

**Supplemental Material to
OPTIMAL TWO-SIDED INVARIANT SIMILAR TESTS FOR
INSTRUMENTAL VARIABLES REGRESSION**

BY

**DONALD W. K. ANDREWS, MARCELO J. MOREIRA
AND JAMES H. STOCK**

COWLES FOUNDATION PAPER NO. 1168S



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

2006

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

The copyright to this Article is held by the Econometric Society. It may be downloaded, printed and reproduced only for educational or research purposes, including use in course packs. No downloading or copying may be done for any commercial purpose without the explicit permission of the Econometric Society. For such commercial purposes contact the Office of the Econometric Society (contact information may be found at the website <http://www.econometricsociety.org> or in the back cover of *Econometrica*). This statement must be included on all copies of this Article that are made available electronically or in any other format.

SUPPLEMENT TO “OPTIMAL TWO-SIDED INVARIANT SIMILAR TESTS FOR INSTRUMENTAL VARIABLES REGRESSION”

(*Econometrica*, Vol. 74, No. 3, May 2006, 715–752)

BY DONALD W. K. ANDREWS, MARCELO J. MOREIRA, AND JAMES H. STOCK

1. INTRODUCTION

THIS PAPER CONTAINS supplemental material to the main paper (hereafter referenced as AMS). Tables of the conditional critical values of the conditional likelihood ratio (CLR) test are given in a separate file. Section 2 provides details concerning the sign-invariant power envelope for similar tests introduced in AMS. Section 3 does likewise for the locally unbiased (LU) power envelope for invariant similar tests. Section 4 reports additional numerical results. The figures for Section 4 are provided in a separate file. Section 5 establishes consistency of the covariance matrix estimator. Section 6 gives proofs of Lemmas 1 and 2. Section 7 proves the claim made in comment (ii) to Corollary 1 of AMS that when $k = 1$, the optimal invariant similar test in terms of two-point weighted average power is the Anderson–Rubin (AR) test (which is equivalent in this case to the Lagrange multiplier (LM) and CLR tests). An Appendix describes numerical methods used in Section 4 of this supplement.

2. POWER ENVELOPE FOR SIGN-INVARIANT TESTS

Here we consider similar tests that satisfy a sign-invariance condition in addition to the invariance condition of (3.1):

$$(S2.1) \quad [S : T] \rightarrow [-S : T].$$

The corresponding transformation in the parameter space is $(\beta^*, \lambda^*) \rightarrow (\beta_2^*, \lambda_2^*)$, where (β_2^*, λ_2^*) is defined in (4.1). This sign-invariance condition is a natural condition to impose to obtain two-sided tests, because the parameter vector (β_2^*, λ_2^*) is the appropriate “other-sided” parameter vector to (β^*, λ^*) for the reasons stated in AMS. The maximal invariant under this sign-invariance condition (plus the invariance conditions in (3.1)) is

$$(S2.2) \quad (S'S, |S'T|, T'T) = (Q_S, |Q_{ST}|, Q_T).$$

The likelihood ratio (LR), LM, and AR test statistics all depend on the data only through this maximal invariant and, hence, satisfy the sign-invariance condition (S2.1).

The density of the maximal invariant $(Q_S, |Q_{ST}|, Q_T)$ at (q_S, q_{ST}, q_T) for $q_{ST} \geq 0$, when the true parameters are (β^*, λ^*) , is

$$(S2.3) \quad \frac{1}{2} [f_{Q_1, Q_T}(q_S, q_{ST}, q_T; \beta^*, \lambda^*) + f_{Q_1, Q_T}(q_S, -q_{ST}, q_T; \beta^*, \lambda^*)].$$

Lemma 3(a) provides an expression for $f_{Q_1, Q_T}(q_S, q_{ST}, q_T)$. Straightforward calculations show that

$$(S2.4) \quad f_{Q_1, Q_T}(q_S, -q_{ST}, q_T; \beta^*, \lambda^*) = f_{Q_1, Q_T}(q_S, q_{ST}, q_T; \beta_2^*, \lambda_2^*)$$

using (4.1). Hence, the density of $(Q_S, |Q_{ST}|, Q_T)$ when the true parameters are (β^*, λ^*) equals $f_{Q_1, Q_T}^*(q_1, q_T; \beta^*, \lambda^*)$ as defined in (4.4). Now, following the same argument as in Section 4 of AMS, this implies that the power envelope for invariant similar tests using the invariance condition (3.1) coupled with (S2.1) is the same as the asymptotically efficient (AE) two-sided power envelope for invariant similar tests given in AMS.

3. POWER ENVELOPE FOR LOCALLY UNBIASED TESTS

3.1. Results

Another approach to constructing a power envelope designed for two-sided alternatives is to impose a necessary condition for unbiasedness—what we call a locally unbiased condition. This approach has a long tradition in the statistics literature and is a standard way to derive optimal tests for two-sided alternatives. In exponential families, uniformly most powerful (UMP) two-sided tests exist among the class of unbiased tests; see Lehmann (1986, Theorem 4.3, p. 147). This is not the case in the curved exponential family testing problem considered here. Nevertheless, one can construct a power envelope for LU invariant similar tests.

