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

Patrik Guggenberger
Pennsylvania State University

Joint work with Frank Kleibergen (University of Amsterdam) and
Sophocles Mavroeidis (University of Oxford)

Indiana University

September, 2018

Overview

- Robust inference on a slope coefficient(s) in a linear IV regression

- "Robust" means uniform control of null rejection probability over all "empirically relevant" parameter constellations

- "Weak instruments"
  - pervasive in applied research (Angrist and Krueger, 1991)
  - adverse effect on estimation and inference (Dufour, 1997; Staiger and Stock 1997)

- Large literature on "robust inference" for the full parameter vector

- Here: Consider **subvector inference in the linear IV model**, allowing for **weak instruments**

- First assume **homoskedasticity**

  – then relax to general **Kronecker-Product** structure

  – then allow for arbitrary forms of **heteroskedasticity**

- Presentation based on two papers; one being "A more powerful subvector Anderson Rubin test in linear instrumental variables regression"

- Focus on the **Anderson and Rubin (AR, 1949) subvector test statistic**:
  - **"History of critical values":**
  - Projection of AR test (Dufour and Taamouti, 2005)
  - Guggenberger, Kleibergen, Mavroeidis, and Chen (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

- Show: controls **finite sample/asymptotic size** & has uniformly **higher power** than method in GKMC

- One additional main contribution : **computational ease**

- Implication: Test in GKMC is "inadmissible"

Presentation

- Introduction: ✓

- finite sample case

  a) $m_W = 1$: motivation, correct size, power analysis (near optimality result)

  b) $m_W > 1$: correct size, uniform power improvement over GKMC

  c) refinement

- asymptotic case:

  a) homoskedasticity

  b) general Kronecker-Product structure

  c) general case (arbitrary forms of heteroskedasticity)

Model and Objective (finite sample case)

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

$$y \in R^n, Y \in R^{n \times m_Y} \text{ (end or ex)}, W \in R^{n \times m_W} \text{ (end)},\ Z \in R^{n \times k} \text{ (IVs)}$$

- **Reduced form:**

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

where $v_y := \varepsilon + V_Y\beta + V_W\gamma$.

- **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 reduced form error satisfies:

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

   for some $\Omega \in R^{(m+1)\times(m+1)}$ s.t. the variance matrix of $(\overline{Y}_{0i}, V_{Wi}')'$ for $\overline{Y}_{0i} = y_i - Y_i'\beta_0 = W_i'\gamma + \varepsilon_i$, namely

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

   is 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 Π_Y and Π_W $\longrightarrow$ allow for weak IV

- Several robust 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), Andrews, Marmer, and Yu (2018), and Chernozhukov, Hansen, and Jansson (2009)

Subvector procedures

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

- **Bonferroni and related techniques:** Staiger and Stock (1997), Chaudhuri and Zivot (2011), McCloskey (2012), Zhu (2015), Andrews (2017),Wang and Tchatoka (2018) ...; often computationally hard, power ranking with projection unclear

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

- GMM models: Andrews, I. and Mikusheva (2016)

- Models defined by moment inequalities: Gafarov (2016), Kaido, Molinari, and Stoye (2016), Bugni, Canay, and Shi (2017), ...

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 distri. as 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 again $\overline{Y}_0 = y - Y\beta_0$.

- Alternative representation (using $\kappa_{\min}(A) = \min_{x,||x||=1} x'Ax$):

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

where $\hat{\kappa}_i$ for $i = 1, ..., p = 1 + m_W$ be roots of characteristic polynomial in κ

$$\left| \kappa I_p - \Omega(\beta_0)^{-1/2} \left(\overline{Y}_0 \vdots W\right)' P_Z \left(\overline{Y}_0 \vdots W\right) \Omega(\beta_0)^{-1/2} \right| = 0,$$

ordered non-increasingly

- When using $\chi^2_{k,1-\alpha}$ critical values, as for projection, trivially, test has correct size;

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

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

- For par space above, 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

$$\Xi \sim N \left(M, I_k \otimes I_p \right),$$

and M is a $k \times p$.

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

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

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 the $p \times p$ matrix

$$\Xi'\Xi \sim W\left(k, I_p, M'M\right),$$

has non-central Wishart with noncentrality matrix

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

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 $M'M$.

- 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 $m_W = 1$, $\kappa = \kappa_1 = \Theta_W'\Theta_W$ is scalar.

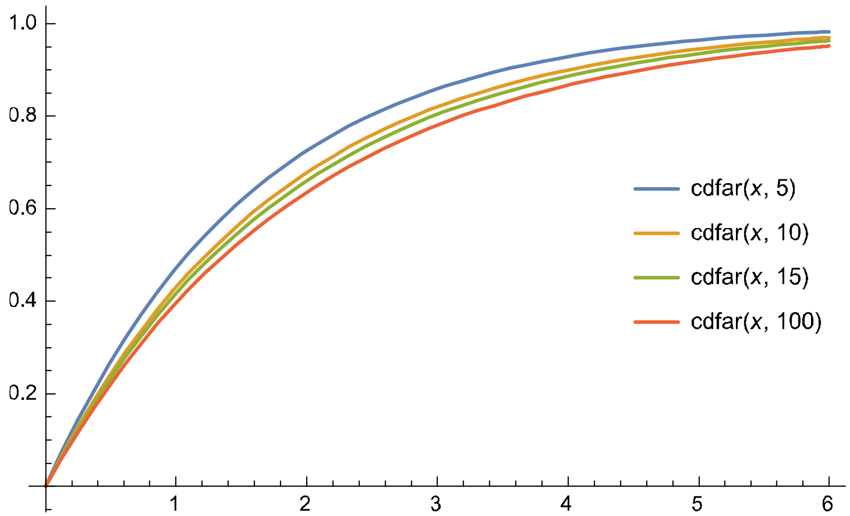


Figure 1: The cdf of the subset AR statistic with $k = 3$ instruments, for different values of $\kappa_1 = 5, 10, 15, 100$

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.

