A uniformly valid test for instrument exogeneity

Prosper Dovonon
Concordia University

Nikolay Gospodinov
Atlanta FED

Indiana University, Nov. 19, 2024

Setup

- We consider the linear regression model with possibly endogenous regressors:

$$y_i = \eta_0 + x_i'\theta_0 + \varepsilon_i,$$

  where $(\eta_0, \theta_0) \in \mathbb{R} \times \mathbb{R}^p$.

- Let $z_i \in \mathbb{R}^m$ be the vector of instruments.
- The goal is to consistently test for:

$$H_0 : E(\varepsilon_i | z_i) = 0, \ \ a.s.$$

  without any knowledge of the instruments' strength.

- The sample $\{(x_i, z_i, y_i) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R} : i = 1, \ldots, n\}$ is i.i.d.
- We do not require that $m \geq p$ but we need the support of z_i to be rich enough.

Instrument's strength

We assume that:

- For some $\delta \geq 0$,

$$E(x_i|z_i) = \mu_x + \frac{C(z_i)}{n^\delta},$$

- and

$$\text{Rk}\left[E(C(z_i)C(z_i)')\right] = p.$$

  - $\delta = 0$: strong instruments.
  - $0 < \delta < 1/2$: semi-weak instruments (Hahn and Kuersteiner, 2002; Antoine and Renault, 2009).
  - $\delta \geq 1/2$: weak instruments (Staiger and Stock, 1997,...).
  - $\delta = \infty$: completely irrelevant instruments.

Instrument's strength/exogeneity: The pitfall

In standard GMM infer. with fixed # of uncond'nal moment restr.:

- If the instruments are exogenous: $E(\varepsilon_i|z_i) = 0, a.s.$,
  - When $\delta = 0$: consistent estimation and asymptotic normality.
  - When $0 < \delta < 1/2$: (slow-rate) consistent estimation and asymptotic normality.
  - When $\delta \geq 1/2$: consistent/convergent estimation is not possible. $\delta = \infty$: no information about the true value.
- If the instruments are not exogenous: $P(E(\varepsilon_i|z_i) = 0) < 1$:
  - If $\delta = 0$: Convergence to a pseudo true value.
  - If $\delta > 0$: the estimator diverges to ∞ regardless of true value.
  - ▶ This may result to a misleading inference; x falsely significant!
- Before parameter inference, need for exogeneity test that is valid irrespective of strength.

Summary of Main Results

- We propose a uniformly valid exogeneity test for instruments on linear IV models.
- This test is a normalized J-test based on an expanding number k_n of moment restrictions.
- We analyze the limiting behavior of the GMM estimator depending on the instruments strength leading to the choice of k_n.
- We show that, regardless of the instruments' strength:
    - The test is asymptotic standard normal under the null of exogeneity
    - The test is consistent under the alternative.
- We revisit some empirical findings on impact of trade on growth.

Outline

- Review of standard J-test for specification.
- Literature review.
- Our testing framework.
- Main results.
- Empirical application.

Is the J-test robust to instrument's strength?

Consider the Hansen-Sargan J-test for correct specification of uncond. moment models:

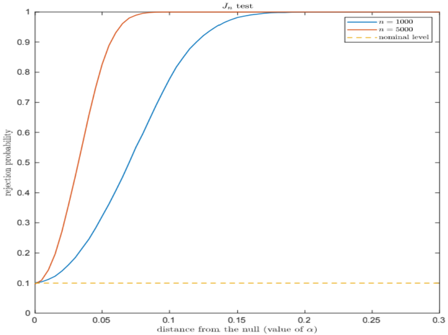
- Fixed number, k, of moment restrictions: $h(z_i) \in \mathbb{R}^k$, $k > p$.
- Apply J-test to: $E(h(z_i)\varepsilon_i) = 0$.
- Outcome when z_i is strong, weak, etc.
- We do this by simulations:

$$y_i = \theta_0 x_i + \alpha_0 f(z_{1i}) + \varepsilon_i, \quad x_i = \rho_{xz} z_i + v_i,$$

  - $z_i = (z_{1i}, z_{2i})' \sim \text{NID}(0, I_2)$, $\theta_0 = 1$,
  - $\rho_{xz} = (0.3\ 0.3)$, for strong instr; $\rho_{xz} = 0$, for compl. irrel. instr.
  - $(\varepsilon_i, v_i)' \sim N(0, (1, 0.5, 1))$.

