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

Patrik Guggenberger
EUI/Pennsylvania State University

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

Toulouse - online!

March 2020

Overview

- **Robust, powerful, computationally fast** 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 inference and estimation (Phillips, 1989; Dufour, 1997; Staiger and Stock, 1997): classical tests overreject true null hypothesis; estimators are biased.

- FIRST assume **cond homoskedasticity** (Guggenberger, Mavroeidis, and Kleibergen, 2019 QE; GKM from now on)
  - THEN: relax to general **Kronecker-Product** structure (KPS)
  - FINALLY: allow for arbitrary forms of **cond heteroskedasticity**

- Large literature on "robust inference" for the **full** parameter vector (Kleibergen, 2002, Moreira, 2003/9)

  Here: Consider robust **subvector** inference in the linear IV model

  Relevant!

- Regarding **power**

Cond Homoskedastic case

We focus on the **Anderson and Rubin (AR, 1949) subvector test statistic**

- **"History of critical values":**

- Projection of full vector 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" (largest quantile) occurs under strong identification

- In GKM (2019) consider a **data-dependent critical value** that adapts to strength of identification

- GKM has uniformly **higher power** than method in GKMC; Implication: Test in GKMC is "inadmissible"

Cond Heteroskedastic case: use KPS AR subvector test or fully robust test depending on outcome of a "pretest" for KPS; suggested test has non-smaller power than certain other robust tests suggested in recent literature

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

- Two related projects: namely GKM (2020) about tests for KPS and GMZ (2020) about optimality properties of a new CLR test

Presentation

- Introduction: ✓

- finite sample case (cond homoskedasticity)

  a) $m_W = 1$: motivation, correct size, uniform power improvement over GKMC

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

- asymptotic case:

  a) cond homoskedasticity

b) general Kronecker-Product structure

c) general case (arbitrary forms of cond heteroskedasticity)

Model and Objective (finite sample, cond homosk. case)

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

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

- **Reduced form:**

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

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

2. 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.

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

Full vector inference

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

- Robust methods: e.g. 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 (2019), and Chernozhukov, Hansen, and Jansson (2009)

- Also: Guggenberger, Mavroeidis, Zhang (2020, working paper); derives optimality results of a new CLR test in a linear IV model with KPS

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), Zhu (2015), Andrews (2017), McCloskey (2018), 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: e.g. Andrews and Cheng (2012), Andrews, I. and Mikusheva (2016), Andrews (2017), Han and McCloskey (2019)

- Models defined by moment inequalities: e.g. Gafarov (2016), Bugni, Canay, and Shi (2017), and Kaido, Molinari, and Stoye (2019)

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 $i = 1, ..., p = 1 + m_W$ are 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

- 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 $k \times p$ matrix.

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

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

which measures the strength of the IVs

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

  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

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

- 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 $\Theta_W' \Theta_W$,

$$\kappa_i, \text{ say } i = 1, \ldots, m_W.$$

- When $m_W = 1$, $\kappa_1 = \Theta_W' \Theta_W$ is a measure of strength of identification of γ; one dimensional nuisance parameter

Theorem: (Perlman and Olkin, 1980). 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.

