Journal of Econometrics 146 (2008) 241–254

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Efficient two-sided nonsimilar invariant tests in IV regression with weak instruments

Donald W.K. Andrews [a], Marcelo J. Moreira [b,c], James H. Stock [b,d,*]

[a] Cowles Foundation for Research in Economics, Yale University, United States
[b] Department of Economics, Harvard University, United States
[c] Graduate School of Economics, Getulio Vargas Foundation, Brazil
[d] The National Bureau of Economic Research, United States

A R T I C L E   I N F O

Article history:
Available online 28 August 2008

JEL classification:
C12
C30

Keywords:
Instrumental variables regression
Invariant tests
Optimal tests
Power envelope
Two-sided tests

A B S T R A C T

As Nelson and Startz [Nelson, C.R., Startz, R., 1990a. The distribution of the instrumental variable estimator and its t ratio when the instrument is a poor one. Journal of Business 63, S125–S140; Nelson, C.R., Startz, R., 1990b. Some further results on the exact small sample properties of the instrumental variables estimator. Econometrica 58, 967–976] dramatically demonstrated, standard hypothesis tests and confidence intervals in instrumental variables regression are invalid when instruments are weak. Recent work on hypothesis tests for the coefficient on a single included endogenous regressor when instruments may be weak has focused on similar tests. This paper extends that work to nonsimilar tests, of which similar tests are a subset. The power envelope for two-sided invariant (to rotations of the instruments) nonsimilar tests is characterized theoretically, then evaluated numerically for five IVs. The power envelopes for similar and nonsimilar tests differ theoretically, but are found to be very close numerically. The nonsimilar test power envelope is effectively achieved by the Moreira [Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71, 1027–1048] conditional likelihood ratio test, so that test is effectively uniformly most powerful invariant (UMPI). We also provide a new nonsimilar test, P^*, which has χ_1^2 critical values, is asymptotically efficient under strong instruments, involves only elementary functions, and is very nearly UMPI.

1. Introduction

In a pair of highly influential papers, Nelson and Startz (1990a,b) provided stark illustrations of the breakdown of the usual large-sample approximation to the distribution of IV statistics, when the instruments have a small correlation with the included endogenous regressor, a case that has come to be known as weak instruments. Nelson and Startz showed that, in this situation, the distribution of the two-stage least squares (TSLS) estimator can be bimodal and the TSLS t-statistic can have a highly skewed distribution, resulting in large size distortions. Indeed, size distortions when instruments are weak is a problem not just for TSLS, but for all k-class estimators.

Spurred by the Nelson–Startz results, there has been a flurry of research over the past decade on methods for inference in IV regression that are robust to weak instruments, i.e., that are valid even when instruments are weak; see Andrews and Stock (2006) for a survey. One way to think about the testing problem is that the null hypothesis is in fact a compound hypothesis, involving both the parameter of interest (the coefficient on the included endogenous regressor) and a nuisance parameter, which governs the strength of the instruments. Early work focused on tests that are not similar, that is, tests that have rejection rates less than the significance level for some values of the nuisance parameter under the null hypothesis, and important contributions along these lines were made by Nelson and coauthors (Wang and Zivot, 1998; Nelson et al., 2006).

More recent work has focused on similar tests. Two test statistics with null distributions that do not depend on the strength of the instrument are the Anderson and Rubin (1949) (AR) statistic, and the LM statistic of Kleibergen (2002) and Moreira (2001). Moreira (2003) provided a general way to conduct inference in the IV regression model, by conducting inference conditional on a complete sufficient statistic for the nuisance parameter under the null. Andrews et al. (2006) (hereafter AMS) imposed an additional condition that the test be invariant to rotations of the matrix of instruments, and characterized (and computed) the power envelope for two-sided invariant similar tests that are asymptotically efficient if instruments are strong. Notably, AMS found that one of the tests proposed by Moreira (2003), the conditional likelihood ratio (CLR) test, effectively lies on this

* Corresponding address: Department of Economics, Harvard University, Cambridge, MA, United States. Tel.: +1 617 496 0502; fax: +1 617 495 7730.
E-mail address: james_stock@harvard.edu (J.H. Stock).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.08.015

power envelope; while in theory the CLR test is not uniformly most powerful among two-sided invariant similar tests, as a numerical matter it is. But these developments in the area of similar tests, while useful, do not address the possibility that one could do better yet, by considering nonsimilar tests, of which similar tests are a subset.

