Efficient Two-Sided Nonsimilar Invariant Tests in IV Regression with Weak Instruments

June 2006 (this revision: May 2007)

Donald W.K. Andrews
Cowles Foundation for Research in Economics, Yale University

Marcelo J. Moreira
Department of Economics, Harvard University

James H. Stock*
Department of Economics, Harvard University and the National Bureau of Economic Research

ABSTRACT

As Nelson and Startz (1990a, 1990b) 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 Moreira's (2003) 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^2_1 critical values, is asymptotically efficient under strong instruments, involves only elementary functions, and is very nearly UMPI.

Key Words: Instrumental variables regression, invariant tests, optimal tests, power envelope, two-sided tests

JEL Code: C12, C30

CORRESPONDING AUTHOR: James H. Stock, Department of Economics, Harvard University, Cambridge MA. Phone: 617-496-0502. Fax: 617-495-7730. Email: james_stock@harvard.edu

*We thank Gary Chamberlain for suggesting the problem considered in this paper and Anna Mikusheva for research assistance. The authors gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-0417911(Andrews), SES-0418268 (Moreira), and SBR-0214131 and SBR-0617811 (Stock), respectively.

1. Introduction

In a pair of highly influential papers, Nelson and Startz (1990a, 1990b) 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 strictly 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 some of Nelson's coauthors (Wang and Zivot (1998)).

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-Rubin (1949) (AR) statistic and the LM statistic of Kleibergen (2001) 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, Moreira, and Stock (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 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 seem because results based on the Gaussian model with a known error covariance matrix apply as asymptotic limits when these assumptions are relaxed using Staiger-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 Hiller (1990) and Chamberlain (2005). 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, we found that numerical analysis is considerably easier in polar coordinates than in the native parameters (smoother surfaces). 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 $[-\infty, \infty]$.

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, we have found a test, the P^{*} test, which is very nearly UMPI, which can be computed using only elementary functions, and which has an asymptotic χ_1^2 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_1^2 critical values, however we do not provide 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, Moreira, and Stock [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,

---

[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 conventional 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, Moreira, and Stock (2007b).

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

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

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

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

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,

$$E(VV'|X, Z) \sim \text{i.i.d. N}(0, I_n \otimes \Omega). \tag{5}$$

Throughout, Ω is treated as known.

Maximal invariant. AMS restrict attention to tests that are invariant to orthonormal transformations of the k IVs, that is, transformations $Z \rightarrow ZF'$, 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 $b_0 = [1\ \ -\beta_0]'$ and $a_0 = [\beta_0\ \ 1]'$. Three test statistics that are functions of Q are the LM statistic of Kleibergen (2002) and Moreira (2001), the Anderson-Rubin (1949) statistic (AR), and Moreira's (2003) conditional likelihood ratio statistic (CLR). Expressed in terms of Q, these test statistics respectively are,

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

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

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

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.

Distributions in native parameters. Under both the null and alternative, S and T are independent and are normally distributed:

$$S \sim \mathrm{N}(c_\beta\mu_\Pi, I_k) \text{ and } T \sim \mathrm{N}(d_\beta\mu_\Pi, I_k), \tag{12}$$

where $c_\beta = (\beta - \beta_0)/(b_0'\Omega b_0)^{1/2}$, $d_\beta = a'\Omega^{-1}a_o/(a_0'\Omega^{-1}a_0)^{1/2}$, 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,

$$f_Q(q; \beta, \lambda) = K_1 e^{-\frac{1}{2}\lambda\left(c_\beta^2 + d_\beta^2\right)} \det(q)^{(k-3)/2} e^{-\frac{1}{2}(q_S + q_T)} \left(\sqrt{\lambda \xi_\beta(q)}\right)^{-\nu} I_\nu\left(\sqrt{\lambda \xi_\beta(q)}\right) \quad (13)$$

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

where $\xi_\beta(q) = [c_\beta \;\; d_\beta]q[c_\beta \;\; d_\beta]'$, $K_1 = [2^{(k+2)/2}\pi^{1/2}\Gamma((k-1)/2)]^{-1}$, $K_2 = ½$, and $I_\nu(z)$ is the modified Bessel function of the first kind,

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

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

$$r^2 = \lambda h'h \quad (16)$$

$$\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} c_\beta / \sqrt{h'h} \\ d_\beta / \sqrt{h'h} \end{pmatrix}. \quad (17)$$

where $h = [c_\beta \;\; d_\beta]'$. 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 the mean vector implied by 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 (2005).

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

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

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

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

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

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

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

It is straightforward to verify that the distribution of Q satisfies

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

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

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

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 $Z'Z/n \xrightarrow{p} D_Z$, then $\lambda/n = \Pi'(Z'Z/n)\Pi \xrightarrow{p}$ $\Pi' D_Z \Pi \equiv \Lambda_\infty$ (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 $r^2/n \xrightarrow{p} R_\infty^2 = \Lambda_\infty a_0'\Omega^{-1}a_0$. It is shown in the appendix that the nestings (23) and (24) are equivalent in a $n^{-1/2}$ neighborhood of the null hypothesis.

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*, –θ*). Thus the POIS test of the null and alternative

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

where equal weight is placed on (r_1, θ_1) and (r_1, –θ_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, ±θ_1) are asymptotically efficient, whereas POIS tests that place equal weight on (λ, ±β) are not. In native coordinates, AE POIS tests have different power against (λ, β) and (λ, -β), whereas in polar coordinates, they have the same power against (r_1, θ_1) and (r_1, -θ_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

$$LR_{r_1,|\theta_1|} = \frac{1}{2}e^{-\frac{1}{2}r_1^2\sin^2\theta}\left[\frac{\left(\sqrt{r_1^2 x_1' q x_1}\right)^{-\nu} I_\nu\left(\sqrt{r_1^2 x_1' q x_1}\right) + \left(\sqrt{r_1^2 \tilde{x}_1' q \tilde{x}_1}\right)^{-\nu} I_\nu\left(\sqrt{r_1^2 \tilde{x}_1' q \tilde{x}_1}\right)}{\left(\sqrt{r_1^2\cos^2\theta q_T}\right)^{-\nu} I_\nu\left(\sqrt{r_1^2\cos^2\theta q_T}\right)}\right] \quad (26)$$

where $\tilde{x}_1 = x(-\theta) = [-\sin\theta_1 \ \ \cos\theta_1]'$, 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,

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

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 (27) 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 $\{r: 0 \leq r < \infty\}$ and let h_Λ be the weighted pdf,

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

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

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

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 $h_{\Lambda'}(q) = \int f_Q(q;r,\theta_0)d\Lambda'(r)$, 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 the 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 (27) 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 in (27).

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.

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

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

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 $z_0 = r_{0,\Lambda}^2 q_T$. Equation (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 $|\theta_1| \neq \pi/2$,

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. 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 following 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.

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 $\theta = \pm\pi/2$.
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 $r_{0,\mathrm{LF}}$.

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 ($\theta_0 = 0$, $0 \le r < \infty$) (if Λ^* is not least favorable the test does not satisfy the size condition (31)). Thus $\Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \underline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right] \ge \Pr_{r_1,\theta_1}\left[NP_{\Lambda^{LF},r_1,|\theta_1|}(q) > \kappa_{\Lambda^{LF},r_1,|\theta_1|;\alpha}\right]$, 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 ($\theta_0 = 0$, $0 \le r < \infty$) 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 $\Pr_{r_1,\theta_1}\left[NP_{r^*,r_1,|\theta_1|}(q) > \bar{\kappa}_{r^*,r_1,|\theta_1|;\alpha}\right] \le \Pr_{r_1,\theta_1}\left[NP_{\Lambda^{LF},r_1,|\theta_1|}(q) > \kappa_{\Lambda^{LF},r_1,|\theta_1|;\alpha}\right]$. Thus

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

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 $\overline{\kappa}_{r^*,r_1,|\theta_1|;\alpha}$. 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 similar tests is also bounded below by the power envelope of AE similar tests.

As mentioned in remark 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 have a power function that is not too far from the power envelope for other alternatives. See King (1988) for a general discussion of this approach and some examples. In this section, we pursue this approach, with three modifications.

The first modification is that the POINS tests in Section 4 need not be asymptotically efficient under strong instruments. Asymptotic efficiency under strong instruments is a desirable property and we impose this as an additional side condition. As stated in remark 2 following Theorem 1, a necessary and sufficient condition for asymptotic efficiency of a POINS test is that (i) $r_{0,LF}^2 = r_1^2 \cos^2 \theta_1$ and (ii) the α-level quantile of the test statistic, as a function of r, must achieve its maximum in the limit $r \to \infty$. Condition (i) is imposed analytically and (ii) is imposed numerically.

The second modification is that we rewrite the POINS test statistic and its approximation so that their SIV-LA distribution is χ_1^2 under the null and a non-central chi-squared under the local alternative.

The third modification is that the POINS test involves the Bessel function, which is not implemented in all statistical software. We therefore provide an approximation to the POINS test statistic that uses only elementary functions.

