Inference with Many Weak Instruments

Anna Mikusheva[1] Liyang Sun[2]

[1,2]MIT

May, 2020

Introduction

- We are interested in Instrumental Variables models where the instruments are:
  1. Many
  2. Potentially Weak
- **Example 1**: Angrist and Krueger (1991) interacts quarter of birth with:
  - year of birth (30)
  - year and state of birth (180)
  - year and state of birth, and their interactions (1530)
- **Example 2**: 'Judges design': assignment to judge is an instrument. Sample size (number of cases) is roughly proportional to the number of judges. Maestas et al. (2013), Sampat and Williams (2015), Dobbie et al. (2018)

- **Example 3**: Factor Pricing

$$Er_{it} = \lambda\beta_i, \text{ where } \beta_i = \Sigma_F^{-1} cov(F_t, r_{it}),$$

where λ is the risk premia, while β_i is risk exposure.

- Common estimation procedure is Fama-MacBeth:
  1. For each i estimate $\widehat{\beta}_i$ by OLS of r_{it} on F_t.
  2. Estimate $\widehat{\lambda}$ by OLS of $\frac{1}{T}\sum_t r_{it}$ on $\widehat{\beta}_i$.
- The problem can be re-formulated as TSLS with the number of instruments proportional to the number of stocks/portfolios.
- In applications many factors are only weakly correlated with all the returns, that results in weak IV.

IV with Many Weak Instruments

- Linear IV model with one endogenous variable, (a small number of) exogenous variables partialled out

$$\begin{cases} Y_i = \beta X_i + e_i, \\ X_i = \pi' Z_i + v_i, \end{cases}$$

  where $Z_i \in \mathbb{R}^K$ is conditioned upon.
- Weak instruments: π is close to zero.
- Many instruments $K \to \infty$ as $N \to \infty$ (up to $K = \lambda N$).
- The errors are heteroscedastic (but independent).
- Many of our results hold for more general non-linear first stage $X_i = \Pi_i + v_i$ with $\Pi_i = E[X_i | Z_i]$

IV with Many Weak Instruments

- When K is fixed, weak identification is defined as $\pi = \frac{C}{\sqrt{N}}$, no consistent estimator exists.
- If $K \to \infty$ and *each* instrument is weak, then in totality there is a lot of information $\pi' Z' Z \pi \asymp K$.
  - Literature on "many weak" mostly concentrated on estimation: TSLS is inconsistent, but consistent estimation is possible.
  - Consistent estimators under some conditions are LIML (Newey, mimeo, 2004), JIVE (Chao et al, ET, 2012), Jackknife LIML (Hausman et al, QE, 2012), CUE (Newey and Windmeijer, ECMA, 2009).
- Where is the knife-edge, below which there is no consistent estimator?

Goals of This Project

1. Define weak identification when $K \to \infty$.
2. Find tests robust to weak identification and to heteroscedasticity.
3. Create a pre-test for weak instruments (à la first stage F).
4. Bonus: power of robust tests. Optimal tests?

Overview

1. What is Weak Identification?
2. Weak IV- Robust Testing
3. Pre-test for weak identification
4. Other tests: Power considerations

What is Weak Identification?

- Model again:
$$\begin{cases} Y_i = \beta X_i + e_i, \\ X_i = \pi' Z_i + v_i, \end{cases}$$
where $Z_i \in \mathbb{R}^K$
- Our answer: if the direction of π is unknown, then $\frac{\pi' Z' Z \pi}{\sqrt{K}} \asymp const$ is the knife-edge case for consistency.
- Negative statement: in the best possible scenario – only π and β are unknown, if $\frac{\pi' Z' Z \pi}{\sqrt{K}} \asymp const$, there exists no asymptotically consistent robust test.
- Positive statement: we construct robust tests that are consistent when $\frac{\pi' Z' Z \pi}{\sqrt{K}} \to \infty$.

What is Weak Identification?

