Efficiency Loss of Asymptotically Efficient Tests in an Instrumental Variables Regression [1]

Marcelo J. Moreira [2] Geert Ridder [3]

January 3, 2019

[1]Preliminary results were presented at seminars organized by BU, Brown, Caltech, Harvard-MIT, PUC-Rio, UCL, USC, University of California (Berkeley, Davis, Irvine, Los Angeles, Santa Barbara, and Santa Cruz campuses), and Yale, and conferences organized by CIREq (in honor of Jean-Marie Dufour), Harvard University (in honor of Gary Chamberlain), Oxford University ('New approaches to the Identification of Macroeconomic Models'), and the Tinberger Institute ('Inference Issues in Econometrics'). We thank Jack Porter, Leandro Gorno, Marinho Bertanha, and Pierre Perron for helpful comments. Mahrad Sharifvaghefi provided excellent research assistance. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

[2]FGV EPGE, Praia de Botafogo, 190, 11th floor, Rio de Janeiro, RJ 22250-040, Brasil. Electronic correspondence: mjmoreira@fgv.br

[3]Department of Economics and USC Dornsife INET, University of Southern California, Kaprilian Hall, Los Angeles, USA, CA 90089. Electronic correspondence: ridder@usc.edu

Abstract

In a model with endogenous regressors, heteroskedastic and autocorrelated (HAC) errors and weak instruments, tests that depend on the data only through the Anderson-Rubin (AR) and Lagrange Multiplier (LM) statistics ignore important information on the regression coefficients. This is in contrast to the homoskedastic case, where these statistics, together with the rank statistic, are one-to-one with the maximal invariant. The information loss with heteroskedastic and/or autocorrelated errors can be so extreme that the LM and conditional quasi-likelihood ratio (CQLR) tests have power close to size when it is trivial to distinguish the null from the alternative hypothesis. The severe loss of power can occur if the Hermitian part of the reduced-form covariance matrix has eigenvalues of opposite signs.

The conditional invariant likelihood (CIL) test proposed by Moreira and Ridder (2018) does not suffer this power loss. On the contrary, when the CQLR and LM tests fail, the CIL test can have power close to 1 under the alternative, even if its size is close to 0. This implies that the total variation distance between the null and subsets of the alternative is large, so that it is actually easy to distinguish between these hypotheses.

We also show that in the HAC-IV model, there are invariant statistics beyond the triad of AR, LM and rank statistics, so that the latter are not maximal invariant in the HAC case. We conclude that the popular LM and CQLR tests use data inefficiently if the equation errors are HAC.

Keywords: Endogenous regressor, Instrumental variable, Score test, Invariant test, HAC errors
JEL classification: C14, C36

1 Introduction

In an instrumental variable (IV) regression with possibly weak instruments, the practitioner currently has a choice of statistics for inference when equation errors are homoskedastic and uncorrelated. In the just-identified case, the Anderson-Rubin statistic (Anderson and Rubin, 1949) is unbiased, and has at least as much power as any other unbiased test; see Moreira (2001, 2009). In the overidentified case, Andrews, Stock, and Sun (2018) suggest using the conditional likelihood ratio (CLR) test of Moreira (2003), from among the other choices. Many others have contributed to inference on the structural parameters that is robust to weak instruments, among them Staiger and Stock (1997), Kleibergen (2002), Moreira (2009), Andrews, Moreira, and Stock (2006), and Mikusheva (2010). Stock, Wright, and Yogo (2002), Dufour (2003), and Andrews and Stock (2007) review weak-instrument robust inference.

Allowing for HAC errors in IV regression is important, because ignoring them results in substantial biases in the inference. That omitted variables are potentially serially correlated in time-series data is obvious. Newey and West (1987) and Andrews (1991) propose non-parametric estimators of the variance matrix of the equation error. In cross-sectional and panel data, HAC errors are also common. In panel data, the usual assumption is that equation errors are correlated in the time-series, but not in the cross-sectional dimension. Under the random-effects assumption, the model can be transformed into a model with homoskedastic and serially uncorrelated errors. If the random-effects assumption is too restrictive, then the robust variance matrix of the regression coefficients can be estimated, as in White (1980). This estimator makes assumptions neither on the error correlation within units nor on the conditional variance of the equation errors. Robust standard errors are routinely used in empirical research; see Angrist and Pischke (2009) and Wooldridge (2001). Betrand, Duflo, and Mullainathan (2003) show that in a dif-in-dif model, ignoring the autocorrelation of the equation errors leads to large biases in the inference. If data is clustered, empirical researchers are keenly aware that standard errors of regression coefficient estimators are underestimated, if both the equation errors and the covariates are positively correlated; see Kloek (1981), Moulton (1986), and Cameron and Miller (2015).

Which test is best for IV regression with heteroskedastic and autocorrelated (HAC) errors is less obvious. In the just-identified case, the Anderson-Rubin (AR) test is still the best choice. For the overidentified case, a number of tests have been proposed by Stock and Wright (2000), Kleibergen (2005), Andrews and Guggenberger (2015), Moreira and Moreira (2015), Andrews (2016), Andrews and Mikusheva (2016), and Moreira and Ridder (2018). Here we show that not all tests suggested for the HAC case are created equal. In particular, we show that tests that depend on the data only through the AR, score, and rank statistics do not use all relevant information in the data, and therefore can have low power if the errors are HAC. A preference for tests that depend on the data just through the AR, score, and rank statistics is informed by the observation that these statistics are equivalent to the maximal invariant in the homoskedastic and uncorrelated case. In a simulation study, we compare the power of tests that are functions of the score, AR, and rank statistics –in particular the popular AR, LM, and conditional quasi-likelihood ratio (CQLR) tests– to that of the conditional invariant likelihood (CIL) test proposed by Moreira and Ridder (2018). The additional information used by the CIL test derives from symmetries in the HAC-IV model that had not previously been exploited. In fact, with HAC errors, there seem to be no model symmetries, because linear transformations of the data change the variance matrix of the errors, so that this variance matrix is not invariant. A key insight of Moreira and Ridder (2018) is that the HAC-IV model has symmetries, if the variance matrix is considered as given, but not fixed. Because the variance matrix is considered to be part of both parameter and data spaces, there are symmetries in the HAC-IV model. By construction, the CIL test is invariant to this transformation (as are the LM and AR statistics). It maintains good power in many cases in which the LM and CQLR tests have power close to the size of the test.

We demonstrate the information loss for statistics that depend on the data only through the AR, LM, and rank statistics in two ways. First, we present simulations in which the CIL test has power close to 1 even if its size is close to 0. Kraft (1955) derives a necessary and sufficient condition for the existence of a test with minimal power over the alternative that exceeds the size by some ε. This condition is that the total variation distance between the convex combinations of the distributions under the null and (subsets of) under the alternative is at least ε. Because for our hypotheses ε is near one, the distance between the hypotheses is large. This means that it is easy to distinguish the null and alternative hypotheses. Since the LM and CQLR tests have power close to the size, these tests perform poorly in an easy problem, and therefore must omit relevant information regarding the regression coefficient. Second, we show that in the HAC-IV model, there exist invariant statistics beyond the triad of the LM, AR, and rank statistics. The LM and AR statistics are therefore not maximal invariant in general, and do not use all relevant information.

Because the power function is a very smooth function of the parameters of the HAC-IV model, the extreme power loss of the LM and CQLR tests extends to neighborhoods of the low power DGP with power remaining near size, while it is still trivial to make inference on the structural parameter. These problematic parameter regions are large in a topological sense, having nonzero Lebesgue measure. Therefore the power loss is sufficiently extensive to dissuade practitioners tof using the LM and CQLR tests, when errors are HAC. We show theoretically that the severe loss of power of the LM and CQLR tests can occur if the Hermitian part of the reduced-form covariance matrix has eigenvalues of opposite signs. In applications, the convex hull of the spectrum of this Hermitian matrix containing the zero value. For example, take the problem of making inference on the intertemporal elasticity of substitution (IES) with weak instruments. Using the data by Yogo (2004), only two of the eleven countries have all eigenvalues of the same sign.

The paper is organized as follows. In Section 2 we introduce the model and the test statistics. In Section 3 we show the information loss of test statistics that depend on the data only through the AR, score, and rank statistics. The asymptotic behavior of the score statistic for sequences of the variance matrix is analyzed in Section 4. DGP for which, according to the asymptotic analysis, the score test has low power are characterized in Section 5. Section 6 relates the loss of power to the correlation matrix of the reduced-form and first-stage errors and shows that power loss occurs in neighborhoods of the DGP delineated in Section 5. Section 7 concludes. Appendix A discusses conditions under which power of the LM test is arbitrarily close to zero. Appendix B contains proofs.

2 Model and test statistics

We consider the instrumental variable (IV) regression model

$$\begin{aligned} y_1 &= y_2\beta + u \\ y_2 &= Z\pi + v_2 \end{aligned}$$

with y_1, y_2 $n \times 1$ vectors of observations on two endogenous variables, and Z an $n \times k$ matrix of non-random instrumental variables. The $n \times 1$ vectors u, v_2 are the mean 0 structural equation and first-stage errors. The variance and covariance matrices of these errors are unrestricted, i.e., we allow for heteroskedastic and autocorrelated (HAC) errors. The errors have a normal distribution. Our objective is to test the null hypothesis $H_0 : \beta = \beta_0$ against the alternative $H_1 : \beta \neq \beta_0$ with π a $k \times 1$ vector of nuisance parameters.

The reduced-form model for $Y = [y_1\ y_2]$ is

$$Y = Z\pi a' + V \tag{2.1}$$

with $a = (\beta\ 1)'$ and $V = [v_1\ v_2]$, $v_1 = u + v_2\beta$.

Let $P = [P_1 : P_2]$ be an $n \times n$ orthogonal matrix with $P_1 = Z\,(Z'Z)^{-1/2}$. By orthogonality of P, we have $P_2'Z = 0$. We pre-multiply (2.1) by P' and define $R = P_1'Y$. The statistic $P_2'Y$ is ancillary, so inference is based on the $k \times 2$ statistic R. The induced model for R is given by

$$R = \mu a' + \widetilde{V} \tag{2.2}$$

with $\mu = (Z'Z)^{1/2}\,\pi$ and $\widetilde{V} = (Z'Z)^{-1/2}\,Z'V$. For finite-sample results, we assume that the vector $vec(\widetilde{V})$ has a normal distribution with variance matrix Σ, so that

$$vec(R) \sim N(vec(\mu a'), \Sigma).$$

As usual, we can drop the normality assumption at the cost of asymptotic approximations.