Transformed POINS test statistic. We now impose $r_{0,LF}^2 = r_1^2 \cos^2 \theta_1$ and rewrite the POINS test statistic so that it has a χ_1^2 SIV-LA null distribution. The function cosh(x) is strictly increasing in $|x|$ and has the inverse function, $\cosh^{-1}(y) = \ln[y + (y^2 - 1)^{1/2}]$. Accordingly, upon imposing $r_{0,LF}^2 = r_1^2 \cos^2 \theta_1$ the expression in Theorem 1(b) can be manipulated to yield,

$$\frac{\cosh^{-1}\left[e^{\frac{1}{2} r_1^2 \sin^2 \theta_1} NP_{r_{0,\Lambda}, r_1, |\theta_1|}(q) \right]}{r_1 \sin \theta_1} = \left| \frac{Q_{ST}}{\sqrt{Q_T}} \right| + o_p(1). \tag{34}$$

Squaring the expression on the left hand side of (34) and reexpressing the result using (32) yields the test statistic,

$$P_{r_1,\theta_1}^{*B} = \frac{\left[\cosh^{-1}(D_{r_1,\theta_1}^{*B}) \right]^2}{r_1^2 \sin^2 \theta_1} \tag{35}$$

where

$$D^{*B}_{r_1,\theta_1} = \frac{1}{2}\frac{\left(\sqrt{z_1}\right)^{-\nu} I_\nu\left(\sqrt{z_1}\right)+\left(\sqrt{\tilde{z}_1}\right)^{-\nu} I_\nu\left(\sqrt{\tilde{z}_1}\right)}{\left(\sqrt{z_0}\right)^{-\nu} I_\nu\left(\sqrt{z_0}\right)} \tag{36}$$

where z_0, z_1, and $\tilde{z}_1$ are defined following (32) and where the superscript B in $P^{*B}_{r_1,\theta_1}$ indicates that the test involves Bessel functions.

The tests based on $P^{*B}_{r_1,\theta_1}$ and $NP_{r_1^2 \sin^2\theta_1, r_1, |\theta_1|}(q)$ are equivalent. However, the formulation $P^{*B}_{r_1,\theta_1}$ in (35) is convenient because $P^{*B}_{r_1,\theta_1}$ has a χ^2_1 null distribution under strong instruments: by (34), under strong instruments

$$P^{*B}_{r_1,\theta_1} = Q^2_{ST}/Q_T + o_p(1) \tag{37}$$

and $Q^2_{ST}/Q_T = LM$ has a central χ^2_1 null distribution and a noncentral χ^2_1 distribution under a local alternative.

An approximation not involving Bessel functions. Olver (1974) provides the approximation to the Bessel function, $I_\nu(\nu x) \sim \left(1+x^2\right)^{-1/4} e^{\nu\xi}/\sqrt{2\pi\nu}$, where $\xi = \sqrt{1+x^2} + \ln\left[x/\left(1+\sqrt{1+x^2}\right)\right]$. Upon applying this approximation to (36) and simplifying, we obtain the following approximation to the transformed POINS test statistic $P^{*B}_{r_1,\theta_1}$ that involves only elementary functions:

$$P^{*}_{r_1,\theta_1} = \frac{\left[\cosh^{-1}\left(D^{*}_{r_1,\theta_1}\right)\right]^2}{r_1^2 \sin^2\theta_1} \tag{38}$$

where

$$D^{*}_{r_1,\theta_1} = \frac{1}{2}\left[\frac{\left(\nu^2+z_0\right)^{1/4} z_0^{\nu/2}}{\left(\nu^2+z_1\right)^{1/4} z_1^{\nu/2}} e^{\phi_1-\phi_0} + \frac{\left(\nu^2+z_0\right)^{1/4} z_0^{\nu/2}}{\left(\nu^2+\tilde{z}_1\right)^{1/4} \tilde{z}_1^{\nu/2}} e^{\tilde{\phi}_1-\phi_0}\right], \tag{39}$$

$$\phi_0 = \sqrt{\nu^2 + z_0} + \nu \ln\left(\frac{z_0}{\nu + \sqrt{\nu^2 + z_0}}\right), \quad (40)$$

and ϕ_1 and $\tilde{\phi}_1$ are defined similarly to ϕ_0 except with z_1 and $\tilde{z}_1$ respectively replacing z_0.[2]

Under SIV-LA asymptotics, the $P^{*B}_{r_1,\theta_1}$ and $P^{*}_{r_1,\theta_1}$ tests are asymptotically equivalent:

$$P^{*B}_{r_1,\theta_1} = P^{*}_{r_1,\theta_1} + o_p(1) = Q_{ST}^2 / Q_T + o_p(1). \quad (41)$$

As discussed in remark 2 following Theorem 1, (41) is one of two conditions for the $P^{*B}_{r_1,\theta_1}$ and $P^{*}_{r_1,\theta_1}$ tests to be efficient under strong instruments. The second condition, the quantile restriction discussed in the remark, places an additional restriction on r_1 and θ_1, and this second condition will be imposed numerically.

6. Numerical Results

Because finding the least favorable distribution is computationally intensive, the invariant nonsimilar power envelope was computed only for k = 5 instruments and tests of level α = 0.05. Less demanding computations, including the power functions of the P^* test, of its approximations, and of similar invariant tests were also performed for k = 2 and k = 10. Because the power functions are symmetric in θ and repeat every π, it suffices to compute power envelopes and functions for $0 \leq \theta \leq \pi/2$. All computations were done on a grid of 34 values of θ and on a grid of $r^2/\sqrt{k}$ = 0.5, 1, 2, 4, 8, 16, 32, 64, a range that spans very weak to strong instruments. The r^2 grid was scaled by $\sqrt{k}$

[2] In an earlier draft of this paper, we used test statistics based on the simpler approximation given in Appendix equation (A.2), which holds for large arguments of the Bessel function, however those approximations are not uniformly as good as Olver's (1974) and the resulting approximating tests have power substantially less than the $P^{*}_{r_1,\theta_1}$ and $P^{*B}_{r_1,\theta_1}$ tests in some parts of the parameter space. To save space those results are not reported here but are available from the authors upon request.

because Andrews and Stock (2006) show that holding $\lambda/\sqrt{k}$ constant as k increases is the sequence that results in a nondegenerate limiting power envelope for tests involving many weak instruments; and in polar coordinates this sequence corresponds to holding $r^2/\sqrt{k}$ constant as k increases.

Numerical considerations. For a given alternative (r_1, $|\theta_1|$), the POINS test was computed by searching numerically for the least favorable distribution. Only one-point distributions were considered. Let $F(r_{0,\Lambda})$ = $\sup_{0\le r<\infty} \Pr_{r,\theta=0}\left[NP_{r_{0,\Lambda},r_1,|\theta_1|}(Q) > \kappa_{r_{0,\Lambda},r_1,|\theta_1|;\alpha}(r_{0,\Lambda})\right] - \alpha$, where $\kappa_{r_{0,\Lambda},r_1,|\theta_1|;\alpha}(r_{0,\Lambda})$ is the $1-\alpha$ quantile of $NP_{r_{0,\Lambda},r_1,|\theta_1|}(Q)$ when the true $r = r_{0,\Lambda}$ and $\theta = 0$. By the theory outlined in Section 4, the function $F(r_{0,\Lambda})$ is nonnegative and if $F(r_{0,\Lambda}) = 0$, then the least favorable distribution places unit mass on $r_{0,\Lambda}$. Thus the one-point least favorable distribution (if it exists) can be found by minimizing $F(r_{0,\Lambda})$. For a given (r_1, $|\theta_1|$) and $r_{0,\Lambda}$, $F(r_{0,\Lambda})$ was computed by (i) computing $\kappa_{r_{0,\Lambda},r_1,|\theta_1|;\alpha}(r_{0,\Lambda})$ as the $1-\alpha$ quantile of the Monte Carlo distribution of $NP_{r_{0,\Lambda},r_1,|\theta_1|}(Q)$; (ii) for a given r, estimating the rejection probability $\Pr_{r,\theta=0}\left[NP_{r_{0,\Lambda},r_1,|\theta_1|}(Q) > \kappa_{r_{0,\Lambda},r_1,|\theta_1|;\alpha}(r_{0,\Lambda})\right]$ by Monte Carlo simulation; and (iii) computing the maximum of this rejection probability on a grid over r, with $0 \le r^2 \le 1000$. Preliminary investigations found that results were numerically stable using a grid that is equal-spaced on a log scale. Computations used either 10,000 Monte Carlo draws in steps (i) and (ii) and 60 grid points in step (iii), or 20,000 Monte Carlo draws and 100 grid points. The reason for different precisions is that some of the least favorable distribution points were easier to compute than others; some seem to require (at least) four-digit accuracy to minimize F. Attempts to find the least favorable distribution were only made when the invariant similar power envelope was less than 0.995.

Upon minimization of F, the critical values for the upper and lower bounds of the nonsimilar test power envelope ($\bar{\kappa}_{r^*_{0,\Lambda},r_1,|\theta_1|;\alpha}$ and $\underline{\kappa}_{r^*_{0,\Lambda},r_1,|\theta_1|;\alpha}$, where $r^*_{0,\Lambda}$ is the minimizer of $F(r_{0,\Lambda})$) were computed by Monte Carlo simulation (10,000 draws, grid size 150). The upper and lower bounds in (33) were then computed by Monte Carlo simulation (10,000) draws.