- We consider gaussian model with homoscedasticity and known covariance matrix for the reduced-form errors.
- Let Ψ be a class of tests for $H_0 : \beta = \beta_0$ that has correct size uniformly over nuisance parameter π.
- Statement: for any fixed $\beta^* \neq \beta_0$

$$\limsup_{n\to\infty} \max_{\psi\in\Psi} \left(\min_{\pi:\frac{\pi' Z' Z\pi}{\sqrt{K}}=const} E_{\pi,\beta^*}\psi \right) < 1.$$

What is Weak Identification?

$$\lim_{n\to\infty} \sup_{\psi\in\Psi} \max \left(\min_{\pi:\frac{\pi' Z' Z\pi}{\sqrt{K}}=const} E_{\pi,\beta^*}\psi \right) < 1.$$

- No fixed alternative β^* can be consistently distinguished from β_0 if the direction of π is unknown and $\frac{\pi' Z' Z\pi}{\sqrt{K}} = const$.
- If K is fixed – it is the usual weak iv embedding.
- When $K \to \infty$ this result is new
  - Stronger statement than Chao and Swanson (2005) where they prove this is the knife-edge case for B2SLS and LIML.

Overview

Weak IV-Robust Tests: Refresher, Fixed K

$H_0 : \beta = \beta_0$. Define $e(\beta_0) = Y - \beta_0 X$.

- AR (Anderson-Rubin) statistics:

$$e(\beta_0)' Z \Sigma^{-1} Z' e(\beta_0) \sim \chi^2_K.$$

  Σ is a covariance matrix of $e'Z$ or a good estimate of it.

  - Size is robust to weak IV.
  - Loss of power if identification is strong.

What Changes with $K \to \infty$?

- Think about AR (Anderson-Rubin) statistics:

$$e(\beta_0)' Z \Sigma^{-1} Z' e(\beta_0) \sim \chi^2_K.$$

  - CLT may not work well when K is growing.
  - χ^2_K is a diverging distribution.
  - Σ is a covariance matrix of $e'Z$ – it is $K \times K$, not well estimated.

What Changes with $K \to \infty$?

- Even under homoscedasticity (need to estimate a scalar σ_e^2), variance estimation mistakes change the limit distribution for $K = \lambda N$.
- AR statistics for fixed K

$$AR(\beta_0) = \frac{e(\beta_0)' P_Z e(\beta_0)}{e(\beta_0)' M_Z e(\beta_0)/(N-K)} \Rightarrow \chi^2_K,$$

$$\frac{\chi^2_K - K}{\sqrt{2K}} \Rightarrow N(0,1) \text{ as } K \to \infty.$$

- However, if $K = \lambda N$, Anatolyev and Gospodinov (ET, 2011) show under the null

$$\frac{AR(\beta_0) - K}{\sqrt{2K}} \Rightarrow N\left(0, \frac{1}{1-\lambda}\right).$$

Our Proposed Ideas

- We wish to construct *weak identification* robust test for *heteroscedastic* model.
- We have several concerns:
  - Estimation of Σ, covariance matrix for $e'Z$, is impossible.
  - Need to re-center statistics such as AR.
  - CLT may not work.
  - Estimation of variances.
- Ideas:
  - Use default Σ $(= (Z'Z)^{-1}$ as for homoscedastic case).
  - Re-center and re-normalize the statistics.
  - Use CLT for quadratic forms.
  - Use cross-fit variance estimation.

Our Proposed Ideas

- Proposed AR uses default homoscedastic weighting $(Z'Z)^{-1}$ and is proportional to $e(\beta_0)'P_Z e(\beta_0) = \sum_{i,j} P_{ij} e_i(\beta_0) e_j(\beta_0)$.
- Re-centering: in heteroscedastic case $Ee'P_Z e = \sum_{i=1}^{N} P_{ii} Ee_i^2$.
- The obvious way to re-center:

$$e(\beta_0)'P_Z e(\beta_0) - \sum_{i=1}^{N} P_{ii} e_i^2(\beta_0) = \sum_{i \neq j} P_{ij} e_i(\beta_0) e_j(\beta_0).$$

- This is the leave-one-out (jackknife) approach! See Angrist et al (JAE, 1999), Chao et al (ET, 2012) and Hausman et al (QE, 2012).
- Similar proposal of AR test- Crudu, Mellace and Sandor (2020)

Central Limit Theorem for Quadratic Forms

Theorem 1 (Chao et al, 2012).