It is convenient to define $R_0 = RB_0$, with B_0 the non-singular 2×2 matrix

$$B_0 = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix}.$$

The induced model for $R_0 = [R_1 : R_2]$ is

$$vec(R_0) \sim N(vec(\mu a_\Delta'), \Sigma_0)$$

with $a_\Delta = (\Delta, 1)'$, $\Delta = \beta - \beta_0$ and $\Sigma_0 = (B_0' \otimes I_k)\Sigma(B_0 \otimes I_k)$. By construction, the first column of R_0 has mean 0 under the null and is a pivotal statistic.

The $k \times 1$ statistics S and T are defined by the 1-1 transformation of R:

$$\begin{aligned} S =& [(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)]^{-1/2}(b_0' \otimes I_k)vec(R) \text{ and} \\ T =& [(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)]^{-1/2}(a_0' \otimes I_k)\Sigma^{-1}vec(R), \end{aligned}$$

with $a_0 = (\beta_0, 1)'$, $b_0 = (1, -\beta_0)'$.

Under joint normality of $vec(R_0)$, the statistics S, T are independent and have distributions

$$\begin{aligned} S \sim & N(\Delta C_{\beta_0}\mu, I_k) \\ T \sim & N(D_\beta \mu, I_k) \end{aligned}$$

with $C_{\beta_0} = [(b_0' \otimes I_k)\Sigma(b_0 \otimes I_k)]^{-1/2}$ and $D_\beta = [(a_0' \otimes I_k)\Sigma^{-1}(a_0 \otimes I_k)]^{-1/2}(a_0' \otimes I_k)\Sigma^{-1}(a \otimes I_k)$. Under the null, T is a complete sufficient statistic for μ. For this reason we consider conditional tests given $T = t$.

The two-sided score test of $\beta = \beta_0$ has test statistic

$$LM = \frac{\left(S' C_{\beta_0} D_{\beta_0}^{-1} T\right)^2}{T' D_{\beta_0}^{-1} C_{\beta_0}^2 D_{\beta_0}^{-1} T} \ , \tag{2.3}$$

while the one-sided score test has test statistic

$$LM_1 = \frac{S' C_{\beta_0} D_{\beta_0}^{-1} T}{(T' D_{\beta_0}^{-1} C_{\beta_0}^2 D_{\beta_0}^{-1} T)^{1/2}} \ . \tag{2.4}$$

Other tests of $\beta = \beta_0$ if the errors are HAC are the Anderson-Rubin (AR) test, the CQLR test, and the general class of conditional linear combination tests (CLC), of which AR and CQLR are special cases (Andrews (2016)). The corresponding test statistics are

$$AR = S'S \tag{2.5}$$

$$CQLR = \frac{AR - T'T + \sqrt{(AR - T'T)^2 + 4LM \cdot T'T}}{2} \tag{2.6}$$

$$CLC = m(T)(AR - LM) + (1 - m(T)) \cdot AR \tag{2.7}$$

with $0 \leq m(T) \leq 1$.

Except for the AR test, the CLC statistics depend in a non-trivial way on the LM statistic. Therefore if the LM test has low power in some regions of the parameter space, the power of all CLC tests is affected. In Section 4, we show that we can find parameter values for which the power of the LM test is arbitrarily close to size. Hence, the power of CLC tests will be bounded by the AR test.

Moreira and Ridder (2018) propose a conditional invariant likelihood (CIL) test that is not in the CLC class of tests. The test is an invariant test for the group $g = (g_1, g_2)$ with g_1 in the group of $k \times k$ non-singular matrices $\mathcal{G}l_k$, and g_2 in the group of 2×2 lower triangular matrices with a positive main diagonal $\mathcal{G}_2^+$. As Moreira and Ridder (2018) argue, it is not obvious that the IV regression model with HAC errors is invariant to g. In particular, the variance matrix Σ changes with the transformation g, and this implies that the HAC-IV model is not invariant with respect to g. The model is invariant to g if we consider the variance matrix as given, but not fixed. To be precise, if we consider the variance matrix as both a parameter and as part of the data, then it is obvious that transforming the data changes the variance matrix (data) such that it is the variance matrix (parameter) of the transformed model. Therefore, if we consider the variance matrix both as data and as a parameter, we can use invariance arguments in the IV-model with HAC errors.

For the just-identified case, the Anderson-Rubin test is uniformly most powerful test, among unbiased (Moreira (2001, 2009) and Moreira and Moreira (2015)) or invariant (Andrews, Moreira, and Stock (2006) and Moreira and Ridder (2018)) tests. For the over-identified case, Moreira and Ridder (2018) derive an invariant test that maximizes the weighted average power (WAP). The test statistic of their conditional invariant likelihood (CIL) test is

$$IL = \int_{-\infty}^{\infty} e^{\frac{vec(R_0)'\Sigma_0^{-1}(a_\Delta \otimes I_k)\left((a_\Delta' \otimes I_k)\Sigma_0^{-1}(a_\Delta \otimes I_k)\right)^{-1}(a_\Delta' \otimes I_k)\Sigma_0^{-1}vec(R_0) - T'T}{2}} \times \left|(a_\Delta' \otimes I_k)\Sigma_0^{-1}(a_\Delta \otimes I_k)\right|^{-1/2} . |\Delta|^{k-2} d\Delta \ . \tag{2.8}$$

The test statistic is an integrated likelihood with weights $|\Delta|^{k-2}$ that make the test statistic invariant to the group g given above.

3 Information loss

Kraft (1955) provides that a necessary and sufficient condition for the existence of a test with a minimal power over the alternative that exceeds the size of the test by $\varepsilon > 0$. Specifically, this happens when the total variation distance between (convex combinations of) distributions under the null and under the alternative is at least ε. We use this result to argue that if there exists a test with size close to 0 and minimum power close to 1, this implies that the total variation distance between (convex combinations of) distributions under the null and under the alternative is close to 1. This implies that the null and alternative hypotheses are so far apart, that there exists a test with type I and type II errors close to 0. In Section 5, we show that the score and CQLR tests have power close to size for certain DGP and a subset of the alternative. The CIL test has power close to 1 for this subset, even if the size is close to 0, so that the total variation distance between the null and alternative hypotheses is close to 1. As a consequence, there exists for each null and alternatives, a test that can perfectly discriminate between these distributions. Tests like the score test that have power much smaller than 1 under the alternative, clearly do not use all relevant information. As a result, the power loss is not restricted to the DGP that we derive in Sections 4 and 6. Indeed, we adapt Theorem 5 of Kraft (1955) to show a large collection of DGP in which the score test behaves as badly as in our impossibility design, described in Section 6.

We also show the information loss by finding statistics that are invariant to the transformation g that we define in Section 2, and that are not a function of the AR and score statistics. If errors are homoskedastic and serially uncorrelated, the maximal invariant consists of the statistics $S'S$, $S'T$, and $T'T$.[1] There is a one-to-one relation between the one-sided score statistic $S'T/(T'T)^{1/2}$, the AR statistic $S'S$, and the statistic $T'T$ on the one hand, and $S'S$, $S'T$, and $T'T$ on the other. Since we condition on T, statistics that only depend on the data through AR and score statistics use all relevant information. With HAC errors, the AR and score statistics

[1] See Andrews, Moreira, and Stock (2006) for a proof. However, they do not rule out the use of the reduced-form variance itself as part of the maximal invariant. Moreira and Ridder (2018) eliminate the reduced-form variance as part of the maximal invariant by finding the largest affine group of transformations that preserves the testing problem.

are no longer part of the maximal invariant that is (much) larger than these statistics. Therefore, a test that depends on the data only through the score and AR statistics is expected to have lower power. The power loss for a class of DGP is shown in Sections 4 and 6. The information loss occurs if the error variance matrix Σ does not have a Kronecker-product structure, i.e., if $\Sigma \neq \Omega \otimes \Phi$, where Ω is a symmetric 2×2 matrix and Φ is a symmetric $k \times k$ matrix.

3.1 Test statistics and the total variation distance between the null and alternative

Theorem 5 of Kraft (1955) connects the possibility of testing a null against an alternative hypothesis to the distance between these hypotheses. Let $\phi(R)$ be a test of H_0 against H_1. The distributions Q of R for the null hypothesis are in a set $\mathbb{Q}_0$. The distributions for the alternative are in a set $\mathbb{Q}_1$. The size and power of the test are

$$\sup_{Q_0 \in \mathbb{Q}_0} \mathbb{E}_{Q_0}(\phi(R))$$

and

$$\inf_{Q_1 \in \mathbb{Q}_1} \mathbb{E}_{Q_1}(\phi(R)) \ .$$

We also define the convex hull $co(\mathbb{Q}_0)$ of $\mathbb{Q}_0$ as the set of all discrete mixtures of distributions in $\mathbb{Q}_0$. We define $co(\mathbb{Q}_1)$ in the same way.

Kraft (1955) gives a necessary and sufficient condition for the existence of a test with power that exceeds the size by $\varepsilon > 0$.

Theorem 1 (Kraft) *For $\varepsilon > 0$ there exists a test ϕ with*

$$\inf_{Q_1 \in \mathbb{Q}_1} \mathbb{E}_{Q_1}(\phi(R)) \geq \varepsilon + \sup_{Q_0 \in \mathbb{Q}_0} \mathbb{E}_{Q_0}(\phi(R))$$

if and only if for all $Q_0 \in co(\mathbb{Q}_0)$ and $Q_1 \in co(\mathbb{Q}_1)$,

$$d(Q_0, Q_1) \geq \varepsilon \ .$$

The distance between two distributions is measured by the total variation (TV) distance

$$d(Q_0, Q_1) = \sup_{B \in \mathcal{B}} |Q_0(B) - Q_1(B)|,$$

where $\mathcal{B}$ is the Borel sigma-algebra. Kraft's theorem has been used to study so-called impossible testing problems, for which the power of any test does not exceed the size of the test. Bertanha and Moreira (2018) discuss such impossible testing problems, and give examples where inference with power greater than size is impossible, because the null and alternative hypotheses are very close.

For the testing problem that we consider, we use the fact that Kraft's theorem gives a necessary and sufficient condition for the existence of a test with power exceeding the size by

ε. In Section 5, we show that the power of the CIL test is close to 1, even if the size of the test is close to 0. By Kraft's theorem, this implies that the total variation distance between the (convex hull of) $\mathbb{Q}_0$ and $\mathbb{Q}_1$ is close to 1, which is the largest value that the total variation distance can take. We will apply Theorem 5 of Kraft (1955) by taking $\mathbb{Q}_0$ as the null hypothesis and $\mathbb{Q}_1$ as subsets of the alternative (because the closure of our alternative contains the null $\beta = \beta_0$, the TV distance is trivially zero).