Critical values for the $P^{*}_{r_1,\theta_1}$ and $P^{*B}_{r_1,\theta_1}$ tests were computed using a grid of 150 points in *r* and 50,000 Monte Carlo draws, and *p*-values for the conditional likelihood ratio test were computed by numerical integration using the algorithm in Andrews, Moreira, and Stock (2007a). Power envelopes for similar invariant tests were computed as described in AMS (except for the different grid in polar coordinates). All power functions were computed by Monte Carlo simulation using either 10,000 or 50,000 draws. In all cases, the covariance matrix Ω is taken as known and the underlying errors are all normally distributed so that these computations are numerical implementations of the exact distribution theory presented in the previous sections.

Results for least favorable distributions. A one-point least favorable distribution was deemed to have been found if the minimized objective function *F* was within one Monte Carlo standard error of zero; this corresponds to a maximum violation of the size condition of 0.0022 when 10,000 Monte Carlo draws are used. Under this criterion, one-point least favorable distributions were found in all cases (all 33 values of $\theta_1 > 0$) for $r_1^2/\sqrt{k}$ = 0.5, 1, and 4, and in all but one case for $r_1^2/\sqrt{k}$ = 2. For $r_1^2/\sqrt{k}$ = 8, a one-point least favorable distribution was not found in 6 cases; for $r_1^2/\sqrt{k}$ = 16, not in 13 cases; for $r_1^2/\sqrt{k}$ = 32, not in 9 cases; and for $r_1^2/\sqrt{k}$ = 64, not in 5 cases. The cases in which the least favorable distribution was not found are for the largest deviations θ from the null. Even in the cases that a least favorable distribution was not found, the largest size discrepancy (largest value of $\min_r F(r)$) was approximately 0.02 (size of 7% for a 5% test).

Figure 1 plots the values of the least favorable point $r^2_{0,LF}$, when it was found, against $r_1^2 \cos^2\theta_1$, on a log-log scale, for $\theta < \pi/2$. For small values of $r_1^2 \cos^2\theta_1$, $r^2_{0,LF}$ departs from the 45° line, and in some cases lies as much as 0.2 log units above the 45° line. When $r_1^2 \cos^2\theta_1$ is large, however, $r^2_{0,LF}$ very nearly equals $r_1^2 \cos^2\theta_1$. These numerical results accord with the calculation in the Appendix, which suggests that $r^2_{0,LF}$ =

$r_1^2 \cos^2 \theta_1$ in the limit $\theta_1 = t_1/r_1$ and $r_1 \to \infty$. In fact, in the extreme case $r_1^2 / \sqrt{k}$ = 64 and $\theta_1 = 0.01\pi$, $r_{0,LF}^2$ exceeds $r_1^2 \cos^2 \theta_1$ by only 0.01%.

Comparison of similar and nonsimilar power envelopes. The lower and upper bounds on the nonsimilar power envelope are plotted for k = 5 in Figure 2 for weak instruments and in Figure 3 for stronger instruments. The power envelope for similar tests is also plotted. Figures 2 and 3 have two remarkable features.

First, when they differ, the lower and upper bounds on the similar power envelopes are still very close in practical terms, rarely diverging by more than 0.01. In this numerical sense, the bounds on the nonsimilar power envelope are tight.

Second, the nonsimilar and similar power envelopes are very close, and in most cases cannot be distinguished visually. In fact, against some alternatives, the similar power envelope exceeds the upper bound of the nonsimilar power envelope. This is theoretically impossible (similar tests are a subset of nonsimilar tests) and reflect limitations in computational accuracy. In no case does the similar power envelope deviate from the nonsimilar power envelope (above or below) by more than one Monte Carlo standard error.

These two findings provide our primary numerical conclusion of the paper: for k = 5 instruments and tests of level α = 0.05, the power envelopes of invariant similar and nonsimilar tests are, as a numerical matter, the same. It is surprising that there is such close agreement between the similar and nonsimilar power envelopes. The reason for this agreement is that the quantile functions of the POINS statistics, as a function of r under the null, are in most cases very nearly (but not exactly) flat, so that the POINS tests are very nearly (but not exactly) similar tests.

AMS compared numerically (in native parameters) the performance of the LM, AR, and CLR tests to each other and to the AEPE for similar tests. Their main finding was that the power function for the CLR test is, for all practical purposes, essentially on the similar power envelope. Because the AEPE for similar tests was found in Figures 2 and 3 to equal the power envelope for nonsimilar tests, the power function for the CLR test also essentially lies on the power envelope for invariant nonsimilar tests. Although there does not exist a UMPI test (similar or nonsimilar) in this problem, as a numerical matter, the CLR test is, in effect, UMPI among nonsimilar tests, and thus also among similar tests, at least when $k = 5$.

The power envelopes are everywhere below one for values of $r^2/\sqrt{k} \leq 8$. This has an intuitive interpretation. The parameter r^2 measure the amount of information (strength of the instruments) and if there is not much information then a test cannot be decisive. This might appear to conflict with the power envelopes in AMS, which all achieve unit power against distant alternatives; but the AMS power envelopes are in native parameters and as β diverges from β_0 holding λ constant (the AMS experiment), r increases (see (16)), so in this sense the strength of the instruments increases with the divergence from the null in the AMS plots. Of course, the plots in this paper and in AMS are simply different renderings or slices of the same power surface so there is no conflict, however when portrayed in polar coordinates the power implications of weak instruments for the testing problem are more obvious.[3]

The computation of the least favorable distributions and the nonsimilar power envelope is more involved numerically than computation of the AEPE for similar tests, so we suspect that the similar power envelope is computed more accurately. Because these two power envelopes are numerically indistinguishable in all the cases we considered (the cases of Figures 2 and 3), the remaining figures report the AEPE for similar tests.

Comparison of LM, AR, and CLR tests to the nonsimilar power envelope. The LM, AR, and CLR tests have been studied extensively by AMS and others in native coordinates, however translation of those results into polar coordinates is not simple so it is of interest to make a direct comparison of these similar tests on the same grid in polar coordinates. The results are presented (also for $k = 5$) in Figure 4. For comparison purposes, Figure 4 also presents the AEPE for similar tests. The power functions in Figure 4 show quite clearly the optimality of the AR test against $\theta = \pi/2$, and the deterioration of the LM test as θ approaches $\pi/2$. The CLR test is essentially on the

[3] For a given k, the power functions in AMS are presented as a function of λ, β, and ρ, where $\Omega = [1\ \rho\ :\ \rho\ 1]$. The domain of the power functions is, however, only two dimensional. In native coordinates power is solely function of $(\sqrt{\lambda}c_\beta, \sqrt{\lambda}d_\beta)$, and in polar coordinates it is solely a function of (r, θ). In this sense, the plots here represent the power functions more concisely than those given as a function of (λ, β, ρ) in AMS.

power envelope in almost all cases; the greatest discrepancy occurs for moderately weak instruments ($r_1^2/\sqrt{k}$ = 4) as θ approaches $\pi/2$.

The P* and P*B tests. Numerical investigations of the $P^{*B}_{r_1,\theta_1}$ test entailed finding values of (r_1, θ_1) such that the test is asymptotically efficient under SIV-LA asymptotics and such that it has good power (is close to the POINS power envelope) for a wide range of true (r, θ). As discussed in the second paragraph of Section 5, imposing the condition that $P^{*B}_{r_1,\theta_1}$ be asymptotically efficient entails restricting attention to (r_1, θ_1) such that the 1 – α quantile of $P^{*B}_{r_1,\theta_1}$ for (r, $\theta = 0$) be maximized in the limit $r \to \infty$, in which case the quantile is the 1 – α quantile of a χ_1^2. This condition for asymptotic efficiency was imposed numerically. Numerical experimentation revealed that setting $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$ results in $P^{*B}_{r_1=(20k)^{1/4},\theta_1=\pi/4}$ being asymptotically efficient for tests at the 5% or 1% level (but not at levels exceeding 9%) and having good overall power properties, at least for $2 \leq k \leq 10$. For convenience, we refer to statistics $P^{*B}_{r_1=(20k)^{1/4},\theta_1=\pi/4}$ and its approximation using (38), $P^{*}_{r_1=(20k)^{1/4},\theta_1=\pi/4}$, simply as P^{*B} and P^*, respectively.

The power functions of P^{*B} and P^* are reported in Figure 5 for k = 5 and in Figures 6 and 7 for k = 2 and 10, respectively. For comparison purposes, all figures also display the power function of the CLR test. In all cases, the P^{*B} and P^* power functions are nearly indistinguishable (curiously, for k = 2 the power function of the approximate test occasionally lies above that of the Bessel function test): evidently, nothing is lost (and perhaps something is gained) by using the approximation P^*. We therefore focus on the simpler of the two nonsimilar tests, the P^* test.

The maximum gaps between the CLR and P^* power functions are .006, .031, and .046 for k = 2, k = 5, and k = 10, respectively. The CLR test is noticeably more powerful than the P^* test only for k = 5 and 10 and only then when instruments are moderately weak ($r_1^2 = 4/\sqrt{k}$). Because the CLR test effectively achieves the AE similar and nonsimilar power envelopes, for k = 2 the P^* power function effectively lies on the nonsimilar power envelope so the P^* test is effectively UMPI. For k = 5 and k = 10, the departures of the P^* test from efficiency are few and small, although it cannot quite be said that the P^* test is effectively UMPI.