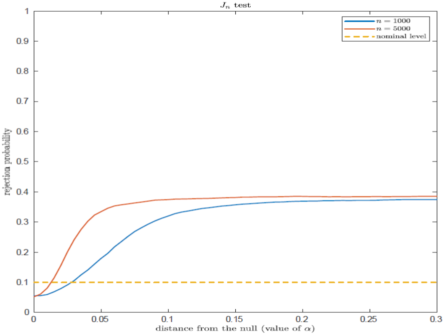
Is the J-test robust to instrument's strength?

Power curve of the J-test when instr. is strong - $f(z_{1i})$ linear



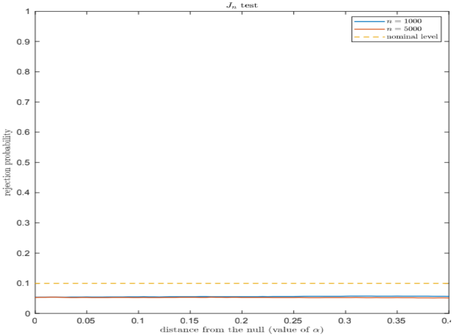
Is the J-test robust to instrument's strength?

Power curve of the J-test when instr. is irrelevant - $f(z_{1i})$ linear



Is the J-test robust to instrument's strength?

Power curve of the J-test when instr. is irrelevant - $f(z_{1i})$ nonlinear



Is the J-test robust to instrument's strength?

- By testing a fixed number of moment restrictions, J-test does not integrate enough information to generate power.
- In a way, it does not fully test for the conditional restriction.
- We will propose a test for $E(\varepsilon_i|z_i) = 0$, $a.s.$ by allowing $g(z)$ to have an increasing size.
- This is essential for uniform validity of the resulting test.

Literature on conditional moment models

- Estimation
  - Carrasco & Florens (2000): Objective function is the norm of moment function on a Hilbert space.
  - Dominguez & Lobato (2004) use the expected square of the ‘integrated regression function’.
  - Chunrong & Chen (2003) propose the seive minimum distance estimator.
  - Lavergne & Patilea (2013): Inference is based on the kernel estimate of the conditional moment.

Literature on conditional moment models

- Specification test
  - Bierens (1982, 84, 87, 90), Bierens & Ploberger (1987), Carrasco & Florens (2000), Dominguez & Lobato (2010): propose CMR tests that are consistent by using an increasing number of instruments.
  - Tripathi & Kitamura (2003) introduce the "smoothed" empirical likelihood based test that Smith (2007) extends to the family of Cressie-Read divergence functions.
  - Delgado, Dominguez & Lavergne (2006): propose a kernel estimation based test.
  - de Jong & Bierens (1994): propose a chi square-type test that converges to a standard normal distribution: Easy to implement and has non trivial power.
  - Dovonon & Gospodinov (2024): Extend dJB-type of tests to dependent data and locally under-identified models.

Exogeneity test for linear IV models with possibly weak instruments

- Parameter inference in linear IV models robust to weak identification: Antoine & Lavergne (2023).
- Specification test:

  Jun & Pinkse (2009,12): propose a non parametric test for instr. exogeneity,

  Doko Tchatoka & Dufour (2023): study the DWH-type of exogeneity tests under weak instruments,

  Dovonon & Gospodinov (2023): propose a dJB-type of cond'al mom. test that is consist. when instr. are compl. irrel.
- This paper is an extension of D&G (2003) to instruments of arbitrary strength!

