2009 Granger Lecture
Granger Center for Time Series
The University of Nottingham

Instrumental Variables Regression, GMM, and Weak Instruments in Time Series

James H. Stock
Department of Economics
Harvard University

June 11, 2009

NBER-NSF Time Series Conference, SMU, Sept. 2005

Outline

1) IV regression and GMM
2) Problems posed by weak instruments
3) Detection of weak instruments
4) Some solutions to weak instruments
5) Current research issues & literature

1) IV Regression and GMM

Philip Wright (1928) needed to estimate the supply equation for butter:

$$\ln(Q_t^{butter}) = \beta_0 + \beta_1 \ln(P_t^{butter}) + u_t$$

Or:

$$y_t = Y_t\beta + u_t, \qquad \text{(generic notation)}$$

β_1 = price elasticity of supply of butter

OLS is inconsistent because $\ln(P_t^{butter})$ is endogenous: price and quantity are determined simultaneously by supply *and* demand…

Philip Wright (1861-1934), MA Harvard, Econ, 1887
Lecturer, Harvard, 1913-1917

Figure 4. Price-output Data Fail to Reveal Either Supply or Demand Curve.

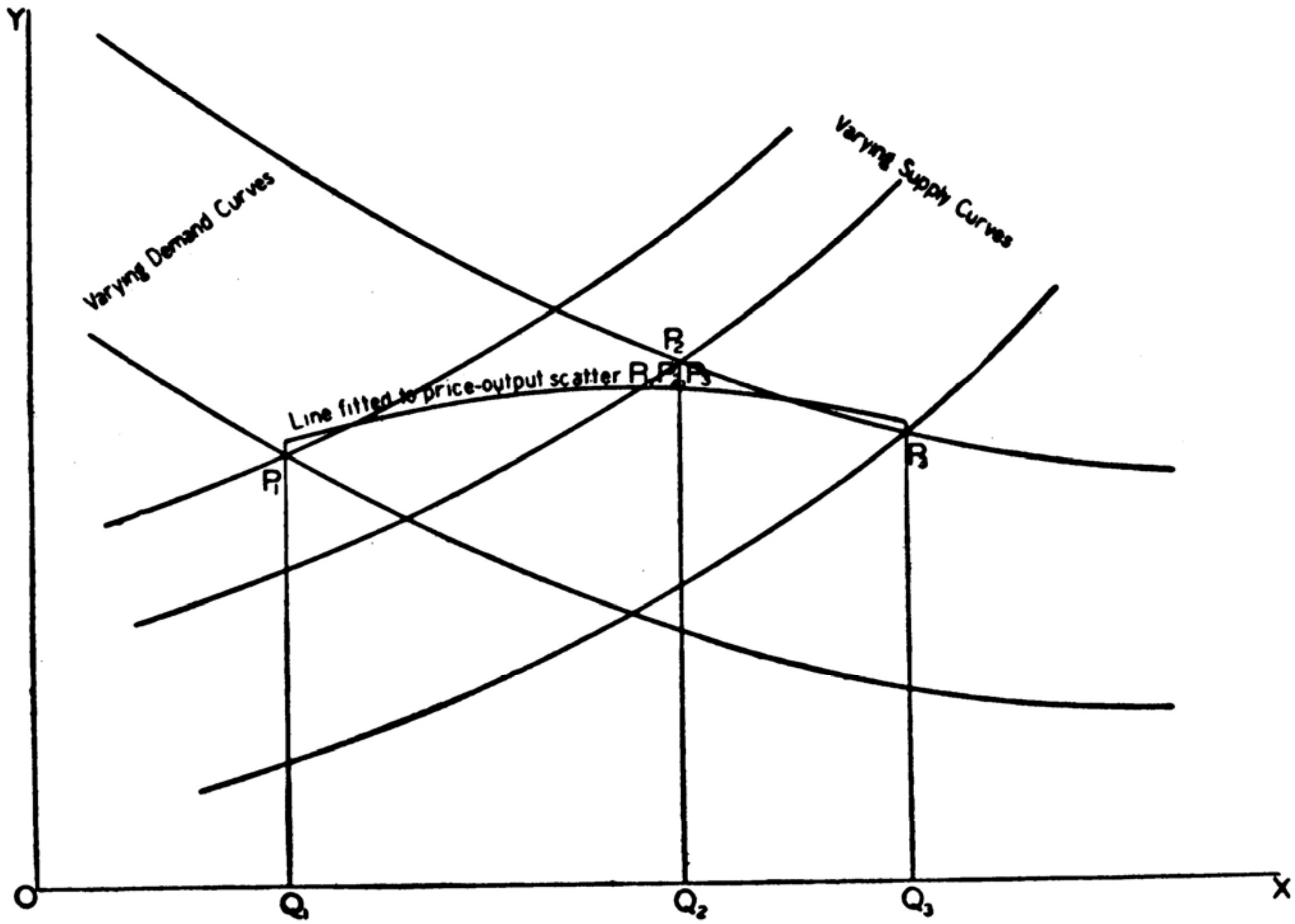


Figure 4, p. 296, Wright (1926), Appendix B

Derivation of the IV estimator in P. Wright (1928, p. 314)

Now multiply each term in this equation by A (the corresponding deviation in the price of a substitute) and we shall have:

$$eA \times P = A \times O - A \times S_1.$$

Suppose this multiplication to be performed for every pair of price-output deviations and the results added, then:

$$e\sum A \times P = \sum A \times O - \sum A \times S_1 \text{ or } e = \frac{\sum A \times O - \sum A \times S_1}{\sum A \times P}.$$

But A was a factor which did not affect supply conditions; hence it is uncorrelated with S_1; hence $\sum A \times S_1 = 0$; and hence $e = \dfrac{\sum A \times O}{\sum A \times P}$.

In modern notation: for an exogenous instrument, i.e. Z_t s.t. $EZ_t u_t = 0$,

$$\mathbf{Z}'(\mathbf{y} - \mathbf{Y}\beta) = \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{Y}\beta = \mathbf{z}'\mathbf{u}, \text{ but } E\mathbf{z}'\mathbf{u} = 0, \text{ so } E\mathbf{Z}'\mathbf{y} - E\mathbf{Z}'\mathbf{Y}\beta = 0,$$

which suggests: $\hat{\beta}^{TSLS} = \dfrac{\mathbf{Z}'\mathbf{y}}{\mathbf{Z}'\mathbf{Y}}$

Classical IV regression model & notation

Equation of interest: $y_t = Y_t\beta + u_t$, $m = \dim(Y_t)$

k exogenous instruments Z_t: $E(u_tZ_t) = 0$, $k = \dim(Z_t)$

Auxiliary equations: $Y_t = \Pi' Z_t + v_t$, $\mathrm{corr}(u_t, v_t) = \rho$ (vector)

Sampling assumption: (y_t, Y_t, Z_t) are i.i.d. (*for now*)

Equations in matrix form: $\mathbf{y} = \mathbf{Y}\beta + \mathbf{u}$ ("second stage")

$\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v}$ ("first stage")

Comments:

- We assume throughout the instrument is exogenous ($E(u_tZ_t) = 0$)
- Included exogenous regressors have been omitted without loss of generality
- Auxiliary equation is just the projection of Y on Z

Generalized Method of Moments

<u>GMM notation and estimator:</u>

GMM "error" term (G equations): $h(Y_t;\theta)$; θ_0 = true value

Note: In the linear model, $h(Y_t;\theta) = y_t - \theta' Y_t$

Errors times k instruments: $\phi_t(\theta) = \overset{G\times 1}{h(Y_t,\theta_0)} \otimes \overset{k\times 1}{Z_t}$

Moment conditions - k instruments: $E\phi_t(\theta) = E[\overset{G\times 1}{h(Y_t,\theta_0)} \otimes \overset{k\times 1}{Z_t}] = 0$

GMM objective function: $S_T(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' W_T \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]$

GMM estimator: $\hat{\theta}$ minimizes $S_T(\theta)$

Efficient (infeasible) GMM: $W_T = \Omega^{-1}$, $\Omega = 2\pi S_{\phi_t(\theta)}(0)$

CUE (Hansen, Heaton, Yaron 1996): $W_T = \hat{\Omega}(\theta)^{-1}$ (each θ)

Weak instruments: four examples

Example #1 (time-series IV): Estimating the elasticity of intertemporal substitution, linearized Euler equation

e.g. Campbell (2003), *Handbook of Economics of Finance*

Δc_{t+1} = consumption growth, t to t+1

$r_{i,t+1}$ = return on i^{th} asset, t to t+1

Log-linearized Euler equation moment condition:

$$E_t(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1}) = 0$$

where

ψ = elasticity of intertemporal substitution (EIS)

$1/\psi$ = coeff. of relative risk aversion under power utility

Resulting IV estimating equation:

$$E[(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1})Z_t] = 0$$

(or use Z_{t-1} because of temporal aggregation)

EIS estimating equations:

$$\Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + u_{i,t+1} \quad \text{(a)}$$

or

$$r_{i,t+1} = \mu_i + (1/\psi)\Delta c_{t+1} + \eta_{i,t+1} \quad \text{(b)}$$

Under homoskedasticity, standard estimation is by the TSLS estimator in (a) or by the inverse of the TSLS estimator in (b).

Findings in literature (e.g. Campbell (2003), US data):

- regression (a): 95% TSLS CI for ψ is (-.14, .28)
- regression (b): 95% TSLS CI for $1/\psi$ is (-.73, 2.14)

What is going on?

Example #2 (cross-section IV): Angrist-Kreuger (1991),

What are the returns to education?

Example #3 (linear GMM): Hybrid New Keynesian Phillips Curve

e.g. Gali and Gertler (1999), where x_t = labor share; see survey by Kleibergen and Mavroeidis (2008).

Example #4 (nonlinear GMM): Estimating the elasticity of intertemporal substitution, nonlinear Euler equation

Hansen, Heaton, Yaron (1996), Stock & Wright (2000), Neely, Roy, & Whiteman (2001)

Working definition of weak identification

θ is ***weakly identified*** if the distributions of GMM or IV estimators and test statistics are not well approximated by their standard asymptotic normal or chi-squared limits because of limited information in the data.

- Departures from standard asymptotics are what matters in practice
- The source of the failures is limited information, not (for example) heavy tailed distributions, near-unit roots, unmodeled breaks, etc.
- The focus is on large T.
- Throughout, we assume instrument exogeneity

Why do weak instruments cause problems?