Assume that P is $N \times N$ symmetric idempotent matrix of rank K with $K \to \infty$ as $N \to \infty$, and $P_{ii} < C < 1$. Let $(U_1, e_1), ..., (U_N, e_N)$ be independent, mean-zero with bounded fourth moments. Then

$$\frac{1}{B_N \sqrt{K}} \sum_{i \neq j} U_i P_{ij} e_j \Rightarrow N(0,1)$$

here

$$B_N^2 = \frac{1}{K} \sum_{i \neq j} P_{ij}^2 (E[U_i U_i'] E[e_j^2] + E[U_i e_i] E[U_j' e_j]).$$

Alternative CLT is in Sølvsten (2017).

AR: Variance Estimation

- The infeasible leave-one-out AR is

$$AR_0(\beta_0) = \frac{1}{\sqrt{K\Phi_0}} \sum_{i \neq j} e_i(\beta_0) P_{ij} e_j(\beta_0),$$

  for $\Phi_0 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2$.
- Rejects for large values of AR.
- Need to estimate the variance.

AR: Variance Estimation

$$\Phi_0 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2$$

- Idea 1: $\widehat{\sigma}_i^2 = e_i^2(\beta_0)$. Crudu, Mellace and Sandor (2020). It gives correct size, robust toward heteroscedasticity, but power is problematic at distant alternatives.
- Idea 2: Residualizing $e(\beta_0)$ with respect to Z $(M = I - P_Z)$

$$\widehat{\sigma}_i^2 = \left(M_i \mathbf{e}(\beta_0)\right)^2, \; E[\widehat{\sigma}_i^2] \neq \sigma_i^2$$

  estimates the wrong quantity.
- Estimation error is a large part of the residual; squaring messes up estimation of variance.

AR: Variance Estimation

$$\Phi_0 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2$$

- Idea 3 is to estimate $\sigma_i^2 = E[e_i^2]$ by a "cross-fit" variance estimator (Newey and Robins (2018), Kline et al (2018)):

$$\widehat{\sigma}_i^2 = \frac{1}{1 - P_{ii}} e_i(\beta_0) M_i e(\beta_0).$$

- Challenge is that we need a double sum:

$$E\left[(e_i M_i e)(e_j M_j e)\right] = (M_{ii} M_{jj} + M_{ij}^2) \sigma_i^2 \sigma_j^2.$$

- Our suggested estimator:

$$\widehat{\Phi} = \frac{2}{K} \sum_{i \neq j} \frac{P_{ij}^2}{M_{ii} M_{jj} + M_{ij}^2} \left[e_i(\beta_0) M_i e(\beta_0)\right] \left[e_j(\beta_0) M_j e(\beta_0)\right].$$

Feasible AR

Feasible $AR(\beta_0) = \frac{1}{\sqrt{K\widehat{\Phi}}} \sum_{i \neq j} e_i(\beta_0) P_{ij} e_j(\beta_0)$ uses

$$\widehat{\Phi} = \frac{2}{K} \sum_{i \neq j} \frac{P_{ij}^2}{M_{ii} M_{jj} + M_{ij}^2} \left[e_i(\beta_0) M_i e(\beta_0) \right] \left[e_j(\beta_0) M_j e(\beta_0) \right].$$

- $\widehat{\Phi}$ is consistent for Φ_0 under the null.
- Feasible test achieves the same local power as the infeasible AR.
- If non-linear first stage ($X_i = \Pi_i + v_i$, $\Pi_i = E[X_i | Z_i]$) for consistency of $\widehat{\Phi}$ we need additional assumption $\Pi' M \Pi \leq \frac{C}{K} \Pi' \Pi$
- Feasible AR is consistent for distant alternatives (though $\widehat{\Phi}$ is not consistent there).

Power of AR

- The leave-one-out AR is

$$AR(\beta_0) = \frac{1}{\sqrt{K\widehat{\Phi}}} \sum_{i \neq j} e_i(\beta_0) P_{ij} e_j(\beta_0).$$

- Under the alternative $\beta = \beta_0 + \Delta$, we have $e_i(\beta_0) = \Delta \Pi_i + \eta_i$.
- Define a leave-one-out concentration parameter:

$$\mu^2 = \sum_{i \neq j} P_{ij} \Pi_i \Pi_j.$$

- Power statement: uniformly over set of local alternative and (reasonably restricted) set of μ^2 :

$$AR(\beta_0) \Rightarrow \Delta^2 \frac{\mu^2}{\sqrt{K\Phi_0}} + \mathcal{N}(0,1).$$

- As soon as $\mu^2/\sqrt{K} \to \infty$, AR is consistent for fixed alternatives.