New critical value for subvector Anderson and Rubin test: $m_W = 1$

- **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".

- Muirhead (1978): Under null, when κ_1 "is large", the larger root $\widehat{\kappa}_1$ (which measures strength of identification) is a sufficient statistic for κ_1

- More precisely: the conditional density of $AR_n(\beta_0) = \hat{\kappa}_2$ given $\hat{\kappa}_1$ can be approximated by

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

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.

- Analytical formula for g

- 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 $\alpha, k - m_W$, and $\hat{\kappa}_1$

- Conditional quantiles can be computed by numerical integration

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

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

- We find, by simulations over fine grid of values of κ_1, that new test

$$1(AR_n(\beta_0) > c_{1-\alpha}(\hat{\kappa}_1, k - m_W))$$

  controls size

- It improves on the GKMC procedure in terms of power

- **Theorem:** Suppose $m_W = 1$. The new conditional subvector Anderson Rubin test has correct size under the assumptions above.

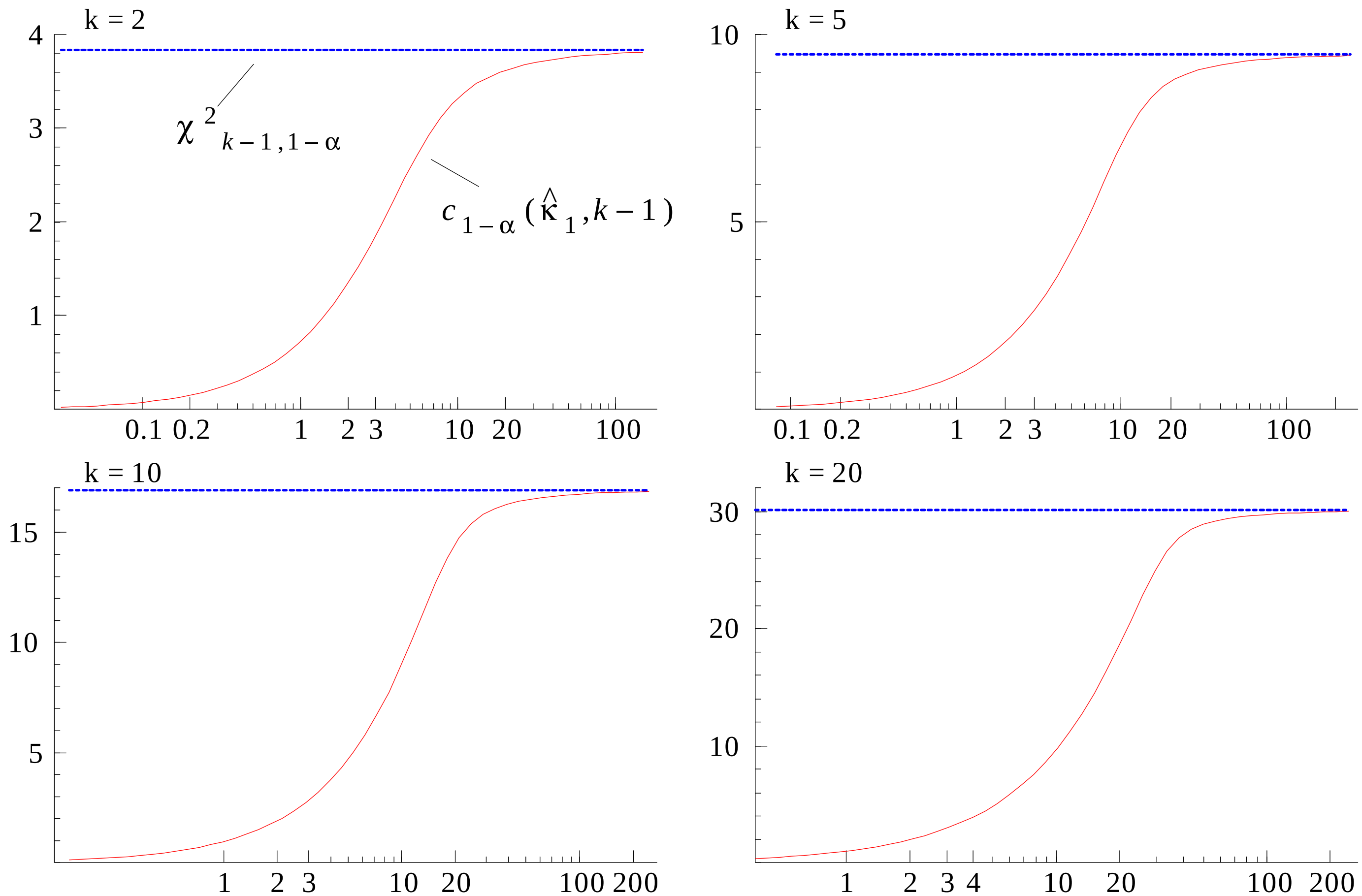
- Proof partly based on simulations; Verified for e.g. $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$.