The testing framework

- We aim to test:

$$H_0 : \quad E(\varepsilon_i | z_i) = 0, \;\; a.s.$$

- We follow D&G (2023,24) and de Jong & Bierens (1994) by testing the equivalent seq. of uncond'al moment restrict's:

$$E\left(g^{(k)}(z_i)\varepsilon_i\right) = 0, \quad k = 1, 2, \ldots,$$

- $g^{(k)}(z) := (g_1(z), \ldots, g_k(z))$ is an $\mathbb{R}^k$-valued function of the enumeration of the:
- basis functions, $(g_l(z))_{l\in\mathbb{N}}$, of $L^2(P_z) := L^2(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), P_z)$.
- Example: $g_l(z) = \cos(l\Psi(z)) + \sin(l\Psi(z)), \;\; l = 1, \ldots.$

The GMM estimator

- Moment condition: with $Z_i := g^{(k)}(z_i)$,

$$E\left((Z_i - \bar{Z})[y_i - \bar{y} - (x_i - \bar{x})'\theta_0]\right) = 0, \;\; k = 1, \ldots$$

- The GMM estimator with weighting matrix $\hat{W}$:

$$\tilde{\theta} = (\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zx})^{-1}\tilde{\mu}'_{zx}\hat{W}\tilde{\mu}_{zy}$$

with $\tilde{\mu}_{ab} := n^{-1}\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})'$

  - The first-step GMM or standard IV estimator use, e.g.,

$$\hat{W} = I_k, \quad \text{or} \quad \hat{W} = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z})(Z_i - \bar{Z})'.$$

  - The 2SGMM uses:

$$\hat{W} = \hat{V}(\tilde{\theta}) := \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i(\tilde{\theta})^2(Z_i - \bar{Z})(Z_i - \bar{Z})',$$

with $\hat{\varepsilon}_i(\theta) := y_i - \bar{y} - (x_i - \bar{x})'\theta$.

The test statistics

- The Hansen-Sargan J-test statistic:

$$J_n := n\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right)' \hat{V}^{-1}\left(\tilde{\mu}_{zy} - \tilde{\mu}_{zx}\hat{\theta}\right),$$

  - where $\hat{\theta}$ is the 2SGMM and $\hat{V} := \hat{V}(\hat{\theta})$.
- Our test statistic:

$$S_{n,k} = \frac{J_n - k}{\sqrt{2k}}.$$

  - The J-test stat. introduced is slightly different than the standard.
  - The change is essential to obtain exact asymptotic normality for $S_{n,k}$.

Key assumptions

Let $a_k := E[(Z_i - \mu_z)C(z_i)'],\ V_z := Var(Z_i)$

Assumption

$\|a_k\|_2 = O(1)$, *and* $Rank(C(z_i)C(z_i)') = p$.

Assumption

(a) x_i, ε_i, *and* Z_i *have up to 8th moment,* *(b)* $\lambda(V_z) \in (\underline{\lambda}, \bar{\lambda})$,
(c) $\lambda(E[w_i^2(Z_i - \mu_z)(Z_i - \mu_z)']) \leq \bar{\lambda}$: $w_i \in \{\varepsilon_i, x_{hi} : h = 1, \dots, p\}$.

Assumption

Assume that there exists W a nonrandom (p, p)-matrix symmetric positive definite such that:
$\|\hat{W} - W\|_2 = o_P(k^{-1/2})$ *and* $\lambda_{\max}(W) \leq \bar{\lambda}$.

Behavior of the GMM estimator

Theorem

Under H_0 and the assumptions above, if $k \to \infty$ with $k = o(n)$, then:

(a) *If* $\delta \geq 1/2$, *then* $\tilde{\theta} = \theta_0 + \mathbb{V}_{1k}^{-1}\mathbb{C}_{1k} + O_P(k^{-1/2})$.