IV regression with one Y and a single irrelevant instrument

$$\hat{\beta}^{TSLS} = \frac{\mathbf{Z}'\mathbf{y}}{\mathbf{Z}'\mathbf{Y}} = \frac{\mathbf{Z}'(\mathbf{Y}\beta + \mathbf{u})}{\mathbf{Z}'\mathbf{Y}} = \beta + \frac{\mathbf{Z}'\mathbf{u}}{\mathbf{Z}'\mathbf{Y}}$$

If Z is irrelevant (as in Bound et. al. (1995)), then $\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v} = \mathbf{v}$, so

$$\hat{\beta}^{TSLS} - \beta = \frac{\mathbf{Z}'\mathbf{u}}{\mathbf{Z}'\mathbf{v}} = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t u_t}{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t v_t} \xrightarrow{d} \frac{z_u}{z_v}, \text{ where } \begin{pmatrix} z_u \\ z_v \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right)$$

Comments:

- $\hat{\beta}^{TSLS}$ isn't consistent (nor should it be!)
- Distribution of $\hat{\beta}^{TSLS}$ is Cauchy-like (ratio of correlated normals) (Choi & Phillips (1992))

- The distribution of $\hat{\beta}^{TSLS}$ is a ***mixture of normals with nonzero mean***: write $z_u = \delta z_v + \eta$, $\eta \perp z_v$ where $\delta = \sigma_{uv}/\sigma_v^2$. Then

$$\frac{z_u}{z_v} = \frac{\delta z_v + \eta}{z_v} = \delta + \frac{\eta}{z_v}, \text{ and } \frac{\eta}{z_v}|z_v \sim N(0, \frac{\sigma_\eta^2}{z_v^2})$$

so **the asymptotic distribution** of $\hat{\beta}^{TSLS} - \beta_0$ is the mixture of normals,

$$\hat{\beta}^{TSLS} - (\beta_0 + \delta) \xrightarrow{d} \int N(0, \frac{\sigma_\eta^2}{z_v^2}) f_{z_v}(z_v) dz_v \text{ (1 irrelevant instrument)}$$

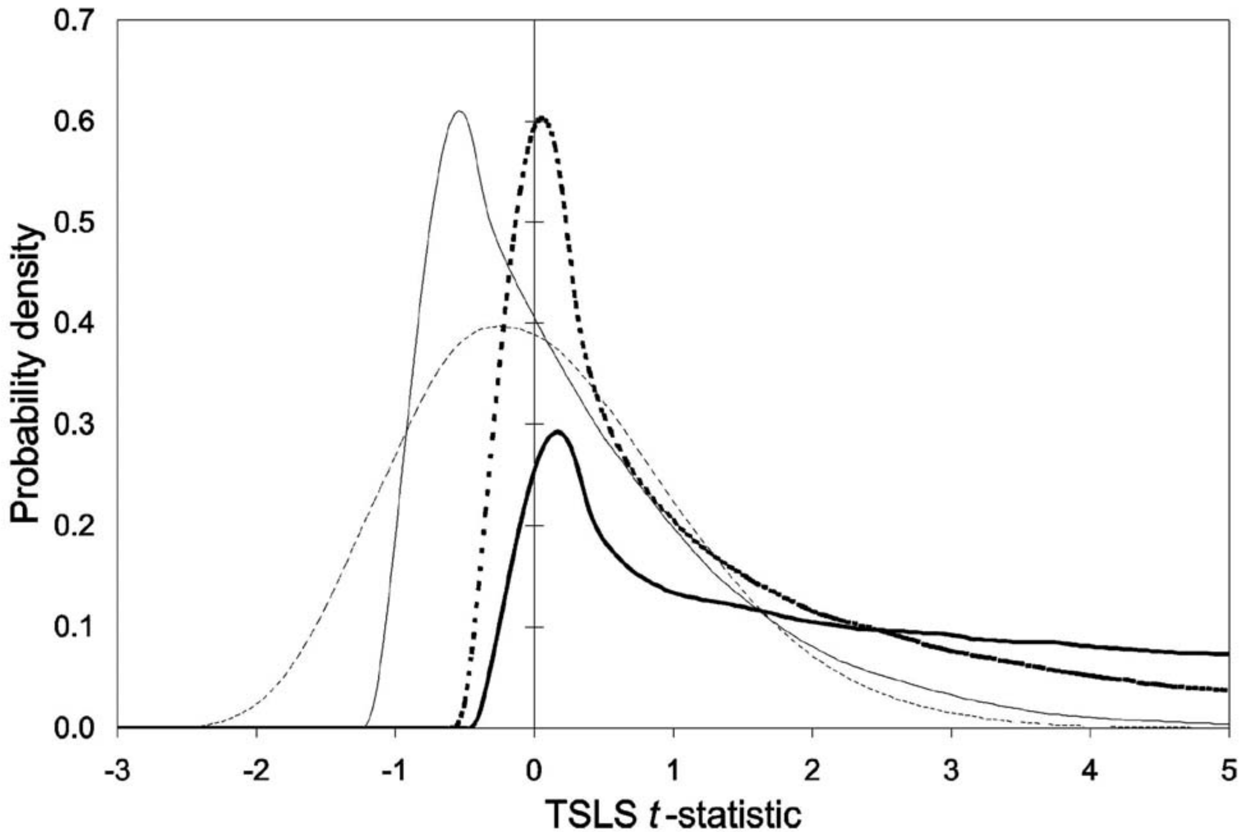
- heavy tails (mixture is based on inverse chi-squared)
- center of distribution of $\hat{\beta}^{TSLS}$ is $\beta_0 + \delta$. But

$$\hat{\beta}^{OLS} - \beta_0 = \frac{\mathbf{Z}'\mathbf{u}/T}{\mathbf{Z}'\mathbf{Z}/T} = \frac{\mathbf{v}'\mathbf{u}/T}{\mathbf{v}'\mathbf{v}/T} \xrightarrow{p} \frac{\sigma_{uv}}{\sigma_v^2} = \delta, \text{ so } plim(\hat{\beta}^{OLS}) = \beta_0 + \delta,$$

so

$$\hat{\beta}^{TSLS} - plim(\hat{\beta}^{OLS}) \xrightarrow{d} \int N(0, \frac{\sigma_\eta^2}{z_v^2}) f_{z_v}(z_v) dz_v \text{ (1 irrelevant instrument)}$$

The unidentified and strong-instrument distributions are two ends of a spectrum. Distribution of the TSLS t-statistic (Nelson-Startz (1990a,b)):



Dark line = irrelevant instruments; dashed light line = strong instruments; intermediate cases: weak instruments. The key parameter is:

$$\mu^2 = \Pi'\mathbf{Z}'\mathbf{Z}\Pi/\sigma_v^2 \text{ (\textbf{\textit{concentration parameter}})}$$

$$= k \times \text{(numerator) noncentrality parameter of first-stage } F \text{ statistic}$$

Weak instrument asymptotics bridges this spectrum.

Adopt nesting that makes the concentration parameter tend to a constant as the sample size increases by setting

$$\Pi = C/\sqrt{T} \quad \text{(weak instrument asymptotics)}$$

- This is the Pitman drift for obtaining the local power function of the first-stage F.
- This nesting holds $E\mu^2$ constant as $T \to \infty$.
- Under this nesting, $F \xrightarrow{d}$ noncentral χ_k^2/k with noncentrality parameter $E\mu^2/k$ (so $F = O_p(1)$)
- Letting the parameter depend on the sample size is a common ways to obtain good approximations – e.g. local to unit roots (Bobkoski 1983, Cavanagh 1985, Chan and Wei 1987, and Phillips 1987)

Weak IV asymptotics for TSLS estimator, 1 included endogenous vble:

$$\hat{\beta}^{TSLS} - \beta_0 = (\mathbf{Y}'\mathrm{P}_{\mathbf{Z}}\mathbf{u})/(\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y})$$

Now

$$\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y} = \left(\frac{(\mathbf{Z}\Pi + \mathbf{v})'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'(\mathbf{Z}\Pi + \mathbf{v})}{\sqrt{T}}\right)$$

$$= \left(\frac{\Pi\mathbf{Z}'\mathbf{Z}}{\sqrt{T}} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2\prime}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2}\left(\frac{\mathbf{Z}'\mathbf{Z}\Pi}{\sqrt{T}} + \frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}}\right)$$

$$= \left[C'\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{1/2} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2\prime}\right]\left[\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{1/2\prime}C + \left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1/2}\frac{\mathbf{Z}'\mathbf{v}}{\sqrt{T}}\right]$$

$$\xrightarrow{d} (\lambda + z_v)'\,(\lambda + z_v),$$

where

$$\lambda = C'Q_{ZZ}^{1/2},\ Q_{ZZ} = EZ_tZ_t',\ \text{and}\ \begin{pmatrix} z_u \\ z_v \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right)$$

Similarly,

$$\mathbf{Y}'P_{\mathbf{Z}}\mathbf{u} = \left(\frac{(\mathbf{Z}\Pi + \mathbf{v})'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u})}{\sqrt{T}}\right)$$

$$= \left(C'\frac{\mathbf{Z}'\mathbf{Z}}{T} + \frac{\mathbf{v}'\mathbf{Z}}{\sqrt{T}}\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}\right)^{-1}\left(\frac{\mathbf{Z}'\mathbf{u}}{\sqrt{T}}\right)$$

$$\xrightarrow{d} (\lambda + z_v)'z_u,$$

so

$$\hat{\beta}^{TSLS} - \beta_0 \xrightarrow{d} \frac{(\lambda + z_v)'z_u}{(\lambda + z_v)'(\lambda + z_v)}$$

- Under weak instrument asymptotics, $\mu^2 \xrightarrow{p} C'Q_{ZZ}C/\sigma_v^2 = \lambda'\lambda/\sigma_v^2$
- Unidentified special case: $\hat{\beta}^{TSLS} - \beta_0 \xrightarrow{d} \dfrac{z_v{}'z_u}{z_v{}'z_v}$ (obtained earlier)
- Strong IVs: $\sqrt{\lambda'\lambda}\,(\hat{\beta}^{TSLS} - \beta_0) \xrightarrow{d} \dfrac{\lambda' z_u}{\sqrt{\lambda'\lambda}} \sim N(0,\sigma_u^2)$ (standard limit)

Summary of weak IV asymptotic results:

- Resulting asymptotic distributions are the same as in the exact normal classical model with fixed Z – but with ***known*** covariance matrices.
- Weak IV asymptotics yields good approximations to sampling distributions uniformly in μ^2 for T moderate or large.
- Under this nesting:
  - IV estimators are not consistent, are nonnormal
  - Test statistics (including the J-test of overidentifying restrictions) do not have normal or chi-squared distributions
  - Conventional confidence intervals do not have correct coverage
- Because μ^2 is unknown, these distributions can't be used directly in practice to obtain a "corrected" distribution for purposes of inference

3) Detection of weak instruments

How weak is weak? Need a cutoff value for μ^2

$$\hat{\beta}^{TSLS} - \beta_0 \xrightarrow{d} \frac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)},$$

where $\mu^2 = \Pi'\mathbf{Z}'\mathbf{Z}\Pi/\sigma_v^2$ (***concentration parameter***)

$= k \times$ (numerator) noncentrality parameter of first-stage F

$\xrightarrow{p} \lambda'\lambda/\sigma_v^2$

For what values of μ^2 does $\dfrac{(\lambda + z_v)' z_u}{(\lambda + z_v)'(\lambda + z_v)} \approx \dfrac{\lambda' z_u}{\lambda'\lambda}$?

Various procedures:

- First stage F > 10 rule of thumb (Staiger-Stock (1997))
- Stock-Yogo (2005a) relative bias method (approximately yields F>10)
- Stock-Yogo (2005a) size method
- Hahn-Hausman (2003) test
- Other methods (R^2, partial R^2, Shea (1997), etc.)

TSLS relative bias cutoff method (Stock-Yogo (2005a))

Some background:

The relative squared normalized bias of TSLS to OLS is,

$$B_n^2 = \frac{(E\hat{\beta}^{\mathrm{IV}} - \beta)'\Sigma_{YY}(E\hat{\beta}^{\mathrm{IV}} - \beta)}{(E\hat{\beta}^{\mathrm{OLS}} - \beta)'\Sigma_{YY}(E\hat{\beta}^{\mathrm{OLS}} - \beta)}$$

The square root of the maximal relative squared asymptotic bias is:

$$B^{max} = \max_{\rho:\, 0 < \rho'\rho \leq 1} \lim_{n\to\infty}|B_n|, \text{ where } \boldsymbol{\rho} = \mathrm{corr}(u_t, v_t)$$

This maximization problem is a ratio of quadratic forms so it turns into a (generalized) eigenvalue problem; algebra reveals that the solution to this eigenvalues problem depends only on μ^2/k and k; this yields the cutoff μ_{bias}^2.