We start by determining two necessary conditions for an invariant test (with respect to (3.1)) to be unbiased. The first condition is similarity and the second condition is the requirement that the power function has zero derivative at the null hypothesis. Otherwise, the power function would dip below the size of the test for some alternatives close to the null.

THEOREM S.1: *An invariant test $\phi(Q)$ is unbiased with size α only if $E_{\beta_0}(\phi(Q)|Q_T = q_T) = \alpha$ and $E_{\beta_0}(\phi(Q)Q_{ST}|Q_T = q_T) = 0$ for almost all q_T .*

COMMENTS: (i) The first condition establishes that all unbiased invariant tests must be similar; the second establishes that the power function of an unbiased test must have zero derivative under H_0 . The two conditions together are what we call the LU condition. (Note that the two conditions are only first-order conditions, not sufficient conditions, for a test's power function to have a local minimum at the null hypothesis.) Obviously, the class of LU tests contains the class of unbiased tests.

(ii) The two conditions in Theorem S.1 are closely related to the conditions used for two-sided alternatives in the classical hypothesis testing theory for exponential families; see Lehmann (1986, Chap. 4).

(iii) The second condition of Theorem S.1 is equivalent to

$$(S3.1) \quad E_{\beta_0}(\phi(Q)Q_{ST}/Q_T^{1/2}) = 0.$$

That is, any LU invariant test statistic $\phi(Q)$ must be uncorrelated with the pivotal statistic $Q_{ST}/Q_T^{1/2}$ under H_0 .¹

The LR, LM, and AR test statistics depend on the data through $(Q_S, |Q_{ST}|, Q_T)$. The following result shows that these tests satisfy the second condition of Theorem S.1.

COROLLARY S.1: *Any similar level α test that depends on the observations through $(Q_S, |Q_{ST}|, Q_T)$ satisfies the LU condition of Theorem S.1.*

Next, we determine the test that maximizes power against any given parameter vector (β^*, λ^*) among the class of LU invariant tests. We do so using the same conditioning argument as in Section 4 of AMS, and using the generalized Neyman–Pearson lemma (see Lehmann (1986, Theorem 3.5, pp. 96–97)). Define

$$(S3.2) \quad \begin{aligned} \text{LR}(q_1, q_T; \beta^*, \lambda^*) &= \frac{f_{Q_1, Q_T}(q_1, q_T; \beta^*, \lambda^*)}{f_{Q_T}(q_T; \beta^*, \lambda^*)f_{Q_1|Q_T}(q_1|q_T; \beta_0)} \\ &= \frac{\psi(q_1, q_T; \beta^*, \lambda^*)}{\psi_2(q_T; \beta^*, \lambda^*)}. \end{aligned}$$

THEOREM S.2: *The test that maximizes power against $(\beta, \lambda) = (\beta^*, \lambda^*)$ among LU invariant tests with significance level α rejects H_0 if*

$$\text{LR}(Q_1, Q_T; \beta^*, \lambda^*) > \kappa_{1\alpha}(Q_T; \beta^*, \lambda^*) + Q_{ST}\kappa_{2\alpha}(Q_T; \beta^*, \lambda^*),$$

where $\kappa_{1\alpha}(Q_T; \beta^*, \lambda^*)$ and $\kappa_{2\alpha}(Q_T; \beta^*, \lambda^*)$ are chosen such that the two conditions in Theorem S.1 hold.

COMMENT: The power of the tests $\text{LR}(Q_1, Q_T; \beta^*, \lambda^*)$ as (β^*, λ^*) varies maps out the power envelope for LU invariant tests.

3.2. Proofs of Local-Unbiasedness Results

PROOF OF THEOREM S.1: By continuity of the power function, which holds by Lehmann (1986, Theorem 2.9, p. 59), any unbiased test $\phi(Q)$ is similar. Hence, the first condition of Theorem S.1 holds by Theorem 2.

¹The second condition of Theorem S.1 clearly implies (S3.1). The converse holds by the completeness of Q_T , because by iterated expectations the left-hand side in (S3.1) can be written as $E_{\beta_0}h(Q_T)$, where $h(Q_T) = E_{\beta_0}(\phi(Q)Q_{ST}|Q_T = q_T)/Q_T^{1/2}$.

Now, for a test to be unbiased, $(\partial/\partial\beta)E_{\beta,\lambda}\phi(Q_1, Q_T)|_{\beta=\beta_0} = 0$ for all values of λ . By interchanging derivatives and integrals (which is justified by Lehmann (1986, Theorem 2.9, p. 59)) and the chain rule, the left-hand side of this equality equals $I_1 + I_2$, where

$$(S3.3) \quad I_1 = \int \int \phi(q_1, q_T) \frac{\partial f_{Q_1|Q_T}(q_1|q_T; \beta_0, \lambda)}{\partial \beta} dq_1 f_{Q_T}(q_T; \beta_0, \lambda) dq_T,$$

$$I_2 = \int \int \phi(q_1, q_T) f_{Q_1|Q_T}(q_1|q_T; \beta_0) dq_1 \frac{\partial f_{Q_T}(q_T; \beta_0, \lambda)}{\partial \beta} dq_T$$

$$= \int \alpha \frac{\partial f_{Q_T}(q_T; \beta_0, \lambda)}{\partial \beta} dq_T = 0,$$

where the second to the last equality holds by the condition for similarity and the last equality holds because $\int f_{Q_T}(q_T; \beta, \lambda) dq_T = 1$ for all β .