(b) *If* $0 \leq \delta < 1/2$, *then:*

- *If* $k \ll n^{1/2-\delta} \vee k \sim n^{1/2-\delta}$, $\tilde{\theta} = \theta_0 + O_P(n^{-1/2+\delta})$.
- *If* $n^{1/2-\delta} \ll k \ll n^{1-2\delta}$, $\tilde{\theta}$ *converges but at a slower rate.*
- *If* $k \sim n^{1-2\delta} \vee k \gg n^{1-2\delta}$, $\tilde{\theta}$ *is inconsistent.*

A surprising implications

- When z_i is completely irrelevant, the GMM estimator converges at the rate $\sqrt{k}$ to a "pseudo-true" value.
- Under the assumption that:

$$E(x_i x_i' | z_i) = V_x, \quad \text{and} \quad E(x_i \varepsilon_i | x_i) = c_x,$$

the theorem shows that

$$\tilde{\theta} \xrightarrow{P} \theta_0 + V_x^{-1} c_x.$$

- Same limit as OLS.
- Simulating arbitrary instruments and doing IV converges to the same limit as OLS.
- $0 \leq \delta < 1/2$ corresponds to a 'phase transition' where where estimation is consistent for the right k.

Robust choice of $k = k_n$

- For $\delta \in [0, 1/2)$, when k grows slowly, this result is consistent with Antoine and Renault (2012) who consider k fixed.
- For $\delta \geq 1/2$, it suffices that $k = o(n)$ for the stated convergence to hold.
- For robustness, k_n must grow slower than n^α for any $\alpha \in (0, 1/2]$.
- A right choice of k is:

$$k := k_n = a(\log n)^b, \qquad \text{for some } a, b > 0.$$

Asymptotic behavior under the null

- For the first step GMM, define:

$$q_{1i} = (Z_i - \mu_z)' W (Z_i - \mu_z), \qquad v_i := x_i - E(x_i | z_i),$$

$$r_{1i} = \varepsilon_i - v_i' \left(E(q_{1i} \cdot v_i v_i') \right)^{-1} \left(E(q_{1i} \cdot \varepsilon_i \cdot v_i) \right),$$

$$V_{1,\delta} = \begin{cases} E(\varepsilon_i^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad 0 \leq \delta < 1/2, \\ E(r_{1i}^2 (Z_i - \mu_z)(Z_i - \mu_z)') & \text{if} \quad \delta \geq 1/2. \end{cases}$$

- For the second step GMM, define:

q_{2i} as q_{1i} but with $V_{1,\delta}$ replacing W;

r_{2i} as r_{1i} but with q_{2i}; and

$V_{2,\delta}$ as $V_{1,\delta}$ but with r_{2i}.

Behavior of the weighting matrix and 2SGMM

Assumption

Assume that $E(r_{1i}v_i|z_i) = 0$ *and* $\lambda(V_{1,\delta}) \geq \underline{\lambda}$.

Theorem

If the assumptions above hold and $k \to \infty$ *with* $k \sim a(\log n)^b$, *for some* $a, b > 0$. *Then, under* H_0, *we have:*

(a) If $0 \leq \delta < 1/2$, $\tilde{V} - V_{1,\delta} = O_P(n^{-1/2+\delta})$.
If $\delta \geq 1/2$, $\tilde{V} - V_{1,\delta} = O_P(k^{-1})$.

(b) If $0 \leq \delta < 1/2$, $\hat{\theta} = \theta_0 + O_P(n^{-1/2+\delta})$.
If $\delta \geq 1/2$, $\hat{\theta} = \theta_0 + \mathbb{V}_{2k}^{-1}\mathbb{C}_{2k} + O_P(k^{-1/2})$.

The test statistic under H_0

Assumption

Assume that $E(r_{2i}v_i|z_i) = 0$ *and* $\lambda(V_{2,\delta}) \geq \underline{\lambda}$.