Critical values

One included endogenous regressor

The 5% critical value of the test is the 95% percentile value of the noncentral χ_k^2/k distribution, with noncentrality parameter μ_{bias}^2/k

Multiple included endogenous regressors

The Cragg-Donald (1993) statistic is:

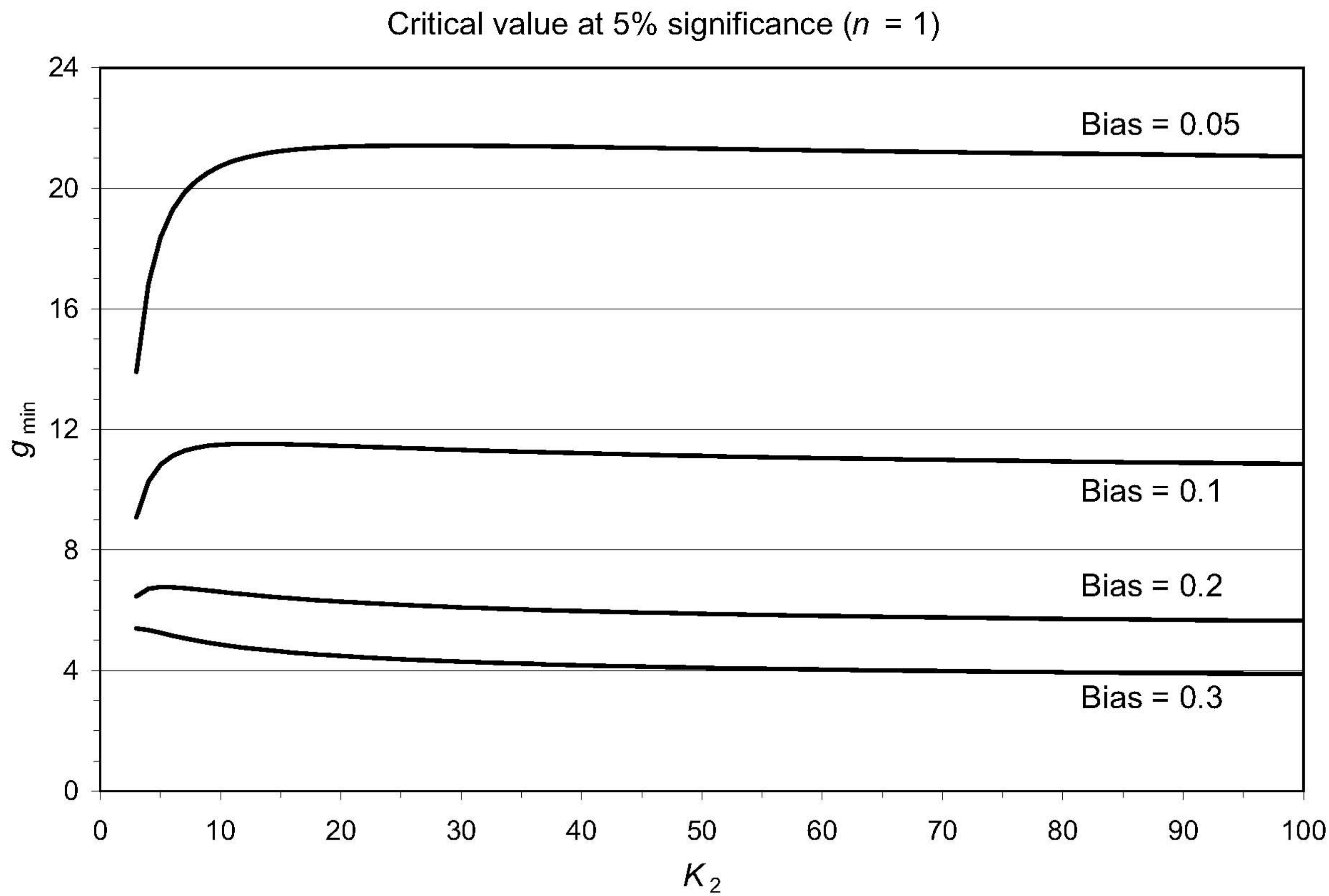
$$g_{min} = \text{mineval}(G_T), \text{ where } G_T = \hat{\Sigma}_{\mathbf{VV}}^{-1/2}{}'\mathbf{Y}'P_{\mathbf{Z}}\mathbf{Y}\hat{\Sigma}_{\mathbf{VV}}^{-1/2}/k,$$

- G_T is essentially a matrix first stage F statistic
- Critical values are given in Stock-Yogo (2005a)

Software

STATA (ivreg2),…

5% critical value of F to ensure indicated maximal bias (Stock-Yogo, 2005a)



To ensure 10% maximal bias, need $F < 11.52$; $F < 10$ is a rule of thumb

Other methods for detecting weak instruments

R^2, partial R^2, or adjusted R^2

- None of these are a good idea, more precisely, what needs to be large is the concentration parameter, not the R^2. An $R^2 = .10$ is small if $T = 50$ but is large if $T = 5000$.
- The first-stage R^2 is especially uninformative if the first stage regression has included exogenous regressors (W's) because it is the marginal explanatory content of the Z's, given the W's, that matters.

Hahn-Hausman (2003) test

- Idea is to test the null of strong instruments, under which the TSLS estimator, and the inverse of the TSLS estimator from the "reverse" regression, should be the same
- Unfortunately the HH test is not consistent against weak instruments (power of 5% level test depends on parameters, is typically ≈ 15-20% (Hausman, Stock, Yogo (2005))

4) Some Solutions to Weak Instruments

There are two approaches to improving inference (providing tools):

Fully robust methods:

- Inference that is valid for any value of the concentration parameter, including zero, at least if the sample size is large, under weak instrument asymptotics
  - For tests: asymptotically correct size (and good power!)
  - For confidence intervals: asymptotically correct coverage rates
  - For estimators: asymptotically unbiased (or median-unbiased)

Partially robust methods:

- Methods are less sensitive to weak instruments than TSLS – e.g. bias is "small" for a "large" range of μ^2

Fully Robust Testing

- Approach #1: use a "worst case" (over all possible values of μ^2) critical value for the TSLS t-stat
  - leads to low-power procedures
- Approach #2: use a statistic whose distribution does not depend on μ^2 (two such statistics are known)
- Approach #3: use statistics whose distribution depends on μ^2, but compute the critical values as a function of another statistic that is sufficient for μ^2 under the null hypothesis.
- Approach #4: "optimal" nonsimilar tests (subsumes 1-3)

Approach #2: Tests that are valid unconditionally – linear IV case

(that is, the distribution of the test statistic does not depend on μ^2)

The Anderson-Rubin (1949) test

Consider H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in $\quad \mathbf{y} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{u}$,

$$\mathbf{Y} = \mathbf{Z}\Pi + \mathbf{v}$$

The Anderson-Rubin (1949) statistic is the F-statistic in the regression of $\mathbf{y} - \mathbf{Y}\beta_0$ on $\mathbf{Z}$.

$$\mathrm{AR}(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

$$\mathrm{AR}(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

Comments

- $\mathrm{AR}(\hat{\beta}^{TSLS})$ = Hansen's (2003) *J*-statistic
- Null distribution doesn't depend on μ^2:

  Under the null, $\mathbf{y} - \mathbf{Y}\beta_0 = \mathbf{u}$, so

  $$\mathrm{AR} = \frac{\mathbf{u}' P_{\mathbf{Z}} \mathbf{u} / k}{\mathbf{u}' M_{\mathbf{Z}} \mathbf{u} / (T - k)} \sim F_{k,n-k} \qquad \text{if } u_t \text{ is normal}$$

  $$\mathrm{AR} \xrightarrow{d} \chi_k^2 / k \qquad \text{if } u_t \text{ is i.i.d. and } Z_t u_t \text{ has 2 moments (CLT)}$$
- The distribution of AR under the alternative depends on μ^2 – more information, more power (of course)
- Difficult to interpret: rejection arises for two reasons: β_0 is false *or* Z is endogenous
- Power loss relative to other tests; inefficient under strong instruments

Kleibergen's (2002) LM test

Kleibergen developed an LM test that has a null distribution that is χ_1^2 - doesn't depend on μ^2.

- Fairly easy to implement
- Is efficient if instruments are strong
- Has very strange power properties (we shall see)
- Its power is dominated by the conditional likelihood ratio test

Approach #3: Conditional tests

Conditional tests have rejection rate 5% for all points under the null (β_0, μ^2) ("similar tests"). Moreira (2003):

- LR tests $\beta = \beta_0$ using the LIML likelihood:

$$LR = \max_{\beta} \text{log-likelihood}(\beta) - \text{log-likelihood}(\beta_0)$$

- Q_T is sufficient for μ^2 under the null
- Thus the distribution of $LR| Q_T$ does not depend on μ^2 under the null
- Thus valid inference can be conducted using the quantiles of $LR| Q_T$, that is, use critical values which are a function of Q_T

Moreira's (2003) conditional likelihood ratio (CLR) test

$$LR = \max_{\beta} \text{log-likelihood}(\beta) - \text{log-likelihood}(\beta_0)$$

After some algebra, this becomes:

$$LR = ½\{\hat{Q}_S - \hat{Q}_T + [(\hat{Q}_S - \hat{Q}_T)^2 + 4\hat{Q}_{ST}^2]^{1/2}\}$$

where

$$\hat{Q} = \begin{bmatrix} \hat{Q}_S & \hat{Q}_{ST} \\ \hat{Q}_{ST} & \hat{Q}_T \end{bmatrix} = \hat{J}_0{}'\hat{\Omega}^{-1/2}\mathbf{Y}^{+}{}'P_{\mathbf{Z}}\mathbf{Y}^{+}\hat{\Omega}^{-1/2}{}'\hat{J}_0$$

$$\hat{\Omega} = \mathbf{Y}^{+}{}'M_{\mathbf{Z}}\mathbf{Y}^{+}/(T-k), \ \ \mathbf{Y}^{+} = (\mathbf{y}\ \ \mathbf{Y})$$

$$\hat{J}_0 = \left[\frac{\hat{\Omega}^{1/2}{}'b_0}{\sqrt{b_0{}'\hat{\Omega}b_0}} \quad \frac{\hat{\Omega}^{-1/2}a_0}{\sqrt{a_0{}'\hat{\Omega}^{-1}a_0}} \right], b_0 = \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} a_0 = \begin{pmatrix} \beta_0 \\ 1 \end{pmatrix}.$$

CLR test: Comments

- More powerful than AR or LM
- In fact, effectively uniformly most powerful among asymptotically efficient similar tests that are invariant to rotations of the instruments (Andrews, Moreira, Stock (2006))
- STATA (condivreg), Gauss code for computing LR and conditional *p*-values exists

but..

- Only developed (so far) for a single included endogenous regressor
- As written here, the software requires homoskedastic errors; extensions to heteroskedasticity and serial correlation have been developed but are not in common statistical software

Approach #4: Nonsimilar tests (Andrews, Moreira, Stock (2008))

Polar coordinate transform (Hillier (1990), Chamberlain (2005)):

$$r^2 = \lambda'\lambda h'h, \qquad h = \begin{pmatrix} c_\beta \\ d_\beta \end{pmatrix} = \begin{pmatrix} (\beta\text{-}\beta_0)/\sqrt{b_0{}'\Omega b_0} \\ a'\Omega^{-1}a_0/\sqrt{a_0{}'\Omega^{-1}a_0} \end{pmatrix}$$

$$x(\theta) = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} = h/\sqrt{h'h} \quad (\text{so } x(\theta_0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix})$$

Mapping:

$\underline{\boldsymbol{\beta}}$	$\underline{\boldsymbol{\theta}}$
β_0	0
$< \beta_0$ ($> \beta_0$)	< 0 (> 0)
∞	$\theta_\infty = lim_{\beta\to\infty}\cos^{-1}[d_\beta/(h'h)^{1/2}]$
$-\infty$	$\theta_{-\infty} = \theta_\infty - \pi$

Compound null hypothesis and two-sided alternative:

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

Strategy (follows Lehmann (1986))

1. Null: transform compound null into point null via weighs Λ:

$$h_\Lambda(q) = \int f_Q(q; r, \theta_0)\, d\Lambda(r)$$

2. Alternative: transform into point alternative via equal weighting of (r_1, $\pm\theta_1$) (this is a necessary but not sufficient condition for nonsimilar tests to be AE):

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

3. <u>Point optimal invariant test of h_Λ vs. g</u>: from Neyman -Pearson Lemma, reject if

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

4. <u>Least favorable distribution Λ</u>: $NP_{\Lambda,r_1,|\theta_1|}(q)$ is POINS for the original distribution if Λ is least favorable, that is, if

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

5. <u>POINS</u> <u>Power envelope</u>.

