A more powerful subvector Anderson and Rubin test in linear instrumental variables regression

Frank Kleibergen
University of Amsterdam

Joint work with Patrik Guggenberger (Pennsylvania State University) and Sophocles Mavroeidis (University of Oxford)

North American Winter Meeting

January 2018

Overview

- Consider **subvector inference in the linear IV model**, allowing for **weak instruments** but assuming **conditional homoskedasticity**

- **Background:**
  - Projection of Anderson and Rubin (AR) test (Dufour and Taamouti, Ecta 2005).

  - Guggenberger, Kleibergen, Mavroeidis, and Chen (Ecta 2012, GKMC) provide power improvement:

  - Using $\chi^2_{k-m_W,1-\alpha}$ as critical value, rather than $\chi^2_{k,1-\alpha}$ still controls asymptotic size.

  - "Worst case" occurs under strong identification.

- **HERE:** consider a **data-dependent critical value** that adapts to strength of identification.

- One main objective: computational ease.

- Show: conditional subvector AR test controls finite sample/asymptotic size & has higher power than method in GKMC.

- Test in GKMC is inadmissible.

- Proposed test has a near optimality property when $m_W = 1$.

Outline

1. Finite sample analysis

   (a) Motivation for conditional subvector AR test

   (b) Size of test when $m_W = 1$

   (c) Power analysis when $m_W = 1$

   (d) Size of test when $m_W > 1$

2. Asymptotics

Model and Objective (finite sample case)

$$
\begin{aligned}
y &= Y\beta + W\gamma + \varepsilon, \\
Y &= Z\Pi_Y + V_Y, \\
W &= Z\Pi_W + V_W,
\end{aligned}
$$

$$y \in \Re^n, Y \in \Re^{n\times m_Y}, W \in \Re^{n\times m_W}, \text{ and } Z \in \Re^{n\times k}.$$

- **Reduced form:**

$$(y \,\vdots\, Y \,\vdots\, W) = Z\,(\Pi_Y \,\vdots\, \Pi_W)\begin{pmatrix} \beta & I_{m_Y} & 0 \\ \gamma & 0 & I_{m_W} \end{pmatrix} + \underbrace{(v_y \,\vdots\, V_Y \,\vdots\, V_W)}_{V}.$$

- **Objective:** test

$$H_0 : \beta = \beta_0 \;\text{ versus }\; H_1 : \beta \neq \beta_0$$

s.t. size bounded by nominal size & "good" power.

Parameter space:

1. The error term is distributed as

$$V_i \sim \text{i.i.d. } N\left(0, \Omega\right), \; i = 1, ..., n,$$

   where $\Omega \in R^{(m+1)\times(m+1)}$ is assumed to be known and positive definite.

2. $Z \in R^{n\times k}$ fixed, and $Z'Z > 0 \; k \times k$ matrix.

- **Note:** no restrictions on reduced form parameters $\longrightarrow$ allow for weak IV.

- Many tests available for **full vector inference**

$$H_0 : \beta = \beta_0, \gamma = \gamma_0 \text{ vs } H_1 : \text{not } H_0$$

  including AR (Anderson and Rubin, 1949), LM, and CLR tests, see Kleibergen (2002), Moreira (2003, 2009).

- **Optimality properties:** Andrews, Moreira, and Stock (2006) and Chernozhukov, Hansen, and Jansson (2009).

Derived subvector procedures

- **Projection:** "inf" over parameter not under test, same critical value $\longrightarrow$ "computationally hard" and "uninformative".

- **Bonferroni:** Staiger and Stock (1997), Chaudhuri and Zivot (2011), McCloskey (2012), Wang and Doko Tchatoka (2017)...; often computationally hard, power ranking with projection unclear.

- **Plug-in approach:** Kleibergen (2004), Guggenberger and Smith (2005)...Requires strong ID of parameters not under test.

- Kleibergen (2015): subvector CLR test with correct size under weak IV and asymptotically efficient under strong IV.

- Power ranking under weak IV is unclear:
  - In just-identified case $k = m_Y + m_W$, subvector LR statistic is equal to the subvector AR statistic, and CLR cv is $\chi^2_{m_Y,1-\alpha}$.
  - Hence, less powerful than the test proposed here.

The Anderson and Rubin (1949) test

- **AR test stat** for **full vector hypothesis**

$$H_0 : \beta = \beta_0, \gamma = \gamma_0 \; vs \; H_1 : \text{not } H_0$$

- AR statistic exploits $EZ_i\varepsilon_i = 0$.