Testing problems with a total variation distance between the null and the alternative equal to one are easy. To give an example, consider distributions that are mixed discrete continuous, so that they are absolutely continuous with respect to the sum of the Lebesgue and counting measures. The support of the distributions is $[0, 1]$. $\mathbb{Q}_0$ is the set of discrete distributions that assign positive probability to 0 and 1, and $\mathbb{Q}_1$ is the set of continuous distributions on the unit interval. The test ϕ rejects the null if we observe a value that is equal to neither 0 nor 1. If the null is correct, we reject the null with probability $Q_0([0, 1] \setminus \{0, 1\}) = 0$ for all $Q_0 \in \mathbb{Q}_0$. If the alternative is correct, we reject the null with probability $Q_1([0, 1] \setminus \{0, 1\}) = 1$ for all $Q_1 \in \mathbb{Q}_1$. Therefore, the type I and type II error probabilities of this test are 0. The total variation distance between $co(\mathbb{Q}_0)$ and $co(\mathbb{Q}_1)$ is 1 if we take $B = [0, 1] \setminus \{0, 1\}$ so that $Q_0(B) = 0, Q_1(B) = 1$ for all $Q_0 \in co(\mathbb{Q}_0)$ and $Q_1 \in co(\mathbb{Q}_1)$ or if we take $B = \{0, 1\}$ so that $Q_0(B) = 1, Q_1(B) = 0$ for all $Q \in co(\mathbb{Q}_0)$ and $Q \in co(\mathbb{Q}_1)$.

Any reasonable test should have type I and type II error probabilities equal to 0, because the total variation distance between the hypotheses is 1. Now consider the test that rejects the null if the observed value is greater than .5. This test has a probability of false rejection of the null equal to .5 and a probability of false rejection of the alternative also equal to .5. Therefore, the probabilities of type I and type II errors are large, because the test ignores the information that, under the null, the support is $\{0, 1\}$. By increasing the number of points in the support under the null, we can create a test with arbitrary size and power over the alternative equal to size. The test ignores the support information, and that loss of information causes its disappointing performance.

If $\mathbb{Q}_0$ and $\mathbb{Q}_1$ are simple hypotheses, then an ideal test is easily found. Let the total variation distance between $\mathbb{Q}_0$ and $\mathbb{Q}_1$ be one, and let B be such that $Q_1(B) = 1$ and $Q_0(B) = 0$. The test that rejects the null if the observed value is in B has both type I and II error probabilities equal to zero.

As shown in Section 5, the power of the CIL test is close to 1 for a subset of the alternative even if the size is close to 0. We conclude from Kraft's theorem that our testing problem is easy, and that an (almost) ideal test exists. The LM and CQLR tests have power close to size in this easy testing problem, because these tests ignore relevant information. The loss of information occurs for all DGP, not just for the DGP in Section 6.

3.2 Invariant statistics in the HAC model that are not a function of LM and AR statistics

As noted in the homoskedastic and not serially correlated case, there is a one-to-one relation between the maximal invariant and the score, AR, and the $T'T$ statistics. Moreira and Ridder (2018) also consider when Σ has a Kronecker product structure $\Sigma = \Omega_0 \otimes \Phi$ with Ω and Φ known, but not fixed. The group $g = (g_1, g_2)$ is the same as in Section 2, i.e., g_1 is in the group of $k \times k$ non-singular matrices $\mathcal{G}l_k$, and g_2 is in the group of 2×2 lower triangular matrices with a positive main diagonal $\mathcal{G}_2$. The group action/transformation of g_1 in the sample space is

$$g_1 \circ (R_0, \Omega_0, \Phi) = (g_1 R_0, \Omega_0, g_1 \Phi g_1') \tag{3.9}$$

and that of g_2 is

$$g_2 \circ (R_0, \Omega_0, \Phi) = (R_0 g_2', g_2 \Omega_0 g_2', \Phi) \ . \tag{3.10}$$

Note that g_1 works transitively on Φ, i.e.,

$$g_1 \Phi g_1' = g_1 \widetilde{\Phi} g_1' \Leftrightarrow \Phi = \widetilde{\Phi}$$

and g_2 works transitively on Ω_0. Therefore Ω_0 and Φ are not part of the maximal invariant. Indeed Moreira and Ridder (2018) show that the maximal invariant is $S'S, T'T, (S'T)^2$ which is the same as that for the homoskedastic and not autocorrelated case, if we also require invariance to changes of sign. Therefore, if the error variance has a Kronecker product structure, then all relevant information is in the AR, LM and $T'T$ statistics.

We now show that in the case of HAC errors, there exist invariant statistics that do not depend on the data only through the score, AR and $T'T$ statistics. If such invariant statistics exist, then the AR and score test statistics are not part of the maximal invariant statistic, that is much larger.

We want to find invariant statistics that are not a function of the score and AR statistics. The statistics S and T depend on the data through $(e_1' \otimes I_k) vec(R_0)$ for S and $(e_2' \otimes I_k) \Sigma_0^{-1} vec(R_0)$ for T (e_1, e_2 are the two-dimensional unit vectors), and S and T also depend on Σ_0, a statistic under the known, but not fixed, variance assumption. We first consider the data transformation on these statistics separately. The transformation applied to $vec(R_0)$ and Σ_0 is $g = g_2 \otimes g_1$ with

$$g_2' = \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix}$$
$$g_2'^{-1} = \begin{pmatrix} g_{11}^{-1} & -g_{11}^{-1} g_{12} g_{22}^{-1} \\ 0 & g_{22}^{-1} \end{pmatrix}$$

and g_1 a non-singular $K \times K$ matrix.

Therefore, if we denote the transformation of the statistic S by $g \circ S$ (and the same for other statistics), then for S

$$\begin{aligned} g \circ (e_1' \otimes I_k) vec(R_0) &= (e_1' \otimes I_k) (g_2 \otimes g_1) \, vec\,(R_0) \\ &= (e_1' g_2 \otimes g_1) \, vec\,(R_0) \\ &= ((g_{11}, 0) \otimes g_1) \, vec\,(R_0) \\ &= g_{11} g_1 \, (e_1' \otimes I_k) \, vec\,(R_0) \end{aligned}$$

and for T (both R_0 and Σ_0 are transformed)

$$\begin{aligned} g \circ (e_2' \otimes I_k) \, \Sigma_0^{-1} vec\,(R_0) &= (e_2' \otimes I_k) \left(g_2'^{-1} \otimes g_1'^{-1} \right) \Sigma_0^{-1} \left(g_2^{-1} \otimes g_1^{-1} \right) (g_2 \otimes g_1) \, vec\,(R_0) \\ &= \left(e_2' g_2'^{-1} \otimes g_1'^{-1} \right) \Sigma_0^{-1} vec\,(R_0) \\ &= \left(\left(0, g_{22}^{-1}\right) \otimes g_1'^{-1} \right) \Sigma_0^{-1} vec\,(R_0) \\ &= g_{22}^{-1} g_1'^{-1} \, (e_2' \otimes I_k) \, \Sigma_0^{-1} vec\,(R_0) \, . \end{aligned}$$

The transformation of Σ_0 is given by

$$\begin{aligned}
&\left(\begin{pmatrix} g_{11} & 0 \\ g_{12} & g_{22} \end{pmatrix} \otimes g_1\right) \cdot \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \cdot \left(\begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} \otimes g_1'\right) \\
&= (I_2 \otimes g_1) \begin{bmatrix} g_{11}\Sigma_{11} & g_{11}\Sigma_{12} \\ g_{12}\Sigma_{11} + g_{22}\Sigma_{21} & g_{12}\Sigma_{12} + g_{22}\Sigma_{22} \end{bmatrix} \left(\begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} \otimes g_1'\right) \\
&= (I_2 \otimes g_1) \begin{bmatrix} g_{11}^2\Sigma_{11} & g_{11}\left(g_{12}\Sigma_{11} + g_{22}\Sigma_{12}\right) \\ g_{11}\left(g_{12}\Sigma_{11} + g_{22}\Sigma_{21}\right) & g_{12}^2\Sigma_{11} + g_{12}g_{22}\left(\Sigma_{21} + \Sigma_{21}\right) + g_{22}^2\Sigma_{22} \end{bmatrix} (I_2 \otimes g_1') \, .
\end{aligned}$$

The transformation of Σ_0^{-1} is the inverse of this matrix

$$\begin{aligned}
&\left(\begin{pmatrix} g_{11}^{-1} & -g_{11}^{-1}g_{12}g_{22}^{-1} \\ 0 & g_{22}^{-1} \end{pmatrix} \otimes g_1'^{-1}\right) \cdot \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \cdot \left(\begin{pmatrix} g_{11}^{-1} & 0 \\ -g_{11}^{-1}g_{12}g_{22}^{-1} & g_{22}^{-1} \end{pmatrix} \otimes g_1^{-1}\right) \\
&= \left(I_2 \otimes g_1'^{-1}\right) \begin{bmatrix} g_{11}^{-1}\Sigma^{11} - g_{11}^{-1}g_{12}g_{22}^{-1}\Sigma^{21} & g_{11}^{-1}\Sigma^{12} - g_{11}^{-1}g_{12}g_{22}^{-1}\Sigma^{22} \\ g_{22}^{-1}\Sigma^{21} & g_{22}^{-1}\Sigma^{22} \end{bmatrix} \left(\begin{pmatrix} g_{11}^{-1} & 0 \\ -g_{11}^{-1}g_{12}g_{22}^{-1} & g_{22}^{-1} \end{pmatrix} \otimes g_1^{-1}\right) \\
&= \left(I_2 \otimes g_1'^{-1}\right) \begin{bmatrix} g_{11}^{-2}\left(\Sigma^{11} - g_{12}g_{22}^{-1}\left(\Sigma^{12} + \Sigma^{21}\right) + g_{12}^2g_{22}^{-2}\Sigma^{22}\right) & g_{11}^{-1}g_{22}^{-1}\left(\Sigma^{12} - g_{12}g_{22}^{-1}\Sigma^{22}\right) \\ g_{11}^{-1}g_{22}^{-1}\left(\Sigma^{21} - g_{12}g_{22}^{-1}\Sigma^{22}\right) & g_{22}^{-2}\Sigma^{22} \end{bmatrix} \left(I_2 \otimes g_1^{-1}\right) .
\end{aligned}$$

We use these results to find the transformation of the statistics

$$\begin{aligned}
C_{\beta_0}^{-1}S &= (e_1' \otimes I_k)vec(R_0) \\
D_{\beta_0}T &= [(a_0' \otimes I_k)\Sigma_0^{-1}(a_0 \otimes I_k)]^{-1}(a_0' \otimes I_k)\Sigma_0^{-1}vec(R) \\
C_{\beta_0}^{-2} &= \Sigma_{11}
\end{aligned}$$

that are

$$g \circ C_{\beta_0}^{-1}S = g_{11}g_1C_{\beta_0}^{-1}S \tag{3.11}$$
$$g \circ D_{\beta_0}T = g_{22}^{-1}g_1'^{-1}D_{\beta_0} \tag{3.12}$$
$$g \circ C_{\beta_0}^{-2} = g_{11}^2g_1C_{\beta_0}^{-2}g_1' \ . \tag{3.13}$$

The statistic we consider is

$$F_1 = \frac{S'C_{\beta_0}^{-1}D_{\beta_0}T}{\left(T'D_{\beta_0}C_{\beta_0}^{-2}D_{\beta_0}T\right)^{1/2}} \ . \tag{3.14}$$

We have

$$g \circ F_1 = \frac{g_{11}S'C_{\beta_0}^{-1}g_1'g_{22}^{-1}g_1'^{-1}D_{\beta_0}}{\left(g_{22}^{-1}T'D_{\beta_0}g_1^{-1}g_{11}^2g_1C_{\beta_0}^{-2}g_1'g_{22}^{-1}g_1'^{-1}D_{\beta_0}\right)^{1/2}} = sign(g_{11}/g_{22})F_1$$

so that F_1^2 is indeed invariant.