The PE of POINS tests of (*) is the envelope of power functions of $NP_{\Lambda^{LF},r_1,|\theta_1|}(q)$, where Λ^{LF} is the least favorable distribution

A closed form, POINS test of $\theta = 0$ (using theoretical results on one-point least favorable distributions + Bessel function approximations):

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

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

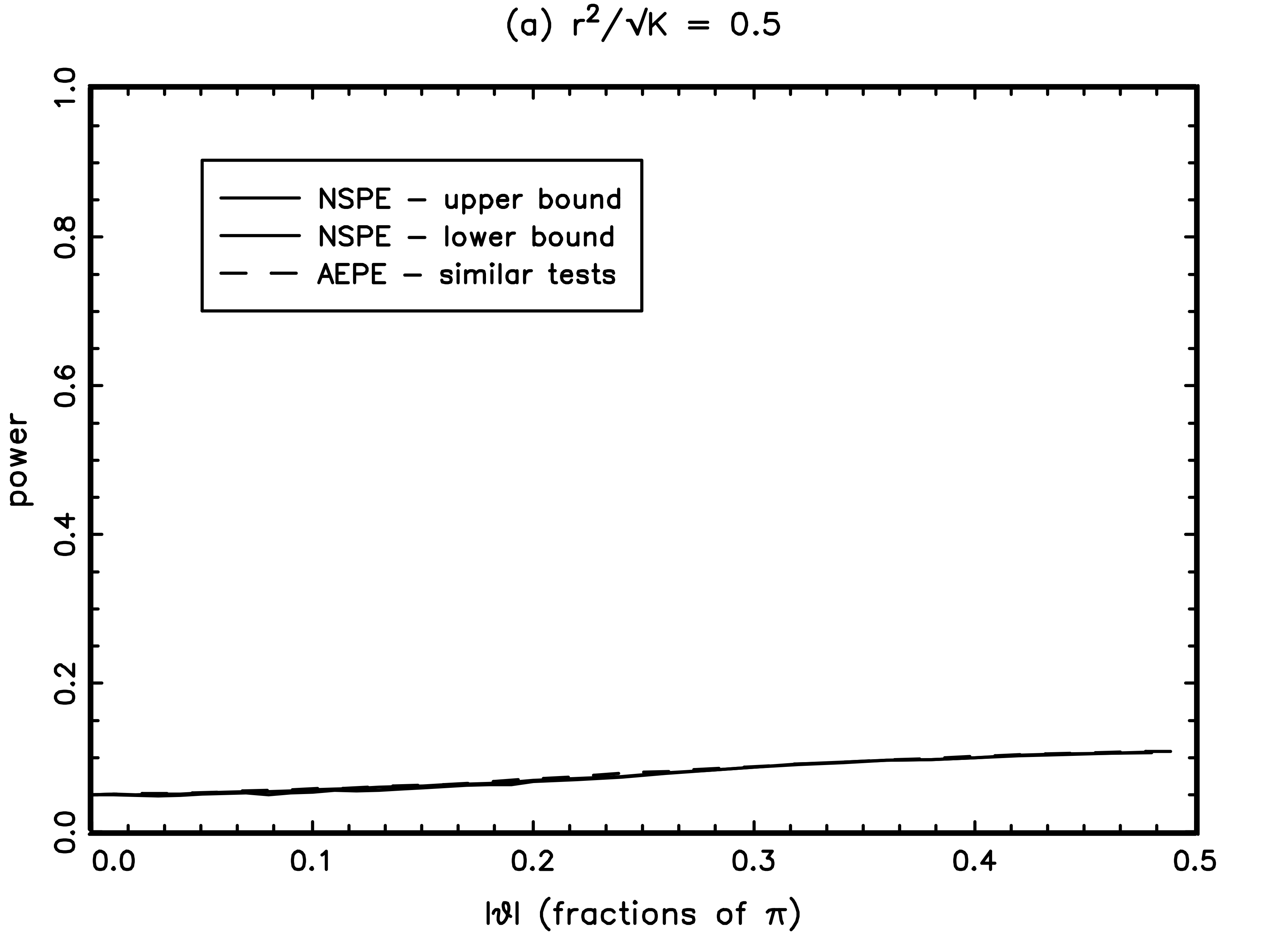
$$\phi_0 = \sqrt{\nu^2 + z_0} + \nu \ln\left(\frac{z_0}{\nu + \sqrt{\nu^2 + z_0}}\right) \text{ (etc.)}, \ \nu = (k-2)/2$$

Numerical search over r_1, θ_1 resulted in $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$

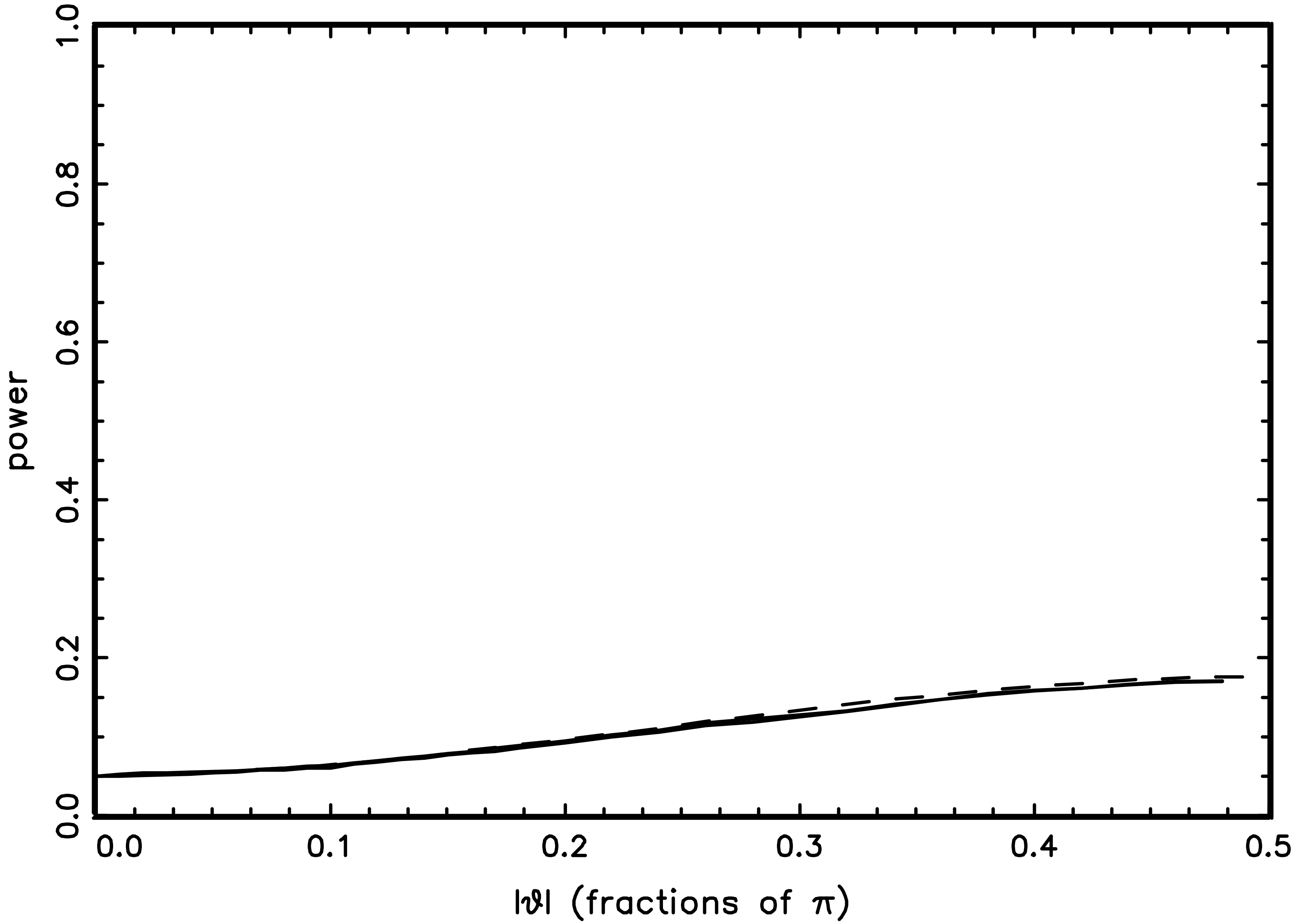
Andrews, Moreira and Stock (2008), Figures 2/3

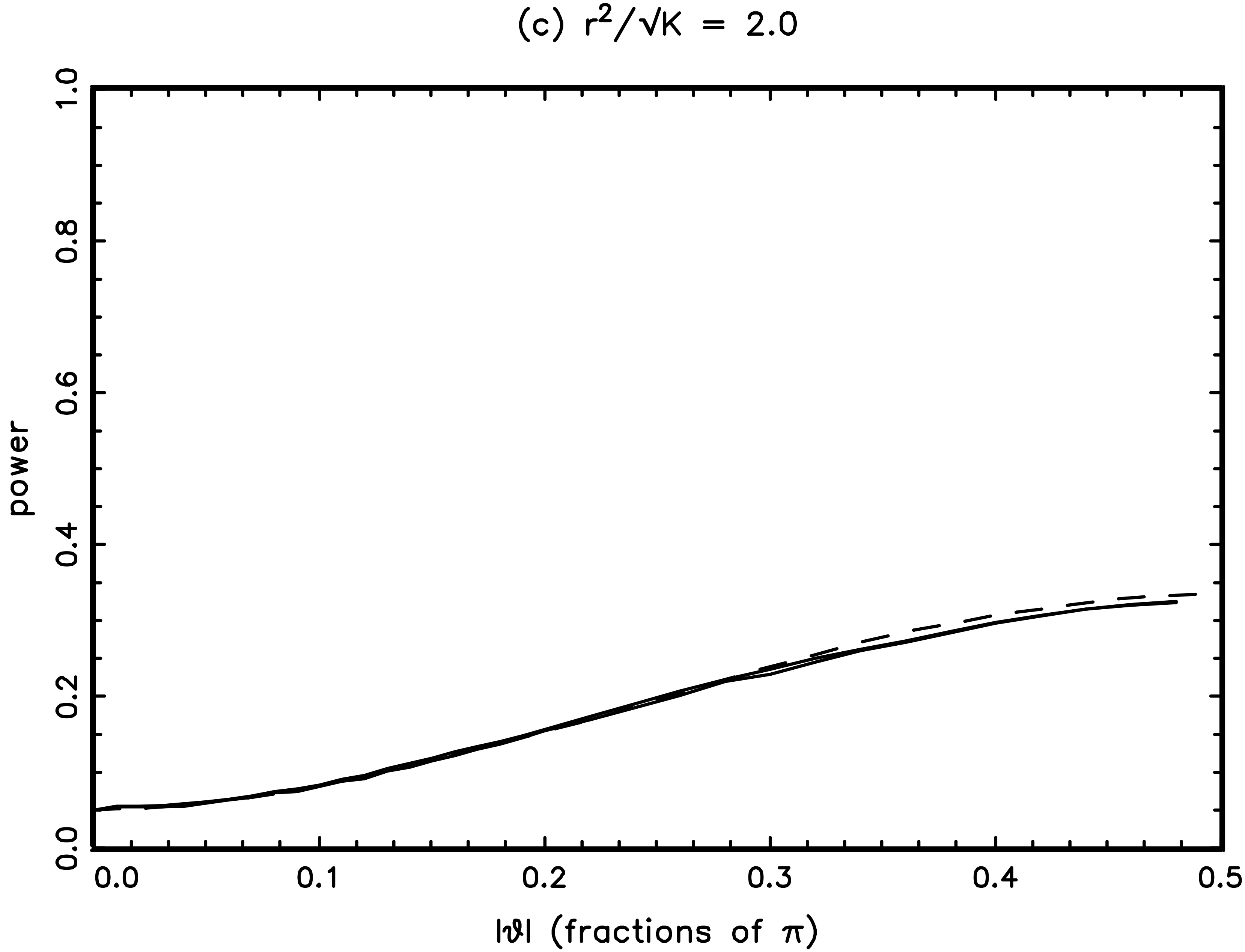
Upper and lower bound on power envelope for nonsimilar invariant tests against (r, $|\theta|$) and power envelope for similar invariant tests against (r, $|\theta|$),