- **AR test stat:**

$$AR_n(\beta_0, \gamma_0) = \frac{(y - Y\beta_0 - W\gamma_0)' P_Z (y - Y\beta_0 - W\gamma_0)}{\left(1 \,\vdots\, -\beta_0' \,\vdots\, -\gamma_0'\right) \Omega \left(1 \,\vdots\, -\beta_0' \,\vdots\, -\gamma_0'\right)'}$$

- AR stat is χ^2_k under null hypothesis; critical value $\chi^2_{k,1-\alpha}$.

- **Subvector AR statistic** for testing H_0 is given by

$$AR_n(\beta_0) = \min_{\gamma \in R^{m_W}} \frac{(\overline{Y}_0 - W\gamma)' P_Z (\overline{Y}_0 - W\gamma)}{(1 \vdots -\beta_0' \vdots -\gamma')\, \Omega\, (1 \vdots -\beta_0' \vdots -\gamma')'},$$

where $\overline{Y}_0 = y - Y\beta_0$.

- Alternative representation: Let $\hat{\kappa}_i$ for $i = 1, ..., p = 1 + m_W$ be roots of characteristic polynomial in κ

$$\left| \kappa \Omega(\beta_0) - \left(\overline{Y}_0 \vdots W\right)' P_Z \left(\overline{Y}_0 \vdots W\right) \right| = 0,$$

ordered non-increasingly, where we define

$$\Omega(\beta_0) = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_W} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ 0 & I_{m_W} \end{pmatrix}.$$

Then

$$AR_n(\beta_0) = \hat{\kappa}_p.$$

- As discussed: When using $\chi^2_{k,1-\alpha}$ critical values, trivially, test has correct size;

- GKMC show that this is also true for $\chi^2_{k-m_W,1-\alpha}$ critical values.

- **Next:** AR statistic is the minimum eigenvalue of a non-central Wishart matrix.

- The roots $\hat{\kappa}_i$ solve

$$0 = \left|\hat{\kappa}_i I_{1+m_W} - \Xi'\Xi\right|, \quad i = 1, ..., p = 1 + m_W,$$

where $\Xi \sim N\left(\mathcal{M}, I_k \otimes I_p\right)$, and $\mathcal{M}$ is a $k \times p$.

- Under H_0, the noncentrality matrix becomes $\mathcal{M} = \left(0^k, \Theta_W\right)$, where

$$\Theta_W = \left(Z'Z\right)^{1/2} \Pi_W \Sigma_{V_W V_W \cdot \varepsilon}^{-1/2},$$
$$\Sigma_{V_W V_W \cdot \varepsilon} = \Sigma_{V_W V_W} - \Sigma'_{\varepsilon V_W} \sigma_{\varepsilon\varepsilon}^{-1} \Sigma_{\varepsilon V_W}$$

and

$$\begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon V_W} \\ \Sigma'_{\varepsilon V_W} & \Sigma_{V_W V_W} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ -\gamma & I_{m_W} \end{pmatrix}' \Omega \begin{pmatrix} 1 & 0 \\ -\beta_0 & 0 \\ -\gamma & I_{m_W} \end{pmatrix}$$

- **Summarizing**, under H_0

$$\Xi'\Xi \sim \mathcal{W}_p\left(k, I_p, \mathcal{M}'\mathcal{M}\right),$$

non-central Wishart, with noncentrality matrix

$$\mathcal{M}'\mathcal{M} = \begin{pmatrix} 0 & 0 \\ 0 & \Theta'_W \Theta_W \end{pmatrix}$$

and

$$AR_n(\beta_0) = \kappa_{\min}(\Xi'\Xi)$$

- The distribution of the eigenvalues of a noncentral Wishart matrix only depends on the eigenvalues of the noncentrality matrix $\mathcal{M}'\mathcal{M}$.

- Hence, distribution of $\hat{\kappa}_i$ only depends on the eigenvalues of $\Theta_W'\Theta_W$, κ_i say, $i = 1, \ldots, m_W$ and $\kappa = (\kappa_1, ..., \kappa_{m_W})'$

- When $m_W = 1$, $\kappa_1 = \Theta_W'\Theta_W$ is scalar (concentration parameter for γ under Null).

Theorem: Suppose $m_W = 1$. Then, under the null hypothesis $H_0 : \beta = \beta_0$, the distribution function of the subvector AR statistic, $AR_n(\beta_0)$, is monotonically decreasing in the parameter κ_1.