There are many more invariant statistics. For instance, quadratic forms of

$$F_{k+1} = \frac{S'C_{\beta_0}^{-1}\left(D_{\beta_0}^2C_{\beta_0}^{-2}\right)^k D_{\beta_0}T}{\left(T'D_{\beta_0}\left(C_{\beta_0}^{-2}D_{\beta_0}^2\right)^k C_{\beta_0}^{-2}\left(D_{\beta_0}^2C_{\beta_0}^{-2}\right)^k D_{\beta_0}T\right)^{1/2}} \text{ for } k = 1, 2, \ldots,$$

are invariant and, in general, not functions of the score statistic.

4 Asymptotic analysis of the score statistic

We analyze the large sample properties of the one-sided score statistic (2.4) under both strong and weak instrument assumptions. Without loss of generality, we set $\beta_0 = 0$, so that

$$S \sim N(\Delta\Sigma_{11}^{-1/2}\mu, I_k) \text{ and } T \sim N\left((\Sigma^{22})^{1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu, I_k\right),$$

with $C_0 = \Sigma_{11}^{-1/2}$, $D_0 = (\Sigma^{22})^{1/2}$, and $\Sigma^{22} = (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}$. Therefore,

$$LM_1 = \frac{S'C_0D_0^{-1}T}{(T'D_0^{-1}C_0^2D_0^{-1}T)^{1/2}} = \frac{S'\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}T}{(T'^{22})^{-1/2}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}T)^{1/2}}$$

The normality of S and T implies that

$$\begin{aligned} S =& \Delta\Sigma_{11}^{-1/2}\mu + U_S \text{ and} \\ T =& (\Sigma^{22})^{1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + U_T \end{aligned}$$

with U_S and U_T being independent random vectors with distribution $N(0, I_k)$. Substitution in LM_1 gives

$$LM_1 = \frac{(\Delta\Sigma_{11}^{-1/2}\mu + U_S)'\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}((\Sigma^{22})^{1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + U_T)}{(((\Sigma^{22})^{1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + U_T)'^{22})^{-1/2}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}((\Sigma^{22})^{1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + U_T))^{1/2}}.$$

First, we consider strong IV assumptions. We assume

$$\frac{Z'Z}{n} \xrightarrow{p} m_Z$$

with m_Z a constant $k \times k$ matrix. This holds if the rows of Z are an i.i.d. sample from a population distribution, but also under other assumptions. The other strong IV assumption is that $\Delta = \delta/\sqrt{n}$ with δ a constant scalar.

Under these assumptions, we have that

$$\mu = \sqrt{n}\left(\frac{Z'Z}{n}\right)^{1/2}\pi = \sqrt{n}(m_Z + o_p(1))\pi = \sqrt{n}(m + o_p(1))$$

with $m = (m_Z)^{1/2}\pi$. Therefore

$$\begin{aligned} \Delta\Sigma_{11}^{-1/2}\mu + U_S &= \frac{\delta}{\sqrt{n}}\Sigma_{11}^{-1/2}\sqrt{n}(m + o_p(1)) + U_S \\ &= \delta\Sigma_{11}^{-1/2}m + U_S + o_p(1) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}}\left[(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + U_T\right] &= \left(I_k - \frac{\delta}{\sqrt{n}}\Sigma_{21}\Sigma_{11}^{-1}\right)(m + o_p(1)) + \frac{U_T}{\sqrt{n}} \\ &= m + o_p(1)\ . \end{aligned}$$

Hence if we divide numerator and denominator by $\sqrt{n}$ we obtain

$$LM1 = \frac{[\delta\Sigma_{11}^{-1/2}m + U_S + o_p(1)]'\Sigma_{11}^{-1/2}[m + o_p(1)]}{\left([m + o_p(1)]'\Sigma_{11}^{-1}[m + o_p(1)]\right)^{1/2}} \xrightarrow{p} \frac{(\delta\Sigma_{11}^{-1/2}m + U_S)'\Sigma_{11}^{-1/2}m}{(m'\Sigma_{11}^{-1}m)^{1/2}} \sim N\left(\delta\left(m'\Sigma_{11}^{-1}m\right)^{1/2}, 1\right) .$$

This derivation is summarized in the following proposition.

Proposition 1 *If $\frac{Z'Z}{n} \xrightarrow{p} m_Z$ and $\Delta = \delta/\sqrt{n}$, then*

$$LM_1 \xrightarrow{d} N\left(\delta\left(m'\Sigma_{11}^{-1}m\right)^{1/2}, 1\right) .$$

It follows that the two-sided LM test has non-centrality parameter $\delta^2 m'\Sigma_{11}^{-1}m$, which is also that of the AR test, that is UMPI if $k = 1$ but not if $k > 1$. The key observation from Proposition 1 is that the mean of the LM_1 statistic depends on δ if $m'\Sigma_{11}^{-1}m > 0$. We next show that this is not necessarily true under weak IV assumptions.

Other authors have pointed out problems and solutions for the LM test under weak-IV assumptions. Moreira (2001) shows that the noncentrality parameter for LM can be zero for a particular alternative with homoskedastic errors, and proposes a switching test based on the AR and LM tests. Andrews (2016) also notes problems with the LM test in the GMM context with heteroskedasticity and autocorrelation, and recommends the use of conditional tests based on linear combinations between the AR and LM statistics. The theory derived below gives more definitive conclusions regarding the low power of the LM statistic, with implications for other tests, including the CQLR test.

Under weak-IV assumptions Δ and μ do not change with the sample size, so that the numerator and denominator do not have probability limits if the sample size increases without bounds. Instead, we consider a small-σ approximation analogous to Kadane (1971). Such an approximation gives us the expected value of the ratio that defines the LM statistic as the ratio of the expected values of the numerator and denominator. To be precise, we assume a variance sequence such that

$$\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2} \to 0 . \tag{4.15}$$

We rewrite the LM_1 statistic as

$$LM_1 = \frac{(\Delta\Sigma_{11}^{-1/2}\mu + U_S)'((\Sigma_{11}^{-1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + \Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T)}{((\Sigma_{11}^{-1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + \Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T)'(\Sigma_{11}^{-1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + \Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T))^{1/2}} .$$

If (4.15) holds, then LM_1 converges in probability to a normal random variable with mean

$$\frac{\Delta\mu'\Sigma_{11}^{-1}\left(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1}\right)\mu}{\left(\mu'\left(I_k - \Delta\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma_{11}^{-1}\left(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1}\right)\mu\right)^{1/2}} .$$

The numerator is of order Δ^2 while the denominator is of order Δ, so that, in general, the absolute mean increases without bounds if $|\Delta| \to \infty$. However, if

$$\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu = 0, \tag{4.16}$$

then the order of the numerator is Δ, the same as that of the denominator. The mean is bounded by an expression of μ and Σ that does not depend on Δ. This implies that if (4.16) holds, then the power of the LM_1 test is less dependent of Δ. In fact, as the following theorem states under slightly stronger conditions, the convergence to a normal random variable is uniform in Δ if (4.16) holds. The mean is bounded from above by a function of μ and Σ that does not depend on Δ.

Theorem 2 *If (4.15) and (4.16) hold and in addition*

$$\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2} \to 0 \tag{4.17}$$

and for such variance sequences

$$\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu > 0 \tag{4.18}$$

then uniformly in Δ

$$LM_1 \xrightarrow{p} \frac{U_S'\Sigma_{11}^{-1/2}(I_k - \Delta\Sigma_{21}\Sigma_{11}^{-1})\mu + \Delta\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}} \sim N\left(\frac{\Delta\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}, 1\right). \tag{4.19}$$

For all Δ the mean has upper bound

$$\frac{\Delta\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}} \leq \frac{\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}. \tag{4.20}$$

5 Impossibility designs

If (4.16) holds, then the mean of the LM_1 test is bounded by a function that depends on the HAC variance Σ and the first stage parameter μ, but not on Δ. Therefore, the test is unable to detect large deviations $|\Delta|$ from the null hypothesis. Worse, if μ and Σ are such that the mean is close to 0 for all Δ, then the power of the test is close to its size. In this section, we show that there is a region of the parameter space of μ and Σ where this occurs. We call designs such that both (4.15) and (4.16) hold as impossibility designs. We give an example of a set of such designs in this section. In Section 6 we discuss the prevalence of these impossibility designs that is related to properties of the covariance matrix of the reduced-form and first-stage errors.

Let J_k be the $k \times k$ matrix with the anti-diagonal equal to 1 and the other components equal to 0. We have $J_k^2 = I_k$. The HAC variance matrix is the $2k \times 2k$ matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{5.21}$$

with $k \times k$ submatrices

$$\Sigma_{11} = c_{11}I_k,\ \Sigma_{12} = c_{12}J_k, \text{ and } \Sigma_{22} = c_{22}I_k. \tag{5.22}$$

The constants c_{11}, c_{12}, and c_{22} are chosen so that the matrix Σ_0 is positive definite. Each one of the eigenvalues of Σ_0,

$$\varsigma_1 = \frac{c_{11} + c_{22} + \sqrt{(c_{11} - c_{22})^2 + 4.c_{12}^2}}{2} \text{ and}$$
$$\varsigma_2 = \frac{c_{11} + c_{22} - \sqrt{(c_{11} - c_{22})^2 + 4.c_{12}^2}}{2}$$

appear with multiplicity k. As long as $c_{11}, c_{22} \geq 0$ and $c_{11}.c_{22} \geq c_{12}^2$, the matrix Σ_0 is semi-positive definite.