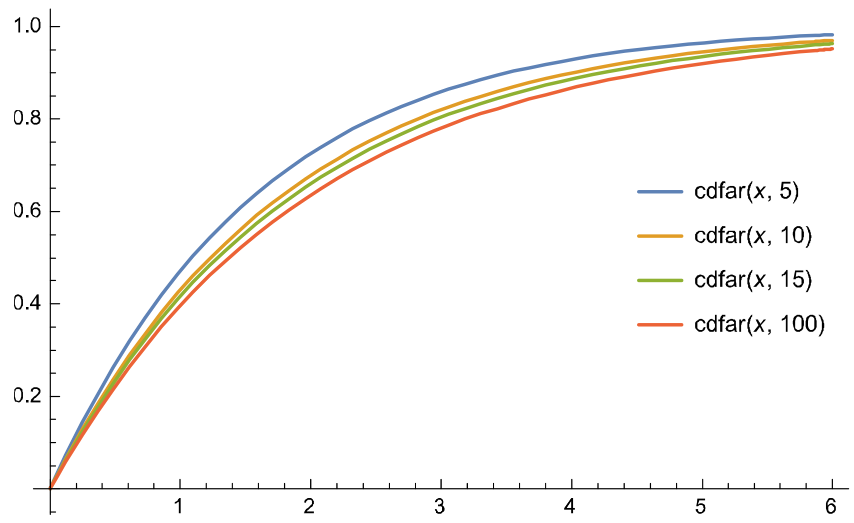


Figure 1: The cdf of the subset AR statistic with $k = 3$ instruments, for different values of $\kappa_1 = 5, 10, 15, 100$

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

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

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.

- Analytical formula for g; confluent hypergeometric function.

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

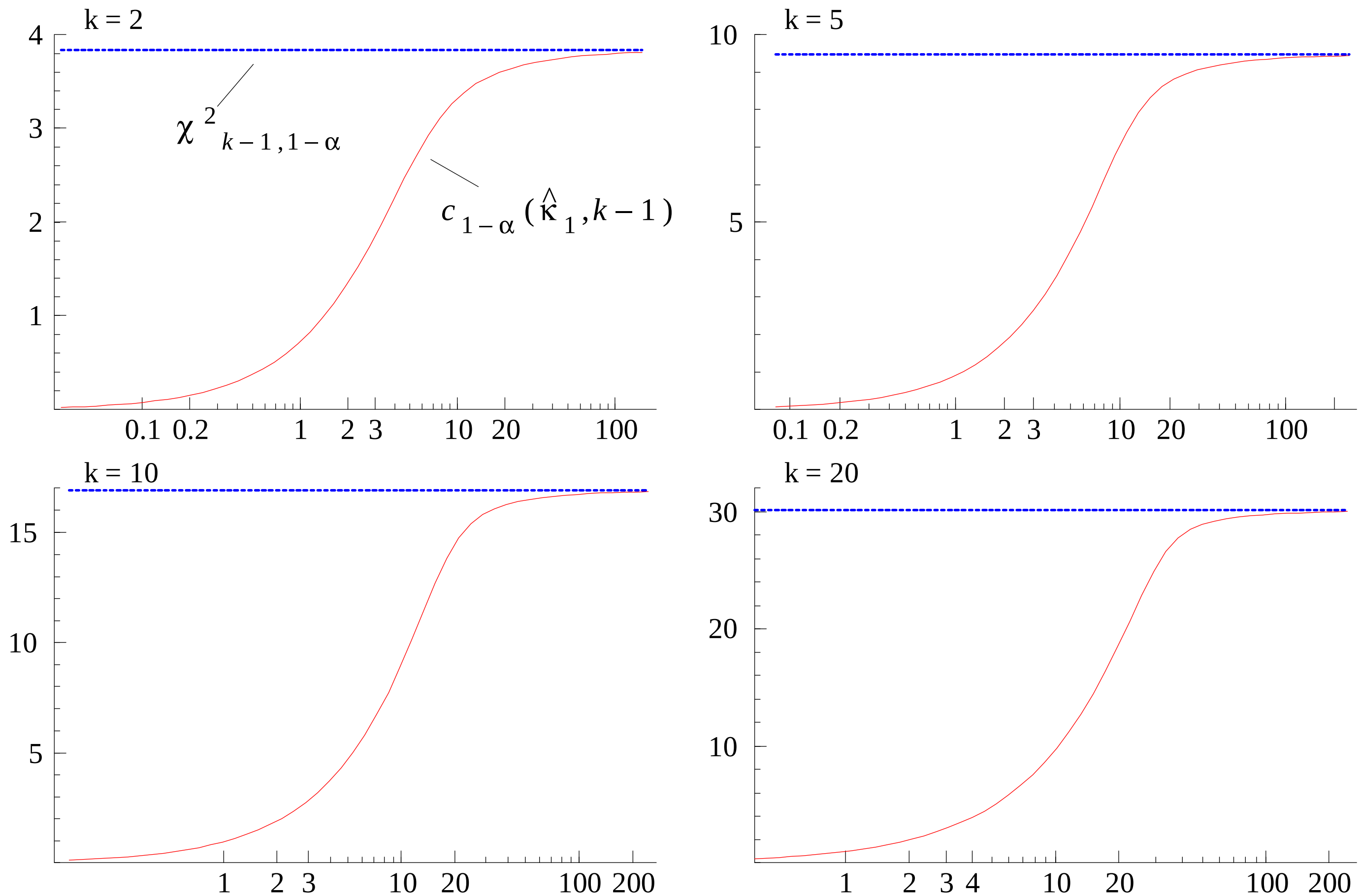
- Proof partly based on simulations; Verified for e.g. $\alpha \in \{1\%, 5\%, 10\%\}$ and $k - m_W \in \{1, ..., 20\}$.

