Inference with Many Weak Instruments

Anna Mikusheva[1] Liyang Sun[2]

[1,2]MIT

Econometrics Session in Honor of Gary Chamberlain

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' (Maestas et al. (2013), Sampat and Williams (2015), Dobbie et al. (2018)): assignment to judge is an instrument. Sample size (number of cases) is roughly proportional to the number of judges.

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.
  - $K \to \infty$ as $N \to \infty$ (up to $K = \lambda N$).
  - The errors are heteroscedastic (but independent).
- When K is fixed, weak identification is defined as $\pi = \frac{C}{\sqrt{N}}$.
- If $K \to \infty$ and *each* instrument is weak, then in totality there is a lot of information measured by $\mu^2 = \pi' Z' Z \pi$.

Goals of This Project

1. Define weak identification when $K \to \infty$ i.e. find the knife-edge, below which there is no consistent estimator.
2. Find tests robust to weak identification and to heteroscedasticity.
3. Create a pre-test for weak instruments (à la first stage F).

Overview

1. What is Weak Identification?
2. Weak IV- Robust Testing
3. Pre-test for weak identification

What is Weak Identification?

- Negative statement: no fixed alternative β^* can be consistently distinguished from β_0 if the direction of π is unknown and $\frac{\mu^2}{\sqrt{K}} = const$.
- Consider the best possible scenario – gaussian model with homoscedasticity, only π and β are unknown.
- Let Ψ be a class of tests for $H_0 : \beta = \beta_0$ that has correct size uniformly over nuisance parameter π. For any fixed $\beta^* \neq \beta_0$

$$\lim\sup_{n\to\infty} \max_{\psi\in\Psi} \left(\min_{\pi: \frac{\mu^2}{\sqrt{K}} = const} E_{\pi,\beta^*}\psi \right) < 1.$$

What is Weak Identification?

- If K is fixed, then $\frac{\mu^2}{\sqrt{K}} = const$ 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.
- Positive statement: we construct robust tests that are consistent when $\frac{\mu^2}{\sqrt{K}} \to \infty$.
- Our answer: if the direction of π is unknown, then $\frac{\mu^2}{\sqrt{K}} \asymp const$ is the knife-edge case for consistency.

Overview

Weak IV-Robust Tests: Refresher, Fixed K

$H_0 : \beta = \beta_0$. Define $e(\beta_0) = Y - \beta_0 X$. We wish to construct *weak identification* robust test for *heteroscedastic* model.

- 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.

- What changes with $K \to \infty$?
  - Estimation of Σ, covariance matrix for $e'Z$, is impossible.
  - Need to re-center statistics because χ^2_K is a diverging distribution.
  - CLT may not work well when K is growing.
  - Even under homoscedasticity (need to estimate a scalar σ^2_e), variance estimation mistakes change the limit distribution for $K = \lambda N$.

Our Proposed 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 $Ee'P_Z e = \sum_{i=1}^N P_{ii} E e_i^2$.
- 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).

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 $(U_1, e_1), \ldots, (U_N, e_N)$ 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

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

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. Under the alternative $\beta = \beta_0 + \Delta$, we have $e_i(\beta_0) = Z_i' \pi \Delta + \eta_i$.
- Estimation error is a large part of the residual when K is large; squaring messes up estimation of variance.

AR: Variance Estimation

- Instead estimate $\sigma_i^2 = E[e_i^2]$ by a "cross-fit" variance estimator (Newey and Robins (2018), Kline et al (2020)):

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

- Challenge is that we need a double sum $\Phi_0 = \frac{2}{K} \sum_{i \neq j} P_{ij}^2 \sigma_i^2 \sigma_j^2$:

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

- Our suggested estimator:

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

Power of AR

- The leave-one-out AR is

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

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

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

Summary

- Positive statement: we constructed a test (leave-one-out AR), that is robust to weak identification (and heteroscedasticity). It becomes consistent as soon as $\frac{\mu^2}{\sqrt{K}} \to \infty$.
- The suggested AR is probably not very powerful if identification is strong.
- How can we distinguish empirically if identification is strong?

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).
- Problems:
  - Valid only under homoscedasticity.
  - May be conservative for large K for properly selected estimator: $EF = \frac{\mu^2}{\sigma_v^2 K} + 1$.

Pre-test for Weak Identification

- Our pre-test is based on the empirical measure for $\frac{\mu^2}{\sqrt{K}}$:

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

  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.
- Our two-step procedure uses the leave-one-out AR test when instruments are weak.

Two-step Procedure

- Our two-step procedure uses what is called JIV2 estimator (Chao et al (2012)) and Wald statistics when instruments are strong:

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

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

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

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

Two-step Procedure

Theorem 2.

If $\frac{\mu^2}{K^{2/3}} \to 0$ *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 small, $Wald(\beta_0)$ has non-standard asymptotic distribution.
- When large then $Wald(\beta_0) \approx \xi^2 = \chi^2_1$.
- This 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]

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.

Conclusions

- We found that the knife-edge case for consistency happens when $\frac{\mu^2}{\sqrt{K}} \asymp const$ (recall $\mu^2 = \pi' Z' Z \pi$).
- We introduced AR test robust to weak id, heteroscedasticity and many instruments.
- We use ideas 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 $K \to \infty$.