Note that $J_k e_1 = e_k$ with e_1, e_k the first c.q. k-th unit vector. Therefore, if we set $\mu = \lambda^{1/2} e_1$, with λ some positive constant, we find that

$$\mu' \Sigma_{11}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \mu = \lambda \frac{c_{12}}{c_{11}^2} e_1' J_k e_1 = \lambda \frac{c_{12}}{c_{11}^2} e_1' e_k = 0$$

so that (4.16) holds for this choice of μ and Σ. We also have

$$\begin{aligned} \Sigma^{22} &= \left(c_{12} I_k - \frac{c_{12}^2}{c_{11}} J_k J_k \right) \\ &= \frac{c_{11}}{c_{11} c_{22} - c_{12}^2} I_k \ . \end{aligned}$$

Now set $c_{11} = 1$ and $c_{22} = c_{12}^2 + c_{12}^{-3}$. For this choice, the matrices in (4.15) and (4.17) in Theorem 2 are

$$\Sigma_{11}^{-1} \left(\Sigma^{22} \right)^{-1/2} = \Sigma_{11}^{-1/2} \left(\Sigma^{22} \right)^{-1/2} = c_{12}^{-3/2} I_k$$

so that (4.15) and (4.17) hold if $c_{12} \to \infty$. Because the conditions of Theorem 2 hold, the upper bound on the mean of the LM_1 statistic is

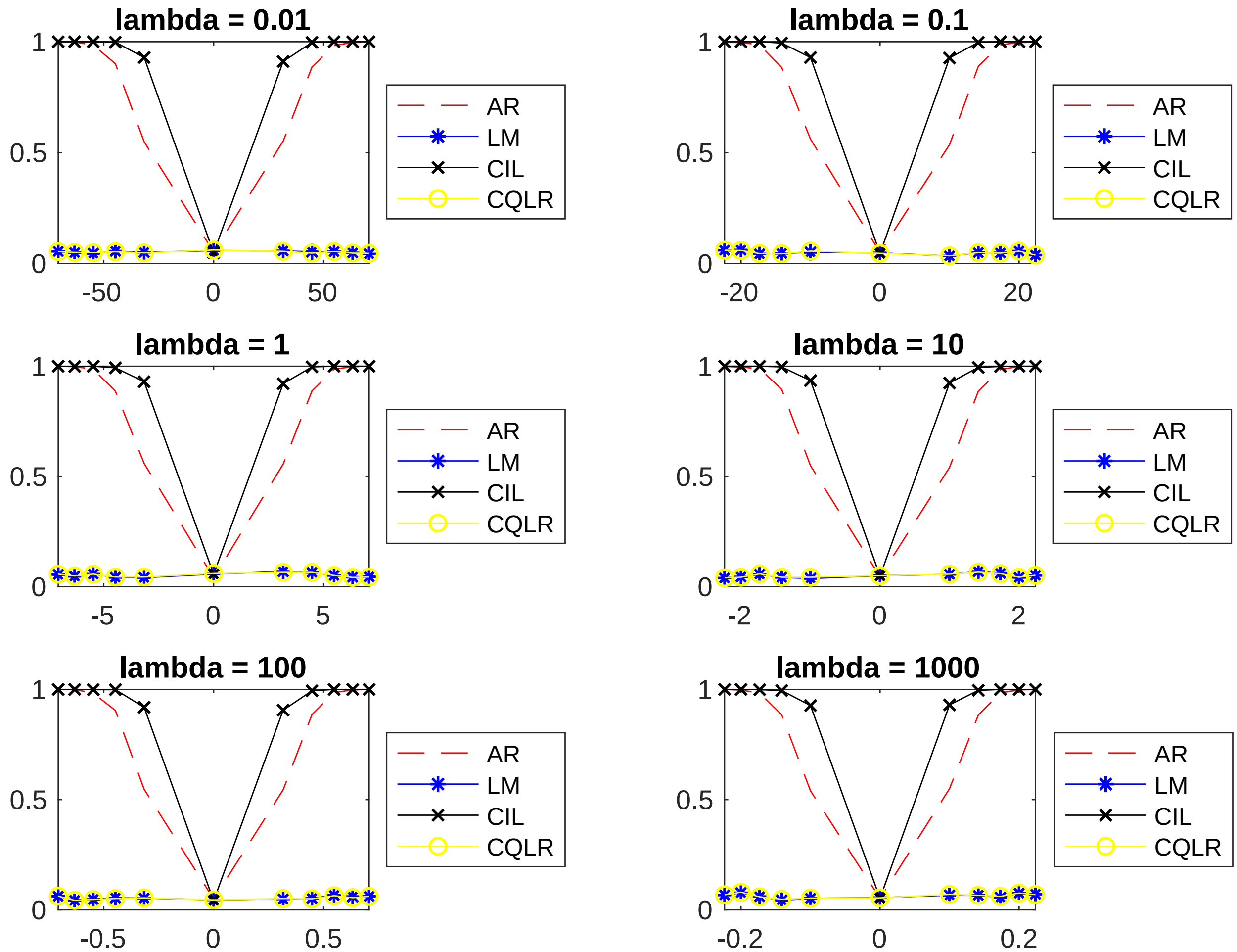
$$\frac{\mu' \Sigma_{11}^{-1} \mu}{\left(\mu' \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \mu \right)^{1/2}} = \frac{\frac{\lambda}{c_{11}}}{\left(\lambda \frac{c_{12}^2}{c_{11}^3} \right)^{1/2}} = c_{11}^{1/2} \lambda^{1/2} c_{12}^{-1}$$

so that if either c_{12} is large or λ is small, the mean of LM_1 is close to 0 under the alternative and the test has power equal to size.

If $c_{12} \to \infty$ the HAC variance matrix Σ converges to a singular matrix. Singularity of the variance matrix implies that a linear transformation of the S and T statistics has variance 0. This should improve the power of any test that depends on the data through S and T. By its structure, however, the power of the LM test is bounded, so that the LM test cannot take advantage of this information.

Above, we identified regions of the parameter space where we expect the score test to have low power, even power close to size. We confirm this suspicion in a simulation experiment. Data are generated as in (2.2) with $k = 10$, i.e., we have 10 instruments. The normal errors are HAC with a variance matrix as in (5.21) and (5.22). As the approximations in Theorem 2 are more accurate if c_{12} is large, we set $c_{12} = 100$. Further, $c_{11} = 1, c_{22} = c_{12}^2 + c_{12}^{-3}$. We test $H_0 : \beta = \beta_0 = 0$.

Figure 1: Power curves AR, LM, CQLR, and CIL tests for model with HAC errors with $c_{12} = 100$, $c_{11} = 1$ and $c_{22} = c_{12}^2 + c_{12}^{-3}$; varying instrument strength λ, $\alpha = .05$.



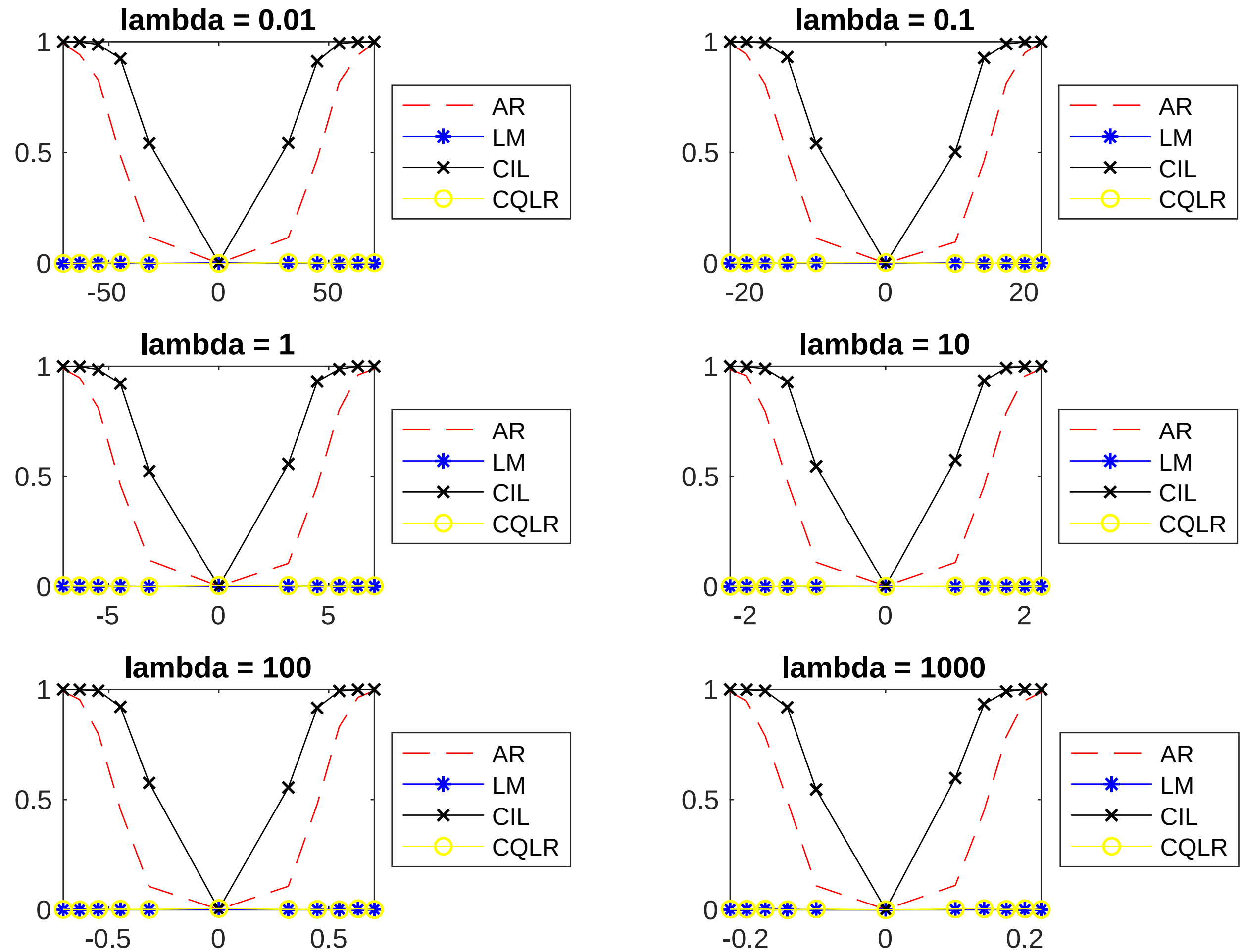
In Figures 1 and 2 we report the power functions for the AR, CQLR, LM, and CIL tests for levels of instrument strength λ ranging from .01 (very weak) to 1000 (borderline strong) and for the 5% significance level.

For values of λ of 1000 or less, the power of the LM test is approximately equal to the size of the test. The AR test does not have the same lack of power for the designs where the LM test fails. However, the power of the AR test is known to deteriorate when the number of instruments increases. All other tests in the class of CLC tests have power smaller than the AR test. This is because they are conditional convolutions between the AR statistic and another statistic that is approximately ancillary.