Because the power functions of the P^{*B} and CLR tests are so close, one might wonder whether they have the same critical regions. Interestingly, they do not. For example, when $k = 5$, $r^2/\sqrt{k} = 4$ and $\theta = 0.04\pi$, the power of the two tests are numerically indistinguishable (power of 0.06), but the conditional probability that P^* rejects, given that CLR rejects, is only 92%. As the power increases, these probabilities increase, for example when $k = 5$, $r^2/\sqrt{k} = 8$ and $\theta = 0.15\pi$, the power of the two tests are again numerically indistinguishable (power of 0.40) and this conditional probability is 96%. Only when the instruments are strong and the alternative is distant, however, does this conditional probability reach one. Thus, the choice faced by a practitioner about which test to use is nontrivial, in the sense that in some realizations the two tests will give different conclusions even though they have the same power.

The P^* test is based on elementary functions and has χ_1^2 critical values (at the 1% and 5% level), it is effectively UMPI for $k = 2$, and it is nearly so for $k = 5$ and $k = 10$. This said, P^* test does not significantly improve upon the power of the CLR test. Moreover, fast code for the computation of *p*-values and confidence sets using the CLR is now available (Mikusheva 2006), while we have not found a closed form expression for the inversion of the P^* test for the computation of confidence intervals. These practical considerations, and the slightly higher power in some regions of the parameter space, lead to our recommending the CLR test for empirical work.

Several open issues remain for future research. First, we found that the nonsimilar and AE similar power envelopes were the same for tests with level 5% for k = 5. Additional numerical work checking this finding for other values of k and other significance levels is in order. Second, we did not try to optimize the performance of the $P^*_{r_1,\theta_1}$ test over r_1 and θ_1; doing so might find values of r_1 and θ_1 such that the resulting $P^*_{r_1,\theta_1}$ test is effectively UMPI for values of k exceeding 2. Third, the $P^*_{r_1,\theta_1}$ test can be inverted to compute confidence intervals, however at this point the only way to do the inversion is numerical and it would be of interest to have a closed-form expression for confidence intervals based on $P^*_{r_1,\theta_1}$.

References

Anderson, T.W. and H. Rubin, 1949, Estimators of the Parameters of a Single Equation in a Compete Set of Stochastic Equations, Annals of Mathematical Statistics 21, 570-582.

Andrews, D.W.K., M. Moreira, and J.H. Stock, 2006, Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression, Econometrica 74, 715-752.

Andrews, D.W.K., M. Moreira, and J.H. Stock, 2007a, Performance of Conditional Wald Tests in IV Regression with Weak Instruments, forthcoming, Journal of Econometrics

Andrews, D.W.K., M. Moreira, and J.H. Stock, 2007b, Optimal One-Sided Tests for Instrumental Variables Regression, manuscript.

Andrews, D.W.K. and J.H. Stock, 2006, Inference with Weak Instruments, in R. Blundell, W. Newey, and T. Persson (eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Cambridge: Cambridge University Press (forthcoming).

Chamberlain, G., 2005, Decision Theory Applied to an Instrumental Variables Model, manuscript, Harvard University.

Hillier, G., 1990, On the Normalization of Structural Equations: Properties of Direct Estimators, Econometrica 58, 1181-1194.

King, M.L., 1988, Towards a Theory of Point Optimal Testing, Econometric Reviews, 6, 169-218.

Kleibergen, F., 2002, Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, Econometrica 70, 1781-1803.

Lebedev, N.N., 1965, Special Functions and their Applications, Englewood Cliffs, NJ: Prentice-Hall.

Lehmann, E.L., 1986, Testing Statistical Hypotheses (second edition), New York: Wiley.

Mikusheva, A., 2006, Robust Confidence Sets in the Presence of Weak Instruments, manuscript, Harvard University.

Moreira, M.J., 2001, Tests with Correct Size When Instruments Can Be Arbitrarily Weak, Working Paper Series 37, Center for Labor Economics, Department of Economics, University of California, Berkeley.

Moreira, M.J., 2003, A Conditional Likelihood Ratio Test for Structural Models, Econometrica 71, 1027-1048.

Nelson, C.R. and 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. and Startz, R., 1990b, Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator, Econometrica 58, 967-976.

Nelson, C.R., Startz, R., and Zivot, E., 2006, Improved Inference in Weakly Identified Instrumental Variables Regression, in Corbae, D., S. N. Durlauf, and B. E. Hansen (eds.), Econometric Theory and Practice (Essays in Honor of Peter C.B. Phillips), Cambridge: Cambridge University Press.

Olver, F.W.J., 1974, Asymptotics and Special Functions, New York: Academic Press.

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

Wang, J. and E. Zivot, 1998, Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments, Econometrica 66, 1389-1404.

Appendix

Local equivalence of SIV-LA nestings in native and polar coordinates

The SIV-LA nesting in native parameters is given by (23). Under (23), $c_\beta =$ $n^{-1/2}\bar{b}/\sqrt{b_0'\Omega b_0}$ and $d_\beta = \sqrt{a_0'\Omega^{-1}a_0} + n^{-1/2}\bar{b}e_1'\Omega^{-1}a_0/\sqrt{a_0'\Omega^{-1}a_0}$, where $e_1 = [1\ \ 0]'$. Thus $h'h = a_0'\Omega^{-1}a_0 + O(1/\sqrt{n})$ and $c_\beta/\sqrt{h'h} = n^{-1/2}\bar{b}/\sqrt{b_0'\Omega b_0 a_0'\Omega^{-1}a_0} + O(1/n)$. By (17), $\sin\theta = c_\beta/\sqrt{h'h}$, so under (23) $\theta = \sin^{-1}\left(n^{-1/2}\bar{b}/\sqrt{b_0'\Omega b_0 a_0'\Omega^{-1}a_0} + O(n^{-1})\right) = t/\sqrt{n} +$ $O(n^{-1})$, where $t = \bar{b}/\sqrt{b_0'\Omega b_0 a_0'\Omega^{-1}a_0}$. An implication of Π and Ω being fixed and $(Z'Z/n) \xrightarrow{p} D_Z$ is that, as stated following (23), $\lambda/n = \Pi'(Z'Z/n)\Pi \xrightarrow{p} \Lambda_\infty$, so $r^2/n = \lambda h'h/n$ $= (\lambda/n)[a_0'\Omega^{-1}a_0 + O(1/\sqrt{n})] \xrightarrow{p} \Lambda_\infty a_0'\Omega^{-1}a_0 = R_\infty^2$. From (13) and (19), the distribution of Q depends only on $(c_\beta,\ d_\beta,\ \lambda)$ or, equivalently, on (θ, r). Thus the asymptotic distribution of Q is the same under the sequence $(\beta = \beta_0 + \bar{b}/\sqrt{n},\ \lambda/n \xrightarrow{p} \Lambda_\infty)$ as it is under the sequence $(\theta = t/\sqrt{n},\ r^2/n \xrightarrow{p} R_\infty^2)$, where $t = \bar{b}/\sqrt{b_0'\Omega b_0 a_0'\Omega^{-1}a_0}$ and $R_\infty^2 =$ $\Lambda_\infty a_0'\Omega^{-1}a_0$; this latter sequence is the sequence in (24).

Proof of Theorem 1

Because Q has the same limiting distribution under the SIV-LA sequences (23) and (24), results in AMS derived under the native local parameterization also apply under the polar coordinate local parameterization, with a change of parameters. In particular, from the proof of Theorem 8 in AMS, we have that, under (24), (i) $Q_T/n \xrightarrow{p} R_\infty^2$; (ii) $Q_{ST}/\sqrt{Q_T} = O_p(1)$; (iii) $Q_S/\sqrt{Q_T} \xrightarrow{p} 0$; (iv) $z_1/z_0 = r_1^2\cos^2\theta_1/r_0^2 + O_p(n^{-1/2})$ and $\tilde{z}_1/z_0 =$ $r_1^2\cos^2\theta_1/r_0^2 + O_p(n^{-1/2})$; and (v) (by AMS, Equation (A.15)),

$$\sqrt{z_1} - \sqrt{z_0} = r_1 \sin\theta_1 \operatorname{sgn}(\cos\theta_1)\frac{Q_{ST}}{\sqrt{Q_T}} + \left(\sqrt{r_1^2\cos^2\theta}_1 - \sqrt{r_0^2}\right)\sqrt{Q_T} + o_p(1)$$

$$\sqrt{\tilde{z}_1} - \sqrt{z_0} = -r_1 \sin\theta_1 \operatorname{sgn}(\cos\theta_1)\frac{Q_{ST}}{\sqrt{Q_T}} + \left(\sqrt{r_1^2\cos^2\theta}_1 - \sqrt{r_0^2}\right)\sqrt{Q_T} + o_p(1).$$

where z_0, z_1, and $\tilde{z}_1$ are defined following (32). The statement that the remainder terms in (iv) and (v) are $O_p(n^{-1/2})$ is slightly stronger than the $o_p(1)$ result given in AMS, however the stronger $O_p(n^{-1/2})$ result follows by a straightforward extension of the proof of Theorem 8 in AMS. We therefore have that, under the SIV-LA sequence,