To compute the derivative of the conditional density of Q_1 given $Q_T = q_T$ with respect to β evaluated at β_0 , it is convenient to write the conditional density of Q_1 given $Q_T = q_T$ as

$$(S3.4) \quad f_{Q_1|Q_T}(q_1|q_T; \beta, \lambda)$$

$$= K_1 K_2^{-1} \exp\left(-\frac{\lambda c_\beta^2}{2}\right) \exp\left(-\frac{q_S}{2}\right) \det(q)^{(k-3)/2} q_T^{-(k-2)/2}$$

$$\times \sum_{j=0}^{\infty} \frac{(\lambda \xi_\beta(q)/4)^j}{j! \Gamma((k-2)/2 + j + 1)} \bigg/ \sum_{j=0}^{\infty} \frac{(\lambda d_\beta^2 q_T/4)^j}{j! \Gamma((k-2)/2 + j + 1)}$$

using Lemma 3(a) and (b) and (4.8).

Tedious algebraic manipulations show that

$$(S3.5) \quad \frac{\partial f_{Q_1|Q_T}(q_1|q_T; \beta_0, \lambda)}{\partial \beta} = \frac{\lambda^{1/2}}{2(d_\beta^2 q_T)^{1/2}} f_{Q_1|Q_T}(q_1|q_T; \beta_0) q_{ST} (\det(\Omega))^{-1/2}$$

$$\times \frac{I_{k/2}(\sqrt{\lambda a'_0 \Omega^{-1} a_0 q_T})}{I_{(k-2)/2}(\sqrt{\lambda a'_0 \Omega^{-1} a_0 q_T})}.$$

The function $I_{k/2}(\cdot)$ arises because

$$(S3.6) \quad \frac{\partial}{\partial \beta} \sum_{j=0}^{\infty} \frac{(\lambda \xi_\beta(q)/4)^j}{j! \Gamma((k-2)/2 + j + 1)} = \frac{\lambda}{4} \frac{\partial \xi_\beta(q)}{\partial \beta} \sum_{s=0}^{\infty} \frac{(\lambda \xi_\beta(q)/4)^s}{s! \Gamma(k/2 + s + 1)}$$

and likewise with $\xi_\beta(q)$ replaced by $(d_\beta^2 q_T)$.

The necessary condition for unbiasedness, (S3.3), and (S3.5) give

$$(S3.7) \quad 0 = \int h(q_T) f_{Q_T}(q_T; \beta_0, \lambda) \frac{I_{k/2}(\sqrt{\lambda a'_0 \Omega^{-1} a_0 q_T})}{I_{(k-2)/2}(\sqrt{\lambda a'_0 \Omega^{-1} a_0 q_T})} dq_T, \quad \text{where}$$

$$h(q_T) = \int \phi(q_1, q_T) q_{ST} f_{Q_1|Q_T}(q_1|q_T; \beta_0) dq_1.$$

By completeness of Q_T under H_0 (see the proof of Theorem 2), it must be the case that $h(q_T)$ is zero for almost all q_T and all $\lambda \geq 0$, which yields the second condition of the theorem. *Q.E.D.*

PROOF OF COROLLARY S.1: Any test that depends on (Q_S, Q_{ST}^2, Q_T) can be written as $\phi(Q_S, S_2^2, Q_T)$, where $S_2 = Q_{ST}/(Q_S Q_T)^{1/2}$. By Lemma 3(e) and (f), Q_S , S_2 , and Q_T are independent under H_0 , and S_2 has a distribution that is symmetric about zero. Hence, we have

$$(S3.8) \quad E_{\beta_0}(\phi(Q_S, S_2^2, Q_T) Q_{ST} | Q_T = q_T)$$

$$= E_{\beta_0}(\phi(Q_S, S_2^2, q_T) S_2 Q_S^{1/2}) q_T^{1/2}$$

$$= \int E_{\beta_0}(\phi(q_S, S_2^2, q_T) S_2) q_S^{1/2} f_{Q_S}(q_S) dq_S \cdot q_T^{1/2} = 0$$

for all q_T , where the last equality holds because $\phi(q_S, S_2^2, q_T) S_2$ is an odd function of S_2 and S_2 is symmetrically distributed about zero. *Q.E.D.*

PROOF OF THEOREM S.2: By the same argument as in Section 4 of AMS, it suffices to find the test that maximizes power against the single alternative conditional density $f_{Q_1|Q_T}(q_1|q_T; \beta^*, \lambda^*)$ conditional on $Q_T = q_T$. Given the restriction to LU tests, we apply the generalized Neyman–Pearson (GNP) lemma (see Lehmann (1986, Theorem 3.5, pp. 96–97)). The GNP lemma implies that the optimal (conditional) test rejects when $\text{LR}(Q_1, q_T; \beta^*, \lambda^*) > \kappa_{1\alpha}(q_T; \beta^*, \lambda^*) + \kappa_{2\alpha}(q_T; \beta^*, \lambda^*) Q_{ST}$ for some $\kappa_{1\alpha}(q_T; \beta^*, \lambda^*)$ and $\kappa_{2\alpha}(q_T; \beta^*, \lambda^*)$ that are chosen such that the two conditions of Theorem S.1 hold.