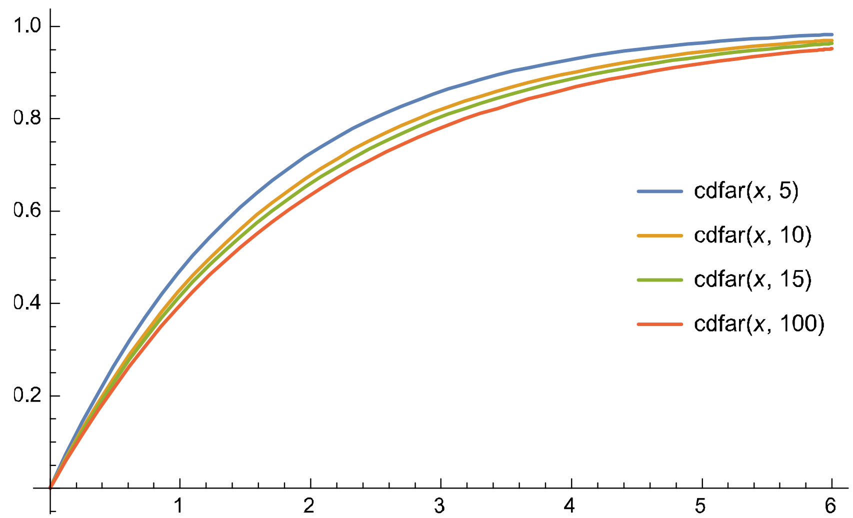


Figure 1: The cdf of the subset AR statistic with $k = 3$ instruments, for different values of $\kappa_1 = 5, 10, 15, 100$, shown in the legend on the right.

New critical value for subvector Anderson and Rubin test

- **Relevance:** If we knew κ_1 we could implement the subvector AR test with a smaller critical value than $\chi^2_{k-m_W,1-\alpha}$ which is the critical value in the case when κ_1 is "large".

- **Intuition for new critical value.** Let's assume $m_W = 1$ for simplicity.

- Under null, when κ_1 "is large", the larger root $\widehat{\kappa}_1$ is a sufficient statistic for κ_1, see Muirhead (1978).

- Muirhead provides approximate, nuisance parameter free, density of $AR_n(\beta_0) = \widehat{\kappa}_2$ given $\widehat{\kappa}_1$ (which measures strength of identification).

- The **new critical value** for the subvector AR-test at significance level $1-\alpha$ is given by

$$1-\alpha \text{ quantile of (approximation of } AR_n \text{ given } \widehat{\kappa}_1)$$

- Denote cv by

$$c_{1-\alpha}(\hat{\kappa}_1, k-m_W)$$

Depends only on α, $k-m_W$, and $\hat{\kappa}_1$.

- We find, by simulations over fine grid of values of κ_1, that test controls size.

- It improves on the GKMC procedure in terms of power.

- **Theorem:** Suppose $m_W = 1$. The subvector Anderson Rubin test that uses the new conditional critical value $c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$ has correct size under the assumptions above.

Details

- Again: $\kappa_1 \geq 0$ is nonzero latent root of $\mathcal{M}'\mathcal{M}$ (nuisance parameter).

- When the root is "large", the conditional density of $AR_n(\beta_0) = \hat{\kappa}_2$ given $\hat{\kappa}_1$ can be approximated by

$$f_{\hat{\kappa}_2|\hat{\kappa}_1}(x) \sim f_{\chi^2_{k-1}}(x)\,(\hat{\kappa}_1 - x)^{1/2}\,g(\hat{\kappa}_1),$$