- **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**

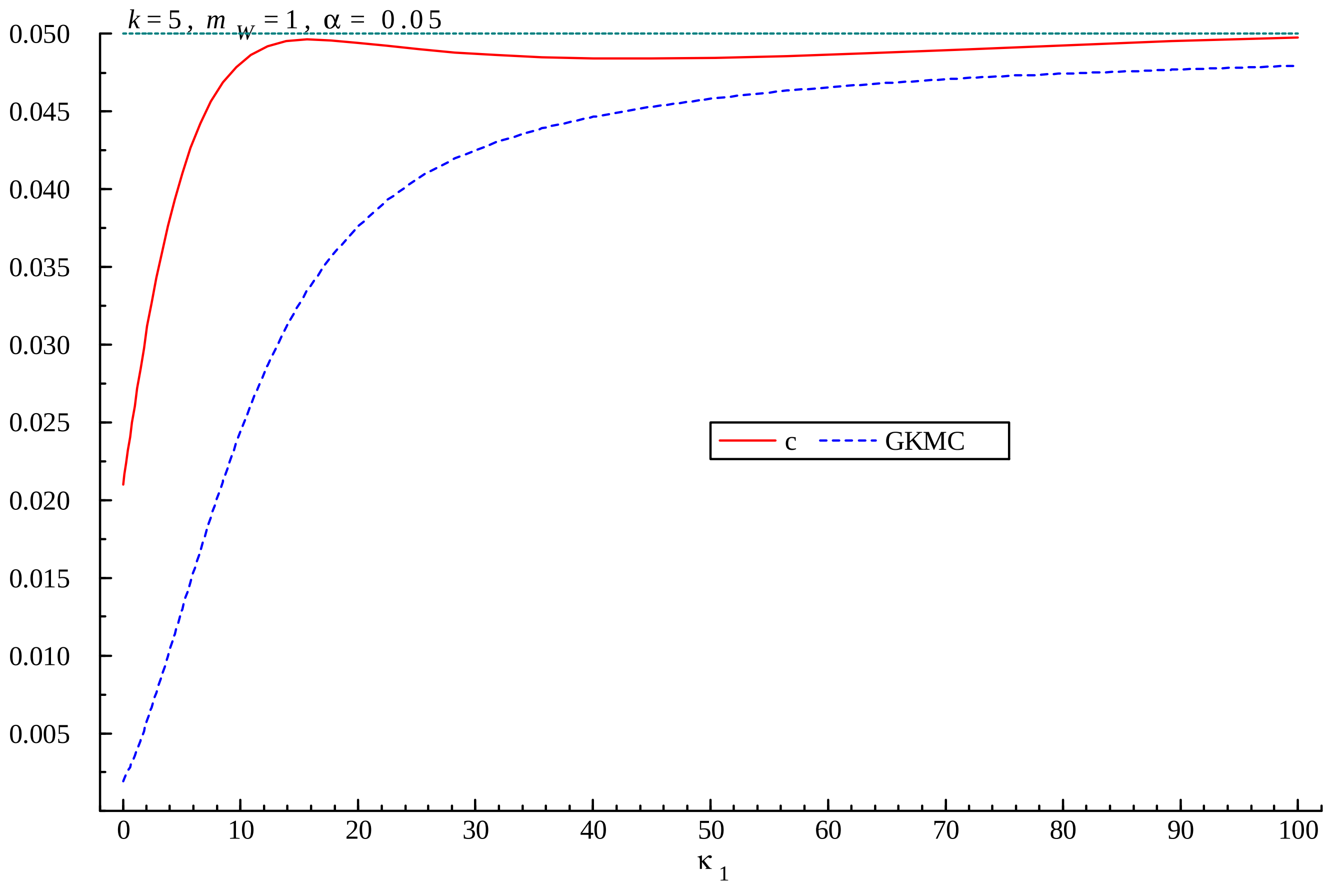


Critical value function $c_{1-\alpha}(\widehat{\kappa}_1, k-1)$ for $\alpha = 0.05$.

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

Asymptotic case: a) cond homoskedasticity (GKM)

- Define **parameter space** $\mathcal{F}_{HOM}$ 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).

$$\begin{aligned}
&\mathcal{F}_{HOM} = \{(\gamma, \Pi_W, \Pi_Y, F)\ s.t.\\
&\gamma \in R^{m_W}, \Pi_W \in R^{k\times m_W}, \Pi_Y \in R^{k\times m_Y},\\
&E_F(||T_i||^{2+\delta}) \leq M,\ \text{for } T_i \in \{vec(Z_iU_i), Z_i, U_i\},\\
&E_F(Z_i(\varepsilon_i, V'_{Wi}, V'_{Yi})) = 0,\\
&E_F(vec(Z_iU'_i)(vec(Z_iU'_i))') = (E_F(U_iU'_i) \otimes E_F(Z_iZ'_i)),\\
&\kappa_{\min}(A) \geq \delta \text{ for } A \in \{E_F(Z_iZ'_i), E_F(U_iU'_i)\}\}
\end{aligned}$$

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

- This particular KPS is slightly less restrictive than cond homoskedasticity

$$E_F(U_iU'_i|Z_i) \text{ does not depend on } Z_i$$

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

- **subvector AR stat** equals smallest solution $\widehat{\kappa}_p$ (slight abuse of notation) 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

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

where $\widehat{\kappa}_1$ is largest solution of above eqn.

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

- Proof based on Andrews and Guggenberger (2017) and Andrews, Cheng, and Guggenberger (2020, JoE, forthcoming)

- As part of the proof relies on the corresponding finite sample result, this proof too is partly 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

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

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

- Covers cond homoskedasticity, but also cases of cond heteroskedasticity; relevant enlargement of parameter space!

Example. Take $(\widetilde{\varepsilon}_i, \widetilde{V}'_{Wi})' \in R^p$ 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

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

has KP structure even though

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

depends on Z_i.

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

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

- 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 $\overline{G}_1 \in R^{p \times p}$, $\overline{G}_2 \in R^{k \times k}$ 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 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, φ_{KP} say, 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 φ_{KP} implemented at nominal size α has asymptotic size, i.e.

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

equal to α.

- In order to make the procedure invariant to transformations $Z_i \to AZ_i$ for nonrandom and nonsingular $A \in R^{k \times k}$ we implement the above procedure replacing Z_i by $(Z'Z)^{-1/2} Z_i$.

- Same disclaimer as above.