The purpose of this paper is to assess whether there is a cost to using similar tests and, if so, whether one can construct a nonsimilar test that has often-better power than the best similar test, the CLR test. Nonsimilar tests have null rejection probabilities below the significance level α for some values of the nuisance parameters. Because of the continuity of the power function, for these values of the nuisance parameters, the power of a nonsimilar test will be less than the power of a similar test for alternatives close to the null hypothesis. However, for other values of the nuisance parameters, or for more distant alternatives, nonsimilar tests can have greater power than similar tests.

Specifically, in this paper we apply the theory of optimal nonsimilar testing to the Gaussian IV regression model with a single included endogenous regressor, for the compound null hypothesis described above and a two-sided (two-point) alternative. We characterize the two-sided point-optimal invariant nonsimilar (POINS) tests. In some cases we find a closed-form solution for the POINS test statistic, but, in general, it must be found numerically. Our analysis focuses on Gaussian errors with a known reduced-form error covariance matrix, a pair of assumptions that permit developing exact most powerful tests. These two assumptions are less restrictive than they might initially seem, because finite-sample distributions in the Gaussian IV model with a known error covariance matrix apply as asymptotic limits when these assumptions are relaxed, using Staiger and Stock (1997) weak instrument asymptotics, see AMS for the details.

We use a formulation of the IV regression model in which the native parameters are transformed to polar coordinates following Hillier (1990) and Chamberlain (2007). This transformation is one-to-one and thus has no substantive effect on the testing problem, however the polar coordinate representation has three advantages. First, it results in symmetric power functions. Specifically, AMS show that tests that, in native parameters, invariant similar tests that place equal weight on symmetric positive and negative point alternatives are not asymptotically efficient under strong instruments, an undesirable property, and a two-sided test that is asymptotically efficient does not have symmetric power functions against negative and positive departures from the null. Said differently, in native parameters, the testing problem is not symmetric in a statistical sense because there is more information about departures in one direction than in the other. In contrast we show that, in polar coordinates, point optimal invariant tests (nonsimilar and asymptotically efficient similar) have power functions that are symmetric in the angular departure from the null. Second, numerical analysis turns out to be easier in polar coordinates than in the native parameters (surfaces are smoother). Third, and least important, by mapping the native coordinates onto the circle, the entire power function is more easily plotted because its domain is $[0, \pi/2]$ instead of the real line.

The theoretical and numerical work yields four main substantive conclusions. First, in the case that we examine exhaustively (tests with five instruments and a 5% significance level), the power envelope for two-sided invariant nonsimilar tests effectively equals the power envelope for asymptotically efficient two-sided invariant similar tests. Thus, while something might be gained in theory by considering nonsimilar tests, it turns out that nothing is gained in practice, at least for the case we study numerically.

Second, we propose a test statistic P^{*B}, which is a member of the family of point optimal invariant nonsimilar tests with very good overall power properties. In particular, P^{*B} is asymptotically efficient under strong instruments and has a power function that is numerically very close to the invariant nonsimilar power envelope. Although a uniformly most powerful invariant (UMPI) test does not exist in this problem, in a numerical sense the P^{*B} test is very nearly UMPI among nonsimilar tests.

Third, the P^{*B} test involves Bessel functions, so we also propose a test, P^*, based on an approximation to the Bessel function which uses only elementary functions. It turns out that the power functions of the P^{*B} and P^* tests are extremely close. Thus, in P^* we have found a test that is very nearly UMPI, can be computed using only elementary functions, and has an asymptotic χ^2_1 distribution under strong instruments.

Fourth, the CLR test is also found to be, in a numerical sense, effectively UMPI among nonsimilar two-sided tests. This strengthens the conclusion of AMS, who found the CLR test to be effectively UMPI among similar two-sided tests.[1]

Because the P^* and CLR tests are both nearly UMPI, the choice between the two in practice is one of convenience. Both involve only elementary functions. The CLR test requires conditional critical values. The P^* test uses χ^2_1 critical values, however we have not found an analytic formula for confidence intervals constructed by inverting the P^* test. Because fast and accurate software now exists for the computation of CLR p-values (using the algorithm in Andrews et al. (2007a)) and confidence intervals constructed by inverting the CLR test (Mikusheva, 2006), the practical advice for empirical work coming out of this research is the same as AMS, that is, to use the CLR test when instruments are potentially weak.

The paper proceeds as follows. Section 2 lays out the model, the maximal invariant statistic, and its distribution in native parameters. Section 3 presents the testing problem in polar coordinates. Section 4 develops the theory of POINS tests and the nonsimilar power envelope. Section 5 presents the new P^{*B} and P^* test statistics. Numerical results are presented in Section 6.