Theorem

Suppose the assumptions above hold, and $k \to \infty$ *with* $k \sim a(\log n)^b$*, for some* $a, b > 0$*. Then, under* H_0*, for any value of* $\delta \in [0, +\infty]$*, we have:*

$$S_{n,k} \xrightarrow{d} N(0,1).$$

Sketch of proof (case $\delta \geq 1/2$)

We show that:

- $\hat{V}$ is equal to:

$$V_{2,\delta} + o_P(\bullet) := E\left(r_{2i}^2 (Z_i - \mu_z)(Z_i - \mu_z)'\right) + o_P(\bullet)$$

- $J_n - k = A_{1n} + A_{2n} + o_P(\bullet)$, with

$$A_{1n} = \frac{1}{n} \sum_{i \neq j} r_{2i} r_{2j} (Z_i - \mu_z)' V_{2,\delta}^{-1} (Z_j - \mu_z),$$

$$A_{2n} = \frac{1}{n} \sum_{i=1}^{n} r_{2i}^2 (Z_i - \mu_z)' V_{2,\delta}^{-1} (Z_i - \mu_z) - k.$$

- We show that $A_{2n} = o_P(\sqrt{k})$ and the CLT for degener. U-stat of Hall (1984) allows to claim that $A_{1n}/\sqrt{2k} \xrightarrow{d} N(0,1)$.

Discussion

- The test is easy to implement: quantiles from $N(0,1)$.
- This asymptotic distribution is valid under H_0 regardless of the strength of z_i.
- Even if the system is not overidentified, i.e., $|z_i| \leq |x_i|$, the test can be performed.
- z_i needs a support that is rich enough (ex: continuous).
- In comparison to Jun & Pinkse (2009,12), our result is sharp: They are conserv. under weak instr.

Asymptotic behavior under the alternative

- The alternative hypothesis amounts to:

$$P\left(E(\varepsilon_i(\eta,\theta)|z_i)=0\right)<1, \quad \forall(\eta,\theta)\in\mathbb{R}^{p+1}$$

- We can show that, for any compact $\mathcal{C}\subset\mathbb{R}^{p+1}$,

$$\exists k_0\in\mathbb{N} \text{ and } \delta_0>0: \qquad \inf_{(\eta,\theta)\in\mathcal{C}}\|E[g^{(k)}(z_i)\varepsilon_i(\eta,\theta)]\|_2>\delta_0.$$

- Thus $c_z := E[g^{(k)}(z_i)\varepsilon_i(\eta_0,\theta_0)] := E[Z_i\varepsilon_i(\eta_0,\theta_0)]\neq 0$, for k large enough.
- Internal consistencies require that $\|c_z\|_2=O(1)$.
- We maintain this condition.

2SGMM under the alternative

Theorem

If some regularity conditions hold and $k \sim a(\log n)^b$, for some $a, b > 0$. Then, under H_1, we have:

(a) *For* $0 < \delta < 1/2$,

$$\hat{\theta} - \theta_0 = O_P(n^\delta).$$

(b) *For* $\delta = 0$,

$$\hat{\theta} - \theta_0 = A_k c_z + O_P\left(\frac{1}{\sqrt{k}}\right).$$

(c) *For* $\delta \geq 1/2$,

$$\hat{\theta} - \theta_0 = O_P\left(\frac{\sqrt{n}}{k}\right).$$

Comments

- The estimator $\hat{\theta}$ explodes when the instruments are not strong.
- This may have adverse consequences on t-statistics if the variance does not explode fast enough.
- Standard inference may be misleading:
  - falsely claim that x is significant,
  - and this, only because z_i is (unknowingly) endogenous.

J_n and $S_{n,k}$ under the alternative

Let $P_\delta = \Sigma^{-1/2} a_k (a_k' \Sigma^{-1} a_k)^{-1} a_k' \Sigma^{-1/2}$, and
$\Delta_k = c_z' \Sigma^{-1/2} (I_k - P_\delta) \Sigma^{-1/2} c_z$.

Theorem

Under the same conditions as above, we have:

(a) *For* $0 < \delta < 1/2$: $\exists C > 0$ *such that,*

$$J_n \geq n^{1-2\delta} \cdot C \cdot (\Delta_k + o_P(1)) \quad w.p.a.1.$$

(b) *For* $\delta = 0$: $J_n \geq n \cdot C \cdot \Delta_k + O_P(n/\sqrt{k}), \quad w.p.a.1.$

$$S_{n,k} \geq \frac{n^{1-2\delta}}{\sqrt{k}} \cdot (C \cdot \Delta_k + o_P(1)), \quad w.p.a.1$$

so that $J_n, S_{n,k} \to \infty$, *in probability as* $n \to \infty$.