Summary

- If $\frac{\pi' Z' Z \pi}{\sqrt{K}} \asymp const$, then we cannot consistently distinguish two values of β. Thus, weak identification.
- We constructed a test (Leave-one-out AR), that is robust to weak identification (and heteroscedasticity). It becomes consistent as soon as $\mu^2/\sqrt{K} \to \infty$.
- The suggested AR is probably not very powerful if identification is strong.
- How can we distinguish empirically if identification is weak?

Overview

F test under fixed K

- To measure the identification strength, it is common practice to conduct a pretest based on the first-stage F statistic.
  - Researchers compare their F statistic to some cut-off (10?) to gauge the degree of weak identification (defined as Wald test has actual size up to 10% for a nominal 5% test).
  - If the F statistic is greater than the cut-off, report the usual TSLS-Wald confidence set. Otherwise, report a robust confidence set (i.e., AR).
  - The resulting two-step test has size at most 15%.
- Huge problem: valid only under homoscedasticity
- Our goal: create two-step procedure robust to heteroscedasticity and weak identification under $K \to \infty$.

F test under fixed K

- First stage F- pre-test may be conservative for large K for properly selected estimator ($EF = \frac{\pi' Z' Z \pi}{\sigma_v^2 K} + 1$):

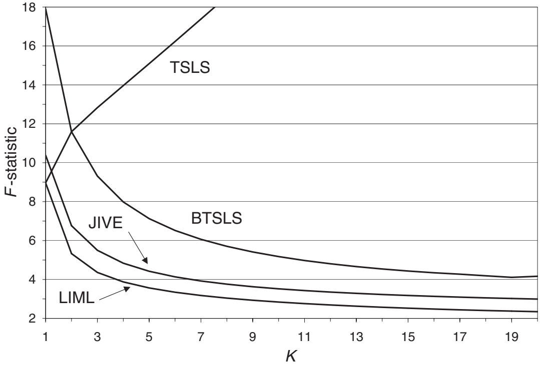


Figure: Stock, Wright, Yogo (JBES, 2002)

Estimation with many instruments

- Different estimators have distinct convergence properties when K is large:
  - TSLS is consistent whenever $\frac{\pi' Z' Z\pi}{K} \to \infty$. (Chao and Swanson (2005), Newey (2004))
  - Under homoscedasticity LIML, BTSLS, JIVE are consistent whenever $\frac{\pi' Z' Z\pi}{\sqrt{K}} \to \infty$. (Newey(2004),Hausman et al (2007))
  - Under heteroscedasticity LIML and BTSLS are consistent whenever $\frac{\pi' Z' Z\pi}{K} \to \infty$, while JIVE is consistent whenever $\frac{\pi' Z' Z\pi}{\sqrt{K}} \to \infty$ (Chao et al (2012))
- First stage F- measures $EF = \frac{\pi' Z' Z\pi}{\sigma_v^2 K} + 1$.

Our pre-test for weak identification

- The goal of a pre-test is to control the size of a test for $H_0 : \beta = \beta_0$.
- If identification seems weak, we will use our new AR test.
- If identification seems strong, we will use the JIVE-Wald test.
- We want asymptotic size of this two step procedure to be not larger than 10%.
- Motivation for pre-test: under strong instruments it gives
  - A simpler procedure
  - A more powerful test

Wald test based on JIVE

- We use what is called JIV2 estimator and Wald statistics (Chao et al (2012)):

$$\widehat{\beta}_{JIVE} = \frac{\sum_i \sum_{j \neq i} P_{ij} Y_i X_j}{\sum_i \sum_{j \neq i} P_{ij} X_i X_j},$$

$$Wald(\beta_0) = \frac{\left(\widehat{\beta}_{JIV} - \beta_0\right)^2}{\widehat{V}},$$

$$\widehat{V} = \frac{\sum_{i=1}^{N} \left(\sum_{j \neq i} P_{ij} X_j\right)^2 \frac{\widehat{e}_i M_i \widehat{e}}{M_{ii}} + \sum_{i=1}^{N} \sum_{j \neq i} \widetilde{P}_{ij}^2 M_i X \widehat{e}_i M_j X \widehat{e}_j}{\left(\sum_{i=1}^{N} \sum_{j \neq i} P_{ij} X_i X_j\right)^2},$$

where $\widehat{e}_i = Y_i - X_i \widehat{\beta}_{JIV}$.