$$NP_{r_0,r_1,|\theta_1|}(Q) = \frac{1}{2}e^{-\frac{1}{2}\left(r_1^2-r_0^2\right)}\left[\frac{\left(\sqrt{z_1}\right)^{-\nu} I_\nu\left(\sqrt{z_1}\right) + \left(\sqrt{\tilde{z}_1}\right)^{-\nu} I_\nu\left(\sqrt{\tilde{z}_1}\right)}{\left(\sqrt{z_0}\right)^{-\nu} I_\nu\left(\sqrt{z_0}\right)}\right]$$

$$= \frac{1}{2}e^{-\frac{1}{2}\left(r_1^2-r_0^2\right)}\left[\left(\frac{\sqrt{z_1}}{\sqrt{z_0}}\right)^{-\left(\nu+\frac{1}{2}\right)} e^{\sqrt{z_1}-\sqrt{z_0}} + \left(\frac{\sqrt{\tilde{z}_1}}{\sqrt{z_0}}\right)^{-\left(\nu+\frac{1}{2}\right)} e^{\sqrt{\tilde{z}_1}-\sqrt{z_0}}\right]\left(1+O_p(n^{-1/2})\right)$$

$$= \frac{1}{2}e^{-\frac{1}{2}\left(r_1^2-r_0^2\right)}\left[e^{r_1\sin\theta_1 \operatorname{sgn}(\cos\theta_1)\frac{Q_{ST}}{\sqrt{Q_T}} + \left(r_1|\cos\theta_1| - r_0\right)\sqrt{Q_T}} + e^{-r_1\sin\theta_1 \operatorname{sgn}(\cos\theta_1)\frac{Q_{ST}}{\sqrt{Q_T}} + \left(r_1|\cos\theta_1| - r_0\right)\sqrt{Q_T}}\right]\left(1+O_p(n^{-1/2})\right)$$

$$= e^{-\frac{1}{2}\left(r_1^2-r_0^2\right)}\left(\frac{r_1^2\cos^2\theta_1}{r_0^2}\right)^{-\frac{1}{2}\left(\nu+\frac{1}{2}\right)} \cosh\left(r_1\sin\theta_1\frac{Q_{ST}}{\sqrt{Q_T}}\right)e^{\left(r_1|\cos\theta_1|-r_0\right)\sqrt{Q_T}}\left(1+O_p(n^{-1/2})\right) \qquad \text{(A.1)}$$

where the first equality restates (32); the second equality uses the exponential approximation to the Bessel function as $x \to \infty$,

$$I_\nu(x) = (2\pi x)^{-1/2}e^x[1 + O(x^{-1})] \qquad \text{(A.2)}$$

(Lebedev (1965, Equation (5.11.10)) and the limiting results (i) and (iv), which imply that $z_0 = O_p(n)$ and z_1 and $\tilde{z}_1$ are $O_p(n^{1/2})$; the third equality in (A.1) uses the limiting approximations provided above; and the final equality uses the definition of the hyperbolic cosine and the limiting result (iv) above.

By (ii), $\cosh\left(r_1 \sin\theta_1 Q_{ST} / \sqrt{Q_T}\right)$ is $O_p(1)$ under (24). The results stated in the theorem obtain by evaluating the limiting behavior of (A.1) in the three cases, $r_1^2 \cos^2\theta_1$ <, =, or > r_0^2 where (by (i)) under SIV-LA asymptotics $Q_T = O_p(n)$.

Proof of Theorem 2

(a) First observe that the SIV-LA limits in Theorem 1 are also limits that hold for n fixed under the sequence of true parameters $r \to \infty$, $\theta = t/r$. This follows by noting that the distribution of Q in (19) depends only on r and θ; so formally the limit in Theorem 1 can be obtained by replacing n in the proof of Theorem 1 by $n = r^2/R_\infty^2$, treating Z as nonrandom, and letting $r \to \infty$. With this substitution, the result in Theorem 1(c) is that, if $r_0^2 < r_1^2 \cos^2\theta_1$, then $r^{-1}\ln\left[NP_{r_0,r_1,|\theta_1|}(q)\right] = \left(r_1\left|\cos\theta_1\right| - r_0\right)\sqrt{Q_T / r} + o_p(1)$, where $o_p(1)$ is interpreted in terms of the sequence $r \to \infty$ and where $Q_T/r^2 = O_p(1)$.

Next note that, for all fixed n, r_0, r_1, and true r, $NP_{r_0,r_1,|\theta_1|}(q)$ is $O_p(1)$ so its quantiles are bounded. Thus a necessary condition for a point $\tilde{r}_0$ to be the sole mass point of Λ^{LF} is that the supremum over r of the quantiles of $NP_{\tilde{r}_0,r_1,|\theta_1|}(q)$ are bounded (else the sup-size condition (31) will not hold). In particular, it must be the case that the quantiles of $NP_{\tilde{r}_0,r_1,|\theta_1|}(q)$ are bounded under the sequence $r \to \infty$, $\theta = t/r$. But given the reinterpretation of the preceding paragraph, if $\tilde{r}_0^2 < r_1^2 \cos^2\theta_1$ then the quantiles of $NP_{\tilde{r}_0,r_1,|\theta_1|}(q)$ are unbounded under the sequence $r \to \infty$, $\theta = t/r$, so Λ^{LF} cannot place unit mass on $\tilde{r}_0$ if $\tilde{r}_0^2 < r_1^2 \cos^2\theta_1$.

(b) In the special case $\theta_1 = \pm\pi/2$, $x(\pi/2) = [1\ \ 0]'$ so the test statistic (32) becomes,

$$NP_{r_0=0,r_1,\theta_1=\pi/2}(Q) = e^{-\frac{1}{2}r_1^2}\left(\sqrt{r_1^2 Q_S}\right)^{-\nu} I_\nu\left(\sqrt{r_1^2 Q_S}\right) \tag{A.3}$$

where (A.3) uses $z^{-\nu}I_\nu(z)|_{z=0} = 1$. Because $z^{-\nu}I_\nu(z)$ is strictly increasing in z, $NP_{r_0=0,r_1,\theta_1=\pi/2}(Q)$ is strictly increasing in the AR statistic Q_S, so the AR test is the POI test of ($r = 0$, $\theta = 0$) vs. (r_1, $\theta_1 = \pi/2$). But the null distribution of Q_S does not depend on r, so (31) is satisfied. Thus $r_{0,LF} = 0$ and the AR test is the POINS test of ($0 \le r < \infty$, $\theta = 0$) against (r_1, $\theta_1 = \pm\pi/2$). The AR statistic does not depend on r_1, so this is true for all r_1.

Additional Calculation

Limiting POINS test along the sequence $r_1 \to \infty$ and $\theta_1 = t_1/r_1$

The following calculation suggests that, under the sequence $r_1 \to \infty$ and $\theta_1 = t_1/r_1$, the test $NP_{r_{0,\Lambda},|\theta_1|}(q)$ with $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$ has a critical value that does not depend on the true r, and thus that the condition (31) is satisfied so that the distribution that places unit mass on $r_0^2 = r_1^2\cos^2\theta_1$ is in fact least favorable. This result is an approximation that holds with arbitrary accuracy so in this sense the bounds (33) based on the one-point least favorable distribution with $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$ will be arbitrarily tight.

Fix $Q = q$. Set $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$ so that

$$\begin{aligned}
\sqrt{z_1} - \sqrt{z_0} &= \sqrt{r_1^2 x_1' q x_1} - \sqrt{r_1^2\cos^2\theta_1 q_T} \\
&= r_1\left(\sqrt{\left(\sin^2\theta_1\right)q_S + 2\left(\sin\theta_1\cos\theta_1\right)q_{ST} + \left(\cos^2\theta_1\right)q_T} - \sqrt{\left(\cos^2\theta_1\right)q_T}\right) \\
&= r_1\sqrt{\left(\cos^2\theta_1\right)q_T}\left[\left(1 + \frac{2\left(\sin\theta_1\cos\theta_1\right)q_{ST}}{\left(\cos^2\theta_1\right)q_T} + \frac{\left(\sin^2\theta_1\right)q_S}{\left(\cos^2\theta_1\right)q_T}\right)^{1/2} - 1\right] \\
&= t_1\frac{q_{ST}}{\sqrt{q_T}} + O(1/r_1)
\end{aligned} \tag{A.4}$$

where the first equality follows by $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$, the second equality follows by the definition of x_1, the third line rearranges the second, and the final equality follows from setting $\theta_1 = t/r_1$ and taking a Taylor series expansion in $1/r_1$. Similarly, $\sqrt{\tilde{z}_1} - \sqrt{z_0} = -t_1 q_{ST}/\sqrt{q_T} + O(1/r_1)$. Also, $z_1/z_0 = 1 + O(1/r_1)$ and $\tilde{z}_1/z_0 = 1 + O(1/r_1)$.

Let q_{min} denote the minimum eigenvalue of q. By (A.2),

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

Substituting $r_{0,\Lambda}^2 = r_1^2\cos^2\theta_1$ and the approximations in (A.4) and subsequently into this expression for $NP_{r_{0,\Lambda},r_1,|\theta_1|}(q)$ and collecting terms yields

$$NP_{r_{0,\Lambda},r_1,|\theta_1|}(q) = e^{-\frac{1}{2}r_1^2\sin^2\theta_1}\cosh\left(t_1\frac{q_{ST}}{\sqrt{q_T}}\right)\left(1+O\left(\frac{1}{r_1 q_{min}}\right)\right). \qquad \text{(A.5)}$$

Thus, for a given value of $q_{min} > 0$, in the limit of the sequence $r_1 \to \infty$ and $\theta_1 = t_1/r_1$, $NP_{r_0,r_1,|\theta_1|}(q)$ is an increasing function of q_{ST}^2/q_T.