Asymptotic case: c) General forms of cond heteroskedasticity

- Let $\mathcal{F}_{Het}$ be the parameter space $\mathcal{F}_{KP}$ above without the condition $R_F = G_1 \otimes G_2$, where

$$R_F := E_F(vec(Z_i U_i')(vec(Z_i U_i'))')$$

- There are tests that have correct asymptotic size for that parameter space, see e.g. Andrews (2017); the objective is to improve on the power.

- **Idea:** use φ_{KP} if data suggests that KPS holds and use a test robust to general forms of cond heteroskedasticity, φ_{Rob} say, otherwise

- Implementation: For some sequence of constants c_n such that

$$c_n \to \infty \text{ and } c_n/n^{1/2} \to 0$$

define the **new suggested test** φ_{c_n} by

$$I(\widehat{K}_n > c_n)\varphi_{Rob} + I(\widehat{K}_n \leq c_n)\varphi_{KP},$$

where

$$\widehat{K}_n := n^{1/2}||\widehat{R}_n^{-1/2}(\widehat{G}_1 \otimes \widehat{G}_2 - \widehat{R}_n)\widehat{R}_n^{-1/2}||_F$$

- Note: we use transformation by $\widehat{R}_n^{-1/2}$ to make procedure invariant against nonsingular transformations of the IV vector

- "Pretesting" is akin to Andrews and Soares (2010); and used in many papers since

- **Asymptotic size?**

- Assume φ_{Rob} is a test has has correct asymptotic size for the parameter space $\mathcal{F}_{Het}$ and satisfies $P_{\lambda_n}(\varphi_{Rob} \leq \varphi_{KP}) \rightarrow 1$ where λ_n is a "local to KPS" sequence

- **Under the above assumptions, φ_{c_n} has correct asymptotic size**

- **Reasons:**

  – Under sequences ("far from KPS") for which

$$n^{1/2} \min_{\overline{G}_1, \overline{G}_2} ||R_{F_n}^{-1/2}(\overline{G}_1 \otimes \overline{G}_2 - R_{F_n})R_{F_n}^{-1/2}|| \rightarrow \infty$$

one can show that $I(\widehat{K}_n > c_n) = 1$ wpa1. Thus φ_{Rob} is chosen wpa1 which has limiting null rejection probability bounded by α

– Under sequences ("local to KPS") for which

$$n^{1/2} \min_{\overline{G}_1, \overline{G}_2} ||R_{F_n}^{-1/2}(\overline{G}_1 \otimes \overline{G}_2 - R_{F_n}) R_{F_n}^{-1/2}|| = O(1)$$

φ_{KP} has limiting null rejection probability bounded by α and given $P_{\lambda_n}(\varphi_{Rob} \leq \varphi_{KP}) \to 1$ the same is true for φ_{c_n}

- We take φ_{Rob} as a certain implementation of an AR/AR type test in Andrews (2017)

- We currently do finite sample experiments to determine good choices for c_n and compare power/size of new test to the ones in Andrews (2017).

- **Alternatively,** one could implement the test by GKM (2020) for KPS in the first stage.

- The role of c_n would then be played by $\beta = \beta_n$, the pretest nominal size.

- We can let $\beta_n = c/n^{1/2}$ and still prove correct asymptotic size of the two step procedure.

- Currently do finite sample experiments to determine which "pretest" performs better

  THE END

Proof of correct size in finite sample case and general m_W

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

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

$\tilde{\Xi}_{11}'\tilde{\Xi}_{11}$ 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

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

by inclusion principle, and thus

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

where first inequality follows from (1) and last inequality from correct size for $m_W = 1$ (by conditioning 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

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

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

- Under H_0', $(\hat{\kappa}_1, ..., \hat{\kappa}_p)$ are distributed as eigenvalues of Wishart $W(k, I_p, M'M)$ 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 $\rho(A) = p$.