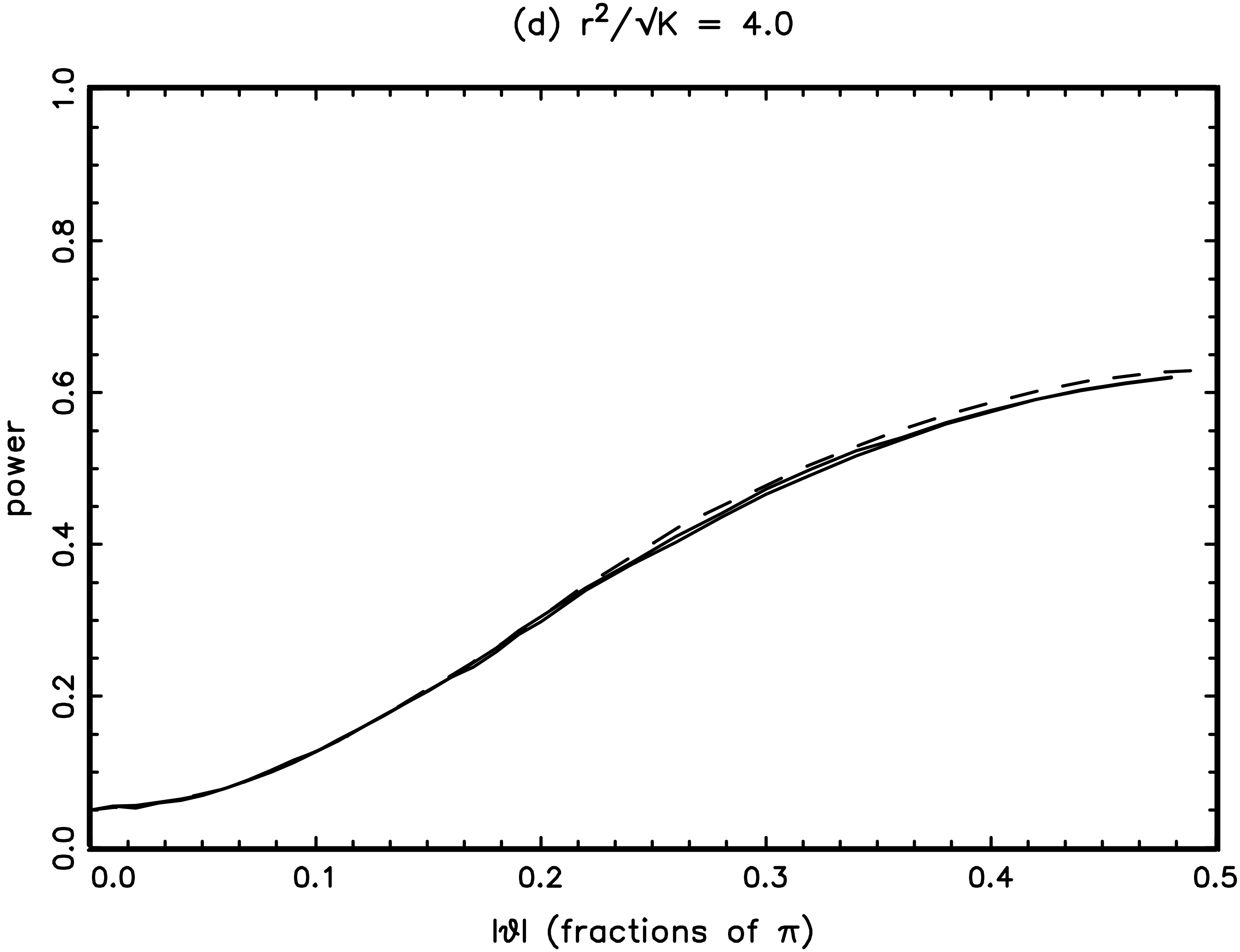
$0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 0.5, 1, 2, 4, 8, 16, 32, 64$; $k = 5$


(a) $r^2/\sqrt{K} = 0.5$
NSPE – upper bound
NSPE – lower bound
AEPE – similar tests
power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
$|\vartheta|$ (fractions of π)

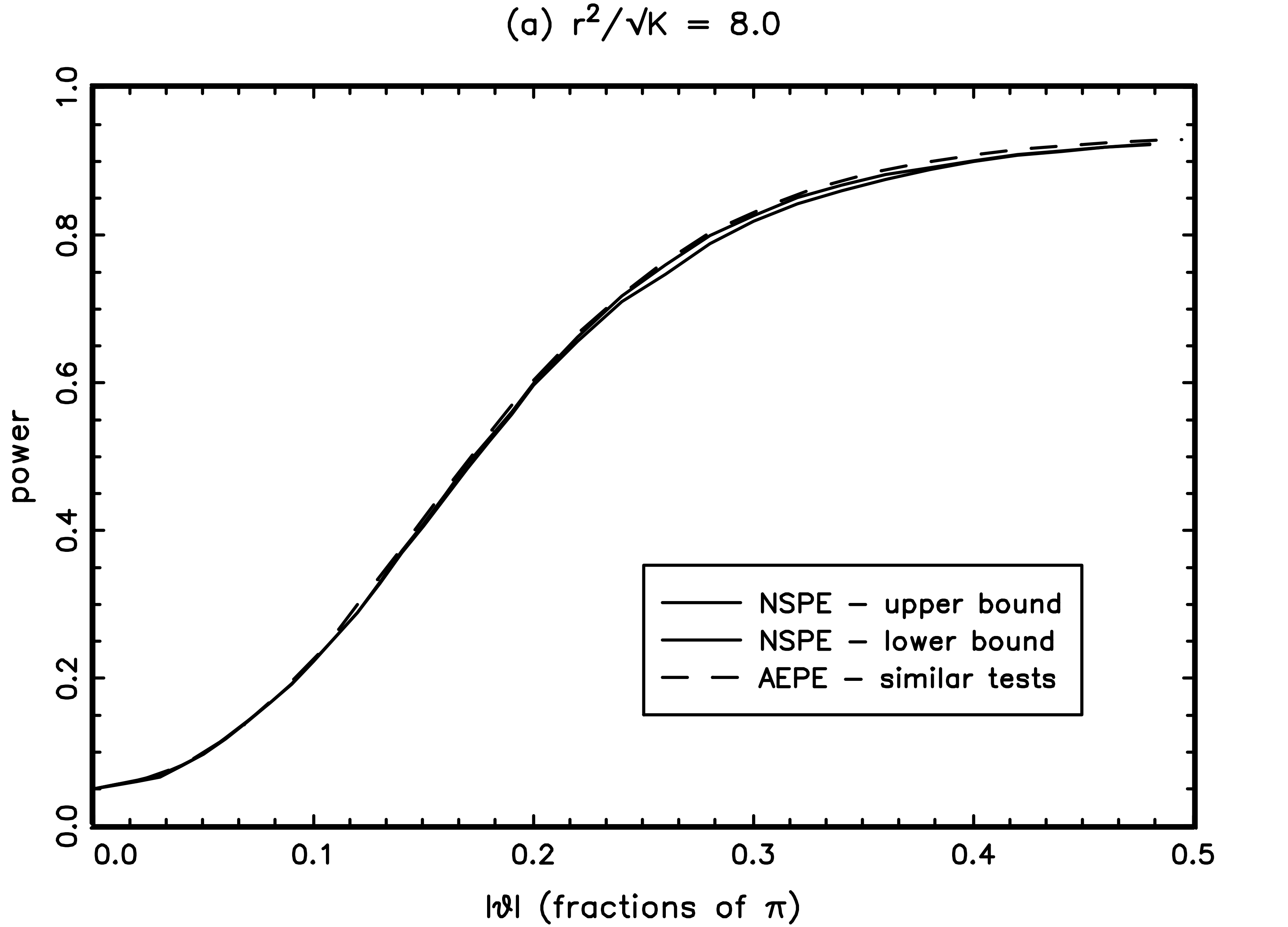
(b) $r^2/\sqrt{K} = 1.0$



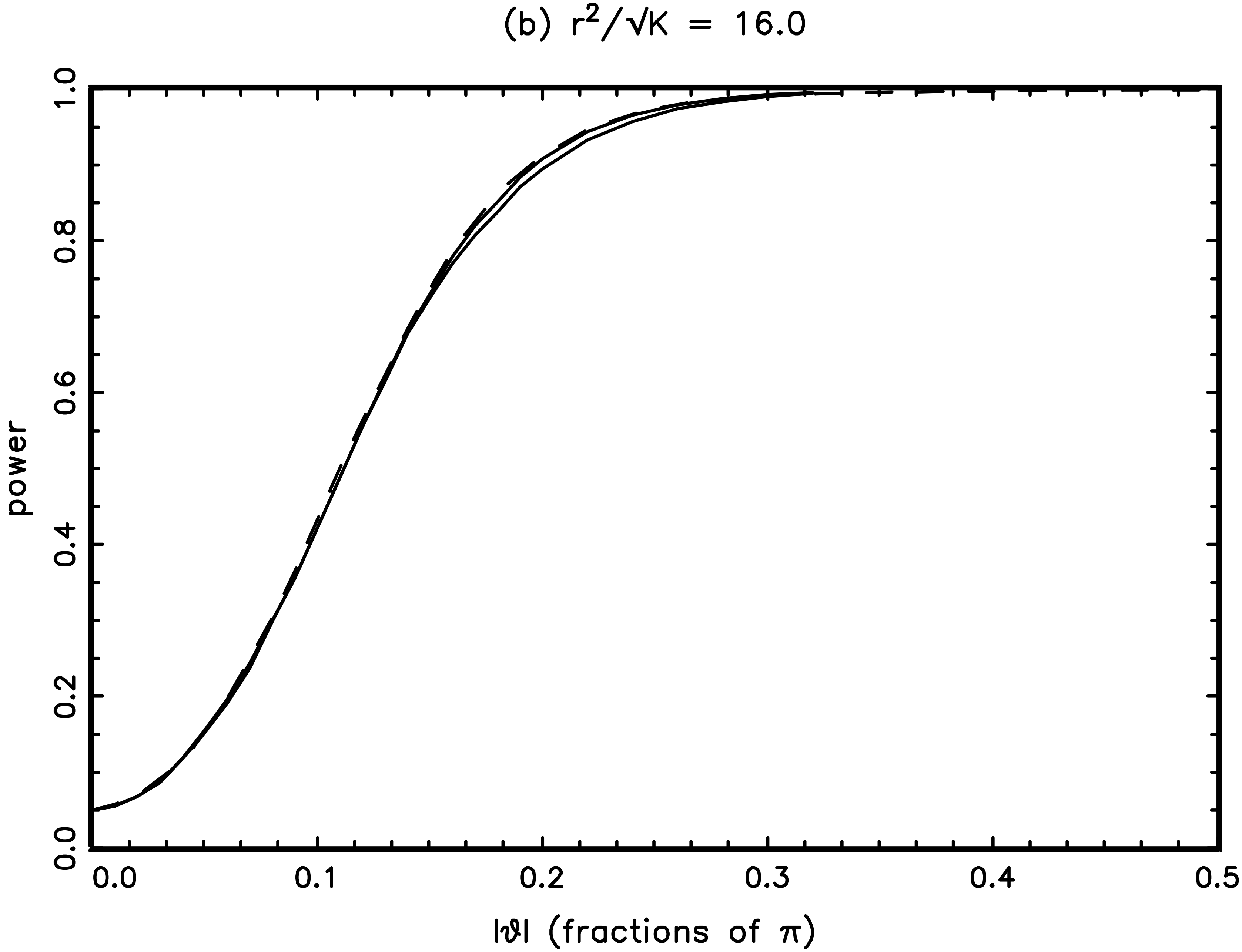

(c) r²/√K = 2.0
power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
|ϑ| (fractions of π)