This approximation (A.5) is uniform over mineval(q) $\geq q_{min} > 0$. Thus on the subset of the support Q for which mineval(Q) $\geq q_{min}$, $NP_{r_0,r_1,|\theta_1|}(Q)$ depends in the limit on $LM = Q_{ST}^2/Q_T$, however the null distribution of LM does not depend on r so (31) is satisfied conditional on mineval(Q) $\geq q_{min}$. Because q_{min} can be chosen to be arbitrarily small the probability that mineval(Q) $\geq q_{min}$ can be made arbitrarily close to one (uniformly in r, θ) so the bounds in (33) can be satisfied with arbitrary precision.

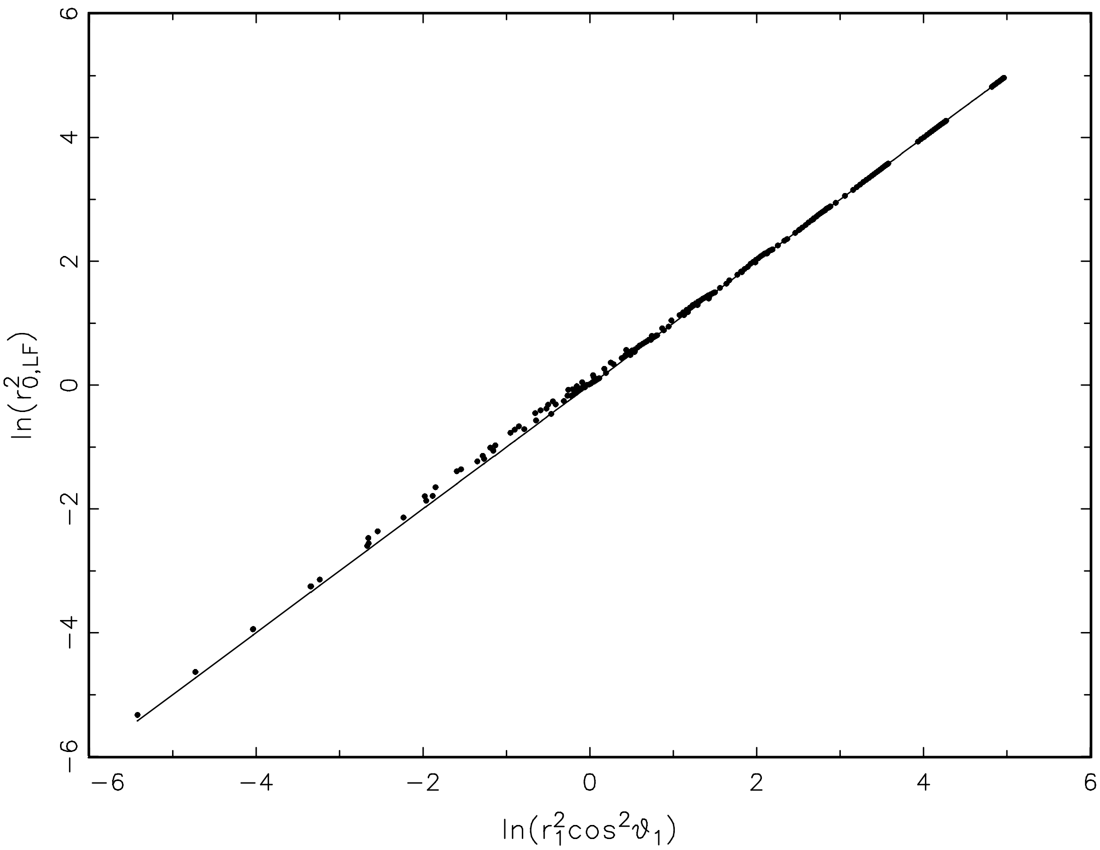

Figure 1. Points $r_{0,LF}^2$ of one-point least favorable distribution vs. limiting theoretical value $r_1^2 \cos^2 \theta_1$, log-log scale, and 45^o line (solid)

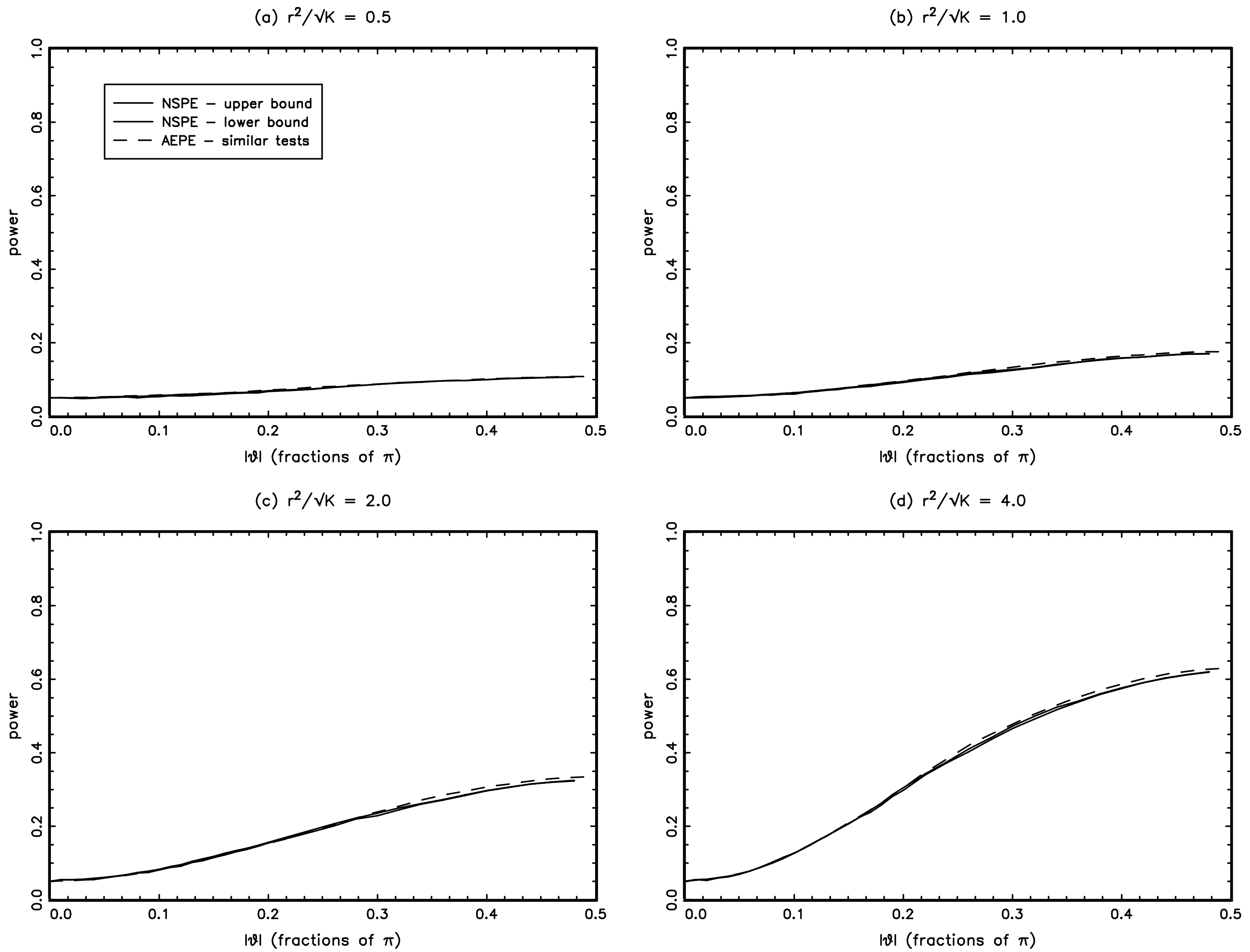

Figure 2 Upper and lower bound on power envelope for nonsimilar invariant tests against $(r, |\theta|)$ and power envelope for similar invariant tests against $(r, |\theta|)$, $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 0.5, 1, 2, 4$, $k = 5$