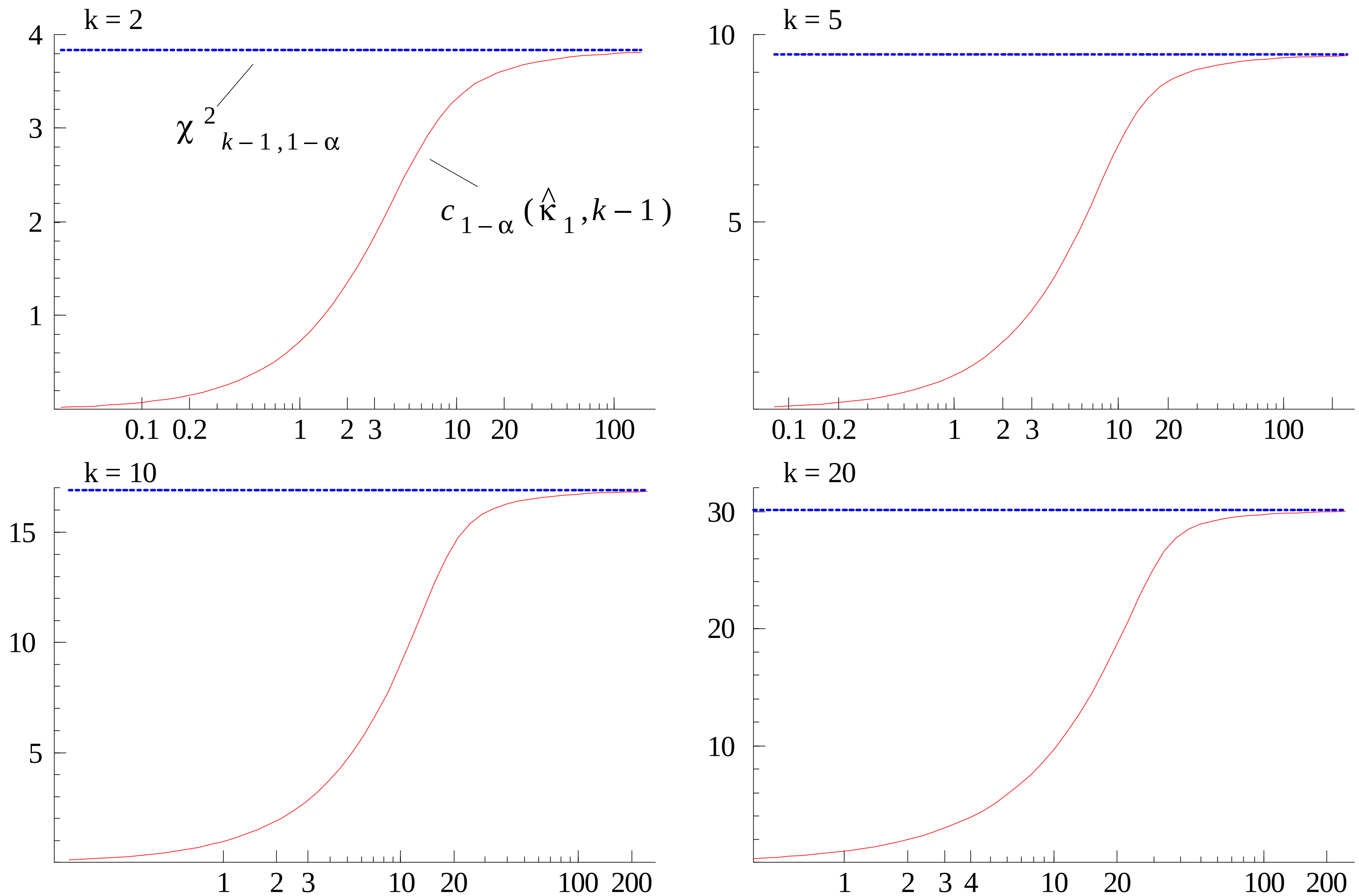
  where $f_{\chi^2_{k-1}}$ is the density of a χ^2_{k-1} and g is a function that does not depend on κ_1. (Muirhead, 1978 due to Leach, 1969).

- Analytical formula for g.

- Conditional quantiles can be computed by numerical integration.

- Conditional critical values can be tabulated $\longrightarrow$ implementation of new test is trival and fast.