- **Summary $m_W = 1$:** the cond'l test rejects when

$$\hat{\kappa}_2 > c_{1-\alpha}(\hat{\kappa}_1, k-1),$$

where $(\hat{\kappa}_1, \hat{\kappa}_2)$ are the eigenvalues of 2×2 matrix $\Xi'\Xi \sim W(k, I_p, M'M)$;

Under the null $M'M$ is of rank 1; **test has size α**

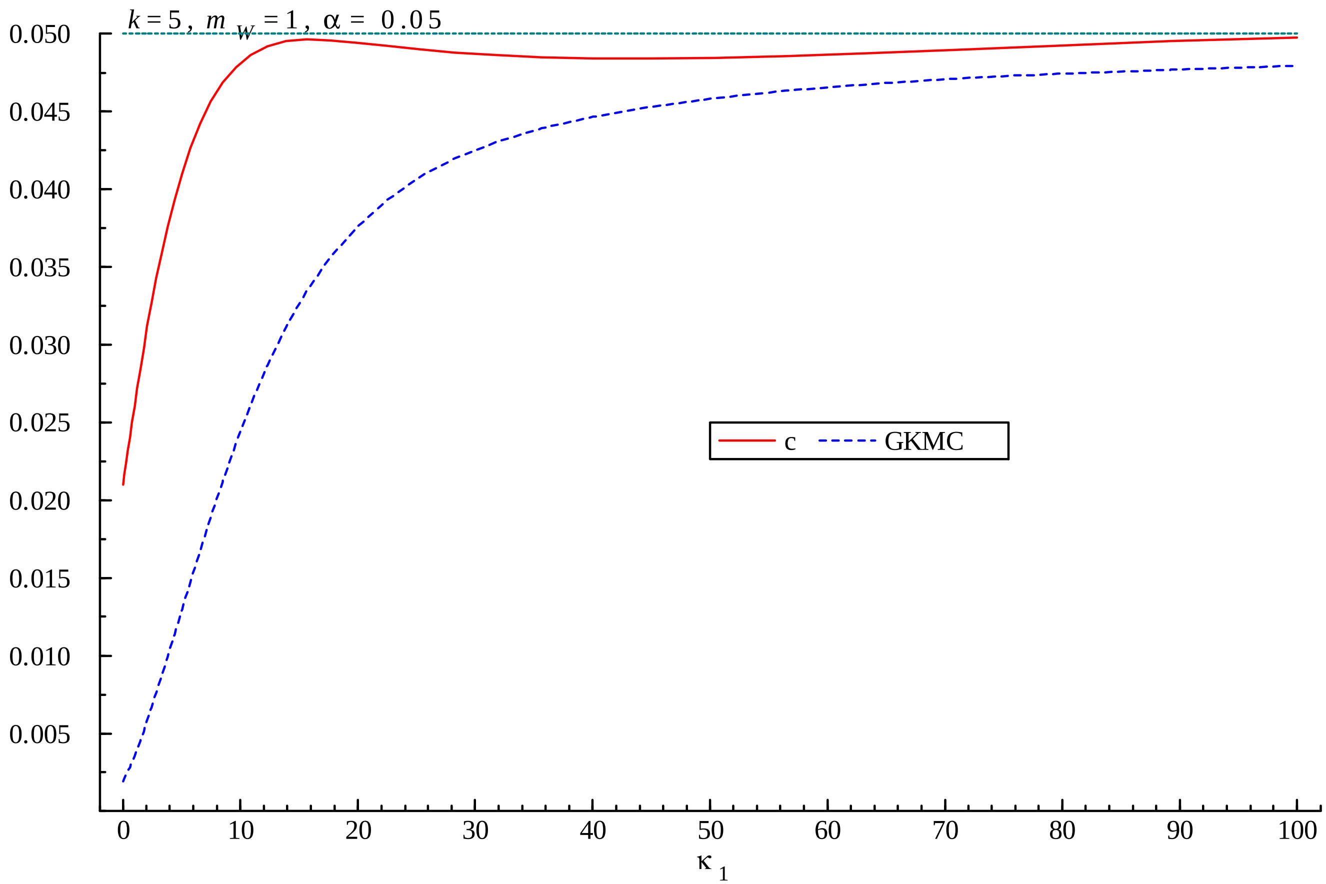


Critical value function $c_{1-\alpha}(\widehat{\kappa}_1, k-1)$ for $\alpha = 0.05$.

Table of conditional critical values cv=$c_{1-\alpha}(\hat{\kappa}_1, k - m_W)$

$\alpha = 5\%, \quad k - m_W = 4$											
$\hat{\kappa}_1$	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

* For simplicity of implementation we suggest linear interpolation of tabulated cvs; we verify resulting test has correct size



Null rejection frequency of subset AR test based on conditional (red) and χ^2_{k-1} (blue) critical values, as function of κ_1.

Extension to $m_W > 1$

We define a new subvector Anderson Rubin 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.

Theorem: The test above has i) correct size and ii) has uniformly larger power than the test in GKMC.