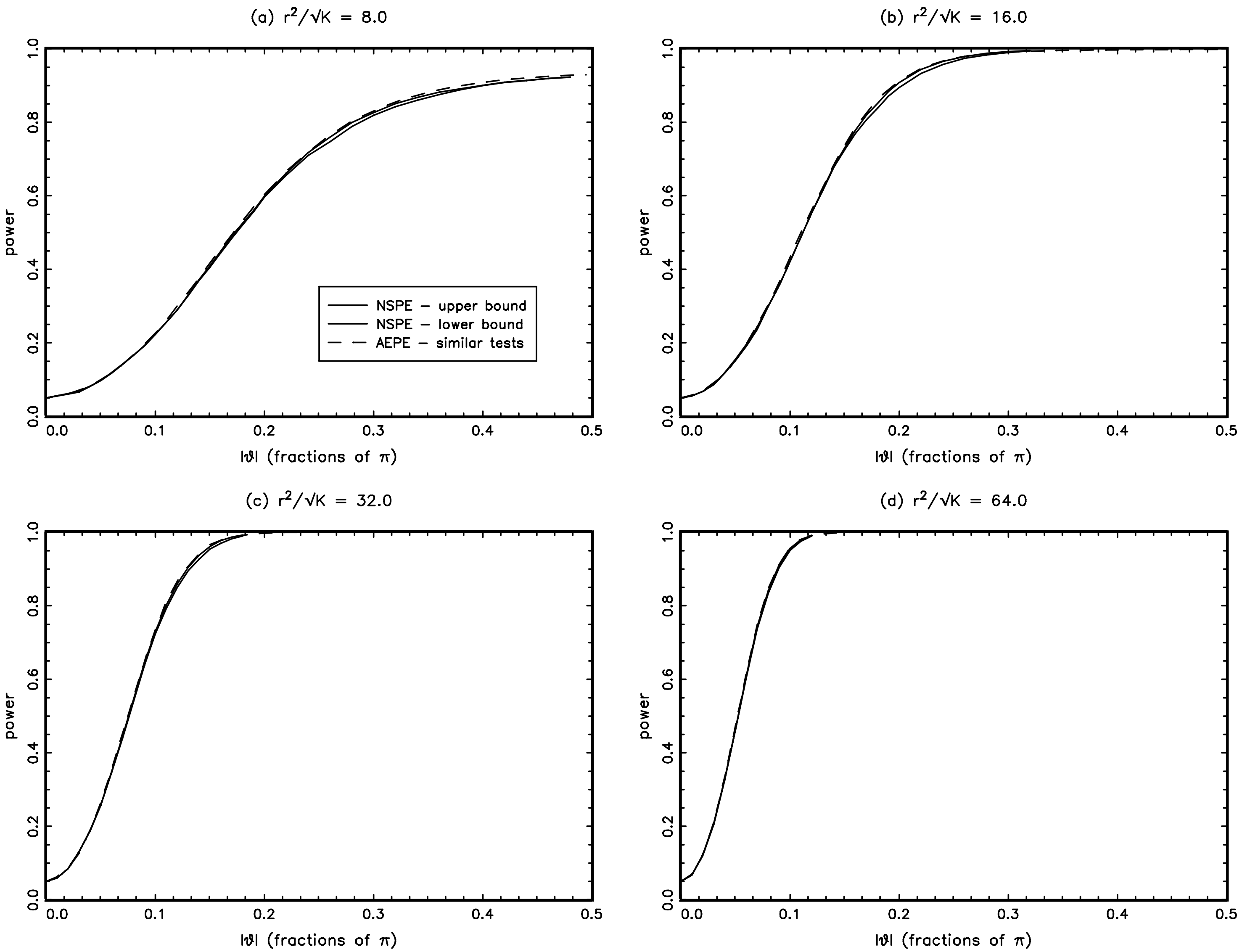

Figure 3 Upper and lower bound on power envelope for nonsimilar invariant tests against $(r, |\theta|)$ and power envelope for similar invariant tests against $(r, |\theta|)$, $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k}$ = 8, 16, 32, 64, $k = 5$,

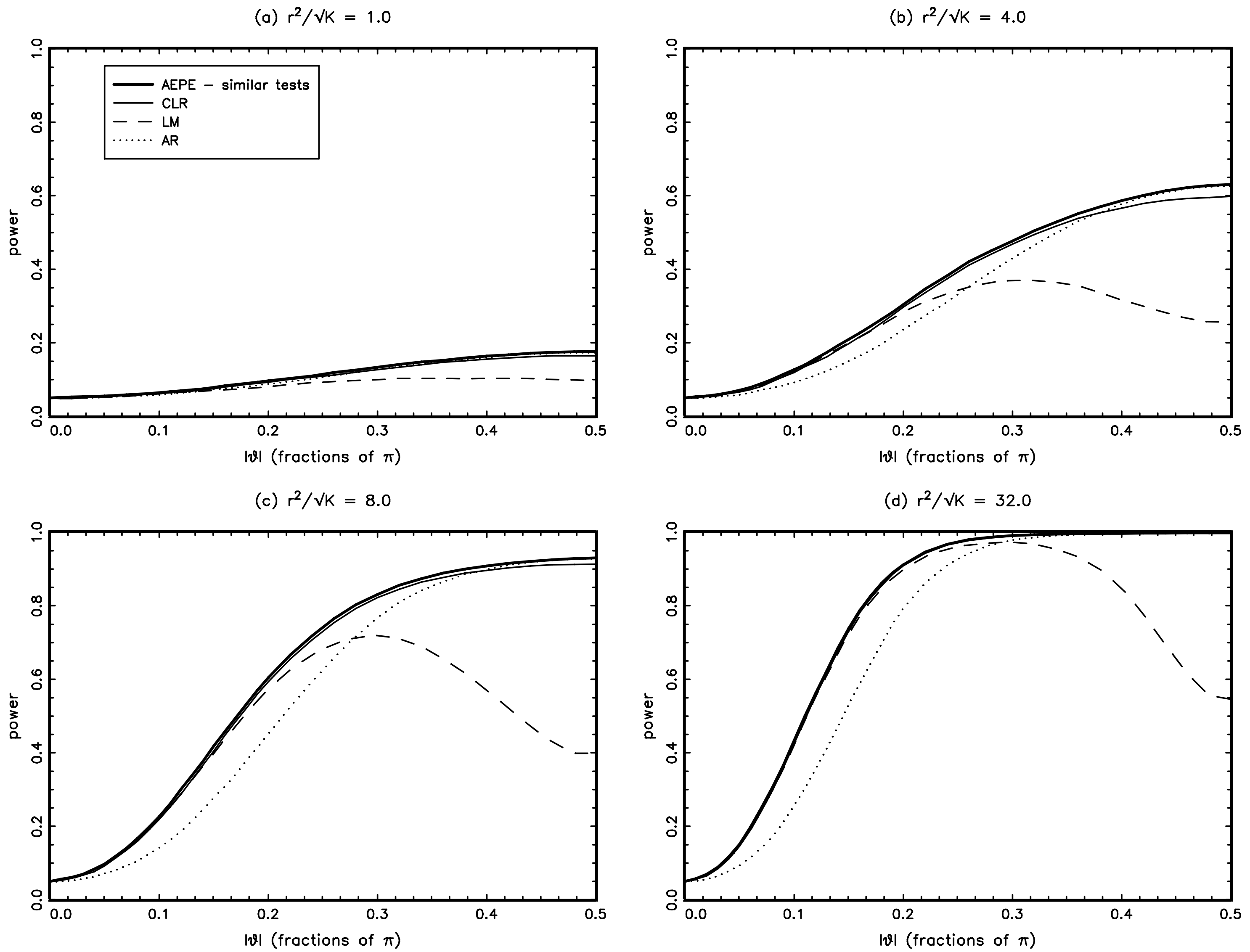

Figure 4 Power envelope for similar invariant tests against $(r, |\theta|)$ and power functions of the CLR, LM, and AR tests, $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, $k = 5$,