The CIL test proposed by Moreira and Ridder (2018) is not in the CLC class of tests, and depends on the data not just through the AR and LM statistics. This test may therefore avoid the loss of power of the LM and CQLR tests. This is confirmed in Figures 1 and 2. The CIL test performs better than the LM and CQLR tests, and even improves on the AR test.

The flat power curves of the LM and CQLR tests are surprising. These tests are often applied

Figure 2: Power curves AR, LM, CQLR, and CIL tests for models with HAC errors with $c_{12} = 100$, $c_{11} = 1$ and $c_{22} = c_{12}^2 + c_{12}^{-3}$; varying instrument strength λ, $\alpha = .001$.



in nonlinear models and heteroskedastic and correlated errors, and their use is so common that the citations number in the thousands.[2] On the other hand, Moreira and Moreira (2015) report power comparisons that show that the CIL, CLR and strongly unbiased (SU) tests perform better than the LM and CQLR tests. One could perhaps argue that that we should not discard these tests yet, because it is possible that there are regions in the parameter space where they outperform the other tests. As we argue next, the problems with these tests are much more fundamental.

Figure 2 shows even more clearly the information loss of the LM and CQLR tests. It turns out we can nearly perfectly distinguish the null from alternatives far enough from the null. To show this we choose a very small probability of a type I error, and we find regions of the alternative where the probability of making a type II error is very small as well. Specifically, we choose the significance level $\alpha = .001$ and show that the power of the CIL test remains

[2]See Finlay and Magnussoni (2009) for a reference on their implementation in Stata.

close to one when β is distant enough from the null. From Andrews (2016), we can show the CIL test gives a lower bound for the total variation distance between the null hypothesis and those regions of the alternative hypothesis. Hence, like testing a Bernoulli against a continuous distribution, the testing problem is easy and we should have no problem distinguishing the null from distant alternatives. Nevertheless, the score and the CQLR tests have power close to size, and cannot separate the null from these alternatives at all.

6 Regions of low power

Regions of low power are present in many designs, not only the specific one presented in Section 5. For example, equation (4.16) is equivalent to the existence of standardized instruments coefficients μ so that $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu = 0$. What properties of the $2k \times 2k$ matrix Σ and μ imply that the noncentrality parameter of the L statistic is bounded for any value of Δ? The next proposition gives necessary and sufficient conditions for this to happen.[3]

Proposition 2 *For a $k \times k$ matrix A, define the Hermitian part of A by the symmetric matrix $H = (A + A')/2$. Then there exists $x \neq 0$ so that $x'Ax = 0$ if and only if the conex hul of H contains the zero value.*

We can apply this proposition to (4.16) by taking $x = \mu$ and $A = \Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$. Proposition 2 shows why it is not possible to have an impossibility design in a Kronecker product design (in particular, when the errors are homoskedastic). When $\Sigma = \Omega \otimes \Phi$ for a positive definite 2×2 matrix Ω and a symmetric positive definite matrix $k \times k$ matrix Φ, the matrix $\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$ is proportional to Φ and positive or negative definite.

This proposition also explains the findings in Section 5. There the matrix $\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$ is symmetric and proportional to the anti-diagonal J_k matrix. The trace of this matrix is 0 or 1 (depending on whether k is even or odd) and the determinant is negative. Therefore, there exist at least one positive eigenvalue and one negative eigenvalue. As a result, we have the reassurance there are coefficients μ so that the noncentrality parameter of the LM statistic is bounded, i.e., (4.16) holds.

Note that, if Σ_{11} is singular, it is in general trivial to separate the null and alternative hypotheses. For example, the AR test has a noncentrality parameter going to infinity under the alternative if one of the eigenvalues of Σ_{11} approaches zero for most parameter choices μ. If Σ_{11} is positive definite, we can apply Proposition 2 to $x = \Sigma_{11}^{-1}\mu$ and $A = \Sigma_{21}$. Hence, the impossibility design holds for all matrices Σ_{21} so that $\Sigma_{21} + \Sigma_{21}'$ does not have all eigenvalues of the same sign. For a given matrix Σ_{21}, we can choose Σ_{11} and Σ_{22} so that the variance matrix Σ is semi-positive definite. Hence, there exists a vast range of designs in which the non-centrality parameter of the LM statistic is bounded. In Appendix A, we also discuss cases in which (4.15) holds. and we have the impossibility design with the bound in (4.20) being arbitrarily close to zero.

[3]We are grateful to Leandro Gorno for suggesting the connection between this problem and the spectral decomposition of a symmetric matrix.

Table 1: Eigenvalues of $\Sigma_{21} + \Sigma_{21}'$ for model and data in Yogo (2004)

Australia	0.0014	0.0008	0.0001	0.0003
Canada	0.0030	0.0009	-0.0001	0.0002
France	-0.0011	0.0001	0.0008	0.0013
Germany	-0.0019	-0.0003	0.0005	0.0002
Italy	-0.0022	0.0013	0.0006	-0.0005
Japan	0.0027	-0.0006	0.0010	0.0008
Netherlands	0.0006	-0.0005	-0.6000	-0.0005
Sweden	0.0008	0.0003	-0.0004	0.0001
Switzerland	0.0005	0.0001	-0.0001	-0.0003
United Kingdom	-0.0032	0.0016	0.0003	0.0001
United States	0.0009	0.0006	0.0003	0.0011

Could we test whether this occurs by estimating π itself? If the instruments are weak, we consider $\pi = h_\pi/\sqrt{n}$. Because the parameter h_π is not consistently estimable, we cannot be sure whether the noncentrality parameter of the LM statistic is bounded for any value of Δ.

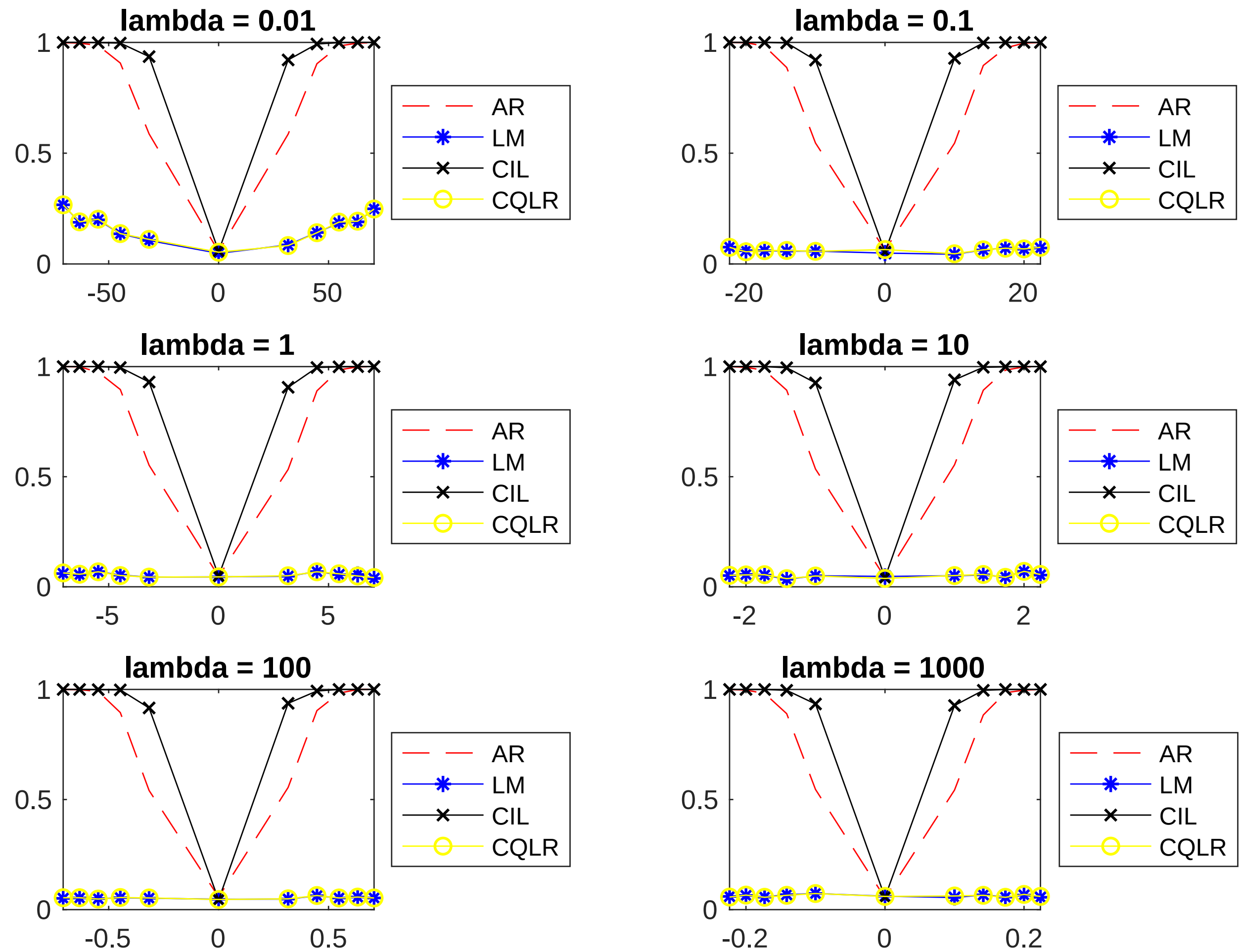
The impossibility designs were constructed such that the commonly used LM and CQLR tests have no power. When errors are homoskedastic, there exists a minimax result justifying the use of those tests. Andrews, Moreira, and Stock (2006) and Chamberlain (2007) implicitly use the Hunt-Stein theorem to justify the focus on tests that depend on the data only through $S'S$, $(S'T)^2$, and $T'T$. Andrews (2016) further shows there is no loss of generality from looking at conditional (on T statistic) tests that are linear combinations of the AR statistic $S'S$ and the LM statistic $(S'T)^2/T'T$. Our total variation argument, building on Kraft (1955), shows that conditional tests depending only on AR and LM statistics are not minimax tests in the general HAC model.

One may wonder if it is usual that empirical estimates of Σ_{21} have eigenvalues of $\Sigma_{21} + \Sigma_{21}'$ that are of opposite signs. As an example we take the estimation of IES in Yogo (2004). He considers four instruments and three different models. As Moreira and Moreira (2015) do, we focus on the model where the endogenous variable is the real stock return and the instruments are genuinely weak. Out of the eleven original countries considered by Yogo (2004) (Australia, Canada, France, Germany, Italy, Netherlands, Sweden, Switzerland, United Kingdom, and the United States), nine have eigenvalues with opposite signs. Eigenvalues of the estimates of $\Sigma_{21} + \Sigma_{21}'$, using the popular Newey-West estimator (Newey and West (1987)), are in Table 1. The condition for the LM non-centrality parameter to be bounded is satisfied for most countries.