Lemma: Under the null $H_0 : \beta = \beta_0$, there exists a random matrix $O \in O(p)$, such that for

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

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

Proof of Theorem:

(i) Note that

$$\begin{aligned} AR_n(\beta_0) &= \kappa_{\min}\left(\Xi'\Xi\right) = \kappa_{\min}\left(\tilde{\Xi}'\tilde{\Xi}\right) \\ &\leq \kappa_{\min}\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right) \leq \kappa_{\max}\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right) \\ &\leq \kappa_{\max}\left(\tilde{\tilde{\Xi}}'\tilde{\tilde{\Xi}}\right) = \kappa_{\max}\left(\Xi'\Xi\right) \end{aligned} \tag{1}$$

and thus

$$\begin{aligned} & P(AR_n(\beta_0) > c_{1-\alpha}(\kappa_{\max}\left(\Xi'\Xi\right), k - m_W)) \\ \leq\ & P(\kappa_{\min}\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right) > c_{1-\alpha}(\kappa_{\max}\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right), k - m_W)) \\ =\ & P(\kappa_2\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right) > c_{1-\alpha}(\kappa_1\left(\tilde{\tilde{\Xi}}_{11}'\tilde{\tilde{\Xi}}_{11}\right), k - m_W)) \\ \leq\ & \alpha, \end{aligned}$$

where first inequality follows from (1) and last inequality from correct size for $m_W = 1$ (by conditionning on O) and the lemma

Recall summary when $m_W = 1$: new test rejects when

$$\hat{\kappa}_2 > c_{1-\alpha}(\hat{\kappa}_1, k-1)$$

where $(\hat{\kappa}_1, \hat{\kappa}_2)$ are the eigenvalues of $\Xi'\Xi \sim W(k, I_2, M'M)$ and $M'M$ is of rank 1 under the null

(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), i.e. the test in GKMC is inadmissible

Power analysis of tests based on $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$

- For $A = E\left[Z'\left(y - Y\beta_0 \vdots W\right)\right] \in R^{k \times p}$, consider

$$H_0' : \rho\left(A\right) \leq m_W \text{ versus } H_1' : \rho\left(A\right) = p = m_W + 1$$

- $H_0 : \beta = \beta_0$ implies H_0' but the converse is not true:
  - H_0' holds iff $\left[\rho\left(\Pi_W\right) < m_W \text{ or } \Pi_Y\left(\beta - \beta_0\right) \in span(\Pi_W)\right]$

- Under H_0', $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ are distributed as eigenvalues of Wishart $W\left(k, I_p, M'M\right)$ with rank deficient noncentrality matrix - a distribution that appears also under H_0

- 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 $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ under H_0 can only have nontrivial power under alternatives $\rho(A) = p$.

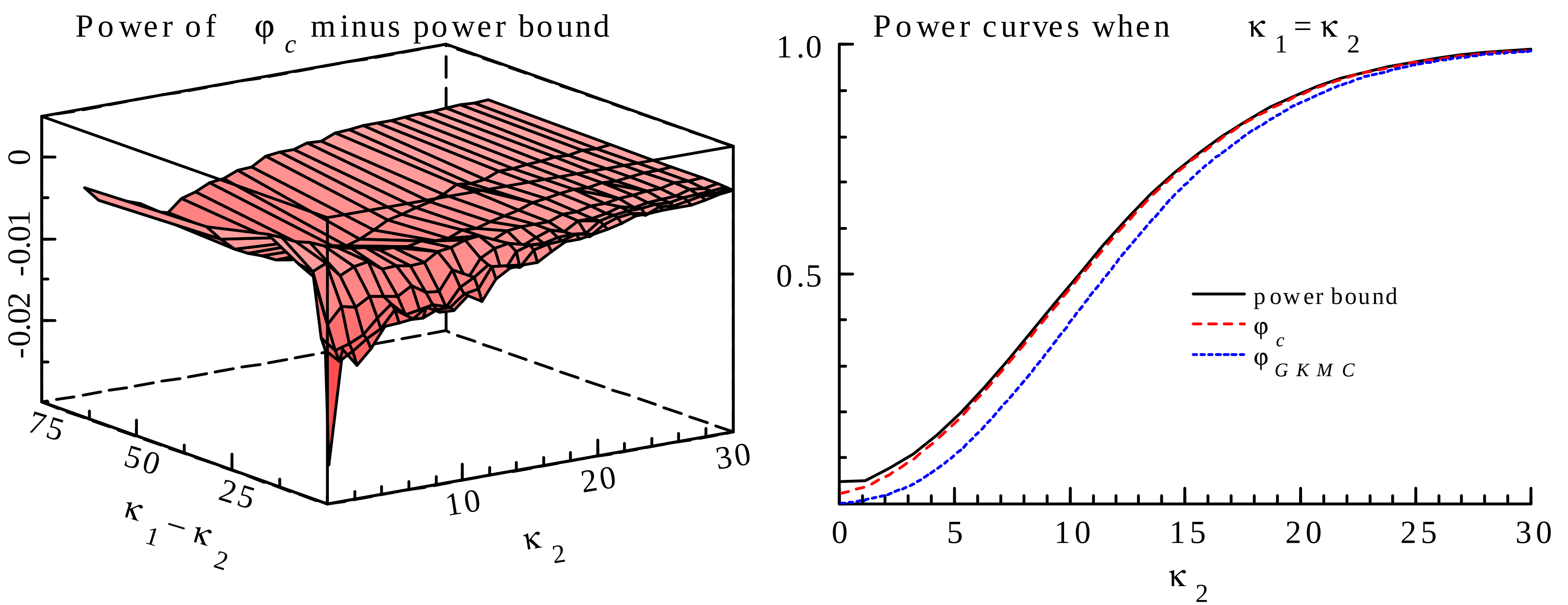
- We use this insight to derive a power envelope for tests of the form $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$.

Power bounds

- Consider only the case $m_W = 1$.