It remains to verify the conditions needed to apply the generalized Neyman–Pearson lemma. Let M be the set of points

$$(S3.9) \quad (E(\phi(Q_1, Q_T) | Q_T = q_T), E(\phi(Q_1, Q_T) Q_{ST} | Q_T = q_T))$$

as ϕ ranges over all possible critical functions. It suffices to show that $(\alpha, 0)$ is an interior point of M , see Lehmann (1986, Theorem 3.5(iv), p. 97).

The set M is convex because the conditional expectation operator is linear. Moreover, M contains $(\alpha, 0)$ by considering the LM test. It also contains points (α, u_α^+) with $u_\alpha^+ > 0$ by considering the one-sided LM test, which re-

jects H_0 when $Q_{ST}/Q_T^{1/2} > c_\alpha$. This follows because the derivative of the conditional power function of this test is an increasing linear transformation of

$$(S3.10) \quad \int \mathbb{1}(q_{ST}/q_T^{1/2} > c_\alpha) q_{ST} f_{Q_1|Q_T}(q_1|q_T; \beta_0) dq_1,$$

which is strictly positive. Likewise, M also contains points (α, u_α^-) with $u_\alpha^- < 0$ by considering the test that rejects H_0 when $-Q_{ST}/Q_T^{1/2} > c_\alpha$ by an analogous argument. This completes the verification that $(\alpha, 0)$ lies in the interior of M . *Q.E.D.*

4. NUMERICAL RESULTS: MODEL WITH KNOWN COVARIANCE MATRIX

This section reports numerical results for power envelopes and power functions of two-sided invariant similar tests. A representative subset of these results is reported in AMS.

Throughout, we focus on tests with significance level 5% and, without loss of generality, on the case $\beta_0 = 0$. As discussed in AMS, the remaining parameters that characterize the distribution of the tests are λ , k , $\rho = \text{corr}(v_{1i}, v_{2i})$, and the alternative, β . (The distribution of Q and thus the power depends on the sample size only through λ .) Results are reported here for $\lambda/k = 0.5, 1, 2, 4, 8$, and 16, which spans the range from weak to strong instruments, for $\rho = 0.95, 0.5$, and 0.2, and for $k = 2, 5, 10$, and 20. Numerical issues are discussed in the Appendix.

The results are summarized in Figures S1–S6 which are provided in a separate file. As in AMS, the horizontal axis is the scaled “local” alternative, $\sqrt{\lambda}\beta$.

Figure S1 presents the asymptotically efficient two-sided power envelope (for invariant similar tests) and the power envelope for LU invariant tests for $k = 5$.

Figure S2 presents the asymptotically efficient two-sided power envelope and the power functions of the two-sided CLR, LM, and AR tests for $k = 2$. These power envelopes and power functions are plotted for $k = 5, 10$, and 20 in Figures S3, S4, and S5, respectively.

Figure S6 presents the asymptotically efficient two-sided power envelope and the power functions of two two-point optimal invariant similar (POIS2) tests, labeled POIS2a and POIS2b. One approach to testing when there is not a UMPI invariant test is to consider a point optimal invariant (POI) test that has a power function tangent to the power envelope at a certain value. If the power functions remain sufficiently close to the power envelope against alternatives other than that for which the test is point optimal, then that particular POI test provides a good practical choice (cf. King (1988)). Specifically, the POIS2a test is the LR* test against $(\beta^* = 0.8, \lambda^* = 5)$, and the POIS2b test is the LR* test against $(\beta^* = 1.45, \lambda^* = 5)$. When $\lambda^* = 5$ and $\rho = 0.95$, the power functions of the POIS2a and POIS2b tests are tangent to the power envelope

at approximately 25% power and 75% power, respectively. Like the power envelope itself, these tests depend on ρ ; these tests are infeasible if ρ is unknown, although a feasible version of these tests could be computed by plugging in a consistent estimator of ρ .

Summary of Findings

(i) In theory, the two-sided asymptotically efficient (AE) power envelope is no higher than the LU power envelope. Numerically, it appears from Figure S1 that the two power envelopes are essentially the same for most values of ρ , λ , and β . In the cases considered, the greatest difference occurs in Figure S1(c) for $\sqrt{\lambda}\beta \cong 2.2$, where the difference between the two power envelopes is approximately 0.05.

(ii) In Figures S2–S5, the two-sided CLR test has a power function that is essentially on the power envelope in every case.

(iii) Figures S2–S5 extend previous findings that the LM statistic can have a nonmonotone power function and has poor power properties compared with the CLR statistic.

(iv) Figures S2–S5 show that the AR statistic can have power well below the AE power envelope and well below that of the CLR statistic. The power gap increases with k . This gap is present for both weak and strong instruments.