Finally, someone can argue that the values of μ for which the LM non-centrality parameter is bounded, i.e., $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu = 0$, are too special. For example, for a given matrix Σ_{21}, the set of μ so that the LM non-centrality is bounded, has Lebesgue measure zero. This defense is questionable. Take for example the theory of limit experiments. If a family of models is locally asymptotically quadratic (LAQ) then it is locally asymptotically mixed normal (LAMN) except for a set with Lebesgue measure 0. Yet, there is a vast literature analyzing those special models of which the weak-IV model itself is a special case.

However, even if μ does not satisfy $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu = 0$, the problems with the LM and CQLR tests remain. Because of Theorem 3, the information loss also occurs in a neighborhood of the impossibility designs.

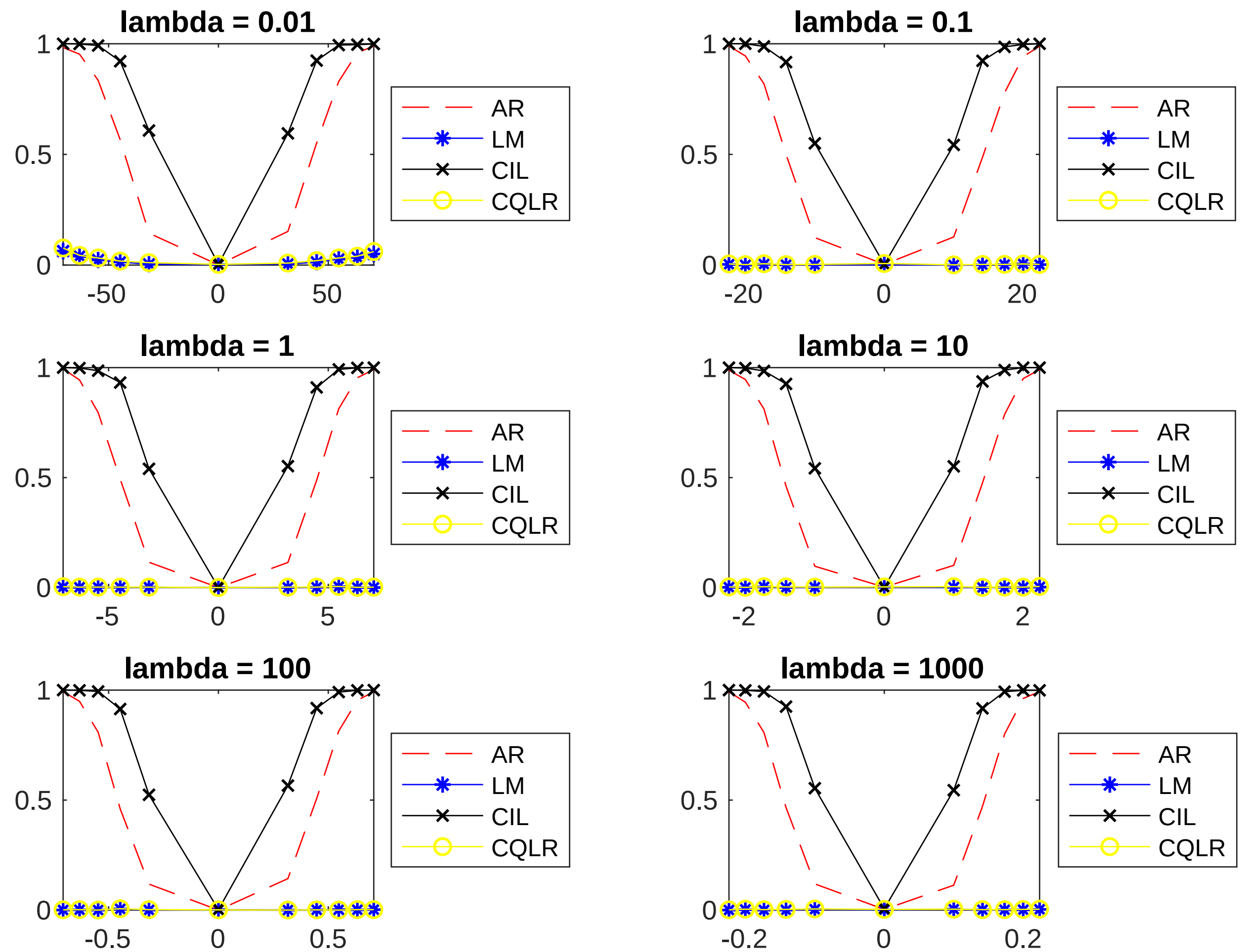
Figure 3: Power curves AR, LM, CQLR, and CIL tests for models with HAC errors with $c_{12} = 100$, $c_{11} = 1$ and $c_{22} = c_{12}^2 + c_{12}^{-3}$; varying instrument strength λ, $\alpha = .05$; perturbed variance matrix.



Theorem 3 *In the HAC-IV model:*
(i) *the power function $\rho_\phi(\beta, \mu)$ for any test ϕ is analytic in (β, μ); and*
(ii) *the power function $\rho_\phi(\beta, \mu)$ is uniformly continuous over any compact set.*

Part (i) shows the power function is analytic for any test. Close inspection of the proof shows that the expectation of any statistic (assuming the expectation exists) is also analytic

Figure 4: Power curves AR, LM, CQLR, and CIL tests for models with HAC errors with $c_{12} = 100$, $c_{11} = 1$ and $c_{22} = c_{12}^2 + c_{12}^{-3}$; varying instrument strength λ, $\alpha = .001$; perturbed variance matrix.



so that the result can be generalized for other (curved) exponential families. This is a stronger result than the differentiability of the expectation obtained by Hirano and Porter (2012, 2015) for such models. We apply part (ii) of Theorem 3 for compact sets containing values of μ such that $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu = 0$ and to values of β large enough that the total variation distance between the null and that specific alternative is close to 1. Small changes in both parameters change the power function by little. Let us consider the CIL and LM tests. The power function over the alternative of the CIL test gives a lower bound on the total variation distance of the hypotheses. Therefore for small changes in the parameters, the total variation distance remains close to one. Applying Theorem 3 to the LM test, the power of that test will be close to size for small changes in β.

In a simulation, we confirm that the LM and CQLR tests have power close to size, if $\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu$ is only close to 0. We use a variance matrix $\Sigma_\delta = \Sigma + \Psi_\delta$ with Σ a variance matrix of the impossibility design and Ψ_δ a small positive definite matrix. The $2k \times 2k$ matrix

Ψ_δ is obtained by drawing the columns of the $2k \times 2k$ matrix $X = [X_1 \ldots X_{2k}]$ independently from $N(0, \delta^2 \cdot I_{2k})$, so that $vec(X) \sim N(0, \delta^2 I_{4k^2})$. Define $P_0 = X\,(X'X)^{-1/2}$, which is by construction an orthogonal matrix. Let λ_δ be the diagonal matrix of the eigenvalues of $X'X$ that are non-negative. We take $\Psi_\delta = P_0 \lambda_\delta P_0'$, which is positive-definite with probability one. For each repetition, we first draw Σ_δ and next draw the equation errors of the HAC model using Σ_δ as their variance matrix. The power plots are averaged over the draws of X. Figures 3 and 4 show power curves for the AR, LM, CQLR, and CIL tests for $\delta^2 = 0.1$ at significance levels 5% and 0.1%, respectively. The power curves for the LM and CQLR tests are no longer flat, but the power problems of these tests remain. This confirms that the smoothness of the power function as stated in Theorem 3, implies loss of power of the LM and CQLR tests in a neighborhood of DGP, as in the impossibility design.

7 Conclusion

In a model with endogenous regressors and HAC errors, Moreira and Ridder (2018) find the largest transformation group that preserves the model and the hypothesis testing problem. They show symmetries in the HAC-IV model exist as long as the variance matrix of the HAC errors is taken as part of both the parameter and the data spaces.

When errors are homoskedastic, the theory simplifies to that of Andrews, Moreira, and Stock (2006) who show that the Anderson-Rubin, score, and rank statistics are one-to-one transformations of the maximal invariant. This finding contrasts with the model with HAC errors. In the HAC-IV model, tests that are a function of just the Anderson-Rubin, score, and rank statistics suffer from information loss. We show this by finding invariant statistics that are not functions of the AR, LM, and rank statistics, so that this triad is not maximal invariant in general. This information loss can be so extreme that the LM and CQLR tests can have power close to size. We give a set of necessary and sufficient conditions for the noncentrality parameter of the LM statistic to be bounded. If the value of the rank statistic is large, the CQLR essentially reduces to the LM statistic and can have power equal to size as well.

Applying the theory of Kraft (1955), we show that if a nearly perfect test exists, the (total variation) distance between the null and alternative is large. Because for a class of DGP the LM statistic has no information on the structural parameter, it fails to distinguish the null from the alternative. In testing contexts in which it is trivial to separate the null from the alternative, it is embarrassing that, if the instruments are weak, the LM and CQLR tests act as if there is no information at all. This efficiency loss in finite samples is particularly striking, because both tests are asymptotically efficient under the usual strong IV asymptotic theory. As a general remark, it is not enough to find tests with asymptotically efficient under the usual asymptotics and with correct size. Our theoretical framework and simulations findings highlight the importance of studying power under weak identification as well.

If we consider conditional tests that are functions of only the Anderson-Rubin and score statistics, their power will be bounded from above by the Anderson-Rubin test itself. To do better, we need to use tests that depend on the data beyond these statistics. Natural choices are the CIL test proposed by Moreira and Ridder (2018), and the SU and CLR tests derived by Moreira and Moreira (2015) for the HAC-IV model.[4] There are some numerical and theoretical advantages of the CIL test over the other tests. First, the CLR test requires numerical optimization to find the LR statistic. The SU test requires additional boundary conditions and linear programming methods to be implemented. Second, the CIL test uses model symmetries that yield a minimax result for HAC errors, in the same way that this literature was built on a minimax result for homoskedastic errors. The CIL test implicitly integrates the likelihood with respect to an associated Haar measure. It nicely connects to Kraft (1955) who establishes an affinity between the maximal smallest power and size of all tests and the total variation distance between the (convex hull of the) null and alternative hypotheses. Third, the CIL test gives more weight to distant alternatives. This can decrease the probability that the confidence set is uninformative, i.e. is the whole real line. These and other advantages of the CIL test will be discussed in greater detail in a separate paper.

References

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

Andrews, D., and P. Guggenberger (2015): "Identification- and Singularity-Robust Inference for Moment Condition Models," Working paper, Yale University.