2. The model, statistics, and distributions in native parameters

We consider the linear IV regression model with a single included endogenous regressor,

$$y_1 = y_2\beta + X\gamma_1 + u, \tag{1}$$

$$y_2 = Z\Pi + X\xi + v_2, \tag{2}$$

where y_1 and y_2 are $n \times 1$ vectors of endogenous variables, X is a $n \times p$ matrix of exogenous regressors, and Z is a $n \times k$ matrix of k instrumental variables. It is assumed that Z is constructed so that $Z'X = 0$.

Our interest is in two-sided tests of the null hypothesis

$$H_0 : \beta = \beta_0 \quad \text{vs.} \quad H_1 : \beta \neq \beta_0. \tag{3}$$

The reduced form of (1) and (2) is

$$Y = Z\Pi a' + X\eta + V, \tag{4}$$

where $Y = [y_1\ y_2]$, $V = [v_1\ v_2]$, $a = [\beta\ 1]'$, and $\eta = [\gamma\ \xi]$, where $v_1 = u + v_2\beta$ and $\gamma = \gamma_1 + \xi\beta$. The reduced-form errors are assumed to be i.i.d. across observations, homoskedastic, and normally distributed with covariance matrix Ω, that is,

[1] One might wonder about the properties of the CLR test among tests that are not invariant to rotations of the instruments, or as a one-sided test. On the former point, the class of non-invariant tests is too large to be useful because the power envelope is constructed using tests in which the population first-stage coefficients are treated as known. (For example, a feasible test is the Neyman–Pearson test constructed assuming that all the first-stage coefficients equal one. If, in fact, all those coefficients *are* one, then this test is efficient, but if they are not, its power could be quite poor.) As discussed in AMS, invariance to rotations of the instruments is a natural additional condition to impose, one that is satisfied by all standard IV tests. On the latter point, the one-sided and two-sided testing problems turn out to be quite different when the instruments are weak, and conclusions in the two-sided problem do not necessarily carry over to the one-sided case, see Andrews et al. (2007b).

$$V|X, Z \sim N(0, I_n \otimes \Omega), \tag{5}$$

where I_n is the $n \times n$ identity matrix. Throughout, Ω is treated as known.

2.1. Maximal invariant

AMS restrict attention to tests that are invariant to orthonormal transformations of the k IVs, that is, transformations $Z \to ZF'$, where F is a $k \times k$ orthonormal matrix. AMS show that the maximal invariant is

$$Q \equiv \begin{bmatrix} Q_S & Q_{TS} \\ Q_{ST} & Q_T \end{bmatrix} = \begin{bmatrix} S'S & T'S \\ S'T & T'T \end{bmatrix}, \tag{6}$$

where Q is 2×2 and S and T are the $k \times 1$ vectors

$$S = (Z'Z)^{-1/2} Z'Yb_0/(b_0'\Omega b_0)^{1/2}, \tag{7}$$

$$T = (Z'Z)^{-1/2} Z'Y\Omega^{-1}a_0/(a_0'\Omega^{-1}a_0)^{1/2}, \tag{8}$$

where $b_0 = \begin{bmatrix} 1 & -\beta_0 \end{bmatrix}'$ and $a_0 = \begin{bmatrix} \beta_0 & 1 \end{bmatrix}'$. Three test statistics that are functions of Q are the LM statistic of Kleibergen (2002) and Moreira (2001), the Anderson and Rubin (1949) statistic (AR), and the Moreira (2003) conditional likelihood ratio statistic (CLR). Expressed in terms of Q, these test statistics respectively are,

$$LM = Q_{ST}^2/Q_T \tag{9}$$

$$AR = Q_S/k \tag{10}$$

$$CLR = \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4Q_{ST}^2}\right). \tag{11}$$

Under the null hypothesis, the LM statistic is distributed χ_1^2 and the AR statistic is distributed χ_k^2/k. The expressions (9)–(11) and these distributions obtain because Ω is treated as known.

2.2. Distributions in native parameters

Under both the null and alternative, S and T are independent and are normally distributed:

$$\begin{pmatrix} S \\ T \end{pmatrix} \sim N(h \otimes \mu_\Pi, I_2 \otimes I_k), \tag{12}$$

where $h = [c_\beta \;\; d_\beta]'$, $c_\beta = (\beta - \beta_0)/(b_0'\Omega b_0)^{1/2}$, $d_\beta = a'\Omega^{-1}a_0/(a_0'\Omega^{-1}a_0)^{1/2}$, and $\mu_\Pi = (Z'Z)^{1/2}\Pi$. AMS show that the distributions of Q and Q_T depend only on (β, λ), where $\lambda = \Pi Z'Z\Pi = \mu_\Pi'\mu_\Pi$. Let $\nu = (k-2)/2$. The noncentral Wishart distributions of Q and Q_T are,