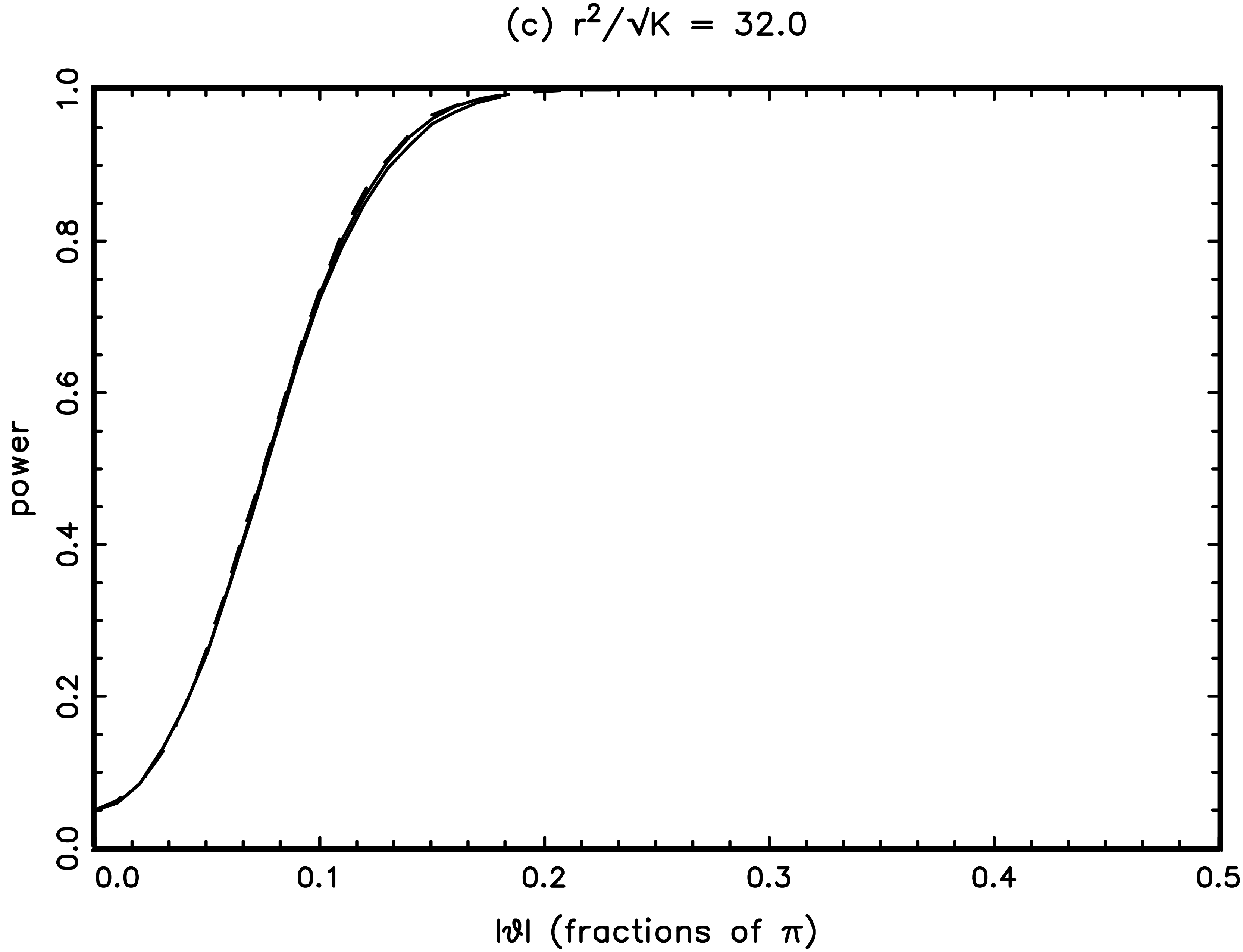
(d) $r^2/\sqrt{K}$ = 4.0
1.0
0.8
0.6
0.4
0.2
0.0
power
0.0
0.1
0.2
0.3
0.4
0.5
$|\vartheta|$ (fractions of π)



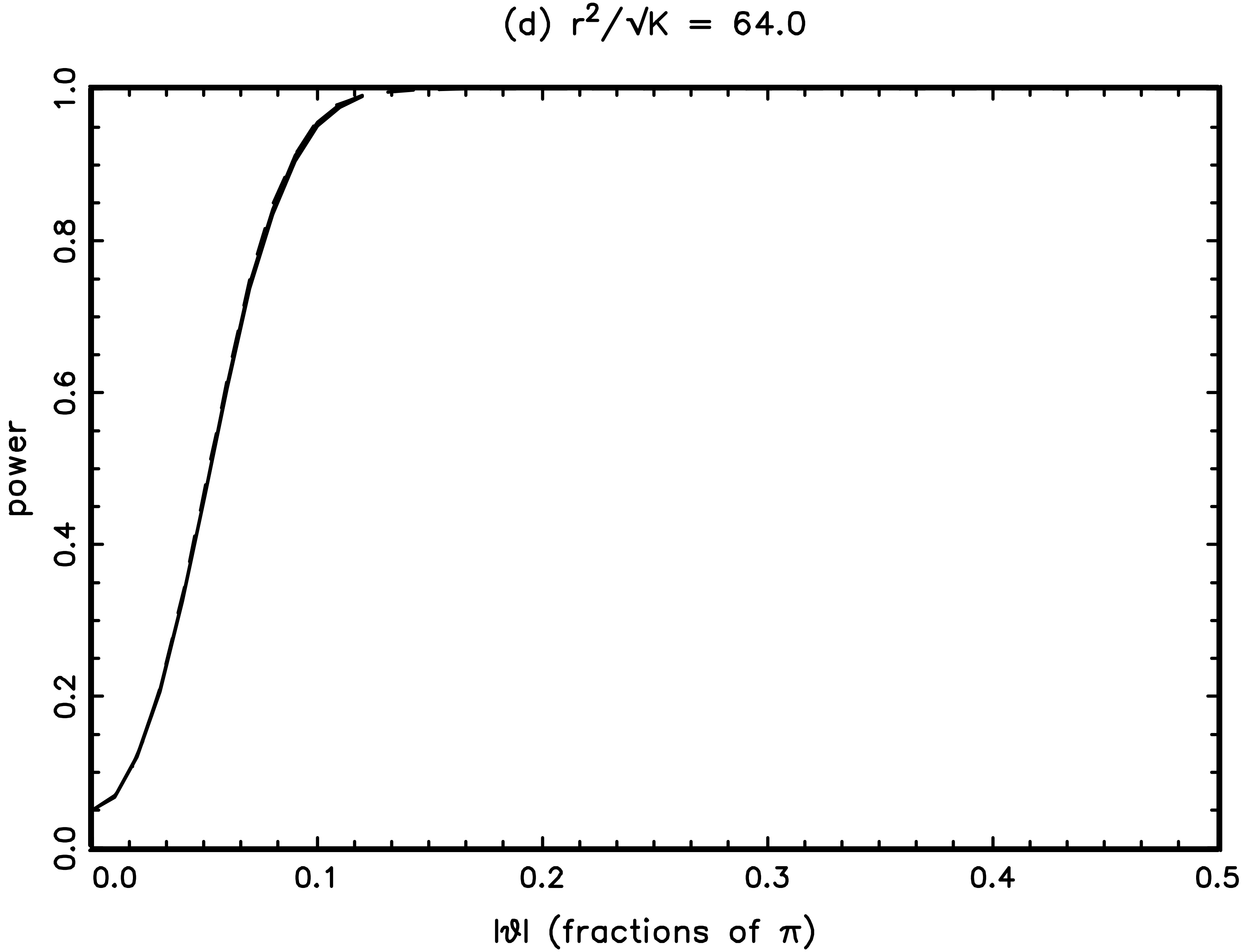
(a) $r^2/\sqrt{K}$ = 8.0
power
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
|ϑ| (fractions of π)
NSPE – upper bound
NSPE – lower bound
AEPE – similar tests



(b) $r^2/\sqrt{K}$ = 16.0
power
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
|ϑ| (fractions of π)



(c) $r^2/\sqrt{K}$ = 32.0
power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
$|\vartheta|$ (fractions of π)


(d) $r^2/\sqrt{K}$ = 64.0
power
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
$|\vartheta|$ (fractions of π)

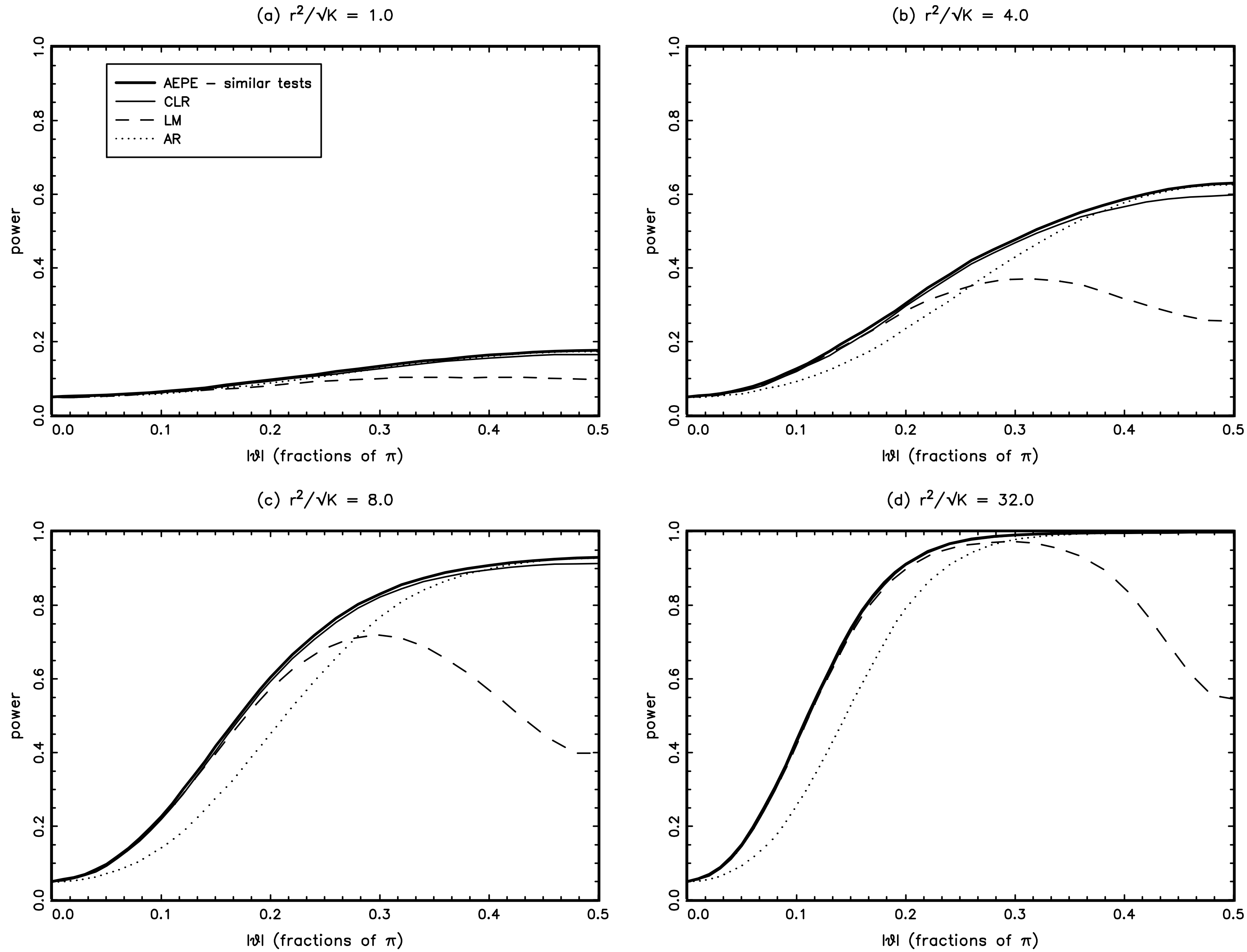


Figure 4 Power envelope for similar invariant tests against $(r, |\theta|)$ and power functions of the CLR, LM, and AR tests, $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, $k = 5$