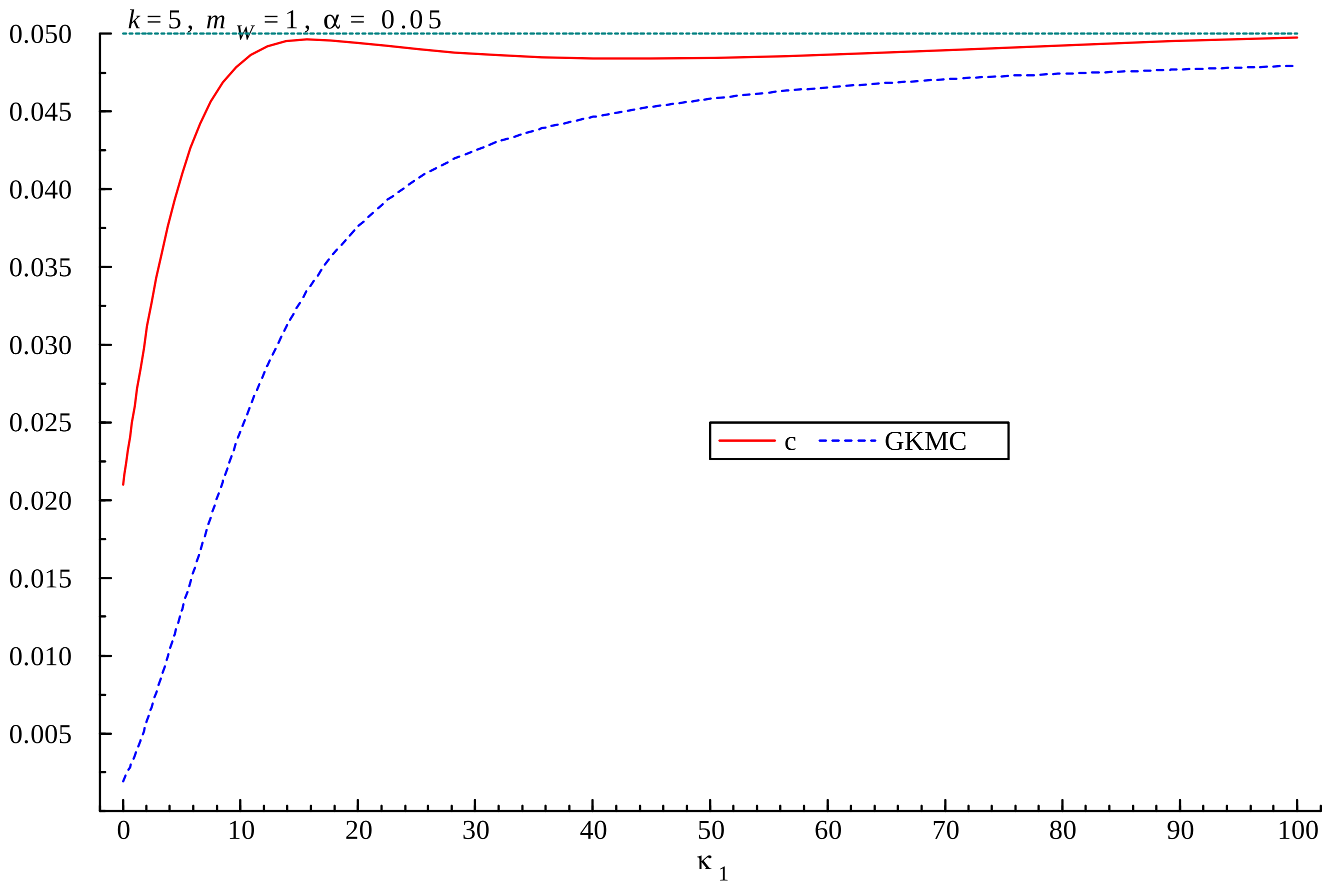
- They are increasing in $\hat{\kappa}_1$ and converging to quantiles of χ^2_{k-1}.



Critical value function $c_{1-\alpha}(\widehat{\kappa}_1, k-1)$ for $\alpha = 0.05$.

Table of conditional critical values

$\alpha = 5\%, \quad k - m_W = 4$											
$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv	$\hat{\kappa}_1$	cv
0.22	0.2	2.00	1.8	3.92	3.4	6.10	5.0	8.95	6.6	14.46	8.2
0.44	0.4	2.23	2.0	4.17	3.6	6.41	5.2	9.40	6.8	15.88	8.4
0.65	0.6	2.46	2.2	4.43	3.8	6.73	5.4	9.89	7.0	17.85	8.6
0.87	0.8	2.70	2.4	4.69	4.0	7.05	5.6	10.42	7.2	20.89	8.8
1.10	1.0	2.94	2.6	4.96	4.2	7.39	5.8	11.01	7.4	26.42	9.0
1.32	1.2	3.18	2.8	5.24	4.4	7.75	6.0	11.68	7.6	39.82	9.2
1.54	1.4	3.42	3.0	5.52	4.6	8.13	6.2	12.44	7.8	114.76	9.4
1.77	1.6	3.67	3.2	5.81	4.8	8.52	6.4	13.35	8.0	+.Inf	9.5



Null rejection frequency of subset AR test based on conditional (red) and χ^2_{k-1} (blue) critical values, as function of κ_1. 10000 MC simulations with importance sampling over a grid of 42 points.

Power

- The subvector AR statistic is the LR statistic for testing $H_0' : \rho(A) \leq m_W$ against $H_1' : \rho(A) = m_W + 1$ for $A = E\left[Z'\left(y - Y\beta_0 \vdots W\right)\right]$, where the data is $Z'\left(y - Y\beta_0 \vdots W\right)$.