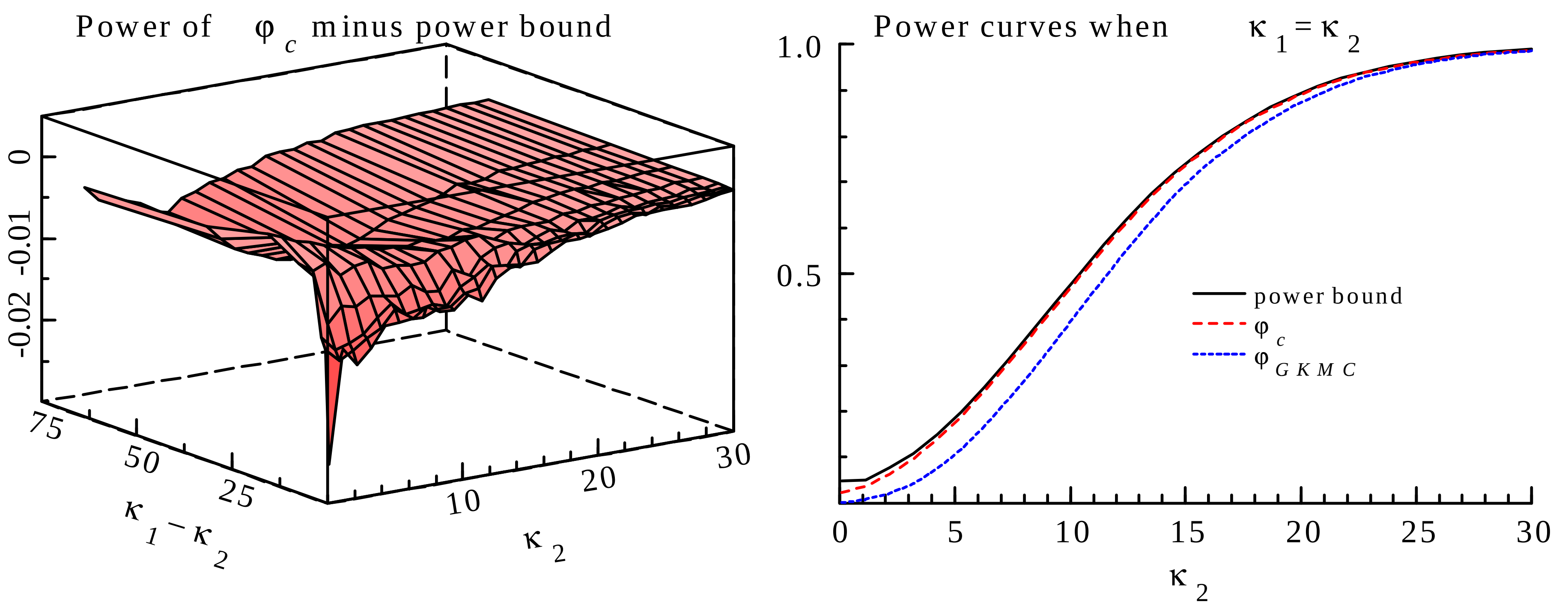
- We use this insight to derive a power envelope for tests of H_0' of the form $\varphi(\hat{\kappa}_1, ..., \hat{\kappa}_p)$.

Power bounds

- Consider only the case $m_W = 1$.

- Can write hyp equivalently as: $H_0' : \kappa_2 = 0,\ \kappa_1 \geq \kappa_2$ 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 no scope at all...:

e.g. **Refinement:** For the case $k = \mathbf{5}$, $m_W = \mathbf{1}$, and $\alpha = \mathbf{5\%}$, let φ_{adj} be the test that uses the critical values in Table above where the smallest 8 critical values are divided by 5. Then this test still has correct size.