J_n and $S_{n,k}$ under the alternative (2)

Theorem

(c) *For* $\delta \geq 1/2$*: There exists a random sequence* $\pi_n \geq 0$ *such that* $\lim_{\epsilon \downarrow 0} P(\pi_n \leq \epsilon) = 0$ *and, with probability approaching 1,*

$$J_n \geq k^2 \cdot \pi_n \cdot \|c_z\|_2^2 + o_P(1),$$

$$S_{n,k} \geq 2^{-3/2} \cdot k^{3/2} \cdot \pi_n \cdot \|c_z\|_2^2 + o_P(1)$$

so that both $J_n, S_{n,k} \to \infty$*, in probability as* $n \to \infty$*.*

Comments on J_n and $S_{n,k}$ under H_1

- J_n and $S_{n,k}$ explode to $+\infty$ under H_1
  - For $0 \leq \delta < 1/2$, the test statistics explode with n as in standard settings.
  - For $\delta \geq 1/2$, only k is responsible for power.
  - In this case, when k fixed, power is not guaranteed.

Intuition for power when $\delta \geq 1/2$

Consider the simple case: $\mu_z = 0$, $|x_i| = 1$, $\mu_x = 0$ and $\mu_y = 0$.

We show that:

- $\hat{\theta} - \theta_0 := \hat{e} \sim \frac{\sqrt{n}}{k} h_n; \quad h_n = O_P(1)$.
- The signal: $\sqrt{n}(\bar{\mu}_{zy} - \hat{\theta}\bar{\mu}_{zy}) \sim \sqrt{n} c_z$.
- $\hat{V} \sim \hat{e}^2 E(x_i^2 Z_i Z_i')$ explodes but slower than the signal.
- As a result,

$$J_n \sim k^2 \cdot h_n^2 \cdot c_z' \left(E[x_i^2 Z_i Z_i']\right)^{-1} c_z.$$

  - Due to expanding k, J_n explodes at rate k^2 and $S_{n,k}$ explodes at rate $k^{3/2}$.

Simulation design

$$y_i = \theta_0 x_i + \alpha_0 z_{1i} + \varepsilon_i, \quad x_i = \Pi(\delta)' Z_i + v_i,$$

$$\theta_0 = 1, \quad \Pi(\delta) = n^{-\delta}\iota,$$

$$Z_i = (z_{1i}, z_{1i}, z_{1i})' \sim \mathrm{NID}(0, I_3),$$

$$(\varepsilon_i, v_i)' \sim N(0, (1, 0.3, 1)).$$

$$g_l(z) = \cos(l\Psi(z)) + \sin(l\Psi(z)), \;\; l = 1, \ldots, \lceil \log n \rceil,$$

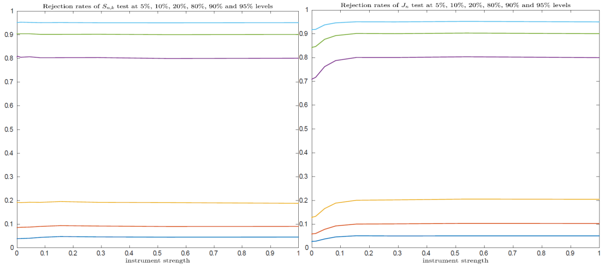
$$\Psi(z) = 2\arctan(z), \quad \alpha_0 \in [0, 0.4], \quad \delta \in \{0, 0.2, 0.5, 1, 100\}.$$

$$n = 500, \;\; \text{nReplications} \; = 100,000.$$

- J-test with 3 instr. $\chi^2(2)$ under the null.