- $H_0 : \beta = \beta_0$ implies H_0' but the converse is not true:

  – H_0' holds iff $\rho\left(\Pi_Y\left(\beta - \beta_0\right) \vdots \Pi_W\right) \leq m_W$, which includes $H_1 : \beta \neq \beta_0$ when $H_0' \backslash H_0$ holds, i.e., if Π_W is rank deficient or $\Pi_Y(\beta - \beta_0) \in span(\Pi_W)$.

- Under H_0', $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ are distributed as eigenvalues of $\mathcal{W}_p\left(k, I_p, \mathcal{M}'\mathcal{M}\right)$ with rank deficient noncentrality.

- Thus, every test $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p) \in [0, 1]$ that has size α under H_0 must also have size α under H_0', so cannot have power exceeding size under alternatives $H_0' \backslash H_0$.

- In other words, size α tests $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ can only have nontrivial power under alternatives $\rho(A) = m_W + 1$.

- We use this insight to derive a power envelope for tests of the form $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$.

- Consider only the case $m_W = 1$.

- Testing $\rho(\mathcal{M}) \leq 1$ against $\rho(\mathcal{M}) = 2$, where $\Xi \sim N(\mathcal{M}, I)$.

- Equivalently, $H_0' : \kappa_2 = 0,\ \kappa_1 \geq \kappa_2$ against $H_1' : \kappa_2 > 0, \kappa_1 \geq \kappa_2$.

- Maximal invariant is $\hat{\kappa}_1, \hat{\kappa}_2$ (Muirhead, 2009, Section 10.2).

- Likelihood (James, 1964)

$$lik(\kappa|\hat{\kappa}) = \exp\left(-\frac{\kappa_1 + \kappa_2}{2}\right)\ {}_0F_1^{(2)}\left(\frac{k}{2}; \frac{1}{4}\begin{pmatrix}\kappa_1 & 0 \\ 0 & \kappa_2\end{pmatrix}, \begin{pmatrix}\hat{\kappa}_1 & 0 \\ 0 & \hat{\kappa}_2\end{pmatrix}\right)$$

- Computed using the algorithms developed by Koev and Edelman (2008), available in C and Matlab.