Pre-test for weak identification

- Our pre-test is based on empirical measure:

$$\widetilde{F} = \frac{1}{\sqrt{K}\sqrt{\widehat{\Upsilon}}} \sum_{i=1}^{N} \sum_{j \neq i} P_{ij} X_i X_j,$$

  here $\widehat{\Upsilon} = \frac{2}{K} \sum_i \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii} M_{jj} + M_{ij}^2} X_i M_i X X_j M_j X$ is an estimate of uncertainty in the first stage
- It has signal-to-noise form normalized by $\sqrt{K}$

Pre-test for weak identification

Theorem 2.

If $\Pi' M \Pi \leq \frac{C\Pi'\Pi}{K}$ *and* $\frac{\Pi'\Pi}{K^{2/3}} \to 0$ *as* $N \to \infty$*, then under* $H_0 : \beta = \beta_0$*,*

$$\left(Wald(\beta_0), \widetilde{F}\right) \Rightarrow \left(\frac{\xi^2}{1 - 2\varrho\frac{\xi}{\nu} + \frac{\xi^2}{\nu^2}}, \nu\right),$$

where $\begin{pmatrix} \xi \\ \nu \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\mu^2}{\sqrt{K}\sqrt{\Upsilon}} \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}\right)$ *and* ϱ *is a correlation parameter.*

- Notice that when $\frac{\mu^2}{\sqrt{K}\sqrt{\Upsilon}}$ is large then $Wald(\beta_0) \approx \xi^2 = \chi_1^2$

Pre-test for weak identification

For different values of $\frac{\mu^2}{\sqrt{K}\sqrt{\Upsilon}}$ we may simulate

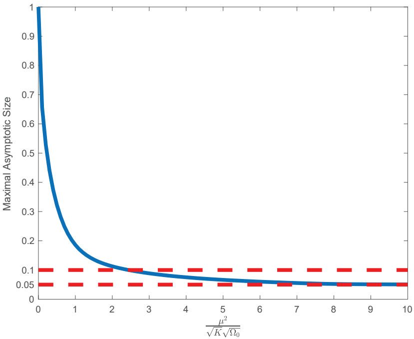
$$\begin{pmatrix} \xi \\ \nu \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \frac{\mu^2}{\sqrt{K}\sqrt{\Upsilon}} \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}\right)$$

and the worst-case asymptotic size of the Wald test

$$\max_{\varrho\in[-1,1]} Pr\left\{\frac{\xi^2}{1-2\varrho\frac{\xi}{\nu}+\frac{\xi^2}{\nu^2}} \geq \chi^2_{1,\alpha}\right\},$$

which is maximized at $\varrho = 1$.

Worst case asymptotic size of JIVE-Wald



- This gives us a definition of (cut-off for) many weak instrument in terms of $\frac{\mu^2}{\sqrt{K}\sqrt{\Upsilon}} < 2.5$, which implies the pre-test: reject many weak instruments if $\widetilde{F} > 4.14$.

Re-visiting Angrist and Krueger (1991)

- Research question: return to education. Y_i is the log weekly wage, X_i is education.
- Instruments: quarter of birth. Justification is related to compulsory education laws:
  - 180 instruments: 30 quarter and year of birth interactions (QOB-YOB) and 150 quarter and state of birth interactions (QOB-POB)
  - 1530 instruments: full interactions among QOB-YOB-POB
- The sample contains 329,509 men born 1930-39 from the 1980 census.
- This paper sparked the weak IV literature. It is a running example for multiple papers.

Re-visiting Angrist and Krueger(1991)

	FF	$\widetilde{F}$	JIVE-Wald	Jackknife AR
180 instruments	2.428	13.422	[0.066,0.132]	[0.008,0.201]
1530 instruments	1.27	6.173	[0.024,0.121]	[-0.047, 0.202]

Table: Angrist and Krueger (1992) Pre-test Results

Notes: Results on pre-tests for weak identification and confidence sets for IV specification underlying Table VII Column (6) of Angrist and Krueger (1991). The confidence set based on jackknife AR is constructed via analytical test inversion.

