymptotic size and asymptotic similarity of the CQLR$_1$ test based on (g^*_{Fi}, G^*_{Fi}) and r for the parameter space $\mathcal{F}^{SR}_{1[r]}$ for $r = 1, ..., k$. As noted above, the dependence of g^*_{Fi} and G^*_{Fi} on F does not cause a problem in the application of Theorem 10.1. This establishes the results of Theorem 8.1 for the SR-CQLR$_1$ test by the argument given above.[55] The proof for the SR-CQLR$_1$ CS is essentially the same, but with θ_0 taking any value in Θ and with $\mathcal{F}^{SR}_{\Theta,1}$ and $\mathcal{F}_{\Theta,1}$, defined in (4.10) and just below (10.1), in place of $\mathcal{F}_1^{SR}$ and $\mathcal{F}_1$, respectively.

This completes the proof of Theorem 8.1 given Theorem 10.1.

[54] We require $\delta \leq \min\{1, 1/(2c)\}$, rather than $\delta \leq 1/(2c)$, because $\lambda_{\min}(E_F g^*_{Fi} g^{*\prime}_{Fi}) = 1$ by (10.31) and $\mathcal{F}_1$ ($\subset \mathcal{F}_{AR}$) requires $\lambda_{\min}(E_F g^*_{Fi} g^{*\prime}_{Fi}) \geq \delta$.

[55] The fact that Z^*_{Fi} depends on θ_0 through $\Pi_{1F}^{-1/2}(\theta_0) A_F(\theta_0)'$ and that $G^*_{Fi}(\theta_0) \neq (\partial/\partial\theta') g^*_{Fi}(\theta_0)$ (because $(\partial/\partial\theta') Z^*_{Fi}$ is ignored in the specification of $G^*_{Fi}(\theta_0)$) does not affect the application of Theorem 10.1. The reason is that the proof of this Theorem goes through even if Z_i depends on θ_0 and for any $G_i(\theta_0)$ that satisfies the conditions in $\mathcal{F}_1$, not just for $G_i(\theta_0) := (\partial/\partial\theta') g_i(\theta_0)$.

References

Ahn, S. C., and P. Schmidt (1995): “Efficient Estimation of Models for Dynamic Panel Data,” *Journal of Econometrics,* 68, 5–27.

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

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

——— (2013a): “GMM Estimation and Uniform Subvector Inference with Possible Identification Failure,” *Econometric Theory,* 30, 1–47.

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

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

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

——— (2014a): “Asymptotic Size of Kleibergen’s LM and Conditional LR Tests for Moment Condition Models,” Cowles Foundation Discussion Paper No. 1977, Yale University.

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

——— (2014c): “Supplemental Material to ‘Identification- and Singularity-Robust Inference for Moment Condition Models’,” Cowles Foundation Discussion Paper No. 1978, Yale University.

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

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

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica,* 62, 1383–1414.

Andrews, D. W. K., and G. Soares (2007): "Rank Tests for Instrumental Variables Regression with Weak Instruments," *Econometric Theory,* 23, 1033–1082.

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

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

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

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

Angrist, J. D., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics,* 106, 979–1014.

——— (1992): "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," NBER Working Paper No. 4067.

Antoine, B., and E. Renault (2009): "Efficient GMM with Nearly Weak Instruments," *Econometrics Journal,* 12, S135–S171.

——— (2010): "Efficient Inference with Poor Instruments, a General Framework," in *Handbook of Empirical Economics and Finance,* ed. by D. Giles and A. Ullah. Oxford: Taylor and Francis.

Arellano, M., and O. Bover (1995): "Another Look at Instrumental Variable Estimation of Error Component Models," *Journal of Econometrics,* 68, 29–51.

Armstrong, T. B. (2012): "Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply," unpublished manuscript, Cowles Foundation, Yale University.

Armstrong, T. B., H. Hong, and D. Nekipelov (2012): "How Strong Must Identification Be for Conventional Asymptotics in Nonlinear Models?", unpublished manuscript, Cowles Foundation, Yale University.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica,* 60, 889–917.

Billingsley, P. (1979): *Probability and Measure.* New York: John Wiley and Sons.

Blundell, R., and S. Bond (1995): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," Working Paper No. W95/17, The Institute for Fiscal Studies, London.