$$f_Q(q; \beta, \lambda) = K_1 \mathrm{e}^{-\frac{1}{2}\lambda h'h} \det(q)^{(k-3)/2} \mathrm{e}^{-\frac{1}{2}(q_S + q_T)} \left(\sqrt{\lambda h'qh}\right)^{-\nu} \times I_\nu\left(\sqrt{\lambda h'qh}\right) \tag{13}$$

$$f_{Q_T}(q_T; \beta, \lambda) = K_2 \mathrm{e}^{-\frac{1}{2}\lambda d_\beta^2} q_T^{(k-3)/2} \mathrm{e}^{-\frac{1}{2}q_T} \left(\sqrt{\lambda d_\beta^2 q_T}\right)^{-\nu} \times I_\nu\left(\sqrt{\lambda d_\beta^2 q_T}\right) \tag{14}$$

where $K_1 = [2^{(k+2)/2}\pi^{1/2}\Gamma((k-1)/2)]^{-1}$, $K_2 = 1/2$, and $I_\nu(z)$ is the modified Bessel function of the first kind,

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{j=0}^{\infty} \frac{\left(z^2/4\right)^j}{j!\,\Gamma(\nu + j + 1)}. \tag{15}$$

3. The testing problem in polar coordinates

The transformation from native parameters to polar coordinates is obtained by letting r and θ be given by

$$r^2 = \lambda h'h \tag{16}$$

$$\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} = \frac{h}{\sqrt{h'h}}. \tag{17}$$

The mapping from (λ, β) to (r, θ) is one-to-one, so the transformation does not change the testing problem. Under the null hypothesis, $c_{\beta_0} = 0$ so $\beta = \beta_0$ corresponds to $\theta = \theta_0 = 0$. Under the alternative, $\beta > \beta_0$ corresponds to $\theta > 0$, while $\beta < \beta_0$ corresponds to $\theta < 0$.

The transformation (16) and (17) has a natural interpretation: r is the norm of the mean vector of (S, T) and θ is the angular departure of that mean vector from its value under the null hypothesis. The radius r can be thought of as the amount of information in the mean vector that is usable for testing the null hypothesis, and θ governs how large the departure is from the null hypothesis, cf. Hillier (1990), Chamberlain (2007).

Two useful reference values of θ correspond to the limits of the range of β:

$$\theta_\infty = \lim_{\beta\to\infty} \cos^{-1}[d_\beta/(h'h)^{1/2}] \quad \text{and} \quad \theta_{-\infty} = \theta_\infty - \pi. \tag{18}$$

The range of θ corresponding to $-\infty < \beta < \infty$ thus is $-\pi \le \theta_{-\infty} = \theta_\infty - \pi < \theta < \theta_\infty \le \pi$. When $\Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\theta_\infty = \cos^{-1}(-\rho)$.

3.1. Distributions in polar coordinates

In the (r, θ) coordinate system, the distributions of Q and Q_T in (13) and (14) are

$$f_Q(q; r, \theta) = K_1 \mathrm{e}^{-\frac{1}{2}r^2} \det(q)^{(k-3)/2} \mathrm{e}^{-\frac{1}{2}(q_S + q_T)} \left(\sqrt{r^2 x'qx}\right)^{-\nu} \times I_\nu\left(\sqrt{r^2 x'qx}\right) \tag{19}$$

$$f_{Q_T}(q_T; r, \theta) = K_2 \mathrm{e}^{-\frac{1}{2}r^2\cos^2\theta} q_T^{(k-3)/2} \mathrm{e}^{-\frac{1}{2}q_T} \left(\sqrt{r^2\left(\cos^2\theta\right)q_T}\right)^{-\nu} \times I_\nu\left(\sqrt{r^2\left(\cos^2\theta\right)q_T}\right) \tag{20}$$

where $x = x(\theta) = [\sin\theta \;\; \cos\theta]'$.

It is straightforward to verify that the distribution of Q satisfies

$$f_Q(q; r, \theta) = f_Q(q; r, \theta \pm j\pi), \quad j = 1, 2, 3, \ldots \tag{21}$$

$$f_Q(q_S, q_{ST}, q_T; r, \theta) = f_Q(q_S, -q_{ST}, q_T; r, -\theta). \tag{22}$$

An implication of (21) is that θ and $\theta \pm j\pi$ are observationally equivalent.

3.2. Strong instrument local asymptotic nestings

AMS (section 7, Assumption SIV-LA) consider the strong instrument local asymptotic (SIV-LA) sequence,

$$\beta = \beta_0 + \bar{b}/\sqrt{n}, \tag{23}$$

where $\bar{b}$, Π and Ω are fixed. Under (23), if $Z'Z/n \to^p D_Z$, then $\lambda/n = \Pi'(Z'Z/n)\Pi \to^p \Pi' D_Z \Pi \equiv \Lambda_\infty$ (the convergence in probability can be replaced by nonrandom convergence if (X, Z) are nonrandom).