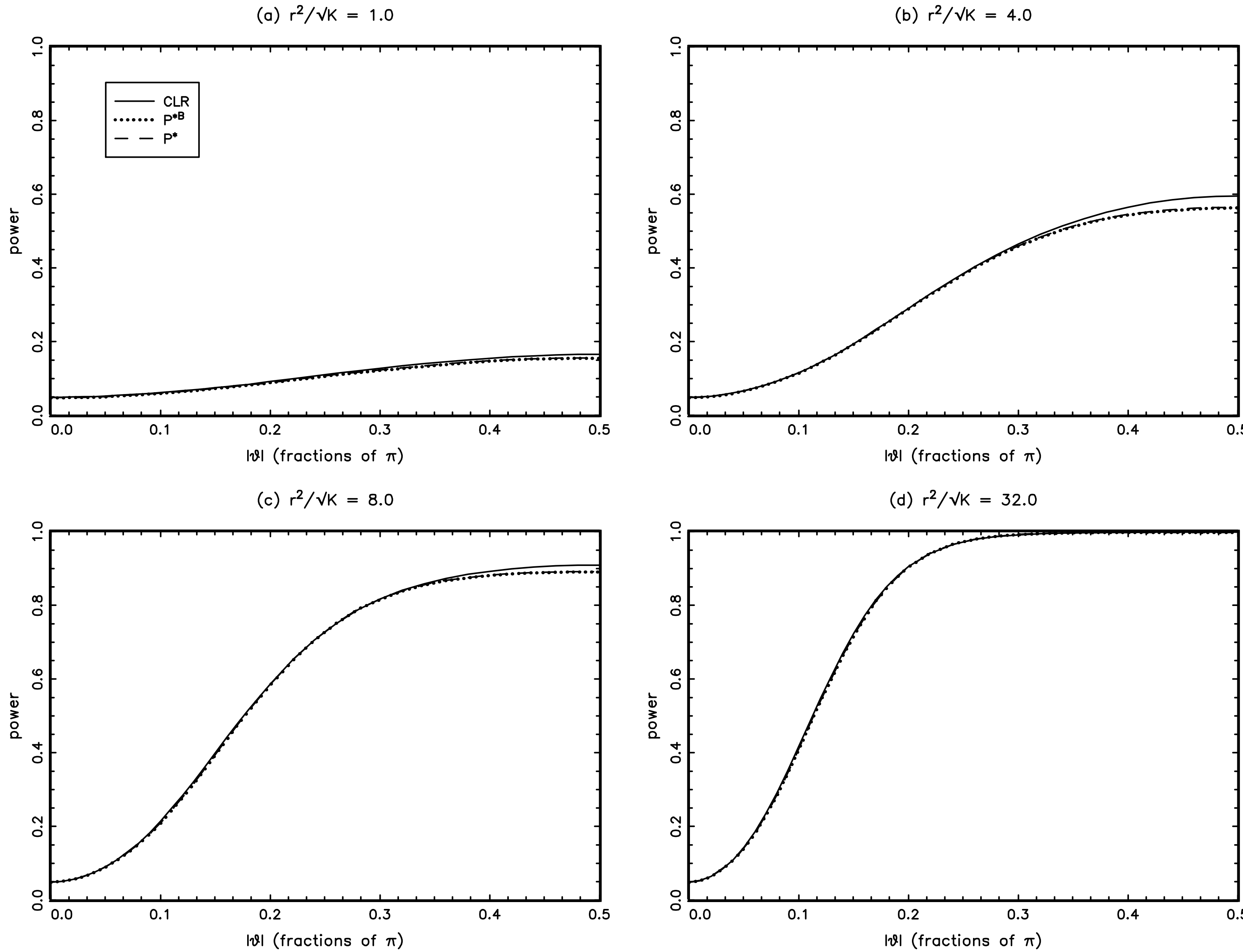


Figure 5 Power functions of the CLR, P^{*B}, and P^{*} tests (in which $r_1^2 = \sqrt{20k}$ and $\theta_1 = \pi/4$), for $0 \leq \theta \leq \pi/2$, $r^2/\sqrt{k} = 1, 4, 8, 32$, and $\boldsymbol{k = 5}$

Confidence Intervals – linear IV case

- Dufour (1997) impossibility result for Wald intervals
- Valid intervals come from inverting valid tests

(1) Inversion of AR test: AR Confidence Intervals

$$95\%\ \mathrm{CI} = \{\beta_0: \mathrm{AR}(\beta_0) < F_{k,T-k;.05}\}$$

- For $m = 1$, this entails solving a quadratic equation:

$$\mathrm{AR}(\beta_0) = \frac{(\mathbf{y}-\mathbf{Y}\beta_0)'P_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\beta_0)/k}{(\mathbf{y}-\mathbf{Y}\beta_0)'M_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\beta_0)/(T-k)} < F_{k,T-k;.05}$$

- For $m > 1$, solution can be done by grid search or using methods in Dufour and Taamouti (2005)
- Sets for a single coefficient can be computed by projecting the larger set onto the space of the single coefficient (see Dufour and Taamouti (2005)), also see recent work by Kleibergen (2008)
- Intervals can be empty, unbounded, disjoint, or convex

(2) Inversion of CLR test: CLR Confidence Intervals

$$95\%\ \text{CI} = \{\beta_0 : \text{LR}(\beta_0) < \text{cv}_{.05}(Q_T)\}$$

where $\text{cv}_{.05}(Q_T)$ = 5% conditional critical value

Comments:

- Efficient GAUSS and STATA (condivreg) software
- Will contain the LIML estimator (Mikusheva (2005))
- Has certain optimality properties: nearly uniformly most accurate invariant; also minimum expected length in polar coordinates (Mikusheva (2005))
- Only available for $m = 1$

What about the bootstrap or subsampling?

A straightforward bootstrap algorithm for TSLS:

$$y_t = \beta Y_t + u_t$$

$$Y_t = \Pi' Z_t + v_t$$

i) Estimate β, Π by $\hat{\beta}^{TSLS}$, $\hat{\Pi}$

ii) Compute the residuals $\hat{u}_t$, $\hat{v}_t$

iii) Draw T "errors" and exogenous variables from $\{\hat{u}_t, \hat{v}_t, Z_t\}$, and construct bootstrap data $\tilde{y}_t$, $\tilde{Y}_t$ using $\hat{\beta}^{TSLS}$, $\hat{\Pi}$

iv) Compute TSLS estimator (and t-statistic, etc.) using bootstrap data

v) Repeat, and compute bias-adjustments and quantiles from the boostrap distribution, e.g. bias = bootstrap mean of $\hat{\beta}^{TSLS} - \hat{\beta}^{TSLS}$ using actual data

- Under strong instruments, this algorithm works (provides second-order improvements).

Bootstrap, ctd.

- Under weak instruments, this algorithm (or variants) does not even provide first-order valid inference

  The reason the bootstrap fails here is that $\hat{\Pi}$ is used to compute the bootstrap distribution. The true pdf depends on μ^2, say $f_{TSLS}(\hat{\beta}^{TSLS};\mu^2)$ (e.g. Rothenberg (1984 exposition above, or weak instrument asymptotics). By using $\hat{\Pi}$, μ^2 is estimated, say by $\hat{\mu}^2$. The bootstrap correctly estimates $f_{TSLS}(\hat{\beta}^{TSLS};\hat{\mu}^2)$, but $f_{TSLS}(\hat{\beta}^{TSLS};\hat{\mu}^2)$ $\neq f_{TSLS}(\hat{\beta}^{TSLS};\mu^2)$ because $\hat{\mu}^2$ is not consistent for μ^2. '

- This story might sound familiar – it is the same reason the bootstrap fails in the unit root model, and in the local-to-unity model.
- Subsampling for these (non-pivotal) statistics doesn't work either; see Andrews and Guggenberger (2007a,b).

Some remarks about estimation

- It isn't possible to have an estimator that is completely robust to weak instruments
- TSLS (2-step GMM) is about the worst thing you can do – from a bias perspective
- LIML (in GMM, CUE) has much better median bias properties but can have large (infinite) variance
- In the linear case, there are alternative estimators, e.g. Fuller's estimator; see Hahn, Hausman, and Kuersteiner (*JAE*, 2006) for an extensive MC comparison

Example #1: Consumption CAPM and the EIS

Yogo (REStat, 2004)

Δc_{t+1} = consumption growth, t to t+1

$r_{i,t+1}$ = return on i^{th} asset, t to t+1

Moment conditions: $E_t(\Delta c_{t+1} - \tau_i - \psi r_{i,t+1}) = 0$

EIS estimating equations: $\Delta c_{t+1} = \tau_i + \psi r_{i,t+1} + u_{i,t+1}$ ("forwards")

or $r_{i,t+1} = \mu_i + (1/\psi)\Delta c_{t+1} + \eta_{i,t+1}$ ("backwards")

Under homoskedasticity, standard estimation is by TSLS or by the inverse of the TSLS estimator (remember Hahn-Hausman (2003) test?); but with weak instruments, the normalization matters

First stage F-statistics for EIS (Yogo (2004)):

TABLE 1.—TEST FOR WEAK INSTRUMENTS

				p-Value			
Country	Sample Period	Variable	F	TSLS Bias	TSLS Size	Fuller-k	LIML
USA	1947.3–1998.4	Δc	2.93	0.93	1.00	0.53	0.37
		r_f	15.53	0.00	0.66	0.00	0.00
		r_e	2.88	0.93	1.00	0.54	0.39
AUL	1970.3–1998.4	Δc	1.79	0.99	1.00	0.81	0.69
		r_f	21.81	0.00	0.14	0.00	0.00
		r_e	1.82	0.99	1.00	0.80	0.68
CAN	1970.3–1999.1	Δc	3.03	0.92	1.00	0.50	0.35
		r_f	15.37	0.00	0.67	0.00	0.00
		r_e	2.51	0.96	1.00	0.64	0.48
FR	1970.3–1998.3	Δc	0.17	1.00	1.00	1.00	1.00
		r_f	38.43	0.00	0.00	0.00	0.00
		r_e	3.09	0.91	1.00	0.49	0.34
GER	1979.1–1998.3	Δc	0.83	1.00	1.00	0.97	0.93
		r_f	17.66	0.00	0.45	0.00	0.00
		r_e	0.69	1.00	1.00	0.98	0.95
ITA	1971.4–1998.1	Δc	0.73	1.00	1.00	0.98	0.95