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

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics,* 56, 431–449.

Carroll, C. D., J. Slacalek, and M. Sommer (2011): "International Evidence on Sticky Consumption Growth," *Review of Economics and Statistics,* 93, 1135–1145.

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

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

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

Chen, X., M. Ponomareva, and E. Tamer (2014): "Likelihood Inference in Some Finite Mixture Models," *Journal of Econometrics,* 182, 87–99.

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

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

Cho, J. S., and H. White (2007): "Testing for Regime Switching," *Econometrica,* 75, 1671–1720.

Choi, I., and Phillips, P. C. B. (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics,* 51, 113–150.

Cragg, J. C., and S. G. Donald (1996): "On the Asymptotic Properties of LDU-Based Tests of the Rank of a Matrix," *Journal of the American Statistical Association,* 91, 1301–1309.

——— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics,* 76, 223–250.

Cruz, L. M., and M. J. Moreira (2005): "On the Validity of Econometric Techniques with Weak Instruments: Inference on Returns to Education Using Compulsory School Attendance Laws," *Journal of Human Resources,* 40, 393–410.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Biometrika,* 64, 247–254.

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

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

Dufour, J.-M., L. Khalaf, and M. Kichian (2006): "Inflation Dynamics and the New Keynesian Phillips Curve: An Identification Robust Econometric Analysis," *Journal of Economic Dynamics and Control,* 30, 1707–1727.

——— (2013): "Identification-Robust Analysis of DSGE and Structural Macroeconomic Models," *Journal of Monetary Economics,* 60, 340–350.

Elliott, G., U. K. Müller, and M. W. Watson (2012): "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis," unpublished manuscript, Department of Economics, Princeton University.

Gomes, F. A. R., and L. S. Paz (2013): "Estimating the Elasticity of Intertemporal Substitution: Is the Aggregate Financial Return Free from the Weak Instrument Problem," *Journal of Macroeconomics,* 36, 63–75.

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

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models, *Quantitative Economics,* 4, 197–229.

Guggenberger, P. (2012): "On the Asymptotic Size Distortion of Tests When Instruments Locally Violate the Exogeneity Condition," *Econometric Theory,* 28, 387–421.

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

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

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

Hannan, E. J. (1982): "Testing for Autocorrelation and Akaike's Criterion," in *Essays in Statistical Science,* ed. by J. Gani and E. J. Hannan, *Journal of Applied Probability,* Special Volume 19A.

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

Hansen, L. P. and J. A. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica,* 63, 767–804.

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

Iskrev, N. (2010): "Local Identification in DSGE," *Journal of Monetary Economics,* 57, 189–202.

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

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

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

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

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests and the New Keynesian Phillips Curve," *Journal of Business and Economic Statistics,* 27, 293–311.

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

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

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

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

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

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

Nason, J. M., and G. W. Smith (2008): "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics,* 23, 525–551.

Neely, C. J., A. Roy, and C. H. Whiteman (2001): "Risk Aversion versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model," *Journal of Business and Economic Statistics,* 19, 395–403.

Nelson, C. R., and R. Startz (2007): "The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models," *Journal of Econometrics,* 138, 47-62.

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

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

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

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

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

Puhani, P. A. (2000): "The Heckman Correction for Sample Selection and its Critique," *Journal of Economic Surveys,* 14, 53–68.

Qu, Z. (2014): "Inference in DSGE Models with Possible Weak Identification," *Quantitative Economics,* 5, 457–494.

Qu, Z., and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models," *Quantitative Economics,* 3, 95–132.

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

Schorfheide, F. (2014): "Estimation and Evaluation of DSGE Models: Progress and Challenges," in *Advances in Economics and Econometrics*: *Theory and Applications, Tenth World Congress,* ed. by D. Acemoglu, M. Arellano, and E. Dekel, Vol. III, 184–230. Cambridge, UK: Cambridge University Press.

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

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

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

Teräsvirta, T. (1994): "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association,* 89, 208–218.

Van Loan, C. F., and N. Pitsianis (1993): "Approximation with Kronecker Products," in *Linear Algebra for Large Scale and Real-Time Applications,* ed. by M. S. Moonen, G. H. Golub, and B. L. R. de Moor, 293–314. Amsterdam: Kluwer Academic Publishers.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics,* 86, 797–810.