(v) The strong instrument local alternative asymptotic results in AMS indicate that the LM and CLR tests have power functions that coincide with the AE power envelope when λ is large. As a numerical matter, in the cases considered in Figures S2–S5, for most values of k , ρ , and β , this coincidence occurs by $\lambda/k = 16$, sometimes by $\lambda/k = 8$.

(vi) Comparing the AE two-sided power envelope across all the panels of Figures S2–S5 reveals that the AE power envelope, viewed as a function of $\sqrt{\lambda}\beta$, takes on similar values regardless of ρ , k , or λ/k . Said differently, $\sqrt{\lambda}\beta$ evidently is the appropriate “local” parameterization.

(vii) The power functions of the POIS2a and POIS2b tests are close to the AE two-sided power envelope. For $\rho = 0.5$, the POIS2a performs slightly better than the POIS2b. The very good performance of the POIS2a suggests that further work on a feasible version of this test (in which ρ is estimated) is a promising way to develop a new test, based on the theory of tangency testing, that has power approaching the AE two-sided power envelope. A theoretical advantage of this test over the CLR test is that it is known to be asymptotically admissible among sign-invariant tests (under weak instrument asymptotics). The practical disadvantage of this test, relative to the CLR test, is that it is numerically more difficult to work with because it involves Bessel functions, whereas the CLR test does not. Moreover, the results in Figures S2–S5 indicate that there is little room for improvement over the CLR test.

5. CONSISTENCY OF THE COVARIANCE MATRIX ESTIMATOR

In AMS, the covariance matrix Ω ($\in \mathbb{R}^{2 \times 2}$) (defined in Assumption 2) is estimated via

$$(S5.1) \quad \widehat{\Omega}_n = (n - k - p)^{-1} \widehat{V}' \widehat{V}, \quad \text{where} \quad \widehat{V} = Y - P_Z Y - P_X Y,$$

where k and p are the dimensions of Z_i and X_i , respectively. Let \widehat{V}_i denote the i th row of \widehat{V} written as a column 2-vector.

Under Assumptions 1–3, the variance estimator is consistent.

LEMMA S.1: *Under Assumptions 1–3, $\widehat{\Omega}_n \rightarrow_p \Omega$.*

COMMENT: The convergence in the lemma occurs uniformly over all true parameters β , C , γ , and ξ no matter what the parameter space is. This can be seen by inspection of the proof of the lemma.

PROOF OF LEMMA S.1: Using the definition $Y = Z\pi a' + X\eta + V$, we obtain $\widehat{V} = V - P_Z V - P_X V$. This and $P_Z P_X = 0$ give

$$(S5.2) \quad n^{-1} \widehat{V}' \widehat{V} - \Omega = (n^{-1} V' V - \Omega) - n^{-1} V' P_Z V - n^{-1} V' P_X V.$$

The first summand on the right-hand side of (S5.2) converges in probability to zero by Assumption 2. The second summand satisfies

$$(S5.3) \quad 0 \leq n^{-1} V' P_Z V \leq n^{-1} V' P_{\bar{Z}} V \\ = n^{-1} (n^{-1/2} V' \bar{Z}) (n^{-1} \bar{Z}' \bar{Z})^{-1} (n^{-1/2} \bar{Z}' V) \rightarrow_p 0,$$

where the second inequality holds because the span of Z is contained in the span of \bar{Z} and the convergence to zero holds by Assumptions 1 and 3. The third summand of (S5.2) converges in probability to zero by an analogous argument. *Q.E.D.*

6. PROOFS OF LEMMAS 1 AND 2

Here, we state Lemmas 1 and 2 and provide proofs of these lemmas.

The two equation reduced-form model can be written in matrix notation as

$$(S6.1) \quad Y = Z\pi a' + X\eta + V, \quad \text{where} \\ Y = [y_1 : y_2], \quad V = [v_1 : v_2], \\ a = (\beta, 1)', \quad \eta = [\gamma : \xi].$$

The distribution of $Y \in \mathbb{R}^{n \times 2}$ is multivariate normal with mean matrix $Z\pi a' + X\eta$, independence across rows, and covariance matrix Ω for each row. The

parameter space for $\theta = (\beta, \pi', \gamma', \xi')'$ is taken to be $\mathbb{R} \times \mathbb{R}^k \times \mathbb{R}^p \times \mathbb{R}^p$.

LEMMA AMS-1: For the model in (S6.1):

- (a) $Z'Y$ and $X'Y$ are sufficient statistics for θ ;
- (b) $Z'Y$ and $X'Y$ are independent;
- (c) $X'Y$ has a multivariate normal distribution that does not depend on $(\beta, \pi')'$;
- (d) $Z'Y$ has a multivariate normal distribution that does not depend on $\eta = [\gamma; \xi]$;
- (e) $Z'Y$ is a sufficient statistic for $(\beta, \pi')'$.

LEMMA AMS-2: For the model in (S6.1):

- (a) $S \sim N(c_\beta \mu_\pi, I_k)$;
- (b) $T \sim N(d_\beta \mu_\pi, I_k)$;
- (c) S and T are independent.