Overview

Weak IV-Robust Tests: Refresher, Fixed K

- Problem: AR is not efficient if identification is strong
- LM intends to test a "powerful" combination of instruments $e'Z\pi$, but $\widehat{\pi}$ is correlated with $e'Z$.
- KLM: define $\widetilde{\pi} = \widehat{\pi} - AZ'e(\beta_0)$, orthogonalized version of estimator $\widehat{\pi}$:

$$KLM = \frac{\left(e(\beta_0)'Z\widetilde{\pi}\right)^2}{\sigma^2_{KLM}} \sim \chi^2_1.$$

  - Robust size when identification is weak.
  - Efficient if identification is strong.
  - Non-monotonic power if identification is weak.
- CLR is doing a smooth switch between AR and KLM: AR for weaker cases, KLM for stronger.

Infeasible leave-one-out LM

- Idealistic LM is based on linear combination $e'(\beta_0)Z\widehat{\pi} = e'(\beta_0)P_Z X$.
- Leave-one-out gives us $LM^{1/2} \propto \sum_{i\neq j} e_i(\beta_0)P_{ij}X_j$.
- $X_j = Z_j'\pi + v_j$

$$LM^{1/2} \propto \sum_{i\neq j} e_i(\beta_0)P_{ij}v_j + \sum_{i} e_i(\beta_0)\left(\sum_{j\neq i} P_{ij}Z_j'\pi\right).$$

- We apply both Lindeberg's CLT and quadratic CLT.
- We need each term to be asymptotically negligible: $\max_i \pi' Z_i Z_i' \pi/\mu^2 \to 0$.
- We do NOT need KLM-correction.

Feasible LM

- The feasible leave-one-out LM is

$$LM^{1/2}(\beta_0) = \frac{1}{\sqrt{K\widehat{\Psi}}} \sum_{i \neq j} e_i(\beta_0) P_{ij} X_j,$$

  with $N(0,1)$ asymptotic distribution under the null and two-sided rejection.
- Asymptotic variance is somewhat more complicated, though still makes use of the ideas of double-cross-fit.

$$\widehat{\Psi} = \frac{1}{K} \sum_i \frac{e_i M_i e}{M_{ii}} (\sum_{j \neq i} P_{ij} X_j)^2 + \frac{1}{K} \sum_i \sum_{j \neq i} \widetilde{P}_{ij}^2 X_i M_i e X_j M_j e.$$

- Variance estimator is consistent uniformly over LM local alternatives.

Power of LM

- The feasible leave-one-out LM is

$$LM^{1/2}(\beta_0) = \frac{1}{\sqrt{K\widehat{\Psi}}} \sum_{i \neq j} e_i(\beta_0) P_{ij} X_j.$$

- Under the alternative $\beta = \beta_0 + \Delta$, we have $e_i(\beta_0) = Z_i' \pi \Delta + \eta_i$:

$$LM^{1/2} \Rightarrow \Delta \frac{\mu^2}{\sqrt{K\Psi}} + \mathcal{N}(0,1),$$

  uniformly over local alternatives.
- LM test has two-sided rejection region.
- As soon as $\mu^2/\sqrt{K} \to \infty$, LM is consistent for fixed alternatives.

Power trade-off

- Under the alternative $\beta = \beta_0 + \Delta$, we have :

$$LM^{1/2} \Rightarrow \Delta \frac{\mu^2}{\sqrt{K\Psi}} + \mathcal{N}(0,1),$$

$$AR \Rightarrow \Delta^2 \frac{\mu^2}{\sqrt{K\Phi}} + \mathcal{N}(0,1),$$

- When $\frac{\mu^2}{\sqrt{K}} \to \infty$, AR and LM are asymptotically consistent for fixed alternatives β.
- When $\frac{\mu^2}{\sqrt{K}} \to \infty$ but $\frac{\mu^2}{K} \to 0$ local alternatives are:
  - for AR $\{\Delta : \frac{\Delta^2\mu^2}{\sqrt{K}} \leq C\}$ i.e. $|\Delta| \propto \sqrt{\frac{\sqrt{K}}{\mu^2}}$,
  - for LM $\{\Delta : \frac{|\Delta|\mu^2}{\sqrt{K}} \leq C\}$ i.e. $|\Delta| \propto \frac{\sqrt{K}}{\mu^2}$,
  - AR has slower speed of detection.