In polar coordinates, the SIV-LA nesting corresponding to (23) is,

$$\theta = t/\sqrt{n}, \tag{24}$$

where $t = \bar{b}/\sqrt{b_0'\Omega b_0 a_0'\Omega^{-1}a_0}$ and where $r^2/n \to^p R_\infty^2 = \Lambda_\infty a_0'\Omega^{-1}a_0$. It is shown in the Appendix that the nestings (23) and (24) are equivalent in a $n^{-1/2}$ neighborhood of the null hypothesis.

3.3. Asymptotically efficient POIS 2-sided tests in polar coordinates

AMS show that POIS tests against a two-point alternative are asymptotically efficient (AE) under the SIV-LA nesting in native parameters (23) if they are against two equally weighted points (λ^*, β^*) and (λ_2^*, β_2^*) which satisfy $\sqrt{\lambda^*}c_{\beta^*} = \sqrt{\lambda_2^*}c_{\beta_2^*}$ and $\sqrt{\lambda^*}d_{\beta^*} = \sqrt{\lambda_2^*}d_{\beta_2^*}$. By (16) and (17), in polar coordinates this corresponds to testing against the two points, (r^*, θ^*) and $(r^*, -\theta^*)$. Thus the POIS test of the null and alternative

$$H_0 : (r_0, \theta = 0) \text{vs.} H_1 : (r_1, \pm\theta_1), \tag{25}$$

where equal weight is placed on (r_1, θ_1) and $(r_1, -\theta_1)$, is AE under the SIV-LA nesting (24) (this is true for all r_0 and $r_1 > 0$).

The results in the previous paragraph provide a precise sense, in which the two-sided testing problem is symmetric in polar coordinates but not native coordinates. POIS tests that place equal weight on $(r_1, \pm\theta_1)$ are asymptotically efficient, whereas POIS tests that place equal weight on $(\lambda, \pm\beta)$ are not. In native coordinates, AE POIS tests have different power against (λ, β) and $(\lambda, -\beta)$, whereas in polar coordinates, they have the same power against (r_1, θ_1) and $(r_1, -\theta_1)$. It is seen in the next section that this symmetry in polar coordinates, but not in native parameters, carries over to POI nonsimilar tests.

The POIS test against the symmetric two-sided alternative (25) rejects for large values of

$$LR_{r_1,|\theta_1|} = \frac{1}{2}e^{-\frac{1}{2}r_1^2\sin^2\theta}\left[\frac{\left(\sqrt{z_1}\right)^{-\nu} I_\nu\left(\sqrt{z_1}\right) + \left(\sqrt{\tilde{z}_1}\right)^{-\nu} I_\nu\left(\sqrt{\tilde{z}_1}\right)}{\left(\sqrt{\bar{z}_0}\right)^{-\nu} I_\nu\left(\sqrt{\bar{z}_0}\right)}\right] \tag{26}$$

where $z_1 = r_1^2 x_1' q x_1$, $\tilde{z}_1 = r_1^2 \tilde{x}_1' q \tilde{x}_1$, $\bar{z}_0 = r_1^2(\cos^2)\theta q_T$, $x_1 = x(\theta_1) = [\sin\theta_1 \ \cos\theta_1]'$ and $\tilde{x}_1 = x(-\theta_1) = [-\sin\theta_1 \ \cos\theta_1]'$, and where the critical value depends on Q_T. The statistic (26) is the statistic LR* in AMS, Corollary 1, written here in polar coordinates. Alternatively (26) can be derived directly as the POIS test that maximizes weighted average power against the two points (r_1, θ_1) and $(r_1, -\theta_1)$, with equal weights. Because the conditional distribution of (Q_S, Q_{ST}) given Q_T does not depend on r under the null, the statistic $LR_{r_1,|\theta_1|}$ does not depend on r_0. The envelope of the power functions of the tests based on (26) is the power envelope of 2-sided asymptotically efficient invariant similar tests, derived originally in AMS in native parameters.

4. Point optimal invariant nonsimilar tests

Now consider the compound null hypothesis and two-sided alternative,

$$H_0 : 0 \le r < \infty, \quad \theta = 0 \quad \text{vs.} \quad H_1 : r = r_1, \theta = \pm\theta_1 \tag{27}$$

where without loss of generality we let θ_1 be positive. Our construction of point optimal invariant nonsimilar (POINS) tests follows Lehmann (1986, Section 3.8). The strategy is to transform the compound null and alternative (28) into simple hypotheses by putting distributions over θ and r under the two hypotheses.