PROOF OF LEMMA AMS-1: Let $Z = [Z_1 : \dots : Z_n]'$ and let $X = [X_1 : \dots : X_n]'$. The distribution of Y is multivariate normal with

$$(S6.2) \quad EY = Z\pi a' + X\eta,$$

independence across rows, and covariance matrix Ω for each row. Hence, the density of Y evaluated at the $n \times 2$ matrix $y = [y_1 : \dots : y_n]'$ is

$$(S6.3) \quad (2\pi)^{-n/2} |\Omega|^{-n/2} \times \exp\left(-\frac{1}{2} \sum_{i=1}^n (Y_i - a\pi'Z_i - \eta'X_i)' \Omega^{-1} (Y_i - a\pi'Z_i - \eta'X_i)\right) \\ = (2\pi)^{-n/2} |\Omega|^{-n/2} \exp\left(-\frac{1}{2} \left[\sum_{i=1}^n Y_i' \Omega^{-1} Y_i - 2\pi' \left(\sum_{i=1}^n Z_i Y_i' \right) \Omega^{-1} a \right. \right. \\ \left. \left. - 2 \operatorname{tr} \left(\left(\sum_{i=1}^n X_i Y_i' \right) \Omega^{-1} \eta' \right) \right. \right. \\ \left. \left. + \sum_{i=1}^n (a\pi'Z_i - \eta'X_i)' \Omega^{-1} (a\pi'Z_i - \eta'X_i) \right] \right).$$

If a density can be factorized as $p_\theta(x) = f_\theta(T(x))h(x)$, then $T(X)$ is a sufficient statistic for θ . In consequence, given that Ω is known, Z_i and X_i are fixed and known, $a = (\beta, 1)'$, and $\eta = [\gamma; \xi]$, sufficient statistics for $\theta = (\beta, \pi', \gamma', \xi')'$ are $\sum_{i=1}^n Z_i Y_i' = Z'Y$ and $\sum_{i=1}^n X_i Y_i' = X'Y$, and part (a) of the lemma holds.

To prove part (b) of the lemma, note that $Z'Y$ and $X'Y$ are (jointly) multivariate normal random matrices and $Z'X = 0$. For any $m_1, m_2 \in \mathbb{R}^2$, we have

$$\begin{aligned}
 \text{(S6.4)} \quad \text{cov}(Z'Y m_1, X'Y m_2) &= \text{cov}\left(\sum_{i=1}^n Z_i Y_i' m_1, \sum_{i=1}^n X_i Y_i' m_2\right) \\
 &= \sum_{i=1}^n Z_i X_i' \text{cov}(Y_i' m_1, Y_i' m_2) = Z'X \cdot m_1' \Omega m_2 = 0,
 \end{aligned}$$

where the second equality uses independence across i and the third equality uses the assumption that the covariance matrix Ω of Y_i does not depend on i . Hence, $Z'Y$ and $X'Y$ are independent.

The distribution of $X'Y$ is multivariate normal with variances and covariances that depend on X and Ω , but not on θ , and with mean

$$\text{(S6.5)} \quad X'EY = X'(Z\pi a' + X\eta) = X'X\eta$$

because $X'Z = 0$. Hence, the distribution of $X'Y$ does not depend on (β, π) and part (c) of the lemma holds.

The distribution of $Z'Y$ is multivariate normal with variances and covariances that depend on Z and Ω , but not on θ , and with mean

$$\text{(S6.6)} \quad Z'EY = Z'(Z\pi a' + X\eta) = Z'Z\pi a'$$

because $Z'X = 0$. Hence, the distribution of $Z'Y$ does not depend on (γ, ξ) and part (d) of the lemma holds.

Part (e) of the lemma follows from parts (b)–(d). *Q.E.D.*

PROOF OF LEMMA AMS-2: The k -vector S is multivariate normal with mean

$$\begin{aligned}
 \text{(S6.7)} \quad ES &= (Z'Z)^{-1/2} Z'EY b_0 \cdot (b_0' \Omega b_0)^{-1/2} \\
 &= (Z'Z)^{-1/2} Z'(Z\pi a' + X\eta) b_0 \cdot (b_0' \Omega b_0)^{-1/2} = c_\beta \mu_\pi,
 \end{aligned}$$

using (S6.2), $Z'X = 0$, and $a'\beta_0 = \beta - \beta_0$. We have

$$\begin{aligned}
 \text{(S6.8)} \quad \text{var}(Z'Y b_0) &= \text{var}\left(\sum_{i=1}^n Z_i Y_i' b_0\right) = \sum_{i=1}^n Z_i Z_i' \text{var}(Y_i' b_0) \\
 &= \sum_{i=1}^n Z_i Z_i' b_0' \Omega b_0 = Z'Z b_0' \Omega b_0.
 \end{aligned}$$

Hence, from the definition of S , $\text{var}(S) = I_k$ and part (a) of the lemma holds.