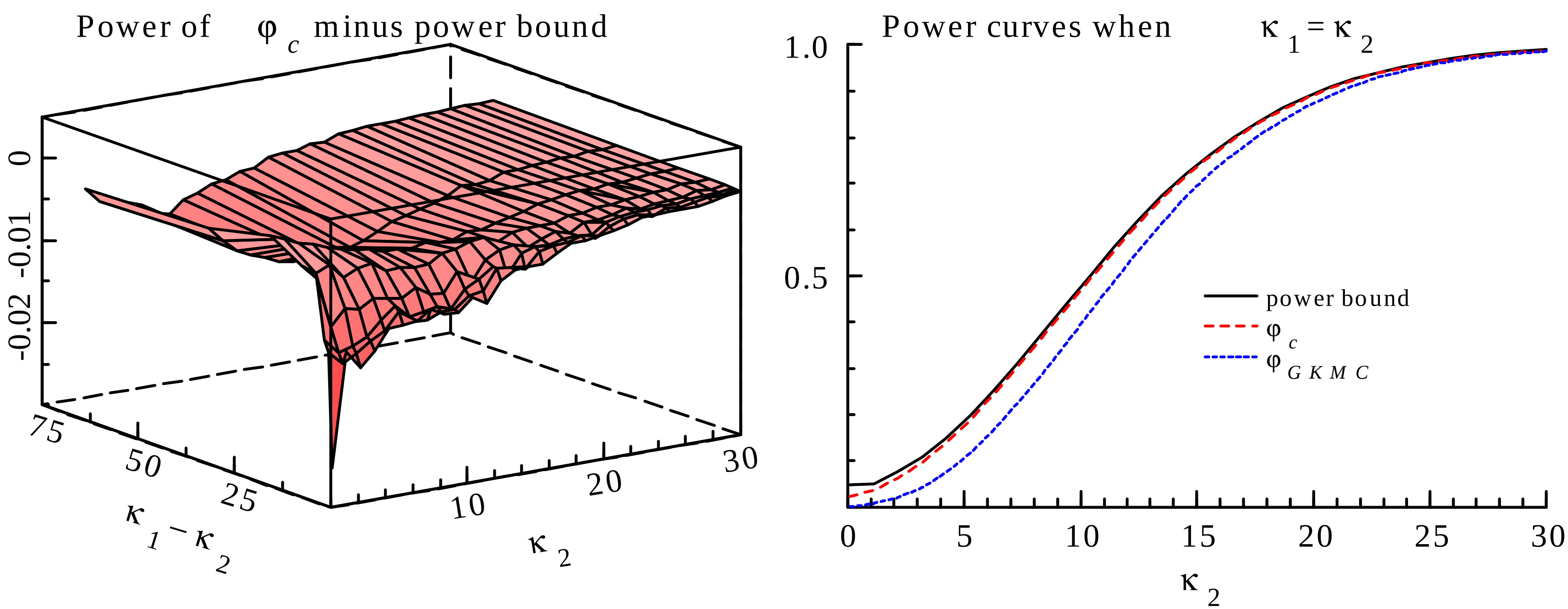
Power bounds

- Point-optimal power bounds for reduced rank testing problem using least favourable distribution Λ^{LF} over nuisance parameter κ_1.

- Two methods: Andrews Moreira and Stock (JoE, 2008, Sec 4.2) – AMS.
  - assumes one-point Λ^{LF}, gives lower and upper bounds on envelope.

- Elliott Mueller and Watson (Ecma 2015, Lemma 1) – ALFD (Approximate LF distn).

- Implementation: 42 points evenly spaced in log-scale between 0 and 99.



Power of conditional subvector AR test $\varphi_c(\hat{\kappa}) = 1_{\{\hat{\kappa}_2 > c_{1-\alpha}(\hat{\kappa}_1, k-1)\}}$ relative to power bound (left) and power of φ_c, $\varphi_{GKMC}(\hat{\kappa}) = 1_{\left\{\hat{\kappa}_2 > \chi^2_{k-1,1-\alpha}\right\}} = 1_{\{\hat{\kappa}_2 > c_{1-\alpha}(\infty, k-1)\}}$ and bound at $\kappa_1 = \kappa_2$ (right) for $k = 5$. Computed using 10000 MC replications.

- Little scope for power improvement over proposed test.

Size for $m_W > 1$

When $m_W = 1$ the new subvector AR test has correct size and uniformly improves the power of the test in GKMC.

$\longrightarrow$ Generalize this result to any m_W.

We define a new subvector AR test that rejects when

$$AR_n(\beta_0) > c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k - m_W).$$

Note: We condition on the LARGEST eigenvalue of the Wishart matrix.

Show now that this test has correct size and has uniformly larger power than the test in GKMC.

Theorem: Under the null $H_0 : \beta = \beta_0$, there exists a random orthogonal matrix O, such that for

$$\tilde{\tilde{\Xi}} = \Xi O \in R^{k\times p}, \text{ and its upper left submatrix } \tilde{\tilde{\Xi}}_{11} \in R^{k-m_W+1\times 2}$$

$\tilde{\tilde{\Xi}}'_{11}\tilde{\tilde{\Xi}}_{11}$ is a non-central Wishart 2×2 matrix of order $k - m_W + 1$ (cond'l on O), whose noncentrality matrix, $\tilde{\mathcal{M}}'_1\tilde{\mathcal{M}}_1$ say, is of reduced rank.

It then follows that

$$\begin{aligned}
\textbf{(i)} \quad AR_n(\beta_0) &= \kappa_{\min}\left(\Xi'\Xi\right) = \kappa_{\min}\left(\tilde{\tilde{\Xi}}'\tilde{\tilde{\Xi}}\right) \\
&\leq \kappa_{\min}\left(\tilde{\tilde{\Xi}}'_{11}\tilde{\tilde{\Xi}}_{11}\right) \leq \kappa_{\max}\left(\tilde{\tilde{\Xi}}'_{11}\tilde{\tilde{\Xi}}_{11}\right) \\
&\leq \kappa_{\max}\left(\tilde{\tilde{\Xi}}'\tilde{\tilde{\Xi}}\right) = \kappa_{\max}\left(\Xi'\Xi\right)
\end{aligned}$$

and thus

$$
\begin{aligned}
& P(AR_n(\beta_0) > c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k - m_W)) \\
& \leq P(\kappa_{\min}\left(\tilde{\Xi}'_{11}\tilde{\Xi}_{11}\right) > c_{1-\alpha}(\kappa_{\max}\left(\tilde{\Xi}'_{11}\tilde{\Xi}_{11}\right), k - m_W)) \\
& \leq \alpha,
\end{aligned}
$$

where the last inequality follows from the case $m_W = 1$ (by conditionning on O).