Conditional Switch Test: CLR

- Use $\widetilde{F}$ for conditioning:

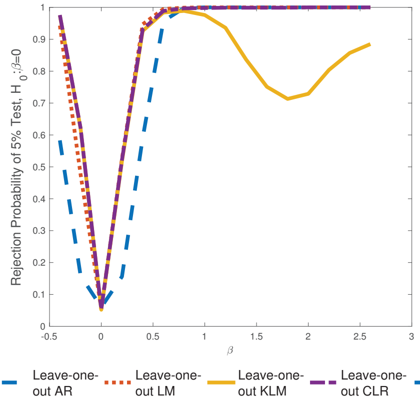
$$\begin{pmatrix} AR(\beta_0) - \Delta^2 \frac{\mu^2}{\sqrt{K\Phi}} \\ LM^{1/2}(\beta_0) - \Delta \frac{\mu^2}{\sqrt{K\Psi}} \\ \widetilde{F} - \frac{\mu^2}{\sqrt{K\Upsilon}} \end{pmatrix} \Rightarrow \mathcal{N}\left(\mathbf{0}, \Sigma\right).$$

- To test $H_0 : \Delta = 0$ against $H_0 : \Delta \neq 0$ we use:

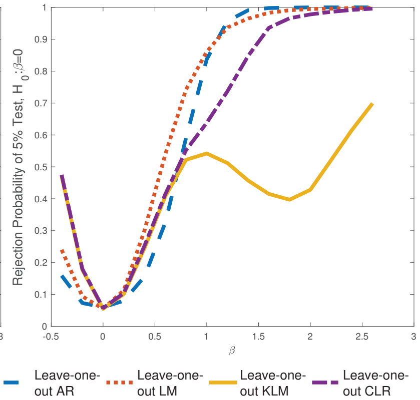
$$LR(AR(\beta_0), LM^{1/2}(\beta_0), \widetilde{F}) = \max_{\Delta,\mu} \ell(\Delta, \frac{\mu^2}{\sqrt{K}}) - \max_{\mu} \ell(0, \frac{\mu^2}{\sqrt{K}}),$$

and simulate critical values conditional on the orthogonalized version of $\widetilde{F}$.

Preliminary Simulation Evidence: switch by CLR

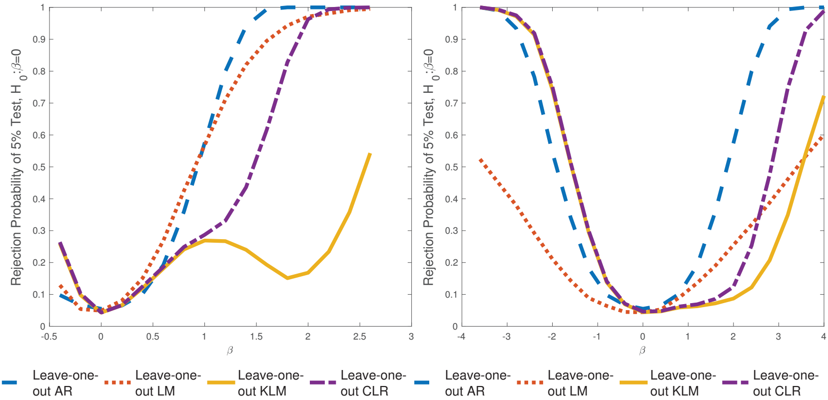


(a) strong normal instruments



(b) weak normal instruments

Preliminary Simulation Evidence: switch by CLR



(c) strong group instruments

(d) weak group instruments

Conclusions

- We found that the knife-edge case for consistency happens when $\frac{\pi' Z' Z \pi}{\sqrt{K}} \asymp const.$
- We introduced AR, LM and CLR tests robust to weak id, heteroscedasticity and many instruments.
- Tests use idea of leave-one-out quadratic forms and cross-fit variance estimation.
- We can create a simple pre-test for weak identification robust to heteroscedasticity when $K \to \infty$.