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

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.

Andrews, D. W. K., and J. H. Stock (2007): "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, ed. by T. P. R. Blundell, W. K. Whitney, vol. 3, chap. 6. Cambridge University Press, Cambridge.

Andrews, I. (2016): "Conditional Linear Combination Tests for Weakly Identified Models," *Econometrica*, 84, 2155–2182.

Andrews, I., and A. Mikusheva (2016): "Conditional Inference with a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612.

Andrews, I., J. H. Stock, and L. Sun (2018): "Weak Instruments in IV Regression: Theory and Practice," Unpublished working paper.

Angrist, J., and J.-S. Pischke (2009): *Mostly Harmless Econometrics*. Princeton University Press.

[4]This theory favors the adaptation of the CLR test by Andrews and Mikusheva (2016) for moment models, in contrast to other GMM versions of the original LM and CLR tests designed for the IV model with homoskedastic errors.

Bertanha, M., and M. J. Moreira (2018): "Impossible Inference in Econometrics: Theory and Applications," arXiv:1612.02024.

Betrand, M., E. Duflo, and S. Mullainathan (2003): "How Much Should We Trust Differences-In-Differences Estimates?," *Quarterly Journal of Economics*, 119, 249–275.

Cameron, A. C., and D. L. Miller (2015): "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50(2), 317–372.

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

Dufour, J.-M. (2003): "Presidential Address: Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics*, 36, 767–808.

Finlay, K., and L. M. Magnussoni (2009): "Implementing Weak-Instrument Robust Tests for a General Class of Instrumental-Variables Models," *Stata Journal*, 9, 398–421.

Hirano, K., and J. Porter (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80, 1769–1790.

——— (2015): "Location Properties of Point Estimators in Linear Instrumental Variables and Related Models," *Econometric Reviews*, 34, 720–733.

Kadane, J. B. (1971): "Comparison of k-Class Estimators When the Disturbances Are Small," *Econometrica*, 39, 723–737.

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

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

Kloek, T. (1981): "OLS Estimation in a Model Where a Microvariable Is Explained by Aggregates and Contemporaneous Disturbances Are Equicorrelated," *Econometrica*, 49(1), 205–207.

Kraft, C. (1955): "SomeConditions for Consistency and Uniform Consistency of Statistical Procedures," *University of California Publications in Statistics*, 2(6), 125–142.

Lehmann, E. L., and J. P. Romano (2005): *Testing Statistical Hypotheses*. Third edn., Springer Series in Statistics.

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

Moreira, H., and M. J. Moreira (2015): "Optimal Two-Sided Tests for Instrumental Variables Regression with Heteroskedastic and Autocorrelated Errors," arXiv:1505.06644.

Moreira, M. (2009): "Tests with Correct Size when Instruments can be Arbitrarily Weak," *Journal of Econometrics*, 152(2), 131–140.

Moreira, M., and G. Ridder (2018): "Optimal Invariant Tests in an Instrumental Variable Regression with Heteroskedastic and Autocorrelated Errors," Discussion paper.

Moreira, M. J. (2001): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," *Center for Labor Economics Working Paper Series*, 37, UC Berkeley.

——— (2003): "A General Theory of Hypothesis Testing in the Simultaneous Equations Model," Harvard Institute of Economic Research Working Papers, no. 1992.

Moulton, B. R. (1986): "Random group effects and the precision of regression estimates," *Journal of Econometrics*, 32(3), 385–397.

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

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

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

Stock, J. H., J. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

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

8 Appendix A: Impossibility Design

In this appendix, we will find impossibility designs with an upper as in (4.20) that is arbitrarily close to zero. We already gave a necessary and sufficient condition on the Hermitian part of the matrix Σ_{12} for (4.16) to hold. We will give here conditions in terms of Σ_{12} itself so that (4.16) holds and the bound in (4.20) to be arbitrarily close to zero.

First of all, let us give necessary and sufficient conditions for Σ_0 to be positive definite. Note that

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} I_k & 0_{k\times k} \\ -\Sigma_{11}^{-1}\Sigma_{12} & I_k \end{bmatrix} \cdot \begin{bmatrix} \Sigma_{11} & 0_{k\times k} \\ 0_{k\times k} & \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix} \cdot \begin{bmatrix} I_k & -\Sigma_{12}'\Sigma_{11}^{-1} \\ 0_{k\times k} & I_k \end{bmatrix}.$$

Therefore, Σ_0 is positive definite if and only if Σ_{11} is positive definite and $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ is positive definite. Consider the Singular Value Decomposition (SVD) of $\Sigma_{12} = C\Lambda_{12}D'$, where C and D are orthogonal matrices and Λ_{12} is a diagonal matrix with singular values of Σ_{12}. Consider the design in which $\Sigma_{11} = C\Lambda_{11}C'$ and $\Sigma_{22} = D\Lambda_{22}D'$, where Λ_{11} and Λ_{22} are the eigenvalues of Σ_{11} and Σ_{22}, respectively. For Σ_0 to be positive definite, we then need the matrices Λ_{11} and $\Lambda_{11}\Lambda_{22} - \Lambda_{12}^2$ only have strictly positive components.

Now, we can easily have $\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2} \to 0$ because

$$\begin{aligned} \Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2} &= \Sigma_{11}^{-1/2}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{1/2} \\ &= C\Lambda_{11}^{-1/2}C'.D\left(\Lambda_{22} - \Lambda_{12}^2\Lambda_{11}^{-1}\right)^{1/2} D'. \end{aligned}$$

For example, we can have Λ_{11} bounded and each component of $\Lambda_{22} - \Lambda_{12}^2\Lambda_{11}^{-1}$ going to zero, or each component of Λ_{11} going to infinity and $\Lambda_{22} - \Lambda_{12}^2\Lambda_{11}^{-1}$ bounded. For the latter case in which Σ_{12} is fixed, we can have the bound

$$\frac{\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}$$

go to zero as well.

9 Appendix B: Proofs

Proof of Proposition 2. The LM_1 statistic is

$$LM_1 = \frac{\begin{matrix} U_S'\Sigma_{11}^{-1/2}\mu - \Delta U_S'\Sigma_{11}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\mu + U_S'\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T + \\ \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu + \Delta\mu'\Sigma_{11}^{-1}\mu + \Delta\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T \end{matrix}}{\left(\begin{matrix} \mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu - 2\Delta\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\mu + 2\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T \\ -2\Delta\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T + U_T'^{22})^{-1/2}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T \end{matrix}\right)^{1/2}}.$$

Divide the numerator and denominator by $1 + |\Delta|$. For all Δ

$$\begin{aligned}
\frac{1}{1+|\Delta|}\left|U_S'\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T\right| &\leq \left|U_S'\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T\right| \\
\frac{|\Delta|}{1+|\Delta|}\left|\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| &\leq \left|\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| \\
\frac{1}{(1+|\Delta|)^2}\left|\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| &\leq \left|\mu'\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| \\
\frac{\Delta^2}{(1+|\Delta|)^2}\left|\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| &\leq \left|\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}(\Sigma^{22})^{-1/2}U_T\right| \\
\frac{1}{1+|\Delta|}\left|\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T\right| &\leq \left|\Sigma_{11}^{-1/2}(\Sigma^{22})^{-1/2}U_T\right| .
\end{aligned}$$

For variance sequences that satisfy (4.15) and (4.17), the bounds are $o_p(1)$, so that the left-hand sides converge to 0 in probability uniformly in Δ. By the continuous mapping theorem that extends to sequences that converge uniformly, we find that uniformly in Δ

$$LM_1 \xrightarrow{p} \frac{U_S'\Sigma_{11}^{-1/2}\mu - \Delta U_S'\Sigma_{11}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\mu + \Delta\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu - 2\Delta\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\mu\right)^{1/2}} \sim$$

$$N\left(\frac{\Delta\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu - 2\Delta\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\mu\right)^{1/2}}, 1\right)$$

because by (4.16)

$$\mu'\Sigma_{11}^{-1/2}(I_k - \Delta\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{11}^{-1/2})(I_k - \Delta\Sigma_{11}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2})\Sigma_{11}^{-1/2}\mu = \mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu$$

The mean is bounded by

$$\left|\frac{\Delta\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\mu + \Delta^2\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)^{1/2}}\right| \leq \frac{\mu'\Sigma_{11}^{-1}\mu}{\left(\mu'\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{11}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\mu\right)}.$$

□

Proof of Proposition 2. Note that $x'Ax = x'A'x$. Therefore,

$$x'Ax = x'Hx, \text{ where } H = \frac{A + A'}{2}.$$

Write $H = C\Lambda C'$, where C is an orthogonal matrix composed by the eigenvectors of H and Λ is a diagonal matrix with the associated eigenvalues λ_i and. By writing $y = C'x$, we have

$$x'Bx = y'\Upsilon y = \sum\nolimits_{i=1}^{k} \lambda_i y_i^2,$$

where y_i is the i-th entry of the vector y. Because C is invertible, $x = 0$ if and only if $y = 0$.

Assume that all $\lambda_i > 0$. Then $\sum_{i=1}^{k} \lambda_i y_i^2 > 0$ for all $y \neq 0$. Analogously, if all $\lambda_i < 0$, then $\sum_{i=1}^{k} \lambda_i y_i^2 > 0$ for all $y \neq 0$.

On the other hand, suppose there exists no $y = (y_1, ..., y_k) \neq 0$ such that $\sum_{i=1}^{k} \lambda_i y_i^2 = 0$. Choose the canonical vectors $e_{(i)}$ for y. Then, each λ_i has to be either larger or smaller than zero. Now, assume that there exists one i and j so that $\lambda_i.\lambda_j < 0$ (that is, they have different signs). Then, we can choose the vector y so that $y_k = 0$ for all $k \neq i, j$, $y_i^2 = 1$ and $y_j^2 = -\lambda_i/\lambda_j > 0$. Contradiction. □

Proof of Theorem 3. The IV model belongs to a curved exponential family. We can write the likelihood of R as

$$f\left(r;\delta,\mu,\Sigma\right) = (2.pi)^{-k} \exp\left\{-\frac{1}{2} vec\left(R - [\delta : \mu]\right)' \Sigma^{-1} vec\left(R - [\delta : \mu]\right)\right\},$$

where $\delta = \mu\beta$. Importantly, the model is well-defined even when $\delta \neq \mu\beta$. The model can the be written as an exponential family for the parameter (δ, μ). From Theorem 2.7.1 of Lehmann and Romano (2005), the power function is analytic (because the test ϕ is bounded between zero and one). Because the transformation $\delta = \mu\beta$ is an analytic function, the power function is analytic as well. This proves part (i). Part (ii) follows from the fact that any continuous function is uniformly continuous over any compact set. □