First consider the null hypothesis. Let Λ be a probability distribution over $\{r : 0 \le r < \infty\}$ and let h_Λ be the weighted pdf,

$$h_\Lambda(q) = \int f_Q(q; r, \theta_0)\mathrm{d}\Lambda(r) \tag{28}$$

where $\theta_0 = 0$ and $f_Q(q; r, \theta)$ is given in (19).

Next consider the alternative hypothesis. We follow the treatment of similar tests in AMS and place equal weight on the alternatives (r_1, θ_1) and $(r_1, -\theta_1)$. The distribution of Q under this equal-weighted alternative is

$$g(q) = \frac{1}{2}\left[f_Q(q; r_1, \theta_1) + f_Q(q; r_1, -\theta_1)\right]. \tag{29}$$

The effect of the weighting is to turn the null and alternative into point hypotheses, so the most powerful test is obtained using the Neyman–Pearson Lemma. Specifically, let ϕ_Λ be the most powerful level- α test of h_Λ against g; this test rejects when

$$NP_{\Lambda,r_1,|\theta_1|}(q) = \frac{g(q)}{h_\Lambda(q)} = \frac{1}{2}\frac{f_Q(q; r, \theta) + f_Q(q; r, -\theta)}{h_\Lambda(q)} > \kappa_{\Lambda,r_1,|\theta_1|;\alpha} \tag{30}$$

where $\kappa_{\Lambda,r_1,|\theta_1|;\alpha}$ is the critical value of the test, chosen so that $NP_{\Lambda,r_1,|\theta_1|}(q)$ rejects the null with probability α under the distribution h_Λ.

If the test ϕ_Λ has size α for the null hypothesis H_0 in (27), that is, if

$$\sup_{0\le r<\infty} \Pr_{r,\theta=0}\left[NP_{\Lambda,r_1,|\theta_1|}(q) > \kappa_{\Lambda,r_1,|\theta_1|;\alpha}\right] = \alpha, \tag{31}$$

then Λ is the least favorable distribution: because size is controlled by (31), ϕ_Λ is a valid test of any other null $h_{\Lambda'}(q) = \int f_Q(q; r, \theta_0)\mathrm{d}\Lambda'(r)$, where Λ' is some other distribution over r, but the power of ϕ_Λ testing $h_{\Lambda'}$ against g cannot exceed that of the Neyman–Pearson test $\phi_{\Lambda'}$. Let Λ^{LF} denote the least favorable distribution. Because $\phi_{\Lambda^{LF}}$ is the most powerful test based on the least favorable distribution over r, $\phi_{\Lambda^{LF}}$ is in fact most powerful for testing H_0 against H_1 (see Lehmann (1986), Section 3.8, Theorem 7 and Corollary 5). Thus the POINS test of the null vs. alternative in (28) rejects for large values of $NP_{\Lambda^{LF},r_1,|\theta_1|}(q)$. The envelope of the power functions of the tests based on $NP_{\Lambda^{LF},r_1,|\theta_1|}(q)$ is the power envelope of invariant nonsimilar tests of H_0 against H_1.

The drawback of this approach is the difficulty of finding the least favorable distribution Λ^{LF}. Given a candidate distribution Λ, condition (31) is readily checked numerically to see if Λ is least favorable. What proves more difficult, however, is searching over Λ to find the distribution that satisfies (31). This difficulty is mitigated by working in polar coordinates, because of the symmetry of the testing problem in θ, which in turn facilitates simple approximations to the least favorable distribution.

4.1. One-point distributions

For tractability, we consider distributions Λ that place unit mass on a single point $r_{0,\Lambda}$. In this case, the test statistic in (30) becomes

$$NP_{r_{0,\Lambda},r_1,|\theta_1|}(q) = \frac{1}{2}e^{-\frac{1}{2}\left(r_1^2 - r_{0,\Lambda}^2\right)} \times \left[\frac{\left(\sqrt{z_1}\right)^{-\nu} I_\nu\left(\sqrt{z_1}\right) + \left(\sqrt{\tilde{z}_1}\right)^{-\nu} I_\nu\left(\sqrt{\tilde{z}_1}\right)}{\left(\sqrt{z_0}\right)^{-\nu} I_\nu\left(\sqrt{z_0}\right)}\right] \tag{32}$$

where $z_1 = r_1^2 x_1' q x_1$, $\tilde{z}_1 = r_1^2 \tilde{x}_1' q \tilde{x}_1$, and $z_0 = r_{0,\Lambda}^2 q_T$. Eq. (32) is derived by substituting (19) into (30) and simplifying.