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

- Obtain point-optimal power bounds using approximately least favorable distribution Λ^{LF} over nuisance parameter κ_1 based on algorithm in Elliott, Müller, and Watson (2015)



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. But not zero scope...:

Refinement: For the case $k = 5$, $m_W = 1$, and $\alpha = 5\%$, let φ_{adj} be the test that uses the critical values in Table above where the smallest 8 critical values are divided by 5

Asymptotic case: a) homoskedasticity

- Define **parameter space** $\mathcal{F}$ under the null hypothesis $H_0 : \beta = \beta_0$.

  Let $U_i := (\varepsilon_i + V'_{W,i}\gamma, V'_{W,i})'$ and F distribution of (U_i, V_{Yi}, Z_i)

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

$$
\begin{aligned}
&\gamma \in R^{m_W}, \Pi_W \in R^{k \times m_W}, \Pi_Y \in R^{k \times m_Y}, \\
&E_F(||T_i||^{2+\delta}) \leq M, \text{ for } T_i \in \{vec(Z_i U_i), Z_i, U_i\}, \\
&E_F(Z_i(\varepsilon_i, V'_{Wi}, V'_{Yi})) = 0, \\
&E_F(vec(Z_i U_i')(vec(Z_i U_i'))') = (E_F(U_i U_i') \otimes E_F(Z_i Z_i')), \\
&\kappa_{\min}(A) \geq \delta \text{ for } A \in \{E_F(Z_i Z_i'), E_F(U_i U_i')\}
\end{aligned}
$$

  for some $\delta > 0$, $M < \infty$

- Note: no restriction is imposed on the variance matrix of $vec(Z_i V'_{Yi})$

- **subvector AR stat** equals smallest solution of

$$\left|\widehat{\kappa} I_{1+m_W} - \left(\frac{\overline{Y}' M_Z \overline{Y}}{n-k}\right)^{-1/2} (\overline{Y}' P_Z \overline{Y}) \left(\frac{\overline{Y}' M_Z \overline{Y}}{n-k}\right)^{-1/2}\right| = 0$$

  where

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

- **Note:** Same as in finite sample case with $\Omega(\beta_0)$ replaced by $\frac{\overline{Y}' M_Z \overline{Y}}{n-k}$

- **critical value** is again

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

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

- **Theorem:** The new subvector AR test has correct asymptotic size for parameter space $\mathcal{F}$.

- Again, part of the proof is based on simulations.

Asymptotic case: b) general Kronecker Product Structure

- For $U_i := (\varepsilon_i + V'_{W,i}\gamma, V'_{W,i})'$, $p := 1 + m_W$, and $m := m_Y + m_W$ let

$$\begin{aligned}\mathcal{F}_{KP} = \{(\gamma, \Pi_W, \Pi_Y, F) : \gamma \in \Re^{m_W}, \Pi_W \in \Re^{k \times m_W}, \Pi_Y \in \Re^{k \times m_Y}, \\ E_F(||T_i||^{2+\delta_1}) \leq B, \text{ for } T_i \in \{vec(Z_i U_i'), vec(Z_i Z_i')\}, \\ E_F(Z_i V_i') = 0^{k \times (m+1)}, \ \mathbf{E}_F(\mathbf{vec}(\mathbf{Z}_i \mathbf{U}_i')(\mathbf{vec}(\mathbf{Z}_i \mathbf{U}_i'))') = \mathbf{G}_1 \otimes \mathbf{G}_2, \\ \kappa_{\min}(A) \geq \delta_2 \text{ for } A \in \{E_F\left(Z_i Z_i'\right), G_1, G_2\}\}\end{aligned}$$

  for pd $G_1 \in \Re^{p \times p}$ (whose upper left element is normalized to 1) and $G_2 \in \Re^{k \times k}$ and $\delta_1, \delta_2 > 0$, $B < \infty$

- Covers homoskedasticity, but also cases of (cond) heteroskedasticity

Example. Take $(\widetilde{\varepsilon}_i, \widetilde{V}_{Wi}')' \in \Re^p$ i.i.d. zero mean with pd variance matrix, independent of Z_i, and

$$(\varepsilon_i, V_{Wi}')' := f(Z_i)(\widetilde{\varepsilon}_i, \widetilde{V}_{Wi}')'$$

for some scalar valued function f of Z, e.g. $f(Z_i) = ||Z_i||/k^{1/2}$. Then

$$\begin{aligned}
&E_F(vec(Z_iU_i')(vec(Z_iU_i'))') \\
&= E_F\left(U_iU_i' \otimes Z_iZ_i'\right) \\
&= E_F\left((\varepsilon_i + V_{W,i}'\gamma, V_{W,i}')'(\varepsilon_i + V_{W,i}'\gamma, V_{W,i}') \otimes Z_iZ_i'\right) \\
&= E_F\left((\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')'(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')\right) \otimes E_F\left(f(Z_i)^2 Z_iZ_i'\right)
\end{aligned}$$

has KP structure even though

$$E_F(U_iU_i'|Z_i) = f(Z_i)^2 E_F(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')'(\widetilde{\varepsilon}_i + \widetilde{V}_{W,i}'\gamma, \widetilde{V}_{W,i}')$$

depends on Z_i.

- **Modified AR subvector statistic**. Estimate $E_F(U_iU_i' \otimes Z_iZ_i')$ by

$$\widehat{R}_n := n^{-1} \sum_{i=1}^{n} f_i f_i' \in \Re^{kp \times kp}, \text{ where}$$
$$f_i := ((M_Z(y - Y\beta_0))_i, (M_Z W)_i')' \otimes Z_i \in \Re^{kp}.$$

- Let

$$(\widehat{G}_1, \widehat{G}_2) = \arg\min ||\overline{G}_1 \otimes \overline{G}_2 - \widehat{R}_n||_F,$$

where the minimum is taken over $(\overline{G}_1, \overline{G}_2)$ for $\overline{G}_1 \in \Re^{p \times p}$, $\overline{G}_2 \in \Re^{k \times k}$ being pd, symmetric matrices, normalized such that the upper left element of $\overline{G}_1$ equals 1. Estimators are unique and given in closed form.