The k -vector T is multivariate normal with mean

$$(S6.9) \quad \begin{aligned} ET &= (Z'Z)^{-1/2} Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \\ &= (Z'Z)^{-1/2} Z'(Z\pi a' + X\eta)\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} = d_\beta \mu_\pi. \end{aligned}$$

From (S6.8) with b_0 replaced by $\Omega^{-1}a_0$, we have $\text{var}(Z'Y\Omega^{-1}a_0) = Z'Za_0'\Omega^{-1}a_0$. Hence, from the definition of T , $\text{var}(T) = I_k$ and part (b) of the lemma holds.

The random vectors S and T are independent because they are nonstochastic functions of $Z'Yb_0$ and $Z'Y\Omega^{-1}a_0$, respectively, and the latter are jointly multivariate normal with covariance given by

$$(S6.10) \quad \begin{aligned} \text{cov}(Z'Yb_0, Z'Y\Omega^{-1}a_0) &= \text{cov}\left(\sum_{i=1}^n Z_i Y_i' b_0, \sum_{i=1}^n Z_i Y_i' \Omega^{-1} a_0\right) \\ &= \sum_{i=1}^n Z_i Z_i' \text{cov}(Y_i' b_0, Y_i' \Omega^{-1} a_0) = \sum_{i=1}^n Z_i Z_i' b_0' \Omega^{-1} a_0 = 0, \end{aligned}$$

using $b_0' a_0 = 0$. Hence, part (c) of the lemma holds.

Q.E.D.

7. RESULTS FOR THE JUST-IDENTIFIED MODEL

Here we prove the claim given in comment (ii) following Corollary 1 that $\psi(q_1, q_T; \beta^*, \lambda^*) + \psi(q_1, q_T; \beta_2^*, \lambda_2^*)$ is increasing in S^2 when $k = 1$. (It is strictly increasing unless $S = 0$.) This claim leads to the result that the AR, LM, and CLR tests (which are equivalent when $k = 1$) maximize average power against (β^*, λ^*) and (β_2^*, λ_2^*) for all (β^*, λ^*) in the class of invariant similar tests. That is, these tests are UMP two-sided invariant similar tests.

By the definition given in Corollary 1, we have

$$(S7.1) \quad \begin{aligned} \psi(q_1, q_T; \beta, \lambda) &= \exp(-\lambda(c_\beta^2 + d_\beta^2)/2)(\lambda\xi_\beta(q))^{-(k-2)/4} I_{(k-2)/2}(\sqrt{\lambda\xi_\beta(q)}). \end{aligned}$$

When $k = 1$, we have

$$(S7.2) \quad \begin{aligned} I_{-1/2}(x) &= x^{-1/2}(2/\pi i)^{1/2}(\exp(x) + \exp(-x))/2 \\ &= x^{-1/2}(2/\pi i)^{1/2} \cosh(x); \end{aligned}$$

see comment (ii) to Lemma 3. When $k = 1$, $Q_{ST} = S \cdot T$. Using this and Equations (4.1) and (4.7), we have

$$(S7.3) \quad \lambda^* \xi_{\beta^*}(q) = \lambda^*(c_{\beta^*} S + d_{\beta^*} T)^2,$$

$$\begin{aligned}\lambda_2^* \xi_{\beta_2^*}(q) &= \lambda^*(c_{\beta^*} S - d_{\beta^*} T)^2, \\ \lambda^*(c_{\beta^*}^2 + d_{\beta^*}^2) &= \lambda_2^*(c_{\beta_2^*}^2 + d_{\beta_2^*}^2).\end{aligned}$$

Combining (S7.1)–(S7.3) gives

$$\begin{aligned}\text{(S7.4)} \quad & \psi(q_1, q_T; \beta^*, \lambda^*) + \psi(q_1, q_T; \beta_2^*, \lambda_2^*) \\ &= \frac{1}{4} \exp(-\lambda^*(c_{\beta^*}^2 + d_{\beta^*}^2)/2) (2/\pi i)^{1/2} \\ & \quad \times (\cosh(\sqrt{\lambda^*}(c_{\beta^*} S + d_{\beta^*} T)) + \cosh(\sqrt{\lambda^*}(c_{\beta^*} S - d_{\beta^*} T))),\end{aligned}$$

using the fact that $\cosh(\cdot)$ is symmetric about zero.

Define

$$\text{(S7.5)} \quad h_1(x) = \exp(x + K) + \exp(-x - K) + \exp(x - K) + \exp(-x + K),$$

where K is a constant. The function $h_1(x)$ only depends on x through $|x|$ because it is symmetric in x . The same is true for K , so without loss of generality assume $K \geq 0$. We show that $h_1(x)$ is increasing in $|x|$ by showing that its derivative is nonnegative for $x \geq 0$. Combining this with (S7.4) and the definition of $\cosh(\cdot)$ gives the desired result that the left-hand side of (S7.4) is increasing in $|S|$.

The derivative of $h_1(x)$ is

$$\text{(S7.6)} \quad h_1'(x) = \exp(x + K) - \exp(-x - K) + \exp(x - K) - \exp(-x + K).$$

Whereas $x \geq 0$ and $K \geq 0$, there are two cases to consider: (i) $0 \leq K \leq x$ and (ii) $0 \leq x \leq K$. For case (i), we have $\exp(x + K) - \exp(-x - K) \geq 0$ and $\exp(x - K) - \exp(-x + K) \geq 0$ because $x + K \geq 0$, $x - K \geq 0$, and $\exp(\cdot)$ is increasing. For case (ii), we have $\exp(x + K) - \exp(-x + K) \geq 0$ and $\exp(x - K) - \exp(-x - K) \geq 0$ because $x \geq 0$ and $\exp(\cdot)$ is increasing. Hence, $h_1'(x)$ is nonnegative. The inequalities are strict if $x > 0$.