Theorem 1 gives the SIV-LA limiting behavior of $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$.

Theorem 1. *Under the SIV-LA sequence (24), if $|\theta_1| \ne \pi/2$,*

(a) *If $r_{0,\Lambda}^2 > r_1^2\cos^2\theta_1$, then $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q) = O_p(e^{-\sqrt{n}})$.*
(b) *If $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$, then $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q) = e^{-\frac{1}{2}r_1^2\sin^2\theta_1}\cosh\left(r_1\sin\theta_1\frac{Q_{ST}}{\sqrt{Q_T}}\right) + o_p(1)$.*
(c) *If $r_{0,\Lambda}^2 < r_1^2\cos^2\theta_1$, then $n^{-1/2}\ln\left[NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)\right] = \left(r_1|\cos\theta_1| - r_{0,\Lambda}\right)\sqrt{Q_T/n} + o_p(1)$.*

Proofs of theorems are given in the Appendix.

Remarks 1. 1. Because cosh is an even and strictly increasing function, in the case of part (b) that $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$, in the SIV-LA limit the $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$ statistic is a strictly increasing function of $LM = Q_{ST}^2/Q_T$. AMS showed that LM is asymptotically efficient under SIV-LA asymptotics among nonsimilar tests (and, because it is similar, also among similar tests). If $r_{0,\Lambda}^2 > r_1^2\cos^2\theta_1$, $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q) = O_p(1)$ for fixed r and thus the nonsimilar test based on $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$ has a critical value that is $O(1)$, so part (a) implies that, in this case, the test will have power zero under SIV-LA asymptotics and thus will not be asymptotically efficient. Part (c) shows that, if $r_{0,\Lambda}^2 < r_1^2\cos^2\theta_1$, the test is asymptotically a function of only Q_T (under SIV-LA asymptotics, $Q_T/n = O_p(1)$, see the Appendix), so in this case the nonsimilar test based on $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$ will not be SIV-LA asymptotically efficient. Thus a necessary condition for $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$ to be efficient under SIV-LA asymptotics is that $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$.

2. Because the test is nonsimilar, a necessary and sufficient condition for asymptotic efficiency under SIV-LA asymptotics is that $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$ and that the α-level quantile of $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$, when $\theta = 0$, attains its maximum at a point, say r_n^*, such that $r_n^* \to \infty$ as $n \to \infty$; the latter condition is not implied by the former.

The next theorem provides some results about the least favorable distribution.

Theorem 2. *Let Λ^{LF} denote the least favorable distribution.*

(a) *For general r_1, if Λ^{LF} places unit mass on $r_{0,LF}$, then $r_{0,LF}^2 \geq r_1^2\cos^2\theta_1$.*

(b) *For the alternative $(\theta_1 = \pm\pi/2, r_1)$, where r_1 is fixed, Λ^{LF} is a one-point distribution that places unit mass on $r_{0,LF} = 0$, and the AR test is POINS against $(\theta_1 = \pm\pi/2, r_1)$ for all values of r_1.*

Remarks 2. 1. Theorem 2(a) differs from Theorem 1, which pertains to the limit when the true value of r increases proportionally to $\sqrt{n}$, whereas Theorem 2 describes the limit for a fixed true r, as a function of the hypothesized alternative r_1.

2. One might suspect that the least favorable distribution would put point mass on the boundary of the parameter space that corresponds to nonidentification, but this is not so: Theorem 2(a) implies that, if a one-point least favorable distribution exists, the point generally is interior and depends on both the strength of the instruments under the alternative and the magnitude of the departure from the null.

3. Part (b) implies that the condition in part (a) is satisfied with equality when $\theta_1 = \pm\pi/2$.

4. AMS showed (in native parameters) that the AR statistic is admissible and most powerful among invariant similar tests against $\theta_1 = \pm\pi/2$, for all r_1. Part (b) extends this result to nonsimilar tests. The alternative $\theta_1 = \pm\pi/2$ is special in the sense that the POINS test against $\pm\pi/2$ (i.e. the AR test) is not asymptotically efficient under strong instruments.

5. Although Theorem 2 does not show the existence of a one-point Λ^{LF} for general values of r_1, it states that such a distribution does exist for all values of r_1 when $\theta_1 = \pm\pi/2$. Additionally, a calculation provided in the Appendix (and discussed further below) suggests that a one-point least favorable distribution will exist against alternatives for which r_1 is large and θ_1 is small. This suggests that a one-point distribution for Λ^{LF} might exist more generally and moreover provides a range of values – specifically, values of $r_{0,LF}^2$ slightly exceeding $r_1^2\cos^2\theta_1$ – in which to search numerically for $r_{0,LF}$.