- The subvector AR statistic, $AR_{KP,n}(\beta_0)$ is defined it as the smallest root $\hat{\kappa}_{pn}$ of the roots $\hat{\kappa}_{in}$, $i = 1, ..., p$ (ordered nonincreasingly) of the characteristic polynomial

$$\left|\hat{\kappa} I_p - n^{-1} \widehat{G}_1^{-1/2} \left(\overline{Y}_0, W\right)' Z \widehat{G}_2^{-1} Z' \left(\overline{Y}_0, W\right) \widehat{G}_1^{-1/2}\right| = 0.$$

- Note: Relative to previous definition,

  $\widehat{G}_1$ replaces $\frac{\overline{Y}' M_Z \overline{Y}}{n-k}$ and $\widehat{G}_2$ replaces $\frac{Z'Z}{n}$

- The conditional subvector AR_{KP} test rejects H_0 at nominal size α if

$$AR_{KP,n}(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k - m_W),$$

  where $c_{1-\alpha}(\cdot, \cdot)$ is defined as above.

Theorem: The conditional subvector AR_{KP} test implemented at nominal size α has asymptotic size, i.e.

$$\lim \sup_{n\to\infty} \sup_{(\gamma,\Pi_W,\Pi_Y,F)\in\mathcal{F}_{KP}} P_{(\beta_0,\gamma,\Pi_W,\Pi_Y,F)}(AR_{AKP,n}(\beta_0) > c_{1-\alpha}(\hat{\kappa}_{1n}, k-m_W))$$

equal to α.

Asymptotic case: c) General forms of Hetero

- Perform a Wald type pretest based on $\widehat{G}_1 \otimes \widehat{G}_2 - \widehat{R}_n$ to test the null of Kronecker Product structure

- If pretest rejects continue with a robust (to hetero and weak IV) subvector procedure, like the AR type tests proposed in Andrews (2017)

- Otherwise, continue with the test AR$_{KP}$ test

- Resulting test has correct asymptotic size no matter what the pretest nominal size is

- Reasons:
  - pretest is consistent against deviations from null for which

    $$n^{1/2}\min||\overline{G}_1 \otimes \overline{G}_2 - E_F(U_iU_i' \otimes Z_iZ_i')|| \to \infty$$

    and the AR type tests in Andrews (2017) have correct asymptotic size
  - when

    $$n^{1/2}\min||\overline{G}_1 \otimes \overline{G}_2 - E_F(U_iU_i' \otimes Z_iZ_i')|| = O(1)$$

    the conditional subvector AR_{KP} test has correct asymptotic size and rejects whenever the AR type test in Andrews (2017) rejects.

Asymptotic Size: General theory

- Distinction between pointwise (asymptotic) null rejection probability and (asymptotic) size

"Discontinuity" in limiting distribution of test statistic

Staiger and Stock (1997): simplified version of linear IV model with one IV

$$\begin{aligned} y_1 &= y_2\theta + u, \\ y_2 &= Z\pi + v \end{aligned}$$

Let $\lambda_n = (\lambda_{1n}, \lambda_{2n}, \lambda_{3n})$ be sequence of parameters s.t. $\lambda_{3n} = (F_n, \pi_n)$

$$\lambda_{1n} = (EZ_i^2)^{1/2}\pi/\sigma_v \text{ and } \lambda_{2n} = corr(u_i, v_i)$$

satisfies

$$h_{n,1}(\lambda_n) = n^{1/2}\lambda_{1n} \rightarrow h_1 < \infty \text{ and } h_{n,2}(\lambda_n) = \lambda_{2n} \rightarrow h_2.$$

We will denote such a sequence λ_n by $\lambda_{n,h}$.

Work out limiting distribution of 2SLS under $\lambda_{n,h}$:

$$\begin{aligned}
\frac{\sigma_v}{\sigma_u}(\widehat{\theta}_{2SLS} - \theta) &= \frac{\sigma_v}{\sigma_u}\frac{y_2' P_Z u}{y_2' P_Z y_2} = \frac{(n^{-1}Z'Z)^{-1/2} n^{-1/2} Z'u/\sigma_u}{(n^{-1}Z'Z)^{-1/2} n^{-1/2} Z' y_2/\sigma_v} \\
&= \frac{(n^{-1}Z'Z)^{-1/2} n^{-1/2} Z'u/\sigma_u}{(n^{-1}Z'Z)^{1/2} n^{1/2}\pi/\sigma_v + (n^{-1}Z'Z)^{-1/2} n^{-1/2} Z'v/\sigma_v} \\
&\rightarrow {}_d \frac{z_{u,h_2}}{h_1 + z_{v,h_2}}, \text{ where}
\end{aligned}$$

$$\begin{pmatrix} z_{u,h_2} \\ z_{v,h_2} \end{pmatrix} \sim N(0, \mathbf{\Sigma}_{h_2}) \text{ and } \mathbf{\Sigma}_{h_2} = \begin{pmatrix} 1 & h_2 \\ h_2 & 1 \end{pmatrix}$$

- Similarly for t test statistic $T_n(\theta_0)$:

$$T_n(\theta_0) \rightarrow_d J_h$$

  for $h = (h_1, h_2)$ under the parameter sequence $\lambda_{n,h}$.