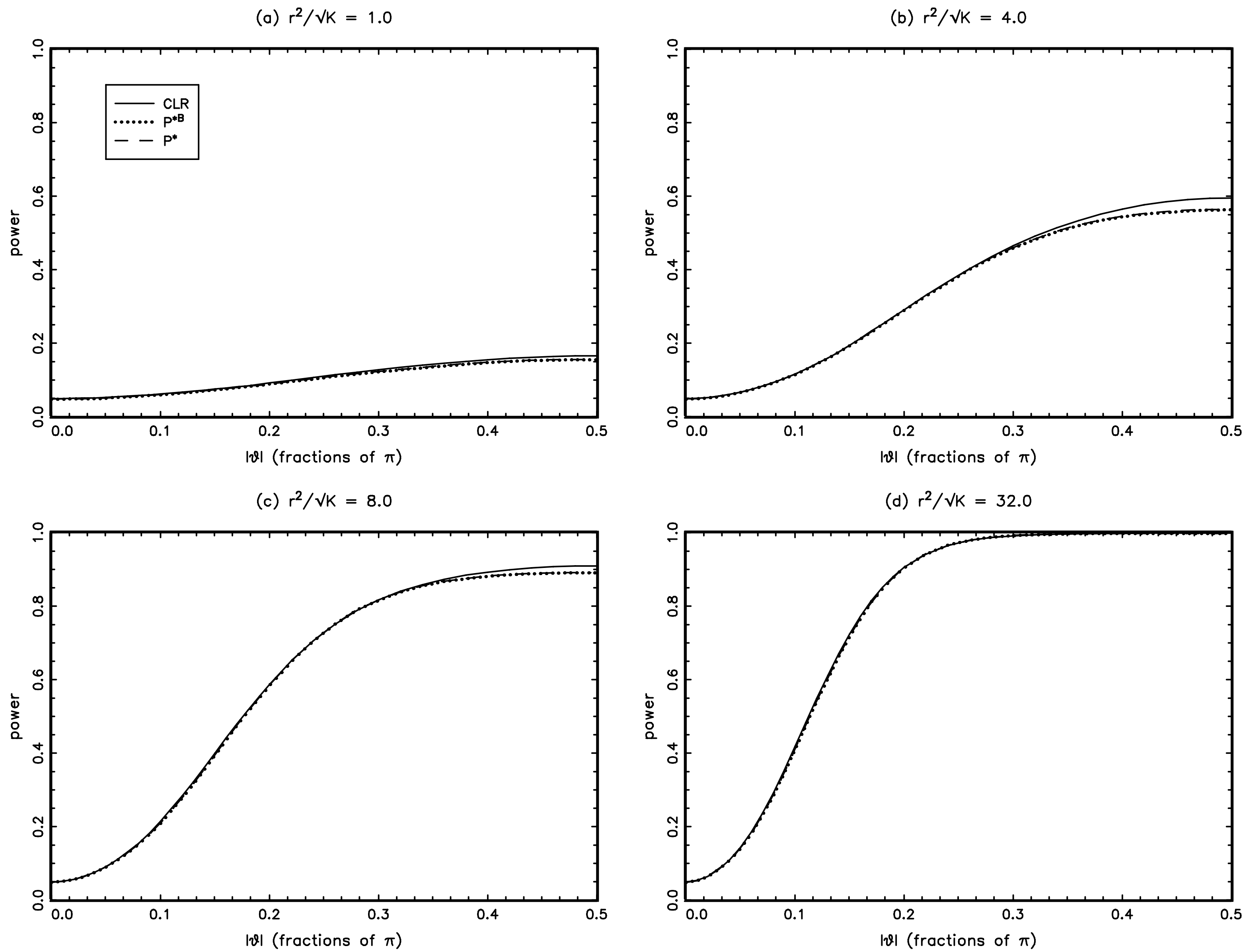

Figure 5 Power functions of the CLR, P^{*B}, and P^{*} tests (in which $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$), for $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, and $k = 5$

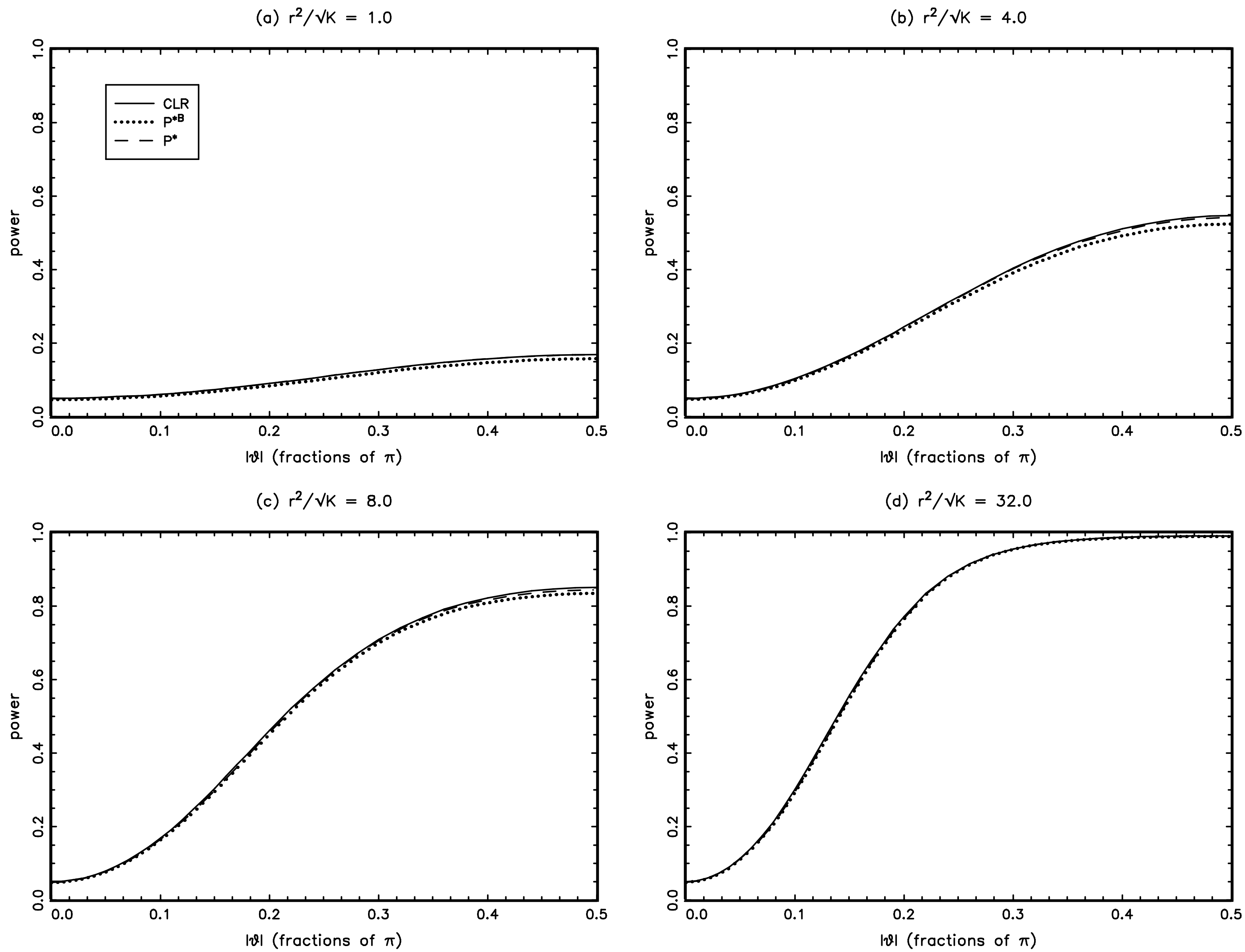

Figure 6 Power functions of the CLR, P^{*B}, and P^{*} tests (in which $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$), for $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, and $k = 2$

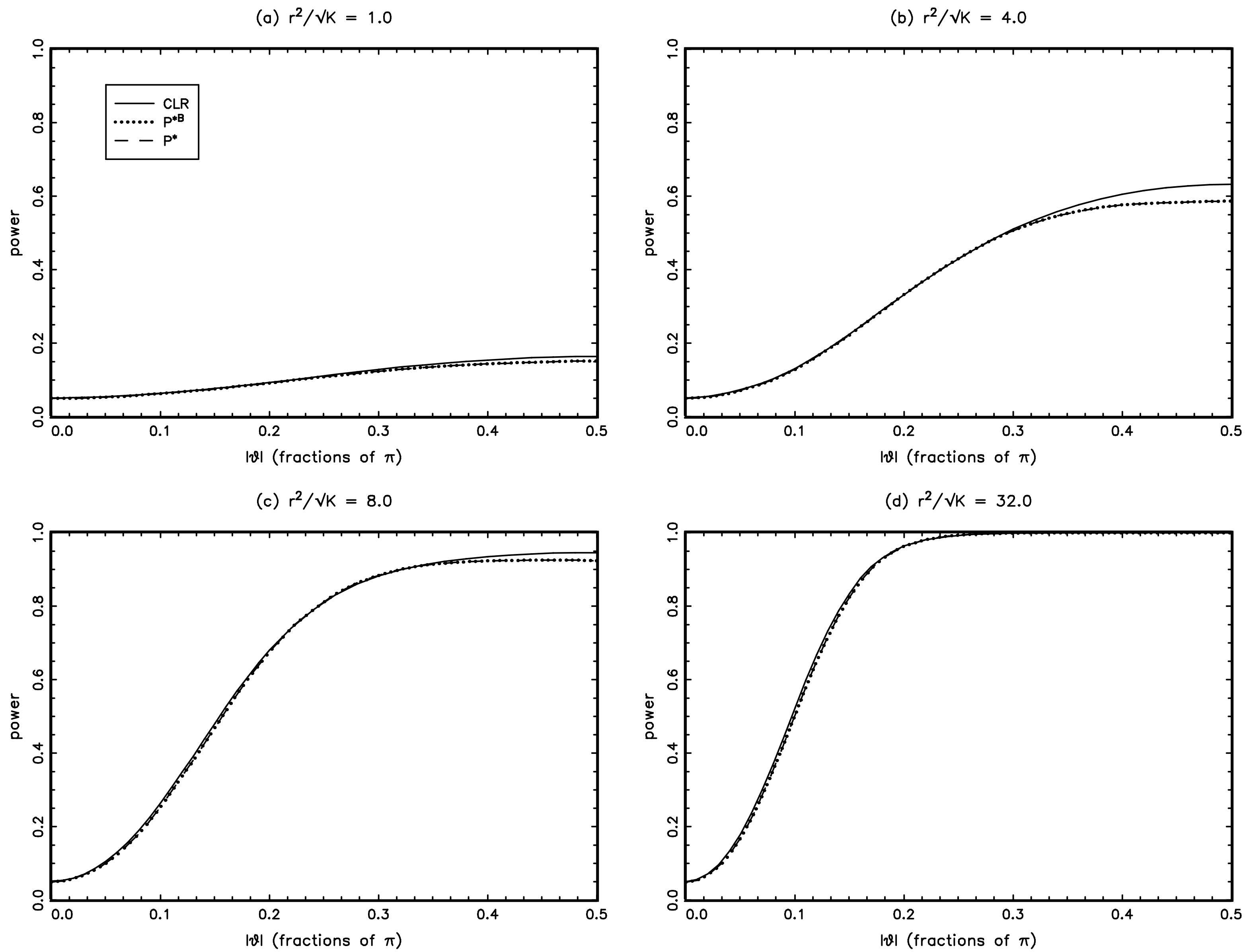

Figure 7 Power functions of the CLR, P^{*B}, and P^{*} tests (in which $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$), for $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, and $k = 10$