Cowles Foundation for Research in Economics, Yale University, 30 Hillhouse Ave., Box 208281, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu,

Dept. of Economics, Harvard University, Littauer Center M-6, 1875 Cambridge St., Cambridge, MA 02138, U.S.A.; moreira@fas.harvard.edu,

and

Dept. of Economics, Harvard University, Littauer Center M-6, 1875 Cambridge St., Cambridge, MA 02138, U.S.A.; James_Stock@harvard.edu.

APPENDIX: NUMERICAL METHODS

A.1. *General Numerical Notes*

(a) All numerical results are based on 5,000 Monte Carlo draws of Q except that CLR, LM, and AR power functions are computed using 10,000 draws, and the asymptotically efficient two-sided power envelopes for $k = 2$ and $k = 10$ are based on 2,000 draws. All computations were done in GAUSS.

(b) The statistics ψ and ψ_2 have a range of many orders of magnitude, so they were computed in logarithms. The statistic ψ_2 is a function of Q_T only so that, in theory, the denominator term in the LR* statistic can be absorbed into the conditional critical value function. However, the conditional critical values of $\ln(\psi(Q_1, q_T; \beta^*, \lambda^*) + \psi(Q_1, q_T; \beta_2^*, \lambda_2^*))$ turn out to depend strongly on q_T , whereas the conditional critical values of this term minus $\ln(\psi_2(Q_1, q_T; \beta^*, \lambda^*) + \psi_2(Q_1, q_T; \beta_2^*, \lambda_2^*))$ typically depend less strongly on q_T . Hence, numerical accuracy is improved by computing the LR* statistic (and its critical values) as the difference of the two log terms.

(c) Bessel functions were computed in logarithms using the GAUSS function `mbesselei`.

A.2. *Computation of Conditional Critical Values for Similar Tests*

(a) For the conditional similar tests that involve Bessel functions, conditional critical value functions were computed on a grid of 150 values of Q_T (125 of which were equispaced on a log scale between the 0.5% percentile of a central chi-squared (k) distribution and $Q_T = 1,000$, plus 25 additional points). The 5% critical values were stored in a lookup table that was then accessed (with linear interpolation) to compute rejection rates under the alternative.

(b) The algorithm for numerical evaluation of the p -values for the CLR statistic is described in Andrews, Moreira, and Stock (2006).

A.3. *Power Envelope for LU Invariant Tests*

According to Theorems S.1, the POI LU test is constructed as the conditional test of the LR form in (S3.2), against a given point alternative, where the conditional critical value function is of the form given in Theorem S.2, and κ_1 and κ_2 are chosen to satisfy the conditions of Theorem S.1.

Specifically, consider the problem of construction of the LU test that is POI against a given (fixed) value of the alternative (β^*, λ^*) for a given value of ρ . Denote this LR test by ϕ . The numerical task is to find κ_1 and κ_2 such that

$$(SA.1) \quad E(\phi|Q_T = q_T) = \alpha,$$

$$(SA.2) \quad E(\phi Q_{ST}|Q_T = q_T) = 0.$$

Note that κ_1 and κ_2 are functions of q_T . This POI LU test is implemented by constructing two lookup tables: one for κ_1 and one for κ_2 (both as a function

of q_T). At each value of q_T , κ_1 and κ_2 can be computed by solving the Equations (SA.1) and (SA.2). These two equations were solved using the following algorithm: (i) For a given value of q_T , compute 5,000 Monte Carlo draws of Q under the null hypothesis. (ii) Select a value of κ_2 . Given this value, compute κ_1 as the 0.05 percentile of $LR - Q_{ST}\kappa_2$; that is, κ_1 is chosen to satisfy (SA.1) (where the expectation is replaced by the summation over the 5,000 draws). (iii) Construct $\sum_{MC \text{ Draws}} \phi Q_{ST}$. Repeat steps (ii) and (iii) for different κ_2 with the objective of minimizing $|\sum_{MC \text{ Draws}} \phi Q_{ST}|$. The minimization was done using a line search and the minimized value of $|\text{corr}(\phi, Q_{ST})|$ was always less than 0.001. This was repeated for each value of q_T on the standard grid of q_T (discussed in Section A.2) to construct the lookup tables for κ_1, κ_2 .

To construct the power envelopes, the null rejection frequency based on 5,000 Monte Carlo draws was calculated for each point alternative, (β^*, λ^*) , for each value of ρ and λ/k considered.

REFERENCES

- ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2006): "Performance of Conditional Wald Tests in IV Regressions with Weak Instruments," *Journal of Econometrics*, forthcoming.
- KING, M. L. (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218.
- LEHMANN, E. L. (1986): *Testing Statistical Hypotheses* (Second Ed.). New York: Wiley.