- So, to implement the test, we should take the $1 - \alpha$-quantile $c_h(1 - \alpha)$ of J_h as the critical value

- If we implement a test using a Wald statistics with chi-square critical values, the asymptotic size is 1, see Dufour (1997)

- Problem: we cannot consistently estimate h; we can only estimate consistently λ_{1n}

- (h_1, h_2) takes on values in $H = (R \cup \{\pm\infty\}) \times [-1, 1]$

- We say the limit distribution of $T_n(\theta_0)$ "**depends discontinuously** on nuisance parameter λ_1" and **continuously** on λ_2

  Continuity: when $x \to x_0$ then $f(x) \to f(x_0)$

  Here $(EZ_i^2)^{1/2}\pi/\sigma_v \to 0$, but limit of $T_n(\theta_0)$ does not just depend on 0

- Situation arises frequently in applied econometrics and leads to size distortion for various "classical" inference procedures:

  weak IVs/identification, use of pretests, moment inequalities, (nuisance) parameters on boundary, inference in (V)ARs with unit root(s)

General Theory: Asymptotic Size of Tests

- $\{\varphi_n : n \geq 1\}$ sequence of tests for null hypothesis H_0
- λ indexes the true null distribution of the observations
- Parameter space for λ is some space Λ
- $RP_n(\lambda)$ denotes rejection probability of φ_n under λ
- The asymptotic size of φ_n for the parameter space Λ is defined as:

$$AsySz = \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} RP_n(\lambda)$$

Formula for Calculation of AsySz

Recall relevance of limits of $h_{n,1}(\lambda_n) = n^{1/2}\lambda_{1n} = n^{1/2}(EZ_i^2)^{1/2}\pi/\sigma_v$ and $h_{n,2}(\lambda_n) = \lambda_{2n} = corr(u_i, v_i)$ for limit distributions of test statistics in weak IV example

Generalizing, let

$$\{h_n(\lambda) = (h_{n,1}(\lambda), ..., h_{n,J}(\lambda))' \in R^J : n \geq 1\}$$

be a sequence of functions on Λ, where $h_{n,j}(\lambda) \in R\ \forall j = 1, ..., J$.

For any subsequence $\{p_n\}$ of $\{n\}$ and $h \in (R \cup \{\pm\infty\})^J$ denote a sequence $\{\lambda_{p_n} \in \Lambda : n \geq 1\}$ such that $h_{p_n}(\lambda_{p_n}) \to h$ by

$$\lambda_{p_n,h}$$

Define

$H = \{h \in (R \cup \{\pm\infty\})^J :$ there is subsequence $\{p_n\}$ and sequence $\lambda_{p_n,h}\}$.

Theorem, Andrews, Cheng, and Guggenberger (2011)

Assume that under any sequence $\lambda_{p_n,h}$

$$RP_{p_n}(\lambda_{p_n,h}) \to RP(h)$$

for some $RP(h) \in [0,1]$. Then:

$$AsySz = \sup_{h \in H} RP(h).$$

Proof. i) Let $h \in H$. To show $AsySz \geq RP(h)$. By definition of H, there is $\lambda_{p_n,h}$. Then

$$\begin{aligned} AsySz &= \limsup_{n \to \infty} \sup_{\lambda \in \Lambda} RP_n(\lambda) \\ &\geq \limsup_{n \to \infty} RP_{p_n}(\lambda_{p_n,h}) \\ &= RP(h) \end{aligned}$$

Proof. (continued)

ii) To show $AsySz \leq \sup_{h \in H} RP(h)$. Let $\{\lambda_n \in \Lambda : n \geq 1\}$ be a sequence such that

$$\limsup_{n \to \infty} RP_n(\lambda_n) = AsySz.$$

Let $\{p_n : n \geq 1\}$ be a subsequence of $\{n\}$ such that $\lim_{n \to \infty} RP_{p_n}(\lambda_{p_n})$ exists and equals $AsySz$ and $h_{p_n}(\lambda_{p_n}) \to h$. Therefore this sequence is of type $\lambda_{p_n,h}$, and thus, by assumption, $RP_{p_n}(\lambda_{p_n}) \to RP(h)$. Because also $RP_{p_n}(\lambda_{p_n}) \to AsySz$, it follows that $AsySz = RP(h)$. $\square$

Specification of λ for subvector Anderson and Rubin test

- Given F let

$$W_F := (E_F Z_i Z_i')^{1/2} \text{ and } U_F := \Omega(\beta_0)^{-1/2}.$$

- Consider a singular value decomposition

$$C_F \Lambda_F B_F'$$

of

$$W_F(\Pi_W \gamma, \Pi_W) U_F$$

- i.e. B_F denote a $p \times p$ orthogonal matrix of eigenvectors of

$$U_F'(\Pi_W \gamma, \Pi_W)' W_F' W_F (\Pi_W \gamma, \Pi_W) U_F$$

and C_F denote a $k \times k$ orthogonal matrix of eigenvectors of

$$W_F(\Pi_W \gamma, \Pi_W) U_F U_F' (\Pi_W \gamma, \Pi_W)' W_F'$$

- Λ_F denotes a $k \times p$ diagonal matrix with singular values $(\tau_{1F}, ..., \tau_{pF})$ on diagonal, ordered nonincreasingly