(ii) new conditional test is uniformly more powerful than test in GKMC (because $c_{1-\alpha}(\cdot, k - m_W))$ is increasing and converging to $\chi^2_{k-m_W,1-\alpha}$ as argument goes to infinity).

Asymptotic case

- **Parameter space** $\mathcal{F}$ under the null hypothesis $H_0 : \beta = \beta_0$. Let $U_i = (\varepsilon_i, V'_{W,i})'$ and F distribution of (U_i, V_{Yi}, Z_i)

  $\mathcal{F}$ is set of all $(\gamma, \Pi_W, \Pi_Y, F)$ s.t.

  $\gamma \in R^{m_W}, \Pi_W \in R^{k \times m_W}, \Pi_Y \in R^{k \times m_Y}$,
  $E_F(||T_i||^{2+\delta}) \leq B$, for $T_i \in \{Z_i\varepsilon_i, vec(Z_i V'_{W,i}), V_{W,i}\varepsilon_i, \varepsilon_i, V_{W,i}, Z_i\}$,
  $E_F(Z_i(\varepsilon_i, V'_{Wi}, V'_{Yi})) = 0$,
  $E_F(vec(Z_i U'_i)(vec(Z_i U'_i))') = (E_F(U_i U'_i) \otimes E_F(Z_i Z'_i))$,
  $\kappa_{\min}(A) \geq b$ for $A \in \{E_F(Z_i Z'_i), E_F(U_i U'_i)\}$

  for some $b > 0$, $B < \infty$, where $\kappa_{\min}(\cdot)$ is smallest eigenvalue, "$\otimes$" Kronecker product, $vec(\cdot)$ column vectorization.

- Subvector AR stat equals

$$AR_n(\beta_0) = \kappa_{\mathsf{min}} \left(\left(\frac{\overline{Y}' M_Z \overline{Y}}{n-k} \right)^{-1/2} (\overline{Y}' P_Z \overline{Y}) \left(\frac{\overline{Y}' M_Z \overline{Y}}{n-k} \right)^{-1/2} \right)$$

where

$$\overline{Y} := (y - Y\beta_0 \vdots W) \in R^{n \times (1+m_W)}$$

- GKMC showed $\varphi_{GKMC} = 1_{\left\{AR_n(\beta_0) > \chi^2_{k-m_W,1-\alpha}\right\}}$ has correct asymptotic size for parameter space $\mathcal{F}$.

- **Current paper:** $\varphi_c = 1_{\{AR_n(\beta_0) > c_{1-\alpha}(\widehat{\kappa}_{\mathsf{max}}, k-m_W)\}}$ has correct asy size.

Asymptotic Size of conditional subvector AR test

- The derivation of asymptotic size follows the method of Andrews Cheng and Guggenberger (2011).

- The complication relative to GKMC is that we need joint limiting distribution of $\widehat{\kappa}_1, ..., \widehat{\kappa}_p$, not just the minimum, $\widehat{\kappa}_p$.

- Fortunately, we can use the results of Andrews and Guggenberger (2015) on limit distribution of eigenvalues of quadratic forms.

- It turns out that joint limit depends only on localization parameters corresponding to the singular values of

$$(E_F Z_i Z_i')^{1/2}(\Pi_W \gamma, \Pi_W)\Omega(\beta_0)^{-1/2},$$

which correspond to singular values of Θ_W (concentration matrix) in the finite sample case.

- Hence, replicates the finite sample, normal, fixed IV, known variance matrix setup.

- Correct asymptotic size then follows from correct finite sample size.

Takeaways

- We can obtain uniform power improvement over the subvector AR test in GKMC by using data-dependent critical values.

- We propose one such test whose conditional cv's are easy to compute and can be tabulated.

- In the case $m_W = 1$, i.e., when there is a single endogenous regressor whose coefficient is unrestricted under H_0, the proposed cv's are an increasing function of a first-stage F statistic for that regressor.

- There is little scope for further power improvement when $m_W = 1$ – our proposed test is nearly optimal.

Current work: Drop assumption of conditional homoskedasticity $\longrightarrow$ allow for **heteroskedasticity**.

- Lee (2014) found an example in which the subvector AR with $\chi^2_{k-m_W,1-\alpha}$ cv's overrejects when the covariance matrix does not have Kronecker product form.

- Importantly, this does not apply to iid data.

- So far, we have found correct size of the heteroskedasticity robust subvector AR test that uses $\chi^2_{k-m_W,1-\alpha}$ cv's when $m_W = 1$ and $k = 2$.

- We are working on generalizing this to higher dimensions.