4.2. *Using the one-point distribution to bound the nonsimilar power envelope*

Although a least favorable distribution must exist, it might not be a one-point distribution. Even so, it is possible to use tests based on one-point distributions to provide upper and lower bounds on the power envelope of nonsimilar tests. Let Λ^* be a distribution that places point mass on r^*, let $NP_{r^*,r_1,|\theta_1|}(Q)$ be the most powerful test statistic (32) for testing h_{Λ^*} against g, let $\kappa_{r^*,r_1,|\theta_1|;\alpha}(r)$ be the $1-\alpha$ quantile of $NP_{r^*,r_1,|\theta_1|}(Q)$ under $(r, \theta_0 = 0)$, let $\underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha} = \kappa_{r^*,r_1,|\theta_1|;\alpha}(r^*)$, and let $\bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha} = \sup_r \kappa_{r^*,r_1,|\theta_1|;\alpha}(r)$. The test that rejects when $NP_{r^*,r_1,|\theta_1|}(Q) > \underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}$ is the most powerful test of h_{Λ^*} against g, however unless Λ^* is least favorable, it is not a valid test of the compound null $(\theta_0 = 0, 0 \leq r < \infty)$ (if Λ^* is not least favorable the test does not satisfy the size condition (31). Thus $\Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right] \geq \Pr_{r_1,\theta_1}\left[NP_{\Lambda^{LF},r_1,|\theta_1|}(q) > \kappa_{\Lambda^{LF},r_1,|\theta_1|;\alpha}\right]$, where the least favorable distribution Λ^{LF} need not be a one-point distribution (this inequality follows from Lehmann (1986, Section 3.8, Theorem 7 (iii)). On the other hand, the test that rejects when $NP_{r^*,r_1,|\theta_1|}(Q) > \bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha}$ is a valid test of the compound null $(\theta_0 = 0, 0 \leq r < \infty)$ because it uses a sup critical value, however unless Λ^* is least favorable, this test will not be the most powerful test of the compound null against $(r_1, |\theta_1|)$; thus $\Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right] \leq \Pr_{r_1,\theta_1}\left[NP_{\Lambda^{LF},r_1,|\theta_1|}(q) > \kappa_{\Lambda^{LF},r_1,|\theta_1|;\alpha}\right]$. Thus

$$\begin{aligned}&\Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right]\\ &\quad\leq \Pr_{r_1,\theta_1}\left[NP_{\Lambda^{LF},r_1,|\theta_1|}(q) > \kappa_{\Lambda^{LF},r_1,|\theta_1|;\alpha}\right]\\ &\quad\leq \Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right].\end{aligned} \tag{33}$$

If the least favorable distribution places point mass on r^*, then the inequalities in (33) are equalities. Otherwise, (33) provides a lower and upper bound on the power envelope. Whether this bound is useful in practice depends on how close are the two quantiles $\underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}$ and $\bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha}$. If these two quantiles are close, then the one-point distribution Λ^* provides a useful approximation to the least favorable distribution in the sense that it provides a tight bound (33) on the power envelope.

Because similar tests are a subset of nonsimilar tests, the power envelope of invariant nonsimilar tests is also bounded below by the power envelope of AE similar tests.

As mentioned in Remark 2.5 following Theorem 2, additional calculations (given in the Appendix) suggest that the limit of the sequence of least-favorable distributions against the sequence of alternatives $(\theta_1 = t_1/r_1, r_1)$, where t_1 is a constant, is a one-point distribution with point mass on $r_{0,LF}^2 = r_1^2\cos^2\theta_1$. If, in fact, $r_{0,LF}^2 = r_1^2\cos^2\theta_1$, then the POINS test statistic $NP_{r_{0,LF},r_1,|\theta_1|}(q)$ in (32) equals the AE POIS test statistic $LR_{r_1,|\theta_1|}$ in (26). In addition, in the Appendix it is shown that, in the limit $\theta_1 = t_1/r_1$, $r_1 \to \infty$ (holding fixed n and the true r and θ), $LR_{r_1,|\theta_1|}$ is equivalent to LM, the distribution of which does not depend on r. Taken together, these results suggest that in the limit $\theta_1 = t_1/r_1$, $r_1 \to \infty$, the POINS test is similar and the POINS and AE POIS power envelopes coincide.

5. An approximate POINS tests

One general approach to testing when there is not a uniformly most powerful test, is to select from the family of point-optimal tests a test that has a power function tangent to the power envelope at some intermediate value of the alternative. This test is optimal against that specific alternative and, it is hoped, might