- Note $\tau_{pF} = 0$

- Define the elements of λ_F to be

$$\begin{aligned}
\lambda_{1,F} &:= (\tau_{1F}, ..., \tau_{pF})' \in R^p, \\
\lambda_{2,F} &:= B_F \in R^{p \times p}, \\
\lambda_{3,F} &:= C_F \in R^{k \times k}, \\
\lambda_{4,F} &:= W_F \in R^{k \times k}, \\
\lambda_{5,F} &:= U_F \in R^{p \times p}, \\
\lambda_{6,F} &:= F, \\
\lambda_F &:= (\lambda_{1,F}, ..., \lambda_{9,F}).
\end{aligned}$$

- A sequence $\lambda_{n,h}$ denotes a sequence λ_{F_n} such that $(n^{1/2}\lambda_{1,F_n}, ..., \lambda_{5,F_n}) \rightarrow h = (h_1, ..., h_5)$

- Let $q = q_h \in \{0, ..., p-1\}$ be such that

$$h_{1,j} = \infty \text{ for } 1 \leq j \leq q_h \text{ and } h_{1,j} < \infty \text{ for } q_h + 1 \leq j \leq p - 1$$

- Roughly speaking, need to compute asy null rej probs under seq's with (i) strong ident'n,(ii) semi-strong ident'n, (iii) std weak ident'n (all parameters weakly ident'd) & (iv) nonstd weak ident'n

- **strong identification:** $\lim_{n\to\infty} \tau_{m_W,F_n} > 0$

- **semi-strong ident'n:** $\lim_{n\to\infty} \tau_{m_W,F_n} = 0$ & $\lim_{n\to\infty} n^{1/2}\tau_{m_W,F_n} = \infty$

- **weak ident'n:** $\lim_{n\to\infty} n^{1/2}\tau_{m_W,F_n} < \infty$

  - **standard** (of all parameters): $\lim_{n\to\infty} n^{1/2}\tau_{1,F_n} < \infty$ as in Staiger & Stock (1997)

  - **nonstandard:** $\lim_{n\to\infty} n^{1/2}\tau_{m_W,F_n} < \infty$ & $\lim_{n\to\infty} n^{1/2}\tau_{1,F_n} = \infty$ includes some weakly/some strongly ident'd parameters, as in Stock & Wright (2000); also includes **joint weak ident'n**

Andrews and Guggenberger (2014): Limit distribution of eigenvalues of quadratic forms

- Consider a singular value decomposition $C_F \Lambda_F B_F'$ of $W_F D_F U_F$

- Define $\lambda_F, h, \lambda_{n,h}$... as above

Let $\widehat{\kappa}_{jn}$ $\forall j = 1, ..., p$ denote jth eigenval of

$$n\widehat{U}_n' \widehat{D}_n' \widehat{W}_n' \widehat{W}_n \widehat{D}_n \widehat{U}_n,$$

where under $\lambda_{n,h}$

$$\begin{aligned} n^{1/2}(\widehat{D}_n - D_{F_n}) &\rightarrow {}_d\overline{D}_h \in R^{k\times p}, \\ \widehat{W}_n - W_{F_n} &\rightarrow {}_p 0^{k\times k}, \\ \widehat{U}_n - U_{F_n} &\rightarrow {}_p 0^{p\times p}, \\ W_{F_n} &\rightarrow h_4,\ U_{F_n} \rightarrow h_5 \end{aligned}$$

with h_4, h_5 nonsingular

Theorem (AG, 2014): under $\{\lambda_{n,h} : n \geq 1\}$,

(a) $\widehat{\kappa}_{jn} \rightarrow_p \infty$ for all $j \leq q$

(b) vector of smallest $p-q$ eigenvals of $n\widehat{U}_n'\widehat{D}_n'\widehat{W}_n'\widehat{W}_n\widehat{D}_n\widehat{U}_n$, i.e., $(\widehat{\kappa}_{(q+1)n}, ..., \widehat{\kappa}_{pn})'$, converges in dist'n to $p-q$ vector of eigenvals of random matrix $M(h, \overline{D}_h) \in R^{(p-q)\times(p-q)}$

- complicated proof;
  - eigenvalues can diverge at any rate or converge to any number
  - can become close to each other or close to 0 as $n \to \infty$

- We apply this result with

$$\begin{aligned} W_F &= (E_F Z_i Z_i')^{1/2}, \widehat{W}_n = (n^{-1} \sum Z_i Z_i')^{1/2}, \\ U_F &= \Omega(\beta_0)^{-1/2}, \widehat{U}_n = \left(\frac{\overline{Y}' M_Z \overline{Y}}{n-k} \right)^{-1/2}, \\ D_F &= (\Pi_W \gamma, \Pi_W), \widehat{D}_n = (Z'Z)^{-1} Z' \overline{Y} \end{aligned}$$

to obtain the joint limiting distribution of all eigenvalues

Joint asymptotic dist'n of eigenvalues

- Recall: test statistic and critical value are functions of $p = 1 + m_W$ roots of

$$\left|\widehat{\kappa} I_{1+m_W} - (\frac{\overline{Y}' M_Z \overline{Y}}{n-k})^{-1/2} (\overline{Y}' P_Z \overline{Y}) (\frac{\overline{Y}' M_Z \overline{Y}}{n-k})^{-1/2}\right| = 0$$

- To obtain joint limiting distribution of eigenvalues, we use general result in Andrews and Guggenberger (2014) about joint limiting distribution of eigenvalues of quadratic forms

Results:

- the joint limit depends only on localization parameters $h_{1,1}, ..., h_{1,m_W}$

- asymptotic cases replicate finite sample, normal, fixed IV, known variance matrix setup

- together with above proposition, correct asymptotic size then follows from correct finite sample size