Details on proof of asymptotic size result

- We use results in Andrews, Cheng, and Guggenberger (2019, JoE, forthcoming) that show that is enough to verify that the limiting null rejection probability of the test is bounded by α under certain drifting sequences $\lambda_{n,h}$

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 $(n^{1/2}\lambda_{1,F_n}, ..., \lambda_{5,F_n}) \to h = (h_1, ..., h_5)$

- 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 (2017): 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 $\lambda_{n,h}$

$$\begin{aligned} n^{1/2}(\widehat{D}_n - D_{F_n}) &\rightarrow {}_d\overline{D}_h \in R^{k\times p}, \\ \widehat{W}_n - W_{F_n} &\rightarrow {}_p0^{k\times k}, \\ \widehat{U}_n - U_{F_n} &\rightarrow {}_p0^{p\times p}, \\ W_{F_n} &\rightarrow h_4,\ U_{F_n} \rightarrow h_5 \end{aligned}$$

with h_4, h_5 nonsingular

Theorem (AG): 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 AG 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

Statistical Lemma:

Recall that $\Xi \sim N\left(\mathcal{M}, I_{k(m_W+1)}\right)$, with M nonstochastic and $\rho(\mathcal{M}) \leq m_W$ under the null.

Partition Ξ as

$$\Xi = \begin{pmatrix} \Xi_{11} & \Xi_{12} \\ \Xi_{21} & \Xi_{22} \end{pmatrix},$$

where Ξ_{11} is $(k - m_W + 1) \times 2$, Ξ_{12} is $(k - m_W + 1) \times (m_W - 1)$, Ξ_{21} is $(m_W - 1) \times 2$, and Ξ_{22} is $(m_W - 1) \times (m_W - 1)$.

Partition M conformably with Ξ. Let μ_i, $i = 1, ..., m_W$, denote the possibly nonzero singular values of M. We can set

$$\mathcal{M} = \begin{pmatrix} \mathcal{M}_{11} & 0^{k-m_W+1 \times m_W-1} \\ 0^{m_W-1 \times 2} & \mathcal{M}_{22} \end{pmatrix},$$

where

$$\mathcal{M}_{11} := \begin{pmatrix} 0^{k-m_W \times 1} & 0^{k-m_W \times 1} \\ 0 & \mu_{m_W} \end{pmatrix}, \text{ and } \mathcal{M}_{22} := diag\left(\mu_1, ...\mu_{m_W-1}\right).$$

Finally, let

$$O := \begin{pmatrix} \left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} & \Xi_{21}'\Xi_{22}^{-1\prime}\left(I_{m_W-1} + \Xi_{22}^{-1}\Xi_{21}\Xi_{21}'\Xi_{2 \\ -\Xi_{22}^{-1}\Xi_{21}\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} & \left(I_{m_W-1} + \Xi_{22}^{-1}\Xi_{21}\Xi_{21}'\Xi_{22}^{-1\prime}\right)^{-}$$

and

$$\begin{aligned}\tilde{M}_{11} &:= \left(\mathcal{M}_{11} - \mathcal{M}_{12}\Xi_{22}^{-1}\Xi_{21}\right)\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2} \\ &= M_{11}\left(I_2 + \Xi_{21}'\Xi_{22}^{-1\prime}\Xi_{22}^{-1}\Xi_{21}\right)^{-1/2}.\end{aligned}$$

Theorem 1 *Suppose that Assumption A holds with* $m_W > 1$*. Denote by* $\tilde{\Xi}_{11} \in \Re^{k-m_W+1\times 2}$ *the upper left submatrix of* $\widetilde{\Xi} := \Xi O \in \Re^{k\times p}$*. Then,* *under the null hypothesis* $H_0 : \beta = \beta_0$

$$\tilde{\Xi}_{11}'\tilde{\Xi}_{11}|O \sim \mathcal{W}_2\left(k - m_W + 1, I_2, \tilde{\mathcal{M}}_{11}'\tilde{\mathcal{M}}_{11}\right),$$

where $\rho(\tilde{M}_{11}'\tilde{M}_{11}) \leq 1$.