Various estimates of the EIS, forward and backward

TABLE 2.—ESTIMATES OF THE EIS USING THE INTEREST RATE

		$1/\psi$			ψ		
Country	Sample Period	TSLS	Fuller-k	LIML	TSLS	Fuller-k	LIML
USA	1947.3–1998.4	0.68	3.30	34.11	0.06	0.03	0.03
		(0.48)	(3.20)	(112.50)	(0.09)	(0.10)	(0.10)
AUL	1970.3–1998.4	0.50	2.37	30.03	0.05	0.04	0.03
		(0.48)	(2.45)	(107.71)	(0.11)	(0.12)	(0.12)
CAN	1970.3–1999.1	−1.04	−2.40	−2.98	−0.30	−0.33	−0.34
		(0.39)	(1.13)	(1.54)	(0.16)	(0.17)	(0.17)
FR	1970.3–1998.3	−3.12	−1.83	−12.38	−0.08	−0.08	−0.08
		(3.75)	(1.72)	(29.61)	(0.19)	(0.19)	(0.19)
GER	1979.1–1998.3	−1.05	−1.38	−2.29	−0.42	−0.43	−0.44
		(0.62)	(0.90)	(1.87)	(0.35)	(0.35)	(0.36)
ITA	1971.4–1998.1	−3.34	−5.82	−14.81	−0.07	−0.07	−0.07
		(1.98)	(4.47)	(18.55)	(0.08)	(0.08)	(0.08)
JAP	1970.3–1998.4	−0.18	−0.86	−21.56	−0.04	−0.04	−0.05
		(0.43)	(1.23)	(106.53)	(0.21)	(0.23)	(0.23)

AR, LM, and CLR confidence intervals for ψ:

TABLE 3.—WEAK-INSTRUMENT-ROBUST CONFIDENCE INTERVALS FOR THE EIS USING THE INTEREST RATE

Country	Sample Period	AR	LM	Cond. LR
USA	1947.3–1998.4	$\varnothing$	$[-0.21, 0.23]$	$[-0.19, 0.22]$
AUL	1970.3–1998.4	$[-0.16, 0.21]$	$[-0.22, 13.74]$	$[-0.22, 0.27]$
CAN	1970.3–1999.1	$[-0.54, -0.14]$	$[-0.73, 14.15]$	$[-0.71, 0.00]$
FR	1970.3–1998.3	$[-0.68, 0.53]$	$[-0.47, 0.31]$	$[-0.48, 0.33]$
GER	1979.1–1998.3	$[-1.57, 0.54]$	$[-1.21, 0.26]$	$[-1.23, 0.28]$
ITA	1971.4–1998.1	$[-0.29, 0.18]$	$[-0.24, 0.11]$	$[-0.24, 0.12]$
JAP	1970.3–1998.4	$[-0.60, 0.49]$	$[-\infty, \infty]$	$[-0.56, 0.45]$
NTH	1977.3–1998.4	$[-0.91, 0.64]$	$[-\infty, \infty]$	$[-0.76, 0.48]$
SWD	1970.3–1999.2	$[-0.30, 0.29]$	$[-\infty, \infty]$	$[-0.22, 0.21]$
SWT	1976.2–1998.4	$[-1.69, 0.37]$	$[-1.19, 0.07]$	$[-1.22, 0.09]$
UK	1970.3–1999.1	$[0.04, 0.28]$	$[-\infty, \infty]$	$[-0.12, 0.43]$
USA	1970.3–1998.4	$\varnothing$	$[-\infty, \infty]$	$[-0.23, 0.23]$
SWD	1921–1994	$[-0.30, 0.40]$	$[-\infty, \infty]$	$[-0.25, 0.35]$
UK	1921–1994	$[-0.05, 0.88]$	$[0.01, 0.70]$	$[0.01, 0.70]$
USA	1891–1995	$[-0.49, 0.46]$	$[-\infty, \infty]$	$[-\infty, \infty]$

The table reports 95% confidence intervals for the EIS, constructed from AR, LM, and conditional LR tests. $\varnothing$ indicates an empty confidence interval. The instruments are twice lagged nominal interest rate, inflation, consumption growth, and log dividend-price ratio.

What about stock returns – should they "work"?

TABLE 5.—WEAK-INSTRUMENT-ROBUST CONFIDENCE INTERVALS FOR THE EIS USING THE STOCK RETURN

Country	Sample Period	AR	LM	Cond. LR
USA	1947.3–1998.4	$[-0.21, -0.02]$	$[-\infty, \infty]$	$[-\infty, \infty]$
AUL	1970.3–1998.4	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
CAN	1970.3–1999.1	$[0.02, 4.03]$	$[0.05, 0.35]$	$[0.04, 0.41]$
FR	1970.3–1998.3	$[-0.28, 0.20]$	$[-\infty, \infty]$	$[-0.16, 0.11]$
GER	1979.1–1998.3	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
ITA	1971.4–1998.1	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
JAP	1970.3–1998.4	$[-0.05, 0.32]$	$[-1.01, 0.20]$	$[-0.02, 0.21]$
NTH	1977.3–1998.4	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
SWD	1970.3–1999.2	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
SWT	1976.2–1998.4	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
UK	1970.3–1999.1	$[-0.51, -0.02]$	$[-\infty, \infty]$	$[-\infty, \infty]$
USA	1970.3–1998.4	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
SWD	1921–1994	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$
UK	1921–1994	$[-0.04, 0.10]$	$[-\infty, \infty]$	$[-0.10, 0.14]$
USA	1891–1995	$[-\infty, \infty]$	$[-\infty, \infty]$	$[-\infty, \infty]$

See notes to table 3.

Extensions to >1 included endogenous regressor

- CLR exists in theory, but difficult computational issues because the conditioning statistic has dimension $m(m+1)/2$ (AMS (2006), Kleibergen (2007))
- Can test joint hypothesis H_0: $\beta = \beta_0$ using the AR statistic:

$$\mathrm{AR}(\beta_0) = \frac{(\mathbf{y} - \mathbf{Y}\beta_0)' P_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / k}{(\mathbf{y} - \mathbf{Y}\beta_0)' M_{\mathbf{Z}} (\mathbf{y} - \mathbf{Y}\beta_0) / (T - k)}$$

  under H_0, AR $\xrightarrow{d}$ χ_k^2/k

- Subsets by projection (Dufour-Taamouti (2005)) or by concentration + bounds (Kleibergen and Mavroeidis (2008, 2009) – 2009 is GMM treatment)

Extensions to GMM

(1) The GMM-Anderson Rubin statistic

(Kocherlakota (1990); Burnside (1994), Stock and Wright (2000)) The extension of the AR statistic to GMM is the CUE objective function evaluated at θ_0:

$$S_T^{CUE}(\theta_0) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]'\hat{\Omega}(\theta_0)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]$$

$$\xrightarrow{d} \psi(\theta_0)'\Omega(\theta_0)^{-1}\Psi(\theta_0) \sim \chi_k^2$$

- Thus a valid test of H_0: $\theta = \theta_0$ can be undertaken by rejecting if $S_T(\theta_0)$ > 5% critical value of χ_k^2.

<u>GMM-Anderson-Rubin, ctd.</u>

In the homoskedastic/uncorrelated linear IV model, the GMM-AR statistic simplifies to the AR statistic (up to a degrees of freedom correction):

$$S_T^{CUE}(\theta_0) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]'\hat{\Omega}(\theta_0)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_0)\right]$$

$$= \left[T^{-1/2}\sum_{t=1}^{T}(y_t - \theta_0'Y_t)Z_t\right]'\left(\frac{\mathbf{Z}'\mathbf{Z}}{T}s_v^2\right)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}(y_t - \theta_0'Y_t)Z_t\right]$$

$$= \frac{(\mathbf{y}-\mathbf{Y}\theta_0)'P_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\theta_0)}{(\mathbf{y}-\mathbf{Y}\theta_0)'M_{\mathbf{Z}}(\mathbf{y}-\mathbf{Y}\theta_0)/(T-k)} = k \times \text{AR}(\theta_0)$$

- The GMM-AR statistic has the same issues of interpretation issues as the AR, specifically, the GMM-AR rejects because of endogenous instruments and/or incorrect θ
- The GMM-AR can fail to reject any values of θ (remember the Dufour (1997) critique of Wald tests)

(2) GMM-LM

Kleibergen (2005) – develops score statistic (based on CUE objective function – details of construction matter) that provides weak-identification valid hypothesis testing for sets of variables

(3) GMM-CLR

Andrews, Moreira, Stock (2006) – extension of CLR to linear GMM with a single included endogenous regressor, also see Kleibergen (2007). Very limited evidence on performance exists; also problem of dimension of conditioning vector

(4) Other methods

Guggenberger-Smith (2005) objective-function based tests based on Generalized Empirical Likelihood (GEL) objective function (Newey and Smith (2004)); Guggenberger-Smith (2008) generalize these to time series data. Performance is similar to CUE (asymptotically equivalent under weak instruments)

Confidence sets

- Fully-robust 95% confidence sets are obtained by inverting (are the acceptance region of) fully-robust 5% hypothesis tests
- Computation is by grid search in general: collect all the points θ which, when treated as the null, are not rejected by the GMM-AR statistic.
- Subsets by projection or by concentration + bounds (see Kleibergen and Mavroeidis (2008, 2009) for an application of GMM-AR confidence sets and subsets)
- Valid tests must be unbounded (contain Θ) with finite probability with weak instruments

Many instruments: a solution to weak instruments?

The appeal of using many instruments

- Under standard IV asymptotics, more instruments means greater efficiency.
- This story is not very credible because
  (a) the instruments you are adding might well be weak (you already have used the first two lags, say) and
  (b) even if they are strong, this requires consistent estimation of increasingly many parameter to obtain the efficient projection – hence slow rates of growth of the number of instruments in efficient GMM literature.

Example of problems with many weak instruments – TSLS

Recall the TSLS weak instrument asymptotic limit:

$$\hat{\beta}^{TSLS} - \beta_0 \xrightarrow{d} \frac{(\lambda + z_v)'z_u}{(\lambda + z_v)'(\lambda + z_v)}$$

with the decomposition, $z_u = \delta z_v + \eta$. Suppose that k is large, and that $\lambda'\lambda/k \to \Lambda_\infty$ (one way to implement "many weak instrument asymptotics"). Then as $k \to \infty$,

$\lambda' z_v/k \xrightarrow{p} 0$ and $\lambda' z_u/k \xrightarrow{p} 0$

$z_v' z_v/k \xrightarrow{p} 1$ and $z_v'\eta/k \xrightarrow{p} 0$ (z_v and η are independent by construction)

Putting these limits together, we have, as $k \to \infty$,

$$\frac{(\lambda + z_v)'z_u}{(\lambda + z_v)'(\lambda + z_v)} \xrightarrow{p} \frac{\delta}{1 + \Lambda_\infty}$$

In the limit that $\Lambda_\infty = 0$, TSLS is consistent for the *plim* of OLS!

Comments

- Strictly this calculation isn't right – it uses sequential asymptotics ($T \to \infty$, then $k \to \infty$). However the sequential asymptotics is justified under certain (restrictive) conditions on K/T (specifically, $k^4/T \to 0$)
- Typical conditions on k are $k^3/T \to 0$ (e.g. Newey and Windmeijer (2004))
- Many instruments can be turned into a blessing (if they are not too weak! They can't push the scaled concentration parameter to zero) by exploiting the additional convergence across instruments. This can lead to bias corrections and corrected standard errors. There is no single best method at this point but there is promising research, e.g. Newey and Windmeijer (2004), Chao and Swanson (2005), and Hansen, Hausman, and Newey (2006))

5) Current Research Issues & Literature

- Detection of weak instruments in general nonlinear GMM model
- Efficient testing in GMM with weak instruments

  Andrews, Moreira, Stock (2006), Kleibergen (2008)
- Subset testing

  Kleibergen and Mavroides (2008, 2009)
- Improved estimation in GMM

  Guggenberger-Smith (2005) (GEL)
- Many instruments
- Breaks in GMM with weak instruments

  Caner (2008)
- Connection between weak ID, HAC estimation, and SVAR identification using long run restrictions

  Gospodinov (2008)
- Weak set identification?