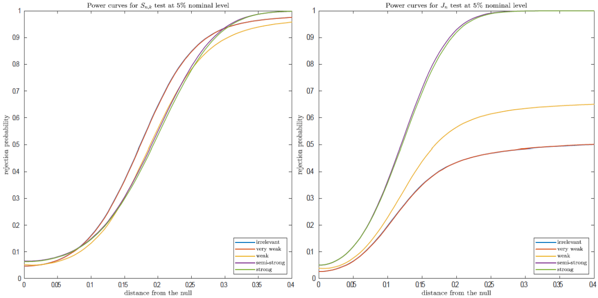
Rejection rates under the null ($\alpha_0 = 0$)

$n = 500$



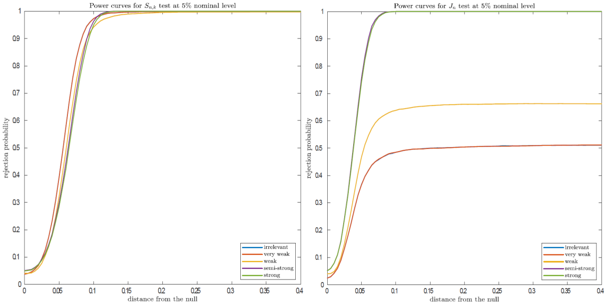
Power curves ($\alpha \in [0; 0.4]$)

$n = 500$



Power curves ($\alpha \in [0; 0.4]$)

$n = 5,000$



Experiment 2

2 endogenous variables, 4 instruments with heterogeneous strength. $n = 500$.

$(\delta_1, \delta_2, \delta_3, \delta_4)$	Panel A: one-sided $S_{n,k}$ test					
	size			power		
	1%	5%	10%	1%	5%	10%
$(0, 0.5, 0.2, 100)$	1.6	5.0	8.6	95.4	97.7	98.5
$(100, 0.3, 0.1, 100)$	1.6	4.7	8.3	89.4	93.2	94.9
$(0, 0.2, 0.5, 0)$	1.5	4.7	8.3	100	100	100
$(0.8, 0.2, 0.5, 0.4)$	1.4	4.2	7.4	93.8	96.1	97.1
$(0.5, 0.4, 0.3, 0.1)$	1.2	3.9	6.8	96.1	97.5	98.2
$(0, 100, 100, 0)$	1.5	4.8	8.3	100	100	100
$(0.1, 0.2, 0.5, 0.5)$	1.2	3.9	6.8	86.9	91.2	93.1
$(0.6, 0.5, 0.2, 1)$	1.2	3.9	7.0	89.8	93.2	94.7

Impact of trade on economic growth

- We revisit some studies of impact of trade on growth, Hausman et al. (2007).
- Country regression of Income per capita on proxies of trade share (ratio of Export or import to GDP): EXPY_i.
- The dependent variable is: Average annual growth in GDP per capita.
- EXPY_i is endogenous: Instruments proposed: country size (population, land area, human capital).
- Are these instruments exogenous?
- We examine this with our test statistic and the standard J-test.
- Dataset: 79 countries, yearly data: 1962-2000.

Result

	Panel A: ten-year sample ($n = 299$)					
	(1u)	(1c)	(2u)	(2c)	(3u)	(3c)
log EXPY	0.092	0.092	0.132	0.074	0.251	0.080
log initial GDP/capita	-0.038	-0.038	-0.054	-0.028	-0.105	-0.031
log human capital	0.004	0.004				
log area			-0.003	-0.002		
log population					-0.009	-0.000
J_n test	11.25		0.369		0.453	
(p-value)	(0.001)		(0.544)		(0.501)	
$S_{n,k}$ test		2.628		3.491		3.255
(p-value)		(0.004)		(0.000)		(0.001)

Conclusion

- We propose a test for instr. exogeneity in lin. regr. models.
- In contrast to existing tests, our test is consistent regardless of instr.'s strength, easy to compute, asymp. normal.
- The test is based on expanding moment condition through basis functions and require that the support of instr. is rich.
- Our test also applies to a variety of configurations including: just-identified and under-identified models.
- Interesting extensions: Non-linear models and models of time dependent data.